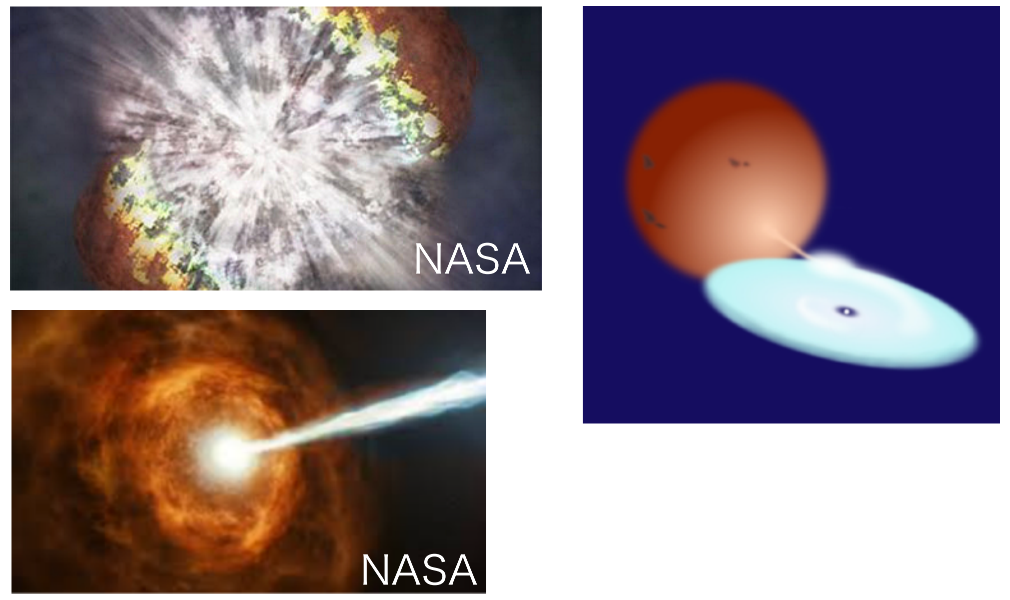
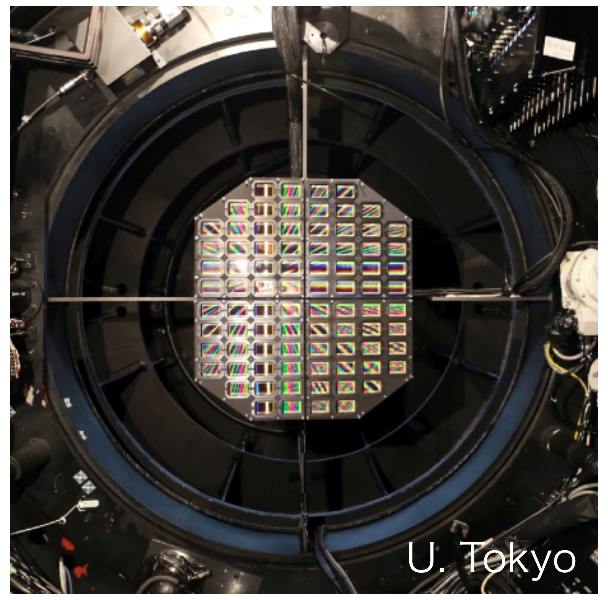
Follow-up observations of galactic transients with astroinformatics

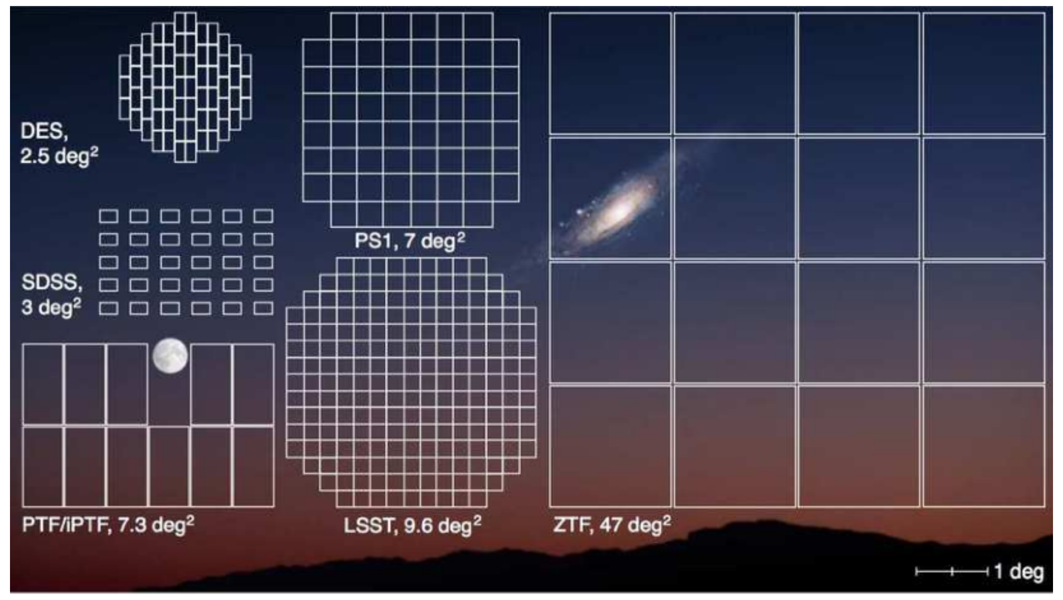
Makoto Uemura (Hiroshima University) @IAU-IAA seminar on 11 Jan. 2022

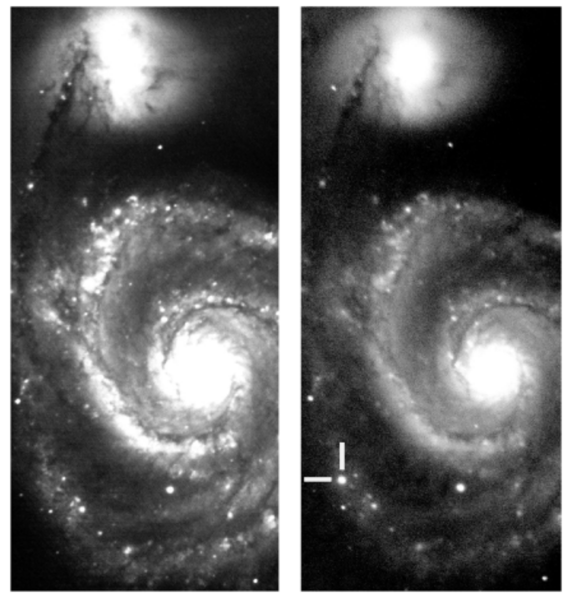
Time Domain Astronomy

Detections and follow-up observations

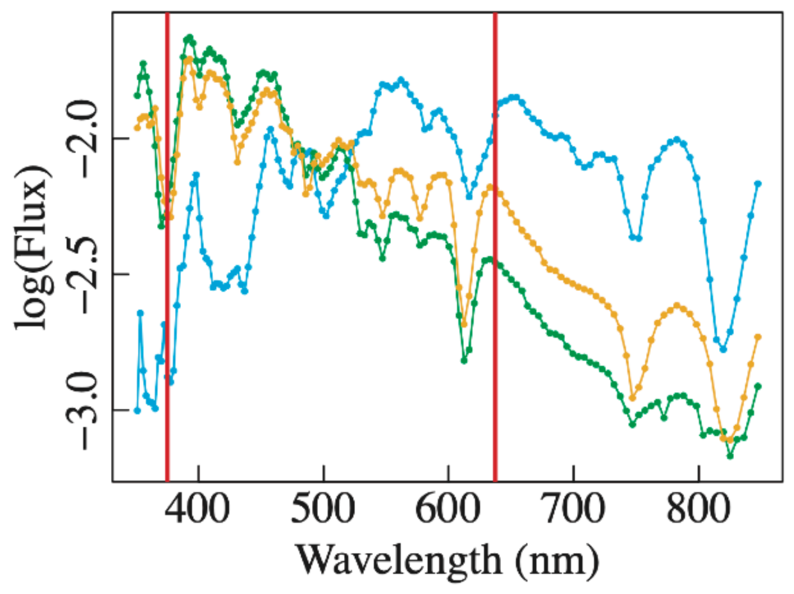


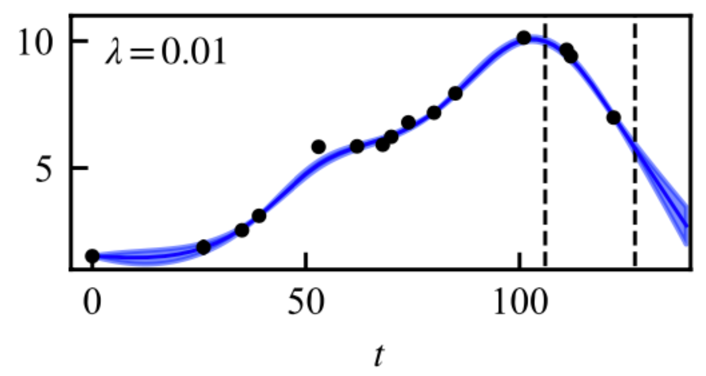






Laher+17





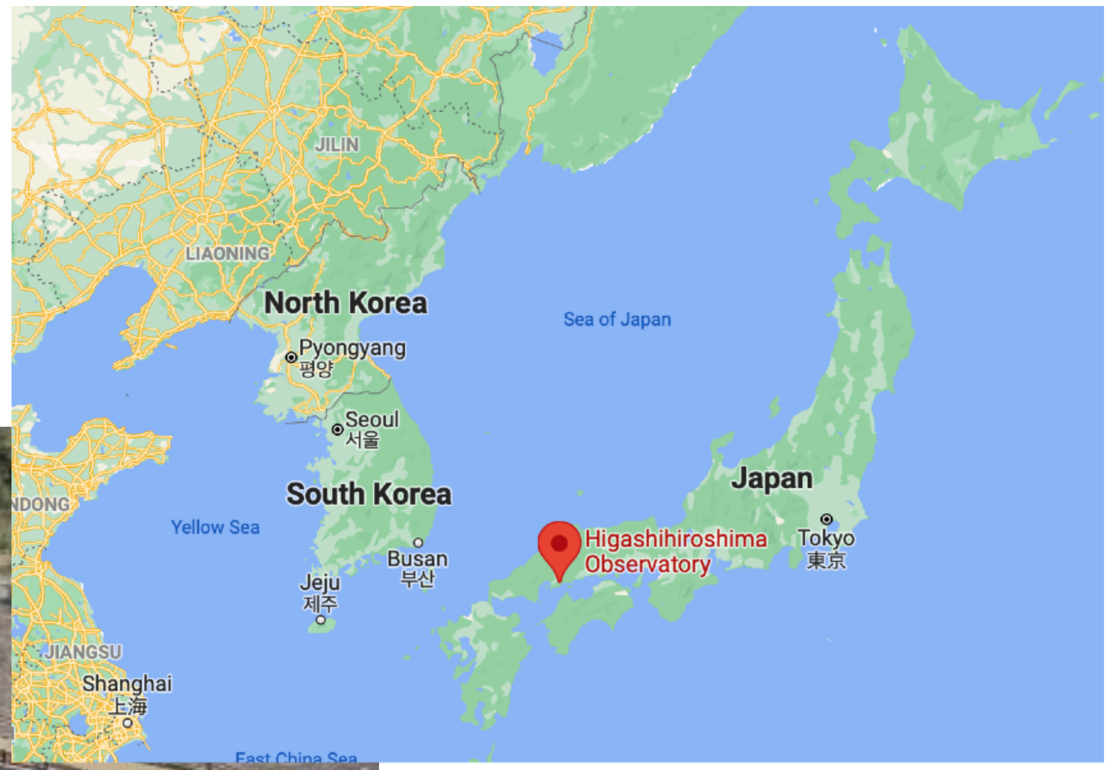
- **Transients**
  - Extragalactic: GRBs, supernovae, GWs, FRBs
  - galactic: novae, dwarf novae, X-ray binaries, stellar flares

- **Detections**
  - ≲90's : Discovered by amateurs
  - Currently: All-sky surveys with large telescopes and detectors
    - Ground-based: ZTF, ASAS-SN, MASTER, ATLAS, Tomo-e, (LSST) etc.
    - Space: Gaia, MAXI, Swift, etc.

- **Follow-up observations**
  - Spectra
  - Multi-band and/or multi-wavelength data
  - Time-series data for short-term variations (< 1d)
  - Long follow-ups.

Our observatory

1.5-m telescope “Kanata” in Hiroshima



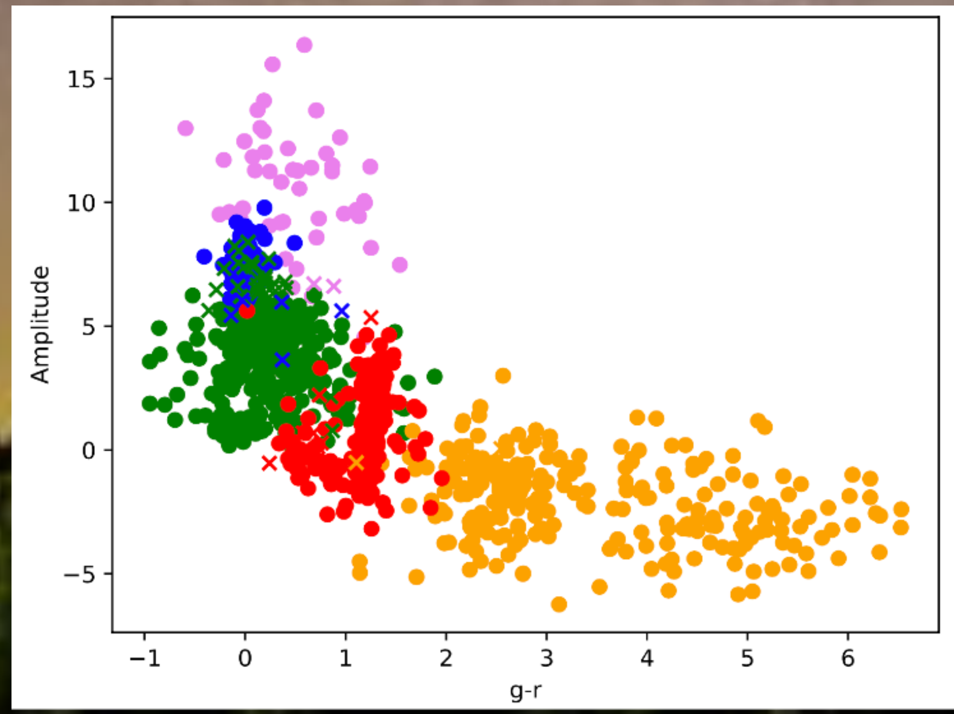
- Used for the follow-up obs. of transients.
    - since 2006
    - GRBs, SNe, GWs, high energy neutrino events, novae, dwarf novae
    - Multi-wavelength team of Fermi
- Simultaneous optical and NIR imaging, low-resolution spectroscopy, and (linear) polarimetry
- Application of modern statistical/machine learning methods —> smart follow-up obs.

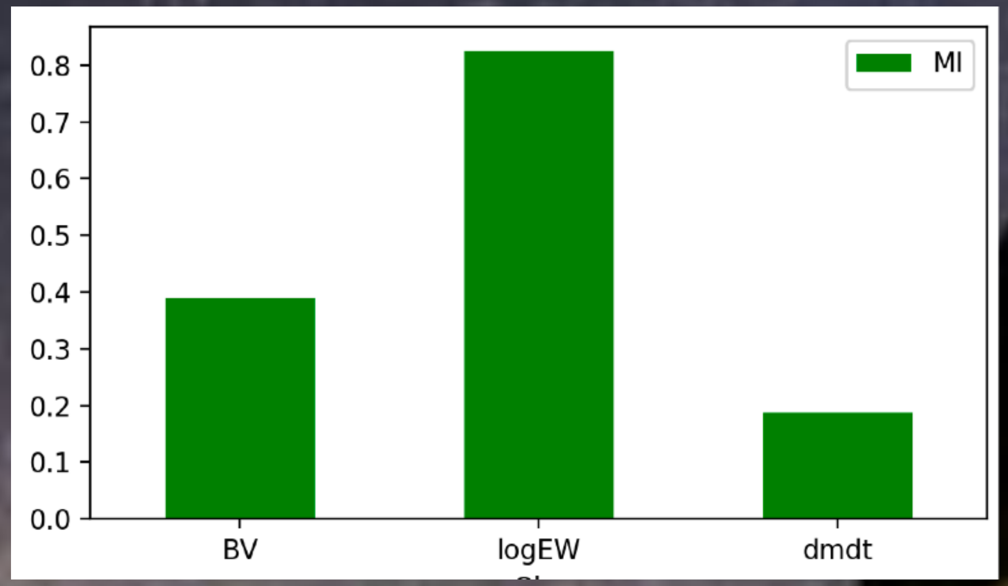
Outline

Case studies of astroinformatics in the follow-up studies of transients

1. Classification of new events

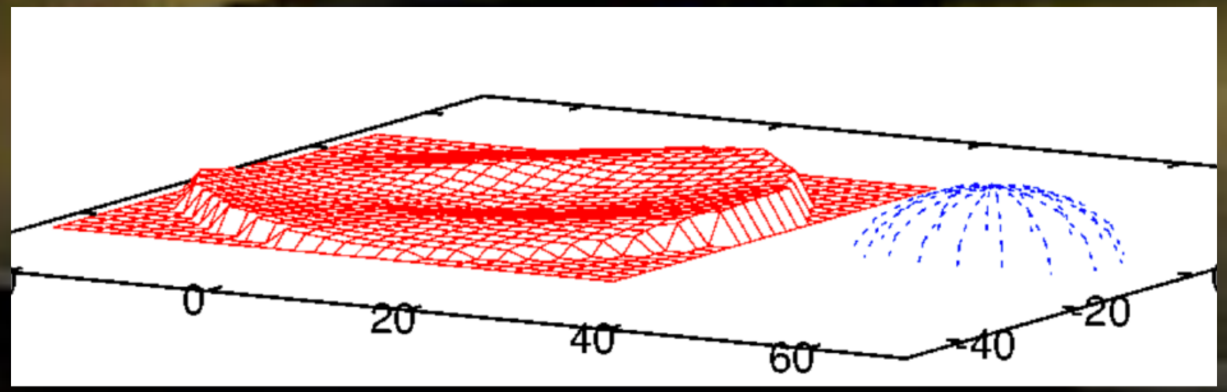
- Sparse multinomial logistic regression
- Generative model





2. Decision making of follow-up obs.

- Information theory, mutual information



3. Study of the accretion disk structure

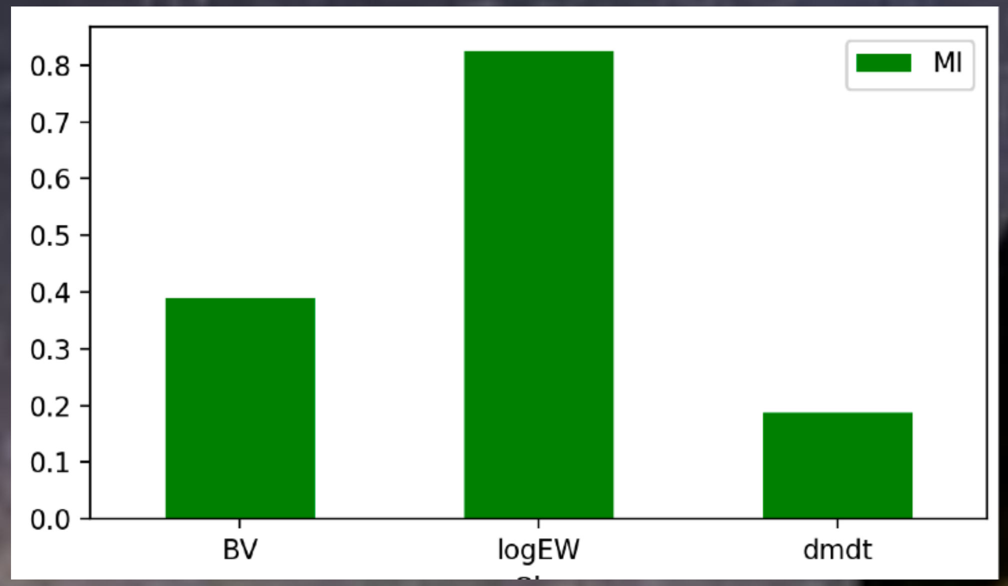
- Tomography, ill-posed inverse problem

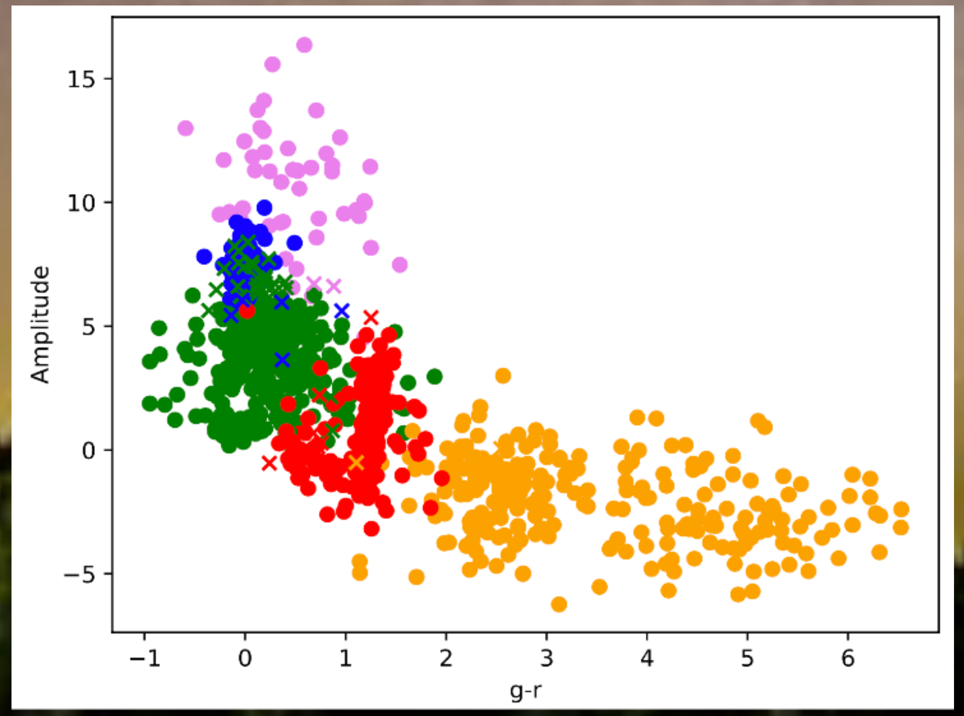
4. (if we have time left,) Tools to support obs.

