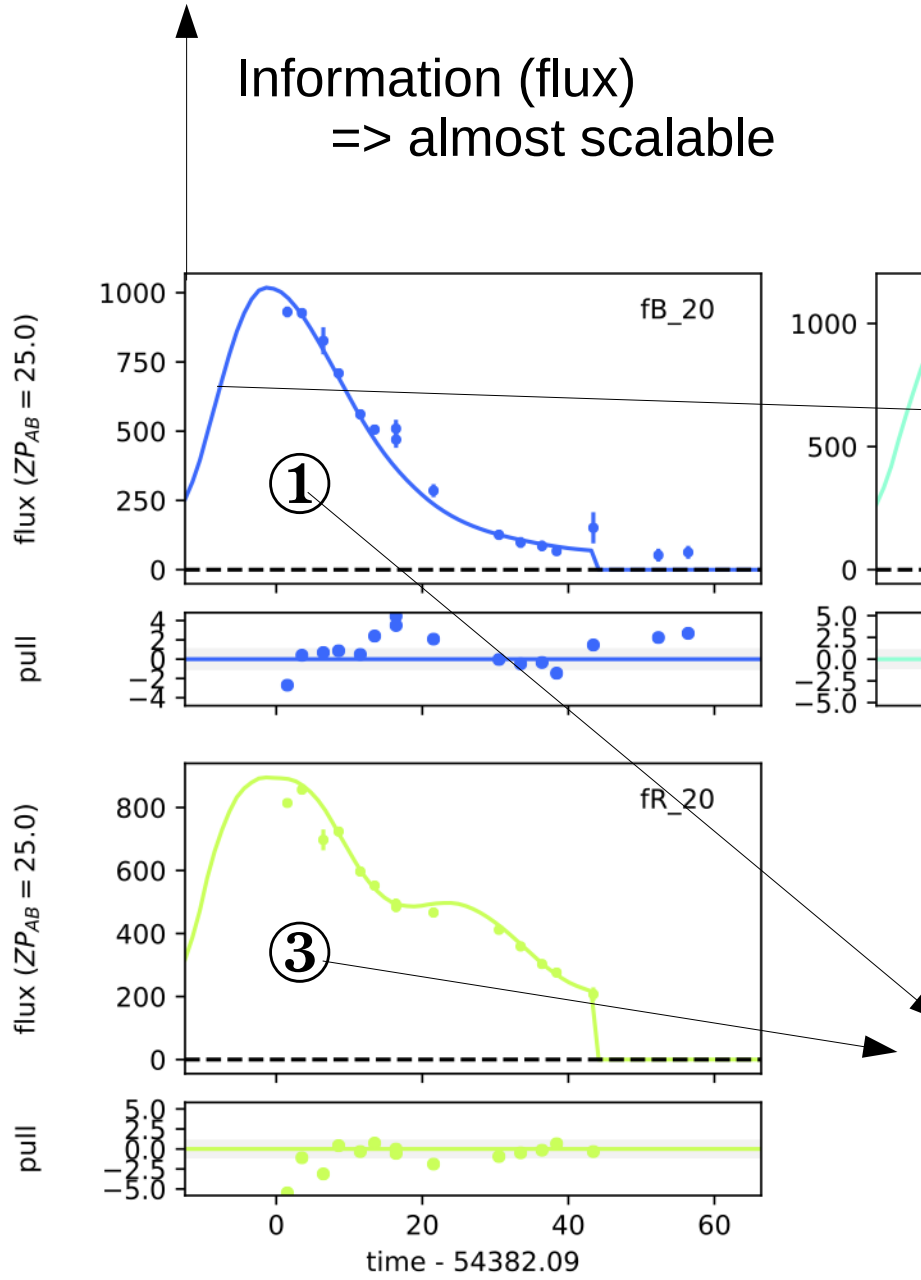
Time Series

An ~~astronomer~~ personal view ...

