

Machine Learning-based Stability Assessment and Change Detection for Geosynchronous Satellites

Phan Dao

AFRL Space Vehicles Directorate

Kristen Weasenforth, Jeff Hollon, Tamara Payne, Kimberly Kinateder

Applied Optimization

Adam Kruchten

University of Pittsburgh

ABSTRACT

Analysts have been able to manually inspect the light curve of a geosynchronous satellite to assess its stability - whether it is stable (three-axis-stabilized) or unstable (tumbling). However, with the large volume of data collected persistently with wide field of view sensors, manual inspection by humans may not be sustainable. It is desirable to automate the stability assessment to (a) classify the satellite as stable or tumbling and when possible (b) pinpoint the moment it transitions from stable to becoming unstable. In this paper, we will show how such an automated system is developed. We found the Random Forest (RF) of Decision Trees classifier to be sufficiently robust and accurate as a solution for (a) when the satellites are either stable or in an established (steady) state of tumbling, as evidenced by the high level of accuracy for stability assessment achieved by the RF. A discussion of the optimal features to use with RF is provided. Once the RF algorithm has detected the first night the satellite becomes unstable, we then aim to pinpoint the precise time of the change. Our trained RF alone is not sufficient for detecting the onset of tumbling because it requires and labels the entire light curve. Also with our existing data set, it was not sufficiently trained to recognize the state in the interim. During that time a combination of two tests is used to recognize a tumbling satellite. The periodicity test determines the significance of periodicity at the detected frequency. The normality test applied to residuals in the signature subintervals detects the presence of aliasing, which is caused by the fundamental tumbling frequency being higher than the observation sampling frequency. In the aftermath of the onset of instability, all three tests – RF, periodicity, and normality – are combined to update the satellite’s status. We show the results of applying this set of algorithms on multi-year high cadence photometry of geosynchronous satellites.

1. INTRODUCTION

High cadence observations of deep space satellites have been a valuable tool to collect long-term photometric information about active and inactive objects left in or near the Geosynchronous Earth Orbit (GEO) belt. Many concepts based on Commercial Off-The-Shelf (COTS) components have been proposed and implemented [1]. While these sensors, which are not based on telescopes, have limited sensitivity, they can measure visual magnitudes of commercial satellites in GEO with useful signal to noise ratios over a significant range of solar phase angles. On the other hand, as fixed sensors with large fields of view, they are capable of persistent operation and afford us the opportunity to produce low-cost and multi-year data sets on many satellites. Even with occasional gaps due to weather events, the nearly continuous coverage allows the user to study the evolution of a satellite’s light curve over a substantial portion of its lifetime. While most satellite operators would communicate important events pertaining to their assets, such as the onset of instability or when they are about to move them to the graveyard orbit and abandon them, the information is often delayed and in some cases not available nor accurate.

We assert that the ability to verify the status of all satellites and objects in GEO and monitor the situation using persistent sensors is important to properly assessing the GEO environment with timeliness. When and if a significant portion of the GEO belt is covered by persistent observation, a global understanding of this environment is even more beneficial. While the collisional risk in GEO is still low [2], it is important to continuously assess the risk of collision in the GEO belt, which can pollute the environment with long-lasting fragments. Therefore, it is important to understand the dynamics of a large sample of the belt population. Persistent, high-cadence monitoring and

automated updating of the environment is crucial to support the abovementioned objectives. We can select a large number of inactive objects to study. Periodicity attributes will be useful in identifying the objects in that collection. Periodicity statistics are also useful for planning future debris removal missions [3]. However, the task of analyzing years of light curves for hundreds, if not thousands, of potential objects in the database is a daunting task if done manually by analysts. We propose to use simple automated algorithms to ingest this large amount of photometric data to automatically determine the status of each satellite with respect to its stability. We are only concerned with determining whether a satellite is active, meaning attitude-controlled, or inactive, as indicated by its tumbling signature. Here, by signature we mean the photometric data recorded from a single satellite over one night by a single sensor. We refer to the two possible states that can be derived from photometry as Stable and Tumbling. To test and evaluate the algorithms, we use a multi-year data set of satellites that can be readily measured over the continental US.

A sequence of events is shown in Fig. 1 for a satellite that was observed for some time before it becomes unstable and exhibits a tumbling signature. The state of this satellite is determined by applying the proposed tests. Phase 1 (blue) is the period when only the Random Forest classifier is used to estimate the satellite state. Phase 1 ends when RF detects an unstable light curve. Since RF is not sensitive enough to classify a light curve that is tumbling for only a portion of the night, in Phase 2 (orange) we activate the periodicity and normality tests to detect the onset of tumbling. If T is the night the RF classifier finds the first tumbling light curve, as shown in Fig. 1, we would start Phase 2 on night $T - 2$. In this example, the tests from Phase 2 identify the onset of a tumbling signature on night $T - 1$ at a given subinterval. We enter Phase 3 (green) the following night. In Phase 3, all tests are applied. The results are combined to improve the confidence levels of the assessment and update the stability status.

