

# Measuring galaxy segregation with the mark connection function

V. J. Martínez<sup>1,2</sup>, P. Arnalte-Mur<sup>1,2</sup>, and D. Stoyan<sup>3</sup>

<sup>1</sup> Observatori Astronòmic, Universitat de València, Apartat de Correus 22085, 46071 València, Spain  
 e-mail: [vicent.martinez@uv.es](mailto:vicent.martinez@uv.es)

<sup>2</sup> Departament d'Astronomia i Astrofísica, Universitat de València, 46100-Burjassot, València, Spain

<sup>3</sup> Institut für Stochastik, TU Bergakademie Freiberg, 09596 Freiberg, Germany

Received 17 July 2009 / Accepted 5 January 2010

## ABSTRACT

**Context.** The clustering properties of galaxies belonging to different luminosity ranges or having different morphological types are different. These characteristics or “marks” permit us to understand the galaxy catalogs that carry all this information as realizations of marked point processes. Many attempts have been presented to quantify the dependence of the clustering of galaxies on their inner properties.

**Aims.** The present paper summarizes methods on spatial marked statistics used in cosmology to disentangle luminosity, color or morphological segregation and introduces a new one in this context, the mark connection function.

**Methods.** The methods used here are the partial correlation functions, including the cross-correlation function, the normalized mark correlation function, the mark variogram and the mark connection function. All these methods are applied to a volume-limited sample drawn from the 2dFGRS, using the spectral type  $\eta$  as the mark.

**Results.** We show the virtues of each method to provide information about the clustering properties of each population, the dependence of the clustering on the marks, the similarity of the marks as a function of the pair distances, and the way to characterize the spatial correlation between the marks. We demonstrate by means of these statistics that passive galaxies exhibit a stronger spatial correlation than active galaxies at small scales ( $r \lesssim 20 h^{-1}$  Mpc), and that the price for galaxies to be close together is in the smaller values of the assigned marks, which means in our case that they are more passive. Through the mark connection function we quantify the relative positioning of different types of galaxies within the overall clustering pattern.

**Conclusions.** The different marked statistics provide different information about the clustering properties of each population. Different aspects of the segregation are encapsulated by each measure, which makes the new one introduced here – the mark connection function – particularly useful for understanding the spatial correlation between the marks.

**Key words.** large-scale structure of Universe – methods: data analysis – methods: statistical

## 1. Introduction

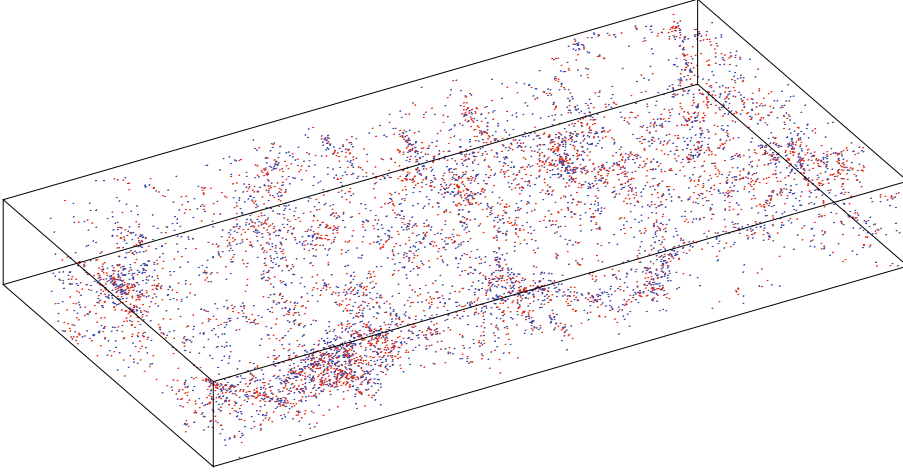
Galaxies of different morphological types show different clustering properties. It is well known, for example, that elliptical galaxies are preferentially found in high density environments, like the centers of rich galaxy clusters (Dressler 1980), while the dominant population of the field are mainly spiral galaxies (Davis & Geller 1976; Dressler 1980). Second order characteristics as the two point correlation function have been used to quantify the clustering of galaxies with different morphologies, different spectral characteristics, different colors or belonging to different luminosity ranges (Phillipps & Shanks 1987; Hamilton 1988; Davis et al. 1988; Loveday et al. 1995; Hermit et al. 1996; Guzzo et al. 1997). Bright galaxies show a stronger spatial correlation than faint ones. Other clustering measures have also been used to quantify the luminosity or morphological segregation: multifractals (Domínguez-Tenreiro & Martínez 1989; Domínguez-Tenreiro et al. 1994), void probability functions (Vogeley et al. 1991; Croton et al. 2004), distributions of the distances to the nearest neighbors (Salzer et al. 1990), etc.

