

# Planetary nebulae in DR2 J-PLUS

## 1 Introduction

### 1.1 Antecedents

In (2020A&A...633A.123G) is presented a number of colour-colour diagrams to discriminate PN from their contaminants of the most efficient way using the novel surveys J-PLUS and S-PLUS. The main disadvantage of this methodology is the great effort and time that is required to find the best combination of colour to do that. After that it is necessary to establish the selection criteria. Another disadvantage consists of that on this method does not use all the filters of the surveys. Note that to create the colour criterion in this works only were used the filters: ..... For this reason, even if with the colour criteria the selection of the sources is done of an automatic way, in this new era of multi-band surveys it is necessary to used techniques even more automatic that use all the bands to guarantee a better identification of our targets.I mean to avoid misclassification.

## 2 Methodology

### 2.1 Dimension reduction algorithm: PCA

The strength of PCA for data analysis comes from its efficient computational mechanism, the fact that it is well understood, and from its general applicability (J. Novakovic, S. Rankov, Vol. VI (2011), No. 2 (June), pp. 317-327). This is a good paper comparing the performance of the classifier methods with and without using PCA.

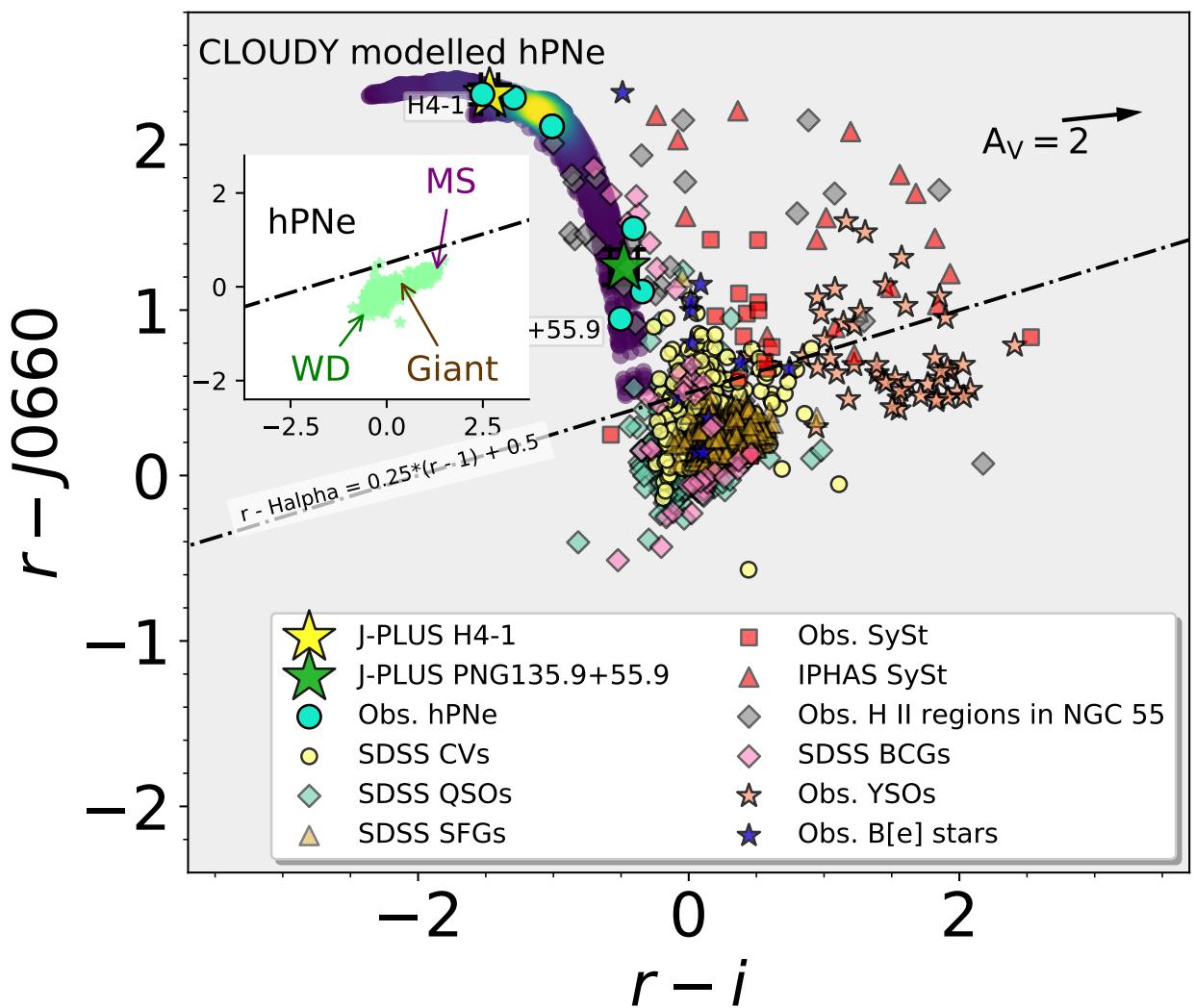
This is especially effective for classification algorithms that do not have any inherent feature selections or feature extraction builds in, such as the nearest neighbour methods or some types of neural networks.

