# Clustering of the AKARI NEP Deep Field mid infrared selected galaxies

Aleksandra Solarz[1], Agnieszka Pollo[1,2],
Tsutomu T. Takeuchi[3] and Katarzyna Małek[1]

1. National Centre of Nuclear Research
   Hoża 69, 00-681 Warszawa, Poland
2. Astronomical Observatory of Jagiellonian University
   Orla 171, 30-244, Kraków, Poland
3. Division of Particle and Astrophysical Science, Nagoya University
   Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan

We present a method of selection of 24 $\mu$m galaxies from the AKARI North Ecliptic Pole (NEP) Deep Field and measurements of their two-point correlation function. We aim to associate different 24 $\mu$m selected galaxy populations with present day galaxies, and to investigate the impact of their environment on the direction of their subsequent evolution. We discuss the use of Support Vector Machines (SVM) algorithms applied to infrared photometric data to perform star-galaxy separation, in which we achieve an accuracy $> 80\%$. We explore the redshift dependance of the correlation function parameters as well as the linear bias evolution (which relates galaxy distribution to the one of the underlying dark matter). We find that the bias parameter increases slowly with redshift, from $b = 0.9$ at $z < 0.5$ to $b \sim 1.9$ at $z \sim 1.1$. Total infrared luminosities ($L_{TIR}$) found for different samples, suggest that galaxies with higher $L_{TIR}$ do not necessarily reside in higher mass dark matter halos. We find that luminous infrared galaxies (LIRGs) at $z \sim 1$ can be ancestors of present day $L_*$ early type galaxies.

## 1 Introduction

Star formation is one of the most influential factors governing the galaxy evolution. The bulk of star formation activity, according to the hierarchical model of structure formation, has shifted from high density regions of the Universe to low density ones and cosmic star formation rate (SFR) has been reported to undergo a sharp decrease since $z \sim 1$. From the optical sky surveys it is known that the distribution of colors is roughly bimodal, divided roughly at stellar masses $\sim 10^{10} M_{\odot}$ into a sequence of blue, star-forming galaxies and passively evolving red galaxies, which exhausted their star formation fuel at earlier epochs. This relation holds through a large observational distance, from the observations of the local Universe up to $z \sim 1.5$, however the fractions of the respective galaxy populations vary. Moreover, there are reports about the differences in densities of red and blue galaxies with cosmic epoch: as the density of red galaxies since $z \sim 2$ has been reported to increase, while the density of star-forming galaxies remains roughly constant. This could indicate that blue galaxies, have somehow quenched their star-formation and shifted from blue to red sequence at a similar pace to which other blue galaxies have maintained the ongoing star formation. Quenching of the star formation can be triggered through many process

(e.g. gravitational interactions with other systems like galaxy mergers harassment, gas stripping; activation of active galactic nucleus, AGN; supernovae feedback), which can lead to morphological changes in the structure of a galaxy and can contribute to removal of gas necessary for fueling the star formation.

Through measurements of clustering of star-forming galaxies appearing at different cosmic epochs it is possible to determine what are the possible masses of dark matter haloes, in which those galaxies reside (e.g. Zehavi et al. 2011). In turn, this implies the typical environment in which those galaxies are found, and therefore can help to constrain the mechanisms which are responsible for the truncation of star formation. Moreover, since the dark matter interacts only gravitationally one can determine the space density and clustering of dark matter haloes as a function of redshift. This means that we can use the galaxy clustering to connect the distant populations of star forming galaxies to galaxies which are observed in the local Universe, and connect them into an evolutionary sequence.