The two-point correlation function  $\xi(r)$  measures the excess probability of finding a neighbor at a distance  $r$  from a given galaxy when compared with that probability for a homogeneous Poisson process. Morphological segregation is encapsulated by the behavior of  $\xi(r)$  when it is calculated separately for different populations of galaxies. Elliptical galaxies show a correlation

function at small scales with steeper slopes and larger amplitudes than spirals (Loveday et al. 1995). A recent analysis of the two degree field galaxy redshift survey (2dFGRS) has shown the same trend when comparing populations for different spectral types, where the two-point correlation function was steeper for passive galaxies than for active galaxies (Madgwick et al. 2003). Also, Zehavi et al. (2002) analyzed the distribution of red and blue galaxies in the Sloan digital sky survey (SDSS) by means of the projected correlation function  $w_p(r_p)$ , showing that red galaxies display a more prominent and steeper real-space correlation function than blue galaxies do.

The galaxy distribution can be considered a realization of a point process. However, in many situations, each galaxy (point in the process) carries additional information regarding a given characteristic (e.g. morphological type) or a given numerical value that measures a given galaxy property: luminosity, color, spectral type. If we attach this characteristic (mark) to the point in the process, we end up at a marked point process, as it is called in mainstream spatial statistics (Stoyan & Stoyan 1994; Martínez & Saar 2002; Illian et al. 2008).

We compare different statistical methods for the study of the marked galaxy distribution. We also introduce – for the first time in this context – the mark connection function. We illustrate the usefulness of these methods by applying them to a volume-limited sample drawn from the 2dFGRS with marks given by the galaxy spectral type. In Sect. 2, we describe the sample and



**Fig. 1.** Tridimensional plot of the galaxy sample used. Red dots correspond to early-type galaxies (population “1”), and blue dots to late-type galaxies (population “2”). The parallelepiped dimensions are  $254 \times 133 \times 31 h^{-1}$  Mpc.

the marks assigned to the galaxies. In Sect. 3 we describe the different statistical methods considered, and in Sect. 4 we show the results of applying them to our galaxy sample. In the conclusions, we stress the capabilities of the mark connection function to characterize the spatial correlation between the marks.

## 2. The samples

To illustrate the different mark clustering measures, we used a nearly volume-limited sample drawn from the 2dFGRS and prepared by the 2dF team (Croton et al. 2004). It contains galaxies with absolute magnitudes in the range  $-20 < M_b < -19$  at redshifts  $z < 0.13$ . In order to avoid the effects of complicated boundaries while using a simple estimator, we selected galaxies inside a rectangular parallelepiped inscribed in the North slice of 2dFGRS. The final sample used contains  $N = 7741$  galaxies and covers a volume of  $V \simeq 10^6 (h^{-1} \text{Mpc})^3$  where  $h$  is the Hubble constant in units of  $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

We characterized the galaxies in the sample using the spectral classification parameter  $\eta$  (Madgwick et al. 2002). Lower values of  $\eta$  correspond to more passive or “early-type” galaxies, while larger values correspond to active or “late-type” ones. In order to avoid negative values of the marks, we defined the used mark as  $m = \eta + 10$ . This shift does not affect our conclusions. Based on this  $\eta$  parameter, we divided our sample in two populations, following Madgwick et al. (2003): population “1” (passive galaxies) with  $\eta \leq -1.4$ , and population “2” (active galaxies) with  $\eta > -1.4$ . These subsamples contain  $N_1 = 3828$  and  $N_2 = 3913$  galaxies, respectively. We show the samples in Fig. 1.

In order to test the existence of mark segregation, we compared the results obtained for the different statistics with random relabeling simulations. In these, we kept the original positions of galaxies, but redistributed the marks randomly among them. This corresponds to a model in which clustering is independent of the mark, or spectral type, of the galaxies. We simulated  $n = 200$  realizations with the random relabeling method, and obtained their maximum and minimum values as a function of the distance  $r$  for each statistic. Deviations of the observed statistics from this range of values correspond to a rejection of the mark-independent clustering model at a pointwise significance of  $1 - \frac{2}{n+1} \simeq 99\%$  (Illian et al. 2008).

## 3. Clustering analysis methods

Recently, the clustering dependence on luminosity, color or morphology has been analyzed by means of the marked clustering statistics, which allow us to study the galaxy clustering as a function of their properties, and moreover provides us with different measures of the correlation between the galaxy properties and the environment (Skibba et al. 2009). The galaxy distribution is interpreted as a realization of a marked point process  $X^M = \{(x_i, m_i)\}$ , where the mark  $m_i$  denotes an intrinsic property of the galaxy located at position  $x_i$ . The mark can be the luminosity, the spectral type, the color, etc. In general, present day galaxy catalogs provide quantitative marks ranging in a continuous interval rather than just a discrete characteristic like if a galaxy is spiral or elliptical. In any case, we shall also show how to use interesting second-order measures to disentangle clustering dependent characteristics of two populations by dividing the sample into two parts using a significant value of the mark as threshold  $m_{\text{thres}}$  and separating the two populations according to the value of the mark: population “1” with  $m_i \leq m_{\text{thres}}$  and population “2” with  $m_i \geq m_{\text{thres}}$ .

