Real-Time Anomaly Detection of Short Time-Scale GWAC Survey Light Curves

Tianzhi Feng*†, Zhihui Du*, Yankui Sun*, Jianyan Wei‡, Jing Bi†, and Jason Liu§

*Department of Computer Science, Tsinghua University, Beijing, China

Email: duzh@tsinghua.edu.cn

†Department of Information, Beijing University of Technology, Beijing, China

‡National Astronomical Observatories of China, Beijing, China

§School of Computing and Information Sciences, Florida International University, Miami, USA

***Abstract*—Ground-based Wide-Angle Camera array (GWAC) is a short time-scale survey telescope that can take images covering a field of view of over $5,000$ square degrees every 15 seconds or even shorter. One scientific missions of GWAC is to accurately and quickly detect anomaly astronomical events. For that, a huge amount of data must be handled in real time. In this paper, we propose a new time series analysis model, called DARIMA (or Dynamic Auto-Regressive Integrated Moving Average), to identify the anomaly events that occur in light curves obtained from GWAC as early as possible with high degree of confidence. A major advantage of DARIMA is that it can dynamically adjust its model parameters during the real-time processing of the time series data. We identify the anomaly points based on the weighted prediction result of different time windows to improve accuracy. Experimental results using real survey data show that the DARIMA model can identify the first anomaly point for all light curves. We also evaluate our model with simulated anomaly events of various types embedded in the real time series data. The DARIMA model is able to generate the early warning triggers for all of them. The results from the experiments demonstrate that the proposed DARIMA model is a promising method for real-time anomaly detection of short time-scale GWAC light curves.**

***Index Terms*—Light curve; ARIMA; real-time analysis; big data processing; anomaly detection**

I. Introduction

Aided by modern instruments, today's astronomy is capable of collecting a large amount of high-resolution data. This large amount of data often needs to be processed and analyzed in real time, for example, for online classification and anomaly detection. The sheer volume of the astronomical data and the speed at which the data needs to be analyzed can easily stretch the limitations of today's computing and storage capabilities.

A case in point is the time domain astronomy [1], where large datasets are generated by sky surveys and represented as time series data. We use the time-series data generated by the ground-based wide angle camera arrays in the Space-based multi-band astronomical Variable Objects Monitor (SVOM) mission as an example. SVOM is a joint space mission between the Chinese National Space Agency (CNSA), the Chinese Academy of Science (CAS), and the French Space Agency (CNES). It is aimed at studying Gamma-Ray Bursts (GRBs) in the next decade [2]. The mission is expected to launch in 2021 and will consist of a medium-size satellite, which carry instruments for detecting and localizing GRBs and measuring GRB afterglows (including two wide field high-energy instruments and two narrow field telescopes), together with a ground segment that consists of a ground-based wide angle optical camera array and two follow-up telescopes.

The Ground-based Wide Angle optical Camera array (GWAC) in SVOM will be used to survey a large field of view for optical transients, before, during and after GRBs. The system consists of 36 wide angle cameras, each with 18 cm in diameter, 22 cm focal length, and $4k \times 4k$ CCD detectors, that are sensitive in the 400-800 nm wavelength range. Altogether, the camera array can take images capable of covering a field of view of over $5,000$ square degrees.

GWAC can produce light curves with high time resolution of millions of objects [3]–[5]. A light curve is a time series of light intensity of a celestial object or region. Many astronomical objects exhibit brightness variability due to different physical processes. The light curve therefore can be used in classification of variable objects, such as variable stars and eclipsing binary. To generate light curves, one needs to process the images in a pipeline that includes steps, such as source extraction, flux calibration, source association.

Once light curves are generated, they can be analyzed for online classification and anomaly detection [6]–[8]. For example, the light curves can be used to detect special astronomical phenomena caused by abnormal brightness warning from transient phenomena, such as short time-scale gravitational microlensing and transits by extrasolar planets. The GWAC camera array can take an image once every 15 seconds (including 10 seconds for exposure and 5 seconds for read-out). Consequently, it can generate a large amount of data, at a rate of 85 MB/s and having more than 6 million sources, which needs to be processed for online anomaly detection.

GWAC is the first short time-scale survey telescope in the world. Existing time-series analysis methods for handling real-time data are limited, given the amount of data processing demand as well as the specific properties of astronomical light curves for anomaly detection. The contribution of this paper can be summarized as follows: (1) We propose an improved time-series analysis method, called DARIMA, which can dynamically adjust its model parameters during real-time processing; (2) We propose a method for anomaly detection of special astronomical events using data from different time windows to calculate the weighted value for identifying anomaly points; and (3) We evaluate the DARIMA model with real light curves and also simulated anomaly events embedded in light curves to demonstrate that our method can quickly and correctly identify all the anomaly points. The proposed DARIMA model is a promising method for real-time anomaly detection of short time-scale GWAC light curves.

