*R*esearch in
*A*stronomy and
*A*strophysics

The tidal tails of globular cluster Palomar 5 based on the neural networks method [*]

Hu Zou[1,2], Zhen-Yu Wu[1], Jun Ma[1] and Xu Zhou[1]

[1] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China; *zywu@bao.ac.cn*
[2] Graduate University of Chinese Academy of Sciences, Beijing 100049, China

Abstract The sixth Data Release (DR6) of the Sloan Digital Sky Survey (SDSS) provides more photometric regions, new features and more accurate data around globular cluster Palomar 5. A new method, Back Propagation Neural Network (BPNN), is used to estimate the cluster membership probability in order to detect its tidal tails. Cluster and field stars, used for training the networks, are extracted over a 40×20 deg^2 field by color-magnitude diagrams (CMDs). The best BPNNs with two hidden layers and a Levenberg-Marquardt (LM) training algorithm are determined by the chosen cluster and field samples. The membership probabilities of stars in the whole field are obtained with the BPNNs, and contour maps of the probability distribution show that a tail extends $5.42°$ to the north of the cluster and another tail extends $3.77°$ to the south. The tails are similar to those detected by Odenkirchen et al., but no more debris from the cluster is found to the northeast in the sky. The radial density profiles are investigated both along the tails and near the cluster center. Quite a few substructures are discovered in the tails. The number density profile of the cluster is fitted with the King model and the tidal radius is determined as $14.28'$. However, the King model cannot fit the observed profile at the outer regions ($R > 8'$) because of the tidal tails generated by the tidal force. Luminosity functions of the cluster and the tidal tails are calculated, which confirm that the tails originate from Palomar 5.

Key words: methods: statistical — galaxy: halo — galaxy: structure — globular cluster: individual (Palomar 5)

1 INTRODUCTION

Globular clusters (GCs) are the oldest populations in the Galaxy. Most GCs, which formed in the early days of the Galaxy, have been destroyed by various mechanisms (Wu et al. 2003). Mass loss from stellar evolution is very important during the first ~ 1 Gyr of cluster evolution, and most low mass clusters were dissolved during this early phase. For surviving clusters, their evolution will be dominated by internal dynamical processes caused by encounters between cluster stars (the two-body relaxation) (Spitzer 1987). GCs in the Galaxy have elliptical orbits, and some of them can move into the central region of the Galaxy with perigalactic distances less than ~ 1 kpc (Wu et al. 2004). When a cluster crosses the bulge or disk of the Galaxy with a timescale shorter than its internal dynamical time, the cluster stars will gain energy and speed up the evaporation. Such an interaction is referred to as tidal

shock (Spitzer 1987). Stars evaporating from the cluster due to two-body relaxation or tidal shock will not leave the cluster and merge into the Galactic field immediately. They will move along the same orbit of the cluster and form the 'tidal tail' of the cluster.

Grillmair et al. (1995) examined the outer structures of 12 Galactic globular clusters using star-count analysis with deep, two-color photographic photometry. They found that most of their sample clusters show extra-tidal wings in their surface density profiles. Two-dimensional surface density maps for several clusters indicate the expected appearances of tidal tails. Leon et al. (2000) used large-field photographic photometry of 20 globular clusters to investigate the presence of tidal tails around these clusters; in this study, star-count analysis and wavelet transform were used to detect the weak structures formed by the stars that were previously members of the clusters; and most globular clusters in their sample display large and extended tidal tails, which exhibit projected directions towards the Galactic center.

The studies of Grillmair et al. (1995) and Leon et al. (2000) are all based on photographic observations covering large areas around the clusters. The low signal-to-noise in the photographic photometry and serious contaminants from the background galaxies make the detected tidal tails in some clusters uncertain. The Sloan Digital Sky Survey (SDSS) can provide large, deep CCD imaging in five passbands covering $10\,000\,\mathrm{deg}^2$ in the sky. The SDSS can also separate the stars and galaxies very well and is very efficient at detecting tidal tails around globular clusters in the Galaxy. Using SDSS data, well-defined tidal tails in some globular clusters have been identified: Palomar 5 (Odenkirchen et al. 2001, 2003; Grillmair & Dionatos 2006), NGC 5466 (Belokurov et al. 2006; Grillmair & Johnson 2006), and NGC 5053 (Lauchner et al. 2006).

In the above mentioned studies, Palomar 5 is a prominent object. It is a remote globular cluster located at a distance of 23.2 kpc from the Sun and has a tidal radius of about 16.3 arcmin (Harris 1996). Using the ESO and SERC survey plates covering about $2.0° \times 2.0°$ in the R and J filters, Leon et al. (2000) searched the tidal tails around this cluster and found that the detected structures outside this cluster are strongly biased by the background galaxy clusters appearing in the field, and it is difficult to derive any conclusions about the genuine location of stars in the tidal tails of this cluster.

Using SDSS data concentrating on Palomar 5 in a region with right ascension $226° \leq \alpha \leq 232°$ and declination $-1.25° \leq \delta \leq +1.25°$, Odenkirchen et al. (2001) searched the tidal tails of this cluster based on the empirical photometric filtering method of Grillmair et al. (1995). They found two well-defined tidal tails emerging from this cluster that stretch out symmetrically to both sides of the cluster and extend out with an angular seperation of $2.6°$ in the sky. Using the new SDSS data (before the public data release DR1) yields complete coverage of a region with a $6.5°$ to $8°$ wide zone along the equator and right ascension from $224°$ to $236°$; based on an optimal contrast filtering method, Odenkirchen et al. (2003) found that the tidal tails of Palomar 5 have a much larger spatial extent and can be traced about an arc of $10°$ on the sky. More recently, using the SDSS DR4 data, Grillmair & Dionatos (2006) applied the optimal contrast filtering method of Odenkirchen et al. (2003) to trace the tidal tails of Palomar 5 in a region $224° < \alpha < 247°$ and $-3° < \delta < +10°$, and found the tidal tails to extend some $22°$ over the sky.