We describe below the different methods we used to obtain information about galaxy clustering segregation. They are the classical partial correlation functions (for two discrete populations), the normalized mark correlation function and the mark variogram (based on the use of continuous marks), and finally the mark connection function (based on the use of discrete marks).

We computed the different statistics based on the estimation of the second-order intensity function for the unmarked point process<sup>1</sup> ( $\lambda_2(r)$ ) presented in Stoyan & Stoyan (1994), Pons-Bordería et al. (1999), and Illian et al. (2008),

$$\hat{\lambda}_2(r) = \frac{1}{4\pi r^2} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \frac{k(r - |\mathbf{x}_i - \mathbf{x}_j|)}{V(W \cap W_{\mathbf{x}_i - \mathbf{x}_j})}, \quad (1)$$

<sup>1</sup> Note that the relation between  $\lambda_2(r)$  and the standard correlation function used in cosmology  $\xi(r)$  is  $\lambda_2(r) = n^2[1 + \xi(r)]$ , where  $n$  is the number density. The function  $g(r) = 1 + \xi(r)$  is known as the pair correlation function in spatial statistics. We use the convention of denoting the estimators by putting a hat  $\hat{\cdot}$  on top of the symbol of a given function to distinguish the estimator  $\hat{\lambda}_2(r)$  from the theoretically defined function  $\lambda_2(r)$ . Although this is not standard in cosmology, it is an extended convention in spatial statistics, and it is quite useful when different estimators of a single function are discussed (see, e.g., Pons-Bordería et al. 1999).

where  $\mathbf{x}_i$  are the positions of the points,  $k(\cdot)$  is a kernel function, and  $V(W \cap W_r)$  is the volume of the window (the parallelepiped in our case) intersected with a version of itself shifted by the vector  $\mathbf{r}$  (see Fig. 1 in Pons-Bordería et al. 1999).

In all our calculations we used the Epanechnikov kernel,

$$k(x) = \begin{cases} \frac{3}{4w} \left(1 - \frac{x^2}{w^2}\right) & \text{for } -w \leq x \leq w \\ 0 & \text{otherwise} \end{cases}$$

with a width of  $w = 1 h^{-1}$  Mpc, and sampled the different functions with a step in  $r$  of  $0.5 h^{-1}$  Mpc. This compact kernel is very well suited for correlation analysis (Pons-Bordería et al. 1999). We note however that the choice of a given kernel is not crucial, while the choice of the bandwidth,  $w$ , is more important and plays the role of the binning in the standard calculation of correlation functions, where a top-hat kernel is typically used as default.

### 3.1. Partial two-point correlation functions

In the standard clustering analysis of the galaxy distribution, the two-point correlation function  $\xi(r)$  measures the clustering in excess ( $\xi(r) > 0$ ) or in defect ( $\xi(r) < 0$ ) relative to a Poisson distribution, for which  $\xi(r) = 0$ . Whenever we want to compare the clustering properties of different populations of galaxies encapsulated by their spatial correlations, we can consider the correlation function restricted to a given population, which is called a *partial* correlation function. In fact, for two populations of interest, one can consider three partial two-point correlation functions, namely  $\xi_{11}(r)$ ,  $\xi_{22}(r)$ , and  $\xi_{12}(r)$ . The first two are those mentioned above for types 1 and 2, while the cross-correlation function (Peebles 1980)  $\xi_{12}(r)$  measures the excess probability of finding a neighbor of the type “1” at a distance  $r$  from a given galaxy of type “2”, or vice versa.

Based on Eq. (1), the partial two-point correlation functions were estimated as

$$\hat{\xi}_{ij}(r) = \frac{1}{4\pi r^2 \hat{n}_i \hat{n}_j} \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} \frac{k(r - |\mathbf{x}_k^{(i)} - \mathbf{x}_l^{(j)}|)}{V(W \cap W_{\mathbf{x}_k^{(i)} - \mathbf{x}_l^{(j)}})} - 1, \quad (2)$$

where  $\mathbf{x}_k^{(i)}$  are the positions of galaxies of a population  $i$ , and  $\hat{n}_i = N_i/V$ .

We estimated the error of the measured correlation functions with the jackknife method (Norberg et al. 2009). We divided the data volume in 32 equal, nearly cubic, sub-volumes. We generated the corresponding “mock” datasets omitting one of these sub-volumes at a time, and calculated the correlation functions for these. The jackknife errors for each scale,  $\sigma_{ij}(r)$ , are then obtained as

$$\sigma_{ij}^2(r) = \frac{N_k - 1}{N_k} \sum_{k=1}^{N_k} \left( \xi_{ij}^k(r) - \bar{\xi}_{ij}(r) \right)^2,$$

where  $\xi_{ij}^k(r)$  is the partial correlation function  $\xi_{ij}(r)$  of the “mock” dataset  $k$ ,  $\bar{\xi}_{ij}(r)$  is the value averaged over these datasets, and  $N_k = 32$ .

