FIBER COLLISION CORRECTION METHOD FOR BOSS POWER SPECTRUM

ChangHoon Hahn[1], Michael R. Blanton[1], Roman Scoccimarro[1]
Draft version March 24, 2015

ABSTRACT

With their ability to collect large amounts of redshift data, fiber-fed multi-object spectroscopic surveys dominate cosmological studies. However, the physical size of the fibers limit the ability of these surveys from successfully collecting redshifts from galaxies within the angular fiber-collision scale. These fiber-collisions ultimately leads to systematic effects on galaxy clustering measurements. Using simulated mock catalogues already extensively used in interpreting large-scale structure clustering measurements of the Sloan Digital Sky Survey, we investigate the effects of fiber-collisions on the power spectrum in particular. We compare current fiber-collision correction methods in the literature and assess their success in reconstructing the true power spectrum for the mock catalogues. We then present our fiber-collision correction method. Our method uses the distribution of the line-of-sight displacements between resolved fiber-collided pairs in order to statistically reconstruct clustering on small scales along with a proper accounting of fiber-collisions in the shot-noise correction term of the power spectrum estimator. Using our method, we reproduce the true power spectrum for the mock catalogues with residuals of $\sim 1\%$ at $k \sim 0.3\ h/\mathrm{Mpc}$. Finally, comparison to other fiber-collision correction methods reveal that the method better reproduces the true power spectrum especially at small scales, with the added advantage that the method can be validated and calibrated in actual data.

Subject headings: cosmology: observations – cosmology: large-scale structure of universe – galaxies: distances and redshift – galaxies: halos – galaxies: statistics

1. INTRODUCTION

Through the tens of millions of redshifts obtained through redshift surveys, cosmological measurements such as galaxy clustering statistics are no longer dominated by uncertainties from statistical precision, but from systematic effects of the measurements. These redshift surveys, such as the 2dF Galaxy Redshift Survey (2dFGRS; Colless 1999) and the Sloan Digital Sky Survey-III Baryon Oscillation Spectroscopic Survey (SDSS-III BOSS; Anderson et al. 2012; Dawson et al. 2013) and future redshift surveys such as the Extended Baryon Oscillation Spectroscopic Survey (eBOSS; CITE CITE) mostly use fiber fed spectrographs.

For each galaxy, a fiber is used to obtain a spectroscopic redshift. However, due to the physical size of the fibers even with efficient targeting strategies (e.g. Blanton et al. 2003 for SDSS), if galaxies are located within the fiber-collision angular scale from each other on the sky, two separate fibers cannot be placed adjacently to separately extract the spectroscopic information. In these cases, the redshifts of fiber-collided galaxies are not individually resolved and only a single redshift is measured.

Furthermore, as we probe to higher redshifts in our cosmological studies, the fiber-collision angular scale correspond to a larger comoving scale thereby affecting our measurements on larger scales. SDSS-III BOSS, in particular, has an angular fiber-collision scale of 62", which corresponds to ~ 0.4 Mpc/h at the center of the redshift bin, and fiber-collided galaxies account for $< 5\%$ of the galaxy sample. Although fiber-collided galaxies

[1] Center for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Place, New York, NY 10003; chh327@nyu.edu

account for a relatively small fraction of the redshifts its effect on clustering measurements such as the power spectrum is significant and must be understood as clustering measurements probe to smaller scales and as we explore higher order statistics such as the three-point correlation function or the bispectrum. Perhaps with the use of robotic fiber positioner technology in future surveys such as the Dark Energy Survey Instrument (DESI; Schlegel et al. 2011; Morales et al. 2012; Makarem et al. 2014) will render fiber-collisions obsolete. However, until then accounting for the systematic effects remains an unavoidable challenge for clustering statistics.

To correct for the systematic effects of fiber collisions, a common method used in clustering statistics is the nearest neighbor method (Zehavi et al. 2002; Berlind et al. 2006; Anderson et al. 2012). For fiber-collided galaxies without resolved redshifts, the method assigns the statistical weight of the fiber-collided galaxy to its nearest angular neighbor. This method has been

- A paragraph that list fiber collision corrections listed in Guo et al. (2012) Zehavi et al. (2002). List all clusteirng statistics measurement that uses this.

- talk about shot-noise correction term adjustments

Recently Guo et al. (2012) proposed a fiber-collision correction method for the two-point correlation function (2PCF) of the BOSS galaxies that is able to SOME PRECISION. Guo et al. (2012) estimates the total contribution of fiber-collided galaxies to the two point correlation function by examining the pair statistics in overlapping tiling regions of the survey, where a smaller fraction of galaxies suffer from fiber-collisions. Unfortunately, this method cannot be applied to the Fourier counter part of the 2PCF due to the complex geometry of the

overlapping regions. Consequently, we propose a fiber-collision correction method that uses the distribution of the line-of-sight displacement between resolved fiber-collided galaxies and proper accounting of fiber-collisions in the shot-noise correction term of the power spectrum estimator.

In Section 2, we present a brief description of the fiber-collided simulated mock catalogues that we use to quantify the systematic effect of fiber-collisions and to test and compare correction methods throughout the paper. After discussing the two components of our correction method in Section 3, we present it in Section 4.1. We also test our correction method on the mock catalogues and compare it to other correction methods in Section 4.1. Finally in Section 5 we summarize and discuss our results.

2. FIBER-COLLIDED MOCK CATALOGUES

For various purposes such as calculating the covariance matrix or estimating cosmic variance, simulated mock catalogues play a crucial role in measuring clustering statistics of observed galaxy data (Scoccimarro & Sheth 2002; Anderson et al. 2012; Manera et al. 2013). They also provide a means of understanding systematic effects such as fiber-collisions (Yoon et al. 2008; Guo et al. 2012 LIST MORE). By using mock catalogues, we can test how these effects influence clustering results and devise correction methods that attempt to account for these effects.

A direct way of understanding the effects of fiber-collisions on clustering statistics in observations is to first apply fiber-collisions to mock catalogues then compare the clustering statistics obtained from mock catalogues with and without the fiber collisions. Afterwards, correction methods for fiber-collisions can be applied to the fiber-collided mocks and the merit of the correction method can be quantified by how successfully they can reproduce the clustering statistics of the original mock catalogues without fiber-collisions. The optimal correction method can then be applied to the observed data with some assurance that it accounts for fiber-collisions and improves the clustering measurements.

When applying the fiber-collisions to the mock catalogues, it is essential to apply them in the same manner they affect the observational data. For BOSS, galaxies within 62" are fiber-collided (Anderson et al. 2012). In reality, this simple criteria is further complicated by the tiling scheme of observing plates that create overlapping regions, which have a higher success rate in resolving galaxy spectra within the fiber-collision angular scale (Guo et al. 2012). Furthermore, fiber-collisions are only one of the systematic effects that influence BOSS data. Systematic effects include the unique geometry of the BOSS survey, the variable completeness in different sectors of the survey, and redshift failures (Anderson et al. 2012).

Effects of fiber-collisions must be understood and interpreted in conjunction with these systematic effects and to ultimately test our fiber-collision correction method we use LasDamas (McBride et al. 2009), Quick Particle Mesh (White et al. 2014), and Tiling mock catalogues. These mocks have already been used in interpreting observational results for BOSS and are generated through different prescriptions. Therefore they provide comple-
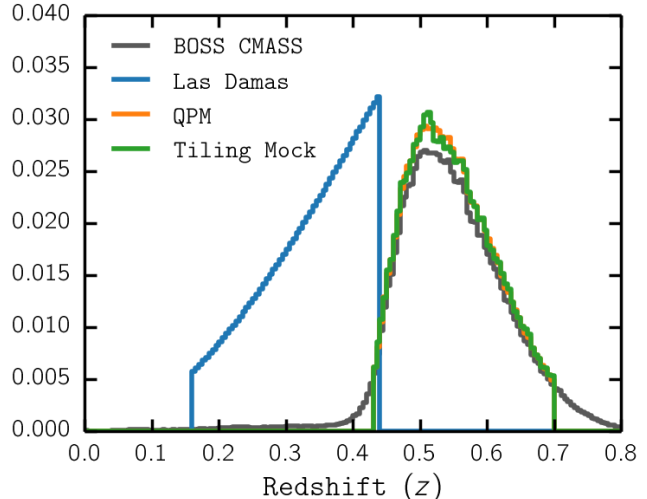

FIG. 1.— Normalized galaxy redshift distribution for Las Damas (blue), QPM (orange), and Tiling Mock (green) mock catalogues. The normalized redshift distribution of BOSS DR12 CMASS is also plotted (black). Bin size of $\Delta z = 0.025$ was used to compute the distributions. Las Damas mock catalogue has a constant number density of galaxies throughout its redshift range ($0.16 < z < 0.44$). QPM and Tiling Mock catalogues, on the other hand, trace the BOSS CMASS redshift distribution ($0.43 < z < 0.7$).

mentary tests fiber-collisions and our correction method tailored for BOSS.