Most of the studies are based on star-count analysis in color-magnitude space. However, Belokurov et al. (2006) gave us a brand new angle of view to recognize the tidal tails of clusters. Their method is based on an intelligent computing technique, Artificial Neural Networks (ANNs), which has been applied to many sorts of areas, such as classification and pattern recognition. In the study of Belokurov et al. (2006), a Back Propagation Neural Network (BPNN)— the most widely applied ANN — was used to estimate the probability of cluster membership for each object in the SDSS 5-band data space. Compared to Odenkirchen et al. (2003), BPNN makes full use of the photometric information, not just one CMD. Therefore, in this paper, we will introduce this method to investigate the tidal tails of Palomar 5, where the DR6 data in a larger region ($40 \times 20\,\mathrm{deg}^2$) are used.

In Section 2, we describe the details of the SDSS DR6 and preprocessing of the observed data. Section 3 presents the general idea of BPNNs. In Section 4, we construct BPNNs with the best performances after being trained with properly selected training data, and then we apply them to the tidal tail detection of Palomar 5. Section 5 discusses the profiles and features of the tails. Brief conclusions are given in Section 6.

2 THE STAR SAMPLE

2.1 Observations

In this study, the photometric data in the SDSS DR6 are used. The SDSS is a photometric and spectroscopic survey, providing detailed optical images covering more than a quarter of the sky and a 3-dimensional map of about a million galaxies and quasars. A dedicated, 2.5-meter telescope is located on Apache Point, New Mexico, equipped with a 120-megapixel camera and a pair of spectrographs fed by optical fibers measuring more than 600 sources in a single observation. There are 30 photometric CCDs with size 2048×2048 pixels for each. The field of view is $3.0°$, and five broad band filters with wavelengths ranging from 3000 Å to 10 000 Å are used when photometric images are taken. So far, subsequent data releases have been published, including Early Data Release (EDR), DR1, DR2, DR3, DR4, DR5, DR6 and DR7.

DR6 (Adelman-McCarthy et al. 2008) is the first release which includes significant changes to the processing software since DR2. For example, calibrations are improved using cross-scans that tie the photometry of the entire survey together. The photometric calibration is improved with uncertainties of roughly 1% in g, r, i and z, and 2% in u, which are substantially better than the ones in previous data releases. In addition, the magnitude limits are 22.0 for u, 22.2 for g and r, 21.3 for i, and 20.5 for z. More importantly, compared with DR4, DR6 includes newly observed regions where we can search for tidal tails of Palomar 5.

Odenkirchen et al. (2003) only considered a limited region covering an area of $87 \deg^2$ and found tails extending about an angular distance of $10°$. No further investigation in the north, where photometric data exist, was made to see whether the tails are longer. Grillmair & Dionatos (2006) detected a $22°$ tidal tail in a larger region. Far from the center of Palomar 5, these newly discovered tails do not clearly appear, since the signals and the background noise are so similar. On the other hand, Grillmair & Dionatos (2006) used DR4, in which a narrow strip ($\alpha > 228.5°$ and $0.5° < \delta < 1.5°$) stretching to the north tidal tail has no photometric data, while DR6 supplements this area. Therefore, in an area of $40 \times 20 \deg^2$ ($220° < \alpha < 260°$ and $-5° < \delta < 15°$), DR6 provides more photometric data. Due to the tremendous number of objects included in the area, we partition the whole field into 4 equal regions: R1 (R.A.: $220° \sim 230°$, Dec.: $-5° \sim 15°$); R2 (R.A.: $229° \sim 240°$, Dec.: $-5° \sim 15°$); R3 (R.A.: $239° \sim 250°$, Dec.: $-5° \sim 15°$) and R4 (R.A.: $249° \sim 260°$, Dec.: $-5° \sim 15°$), where any two contiguous regions overlap by an area of $1 \times 20 \deg^2$ in order to avoid bad smoothing at the edges (see Sect. 4).

The information, which we need to detect the tidal tails of Palomar 5, includes: coordinate (J2000), point spread function (PSF), magnitude (r_{psf}) and exponential model magnitude (r_{exp}) of r band, the reddening values and the type of each source. Details on the way to separate stars from galaxies, the definitions of r_{psf} and r_{exp}, photometric and astrometric data reduction, and relative information can be found in the works (Lupton et al. 2001; Stoughton et al. 2002; Abazajian et al. 2004). We use so-called Cmodel magnitudes (the default provided values of $ugriz$) as our photometric data, because a Cmodel magnitude is the best fit of the exponential and de Vaucouleurs models in each band, and it agrees excellently with both the PSF magnitude of stars and Petrosian magnitude of galaxies. Even though we only extract stars to check the tidal tails, it is impossible for us to guarantee that there are no miscellaneous galaxies which cannot be distinguished by SDSS data reduction pipelines. Thus, for uniformity and the validity of the photometry of both stars and galaxies, we choose the universal magnitudes (Cmodel). Reddening corrections for each object are deducted based on the reddening values from Schlegel et al. (1998).

In our selected sky field, we obtain 15 305 060 sources in the catalog, where there are 7 410 896 stars and 7 894 164 galaxies classified by the SDSS pipelines. There are 5 458 077 sources in R1, 5 508 433 in R2, 4 164 575 in R3 and 173 975 in R4, respectively.

2.2 Data Preprocessing

For source type determination in the SDSS, there are some flaws. For example, SDSS pipelines occasionally fail to distinguish blenders and pairs of stars with small separations. Sometimes, the classification scheme regards Seyfert galaxies or QSOs as stars (Stoughton et al. 2002), and overflows of very bright stars are identified as galaxies. Furthermore, due to variations of observing conditions and natural differences in diverse fields, the completeness of object detection is fluctuant. Considering magnitude limits and the situations depicted above, it is necessary to give cutoffs of magnitude to avoid unnecessary impurities. So, we select the sample stars with a magnitude scope $14 \leq r \leq 22$. Figure 1 shows a visual impression of the selected star sample, and the photometric boundaries. In Figure 1, only $1/30$ of the sample stars are randomly selected to be drawn in order not to be too black. M5 and Palomar 5 are also indicated in Figure 1. The subplot in this figure is the part including M5 and Palomar 5, which is enlarged. The galactocentric distance of M5 is about 6.2 kpc, away from Palomar 5, whose galactocentric distance is about 18.6 kpc. So, M5 hardly has any effect on Palomar 5.