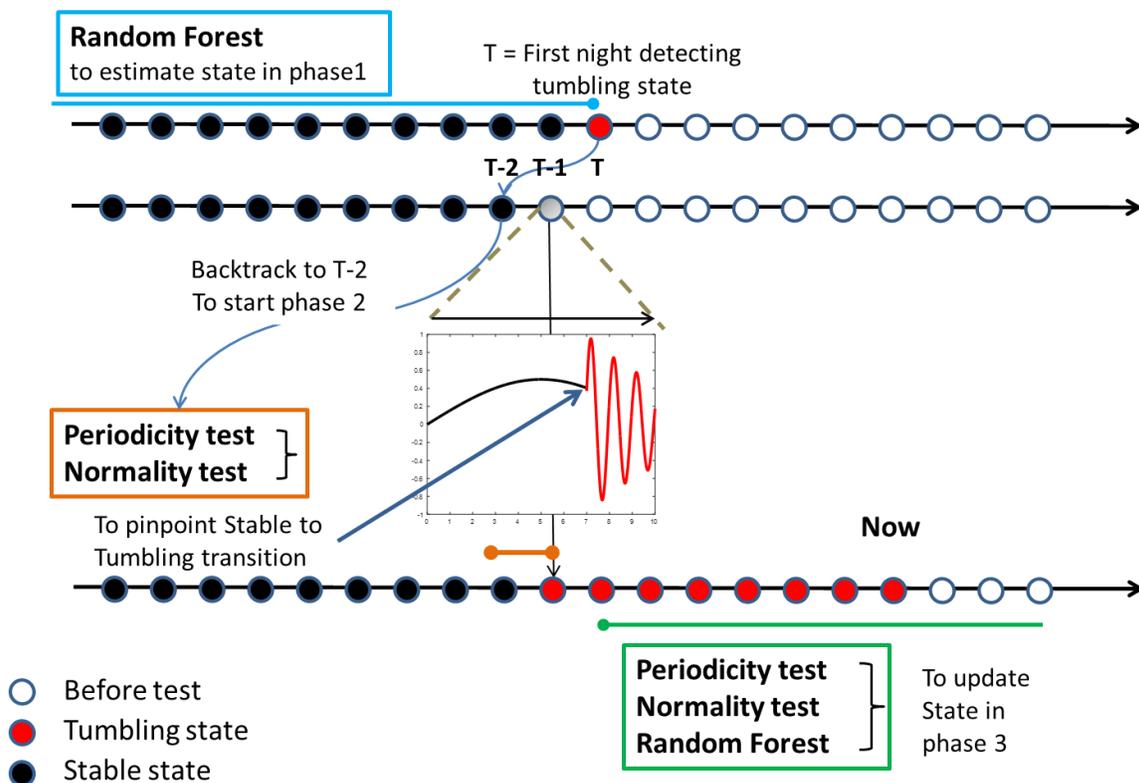


Fig. 1. Sequence of events and application of tests on a satellite that went unstable (tumbling) on $T-1$ and tagged as such by the RF test on T . Each dot represents an observation night.

Nominally, Phase 1 is the period when the satellite is active. Phase 3 is the period after it becomes unstable and begins to tumble. Obviously, this simple scenario does not include a complicated interim phase that may consist of owner/operator attempts to reestablish attitude control or an orderly preparation to retire the satellite. With the proposed techniques that are described here, we are able to assess all phases of this transition. At this time, we only have a few examples of satellites that were observed in all three phases. Most of the data consists of satellites that are always stable or always unstable. Someday, when persistent observations are a common practice for the space community, all GEO satellites will be monitored throughout their entire lifetime in space, and each will be represented by a data set with all possible phases similar to the scenario diagramed in Fig. 1.

2. RANDOM FOREST FOR STABILITY CLASSIFICATION

The Random Forest (RF) algorithm [4] is a type of supervised machine learning. The algorithm typically creates a forest of “bagged” (bootstrap aggregated) decision trees. “Bagged” means that each tree is created using random subsets of the training data with replacement. “With replacement” means that after a subset of training data is used to train one tree, that subset is returned to the total population of training data and so is available to the next tree for sampling. In addition, part of the RF approach is that each tree is also using a random subset of the features with replacement. The features are derived from the observation data and typically a random $\sqrt{\text{total \# of features}}$ are used per tree, which helps de-correlate the trees [5]. One difference between trees in a RF and single decision trees is that decision trees will choose to split the data on the feature that is the most effective classifier first and continue splitting on the next best classifiers. Trees within a RF will choose a random feature at each split of data. This aspect of RFs contributes to their robustness and removes the necessity of identifying which feature is most effective. The votes from decision trees are all used for the final prediction. Labeled data is used to train the RF classifier. Our training data consists predominantly of two classes of satellite: those always stable and those always unstable. The class of satellites that are observed as stable at one point and unstable at a later point is severely underrepresented in the available data sets, therefore we do not have enough data from this class to train the RF to identify it. Because RF cannot be properly trained to label the light curves in this transition phase and cannot pinpoint the time of transition, we rely on two additional tests for those specific purposes. For our training set, we used 5245 signatures from contributing partners, all operated from the same location. This provides a set of 4606 signatures of stable GEO objects and 639 signatures of unstable/tumbling GEO objects.

2.1. FEATURES FOR INPUT INTO RANDOM FOREST

In order to utilize photometric data in the RF algorithm, it is necessary to condense the entire light curve into a vector of features. In our case, all components are real numbers. Many of the features focus on the goodness of fit of a functional fit to the data. The idea being that the stable signatures are well behaved whereas the unstable signatures are not, which should result in much better fits (higher coefficients of determination) for stable signatures. We did this for a quadratic, cubic, and cubic spline least squares fit. Another group of features are taken from a Fourier regression to the data. These features are used because we expect there to be a higher frequency sinusoid in the unstable signatures than in the stable signatures. Other features focus on the distribution of magnitude values. We expect the distribution of magnitude values to look different from each other for the two cases. Along with this idea, we have also made features out of the three largest peaks and the total number of peaks present in an individual signature. All of the features used in the RF are summarized in Table 1 below.

Table 1: Features used as input for RF to classify the stability status of a satellite.

Feature	Description
<i>correlation</i>	correlation coefficient between longitudinal phase angle (LPA) and apparent magnitude
R_{quad}^2	coefficient of determination for a quadratic regression between LPA and magnitude
R_{cubic}^2	coefficient of determination for a cubic regression between LPA and magnitude
μ	mean of magnitude values
σ	standard deviation of magnitude values
<i>skewness</i>	skewness of the distribution of magnitude values
<i>kurtosis</i>	kurtosis of the distribution of magnitude values

$range_{mag}$	range of magnitude values
$freq_{max}$	greatest frequency collected from an L1 regularized Fourier regression
$freq_{min}$	least frequency collected from an L1 regularized Fourier regression
$freq_{most}$	frequency with greatest amplitude collected from an L1 regularized Fourier regression
$peak_1$	magnitude value of the largest peak
$peakLoc_1$	LPA value of the largest peak
$peak_2$	magnitude value of the second largest peak
$peakLoc_2$	LPA value of the second largest peak
$peak_3$	magnitude value of the third largest peak
$peakLoc_3$	LPA value of the third largest peak
R^2_{spline}	coefficient of determination for a cubic spline regression between LPA and magnitude
$numPeaks_{norm}$	number of peaks in a signature divided by total range of LPA values

