A Practical Technique for Discriminating Manoeuvres and Observational Anomalies from Precision Sequential Estimates of Orbits

Tommy Fryer, Robert Arthur, Paul Walker
Space, Defence and Intelligence, CGI IT UK Ltd

ABSTRACT

In order to monitor the health of our Kalman Filter, which may be affected by anomalous observations, changes in motion or poorly tuned process noise, we define a statistic that is able to report adverse conditions. In this paper we assess techniques for detecting manoeuvres and discriminating them from observation anomalies using this metric.

We assess the performance of three classes of statistical tests to detect deviation from the null hypothesis which is that nothing has changed from a baseline: Cramer-von-Mises χ^2 test, two-sample tests against an empirical distribution function and testing the sample mean and variance against baseline empirical values. For synthetic models with realistic process noise, we find that impulse manoeuvres of a typical size can be detected a short time after they occur, we propose to distinguish them from anomalies by rejecting anomalous Kalman Updates and accruing evidence of filter anomaly from more than one observer. For impulses on an otherwise inert object, we find that the Cramer-von-Mises χ^2 test performs best, requires the least computation, requires no baseline data and is as robust to excess process noise as the empirical distribution function tests.

For small continuous thrusts where it is not possible to identify a change in the anomaly metric from the baseline thrust in a single pass, we are able to identify the change by accruing the statistics over multiple passes. The most promising test was the Bootstrap sample variance test which was most robust to sub-optimal process noise and able to detect changes even when baselined against a continuously thrusting object. The Empirical Distribution function tests performed less well but were much improved over the χ^2 tests when process noise was not tuned well or when the baseline period includes thrust.

1. INTRODUCTION

1.1 Problem Description

Precision Orbit Determination (POD) is essential to the Conjunction Analysis problem for the provision of refined orbital trajectories with realistic uncertainty [1]. Ensuring that these estimates are as precise, accurate and realistic as possible given the history and density of observations that are available for each Resident Space Object (RSO) is one of the primary responsibilities of an orbital analyst in a Space Operations Centre.

In support of Space Domain Awareness, POD can detect manoeuvres and could potentially extract changes in the pattern of life of high interest, high revisit rate RSOs more quickly and with greater fidelity than relying on derived approximate orbits such as Two Line Element sets.

Whatever the objective of the sequential POD, the inclusion of erroneous observations or observations that follow a significant event that is not modelled such as an impulse manoeuvre typically results not only in a bad state estimate, but the state estimate may not quickly recover through continual updates, and instead the process should be actively restored. It is therefore important to provide a reliable means to identify that this has occurred; reset the estimation process and to notify interested parties of the potential manoeuvre.

Subtle changes such as a change of attitude may result in a net non-modelled force, but no clear indication from the sequential estimator the something has gone wrong in the estimation process. We investigate whether the same metrics used for impulse detection can be analysed in more detail to provide evidence of an change of non-modelled force. To this end we synthesize very low impulse continuous thrusts.

The focus of this work is to assess the reliability and sensitivity of various statistical measures of sequential estimation update metrics for two distinct scenarios:

1. To continually monitoring the performance of a sequential POD estimate to detect anomalies and impulse manoeuvres;

2. To uncover subtle but enduring changes such as small continuous thrust.

The analysis considers both observations of position in Low Earth Orbit and angles only in Deep Space, and considers how well the estimation process is tuned impacts the ability of each of the assessed measures to reliably discriminate impulses of various ΔV and continuous thrusts producing various accelerations.

1.2 Estimation Process

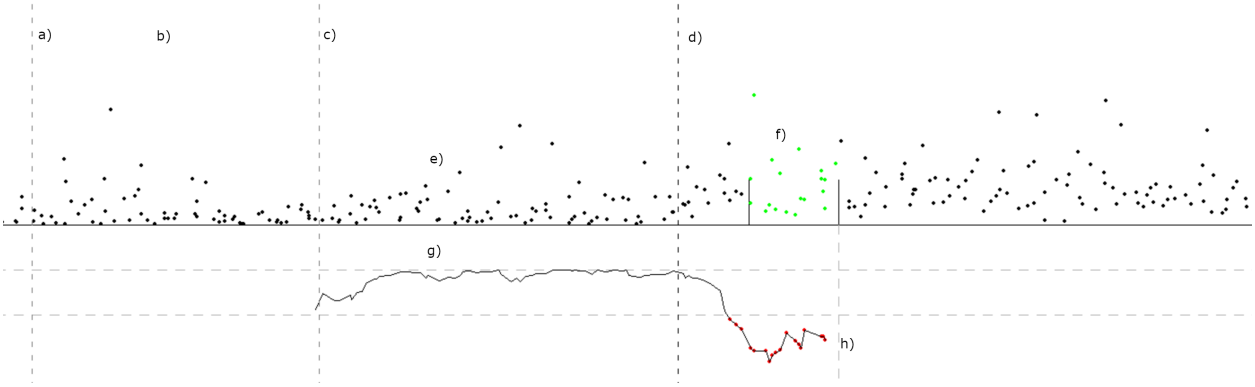

Fig. 1: Diagram showing the state covariance initialisation and estimation routine with a statistical test applied using a rolling sample period (time elapsed on x axis). a) start of the batch estimation period, b) baseline Kalman epoch, c) end of batch estimation period, d) epoch of change in Kalman behaviour, e) input statistic (Mahalanobis distances), f) rolling test sample period, g) statistical test result, h) statistical test threshold exceeded.