The rest of the paper is organized as follows. In Section II we provide the background information. We present the detail of the proposed dynamic auto-regressive integrated moving average model for real-time time-series analysis in Section III. We evaluate the effectiveness of the our model and present the results in Section IV. Finally we conclude the paper in Section V.

II. Background

Astronomy is immensely rich with data. Observations, such as sky surveys, can generate and archive enormous quantities of data, which may reach tens or hundreds of terabytes, with billions of detected sources, each with hundreds of measured attributes. Data mining is the technique used for converting the observed data into useful information, which can then be interpreted with theory or hypothesis, and used for knowledge discovery.

A. *Time-Series Data Mining Methods for Astronomy*

Many tasks have been considered for mining time series data, including indexing (i.e., query by content), clustering, classification, prediction (forecast), summarization, anomaly detection, and segmentation [9]. Prediction, in particular, aims at forecasting future values from past and current data based on historical trends and statistics. In comparison, anomaly detection is to identify previously unknown patterns in data that deviate from "normal" observations. Many techniques can be used for time series forecast and anomaly detection. Some of the major methods are discussed below:

- Support Vector Machine (SVM) is a supervised learning algorithm used for classification and regression analysis [10]. It has been shown to be effective for pattern recognition of non-linear and high-dimension data with a small sample size.
- Random Forest is an ensemble learning algorithm, which constructs multiple decision trees during training and uses the mode or mean output for classification and regression analysis [11], [12]. The method runs efficiently on large databases and is effective in estimating missing data and balancing errors from unbalanced datasets. It is also useful for feature selection in addition to being an effective classifier. As such, it can play an important role as a part of general ensemble methods that can combine the predictions from several base estimators and improve the robustness of a single estimator.
- Artificial Neural Network (ANN) is the most widely used machine learning algorithm [13], [14]. The algorithm constructs a series of interconnected nodes organized in layers with weighted connections. Each node is associated with an activation function (such as sigmoid with a simple threshold). The weights of the connections are adjusted by the training algorithm. Back propagation is one of the best known training algorithms. ANN is very effective at estimating non-linear relationships within the data, and has been used widely in pattern recognition, voice recognition, intelligent control, and nonlinear optimization, and has recently found significance in deep learning.
- Auto-Regressive Moving Average (ARMA) model is an import statistical method for specifically analyzing time-series data and for predicting future values. The ARMA model considers the time series as a stochastic process which can be described as two polynomials, one for the auto-regression (AR) and the other for the moving average (MA). The AR component is the linear combination of observable values while the MA component is the linear combination of the unobservable white noise disturbance terms. The ARMA model assumes that the time series is a stationary stochastic process. If the time series is not stationary, a technique called differencing can be applied to transform the non-stationary time series to a stationary one. The combined method is called ARIMA (Auto-Regressive Integrated Moving Average).

All the data mining methods mentioned above can be applied for time-series data. We develop our method based on ARIMA for processing the astronomical light curves. We extend the ARIMA model for real-time prediction and for anomaly detection. We first discuss the ARIMA method in more details in the next section.

The light curve analysis focuses on examining the photometrically obtained light intensity of a celestial object, which we call the light source. Light curve analysis is very common in astronomy. For example, Harikrishnan et al. [6] studied the black hole system GRS 1915+105 using nonlinear time series analysis of the light curves, including the correlation dimension, the correlation entropy, singular value decomposition, and the multifractal spectrum. Tarnopolski [7] attempted to estimate the maximal Lyapunov Exponent (mLE) from light curves of Hyperion, Saturns seventh moon. McWhirter et al. [8] proposed a data processing framework designed to utilize multiple intelligent agents that can be distributed across multiple machines. The intelligent agents conduct time-domain analysis of the time-series light curves of astronomical objects for automated classification. In this paper, we propose a method for analyzing light curves for fast and accurate anomaly detection and early warning so that more computationally intensive processing can follow up and be applied for specific objects at specific time.

B. *ARIMA*

Auto-Regressive Integrated Moving Average (ARIMA) models is a forecasting technique that projects the future values of a series based entirely on the observed values in the past. The main application of the ARIMA model is in short-term forecasting and can generally work well when the time-series data exhibits a consistent pattern over time with a minimum amount of outliers. ARIMA is also called the Box-Jenkins approach, named after its original authors George P. Box and Gwilym Jenkins [15].

The basic idea behind ARIMA is to consider the time series as a stochastic process. ARIMA extends the Auto-Regressive Moving Average (ARMA) model, discussed in the previous section, by applying the differencing method to the original data series. The first step in applying ARIMA is to check for stationarity. A random variable that is a time series is stationary if its statistical properties are constant over time. In other words, a stationary stochastic process has no trend, and its variations have a constant amplitude. The short-term random time patterns remain the same in a statistical sense. That is, its autocorrelations remain constant over time.