« time » series



Data : **Sparse**
Error bars

Function :
Model
Interpolation

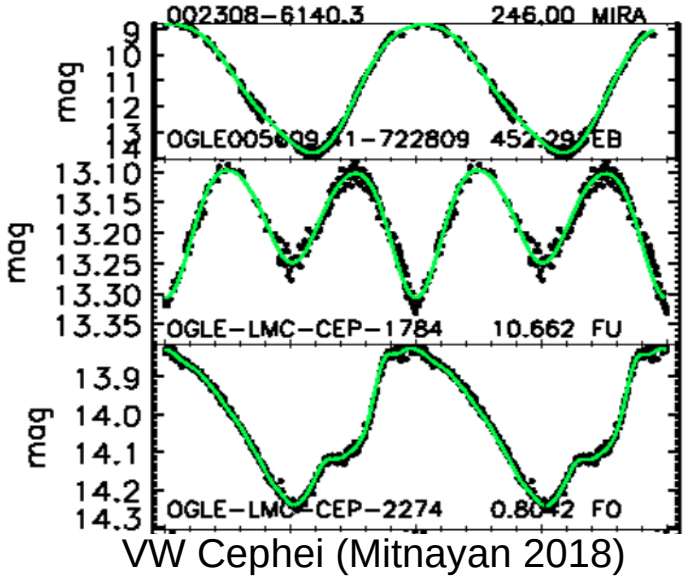
Time
axis

Origin unknown

3 filters (3 timelines)
data may « leak » between filters

Different time series

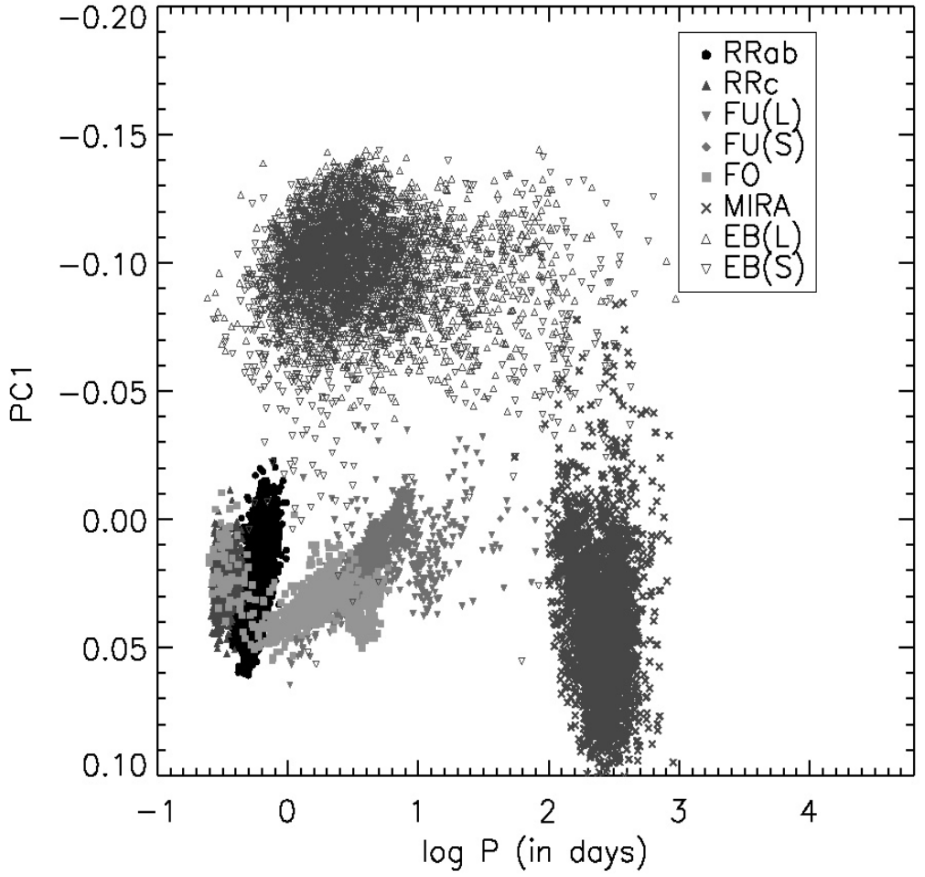
Periodic



VW Cephei (Mitnayan 2018)