Outline

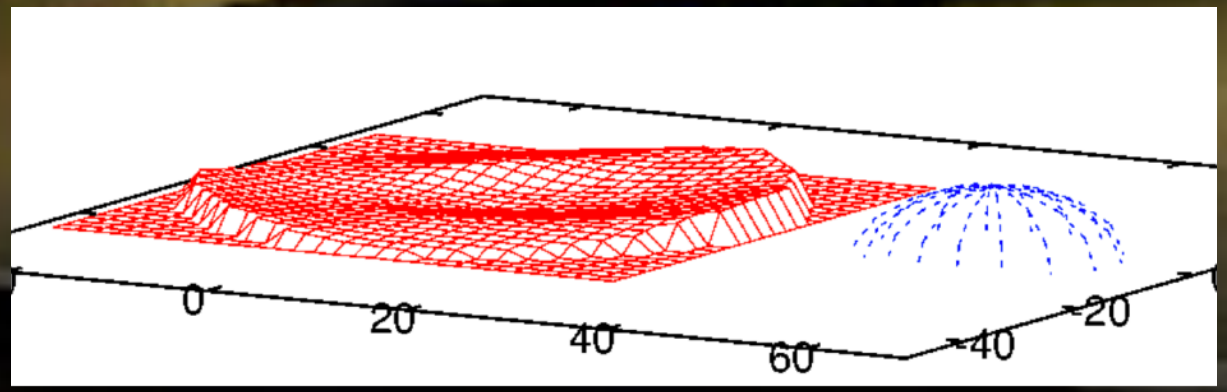
Case studies of astroinformatics in the follow-up studies of transients

1. Classification of new events
   - Sparse multinomial logistic regression
   - Generative model





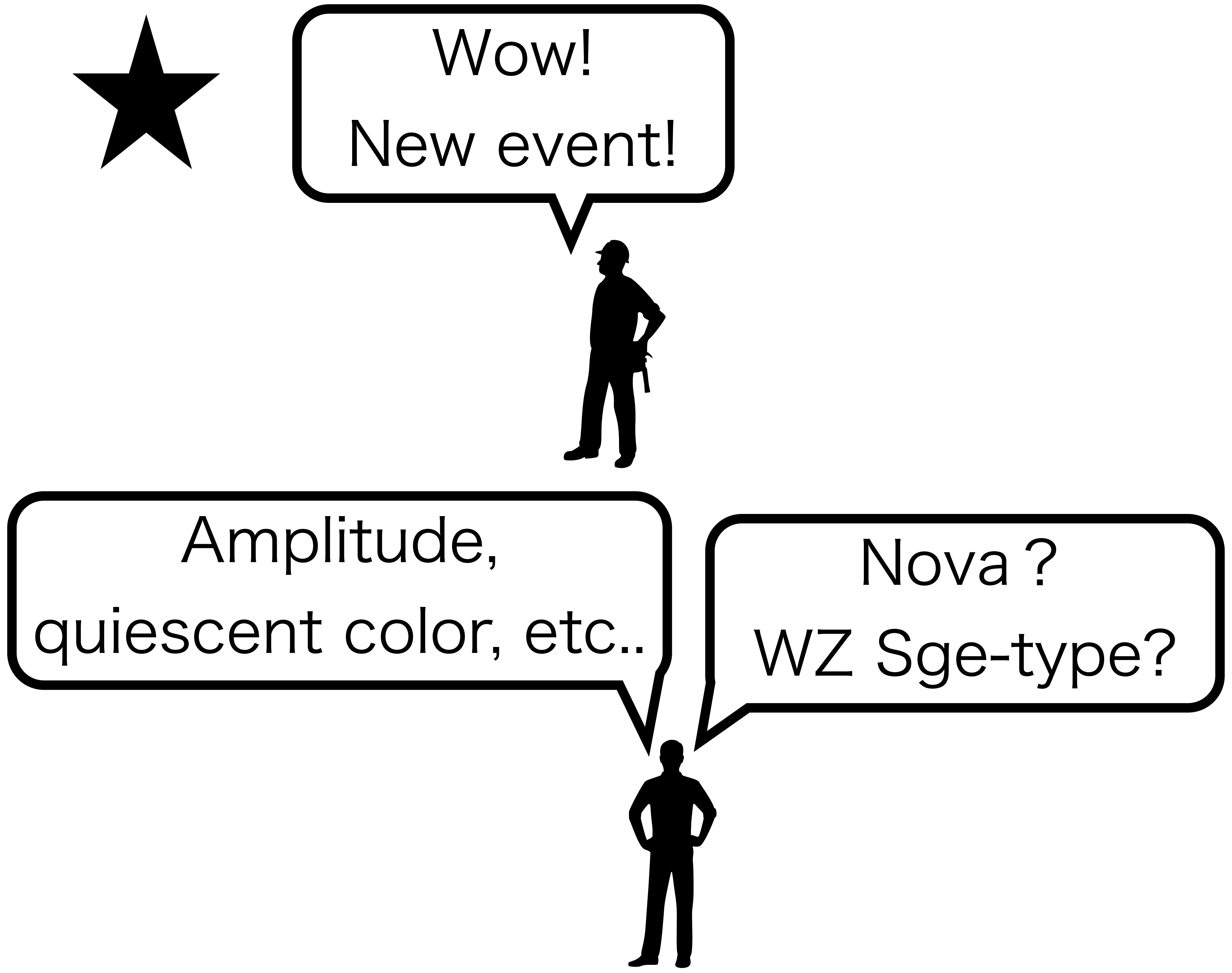
2. Decision making of follow-up obs.
   - Information theory, mutual information



3. Study of the accretion disk structure
   - Tomography, ill-posed inverse problem

4. (if we have time left,) Tools to support obs.

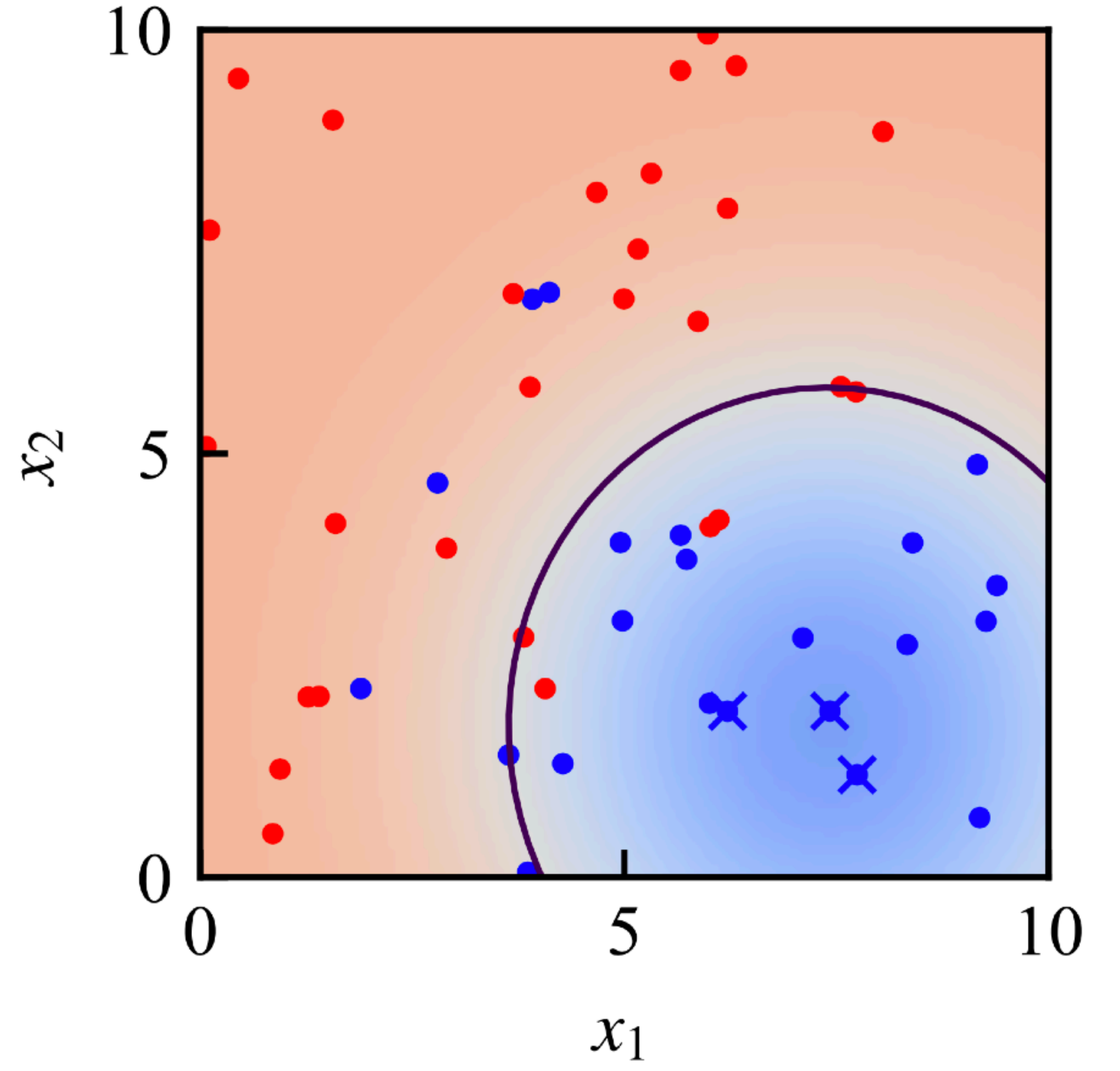
Classification model of transients



- Previously,
  - Domain expert judgement based on knowledge
- Currently,
  - Machine learning with training data

- Our classification problem with 5 classes
  - (classical/recurrent) Novae (N)
  - (normal) Dwarf Novae (DN)
  - WZ Sge-type dwarf novae (WZ)
  - Mira
  - Flare Stars (UV Ceti-type variables, UV)
- Popular classification models
  - Support Vector Machine (SVM)
  - Random forest
  - Generative model
  - Sparse multinomial logistic regression (SMLR)

Sparse multinomial logistic regression



Multinomial logistic regression

Bayes' theorem
$$p(C_k|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|C_k)p(C_k)}{\sum_j p(\boldsymbol{x}|C_j)p(C_j)}$$

Softmax function
$$= \frac{\exp(a_k)}{\sum_j \exp(a_j)}$$

$$a_k = \log(p(\boldsymbol{x}|C_k)p(C_k)) \quad = \quad \boldsymbol{w}_k^T \boldsymbol{x}$$

(joint) Posterior
$$p(\boldsymbol{Y}|\boldsymbol{W}) = \prod_i \prod_k \left\{ \frac{\exp(\boldsymbol{w}_k^T \boldsymbol{x}_i)}{\sum_j \exp(\boldsymbol{w}_j^T \boldsymbol{x}_i)} \right\}^{y_{ik}}$$

Loss function (-log P)
$$E(\boldsymbol{W}) = -\sum_i \sum_k y_{ik} \log \frac{\exp(\boldsymbol{w}_k^T \boldsymbol{x}_i)}{\sum_j \exp(\boldsymbol{w}_j^T \boldsymbol{x}_i)}$$

Kernel + Sparse

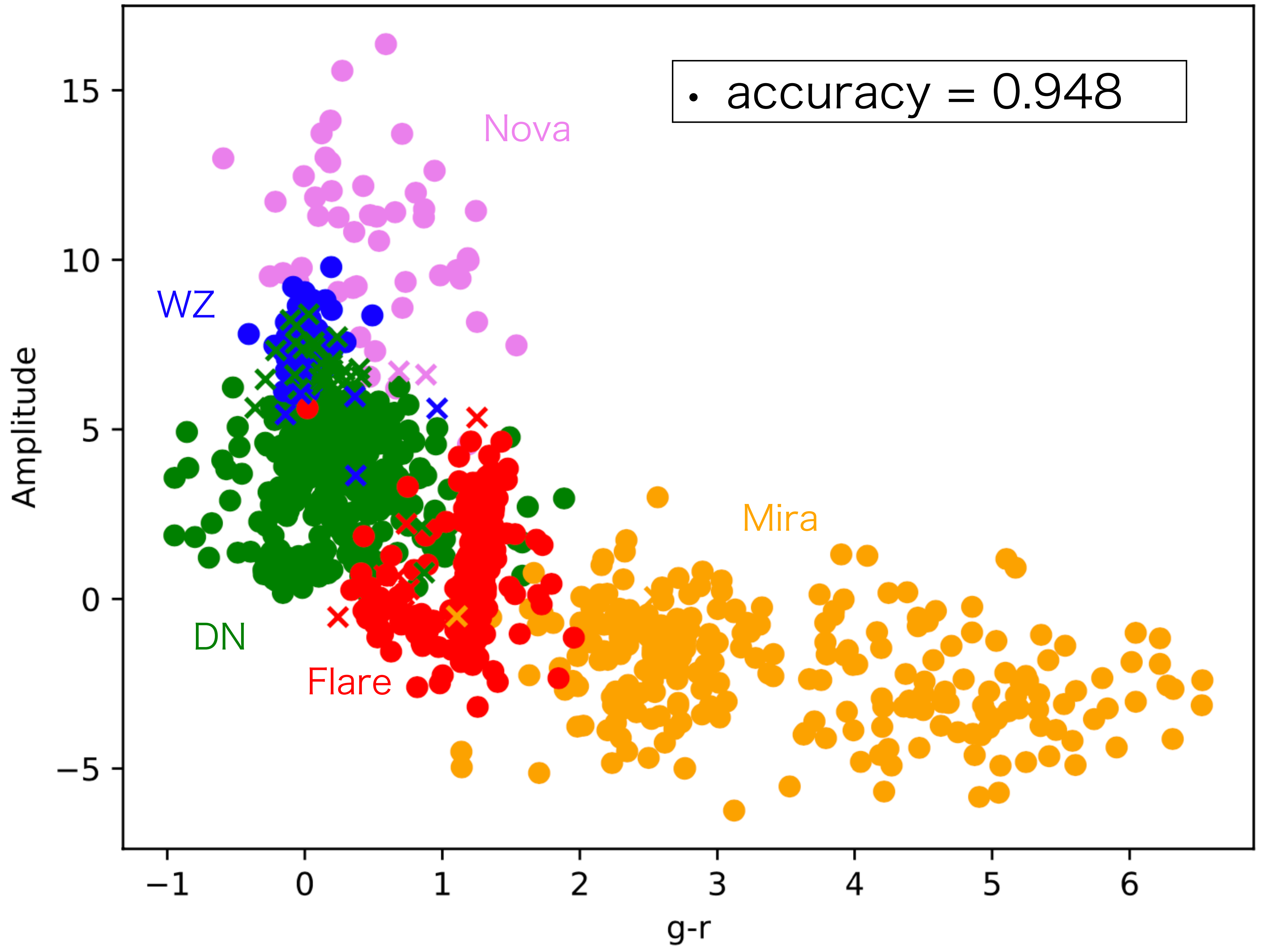
$$\begin{aligned} p(y_i = 1|\boldsymbol{\phi}_i) &= \sigma(\boldsymbol{w}^T \boldsymbol{\phi}_i) \\ &= \sigma\left((\sum_j \alpha_j \boldsymbol{\phi}_j)^T \boldsymbol{\phi}_i \right) \\ &= \sigma\left(\sum_j \alpha_j \boldsymbol{\phi}_j \boldsymbol{\phi}_i \right) \\ &= \sigma\left(\boldsymbol{\alpha} \boldsymbol{k}(\boldsymbol{x}_i, \boldsymbol{x}_j) \right) \end{aligned}$$

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2}{2\sigma^2} \right)$$ RBF kernel

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \{E(\boldsymbol{w}) + \lambda \|\boldsymbol{w}\|_1\}$$ SMLR solution

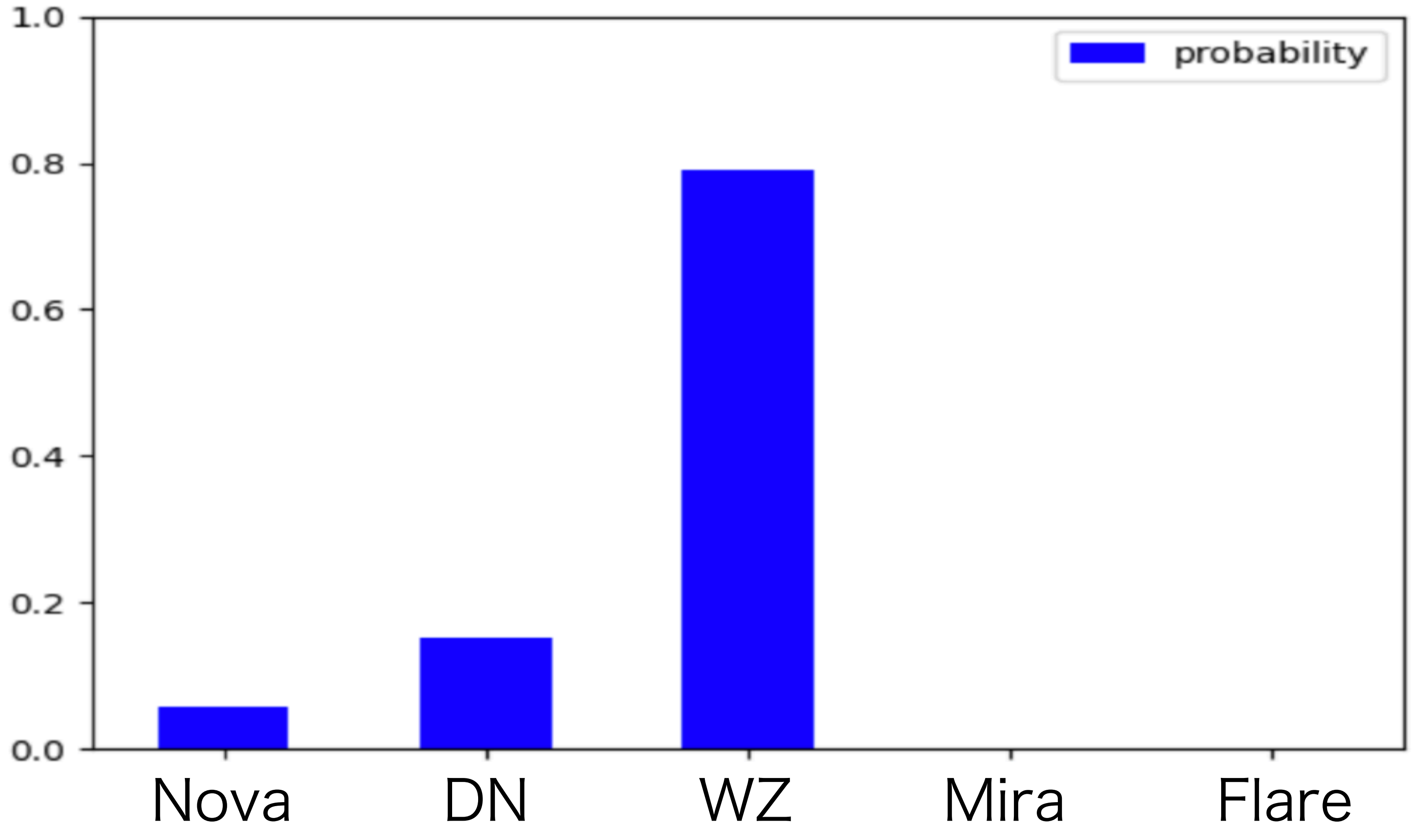
Examples

- 10 features (l, b, d, gal_h, abs_mag_out, abs_mag_qui, amplitude, g-r, r-i, i-z)
- 1045 samples