If the time series does not seem to be stationarity, we can transform a non-stationary series to a stationary one by applying "differencing", that is, by subtracting the observation in the current period from the previous one. Applying this transformation once ("first differencing") can eliminate the trend of the data if the series is growing at a fairly constant rate. Applying this transformation the second time ("second differencing") if the series is growing at an increasing or decreasing rate. Such transformation can be applied more times if so necessary.

After stationarization, ARIMA can then describe the movements in the derived stationary time series using ARMA. Consider a time series data X_t where t is an integer index and the X_t's are real numbers. The derived time series Y_t are the differenced values of the original series X_t. The parameter d is the degree of differencing needed for stationarity. If $d = 0$, $Y_t = X_t$; if $d = 1$, $Y_t = X_t - X_{t-1}$; if $d = 2$, $Y_t = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = X_t - 2X_{t-1} + X_{t-2}$; and so on.

The ARIMA forecasting equation for the "stationarized" time series Y_t is a linear equation in which the predictors consist of lags of the dependent variable and lags of the forecast errors. More formally, the general forecasting equation can be described as follows:

$$\hat{Y}_t = \sum_{i=1}^{p} \beta_i Y_{t-i} + Z_t \tag{1}$$

And the error term Z_t can be described as follows:

$$Z_t = \epsilon_t + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j} \tag{2}$$

where $\beta_1, \beta_2, \cdots, \beta_p$ are auto-regressive parameters, $\alpha_1, \alpha_2, \cdots, \alpha_q$ are moving average parameters, and ϵ_t's are independent identically distributed error terms with zero mean.

This model is called ARIMA(p, d, q) model of X_t, where p, q, and d are non-negative integers: p is the order of the autoregressive (AR) model, or the number of lagged values of Y_t; q is the order of the moving-average (MA) model, or the number of lagged values of the error term; and d is the degree of differencing, i.e., the number of times the data have had past values subtracted.

The parameters of the ARIMA model can be determined empirically. To determine d, one needs to test the stationarity of the time series. If the original series is stationary, we set $d = 0$. Or, we increase d until stationarity is satisfied. To determine stationarity, we can first simply visualize the time series to identify the existence of possible patterns, trends, cycles and seasonality in the data. We can also test the stationarity using the Dickey-Fuller test. Stationary testing and converting a series into a stationary series are the most important steps using the ARIMA model.

There is a systematic approach for determining the values of p and q used in the equation for predicting the stationarized series Y_t. An easy approach is based on inspecting the plots of the autocorrelations and partial autocorrelations of the series. One can inspect ACF and PACF plots and look for a clear AR or MA signature. Several well-known formal criteria have also been developed for the order selection, which include the Akaike's Information Criterion (AIC) [16], the Bayesian Information Criterion (BIC) [17], and the Hannan-Quinn Criterion (HQC) [18]. These criteria are alternative ways for order selection. They introduce a penalty term for the number of parameters being estimated in the model. In general, more parameters would suggest more complicated models. The penalty term is in place so that, for a given level of "fit", a more parsimonious model is preferred over a more complex model. Changing the form of the penalty term gives rise to different information criteria.

III. DARIMA

ARIMA is a powerful forecasting model for time series data and can be especially effective for analyzing astronomical data. In this paper, we make changes of this forecasting model to improve its effectiveness for real-time analysis and anomaly detection.

Our goal is to use the improved model to analyze the light curves generated from GWAC once it is fully implemented. To illustrate the effectiveness of our approach, we use real dataset, called mini-GWAC, which is obtained from the National Astronomical Observatories of China. In particular, we focus on a subset of the data containing three days of observations of 978 astronomical objects. In this paper, we illustrate the results of an object with the catalog ID (catID) of 1114 as an example.

An important step in the traditional ARIMA model is to determine the parameters. However, once these parameters are determines, they cannot be adjusted dynamically. ARIMA uses the differencing technique to remove the obvious patterns, trends, cycles and seasonality in the data. The result time series data is supposed to be stationary—the statistical properties of the time series (e.g., autocorrelations) are expected to remain constant over time. This assumption, however, may not be always true if we deal with real-time analysis of data for forecasting and anomaly detection. In this case, the model needs to adapt to the possible shifts in the dataset over time.

To deal with this problem, we propose an improved model, which we call Dynamic Auto-Regressive Integrated Moving

Average (DARIMA). The model compares the Bayesian Information Criterion (BIC) value of different parameters, p and q, and uses the most appropriate ones in forecast model during the next step. We compute the parameters continuously every time when the data is updated so that the model is able to keep track of the changes in the time series data.

