

中華民國第 64 屆中小學科學展覽會
作品說明書

高級中等學校組 物理與天文學科

佳作

051803

運用主成分分析探尋宇宙中的互繞雙類星體

學校名稱：臺北市立第一女子高級中學

| | |
|---------------------------------|------------------|
| 作者： 高二 林睦潔 高二 李榛昀 | 指導老師： 張清俊 |
|---------------------------------|------------------|

關鍵詞：雙類星體、主成分分析、光譜

摘要

探尋互繞雙類星體是天文研究中重要的課題，互繞雙類星體是星系合併的產物，也是產生重力波的來源之一，研究其性質能使我們了解星系合併的機制，而光譜是最直接研究的面向。雖然目前已經有大量數據，但相關的研究並不多，雙類星體的性質研究也尚不成熟。在本研究中，我們發展出能從類星體光譜數據中尋找雙類星體的方法，並進一步探究其性質。本研究運用機器學習中的主成分分析（Principal Component Analysis, PCA）擷取雙類星體的特徵光譜，結合 PCA 係數分析、最鄰近搜索和支援向量機（support vector machine），精準的篩選雙類星體。我們將此方法運用在 Sloan Digital Sky Surveys 所提供的類星體光譜資料庫，找出並匯集約 500 個新發現的候選雙類星體，並使用此數據進一步探索目前未被發現的雙類星體物理性質。

壹、前言

一、研究動機

雙類星體（Binary Quasars）尋找和研究是相對新的領域，在天文物理上可以提供關於宇宙的諸多資訊，例如關於星系形成與演化、重力波等。不過目前對於雙類星體的研究不多，也鮮少使用系統性數據分析的方式了解雙類星體在整個宇宙中的數量和分布，因此難以對其做出更完整的性質分析。

為了更了解雙類星體，我們讓電腦自動化從大量類星體光譜資料中，辨識出雙類星體，而主成分分析具有降維多變量數據和提取特徵等功能，可以透過發現數據中的主要變異性來簡化複雜的數據集，使研究者更容易理解數據的結構和模式，因此我們決定運用 PCA，自動化且系統性從大量類星體中辨識出雙類星體，再進一步探索其物理性質。

二、研究目的

類星體是星系中心的黑洞周圍環繞著許多氣體，這些氣體原子會發出發射光譜，當兩個類星體互繞時會產生相對速度，根據都卜勒原理，我們觀測到這兩個類星體周圍的氣體發出的發射光譜波長可能有所不同，導致我們觀測到的光譜形狀可能有雙峰特徵。目前天文學界對於雙類星體有別於一般類星體的特徵了解不多，而目前只有類星體周圍氣體發出的雙峰發射譜線曾被作為用於分辨出雙類星體的依據。兩個類星體互繞時產生的相對速度會使其周圍的氣體發出的發射光譜波長有所不同，導致我們觀測到的光譜形狀可能有雙峰特徵。過去的研究（例如：Smith et al. 2010）使用肉眼檢查方式從光譜資料中找出這些雙峰，但目前觀測

類星體光譜的數據量非常大，卻較少用系統性的方式瞭解這些類星體，因此我們希望能建立一個能快速從大量數據中找到雙類星體的方式。若能發展出一套系統性找尋雙類星體的方式，能使已知的雙類星體樣本數量大幅增加，有助於我們研究雙類星體的物理性質，並進一步找出雙類星體獨有的其它特徵。

我們將根據下列目的來進行研究：

- (一) 自動化且系統性尋找雙類星體
- (二) 研究雙類星體光譜有別於一般類星體的物理性質與特徵

貳、研究設備及器材

一、數據來源

(一) 史隆數位化巡天 (Sloan Digital Sky Survey, SDSS)

本研究使用 SDSS DR16 (Data Release 16, 於 2019 年發行) 類星體的星表 (Lyke et al. 2020), 並選用 SDSS DR17 (Data Release 17, 於 2021 年發行) 的類星體光譜。史隆數位化巡天擁有廣泛的觀測範圍, 是現今最大規模的巡天項目之一, 進行可見光波段的觀測, 提供天文學數據, 並使用先進的望遠鏡與數位攝影技術, 收集大量光譜及影像數據, 且其數據對全球科學界開放, 為公眾提供完整的數據庫。DR16 所觀測到的類星體 (Lyke et al. 2020) 共有 75 萬筆, 其中共有約 7.6 萬筆資料為紅移值小於 0.8 的類星體。在本研究中, 我們以系統抽樣 (等距抽樣), 每 28 筆資料取一筆資料, 共得到 2733 筆紅移值小於等於 0.8 的類星體光譜作為訓練 PCA 的數據集, 後續我們也將所有光譜都運用在分析上。

(二) 論文 (Smith et al. 2010) 內的星表

此篇論文 (Smith et al. 2010) 於 2010 年發表, 其內使用的星表與光譜皆是 2008 年發行的 SDSS DR7 (Data Release 7), 在約 2.1 萬個紅移小於 0.7 的類星體光譜中以肉眼觀察, 共找到 148 個具有雙峰特徵的光譜。本研究將論文 (Smith et al. 2010) 內的 148 個光譜進行篩選並刪除 4 個不適用的光譜, 剩餘的 144 個做為雙類星體的範例, 用於研究中的訓練分類方法的數據。

(三) DESI Legacy Surveys

本研究使用 DESI Legacy Surveys DR10 (Data Release 10, Dey et al. 2019) 的類星體影像分析類星體的性質。DESI Legacy Surveys 包含 Mayall z-band Legacy Survey (MzLS)、DECam Legacy Survey (DECaLS)、Beijing-Arizona Sky Survey (BASS)

三個巡天計畫的觀測資料，其 DR10 合併了來自 NOIRLab 的附加 DECam 資料來擴大觀測範圍，包含 >20000 平方度的銀河外天空中 g、r、i、z 四個波段的觀測資料。在我們的研究中，因為其觀測範圍廣大，我們能從其中獲得幾乎所有 SDSS DR16 類星體的影像。

(四) Hyper Suprime-Cam Subaru Strategic Program (HSC SSP)

本研究使用 HSC SSP DR3 (Data Release 3, Aihara et al. 2022 於 2021 年發行) 的類星體影像分析類星體的性質。HSC SSP 使用 Subaru 望遠鏡進行大範圍的多波段成像調查，具有不同的區域覆蓋範圍和深度，旨在解決各種天體物理問題，其 DR3 測量了超過 5 億個天體的精確位置、亮度和形狀，包含 670 平方度的天空中 u、g、r、i、z 五個波段的觀測資料。雖然其觀測範圍較 DESI Legacy Surveys 小，但解析力較佳，因此能提供部分類星體更清楚的影像。

(五) Very Large Array Sky Survey (VLASS)

本研究使用 VLASS 的類星體影像分析類星體的性質。VLASS 旨在使用 Very Large Array (VLA) 進行射電巡天。VLASS 觀測工作於 2017 年 9 月開始，預計於 2024 年完成觀測。最終其觀測範圍將覆蓋 VLA 可見的整個天空共 33885 平方度。VLASS 頻率範圍為 2 - 4GHz，角度解析度約為 2.5 角秒。其影像可以使我們了解類星體在無線電波段的影像。

二、軟體及工具環境

(一) 作業系統：Windows 11

(二) 虛擬環境：Anaconda

(三) 編輯介面：Jupyter Notebook

(四) python：具有高效能的高階資料結構，直譯式且物件導向的程式語言

1. sklearn：提供資料分析與模擬的工具，包含各種機器學習的演算法與資料前處理技術（本研究使用 sklearn.neighbors、svm 及 sklearn.metrics）

(1) svm：用於分類、回歸與異常檢測等任務，sklearn 提供了幾種不同的 svm 功能與不同的核函數，也可以調整參數來優化模型的表現。本研究使用非線性 SVM 中的徑向積函數核（Radial basis function kernel, RBF kernel）作為對類星體進行分類的工具。

(2) sklearn.metrics：提供一系列用於評估模型性能的工具，包含了分類、回歸、聚類、排名與綜合評估等多項任務。本研究用於量化類星體分類模型的表現。

(五) topcat：可以載入並分析星表、繪製圖表的天文專用軟體。本研究利用 topcat 篩選星表內的數據，限制紅移大小，系統抽樣光譜數據。

(六) Cutout Image List Tool for DESI Legacy Imaging Survey：由猶他大學物理與天文學助理教授模姚媛製作的網頁，能以 RA、DEC 數據快速瀏覽各觀測計畫的影像，並快速連結至天體的其他相關數據，我們使用此查詢工具獲得類星體的影像與相關資訊。

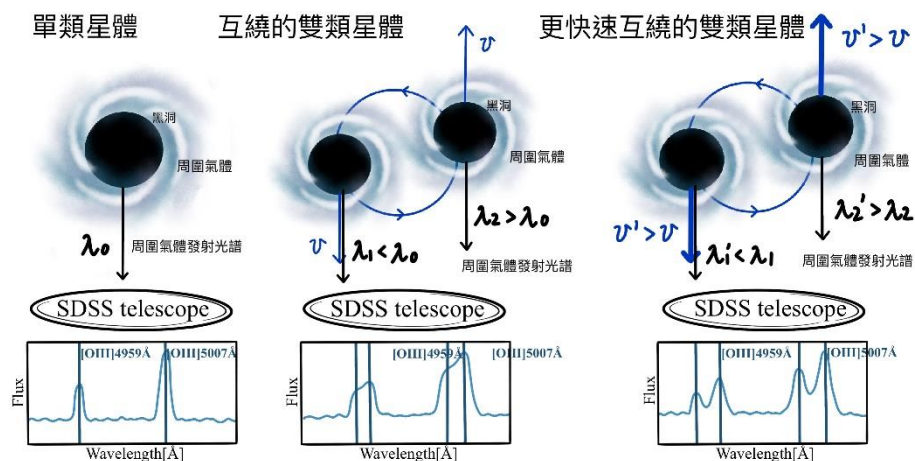
參、研究過程或方法

一、名詞解釋

(一) 類星體：每個星系中央都有一個超大質量黑洞，其引力會使氣體在其周圍高速環繞，摩擦產生巨大能量，使人們觀測到類似恆星卻比恆星亮非常多的天體，因此稱為類星體。簡言之，類星體是星系中央超大質量黑洞和其周圍氣體的總稱。

(二) 雙類星體：星系合併是星系演化過程中的常見現象，許多大型星系都經歷過至少一次星系合併，因此我們應該能在宇宙中觀測到正在合併的星系。合併中的兩星系的類星體會被彼此的引力拉扯，產生互繞的現象，即為雙類星體。根據 Smith et al. 2010，雙類星體之間的距離，近者約 1~10 千秒差距 (kpc, 1000 秒差距 \approx 3.26 光年)，亦有兩者相距大於 10 千秒差距的雙類星體，但大多約相距 10 千秒差距，互繞的相對速度約每秒數百公里。

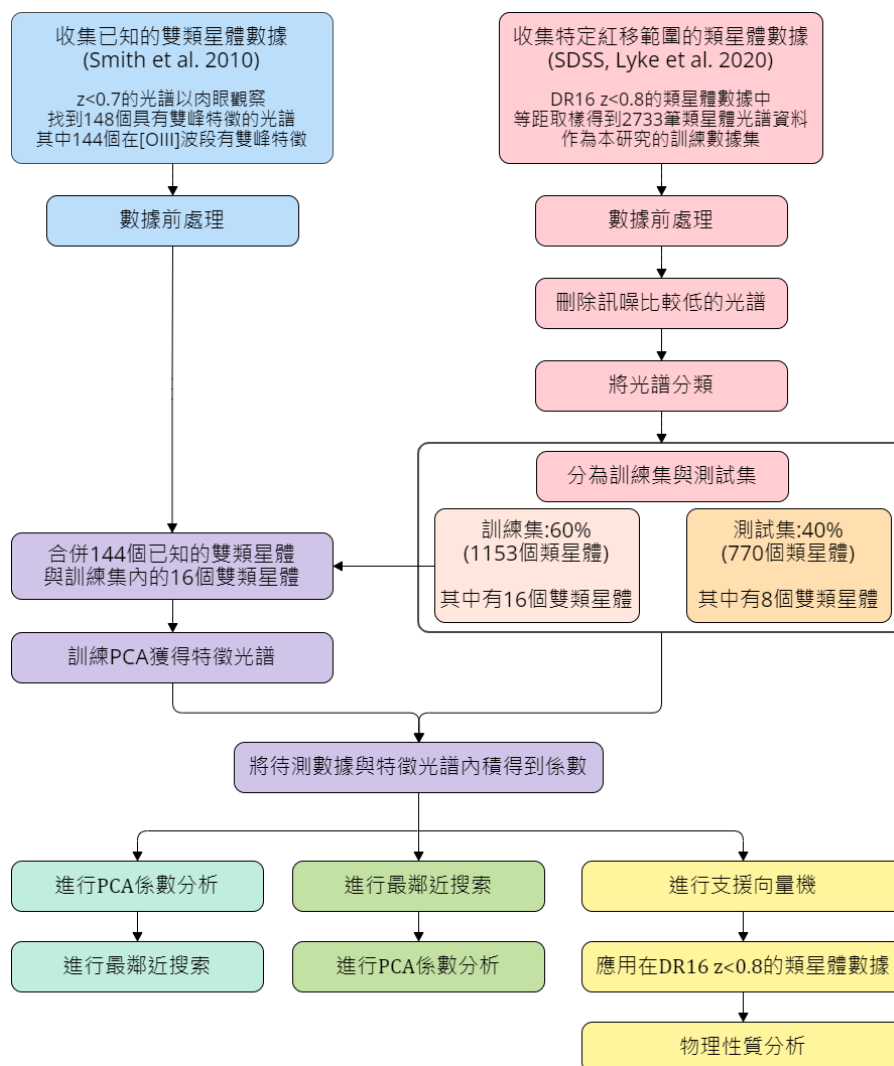
(三) 雙峰發射譜線 (Double-peak Emission Line)：類星體周圍的氣體原子會發出發射光譜，同一種元素產生的發射譜線應在固定的波長，然而被我們的觀測到的波長會跟類星體紅移有關。當兩個類星體互繞時會產生相對速度，根據都卜勒原理，因兩類星體徑向上的速度差，造成紅移的些許差異，而在不只一個波長觀測到相同元素的發射譜線，此即為雙峰發射譜線。(圖一)為雙峰特徵光譜的原理示意圖。



(圖一) 雙峰特徵光譜的原理示意圖。本圖由作者親自製作。

- (四) 最鄰近搜索 (Nearest Neighbor Search, NNS)：最鄰近搜索是機器學習中一種資訊檢索的方法，在大量數據及中可以用於在給定數據集中查找與目標數據最相似的鄰近數據，會通過計算各數據與目標數據的距離來識別最接近目標數據的訓練集中的數據。本研究使最鄰近搜索在PCA模型得到的係數數據集中，以範例雙類星體係數為目標數據，計算待測類星體係數與雙類星體係數的距離。我們再設定欲挑選出的待測類星體數量，使最鄰近搜索挑選出與雙類星體係數較接近的待測類星體。
- (五) 支援向量機 (Support Vector Machine, SVM)：支援向量機是一種用於分類和回歸分析的監督學習模型，能夠在多維的數據中找到一個最佳的超平面將不同類別的數據點分開。支援向量機的目標是最大化不同類別之間間隔，間隔指距離超平面最近的數據點到超平面的距離，透過最大化間隔，可以提高支援向量機對新數據的分類準確性。本研究使用支援向量機在PCA模型得到的係數數據集中，以訓練集的雙類星體係數和分類為目標數據，將測試集的光譜數據進行分類。

二、研究構思與架構



(圖二) 研究流程示意圖。本圖由作者親自製作。

三、研究過程及方法

(一) 取得數據

1. 取得已知雙類星體數據

- (1) 下載 Smith et al. 2010 星表，依星表中 fiber、MJD、plate 在 SDSS DR17 中下載相應的類星體光譜，共取得 148 個 Smith et al. 2010 定義的雙類星體光譜。
- (2) 刪除 4 個 [O III] 發射譜線波長沒有雙峰特徵的光譜，保留 144 個已知的雙類星體光譜。

2. 取得特定紅移範圍的類星體數據

- (1) 下載 SDSS DR16 (Data Release 16) 類星體的星表 (Lyke et al. 2020)
- (2) 用 topcat 挑選紅移小於 0.8 的類星體。因我們主要觀察的雙峰為雙電離氧 [O III]，已知其兩條譜線的波長分別為 4959 Å 及 5007 Å，而史隆數位化巡天 (SDSS) 觀測範圍約為 3800 Å 至 9200 Å，依紅移值公式 (式一) 計算：

$$\lambda_{obs} = \lambda(1 + z) \text{ (式一)}$$

註: z 為紅移值， λ_{obs} 表望遠鏡觀察到的光波波長， λ 表發光體原始的光波波長

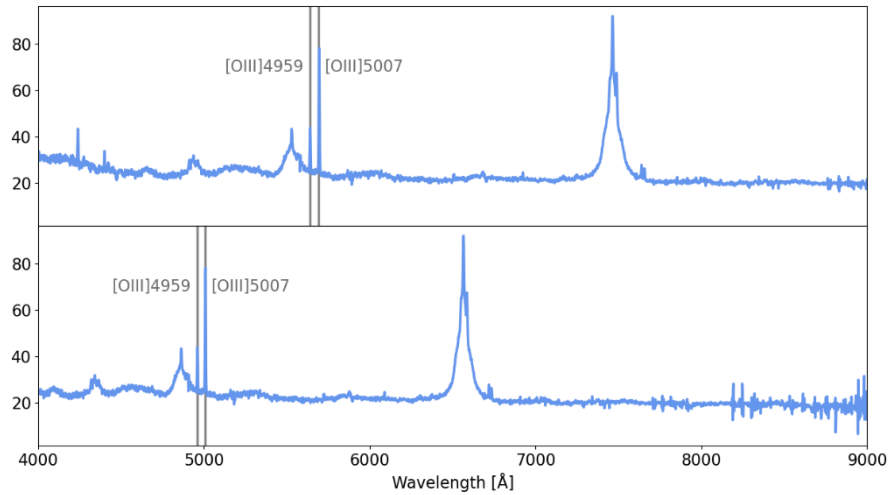
可知若欲使研究中的所有星體都擁有該區段的數據資料，其紅移值須在約 0.83 以下，因此我們選擇以紅移小於 0.8 的類星體進行研究。