2.2. ASSESS ACCURACY OF RANDOM FOREST CLASSIFIER

Using these features, we train the RF on a subset of the training data set and test the accuracy of the algorithm on a holdout set, not used in the training process. With about a 50-50 split of the data into training and testing sets, the RF classifies signatures with about 98% accuracy. The confusion matrix in Fig. 2a shows the RF's accuracy of correctly classifying a signature as stable versus unstable. Starting from the top left box and going clockwise, the confusion matrix values can be interpreted as the true negative rate, false positive rate, true positive rate, and false negative rate, respectively. The false positive rate (top right), also known as the Type I error, is about 3% on the training data. The false negative rate (bottom left), also known as the Type II error, is about 10% for the training set. This means that the algorithm has a slightly higher error rate when classifying unstable signatures.

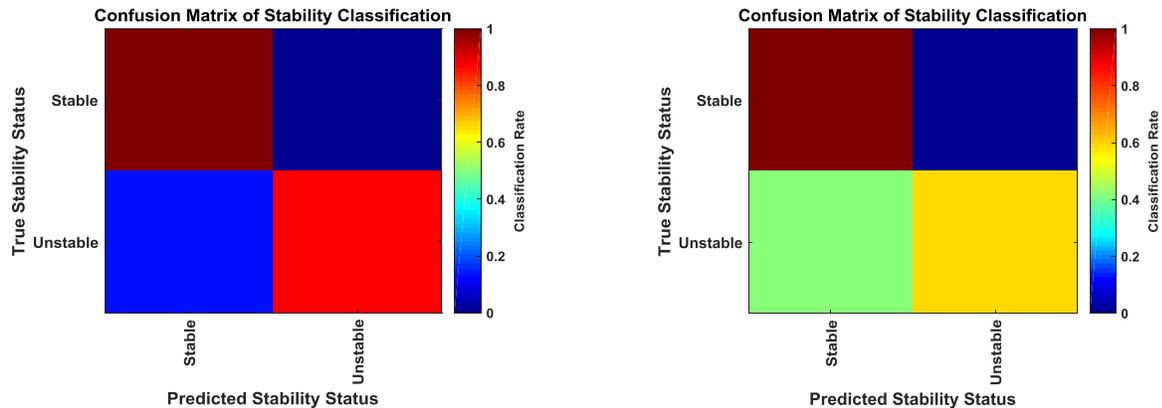


Fig. 2a. Left panel. Confusion matrix displaying the RF's accuracy of classifying signatures as stable versus unstable for holdout dataset from training data.

Fig. 2b. Right panel. Confusion matrix displaying the RF's accuracy at classifying signatures as stable versus unstable for dataset completely separate from training data.

We also applied the RF to an unseen set of data. The new testing set had 5609 signatures total with 4593 stable signatures and 1016 unstable signatures. The RF performs at about 92% accuracy on this new test set and did noticeably worse at classifying unstable signatures than in the previous set. Fig. 2b shows the accuracy of the RF on this new data set. The decrease in accuracy for the unstable class is likely due to the fact that the previously trained RF tends to not perform well on unstable light curves with non-sinusoidal periodicities typically found in signatures initially after a loss of stability. These types of signatures are present in the new test set. Since the RF was trained on unstable light curves with sinusoidal periodicity, it is understandable that it does not perform as well on the non-sinusoidal subclass of unstable light curves. The Type I error was 0.52% (top right) but the Type II error was 42% (bottom left).

Two types of non-sinusoidal tumbling signatures are identified: 1) aliased, nearly sinusoidal signatures (at 1/37 second cadence) and 2) fine-scale specular glints. We will discuss how a new test is devised to detect the manifestation of these signatures and complement the RF test. On its own, the RF is not sufficiently robust to assign unstable and detect the onset of time of instability, as evidenced by the high Type II error. We were able to test the RF on a satellite that went unstable while under observation. Echostar 3 went unstable on August 2, 2017. Though we do not have data from the night the satellite lost stability, we do have data from a few years before and a few months after. The RF performed well on this satellite, catching the first day after the satellite went unstable and only falsely marking one night after that as stable. The RF achieved 99% overall accuracy for classifying the stability status of this satellite.

3. PERIODICITY TEST

A test for periodicity was introduced to identify when a satellite has exhibited behavior typical of an unstable tumbling object through its photometric data. The main difference between the light curve of an unstable satellite from that of a stable satellite is the presence of sinusoidal or non-sinusoidal periodic waveforms. An example of a stable signature can be seen in Fig. 3 and an example of an unstable signature can be seen in Fig. 4. These figures show the brightness values of the satellite plotted against LPA. The figures in this paper show brightness in radiant intensity (RI with the units of Watts per steradian (W/sr) and LPA has units of degrees.

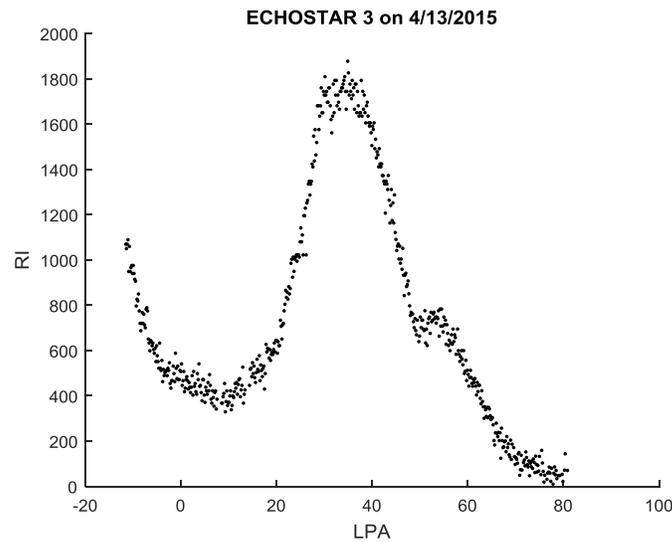


Fig. 3. Example of a stable signature.