GWAC can produce an image every 15 seconds; the image is fed to a pipeline for preprocessing, quality control, source extraction, flux calibration, source association, and eventually light-curve creation. A light curve is a time series of light intensity of a celestial object or region. The dynamic ARIMA model is applied to analyze this time series in real time in order to detect sudden changes in the light source, which can signify special astronomical phenomena, such as gravitational microlensing or transits of extrasolar planets. Anomaly detection in the light curves serves as a precursor to more intensive data processing needed for these special astronomical events.

In the remainder of this section, we describe the important steps of the DARIMA time-series analysis of the GWAC light curves for anomaly detection.

A. Stationarity Test and Order Determination

There are two ways to check the stationarity of the time series data: one way is through visual inspection of the relevant plots, and the other way is via formal statistical tests. The former method is simple and straightforward and however is subjective to interpretation. The latter method is comparatively more complicated and yet can provide a more objective and statistically definitive answer.

The easiest method for checking stationarity is to visualize the time series plot. Non-stationary series typically shows patterns, trends, or seasonality in the time series plot. The mean and variance of a stationary series should be constant, as opposed to changes that depend on time.

Another visual method is to inspect the correlations of the time series using the ACF and PACF plots. The Auto-Correlation Function (ACF) plot shows the autocorrelation coefficients at different lags. The autocorrelation of the time series X_t at lag k is the correlation between the time series and itself lagged by k periods, i.e., it is the correlation between X_t and X_{t-k}. The Partial Auto-Correlation Function (PACF) plot shows the partial autocorrelation for different lags. The partial autocorrelation of X_t at lag k is the coefficient of lag k in a regression of autocorrelation of X_t at lag 1, lag 2, ..., and up to lag k. One way to interpret the partial autocorrelation at lag k is that it is the amount of correlation at lag k not explained by lower-order autocorrelations. A stationary series should only show short-term correlations (with small lags). The correlation decreases rapidly for larger lags, although it may wiggle around zero randomly. The correlation for a non-stationary series may also decrease but at a much slower rate.

We show an example light curve in Fig. 1. From the time series plot, we see that there are no obvious trends or cycles. Fig. 2 shows the autocorrelation function. The coefficients drop sharply at small lags and remain small around zero as the lag increases. It appears that the time series is stationary.

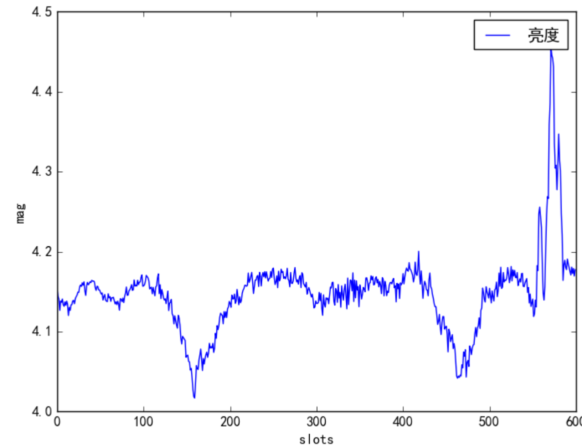

Fig. 1. Time series data of a light curve.

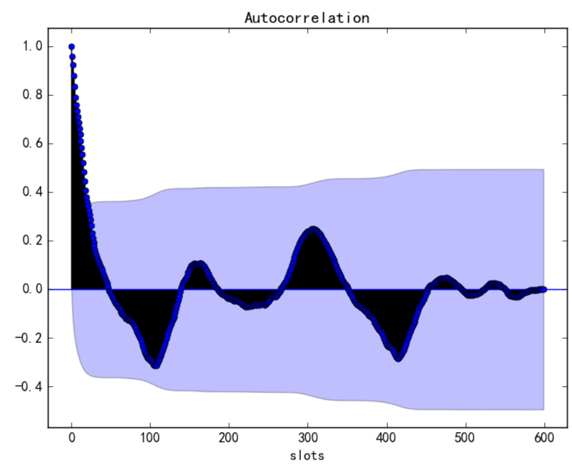

Fig. 2. Autocorrelation of the original time series.

One way to determine more objectively the stationarity of a time series is to use a unit root test. These are statistical hypothesis tests designed to check the stationarity and determine whether differencing is necessary. There are a number of unit root tests available and the Augmented Dickey-Fuller (ADF) test is the most popular test [19]. The null-hypothesis for an ADF test is that the time series is non-stationary. Large p-values are indicative of non-stationarity, and small p-values suggest stationarity. If 5% is the threshold, differencing is required if the p-value is greater than 0.05.