The operational process that this analysis supports is based upon a Self-Starting Filter which proceeds as follows (as shown in Figure 1):

1. Perform a Batch Estimation, including fitting extended state variables as required for the start of the batch period. Shown in Figure 1 period a) to c).
2. Using the Batch Solution and an initial covariance (usually re-scaled from the Batch estimate), run the required Filter over the Batch period.
3. Gather the necessary information from this exercise to initialise the measures considered here. For Empircal Distribution Functions and Bootstrap methods we use a prototypical region after a short initialisation period (shown approximately by b)).
4. Use the resultant state to initialise subsequent Filter updates as observations arise, as shown by period c) onwards.
5. Return to the start of the process following an indication that the Filter is no-longer performing and the process should be re-started as shown by the threshold breaches at epoch h).

The analysis presented here is performed using synthetic data with the objective of down-selecting techniques for operational assessment with real data and the new concerns that real data brings. The absolute magnitude of effects detected in this analysis are not necessarily indicative for all cases and will depend on the specific force model and sensor configuration used. It is however thought that the relative performance of statistics is indicative, particularly at the thresholds of effect detection.

The application of various formulations of Kalman Filters as a scalable technique to continually and rapidly update the precision orbital trajectories of RSOs is well established [2] as is the need to add process noise to both control the Kalman Filter [2] and to present realistic predicted covariance [3]. Throughout this analysis we employ a Square Root Unscented Kalman Filter [4] with a process noise model described in section 1.2.1.

1.2.1 Process Noise Model

In order to appropriately model the state solution uncertainty a process noise matrix $Q(t)$ is applied prior to each observation. Our formulation is similar to [3, 5], but differs in some important details. In the interest of space we do not describe the full non-linear noise accumulation and integration model, but to a linear approximation it is equivalent to:

$$Q^{\Delta}(t_0) = \mathbf{0} \tag{1}$$

$$Q(t_n) = \sum_{k=1}^{n} \Phi(t_n, t_k) Q^{\Delta}(\Delta t_{k-1}) \Phi(t_n, t_k)^T \tag{2}$$

Because the time step sizes are variable it is essential to use a model for $Q^{\Delta}(\Delta t)$ that displays independence from the size of the time step (provided it is small enough). We found that an effective approach, consistent with a Markov assumption for random along track accelerations, is to include only a linear growth in along track velocity, that is in the Velocity direction in the Transverse-Velocity-Normal (TVN) frame at the end of the Δt and to set a maximum Δt. We do not add the higher order terms in Δt to the position or position-velocity cross correlation based upon a free space integration as we have not verified that it satisfies the time-step independence properties described above.

In the experiments here, we have tuned the noise growth rate to provide a good quality estimate that also provides consistent long time (several day) covariance growth. The process for achieving this is not described in detail but is similar to [5] except that we tune directly against posterior observations as truth rather than a smoothed posterior Kalman solution and aim to achieve homoskedacity in the Mahalanobis distance of the prediction as it is propagated further ahead. In our model we do not suffer the sensitivity to the horizon over which a prediction was required seen in [6].

2. TRUTH AND ESTIMATION MODELS

In this report each truth trajectory was generated with a special perturbations propagator with additional accelerations not included in the estimator models in order to emulate the force model disparity experienced in an operational system against live data.

2.1 Astrodynamics model

The Low Earth force model included the Earth gravity field with 20 degrees of spherical harmonics and a Harris-Priester atmospheric profile, in addition to which the truth model alone has a 2.5e-8ms^{-2} amplitude periodic acceleration in the tangential direction, with a period of a single orbit.

The Deep Space force model included the Earth gravity field with 10 degrees of spherical harmonics and a solar radiation pressure force with constant solar flux, in addition to which the truth model alone has a 2.5e-8ms^{-2} amplitude periodic acceleration in the direction of the sun, with period of a single orbit.

The LEO and IGSO test satellites were initialised with the orbital elements as defined in Table 1. We selected IGSO over GEO as it worked with existing observation pass detection.

The estimation model sets the extended state parameter (drag coefficient in LEO, solar radiation coefficient in IGSO) as a least squares free parameter over the Least Square Estimation Period and this remains fixed for the sequential updates.

Component	LEO	IGSO
Apogee (km)	802	35,594
Perigee (km)	812	35,900
Inclination (°)	98.9	60.1
Semi-Major Axis (km)	7,178	42,150

Table 1: A table showing the orbital components of the test satellites in LEO and IGSO

The process noise was tuned to be be of the order 1e-12m^2s^{-3} in the LEO case, and 1e-11m^2s^{-3} in the IGSO case.

Both of these values are consistent with process noises observed in operational contexts, which is how the size of the non-modelled force was selected.

2.2 Observation Model

The observations are generated from the truth trajectory by selecting a typical coverage of sensor sites for a given RSO and observing the hypothetical observed values from these sites. The observations are generated with random Gaussian errors with archetypal sensor precision as shown in table 2. The LEO scenario generates Angular Azimuth/Elevation and Range observations, in IGSO only Angular observations are generated since Range observation availability in IGSO is rare. The daytime limitations in IGSO are not simulated.

Regime	Angle 1 σ ((°))	Angle 2 σ ((°))	Range σ (m)	Lat ((°) +N)	Long ((°) +W)
LEO	0.004	0.005	2	9	167
	0.07	0.02	7	-22	-114
	0.02	0.02	6	54	-359
	0.002	0.002	2	43	-290
IGSO	0.0005	0.0004	-	-7	-72
	0.0002	0.0001	-	37	-354

Table 2: A table showing the precision and position of sensors used in this study

3. DETECTION OF ANOMALIES

3.1 Metric of Anomaly

To detect anomalous observations or observations of anomalous tracks we define an Anomaly Metric Ψ_D as a square Mahalanobis distance obtained from elements necessary for the computation of the Kalman Update step which imposes very low additional computational cost. The metric Ψ_D is defined as a scalar obtained from the D dimensional expression of the general equation (3) where D is the dimension of the observation .

$$\Psi_D = (\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu}) \tag{3}$$

The covariance S is the sum of the observation covariance and the projection of the predicted state covariance into the D dimensional observation space, $\vec{\mu}$ is the projection of the predicted state vector into the observation space and $\vec{x}$ is the observed measurement.