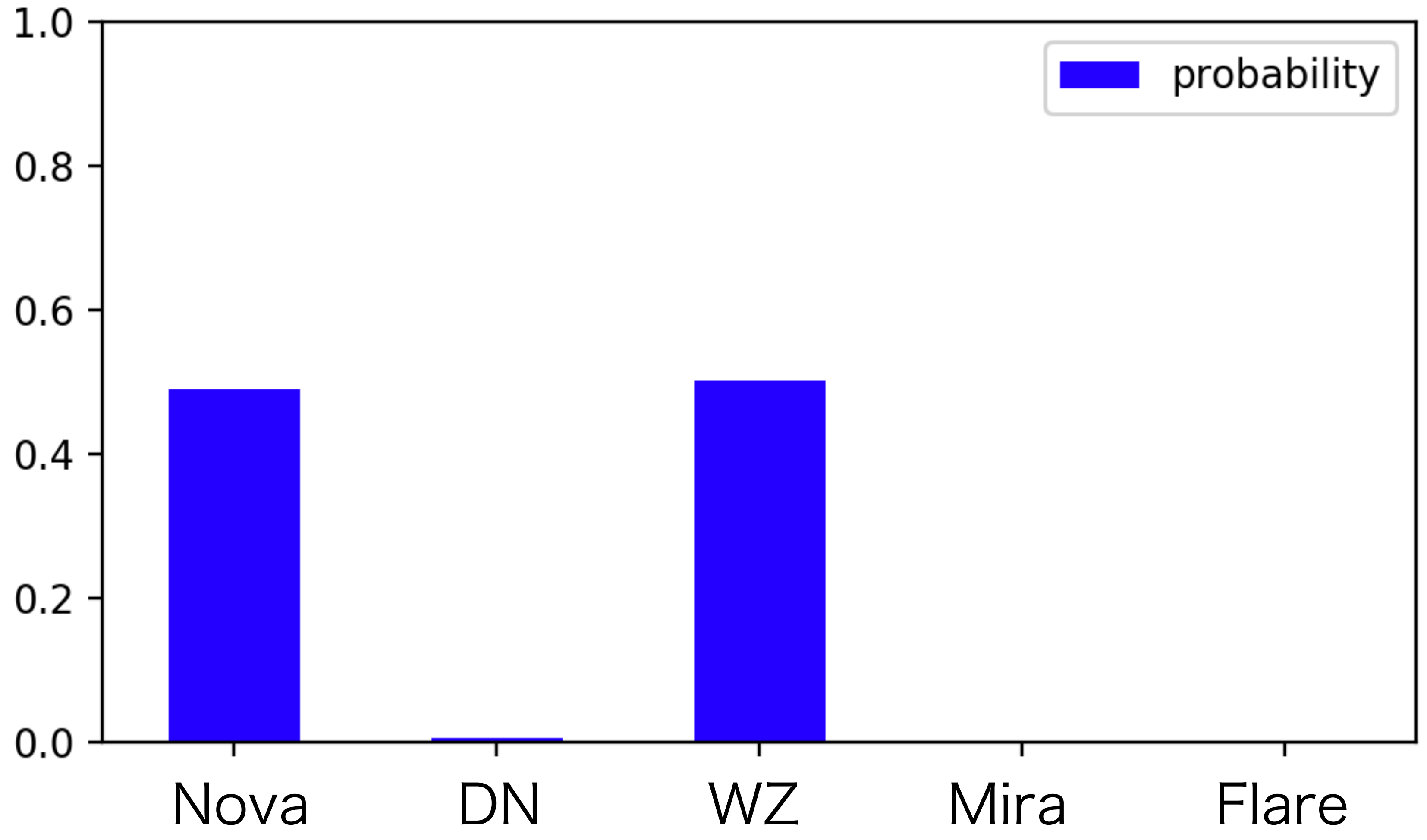
XM78HT

6 features (l, b, amplitude, g-r, r-i, i-z), 1427 samples



TCP J20034647+1335125

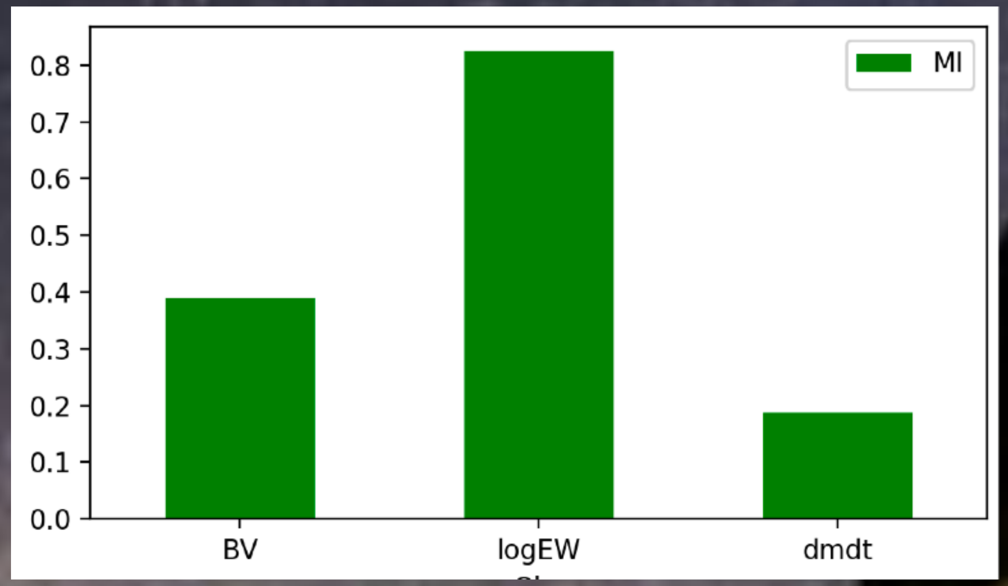
5 features (l, b, amplitude, g-r, r-i), 1492 samples

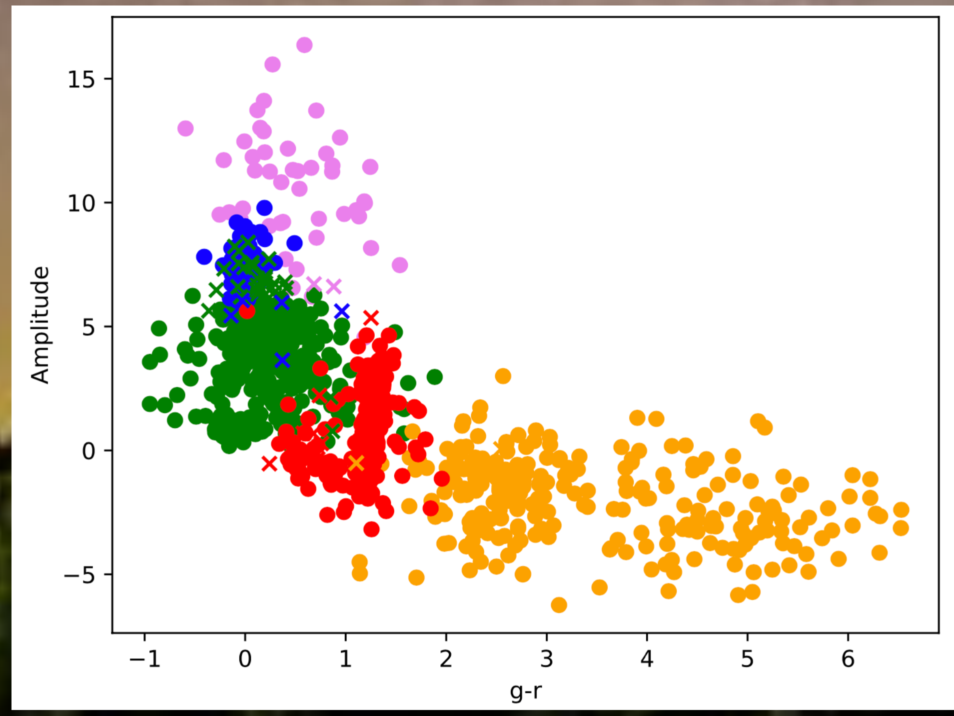


Outline

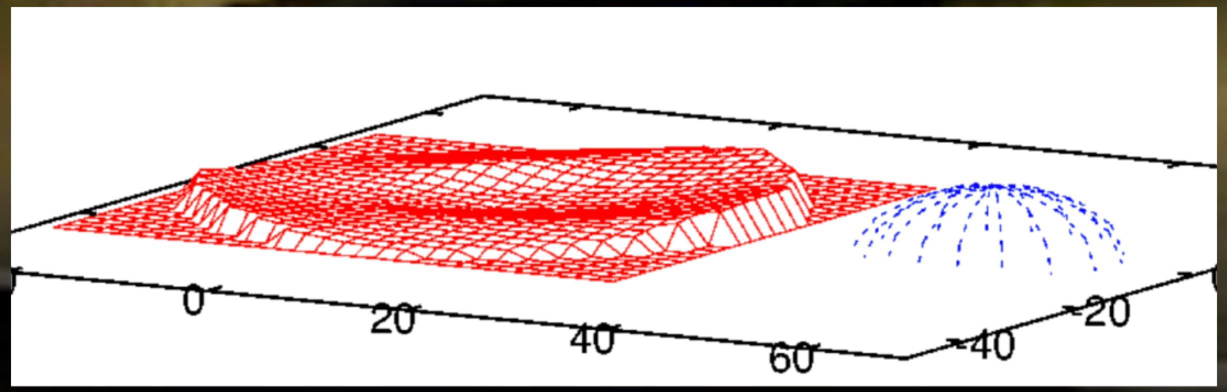
Case studies of astroinformatics in the follow-up studies of transients

1. Classification of new events
    - Sparse multinomial logistic regression
    - Generative model





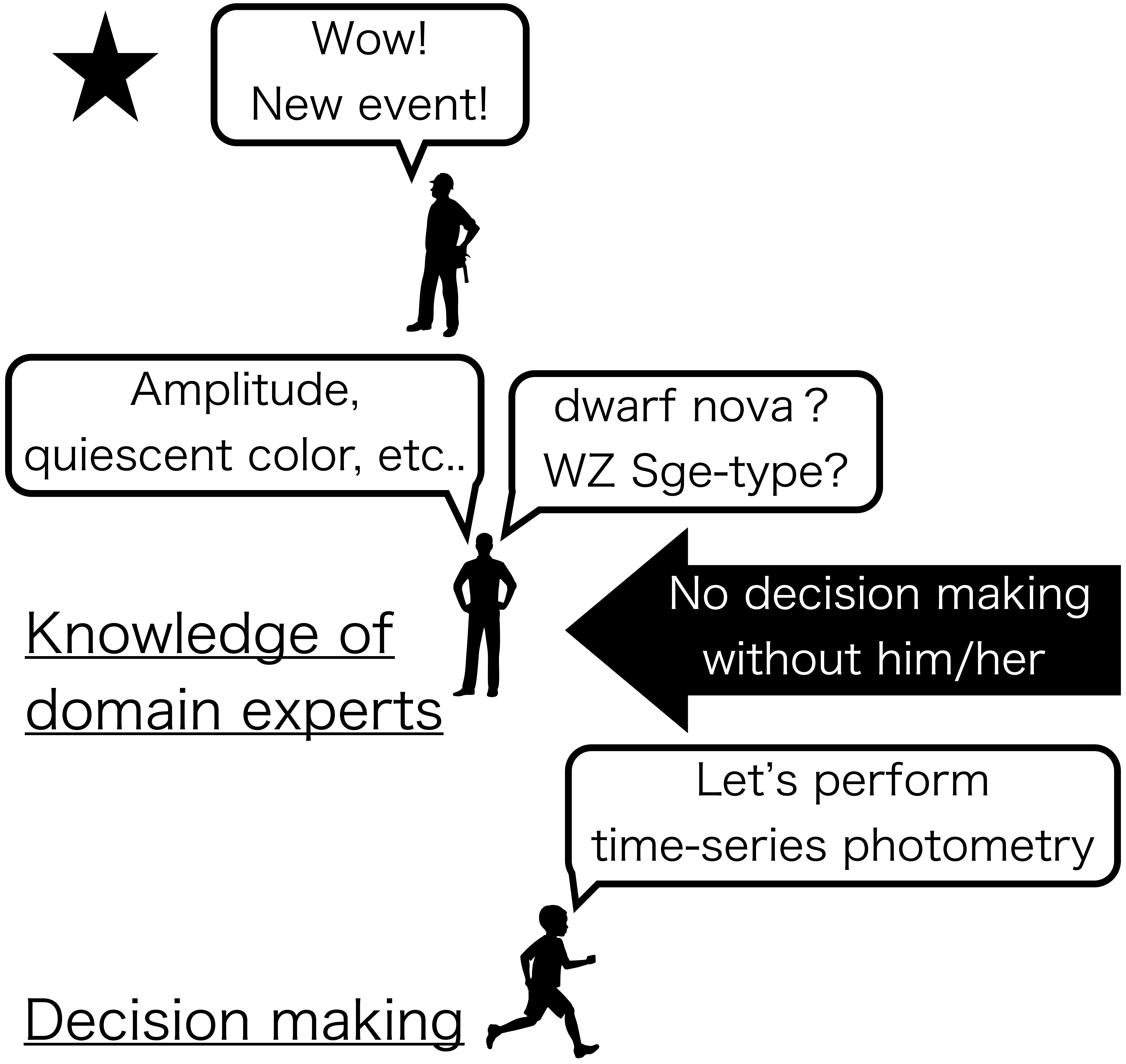
2. Decision making of follow-up obs.
    - Information theory, mutual information



3. Study of the accretion disk structure
    - Tomography, ill-posed inverse problem

4. (if we have time left,) Tools to support obs.

Automatic decision making for follow-up obs.

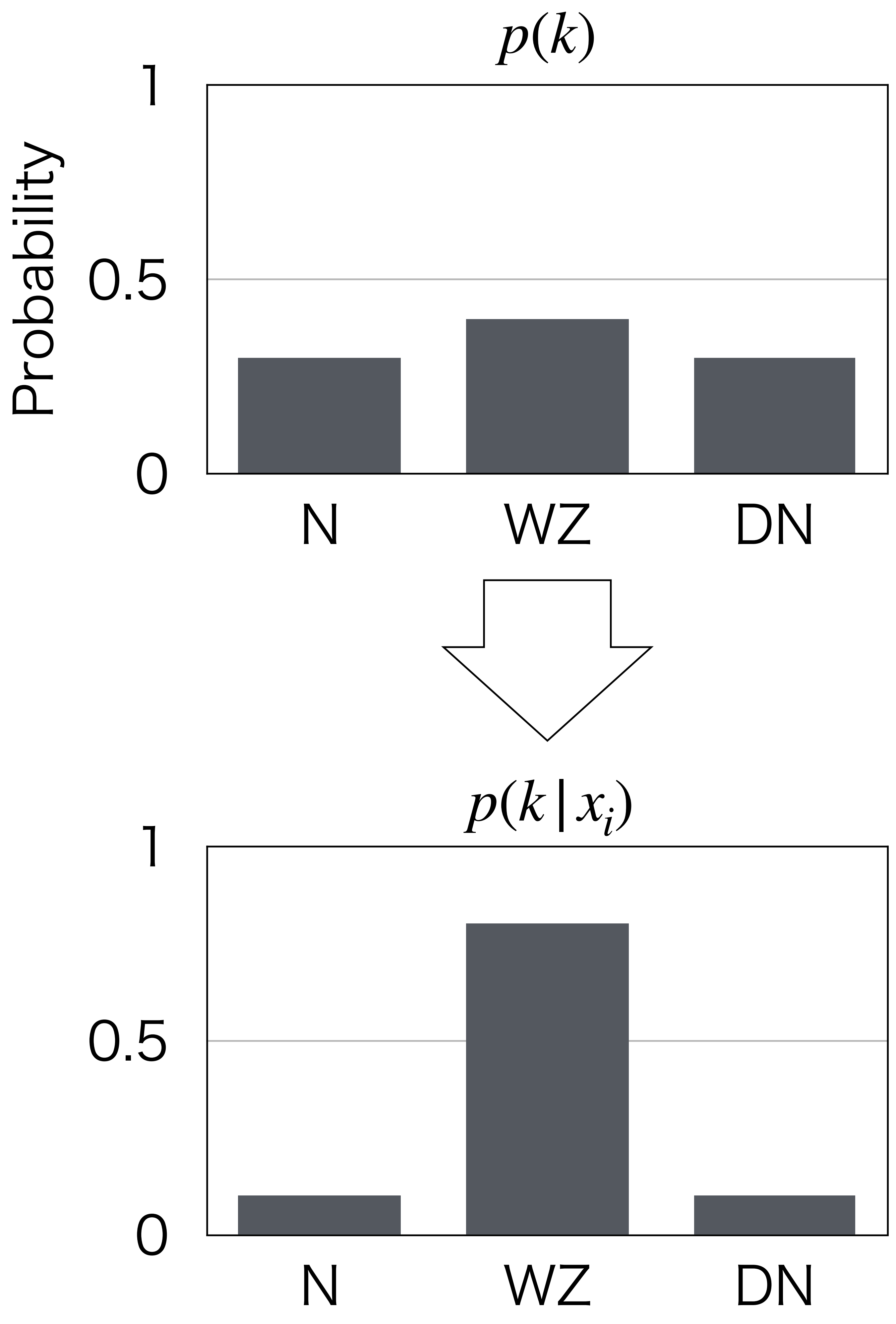


- Simple task
  - GRB: First, take an image. If it is bright, take a spectrum.
  - SN: If it is bright, take a spectrum. If not, take images.
- Complicated task
  - Nova? DN? WZ-Sge-type?: Spectrum? Time-series imaging?
- Previously, experts had made the decision.
  - No data without the experts at observatory

—> automatic decision making for follow-up obs.

Decision making with mutual information

c.f. Mahabal+08, Djorgovski+11



- Information entropy before follow-up obs. : $S_o = -\sum_k p(k) \log_2 p(k)$
- Modes of follow-up obs. : i
  - i=0: multi-band imaging to obtain B-V
  - i=1: spectroscopy to measure the EW of Hα
  - i=2: time-series photometry to search for short-term variations
- Information entropy after follow-up obs. and get x_i : $S_{x_i} = -\sum_k p(k|x_i) \log_2 p(k|x_i)$
- Expected S_{x_i} (= conditional entropy):

$$S_i = \int p(x_i) S_{x_i} dx_i = -\int \sum_k p(k|x_i) p(x_i) \log_2 p(k|x_i) dx_i = -\int \sum_k p(x_i|k) p(k) \log_2 \frac{p(x_i|k)p(k)}{p(x_i)} dx_i$$

(using Bayes' theorem, $p(k|x_i) = \dfrac{p(x_i|k)p(k)}{p(x_i)}$, and $p(x_i) = \sum_k p(x_i|k) p(k)$)

- Perform follow-up obs. having maximum mutual information $M = S_o - S_i$

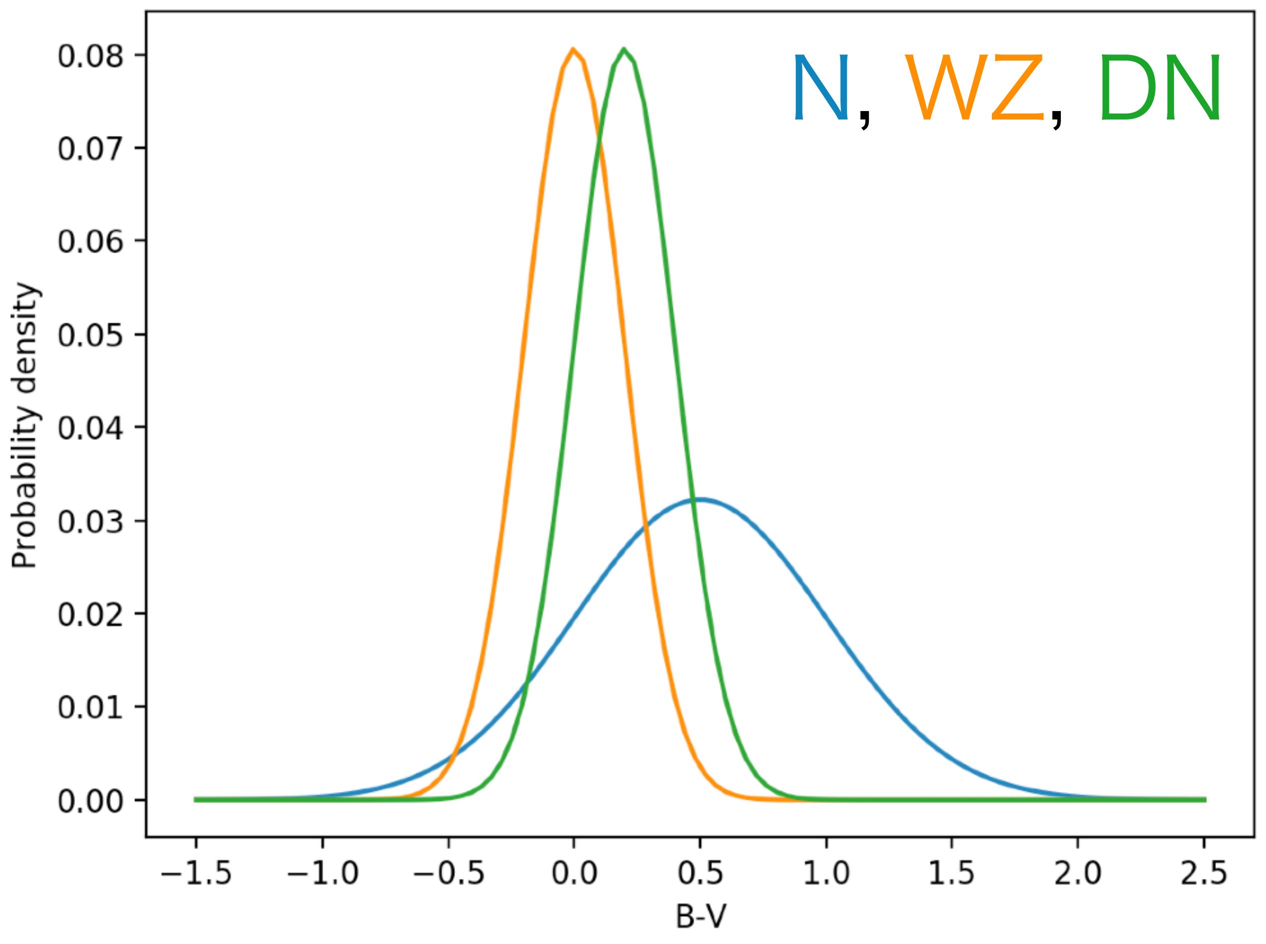
Likelihood, $p(x_i \mid k)$

The conditional probability of the follow-up measurements for each class

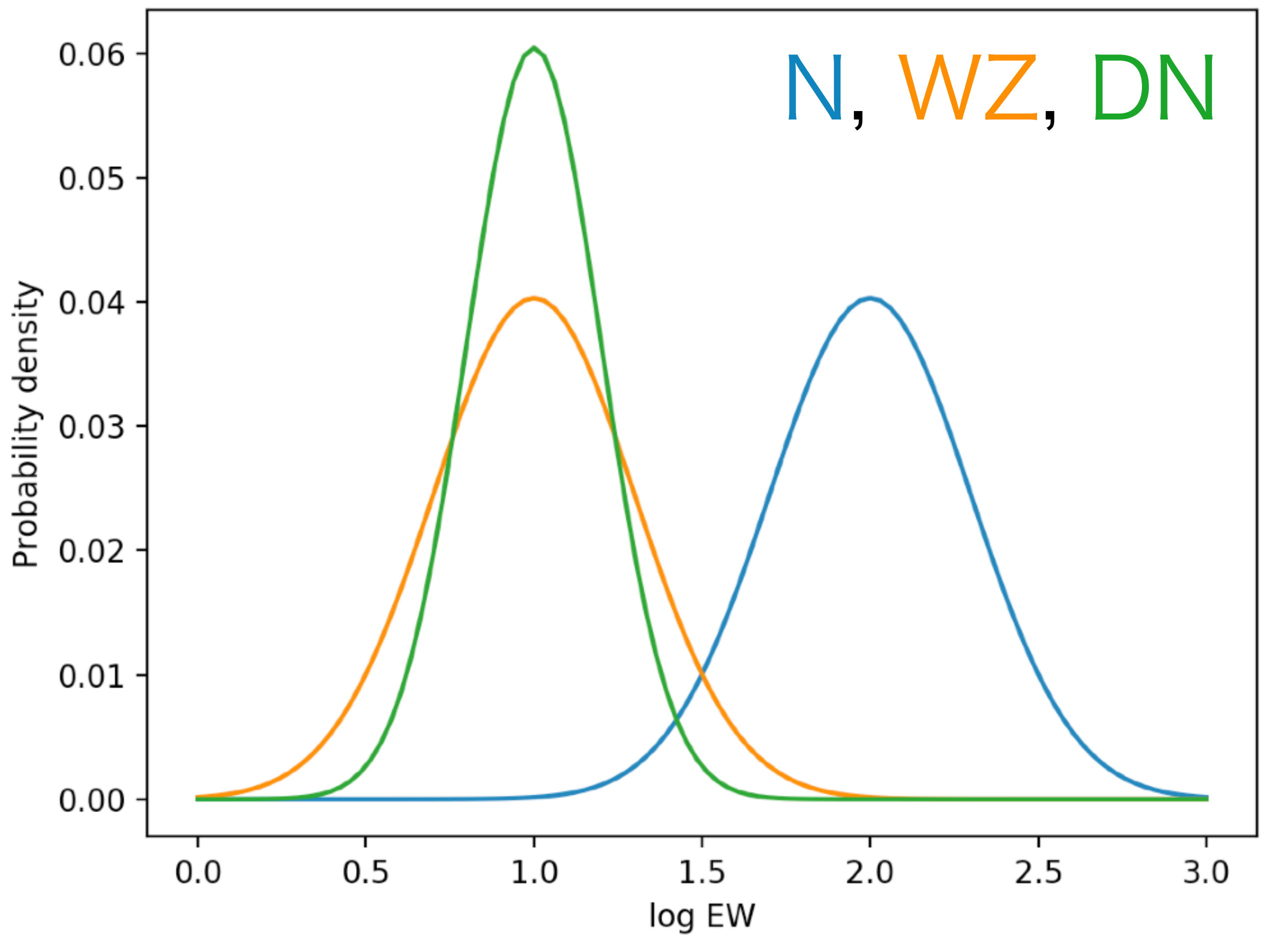
- Should be estimated from data.
- Currently, we use simple Gaussians.