Each of these mock catalogues use some sort of N-body dark matter simulation along with a HOD prescription to populate the dark matter halos with "galaxies". More specifically, the LasDamas Galaxy mock catalogues is constructed from the N-body Large Suite of Dark Matter Simulations (LasDamas). The N-body LasDamas simulations uses a spatially flat ΛCDM cosmology with $\Omega_{\mathrm{m}} = 0.25$, $\Omega_{\Lambda} = 0.75$, $\Omega_{\mathrm{b}} = 0.04$, $\sigma_8 = 0.8$, $n_{\mathrm{s}} = 1$ and $h = 0.7$. The dark matter halos are then identified using a friends-of-friends (FOF) algorithm (Davis et al. 1985) with a linking length $b = 0.2$ times the mean interparticle separation. The over-densities of dark matter halos are then populated with galaxies using the halo occupation distribution (HOD) framework to construct the galaxy mock catalogues (McBride et al. 2009). The HOD prescription is specified so that the galaxy mock catalogues reproduce the correlation function of the SDSS sample.

- above description is not entirely correct because that was LasDamas for Guo+2011

- SDSS-II Geometry

- lower redshift range

Next, the QPM mock galaxy catalogues uses a "quick particle mesh" method, which uses a low resolution particle-mesh N-body solver to evolve particles within a period simulation volume. The particles are assigned halo masses in order to match the halo mass function and large-scale bias of halos of high resolution simulations. Afterward the Tinker et al. (2012) HOD parameterization is used to populate the halos. The mock galaxy sample is then trimmed to the BOSS CMASS survey footprint, downsampled based on angular sky completeness (sector completeness) and radial selection. The QPM mock galaxy catalogue is generated using the following

ΛCDM cosmology: $\Omega_{\rm m} = 0.25$, $\Omega_{\Lambda} = 0.25$, $\Omega_{\rm b} = 0.04$, $\sigma_8 = 0.8$, $n_{\rm s} = 1$ and $h = 0.7$. For a detailed description of the QPM galaxy mock catalogues we refer readers to White et al. (2014).

Finally the Tiling Mock catalogue is generated using

• documentation for tiling mock?

In Figure 2, we plot the redshift distribution for Las-Damas, QPM, and Tiling Mock catalogues along with the redshift distribution of BOSS DR12 galaxies. The redshift distributions illustrate the constant $\bar{n}(z)$ for the LasDamas mock catalogue, while the QPM and Tiling mock $\bar{n}(z)$ distributions trace the observed BOSS redshift distribution.

In order to measure the power spectrum for these mock catalogues, we use the standard Feldman et al. (1994) (FKP) estimator. Throughout the paper we use the following formalisms from Feldman et al. (1994).

$$P(\mathbf{k}) = <|F(\mathbf{k})|^2> - \frac{(1+\alpha) \int d^3 r \, \bar{n}(\mathbf{r}) w_{\rm FKP}^2(\mathbf{r})}{\int d^3 r \, \bar{n}^2(\mathbf{r}) w_{\rm FKP}^2(\mathbf{r})}, \quad (1)$$

where

$$F(\mathbf{r}) = \frac{w_{\rm FKP}(\mathbf{r})[n_g(\mathbf{r}) - \alpha n_r(\mathbf{r})]}{[\int d^3 r \, \bar{n}^2(\mathbf{r}) w_{\rm FKP}^2(\mathbf{r})]^{\frac{1}{2}}} \quad (2)$$

and the second term represents the constant shot-noise contribution to the power due to the discrete density field of our galaxies. More specifically, n_g is the galaxy density, n_r is the density of the synthetic random catalogue, α is the ratio of the number of galaxies over the number of synthetic random galaxies, $\bar{n}$ is the mean density of the galaxies, and $w_{\rm FKP}$ is the minimum variance weight derived in Feldman et al. (1994):

$$w_{\rm FKP}(\mathbf{r}) = \frac{1}{1 + \bar{n}(\mathbf{r}) P_0} \quad (3)$$

where P_0 is the power spectrum amplitude at which the error is minimized. We use $P_0 = 20000 \ {\rm Mpc}^3/h^3$ for our analysis, which corresponds to $k \sim 0.1 \ h/{\rm Mpc}$.

For LasDamas, QPM, and Tiling mock catalogues we use the above FKP estimator to measure the power spectrum, Figure 2. For the $P(k)$ in Figure 2, we do not apply the fiber-collisions to the catalogues. Therefore, there are no complications from using the exact formulation of Equation 1. However, when we calculate $P(k)$ for BOSS DR12 data in Figure 2 (black), the systematic weights assigned to the galaxies that account for sector completeness, redshift failures, and fiber-collisions,

$$w_{\rm tot} = w_{\rm sys}(w_{\rm rf} + w_{\rm fc} - 1), \quad (4)$$

must be taken account (Anderson et al. 2012; Beutler et al. 2014). Hence, instead of the galaxy density $n_g(\mathbf{r})$ we use a weighted galaxy density $n_g'(\mathbf{r}) = n_g(\mathbf{r}) w(\mathbf{r})$ and instead of $\alpha = N_{\rm gal}/N_{\rm ran}$ we use $\alpha' = N_{\rm gal}'/N_{\rm ran}$ where $N_{\rm gal}' = \sum_{\rm gal} w_{\rm tot}$. For LasDamas and QPM, which have multiple mock catalogues, we compute the sample variance of the power spectrum

$$\Delta P(k) = \sqrt{\frac{1}{N_{\rm mock}} \sum_{i}^{N_{\rm mock}} (P_i(k) - \overline{P(k)})^2}. \quad (5)$$
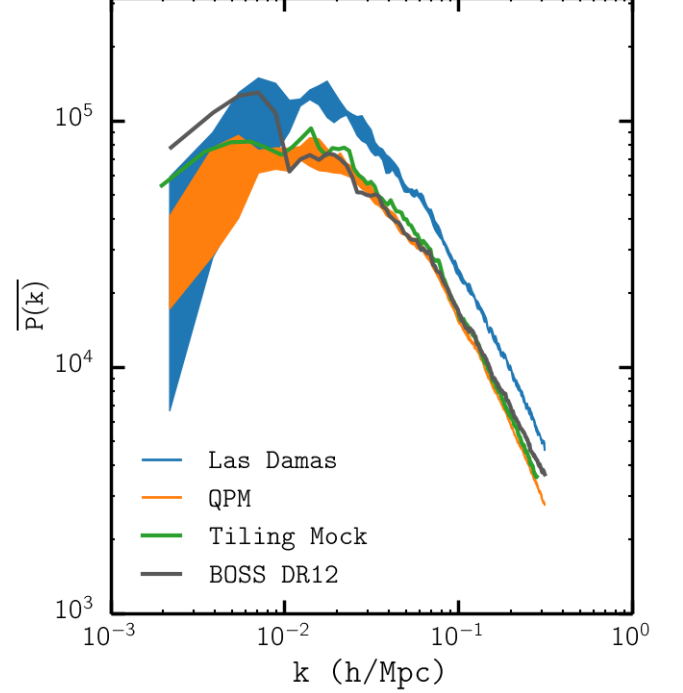

FIG. 2.— Average power spectrum $\overline{P(k)}$ for BOSS DR12 (black) and the mock catalogues Las Damas (blue), QPM (orange), and Tiling Mock (green), listed in Section 2 using the FKP $P(k)$ estimator. The width of the $P(k)$ for Las Damas and QPM represent the sample variance ($\Delta P(k)/\overline{P(k)}$) calculated from the multiple mock catalogues. We emphasize that fiber-collisions were not applied to the mock catalogues for the $P(k)$ measurements. Meanwhile, galaxy weights (Equation 4) are applied to the BOSS DR12 $P(k)$ measurement. We remind readers that the $\overline{P(k)}$ of the LasDamas mock catalogues is an outlier because of the different redshift range compared to the other mock catalogues.

