Modeling Light Curves for Improved Classification

Julian Faraway\textsuperscript{1}, Ashish Mahabal\textsuperscript{2}, Jiayang Sun\textsuperscript{3}, Xiaofeng Wang\textsuperscript{4}, Yi (Grace) Wang\textsuperscript{5}, and Lingsong Zhang\textsuperscript{6}

\textsuperscript{1}University of Bath  
\textsuperscript{2}California Institute of Technology  
\textsuperscript{3}Case Western Reserve University  
\textsuperscript{4}Cleveland Clinic Lerner Research Institute  
\textsuperscript{5}SAMSI and Duke University  
\textsuperscript{6}Purdue University

January 15, 2014

Abstract

Many synoptic surveys are observing large parts of the sky multiple times. The resulting lightcurves provide a wonderful window to the dynamic nature of the universe. However, there are many significant challenges in analyzing these light curves. These include heterogeneity of the data, irregularly sampled data, missing data, censored data, known but variable measurement errors, and most importantly, the need to classify in astronomical objects in real time using these imperfect light curves. We describe a modeling-based approach using Gaussian process regression for generating critical measures representing features for the classification of such lightcurves. We demonstrate that our approach performs better by comparing it with past methods. Finally, we provide future directions for use in sky-surveys that are getting even bigger by the day.

Keywords: Classification, Feature Selection, Gaussian Process Regression, Irregular Sampling, Missing Data.

1 Introduction

In the last few decades we have seen advances in imaging technology, and the storage, transfer and processing of data. As a result, astronomy has moved from taking static, sporadic snapshots of the sky to obtaining high-cadence, deep and large images, almost akin to making movies of the sky. This, in turn, has resulted in opening up the field of studying the dynamic nature of the universe, in particular, the cataloging of different types of objects, both within our Galaxy, and well beyond it all the way to the early Universe. Cataloging goes well beyond stamp-collecting, since it reveals the time scales over which various phenomena occur, directly relating to the physical processes behind the brightness changes in astronomical objects, and allowing us to connect the different families of objects in various ways. A bonus is also the ability to look for connections missing so far, as well as fringe members of different classes.

Much of characterizion or classification for cataloging is done, or at least begun, through the study of variability of objects. Most astronomical objects, be they stars or planets or galaxies, or any of their subclasses, vary in brightness either intrinsically through some physical process
such as explosion, merging or infall of matter, or through an extrinsic process such as eclipse or rotation. For a small fraction of objects the variation can happen over a fraction of a second to hundreds of days depending on the phenomenon. For a majority of objects, the changes are much slower and smaller as the objects evolve through the proverbial astronomical time-scales. We can observe large parts of the sky multiple times at different wavelengths, yet these observations are far from continuous, all-sky, or panchromatic. For each part of the sky, and in particular for each object in the part of the sky we image, we get a time-series of flux. While all objects vary to an extent, for a vast majority of objects, the variations are non-discernible during the rather sparse sequence (tens to hundreds of epochs) of short exposures (less than a minute) that we have, and over the time-scales over which observations occur (few years). That is precisely the reason, for instance, that when we glance at the night sky we do not find stars suddenly changing their brightness.

This leads to most astronomical objects seeming non-variable. When we can discern the variability, e.g. a periodic variation, or a stochastic variation, or even a single sudden jump in brightness, the object could then be called a variable. This functional definition would of course change based on many factors, such as, total interval of observation, type of phenomena involved etc. An extreme case of a variable object is a transient - the brightness of which varies by several standard deviations in a much shorter time, of the order of seconds to minutes. It is the study of these types of objects that has really become possible due to high-cadence wide-field surveys.

In order to understand and classify transients, it is important to understand variability at all levels, including mostly non-variable astronomical sources. Past attempts have included analyses for denser lightcurves from Kepler, as in Blomme et al. [2010], Ciardi et al. [2011] or using brighter objects as in Richards et al. [2011] and general frameworks based on such approaches as in Mahabal et al. [2011], Djorgovski et al. [2012]. It is important to design measures that can isolate specific classes but are also derivable based on the available cadence of observations. Our aim is to present new measures based on object lightcurves which help in better discriminating between variables and non-variables, and among the different transient types.

