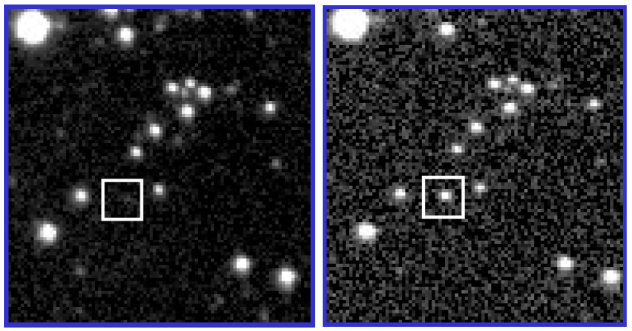
# A Stochastic Structure Function for Light Curves



Baback Moghaddam

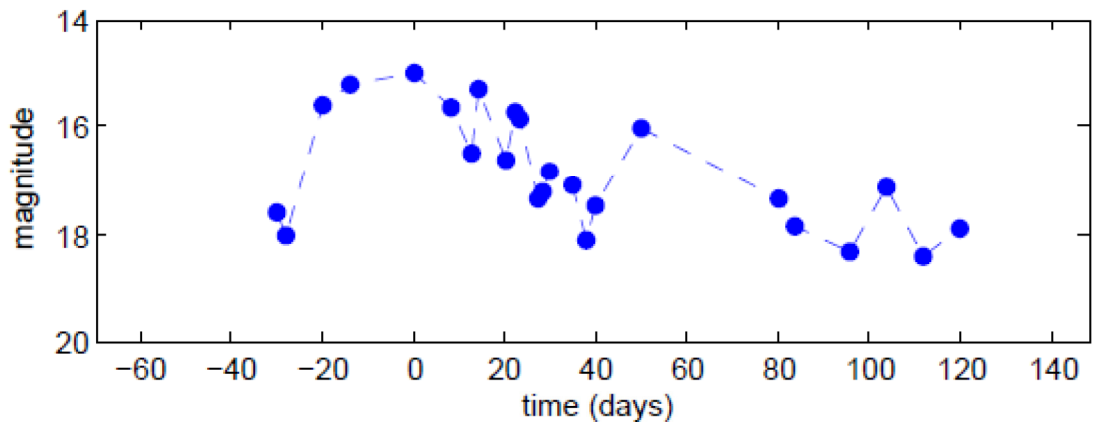
baback@jpl.nasa.gov

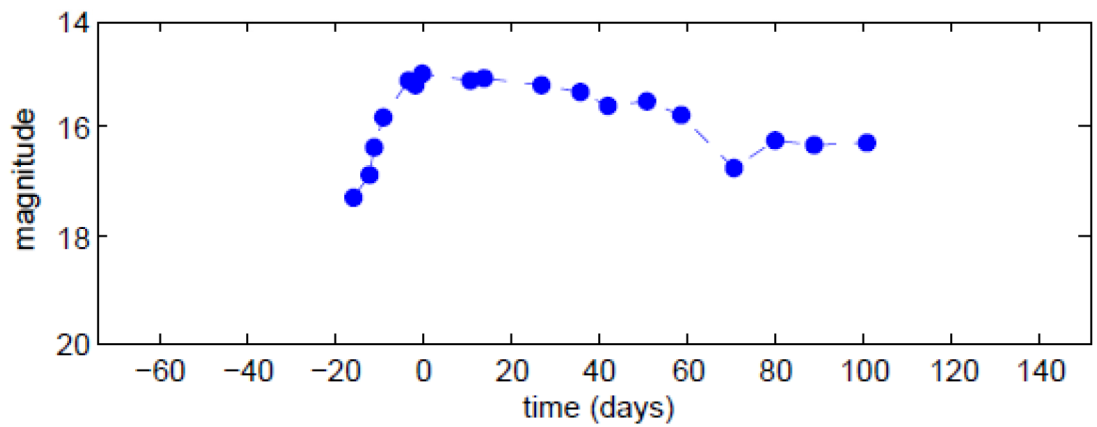
Machine Learning Group
Jet Propulsion Laboratory
California Institute of Technology

# Outline

- Light Curves
- Probabilistic Models
- Classification Results
- Structure Functions
- Conclusions

