時系列データの機械学習的アプローチ

植村誠（広島大学）

共同研究者：安部太晴、山田悠梨香、深澤泰司（広島大学）、
板良房（東北大学）、松永典之（東京大学）、池田思朗（統計数理研究所）

天文学におけるデータ科学的方法　@統数研　2017.5.31

自己紹介

- 突発天体現象の即時観測
  - 降着円盤、ジェット、光・赤外線天文学、多波長、偏光
- 統計・情報系のデータ科学的手法の応用
  - 2013年～　科研費新学術「スパースモデリング」計画　研究「天文班」

Outline

「機械学習」をタイトルに入れてしまいましたが、所謂「機械学習」的な内容が少ないことに後で気が付いてしまいました。すいません。

- 最近の取り組み　その１

  「AGB星の炭素過剰星・酸素過剰星の測光データを用いた機械分類

  ースパースロジスティック回帰ー」

  （安部、植村、板、松永、池田）　6枚

- 最近の取り組み　その２

  「ブレーザーのSEDモデルパラメータのMCMCによる推定」

  （山田、植村、深澤）6枚

- 本題

  「ブレーザーの偏光時系列データからジェット内で起きていることに迫る」　17枚

最近の取り組み　その１

AGB星の炭素過剰星・酸素過剰星の測光データを用いた機械分類

ースパースロジスティック回帰ー

安部、植村（広島大）、板（東北大）、松永（東大）、池田（統数研）

日本天文学会2017年春季年会@九大　で発表

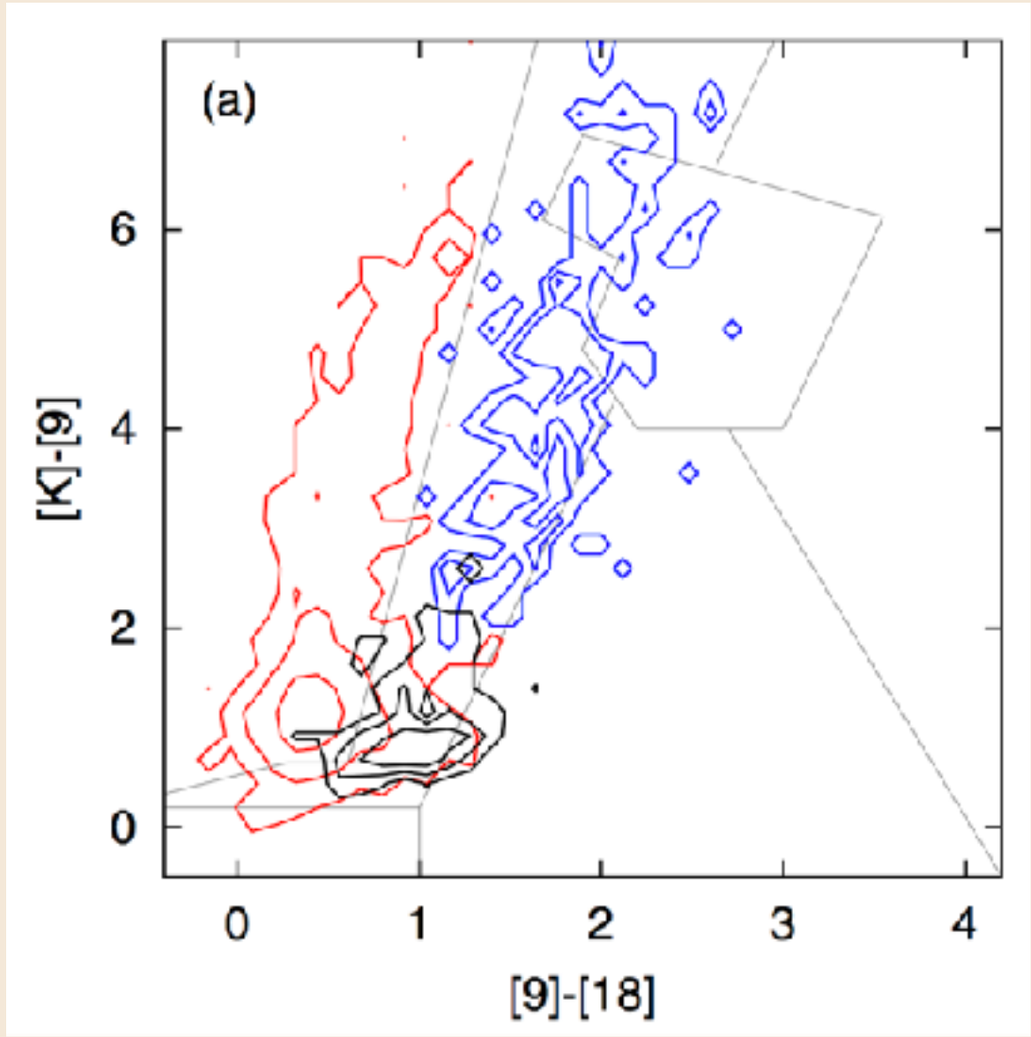
一般的な背景

- 天文では「回帰」の問題は多く語られるが、「分類」の問題はあまり語られない。何かの「境界線」は「目で見て決める」ことも多い。

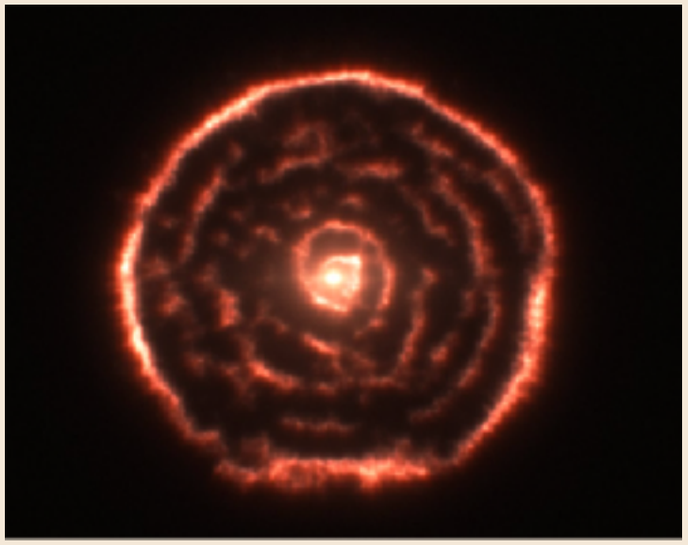
- ビッグデータの時代になり、目的に合った機械分類のノウハウを持っているかどうかで、できる研究・できない研究が分かれそう。

- 勉強しつつ、何か仕事しておきたい。

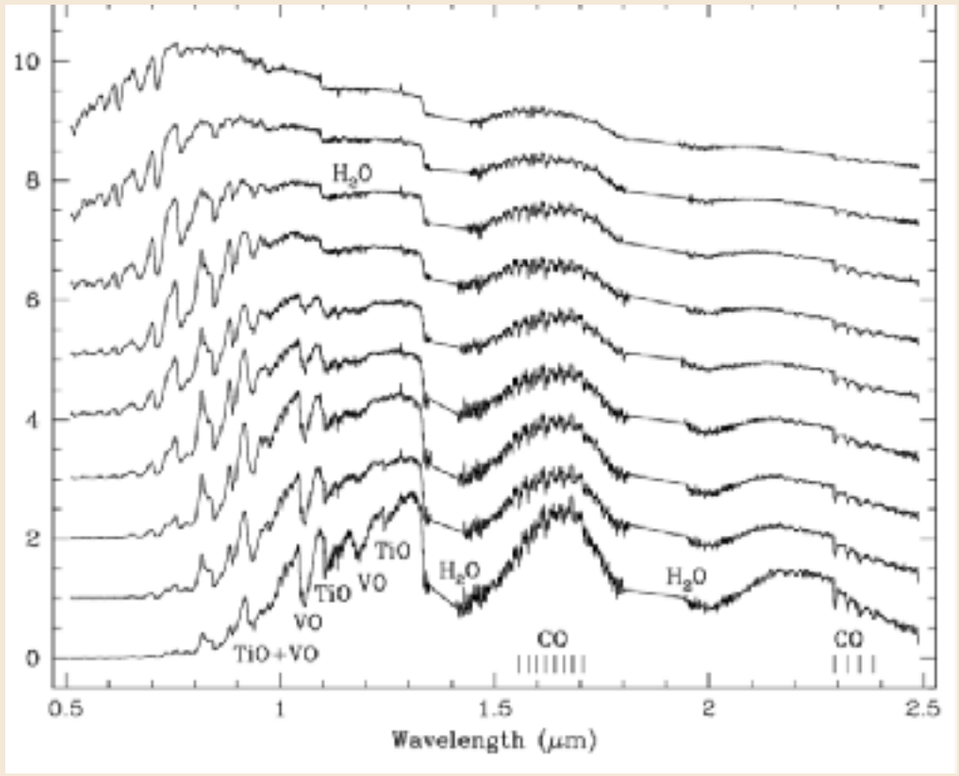
AGB星の炭素過剰・酸素過剰星の分類
(Ishihara+11)

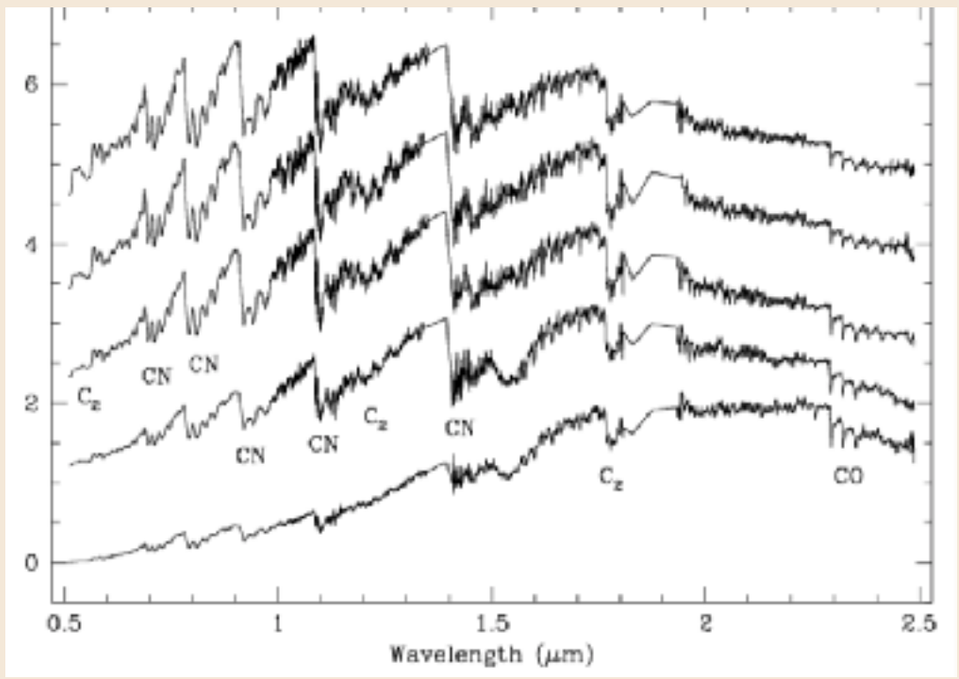


今回の具体的な問題



AGB星 R Scl
(©ESO)





- AGB星（Asymptotic Giant Branch星＝軽い星の進化の最終段階）の酸素過剰(O-rich)星、炭素過剰(C-rich)星の分類

- O-rich, C-rich の分類の定義＝スペクトルの特徴 →スペクトルを撮るのは時間がかかる＝少ないサンプル

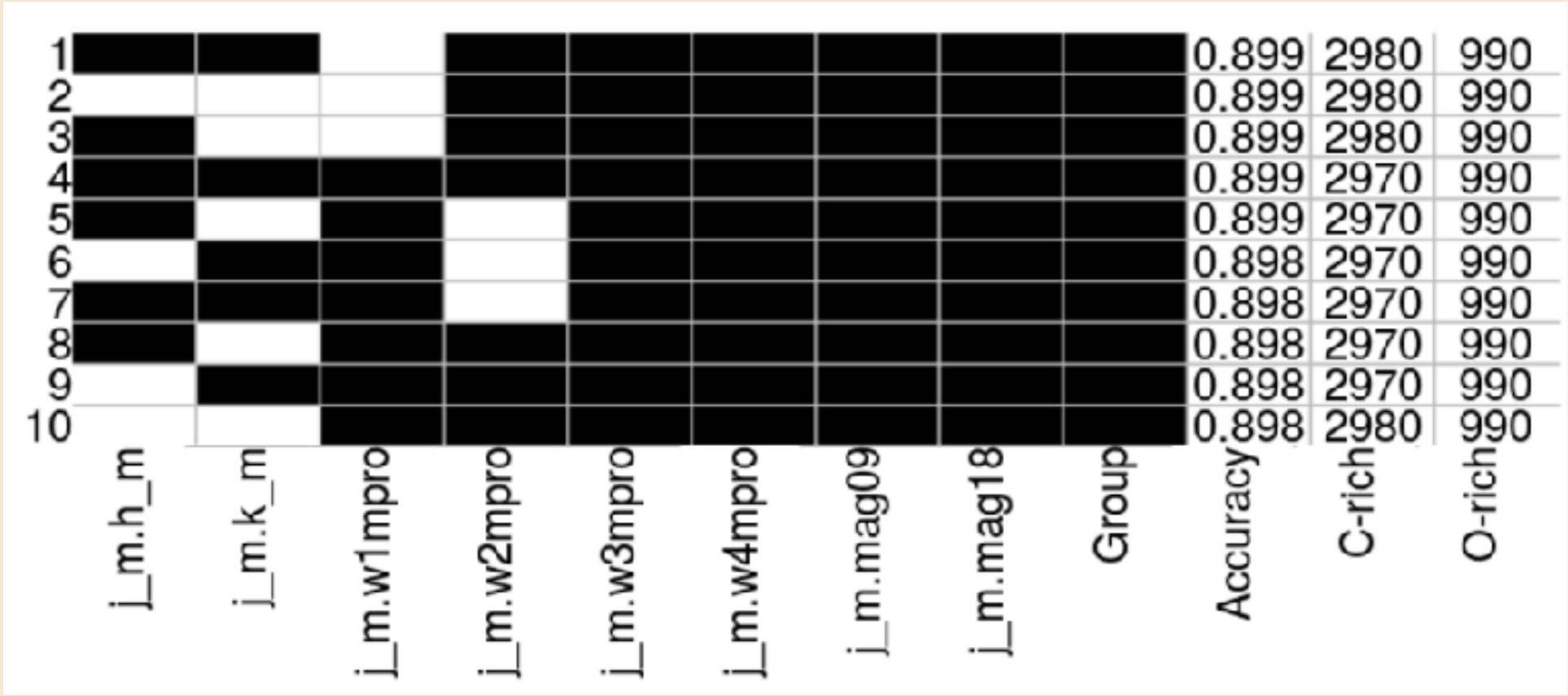
- 撮像＝測光観測だけから分類したい→先行研究：
  Ishihara+11

- 本当にその波長のデータが良いのか？本当にその境界が良いのか

- 怪しい天体、変な天体も探したい→あるクラスに入る確率を与えてくれるものが良い

2MASS＋あかり＋WISEによる線形判別

- データ：スペクトルから既に分類が分かっている C-rich, O-rich 約3000天体の、１μmから22μmまで、9バンドの測光データ＝8個の色指数データ
- 分類手法：フィッシャーの線形判別（最も単純なもの。勉強のため。）
- 8個の色指数データ全ての組み合わせ（2^8-1）を試行。交差検証法で最も正答率の良い組み合わせを探索。
- 結果：9,18μm(AKARI)、12,22μm(WISE)の組み合わせによる分類で、正答率 0.899



	i_m.h_m	i_m.k_m	i_m.w1mpro	i_m.w2mpro	i_m.w3mpro	i_m.w4mpro	i_m.mag09	i_m.mag18	Group	Accuracy	C-rich	O-rich
1	■	■		■	■	■	■	■	■	0.899	2980	990
2				■	■	■	■	■	■	0.899	2980	990
3	■			■	■	■	■	■	■	0.899	2980	990
4	■	■	■	■	■	■	■	■	■	0.899	2970	990
5	■		■		■	■	■	■	■	0.899	2970	990
6		■	■		■	■	■	■	■	0.898	2970	990
7	■	■	■		■	■	■	■	■	0.898	2970	990
8	■		■	■	■	■	■	■	■	0.898	2970	990
9		■	■	■	■	■	■	■	■	0.898	2970	990
10			■	■	■	■	■	■	■	0.898	2980	990