$\Delta P(k)$ is represented by the width of the $P(k)$ in Figure 2.

3. FIBER-COLLISION CORRECTION METHOD

To account for the systematic effects of fiber-collisions Anderson et al. (2012), for the BOSS galaxy data, most recently used a nearest-neighbor galaxy weighting scheme. For galaxies without resolved spectroscopic redshifts due to fiber-collisions, the entire statistical weight of the galaxy was assigned to its nearest-neighbor. This method assumes that all galaxies with angular separations $< 62"$ are correlated. In other words, that galaxies within the fiber-collision angular scale reside in the same group or cluster. Consequently, galaxies that have angular separations $< 62"$ purely by coincidence (hereafter referred to as "chance alignments") are incorrectly assumed to be gravitationally correlated and displaced significantly from its true position. Even for fiber-collided galaxies that reside in the same group or cluster, up-weighting the nearest neighbor ignores halo-scale line-of-sight displacements.

While the nearest-neighbor method provides a reasonable estimate at large scales, in Section 3.1 we demonstrate that at small scales it's effects on the power spectrum are significant and outweigh sample variance. Furthermore, in Section 3.1, we present an improved method to account for correlated fiber-collided galaxies
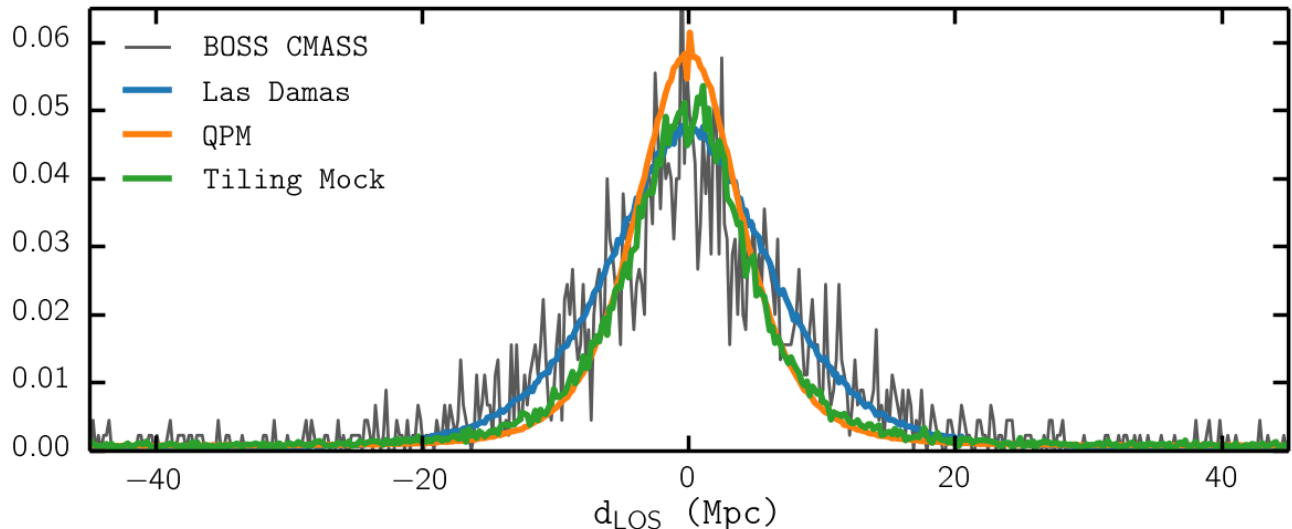
FIG. 3.— Normalized distribution of d_{LOS} for Las Damas (blue), QPM (orange), and Tiling Mock (green) mock catalogues. The normalized d_{LOS} distribution of BOSS DR12 is also plotted (black). Bin size of $\Delta d = 0.2$ was used to compute these d_{LOS} distributions. We focus on the peak of the distribution (discussed in 3.1).

using the line-of-sight displacement between close galaxy pairs. Then in Section 3.2, we present corrections to the power spectrum estimator that further accounts for fiber-collisions through the shot-noise correction term in the estimator.

3.1. *Line-of-Sight Displacement of Fiber-collided Pairs*

It is impossible to determine in observed galaxy data whether fiber-collided galaxies without resolved spectroscopic redshift are either correlated or chance alignments. Fortunately, using just the close galaxy pairs with resolved redshifts (especially in the overlapping regions mentioned in Section 2), it is possible to model the fraction of galaxy pairs that are corrected and uncorrelated.

Meanwhile, for simulated mock catalogues, described in Section 2, modeling the fiber-collided galaxy pairs is a trivial task since fiber-collisions are post-processed after the simulated galaxy positions are generated. Hence all redshift values are resolved regardless of angular proximity. With both redshifts of the fiber-collided pair galaxies, we are able to emulate the observed data by assigning nearest neighbor galaxy weights (as done in Anderson et al. 2012). So for a close galaxy pair, one galaxy ("the nearest neighbor") will have the statistical weight of both galaxies, while the other has a statistical weight of 0. Furthermore, having the redshift values of close galaxy pairs in mock catalogues, allows us to model the correlated and uncorrelated fiber-collided pairs.

In order to determine whether or not fiber-collided pairs are correlated, we examine the comoving line-of-sight displacement (d_{LOS}) of fiber-collided galaxy pairs. We calculate the d_{LOS} by taking the difference of the line-of-sight comoving distance from the resolved redshifts:

$$d_{\text{LOS}} = D_{\text{C}}(z_1) - D_{\text{C}}(z_2). \tag{6}$$

$D_{\text{C}}(z)$ is the line-of-sight comoving distance at redshift z (Hogg 1999). z_1 and z_2 represent the resolved redshifts of the galaxies in the fiber-collided pair.

After calculating the d_{LOS} for each of the fiber-collided pairs, in Figure 3 we present the distribution of d_{LOS} for

LasDamas (blue), QPM (orange), Tiling Mock (green), and BOSS DR12 (black). All of the d_{LOS} distributions contain two distinct components: a peak approximately within the range -10 Mpc $< d_{\text{LOS}} < 10$ Mpc and a flat component (hereafter "tail" component) that extends to d_{LOS} that encompasses the survey depth (entire range of the distribution not displayed in Figure 3). We note that for the BOSS data, since we cannot compute the d_{LOS} for fiber-collided pairs without resolved spectra, the d_{LOS} distribution represents the d_{LOS} from close galaxy pairs with resolved spectroscopic redshifts, often from overlapping regions of the BOSS survey.

Galaxies that are in the same groups or clusters, due to their gravitational interactions at halo-scales, are more likely to be in close angular proximity with each other. These galaxies in over-dense regions cause the peak in the d_{LOS} distribution. While the "tail" component is composed of galaxy pairs that are arbitrarily in close angular proximity even when they are not correlated.

Focusing on the peak of the distribution, we note that it closely traces an exponential function. Therefore, we fit an exponential of the form,

$$p(d_{\text{LOS}}) = A \, e^{-|d_{\text{LOS}}|/\sigma} \tag{7}$$

for a mathematical prescription of the d_{LOS} distribution peak as a function of the displacement for each of the data catalogues in Figure 3. The best-fit σ and A parameters are obtained by fitting Equation 7 to the d_{LOS} distribution peaks using MPFIT (Markwardt 2009). We list the values of σ obtained for the mock catalogues in Table 1. Both Table 1 and Figure 3 illustrate that the d_{LOS} distributions for the mock catalogues closely trace the distribution of the BOSS DR12, which further goes to support our use of mock catalogues to test our fiber-collision correction method.