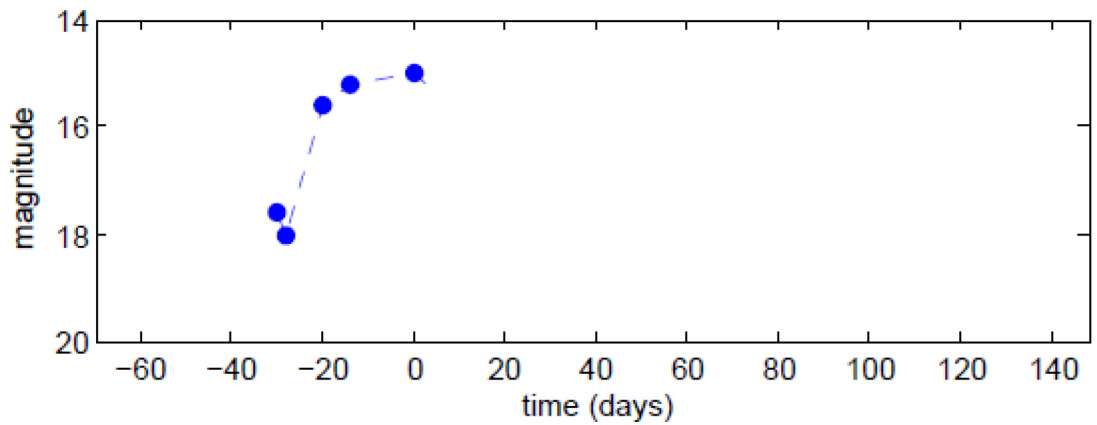
# Light Curves : SN-I *vs* SN-II

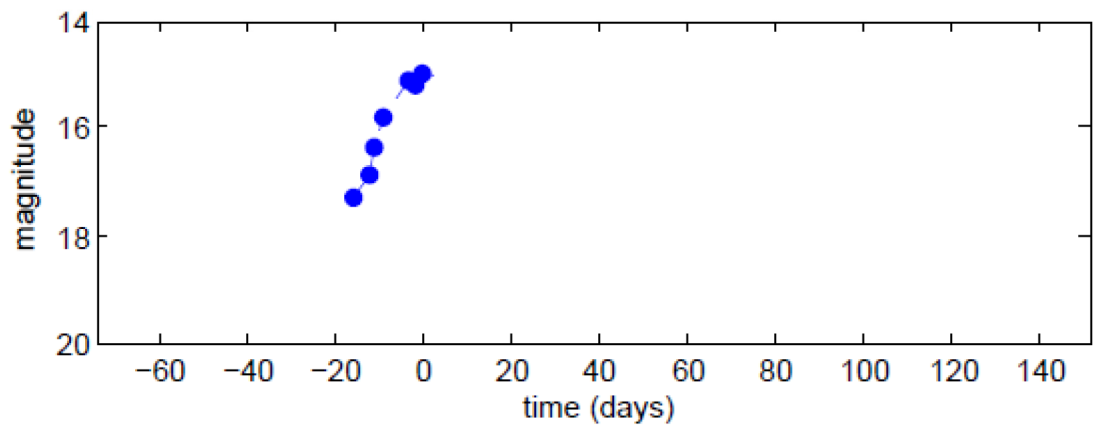



Classifying LCs with adequate # of obs is ~ “easy”

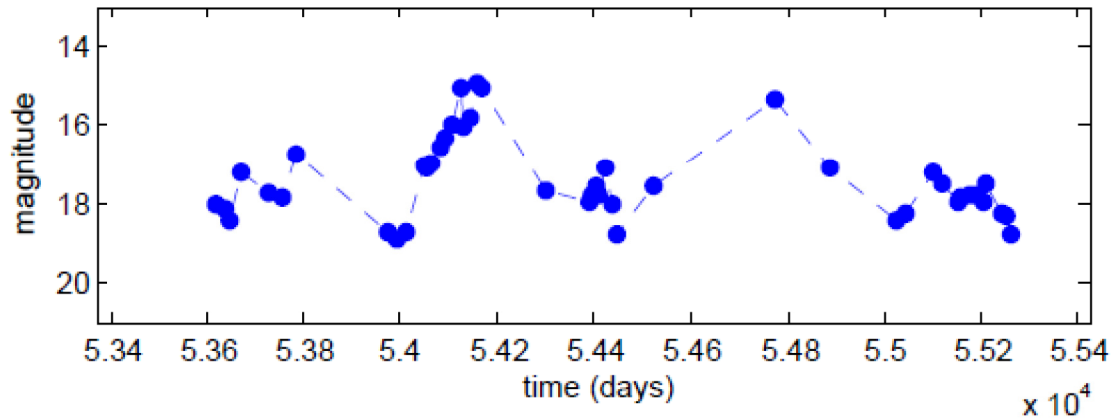
# Light Curves : SN-I *vs* SN-II

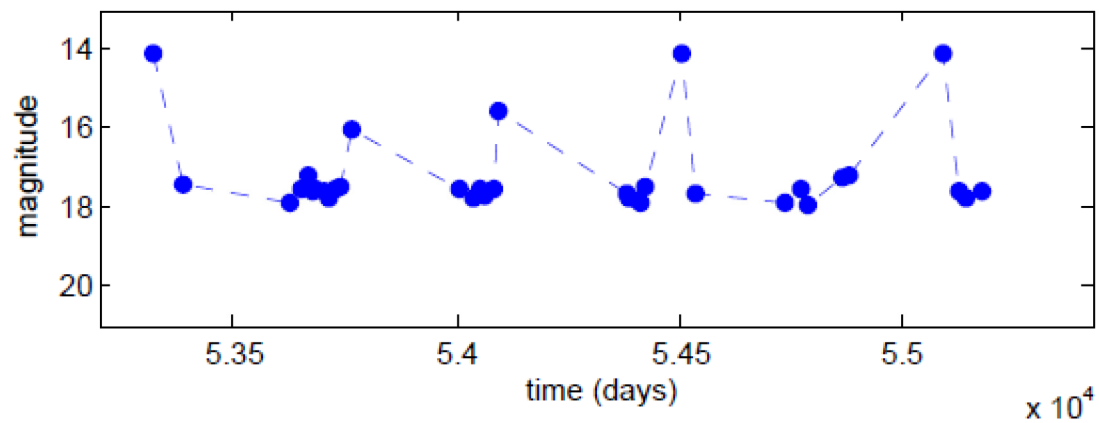



It's much harder yet more critical to classify them "early"

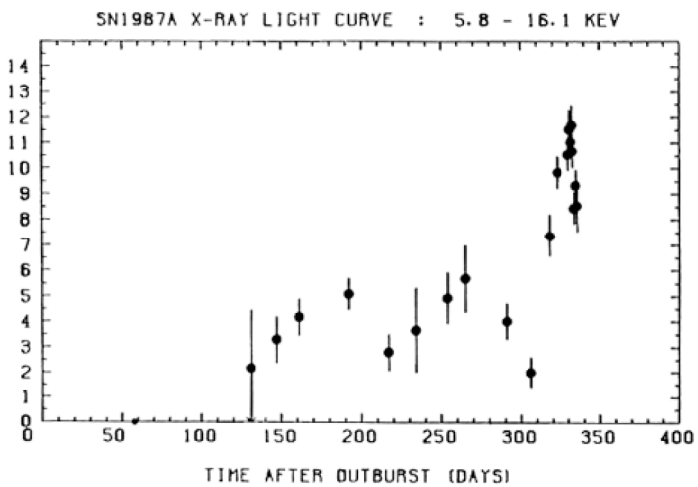
# Light Curves : Blazar *vs* CV




Some are intrinsically hard no matter how many obs

# Classifying LCs

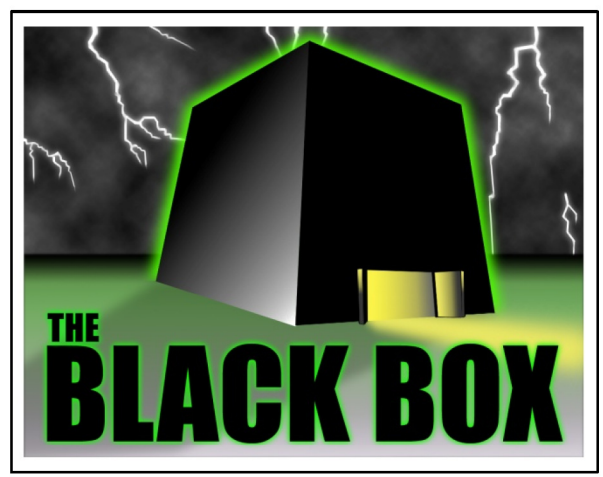



{ SN-Ia, CV, Blazar, … }

Classifier

Feature Extraction

$$\begin{bmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ f_n \end{bmatrix}$$