Sparse Multinomial Logistic Regression

多クラスロジスティック回帰

$$P\left(y^{(i)}=1\middle|\boldsymbol{x},\boldsymbol{w}\right)=\frac{\exp(\boldsymbol{w}^{(i)^T}\boldsymbol{x})}{\sum_{j=1}^{m}\exp(\boldsymbol{w}^{(j)^T}\boldsymbol{x})}$$

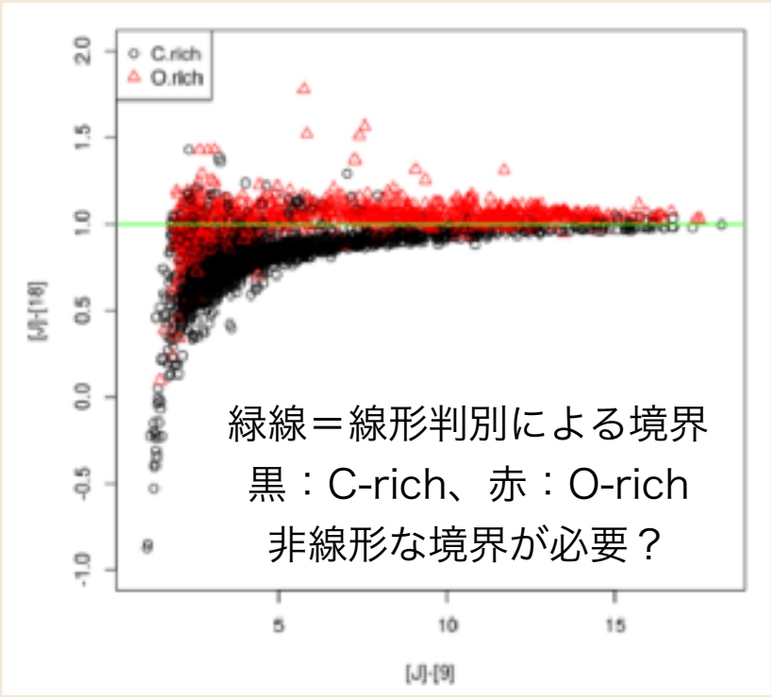
に基づいて、重み関数wの最尤推定を行なう。($\hat{\boldsymbol{w}}_{\mathrm{MAP}}$)

$$l(\boldsymbol{w})=\sum_{j=1}^{n}\log P(\boldsymbol{y}_j|\boldsymbol{x}_j,w)$$

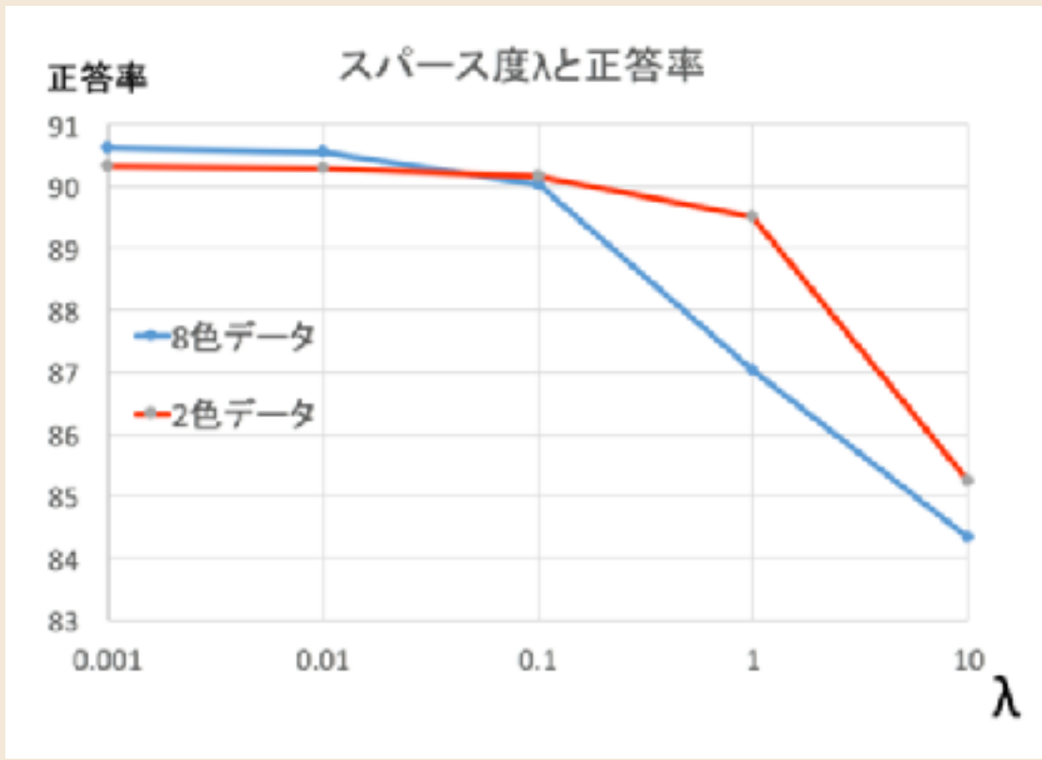
$$\hat{\boldsymbol{w}}_{\mathrm{MAP}}=\arg\max \mathrm{L}\,(\boldsymbol{w})=\arg\max[l(\boldsymbol{w})+\log p(w)]$$

ここで $\quad p(\boldsymbol{w})\propto\exp(-\lambda\left|\left|\boldsymbol{w}\right|\right|_1)$

B Krishnapuram et al.(2005)



- カーネルを用いて非線形境界へ
- ロジスティック回帰＝各クラスに分類される確率
- 重み関数 w の 1 次ノルムを最小化させることで過学習しない判別機ができる
- 結果：正答率～0.9で頭打ち。非線形効果による劇的な改善はなかった。



今後の展開

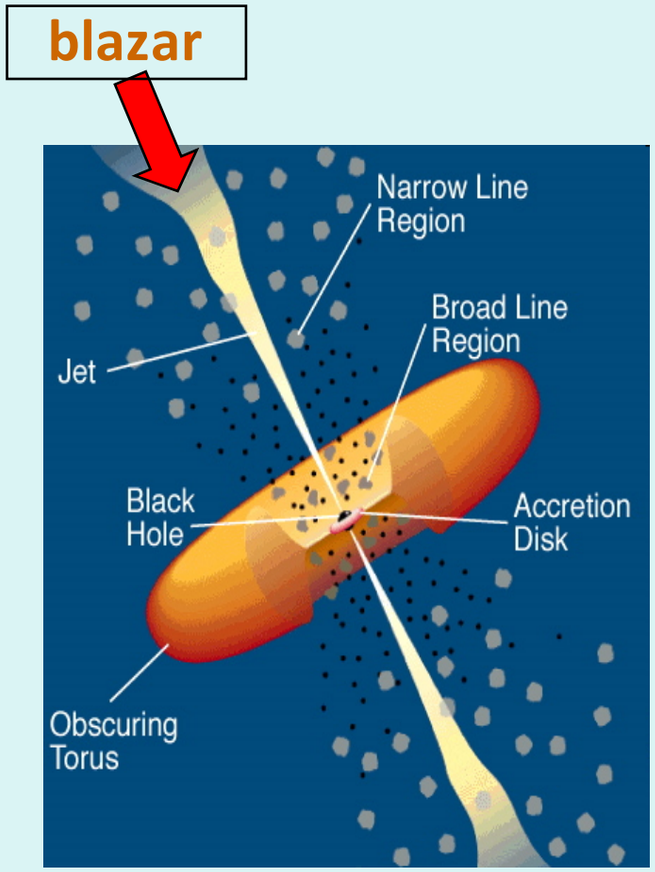
- 機械分類に関して良い勉強＆ツールを得られた。
- AGB星の分類は一旦これでお終い（最終的な目的次第で再度検討するかも）
- 不規則な変動をする変光星データ（KISO-GP計画：松永）の分類を目指したい。

最近の取り組み　その２

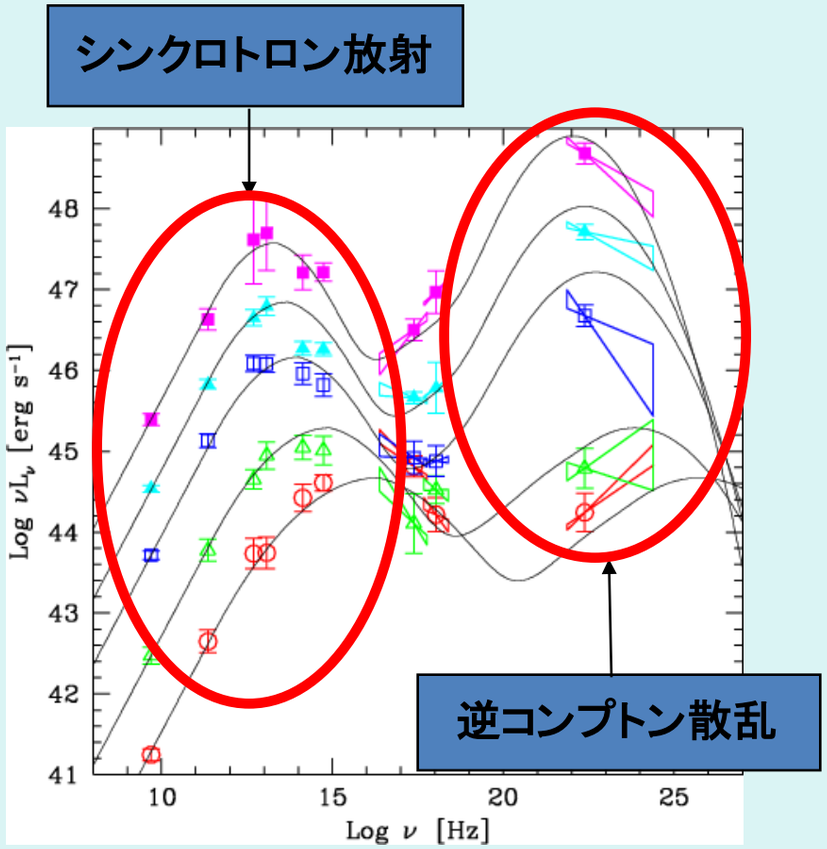
ブレーザーのSEDモデルパラメータのMCMCによる推定

山田、植村、深澤（広島大学）

活動銀河核とブレーザー天体



- 活動銀河核：中心の巨大ブラックホール付近で非常に明るい銀河
- ブレーザー：活動銀河核ジェットを真正面から見た天体
  - 相対論的ビーミング効果によりジェットからの放射が卓越　→ジェットの研究に適している
- 電波からγ線まで幅広い周波数領域で観測
- 強い偏光（シンクロトロン放射）



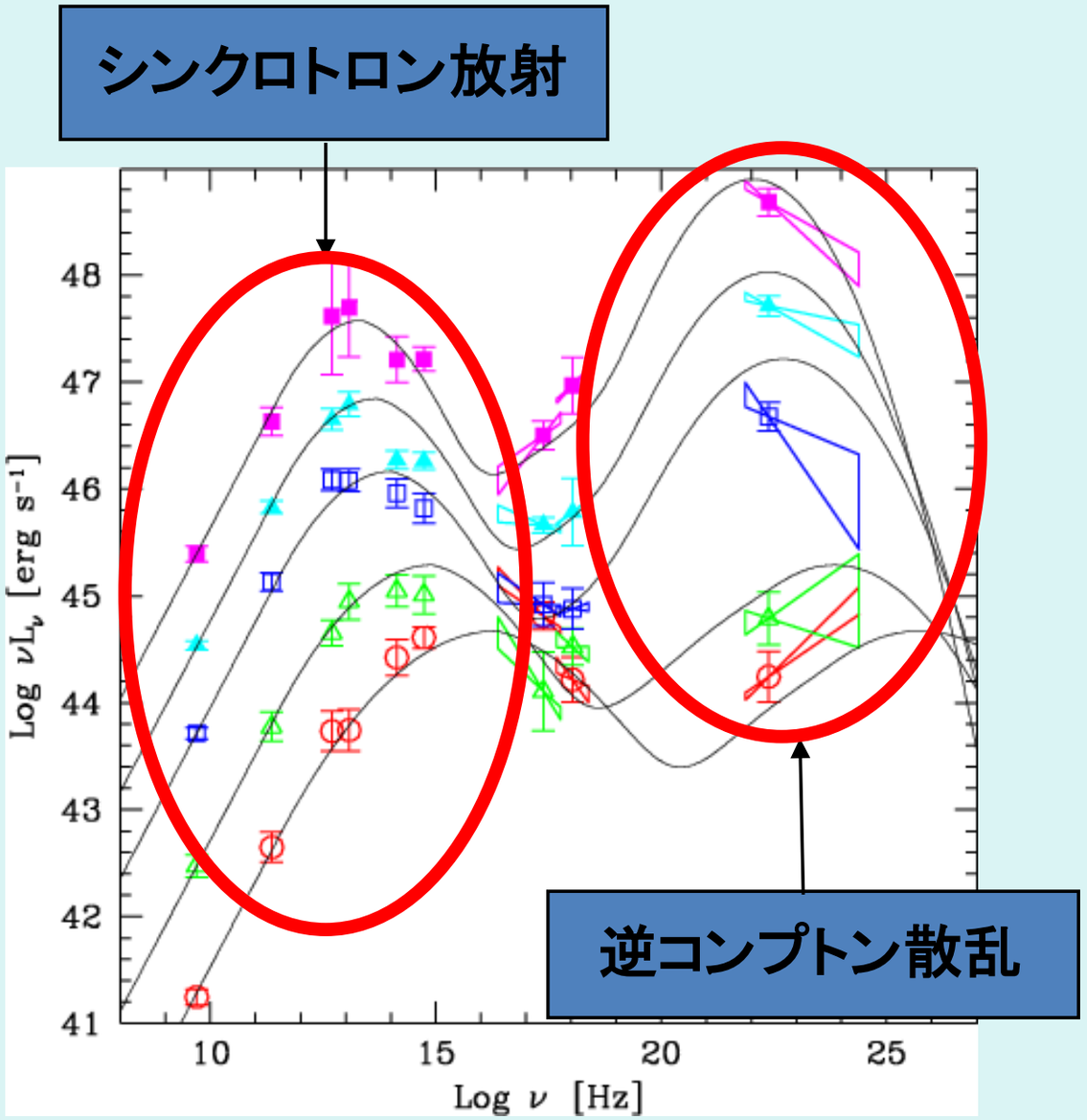
SEDモデルパラメータの推定

- シンクトロン放射と逆コンプトン散乱放射

$$f_\epsilon^{\rm syn} = \frac{\sqrt{3}\delta_D^4 \epsilon' e^3 B}{4\pi h d_L^2} \int_0^\infty d\gamma' N_e'(\gamma') R(x)$$

$$f_{\epsilon_s}^{\rm SSC} = \frac{9}{16} \frac{(1+z)^2 \sigma_T \epsilon_s'^2}{\pi \delta_D^2 c^2 t_{v,\rm min}^2} \int_0^\infty d\epsilon' \frac{f^{\rm syn}}{\epsilon'^3} \times \int_{\gamma'_{\rm min}}^{\gamma'_{\rm max}} d\gamma' \frac{N_e'(\gamma')}{\gamma'^2} F_C(q,\Gamma)$$

$$N_e(\gamma) = K_e \gamma^{-p_0} \quad (\gamma_1 < \gamma < \gamma_2)$$



- 変数(7～9個)：磁場B, ドップラー因子δ, タイムスケール tv（放射領域のサイズ）, 電子のエネルギー分布：定数 Ke, ベキ p0, 最小・最大値 $\gamma 1$, $\gamma 2$