$p(x_i \mid k)$

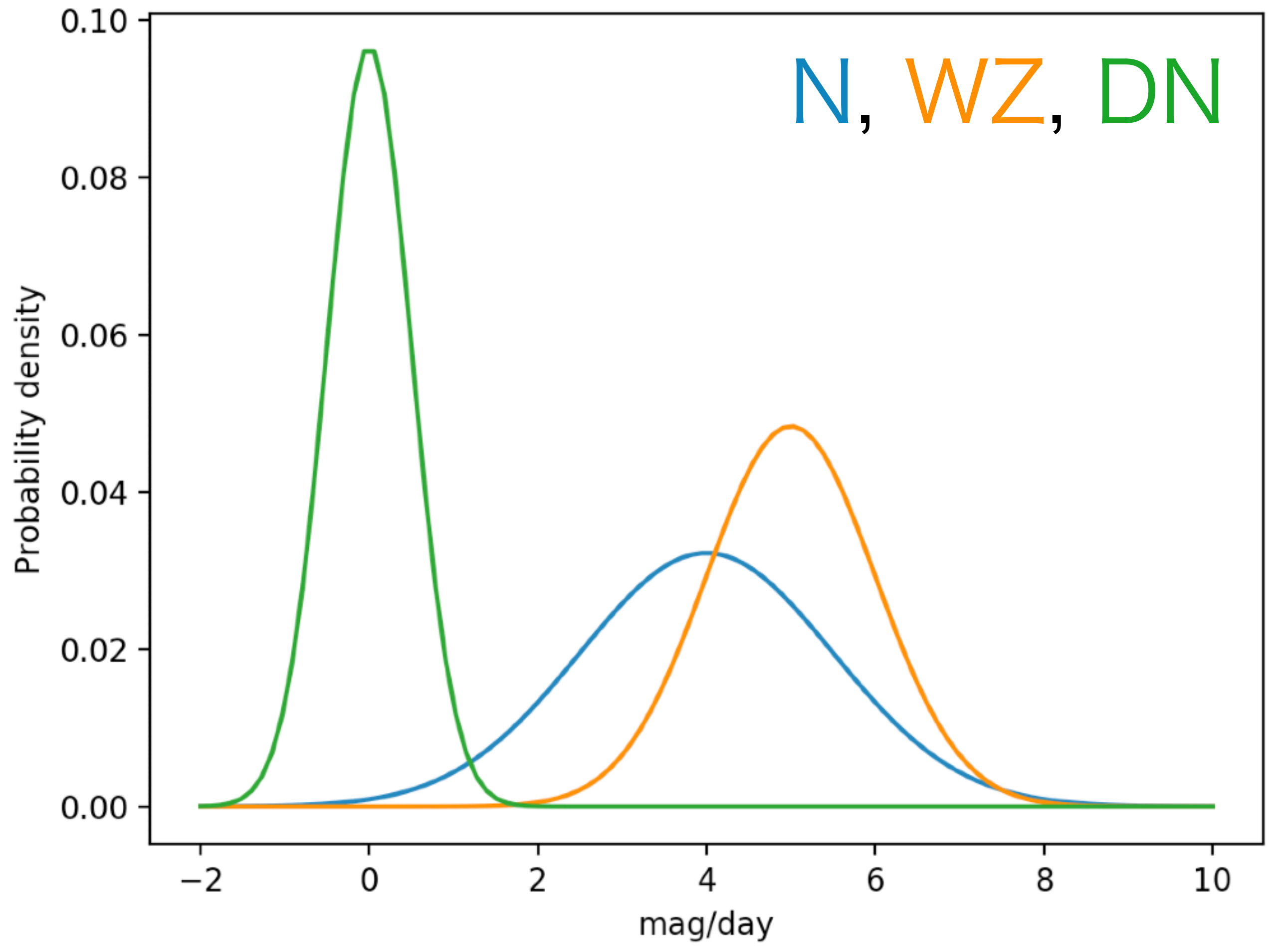
k=0 (color, B-V)



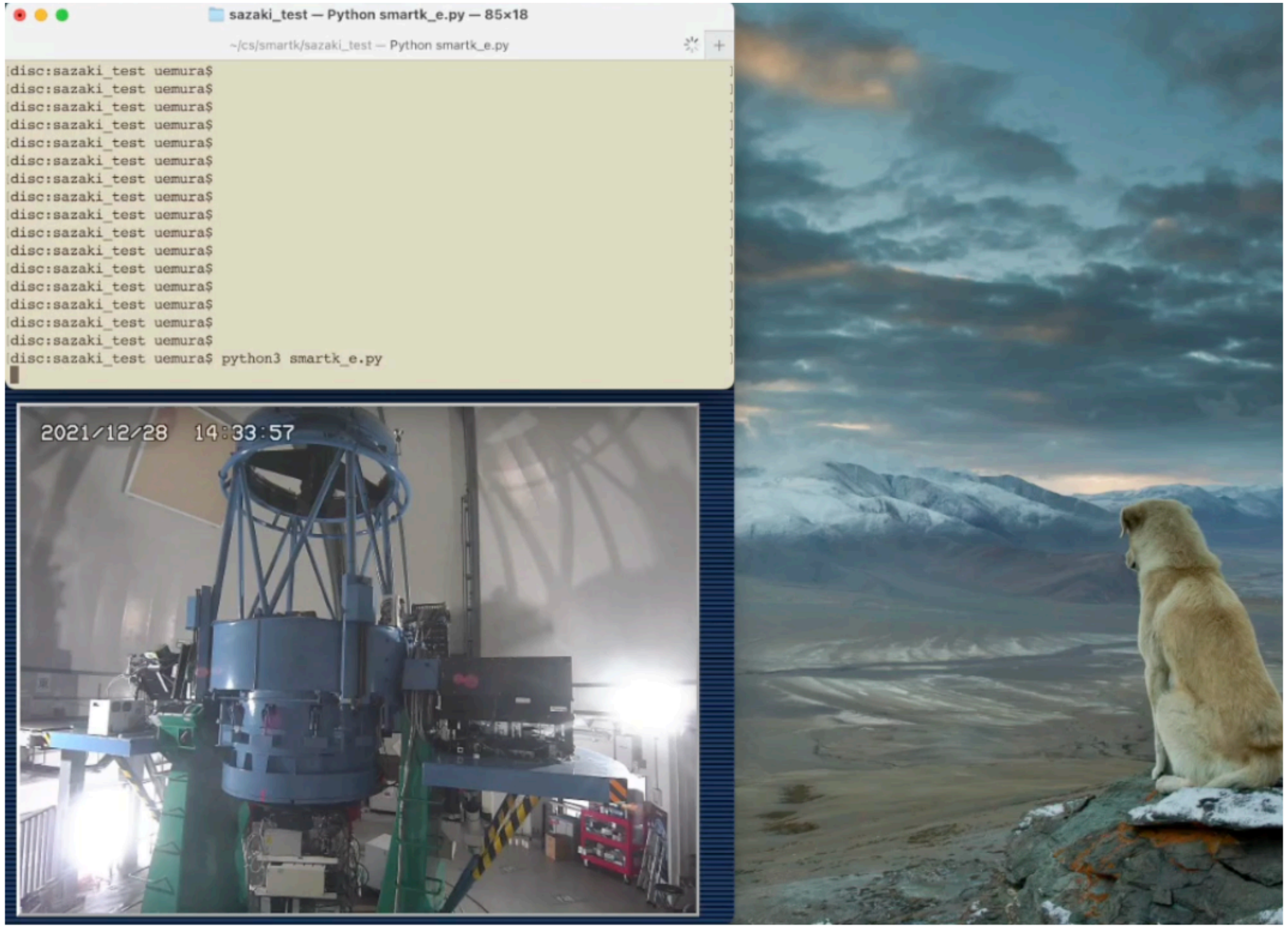
k=1 (EW of Hα)



k=2 (rising rate, mag/day)



Demo. video (Smart Kanata Project)