# Outline

- Light Curves
- **Probabilistic Models**
- Classification Results
- Structure Functions
- Conclusions

# "Gap Histograms" (SSF)

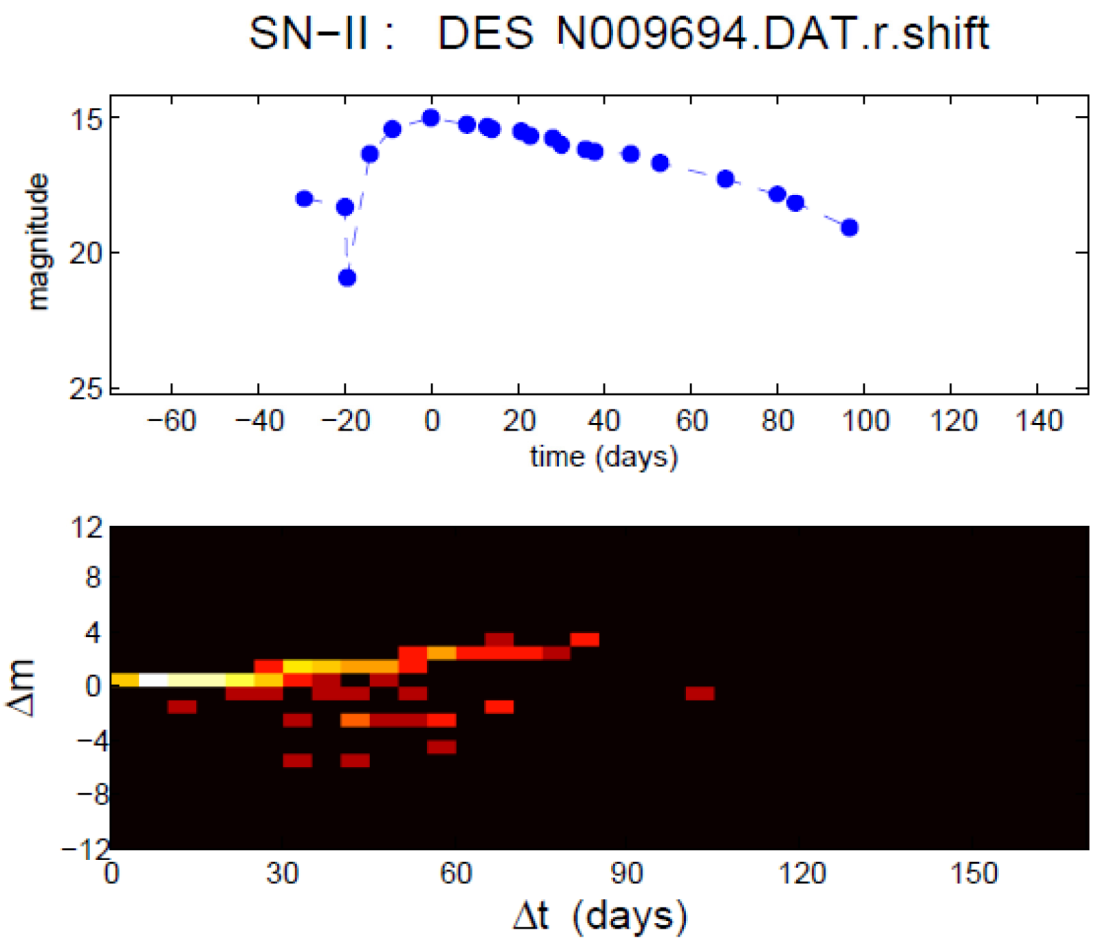
- Model distribution of "gaps" $(\Delta t, \Delta m)$ in LC

  $$LC(t, m) \rightarrow \{ (\Delta t, \Delta m) \} \rightarrow \text{density model}$$

- Causal gaps $(t_k - t_j, m_k - m_j)$ where $t_k > t_j$

- Key advantages:
  - No LC pre-processing (alignment) required
  - Minimal storage requirements for each LC
  - Fast classification: histogram likelihoods
  - Every object class is modeled *in the same way*
  - Can incorporate "error bars" in magnitude obs