Fourier Transform

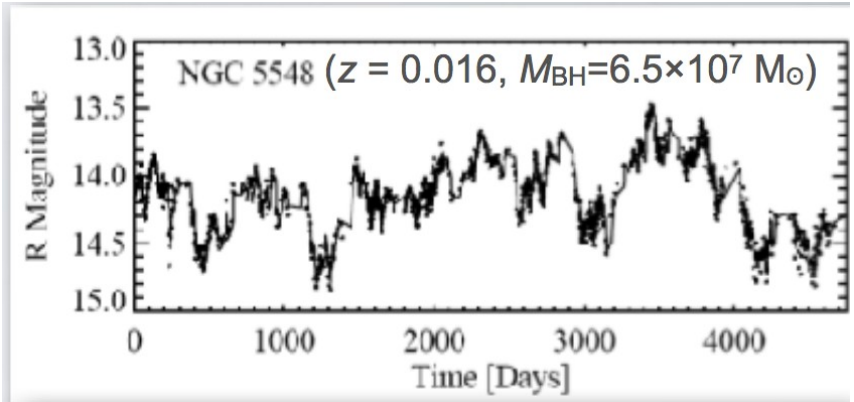
Then :
PCA+Clustering

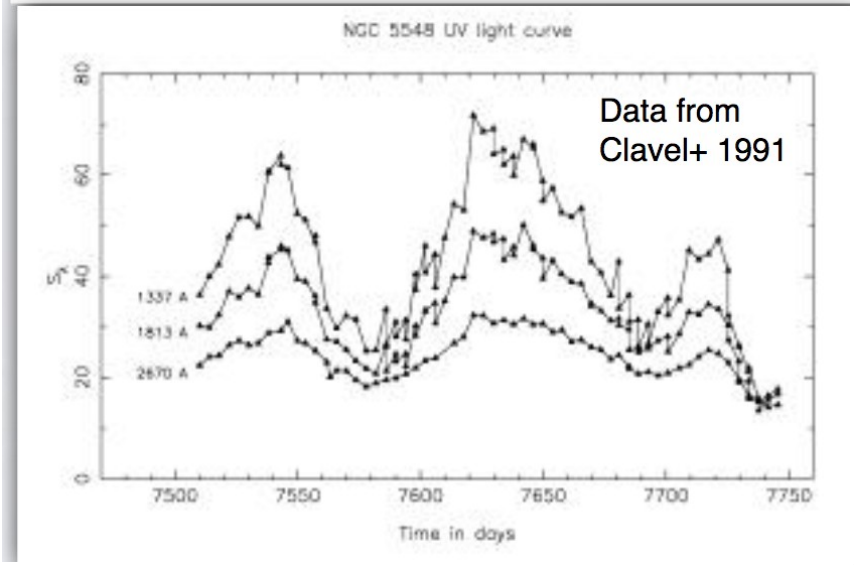


Deb & Singh 2009

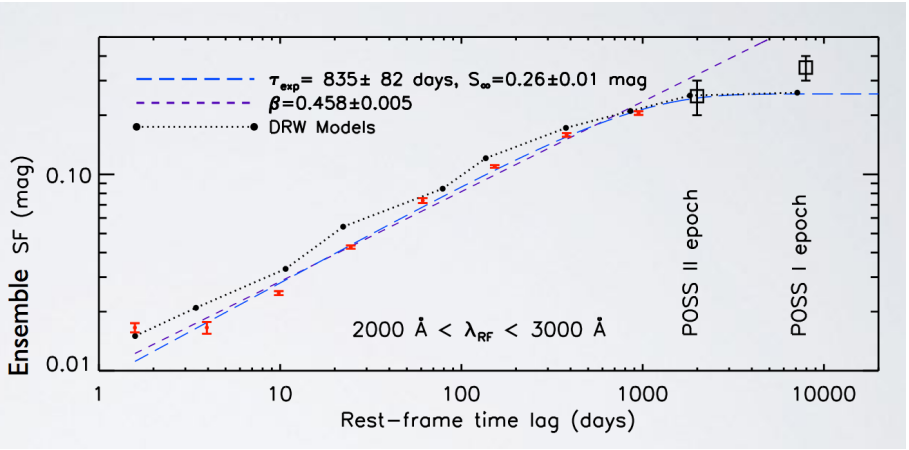
Different time series

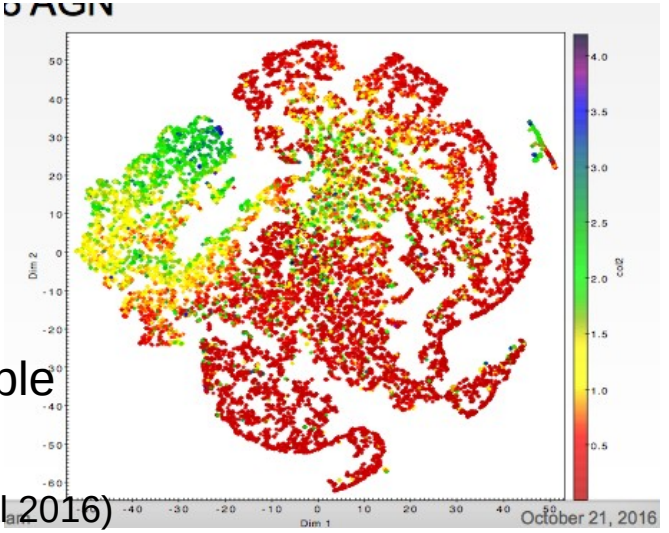
Aperiodic (dense)





Structure Function Analysis
RMS(Δm) at Δt





Classification possible
(Continuous
Autogregressive
Moving Average; Kasliwal 2016)
t-SNE

Transients :

Heterogeneous sparse time sampling

- Intermediate representation (features)
  - Template fitting