The test to confirm the significance of a periodic component is Fisher's Exact Test for Periodicity. The test statistic is found from the most significant frequencies of the Lomb-Scargle periodogram of the signal. The test's null hypothesis, H_0 , claims that there is no significant periodic component and the power density is all due to noise. The alternative hypothesis, H_A , claims that there is a significant periodic component in the signal. The Lomb-Scargle periodogram is used because it does not require observations to be uniformly spaced, which is rare in our set of photometric data.

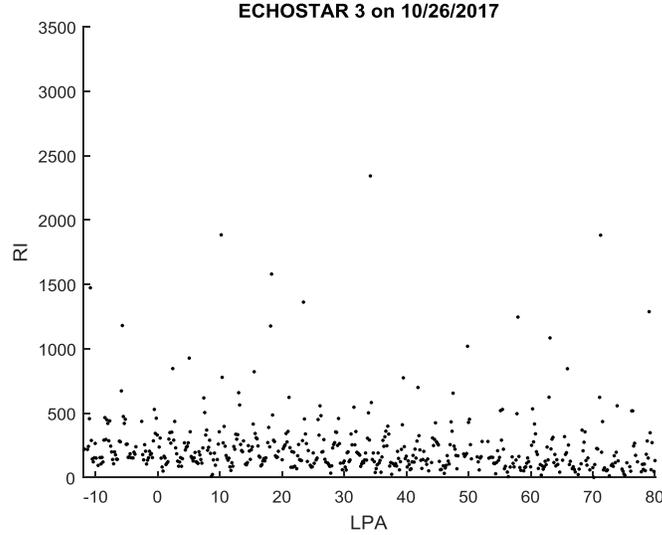


Fig. 4. Example of an unstable signature.

Let $P(f_l)$ be the periodogram of the time series, evaluated at frequencies $f_l = \frac{2\pi l}{n}$, $l = 1, \dots, a$. Then the test statistic, called the g-statistic, takes the following form:

$$g = \text{Fisher's - statistic} = \frac{\max_l P(f_l)}{\sum_{l=1}^a P(f_l)}$$

where $a = \frac{n}{2}$ and $n = \# \text{ of observations in the time series}$. We have modified the g-statistic expression to account for the contribution of harmonics in the periodogram that are present in signatures with periodic sharp glints. To find the p-value for this test, we use the probability distribution of the g-statistic. Part of the assumption of the null hypothesis that there is no significant periodic component, i.e., that the signal is Gaussian noise. Therefore, under the null hypothesis of Gaussian noise, if the value x is observed for the g-statistic, g , the p-value $P(g > x)$ can be computed to be:

$$P(g > x) = \sum_{k=1}^b (-1)^{k-1} \binom{a}{k} (1 - kx)^{a-1}$$

where b is the largest integer less than $1/x$ and x is the observed value of the g-statistic. If p-value $< \alpha$ for a chosen significance level, α , then a significant periodic component was observed and we can claim that the time series is periodic [6].

When initially running the Fisher Exact Test of Periodicity on a set of signatures, undesired results were observed. Because stable signatures have a single peak, it is likely that the test picked up a periodic component with a very long period and therefore thought the stable signatures were periodic. Since we only want to identify signatures from tumbling satellites as periodic, we had to make adjustments as to how the Fisher Exact Test of Periodicity was run, in order to reduce the number of false positives (stable signatures identified as periodic).

3.1. DE-TRENDING AND REMOVING OUTLIERS

Instead of removing a linear trend as is often done before spectral analysis, we fit a Tikhonov-regularized cubic spline model [7] to the flux data and analyze the residuals. Outliers are inevitable in satellite photometry, so we use Cook's Distance to remove some outliers before the de-trending. After running Cook's Distance the first time, we determine whether any outliers remain by testing whether any brightness values exceed 3σ , where σ is the standard

deviation of the brightness values. If any observations satisfy this inequality, we automatically run Cook's Distance a second time.

Cook's distance estimates the influence a single point has on a regression model by essentially implementing a "leave one out" form of cross validation. The regression is done on every point except for the one for which the Cook's distance is calculated. Repeat the procedure for each point and calculate the distance as:

$$D_i = \frac{r_i^2}{p \cdot MSE} \left(\frac{h_{ii}}{(1 - h_{ii})^2} \right)$$

Where r_i is the i^{th} residual and h_{ii} is the i^{th} diagonal of the hat matrix for the regression model using all of the points [8]. An outlier tends to have a high Cook's Distance, D_i , so we remove points such that $D_i > 10 \cdot \mu_{D_i}$ where μ_{D_i} is the mean of all D_i values.

3.2. REJECTED FREQUENCIES

There are some frequencies that we do not expect in the light curve from a tumbling satellite. For instance, if the fundamental frequency of a light curve is too low, that frequency corresponds to one or fewer periods occurring within the observation window. Then the frequency is likely to not correspond to a tumbling satellite because we expect a tumbling satellite to have a sinusoidal light curve with several periods occurring within the observation period. On the other hand, a fundamental frequency that is too high may be close to the Nyquist limit, which can cause aliasing within the light curve. Therefore, we choose to have both an upper and lower bound on the frequencies that we accept as possible for a tumbling satellite's light curve. To do this, all light curves whose fundamental frequencies are outside of the acceptable bounds are considered as not periodic. Note the Nyquist limit is not strictly observed by an irregularly sampled signature, but also when the object brightness is sampled at a low cadence.

These modifications to the base Fisher's Exact Test for Periodicity greatly improved our Type I and Type II errors for correctly identifying a stable versus unstable signature. Results of using the modified periodicity test to identify whether a satellite's signature is stable or unstable are given in Section 5.1.