# LC ⇨ Gap Histogram



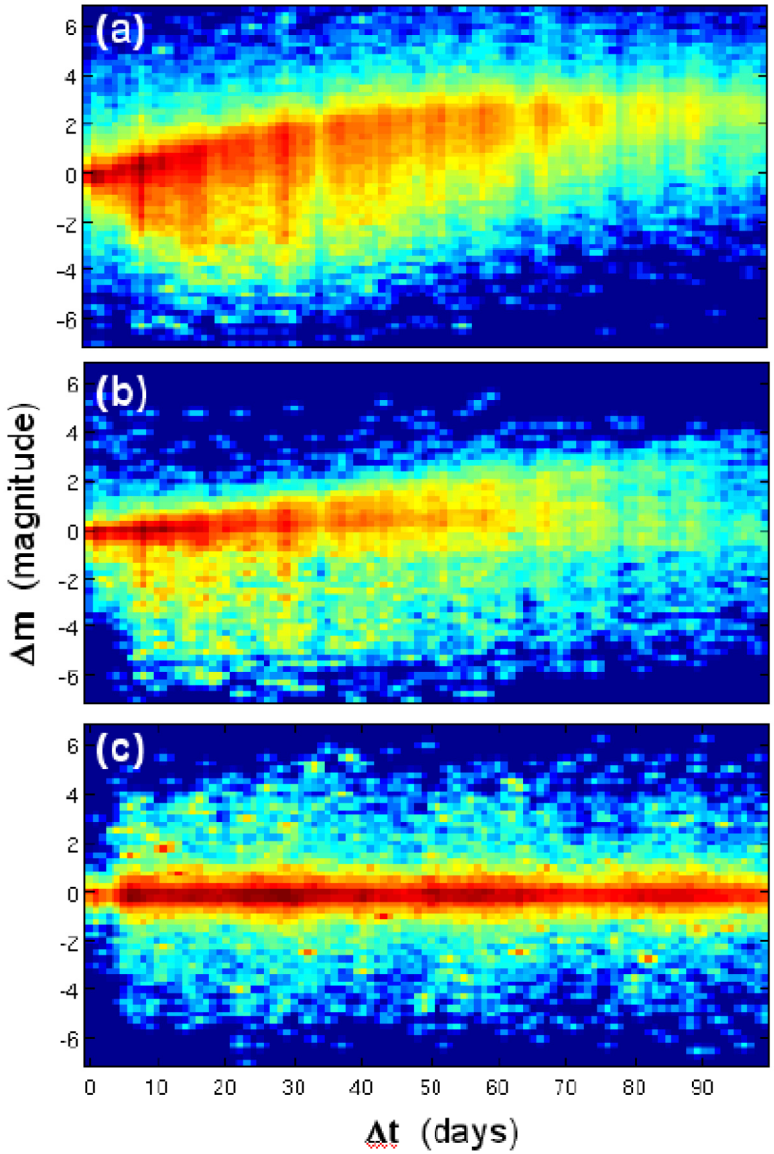

Gap histograms let you work with LC fragments and to also *pool* multiple LCs

The underlying density model is *invariant* to absolute magnitude and time shifts (no alignment)

Gap histograms can model not just error bars but also bounded flux observations

*e.g.* under poor seeing, we may only have bounds such as **m** > 18

# Class Prototype Histograms




**SN-Ia** (879 LCs)

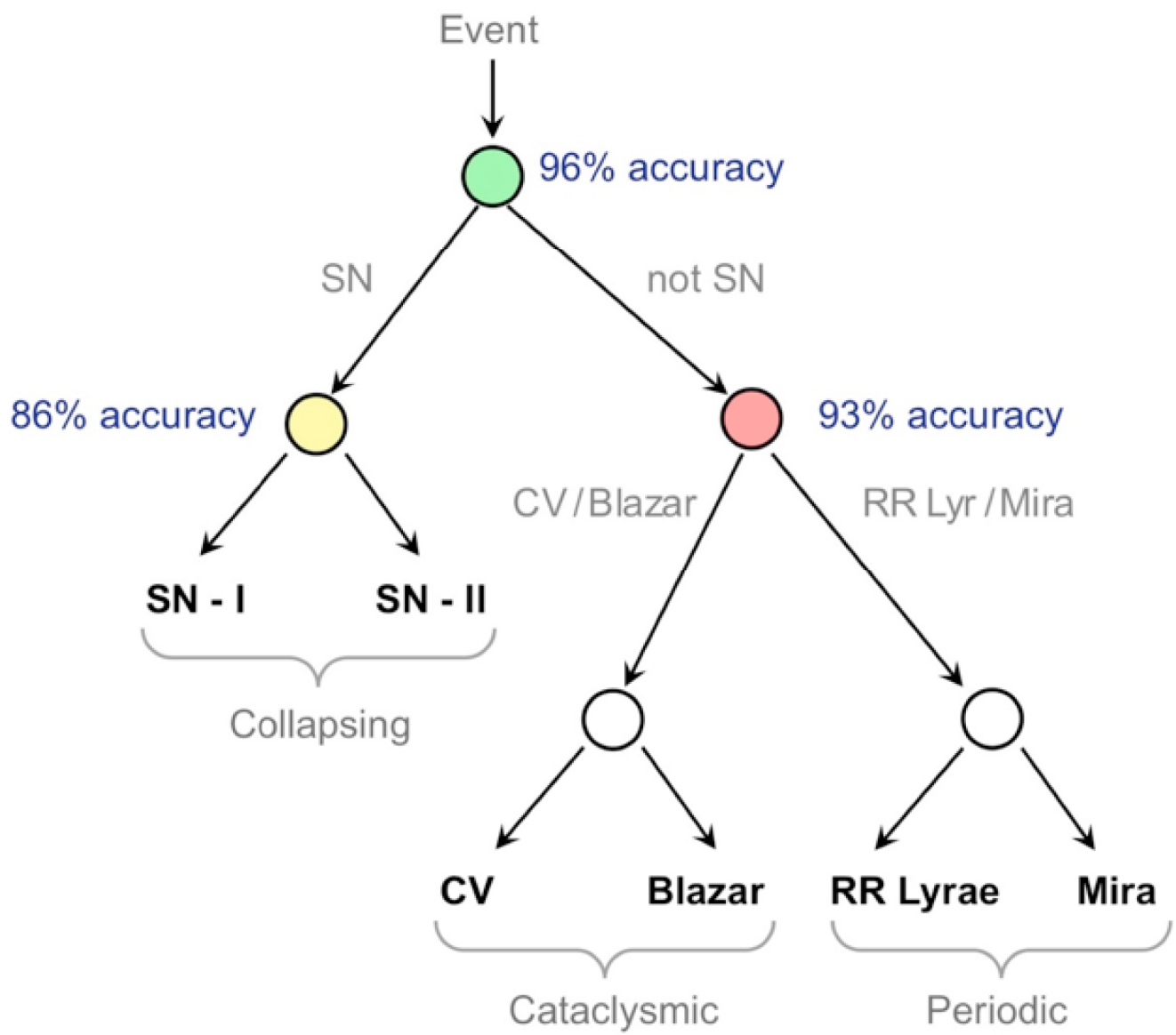
**SN-IIp** (282 LCs)

**CV** (426 LCs)

# Outline

- Light Curves
- Probabilistic Models
- **Classification Results**
- Structure Functions
- Conclusions

# Decision Tree Classifier



# Decision Tree Nodes

- each nodes uses a gap histogram based classifier
  - currently independently optimized
- histogram density model for LC gaps ($\Delta t, \Delta m$) is built for the two classes being discriminated
- classify LCs based on standard inferential rules