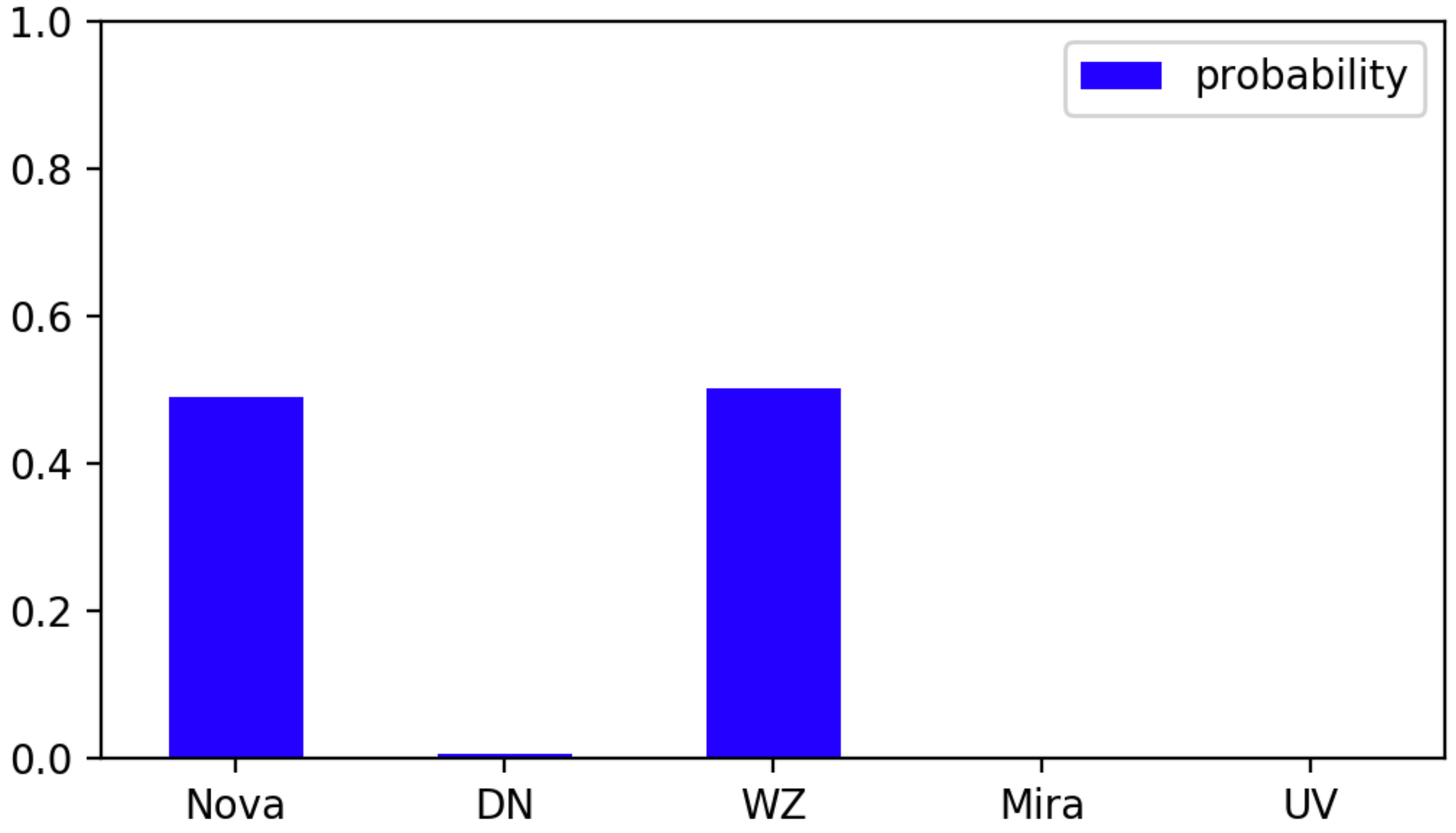


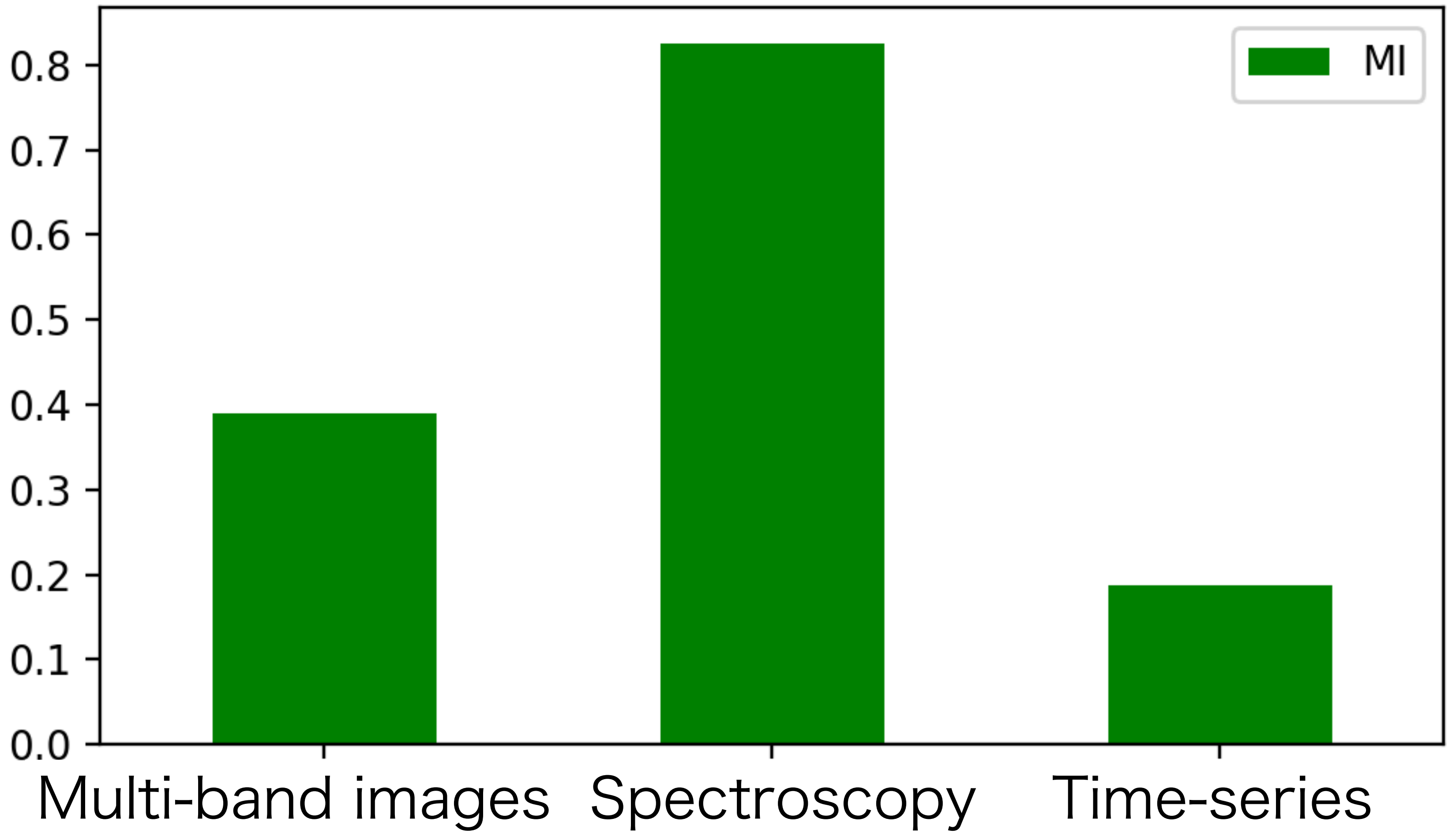
https://vimeo.com/665042022

Examples

TCP J20034647+1335125

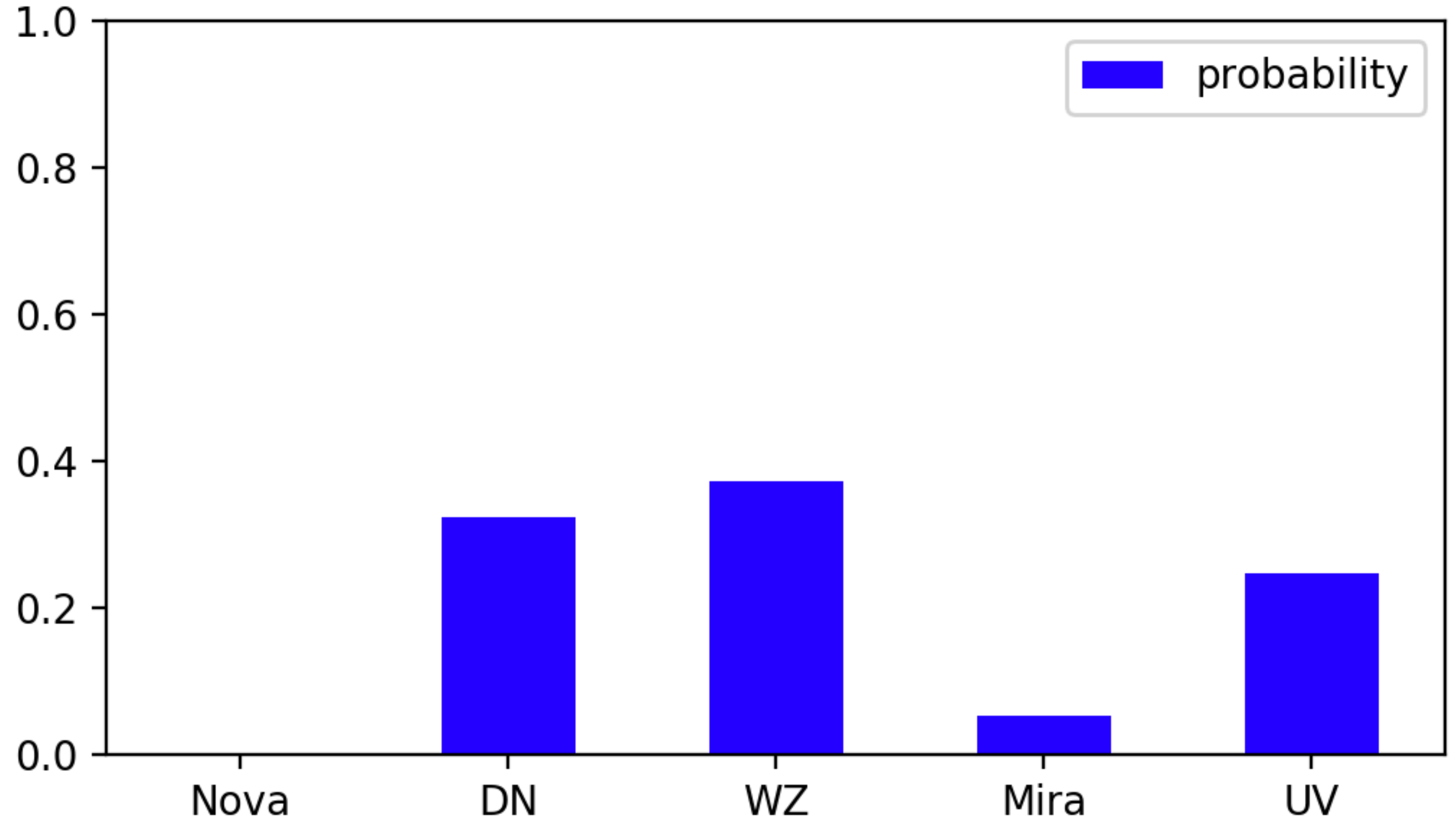
low galactic latitude & large amplitude

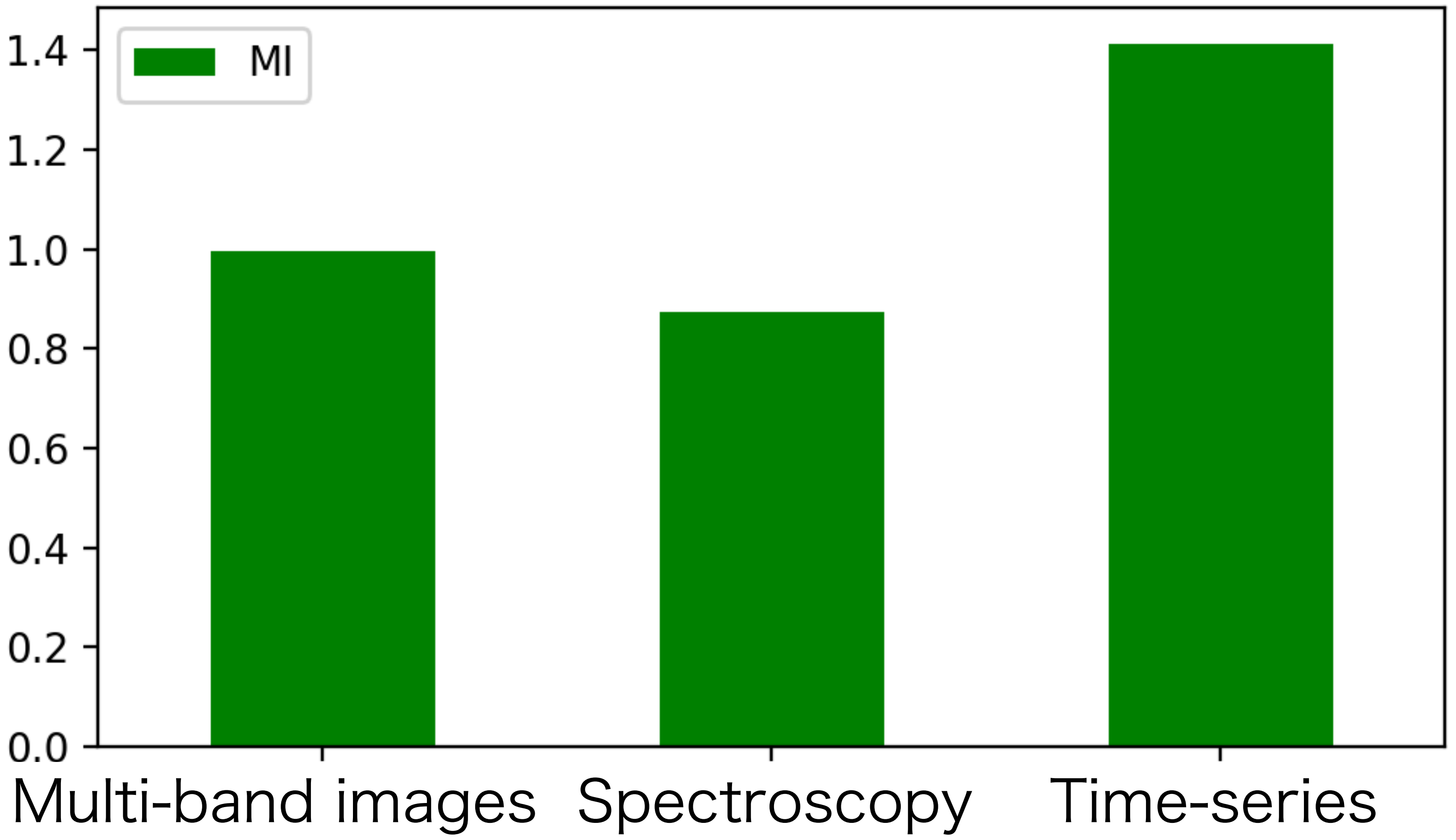




TCP J21130799-6242222

Nearby & low luminosity

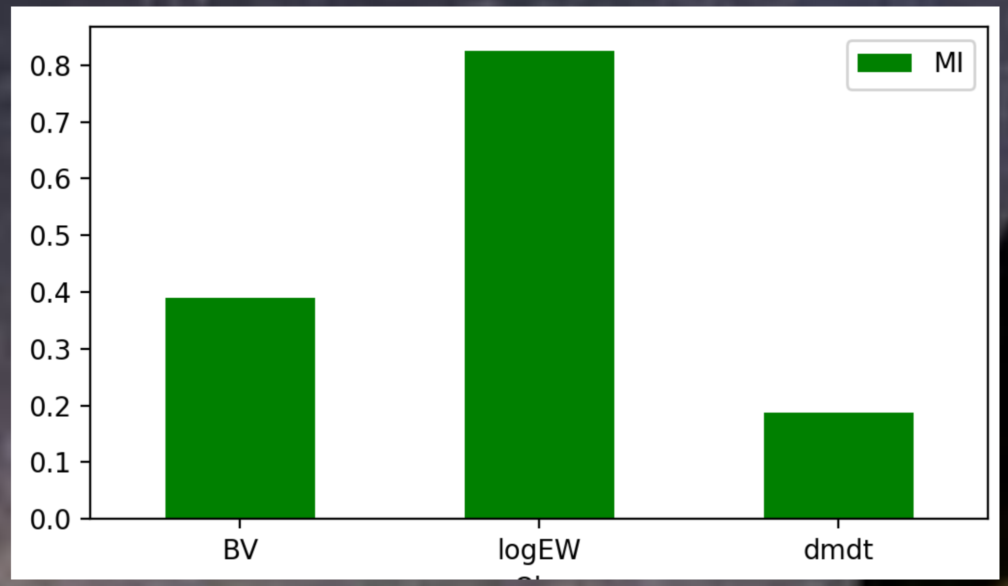


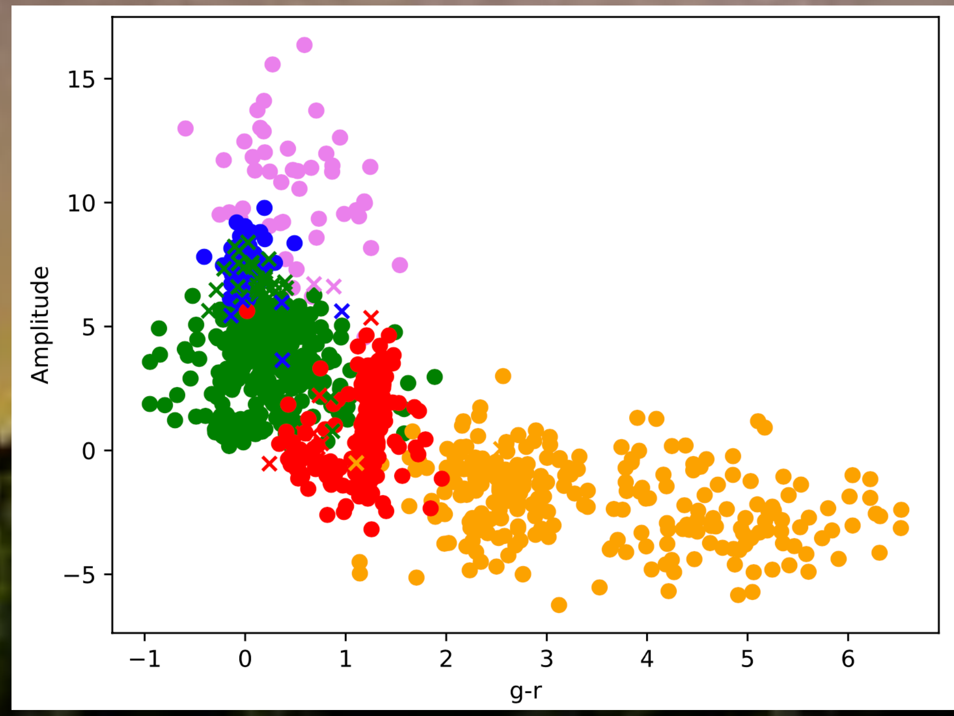


Outline

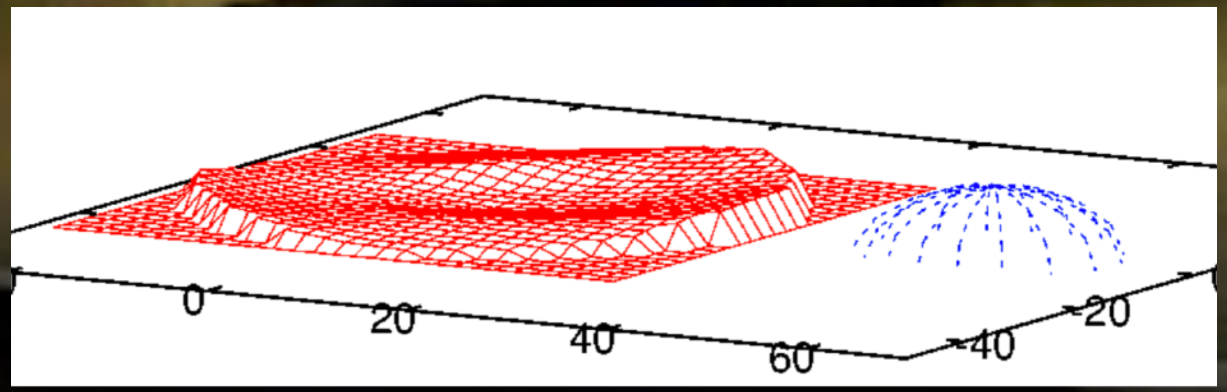
Case studies of astroinformatics in the follow-up studies of transients

1. Classification of new events
   - Sparse multinomial logistic regression
   - Generative model





2. Decision making of follow-up obs.
   - Information theory, mutual information

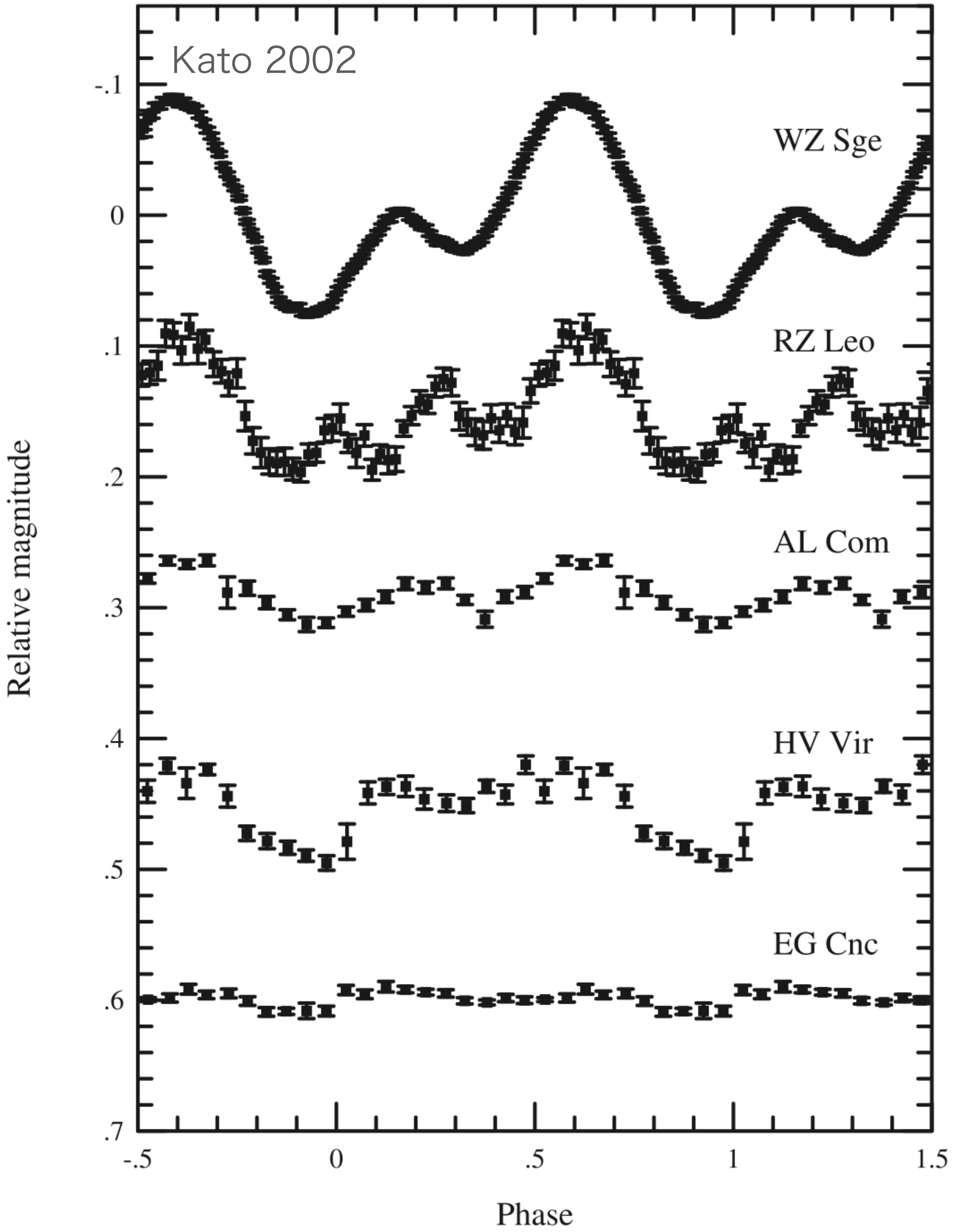


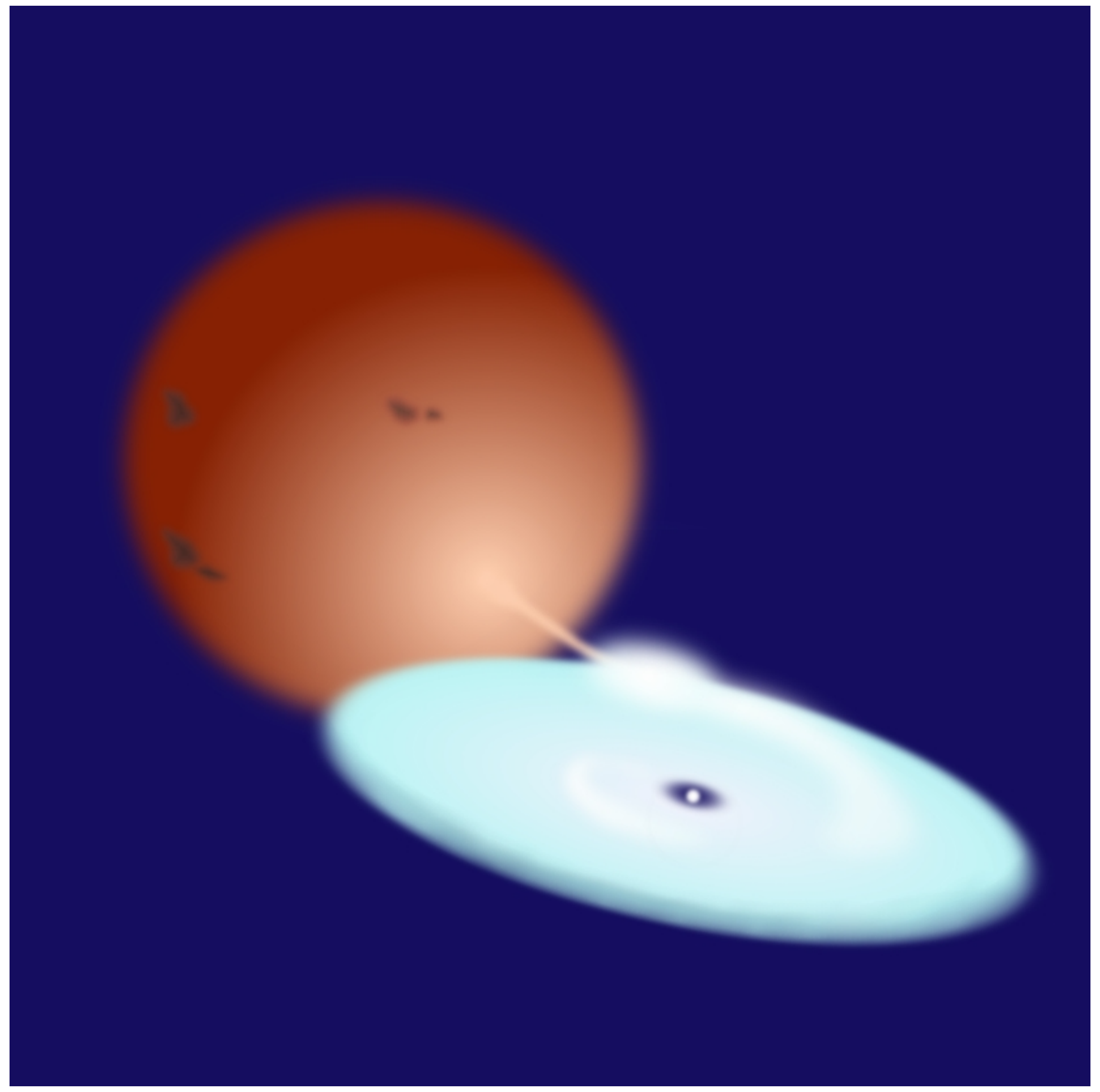
3. Study of the accretion disk structure
   - Tomography, ill-posed inverse problem

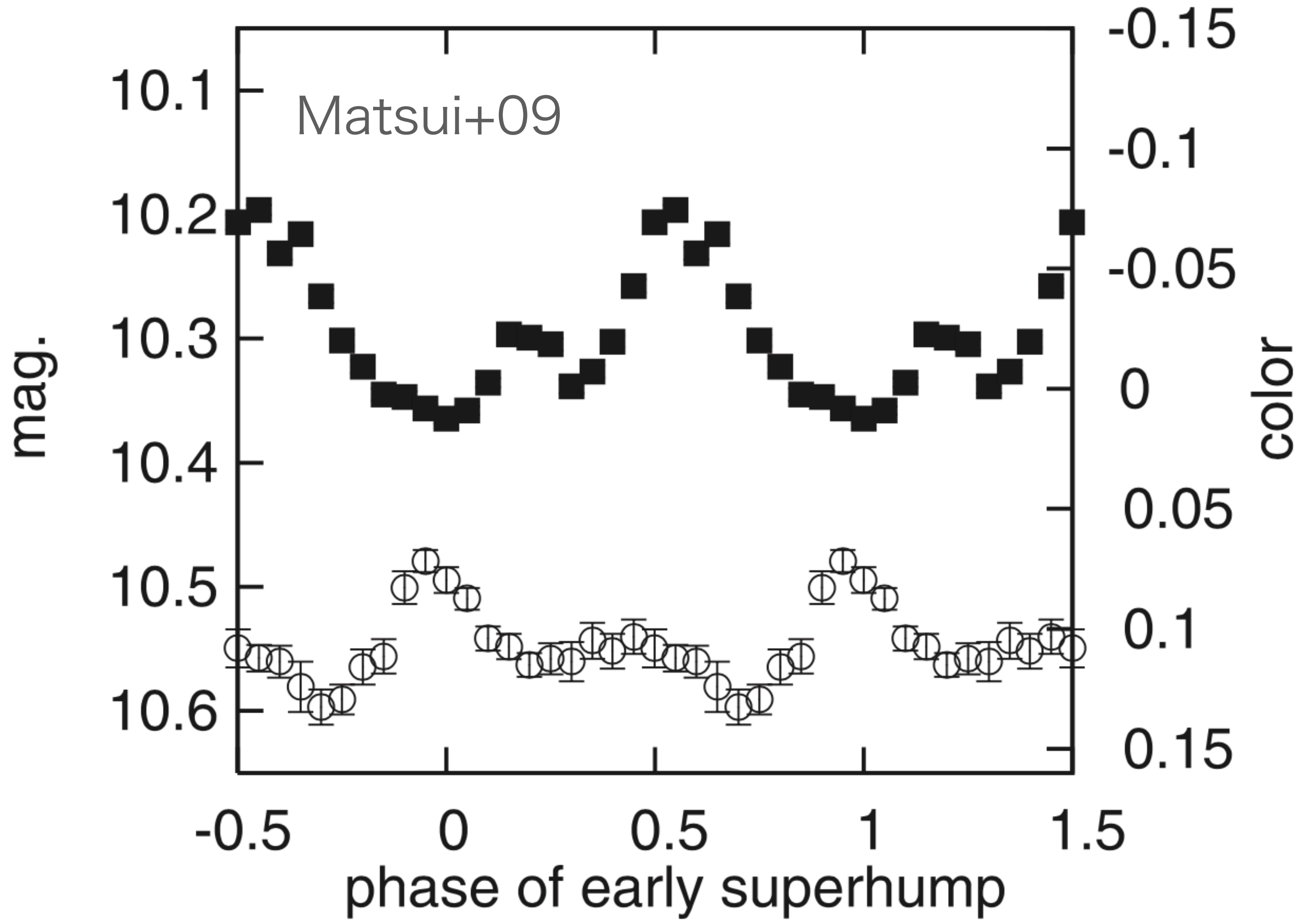
4. (if we have time left,) Tools to support obs.