- (3) 依星表中的 fiber、MJD、plate 在 SDSS DR17 中下載相應的類星體光譜，共取得 76528 個紅移小於 0.8 的類星體光譜。
- (4) 在之後訓練能篩選出雙類星體的模型時，我們需要有經過雙類星體與非雙類星體分類的光譜數據集，才能量化模型篩選雙類星體的表現。因此我們需要先從 76528 個類星體光譜中，用等距取樣的方式，以 28 個光譜為間距，取得 2733 個光譜，這樣的數量讓我們能對每個類星體進行分類，同時也能代表整個類星體的數據集。

(二) 數據前處理

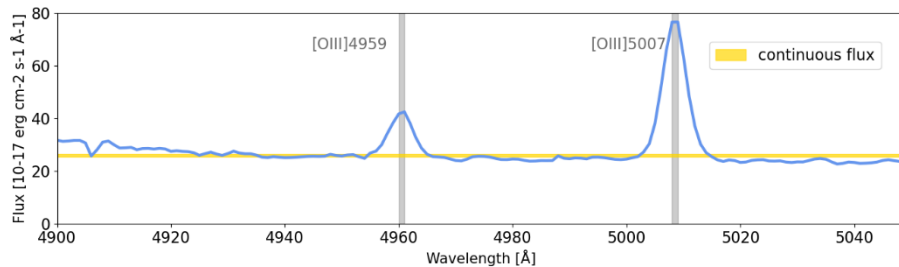
1. 為了排除紅移、光譜數據大小、類星體連續光譜和發射譜線強度的影響，讓電腦機器學習演算法能著重在分析雙電離氧 [O III] 發射譜線形狀，我們對光譜進行去除紅移、選取特定波長範圍並對光譜內插、去除連續光譜、發射線強度標準化的處理。

- (1) 去除紅移：每個類星體依其距離遠近有不同的紅移量值，若要去掉類星體紅移不同對波長的影響，可依紅移值公式 (式一)，將每個類星體光譜的所有波長都除以該類星體的 $1+z$ 。(圖三) 為去除紅移的過程示意圖。



(圖三) 去除紅移的過程示意圖，上圖為去除紅移前的光譜，下圖為去除紅移後的光譜，灰線標記為我們要分析的[OIII]發射譜線。本圖由作者親自製作。

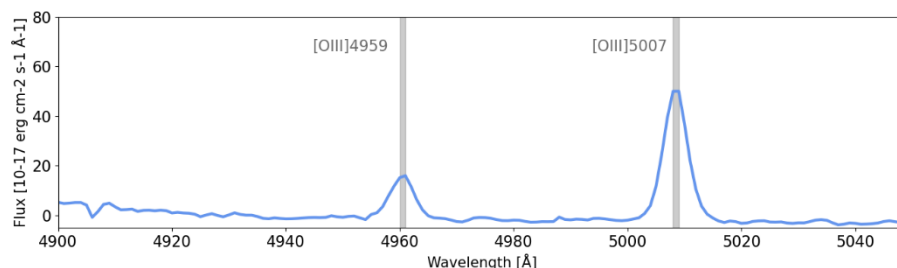
- (2) 限制波長範圍：我們取的波長範圍是 $4900\text{\AA} \sim 5049\text{\AA}$ ，即我們的分析對象雙電離氧[OIII]去除紅移後所在波長範圍。
- (3) 波長-光通量數據內插：因為每個類星體的紅移不同，處去紅移會導致不同類星體的數據點間隔大小不同，造成相同波長範圍的光通量數據量不同，但PCA分析需要所有的光譜在統一的波長區間和一致的間隔，而內插是一種通過已知的、離散的數據點，在範圍內推求新數據點的過程或方法，可以使每個類星體在特定波長範圍有相同的光通量數據量。我們在波長範圍是 $4900\text{\AA} \sim 5049\text{\AA}$ 以 1\AA 為間隔統一取 150 筆光通量資料。(圖四) 為限制波長範圍與內插後的光譜示意圖。



(圖四) 限制波長範圍與內插後的光譜示意圖，波長範圍為 $4900\text{\AA} \sim 5049\text{\AA}$ ，黃線標記為類星體連續光譜，將在下一個部分進行去除。本圖由作者親自製作。

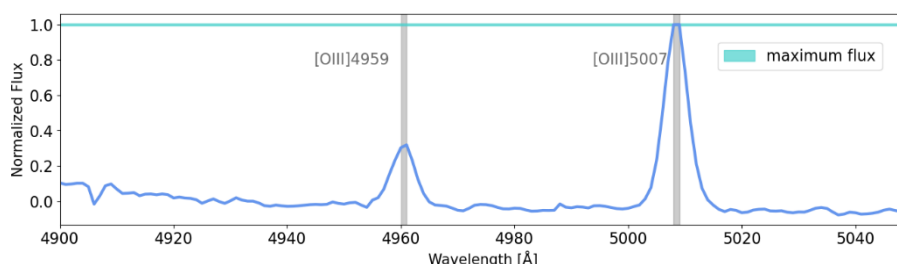
- (4) 去除連續光譜訊號：每個類星體離觀察者的遠近不同，發出的能量大小也不同，每個類星體發出的能量不同導致不同的連續光譜訊號。但後續以機器學習對其光譜行進行分析時，目標為特定發射譜線的形狀，背景光通量可能影響電腦在判斷光譜形狀時的結果，故資料前處理時需去除連續光譜訊號。由(圖六)可見 $4900\text{\AA} \sim 4939\text{\AA}$ 和 $5030\text{\AA} \sim 5049\text{\AA}$ 不包含我們要分析的雙電離氧[OIII]譜線，而且幾乎代表此類星體的連續光譜訊號(上圖背景光通量約 $25 \sim 30(10^{-17}\text{erg cm}^{-2}\text{s}^{-1}\text{\AA}^{-1})$)，在這些有兩個明顯的雙電離氧

[O III] 譜線的類星體數據中皆是如此，所以我們以 4900 Å~4939 Å 和 5030 Å~5049 Å 的光通量的平均為連續光譜訊號，並將所有類星體的連續光譜訊號數據都減掉自己的連續光譜訊號。（圖五）為去除類星體連續光譜後的光譜示意圖。



（圖五）去除類星體連續光譜後的光譜示意圖。本圖由作者親自製作。

- (5) [O III] 發射譜線強度標準化：每個類星體周圍氣體發出的能量不同，相同放射線在每個類星體的光通量大小差異可能很大，但本研究主要聚焦在以放射線的形狀作為判斷雙類星體的依據，而非光通量大小。為了使所有 [O III] 放射線強度有相同標準，將所有類星體的光通量除以自己在該範圍內的光通量最大值，使每筆數據的 [O III] 5007 放射線光通量最大值皆為 1。（圖六）為 [O III] 發射譜線強度標準化後的光譜示意圖。

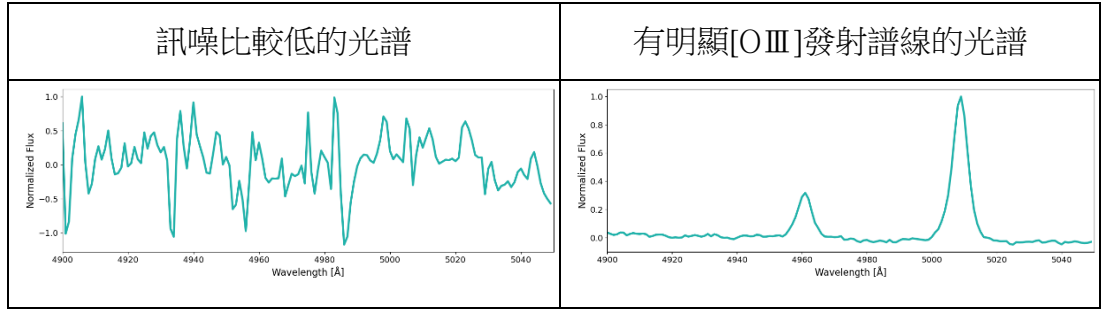


（圖六）[O III] 發射譜線強度標準化後的光譜示意圖，藍線發射譜線光通量最大值。本圖由作者親自製作。

(三) 刪除訊噪比較低的光譜

1. 篩選雙類星體時，我們以 [O III] 發射譜線的特徵做為判斷依據，並研究 [O III] 發射譜線的特徵，因此我們希望在進行分析前，先把訊噪比較低的光譜刪除，減少沒有訊號的光譜對模型的影響，使 [O III] 發射譜線的特徵更容易判斷。
2. 在 Smith et al. 2010 的 148 個雙類星體數據中，大多數光譜皆為有明顯 [O III] 發射譜線特徵的光譜，僅有四個光譜的 [O III] 發射譜線不明顯，因此將其刪除，保留 144 個雙類星體光譜，進行後續的數據分析。
3. SDSS DR16 等距抽樣取得的 2733 個光譜中，有許多訊噪比較低的光譜，訊噪比較低的光譜與有明顯 [O III] 發射譜線的光譜範如（表一）：

(表一) 左圖為訊噪比較低的光譜，右圖為有明顯[O III]發射譜線的光譜。表中圖片由作者親自製作。



4. 訊噪比的計算：SDSS DR16 觀測光譜時儲存光通量資料方式為一個光通量數值對應到一個像素，在 SDSS DR16 的光譜資料中，光譜的每一個光通量數值都有對應的逆方差 (Inverse variance, 在 SDSS 的資料中以 *ivar* 表示)，我們可以由 (式二) 計算出每一個光通量的訊噪比 (signal to noise, *S/N*)：

$$S/N = flux \times \sqrt{ivar} \quad (\text{式二})$$

註： $ivar = 1/noise^2$ ，而 *flux* 即代表光譜的訊號 (*S*)

影響光譜訊噪比的因素：由 (式二) 可知，訊噪比與光通量量值有關，以下為影響訊噪比量值的因素：

- (1) 類星體連續光譜：若類星體的連續光譜較強，但沒有明顯的[O III]發射譜線，仍會使該光譜整體有較高的訊噪比。
- (2) 發射譜線寬度：若發射譜線較寬，有較高光通量的像素較多，發射譜線較窄，有較高光通量的像素較少，因此在有明顯[O III]發射譜線的情況下，發射譜線較窄的光譜，整體訊噪比較低。

為避免沒有[O III]發射譜線特徵卻因連續光譜較強而被保留，或有明顯[O III]發射譜線特徵卻因光譜較窄而被刪除，我們在前述數據前處理時，去除類星體連續光譜後，再計算訊噪比，並以訊噪比最大值 ≥ 5 為保留在數據集內的條件。

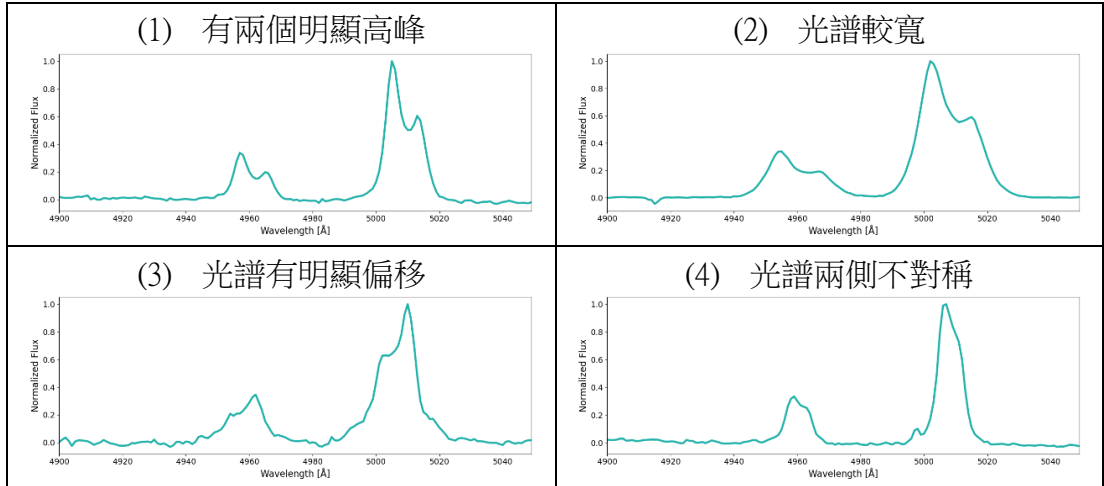
5. 以訊噪比篩選結果：從 SDSS DR16 等距抽樣取得的 2733 個光譜中，我們刪除 810 個訊噪比較低的光譜，留下 1923 個光譜做為之後分析使用的數據集。

(四) Smith et al. 2010 雙類星體數據的定義

1. 我們將 Smith et al. 2010 雙類星體數據作為本研究中分析的雙類星體光譜範例，因此我們必須先了解其對雙類星體光譜的定義，才能定義出我們從待測類星體中篩選出的光譜是否為雙類星體光譜，然而我們並不清楚當時作者對雙類星體光譜的篩選標準，只能從 Smith et al. 2010 中在此波段有明顯雙峰特徵的 144 個雙類星體加以觀察並推敲其篩選標準。
2. 論文中多數雙類星體光譜在[O III]4959 Å 及[O III]5007 Å 有明顯的兩個高峰，而

部分光譜雖沒有兩個明顯高峰，亦可從光譜較寬、有明顯偏移或兩側不對稱等特徵判斷出雙發射譜線，（表二）為論文雙類星體光譜的範例。

（表二）論文中雙類星體光譜的範例。表中圖片由作者親自製作。



(五) 待測類星體光譜的定義與分類

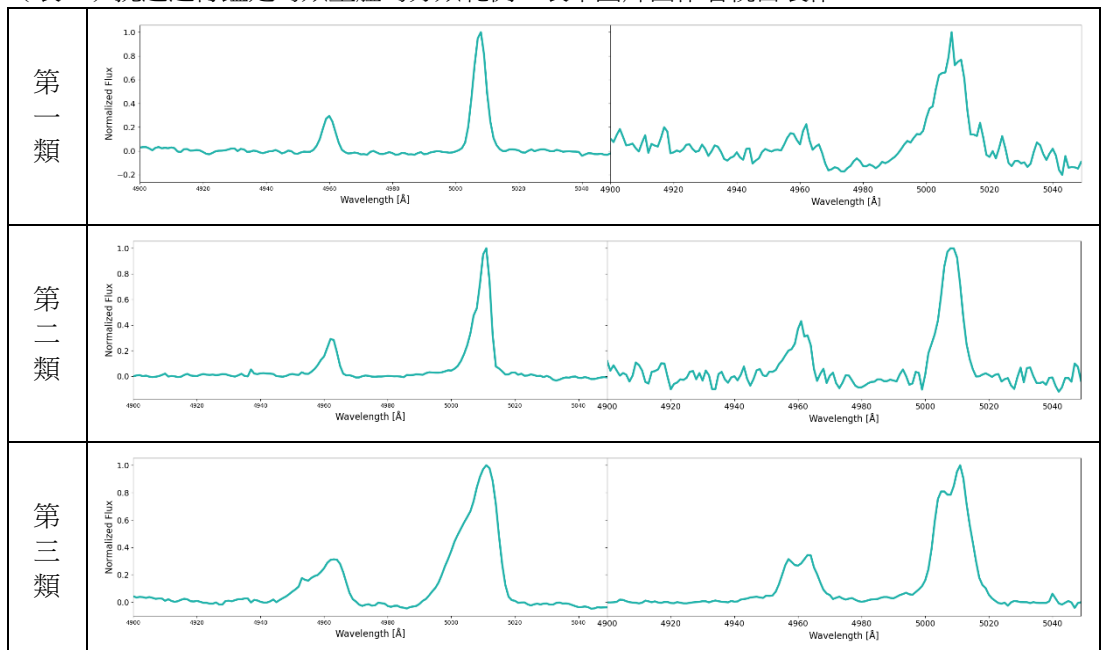
1. 為了在分析過程中以同樣的標準檢視每個分析方式篩選出的類星體是哪一種類型，我們參考 Smith et al. 2010 對類星體光譜的定義並將其分類，再用人工檢視的方式，將所有待測類星體進行分類，如此即可量化每種分析篩選出的類星體類別。（表三）我們將所挑選進行鑑定的類星體進行分類，共分為三類：

第一類：發射譜線不明顯或為單峰光譜（1620 個）

第二類：發射譜線較寬、兩側不對稱等可能具有雙類星體特徵的光譜（279 個）

第三類：發射譜線有兩個高峰、明顯偏移及上述與論文雙類星體光譜相似的光譜（24 個）

（表三）挑選進行鑑定的類星體的分類範例。表中圖片由作者親自製作。



(六) 主成分分析 (PCA)

1. PCA 是一種可以簡化數據集，並保留最大特徵的統計方法，能使電腦從大量的數據中快速辨認出雙類星體的光譜特徵。我們將光譜數據放入 PCA 模組，得到平均光譜、特徵光譜（主成分，Principal Components）、係數（Coefficients）、解釋比例，並以平均光譜、特徵光譜與係數組合出重建光譜。
2. 訓練 PCA：將光譜放入 PCA 中訓練後，會得到平均光譜以及特徵光譜，平均光譜代表所有光譜的平均，而特徵光譜代表所有光譜的特徵以及光譜形狀的特定變化模式。將其它光譜與特徵光譜內積可以得到係數，而係數代表其它光譜與該特徵光譜的關聯性，因此可以作為判斷其它光譜特徵的依據。由上述可知，當特徵光譜代表了雙峰光譜的特徵時，其它光譜與特徵光譜內積得到的係數就可以代表光譜與雙峰特徵的關聯性。