The AKARI satellite was designed to carry out IR observations with a sensitivity and resolution higher than that of preceding missions. It was launched by JAXA's MV8 vehicle on February 22, 2006, carrying out, amongst others, a deep survey of the north ecliptic pole (NEP) region, which we aim to use to explore the MIR properties of galaxies, in particular, the evolution of clustering. The NEP Deep Sky survey covers an area of 0.4 sq. deg around the NEP Matsuhara et al. (2006). The data were obtained by the Infrared Camera (IRC; Onaka et al. 2007) through nine near-IR (NIR) and mid IR (MIR) filters, ranging from at 2 $\mu$m to 24 $\mu$m. With a dense IR wavelength coverage, AKARI is well suited for delivering an independent look at the evolution of composition and clustering of different IR galaxy populations. In this research, we study the clustering properties of 24 $\mu$m galaxies from a perspective of an infrared satellite other than Spitzer, utilizing a new source selection method based on the usage of IR data alone.

## 2 Sample Selection

To separate stars and galaxies we employ the Support Vector Machines (SVM) algorithm, a supervised method based on kernel methods (Shawe-Taylor & Cristianini 2004), allowing for pattern recognition within the provided data. The advantage that the SVM algorithm has over other algorithms is the ability to use all available information simultaneously. The first attempt to separate stars and galaxies within the NEP data was presented in Solarz et al. 2012, where we exploited the Support Vector Machines (SVM) algorithm using all available color information for AKARI objects. Here, we employ the same method, this time aiming at selection of a pure 24 $\mu$m galaxy sample, which we use later on to explore the clustering properties of these kinds of objects.

Here we outline the general procedure which was followed to obtain the catalog via SVM: To create a catalog of 24 $\mu$m AKARI galaxies, we performed a typical four-step routine for the application of SVMs to the classification task, as follows.

- Manual selection of subsets of objects, which are representative to their predefined class, and serve as a training basis for creating a classifier.
- Each training source has to be described by its discriminating properties; in other words, for each training example there has to be a corresponding feature vector.

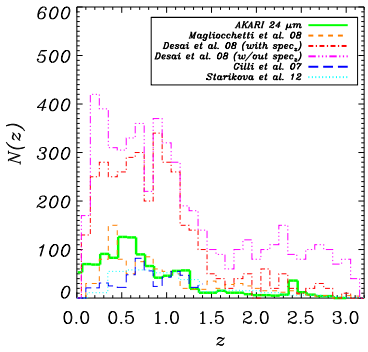


Fig. 1: Redshift distribution of the AKARI NEP Deep Field MIR selected galaxies derived by CIGALE code (solid line) together with those published in the literature (dashed: Magliocchetti et al. 2008; dash-dot-dashed and dash-triple-dotted: Desai et al. 2008, for objects with and without spectroscopic redshifts; large-dashed: Gilli & Daddi 2007; and dotted: Starikova et al. 2012).

- Selection of the kernel function. In this research we only use the radial basis function.
- Training the algorithm to learn how to distinguish objects by creating a separation hyperplane described by two adjustable parameters $(C, \sigma)$.

The final pair of parameters, $(C, \sigma) = (10^0, 10^{-1})$, was chosen based on the best accuracy of correctly classifying stars and galaxies within the training sample. The accuracy is defined as a ratio of all galaxies and stars whose nature was properly recognized by the classifier to the total number of considered objects. The final classifier was applied to all sources detected in the 24 $\mu$m band (2208 objects) to infer their respective classes. The resultant classifier exhibited a 83.89 % total accuracy, with the true galaxy rate ($TGR = TG/(TG + FS)$) and true star rate ($TSR = TS/(TS + FG)$) for selecting stars and galaxies as 97.69 % and 66.40 %, respectively. Here $TS$ stands for a true star (actual star identified as a star by SVM), $TG$ is a true galaxy (actual galaxy identified as a galaxy by SVM), $FG$ is a false galaxy (actual star classified as a galaxy by SVM) and $FS$ is a false star (actual galaxy classified as a star by SVM).

The redshift distribution was obtained by implementing the *Code Investigating GALaxy Emission* (CIGALE; Noll et al. 2009), a spectral energy distribution (SED) fitting routine; final redshift distribution is presented in Fig. 1.