Early superhumps and disk geometry

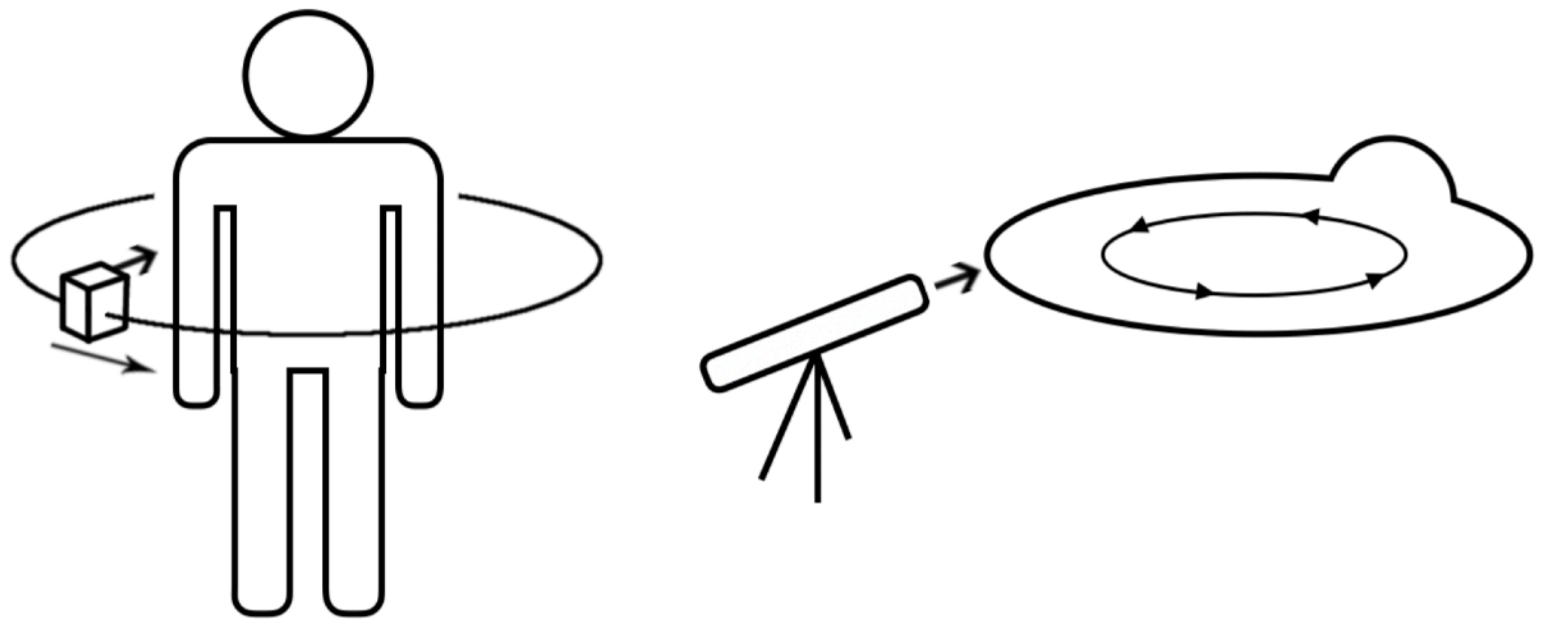
- Dwarf novae
  - a subclass of cataclysmic variables
  - white dwarf, accretion disk, and companion star
- Early superhump
  - only observed in the first few days of hte WZ Sge-type outburst
  - geometrical effect of the disk







Reconstruction of the disk



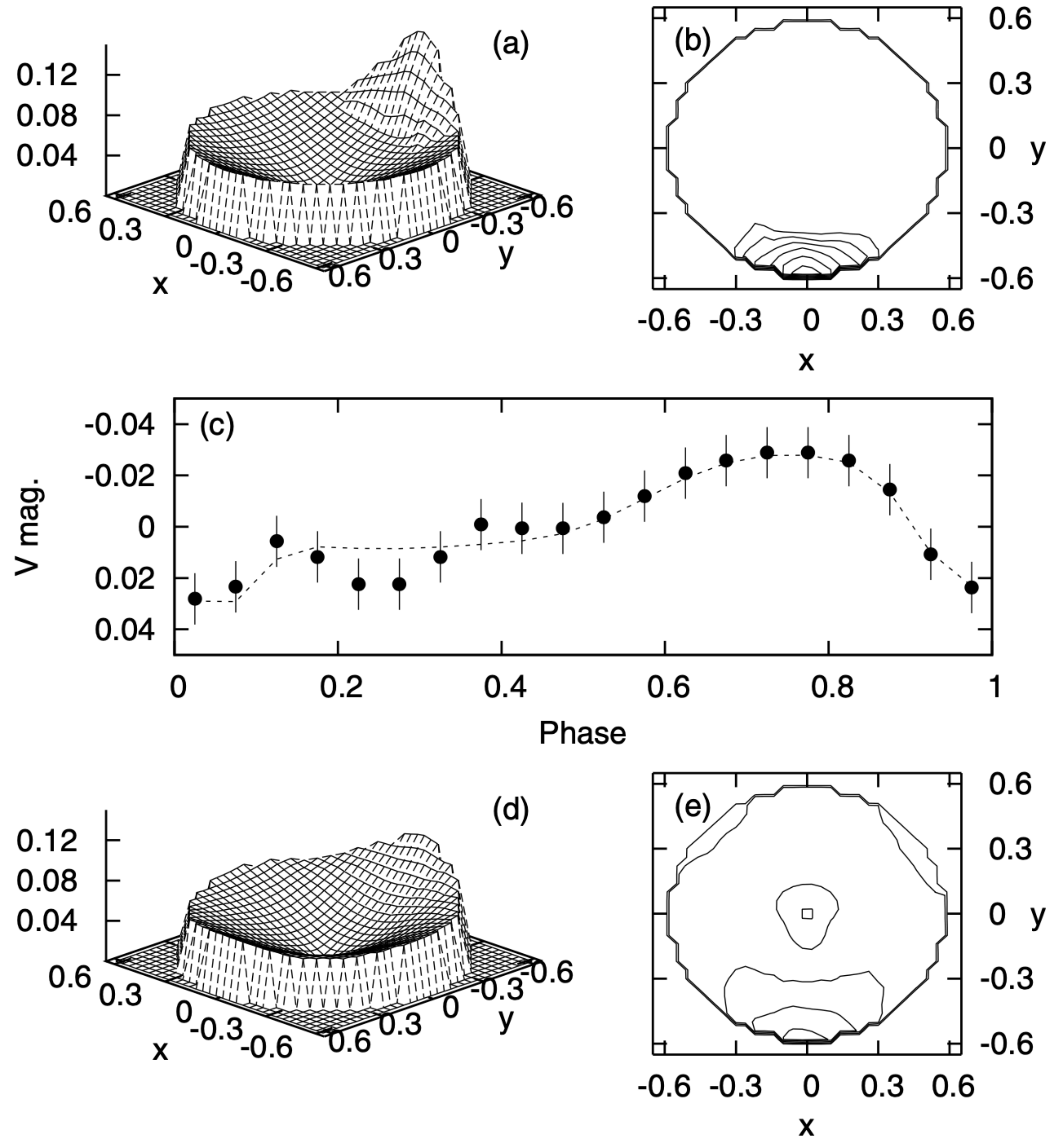
- Estimation of the disk height at each cell
  - 16 radial bins & 20 azimuthal bins
- height structure —> light curve
- Model
  - Gaussian likelihood
  - smooth prior
  - prior for the standard disk structure

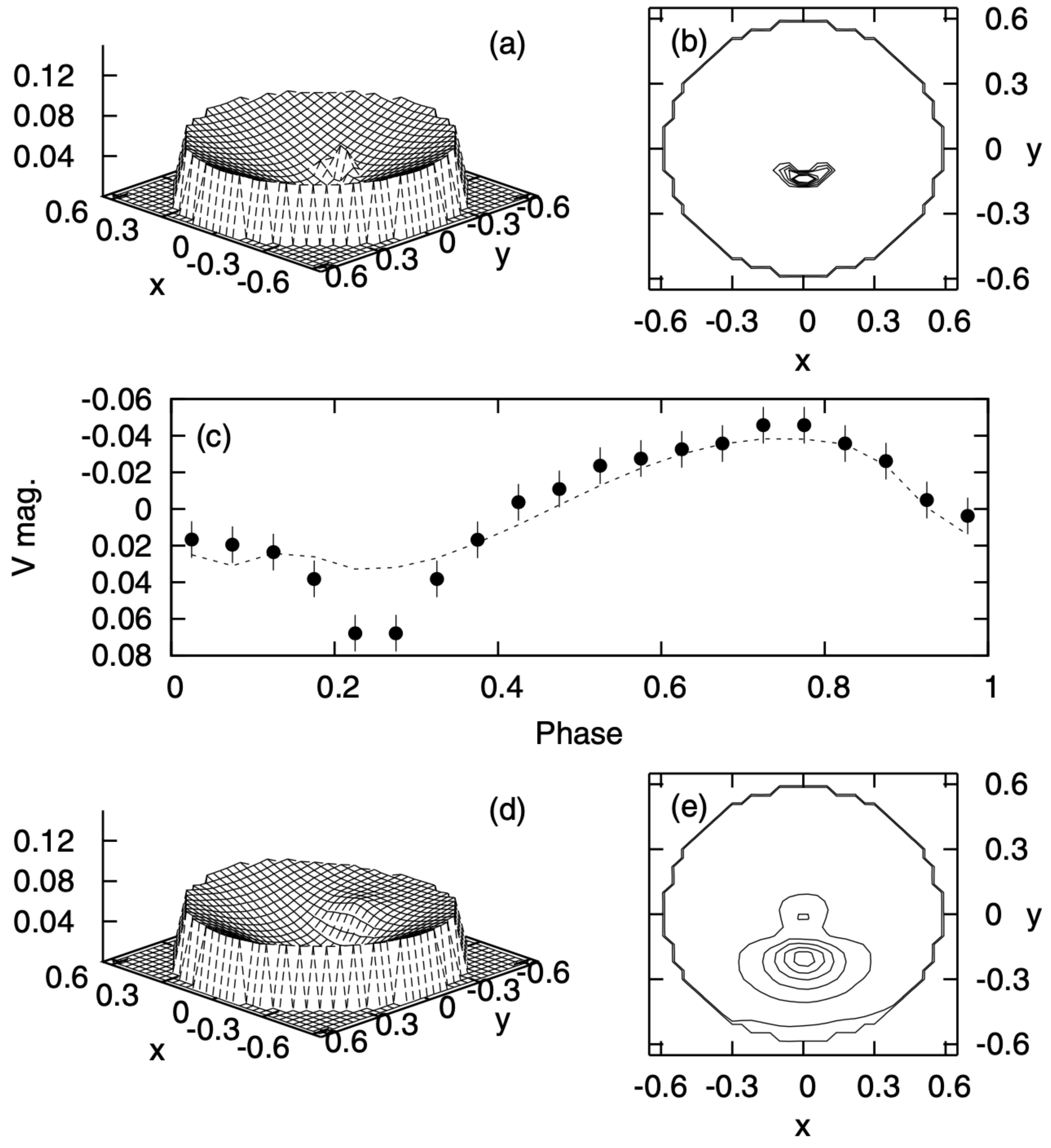
$$L = \prod_{i,j} \frac{1}{\sqrt{2\pi\sigma_{i,j}^2}} \exp\left\{ -\frac{[f_{\nu_i,\mathrm{obs}}(\phi_j) - f_{\nu_i,\mathrm{model}}(\phi_j)]^2}{2\sigma_{i,j}^2} \right\}$$

$$\pi_{\mathrm{smooth}} = \frac{1}{\sqrt{2\pi w^2}} \prod_{l,m} \left\{ \exp\left[-\frac{(h_{l,m} - 2h_{l-1,m} + h_{l-2,m})^2}{2w^2} \right] \exp\left[-\frac{(h_{l,m} - 2h_{l,m-1} + h_{l,m-2})^2}{2w^2} \right] \right\}, \quad (4)$$

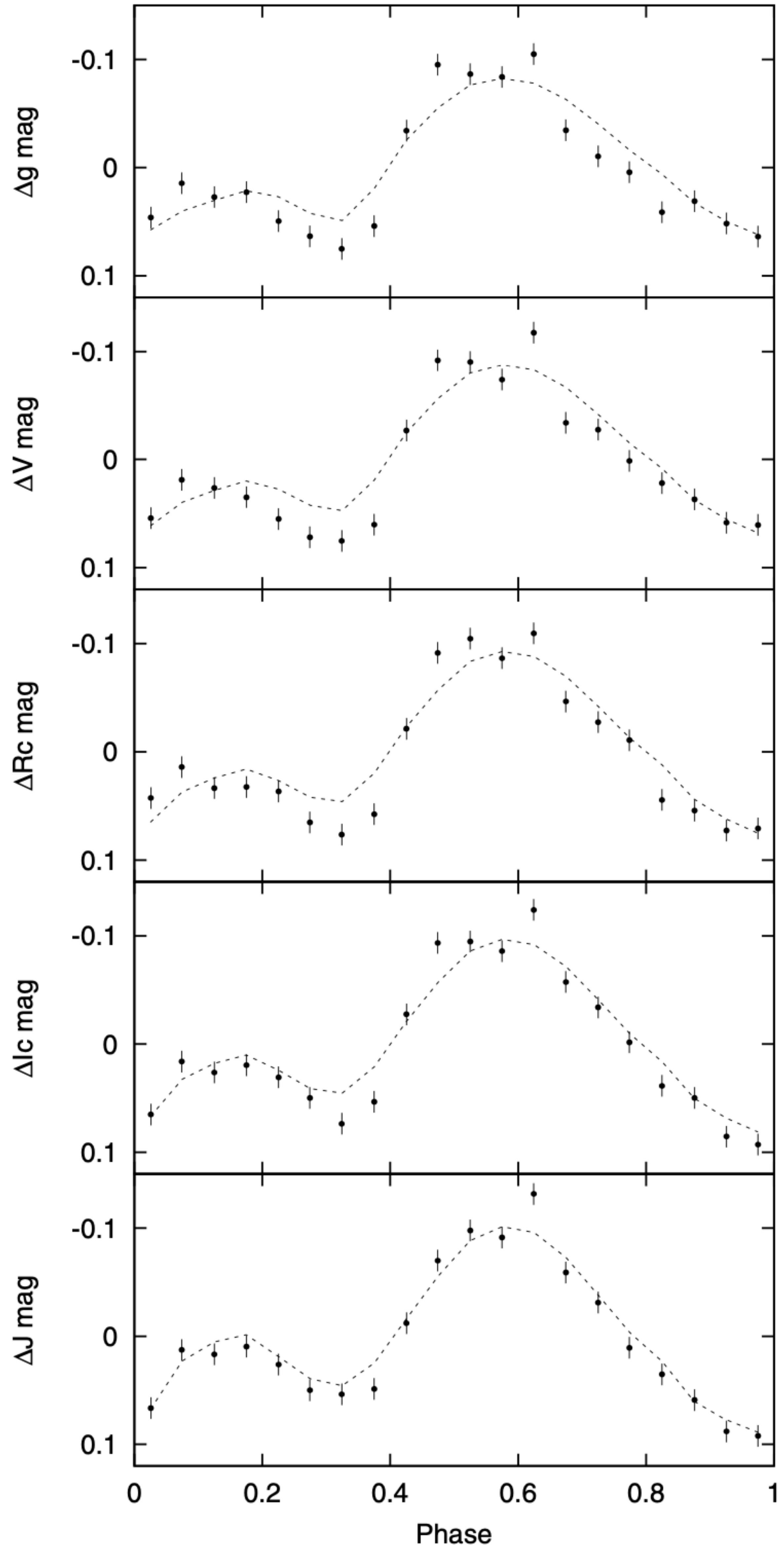
$$\pi_{\mathrm{disk}} = \begin{cases} \prod_{l,m} \frac{1}{\sqrt{2\pi h_{\mathrm{disk},l,m}^2}} \exp\left[-\frac{(h_{l,m} - h_{\mathrm{disk},l,m})^2}{2h_{\mathrm{disk},l,m}^2} \right] & (h_{l,m} \geq 0) \\ 0 & (h_{l,m} < 0) \end{cases}$$

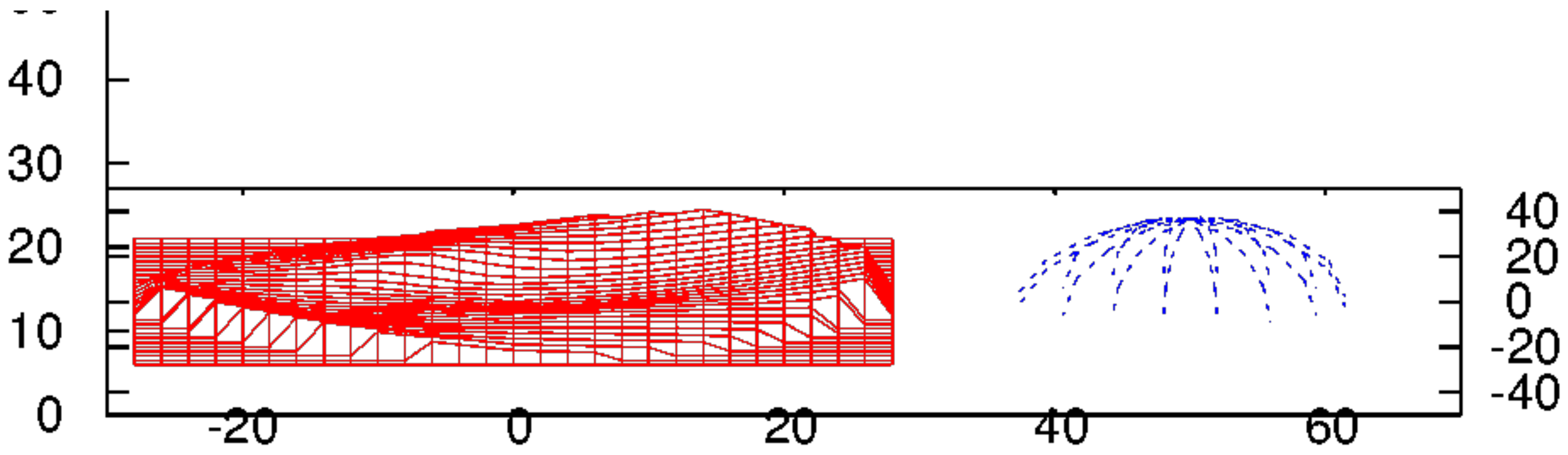
Experiments



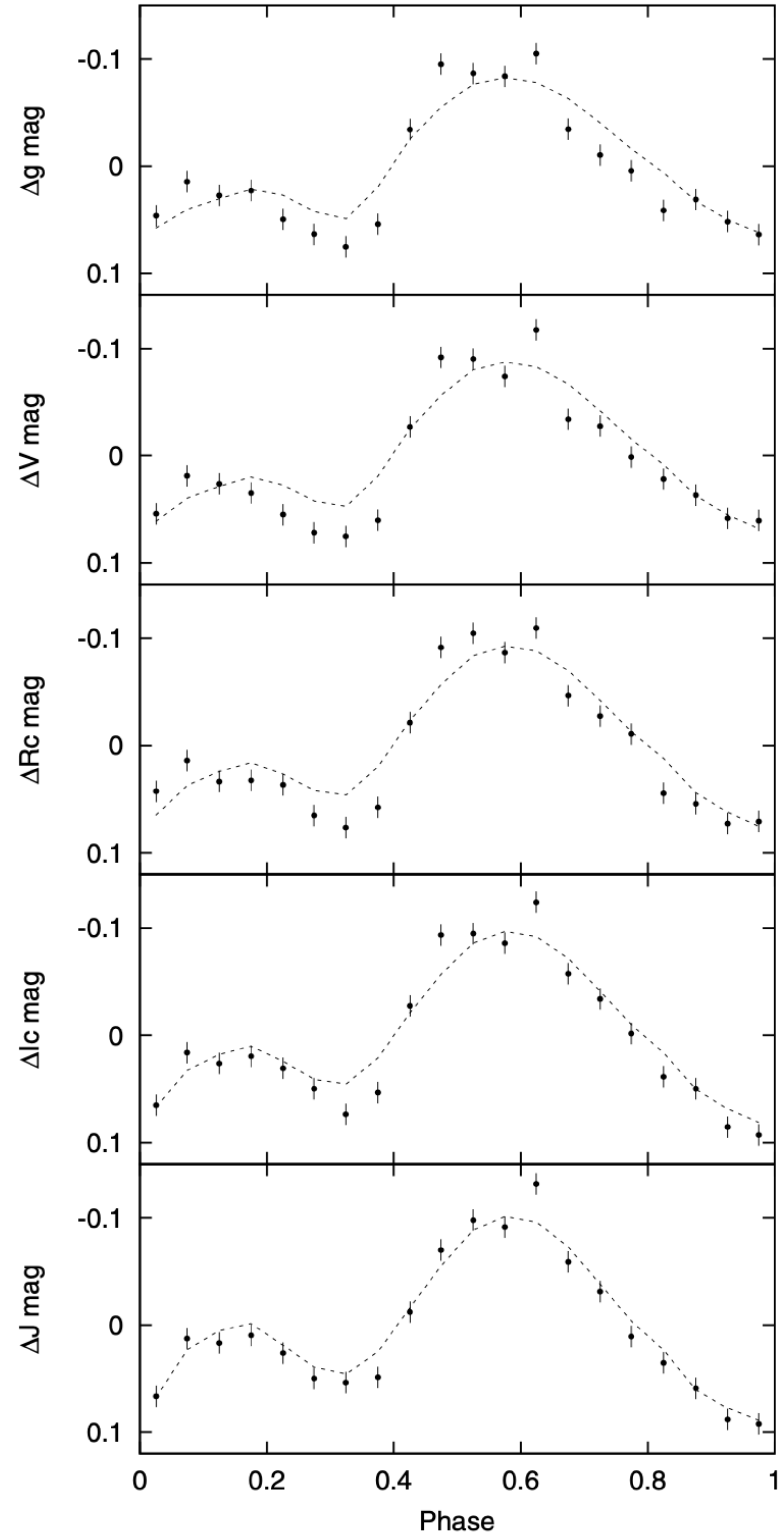


Results (movie)