  - Non-parametric interpolation / Discretization
  - Basis of functions
  - Dealing with « holes »

    → Then : usual classifier
- Direct approach
  - e.g. Recurrent Neural Networks

Application to SN data

Lochner 2016

ArXiv 1603:00882

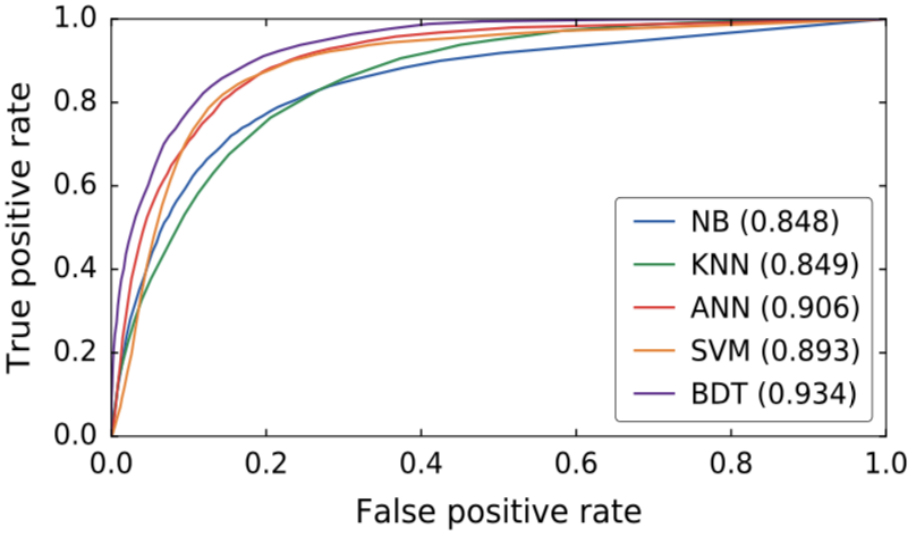
3 approaches :

- Template fitting (1 Class has a reliable template)
- Parametric approach (Light-curve as a parametrised function)
- Wavelet decomposition
  - Gaussian process interpolation
  - Wavelet decomposition (à trous/symlet)
  - Dimensionality reduction (PCA)