Fig. 1 Distribution of all point sources within the magnitude range $14 \leq r \leq 22$ in the 40×20 deg^2 area. This area is divided into four overlapping regions (R1, R2, R3 and R4) to be processed in batches because of the very large number of objects. M5 (center: $\alpha = 229.6°, \delta = 2.1°$) and Palomar 5 (center: $\alpha = 229.02°, \delta = -0.11°$) are the two star-focus regions. The smaller box ($228.5° < \alpha < 230.5°, -0.5° < \delta < 2.5°$) encloses these two objects and the bigger one (subplot) shows a magnified image of the smaller one. The blank areas are regions which may be very dense clusters, bright stars or regions the SDSS has not covered.

Figure 2 presents the derived interstellar extinction distribution based on the values of $E(B - V)$ from Schlegel et al. (1998). The resolution of the distribution is about 6 arcmin. On the whole, there is smaller extinction in the northwest of the sky, but larger extinction in the southeast. The reddening correction in magnitude $E(B - V)$ is 0.057 on average, and the maximum and minimum are about 0.384 and 0.016, respectively. All the sources in the sample are dereddened.

Following Belokurov et al. (2006), the distributions of the magnitude difference, $r_{\text{psf}} - r_{\text{exp}}$, of both stars and galaxies are plotted in Figure 3. It is clear that stars are tightly concentrated around zero, whereas galaxies reveal a significant positive excess. Except for the cases mentioned in the previous

Fig. 2 Distribution of interstellar extinction from the Milky Way. The resolution of this image is 6 arcmin, and the $E(B - V)$ values used by the SDSS databse are derived from Schlegel et al. (1998). As shown in Fig. 1, the blank regions are M5, bright stars and areas not observed by the SDSS.

Fig. 3 Distribution diagrams of $r_{psf} - r_{exp}$ for both stars and galaxies. Kernel density estimation is used to fit the distributions with a normal kernel and 0.02 bandwidth (Bowman & Azzalini 1997). The two distributions are normalized to the peak value.

section, classification is unauthentic when galaxies are point-like. PSF is the best fit model to estimate the magnitude of a point source (r_{psf}), while the exponential model is one of the models used to calculate the magnitude of an extended one (r_{exp}). The SDSS constructs a simple classifier with an analogous difference between r_{psf} and r_{exp}. In Figure 3, $r_{psf} - r_{exp}$ is concentrated at zero for point-like sources, while it is far away from zero for extended sources. Both of them have an intersection where stars and galaxies cannot be distinguished. Also following Belokurov et al. (2006), we take a threshold $r_{psf} - r_{exp} = 0.05$ as the division between stars and galaxies. As a result, a sample of 4 082 662 point sources remains, which includes 1 079 301 in R1, 1 478 788 in R2, 1 378 560 in R3 and 146 013 in R4.

3 BACK PROPAGATION NEURAL NETWORK

BPNNs are now applied to various areas including astronomy, pattern classification, face and speech recognition and finance (Hangan et al. 1996; Haykin 1998). In astronomy, BPNN is used as a classifier in both photometric and spectral applications (von Hippel et al. 1994; Folkes et al. 1996), a morphological recognizer (Odewahn et al. 1992; Naim et al. 1995) in imaging, or a value estimator (Bailer-Jones 2000) in determining theoretical models and physical parameters. Virtually, the specific mechanism of BPNN is as follows: first, provide a train-test data set and train the configured BPNN with it, just as a teacher teaches a student to tell him which is a cluster star and which is a field one; then the BPNN learns the knowledge again and again to modify its inner configuration and adapts itself to perform better. During the course, a BPNN will be judged on its prior learned knowledge with test data; finally, through this kind of repeated train-test-modify cycle, the classifier (BPNN) gains the features and has the best-learned experience to challenge new things. A data processing tool MATLAB, which provides a special neural network toolbox to design and realize all kinds of ANNs, is introduced in our work. The definition of neural networks and other technical terms as well as the specified process of BPNNs are described in the appendix.

As mentioned previously, Belokurov et al. (2006) used neural networks to reconstruct the probability distribution of cluster stars with the SDSS $ugriz$ photometric data. The idea of this approach is very simple: with $ugriz$ 5-band photometric data of cluster members and field stars as inputs to a BPNN, we get an estimate of the cluster membership probability as the output after the network is well trained. This method makes full use of photometric information and constructs a probability estimator in high dimensional data space with limited resources. When being trained, the BPNN can pick out bad sources automatically to form an accurate separator.

With the sample strictly picked out, the first step in detecting tidal tails is to figure out all the cluster members of Palomar 5 in the selected field. A lot of pattern recognition techniques are available, such as the Bayes classifier, template matching, clustering analysis and artificial neural networks (see, e.g., Sergios & Konstantinos 2006, and references therein). In the present study, the only thing we need to do is to estimate the membership probability of an object. BPNNs can measure the posterior probability $P(C|\boldsymbol{x})$ in high dimensional space, where C denotes the cluster member class and $\boldsymbol{x}$ is the photometric data vector.

Finally, we make a summary of the basic parameters and components of the BPNN used in our paper. First, the dimension of the input layer is 5 (photometric magnitudes), and the transfer function of each neuron is a log-sigmoid function. Then, the dimension of the output vector is one, which yields 0 (for field stars) or 1 (for cluster stars). Mean squared errors (MSE) are used to calculate the deviations between the output of the BPNN and the real object type value. Finally, the Levenberg-Marquardt Back-progation (LMBP) algorithm is used to train the network to minimize the performance function MSE. The initial state (such as initial weights and biases, condition of termination, and parameters of the LMBP algorithm) is given automatically by MATLAB.

4 TIDAL TAIL DETECTION BASED ON BPNNS

4.1 Data Set Selection for Training and Testing

In order to apply the BPNN method to the tidal tail detection of Palomar 5, first of all, a training and test data set should be chosen to guide the BPNN to learn the knowledge of the cluster, so that it has the ability to figure out the probability of cluster membership. As Belokurov et al. (2006) suggested, we select cluster stars from candidates using color-magnitude diagrams.

Along the main-sequence in the CMD, stars around the center of a cluster within a proper radius are likely to be members. Figure 4 shows the radial number density distribution of stars around Palomar 5. We can see that the density decreases from the center to the external part of the cluster. As $R > 0.25°$, the average number density becomes about $1.0958 \pm 0.0497 \times 10^4$ deg^{-2}. A radius $R = 0.13°$, where the density is a little higher than the average to avoid excessive field stars, is used to select candidates of cluster stars. About 1523 cluster member candidates are reserved.