We apply ADF unit root test for the mini-GWAC data. The following shows the results:

```
adf:            -2.9258175081947511
pvalue:         0.04242709158188341
usedlag:        18
nobs:           581
critical values (10%): 2.5694261699959413
critical values (5%): -2.8665274458710064
critical values (1%): -3.4416553818946145
```

We observe that the ADF value is less than 5% and the p-value is less than 0.05. That is, the data is stationary. Furthermore, we test the white noise errors. The results is 9.04803104e-123. The p-value is much smaller than 0.05; the data is not white noise. Therefore we can conclude that the time series is mostly stationary with white noise errors, and therefore can be analyzed using the ARIMA model. Note that since the original time series of the light curves data is stationary, there is no need to apply differencing. In our analysis, we set $d = 0$.

From the ACF and PACF plots, we can also determine the other ARIMA parameters. There are explicit rules for determining the parameters p and q by looking for specific AR or MA signatures in the plot of autocorrelations and partial autocorrelations as a function of the lag. For example, we choose the AR(p) model, if the PACF plot cuts off after p lags but the ACF plot decreases gradually; we choose MA(q) model, if the PACF plot decreases gradually but the ACF plot cuts off after q lags. If both ACF and PACF plots tail off, we can choose different combinations of p and q such as either Akaike's Information Criterion (AIC) [16] or the Bayesian Information Criterion (BIC) [17] has the lowest value.

In our study we use BIC. BIC is a criterion for model selection, and in our case, for measuring the efficiency of the parameterized model in terms of predicting the data. BIC is based in part on the likelihood function. When fitting models, it is possible to increase the likelihood by adding parameters, but in doing so we may cause over-fitting. BIC has a penalty term that penalizes the complexity of the model, where complexity refers to the number of parameters in the model. We choose p and q so the BIC value is minimized.

B. *Window Size Selection*

In this paper, we use ARIMA to analyze light curves for forecasting and anomaly detection so as to provide early warning of special astronomical phenomena. It is therefore important to ensure accurate forecast and early warning. Selecting proper window size may affect the accuracy of the forecast and detection. In this section, we compare the effect of typical window sizes on the accuracy of the prediction. A window here suggests the length of the data we use for predicting the next data.

Table I shows the forecast errors—the min, the max, the mean and the variance, for three window sizes: 10, 50, and 100. From the table, we can see that the window size 50 is better than others. Therefore we use this window to predict the data.

C. *DARIMA for Early Warning Anomaly Detection*

Once we have determined the parameters for the ARIMA model, we apply the dynamic ARIMA method for early warning anomaly detection. Our goal is to detect short time-scale astronomical anomalies, such as gravitational microlensing, and provide early warning so as to introduce further processing necessary to analyze the specific astronomical anomalies.

Gravitational microlensing is an astronomical phenomenon caused by gravitational lens effect [20]. According to Einstein's general theory of relativity, a massive object can bend the light of a bright background object (a distant star or quasar) due to its gravitational field, and, as a result, can generate a distorted, magnified, and brightened image of the background source. Since microlensing observations do not rely on radiation received from the foreground object, it allows astronomers to study objects regardless of their light emission. It is thus an ideal technique to study faint or dark objects, such as brown dwarfs, red dwarfs, planets, and black holes.

Compared to the galaxy-scale gravitational lens effect (for galaxies or galaxy clusters), in which case the displacement of light by the lens can be resolved using high-resolution telescopes, such as the Hubble Space Telescope, with microlensing, the foreground objects (a planet or a star) make up only a minor portion of the mass for the displacement of light to be observed easily.

A microlensing event is also a transient phenomenon (in seconds to years in human time-scale as opposed to millions of years), the duration of which depends on the mass of the foreground object as well as on the relative proper motion between the background and background objects. A microlensing event can be detected in the sudden rise and fall of the source brightness in the light curve.

In our study, we focus on anomaly detection and early warning of gravitational microlensing. Every time we apply the prediction model, we make a prediction of values for 5 times after the current time. It can make the predicted value more accurate. They are stored and compared with the real light curve data that arrives continuously. We calculate the average prediction error and compare it with the maximum prediction error during the previous time window in order to determine whether significant deviation has occurred and if so, we treat it as an astronomical anomaly.

IV. Experiment Results

We apply our DARIMA model on the real time series data from the mini-GWAC dataset. Fig. 5 shows the light curve from a particular light source (catID=1114) on October 13, 2015, which contains two potential anomalies, one around the time slot 570 and the other around 790.

The result of our anomaly detection using DARIMA is shown in Fig. 6. The blue dots show the original time series and the red dots indicate the triggered warnings. Our model was able to detect the anomaly correctly.