- いくつかの変数が縮退していて全て独立には決まらない。

- 従来法：別の観測からいくつかの値を固定して、他のパラメータを推定→推定結果の不定性は通常議論しない（固定した値に依存するため）

- 提案法：1つの値に固定→事前分布、にしてMCMCで推定。事後確率からより良い不定性込みの議論ができるように。

適応的メトロポリス法

- メトロポリス法において、提案分布をMCMCサンプルから学習しながら効率の良いサンプリングに最適化していく。

（「適応的マルコフ連鎖モンテカルロ法入門」荒木貴光（産総研）@チュートリアル講演「マルコフ連鎖モンテカルロ法とデータ駆動科学」2016年3月6日@神戸)

提案分布は多変量正規分布

$$f(\boldsymbol{x}) = \frac{1}{(\sqrt{2\pi})^p \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right)$$

変数の平均、分散共分散行列をMCMCステップ毎に更新していく

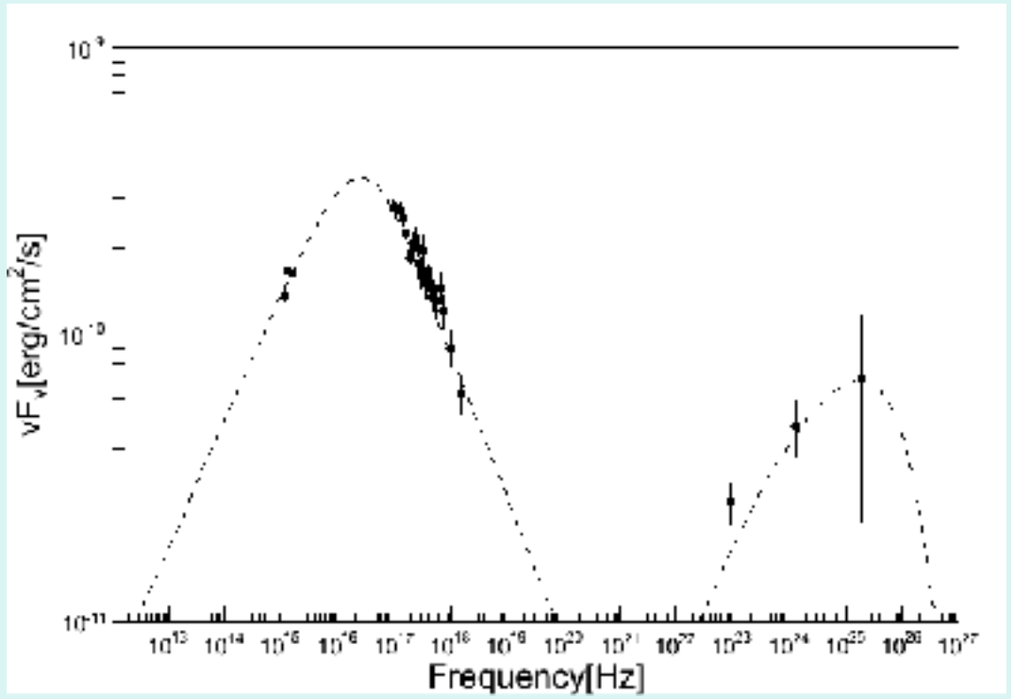
$$\boldsymbol{\mu}_n \leftarrow \boldsymbol{\mu}_{n-1} + h_n(\boldsymbol{x}_n - \boldsymbol{\mu}_{n-1})$$

$$\Sigma_n \leftarrow \Sigma_{n-1} + u_n\left((\boldsymbol{x}_n - \boldsymbol{\mu}_{n-1})(\boldsymbol{x}_n - \boldsymbol{\mu}_{n-1})^T - \Sigma_{n-1}\right)$$

$$\sigma^2{}_n \leftarrow \sigma^2_{n-1} + s_n(FA_n - \alpha)$$

→　採択率αに収束

結果の一例



ブレーザー Mrk 421

データ

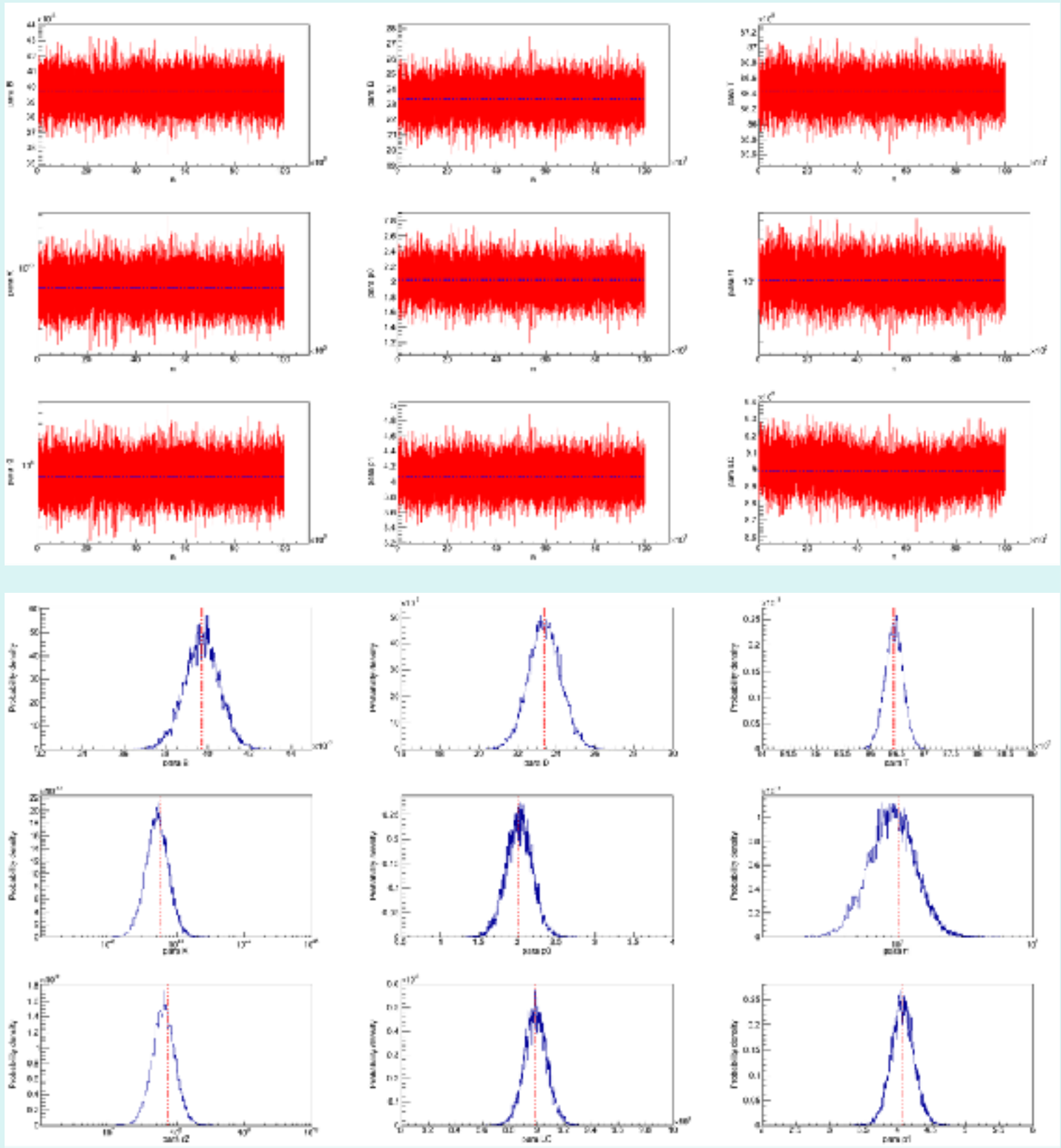
可視光：かなた望遠鏡

X線：XRT/Swift

γ線：LAT/Fermi

モデル

ドップラー因子は電波(VLBI)観測からの制約を利用して、平均20.0、標準偏差10.0の正規分布を事前分布としている



現在の課題

- MCMC 1step の計算に時間が掛かるため、実験に時間が掛かる→並列化して少しマシに。

- 提案分布＝多次元正規分布の分散共分散行列の初期値として強い相関を与えている（物理的にパラメータ間の相関は予想できるため）→学習後も強い相関が残る。適切か？

- 定常分布に収束するのが遅かったため、交換法の実装を検討→局所解は無さそうだったのでペンディング

- 複数のSEDデータ（SED時系列データ）に時間方向の制約も加えてMCMCで同時に解いてみたい。

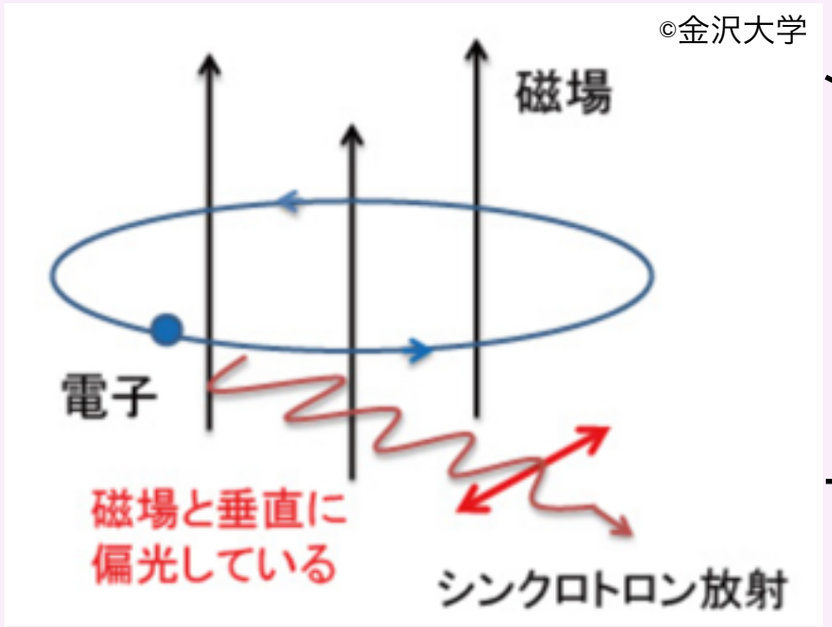
    - 1つ1つのSEDのデータが少なくとも、パラメータの変動の連続性や変動するパラメータの数が少ないこと、などを事前分布に入れて、妥当な解が得られるか。

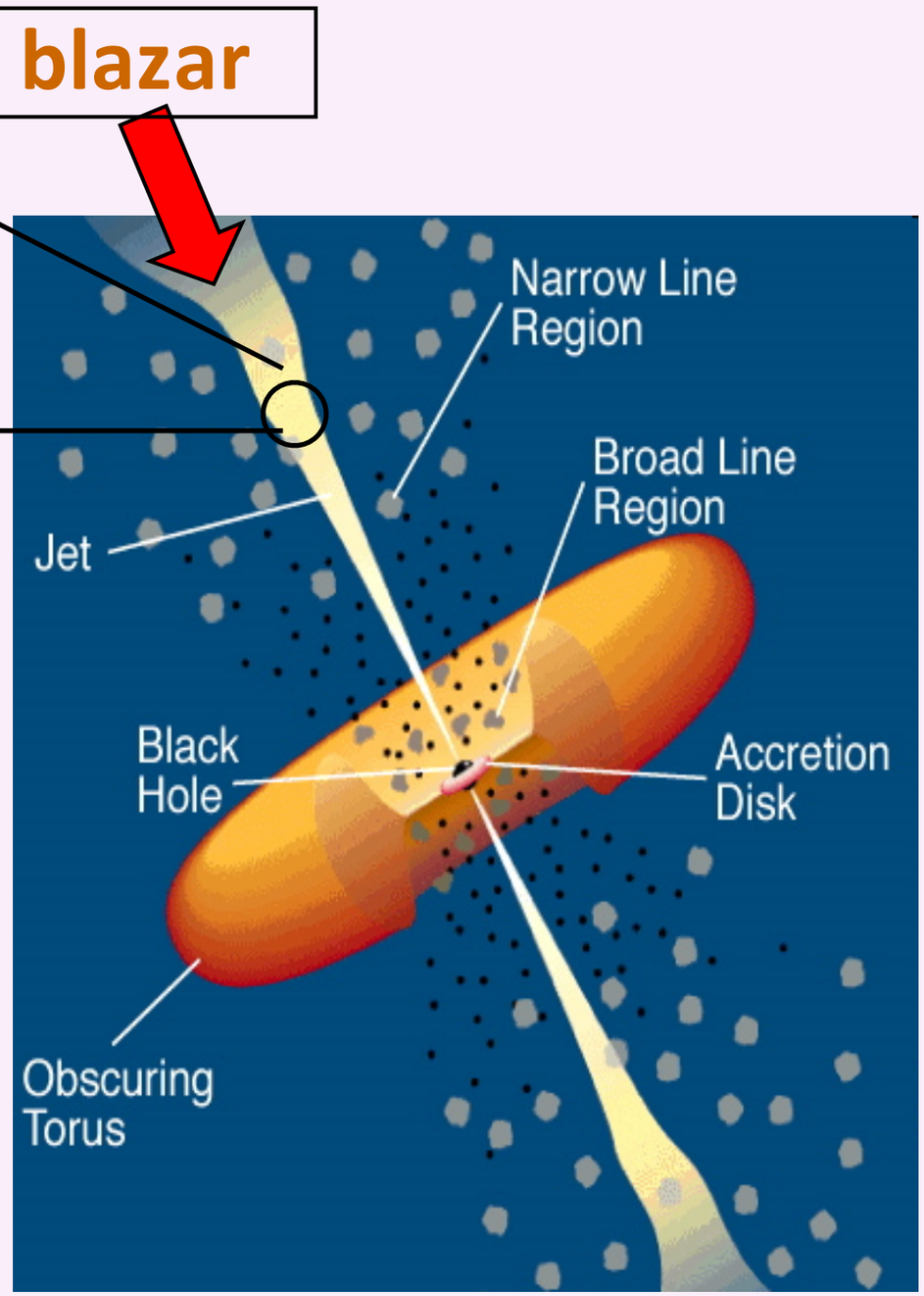
本題

ブレーザーの偏光の時系列データからジェット内で起きていることに迫る

(Uemura+10, PASJ, 62, 69)