Fig. 4 Radial number density distribution around the center of Palomar 5. The solid line shows that the density declines as the radius increases. The dashed line gives a threshold of $R = 0.13°$. Stars within the threshold radius are considered as cluster member candidates.

The next step is to pick out the most probable cluster stars from the 1523 candidates. Figure 5 demonstrates the process of selecting cluster stars as a part of the training and test data set. Objects are reserved by encircling the CMD of r vs. $r - i$ (Fig. 5a) and g vs. $g - r$ (Fig. 5b) with proper enclosures. These objects may be main-sequence stars, red giants and blue horizontal branch stars belonging to Palomar 5. In this way, contaminations from remanent galaxies and field stars (crosses in both CMDs) are almost eliminated. From using the intersection of these two CMDs, 957 objects are pared down to form the distribution of cluster stars in Figure 5d (bigger black dots).

For field-star selection, $R > 1°$, which is far enough from the center of Palomar 5, is adopted to choose candidates of field stars. Because the number of candidates is tremendously high, it is impossible to preserve all of them as our data set. Ideally, we hope that roughly equal numbers of objects should lie at both sides of the boundary of data space so that they can fully represent these two classes. Accordingly, we choose field stars randomly in the sky, satisfying roughly equal numbers of both field and cluster stars near the main-sequence turnoff in the CMD. Thus, a box (a white rectangle with four hollow circles at the vertexes in Fig. 5d) with $19.5 \leq r \leq 21$ and $0.05 \leq r - i \leq 0.18$ is designed to form a region encircling the turnoff. There are about 320 stars in the box of the CMD for the cluster and field, respectively. As a result, 6991 field stars (smaller dots in Fig. 5c) and 957 cluster stars constitute the whole train-test data set (Fig. 5d). As mentioned previously, these field stars with target value 0 and cluster stars with value 1 are transmitted to train a BPNN. The BPNN will configure itself to estimate the cluster membership probability for each object. Although the remaining data set may contain impurities more or less, they are negligible relative to the total group sizes. Also, at the same time, BPNN will get rid of them automatically through being trained again and again, which is one of the reasons why we consider BPNN as our method.

4.2 Network Structure Determination

In order to normalize the input data, we subtract the mean magnitudes of all data and then divide them by their standard deviations. Furthermore, for the sake of determining the number of layers and neurons in each layer, the star sample is segmented into training and test data sets. The training set is used to train a BPNN, while the test set is used to measure its performance. Since cluster stars are relatively

Fig. 5 Color-magnitude diagrams used to select cluster and field stars. a) The CMD of r vs. $r - i$. Bold dots are the possible cluster members and crossings denote the excluded non-cluster sources. b) The CMD of g vs. $g - r$, used to assist r vs. $r - i$ CMD to pick out cluster stars from candidates. c) The distribution of the selected field stars in r vs. $r - i$ CMD. d) r vs. $r - i$ CMD, including both the extracted cluster members and field stars. The smaller dots are field stars and the larger dots are cluster members. The white box with four circles shows the enclosure that decides the number of field star samples.

scarce, all of them are placed into the training set. Half the field stars are laid aside stochastically into the training data set and the others are regarded as the test set.

Ten experiments are implemented for each designed BPNN to calculate the average output as the cluster membership probability. In this way, some random influences from initial weights, the direction of modifying weights and the algorithm terminating conditions are weakened. Here, one hidden layer and two hidden layers with 1, 10, 20, 30 and 40 neurons in the specified networks are investigated. Their performance is shown in Figure 6. Each network is trained and tested ten times with the data set gained in the previous section. We stop the algorithm each time when the test MSE reaches its minimum. As Figure 6 indicates, with the increasing complexity of network configuration, the training and test MSEs decrease as a whole. Also, in the same network group (for example Layer $1 = 10$), the test MSE always decreases at first and then increases when the number of neurons in the second hidden layer increases. No larger difference is observed among the BPNNs with 2 hidden layers when the number of neurons in the first hidden layer is larger than 10. Considering the test performance and the complexity of the configuration relevant to the speed of training, Net[5:10,10,1], which has 5 input elements, 10 neurons in the first hidden layer, 10 in the second hidden layer and 1 output, serves as our final model.

Fig. 6 Mean training and test MSEs of various networks including 1 or 2 hidden layers. The darker solid line shows the training MSEs, while the light dotted line shows the test MSEs. The dashed horizontal line (MSE = 0.055) presents one level so that it can help us to check the network with the best performance.

4.3 Detecting Tidal Tails with the Best Trained Network

It is necessary to take into account the effectiveness of the trained networks. So, we construct a function of the magnitude (r_{psf}) as the ratio of the mean output for field and cluster stars in the training and test data sets. Ideally, the ratio should be around zero as long as the networks are trained well. In Figure 7, the ratio goes up gradually when r_{psf} becomes fainter. No significant deviation is shown above 0.1 near the magnitude limit. This indicates that the selection of the training and test data sets and BPNNs are determined relatively well. However, in order to trace out the profile of the tidal tails more evidently, we cut off all the sources with a truncation $r_{psf} = 21$ with smaller photometric errors.

All the sources with normalized $ugriz$ magnitudes are imported into the ten trained networks, and the mean cluster membership probability for each star from the output layer is calculated. In order to present a panorama of the distribution of cluster stars, we divide the whole field into small square bins, whose sizes are $6' \times 6'$. About 90 objects are included in each bin, which is enough for statistical analysis. Here, the mean probability in each bin is computed. At the same time, Gaussian smoothing and median filtering are used to get rid of noise when detecting the tails in the field. In addition, in order to enhance the resolution of the distribution, cubic spline interpolation is also employed.

A lot of experiments are implemented to investigate the factors impacting the distribution. These tests include the parameters used in the smoothing tools, the selection of cluster stars and the field stars chosen in different regions of the field. As a result, cluster stars within the radius $R < 0.13°$ are appropriate for training the networks. Larger R will bring in pollution from field stars, while smaller R values yield fewer cluster stars which cannot provide enough information about Palomar 5 members. Thus, the selection criterion of cluster stellar candidates in Section 4.1 is reasonable. Field stellar samples chosen from different regions make no difference unless they cover the possible position of tidal tails. However, this does not affect the detection of the remaining parts.