(1) 區分訓練集與測試集：

本研究使用 PCA 係數分析、最鄰近搜索 (NNS)、支援向量機 (Support vector machine, SVM) 篩選雙類星體，SVM 篩選雙類星體時須要訓練集與測試集，依據訓練集的類星體類別預測測試集的類星體類別。為了能以相同標準量化，以所有方式進行篩選時都統一使用已經區分為訓練集與測試集的數據。

我們先將 1923 個類星體依據索引值 (index)，以等距取樣的方式分為 60% 的訓練集與 40% 的測試集，區分方式如下：

$index \div 5 \dots 0 \text{ or } 2 \text{ or } 4$ ：訓練集 (1153 個類星體，佔所有類星體的 60%)

$index \div 5 \dots 1 \text{ or } 3$ ：測試集 (770 個類星體，佔所有類星體的 40%)

(2) 訓練數據集：

訓練集的 1153 個類星體中，有 16 個雙類星體光譜，我們將其與 144 筆來自論文 (Smith et al. 2010) 的已知雙類星體光譜合併為 160 個雙類星體光譜，再以其訓練 PCA。因為他們是標準且特徵明顯的雙峰光譜，用來訓練 PCA 可使前幾個特徵光譜代表雙峰特徵，使其它光譜與之內積得到的係數能表現雙峰特徵與光譜的關聯性。

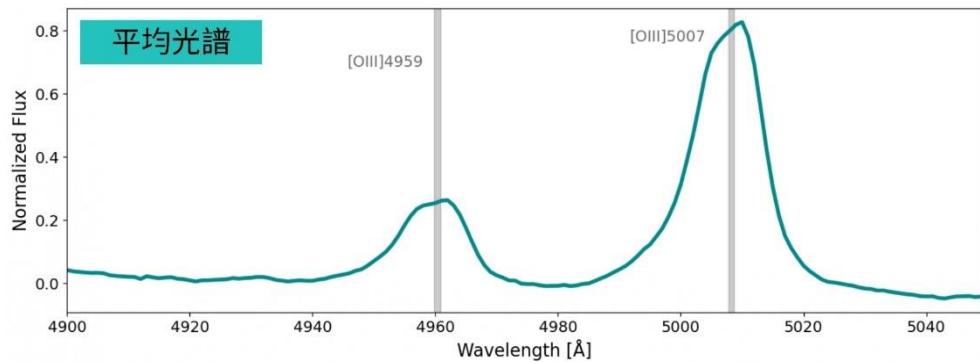
(3) 平均光譜：160 個雙峰光譜的平均 (圖七)。

(4) 特徵光譜：共 150 個特徵光譜 (依 PCA 模型取得特徵光譜的方式，特徵光譜數量與每個光譜光通量數據相同，共 150 個)。為了瞭解特徵光譜代表

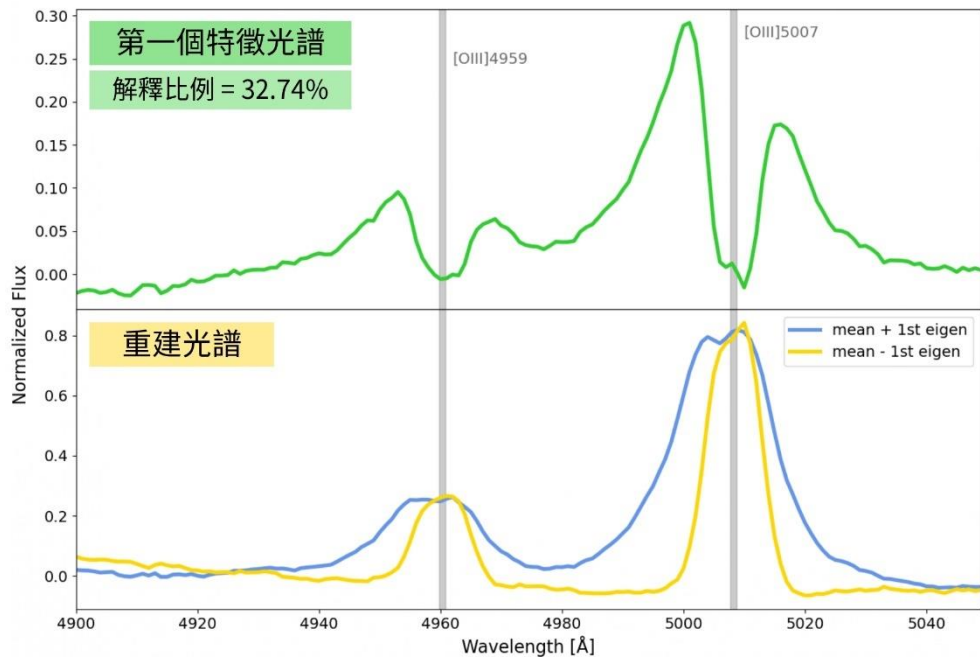
的變化模式，我們可以將特徵光譜乘上一個係數，加到平均光譜上，以模擬該特徵光譜在原始光譜中的特定變化模式。

- (5) 重建光譜：重建光譜是平均光譜加上不同表現程度的特徵光譜，每個特徵光譜的表現程度由係數決定，重建光譜的方法如（式三）。（圖八）、（圖九）、（圖十）中藍色線代表平均光譜加上對應的特徵光譜，橘黃色的線則代表平均光譜減去對應的特徵光譜。以下以前三個特徵光譜為例，呈現特徵光譜的變化模式。

$$reconstruct_n = mean + \sum_{k=0}^{150} component_k \times coefficient_{n,k} \quad (\text{式三})$$

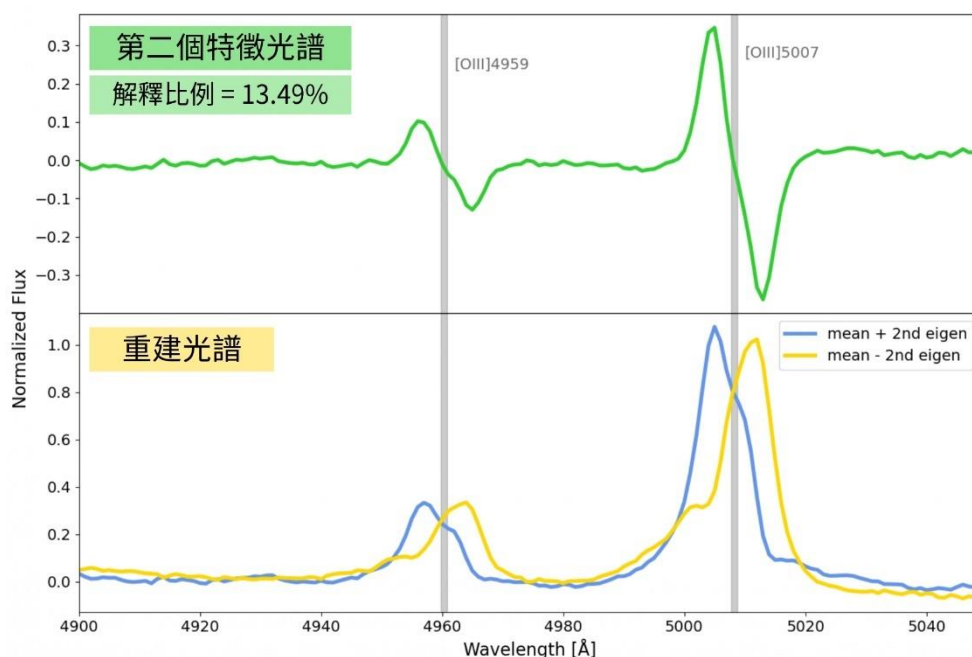


（圖七）160 個雙峰光譜的平均。本圖由作者親自製作。



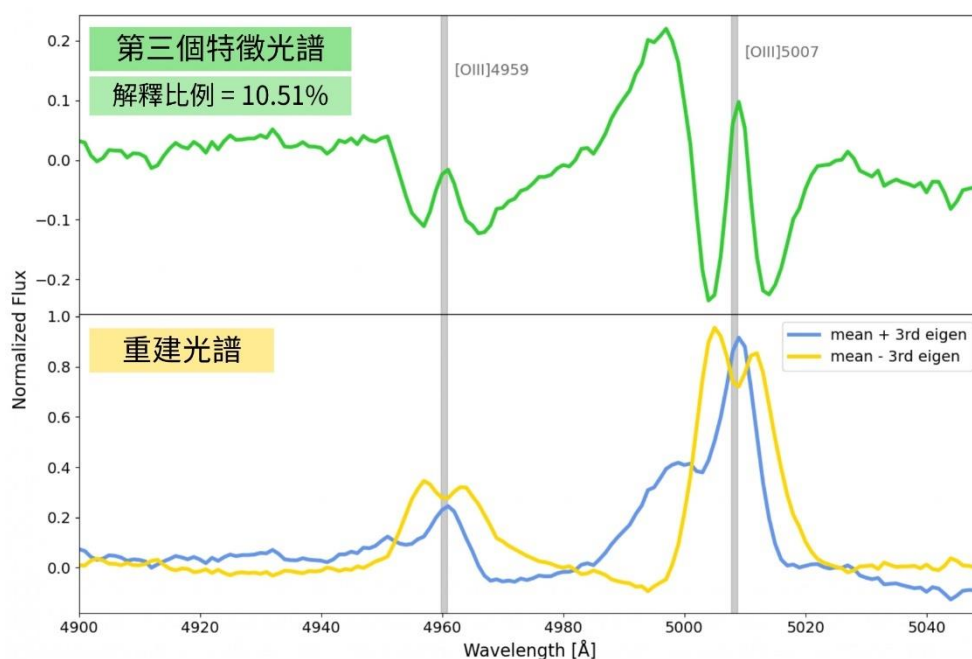
（圖八）藍色的線代表平均光譜加上第一個特徵光譜，橘黃色的線則代表平均光譜減去第一個特徵光譜。本圖由作者親自製作。

第一個特徵光譜的重建光譜中，平均光譜加上第一個特徵光譜的發射譜線寬度較平均光譜減去第一個特徵光譜的更寬，由此可知當係數為正時，第一個特徵光譜可以表現出雙峰光譜的較寬特徵。



(圖九) 藍色的線代表平均光譜加上第二個特徵光譜，橘黃色的線則代表平均光譜減去第二個特徵光譜。本圖由作者親自製作。

第二個特徵光譜的重建光譜中，平均光譜加上第二個特徵光譜與平均光譜減去第二個特徵光譜分別向右、向左偏移，由此可知正負係數接可以呈現雙峰光譜偏移的特徵。



(圖十) 藍色的線代表平均光譜加上第三個特徵光譜，橘黃色的線則代表平均光譜減去第三個特徵光譜。本圖由作者親自製作。

第三個特徵光譜的重建光譜中，平均光譜加上第三個特徵光譜會使光譜不對稱，減去第三個特徵光譜可以產生雙峰特徵，但加上第三個特徵光譜時主要發射線沒有偏移，由此可知負的係數貢獻的雙峰特徵較顯著。

(6) 解釋比例：代表特徵光譜在光譜中的貢獻程度，越前面的特徵光譜解釋比例越大，越能呈現光譜的特徵。

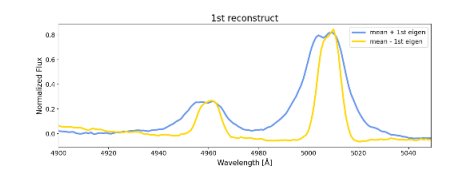
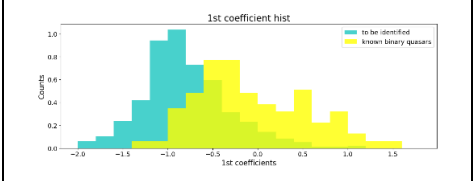
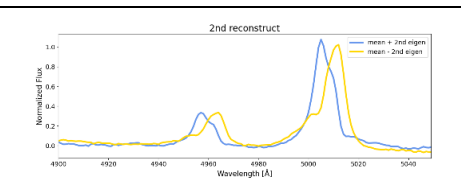
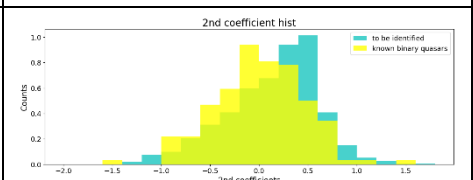
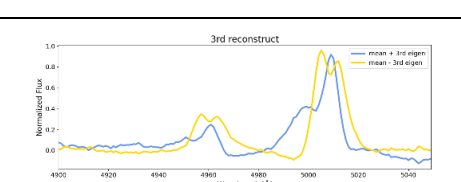
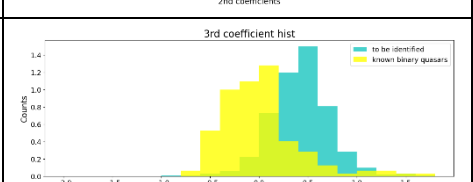
(七) 數據分析

將 1923 個待測的類星體與上步驟所得到已知雙類星體的特徵光譜（主成分）內積後，我們可以得到每個光譜在特徵光譜相對應的係數（共 1923x150 個）。進行以下四種方式篩選雙類星體時，我們以 160 個雙類星體係數為依據，從測試集的 770 個待測類星體中篩選雙類星體。

1. PCA 係數分析

我們對內積得到的係數進行 PCA 係數分析。為了比較係數在已知雙類星體和待測類星體之間的差異，我們將這些係數畫成直方圖進行比較。以係數作為從待測類星體中篩選出可能雙類星體的依據，係數與已知雙類星體越接近者，就越有可能是雙類星體。（表四）是前三個特徵光譜的重建光譜與係數直方圖，為了使數量分布能明顯呈線，直方圖皆使用標準化數量繪製；係數直方圖中，藍色直方代表待測類星體係數與標準化數量的關係，黃色直方則代表論文中已知雙類星體係數與標準化數量的關係。

（表四）前三個特徵光譜的重建光譜與係數直方圖。表中圖片由作者親自製作。

| 重建光譜 | 特徵光譜 | 係數直方圖 |
|---|------|--|
|  | 第一個 |  |
|  | 第二個 |  |
|  | 第三個 |  |

由（表四）可以看到，第一個特徵光譜係數為正時可以使光譜更寬，而且我們已知雙峰光譜的發射譜線普遍上比單峰光譜更寬，同時也可以從第一個特徵光譜的係數直方圖中觀察到，論文中已知雙類星體的第一個係數整體來說有比待測類星體的更大的趨勢，因第一個係數對雙類星體與非雙類星體有顯著差異，因此我們就將它作為篩選的目標之一。第二個特徵光譜的重建光譜中，雖然雙

峰光譜的發射譜線會向左右偏移，但偏移的方向會因為兩個類星體的位置、相對速度等性質而有所差異，也可以從第二個特徵光譜的係數直方圖中觀察到，論文中已知雙類星體的第二個係數與待測類星體的比例差距較小，因此第二個特徵光譜的係數無法作為篩選目標。

篩選條件：因為前 10 個特徵光譜可以看出明顯的光譜特徵，因此我們先使用前 10 個特徵光譜進行篩選，但篩選後檢視被選到的候選雙類星體中仍有許多並非被定義為雙類星體的類星體，因此我們設定更嚴格的條件，並且也再次觀察直方圖，發現雖然第 11 到第 15 個特徵光譜沒有明顯的光譜特徵，解釋比例也很小，但已知的雙類星體光譜係數大部分都集中在 ± 0.1 ，因此我們將第 11 到第 15 個特徵光譜的係數也加上篩選條件，獲得更好的篩選效果。

篩選結果：從 770 個類星體中篩選出 167 個候選雙類星體

（表五）PCA 係數分析選取到的各類別類星體數量。

| 第一類（641 個） | 第二類（121 個） | 第三類（8 個） |
|------------|------------|----------|
| 104 | 55 | 8 |

2. 最鄰近搜索

同樣使用內積得到的係數，我們使用最鄰近搜索篩選這些數據。最鄰近搜索以電腦統計所有係數中待測類星體係數與已知雙類星體係數相近程度，可調整欲選擇的數量或距離，是電腦自動化與精準化的係數分析。

篩選結果：從 770 個類星體中篩選出 167 個候選雙類星體

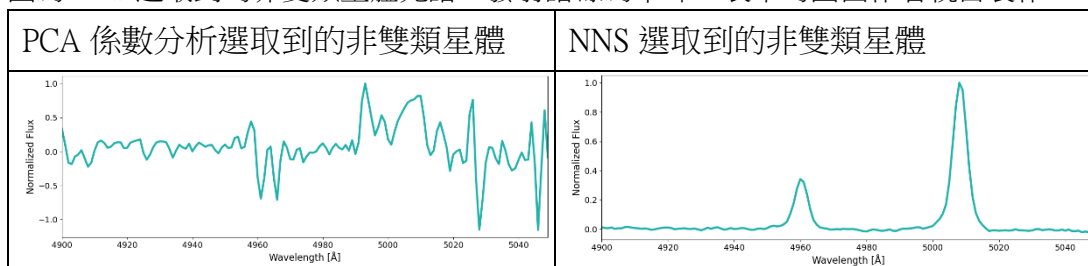
（表六）最鄰近搜索選取到的各類別類星體數量。

| 第一類（641 個） | 第二類（121 個） | 第三類（8 個） |
|------------|------------|----------|
| 90 | 69 | 8 |