ブレーザー天体とジェットと偏光





- ジェット←磁場が重要な役割

- 偏光方向と磁場の方向が垂直
  →直接は測ることができない磁場の情報を得ることができる

偏光の時間変化

偏光の先行研究

- だいたいは明らかな相関なし、めちゃくちゃな変動とされる

  Moore et al. (1982), Jones et al. (1985, 1988), その他たくさん

- ただし、光度曲線や色と相関する例もいくつか

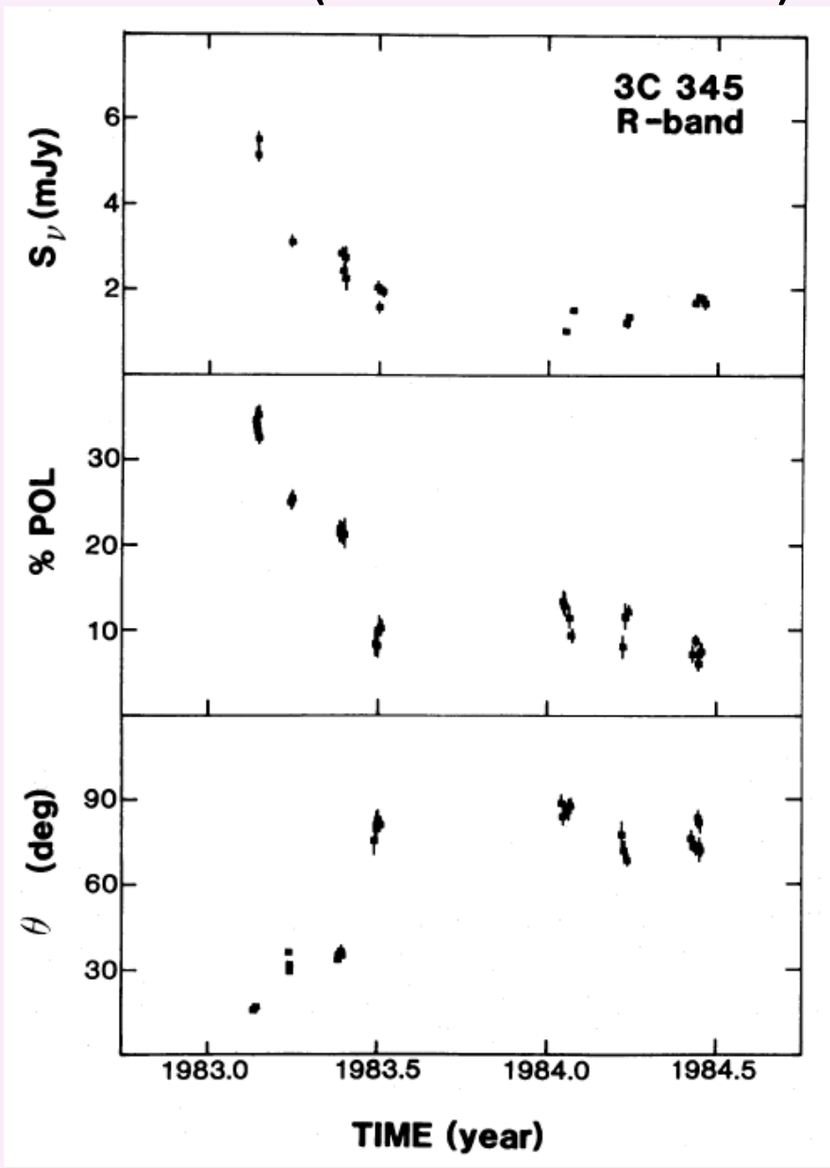
  Smith et al., (1986), Tosti et al. (1998), Efimov & Shakhovskoy (1998), Fan et al. (2000), Cellone et al. (2007),

- 特に数日〜数週間というスケールの密な継続観測は少ない

  本当にランダムなのか？　法則性が見落とされていないか？

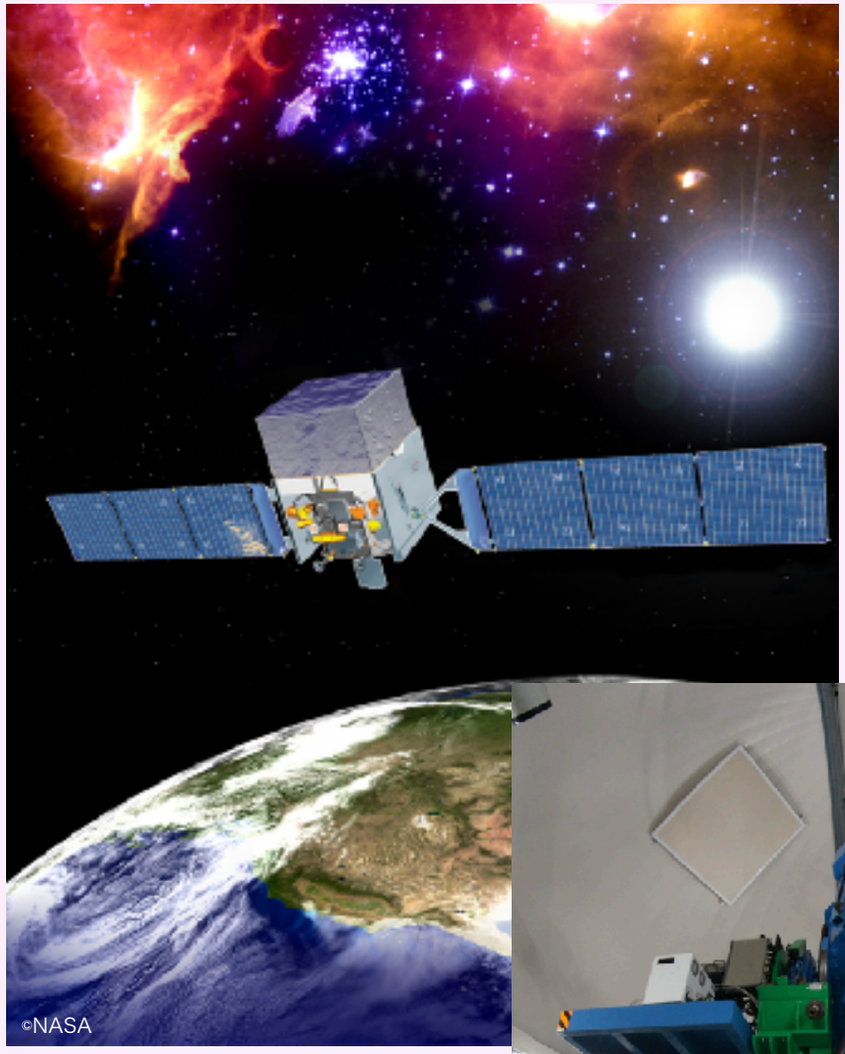
シンプルで普遍的な観測的特徴をとらえたい！

偏光度と光度が相関した例
3C 345 (Smith et al. 1986)



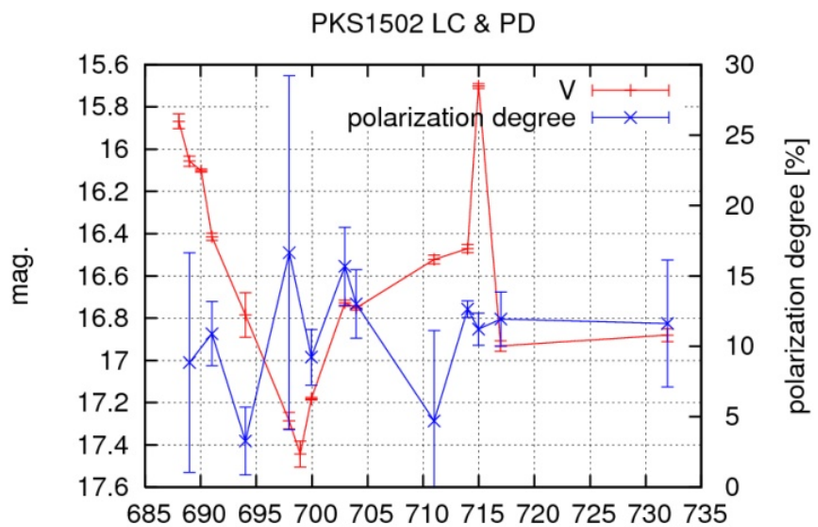
広島大学「かなた望遠鏡」とフェルミγ線宇宙望遠鏡

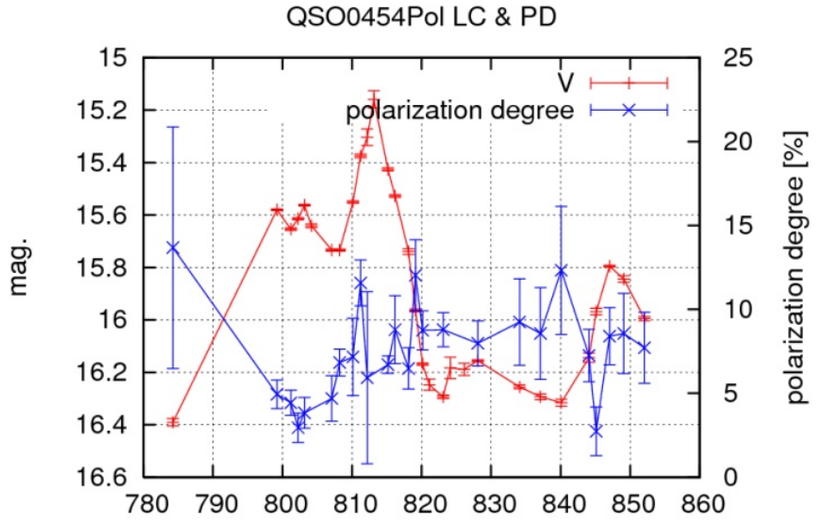
- **フェルミ衛星**
  - γ線宇宙望遠鏡
  - 2008～現在
  - NASAを中心とした国際共同プロジェクト。
    広島大学が日本代表機関
  - γ線が検出される天体の中で最も数が多いのが「ブレーザー」

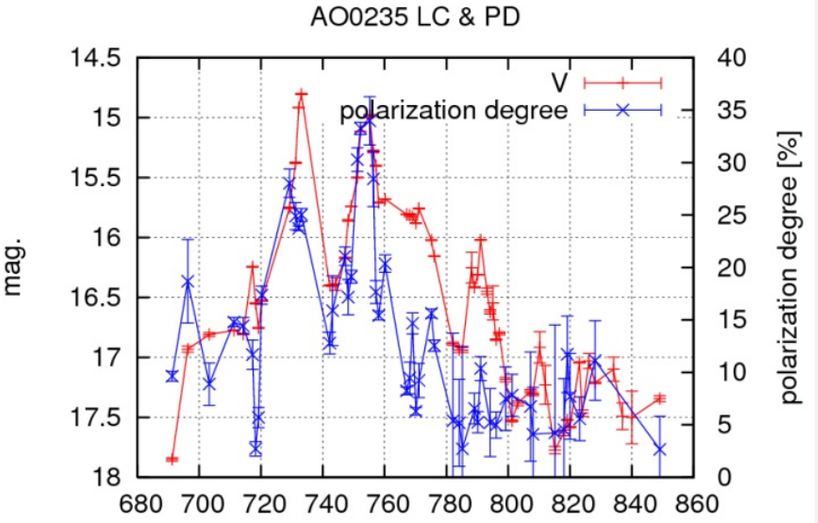


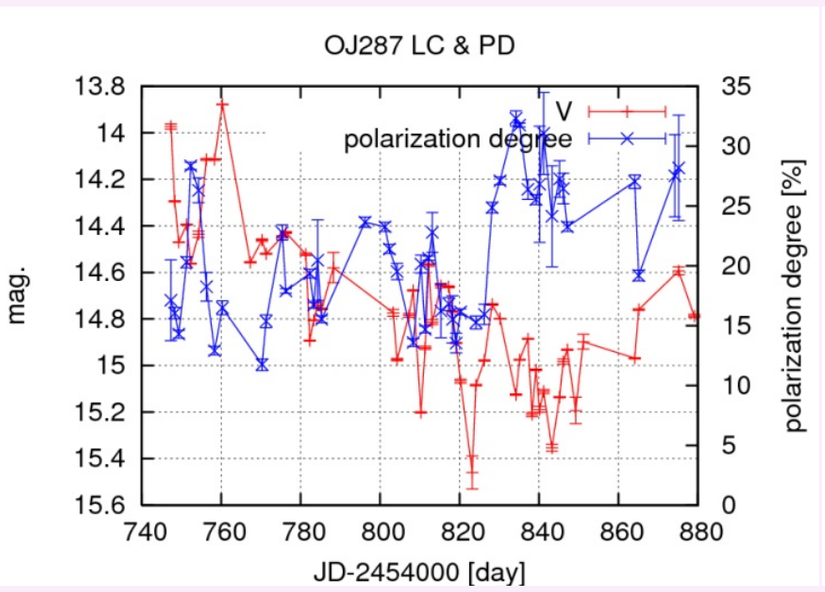
- **かなた望遠鏡**
  - 可視光ー近赤外線望遠鏡
  - 2006～現在
  - フェルミ衛星が多くのブレーザーを検出することを見越して、可視光とγ線で同時観測する目的で開発された。
  - 2008年よりブレーザーの偏光モニター観測（40天体ほど）2008年は９割の観測時間を本プロジェクトに使用　→結果一覧へ