The mini-GWAC data contains real anomalies and yet is limited in the number of occurrences and their diversity. We would like to test the robustness of our approach under various situations. To do that, we use the original time series from mini-GWAC and artificially insert four different types of functions—a sine curve, a tangent curve, a line, and an impulse—into the original time series. Fig. 3 shows an example of the artificially inserted anomalies. This method is particularly useful for generating different anomaly situations. It can also be considered reasonably realistic as it is based on

TABLE I
THE EFFECT OF DIFFERENT WINDOW SIZES ON PREDICTION ERROR

Window Size	Max	Min	Mean	Variance
10	2.263295984	7.1222617e-05	0.0698544648116	0.0397617623294
50	2.395200678	1.3391953e-05	0.0636541312058	0.0365042456634
100	2.440352563	1.4721013e-05	0.0668715443695	0.0407274415606

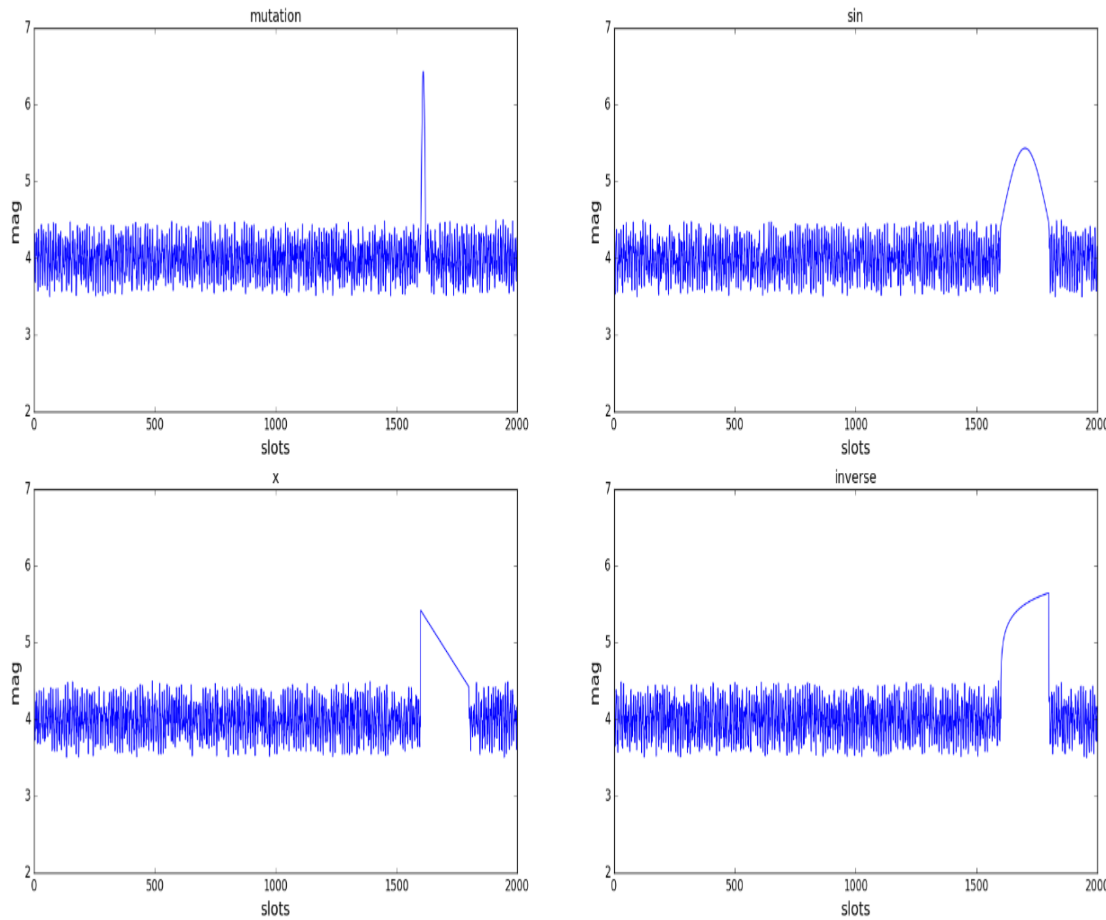

Fig. 3. Mini-GWAC time series with simulated anomalies.

the observation that, for the astronomy data we are interested in, the obtained light curves are mostly stationary, especially for invariable stars.

The results of the anomaly detection using our dynamic ARIMA model are shown in Fig. 4. Again, the blue dots are the original time series and the red dots are the triggered warnings. Our model was able to correctly detect the on-ramp of the anomalies. This would be sufficient for the purpose of early warning. The forecast also generated a few false positives. In this situation, it is possible that one can apply astronomy domain knowledges and use additional default thresholds to filter out or reduce these occurrences according to the particular situation.

V. CONCLUSION

This paper proposed a novel real-time anomaly detection model to analyze and forecast the data generated by mini-GWAC. The dynamic feature makes it possible to predict different kind of anomalies appeared in the light curves at different time periods. Experiments show that the DARIMA model can effectively predict the abnormal events in light curves, and can generate warning triggers in the very early stage of an abnormal events.