Here we use data from the Catalina Real-time Transient Survey (CRTS) (Drake et al. [2009]). CRTS is based on the Catalina Sky Survey (CSS) which has been designed to look for near-Earth asteroids. One way to look for asteroids is by looking for motion of the asteroids with the backdrop of mostly non-moving stars in the night sky. The cadence used for this is four images taken 10-minutes apart. Thus, the CRTS lightcurves have 4 points obtained within 30 minutes. The next such set could be the next night, the next week or even a month later. The sparse and non-uniform nature of the lightcurves presents classification challenges and also allows development of new statistical techniques. CRTS also includes data from the Mt. Lemon Survey (MLS) which covers mainly a narrow region of the sky near the ecliptic, and Siding Spring Survey (SSS) which covers the Southern hemisphere. We have not included data from MLS or SSS in the current study, but all methods are equally applicable to them as well. About 75% of the sky is covered by CRTS, with parts near the poles and near the plane of our galaxy excluded. Despite the relative sparsity of the CRTS lightcurves, a strength of the survey is its longevity - we have data where the epochs are spread over 10 years and hence there are parts of the sky with several hundred observations making CRTS one of the richest synoptic datasets. The techniques we develop will be applied to the 500 million lightcurves that CRTS now has and that in turn can lead to the discovery of newer classes as well as rarer counterparts of known classes of astronomical objects.

Our strategies in deriving these critical measures are 1) selecting a collection of relatively balanced, representative light curves of various types and scales (§2); 2) exploring features and signatures of these light curves (also in §2); 3) developing a Gaussian process regression model
Functionally transient

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

Table 1: Number of light curves for each transient type and non-transient objects

for the light curves with appropriate priors using astronomical information and an empirical Bayesian approach (§3); and then 4) deriving new measures representing critical features of the light curves using the posterior mean regression curve and residuals. These model-based measures complement existing measures. To examine the power of the new measures in classifying light curves in comparison to existing measures, five popular classification procedures are used in four schemes of classification problem in §4: Linear Discriminant Analysis, Decision Trees, Support Vector Machines, Neural Networks, and Random Forests. The results show that our measures perform better than the existing measures. A discussion for why our approach works better is given in §5. Although our modeling approach has been used for an astronomy application, the method could be valuable for other applications involving the classification of sparse, irregularly sampled time series with missing data.

2 Data

We have selected a moderate sample of light curves with which to illustrate our methods. The measures available in each lightcurve are Right Acension (RA) and Declination (Dec) which provide the position of the object on the sky, Epoch as Julian Date, magnitude (negative logarithm of flux), an error estimate on the magnitude. The total number of light curves considered is 3720. The selection is described below.

We started with just the transients detected by CRTS in real-time over about five years. These include Active Galactic Nuclei (AGNs), Blazars, Cataclysmic Variables (CV), Flare stars and Supernovae (SNe), representing five very different types of light curves (e.g. Djorgovski et al. [2011]). We also included a set of 15 random pointings and objects within 3' of those pointings. These objects are assumed to be non-transients because any transients in there would have been detected earlier. The transients tend to be fainter than typical objects (by definition - it is easier to catch objects that are not normally seen but brighten and become visible for a short duration). In order to offset that, two classes of brighter variable objects were included i.e. Cataclysmic Variables from the Downes set (Downes et al. [2005]) and RR Lyrae which are periodic variables with a period of ∼1 day. For the purposes of this article, we have one class called non-transient and seven classes which we call transients viz. AGN, Blazars, CV, Flares, SNe, CV Downes, and RR Lyrae. Note that among the labelled types we have considered here, only the RR Lyrae are periodic. There are methods for distinguishing periodic objects from non-periodic ones but these are not addressed in this article.

The number of observations for each light curve varied greatly with a minimum of 5 and a maximum of 641. The median length was 52. We excluded lightcurves with fewer than 5 observations as these cannot reasonably be classified. The earliest date for any of our lightcurves was 53464 Julian Day (JD) and the latest was 56228, i.e. 5th April 2005 to 28th October 2012. We have used 53464 as our zero-point and referred to all dates as number of days beyond this. Our set spans 2764 days.

An examination of the data is helpful in deciding which methods of analysis may be appropriate. In Figure 1 we see four examples of light curves. The objects are identified by their catalogue numbers for reference. The magnitude is the negative logarithm of flux so in keeping with standard practice, we plot the magnitudes on a reversed scale because smaller magnitudes
represent brighter objects.