It is possible that observation sequences for an RSO could be of differing dimensionality, and it is further possible to choose a reduced observation space. In current operation, range rate observations are integrated in a separate step, so the natural metric for us to use is position only. For the LEO simulations considered here the observations were 3D, and for IGSO they were 2D.

3.2 Statistical Properties of the Anomaly Metric

For Gaussian random errors, the Anomaly Metric Ψ_D should conform to a χ^2_D distribution, where the degree D is the dimension of the observation. Henceforth we will generally elide the D subscript.

There are numerous objections to expecting conformance to a χ^2 distribution.

- Observations from a single site in rapid succession will posses strongly correlated errors due to the trajectory error. This might reasonably wash out over many different observations, but convergence towards a theoretical distribution could be slow;
- Slowly varying observational errors from a single sensor may cause a similar concern;
- Poor quality assessment of the orbit uncertainty will impact the metric, for example an excessively pessimistic state covariance will shrink the metric, most notably for the first observation in a sequence due to the longer time over which the noise is integrated.

Acknowledging these concerns, we perform several different statistical analyses on the Anomaly Metric data and contrast them with a straightforward Cramer-von-Mises χ^2 test.

3.3 Alternative Measures

Vasquez *et al* [7] considered a closely related measure to detect manoeuvres in LEO which they refereed to as Algorithm 1. They applied the same metric from the Kalman Update, but prior to performing the Kalman Update they interpolated all of the measurements in a pass into a single combined value which they call an attributable.

The process of combining data into an attributable reduces observational noise, but it is less clear what the effect is that it has on the discrimination of the state uncertainty throughout the pass or on the statistical behaviour of the metric. For LEO the track transits the sky quickly and the state uncertainty may explore a range of orientations with respect to the observer. In Geostationary orbit on the other hand, the geometry is largely fixed and there isn't a short period over which to accumulate the attributable, so bounds would need to be defined.

Out results should be present and interesting comparison with Algorithm 1 as a simpler less compute intensive option, that also translates readily to Geosynchronous orbits.

4. IMPULSE DETECTION

The criteria for anomaly handling defined here are intended to be incorporated into operational processes for automated handling of both observation anomalies, and anomalies due to RSO manoeuvre.

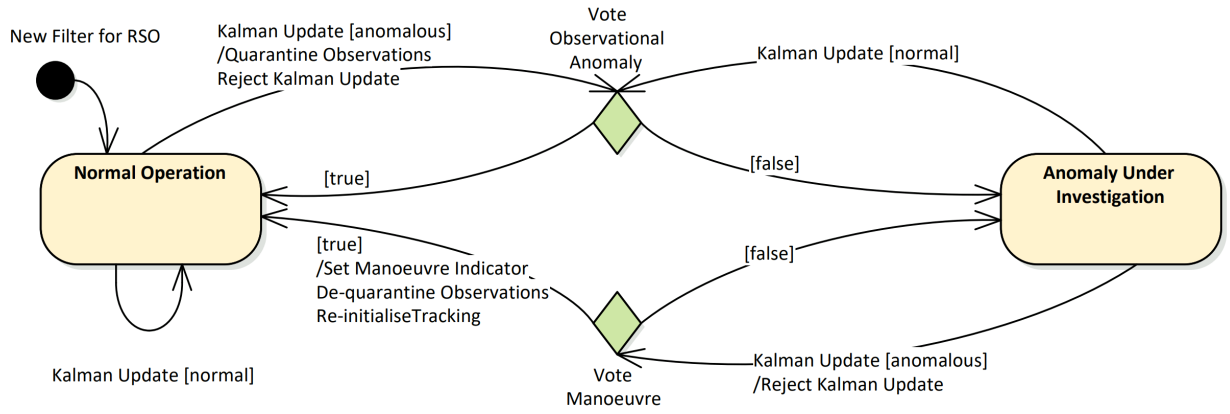

Fig. 2: Anomaly Investigation State Machine Diagram detailing the process of distinguishing changes in state and observational anomalies.

The operational process for applying the criteria defined relies on voting across two or more observed anomalies when they cannot be immediately rejected as an observational error:

1. Perform the Kalman Update for a supplied tracking data message, extracting the anomaly metric for each observation.
2. If the metric violates the criteria identified here then check if the RSO currently has an outstanding anomaly indicator:
   - If there are no ongoing anomaly indicators:
     - A putative Initial Orbit Determination may optionally be performed from the new tracking data, which could be used to provide additional alternative pointing, or potentially to identify an anomaly immediately,
     - if there remains doubt over the anomaly cause, the tracking data is quarantined and linked to a new active anomaly indicator and the state estimate reverted to its preceding value,
     - if the anomaly is rejected immediately as observational error, it is immediately invalidated and no anomaly indicator is set.

- If there is a currently active anomaly indicator for the RSO, then a decision is made to establish if this represents enough evidence for a manoeuvre.
  - if so, a manoeuvre event is indicated, the recent sequence of tracking data messages are de-quarantined and the Estimation process is reset from the first anomalous observation,
  - if not the observation is again quarantined and the filter result rejected as in the prior step.

3. Following an Kalman update which does not present anomalous results, any ongoing anomaly indicators are analysed to establish if there is enough evidence to close the indicator as having been triggered by an observational anomaly, which is now safely quarantined. The indicator may not be closed immediately to account for observability limitations.