## 3 Methodology

Galaxy clustering is traditionally analyzed in the first place by the means of the two-point correlation function, $\xi(r)$, which is treated as an above random excess probability of finding a galaxy within a certain distance $r$ from another galaxy. Often it is approximated by a power law $\xi(r) = (r/r_0)^{(-1/\gamma)}$. To compute the angular correlation function, we use the Landy-Szalay (1993) estimator:

$$\omega(\theta) = \frac{\mathrm{DD}(\theta) - 2\mathrm{DR}(\theta) + \mathrm{RR}(\theta)}{\mathrm{RR}(\theta)}, \tag{1}$$

where DD($\theta$) is the number of galaxy-galaxy pairs, RR($\theta$) is the number of random-random pairs, and DR($\theta$) is the number of galaxy-random pairs within an angular bin of separation on the sky $\theta$. We use $10^5$ homogeneously generated random points within the FOV and overlay the photometric mask with the identical features as the real FOV.

Using the measurements of the angular clustering, we can infer the three-dimensional clustering properties based on the known redshift distribution via Limber's equation Limber (1953). When dealing with small scales both angular and spatial correlation functions are well described by power laws and therefore Limber's equation can be presented in the form (Efstathiou et al. 1991),

$$\mathrm{A}_\omega = \mathrm{C}_\gamma \mathrm{r}_0^\gamma \frac{\int \mathrm{d_A^{1-\gamma} x^{-1}(z)(dN/dz)^2 dz}}{[\int (\mathrm{dN/dz})]^2}, \tag{2}$$

where $d_A$ is the angular diameter distance, $x(z)$ is the derivative of proper distance with redshift, $C_\gamma = \frac{\sqrt{\pi}\Gamma(\frac{\gamma-1}{2})}{\Gamma(\gamma/2),}$ and $dN/dz$ is the redshift selection function, which in case of our study is derived empirically from SED fitting (see Fig. 1).

In order to relate galaxy clustering to dark matter clustering, a bias parameter was introduced; a quantity describing the differences between the clustering of a galaxy field and the underlying mass distribution, i.e., $b^2(r, z, M) = \xi_g(r, z, M)/\xi_m(r, z)$, where $\xi_g(r, z, M)$ is the correlation function of the investigated galaxy population, and $\xi_m(r, z)$ is the correlation function of the dark matter. Bias is dependent on scale ($r$), redshift ($z$) and objects' mass ($M$). Following Magliocchetti et al. (2008), a linear bias can be defined as:

$$b_8 = \frac{\sqrt{C_\gamma (r_0(z)/8)^\gamma}}{\sigma_{8,mass(z)}}, \quad C_\gamma = 72/((3-\gamma)(4-\gamma)(6-\gamma)2^\gamma).$$

here a linear evolution of the rms mass fluctuations with cosmic time is assumed: $\sigma_{8,mass}(z) = \sigma_{8,mass}(z=0)D(z)$, where $D(z)$ is a linear growth factor.

## 4 Results and discussion

The final galaxy catalog was split into four redshift ranges according to the shape of redshift distribution: 1) the low-z samplecontaining galaxies with $z \leq 0.5$; 2) the low-intermediate sample with $0.5 < z \leq 0.9$, which includes the primary and most pronounced peak of the distribution which could arise from 16.3 $\mu$m and/or 17 $\mu$m PAH emission features passing through 24 $\mu$m passband; 3) the intermediate sample with $0.9 < z \leq 1.3$, straddling the secondary peak (attributed to the 12.7 $\mu$m PAH and/or 12.8 $\mu$m $[Neii]$ emission features, which at these redshifts would appear in the 24 $\mu$m passbands); and 4) the high-z sample with $z > 1.3$, a range, which includes the peak at$z \sim 2.4$, which is statistically insufficient, as we expect a significant incompleteness due to sources not reaching the detection limits. Deep silicate absorption features at $\sim 9$. $\mu$m at $z \sim 1.5$ would appear in $\sim$24 $\mu$m passbands, making otherwise potentially luminous galaxies faint at those wavelengths. For the subsamples divided in this way, we estimated their respective angular correlation functions, spatial correlation lengths, $L_{TIR}$ (as calculated by the CIGALE code), and linear bias (see Table 1). In order to infer the minimum mass of the dark matter haloes we have used models of Sheth & Tormen (1999) and matched them with the $b_8$ parameter (see Fig. 2[a]). For $z \leq 0.5$ subsample of galaxies with $L_{TIR} \sim 10^{10} L_\odot$ we found the