3. 交替使用 PCA 係數分析與最鄰近搜索

我們在進行以上兩種找尋方式後，檢視被選到但並非我們所定義的雙類星體的光譜，發現 PCA 係數分析所選到的較多發射譜線不明顯的光譜，最鄰近搜索所選到的則有較多單峰光譜，如下表所示。

(表七) 左圖為 PCA 係數分析選取到的非雙類星體光譜，其發射譜線既不明顯，右圖為 NNS 選取到的非雙類星體光譜，發射譜線為單峰。表中的圖由作者親自製作。



因此我們猜測，若能分別對一筆數據進行 PCA 係數分析與最鄰近搜索，就能將發射譜線不明顯的光譜與單峰光譜都篩選掉，因此我們將已經做完最鄰近搜索的 167 個候選雙類星體再進行 PCA 係數分析。

篩選條件：與 PCA 係數分析的篩選條件相同。

篩選結果：從 167 個類星體中篩選出 86 個候選雙類星體

(表八) 先最鄰近搜索再 PCA 係數分析選取到的各類別類星體數量。

| 第一類 (90 個) | 第二類 (69 個) | 第三類 (8 個) |
|------------|------------|-----------|
| 35 | 43 | 8 |

同時，我們也將已經做完 PCA 係數分析的 167 個候選雙類星體再進行最鄰近搜索，並和上述方式進行綜合比較。

篩選結果：從 167 個類星體中篩選出 86 個候選雙類星體

(表九) 先 PCA 係數分析再最鄰近搜索選取到的各類別類星體數量。

| 第一類 (104 個) | 第二類 (55 個) | 第三類 (8 個) |
|-------------|------------|-----------|
| 37 | 41 | 8 |

我們以前述四種方式盡可能選到所有雙類星體 (第三類)，並在此條件下綜合比較四種方式篩選雙類星體的效果。交替使用 PCA 係數分析與最鄰近搜索後選到的非雙類星體 (第一、二類) 數量比單一使用 PCA 係數分析或最鄰近搜索少了一半以上，然而單看交替使用兩種方式的篩選結果，有約 90% 是非雙類星體。考慮到光譜特徵的變化性大，數據集複雜，PCA 係數分析與最鄰近搜索的篩選方式對於複雜的數據集而言精細度不足，因此我們需要一種能增加光譜特徵差異的演算法，使係數之間的區別更明顯，增加篩選的準確性。

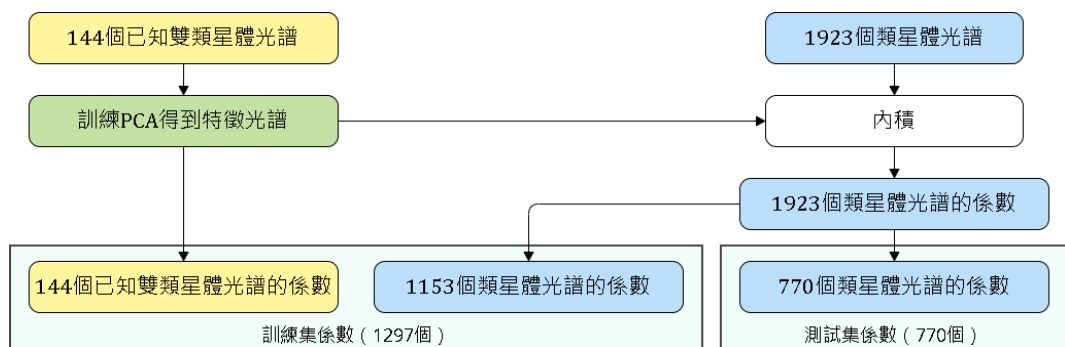
(八) 支援向量機 (Support Vector Machine, SVM)

1. 一開始我們使用 PCA 係數分析以及最鄰近搜索兩種方式，雖然已經能有不錯的篩選結果，但仍有各自的缺點，例如 PCA 係數分析選取到較多低訊噪比的光譜，而最鄰近搜索選取到較多單峰光譜，交替使用兩種方式篩選能有更好的

篩選效果，但仍有進步空間，因此我們想運用彈性更大、有更多參數可調整的支援向量機，再進一步優化我們的成果。檢查前述四種方式的篩選效果時，我們發現光譜的特徵複雜，可能是特徵光譜的非線性組合，然而使用 PCA 係數分析和最鄰近搜索計算係數的相似度無法考量到特徵光譜的非線性組合。因此我們使用 SVM 作為篩選工具，其中的徑向積函數核（Radial basis function kernel, RBF kernel）可以提供非線性的係數相似度計算方式，更明顯的區分係數間的差異，進而增加篩選的準確度。

2. 使用支援向量機分類雙類星體與非雙類星體時，需要先給予包含每個類星體係數與類別的訓練集與測試集，再使用適當的函數計算訓練集與測試集類星體間的係數相似度，並調整適當的參數，使支援向量機依據類星體的係數相似度與訓練集中每個類星體的類別，來預測訓練集中每個類星體的類別。

我們以等距取樣的方式，將 1923 個已經分類好的類星體分為訓練集與測試集。1153 個類星體中，只有 16 個雙類星體，以此做為訓練集的雙類星體樣本數太少，導致 SVM 在後續進行類別預測時無法判別出雙類星體。因此我們在 1153 個類星體中加入 144 筆來自論文（Smith et al. 2010）的已知雙類星體，作為最終的訓練集（共 1297 個類星體）。我們的訓練集包含 1297 個類星體的係數（coefficients train）與類別（types train）；測試集包含 770 個類星體的係數（coefficients test）與類別（types test），（圖十一）為訓練集與測試集係數取得方式的流程圖：



（圖十一）訓練集與測試集係數取得方式的流程圖。本圖由作者親自製作。

3. 之後選用 RBF 核函數計算係數的相似度，並使 SVM 預測測試集類星體的類別。以下將逐步說明 SVM 計算相似度與預測類別的過程：
 - (1) RBF 會分別計算出訓練集裡每一個類星體係數（coefficients train，以下用 x 表示）與所有測試集類星體係數（coefficients test，以下用 x' 表示）的相似度，RBF 核函數的形式如（式四）：

$$K(x, x') = \exp(-\gamma \|x - x'\|_2^2) \quad (\text{式四})$$

$\|x - x'\|_2^2$: x 與 x' 的平方歐式距離，以訓練集的其中一個類星體 A 與測試集的其中一個類星體 B 為例， x 包含類星體 A 的所有係數(a_1, a_2, \dots, a_{144})， x' 包含類星體 B 的所有係數(b_1, b_2, \dots, b_{144})，故：

$$\|x - x'\|_2^2 = (a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_{144} - b_{144})^2 \quad (\text{式五})$$

γ : 是一個能自由調整的參數，可以控制模型對訓練集的擬合度。

RBF 函數的值隨兩個類星體的係數距離增大而變小，且其值介於 0 與 1 之間，是一種相似度計算方式，對於我們的研究而言，它可以更準確的依據係數度量類星體之間的相似度。

- (2) 預測 770 個測試集類星體的類別：使 SVM 進行類別預測時，我們將前述對 1923 個類星體分類時的第一類與第二類標示為 0，代表非雙類星體，第三類標示為 1，代表雙類星體，之後手動調整不同的 γ ，希望 SVM 選到最多雙類星體與最少非雙類星體。SVM 依據計算出的係數相似度與訓練集類星體的分類，預測 770 個測試集類星體的類別，在 $\gamma = 3.5$ 時有最好的效果，結果如（表十）：

（表十）SVM 的類別預測結果，左欄為 SVM 的預測類別，右欄為 SVM 預測的兩個類別中的類星體實際類別。

| SVM 預測 | | 實際類別 | |
|--------|-----|-------|-----|
| 非雙類星體 | 760 | 非雙類星體 | 759 |
| | | 雙類星體 | 1 |
| 雙類星體 | 10 | 非雙類星體 | 3 |
| | | 雙類星體 | 7 |

770 個測試集類星體中，有 8 個雙類星體，在最佳情況下，SVM 能選到 7 個雙類星體與 1 個非雙類星體，相較前述的四種方式效果明顯較佳，因此我們之後將 SVM 運用在 SDSS DR16 $z < 0.8$ 的所有類星體數據中，希望從該數據集的 76528 個類星體中找出新的雙類星體。

(九) 運用 SVM 從 SDSS DR16 $z < 0.8$ 的所有類星體中找出雙類星體

1. 取得數據：

進行前述分析時，我們已經依 SDSS DR16 星表中的 fiber、MJD、plate 在 SDSS DR17 中下載相應的類星體光譜，共取得 76528 個紅移小於 0.8 的類星體光譜。

我們將用我們認為效果最好的方式（SVM），從其中找出新的雙類星體。

2. 數據前處理：

進行與前述分析相同的數據前處理，將光譜去除紅移、選取特定波長範圍（4900Å~5049Å）、內插、去除連續光譜、發射線強度標準化。之後用與前述分析相同的方式，以去除連續光譜後的光通量計算訊噪比，並以訊噪比最大值 ≥ 5 為保留在數據集內的條件，刪除 22513 個訊噪比較低的光譜，留下 54375 個類星體光譜，作為之後分析的數據集。

3. 運用 SVM 選出雙類星體：

在前述分析中，以 SVM 從 770 個類星體的測試集中選出雙類星體並量化模型的表現後，我們得到篩選效果最好的 SVM 模型，之後我們運用 python 中的功能，直接將該模型套用在 54375 個類星體光譜上，選出雙類星體。

之前我們對 1923 個類星體進分類，其中有 24 個被我們定義為雙類星體，由此可知雙類星體在所有類星體中的比例約為 1%。由於 SVM 在預測類星體的類別時，被其預測為雙類星體的數量與參數 γ 有關，為了使 SVM 選出的雙類星體比例與我們前述的定義盡量一致，我們將 γ 調整為 3.5，此時 SVM 從 54375 個類星體中選出 554 個雙類星體。

肆、研究結果

一、量化各篩選方式的表現

我們已經將用於分析的 1923 個類星體進行分類，因此我們可以對前述每一種方式篩選雙類星體的表現進行量化。因為我們希望能盡量選到最多雙類星體與最少非雙類星體，故我們量化篩選方式的概念為：在該方式選取到幾乎全部的雙類星體時，會選取到多少非雙類星體。而此概念可以用完備性（completeness）與純度（purity）表示：

$$\text{完備性(completeness)} = \frac{\text{選取到的雙類星體數量}}{\text{所有雙類星體的數量}}$$

$$\text{純度(purity)} = \frac{\text{選取到的雙類星體數量}}{\text{選取到的類星體數量}}$$

為了能有一致的比較標準，進行量化時，我們盡可能要求完備性接近 100%，此時純度越高，代表該方式篩選的效果越好，亦即完備性為控制變因，純度為應變變因。

(一) PCA 係數分析、NNS，以及交替使用兩種方式

(表十一) 上表為 PCA 係數分析、NNS、交替使用兩種方式以完備性與純度量化的結果，以及各方式選取到的雙類星體與非雙類星體數量。

| | 完備性 | 純度 | 雙類星體數量 | 非雙類星體數量 |
|--------------|---------|-------|--------|---------|
| PCA 係數分析 | 100.00% | 4.79% | 8 | 159 |
| NNS | 100.00% | 4.79% | 8 | 159 |
| NNS→PCA 係數分析 | 100.00% | 9.30% | 8 | 78 |
| PCA 係數分析→NNS | 100.00% | 9.30% | 8 | 78 |

由(表十一)可以發現，在幾乎都選到所有雙類星體的情況下，PCA 係數分析和最臨近搜索的純度與非雙類星體數量差異不大；但若是能交替使用兩種方式，純度幾乎翻倍成長，非雙類星體的數量也下降到單純使用一種方法的一半，可見得交替使用兩種方式的效果較僅使用一種方法更好。

(二) SVM

(表十二) SVM 以完備性與純度量化的結果，以及該方式選取到的雙類星體與非雙類星體數量。

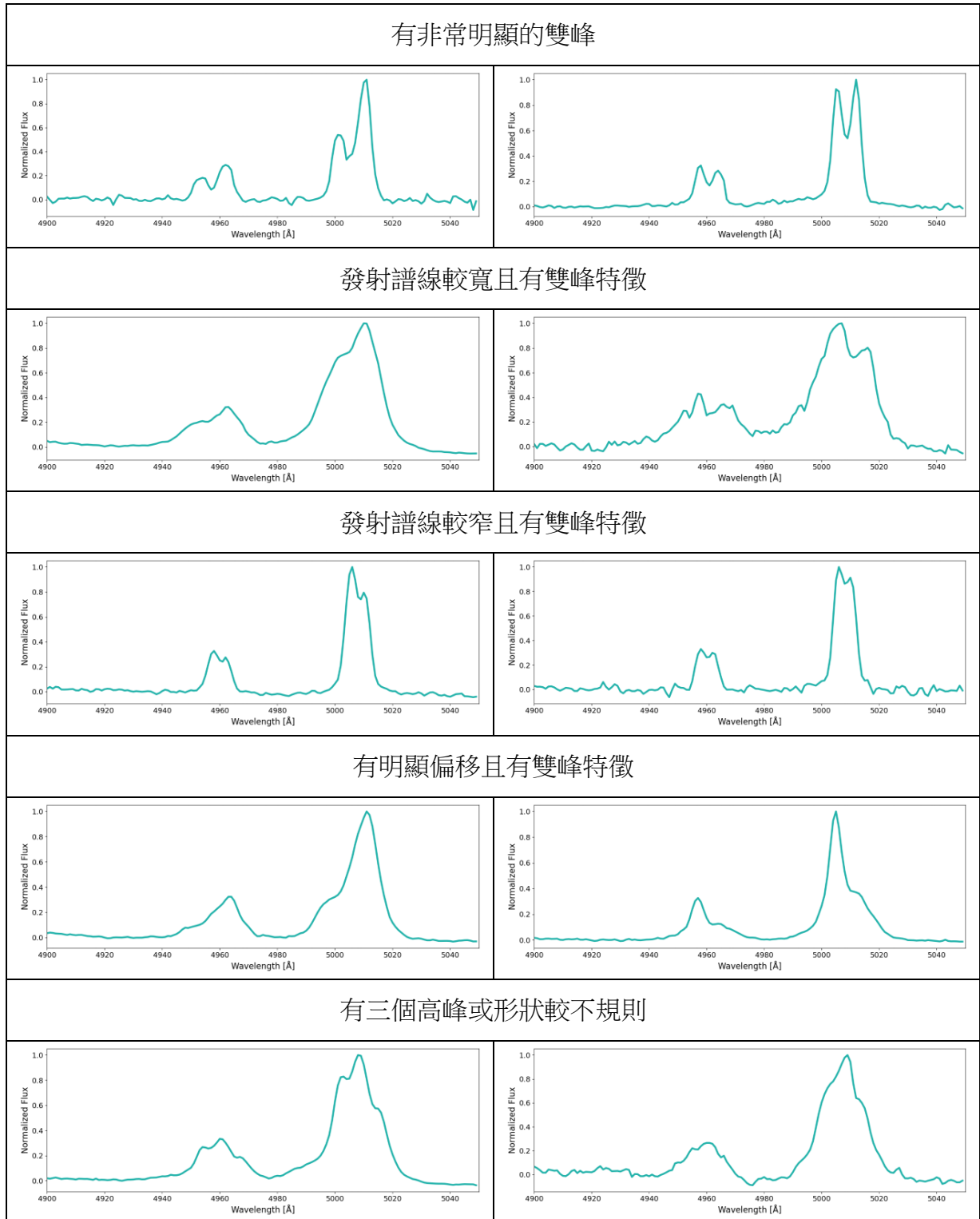
| 完備性 | 純度 | 雙類星體數量 | 非雙類星體數量 |
|--------|--------|--------|---------|
| 87.50% | 70.00% | 7 | 3 |

因 SVM 篩選時的雙類星體樣本數較少，少選取一個雙類星體對完備性影響大，然而其純度 70% 相較前面四種方式有顯著提升。整體而言，SVM 能選到最多雙類星體與最少非雙類星體，而且此方式能自由調整參數，使模型在不同的數據集中有最佳的篩選效果，為本研究中篩選雙類星體最佳方式。

二、找出新的雙類星體

我們運用 SVM 從 SDSS DR16 $z < 0.8$ 的所有類星體中找出 554 個新的雙類星體，檢視被 SVM 找出的候選雙類星體光譜後，我們認為幾乎都是有呈現出雙類星體光譜應有特徵的光譜，除了有明顯雙峰的光譜外，也有雙峰較不明顯，但有明顯偏移、兩側不對稱等足以表示其為雙類星體光譜的特徵，除此之外，我們也找到先前未見，有三個高峰或形狀較不規則的光譜，可見 SVM 已經能夠選出包含多樣特徵的雙類星體光譜，以下為 SVM 找到的雙類星體在 4900Å~5049Å 的光譜的範例：