Using the exponential fits to the peak of the d_{LOS} distribution, we can also estimate the fraction of fiber-

TABLE 1
d_{LOS} DISTRIBUTION BEST-FIT PARAMETERS

Catalogue	σ (h^{-1}Mpc)	f_{peak}
Las Damas	6.85	0.71
QPM	4.98	0.65
Tiling Mock	5.38	0.57
BOSS DR12	7.6	0.65

Notes: Best-fit parameter σ and peak fraction f_{peak} for the d_{LOS} distributions in Figure 3.

collided pairs that are within the peak:

$$f_{\mathrm{peak}} = \frac{\sum\limits_{\mathrm{peak}} p(d_{\mathrm{LOS}})}{N_{\mathrm{pairs}}}, \qquad (8)$$

where N_{pairs} is the total number of fiber-collided pairs. As mentioned above, this portion of fiber-collided pairs are galaxy pairs that are correlated. We list the f_{peak} values for each of the data catalogues in Table 1. The f_{peak} values for the mock catalogue are also in good agreement with the BOSS DR12 f_{peak}.

As stated earlier, the current method of accounting for fiber-collisions uses the nearest-neighbor upweighting scheme (hereafter NN-upweighting). For this method to be completely correct, the d_{LOS} distribution would have to be a delta function, which as Figure 3 demonstrates is not the case. By simply incorporating the peak of the d_{LOS} distribution, we are able to significantly improve clustering statistics on small scales. Rather than placing the fiber-collided galaxy on top of its nearest neighbor, as NN-upweighting does, placing the fiber-collided galaxy at displacement sampled from the d_{LOS} distribution, away from the nearest neighbor better accounts for the small scale clustering measurements. On mock catalogues, we test this by first applying the NN upweighting in the same way as it is applied in the BOSS DR12 catalogue to the mock catalogues. We then sample a displacement d_i from $p(d_{\mathrm{LOS}})$ (Equation 7) for each fiber-collided pair. Afterwards we "place" the fiber-collided galaxy a distance d_i from the upweighted galaxy and reduce the statistical weight of the upweighted galaxy (the nearest-neighbor) by 1. We refer to this method as "d_{LOS}-peak.

The d_{LOS}-peak correction method is only part of our correction scheme we present in this paper. Nonetheless, we test it by applying it to the mock catalogues and comparing the resulting power spectrums. For each of the mock catalogues (Las Damas, QPM, and Tiling Mock) we use the FKP power spectrum estimator from before to compute the power spectrums of the original mock data, the mock data with NN-upweighting, and the mock data with the d_{LOS}-peak correction. We note that both the NN-upweighting and the d_{LOS}-peak correction methods alter the redshift distribution of our samples by a negligible amount $< 1\%$. Therefore, the synthetic random catalogues and the $\bar{n}(z)$ values do not need to be adjusted in the FKP $P(k)$ calculations.

Since we are interested in how well these fiber-collision considerations can reproduce the power spectrum with-out fiber collisions (the "true" power spectrum, unsullied by systematic effects), in Figure 4 we plot the ratio of the average power spectrum $P(k)$ computed from corrected mock catalogues over the average $P(k)$ from the original mock catalogues, $\overline{P_{\mathrm{corr}}(k)}/\overline{P_{\mathrm{true}}(k)}$. The ratio for the NN upweighting scheme is plotted in blue while the ratio for the d_{LOS}-peak correction scheme is plotted in orange. Also plotted is $\Delta P_{\mathrm{true}}(k)/\overline{P_{\mathrm{true}}(k)}$ (black; dashed) in order to compare the ratios to sample variance of the mock catalogues. For all mocks, the $P(k)$ from the NN upweighting scheme is greater than $P_{\mathrm{true}}(k)$ by $> 10\%$ at $k \sim 0.3$ h/Mpc. In comparison, $P(k)$ from the d_{LOS}-peak correction method is $\sim 5\%$ greater than $P_{\mathrm{true}}(k)$ - a significant improvement.

Comparing the NN-upweighting and the d_{LOS}-peak methods to the sample variance ($\Delta P_{\mathrm{true}}(k)/\overline{P_{\mathrm{true}}(k)}$), reveal that neither of these methods provide a sufficient correction for the fiber-collision at $k > 0.1$ h/Mpc. The $P(k)$s obtained from these methods are dominated by systematics on small scales. Furthermore, the d_{LOS}-peak method only statistically accounts for correlated pairs and does not consider uncorrelated pairs. Using the entire d_{LOS} distribution in order to account for uncorrelated pairs results in an overall reduction of $P(k)$ since the method displaces some close pairs that are actually correlated pairs as uncorrelated pairs and the other way around. Therefore, a more robust fiber-collision correction method is required.

3.2. *Shot Noise Correction*

Both the NN-upweighting and the d_{LOS}-peak correction schemes do not sufficiently account for the systematic effects of fiber-collisions in the power spectrum. In this section we introduce the other part of our fiber-collision correction method that involve the shot-noise correction term of the power spectrum estimator.

In Section 2, we described the FKP power spectrum estimator used to compute the power spectrums of Figure 2. In their estimator, to account for the discreteness of the density field, they subtract the shot-noise contribution to the power spectrum (Equation 1):

$$P_{\mathrm{shot}} = \frac{(1+\alpha) \int d^3r\, \bar{n}(\mathbf{r}) w^2(\mathbf{r})}{\int d^3r\, \bar{n}^2(\mathbf{r}) w^2(\mathbf{r})}. \qquad (9)$$

In practice, the integrals in Equation 9 can be written in terms of discrete sums over the random catalogue. The spacial integral $\int d^3r\, \bar{n}(\mathbf{r})...$ can be computed as the discrete sum $\alpha \sum_{\mathrm{ran}} ....$ With this conversion,

$$P_{\mathrm{shot}}^{\mathrm{FKP}} = \frac{(1+\alpha)\alpha \sum\limits_{\mathrm{random}} w_{\mathrm{FKP}}^2(\mathbf{r})}{\alpha \sum\limits_{\mathrm{random}} w_{\mathrm{FKP}}^2(\mathbf{r})}. \qquad (10)$$

In the FKP derivation, they do not take systematic weights of galaxies into account.