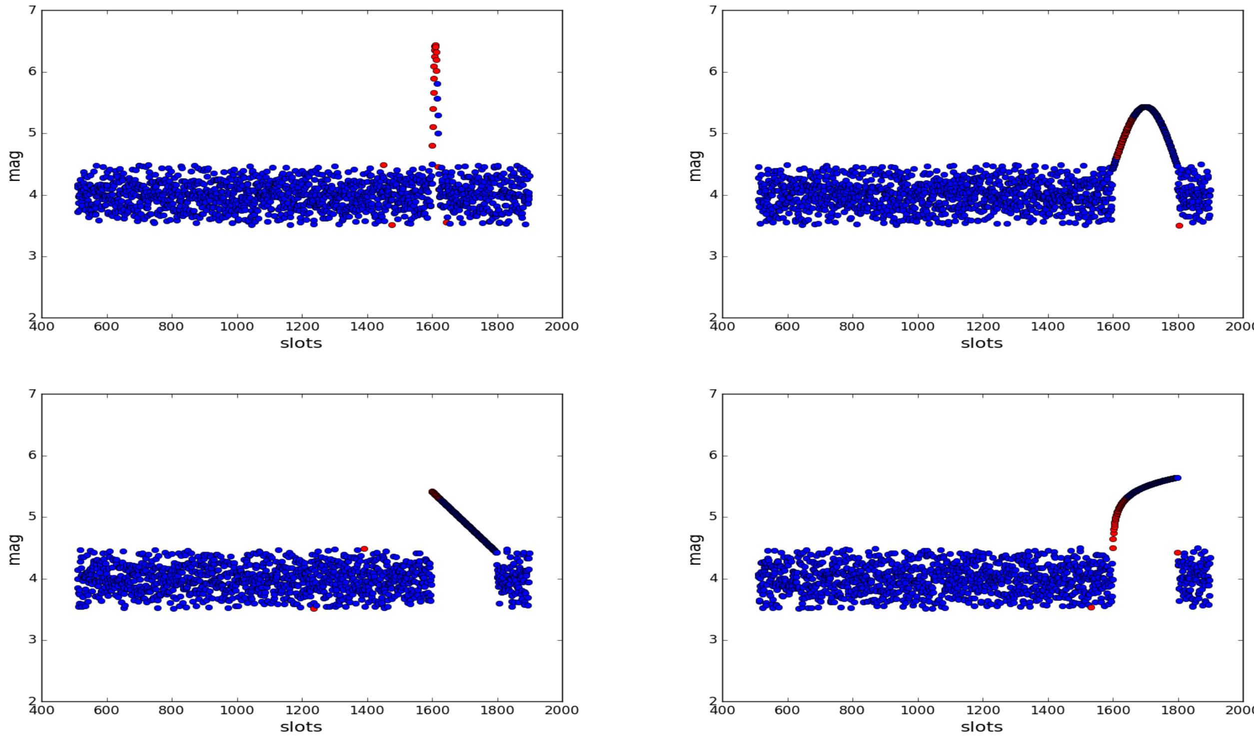

Fig. 4. Anomaly detection results of the simulated time series.

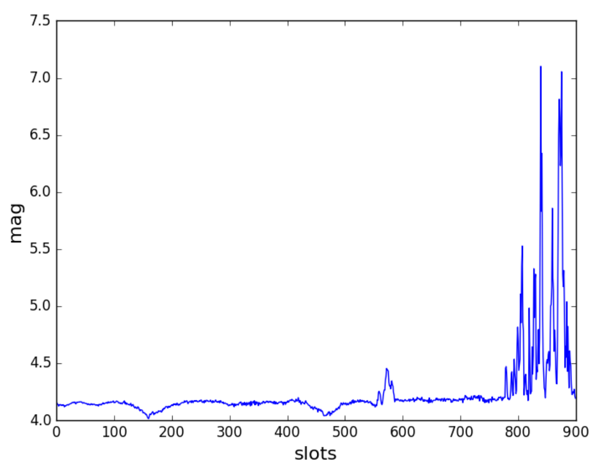

Fig. 5. Mini-GWAC time-series data with anomalies.

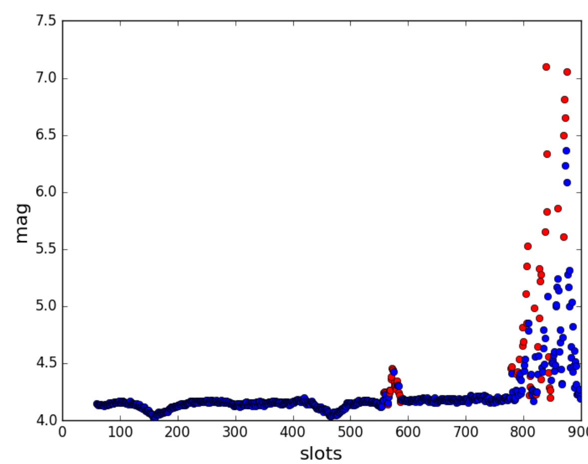

Fig. 6. Mini-GWAC anomaly detection results.