(表十三) 從 SDSS DR16 $z < 0.8$ 的所有類星體中找出 554 個新的雙類星體光譜的範例。表中圖片由作者親自製作。



三、雙類星體色指數分析

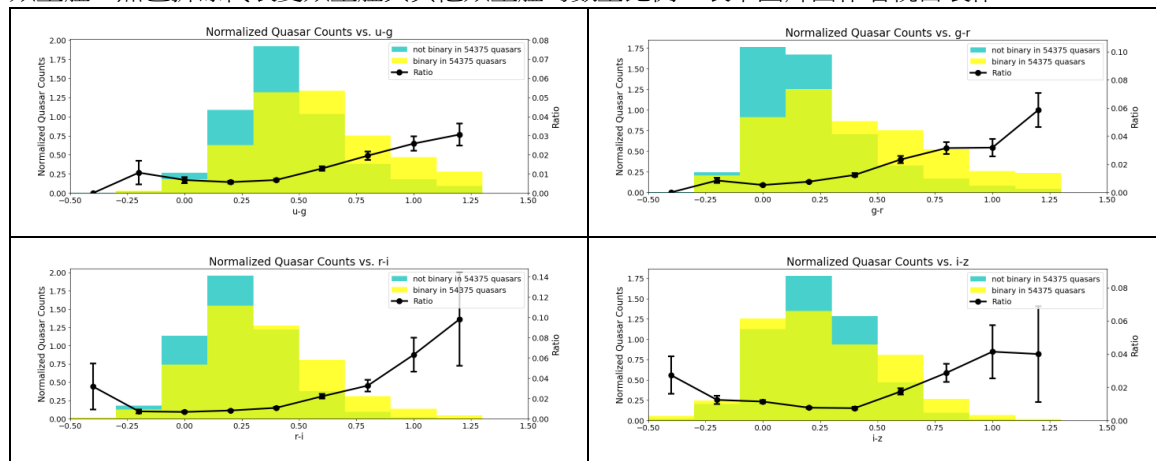
(一) 為了分析雙類星體與一般類星體在顏色上的性質差異，我們使用 SDSS 的 PSFMAG 資料來計算類星體的色指數。SDSS 觀測到的類星體點光源會在 CCD 上呈現一個高斯分布的亮斑，SDSS 利用點擴散函數 (PSF) 計算出類星體的總光通量 (flux, 在 SDSS 資料中以 PSFFLUX 表示)，之後利用星等公式將總光通量轉換成視星等

(magnitude, 在 SDSS 資料中以 PSFMAG 表示)。SDSS 使用 u、g、r、i、z 五個濾光片測量類星體光譜分別在 u (紫外線, 3543Å)、g (綠光, 4770Å)、r (紅光, 6231Å)、i (近紅外線, 7625Å)、z (紅外線, 9134Å) 五個波段的總光通量和視星等, 因此 PSFMAG 在五個波段都有對應的值。

(二) 在天文學上, 發光天體的顏色通常用相鄰兩個波段的星等相減來表示。我們分別將 PSFMAG 的 u 星等減 g 星等, g 星等減 r 星等, r 星等減 i 星等, i 星等減 z 星等, 得到 u-g、g-r、r-i、i-z 色指數。因為後面的濾光片觀測較長的波長, 且視星等量值越小代表類星體在該波段的亮度越大, 因此色指數越大代表類星體的顏色越紅, 反之則越藍。

(三) 我們以直方圖分析雙類星體與一般類星體在四個色指數中不同的色指數量值的數量分布。因為雙類星體的數量只有一般類星體的 1%, 所以我們繪製直方圖時以標準化的數量呈現, 使雙類星體與一般類星體直方的面積總和皆為 1。在直方圖中, 黃色代表雙類星體的標準化數量, 藍色代表一般類星體的標準化數量, 黑線代表雙類星體與一般類星體在不同的色指數量值的數量比例。

(表十四) 以上四張圖為我們找出的 554 個新雙類星體與數據集中的其他 53821 個類星體的標準化數量對應 u-g、g-r、r-i、i-z 色指數的直方圖, 黃色代表 554 個新雙類星體, 藍色代表其他 53821 個類星體, 黑色折線代表雙類星體與其他類星體的數量比例。表中圖片由作者親自製作。



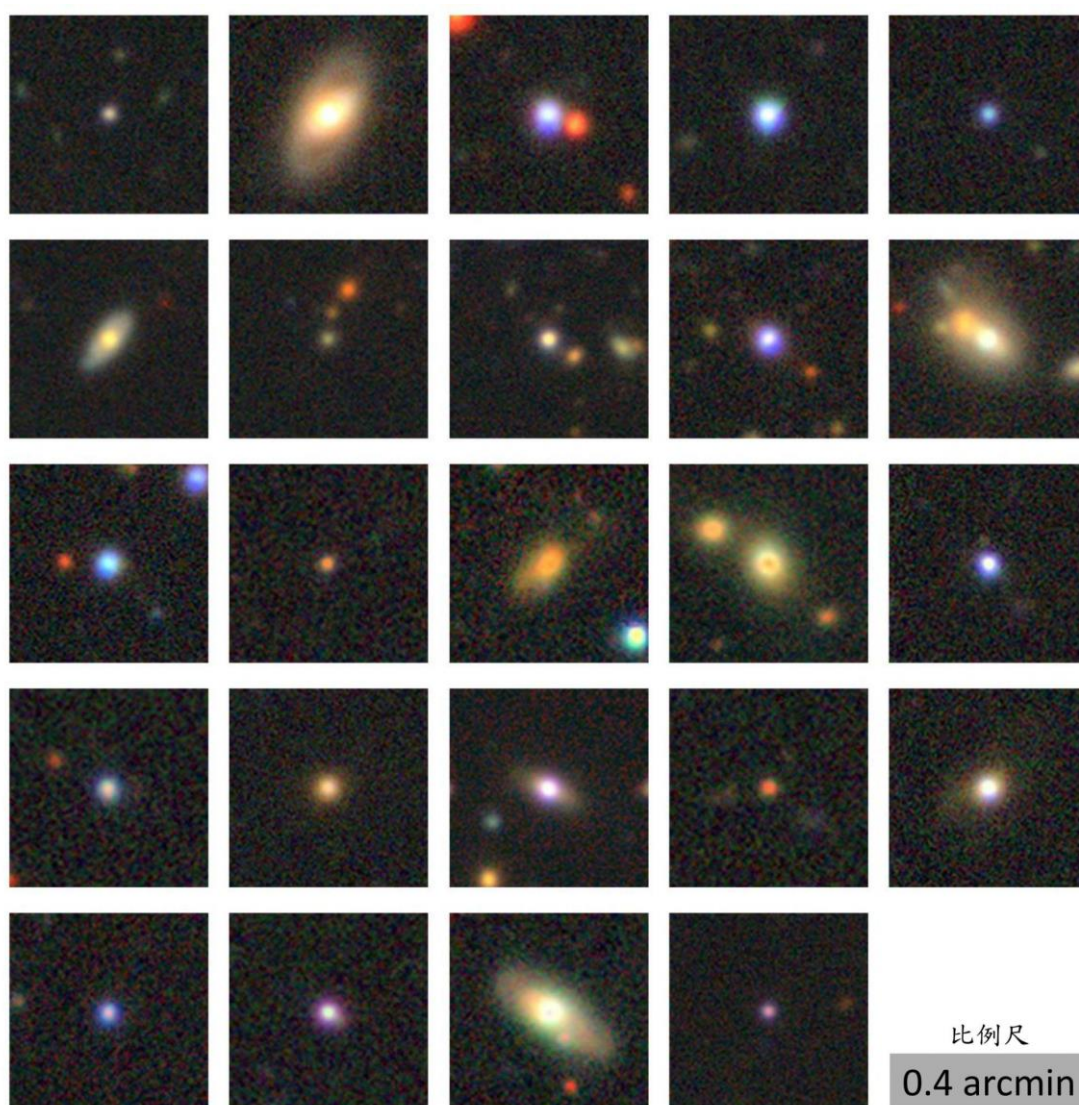
從 (表十四) 中, 我們觀察到在每個色指數中, 色指數量值越大, 雙類星體的比例越高, 所以雙類星體的顏色相較一般類星體而言有偏紅的趨勢, 而我們認為雙類星體的顏色較一般類星體紅與雙類星體的性質有關。

伍、討論

一、雙類星體影像與互繞情形分析

(一) 雙類星體影像

我們希望從類星體的影像中尋找雙類星體有別於一般類星體的物理性質，因此我們參考了 DESI Legacy Surveys 的可見光影像、HSC SSP 的可見光影像、VLASS 的無線電波影像。因為 DESI Legacy Surveys 在天球上的觀測範圍最廣，幾乎能對應到所有我們類星體數據集中的類星體影像，且可見光影像可以直觀的呈現類星體外觀。因此我們先分析 1923 個類星體中，我們定義的 24 個雙類星體在 DESI Legacy Surveys 中的影像，以下為 24 個雙類星體在 DESI Legacy Surveys 中的影像：



(圖十二) 此圖為 24 個雙類星體在 DESI Legacy Surveys 中的可見光影像，本圖引自 Cutout Image List Tool for DESI Legacy Imaging Survey。

從(圖十二)中我們發現，有七個影像中的類星體有類似塵埃的模糊影像分布在其

周圍，有兩個影像中呈現兩個非常靠近的類星體光源，而大多數的影像為單一的類星體點光源，並沒有呈現出明顯的兩個鄰近的光源。

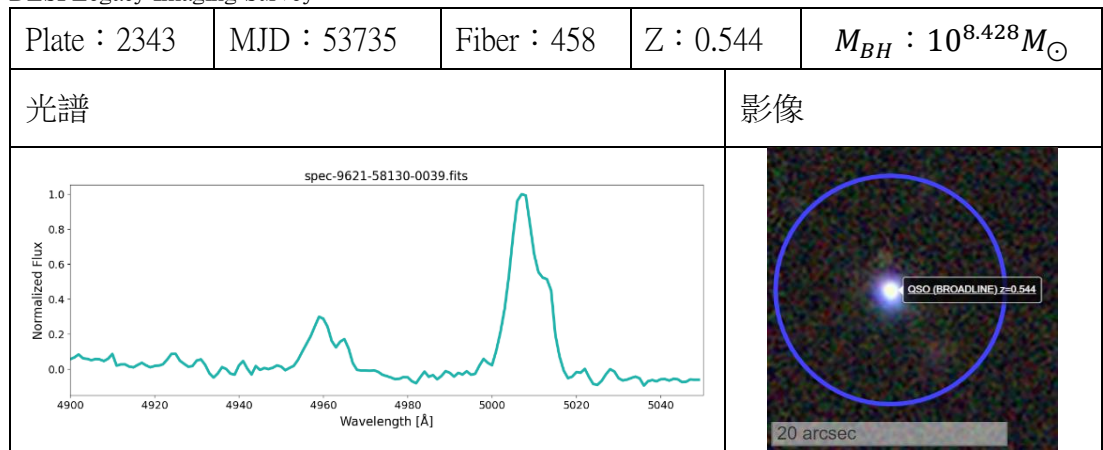
(二) 雙類星體互繞情形分析

觀察影像後，我們分析兩個類星體受彼此的重力場互繞時的距離，推測互繞的雙類星體無法在影像上呈現兩個光源的可能性，以下為我們的假設、推論與計算方式：

1. 選擇雙類星體並取得計算所需數據：

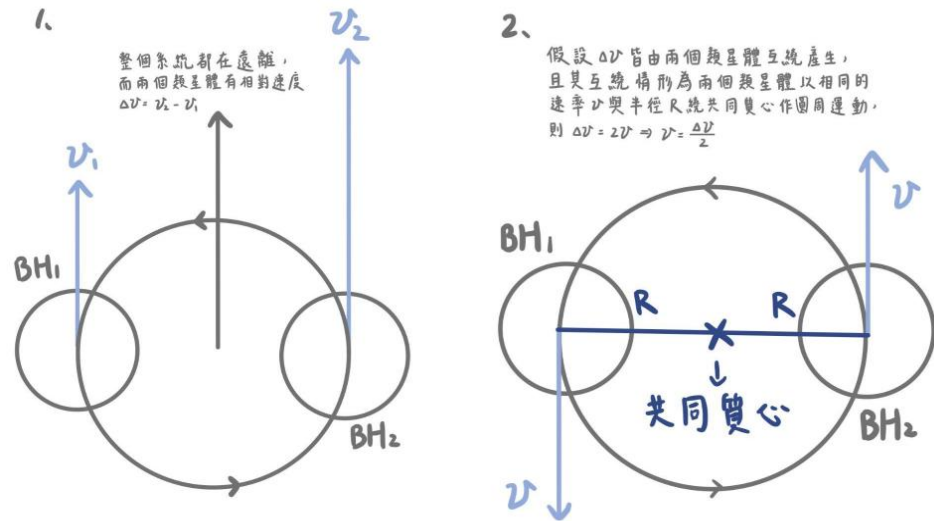
在 24 個雙類星體中，第 15 個類星體的光譜有明顯雙峰特徵，而其影像僅呈現單一的類星體光源，因此我們分析其雙類星體互繞時彼此之間的距離，並以該雙類星體系統的紅移推算在該紅移下，這樣的間距是否能被望遠鏡辨別。類星體互繞時彼此的距離與其相對速度與質量有關，我們可以從類星體雙峰發射譜線的波長差異得到類星體的相對速度，而 Shen et al. 2011 的研究提供了 SDSS DR7 部分類星體的質量，我們使用其提供的上述第 15 個雙類星體的質量。

（表十五）左圖為我們分析的雙類星體標準化後的光譜，此圖由作者親自製作，右圖為我們分析的雙類星體在 DESI Legacy Surveys 中的可見光影像，此圖引自 Cutout Image List Tool for DESI Legacy Imaging Survey。



2. 假設：

- (1) 兩個互繞類星體質量皆設為 Shen et al. 2011 提供的 $10^{8.428} M_{\odot}$
- (2) 因宇宙膨脹造成的紅移可能影響兩個類星體的相對速度，故先以尚未去除紅移的原始光譜中，[O III] 雙峰發射譜線的兩個發射譜線波長，分別計算出兩個類星體在宇宙中的速度，而互繞的雙類星體在宇宙中的位置幾乎相同，因此彼此之間的速度差應為兩者因互繞所產生的相對速度。
- (3) 為簡化計算，假設一理想化的情況，兩個類星體以相同半徑繞共同質心做圓周運動，互繞的相對速度即為我們在光譜上由雙峰發射譜線波長差異所觀察到的相對速度。其互繞情形示意圖如（圖十三）。



(圖十三) 左圖為雙類星體在宇宙中的運動情形示意圖，右圖為單純考慮該雙類星體系統內的互繞情形示意圖，本圖由作者親自製作。

v_1 、 v_2 ：兩個類星體分別在宇宙中的速度， $\Delta v = v_2 - v_1$

v ：兩個類星體互繞的系統內，兩個類星體繞共同質心作圓周運動的速率

R ：兩個類星體互繞的系統內，兩個類星體繞共同質心作圓周運動的半徑

3. 計算：

(1) 以雙峰光譜波長差異計算相對速度：

首先標示出原始光譜中，雙峰發射譜線的兩個高峰所在波長，並以 OIII 發射譜線的雙峰波長差，計算兩個類星體的紅移差。



(圖十四) 此圖為我們分析的雙類星體光譜，本圖引自 SDSS Science Archive Server。

(圖十四) 為 SDSS DR16 原始光譜的 [O III] 發射譜線，我們以圖中的兩條黃線標示 [O III] 5007 發射譜線波長 λ_1 、 λ_2 ， λ_1 為 BH_1 的 [O III] 5007 發射譜線波長 (圖中波長較短的發射線)， λ_2 為 BH_2 的 [O III] 5007 發射譜線波長 (圖中波長較長的發射線)，依據我們的判斷， $\lambda_1 = 7729\text{\AA}$ ， $\lambda_2 = 7738\text{\AA}$ 。依據紅移公式，我們可以計算出 BH_1 的紅移 Z_1 (式七)、 BH_2 的紅移 Z_2 (式八)，再依 (式六) 得兩個類星體在宇宙中的相對速度 Δv 。

$$\Delta v = \frac{Z_1 - Z_2}{(1+Z)} \cdot c \quad (\text{式六})$$

$$Z_1 = \frac{7729}{5007} - 1 \quad (\text{式七})$$

$$Z_2 = \frac{7738}{5007} - 1 \quad (\text{式八})$$

$$\Delta v = 349.35 \text{ km/s}$$

我們計算出 $\Delta v = 349.35 \text{ km/s}$ ，前面的假設為兩個類星體質量相同，故互繞的速率應相同，因此我們考慮兩個類星體繞共同質心做圓周運動的情形時，以 $v = \frac{\Delta v}{2} \approx 174.675 \text{ km/s}$ 當作兩者受彼此重力影響做圓周運動的速率。

(2) 以萬有引力與圓周運動的關係推算兩個類星體的距離：

兩個類星體以彼此的萬有引力作為向心力，以速率 v 與半徑 R 作圓周運動。已知類星體質量 $M_{BH} = 10^{8.428} M_{\odot}$ (M_{\odot} 約為 $2 \cdot 10^{30} \text{ kg}$)，可以推算出兩者做圓周運動的半徑 R 如 (式九)：

$$F_c = \frac{G \cdot M_{BH}^2}{(2R)^2} = \frac{M_{BH} \cdot v^2}{R} \Rightarrow R = \frac{G \cdot M_{BH}}{4 \cdot v^2} \quad (\text{式九})$$

(註： G 為萬有引力常數，約為 $6.6743 \cdot 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$)






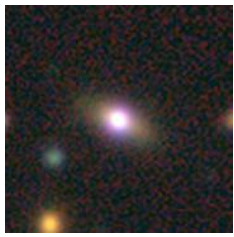

我們計算出 $R \approx 9.978 \text{ pc}$ ，故兩個類星體間距為 $2R = 19.956 \text{ pc}$ 。已知該雙類星體系統的紅移為 0.5436，依據 Wright 2006 提供的方程式，我們可以得知在該紅移的距離與視差關係為 6.444 kpc/arcsec ，亦即觀測上的一角秒在宇宙中的實際距離為 6.444 kpc ，由此可知前述計算的雙類星體間距 19.956 pc 在觀測上約為 $3.097 \cdot 10^{-3} \text{ arcsec}$ 。而這樣的視差對目前地面觀測技術而言幾乎不可分辨。