Broadly speaking two subsequent bad tracking data messages indicate a manoeuvre, but some caution may be applied if the results are correlated such as originating from the same sensor site.

The remainder of this section analyses the effectiveness of four statistical measures on the metric of anomaly for the detection of impulse manoeuvres. We do not present here single anomaly results as for practical anomalies that we tried, all approaches identified the anomaly and the above procedure would exclude it without error. So the focus is on detecting the manoeuvre. We note that if an observation is rejected as potentially anomalous, the Kalman Filter will require a longer time update step from the last accepted state to the next observation than usual. We allow in the analysis for a wide range of lead-in times and we also look at detection after multiple passes.

4.1 Statistical Measure of Anomaly

For trajectory anomalies caused by an impulse, the effect is often quite large and even a rough indicator is useful. Acknowledging concerns around the statistical properties of the Anomaly Metric, we anticipate that it is somewhat close to an appropriate χ^2 distribution, and therefore deviations from this could provide a well normalised measure of anomaly.

The baseline test is therefore to individually assess the likelihood that each observation was drawn from the appropriate theoretical χ^2 distribution using the Cramer-von-Mises test. We are particularly interested in how small an impulse can be reliability detected in our synthetic scenarios, and how robust the approach is to poorly tuned process noise; operationally it can be difficult to obtain optimal tuning, due to lack of processing time, available data or indeed the operating conditions of the manoeuvrable RSO.

Recognising that poor process noise tuning impacts the distribution of the anomaly metric, we also introduce tests against the Empirical Distribution Function (EDF) of the Anomaly Metric which is obtained over the prototypical batch period. To this end, three tests of conformance to the EDF are trialled:

- Anderson-Darling two-sample test (AD),
- Cramer-Von-Mises two sample test (CvM),
- Kolmogorov-Smirnov test (KS).

4.2 Effectiveness Tests

We report both the the true and false positive results and a combined statistic called the Matthews Correlation Coefficient (MCC) which is described in 7.

The MCC measures correlation of outcome with truth, where 1.0 is perfect correlation, and -1.0 is perfect anti-correlation. This statistic was chosen because it is considered [8] robust to the variable populations of true and false tests. Nominally the result with the best MCC provides the best result, however, it treats false positives and false negatives equally.

Whilst we would caution against choosing measures of quality after observing the result, we acknowledge that in different operational circumstances, false positive and false negatives may carry different weight.

In each test we define a tolerance to false acceptance of the null hypothesis (which is that there was no anomaly), we trial two tolerances, 1e-3 and 1e-4. We cannot reasonably consider smaller tolerances for the amount of data we have, and since we anticipate this to be an upper bound on performance, we wanted to avoid greater than 0.1% decision error to be baked in.

	Tolerance	1e-3			1e-4		
Detection Period (n)	Statistic/Noise	1.00E-13	1.00E-12	1.00E-11	1.00E-13	1.00E-12	1.00E-11
1 Pass	CvM χ^2	0.13	0.56	0.60	0.11	0.65	0.50
	Anderson-Darling	0.14	0.54	0.57	0.15	0.66	0.59
	CvM 2-Sample	0.13	0.65	0.48	0.18	0.57	0.45
	Kolmogorov-Smirnov	0.14	0.60	0.45	0.18	0.61	0.40
4 Passes	CvM χ^2	0.15	0.80	0.81	0.18	0.84	0.81
	Anderson-Darling	0.17	0.78	0.78	0.21	0.85	0.77
	CvM 2-Sample	0.21	0.84	0.75	0.30	0.84	0.67
	Kolmogorov-Smirnov	0.19	0.76	0.73	0.26	0.84	0.67

Table 4: A table showing the Matthews Correlation Coefficient (MCC) of tests performed on the n passes after a $1cms^{-1}$ impulse manoeuvre has been performed. Note that a single detection in 4 passes would count as a positive detection in the n = 4 case. The best process noise tuning is 1e-12.

4.3 LEO Results

The focus on the analysis is detecting anomalies due to impulse manoeuvres of 0.1, 1 and 10 cm/s ΔV in LEO covering the range from unusually small to unusually large which compare to the scenarios in [7] of 0.5, 3 and 12 cm/s. We did not find analysing simulated anomalies to be very useful, as the scale of anomaly was somewhat arbitrary. Instead we focus on observably outlying Kalman updates and the voting principle described above to validate a manoeuvre.

The manoeuvre could occur at any time before the first post-manoeuvre pass, so in some cases it is reasonable to expect the effect of the manoeuvre not to accrued enough to be detected at the first pass so we look at 1, 2 and 4 passes after the manoeuvre to see if at least one indication of anomaly has been effected. The result of 20 scenarios for an optimal process noise tuning of 1e-12 are summarised in table 3. It can be seen that all the statistics performed similarly well for a detection threshold of 1e4 with the CvM χ^2 reference statistic performing marginally better and the Anderson-Darling statistic the best of the rest.

Tolerance	Stat	First 1 Pass Detected	First 2 Pass Detected	First 4 Pass Detected
$10cms^{-1}$	CvM χ^2	1.0	1.0	1.0
	Anderson-Darling	1.0	1.0	1.0
	CvM 2-Sample	1.0	1.0	1.0
	Kolmogorov-Smirnov	1.0	1.0	1.0
$1cms^{-1}$	CvM χ^2	0.6	0.7	0.9
	Anderson-Darling	0.55	0.65	0.85
	CvM 2-Sample	0.4	0.6	0.8
	Kolmogorov-Smirnov	0.45	0.65	0.8
$0.1cms^{-1}$	CvM χ^2	0.05	0.1	0.1
	Anderson-Darling	0.05	0.1	0.1
	CvM 2-Sample	0	0.05	0.05
	Kolmogorov-Smirnov	0	0.05	0.05

Table 3: A table showing the 1, 2 and 4 pass detection rates of differing scales of tangential impulse. 20 Individual test cases for each thrust scale. All cases use a detection threshold of 1e-4.