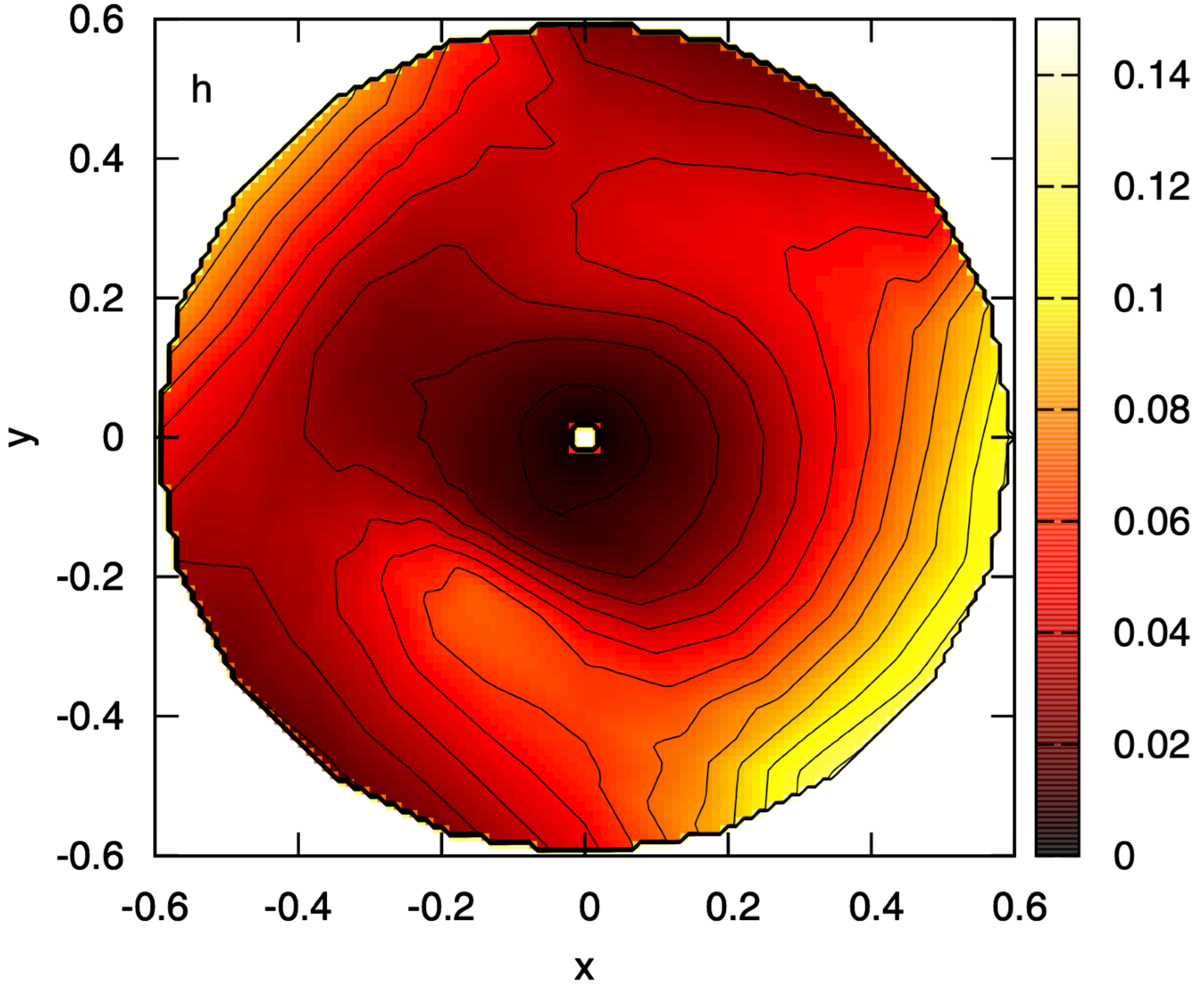




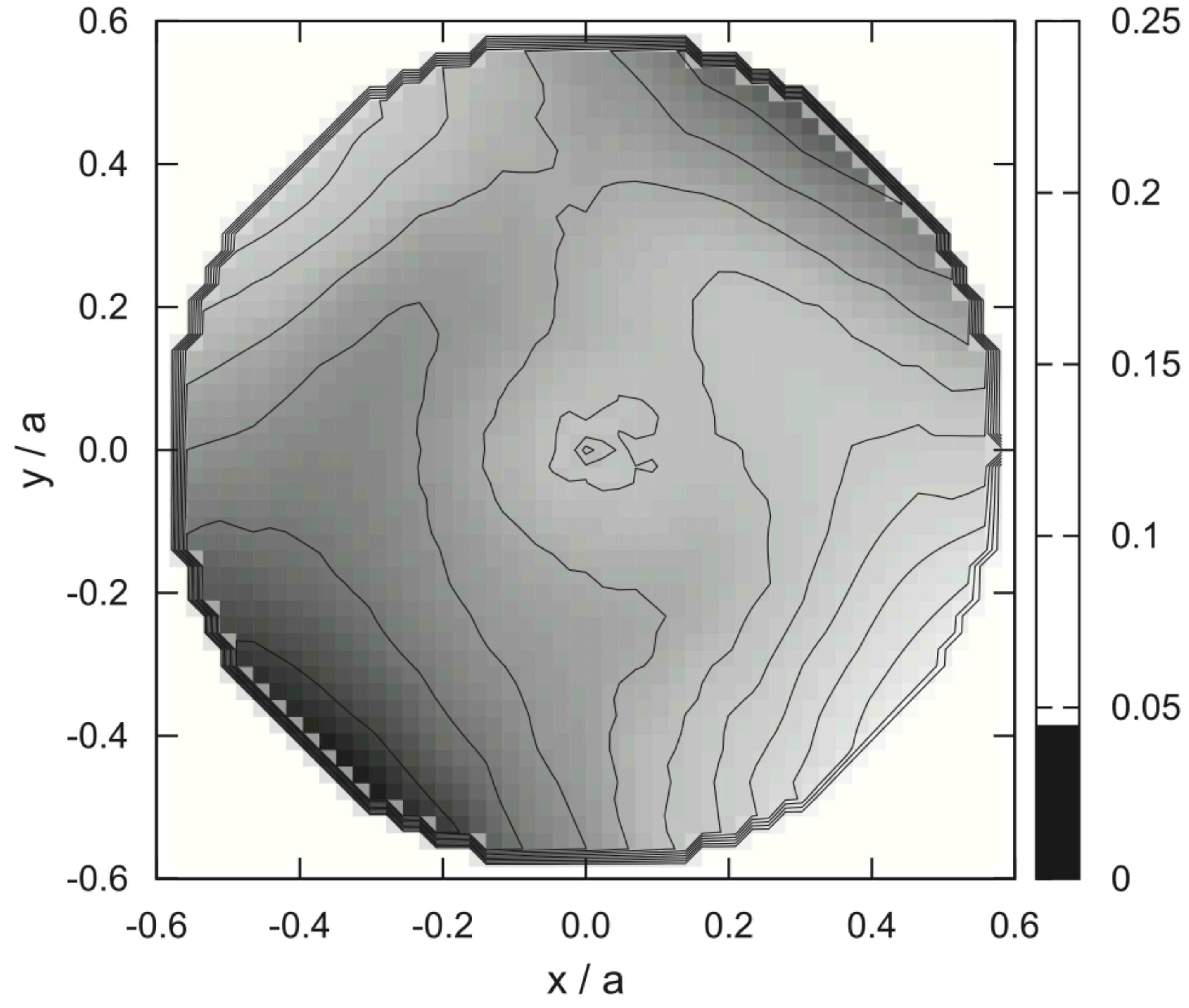
Results



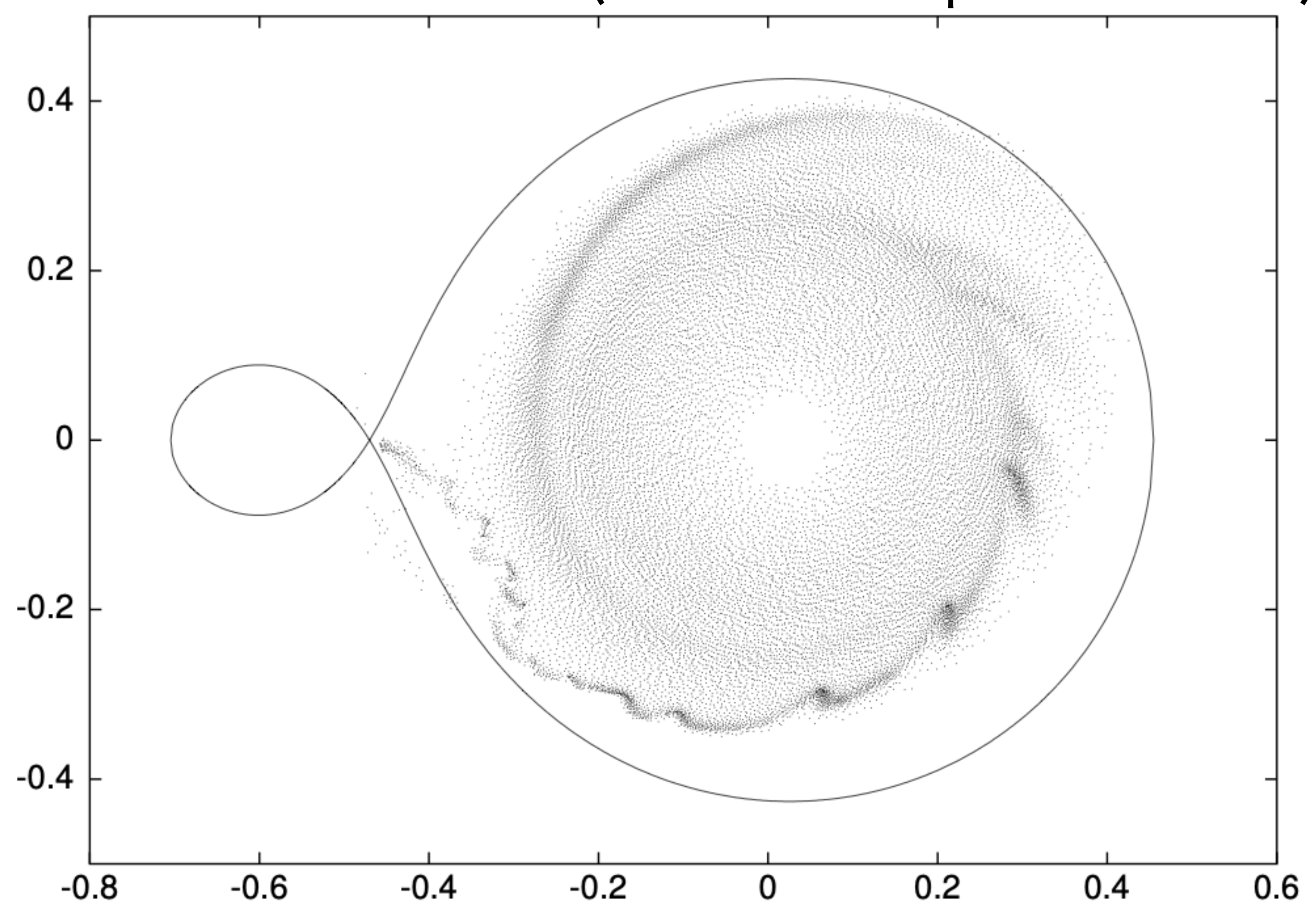
V455 And (Uemura+12)



J0120+32 (Nakagawa+13)



SPH Simulation (Kunze & Speith 2005)



Reconstruction of structures/images

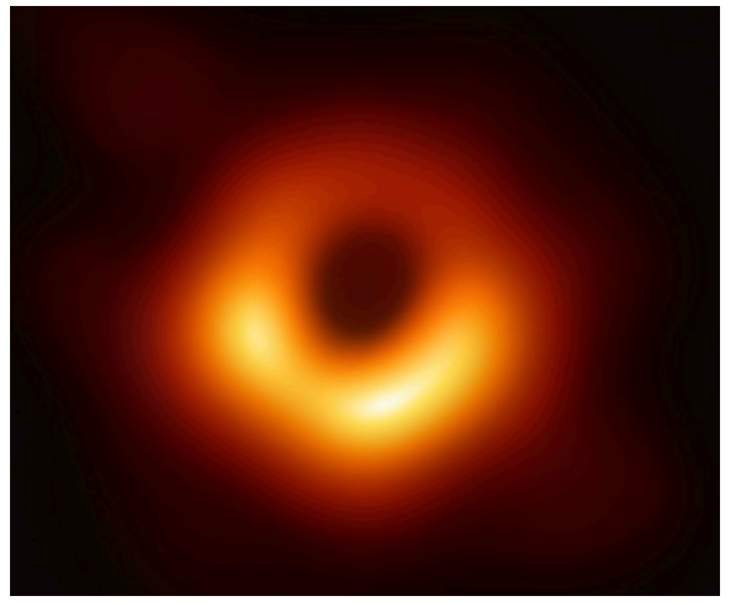
ill-posed inverse problems

- Power spectrum estimation
- Tomography
- Image reconstruction of radio interferometer

$$\left.\begin{array}{l}\\ \\ \\ \end{array}\right\} \Rightarrow \min_{\boldsymbol{\theta}} E(\boldsymbol{\theta}) + \lambda\Phi(\boldsymbol{\theta})$$

	Φ	
Maximum Entropy	$-\sum_i \beta_i \log \beta_i/m_i$	All elements have the same value.
l1	$\lambda\|\boldsymbol{\beta}\|_1$	Making zero elements as many as possible
Total Variation (TV)	$\lambda\sum_i \|\beta_{i+1} - \beta_i\|$	Making zero differences as many as possible
Squared TV	$\lambda\sum_{\ell,m}\left[(\beta_{\ell+1,m} - \beta_{\ell,m})^2 + (\beta_{\ell,m+1} - \beta_{\ell,m})^2\right]$	Locally smooth changes

- Doppler tomography with TV minimization (Uemura+15)
- Event Horizon Telescope (EHT)
  - $\min E(\beta) + \lambda_1\|\beta\|_1 + \lambda_2 TV(\beta) + \lambda_3 STV(\beta)$

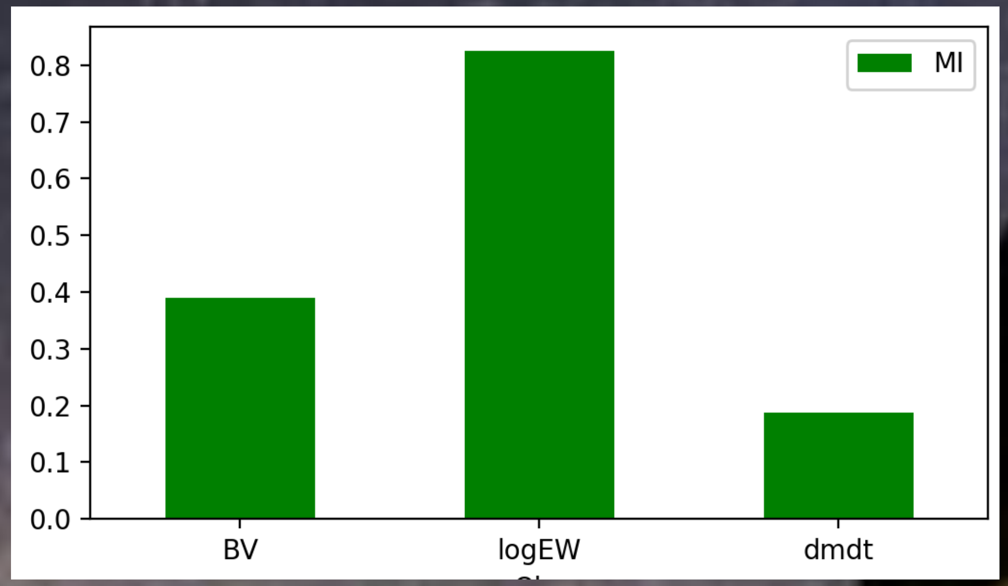


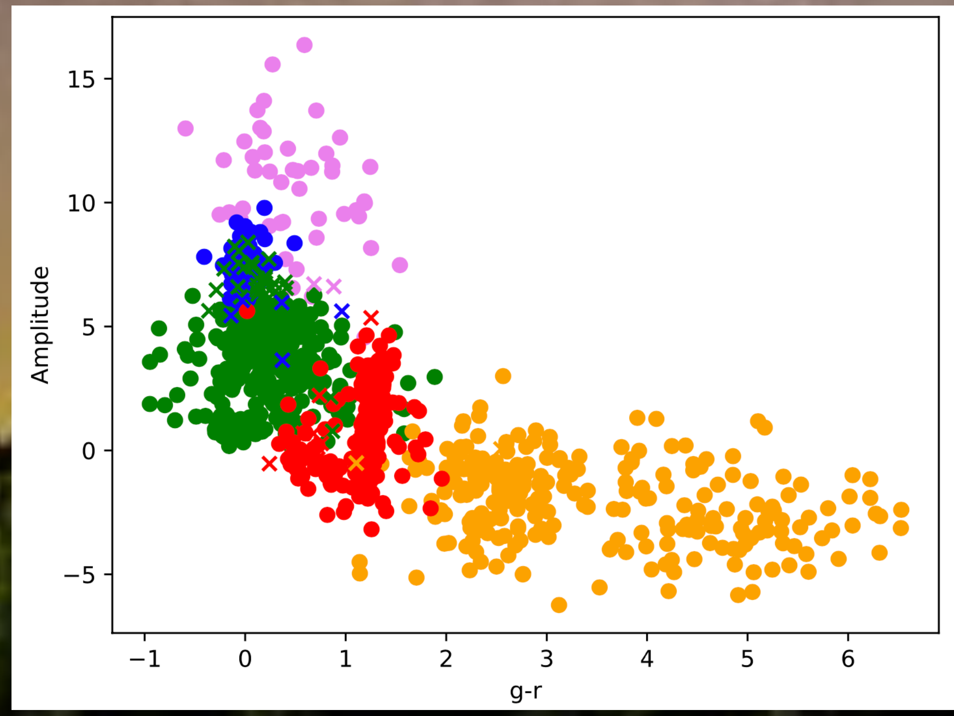
EHT Collaboration

Outline

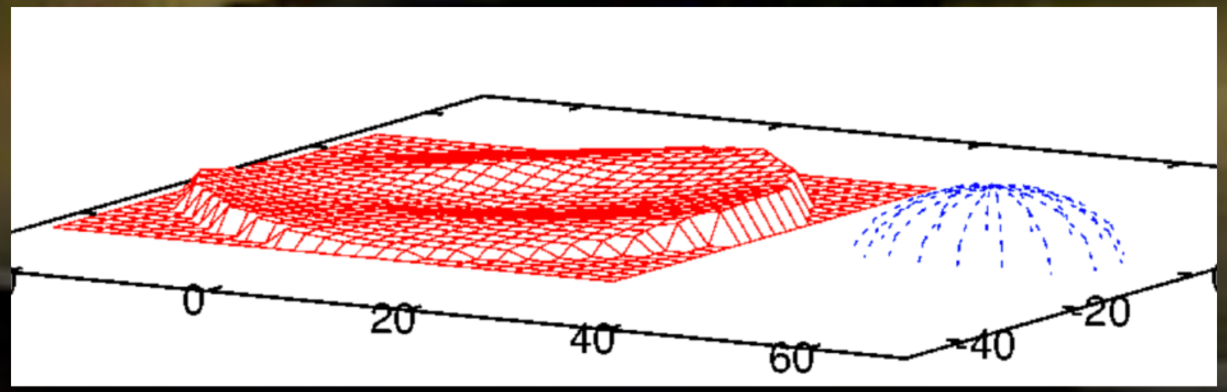
Case studies of astroinformatics in the follow-up studies of transients

1. Classification of new events
   - Sparse multinomial logistic regression
   - Generative model





2. Decision making of follow-up obs.
   - Information theory, mutual information



3. Study of the accretion disk structure
   - Tomography, ill-posed inverse problem

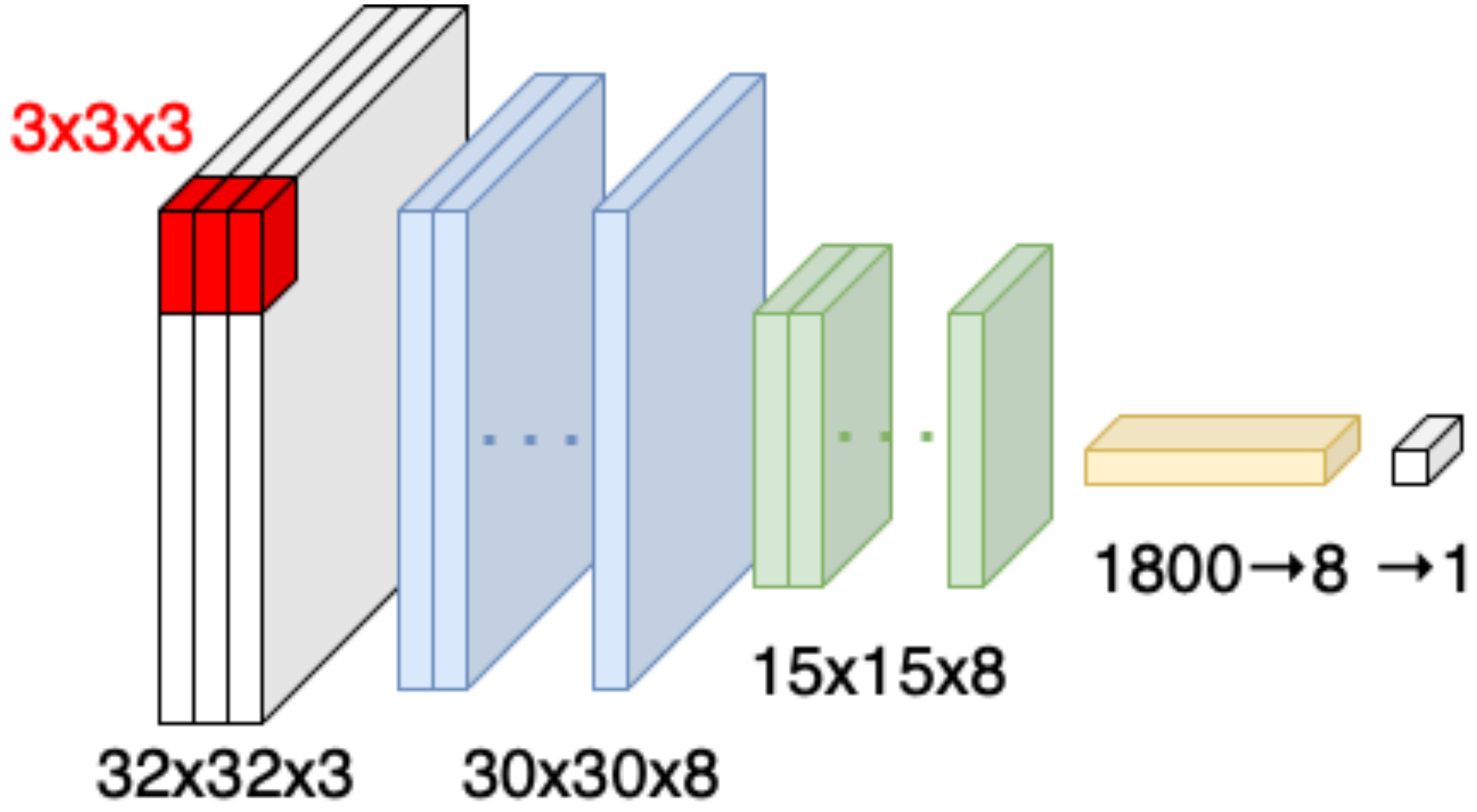
4. (if we have time left,) Tools to support obs.

Decision of the sky condition with CNN

- Weather & sky condition
  - unstable in Hiroshima, Japan
  - 1 or 2 photometric nights/month
  - cloudy or partly cloudy in many nights
  - automated judgement system needed.

