Modeling Light Curves for Improved Classification by Julian Faraway et. al (2014)

Andy Lawler

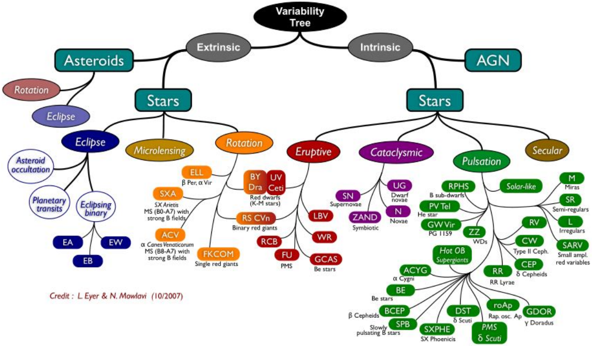
Baylor University

November 16, 2016

Background

All objects in the sky vary in brightness, but for the vast majority of objects the variations are not detectable due to the phenomena such as

1. Long period of variation (tens to hundreds of epochs)
2. Short observation window (seconds to minutes)
3. Short time scales over which the observations occur (a few years to a decades)



Background

- In order to better and understand and classify non-variable, variable, and transient objects we need to understand as many sources of variability as possible.
- Trick is to design measures that can aid in identifying certain classes **and** are derivable based on given cadence of observations.

Aim of paper

Present **new measures** based on object lightcurves which better classify 1) variables and non-variables and 2) among different transient types.

Data

- Four images taken ten minutes apart. This translates to four observations in a 30 minute time frame.
  - The next set of four images varies from the next night or week or even the next month.
  - Sometimes only one or two or three reliable observations can be made in the 30 minute window (presumably due to weather or technical issues).
  - Sometimes the observation is censored due to the instrument detection limit of around 20.5 magnitude. In this case only an upper limit on the magnitude can be specified.

Data

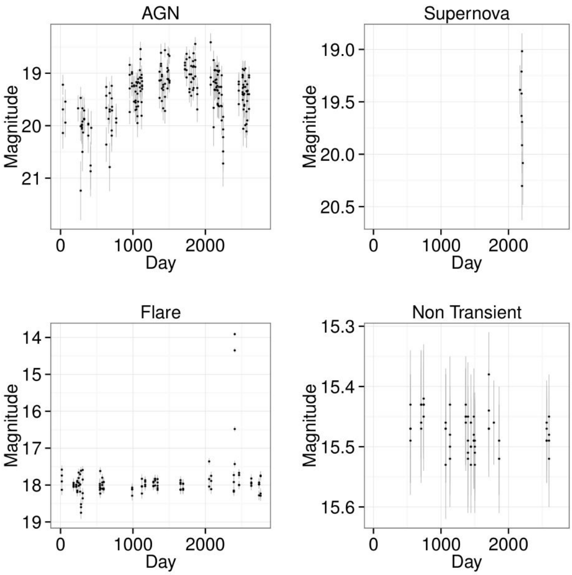
- **3720 lightcurves** selected between April 2005 to October 2012, spanning 2764 days
- The number of observations for each lightcurve varied from 5 to 641, with a median length of 52.
  - Objects that had fewer than 5 observations were excluded since they couldn't reasonably be classified.
- This sample is not representative of real world data. A simple random sample of astronomical objects would yield many many fewer non-variable objects.
- Analysis assumes all objects may be surveyed through period of study BUT failures to observed have not been recorded.

Data

- Started with transients detected by CRTS in real-time over about five years:
  - Active Galactic Nuclei (AGNs)
  - Blazars
  - Cataclysmic Variables (CV)
  - Flare stars
  - Supernovae (SNe)
- Added a set of 15 random pointings and objects within 3 arc minutes. All are assumed to be non-transients.
- Added two classes of brighter variable objects:
  - Cataclysmic Variables from the Downes set (Downes et al. [2005])
  - RR Lyrae (periodic variable stars with a period of about 1 day)

Functionally transient							
Transients					Bright variables		
AGN	Blazar	CV	Flare	SNe	CV Downes	RR-Lyrae	non-transient
140	124	461	66	536	376	292	1971

Data


AGN
Magnitude
19
20
21
0
1000
2000
Day
Supernova
Magnitude
19.0
19.5
20.0
20.5
0
1000
2000
Day
Flare
Magnitude
14
15
16
17
18
19
0
1000
2000
Day
Non Transient
Magnitude
15.3
15.4
15.5
15.6
0
1000
2000
Day

- The paper uses 16 of the measures presented by Richards et al. (2011). These measures were chosen because they were
  - Helpful in distinguishing objects
  - Could be applied quickly and reliably for both short and long lightcurves
- They are considered as the baseline in this analysis (since they have been widely tested, at least for brighter data.)