### 3.2. Normalized mark correlation function

Stoyan & Stoyan (1994) introduced the normalized mark correlation function. To define this function, let us first define the quantity

$$\lambda_2^M[(\mathbf{x}_1, m_1), (\mathbf{x}_2, m_2)] dV_1 dm_1 dV_2 dm_2 \quad (3)$$

as the joint probability that in the volume element  $dV_1$  lies a galaxy with the mark in the range of  $[m_1, m_1 + dm_1]$  and that another galaxy lies in  $dV_2$  with the mark in  $[m_2, m_2 + dm_2]$  (Martínez & Saar 2002). The normalized mark correlation function is

$$k_{mm}(r) = \frac{1}{\bar{m}^2 \lambda_2(r)} \int \int m_1 m_2 \lambda_2^M((\mathbf{x}_1, m_1), (\mathbf{x}_2, m_2)) dm_1 dm_2, \quad (4)$$

for  $\lambda_2(r) \neq 0$ , where  $\bar{m}$  is the mean of the marks.

Despite its name the mark correlation function is not a strict correlation function (Schlather 2001), but it describes important aspects of the spatial correlations of marks. A true mark correlation is a function given by Eq. (4), but replacing the product  $m_1 m_2$  by the product of the differences  $(m_1 - \bar{m})(m_2 - \bar{m})$ . The normalizing denominator  $\bar{m}^2$  must then be replaced by  $\sigma_m^2$ , the variance of the marks. In any case,  $k_{mm}(r) < 1$  represents the inhibition of the marks at the scale  $r$ . For example, in forests it is typically found that trees with a larger stem diameter (mark) tend to be isolated, since they make use of much more ground and sun-light resources than smaller trees. Using luminosity as the mark, the opposite effect has been found for the galaxy distribution, i.e.,  $k_{mm}(r) > 1$  at small scales (Beisbart & Kerscher 2000), implying stronger clustering of brighter galaxies at small separations, in agreement with previous results showing this kind of segregation (Hamilton 1988).

We estimated the normalized mark correlation function as

$$\hat{k}_{mm}(r) = \frac{1}{4\pi r^2 \bar{m}^2 \hat{\lambda}_2(r)} \sum_{i=1}^N \sum_{j=1, j \neq i}^N \frac{m_i m_j k(r - |\mathbf{x}_i - \mathbf{x}_j|)}{V(W \cap W_{\mathbf{x}_i - \mathbf{x}_j})}. \quad (5)$$

### 3.3. Mark variogram

The mark variogram,  $\gamma(r)$  (Walder & Stoyan 1996; Beisbart & Kerscher 2000), is a measure of the similarity of the marks depending on the distance between galaxies. It is defined as

$$\gamma(r) = \frac{1}{2\lambda_2(r)} \int \int (m_1 - m_2)^2 \lambda_2^M((\mathbf{x}_1, m_1), (\mathbf{x}_2, m_2)) dm_1 dm_2. \quad (6)$$

When the clustering properties of a marked point pattern are independent of the marks, the mark variogram  $\gamma(r)$  is constant and takes, naturally, the value of the variance,  $\sigma_m^2$ , of the mark distribution. In the presence of segregation,  $\gamma(r) > \sigma_m^2$  indicates that galaxy pairs at a distance  $r$  tend to have different marks, while the contrary,  $\gamma(r) < \sigma_m^2$ , is an indication that these galaxy pairs tend to have similar marks.

We estimated the mark variogram as

$$\hat{\gamma}(r) = \frac{1}{8\pi r^2 \hat{\lambda}_2(r)} \sum_{i=1}^N \sum_{j=1, j \neq i}^N \frac{(m_i - m_j)^2 k(r - |\mathbf{x}_i - \mathbf{x}_j|)}{V(W \cap W_{\mathbf{x}_i - \mathbf{x}_j})}. \quad (7)$$

### 3.4. Mark connection function

A statistical tool to characterize the spatial correlation between the marks of a point pattern with discrete marks is the mark connection function  $p_{ij}(r)$ , which represents the conditional probability to find two galaxies of type  $i$  and  $j$  at positions separated by a distance  $r$ , under the condition that at these positions there are indeed galaxies. This function yields information different to that from the partial correlation functions,  $\xi_{ij}(r)$ , as shown, for example, in Illian et al. (2008). By its definition it gives the relative frequencies of mark pairs  $(i, j)$  of a distance  $r$ . While  $\xi_{ij}(r)$

takes high values if there are many  $(i, j)$ -pairs at distance  $r$ ,  $p_{ij}(r)$  is high if the proportion of  $(i, j)$ -pairs in all pairs at a distance  $r$  is high. So it may happen that for some  $r$ ,  $\xi_{ij}(r)$  has a minimum, but  $p_{ij}(r)$  has a maximum, if there is only a small number of point pairs at a distance  $r$  in the whole pattern, but many of them are exactly  $(i, j)$ -pairs. Experience shows that often  $p_{ij}(r)$  is able to find finer structures in point patterns than  $\xi_{ij}(r)$ , because of the nature of  $p_{ij}(r)$  as a conditional probability.