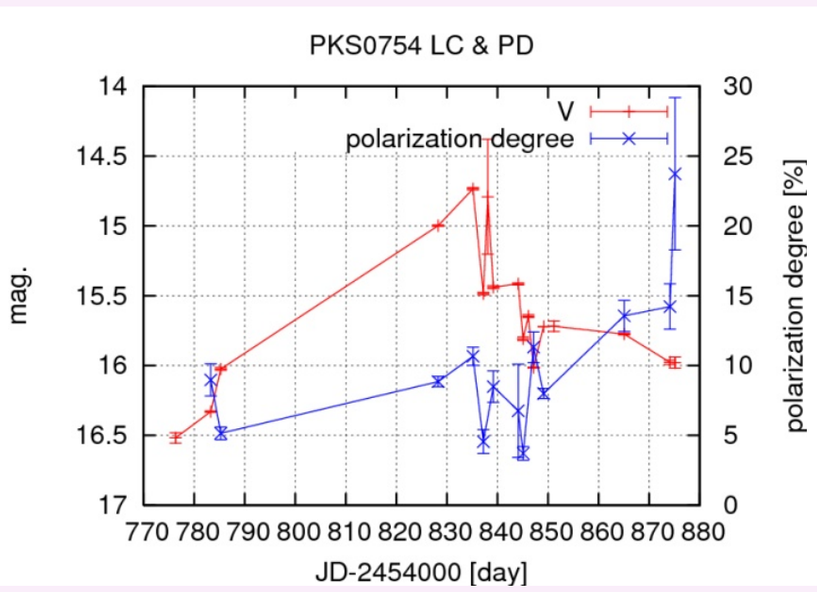
結果

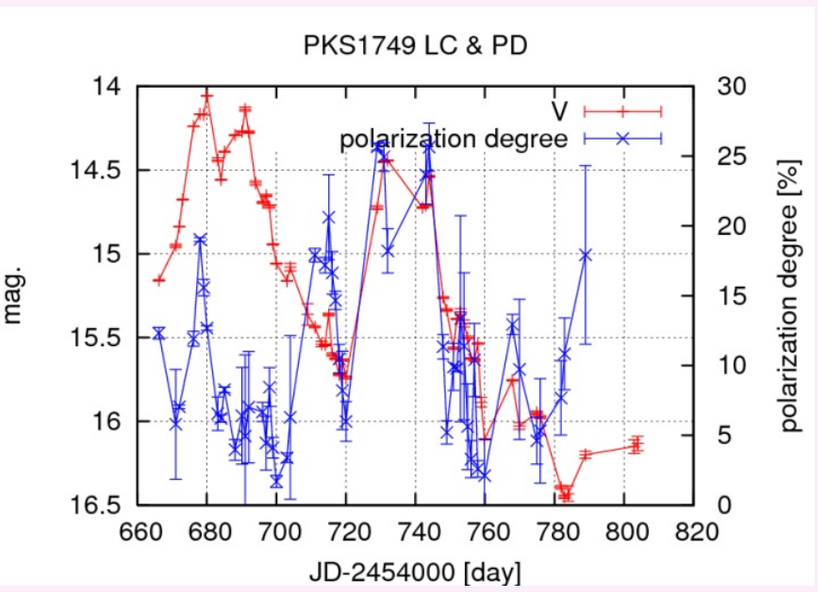

PKS1502 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]

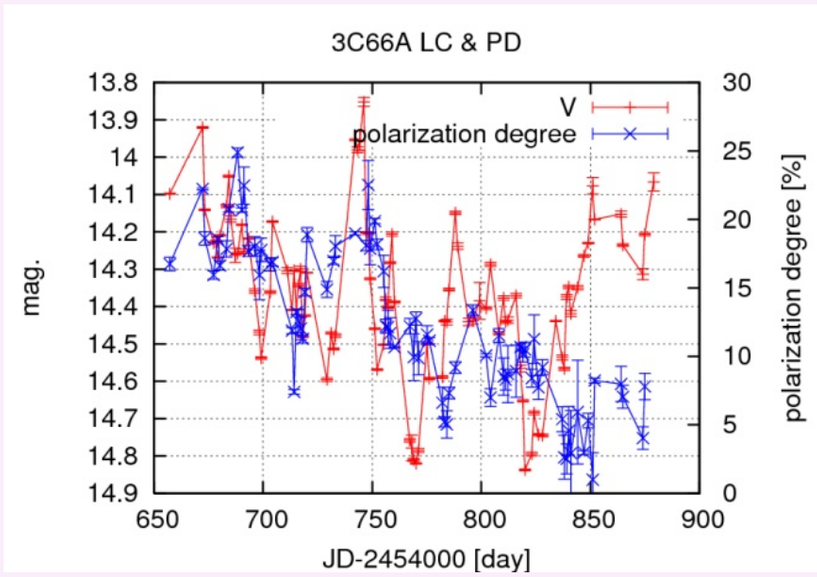

QSO0454Pol LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]

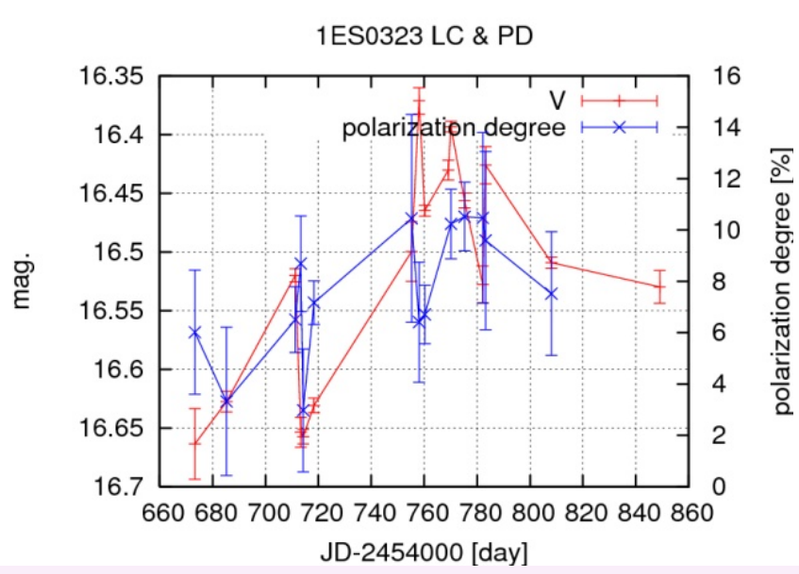

AO0235 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]

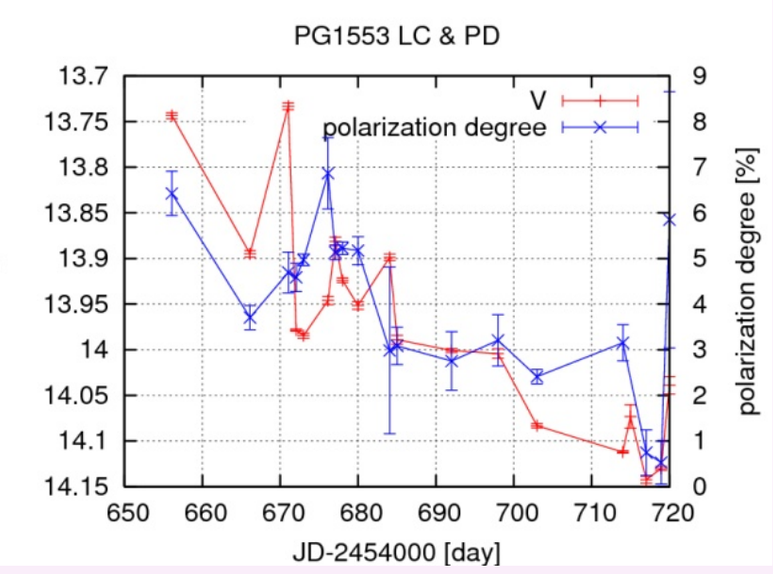

OJ287 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]

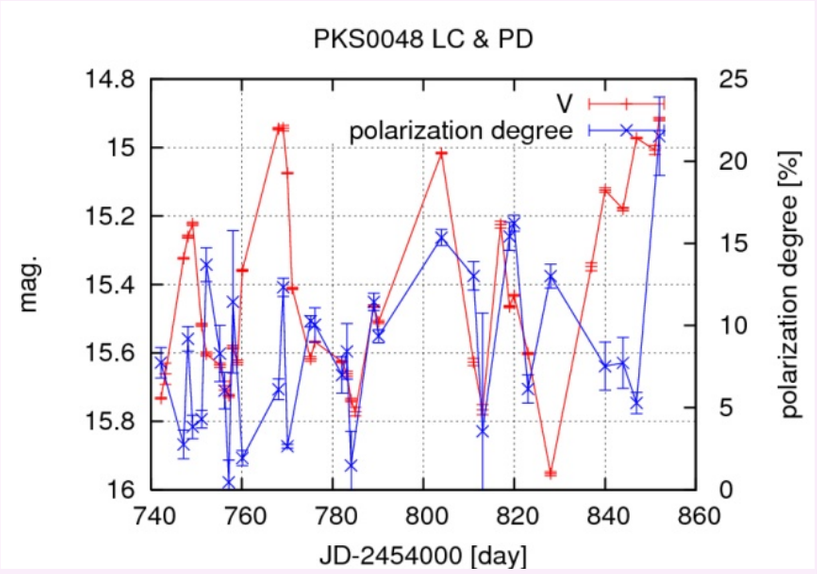

PKS0754 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]

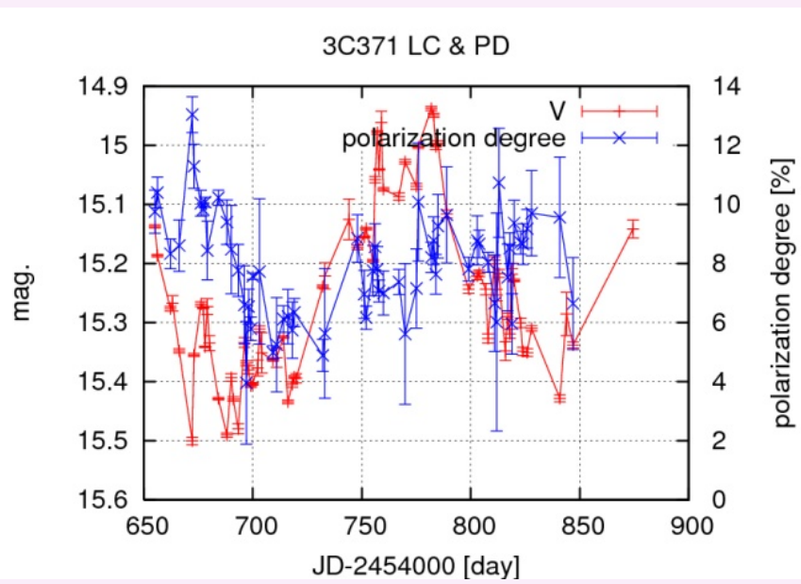

PKS1749 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]

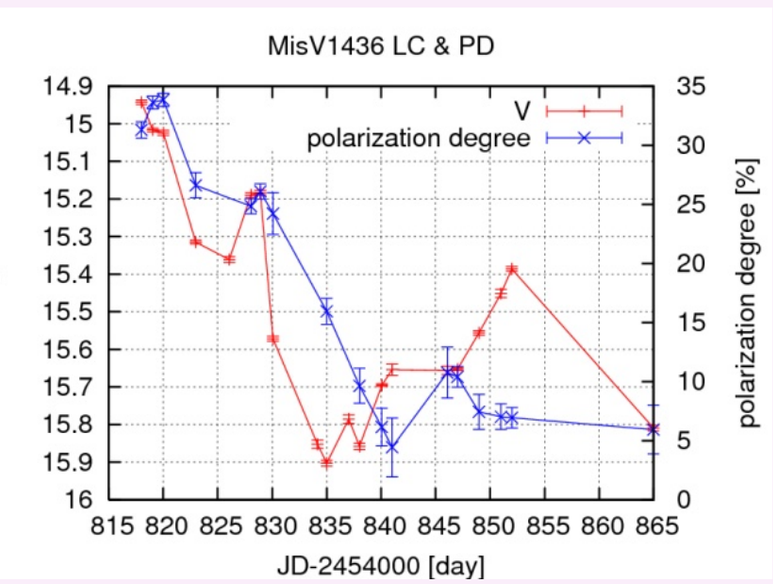

3C66A LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]

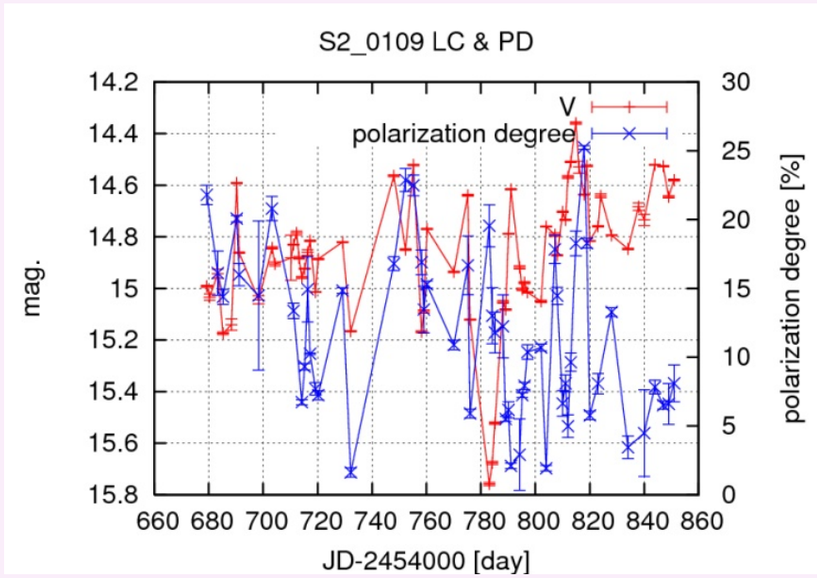

1ES0323 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]

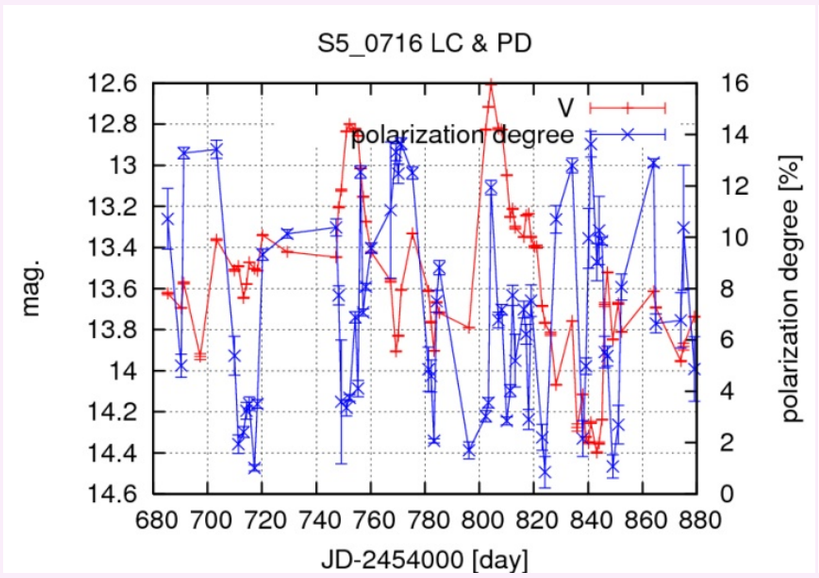

PG1553 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]