Decision of the sky condition with CNN

Simple CNN model

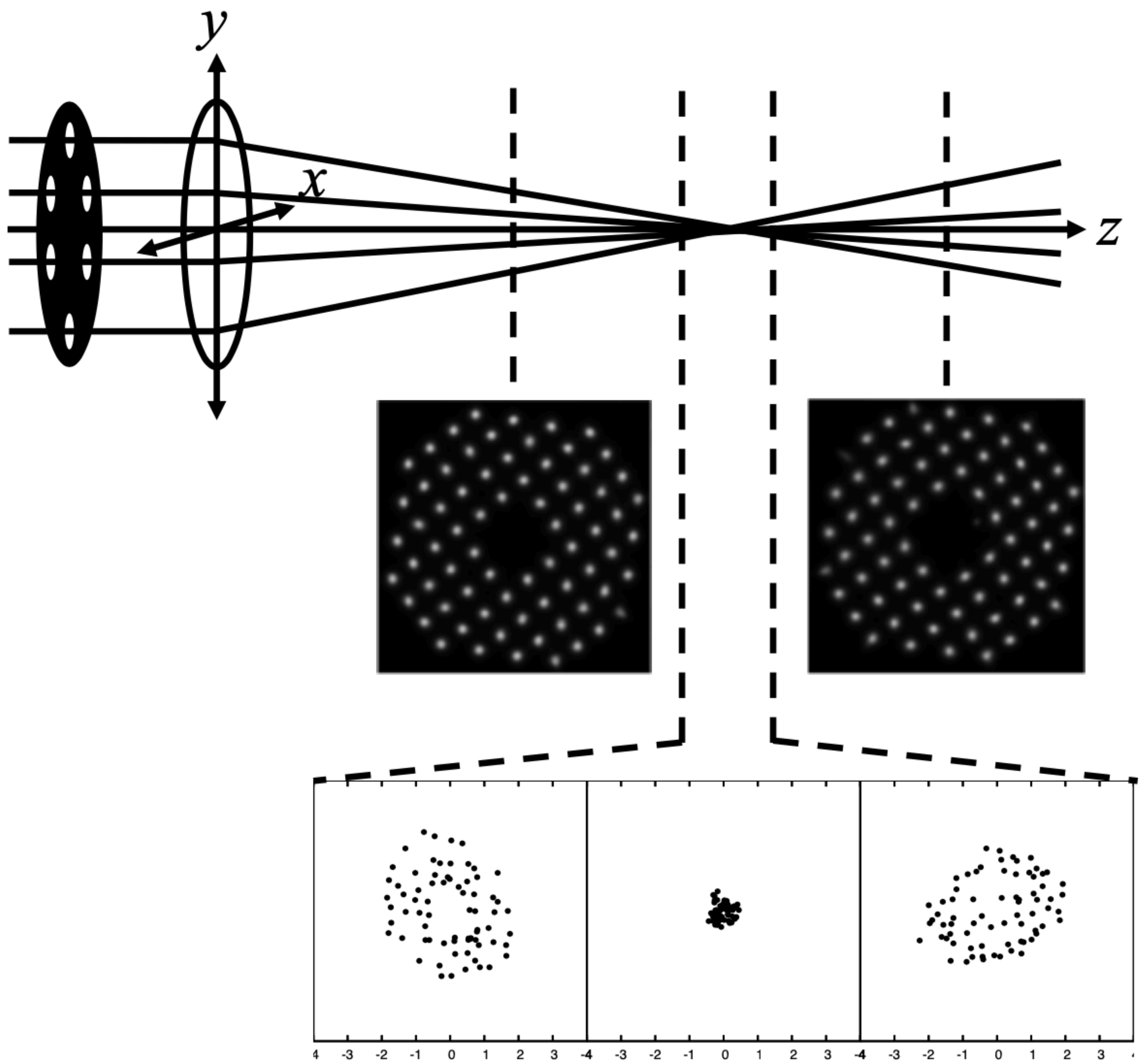


- ~20,000 parameters, 50% dropout
- ~20,000 input images, 80% for training, 20% for validation
- early stop with the validation loss

https://vimeo.com/665043034

Maintenance of the telescope optics
Haltmann test

- Search for the best position of the secondary mirror
- Hartmann test
    1. set the mirror to be a given position
    2. take off-focus images
    3. calculate the Hartmann constant
    - continue 1—3 steps until finding the position with the minimum Hartmann constant

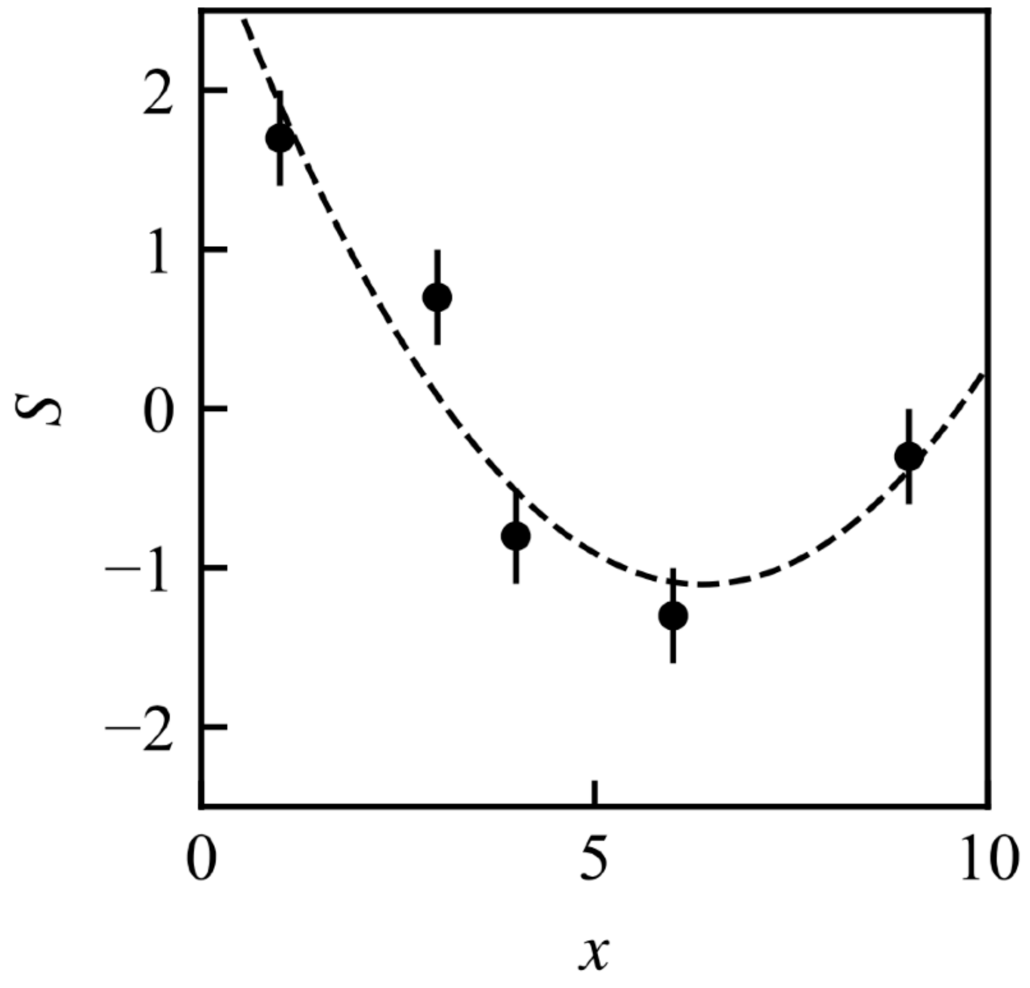


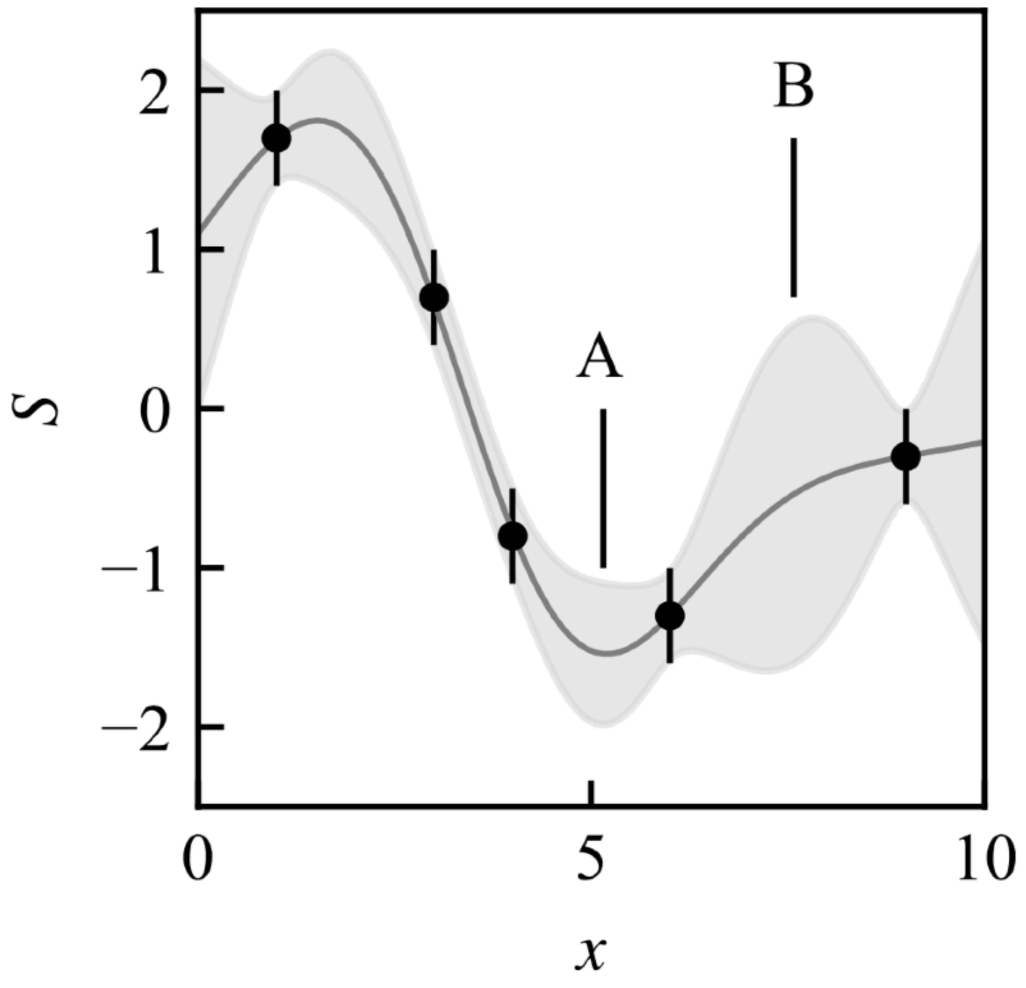
We want to search for the best position with as few measurements as possible.

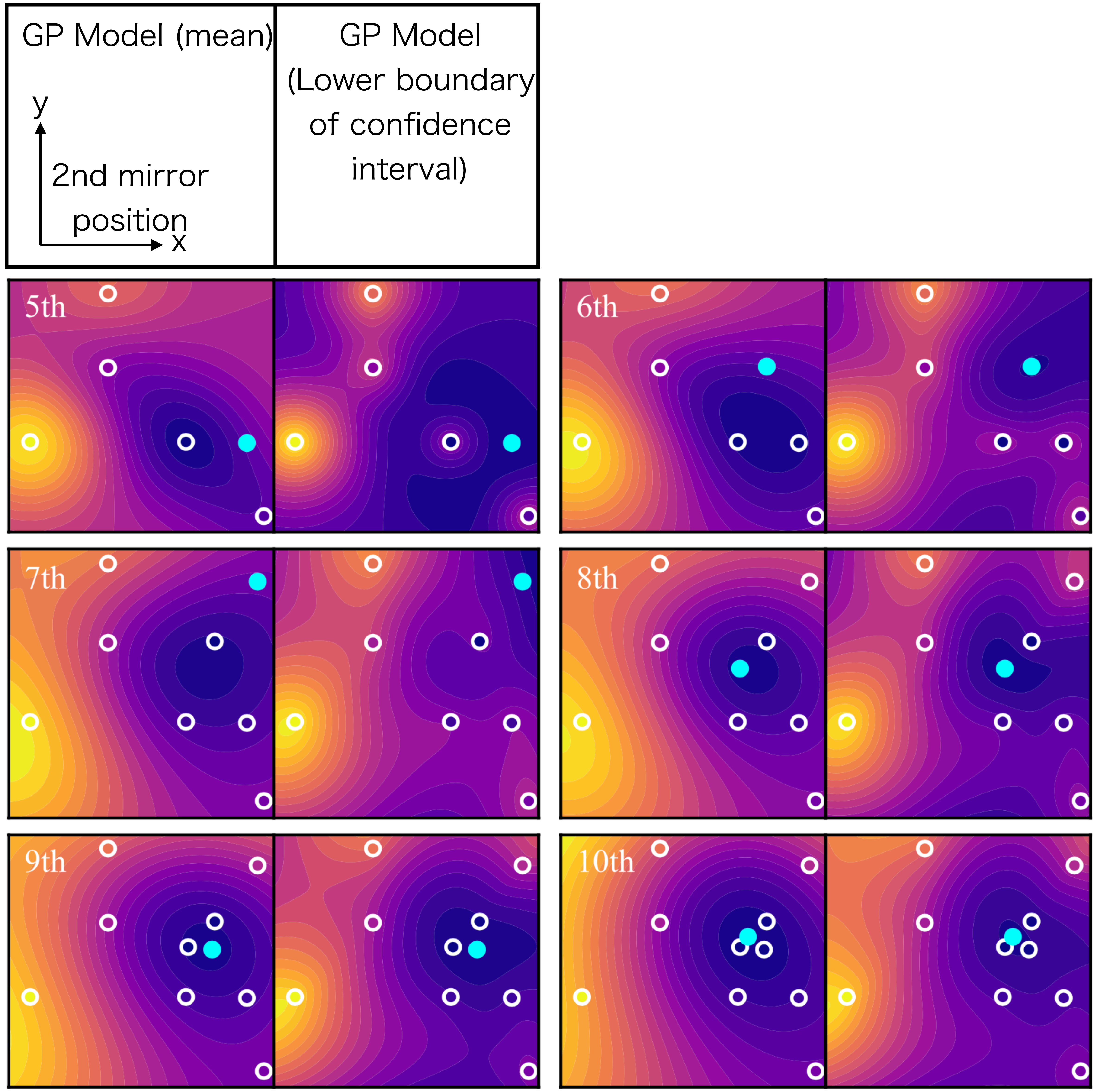
Maintenance of the telescope optics

Bayesian optimization

- Gaussian process (GP) regression —> decide the next point of measurement

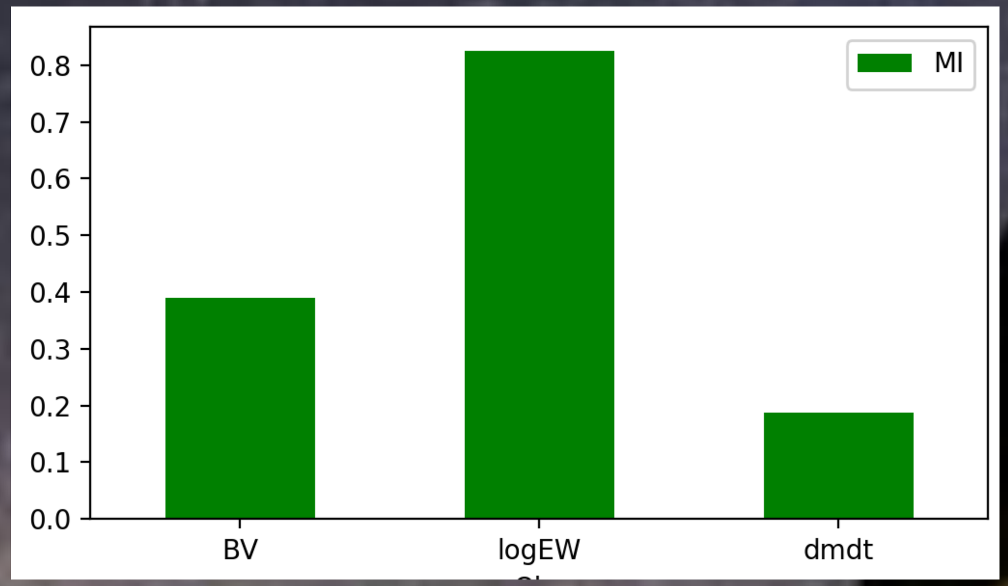




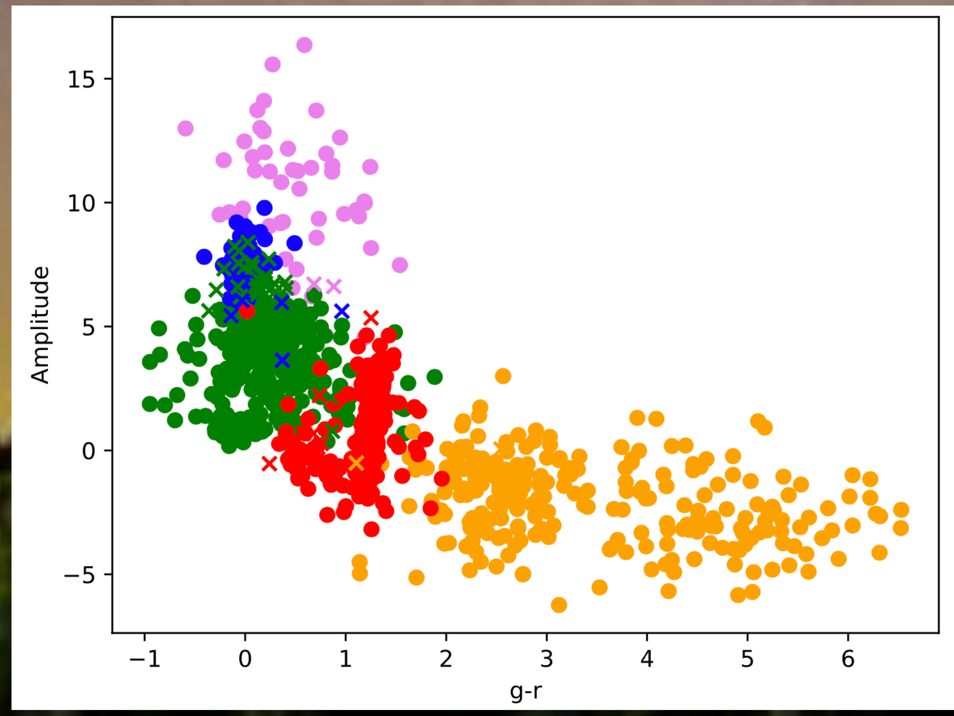


Summary

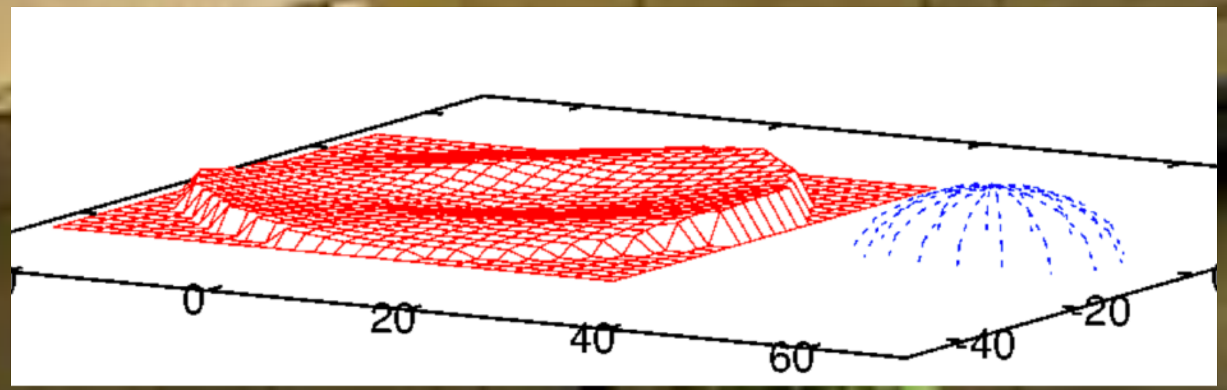
Case studies of astroinformatics in the follow-up studies of transients



1. Classification of new events
   - Sparse multinomial logistic regression
   - Generative model



2. Decision making of follow-up obs.
   - Information theory, mutual information



3. Study of the accretion disk structure
   - Tomography, ill-posed inverse problem

4. (if we have time left,) Tools to support obs.
   - Convolutional neural network
   - Gaussian process & Bayesian optimization

Collaborators and References

- Classification & Decision making
  - S. Ikeda (The Institute of Statistical Mathematics), Y. Koga, R. Sazaki, T. Nakaoka (Hiroshima Univ.), and the VSNET team (Kyoto Univ.)
  - Uemura+22 in prep.
- Disk height mapping
  - T. Kato, T. Ohshima (Kyoto Univ.), and H. Maehara (NAOJ)
  - “Reconstruction of the Structure of Accretion Disks in Dwarf Novae from the Multi-Band Light Curves of Early Superhumps,” PASJ, 64, 92, 2012

back-ups

Problems, and generative model

New object #1

Object	Class	l	b	d (kpc)	g-r	j-h
New object	—	336.5	3.6	—-	—-	—-

New object #2

Object	Class	l	b	d (kpc)	g-r	j-h
New object	—	10.5	13.0	0.42	0.03	—-

New object #3

Object	Class	l	b	d (kpc)	g-r	j-h
New object	—	96.5	40.2	0.31	0.12	0.21

Training dataset

Object	Class	l	b	d (kpc)	g-r	j-h
HY Lup	Nova	318.5	8.6	—-	—-	—-
V3663 Oph	Nova	0.1	7.2	—-	1.29	—-
V1148 Sgr	Nova	5.2	-3.0	2.32	0.47	0.22
FN Anc	DN	127.6	-27.4	0.53	0.22	0.72
V767 Cyg	DN	88.4	9.9	—-	0.09	0.16
V952 Oph	DN	33.1	12.2	0.98	—-	0.56
AL Com	WZ	282.9	76.5	0.57	0.04	—-
V592 Her	WZ	38.8	40.0	—-	0.00	—-
QZ Lib	WZ	356.9	36.4	0.19	0.09	—-
CL Mon	Mira	207.7	3.83	1.39	5.05	1.67
V593 Pup	Mira	243.7	9.0	2.56	2.26	0.85
EU Cen	Mira	309.8	10.8	—-	—-	1.08
UV Cet	Flare	175.5	-75.7	—-	—-	—-
V1101 Tau	Flare	174.7	-18.7	1.09	0.70	0.39
V513 Gem	Flare	194.1	23.3	0.51	1.28	0.66

2704 training samples

—> 1045 (with d & opt. quiescence, 10 features)

—> 724 (with NIR, 14 features)

—> 88 (with X-ray, 17 features)

- Generative model

$$p(C_k|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|C_k)p(C_k)}{\sum_j p(\boldsymbol{x}|C_j)p(C_j)}$$

  - giving all prior and likelihood with Gaussian? other distributions?