Figure 8 exhibits the contours of the 2-dimensional probability distribution, where four regions are processed by the same best trained network and smoothing techniques. Although smoothed, the

Fig. 7 Relationship between the ratio of mean outputs for field and cluster stars and the r_{psf} magnitude.

distribution is fluctuant in the whole field. We find that $P(C)s$, the cluster membership probabilities in all bins, obey a Gaussian distribution with mean value 0.0488 and standard deviation $\sigma = 0.0087$. In order to check the tails more widely and to find possible longer tails, the contour levels are larger than 1σ above the mean. In this figure, M5 is detected at the same time, due to the similarity of the color-magnitude diagrams. The black solid line traveling through R2 and R3 provides us with a possible tidal tail away from the main tails in R0 ($226° < \alpha < 236°$ and $-3° < \delta < 5°$). We are not sure that the tail along this line is a real extension of the tails of Palomar 5, because the foreground noise in the residual regions is so similar to it. Figure 9 shows contours of the probability distribution, where the contour levels are higher than 1.5σ above the mean. The possible extension vanishes, but some debris still coexists with foreground noise. However, Grillmair & Dionatos (2006) insist that the extension is the real tail. Therefore, only the region R0, which encloses the clear tidal tails of Palomar 5, is considered as our region of investigation in the next discussion. The sub-figure in Figure 9 shows a smoothed probability distribution with 1.5σ in R0. There are some differences between the two plots which can be ignored. These petty changes are caused by different bin coordinates and by region smoothing.

5 PROPERTIES OF THE TIDAL TAILS AND THE CLUSTER

5.1 The Profiles of the Tails

The subplot in Figure 9 presents the holistic smoothed distribution of Palomar 5 members. The whole tail lies away from the dense regions of Galactic extinction, which implies that the extinction has little effect on our detection. Two tails extend from the core region of Palomar 5 towards the southwest and northeast directions, which we call the South Tail (ST) and North Tail (NT), respectively. ST is the leading part facing toward the Galactic disk, while NT is the trailing part. The angular distances are $5.42°$ for NT and $3.77°$ for ST. NT is a little shorter than the $5.8°$ northern tail detected by Odenkirchen et al. (2003), while the lengths of both STs are approximately equal. One reason for this may be that the smoothing flattens the distribution. In addition, the longer debris along the north tail detected by Odenkirchen et al. (2003) may be the disturbance or noise of field stars. However, the main tails we find are very similar to the results of Odenkirchen et al. (2003), while we do not support the discovery of any longer tails to the north of the sky that Grillmair & Dionatos (2006) reported to be a 22 ° tail of Palomar 5, because the more extensive debris is so similar to background noise. There are no photometric data in the south, so we cannot cover larger southern regions to check for an even longer ST.

Fig. 8 Contours of the smoothed probability distribution of cluster membership with 1σ higher than the mean. The solid line shows the possible extension, and dashed rectangle R0 is the target region we will discuss. The overlapping gray background is the reddening distribution of the Galaxy.

Fig. 9 Contours of the smoothed probability distribution of cluster members with 1.5σ higher than the mean. The upper-right plot shows the recomputed distribution in the region R0. Refer to the text for detailed specifications.

Another fact we can see from the contour map in Figure 9 is that the distribution of cluster members is more dense in NT than that in ST. The possible reason causing this situation may be that the orbital direction of NT is nearer to us than ST; given that the components of both tails are the same, because of their large extension, the average magnitude of NT ought to be brighter than that of ST. So, due to the detection limit, quite a few cluster members in ST cannot be observed or excluded by data preprocessing, although they are detected. However, when investigating all the detected members in both NT and ST, we do not find any large magnitude shift by comparing the mean magnitudes of the two tails in any

places having the same distance from the cluster center. The maximum magnitude difference (between the northeast and southwest ends) is about 0.1 mag (NT is a little brighter than ST). Qualitative analysis indicates that even if the inclination angle of the tails is large, the magnitude difference is small when considering a specified star lying in a different position in the tails. The above fact tells that it may be true that the track of NT is nearer to us, but this cannot obviously change the distinct density difference of both tails. Therefore, this kind of situation should refer to some dynamical processes, which will be studied in the future.

We convert the probability distribution to a surface density distribution. First, we count the numbers of stars in all square bins. Then, the areas of all bins are calculated. In this way, with the smoothed probabilities of cluster members, we get a smoothed surface density distribution. Figure 10a gives the transformed surface density contour map which is interpolated by the technique of cubic spline interpolation. Figure 10a indicates that there are no obvious changes compared with the probability distribution contours. Furthermore, the noise around the tails and M5 is removed from the map. From Figure 10a, we can see that there is no geometrical symmetry between ST and NT, and it assumes an S shape near the center of Palomar 5. In fact, Dehnen et al. (2004) have illustrated this structure in their simulations. Figure 10b shows the radial surface density profiles of both ST and NT. In Figure 10b, the numerical values originate from the smoothed density. Thus, the density is lower than the density profile along the tails discussed by Odenkirchen et al. (2003). Both density profiles decrease quickly from the center of Palomar 5. However, it seems to be that the density of NT does not decrease as quickly as that of ST. In addition, the trailing tail seems to lag behind the leading tail, because the density of ST reaches its peak at $R = 1.7°$, while NT achieves its minimum. This phenomenon can be slightly seen in Odenkirchen et al. (2003) during the discussion of the radial profiles of both tails, although it was not mentioned. We cannot yet explain what causes this phenomena, but it is an interesting result. Possibly, this kind of lag is closely related to the evolution of the cluster and the interaction between Palomar 5 and the Galaxy. Moreover, several stellar clumps emerge in both tails to form substructures of the cluster. Some of them lie at around $R = 0.90°$ (229.62°, 0.56°), 2.57° (230.94°, 1.6°), 3.28° (231.56°, 19.6°) and 5.13° (233.28°, 2.76°) in the north, and 1.72° (227.8°, −1.32°) and 3.50° (226.5°, −2.54°) in the south. They may be these kinds of substructures and radial density profiles that are formed by dynamical processes between Palomar 5 and the Galaxy. One possible explanation may be as follows: Palomar 5 has experienced several encounters with the Galactic disk or bulge since its birth; its body was heated by dynamical shocks; then the cluster stars were accelerated, and gradually disrupted and extended along the moving direction; as a result, some stars with small mass might escape from the tails, and the residual ones constitute the substructures and wait for the next shock. In fact, Odenkirchen et al. (2003) have indicated that Palomar 5 will be totally destroyed after the next disk crossing within about 100 Myr.