If the marking is independent of clustering, then  $p_{ij}(r)$  are constant,

$$p_{ij}(r) = \begin{cases} 2p_i p_j & \text{if } i \neq j \\ p_i^2 & \text{if } i = j. \end{cases} \quad (8)$$

Here  $p_i$  is the probability that a randomly chosen galaxy is of the type  $i$ . The  $p_i$  are estimated as

$$\hat{p}_i = \frac{N_i}{N}.$$

We calculated  $p_{ij}(r)$  based on the estimation of the partial correlation functions as

$$\hat{p}_{ij}(r) = \hat{p}_i \hat{p}_j \frac{\hat{\xi}_{ij}(r) + 1}{\hat{\xi}(r) + 1},$$

where  $\hat{\xi}(r)$  is the two-point correlation function of the full sample.

## 4. Results

### 4.1. Partial two-point correlation functions

Figure 2 shows the three corresponding partial two-point correlation functions, estimated according to Eq. (2). All three clearly show the high degree of clustering within the pattern of galaxies. It is obvious that the correlation function for the type “1” passive galaxies is steeper than for the type “2” active galaxies as well as for the (1,2) pairs. This result corroborates the spectral segregation detected by Madgwick et al. (2003) for the 2dFGRS.

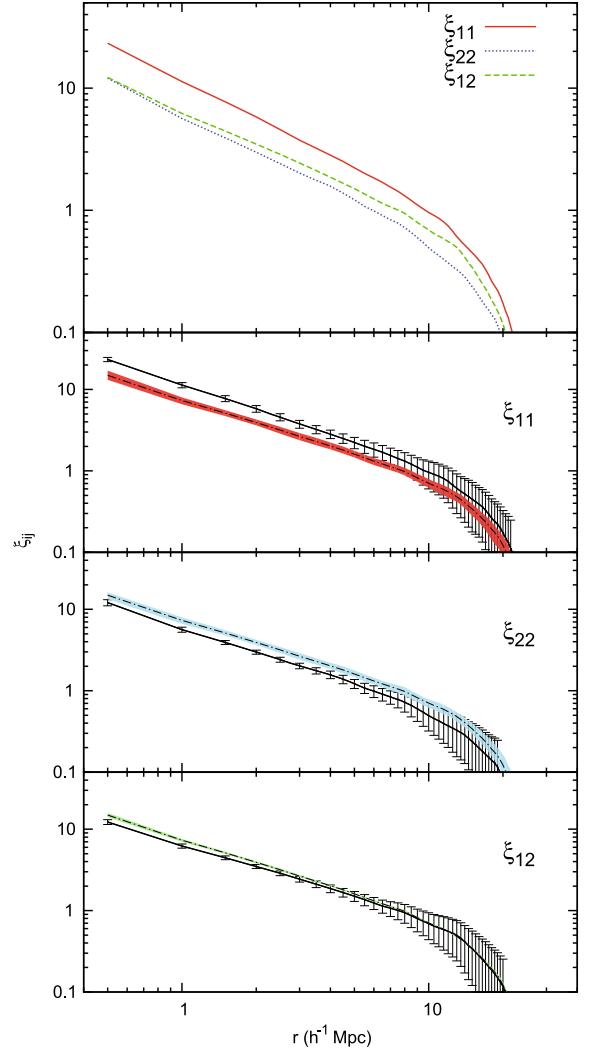
### 4.2. The normalized mark correlation function

The  $k_{mm}(r)$  for our sample, estimated according to Eq. (5), is shown in Fig. 3. The curve for  $k_{mm}(r)$  shows a weak negative correlation or spatial inhibition:  $k_{mm}(r) < 1$ . The range of correlation is about  $20 h^{-1}$  Mpc, where  $k_{mm}(r)$  gets values close to 1. It is interesting to compare this result with the  $k_{mm}(r)$ -function shown in Beisbart & Kerscher (2000) using the galaxy absolute luminosity  $L$  as the mark. They obtain an increasing behavior of  $k_{mm}(r)$  at small scales with  $k_{mm}(r) > 1$  for  $r < 12 h^{-1}$  Mpc, showing that bright galaxies are stronger correlated than faint ones. In our case, the tendency of the values of  $k_{mm}(r)$  to be smaller than 1 at short scales indicates that the price for galaxies to be close together is to have reduced values of the marks, i.e., to be more passive.

### 4.3. The mark variogram

In Fig. 4 we show the mark variogram for our sample, obtained according to Eq. (7). This function is monotonously increasing. In this case the interpretation is straightforward:  $\gamma(r)$  shows that for separations of  $r \lesssim 10 h^{-1}$  Mpc, galaxy pairs tend to have similar marks, that is, similar spectral type.