Moment-based Features	skew	Skew of the fluxes
	kurtosis	Kurtosis of the fluxes, reliable down to a small number of epochs
	std	Standard deviation of the fluxes
	beyond1std	Percentage of points beyond st. dev. from the weighted mean
Magnitude Features	max_slope	Maximum absolute flux slope between two consecutive observations
	amplitude	Half the difference between the maximum and the minimum magnitude
	mad	Median discrepancy of the fluxes from the median flux
	medbuf	Percentage of fluxes within 20% of the amplitude from the median
	pairslope	Percentage of all pairs of consecutive flux measurements that have positive slope
	rcorbor	Fraction of observations greater than 1.5 more than the median
Percentile Features	fpr20	Ratio of flux percentiles (60th - 40th) over (95th - 5th)
	fpr35	Ratio of flux percentiles (67.5th - 32.5th) over (95th - 5th)
	fpr50	Ratio of flux percentiles (75th - 25th) over (95th - 5th)
	fpr80	Ratio of flux percentiles (90th - 10th) over (95th - 5th)
	preamp	Largest percentage difference between either the max or min magnitude and the median
	pdfp	Difference between the 2nd & 98th flux percentiles, converted to magnitude

Modeling

Establishing the lightcurve function

For lightcurve i, we posit a true underlying curve $f_i(t)$ that we would see if we could observe the object continuously without error. However, we are able to observe the object only at times t_{ij} for $j = 1, ...n_i$. Note that the times of measurement may be almost the same for objects close in the sky but quite different for objects which are farther apart.

We observe only y_{ij} for $j = 1, ...n_i$. We assume

$$y_{ij} = f_i(t_{ij}) + \epsilon_{ij}$$

where the errors ϵ_{ij} are normal with mean zero but will be correlated.

Benefits of GPR

1. We have censored data - the lightcurve can fall below the detection limit of around 20.5 during the range of observation.
2. The number of observations for each lightcurve varies greatly.
3. The measurement error is known.

Other modeling techniques may accomodate only one or two of these issues, but GPR accomodates all three.

Specification of the Model

The method requires a prior specification for the Gaussian process: $f(\mathbf{x}) \sim GP(\phi(\mathbf{x}), k(\mathbf{x}, \mathbf{x'}))$. The paper chooses the popular squared covariance kernel:

$$k(\mathbf{x}, \mathbf{x'}) = \sigma_f^2 \text{exp}\left(-\frac{1}{2l^2}(\mathbf{x} - \mathbf{x'})^2\right) + \sigma_n^2 \delta(\mathbf{x} - \mathbf{x'})$$

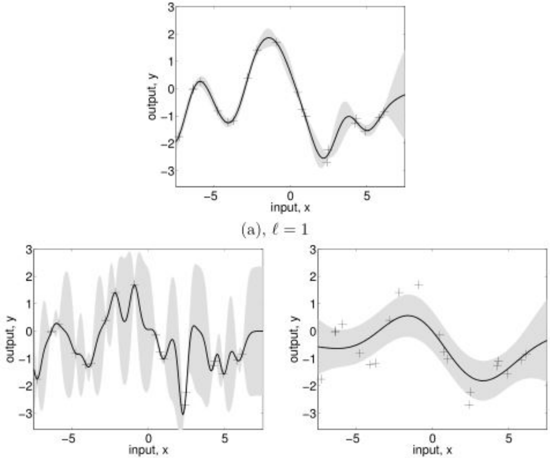
where $\delta(x)$ is the Kroneker delta and is 0 when $x \neq 0$ and 1 when $x = 0$. Also, $\mathbf{x}$ and $\mathbf{x'}$ are either in the training or test sets. Lastly, $k(\mathbf{x}, \mathbf{x'})$ is by definition equivalent to $E[(f(\mathbf{x}) - \phi(\mathbf{x}))(f(\mathbf{x'}) - \phi(\mathbf{x'}))]$.