The result contrasts with [7] where they reported difficulty observing the manoeuvre for anything other than 12 cm/s, but we caution that their process noise tuning was considerably higher. As explained in the introduction, process noise we use is consistent with operational values.

Because of the challenge of operationally ensuring a perfectly tuned filter, it is important to assess the robustness of the statistic to poor tuning. To this end we consider sub-optimal process noise growth rates, and report the results in Table 4. All the tests perform similarly: they are all broadly robust to excess noise, with only slight reduction in correlation for 1e-4 tolerance, and they all perform very badly with insufficient noise.

To unpick the correlation further, we show the false positive rates in Table 5. The reduction in truth correlation with

insufficient noise is a result of a significant increase in false positive rates.

We have observed that with such low noise the Anomaly metric is stable with very little change for many updates, until relatively rare events, possibly driven by the geometry of the observation, which then cause new reduction in the covariance and thus a step increase in the Anomaly Metric. We might have hoped that applying the EDF would resolve this, but it did not. Whether a larger prototypical sample would improve this, or whether the covariance is fundamentally heteroskedastic is hard to say.

The reduction in sensitivity of the test with excess noise is reasonable, as anomalies are more likely to be lost in the noise, noting the de-correlation from truth is due to an increase in false negatives.

Tolerance	1e-3			1e-4		
Statistic/Noise	1.00E-13	1.00E-12	1.00E-11	1.00E-13	1.00E-12	1.00E-11
CvM χ^2	1.0E-1	4.0E-3	1.5E-4	7.8E-2	7.5E-4	0.0
Anderson-Darling	9.0E-2	2.0E-3	3.0E-4	6.0E-2	4.5E-4	0.0
CvM 2-Sample	6.0E-2	7.5E-4	3.0E-4	2.8E-2	3.0E-4	0.0
Kolmogorov-Smirnov	7.0E-2	1.0E-3	4.5E-4	4.0E-2	3.0E-4	1.5E-4

Table 5: A table showing the false positive rate of the test statistics for single pass sample trials. The rates are calculated over 6698 tests in LEO over 40 3-week clear Kalman estimation periods.

4.4 IGSO Results

The same analysis was performed for the IGSO simulation with the same overall outcome. Only a few results are shown to save space.

Thrust Magnitude	Stat	First 1 Pass Detected	First 2 Pass Detected	First 4 Pass Detected
10cm	CvM χ^2	0.60	1.00	1.00
	Anderson-Darling	0.60	1.00	1.00
	CvM 2-Sample	0.55	1.00	1.00
	Kolmogorov-Smirnov	0.55	1.00	1.00
1cm	CvM χ^2	0.00	0.25	0.35
	Anderson-Darling	0.00	0.35	0.45
	CvM 2-Sample	0.00	0.10	0.20
	Kolmogorov-Smirnov	0.00	0.10	0.20

Table 6: A table showing the 1, 2 and 4 pass detection of differing scales of tangential impulse of the IGSO satellite. 20 Individual test cases for each thrust scale. All cases use a detection threshold of 1e-4.

4.5 Conclusion for Impulse Detection

In conclusion the simple Cramer-von-Mises test against a theoretical χ^2 distribution performed as well or better than empirical distribution function tests, and provided clear indicator of impulse of practical size in both LEO and IGSO regimes.

The result is robust to one order of magnitude excess process noise growth rate, but produces many false positives if the process noise is too small. So much so that a high false positive rate in practice may be a suitable indication that process noise requires increasing.

This result is considerably more positive in the ability to discriminate practical impulse manoeuvres in LEO than Vasquez [7], though it is not clear if that is due to the algorithm they employ, or the need for greater process noise in that scenario.

5. DETECTION OF PATTERN OF LIFE CHANGES

In this section we analyse the performance of the same Anomaly Metric in detecting subtle but enduring changes to the pattern of life of an RSO by gathering data over longer periods to support more refined statistics.

The operational process trialled here is to evaluate the prototypical behaviour during the batch initialisation procedure, and use this prototypical behaviour as a baseline to assess the output of the sequential updates for consistency. Obtaining the baseline in this way is adopted for simplicity and fits well with current operational data gathering structures; however, a library of prototypical baselines could instead be established from historical data. It may be possible to compare to multiple prototypical baselines, but we caution that the tests we perform are binary hypothesis tests seeking to safely reject the hypothesis that nothing has changed. With multiple baselines that are too similar, many of them may be accepted at the same time.

For the analysis, we focus on extremely small continuous thrusts ranging from 5e-8ms^{-2} (5μN for a 100kg object) to 1e-6ms^{-2}, but same result is anticipated to read across to realistic operational changes to attitude or continual but not continuous non-modelled thrusts.

5.1 Analysis Approach

We consider a rolling window of multiple passes with the objective of accruing more evidence of change; we present most results for windows of 1, 4 and 8 passes, with a single extension point to 12 passes. The reason analyse longer periods is to understand if it offers an operational trade-off that while gathering information over longer periods may expose subtler effects, there is also a greater risk that the pattern being sought changes again within that period.

We trial the same measures of conformity to statistical distributions as assessed in the impulse analysis. In addition, we also consider tests of the mean and variance of the anomaly metric between the sliding window for the ongoing filter and the baseline sample. The mean and variance tests employ the Bootstrap technique described in 7 which employ re-sampling to extract closer to normally distributed statistics.