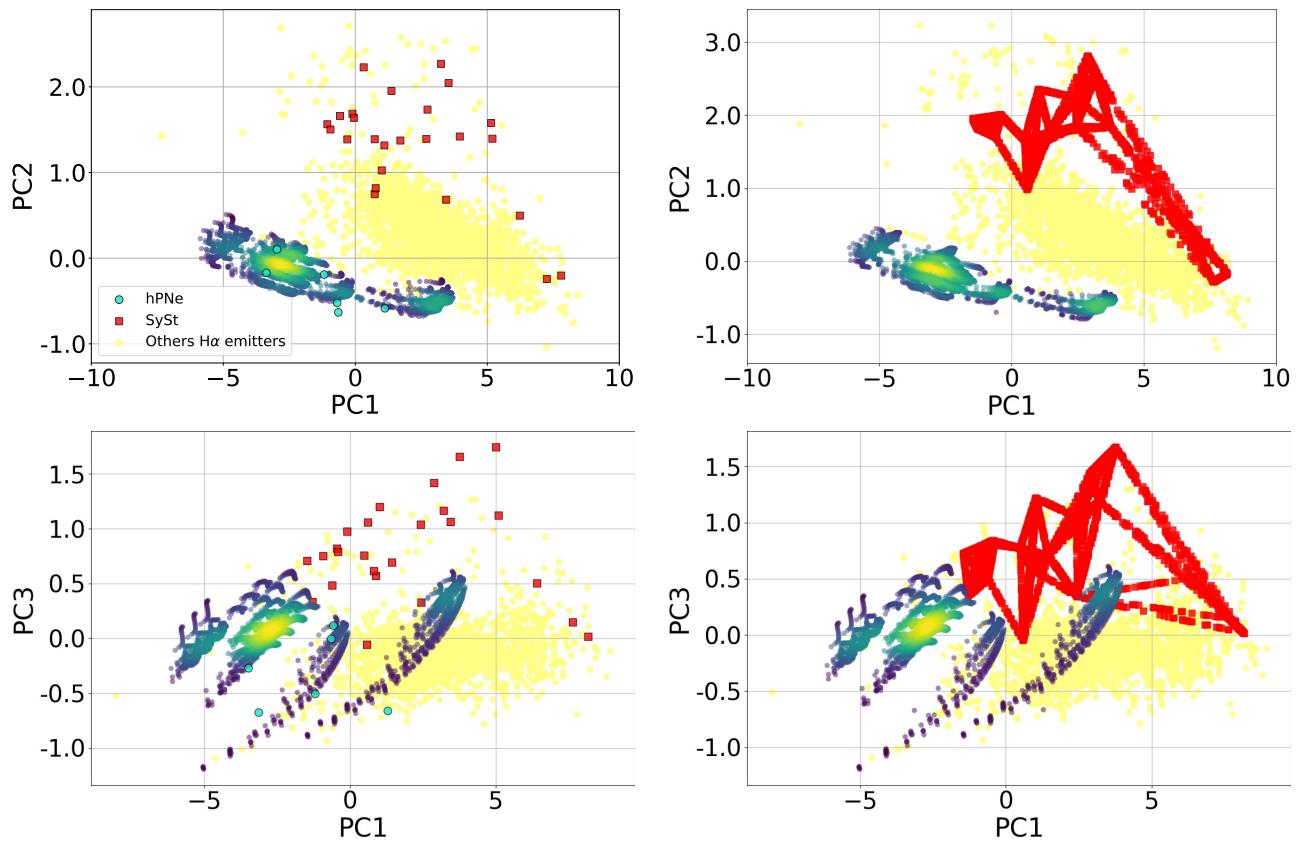
Maybe it will be good idea to combine PCA with the Bayes algorithm classifier.