5. 其它雙類星體の間距：

我們將上述計算方式整理成 (式十) 並在程式中執行，計算出 24 個雙類星體中，Shen et al. 2011 有提供質量資料的 7 個雙類星體間距 $2R(\text{pc})$ ，再以 Wright 2006 提供的方式計算出兩個類星體的視差 p'' ，結果如 (表十六)：

$$2R(\text{pc}) = \frac{4 \cdot G \cdot 10^{(\text{LOG} BH + 14)}}{3 \cdot [c \cdot (\frac{\lambda_1 - \lambda_2}{5007(1+Z)})]^2} \quad (\text{式十})$$

(表十六) 24 個雙類星體中 Shen et al. 2011 有提供質量資料的 7 個雙類星間距與視差分析結果，表中的圖引自 Cutout Image List Tool for DESI Legacy Imaging Survey。

| | | | | |
|----------|---|---|--|---|
| 影像 |  |  |  |  |
| $2R(pc)$ | 181.026 | 3.874 | 33.904 | 161.460 |
| $p(")$ | $3.114 \cdot 10^{-2}$ | $1.169 \cdot 10^{-3}$ | $5.297 \cdot 10^{-3}$ | $4.817 \cdot 10^{-2}$ |
| 影像 |  |  |  | |
| $2R(pc)$ | 19.956 | 23.071 | 2.891 | |
| $p(")$ | $3.097 \cdot 10^{-3}$ | $6.166 \cdot 10^{-3}$ | $1.123 \cdot 10^{-3}$ | |

(表十六) 中的 7 個雙類星體中，我們推算出的視差最大為 $4.817 \cdot 10^{-2}(")$ ，最小為 $1.169 \cdot 10^{-3}(")$ ，而我們觀察的影像來源 DESI Legacy Surveys 的解析力約為 $1.01(") \sim 1.61(")$ (Arjun Dey et al. 2019)，因此我們認為雙類星體在 DESI Legacy Surveys 的影像中幾乎不可能被辨別，應只能呈現出一個點光源。

陸、結論

一、自動化且系統性找尋雙類星體的方式

我們開發出自動化且系統性找出雙類星體的方式，利用支援向量機 (SVM) 判斷雙峰光譜的特徵並選出雙類星體，因為其核函數提供了非線性的特徵相似度計算方式，亦可精準的調整參數，從量化結果顯示，SVM 相較本研究中的 PCA 係數分析、最鄰近搜索 (NNS) 等四種方式效果明顯較佳，在幾乎選到所有雙類星體時亦有相當高的純度，是我們認為能找出雙類星體的最佳方式。

我們已將此方式應用在 SDSS DR16 $z < 0.8$ 的所有類星體數據中，從 7.6 萬個類星體中找出 554 個新的雙類星體，並對其作進一步的物理性質分析。

二、雙類星體的物理性質

過去對雙類星體的了解不多，僅能以[O III]發射譜線的雙峰特徵作為判斷雙類星體的指標，然而雙峰光譜特徵變化複雜，且目前的觀測數據已達到數十萬筆，過去人工檢視光譜雙峰特徵的方式難以獲得足夠的資料以分析雙類星體的物理性質。在我們運用 SVM 系統性的找出雙類星體後，分別對其顏色、影像、互繞情形進行物理性質分析。

(一) 色指數：

我們發現雙類星體在 u-g、g-r、r-i、i-z 色指數中，顏色皆有較其他類星體整體偏紅的趨勢，可能是雙類星體有別於一般類星體的特徵，未來我們將進一步探討顏色偏紅現象的可能成因。

(二) 互繞情形：

在我們已經分類好的 24 個雙類星體的可見光影像中，我們看到多數皆呈現一個點光源，難以從影像上判斷雙類星體與單類星體，因此我們利用現有的雙峰發射譜線、類星體質量資料，推測其中 7 個雙類星體互繞時的間距，再由其紅移得知其在觀測影像上的視差。在我們的推測結果中，雙類星體互繞時的間距約為 $10\sim 1000pc$ ，觀測影像上的視差約為 $10^{-3}\sim 10^{-1}arcsec$ ，而我們使用的 DESI Legacy Surveys 的可見光影像解析角約為 $1.01\sim 1.61arcsec$ ，由此可知我們推算出的雙類星體視差角皆小於可見光影像能辨別的範圍。

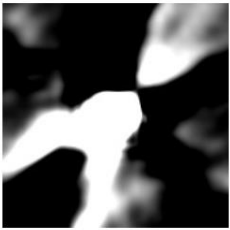

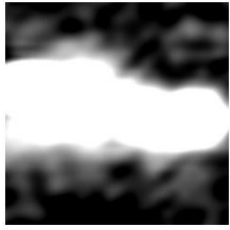
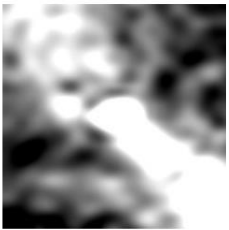
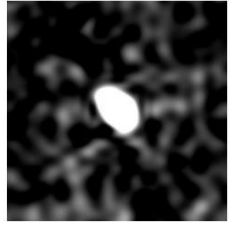
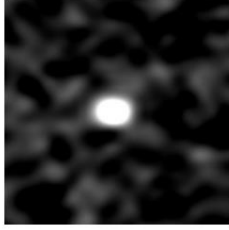
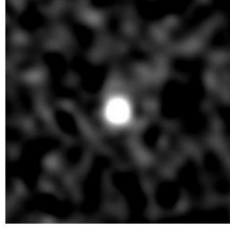
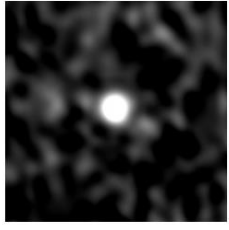
然而在我們推算的假設中，兩個類星體質量皆設為 Shen et al. 2011 提供的類星體質量，且雙類星體互繞的相對速度直接由雙峰譜線的波長差異計算。然而實際情況中，雙類星體互繞呈現的相對速度與我們觀測的角度有關，類星體的質量也難以被準確的計算，因此我們的推算與實際情形能有落差。即便如此，我們能可從數量級看出，雙類星體受彼此重力牽制而互繞時會非常接近彼此。

三、未來展望

(一) 類星體在其他波段的影像

在研究過程中，我們觀察到有部分的雙類星體在 VLASS 的無線電波影像中有明顯的訊號，且有不同的特徵，以下為我們歸納出各種無線電波影像的範例：

(表十七) 544 個新的雙類星體的無線電波影像範例。表中的圖引自 Cutout Image List Tool for DESI Legacy Imaging Survey。

| 輻射狀發出 | | 形狀不規則 | |
|---|---|--|---|
|  |  |  |  |
| 橢圓形光源 | | 圓形光源 | |
|  |  |  |  |

未來我們將探討無線電波源對雙峰發射譜線是否有影響，以及無線電波源在雙類星體與一般類星體影像中出現的數量比例，因為類星體常見的噴流會在無線電波段有明顯的影像，而噴流很可能是產生雙峰發射譜線的原因之一，Smith et al. 2010 也認為無線電波源與雙峰發射譜線的特徵有關。運用我們的方式能系統性的了解無線電波源在雙類星體與一般類星體中出現的數量與型態，便能進一步了解無線電波源與雙類星體的關聯。

(二) 更高解析力的影像

我們需要更高解析力的可見光、無線電波影像，使我們能夠確認找到的雙類星體雙峰發射譜線是否確實為雙類星體互繞所造成，或是被噴流干擾的雙峰發射譜線。

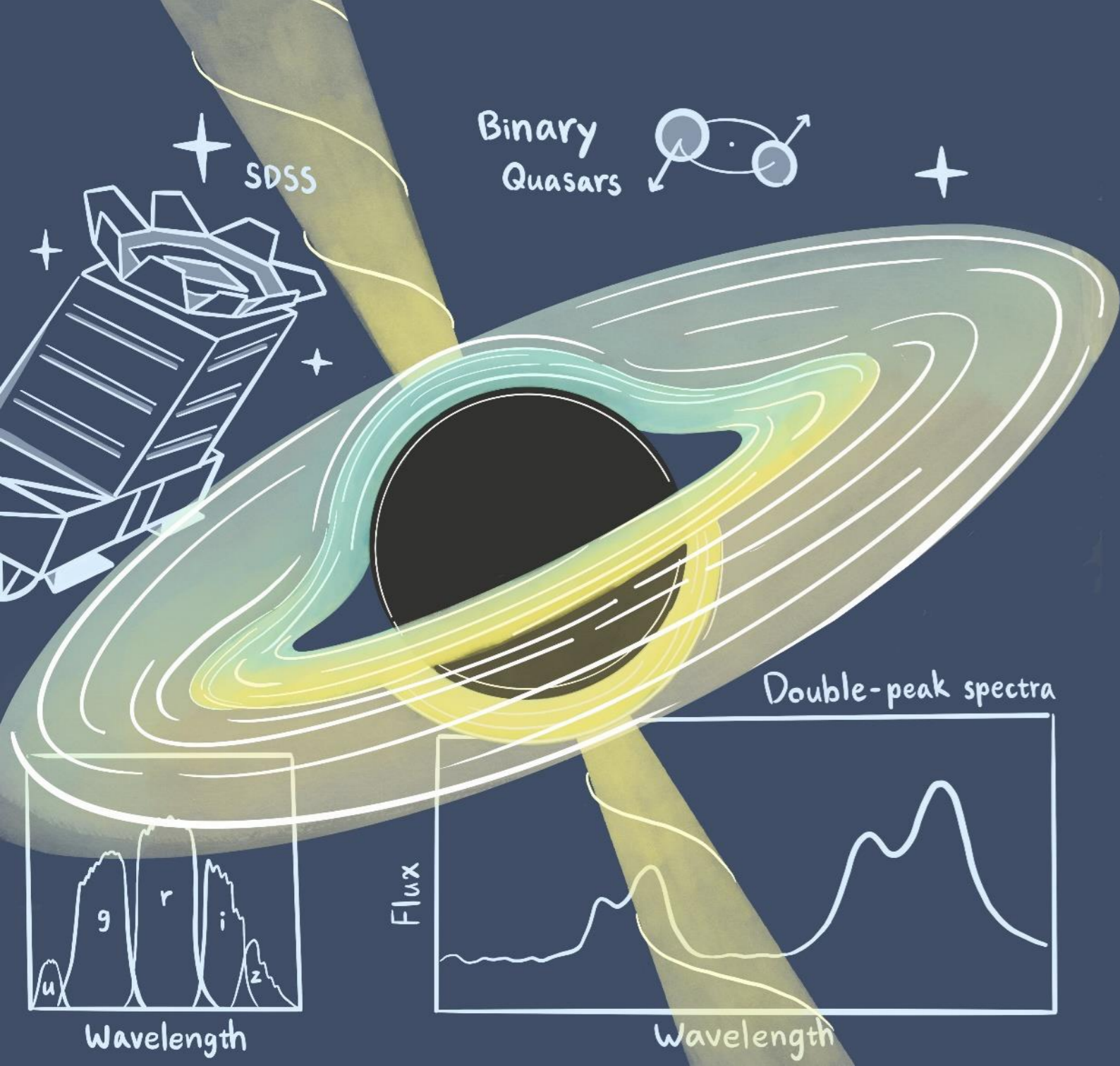
柒、參考文獻資料

- 一、Smith, K. L., Shields, G. A., Bonning, E. W., et al. 2010, *apj*, 716, 866. doi : 10.1088/0004-637X/716/1/866
- 二、Suzuki, N. 2006, *ApJS*, 163, 110. doi : 10.1086/499272
- 三、Lyke, B. W., Higley, A. N., McLane, J. N., et al. 2020, *ApJS*, 250, 8. doi : 10.3847/1538-4365/aba623
- 四、Wright, E. L., 2006, *PASP*, 118, 1711. doi : 10.1086/510102
- 五、Arjun Dey, et al. 2019, *AJ*, 157, 168. doi : 10.3847/1538-3881/ab089d
- 六、Shen, et al. 2011, *ApJS*, 194, 45. doi : arxiv.org/abs/1006.5178v2

【評語】 051803

此研究旨在發展一種自動化方法，系統性地分析類星體光譜數據，以辨識雙類星體並探究其性質。雙類星體是天文研究中的重要課題，因為它們是星系合併的產物，也是重力波的來源之一。因此，研究雙類星體的性質能使我們更好地了解星系合併的機制。此研究利用主成分分析（PCA）來確認雙類星體的特徵光譜，並結合 PCA 係數分析、最鄰近搜索和支援向量機（SVM）等數值方法，成功篩選出約 500 個新的候選雙類星體。研究結果顯示，這種方法能有效率地辨識出足夠數量的雙類星體，使得未來進一步分析雙類星體的形成環境成為可能。然，作者對於其他造成光譜雙峰的成因缺乏討論，若能將其他因素排除，將是一個天文上的重要成果。

作品簡報



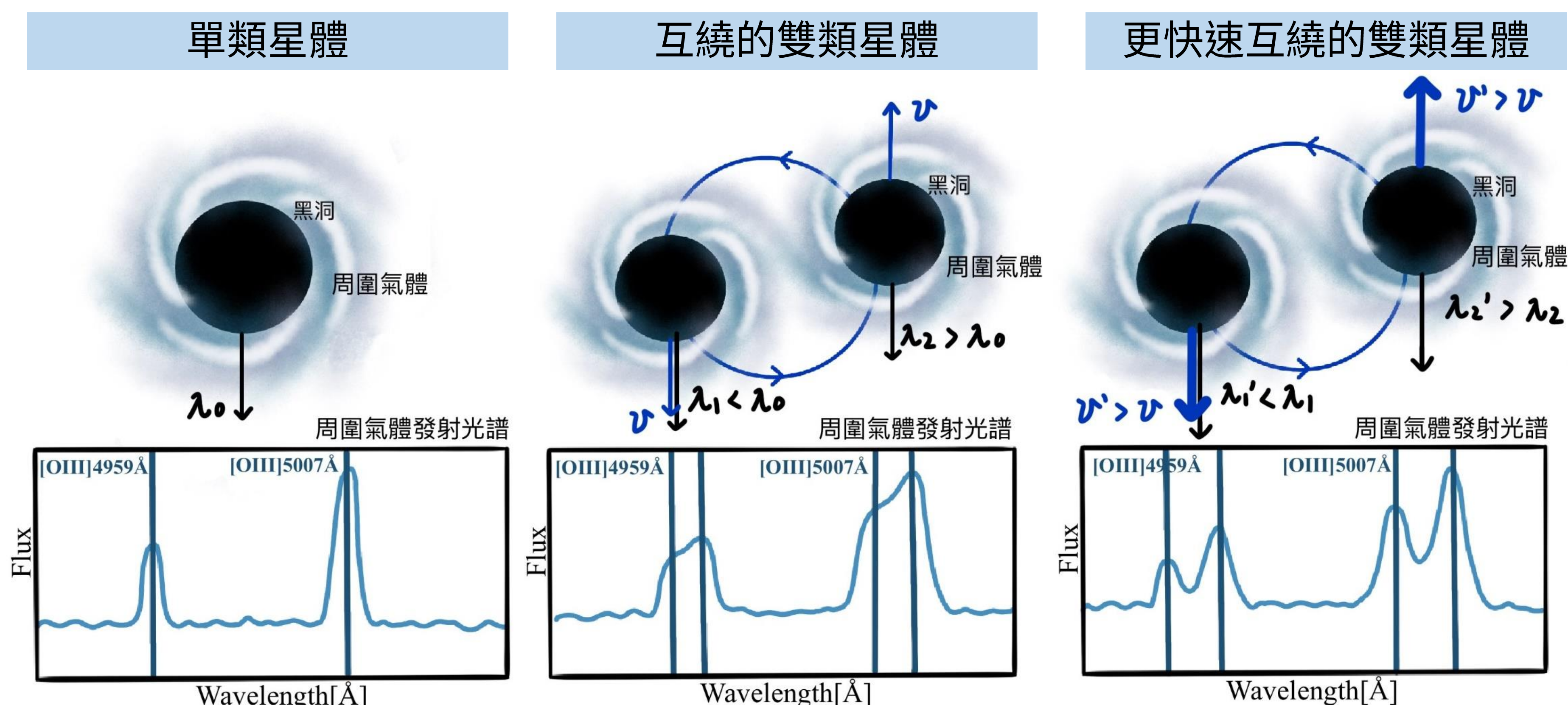
運用主成分分析 探尋宇宙中的互繞雙類星體

摘要

探尋互繞雙類星體是天文研究中重要的課題，互繞雙類星體是星系合併的產物，同時也是產生重力波的來源之一，研究其性質能使我們了解星系合併的機制，而光譜是最直接研究的面向。雖然目前已經有大量天體數據，但相關的研究並不多，雙類星體的性質研究也尚不成熟。在本研究中，我們發展出能從類星體光譜數據中尋找雙類星體的方法，並進一步探究其性質。本研究運用機器學習中的主成分分析擷取雙類星體的特徵光譜，結合支援向量機，精準的篩選雙類星體。我們將此方法運用在 Sloan Digital Sky Surveys 所提供的類星體光譜資料庫，找出並匯集約500個新發現的候選雙類星體，並使用此數據進一步探索目前未被發現的雙類星體物理性質。