3.3. PERIODICITY TEST ON AN EXPANDING INTERVAL

In addition to determining whether a satellite is stable or unstable for a full night, we also want to use the periodicity test on the nights leading up to the first night of an unstable RF classification to determine the time of night that the satellite went unstable. However, by testing the whole signature for periodicity, we do not get information on the time of night that periodicity was observed. Therefore, in order to pinpoint the exact time a satellite went unstable on the first night of instability, we need to additionally run the test on subintervals. We first tried dividing the signatures up into a set of disjoint subintervals, but found that these subintervals were generally too small to catch enough periods of the sinusoid present in a tumbling satellite's signature for the test to deem the interval as periodic. Instead, we tried using an expanding interval.

For this expanding interval approach, we still split the signature into disjoint subintervals, nine subintervals to be specific. Then the first expanding interval would be the first subinterval of the signature. The second expanding interval would be the first two subintervals of the signature, the third expanding interval would be the first three subintervals of the signature, and so on, with the last and ninth expanding interval contains the whole signature. An example of how this expanding interval works can be seen in Fig. 5.

To represent these results, a line is plotted across the bottom of the signature. If non-periodicity was rejected in interval, then that portion of the line is colored red. If non-periodicity could not be rejected in that interval, then that portion of the line is colored green. For the expanding intervals, the n^{th} subinterval of the signature is colored appropriately to represent the results from the n^{th} expanding interval. For example, only the last subinterval in the last expanding interval of the signature is colored red if the hypothesis is rejected, even though that expanding interval contains the whole signature. In Fig. 5, the periodicity test could not reject non-periodicity in any of the

intervals, thus all intervals are colored green. Results of this test’s ability to pinpoint the time of stability loss are discussed in Section 5.1.

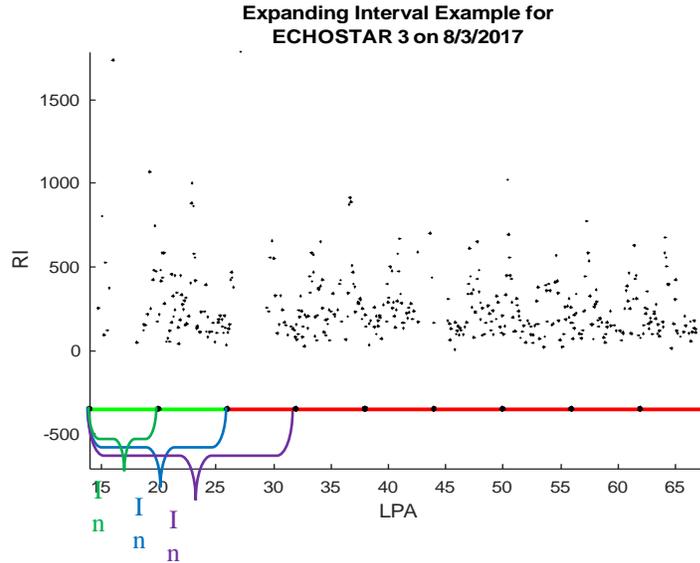


Fig. 5. Example of expanding interval and results of periodicity test on an expanding interval.

4. NORMALITY TEST

When examining unstable signatures, we found several of them to be aliased, i.e. the photometry is varying at a higher rate than the sensor sample rate. While most observations appear normally distributed around the local mean of data, observations from an aliased signatures behaved much differently. Aliased signatures contained one or more intervals of phase angles in which the signature appeared to fall both above and below the local mean in manner that causes the residuals to be bi-modal, an example of such phenomena can be seen in Fig. 6. Such behavior would not be consistent with normally distributed residuals about a local mean. To this end, we developed a hypothesis test to determine whether a signature was aliased by testing whether the residuals from a fitted model are normally distributed in the subintervals. We found that testing normality of residuals within subintervals performed better than testing whether the residuals from an entire signature were normal. One of the reasons why the test on the entire signature performed less well is that the cubic spline method we used to find the residuals fit so well to the observations that the distribution of residuals had a narrow peak around their zero mean. This peak was just non-normal enough for the test to reject normality overall. To understand the test, we will first examine the normality test we used to test individual subintervals, then we will analyze the hypothesis test to achieve an overall result for a signature based on the results from all subintervals.

4.1. KOLMOGOROV-SMIRNOV TEST

The Kolmogorov-Smirnov (KS) test can be used to test the normality of a dataset. The test has the following hypotheses:

H_0 : The dataset is from a population with a standard normal distribution.

H_A : The dataset is not from a population with a standard normal distribution.

The KS test works by comparing the Cumulative Distribution Function (CDF) of the data set to the CDF of a standard normal distribution. This comparison is made by finding the largest difference in the two CDFs. If the resulting p-value is below the selected significance level, α , then the test rejects the null hypothesis of normality with $100(1 - \alpha)\%$ confidence. Otherwise, there is not enough evidence to reject normality.

For our purposes, since we seek to know if the observations were normally distributed around the local mean of the data, we fit a cubic spline to the given signature and test its residuals for normality using the KS test. However, since the KS test specifically tests against a standard normal distribution, we also normalize the residuals by subtracting the means of the residuals and dividing by the standard deviation of the residuals before performing the KS test.

4.2. HYPOTHESIS TEST FOR NORMALITY

For a well-behaved signature, if the signature follows the model, $Y(i) = f(i) + \varepsilon(i)$, where $f(i)$ denotes the true signature value at phase angle i and $\varepsilon(i)$ represents zero-mean, normally-distributed noise, then we expect that the residuals are normally distributed about f with a mean of zero and a constant variance. Thus any subset of residuals is expected not to fail a hypothesis test for normality.

We begin by setting a desired overall level of significance. Next, we choose a number of subintervals, n , in which to test normality of residuals and a significance level, α_N , at which to perform the n hypothesis tests. Empirically, we found that $n = 9$ yielded the best results. Let K be the number of subintervals out of n in which normality is rejected. A large value of K will suggest that a signature is aliased. The hypotheses for our test are:

H_0 : Satellite is stable.

H_A : Satellite is not stable .