5.2 King Model Fitting and the Luminosity Function

For the density profile near the center, due to the small size of Palomar 5, we take $R < 50'$ as the range for discussion. Cluster stars are counted in all bins (from $2'$ to $6.8'$, the bin size is $0.3'$; from $8'$ to $15'$, the bin size is $1'$; the bin size of the rest is $5'$), and the area of each annulus is calculated. Figure 11 shows the radial surface density profile, and the best fitting of the King model (King 1962). The King model is expressed as:

$$\rho = k \left\{ \frac{1}{[1 + (R/r_c)^2]^{\frac{1}{2}}} - \frac{1}{[1 + (r_t/r_c)^2]^{\frac{1}{2}}} \right\}^2, \tag{1}$$

where k is a constant, r_c is the core radius, r_t is the tidal radius and R is the distance away from the center of the cluster. Thus, the centric density is

$$\rho_0 = k \left\{ 1 - \frac{1}{[1 + (r_t/r_c)^2]^{\frac{1}{2}}} \right\}^2. \tag{2}$$

Fig. 10 (a) The surface density contour map of Palomar 5, converted from the smoothed probability distribution map. Here, M5 and noise around the tails are erased. (b) The radial profiles of surface density ($\deg^{-2}$) for both tails. The interval of the radial distance is $0.05°$.

Fig. 11 Radial surface distribution near the center of Palomar 5. The circles represent the surface densities in all bins calculated by counting the cluster stars. The King model with the mean background number density (0.11) is shown by the solid line. The coordinate system is a logarithmic system and the unit of density is arcmin^{-2}.

$c = r_t/r_c$ is the concentration ratio. We fit the data subtracted from the background density (about 0.11) and obtain $k = 114.9$, $r_c = 1.60'$ and $r_t = 14.29'$. Therefore, the concentration ratio c is 8.95 and ρ_0 is 90.81 arcmin^{-2}. We find that the radii obtained in this paper are smaller than those of Harris (1996): $r_c = 3.25'$ and $r_t = 16.28'$. From Figure 11, we can clearly see that the King model cannot fit the observed profile at the outer regions ($R > 8'$) because of the tidal tails.

Fig. 12 (a) The boundaries used to calculate the LFs. C is the boundary of the cluster itself with $R < 10'$; N is the boundary of the north tail; S1+S2 is the boundary of the south tail. The boundaries of both NT and ST are far away from the center with $R > 0.5°$. (b) Different luminosity functions. All LFs are rescaled by some factors to match the LF of the cluster.

The luminosity function (LF) of the cluster reflects its mass distribution. We examine whether the luminosity functions of the cluster itself and the tidal tails are the same, and try to verify that the tidal tails come from Palomar 5. We examine the luminosity functions in three regions: the north tail (N), the south tails (S1 and S2) and Palomar 5 (C). Figure 12a shows the boundaries of these regions. Radius $R = 10'$ is adopted for Palomar 5, and is near enough to the center of the cluster to eliminate contamination from the tails. Radius $R > 0.5°$, far enough away from the center, separates the tails into N for the north tail and S1 and S2 for the south tails. Star counts are taken to deduce the LFs of the ST and NT in these regions. We only consider stars with a high cluster membership probability (> 0.5), which subsequently subtracts the contamination of foreground field stars. Consequently, in Figure 12b, there are four LFs of ST, NT, total tails and the cluster itself, and all of them are rescaled to match the LF of the cluster. In the magnitude range of $19.0 \leq r_{\rm psf} \leq 20.0$, the LFs are rescaled by factors of 19.41, 22.43 and 20.35 for the NT, ST and total tails, respectively. On the whole, there is little difference among the LFs when $r_{\rm psf} < 20.5$. As $r_{\rm psf} \geq 21$, the LF of the cluster lies lower than those of the tails. This case confirms the so-called 'mass segregation effect' (Koch et al. 2004), which shows that cluster members with big masses will accumulate near the center because of the loss of kinetic energy when colliding with others, while stars with small masses would escape from the cluster into the tails. Thus, these luminosity functions reveal the fact that the stars in the tails come from Palomar 5, and some relevant physical properties have not changed much in its history.

6 CONCLUSIONS

In this paper, we present a new method, Back Propagation Neural Network, to detect the tidal tails of globular cluster Palomar 5. Although some approaches, such as the matched-filter method (Rockosi et al. 2002), are widely applied in identifying the tails, we choose BPNN as our model to find the exact and distinct tails of Palomar 5. The photometric magnitudes of five bands ($ugriz$) in the SDSS DR6 are the unique inputs and consequently the cluster membership probabilities are the output in a well trained BPNN. BPNN resembles a black box, and we need not consider its detailed inner structure. The only thing we should do is to give it a set of well-selected cluster and field stars (they may not be completely accurate) as a teacher to make BPNN learn the knowledge. After being trained, BPNN can estimate the the cluster membership probabilities.

First of all, we obtain about $15\,305\,060$ objects in a $40 \times 20\,\mathrm{deg}^2$ field ($220° \leq \alpha \leq 260°$ and $-5° \leq \delta \leq 15°$). Considering the effectiveness of star/galaxy classification in the SDSS and the completeness of the observation, we reserve about $4\,082\,662$ point sources (stars) with $14 \leq r_{\mathrm{psf}} \leq 22$ after reddening correction and eliminating the pollution from galaxies as much as possible with the help of the distribution map of $r_{\mathrm{psf}} - r_{\mathrm{exp}}$. Next, we make use of surface density and CMDs to extract cluster stars from candidates, which lie in the circle where $R < 0.13°$. Also, field stars used for training are chosen by making equal numbers of both cluster and field stars around the turnoff of the main-sequence of Palomar 5. In this way, about 960 cluster stars and 6800 field stars are kept aside as training and test data sets to provide their inherent characteristic information for BPNN. With the training and test data, the best parameters and structures of BPNN are determined. Then, the best trained BPNN, with 5 nodes in the input layer, 10 neurons in the first hidden layer, 10 neurons in the second hidden layer and 1 neuron as the output, is gained to compute the cluster member probability estimate for each point source. We divide the field into bins with size $6' \times 6'$ to calculate the mean probability distribution. The impact of the selection of field stars is also investigated and their effect is not important.