**Posterior odds = (Likelihood ratio) x (Prior odds)**

# DT Classifier Results (1)

## LC Test Data

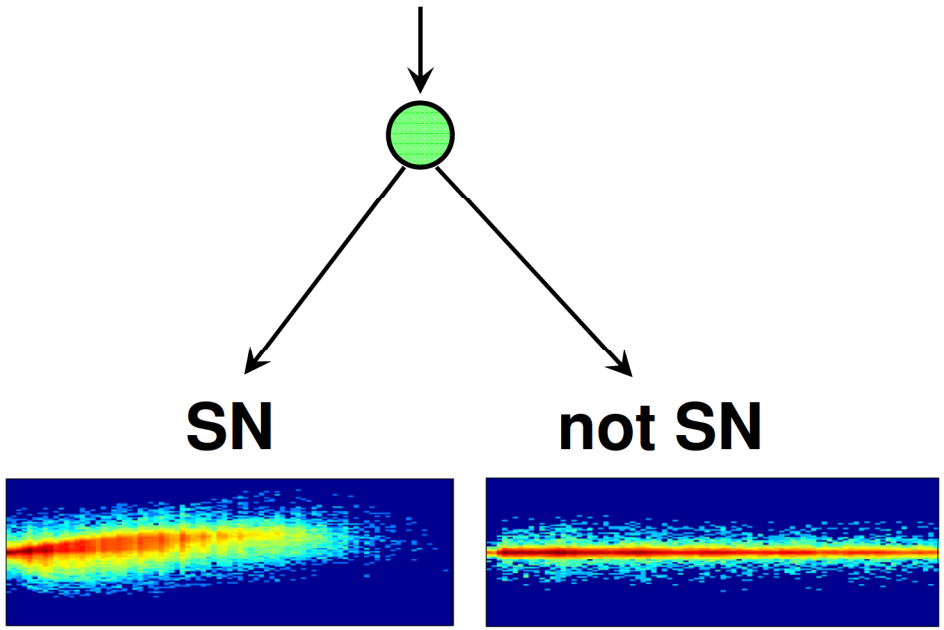
Class 0 : 1236 SN

(908 SN-Ia + 328 SN-II)

Class 1 : 932 “not SN”

(121 Blazars + 417 CVs

+ 29 Miras + 355 RR Lyrae)



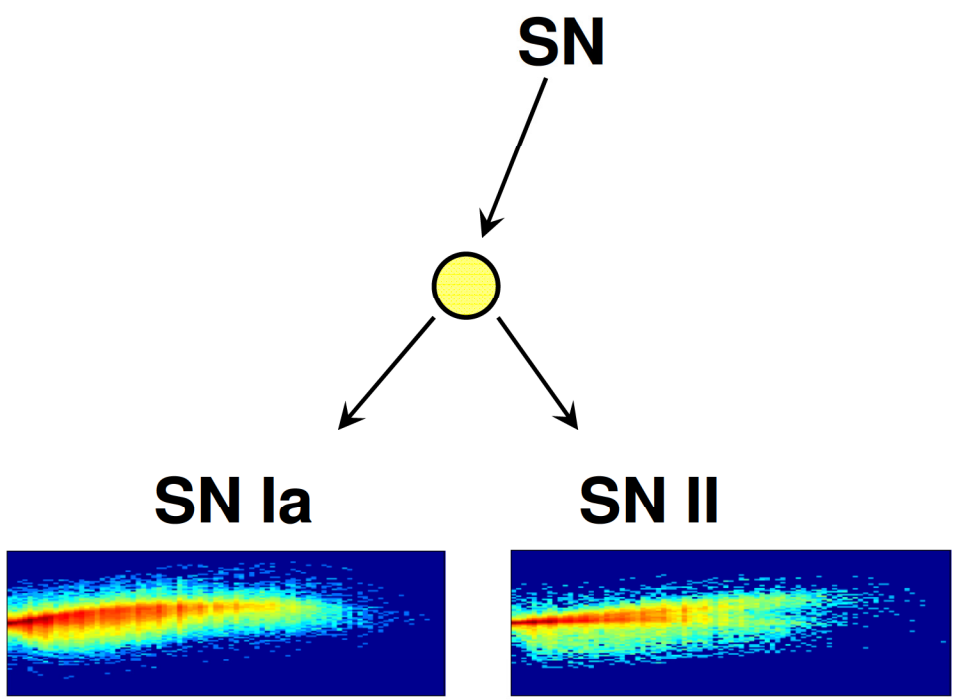

Accuracy = **96** % (assuming equal priors)

# DT Classifier Results (2)

## LC Test Data

Class 0 : 908 SN Ia

Class 1 : 328 SN II




Accuracy = 86 % (assuming equal priors)

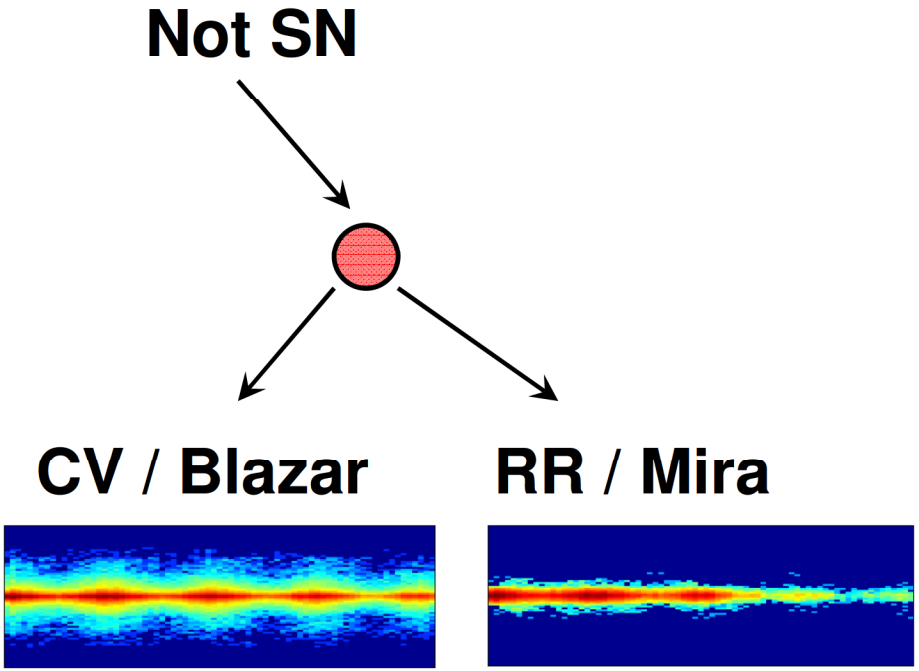
# DT Classifier Results (3)