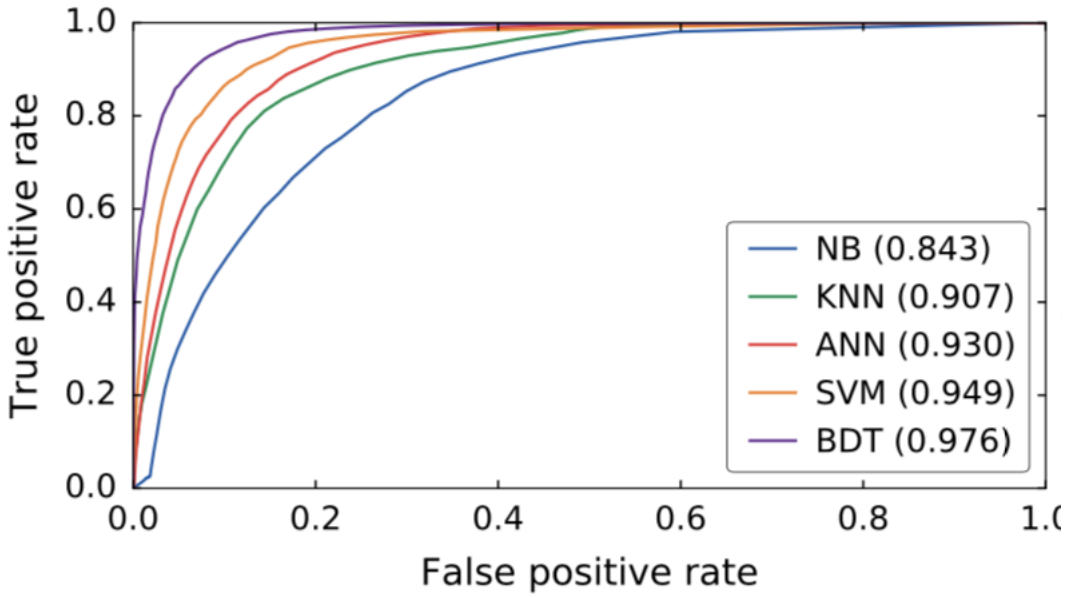
Then : Machine learning (Naive Bayes, KNN, ANN, BDT, SVM)

(Deep learning : not competitive for this case)