This result is partially explained by the previous one shown by the  $k_{mm}$  function: galaxies close together exhibit smaller values of the attached mark (spectral type).



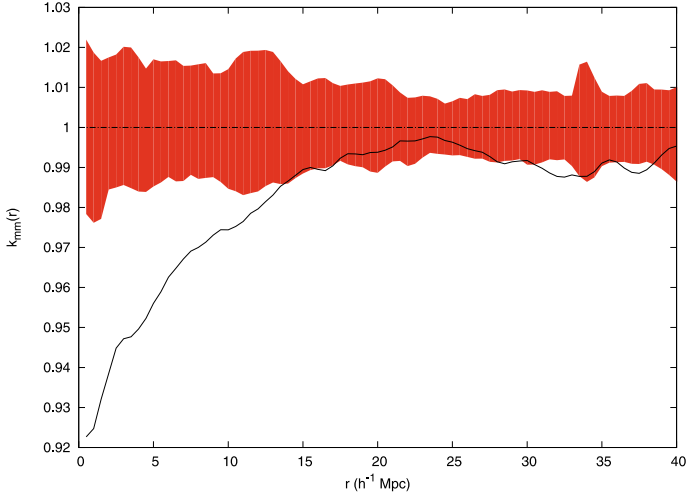
**Fig. 2.** The partial two-point correlation functions  $\xi_{ij}(r)$  estimated for population “1” (early-type) and population “2” (late-type) galaxies in our sample. The top panel shows the three functions together. The three lower panels show each of them separately (solid lines with error bars estimated using the jackknife method), together with a shaded band showing the minimum and maximum values for the 200 realizations of the random relabeling simulation. The dot-dashed lines correspond to  $\xi(r)$  for the full sample, which is the expected value of all  $\xi_{ij}(r)$  in the absence of segregation.

### 4.4. The mark connection function

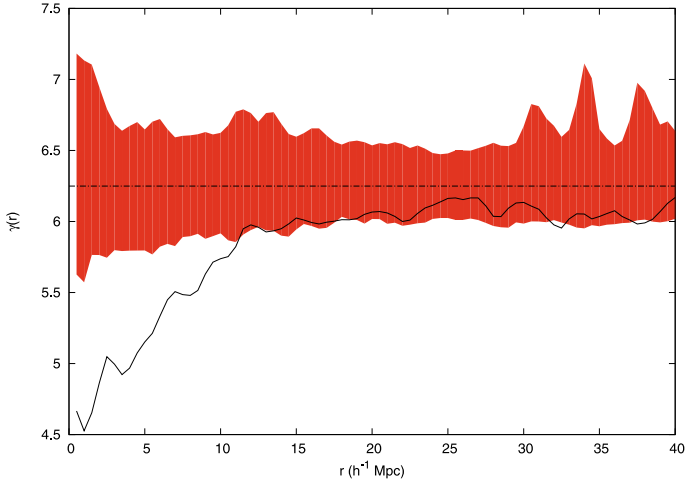
We show the  $p_{ij}(r)$  obtained for the 2dFGRS galaxies together with the results of our random relabeling simulations in Fig. 5. The first panel shows very neatly that for scales of  $r \lesssim 20 h^{-1}$  Mpc the clustering of early-type galaxies is stronger than the clustering of late-type galaxies. The three bottom panels show that the deviation of the observed  $p_{ij}(r)$  from the case of random labeling is significant at these scales.

Moreover, the figure shows clear differences in the spatial correlations of galaxies of the two types. In an overall clustering of all galaxies, we can outline that:

1. Galaxies of the type “1” (passive or early-type) are strongly clustered up to distances of  $20 h^{-1}$  Mpc.
2. The conditional probability to find two galaxies of the type “2” (active or late-type) at two positions separated by a distance  $r$  (under the condition that at these locations are



**Fig. 3.** Normalized mark correlation function  $k_{mm}(r)$  for our sample (solid line). The shaded band shows the minimum and maximum values for the 200 realizations of the random relabeling simulation, while the dot-dashed line corresponds to the value for the case with no segregation,  $k_{mm}(r) = 1$ .