Figure 1: Examples of four light curves: (i) CSS071216:110407-045134, an AGN (ii) CSS110405:141104+01115, a Supernova (iii) CSS111103:230309+40060, a Flare and (iv) 301904800767, a non-transient.

Understanding the pattern of measurement is crucial to proper modeling of these curves. The first example shows some gaps in an otherwise dense sequence of measurements. No observations were taken during these periods because the orbit of Earth precluded it. In the second example, there are no observations outside of a narrow range. Observations were attempted at other times but the object was too faint to be observed with a magnitude less than the current astronomical detection limit of around 20.5. In the third example, we are fortunate that the spike in brightness was observed as this occurs during a brief period of time. In the fourth example, there are quite long periods with no observations but it seems reasonable to assume that no substantial variations in magnitude occurred during these periods given the nearly constant values of magnitude.

It would be useful to know exactly when observations were attempted for given objects
while below detection limit. For the purposes of this analysis, we shall assume that all the objects may be surveyed throughout the period of the study but failures to observe have not been recorded. It will be clear how this information could be incorporated into our methods and that this would improve our results in §3.

3 Methods

The nature of the data and the requirements of object classification impose some constraints on what methods are practical. The problem could be viewed as one of functional data analysis (see Ramsay and Silverman [2005]). However, there are several obstacles to pursuing this approach. The observations on the light curves are very irregular, both in time and in number. There are methods for dealing with such data but there is a more serious obstacle in that there is little sense in which the curves can be registered or aligned. Excepting the rare case where objects are close in the sky and measurements are likely to be correlated due to atmospheric conditions, light curves are independent. This prevents us from using the “borrowing of strength” that registration would allow.

This leads us to another style of analysis based on sample summary statistics. Judgement is used to devise statistics that measure various features of the observed curves which we may believe important in distinguishing them. We prefer that these statistics be relatively simple so that they can be applied quickly and reliably for both short and long light curves.

About 20 measures are presented in Richards et al. [2011] that are mostly derived from previous articles. They found these various measures to be helpful in distinguishing objects. Since these measures have been widely tested, at least for brighter data, we use these as a baseline for our analysis. Our objective is to find additional measures that improve the classification accuracy beyond this set. For ease of reference, we will call this set the Richards measures. The specific measures we have used from Table 5 of Richards et al. [2011] are amplitude, beyond1std, fpr20, fpr35, fpr50, fpr80, maxslope, mad, medbuf, pairslope, peramp, pdfp, skew, kurtosis, std and rcorbor. We have no Quasars in our example data so we have not used measures designed to detect these. We also omitted the linear trend measure as this was only large for light curves with few observations so it becomes a substitute for a short curve measure. As it happens, including it would not make much difference to the results we present later. We coded these measures from the definitions in Richards et al. [2011].

Although the Richards measures encompass a wide variety of features, they do not use any concept of modeling the curves. The primary innovation of this paper is to use such modeling to generate additional measures. For light curve $i$, 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 $j = 1, \ldots, n_i$. 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. Furthermore, there is an error process $\epsilon_i(t)$ which means we observe only $y_{ij}$ for $j = 1, \ldots, n_i$. We assume

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

where $\epsilon_i(t)$ is a Gaussian process with mean zero and covariance function $\Sigma$.

We considered several methods for estimating $f$. But most such methods are unable to use the censoring of information caused by invisibility of objects that sometimes fall below the detection limit. Standard methods such as local polynomial regression or smoothing splines cannot easily incorporate this information. In contrast, Gaussian Process regression allows us to build in the censoring when necessary by setting a prior which incorporates this information as we will see in our example.
3.1 Gaussian Process Regression

See [Rasmussen and Williams 2006] for a general introduction. This method requires that we specify a prior for the Gaussian process: \( f(x) \sim GP(\psi(x), k(x, x')) \). We choose a particular form for the covariance:

\[
k(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2\ell^2} (x - x')^2\right) + \sigma_n^2 \delta(x - x') \tag{2}
\]

where \( \delta(x) \) is 0 when \( x \neq 0 \) and 1 when \( x = 0 \). One advantage of the Gaussian Process approach is that an explicit solution for the posterior is available that can be rapidly calculated. We will need to classify large numbers of future light curves as the measurements are collected so we need efficient methods.