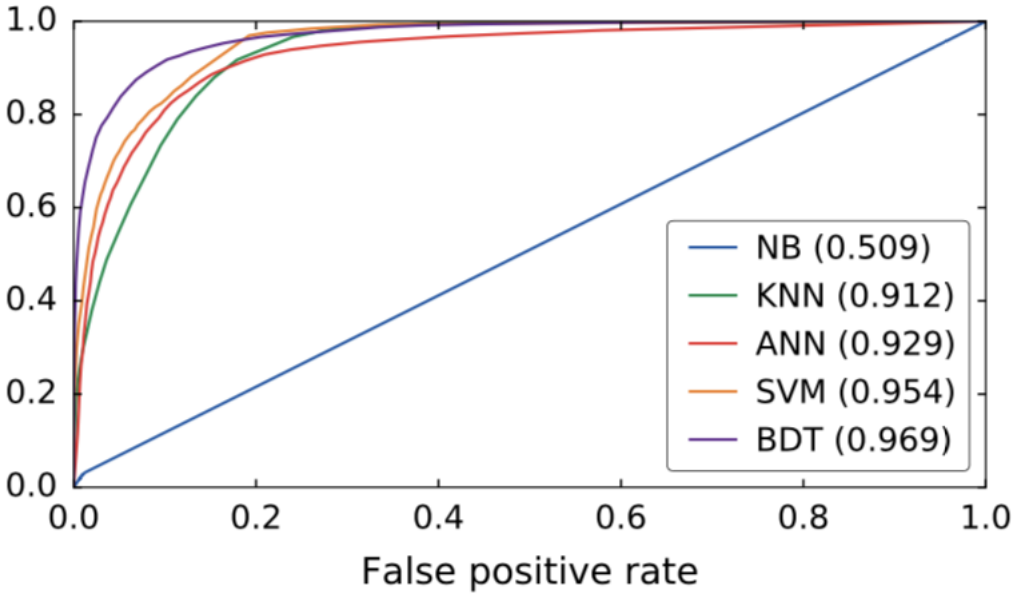
Results



(c) Parametric model 1, no redshift

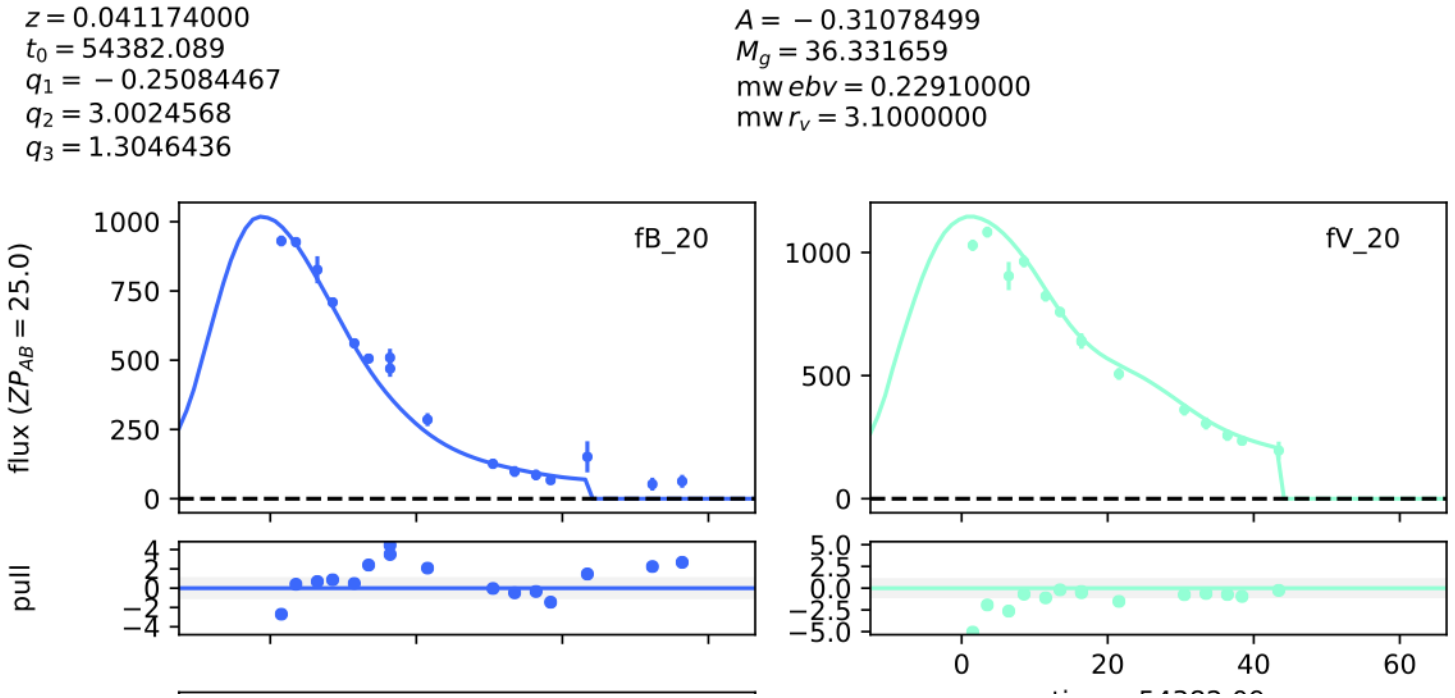


(a) SALT2 model, no redshift



(g) Wavelets, no redshift

Template fitting



Pro : parametrization well known for 1 class
- low dimensionality output
- statistical goodness of fit
- ingredient of likelihood ratio

To be improved
- template for the other class(es)
- automatic template generation ?

Parametric approach

$$F(t) = A\left(\frac{t-\phi}{\sigma}\right)^k \exp\left(-\frac{t-\phi}{\sigma}\right) k^{-k} e^k + \Psi(t),$$

where

$$\Psi(t) = \begin{cases} 0 & -\infty < t < \infty \\ \text{cubic spline} & \phi < t < \tau \\ \psi & \tau < t < \infty \end{cases},$$

To be improved :

- result depends on function description (can be poor)
- is it possible to automate the search for a few families of functions ?

Is it worth investigating more ?

Wavelet decomposition

Pro :

Basis of function provides natural feature extraction scheme
(remark : templates are based on basis of functions - splines)

Some questions :

Is there a need for interpolating the data before projecting on the basis ?
→ null values when no data available
→ null values already present in original data…
Errors may be poorly handled

Some care required for time translation invariance

Is this the best way to go ?

Last remarks

- Feature extraction may depend on classes to be found
  - Physical classes (Transient/Variable, Supernova types, …)
  - Empirical classes (clustering of metadata…)
  - For instance, GP use an « average » function
- Can we skip the time interpolation ?
  - Lot of missing values…
  - Other formulations (RNN)
- The ideal solution should be robust / reprensentativeness issue
  - Templates are ; more difficult for NN