### 2.2 Classifier algorithms: logistic regression

### 2.3 Synthetic Oversampling: SMOTE

The class distribution in many applications on classifier algorithms is not balanced. Our sample used here is highly imbalanced 3,700 PNe (including modeled and observed), 25 SySt, 25 H II regions and approximately 2,385 all else. This means that SySt accounts for 0.4% of all sample. The straightforward application of classification algorithms may lead to misleading results because of the preponderance of the PN class. A Bayesian classifier will have biased priors that favor the majority class in our case the PN sample represents the majority class.





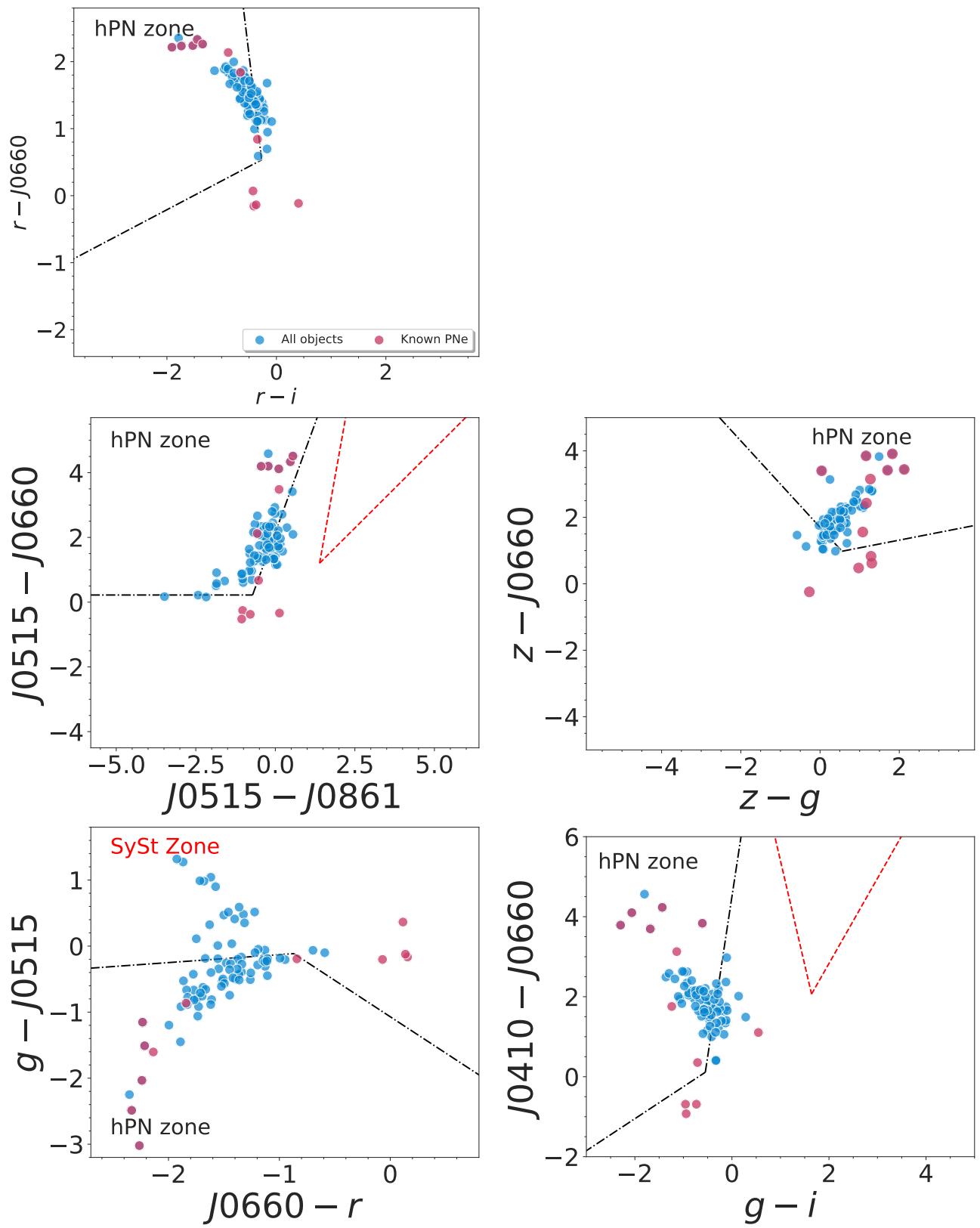
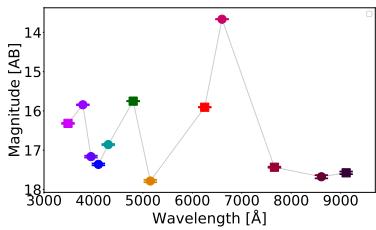
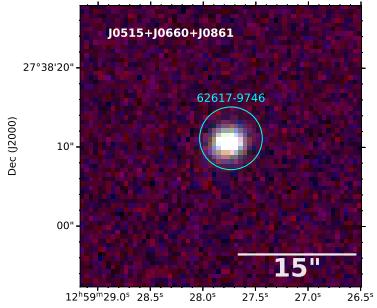
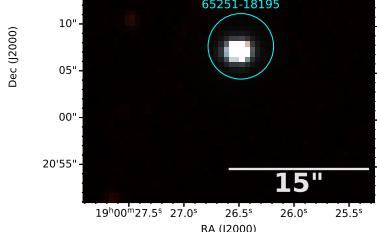
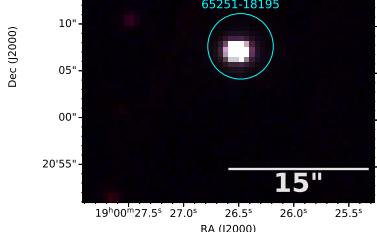
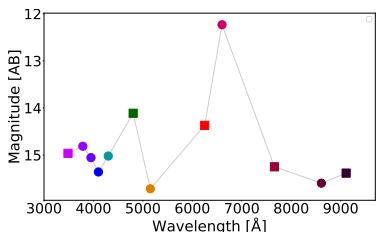
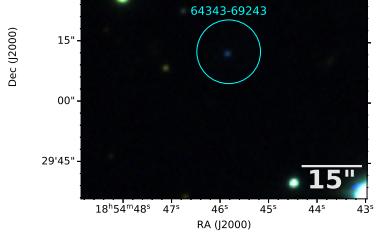
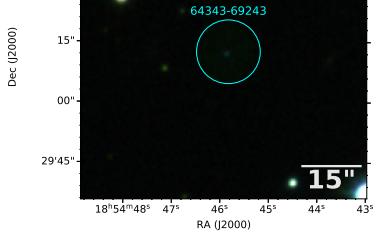