Table 1: Clustering of 24 $\mu$m galaxies in AKARI NEP Deep Field.

| $z$ range | $N_{gal}$ | $\gamma$ | $r_0$ [$h^{-1}$Mpc ] | $L_{TIR}$ | b |
|---|---|---|---|---|---|
| $z \leq 0.5$ | 168 | $1.99 \pm 0.06$ | $3.62 \pm 0.76$ | $10.276 \pm 0.81$ | $0.89 \pm 0.20$ |
| $0.5 < z \leq 0.9$ | 207 | $1.69 \pm 0.03$ | $6.21 \pm 0.78$ | $11.15 \pm 0.315$ | $1.72 \pm 0.25$ |
| $0.9 < z \leq 1.3$ | 159 | $1.57 \pm 0.10$ | $5.86 \pm 0.69$ | $11.84 \pm 0.36$ | $1.91 \pm 0.21$ |
| $z > 1.3$ | 77 | $1.85 \pm 0.12$ | $7.23 \pm 0.87$ | $12.43 \pm 0.44$ | $3.53 \pm 0.51$ |

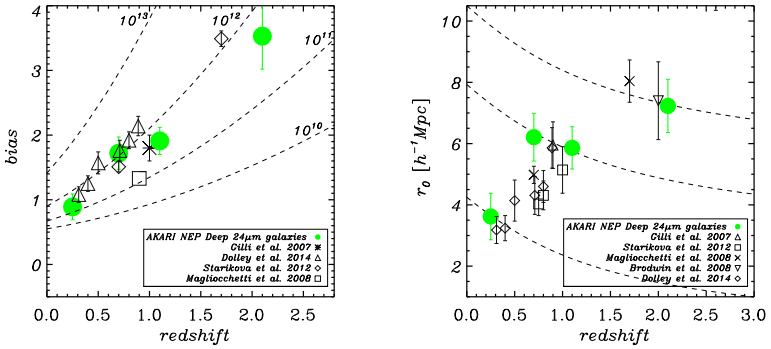


Fig. 2: Linear bias [a] and correlation length as a function of redshift for AKARI photometric redshift samples (filled circles). Dashed curves on [a] represent the theoretical linear halo bias evolution of dark matter halos of minimal masses: $10^{10}$, $10^{11}$, $10^{12}$, and $10^{13}$ (from bottom to top); on [b] dashed curves represent evolutionary tracks as expected from the linear scenario (dashed lines) compared with other research.

correlation length $r_0 = 3.62 \pm 0.79\ h^{-1}$ Mpc. This indicates that this redshift interval is mostly composed of normal SF galaxies. Their bias parameter is equal to $0.93 \pm 0.11$ and their estimated minimal DMH mass is $M_h > 10^{11} M_\odot$. For $0.5 < z \leq 0.9$ and $0.9 < z \leq 1.3$ subsamples we found correlation lengths are $6.21 \pm 0.78\ h^{-1}$ Mpc and $5.86 \pm 0.69\ h^{-1}$ Mpc, respectively. Moreover, galaxies contained in those redshift intervals exhibit very similar, relatively high values of bias parameter ($\sim 1.7 - 1.9$). The sample at $0.9 < z \leq 1.3$, composed of brighter galaxies ($L_{TIR} \sim 10^{11.84} L_\odot$), seems to be residing in lower minimal mass DMH ($M_h < 10^{12} M_\odot h^{-1}$) than the lower luminosity (with $L_{TIR} \sim 10^{11.15} L_\odot$ and $M_h > 10^{12} M_\odot h^{-1}$) galaxies at $0.5 < z \leq 0.9$. This could indicate that brighter infrared galaxies do not necessarily reside in more massive halos. This means that despite similar clustering properties, we are dealing with two different populations of star-forming galaxies. The redshift distribution shows two distinct peaks (at $\sim 0.6$ and $\sim 1.2$): the primary peak in the redshift distribution at $\sim 0.6$ could contain a mix of both normal star-forming galaxies and LIRGs. The dip in the distribution between the peaks at $\sim 0.9$ could mean that normal galaxies slowly fade out from the field of view to reveal a more prominent LIRG population at redshifts $\sim 0.9$ and higher.