S-shaped tidal tails are detected, which subtend towards the northeast and southwest from the center of the cluster: the trailing tail and the leading tail, respectively. The angular distances are $5.42°$ for the north tail and $3.77°$ for the south one. At the same time, there are some density clumps which are substructures of Palomar 5 in both tails. We cannot find any longer stretch for the NT if we do not regard the extension far away as tails, and we cannot confirm whether the ST has a longer spread because no photometric data are available outside towards the southwest. We also find an interesting phenomenon from the radial profile of the density: the NT seems to lag behind the ST like wave propagation, which may be caused by tidal shocks when Palomar 5 crossed the Galactic disk or bulge during its history. In addition, we fit the radial density profile near the cluster center with the King model and find that the model can fit this kind of remote globular cluster with low density very well when the radial distance is less than $8'$. However, when the radial distance becomes larger, the density drops more slowly due to the tidal tails. The tidal radius obtained in this paper is $14.29'$ and the core radius is $1.68'$. Luminosity functions of both tails and the cluster are also determined. We find that there is little difference between the LFs of both tails coming from the original cluster, and their properties have not obviously changed during their lives.

Appendix A: THE MECHANISM OF BPNN

For clarity and continuity of our work, a BPNN with two hidden layers is demonstrated below (Fig. A.1). This network generally contains an input layer (I in the figure), hidden layers ($L1$ and $L2$) and an output layer (O). The input layer reads training or test patterns (input patterns), which are offered to be processed by hidden layers and the output layer yields the relevant output results. In our paper, input patterns correspond to 5-band magnitudes of cluster and field stars. The desired output patterns (target patterns) placed at T in Figure A.1 are 1 for cluster stars or 0 for field ones. The output of BPNN gives the cluster membership probability for each star.

The network is executed in two phases: a training phase and a test phase. In the training phase, input and target patterns are submitted to the network. Two processes, feeding forward and error back propagation, are then performed. After we endow this network with an initial state, an input sample (pattern) travels from the input layer. By being treated by intermediate layers, the information stored by weights is processed and a corresponding result emerges at the output layer. There, comparing the network output result with the relevant target pattern, an error performance is calculated. By using this error estimate, error back propagation is carried out from the output to the input layer to modify the connected weights and biases (in Fig. A.1), and at the same time store learned knowledge. Then, the remaining patterns act in the same way and the iteration continues until satisfying the preplanned error limit. There are two modes of updating weights and biases: one is the incremental mode in which the weights and biases update when the errors of patterns back-propagate one by one as presented above and the other is batch mode in which all patterns travel through the network and the total error is counted, then the weights and biases are renewed once in an iteration. We call an iteration of processing all patterns as one 'epoch'. In the test phase, patterns, which have not been seen by the network, are given to check the efficiency and accuracy of the network configured in the training phase.

A detailed description of the network configuration and concise mathematics of training it are presented below. In the left panel of Figure A.1, the input layer and target segment are divided by dashed lines, which are linked to an exoteric environment. The nodes in the hidden and output layers are called neurons. The right panel of Figure A.1 gives the delicate structure of one neuron. There are p neurons from the previous layer that act as inputs connecting the enclosed neuron by a dashed rectangle, where each connection has a weight w. In the neuron, an adder $\sum$ performs the sum of all the input values and bias θ and transmits the result v to a transfer function f, which is to produce an output y. Expressions can be presented as

$$v = \sum_{i=1}^{p} w_i x_i - \theta = \boldsymbol{w}' \boldsymbol{x} - \theta, \tag{A.1}$$

$$y = f(v), \tag{A.2}$$

where x_i is the output of the ith node from the previous layer and w_i is the corresponding weight, and $\boldsymbol{x} = [x_1, x_2, \ldots, x_p]'$, $\boldsymbol{w} = [w_1, w_2, \ldots, w_p]'$. To express this more vividly, the stimulus v goes beyond the bias (θ), and the neuron will be activated to release an output signal to the next neurons. Here, θ can be arranged into the weight vector $\boldsymbol{w}$, as long as we take into account another constant input of the node. That is to say, we introduce $x_0 = -1$ and $w_0 = \theta$ and let the network adjust θ just like weights, so that v has the form of $v = \boldsymbol{w}' \boldsymbol{x}$, where $\boldsymbol{x} = [x_0, x_1, \ldots, x_p]'$, and $\boldsymbol{w} = [w_0, w_1, \ldots, w_p]'$. Besides, the transfer function f has various forms, such as:

$$\text{Linear}: \ f = v, \tag{A.3}$$

$$\text{Log} - \text{sigmoid}: \ f = \frac{1}{1 + e^{-v}}, \tag{A.4}$$

$$\text{Tan} - \text{sigmoid}: \ f = \frac{2}{1 + e^{-2v}} - 1, \tag{A.5}$$

where $-\infty < v < +\infty$. Among these functions, the log-sigmoid transfer function, which yields results in the range from 0 to 1, is commonly used in back-propagation networks partly because of its unlimited differentiability. We will adopt this kind of transfer function (Eq. (A.4)) in our present study.

Now, turn back to training the network. We will select the batch training mode in this paper. In this mode, weights and bias will be updated only after the entire inputs and targets are submitted, and as a result, the gradients (quantitative changes of weights and bias) are averaged together to produce more accurate estimates. In this case, a performance function, known as the mean squared error, is used to evaluate the outputs of the network. The MSE is expressed as a formula

$$E(\boldsymbol{w}) = \frac{1}{N} \sum_{i=1}^{h} (T_i - O_i)^2 = \frac{1}{N} (\boldsymbol{T} - \boldsymbol{O})'(\boldsymbol{T} - \boldsymbol{O}) = \frac{1}{N} \|\boldsymbol{T} - \boldsymbol{O}\|^2, \tag{A.6}$$

Fig. A.1 *Top*: A multilayer feed-forward back propagation network with 2 hidden layers. *Bottom*: A detailed model of one neuron in BPNN.

where N is the number of input-target pairs, h is the dimension of output vector $\boldsymbol{O}$, $\boldsymbol{T}$ is the target pattern vector, O_i and T_i are the components of $\boldsymbol{O}$ and $\boldsymbol{T}$, and $\boldsymbol{w}$ contains all the weights and biases in the BPNN. Additionally, target patterns are 1 or 0. Only one output neuron is needed, so $h = 1$ in this paper.