There are four components of the prior which must be specified:

1. \( \sigma_f^2 \) is the signal variance. A very large fraction of objects to be classified in the future will be non-transients. These non-transients vary in signal but not very much. For this reason, we set \( \sigma_f^2 \) to the median observed variance in the non-transients.

2. \( \sigma_n^2 \) is the noise variance. Although it is uncommon in other applications, for astronomical data we are often able to estimate the measurement error. In this example, the measurement error varies a little from case to case. For simplicity, we take the mean observed value of the measurement variance for \( \sigma_n^2 \).

3. \( \ell \) is sometimes called the length-scale. It controls the amount of correlation and therefore the amount of smoothness in the resulting posterior fit. We use a value of 140 days as seen in Figure 2. This choice is based on a subjective assessment on how much smoothness should be expected in these curves. The classification performance is not very sensitive to this choice.

4. \( \psi(x) \) is the prior mean. This choice is problematic and requires further discussion below.

   We take an empirical Bayesian approach.

   We illustrate the issues in setting the prior mean in Figure 2. What values are expected for the curve in regions where there are no measurements? One answer is that we might expect about the same magnitude as that seen elsewhere for this object. This suggest setting the prior to the median magnitude for the object. This choice can be seen in the solid line fits in Figure 2. This works well enough in three cases but fails for the supernova example because we do not expect this curve to follow a similar magnitude at other times. If it did, it would have been seen. So an alternative approach is to set the prior to the detection limit at a magnitude of 20.5. This gives the dashed line fits as seen in Figure 2. This works well for the supernova case but is problematic for the other three curves. In regions of no measurement, the fitted curve is drawn down towards the detection limit. We can counteract this by increasing the length-scale (i.e. increasing the smoothness in the prior) but tends to attenuate real effects and still does not work well for relatively sparsely measured curves (such as the non-variable in this example).

   Our solution is to use an adaptive prior. When there is less than one year of observations, we use the detection limit, otherwise we use the median magnitude. The choice of a year is large enough that sparse but widely measured curves such as the fourth example do not use the detection limit. But the choice is small enough that the detection limit is used in cases like the third example. Now using the data to select the prior may make our method at least partly empirical Bayes rather than pure Bayes, but we need to judge the method by its classification performance which is improved by this choice.

   More specific information about times when observations were attempted at a location but nothing was observed would allow us to refine this prior.
We also compute the scaled residuals from the fit and use the maximum in absolute value, called $outl$, as a measure.

### 3.2 Curve Measures

Given the posterior mean $\hat{f}$, we can compute fitted curve measures from $\hat{f}_i$ for curve $i$ computed on an evenly spaced grid of values on the range of observation $u_j$ for $j = 1, \ldots, m = 300$. We have chosen the following, where the italicized word is the name of the variable for future reference:

- **totvar** total variation: $\sum_j |\hat{f}_i(u_{i,j+1}) - \hat{f}_i(u_{i,j})|/m$
- **quadvar** quadratic variation: $\sum_j (\hat{f}_i(u_{i,j+1}) - \hat{f}_i(u_{i,j}))^2/m$
- **famp** amplitude of fitted function: $\max_t |\hat{f}_i|$ 
- **fslope** maximum derivative in the fitted curve: $\max_t |\hat{f}_i'|$

We also compute the scaled residuals from the fit and use the maximum in absolute value, called $outl$, as a measure.
Another feature of this data is the clustering of times of measurement which can occur in
groups of up to 4 where observations are spaced by ten minutes within a thirty minute period.
The Gaussian process regression is not able to model the variation at this finer scale because
setting the length-scale \( l \) to a much smaller value would result in too rough a fit overall. We
need another set of measures to capture the features at this scale of measurement.

We compute the mean within each of these groups as \( \tilde{f}_{ij} \) and then compute the following measures:

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

The last measure is motivated by scoring methods used to judge prediction performance.

There are also some gaps within the Richards measures set of sample curve summary mea-
sures. We add the following:

- **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 mag-
nitude for curve \( i \) and \( s_{ij} \) is the observed measurement error at \( t_{ij} \).