The spatial integral $\int d^3r\, \bar{n}(\mathbf{r})...$ can also be computed using the mock catalogues rather than the synthetic random catalogues, as $\sum_{\mathrm{gal}} w_g...$ where w_g represents the statistical weight assigned to the galaxies (Cole et al. 2005; Yamamoto et al. 2006; Beutler et al. 2014; Gil-Marín et al. 2014). Then the shot-noise component can
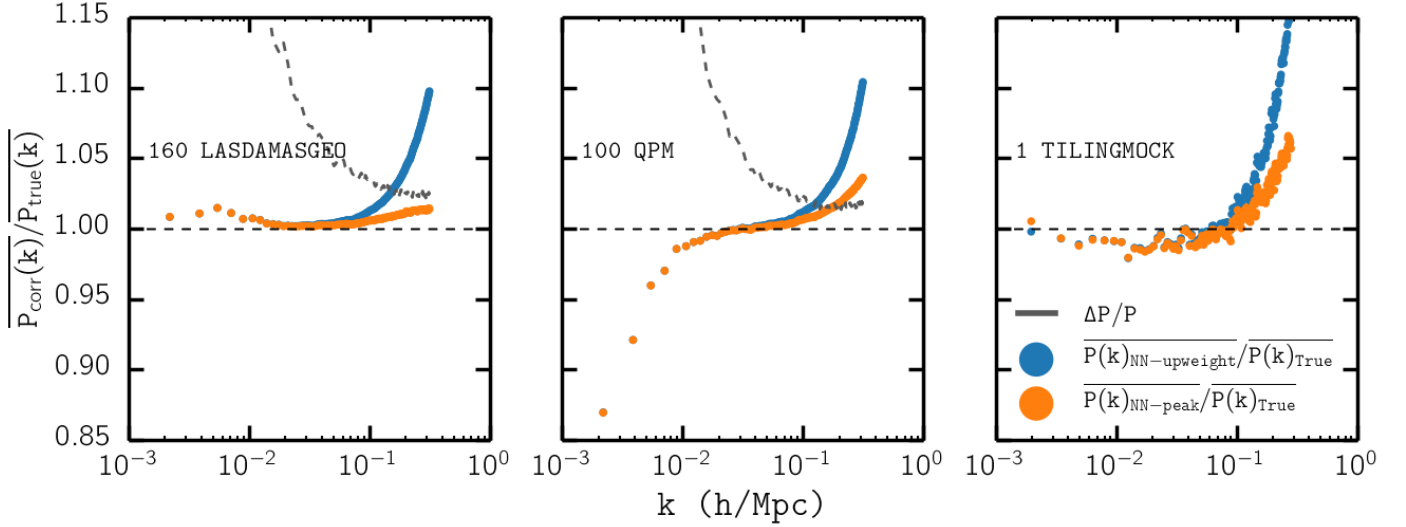
FIG. 4.— The ratio of $P(k)$ using the nearest-neighbor upweighting (blue) and $d_{\rm LOS}$ distribution peak fiber-collision correction (orange) over the true $P(k)$ for LasDamas (left panel), QPM (middle panel), and Tiling Mock (right panel) catalogues.
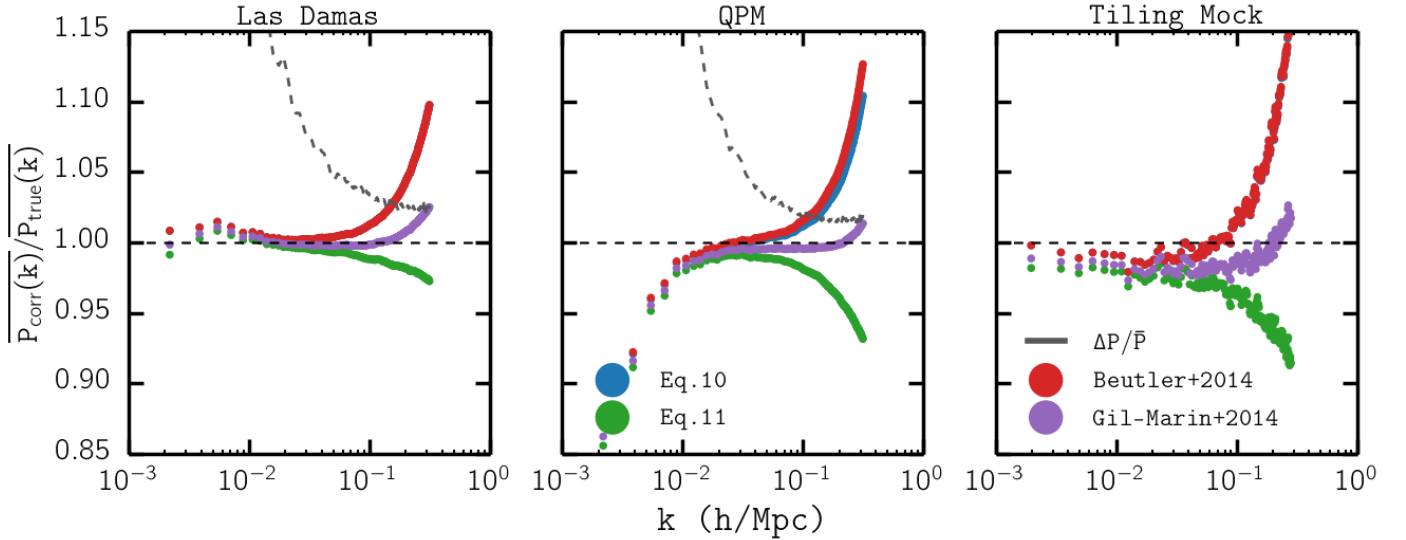

FIG. 5.— The ratio of $\overline{P_{\rm SN}(k)}$ computed using the shot-noise correction term given by Eq. 9 (blue), Eq. ?? (green), $P_{\rm shot}^{\rm Beutler}$ (red), and $P_{\rm shot}^{\rm Gil-Marin}$ (purple) over $\overline{P_{\rm true}(k)}$ for LasDamas (left panel), QPM (middle panel), and Tiling Mock (right panel) catalogues with nearest-neighbor fiber-collision weights. Also plotted for comparison, is $\Delta P_{\rm true}/P_{\rm true}(k)$. For fiber-collisions $\overline{P_{\rm SN}(k)}$ computed using $P_{\rm shot}^{\rm Beutler}$ only marginally improves the $\overline{P_{\rm SN}(k)}$ computed using $P_{\rm shot}^{\rm FKP}$. Both $\overline{P_{\rm SN}(k)}$s significantly overestimates the power spectrum. On the other hand $\overline{P_{\rm SN}(k)}$ computed using Eq. ?? significantly underestimates the power spectrum. At $k \sim 0.3\ h/{\rm Mpc}$, only $\overline{P_{\rm SN}(k)}$ computed using $P_{\rm shot}^{\rm Gil-Marin}$ reduces the effect of fiber-collisions below sample variance.

be expressed as:

$$P_{\rm shot} = \frac{\sum\limits_{\rm galaxy} w_{\rm FKP}^2 w_g^2(\mathbf{r}) - \alpha^2 \sum\limits_{\rm random} w_{\rm FKP}^2(\mathbf{r})}{\alpha \sum\limits_{\rm random} w_{\rm FKP}^2(\mathbf{r})}. \quad (11)$$

In this form, systematic weights for galaxies can be accounted for in w_g and α.

Following the latter formulation, most recently, both Beutler et al. (2014) and Gil-Marín et al. (2014) formulate $P_{\rm shot}$ to account for systematic bias in their analysis of BOSS data. However, they treat systematic weights differently in their derivation. Keeping in mind the weights listed in Eq. 4 for the BOSS data, Beutler et al. (2014) derives

$$P_{\rm shot}^{\rm Beutler} = \frac{\sum\limits_{\rm galaxy} w_{\rm FKP}^2 w_{\rm tot}(\mathbf{r}) w_{\rm sys}(\mathbf{r}) - \alpha'^2 \sum\limits_{\rm random} w_{\rm FKP}^2(\mathbf{r})}{\alpha' \sum\limits_{\rm random} w_{\rm FKP}^2(\mathbf{r})}.$$

$$(12)$$

$\alpha' = \sum_{\rm gal} w_{\rm tot}/N_{\rm ran}$.

On the other hand, Gil-Marín et al. (2014), in a formulation designed to account for fiber-collisions derives

$P_{\rm shot}$ using two separate components: one for correlated pairs and the other chance alignments. For the shot-noise contribution to the power from correlated pairs (referred to as "true pairs" in Gil-Marín et al. 2014) is equivalent to $P_{\rm shot}^{\rm Beu2tler}$ in Beutler et al. (2014). For chance alignments (referred to as "false pairs" in Gil-Marín et al. 2014), the shot-noise contribution is derived as,

$$P_{\rm shot}^{\rm False} = \frac{\sum\limits_{\rm galaxy} w_{\rm FKP}^2 w_{\rm tot}^2(\mathbf{r}) - \alpha'^2 \sum\limits_{\rm random} w_{\rm FKP}^2(\mathbf{r})}{\alpha' \sum\limits_{\rm random} w_{\rm FKP}^2(\mathbf{r})}. \quad (13)$$

Then the total $P_{\rm shot}$ is calculated as a combination of $P_{\rm shot}^{\rm True}$ and $P_{\rm shot}^{\rm False}$:

$$P_{\rm shot}^{\rm Gil-Marin} = (1 - x_{\rm PS})P_{\rm shot}^{True} + x_{\rm PS}P_{\rm shot}^{False} \quad (14)$$

For the BOSS $P(k)$ Gil-Marín et al. (2014) quotes $x_{\rm PS} = 0.58$.