For each analysis technique the true positive rates are assessed for Early and Regular Detection, which combined with false positive rates support the truth correlation analysis with the MCC metric. The detection regimes we define are:

- Early Detection: True Positive within the first 2-days after anomaly start,
- Regular Detection: True Positive 2-4 days after anomaly start.

The number of passes per day is quite variable because we select passes randomly, but there are on average between 5 and 6 passes per day. For the Early Detection, many samples overlapped the onset of thrust, whereas for the Regular Detection Period only the 12 pass case had any sliding windows overlapping the onset of thrust.

5.2 Continuous Manoeuvre Scenarios

The results presented here are established over a number of scenarios, as elaborated in Table 7.

				Trial Counts			
Regime	Manoeuvre	Magnitude	Scenario Count	0-2 Day	2-4 Day	4-6 Day	Negative
LEO	Large	1e-6ms^{-2}	4	37	37	-	-
	Medium	1e-7ms^{-2}	20	195	196	182	-
	Small	5e-8ms^{-2}	10	93	102	-	-
	Clear	0.0	40	-	-	-	800
IGSO	Large	1e-6ms^{-2}	4	24	24	-	-
	Small	1e-7ms^{-2}	10	59	60	-	-

Table 7: A table showing the manoeuvre scenarios performed and volume of observation passes during the thrust periods. All manoeuvres are performed in the Velocity direction of the TVN frame.

5.3 Pattern of Life Detection in LEO

5.3.1 Detection Performance

The results are expressed in terms of the MCC truth correlation which show that for both the small (Table 8) and medium (Table 9) continuous thrusts, single pass analysis performs badly, but by including 4 or 8 passes the Bootstrap Variance performs well and achieves the highest detection rate for all tolerance levels. The χ^2 test performs well in this case, but we shall see later it has other shortfalls.

For the largest scale of thrust (Table 10) nearly all indicators achieve near perfect correlation rates in the Regular Detection Period. For early detection the impact was large enough to identify with the distribution conformance test in a single pass, so longer sliding windows harmed the result due to straddling the onset of thrust. Moreover the benefits of bootstrapping over longer windows don't quite overcome these disadvantages.

This conclusion from these tests alone would be that to detect subtle persistent effects the Bootstrap Variance performs better than the distribution conformance tests, but that for larger effects, the impulse detection technique provides quicker results.

Unpacking the correlation coefficient we find that the Bootstrap method presents a more noticeable improvement in the detection rate, but slightly higher false positive rates temper the overall correlation. False positive rates are inspected in greater detail in section 5.5 where sub-optimal process noise tuning is assessed.

	Samples	8-Pass		4-Pass		1-Pass	
	Stat	Day 0-2	Day 2-4	Day 0-2	Day 2-4	Day 0-2	Day 2-4
1e-3	CvM χ^2	0.42	0.75	0.34	0.49	0.20	0.21
	Anderson-Darling	0.26	0.49	0.31	0.38	0.24	0.32
	CvM 2-Sample	0.20	0.45	0.22	0.31	0.16	0.20
	Kolmogorov-Smirnov	0.37	0.56	0.39	0.37	0.22	0.21
	Bootstrap Variance	0.60	0.81	0.50	0.52	0.09	0.12
	Bootstrap T-test	0.56	0.77	0.39	0.39	0.00	0.00
1e-4	CvM χ^2	0.36	0.63	0.27	0.39	0.11	0.10
	Anderson-Darling	0.14	0.38	0.25	0.29	0.17	0.15
	CvM 2-Sample	0.14	0.33	0.18	0.17	-0.00	-0.00
	Kolmogorov-Smirnov	0.28	0.34	0.25	0.29	0.26	0.13
	Bootstrap Variance	0.51	0.75	0.49	0.44	0.07	0.12
	Bootstrap T-test	0.44	0.60	0.29	0.26	0.00	0.00

Table 8: 0-2 and 2-4 day MCC Results for a 4-day 5e-8ms^{-2} continuous thrust manoeuvre with a baseline sample of 10 passes

Tolerance	Samples	8-Pass		4-Pass		1-Pass	
	Stat	Day 0-2	Day 2-4	Day 0-2	Day 2-4	Day 0-2	Day 2-4
1e-3	CvM χ^2	0.71	0.99	0.71	0.94	0.47	0.55
	Anderson-Darling	0.56	0.96	0.62	0.86	0.48	0.56
	CvM 2-Sample	0.487	0.92	0.54	0.80	0.35	0.51
	Kolmogorov-Smirnov	0.54	0.93	0.61	0.81	0.42	0.50
	Bootstrap Variance	0.73	0.96	0.74	0.88	0.32	0.24
	Bootstrap T-test	0.64	0.95	0.52	0.62	0.07	0.20
1e-4	CvM χ^2	0.66	0.99	0.67	0.90	0.43	0.50
	Anderson-Darling	0.43	0.88	0.54	0.77	0.46	0.52
	CvM 2-Sample	0.38	0.83	0.43	0.68	0.29	0.39
	Kolmogorov-Smirnov	0.42	0.85	0.45	0.67	0.37	0.45
	Bootstrap Variance	0.71	0.95	0.71	0.83	0.26	0.22
	Bootstrap T-test	0.55	0.81	0.31	0.46	0.07	0.00

Table 9: 0-2 and 2-4 day MCC Results for a 4-day 1e-7ms^{-2} continuous thrust manoeuvre with a baseline sample of 10 passes