There are measures that may be informative for the current data we are analyzing but may
not have predictive value in future examples. We have avoided using such measures. They fall
into three categories:

1. Measures based on the number of observations in a light curve. Some phenomena, such as
  supernovae, are not recurrent and subsequent observations may fall below the observable
  limit. Light curves in such cases can be quite short but this is known only in retrospect
  so this is not usefully predictive. The number of observations does have some impact on
  the choice of prior and in scaling some of the measures, but we refrain from using this
  number (or anything closely related) as a direct measure for classification purposes.

2. The classification of an object should be invariant to the addition of a constant to the
  observed magnitudes. But some biases in the way that our example data was extracted
  would cause, say, the mean magnitude, to be an effective discriminator among the types.
  This mean magnitude will not be reproducible in future samples so we do not use this
  measure or anything related to it.

3. Location in the sky. The method of constructing our example dataset would mean that
  location would become useful discriminator. As it happens, location would provide some
  usable information for classification as some types of objects are more likely to appear in
  some locations than others but we must refrain from using this information here.

There is additional information such as the nearest radio source or the nearest galaxy which
could also be useful in classification but we do not use here. We experimented with a larger set
of additional measures but we have presented only those that appear to have some additional value for classification.

Given this set of measures, we can use any number of classification methods to distinguish objects using light curves. We demonstrate the use of our measures using five popular classification methods. We will generally use the default choice of options for the particular implementation in $\textit{R}$. Our objective is to show that our measures represent an improvement over using the Richards measures alone. It is likely that the classification methods could be better tuned to obtain a better result or that the reader may favor another classification method. But that is not the point of this article. We are not trying to claim one classification method is better than another, just that our measures are better.

The methods we have used are:

- **LDA** Standard linear discriminant analysis method as implemented in $\textit{R}$ by Venables and Ripley [2002].
- **TREE** Recursive partitioning as implemented in the $\textit{rpart}$ package of $\textit{R}$ by Therneau et al. [2013].
- **SVM** Support Vector Machines as implemented in the $\textit{kernlab}$ package of $\textit{R}$ by Karatzoglou et al. [2004].
- **NN** Neural network as implemented in the $\textit{nnet}$ package of $\textit{R}$ by Venables and Ripley [2002].
- **RF** Random forest ensemble package by Liaw and Wiener [2002].

We log-transformed several of the measures to reduce extreme skewness in order to improve classification performance. The same transformations were used in all the comparisons below. Without these transformations, both sets of measures would perform less well in general for methods LDA, SVM and NN. The partitioning-based methods, TREE and RF, are invariant to monotone transformations. Explicit details of the implementation may be found in the Appendix.

4 Results

Classification methods usually do not perform as well as expected when applied to new data. When the same data are used to both fit and evaluate a method, the classification rate is inflated. To avoid this problem, we randomly split the data into 2/3 for training i.e. used to develop the classification rule and 1/3 for testing, that is to evaluate how well the rule performs. Since we are only interested in the relative performance of the classification measures and methods and because the sample size is relatively large, there is little advantage in considering more than one random split.

4.1 Classification Performance

We considered four different types of classification problem with bold labels used for future reference:

- **All** The overall problem of classifying eight types — the non-transients and the seven transient types.
- **Transient or not** Perhaps the first step in any light curve classification process will be to determine which objects are transient.
### Table 2: Percentage correctly classified using the Richards measures.

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

| Method               | 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|

Transient only Having separated out the non-transients, the next step might be to identify the type of the transient. For this problem, we delete the non-transients from both the test and training data.

Heirarchical An alternative approach to classifying all objects directly is to first classify objects into transient or not transient, then if transient to classify among the seven available types.

We show the percentages correctly classified using the Richards measures in Table 2 and using our measures (which incorporate the Richards set) in Table 3.

The standard error for the classification rate is just less than 1% which is helpful in judging which differences are notable in these tables. Our measures provide a significant improvement to the Richards measures alone which might be regarded as the previous state of the art. Of course, adding additional measures can only improve the fit of a model, but we are using an independent test set so we can be sure the improvement is more than illusory. There is little to distinguish the heirarchical approach from the one-step method although we would recommend the heirarchical approach on a new and unbiased sample of light curves.