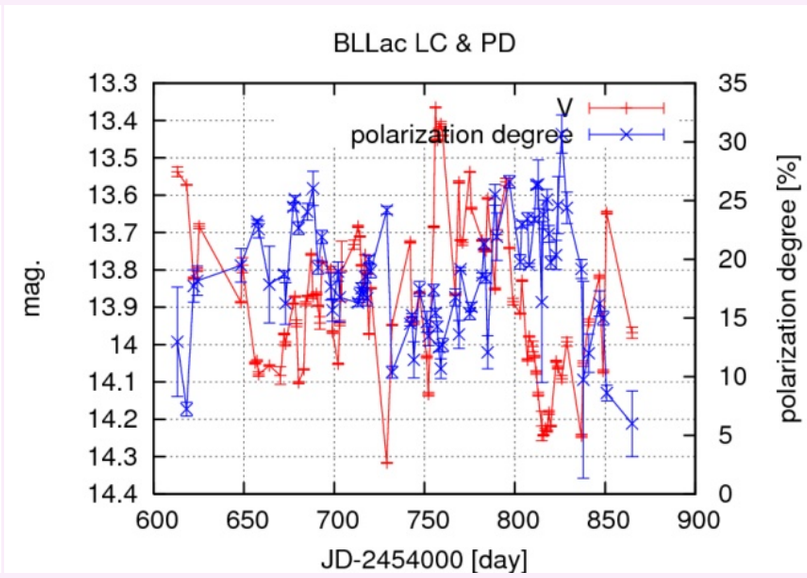

PKS0048 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]

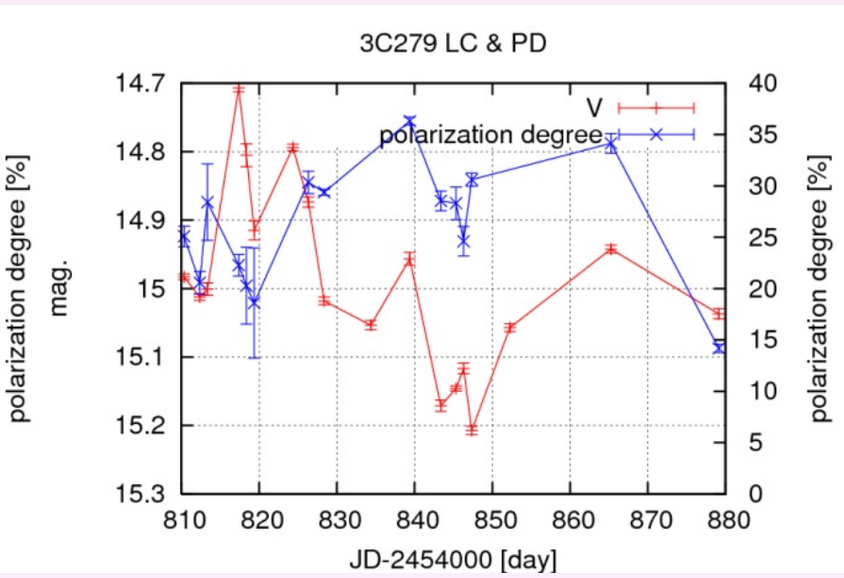

3C371 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]

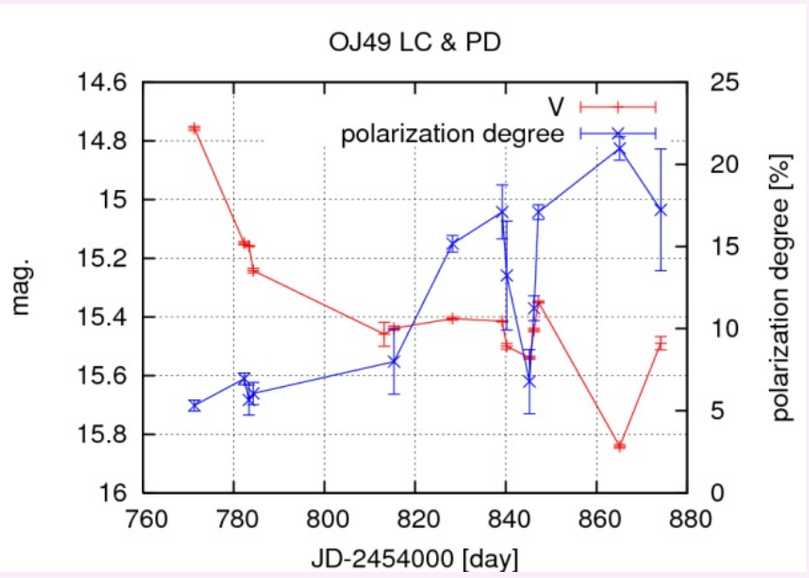

MisV1436 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]

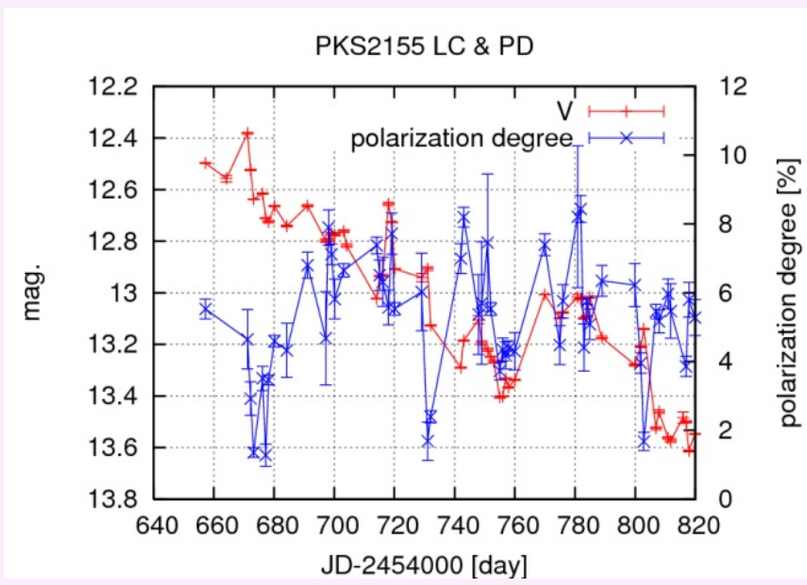

S2_0109 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]

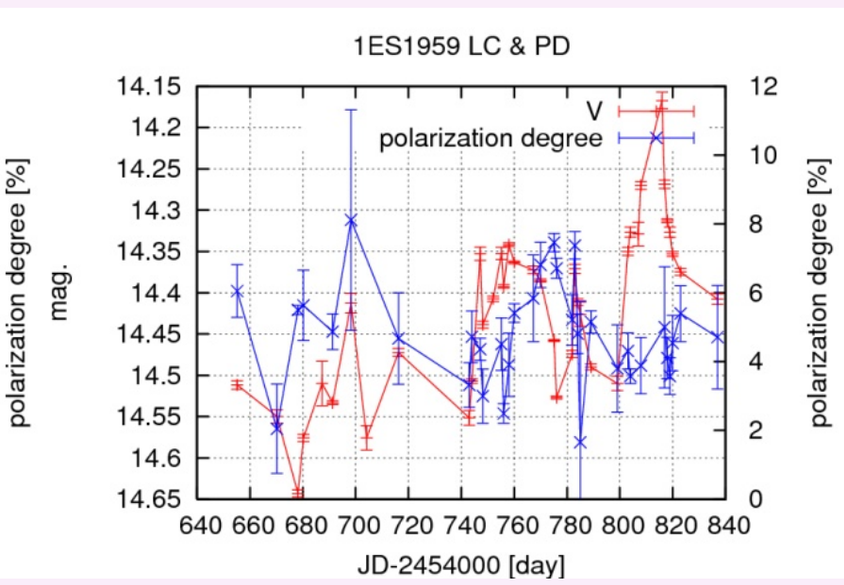

S5_0716 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]


BLLac LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]


3C279 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]


OJ49 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]


PKS2155 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]


1ES1959 LC & PD
V
polarization degree
mag.
polarization degree [%]
JD-2454000 [day]

光度と偏光度の相関関係

めちゃ

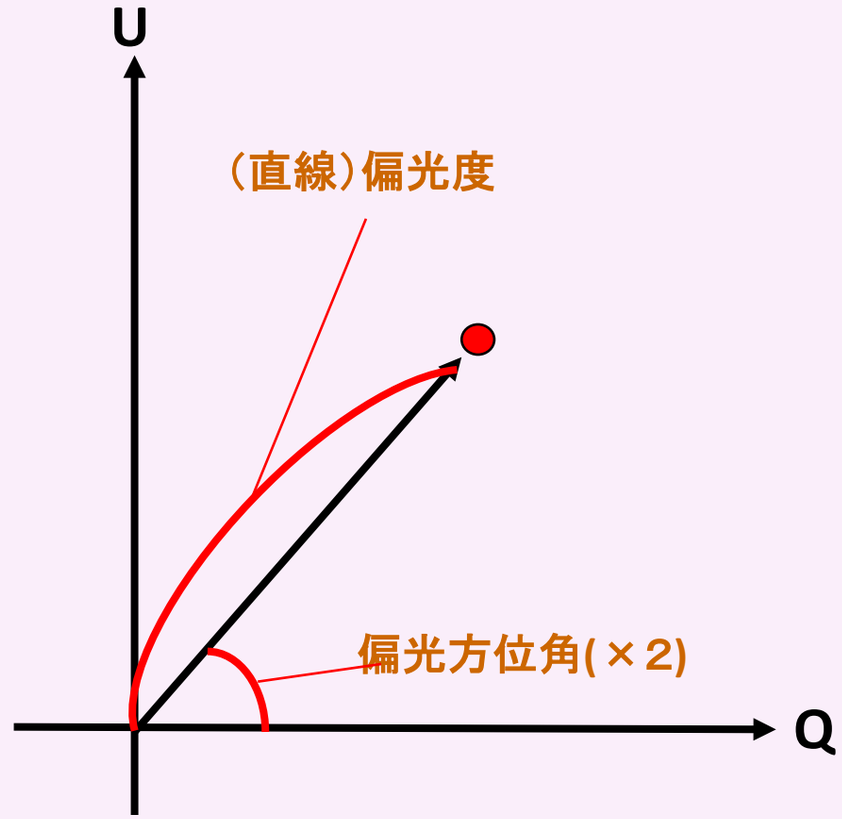
くちゃ

なぜ綺麗な関係性が見えないのか？

- そもそも綺麗な関係性などない　→　不幸
- 複数の放射源からの寄与が混じっているので、綺麗な関係性が見えなくなっている　→検討する価値あり

偏光のストークスパラメータ

- (I, Q, U, V)
- Vは円偏光で、今回は無視
- 偏光度・偏光方位角が加算性の無い量であるのに対し、I, Q, Uは加算性のある量（両者の関係は右図）
- 複数成分の重ね合わせを考える際には Stokes パラメータで考える方が良い。