In order to perform the n tests, we fit a mean curve to the signature using a cubic spline. The signature is divided into n subintervals of equal length in LPA. Next, we find the test statistic, K , that represents the number of subintervals for which the null hypothesis of normality for residuals is rejected. We use the Kolmogorov-Smirnov test for normality at significance level, α_N , on each subset of residuals. The critical region for the rejection of H_0 is found by determining the critical value, k_0 , such that

$$P(K \geq k_0 | \text{satellite is stable}) < \alpha.$$

4.3. DEFINING TEST STATISTIC AND DETERMINING CRITICAL VALUE

To empirically determine the critical value of this hypothesis test, we ran the normality test on a large number of signatures. Our data set was 2879 stable signatures, 141 aliased signatures, and 721 unstable, but non-aliased signatures. The stable signatures were from satellites NORAD 23754, 23846, 24315, 24812. The aliased signatures came from satellites 22653, 23305, 23313, 23461, 23598, 27820, and the non-aliased unstable signatures came from satellites 22653, 23305, 23313, 23461, 23528, 23598, 25004, 27820.

In order to determine the critical value, we first needed to determine how many subintervals we wanted to divide each signature into. To do this, we ran the normality test on $n = 2, \dots, 13$ subintervals and recorded the number of subintervals that rejected normality for each signature and for each n . We observed that the individual normality test results for consecutive subintervals were not independent of one another. For this reason, we could not assume that K has a Binomial distribution. Because of this, the empirical distribution was used to find the critical value for the test. Using that information, we looked at the error rates for our hypothesis test for different critical values k_0 and values of n and determined that $n = 9$ provided us with the best Type I and Type II error rates overall. Depending on whether or not we want the test to minimize Type I or Type II error, we choose either $k_0 = 2$ or $k_0 = 1$ respectively for the critical value.

5. RESULTS

To confirm the effectiveness of each of our tests, we used a large set of stable and unstable signatures from several different satellites. The dataset we used is the same set mentioned in Section 4.3. This large test set was chosen to obtain the error rates for our tests because reliance on signatures from a single satellite could skew the results.

5.1. RESULTS OF PERIODICITY TEST

The periodicity test performs best on typical tumbling signatures, characterized by sinusoidal or non-sinusoidal, but a periodic time series. Many unstable signatures in the data set from a persistent sensor with modest cadence do not

have this pattern due to aliasing. Thus, they tend to not be marked as unstable by the periodicity test. However, due to our modifications to the Periodicity Test, the chance of falsely identifying a stable signature as periodic is low.

5.2. RESULTS OF NORMALITY TEST

After testing different significance levels of α_N and critical values of k_0 used in the KS tests for normality in the nine subintervals, we discovered a trade-off between the empirically derived Type I and Type II error (equivalently, probabilities of false positives and false negatives). The Normality test performed with a smaller α_N and a critical value of $k_0 = 2$ favored a small Type I error, while a larger α_N and critical value $k_0 = 1$ tended to promote a smaller Type II error. The error rates were managed when the test was combined with the periodicity test. This obtained the overall desired level of confidence.

Since the Normality test was designed to identify regions of instability, we will show graphical results applied at $\alpha_N = 0.125$ and critical value $k_0 = 1$ on data of satellite AMC-9, which has several unstable signatures. For each of the plots below, the signature is represented by black dots, the spline fitted to the signature is plotted in blue, and along the bottom, a green or red line is plotted for each of the 9 subintervals showing the results of the normality test. Green is plotted if the test did not reject normality for the residuals of the subinterval, and red is plotted if the test determined that residuals of the subinterval were not normal.

Fig. 6 is an example of an aliased unstable signature. This type of aliasing is due to the glints being smaller than the integration window. The normality test correctly marks almost every subinterval containing aliased behavior. If we zoom into one of the subintervals with aliasing present, we can see why the test rejects normality in that region. For example in the region from -10° to 15° LPA, the spline fit was between the two main trends of observations. In this case, the rejection was due to the distribution of the residuals being bi-modal, and therefore clearly, not normally distributed.

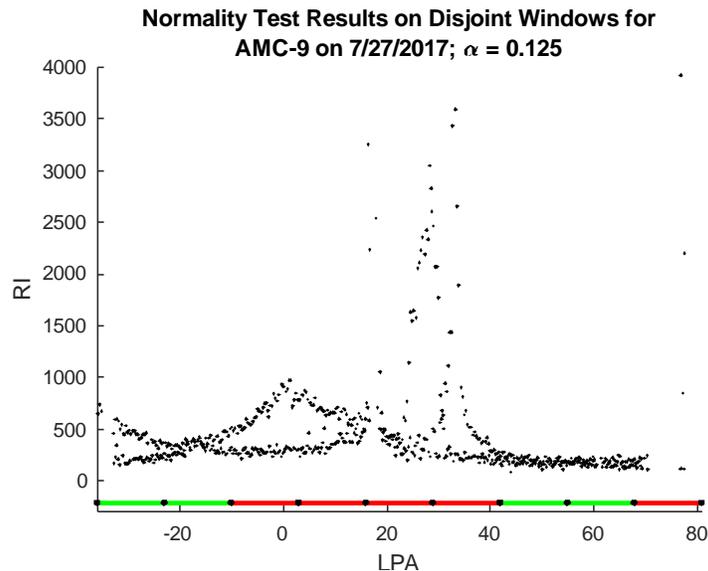


Fig. 6. Signature from AMC-9 taken on July 27, 2017. Normality test was run with significance level $\alpha_N = 0.125$.

5.3. RESULTS OF NORMALITY AND PERIODICITY TEST

The Normality and Periodicity tests were developed to identify when a satellite transitions from stable to unstable. Both hypothesis tests contribute to this decision. The Periodicity test determines if a satellite is tumbling and the Normality test determines if the satellite signature does not follow a model that is normally distributed about a trend function. The latter implies aliasing and therefore also implies the satellite is tumbling. Together, the tests can identify an unstable satellite under either condition, and so we developed a method that combines the results of these tests.

Since the Normality test is run on nine subintervals and the Periodicity test is run on full signatures, we require a method of combining the individual test results and characterizing their combined error rates. Since the Normality test provides an assessment of the whole night depending on the value of K , the number of subintervals that reject normality, we can use the Normality test's overall assessment with the Periodicity test's result since they both determine stability for the whole night. Our combined test method considers a signature to be unstable if the signature fails *either* test.