In our $P(k)$ calculations we use the FKP estimator with the shot-noise correction term computed as Eq. 11 with $w_{\rm g} = w_{\rm tot}$, which gives us,

$$P_{\rm shot}^{\rm Hahn\ et\ al.} = \frac{\sum\limits_{\rm galaxy} w_{\rm FKP}^2 w_{\rm tot}^2(\mathbf{r}) - \alpha'^2 \sum\limits_{\rm random} w_{\rm FKP}^2(\mathbf{r})}{\alpha' \sum\limits_{\rm random} w_{\rm FKP}^2(\mathbf{r})}. \quad (15)$$

We note that this equation is equivalent to the shot-noise contribution from "false pairs" in Gil-Marín et al. (2014).

In order to compare the different $P_{\rm shot}$ derivations, we apply each of them ($P_{\rm shot}^{\rm FKP}$, $P_{\rm shot}^{\rm Hahn}$, $P_{\rm shot}^{\rm Beutler}$, and $P_{\rm shot}^{\rm Gil-Marin}$) to the mock catalogues with nearest-neighbor fiber-collision weights applied to them. This comparison ignores a number of systematic effects listed in Eq. 4 ($w_{\rm sys}$ and $w_{\rm rf}$); however, we are particularly interested in fiber-collisions. So in Equations 9-15, we use $w_{\rm sys} = 1$ and $w_{\rm tot} = w_{\rm fc}$. We compute the $P_{\rm SN}(k)$s using the various $P(k)$ estimators with different $P_{\rm shot}$ equations for each of the mock catalogues. Afterwards we compute the $\overline{P_{\rm SN}(k)}$s for the mock catalogues and take the ratio of them over the average $P_{\rm true}(k)$ using the standard FKP $P(k)$ estimator. Without systematic weights, the different $P_{\rm shot}$ equations are equivalent, so our choice of the standard FKP $P(k)$ estimator for $\overline{P_{\rm true}(k)}$ does not affect the comparison.

In Figure 5, we present $\overline{P_{\rm SN}(k)}/\overline{P_{\rm true}(k)}$ computed using different shot noise correction term equations: $P_{\rm shot}^{\rm FKP}$, $P_{\rm shot}^{\rm Hahn}$, $P_{\rm shot}^{\rm Beutler}$, and $P_{\rm shot}^{\rm Gil-Marin}$. The $\overline{P(k)}$ of the NN-upweighted mocks using the standard FKP estimator is the same quantity as presented in Figure 4 and as we noted earlier, overestimates the $\overline{P_{\rm true}(k)}$. $\overline{P_{\rm FKP}(k)}$ is greater than $\overline{P_{\rm true}(k)}$ by $> 15\%$ at $k \sim 0.3$ h/Mpc. The $\overline{P(k)}$ measured using $P_{\rm shot}^{\rm Beutler}$ does not improve this overestimation in any of the mocks. On the other hand, the $\overline{P(k)}$ using just our shot noise equation significantly underestimates the $\overline{P_{\rm true}(k)}$. At $k \sim 0.3$ h/Mpc, $\overline{P(k)/P_{\rm true}(k)} - 1 = \sim -0.03\%, \sim -0.08\%$, and ~ -0.09 for LasDamas, QPM, and Tiling Mock respectively. Finally, the $\overline{P(k)}$ measured using $P_{\rm shot}^{\rm Gil-Marin}$ reproduces $\overline{P_{\rm true}(k)}$ most accurately.

As we plotted in Figure 4, we plot the sample variance

($\Delta P/\bar{P}$) for each of the mock catalogues in Figure 5. Only $\overline{P(k)}$ computed using $P_{\rm shot}^{\rm Gil-Marin}$ is able to reduce the effects of fiber-collisions below the sample variance.

With these alterations to the power spectrum estimator in mind, we present our fiber-collision correction method in the following section .

4. RESULTS

4.1. Fiber-Collision Correction Method

In Section 3.1 we derive a method to account for fiber-collisions using the peak of the $d_{\rm LOS}$ distribution. Then in Section 3.2 we compare the various formulations for the shot-noise correction term and presented our formulation. We now utilize both methods to devise our fiber-collision correction method.

We begin with the nearest-neighbor upweighted mock catalogues. With the nearest-neighbor fiber-collision weights, these mock catalogues accurately imitate the effects of fiber-collision on the real BOSS data. Then as described in Section 3.1, we calculate the $d_{\rm LOS}$ distribution of all fiber-collided pairs in each mock catalogue. Using the $d_{\rm LOS}$ distribution, we compute the fraction of close pairs that are contained within the peak of the distribution ($f_{\rm peak}$; Section 3.1). We then select $f_{\rm peak}$ of the close pairs and treat them as correlated pairs that lie in the peak of the $d_{\rm LOS}$ distribution. We refer to these close pairs as "peak-assigned" close pairs.

Each of these "peak-assigned" close pairs, due to the nearest-neighbor up-weighting, consist of a galaxy with $w_{\rm fc} > 1$ (the "nearest-neighbor" galaxy) and another galaxy with $w_{\rm fc} = 0$ (the "collided" galaxy). We disregard the collided galaxy. For each of the nearest-neighbor galaxies, we sample a value d from the peak of the $d_{\rm LOS}$ distribution and place a galaxy with $w_{\rm fc} = 1$ at a comoving distance d away from the nearest-neighbor galaxy. At the same time, we reduce the fiber-collision weight of the nearest-neighbor galaxy by 1. This process is repeated until the nearest-neighbor galaxy has $w_{\rm fc} = 1$.

For the remaining close pairs that are not "peak-assigned", we keep the nearest-neighbor fiber collision weights. The resulting mock catalogue has $f_{\rm peak}$ fewer galaxies with $w_{\rm cp} > 1$ compared to the NN-upweighted mock catalogues that we started from; however, the total statistical weight of the mock catalogue is conserved.

Next, we use the FKP estimator with the shot-noise correction term described in Section 3.2 Eq. 15 to calculate the $P(k)$ for our newly corrected mock catalogues. In Figure 6 we present the ratio of the $\overline{P(k)}$ computed using our fiber-collision correction method over $P_{\rm true}(k)$. To assess the merit of the fiber-collision correction scheme, we once again over-plotted the sample variance of the mock catalogues. For Las Damas, at $k \sim 0.3$ h/Mpc, $P(k)/P_{\rm true}(k) - 1 \sim 0.5\%$ and throughout the k range, $0.3\% < P(k)/P_{\rm true}(k) - 1 < 1.3\%$. For QPM, at $k \sim 0.3$ h/Mpc, $P(k)/P_{\rm true}(k) - 1 \sim 0.04\%$ and throughout the k range, $0.3\% < P(k)/P_{\rm true}(k) - 1 < 1.3\%$. Furthermore for the Las Damas and QPM catalogues, the $P(k)$ ratio is significantly below the sample variance throughout the entire k range probed.