Tolerance	Samples	8-Pass		4-Pass		1-Pass	
	Stat	Day 0-2	Day 2-4	Day 0-2	Day 2-4	Day 0-2	Day 2-4
1e-3	CvM χ^2	0.87	0.96	0.85	0.93	0.83	0.88
	Anderson-Darling	0.82	0.97	0.87	0.96	0.81	0.86
	CvM 2-Sample	0.83	0.99	0.87	0.96	0.89	0.92
	Kolmogorov-Smirnov	0.87	1.00	0.93	1.00	0.86	0.91
	Bootstrap Variance	0.85	0.86	0.88	0.89	0.36	0.40
	Bootstrap T-test	0.61	0.91	0.54	0.51	0.28	0.00
1e-4	CvM χ^2	0.87	0.99	0.91	1.00	0.93	0.94
	Anderson-Darling	0.83	1.00	0.90	0.99	0.96	0.96
	CvM 2-Sample	0.80	1.00	0.87	0.99	0.96	0.92
	Kolmogorov-Smirnov	0.82	1.00	0.91	1.00	0.97	0.96
	Bootstrap Variance	0.87	0.94	0.90	0.96	0.41	0.42
	Bootstrap T-test	0.52	0.79	0.49	0.40	0.00	0.00

Table 10: 0-2 and 2-4 day MCC results for a 4-day 1e-6ms^{-2} continuous thrust manoeuvre with a baseline sample of 10 passes

5.4 Pattern of Life Detection in IGSO

When considering angular observation only scenarios in IGSO, the same conclusion with respect to statistic choice hold. The correlation performance of the Bootstrap variance in the 1e-7ms^{-2} case yields a significant improvement. There is limited difference between methods in the overall detection when the effect is sufficiently large, although bootstrapping has a small advantage when the test sample only partially includes artifacts of the thrusting effect as is the case in LEO.

Tolerance	Stat	Day 0-2	Day 2-4
1e-3	CvM χ^2	0.16	0.74
	Anderson-Darling	0.16	0.67
	Bootstrap Variance	0.28	0.84
1e-4	CvM χ^2	0.16	0.64
	Anderson-Darling	0.00	0.60
	Bootstrap Variance	0.22	0.80

Table 11: 0-2 and 2-4 day MCC Results for a 4-day 1e-7ms^{-2} continuous thrust manoeuvre with a baseline sample of 10 passes in IGSO.

Tolerance	Stat	Day 0-2	Day 2-4
1e-3	CvM χ^2	0.74	1.00
	Anderson-Darling	0.74	1.00
	Bootstrap Variance	0.85	0.95
1e-4	CvM χ^2	0.68	1.00
	Anderson-Darling	0.68	1.00
	Bootstrap Variance	0.77	1.00

Table 12: 0-2 and 2-4 day MCC Results for a 4-day 1e-6ms^{-2} continuous thrust manoeuvre with a baseline sample of 10 passes in IGSO.

5.5 Robustness to Compromised Process Noise Tuning

To analyse the robustness of each technique to compromised process noise tuning, a medium thrust was considered, and noise values were assessed around the near-optimal process noise tuning rate, which for this force case is 1e-12m^2s^{-3}. We show in Table 13 that with excess noise the sensitivity drops, whereas for too little noise the detection

rate is high. However, as can be seen in Table 14 the truth correlation is hampered by an increase in false positive results. Overall the Table 14 shows:

- Whilst the χ^2 statistic performs very well for optimal tuning, it tails off in precision for both excess (low sensitivity) and insufficient noise (high specificity), calling in to question its viability for operational use.
- The Empirical distribution functions correlate best with truth for insufficient noise in LEO, so a viable approach might be to use this technique with slightly low noise, and to turn up the noise very carefully to reduce false positives. However the risk of tailing off sensitivity due to excess noise is very notable.
- The Bootstrap variance statistic performs consistently well at excess noise and the reduction in correlation at low noise is cause by an increased number of false positives. In this case, tuning the noise up slightly to reduce false positives has less of an impact on the pattern of life detection.

	Process Noise Rate	1e-13		1e-12		1e-11	
Tolerance	Stat	Day 0-2	Day 2-4	Day 0-2	Day 2-4	Day 0-2	Day 2-4
1e-3	CvM χ^2	0.88	1.00	0.51	0.93	0.00	0.05
	Anderson-Darling	0.67	1.00	0.33	0.81	0.00	0.00
	Bootstrap Variance	0.85	1.00	0.65	1.00	0.16	0.39
1e-4	CvM χ^2	0.79	1.00	0.39	0.89	0.00	0.00
	Anderson-Darling	0.62	1.00	0.25	0.71	0.00	0.00
	Bootstrap Variance	0.80	1.00	0.57	1.00	0.07	0.27

Table 13: A table showing the detection rate of a 1e-7ms^{-2} continuous thrust at a range of different tunings using 8 sample passes.

Period	Tolerance	Statistic/Noise	1.00E-13	3.20E-13	**1.00E-12**	3.20E-12	1.00E-11
0-2 Day	1e-3	CvM χ^2	0.41	0.66	0.71	0.39	0.09
		Anderson-Darling	0.57	0.67	0.56	0.20	-0.03
		Bootstrap Variance	0.44	0.64	0.73	0.62	0.40
	1e-4	CvM χ^2	0.47	0.83	0.66	0.28	0.00
		Anderson-Darling	0.64	0.66	0.43	0.11	0.00
		Bootstrap Variance	0.48	0.68	0.71	0.56	0.26
2-4 Day	1e-3	CvM χ^2	0.41	0.77	0.99	0.84	0.18
		Anderson-Darling	0.71	0.96	0.96	0.43	0.04
		Bootstrap Variance	0.51	0.79	0.96	0.88	0.65
	1e-4	CvM χ^2	0.50	0.99	0.99	0.65	0.00
		Anderson-Darling	0.83	0.98	0.88	0.29	0.00
		Bootstrap Variance	0.60	0.87	0.95	0.81	0.43

Table 14: A table showing the Matthews Correlation Coefficient (MCC) for the detection of a 4-day 1e-7ms^{-1} continuous thrust for a range of tuning. 8 Sample passes are used throughout.