## LC Test Data

<u>Class 0</u> : 417 CV + 121 Blazars

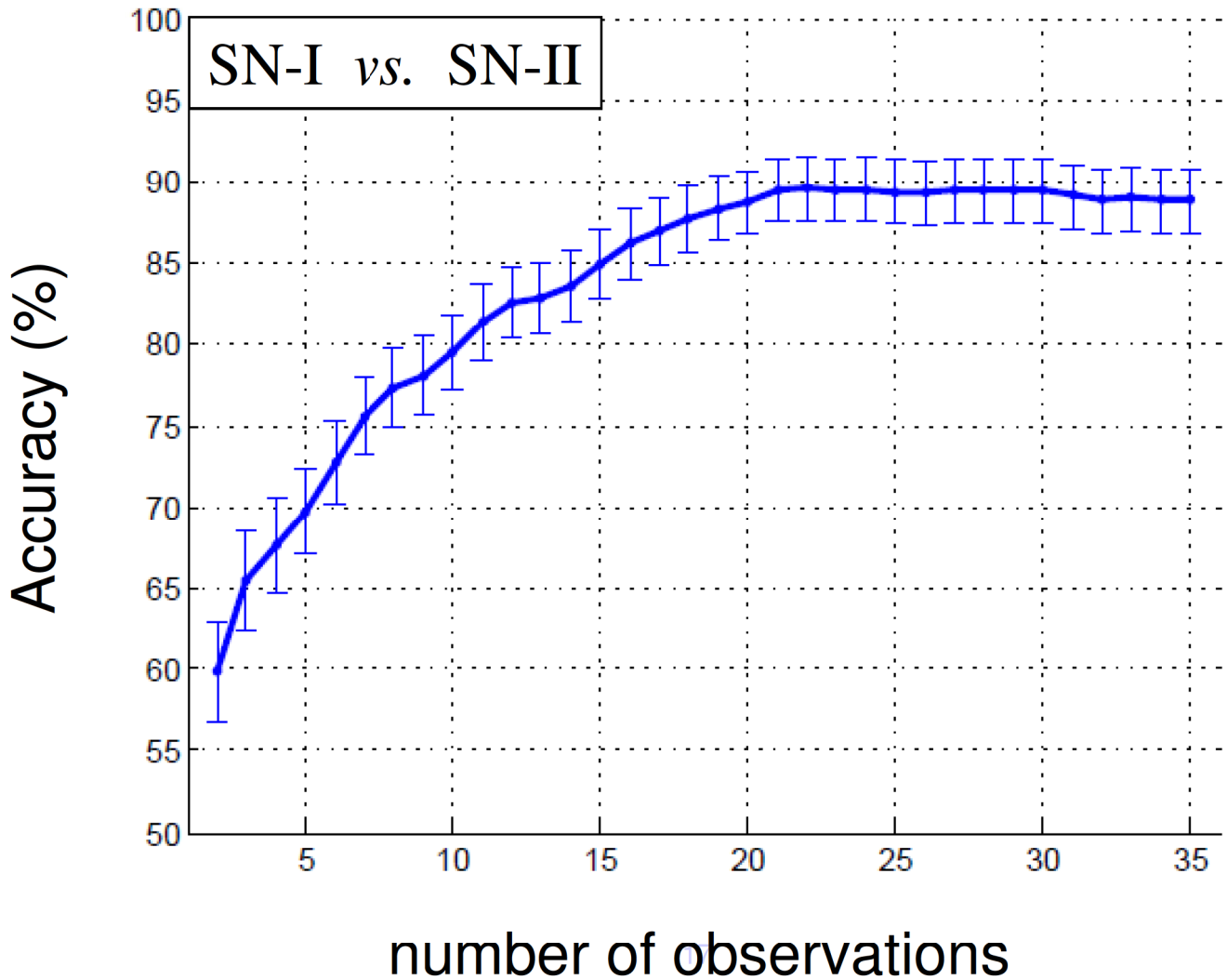
<u>Class 1</u> : 355 RR + 29 Miras

.

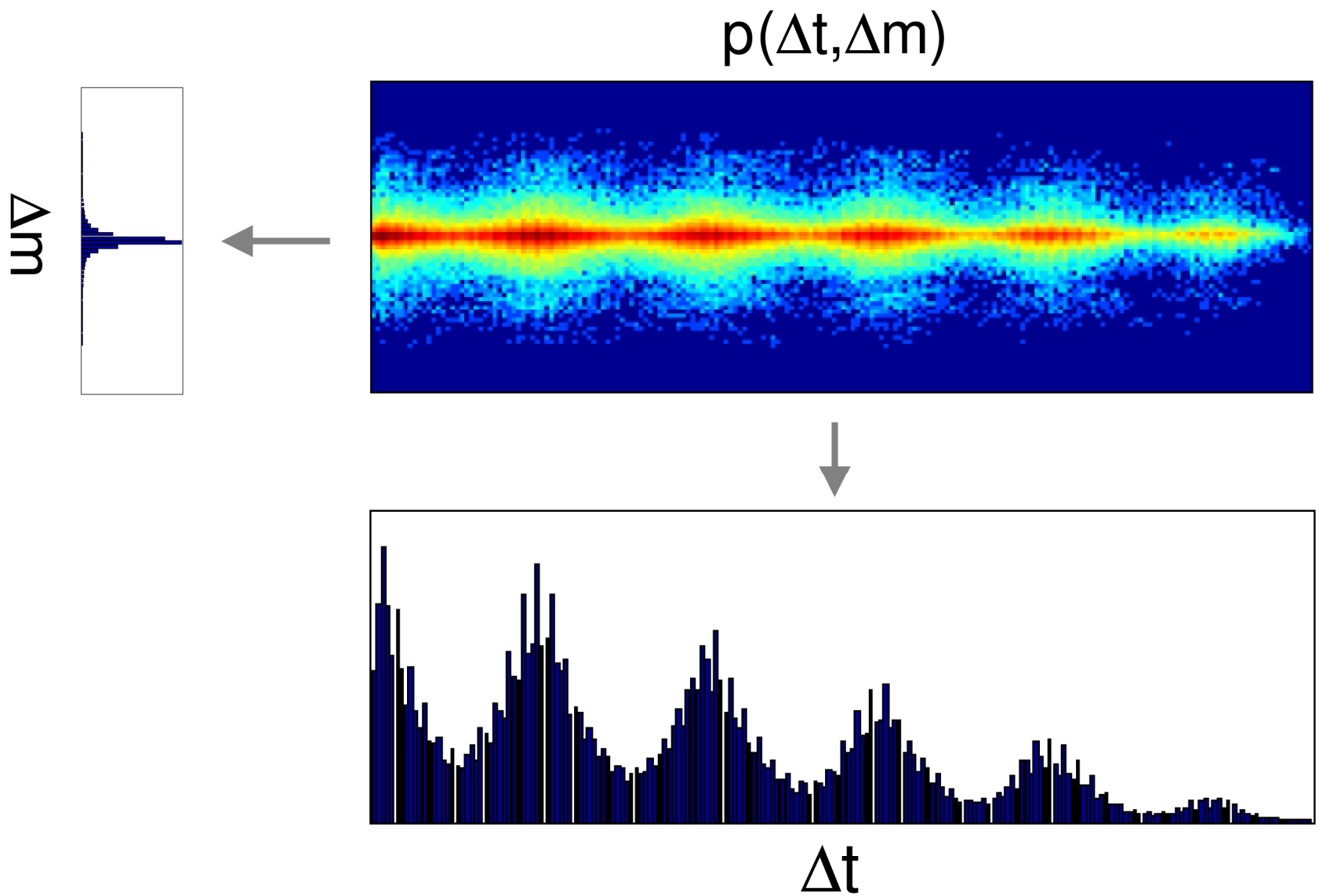



Accuracy = **93** % (assuming equal priors)