Through testing on the set of stable and unstable signatures, we obtained empirical error rates for specified values of α_N and k_o . If we run the Normality test with parameters, $\alpha_N = 0.125$ and $k_o = 1$, and the Periodicity test with significance level, $\alpha = 0.05$, a Type II error of 10% is achieved. That is, 90% of the time the combined tests correctly label an unstable signature. On the other hand, if we run the Normality test with parameters, $\alpha_N = 0.005$ and $k_o = 2$, and the Periodicity test with significance level, $\alpha = 0.01$, a Type I error of 10% is achieved. Depending on which error rate (either Type I or Type II error) we are more concerned with, we can alter the parameters for the individual tests to optimize that error rate.

We show the results of the tests on a typical tumbling signature and an aliased signature since they are both examples of unstable signatures in Fig. 7 and Fig. 8. The results for each interval of the Normality test are shown as the top colored line at the bottom of the figures and the results of the Periodicity test are shown beneath it. If the test's null hypothesis was rejected for an interval, then that interval is colored red. The interval is colored green if there is not enough evidence to reject the null hypothesis. A signature from the day after Echostar 3 went unstable is shown in Fig. 7. Since K was observed to be equal to 7 for the Normality test on this signature and the Periodicity test also identified a significant periodic component, the combined test results determine that the signature is unstable. In the next plot, Fig. 8, an aliased unstable signature from AMC-9 is shown. The regions of aliasing are detected by the Normality test, resulting in an observed test statistic of $K = 3$. Therefore, this signature also failed both tests and is considered unstable by the combined test.

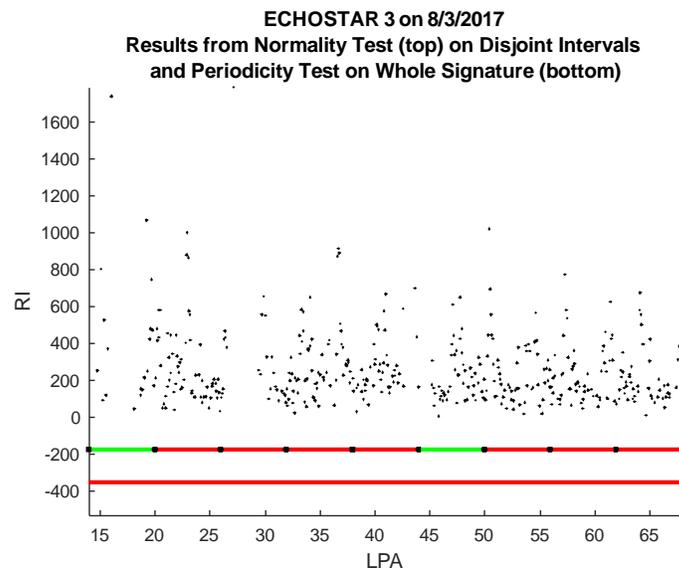


Fig. 7. Results of the Normality and Periodicity tests on an unstable signature from Echostar 3.

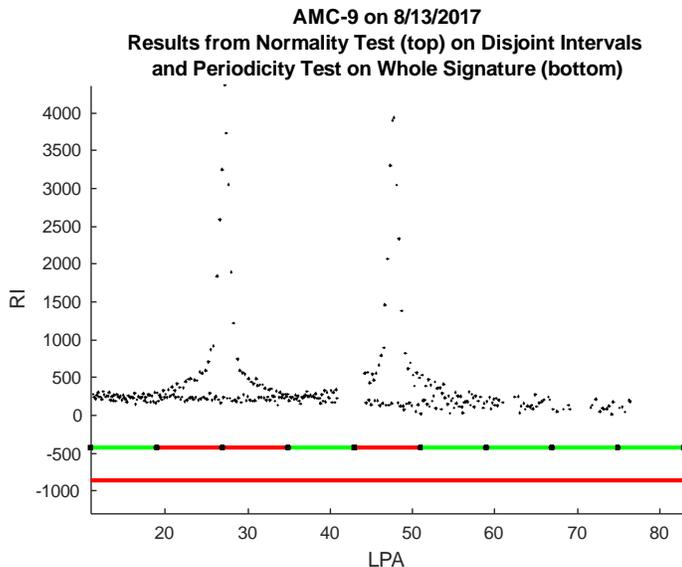


Fig. 8. Results of the Normality and Periodicity tests on an aliased, unstable signature from AMC-9.

6. DISCUSSION AND CONCLUSIONS

Our work is at the early phase of implementing a fully automated, analytical procedure to label the stability status of GEO objects using long-term, high cadence photometric data. While Machine Learning classifiers such as RF can reliably classify the stability status of a GEO satellite using features derived from its photometric data, typical data sampling rates of persistent sensors make the determination of instability more complicated since aliasing can occur when the tumbling rate is fast compared to the data collection cadence. Another complication is that the RF training set is currently not sufficient for satellites in the transition phase between the time that steady-state stability ends and the time that steady-state tumbling is achieved. To mitigate these problems, we added tests that detect “aliasing” and/or periodicity. The results of those tests are integrated with probabilistic considerations.

As the photometry database from persistent sensors becomes more comprehensive and signatures from the transition phase are well represented, we may consider using RF over the entire lifetime of the satellite. On the other hand, the use of the Periodicity/Normality test probabilities can be combined with RF results as observables of a hidden Markov model (HMM) to reduce Type I error. The Bayesian formalism can take advantage of the sequence of observables to update the stability state with higher confidence.

7. FUTURE WORK

The majority of our work so far addresses Phases 1 and 3 from Fig. 1. However, we have plans to address Phase 2 and improve upon Phase 3. These plans are focused on pinpointing the time of the loss of stability and combining additional tests to update the stability status. We will work on implementing a Bayesian Generalized Lom-Scargle technique [9] to determine the fundamental photometric period if periodicity is confirmed.