Gaussian Process Regression

$$k(\mathbf{x}, \mathbf{x'}) = \sigma_f^2 \exp\left(-\frac{1}{2l^2}(\mathbf{x} - \mathbf{x'})^2\right) + \sigma_n^2 \delta(\mathbf{x} - \mathbf{x'})$$

There are **four prior components** which must be specified:

- σ_f^2 is the signal variance. We set σ_f^2 to be the median observed variance in the non-transients. This is because the vast majority of future objects to be classified will be non-transients, and these objects don't vary in signal much.
- σ_n^2 is the noise variance. We take the mean observed value of the measurement variance for σ_n^2 for simplicity sake and because the measurement error doesn't vary much from case to case.
- l is known as the length-scale and controls the amount of correlation and therefore smoothness in the posterior fit. A value of 140 days is used and is a subjective assessment. The classification performance is not very sensitive to this choice.
- $\phi(\mathbf{x})$ is the prior mean. An empirical Bayes approach is used.



(a), $\ell = 1$

(b), $\ell = 0.3$

(c), $\ell = 3$

Gaussian Process Regression

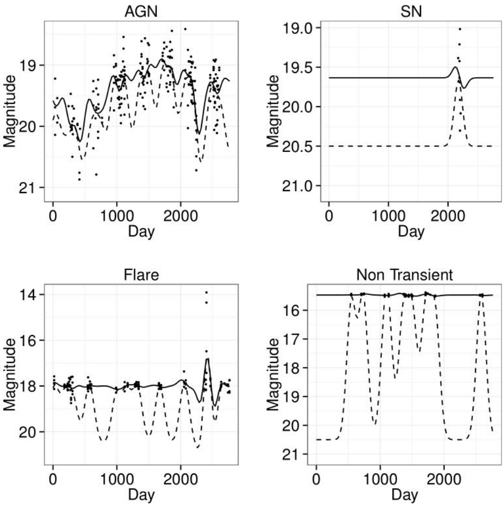


Figure 3: Gaussian process regression fits to lightcurve data. The same four cases as in Figure 2. The solid line fit derives from a prior mean set at the median magnitude while the dashed line fit corresponds to a prior set at a magnitude of 20.5.

Gaussian Process Regression

Solution: Use An Adaptive Prior for the Mean

- When there is less than one year of observations, we use the detection limit
- When there is more than one year of observations, we use the median magnitude.
  - The choice of a year is large enough that sparse but widely measured curves such as the non-transient do not use the detection limit.
  - The choice is small enough that the detection limit is used in cases like the supernova

Curve Measures

Next Step: Now we can compute fitted curve measures from $\hat{f}_i$ for curve i computed on the range of observation u_j for $j = 1, ..., m = 300$. (Note that $\hat{f}$ is the posterior mean.)

- *totvar* total variation: $\sum_j |\hat{f}_i(u_{i,j+1}) - \hat{f}_i(u_{i,j})|m$
- *totvar* total variation: $\sum_j (\hat{f}_i(u_{i,j+1}) - \hat{f}_i(u_{i,j}))^2 m$
- *famp* amplitude of fitted function: $\max_t \hat{f}_i - \min_t \hat{f}_i$
- *fslope* maximum derivative in the fitted curve: $\max_t |\hat{f}_i'|$
- *outl* maximum in absolute value of the scaled residuals from the fit

Curve Measures

More curve measures: The GPR is not able to model variation at the finer time scales of 4 observations every 30 minutes, so additional measures are needed to capture the characteristics at this scale. The mean within each of the groups of up to four observations is computed as $\check{f}_{ij}$ and then the following measures are computed:

- *lsd*: the log of the standard error, $\check{\sigma}$, computed using the residuals from these group mean fits.
- *gtvar*: The group total variation $\sum_j |\check{f}(t_{i,j+1}) - \check{f}(t_{i,j})|/n_i$
- *gscore* $\sum_j \phi((\check{f}_{ij} - \bar{f}_i)/\check{\sigma})/n_i$ where ϕ is the standard normal density, $\bar{f}$ is the mean of the fitted group means.

Curve Measures

Even More curve measures: There are some gaps in the Richards curve summary measures, so the following measures are added:

- *shov* mean of absolute differences of successive observed values: $\sum_j |y_{i,j+1} - y_{ij}|/n_i$
- *maxdiff* the maximum difference of successive observed values: $\max_j |y_{i,j+1} - y_{ij}|$
- *dscore* the density score: $\sum_j \phi((y_{ij} - \tilde{f}_i)/s_{ij}))/n_i$ where $\tilde{f}_i$ is the median observed magnitude for curve i and s_{ij} is the observed measurement error at t_{ij}.

Classification

Using these 11 new curve measures in addition to the 16 Richards curve measures, we can classify according to the following five methods:

- **LDA** Standard linear discriminant analysis
- **TREE** Recursive partitioning
- **SVM** Support vector machines
- **NN** Neural network
- **RF** Random forest ensemble

Note that measures that exhibited extreme skewness were log-transformed to improve classification performance. (The partitioning methods, TREE and RF, are invariant to monotome transformations.) Lastly, 2/3 of the data was randomly split for training and the remaining 1/3 was for testing.