4.2. Goodness-of-fit

The correction method we present in this paper corrects for the effect of fiber-collisions to the extent where
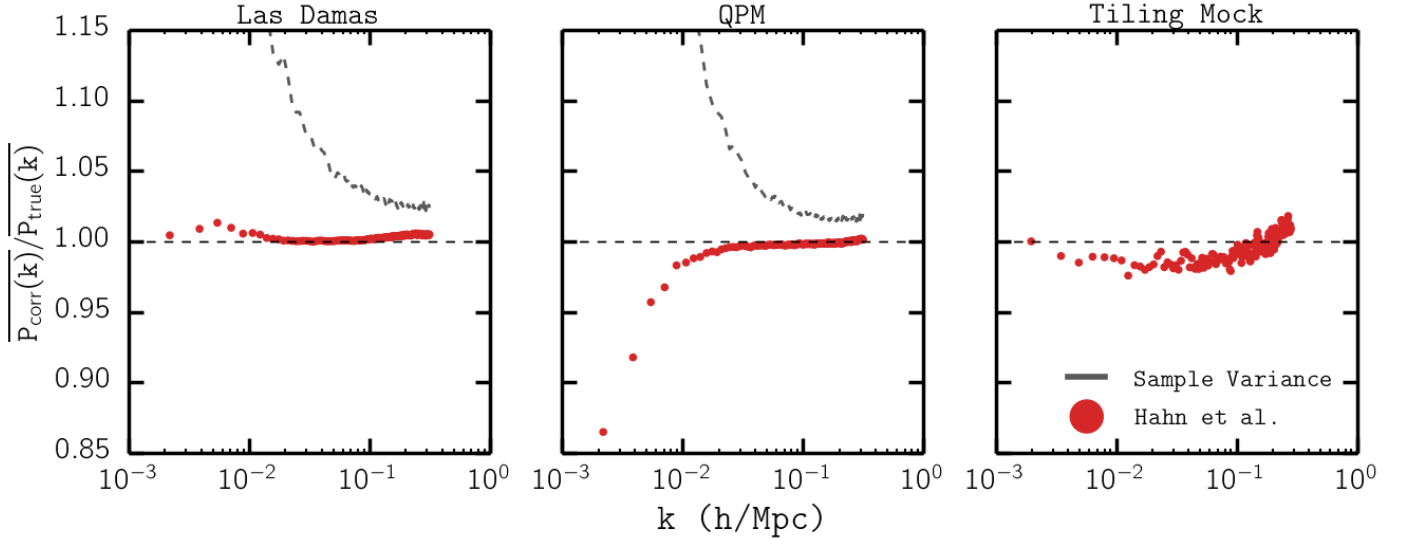
FIG. 6.— The ratio of $\overline{P(k)}$ computed using our fiber-collision correction method described in Section 4.1 over $\overline{P_{\mathrm{true}}(k)}$ for LasDamas (left panel), QPM (middle panel), and Tiling Mock (right panel) catalogues with nearest-neighbor fiber-collision weights (red). Also plotted for comparison, is $\Delta P_{\mathrm{true}}/P_{\mathrm{true}}(k)$. We note that the $\overline{P(k)}$ ratio for our correction method is significantly lower than the $\Delta P_{\mathrm{true}}/P_{\mathrm{true}}(k)$.
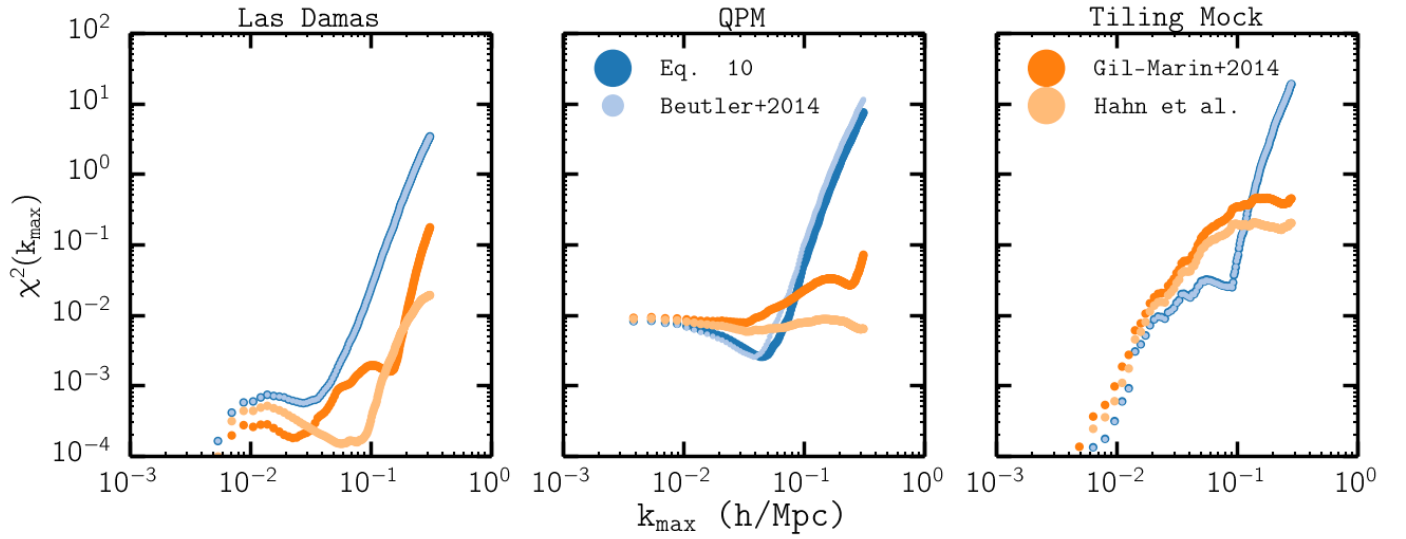

FIG. 7.— $\chi^2(k_{\mathrm{max}})$ for $P_{\mathrm{NN-upweight}}$ (blue), P_{Beutler} (light blue), $P_{\mathrm{Gil-Marin}}$ (orange), and P_{Hahn} (yellow) using the Las Damas (left panel), QPM (centerl panel), and Tiling (right panel) mock catalogues. Equation 16 was used to compute the $\chi^2(k_{\mathrm{max}})$.

$P_{\mathrm{corr}}(k)$ measurements are no longer dominated by systematic effects at $k > 0.2\,h/\mathrm{Mpc}$. With this in mind, we quantify the goodness-of-fit for all the corrected $P(k)$s to $P_{\mathrm{true}}(k)$ by comparing the the following χ^2 to one another:

$$\chi^2(k_{\mathrm{max}}) = \frac{1}{N_{k_{\mathrm{max}}}} \sum_{k < k_{\mathrm{max}}} \frac{(\overline{P_{\mathrm{corr}}(k)} - \overline{P_{\mathrm{true}}(k)})^2}{\Delta P_{\mathrm{true}}(k)^2}. \quad (16)$$

$N_{k_{\mathrm{max}}}$ is the number of k bins in the $P(k)$ calculation with $k < k_{\mathrm{max}}$. $\overline{P_{\mathrm{corr}}(k)}$ is the average fiber-collision corrected powerspectrum of all realizations for each catalogue. $\Delta P_{\mathrm{true}}(k)$ is the standard deviation of $P_{\mathrm{true}}(k)$ computed from all realizations of each mock catalogue. We calculate χ^2 as a function of k_{max} in order to deter-

mine the accumulated χ^2 to a specific scale (k_{max}). In doing so, we determine the scale to which a power spectrum analysis can be extended to without suffering the significant effects of fiber-collisions.

For the Tiling Mock catalogue, since there is only one realization, the variance ($\Delta P_{\mathrm{true}}(k)$) cannot be computed independently. However, for an estimate of the χ^2, we use $\Delta P_{\mathrm{true}}^{\mathrm{Tiling\ Mock}}(k)$ from $\Delta P_{\mathrm{true}}(k)$ of the QPM mock catalogue scaled by the ratio of the power spectrum:

$$\Delta P_{\mathrm{true}}^{\mathrm{Tiling}}(k) = \Delta P_{\mathrm{true}}^{QPM}(k) \times \frac{P_{\mathrm{true}}^{\mathrm{Tiling}}(k)}{P_{\mathrm{true}}^{\mathrm{QPM}}(k)}. \quad (17)$$

In Figure 7, we present the $\chi^2(k_{\mathrm{max}})$ for $P(k)$ using

the following fiber collision correction methods: NN-upweight, Beutler et al. (2014), Gil-Marín et al. (2014), and our method from Section 4.1. The comparison is made using the Las Damas (left panel), QPM (center panel) and Tiling (right panel) mock catalogues. First in the Las Damas panel, at scales $k < 5 \times 10^{-2}$ h/Mpc, $\chi^2 < 10^{-3}$ for all correction methods, which suggests that all correction methods sufficiently correct for fiber-collisions and provide a reasonable fit to $P_{\text{true}}(k)$ at these larger scale. In approaching smaller scales the χ^2 for all correction methods increase. However, throughout the entire k_{max} range, our correction method provides a notably lower χ^2 in comparison to the other correction methods. This is especially clear at $k \sim 0.3$ h/Mpc, where $\chi^2_{\text{NN-upweight}} \approx 0.33$, $\chi^2_{\text{Beutler}} \approx 0.33$, and $\chi^2_{\text{Gil-Marin}} \approx 0.17$ while $\chi^2 \approx 0.019$ for our method.