# Performance *vs.* $N_{obs}$



# Marginal Distributions



# Outline

- Light Curves
- Probabilistic Models
- Classification Results
- **Structure Functions**
- Conclusions

# Structure Functions

- "1st order" SF

$$SF(\tau) = \frac{1}{N(\tau)} \sum_{i=1}^{N} w(t_i) w(t_i + \tau) [x(t_i + \tau) - x(t_i)]^2$$

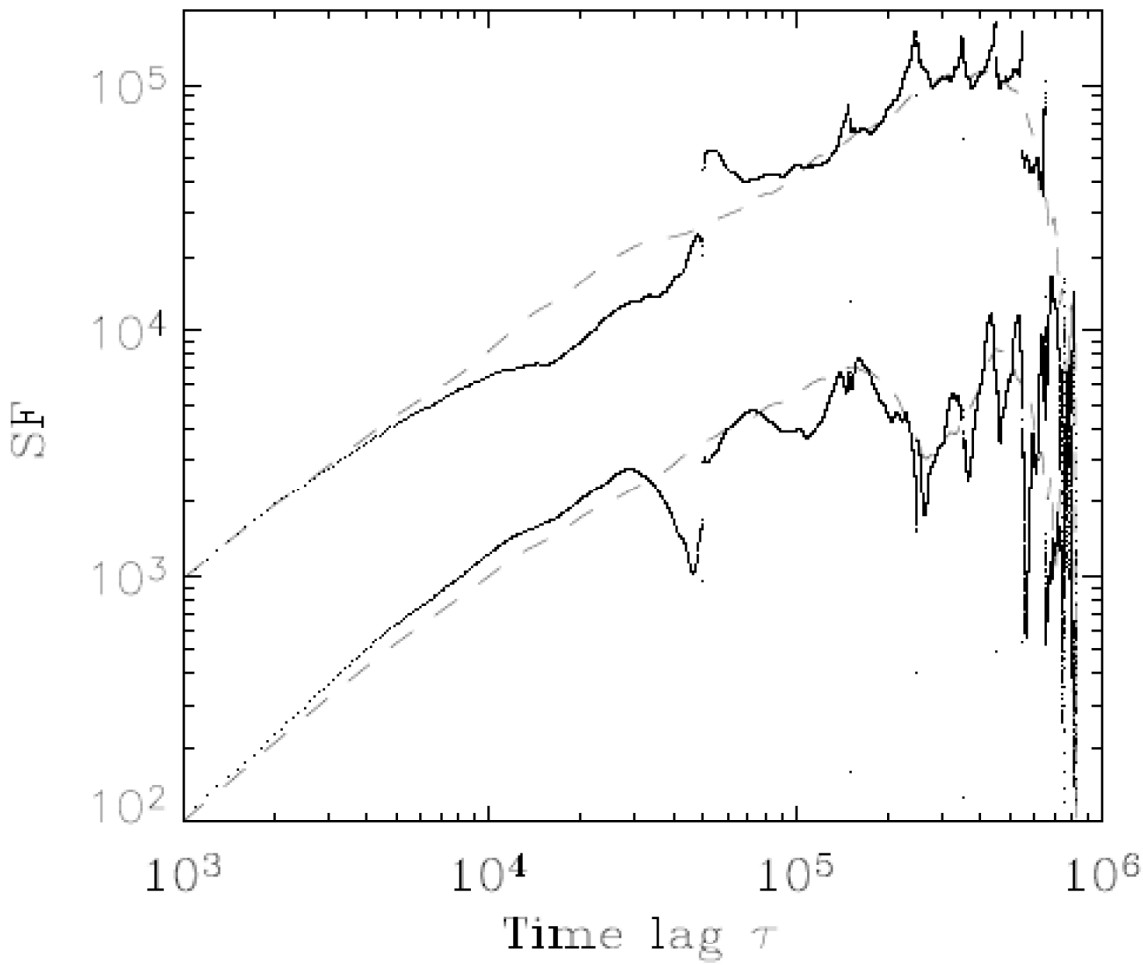
- "variance" of increments with time lag Δt