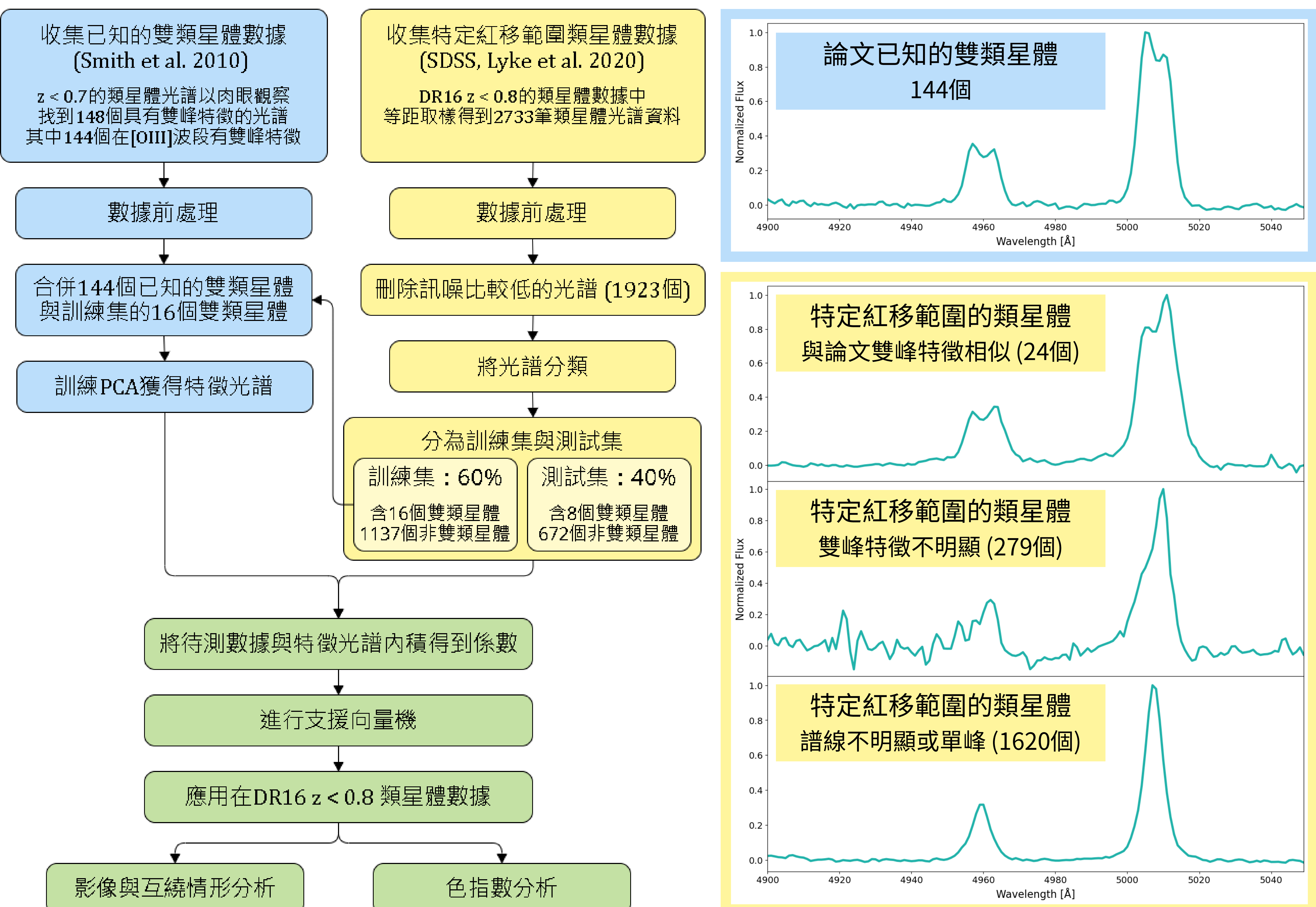
研究目的

星系中央有超大質量黑洞，當其吸積氣體時會發出發射光譜，這種天體被稱為類星體，當兩個類星體互繞時會產生相對速度，在光譜產生雙峰特徵。雙類星體 (Binary Quasars) 尋找和研究是較新的領域，在天文物理上可以提供關於星系的演化、重力波等的資訊，不過目前對於雙類星體的研究不多，也鮮少使用系統性數據分析的方式了解雙類星體在整個宇宙中的數量和分布，因此難以對其做出更完整的性質分析。過去的研究 (e.g., Smith et al. 2010) 使用肉眼辨識雙峰光譜，但現在光譜數據量非常大 (e.g., SDSS)，卻沒有系統性的方式偵測雙類星體，因此我們目標建立一個自動化且系統性尋找雙類星體的方式，並研究其有別於一般類星體的物理性質。



▲ 圖一：雙峰特徵光譜原理示意圖

研究架構

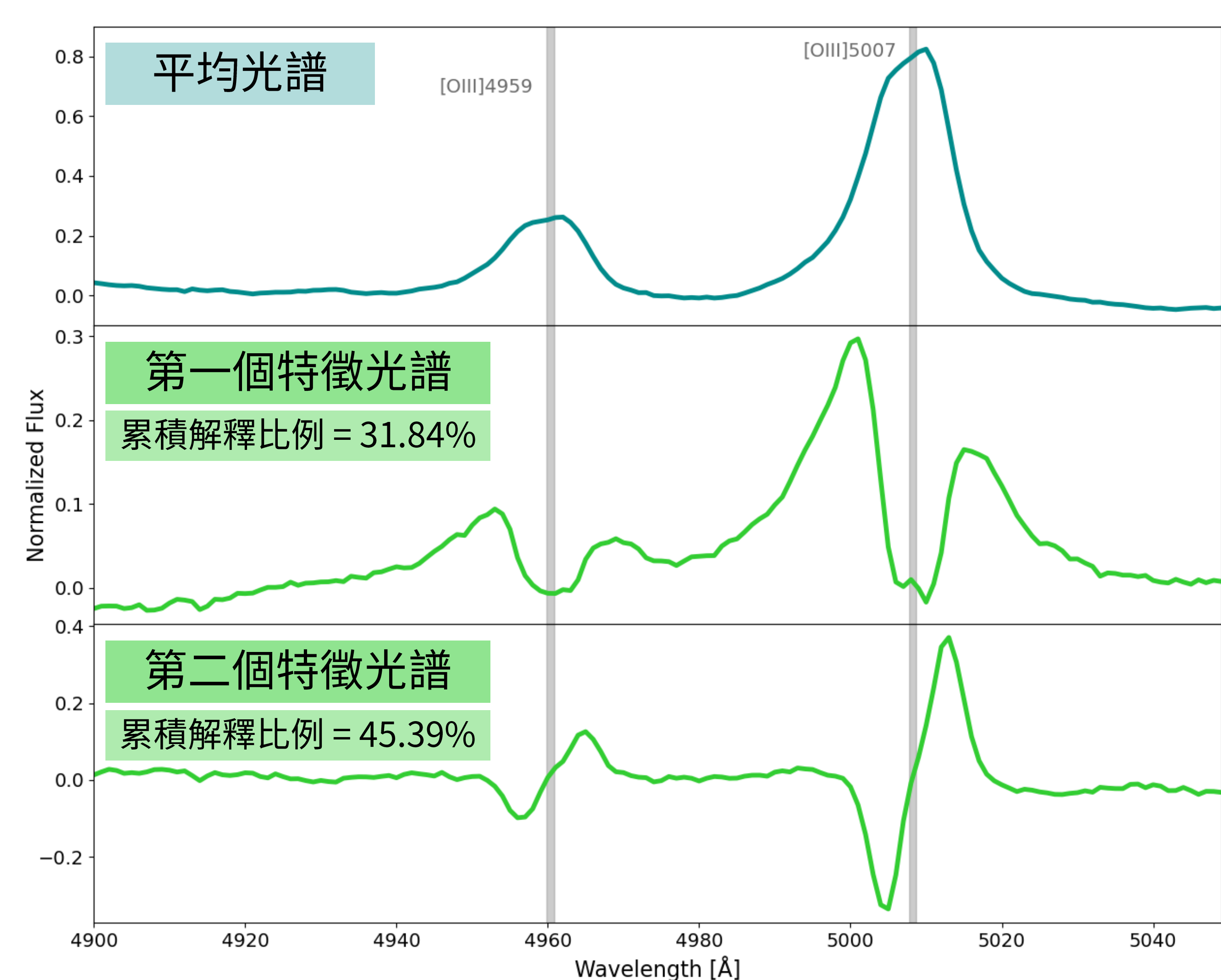


數據分析

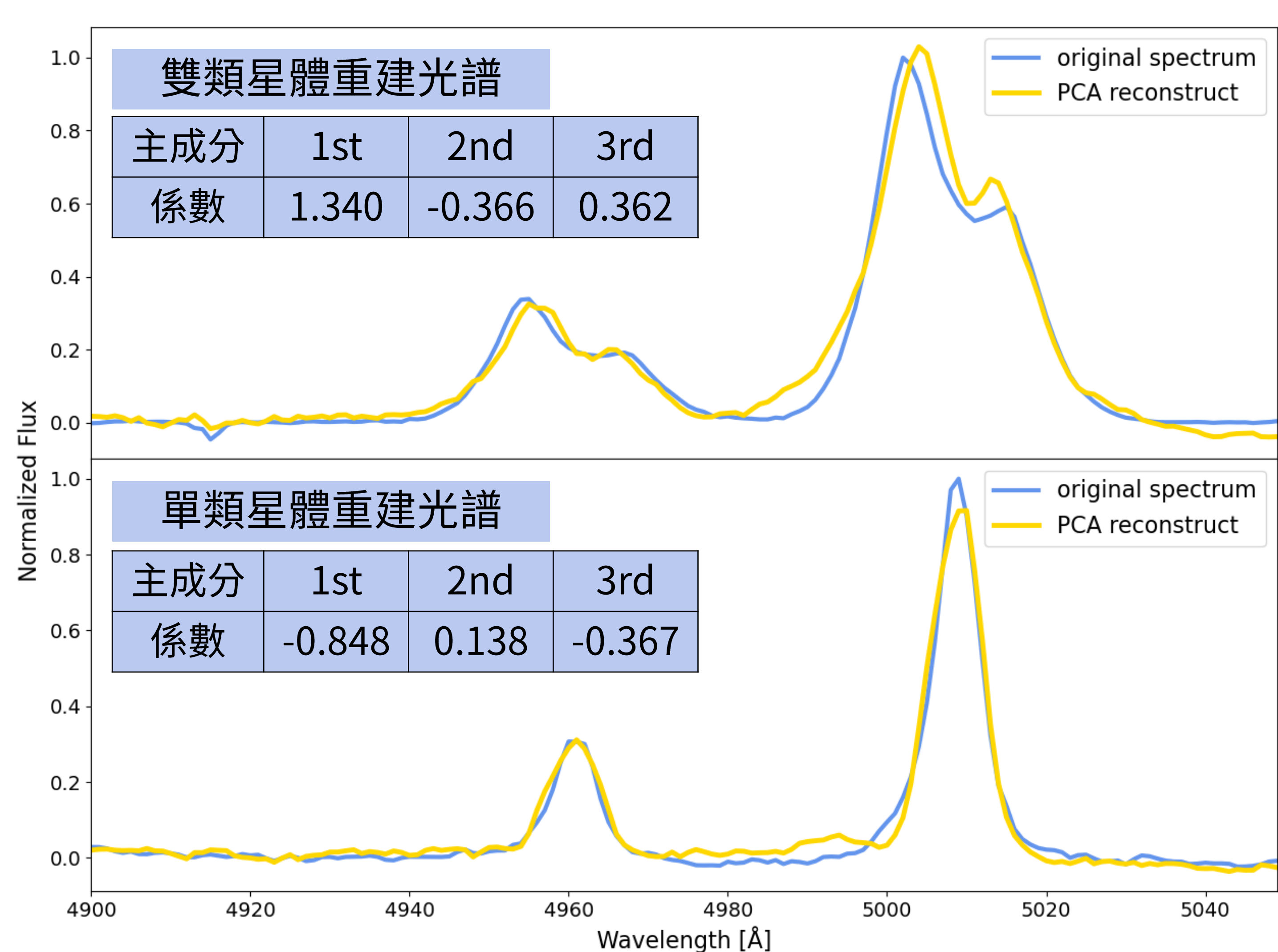
一、主成分分析 (PCA)：可以簡化數據集，同時保留最大特徵的統計方法

1. 特徵光譜 (主成分)：特徵光譜是原始光譜中的貢獻和特定變化模式，可用來辨識光譜裡的特徵峰，或比較光譜間的相似性與差異性 (e.g., Suzuki 2006)。
2. 重建光譜：重建光譜是平均光譜加上不同表現程度的特徵光譜，每個特徵光譜的表現程度由係數決定。

$$\text{第}n\text{個重建光譜} = \text{平均光譜} + \sum_{k=0}^{150} \text{第}k\text{個特徵光譜} \times \text{第}n\text{個光譜第}k\text{個係數}$$

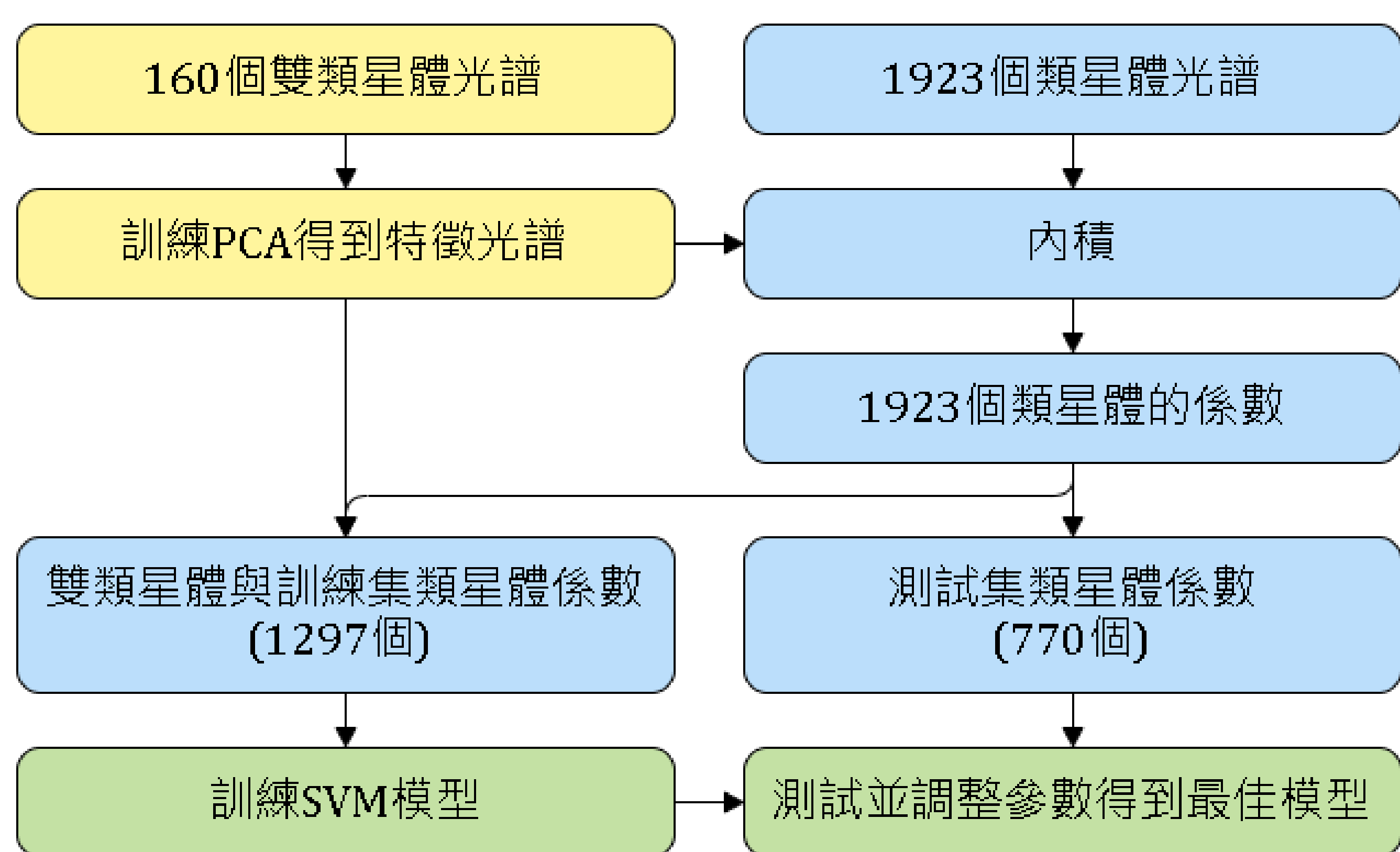


▲ 圖二：進行PCA後得到的平均光譜與前兩個特徵光譜



▲ 圖三：單峰與雙峰前三個主成分的重建光譜

二、支援向量機 (SVM)：以非線性的方式準確計算類星體係數相似度，並依係數相似度對類星體進行分類



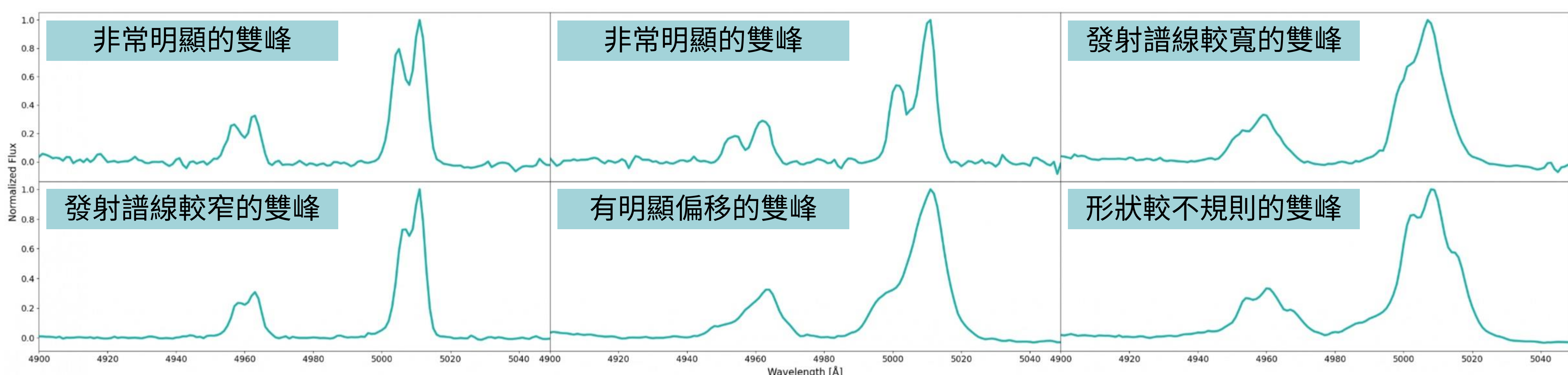
將論文的144個雙類星體與1923個類星體的訓練集中分類出的16個雙類星體合併成160個雙類星體，訓練PCA得到特徵光譜，再將其與1923個類星體內積後得到係數。其中60%與160個雙類星體合併為訓練集，40%為測試集，先以訓練集訓練SVM模型再以測試集測試，量化模型表現並調整參數獲得最佳模型。

| | |
|---|-------|
| 完備性 (completeness) = $\frac{\text{選取到的雙類星體數量}}{\text{所有雙類星體的數量}}$ | 87.5% |
| 純度 (purity) = $\frac{\text{選取到的雙類星體數量}}{\text{選取到的類星體數量}}$ | 70.0% |