Next in the central QPM panel, at large scales ($k < 2 \times 10^{-2}$ h/Mpc) all corrected $P(k)$ once again provide a reasonable reconstruction of $P_{\text{true}}(k)$ with $\chi^2 \approx 10^{-2}$. Then for $2 \times 10^{-2} < k_{\text{max}} < 5 \times 10^{-2}$ h/Mpc, NN-upweight and Beutler et al. (2014) correction methods have the lower χ^2s than Gil-Marín et al. (2014) and our correction methods. However, at smaller scales, the χ^2s of NN-upweight and Beutler et al. (2014) quickly increase beyond the χ^2s of Gil-Marín et al. (2014) and our methods. Meanwhile, throughout the entire k_{max} range our correction method has a lower χ^2 than the correction method of Gil-Marín et al. (2014). The difference in χ^2 is again most remarkable at the smallest scales: $\chi^2(k_{\text{max}} \sim 0.3) \approx 7.3$, 11, 0.068, and 0.0063 for NN-upweight, Beutler et al. (2014), Gil-Marín et al. (2014) and our correction method respectively.

Finally in the right panel (Tiling Mock), we find that $\chi^2(k_{\text{max}})$ increases in a near monotonic fashion as a function of k_{max}. For $k_{\text{max}} < 4 \times 10^{-2}$ h/Mpc, there are no significant differences between the χ^2s of all the correction methods. Then from $4 \times 10^{-2} < k_{\text{max}} < 1.5 \times 10^{-2}$ h/Mpc, χ^2s of NN-upweight and Beutler et al. (2014) are notably lower than the χ^2s of Gil-Marín et al. (2014) and our method. However at smaller scales, the χ^2s of NN-upweight and Beutler et al. (2014) quickly rise beyond the other χ^2s, ultimately both reach $\chi^2(k_{\text{max}} \sim 0.3) \approx 13.5$. In the meantime, $\chi^2(k_{\text{max}} \sim 0.3) \approx 0.32$ for Gil-Marín et al. (2014) and $\chi^2(k_{\text{max}} \sim 0.3) \approx 0.15$ for our correction method. Moreover, through the entire k_{max} our method has a lower χ^2 than Gil-Marín et al. (2014). With only one realization of the Tiling Mocks and scaled ΔP, a detailed comparison of the χ^2 values is impossible. Yet the $\chi^2(k_{\text{max}})$ results for our correction provide realistic assurance that the correction method can be applied to the actually BOSS data.

Overall, regardless of catalogue, for scales $k_{\text{max}} < 3 \times 10^{-2}$ h/Mpc, the χ^2 for all fiber-collision corrected $P(k)$ are negligible. Then, with the exception of the Las Damas catalogue, at intermediate scales of $3 \times 10^{-2} < k_{\text{max}} < 0.15$ h/Mpc the NN-upweight and Beutler et al. (2014) methods show the lowest χ^2s. At the smallest scales, however, these corrections methods no longer provide a good fit to $\overline{P_{\text{true}}(k)}$. A detailed comparison at the small scales ($k > 1.5 \times 10^{-1}$ h/Mpc) demonstrates that the fiber-collision correction we present in this paper significantly better reproduces the true power spectrum than any of the other fiber-collision correction methods. Furthermore, Figure 7 illustrates that throughout the entire k_{max} is able to reasonably reproduces $\overline{P_{\text{true}}(k)}$ at all scales.

5. SUMMARY AND DISCUSSION

Using simulated mock catalogues for SDSS data with realistically imposed fiber-collisions, we quantify the systematic effects of fiber-collisions on galaxy clustering statistics, in particular the power spectrum. Although fiber-collisions have little significant effect on the power spectrum at large scales, its effect quickly overtakes the sample variance at scales smaller than $k \sim 0.1$ h/Mpc. At the smallest scales that we explore in this paper, $k \sim 0.3$, fiber-collisions have over a 10% effect on the power spectrum. Consequently, at these small scales the uncertainties for power spectrum measurements are dominated by systematic effects from fiber-collisions.

Fortunately, through the fiber-collided mock catalogues we are able to model the distribution of the line-of-sight displacement between fiber-collided pairs. Combining this model with a modified shot-noise term in the power spectrum estimator, we devise a new fiber-collision correction method that is able to reconstruct the true power spectrum from fiber-collided data even to $k \sim 0.3$ h/Mpc. Throughout the entire k range explored in our analysis, our correction method reconstructs $P_{\text{True}}(k)$ safely within the sample variance.

Furthermore, we compare our method to the most common nearest-neighbor correction method and recent correction methods presented in Beutler et al. (2014) and Gil-Marín et al. (2014) to demonstrate that our method most successfully reproduces the true power spectrum at small scales for all the Las Damas, QPM, and Tiling mock catalogues. As an added advantage, our correction method can also be validated by actual data rather than only through the mock catalogues thereby making it's application to actual data further reliable.

The fiber-collision correction method we present will enable us to extend our galaxy clustering analysis to smaller scales for SDSS-III BOSS and future surveys such as eBOSS or any other large fiber-fed surveys. Our fiber-collision correction method can also be extend to higher order clustering statistics such as the bispectrum, which we will explore in future work.

REFERENCES

Anderson, L., Aubourg, E., Bailey, S., et al. 2012, MNRAS, 427, 3435

Berlind, A. A., Frieman, J., Weinberg, D. H., et al. 2006, ApJS, 167, 1

Beutler, F., Saito, S., Seo, H.-J., et al. 2014, MNRAS, 443, 1065

Blanton, M. R., Lin, H., Lupton, R. H., et al. 2003, AJ, 125, 2276

Cole, S., Percival, W. J., Peacock, J. A., et al. 2005, MNRAS, 362, 505

Colless, M. 1999, Royal Society of London Philosophical Transactions Series A, 357, 105

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10

Feldman, H. A., Kaiser, N., & Peacock, J. A. 1994, ApJ, 426, 23

Gil-Marín, H., Noreña, J., Verde, L., et al. 2014, ArXiv e-prints, arXiv:1407.5668

Guo, H., Zehavi, I., & Zheng, Z. 2012, ApJ, 756, 127

Hogg, D. W. 1999, ArXiv Astrophysics e-prints, astro-ph/9905116

Makarem, L., Kneib, J.-P., Gillet, D., et al. 2014, A&A, 566, A84

Manera, M., Scoccimarro, R., Percival, W. J., et al. 2013, MNRAS, 428, 1036

Markwardt, C. B. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 411, Astronomical Data Analysis Software and Systems XVIII, ed. D. A. Bohlender, D. Durand, & P. Dowler, 251

McBride, C., Berlind, A., Scoccimarro, R., et al. 2009, in Bulletin of the American Astronomical Society, Vol. 41, American Astronomical Society Meeting Abstracts #213, #425.06

Morales, I., Montero-Dorta, A. D., Azzaro, M., et al. 2012, MNRAS, 419, 1187

Schlegel, D., Abdalla, F., Abraham, T., et al. 2011, ArXiv e-prints, arXiv:1106.1706

Scoccimarro, R., & Sheth, R. K. 2002, MNRAS, 329, 629

Tinker, J. L., Sheldon, E. S., Wechsler, R. H., et al. 2012, ApJ, 745, 16

White, M., Tinker, J. L., & McBride, C. K. 2014, MNRAS, 437, 2594

Yamamoto, K., Nakamichi, M., Kamino, A., Bassett, B. A., & Nishioka, H. 2006, PASJ, 58, 93

Yoon, J. H., Schawinski, K., Sheen, Y.-K., Ree, C. H., & Yi, S. K. 2008, ApJS, 176, 414

Zehavi, I., Blanton, M. R., Frieman, J. A., et al. 2002, ApJ, 571, 172