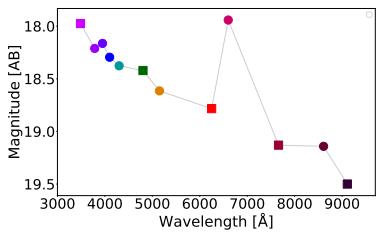
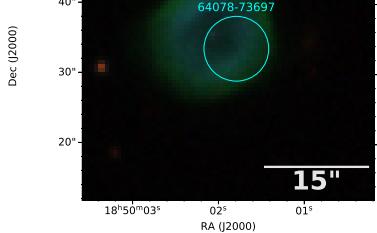
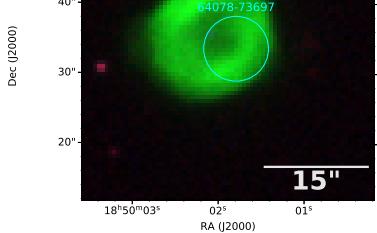
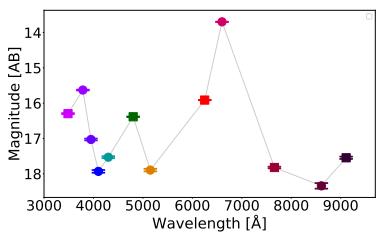
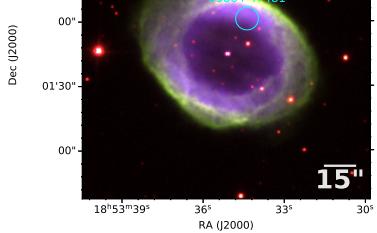
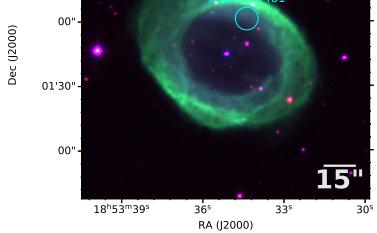
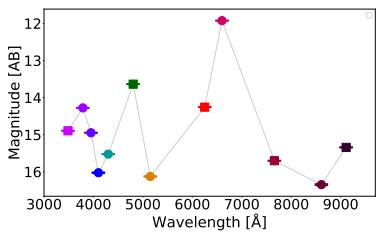
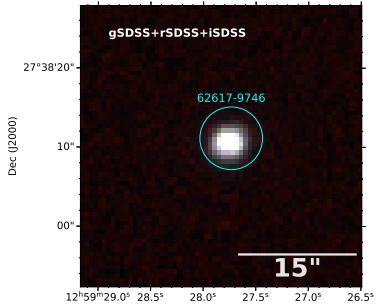
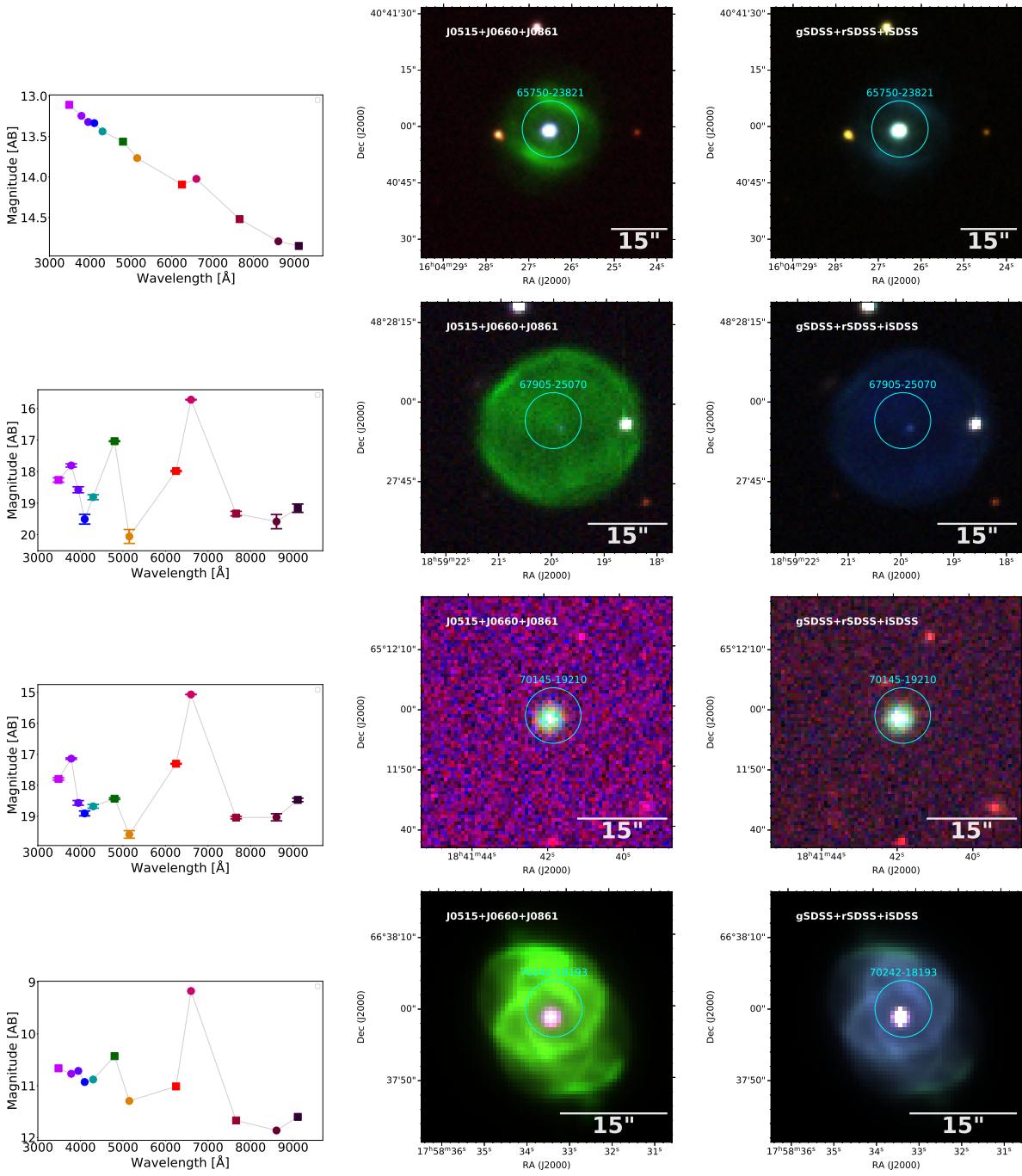


Figure 1:

**Aper (6 ")****J0515+J0660+J0861****gSDSS+rSDSS+iSDSS**



## 2.4 J-spectra and images of the selected objects