▲ 表一：支援向量機最佳模型效能的量化結果

應用與性質分析

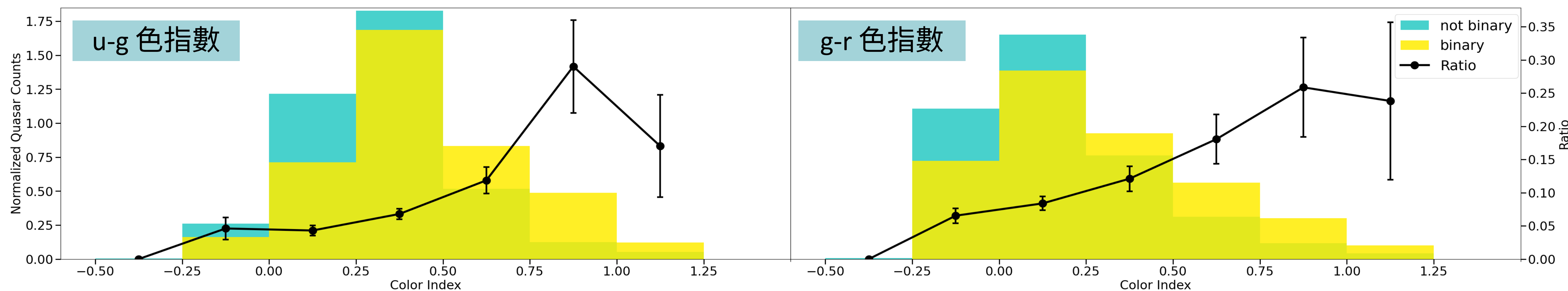
一、找出新的雙類星體：我們將訊噪比、主成分分析結合支援向量機的方法運用在SDSS DR16 $z < 0.8$ 所有類星體光譜 (共約76000筆)，並從中找出554個雙類星體，其中524個為新的雙類星體。



▲ 圖四：524個新的雙類星體光譜範例

二、雙類星體可見光與無線電波分析：分析雙類星體的可見光與無線電波有別於一般類星體的特徵。

1. 控制變因：比較類星體的可見光與無線電波，須控制相同紅移與質量，故我們使用 Shen et al. 2011提供質量數據的200個雙類星體，在每個雙類星體紅移 (z) ± 0.08 、質量 ($\log M_{\odot}$) ± 0.4 範圍內取10個一般類星體，共取2000個類星體為控制組。
2. 色指數分析：Glikman et al. 2015研究約10個紅色類星體，認為星系合併特定階段與類星體偏紅有關。而我們在200個雙類星體中找到雙類星體顏色較一般類星體有偏紅的趨勢，而雙類星體是星系合併的產物，故星系合併過程可能使雙類星體偏紅。



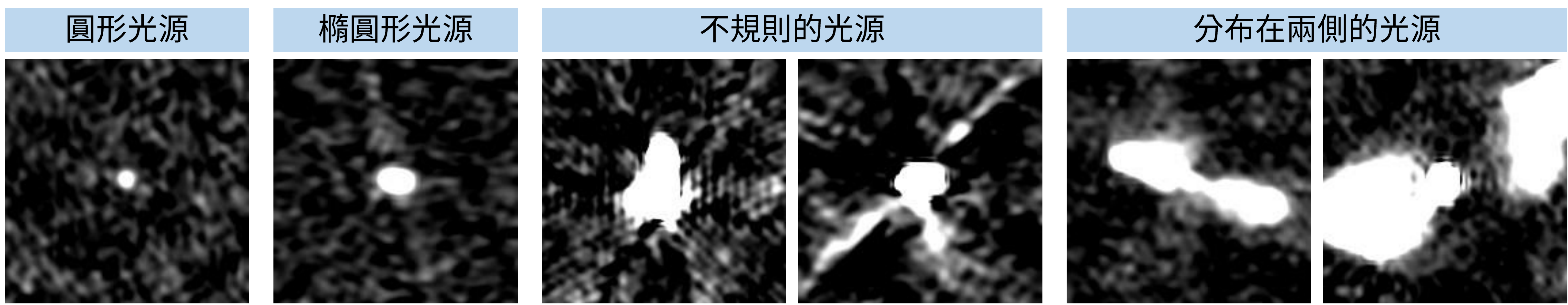
▲ 圖五：雙類星體 (黃色) 與一般類星體 (藍綠色) 的u-g、g-r色指數量值與標準化數量關係 (黑色折線為雙類星體與一般類星體數量比例)

3. 無線電波分析：我們使用無線電波觀測資料 (FIRST survey, White et al. 1997)，計算出200個雙類星體的無線電波源比例 (Radio detected / total) 為43%，而2000個一般類星體控制組則為14%，雙類星體產生無線電波訊號的機率為控制組的三倍，顯示雙類星體跟無線電波訊號有相關性。類星體無線電波訊號通常來自噴流，有研究顯示 (e.g, Breiding et al. 2024) 噴流與黑洞自旋速度有關，且雙類星體互繞時可能會加速黑洞自旋，進而加強了噴流。我們觀測到雙類星體跟無線電波訊號的相關性，有可能是成因於黑洞加速自旋的機制。

| | 無線電波源比例 |
|-------|-----------------|
| 雙類星體 | 43% \pm 5.54% |
| 一般類星體 | 14% \pm 0.89% |

▲ 表二：雙類星體與一般類星體的無線電波源比例

我們觀察類星體的無線電波影像，發現不同的無線電波源特徵，各種特徵的無線電波源影像範例如下：

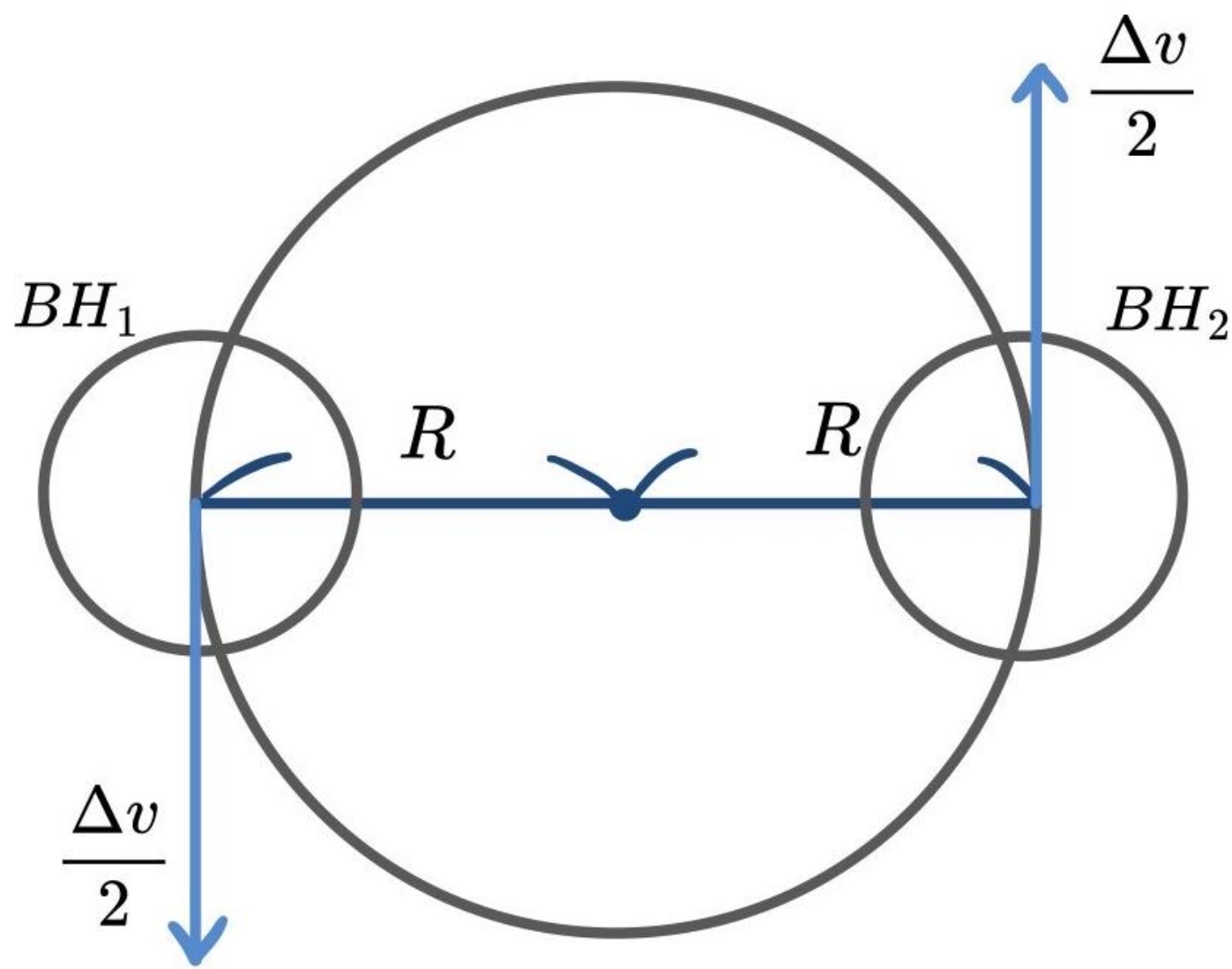


▲ 圖七：類星體在VLASS中的無線電波影像（圖片引自The DESI Legacy Imaging Surveys）

雙類星體的影像與互繞情形

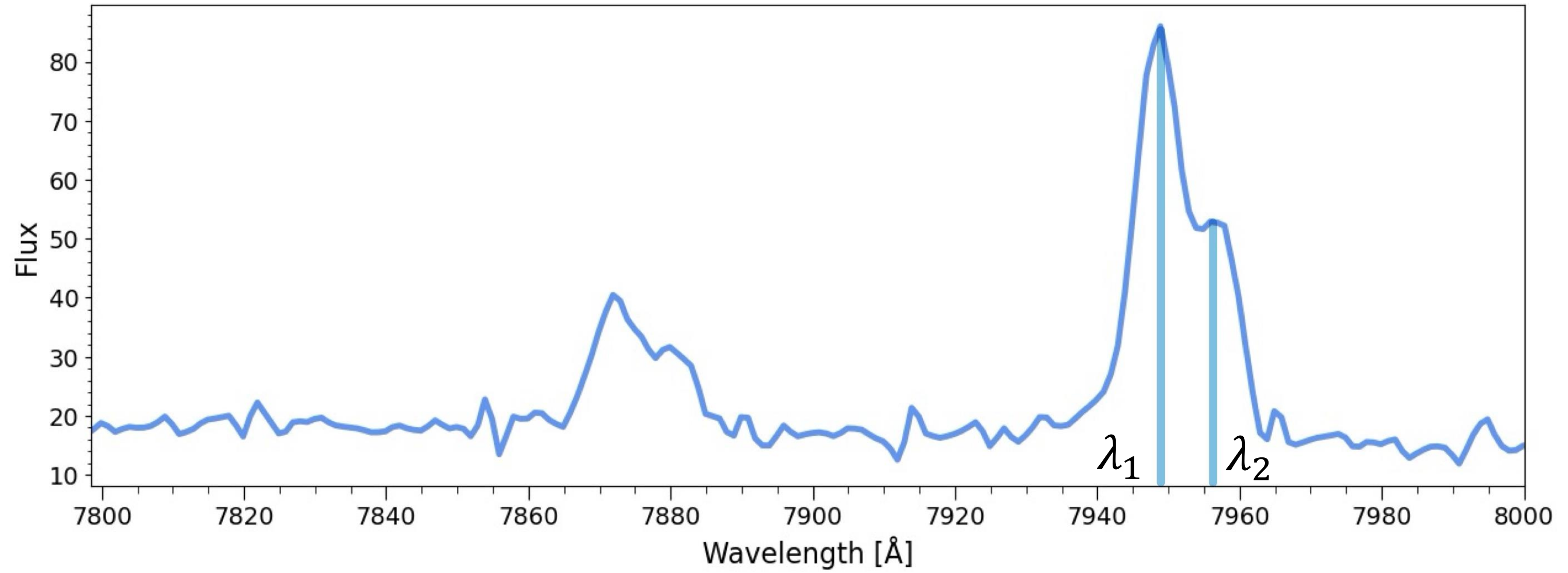
我們觀察554個雙類星體中，Shen et al. 2011有提供質量資料的200個雙類星體在The DESI Legacy Surveys的可見光影像，發現多數並無呈現明顯的兩個鄰近光源，因此我們分析互繞雙類星體受到彼此重力場互繞時的距離，推測其無法在影像上呈現兩個光源的可能性。假設雙類星體互繞的相對速度即為光譜呈現的相對速度，計算出相對速度後，以萬有引力與圓周運動的關係推算兩個類星體的間距（ $2R$ ）與視差（ p ）。

雙類星體互繞情形示意圖



▲ 圖八：雙類星體以相同速率作圓周運動

雙類星體光譜



▲ 圖九：雙類星體的[OIII]發射光譜

計算相對速度

$$\begin{aligned} Z_1 &= \frac{\lambda_1}{5007} - 1 \\ Z_2 &= \frac{\lambda_2}{5007} - 1 \end{aligned} \quad \rightarrow \quad \Delta v = \frac{Z_2 - Z_1}{(1+z)} \cdot c$$

以圓周運動與萬有引力關係得知互繞雙類星體間距

$$F_c = \frac{G \cdot M_{BH}^2}{(2R)^2} = \frac{M_{BH}}{2} \cdot \frac{(\frac{\Delta v}{2})^2}{R} \quad \rightarrow \quad R = \frac{G \cdot M_{BH}}{4 \cdot (\frac{\Delta v}{2})^2}$$

註： M_{BH} 為雙類星體系統的黑洞質量（ M_{\odot} ）

| 可見光影像 | | | | | |
|-----------|----------------------|----------------------|----------------------|----------------------|----------------------|
| $2R$ (pc) | 1077.4 | 579.4 | 651.9 | 353.8 | 289.9 |
| p (") | 1.9×10^{-1} | 1.3×10^{-1} | 9.5×10^{-2} | 4.7×10^{-2} | 4.3×10^{-2} |

▲ 表三：視差最大的前五個雙類星體影像與距離、視差（圖片引自The DESI Legacy Imaging Surveys (Dey et al. 2019))

根據計算結果，即使是視差最大的五個雙類星體，以目前觀測技術而言需要哈伯望遠鏡（HST）或韋伯望遠鏡（JWST）等不受大氣影響且可見光解析力約 $10^{-1} \sim 10^{-2}$ 角秒的天文望遠鏡才能辨別。

結論

一、自動化且系統性尋找雙類星體的方式

我們開發出自動化且系統性找尋雙類星體的方式，結合主成份分析（PCA）和支援向量機（SVM）判斷雙峰光譜特徵並選出雙類星體，支援向量機模型在幾乎選到所有雙類星體時仍有高的純度，是辨識雙類星體的有力方式。我們將此方式應用在SDSS DR16 $z < 0.8$ 所有類星體中，從7.6萬個類星體中找出524個新的雙類星體，我們找到的雙類星體數量是之前的四倍，且此方法可以直接運用在新的光譜資料庫，這是以往未達成的結果。

二、雙類星體有別於一般類星體的物理性質：兩者的性質有顯著不同，包含顏色和無線電波訊號

1. 色指數：我們發現雙類星體在u-g、g-r、r-i、i-z色指數中，顏色皆有較其他類星體整體偏紅的趨勢，是雙類星體有別於一般類星體的特徵，我們認為星系合併的過程可能使雙類星體偏紅。
2. 無線電波：雙類星體有無線電波源的比例約為一般類星體的三倍，此觀測現象可能與雙類星體互繞加速黑洞自旋有關。
3. 互繞情形：我們計算出互繞雙類星體間距最大可能為 1077.4 pc，觀測影像上視差約為0.19"，因此雙類星體互繞時會非常接近彼此，需要不受大氣干擾且解析力高的太空望遠鏡才有可能辨別。

參考文獻資料

1. Smith, K. L., Shields, G. A., Bonning, E. W., et al. 2010, ApJ, 716, 866
2. Lyke, B. W., Higley, A. N., McLane, J. N., et al. 2020, ApJS, 250, 8
3. Suzuki, N. 2006, ApJS, 163, 110
4. Shen, Y., Richards, G. T., Strauss, M. A., et al. 2011, ApJS, 194, 45
5. Glikman, E., Simmons, B., Mailly, M., et al. 2015, ApJ, 806, 218
6. White, R. L., Becker, R. H., Helfand, D. J., et al. 1997, ApJ, 475, 479
7. Breiding, P., Chiaberge, M., Lambrides, E., et al. 2024, ApJ, 963, 91
8. Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168