7.1 PINPOINTING LOSS OF STABILITY

Pinpointing the time a satellite went unstable is part of Phase 2, as shown in Fig. 1. The idea is that, once the RF algorithm determines the satellite went unstable, we will backtrack a couple days to start the process of determining the time the satellite lost stability. To do this, we introduced a method of running the Periodicity test on an expanding interval. This method of running this test in combination with running the Normality test on disjoint subintervals can provide us with an estimate of the time of night that a satellite went unstable. The first interval that detects instability from either test will signal the time when the satellite lost stability. Currently, we do not have an abundance of data from satellites that went from stable to unstable while under observation. In fact, in our dataset only two satellites were observed before and after the loss of stability. Additionally, only one of those satellites was

observed on the night the satellite went unstable. Due to limited observational evidence, we have not yet been able to extensively test how well we can pinpoint the instability onset time.

We do have the results of our approach on one example of such a signature in our dataset. AMC-9 went unstable midway through the night on 6/17/2017 UTC. Fig. 9 shows the results of running our method to pinpoint the time the satellite’s stability status changed. In this case, the Normality test is run using significance level, $\alpha_N = 0.01$, and the significance level of the Periodicity test is run at $\alpha = 0.01$. The Normality test begins to detect non-normality in the subinterval ($20^\circ - 30^\circ$ LPA) before the satellite loses stability around 32° LPA. However, since the unstable portion of the signature is not particularly sinusoidal or periodic, the Periodicity test does not pick up on the change.

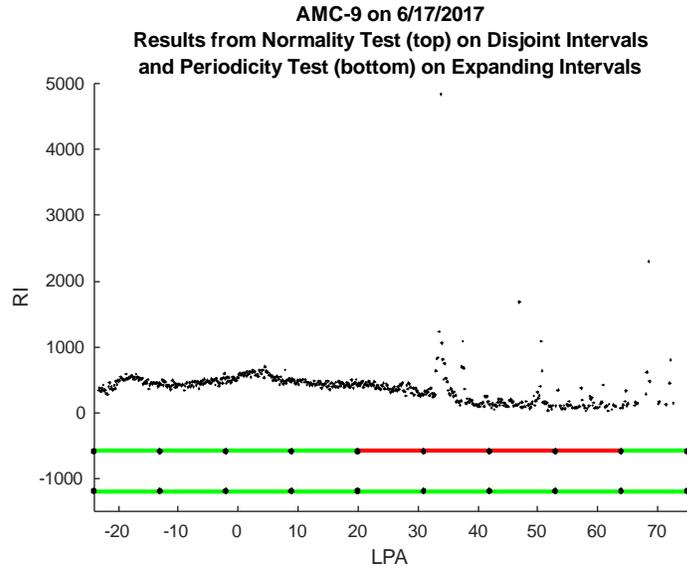


Fig. 9. Results of tests to pinpoint time of stability status change for the night AMC-9 went unstable.

Though the Periodicity test on an expanding interval was incapable of identifying the change in stability for this particular example, this result is not unexpected considering the lack of a periodic pattern in the data. The Normality test was able to perform as intended to identify when the instability began, however. We expect with more observations of these types of cases, we can test these algorithms further and show their ability to detect transitions between stable and unstable. Anticipating a lack of real data, we plan to produce simulated signatures of the type shown in Fig. 9 for further development of these tests.

7.2. UPDATING STABILITY STATUS USING 3 TESTS TOGETHER

Confirming the satellite status is the goal of Phase 3, as shown in Fig. 1. On day $T + 1$, our approach calls for running the Normality and Periodicity tests together with RF to detect whether a return to stability occurs. Such a return would occur with very low probability and only happens if the operator manages to save the satellite. We plan to use all three tests – RF, Periodicity, and Normality – to evaluate the satellite status in an automated way. Since we want the tests in this application to be as sensitive as possible, we will choose testing parameters (such as choice of α_N and k_o in the Normality test) that favor a small Type II error. In order to achieve an overall desired low error rate, each test can be performed at one third of the desired error rate in order to achieve the desired family-wise error rate.

A conditional probability approach can also be taken to confirm the status of the satellite on day $T + 1$. Such an approach would utilize assumed probabilities of the satellite transitioning from night to night between stable and unstable, or the other way around. Given the hypothesis that the satellite is unstable, a Bayesian or conditional probability calculation could be made to quantify the probability of the satellite experiencing a particular sequence of transitions from stable to unstable or vice versa. This probability calculation would then be used to decide

whether or not to confirm the unstable status. The conditional probability approach and the Hidden Markov model as previously described in Section 6 are being considered.

8. REFERENCES

1. McGraw, J. T. (2013). Lens Systems for Sky Surveys and Space Surveillance. *Proceedings of the 2013 AMOS Technical Conference*. Maui.
2. Howard, S. M.-K. (2014). GEO Collisional Risk Assessment Based on Analysis of NASA-WISE Data and Modeling. In S. Ryan (Ed.), *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference*. Maui: The Maui Economic Development Board.
3. Andersen, K. (2014, June). Analysis of tumble rate and spin axis orientation for rocket bodies near geostationary altitude. California Polytechnic State University.
4. Breiman, L. (2001). *Random Forests*. Berkeley: University of California.
5. James, G. (2017). *An Introduction to Statistical Learning: with Applications in R*. New York City: Springer.
6. Liew, A. W.-C. (2009). Statistical Power of Fisher Test for the Detection of Short Periodic Gene Expression. *Pattern Recognition*, 549-556.
7. Stephanie. (2017, June 22). *Tikhonov Regularization: Simple Definition*. Retrieved from Statistics How To: <http://www.statisticshowto.com/tikhonov-regularization/>
8. *Cook's Distance*. (2018). Retrieved from Mathworks: <https://www.mathworks.com/help/stats/cooks-distance.html>
9. Mortier, A. (2015). BGLS: A Bayesian formalism for the generalised Lomb-Scargle periodogram. *A&A* 573, A101 (2015). [arXiv:1412.0467v1](https://arxiv.org/abs/1412.0467v1).

ACKNOWLEDGEMENTS

We acknowledged the support of the Spacecraft Object Tracking and Characterization program and its program manager Virginia Wright. Dr. Dave Monet's comments and assistance are well appreciated.