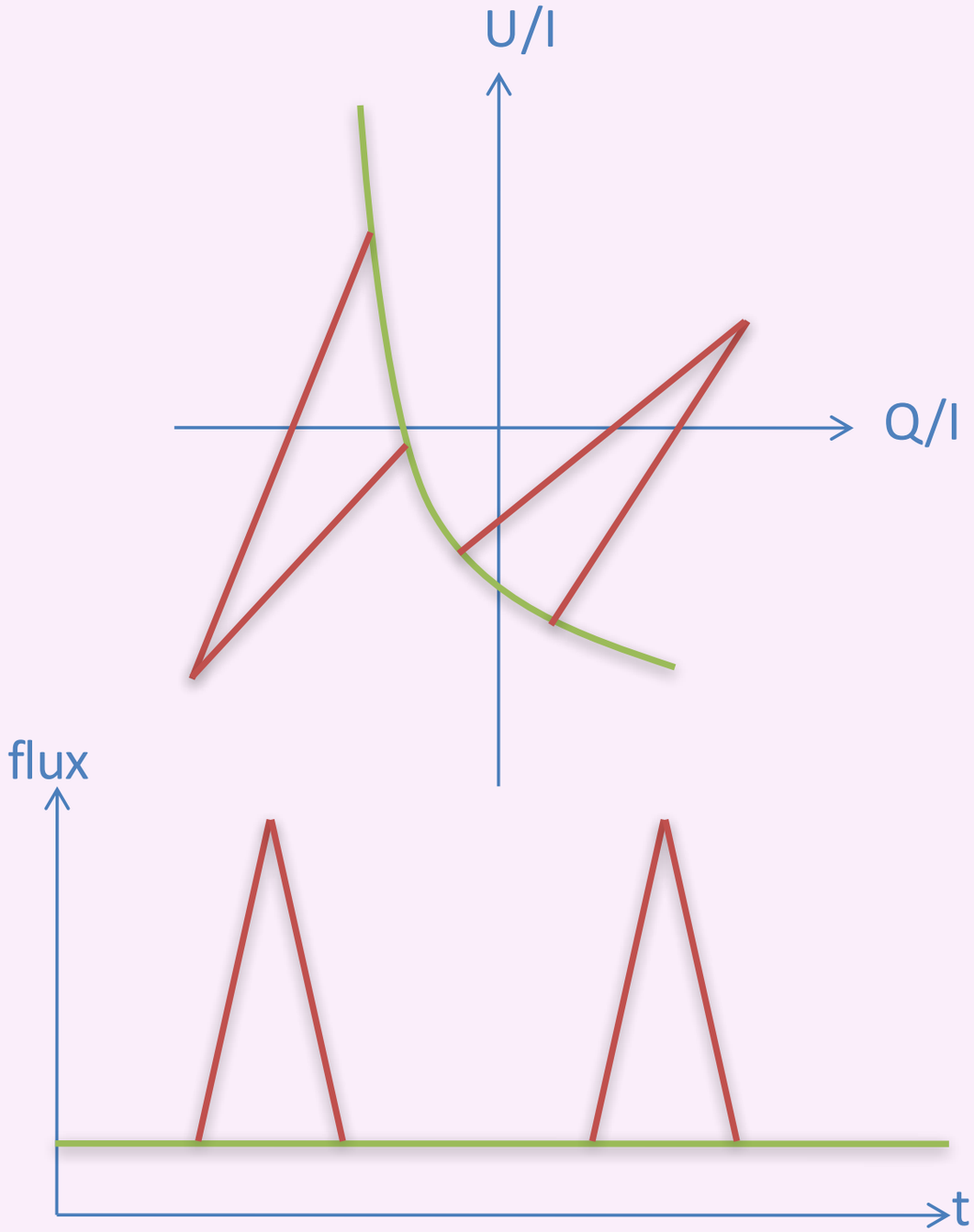


短いフレアと長いトレンドを分離したい

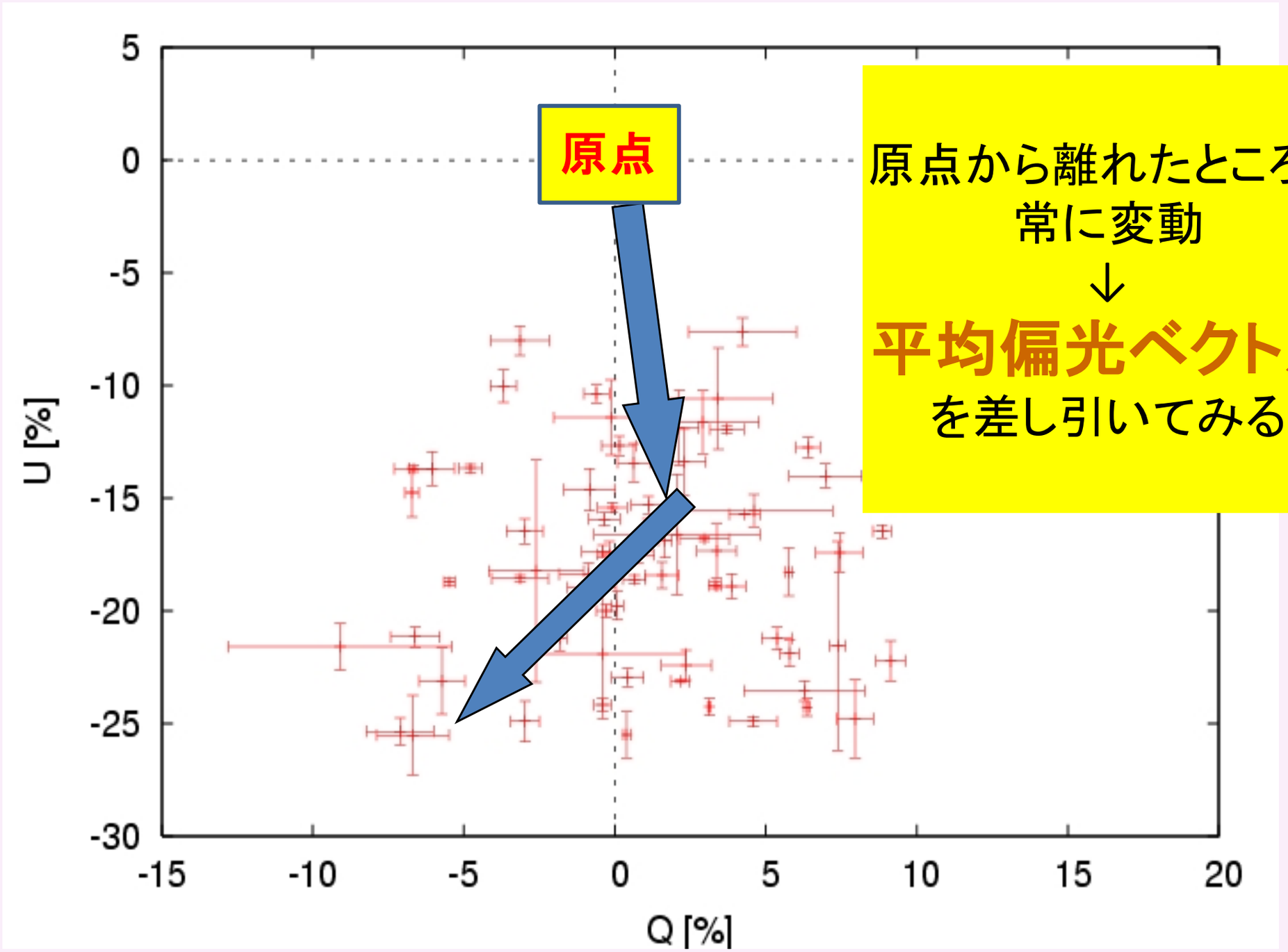
- 仮定：短いフレア（光度も偏光も変わる）が長いトレンドに乗っている

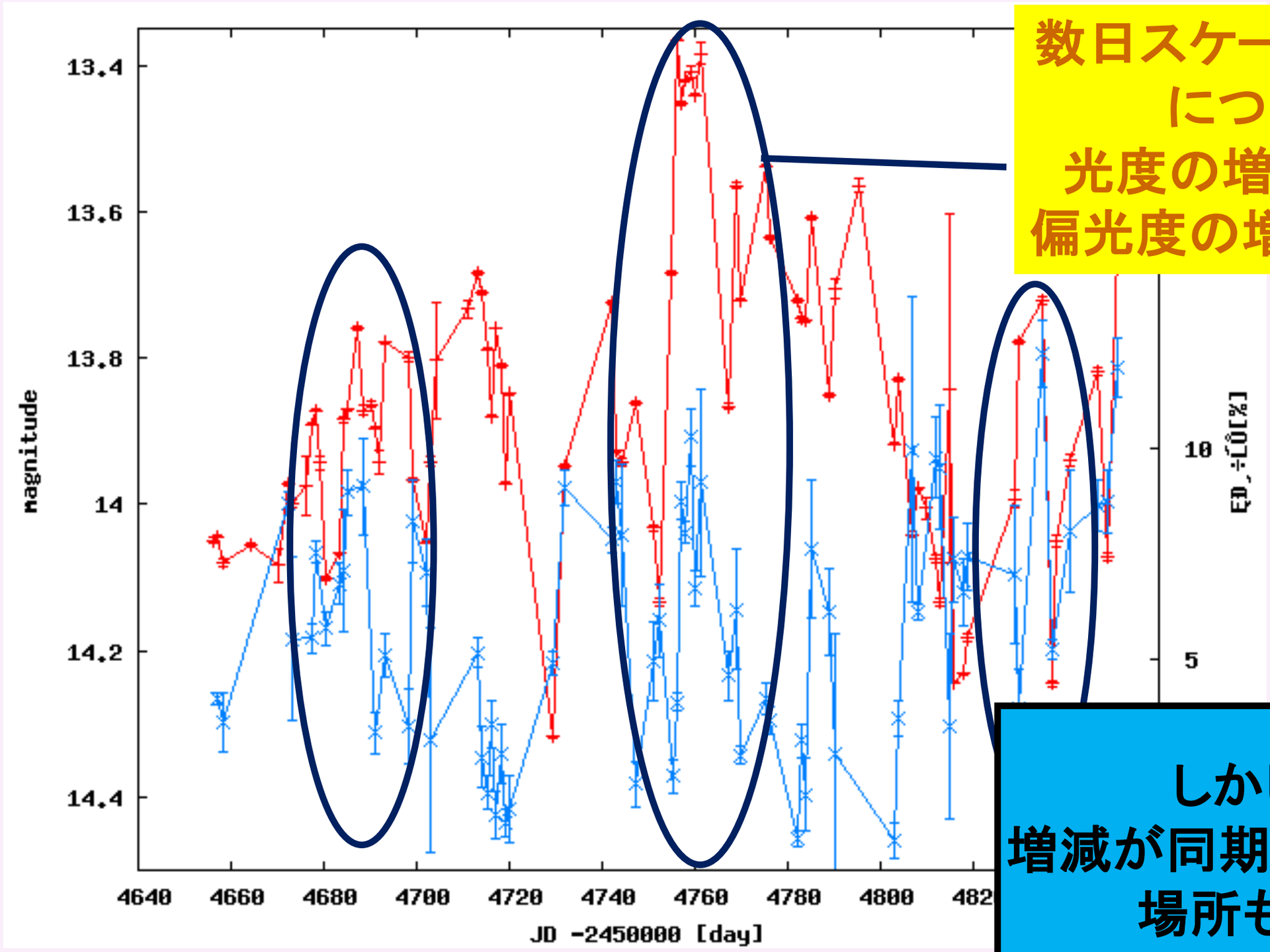
- 実際は特定のタイムスケールが認められるブレーザーは少ない（パワースペクトルはベキ型）

- 一方で、データを撮る間隔（～数日）以下の現象は「ノイズ」としてしか見えず、ある程度以上のタイムスケールしか実際は議論できない



例えば、代表的なブレーザー「BL Lac」の場合





数日スケールの変動
について
光度の増減と共に
偏光度の増減が同期

しかし、
増減が同期していない
場所もあり

→　長期的なトレンドがあるため？

観測値から long-term trend を推定する

Long-term trendを特定の関数で近似する

例：全体の平均で近似。期間毎の平均で近似。

BL Lacの研究（先本、他、天文学会2009春年会）

→　相関らしいのが見えたり、見えなかったり。

→　long-term trendの推定がまずいから？？？

特定の関数に依存しないやり方

事前情報はそれなりにある

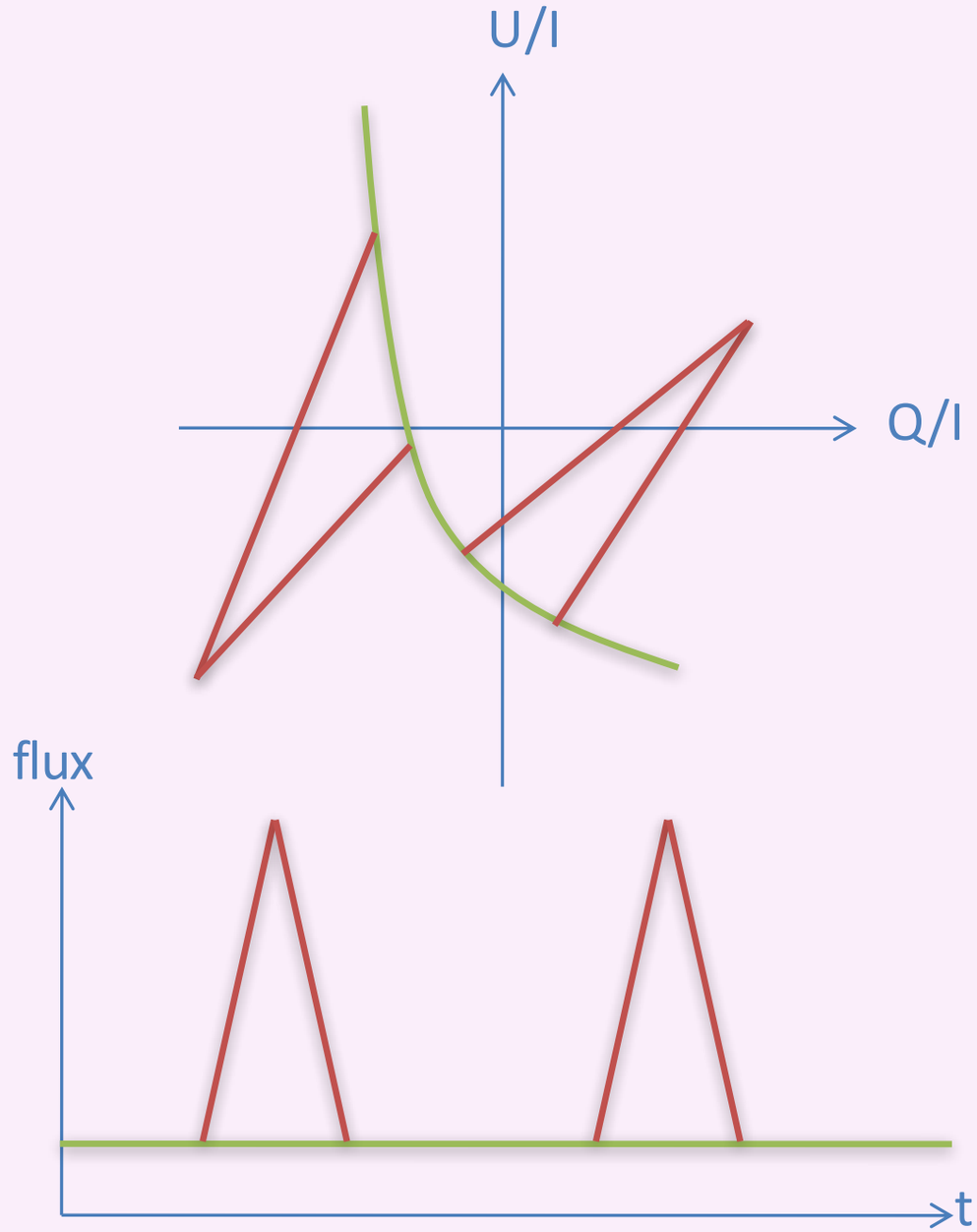
Short-term フレアは固有の偏光成分をもつ（仮説）

→　光度曲線と相関するように long-term trendを推定

→　ただし、それだけだと一意には決まらない

Long-term trendはフレアと比較して「ゆっくり」

「滑らかに」変化している



Long-term trendをベイズ的に推定

偏光は長期トレンドと短期フレアの２成分

$$\begin{cases} Q_{\rm obs} = Q_{\rm L} + Q_{\rm S}, \\ U_{\rm obs} = U_{\rm L} + U_{\rm S}, \end{cases}$$

短期フレアの偏光フラックス

$$\mathrm{PF_S} = \sqrt{(Q_{\rm S}^2 + U_{\rm S}^2)}.$$

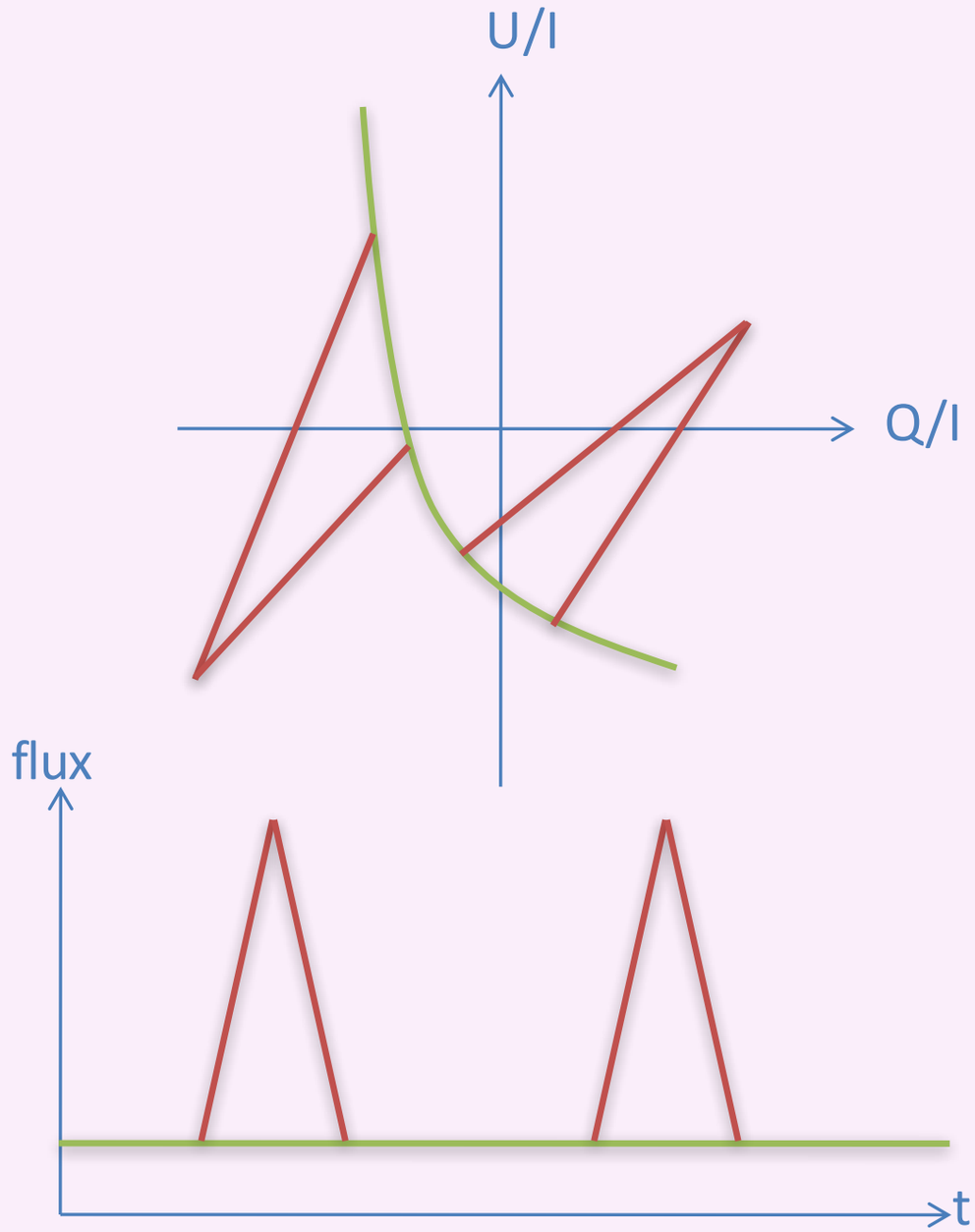
尤度関数：短期フレアの偏光フラックスが総光度に比例

$$L(\mathbf{y}|\mathbf{x}) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2_{\mathrm{PF}',i}}} \exp\left[-\frac{(I'_{{\rm obs},i} - \mathrm{PF}'_{{\rm S},i})^2}{2\sigma^2_{\mathrm{PF}',i}}\right]$$