$$\boldsymbol{SF(\Delta t) = \langle [m(t + \Delta t) - m(t)]^2 \rangle}$$

Simonetti J.H., Cordes M.J., Heeschen D.S., 1985, ApJ, 296, 46

# Structure Functions

Two Red Noise Sources

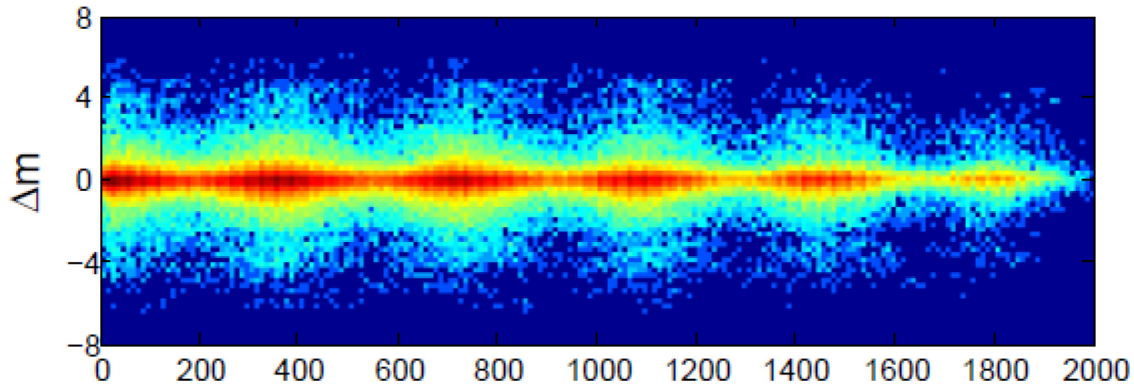



*P. Arevalo*

# “Gap Histogram” = Stochastic SF

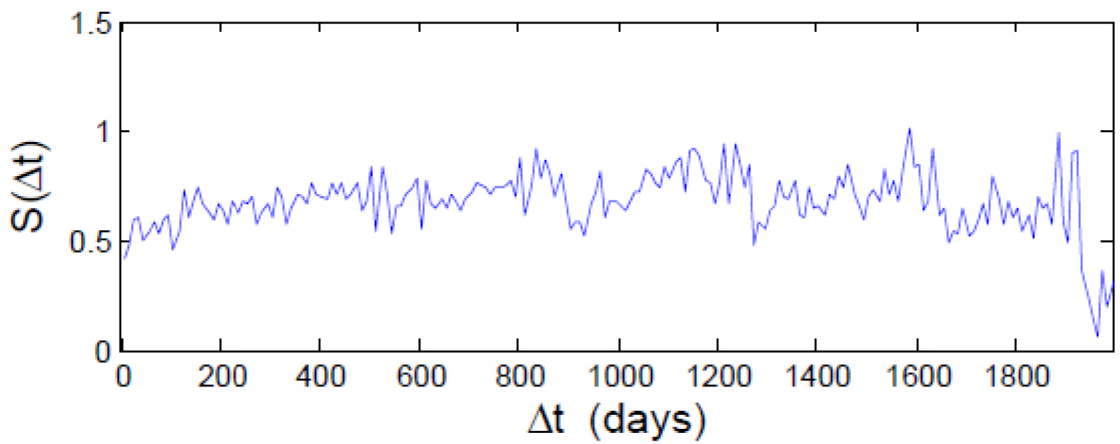
**CV**

**SSF**




Instead of a scalar var(Δm) at each lag Δt SSF gives a full pdf on Δm

**SF**

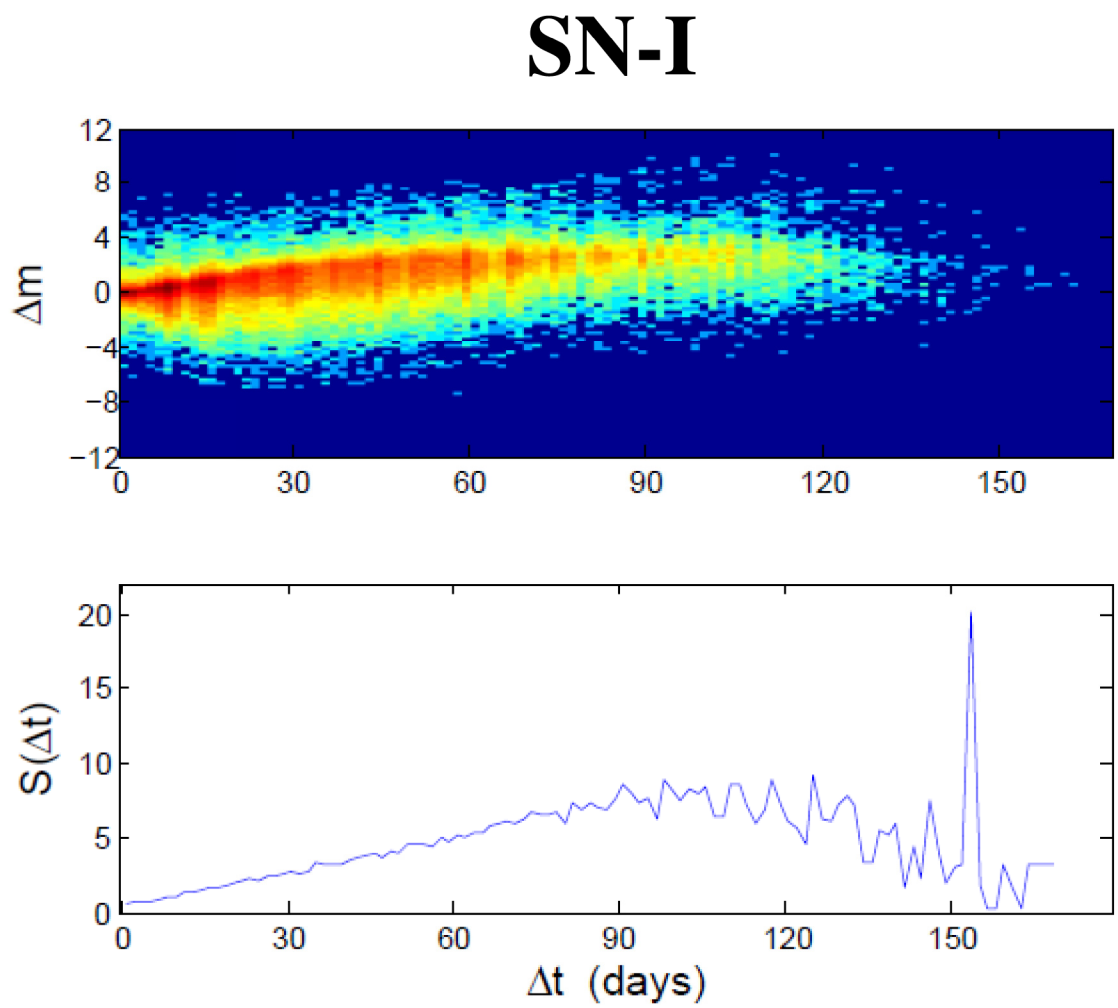



A standard SF can be trivially computed from the SSF

# “Gap Histogram” = Stochastic SF

SSF

SF




Instead of a scalar var(Δm) at each lag Δt SSF gives a full pdf on Δm

A standard SF can be trivially computed from the SSF

# Outline

- Light Curves
- Probabilistic Models
- Classification Results
- Structure Functions
- **Conclusions**

# Conclusions

- SSF is a fast, cheap, flexible, probabilistic method
- Handles LC fragments, error bars, bounds, *etc.*
- Optimal Binning of $\Delta t$-$\Delta m$ space (with Jeff Scargle)
- Can do full multi-class classification (not just binary)
- SSF can be informative for *optimal sampling*
- A standard SF can be easily derived from a SSF

# *The End*

© 2011 California Institute of Technology
Government sponsorship acknowledged