Given the performance function, a training algorithm should be provided to teach the BPNN to learn how to classify. During training, our aim is to make the outputs approach the target patterns as much as possible, resulting in decreasing the value of MSE. That is, in order to adjust weights for better learning, we need to decrease the value of MSE or to minimize this performance function epoch by epoch. Consequently, a universal training scheme to update $\boldsymbol{w}$ comes up as

$$\Delta \boldsymbol{w}^{(k)} = \boldsymbol{w}^{(k+1)} - \boldsymbol{w}^{(k)} = \eta^{(k)} \boldsymbol{d}^{(k)} \tag{A.7}$$

or

$$\boldsymbol{w}^{(k+1)} = \boldsymbol{w}^{(k)} + \eta^{(k)} \boldsymbol{d}^{(k)}, \tag{A.8}$$

where k denotes the k_{th} epoch of training, the positive η is the learning rate which decides the step length of changes of $\boldsymbol{w}$, and $\boldsymbol{d}$ is the search direction where $\boldsymbol{w}$ moves. All the training algorithms of the back propagation network are variations of the above form. For example, the most basic algorithm, Steepest Descent BP (SDBP), is based on the negative gradient of $E(\boldsymbol{w})$ as $\boldsymbol{d}$.

Thus, the learning rule becomes

$$\Delta \boldsymbol{w}^{(k)} = -\eta^{(k)} \nabla E(\boldsymbol{w}) \mid_{\boldsymbol{w}=\boldsymbol{w}^{(k)}}, \tag{A.9}$$

where $\nabla E(\boldsymbol{w})$ is the differential of E. In this training algorithm, the detailed modifying formula of $\boldsymbol{w}$ in each layer is obtained by the chain rule (Hangan et al. 1996; Haykin 1998, sect. 11.9), and the relative learning rate can be optimized (Hangan et al. 1996, sect. 9.6 and 12.12). There are other training algorithms to train BPNN: Back-propagation with Momentum (MOBP) (Hangan et al. 1996, sect. 12.9), Conjugate Gradient Back-propagation (CGBP) (Hangan et al. 1996, sect. 9.15 and 12.15), Newton Method using the Hessian matrix (second derivatives) of the performance as direction (Hangan et al. 1996, sect. 9.10), Levenberg-Marquardt Back-progation (LMBP) (Hangan et al. 1996, sect. 12.19) and so on. Here, we chose LMBP as our training algorithm because it has the speediest convergence. Detailed algorithm information about LMBP can also be found in Ball et al. (2004).

References

Abazajian, K., et al. 2004, AJ, 128, 502

Adelman-McCarthy, J. K., et al. 2008, ApJS, 175, 297

Bailer-Jones, C. A. L. 2000, A&A, 357, 197

Ball, N. M., Loveday, J., Fukugita, M., Nakamura, O., Okamura, S., Brinkmann, J., & Brunner, R. J. 2004, MNRAS, 348, 1038

Belokurov, V., Evans, N. W., Irwin, M. J., Hewett, P. C., & Wilkinson, M. I. 2006, ApJ, 637, L29

Bowman, A. W., & Azzalini, A. 1997, Applied Smoothing Techniques for Data Analysis (New York: Oxford University Press)

Folkes, S. R., Lahav, O., & Maddox, S. J. 1996, MNRAS, 283, 651

Dehnen, W., Odenkirchen, M., Grebel, E. K., & Rix, H.-W. 2004, AJ, 127, 2753

Grillmair, C. J., Freeman, K. C., Irwin, M., & Quinn, P. J. 1995, AJ, 109, 2553

Grillmair, C. J., & Johnson, R. 2006, ApJ, 639, L17

Grillmair, C. J., & Dionatos, O. 2006, ApJ, 641, L37

Hangan, M. T., Demuth, H. B., & Beale, M. H. 1996, Neural Network Design (Boston, MA: PWS Publishing)

Harris, W. E. 1996, AJ, 112, 1487

Haykin, S. 1998, Neural Networks: A Comprehensive Foundation (2nd ed.; Harlow, En: Prentice Hall)

King, I. 1962, AJ, 67, 471

Koch, A., Grebel, E. K., Odenkirchen, M., Martínez-Delgado, D., & Caldwell, J. A. R. 2004, AJ, 128, 2274

Lauchner, A., Powell, W. L. J., & Wilhelm, R. 2006, ApJ, 651, L33

Leon, S., Meylan, G., & Combes, F. 2000, A&A, 359, 907

Lupton, R., Gunn, J. E., Ivezić, Z., Knapp, G. R., & Kent, S. 2001, Astronomical Data Analysis Software and Systems X, 238, 269

Naim, A., et al. 1995, MNRAS, 274, 1107

Odewahn, S. C., Stockwell, E. B., Pennington, R. L., Humphreys, R. M., & Zumach, W. A. 1992, AJ, 103, 318

Odenkirche, M., et al. 2001, ApJ, 548, L165

Odenkirche, M., et al. 2003, AJ, 126, 2385

Rockosi, C. M., et al. 2002, AJ, 124, 349

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Sergios, T., & Konstantinos, K. 2006, Pattern Recognition (3rd ed.; Athens: Academic Press)

Spitzer, L. 1987, Dynamical Evolution of Globular Clusters (Princeton, NJ: Princeton University Press)

Stoughton, C., et al. 2002, AJ, 123, 485

von Hippel, T., Storrie-Lombardi, L. J., Storrie-Lombardi, M. C., & Irwin, M. J. 1994, MNRAS, 269, 97

Wu, Z.-Y., Shu, C.-G., & Chen, W.-P. 2003, Chin. Phys. Lett., 20, 1648

Wu, Z.-Y., Zhou, X., & Ma, J. 2004, Chin. Phys. Lett., 21, 418