Classification

Four different types of classification problems were considered:

- **All** Classifying eight types - the non-transients and the seven transient types.
- **Transient or not**
- **Transient only** For this problem, the non-transients were removed from the training and test data.
- **Hierarchical** An alternative approach is to first classify the objects into transient or not, then if transient to classify among the seven available types.

Classification Performance

	LDA	TREE	SVM	NN	RF
All	56.7	58.6	66.1	63.3	67.3
Transient or not	74.7	79.5	81.0	75.2	82.5
Transient only	54.5	58.9	64.4	60.1	62.9
Heirarchical	56.4	60.4	64.7	58.8	65.6

Table 2: Percentage correctly classified using the Richards measures.

	LDA	TREE	SVM	NN	RF
All	76.0	71.9	80.2	79.6	80.5
Transient or not	90.4	88.4	92.0	91.6	91.8
Transient only	70.1	65.1	74.3	72.3	74.2
Heirarchical	76.0	72.7	79.9	78.5	79.8

Table 3: Percentage correctly classified using our measures in addition to the Richards set.

Note that the standard error for the classification rate is just under 1%.

Classification Performance

Predicted	AGN	Blazar	CV	Downes	Flare	NT	RR-Lyrae	SNe
	Actual types							
AGN	0	0	0	0	0	0	0	0
Blazar	0	0	0	0	0	0	0	0
CV	5	26	95	53	4	20	3	27
CV Downes	0	0	0	0	0	0	0	0
Flare	0	0	0	0	0	0	0	0
non-transient (NT)	31	7	26	73	16	497	47	80
RR-Lyrae	0	1	2	3	0	12	53	3
SNe	8	5	22	6	4	29	0	82

Table 4: Confusion matrix for the Richards set using random forest classification of all eight types. The rows are the predicted types while the columns are the actual types.

Predicted	AGN	Blazar	CV	Downes	Flare	NT	RR-Lyrae	SNe
	Actual types							
AGN	31	3	0	2	0	2	0	2
Blazar	0	27	3	7	0	0	0	0
CV	2	4	93	26	0	4	2	14
CV Downes	1	2	15	58	0	7	5	0
Flare	0	0	0	3	8	0	0	0
non-transient (NT)	8	0	9	25	15	541	1	16
RR-Lyrae	0	1	0	7	0	0	95	0
SNe	2	2	25	7	1	4	0	160

Table 5: Confusion matrix for our measures using random forest classification of all eight types. The rows are the predicted types while the columns are the actual types.

Classification Performance

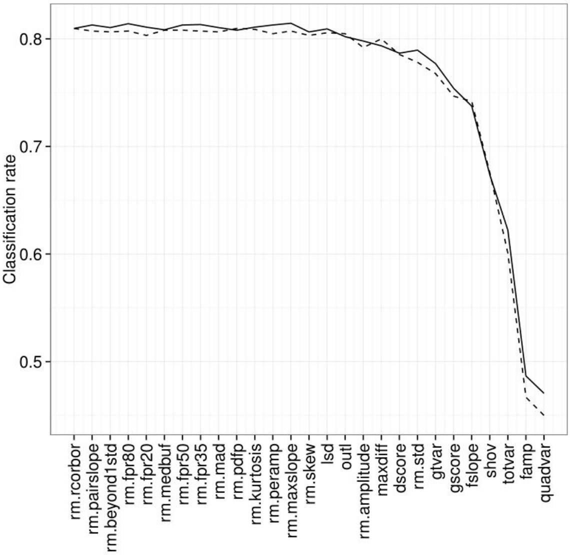
	LDA	TREE	SVM	NN	RF
Richards	60.8	64.0	72.6	71.1	73.1
Ours+Richards	74.6	67.4	78.0	77.2	79.0

Table 6: Percentage of the new set of five transient types correctly classified.

	LDA	TREE	SVM	NN	RF
Richards	75.5	79.4	85.3	76.8	85.5
Ours+Richards	97.5	89.2	98.4	98.3	96.8

Table 7: Percentage of 50,000 variables and non-variables correctly classified.

Variable Selection



Conclusion

Concluding Remarks

- The sample used heavily over-represents transients which constitute less than 1% of an unbiased sample.
- Nonetheless, the proposed measures perform particularly well in classifying into transient vs non-transient, halving the previous error rate.
- Particularly *famp*, *totvar*, and *fslope* are useful classification variables.