Extrapolating the AKARI data to the present day epoch (Fig. 2[b]) has allowed us to investigate the approximate descendants of each galaxy subsample. Working under the assumption that galaxy interactions do not affect the $\sigma_m$ value at scales as large as 8 Mpc, 24 $\mu$m galaxies at redshifts $0.5 < z < 1.3$ could evolve into current

galaxies with $r_0 \sim 8\ h^{-1}$Mpc, a value measured for ellipticals with $L \sim L_*$.

Galaxies at $z > 1.3$ display a strong clustering signal at small scales despite the large size of the redshift range. This could indicate an existence of separate, highly clustered population(s). However, because of small number statistics, these results remain to be confirmed.

*Acknowledgements.* This work is based on observations with AKARI, a JAXA project with the participation of ESA. AS has been supported by the National Science Centre grant UMO-2015/16/S/ST9/00438. AP has been supported by the National Science Centre grants UMO-2012/07/B/ST9/04425 and UMO-2013/09/D/ST9/04030. This research was partially supported by the project POLISH-SWISS ASTRO PROJECT co-financed by a grant from Switzerland through the Swiss Contribution to the enlarged European Union. TTT has been supported by the Grant-in-Aid for the Scientific Research Fund (23340046, and 24111707), for the Global COE Program Request for Fundamental Principles in the Universe: from Particles to the Solar System and the Cosmos commissioned by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, and for the JSPS Strategic Young Researcher Overseas Visits Program for Accelerating Brain Circulation, "Construction of a Global 7 Platform for the Study of Sustainable Humanosphere."

## References

Desai, V., et al., *Redshift Distribution of Extragalactic 24 μm Sources*, ApJ **679**, 1204 (2008), 0802.2489

Efstathiou, G., et al., *The clustering of faint galaxies*, ApJ **380**, L47 (1991)

Gilli, R., Daddi, E., *The Spatial Clustering of Mid-IR Selected Sources in the GOODS Fields*, in J. Afonso, H. C. Ferguson, B. Mobasher, R. Norris (eds.) Deepest Astronomical Surveys, *ASP Conference Series*, volume 380, 409 (2007)

Limber, D. N., *The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field.*, ApJ **117**, 134 (1953)

Magliocchetti, M., et al., *On the evolution of clustering of 24-μm-selected galaxies*, MNRAS **383**, 1131 (2008), 0710.3136

Matsuhara, H., et al., *Deep Extragalactic Surveys around the Ecliptic Poles with AKARI (ASTRO-F)*, PASJ **58**, 673 (2006)

Noll, S., Burgarella, D., Giovannoli, E., et al., *Analysis of galaxy spectral energy distributions from far-UV to far-IR with CIGALE: studying a SINGS test sample*, *A&A* **507**, 1793 (2009)

Onaka, T., et al., *The Infrared Camera (IRC) for AKARI – Design and Imaging Performance*, PASJ **59**, 401 (2007)

Shawe-Taylor, S., Cristianini, N., Kernel Methods for Pattern Analysis, Cambridge, UP, Cambridge, UK (2004)

Sheth, R. K., Tormen, G., *Large-scale bias and the peak background split*, MNRAS **308**, 119 (1999)

Solarz, A., et al., *Star-galaxy separation in the AKARI NEP deep field*, A&A **541**, A50 (2012), 1203.1931

Starikova, S., et al., *Clustering of Star-forming Galaxies Detected in Mid-infrared with the Spitzer Wide-area Survey*, ApJ **751**, 126 (2012), 1205.2045

Zehavi, I., et al., *Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity*, ApJ **736**, 59 (2011), 1005.2413