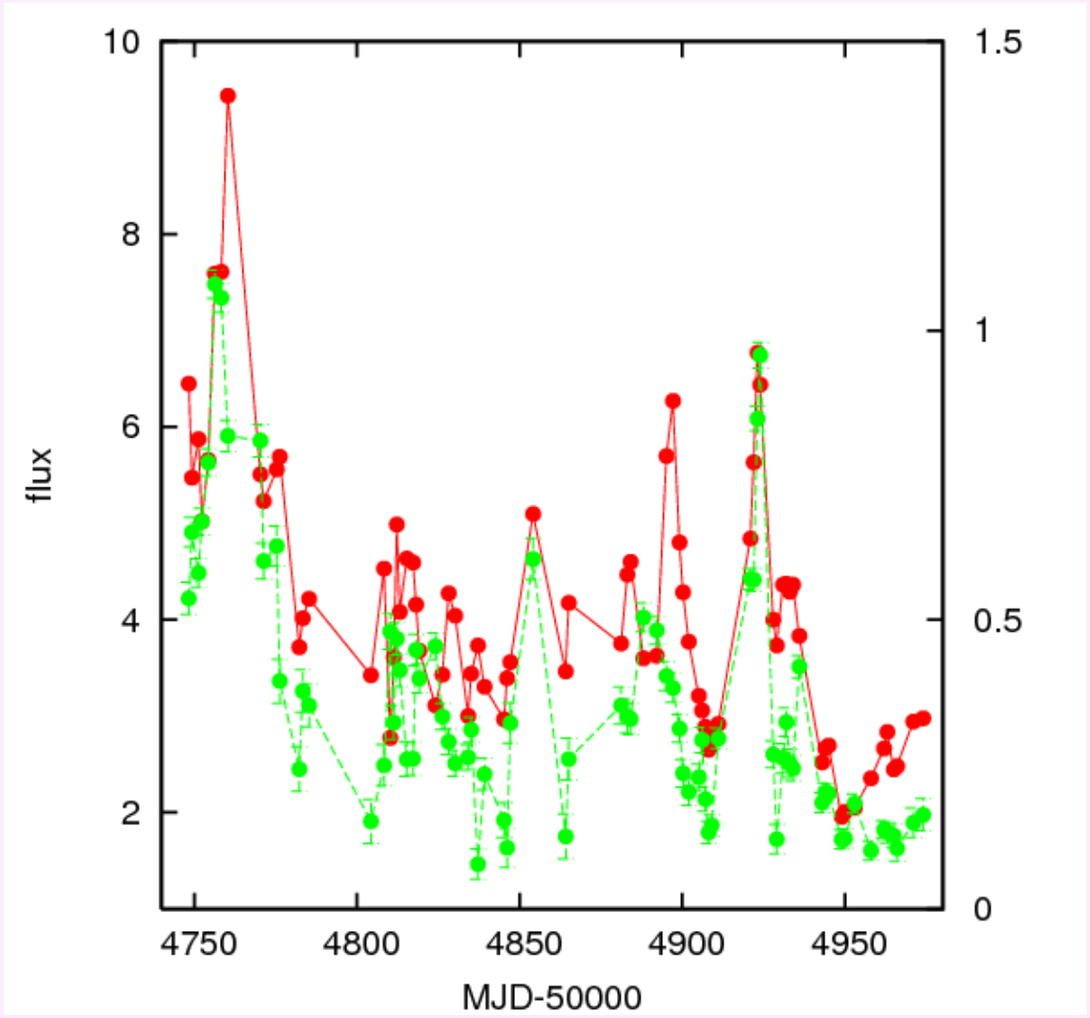
事前分布：長期トレンドは滑らかにゆっくり動く

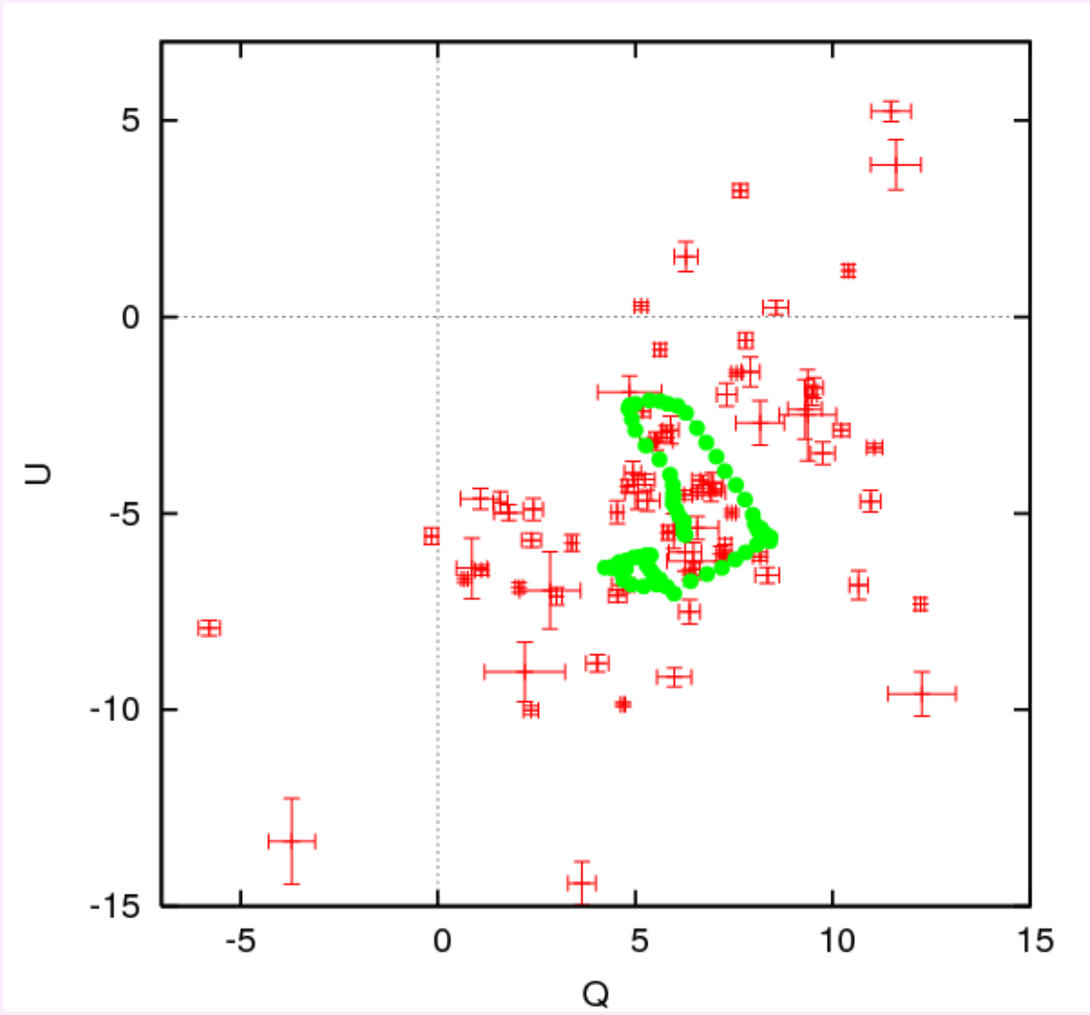
$$\pi(\{Q_{{\rm L},i}\}) = \prod_i \frac{1}{\sqrt{2\pi w^2}} \exp\left[-\frac{(Q_{{\rm L},i} - Q_{{\rm L},i-1})^2}{2w^2}\right],$$

$$\pi(\{U_{{\rm L},i}\}) = \prod_i \frac{1}{\sqrt{2\pi w^2}} \exp\left[-\frac{(U_{{\rm L},i} - U_{{\rm L},i-1})^2}{2w^2}\right].$$



例：ブレーザー
OJ 287の場合





- 長期成分は一定の角度内を振動
- 補正すれば偏光フラックスと光度曲線とよく相関
- 2成分モデルの理想的な例

課題

- MCMCで最適化してるが、収束性が悪い。多分モデルが良くない。

- 光度曲線とフレアの偏光の相関が高いことを尤度関数として扱っていて、モデルとして無理矢理な感じ。→より適切な統計モデルにしたい

- 光度曲線とフレアの偏光が相関する＝偏光度が常に一定であることを仮定していて、現実的でない。

- 長期トレンドの光度変化は無視している。

今日話したこと

- 最近の取り組み　その１

「AGB星の炭素過剰星・酸素過剰星の測光データを用いた機械分類

ースパースロジスティック回帰ー」

（安部、植村、板、松永、池田）

- 最近の取り組み　その２

「ブレーザーのSEDモデルパラメータのMCMCによる推定」

（山田、植村、深澤）

- 本題

「ブレーザーの偏光時系列データからジェット内で起きていることに迫る」

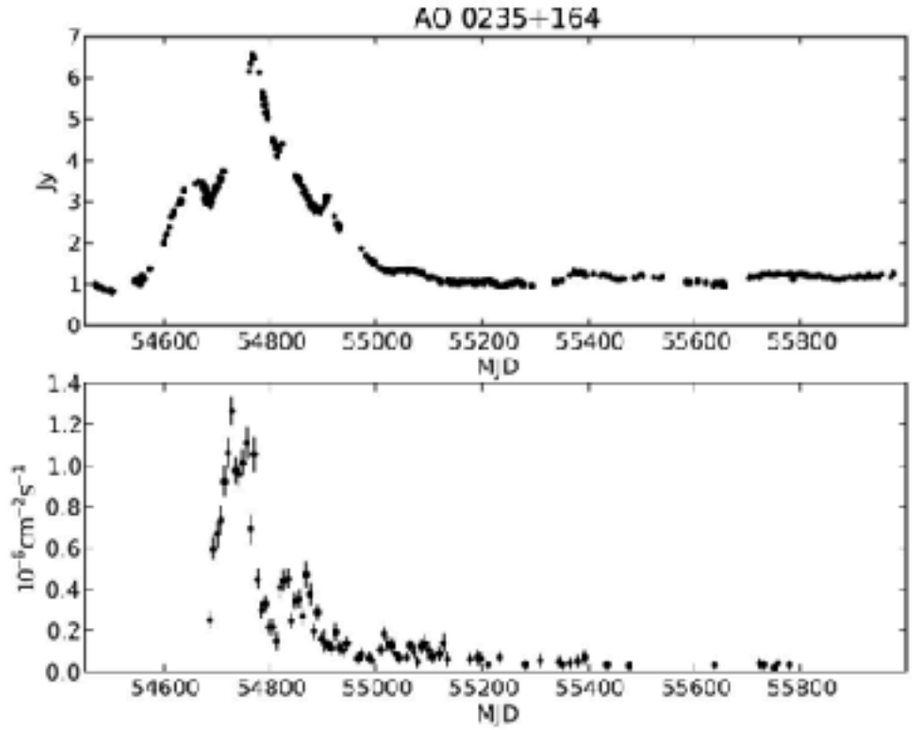
今後やりたいこと
不規則な変動をする天体のデータ処理

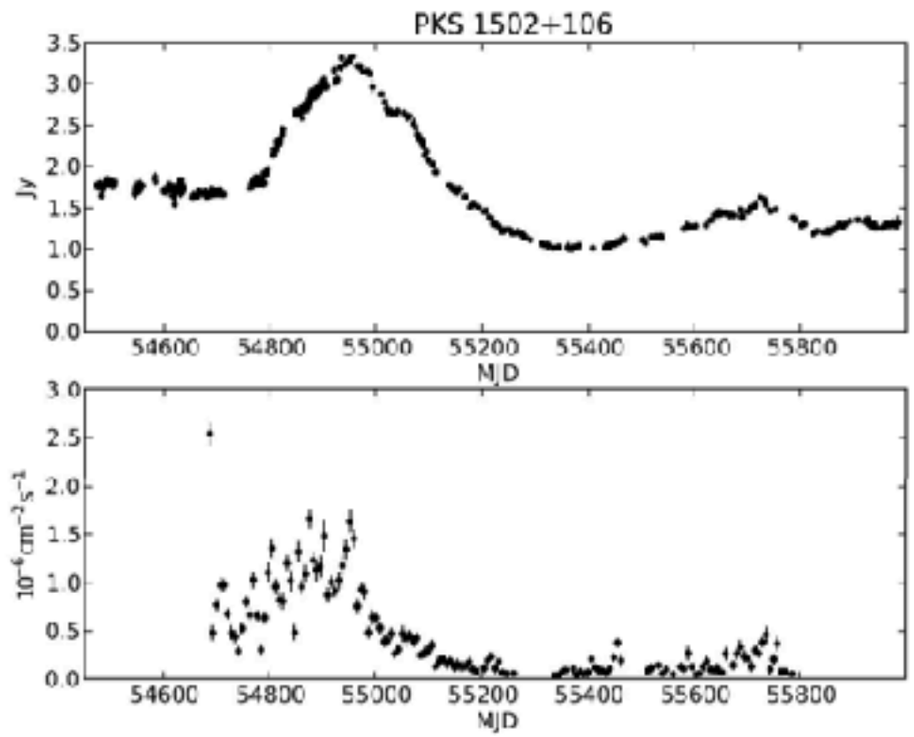
- 不規則変光星の光度曲線から特徴量抽出
  →機会学習による判別

- ブレーザーの時系列SEDのモデルパラメータ推定

- ブレーザーの偏光データの成分分離

- タイムラグのある相関（木邑さんトーク）
  電波ーγ線のタイムラグ、スケールの違い込みで。





Max-Moerbeck+14