In the future, we plan to evaluate the DARIMA model with the GWAC data when GWAC is ready to operate on line.

ACKNOWLEDGMENT

This research is supported in part by the National Key Research and Development Program of China (No. 2016YFB1000602), the National Natural Science Foundation of China (Nos. 61440057, 61272087, 61363019, 61073008, and 11690023) and the MOE Research Center for Online Education Foundation (No. 2016ZD302).

REFERENCES

[1] S. Vaughan, "Random time series in astronomy," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1984, 2012.

[2] B. Cordier, J. Wei, J.-L. Atteia, S. Basa, A. Claret, F. Daigne, J. Deng, Y. Dong, O. Godet, A. Goldwurm, D. Gtz, X. Han, A. Klotz, C. Lachaud, J. Osborne, Y. Qiu, S. Schanne, B. Wu, J. Wang, C. Wu, L. Xin, B. Zhang, and S.-N. Zhang, "The SVOM gamma-ray burst mission," in *Proc. of the Conference Swift: 10 Years of Discovery*, 2014, arXiv:1512.03323.

[3] M. Wan, C. Wu, Y. Zhang, Y. Xu, and J. Wei, "A pre-research on GWAC

massive catalog data storage and processing system," *Astronomical Research and Technology*, vol. 13, no. 3, pp. 373–381, 2016.

[4] M. Wan, C. Wu, J. Wang, Y. Qiu, L. Xin, S. Mullender, H. Mhleisen, B. Scheers, Y. Zhang, N. Nes, M. Kersten, Y. Huang, J. Deng, and J. Wei, "Column store for GWAC: A high-cadence, high-density, large-scale astronomical light curve pipeline and distributed shared-nothing database," *Publications of the Astronomical Society of the Pacific*, vol. 128, no. 969, p. 114501, 2016.

[5] W. Meng, W. Chao, and Y. Z. et al, "Study on selection of GWAC database system catalog massive data processing," *Astronomical Research and Technology*, vol. 3, pp. 373–381, 2016.

[6] K. P. Harikrishnan, R. Misra, and G. Ambika, "Nonlinear time series analysis of the light curves from the black hole system GRS1915+105," *Res. Astron. Astrophys.*, vol. 11, pp. 71–90, 2011.

[7] M. Tarnopolski, "Nonlinear time-series analysis of Hyperion's lightcurves," *Astrophysics and Space Science*, vol. 357, no. 2, pp. 1–12, 2015.

[8] P. R. McWhirter, S. Wright, I. A. Steele, D. Al-Jumeily, A. Hussain, P. Fergus, D.-S. Huang, V. Bevilacqua, and P. Premaratne, "A dynamic, modular intelligent-agent framework for astronomical light curve analysis and classification," in *Proceedings of the 12th International Conference on Intelligent Computing Theories and Application (ICIC 2016)*, 2016.

[9] C. A. Ratanamahatana, J. Lin, D. Gunopulos, E. Keogh, M. Vlachos, and G. Das, "Mining time series data," in *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*, O. Maimon and L. Rokach, Eds. Kluwer Academic Publishers, 2010, ch. 56, pp. 1049–1077.

[10] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[11] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, Aug 1998.

[12] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: http://dx.doi.org/10.1023/A:1010933404324

[13] C. M. Bishop, *Neural networks for pattern recognition*. Oxford university press, 1995.

[14] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015, published online 2014; based on TR arXiv:1404.7828 [cs.NE].

[15] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[16] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Selected Papers of Hirotugu Akaike*. Springer, 1998, pp. 199–213.

[17] ——, "A bayesian analysis of the minimum aic procedure," in *Selected Papers of Hirotugu Akaike*. Springer, 1998, pp. 275–280.

[18] E. J. Hannan and B. G. Quinn, "The determination of the order of an autoregression," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 190–195, 1979.

[19] Z. Xiao and P. C. Phillips, "An ADF coefficient test for a unit root in ARMA models of unknown order with empirical applications to the US economy," *The Econometrics Journal*, vol. 1, no. 2, pp. 27–43, 1998.

[20] S. Refsdal and H. Bondi, "The gravitational lens effect," *Monthly Notices of the Royal Astronomical Society*, vol. 128, no. 4, pp. 295–306, 1964, dOI: https://doi.org/10.1093/mnras/128.4.295.