Our sample heavily over-represents transients which would constitute less than 1% of an unbiased sample. Hence, even a null method which classified randomly based on prior proportions would achieve around 99% accuracy. Certainly any sensible method will do even better than this and it would take a very large unbiased sample to distinguish different methods. This explains why we have used a more balanced, although biased, representation of the eight types. Similar strategies are used in case-control studies.

Because these classification methods will be applied to very large numbers of objects, even quite low error rates will result in large numbers of misclassified objects resulting in wasted resources or missed opportunities. For this reason, the primary classification into transient against non-transient is particularly important. We can see our proposed measures perform well in this respect, halving the previous error rate, although there remains further room for improvement.

### 4.2 Feature Selection

The random forest method provides a means of determining the worth of predictors by measuring how much the Gini index decreases when a predictor is removed. We can use this mechanism to sequentially remove predictors based on the least decrease. See also Donalek et al. 2013.
At each stage, we refit the model and recompute the importance of the predictors. The results of this analysis are shown in Figure 3 for the problem of classifying all 8 types.

There is little difference in classification accuracy between the training and test datasets which is a good indication that we are not over fitting. We see that this selection process removes most of the older non-model based measures without any noticeable loss in classification accuracy. The fitted curve measures quadvar, famp, totvar and fslope are among the most useful classification variables. Hence we can see that deriving measures based on our model is a good place to start and not merely a way to supplement existing older measures. The message from the other three classification problems is similar.

Figure 3: Stepwise selection of features. The classification rates (solid for the training set and dashed for test set) are shown after the named predictor is removed from the model. Features from the previous set have the prefix \textit{rm}.

5 Discussion

We have presented two advances. In Statistics, we have shown how Gaussian Process regression can be adapted and the corresponding priors developed to deal with data of varying sampling
density, structures and scales. We are able to deconvolute the underlying curves \( f_i \) and measurement noise \( \epsilon_i \) to distinguish the objects rather than use summary statistics that use samples \( y_{ij} \) with mixed up noise and signal. With some further effort we might show that the measures based on our estimated curves, under some regularity conditions, would be consistent for the features of the true underlying curves. A summary statistic-based approach can be biased or inconsistent for these features. In Astronomy, we have demonstrated a new method of generating measures representing features of light curves that are significantly better in classifying objects than previous methods.

There is further scope for improvement in performance by optimising the classification using routine methods. With more detailed information about when locations were surveyed but no object observed, we can further refine our priors to obtain superior results. The measures we have developed are now being used for several purposes. We can apply the method for new data where the location has only been surveyed for a shorter period of time. The measures can all be scaled appropriately. We have experimented by taking time-wise subsets of this data and have found that although the absolute performance drops with shorter curves, the relative performance over the older set of measures remains. Furthermore, the measures provide the means to detect objects of unknown type. By adding a richer and more powerful set of measures, we have increased the potential for such interesting discoveries. Some of our measures have already been used in classifying new light curves.

**Acknowledgements**

This work is one of results from the Imaging Working Group at the *SAMSI's 2012-13 Program on Statistical and Computational Methodology for Massive Datasets*. We thank SAMSI for bringing us together and for their financial support. The CSS survey is funded by the National Aeronautics and Space Administration under Grant No. NNG05GF22G issued through the Science Mission Directorate Near-Earth Objects Observations Program. The CRTS survey is supported by the U.S. National Science Foundation under grants AST-0909182 and AST-1313422. We are also thankful to the Keck Institute of Space Studies, and part of the work was supported through the Classification grant, IIS-1118041. We thank SG Djorgovski for useful comments.

**Appendix**

Our data, code and detailed results are available as a supplement to be found at [people.bath.ac.uk/jjf23/modlc](people.bath.ac.uk/jjf23/modlc)

**References**

J. Blomme et al. Automated Classification of Variable Stars in the Asteroseismology Program of the Kepler Space Mission. *The Astrophysical Journal*, 713, 2010.

D. Ciardi et al. Characterizing the Variability of Stars with Early-release Kepler Data. *ApJ*, 141, 2011.

S. Djorgovski, A. Drake, A. Mahabal, M. Graham, C. Donalek, et al. The Catalina Real-Time Transient Survey (CRTS). *arXiv:1102.5004*, 2011.

S. G. Djorgovski et al. Flashes in a star stream: Automated classification of astronomical transient events. *arXiv:1209.1681*, 2012.