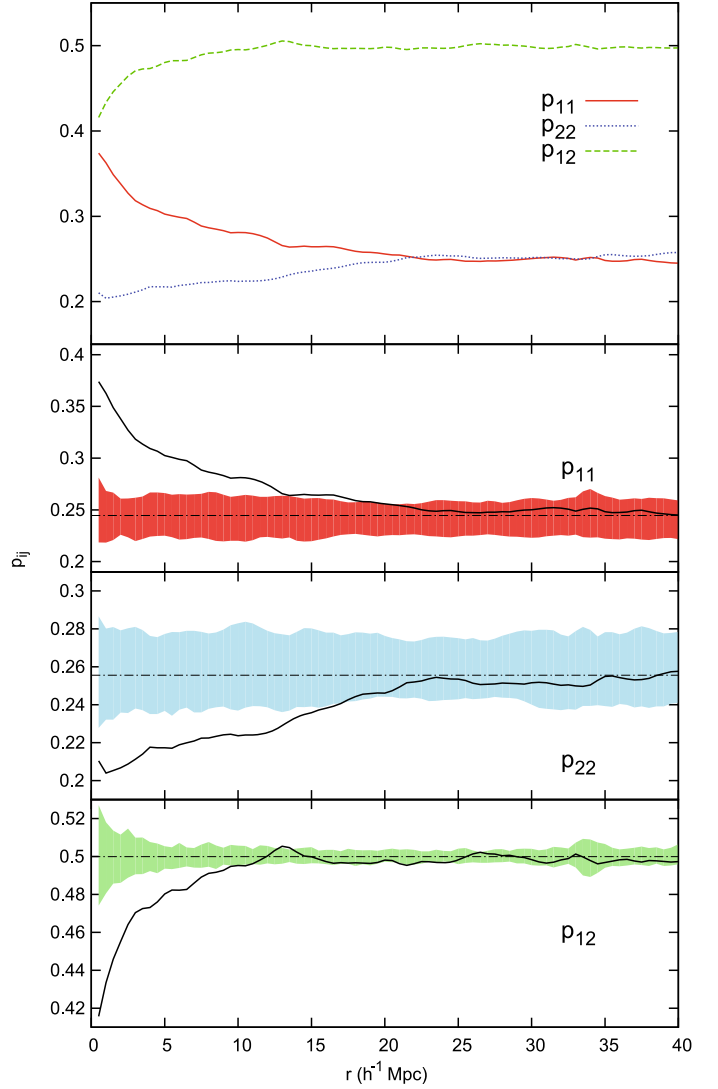
**Fig. 4.** Mark variogram  $\gamma(r)$  for our sample (solid line). The shaded band shows the minimum and maximum values for the 200 realizations of the random relabeling simulation, while the dot-dashed line corresponds to the value for the case with no segregation,  $\gamma(r) = \sigma_m^2 = 6.25$ .

galaxies) is smaller than the same probability for random labeling of the marks for scales of  $r \lesssim 20 h^{-1} \text{ Mpc}$ .

- Galaxy pairs that have one member of the type “1” and the other member of type “2” are less frequent than for random labeling up to distances of  $10 h^{-1} \text{ Mpc}$ .

In summary, all galaxies form a highly clustered pattern. In this pattern, the passive galaxies tend to be close to other passive galaxies, while positioning of active galaxies is less affected by other active galaxies. However, they tend to avoid positions close to passive galaxies.

This clearly shows the power of the mark connection function as an analytical tool in comparison to the partial pair correlation function. While for the untrained eye the curves in Fig. 2 are quite similar and show little structure, the curves in Fig. 5 give valuable information about the inner structure of the mark distribution. Obviously, the idea to consider characteristics of the nature of conditional probabilities helps to divulge structural details which would be otherwise overlooked.



**Fig. 5.** Mark connection functions  $p_{ij}(r)$  obtained for “early-type” (population “1”) and “late-type” (population “2”) galaxies in our sample. The top panel shows the three functions together. The three bottom panels show  $p_{11}(r)$ ,  $p_{22}(r)$ , and  $p_{12}(r)$  separately (solid lines), together with the shaded band showing the minimum and maximum values for the 200 realizations of the random relabeling simulation. The dot-dashed lines correspond to the expected values for the random labels case according to Eq. (8).

The problem are the mutual positions, given the positions of all galaxies without mark information. Since the three partial two-point correlation functions shown in Fig. 2 are different for a large range of scales, the marking with marks 1 and 2 cannot be an independent marking, where every galaxy obtains its mark randomly, independent of the other galaxies. In contrast, there must exist a spatial correlation between the marks. As it was shown in Fig. 5, the mark connection function is the appropriate tool to measure this correlation.

## 5. Conclusions

We used a volume-limited galaxy sample from the 2dFGRS to test different statistical measures used to disentangle mark segregation in the distribution of the galaxies. The mark attached to each galaxy of the sample was its spectral type  $\eta$ . For some of the statistics, the value of the mark enters directly into the

functions used to measure segregation: the normalized mark correlation function  $k_{mm}(r)$  and the mark variogram  $\gamma(r)$ . For other functions, like the partial correlation functions or the mark connection function, the sample has been split into two populations corresponding to passive or early-type galaxies with  $\eta \leq -1.4$  and active or late-type galaxies with  $\eta > -1.4$ . We summarize our results below:

1. The partial correlation functions, including the cross-correlation function, inform us about the degree of clustering of each population separately. It shows that passive galaxies exhibit stronger clustering at small separation. Nevertheless, there is no information about the spatial correlation between the marks.
2. The normalized mark correlation function shows that smaller values of the marks, i.e., smaller values of spectral type (being more passive), is a clear condition for galaxies to be close to each other in the overall clustering pattern.
3. The mark variogram shows in addition that at small separations galaxy pairs tend to have similar marks.
4. The mark connection function has been introduced here for the first time in the analysis of the marked galaxy distribution. The function  $p_{ij}(r)$  measures the conditional probability to find at two positions, separated by a distance  $r$ , a galaxy of the type “ $i$ ” and a galaxy of type “ $j$ ” under the condition that at these positions there are indeed galaxies. This function yields information different from that of the partial correlation functions  $\xi_{ij}(r)$ . This more sophisticated measure, having a nature of conditional quantities, is an efficient statistical tool to characterize the spatial correlation between the marks, filtering out the relative frequencies of the mark pairs  $(i, j)$  at a distance  $r$ .

Applied on the 2dFGRS volume-limited sample, the mark connection function clearly shows that passive galaxies are clustered up to distances of  $20 h^{-1}$  Mpc, while active galaxies exhibit weak spatial anticorrelation of the mark up to distances of  $20 h^{-1}$  Mpc. Mixed pairs are less frequent up to distances of  $10 h^{-1}$  Mpc.

*Acknowledgements.* First, we thank the anonymous referee for detailed and constructive criticism and suggestions. We are pleased to thank the 2dFGRS Team

for the publicly available data releases. We thank D. Croton for the 2dFGRS samples and the mask data and M. J. Pons-Bordería for comments and suggestions. This work has been supported by the Spanish Ministerio de Ciencia e Innovación CONSOLIDER projects AYA2006-14056 and CSD2007-00060, including FEDER contributions, and by the Generalitat Valenciana project of excellence PROMETEO/2009/064. PAM acknowledges support from the Spanish Ministerio de Educación through a FPU contract.

## References

- Beisbart, C., & Kerscher, M. 2000, *ApJ*, 545, 6
- Croton, D. J., Colless, M., Gaztañaga, E., et al. 2004, *MNRAS*, 352, 828
- Davis, M., & Geller, M. J. 1976, *ApJ*, 208, 13
- Davis, M., Meiksin, A., Strauss, M. A., da Costa, L. N., & Yahil, A. 1988, *ApJ*, 333, L9
- Domínguez-Tenreiro, R., & Martínez, V. J. 1989, *ApJ*, 339, L9
- Domínguez-Tenreiro, R., Gómez-Flechoso, M. A., & Martínez, V. J. 1994, *ApJ*, 424, 42
- Dressler, A. 1980, *ApJ*, 236, 351
- Guzzo, L., Strauss, M. A., Fisher, K. B., Giovanelli, R., & Haynes, M. P. 1997, *ApJ*, 489, 37
- Hamilton, A. J. S. 1988, *ApJ*, 331, L59
- Hermit, S., Santiago, B. X., Lahav, O., et al. 1996, *MNRAS*, 283, 709
- Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. 2008, *Statistical Analysis and Modelling of Spatial Point Patterns*, Statistics in Practice (Chichester: John Wiley & Sons)
- Loveday, J., Maddox, S. J., Efstathiou, G., & Peterson, B. A. 1995, *ApJ*, 442, 457
- Madgwick, D. S., Lahav, O., Baldry, I. K., et al. 2002, *MNRAS*, 333, 133
- Madgwick, D. S., Hawkins, E., Lahav, O., et al. 2003, *MNRAS*, 344, 847
- Martínez, V. J., & Saar, E. 2002, *Statistics of the Galaxy Distribution* (Boca Raton: Chapman & Hall/CRC)
- Norberg, P., Baugh, C. M., Gaztañaga, E., & Croton, D. J. 2009, *MNRAS*, 396, 19
- Peebles, P. J. E. 1980, *The large-scale structure of the universe* (Princeton: Princeton University Press)
- Phillipps, S., & Shanks, T. 1987, *MNRAS*, 229, 621
- Pons-Bordería, M., Martínez, V. J., Stoyan, D., Stoyan, H., & Saar, E. 1999, *ApJ*, 523, 480
- Salzer, J. J., Hanson, M. M., & Gavazzi, G. 1990, *ApJ*, 353, 39
- Schlather, M. 2001, *Bernoulli*, 7, 99
- Skibba, R. A., Bamford, S. P., Nichol, R. C., et al. 2009, *MNRAS*, 399, 966
- Stoyan, D., & Stoyan, H. 1994, *Fractals, Random Shapes and Point Fields*, Probability and Mathematical Statistics (Chichester: John Wiley & Sons)
- Vogeley, M. S., Geller, M. J., & Huchra, J. P. 1991, *ApJ*, 382, 44
- Wälder, O., & Stoyan, D. 1996, *Biometrical J.*, 38, 895
- Zehavi, I., Blanton, M. R., Frieman, J. A., et al. 2002, *ApJ*, 571, 172