Period	Tolerance	Statistic/Noise	1.00E-13	3.20E-13	**1.00E-12**	3.20E-12	1.00E-11
False Positive Rate	1e-3	CvM χ^2	4.9E-01	1.2E-01	3.8E-03	0.0E+00	1.3E-03
		Anderson-Darling	1.6E-01	1.5E-02	2.5E-03	1.3E-03	3.8E-03
		Bootstrap Variance	3.5E-01	1.1E-01	1.5E-02	5.0E-03	2.5E-03
	1e-4	CvM χ^2	3.7E-01	3.8E-03	1.3E-03	0.0E+00	0.0E+00
		Anderson-Darling	8.3E-02	6.3E-03	0.0E+00	0.0E+00	0.0E+00
		Bootstrap Variance	2.7E-01	5.9E-02	6.3E-03	0.0E+00	0.0E+00

Table 15: A table showing the false positive rate of the test statistics for a range of process noise tuning in LEO. 8 Sample passes are used throughout.

Period	Tolerance	Statistic/Noise	1.00E-12	**1.00E-11**	1.00E-10
0-2 Day	1e-3	CvM χ^2	0.33	0.16	-0.02
		Anderson-Darling	0.40	0.16	0.00
		Bootstrap Variance	0.57	0.28	-0.05
	1e-4	CvM χ^2	0.38	0.16	-0.05
		Anderson-Darling	0.35	0.00	0.00
		Bootstrap Variance	0.49	0.22	0.00
2-4 Day	1e-3	CvM χ^2	0.75	0.74	-0.08
		Anderson-Darling	0.90	0.67	0.00
		Bootstrap Variance	0.89	0.84	0.10
	1e-4	CvM χ^2	0.84	0.64	-0.05
		Anderson-Darling	0.92	0.60	0.00
		Bootstrap Variance	0.92	0.80	0.00

Table 16: A table showing the Matthews Correlation Coefficient (MCC) for the detection of a 4-day 1e-7ms^{-1} continuous thrust performed in IGSO for a range of tuning. 8 Sample passes are used throughout.

Period	Tolerance	Statistic/Noise	1.00E-12	**1.00E-11**	1.00E-10
False-Positive	1e-3	CvM χ^2	1.9E-01	0.0E+00	2.2E-02
		Anderson-Darling	5.8E-02	0.0E+00	0.0E+00
		Bootstrap Variance	7.2E-02	1.4E-02	7.2E-03
	1e-4	CvM χ^2	1.2E-01	0.0E+00	7.2E-03
		Anderson-Darling	4.3E-02	0.0E+00	0.0E+00
		Bootstrap Variance	5.0E-02	0.0E+00	0.0E+00

Table 17: A table showing the false positive rate of the test statistics for a range of process noise tunings in IGSO

5.6 Longer Detection Periods

To explore whether the Bootstrap Variance technique can provide more sensitive thrust detection by considering longer passes, we compare 8 and 12 pass assessments for a variety of thrust magnitudes and process noise tuning in Table 18. About 90% of the 12 pass tests had the entire 12 pass window after the thrust started.

It is clear that increasing the number of passes from 8 to 12 increases the truth correlation for the regular detection regime. What is perhaps more surprising is that this increase in sample size causes a positive or neutral change for early detection, during which the majority of sample periods will not be encompassed by the pattern of life effect.

Thrust Accelerationms^{-2}	Process Noise	Sample Period	0-2 Day Rate	2-4 Day Rate
1e-7	1e-11	8 Pass	0.08	0.37
		12 Pass	0.37	0.70
1e-7	3e-12	8 Pass	0.46	0.83
		12 Pass	0.44	0.94
5e-8	1e-12	8 Pass	0.33	0.64
		12 Pass	0.37	0.85

Table 18: A table showing the increase in true positive rate for estimation of a sample of 4-day continuous thrust periods when including a higher number of samples in the variance Bootstrap. Detection tolerance is 1e-4 for all cases.

5.7 Baseline Kalman During Thrust Period

We explore a pattern of life change that is transitioning from an already poorly modelled scenario. This initial period will be used as our EDF and Bootstrap baseline to detect change. Clearly the χ^2 test will not perform well as it implicitly has a perfectly modelled baseline.

As a simple test of this, we obtain the baseline pattern over a period with a baseline non-modelled thrust, and then test against that baseline for a period with a change of thrust. We report the MCC in Table 5.7. For the Bootstrap test we extend the test to detect both increase and decrease in variance. The Empirical two sample tests and the Bootstrap Variance tests perform similarly well.

Baseline Thrust ms^{-2}	Test Thrust ms^{-2}	Stat	0-2 Days	2-4 Days
1e-6	1e-7	CvM χ^2	0.00	0.00
		Anderson-Darling	0.41	0.91
		Bootstrap Variance	0.32	0.97
1e-7	1e-6	CvM χ^2	0.21	0.22
		Anderson-Darling	0.73	1.00
		Bootstrap Variance	0.79	1.00
1e-7	0.0	CvM χ^2	-0.49	-1.00
		Anderson-Darling	0.34	0.78
		Bootstrap Variance	0.34	0.97

Table 19: A table showing the MCC values for a series of baseline thrusting periods with a change in thrust magnitude in LEO.

5.8 Pattern of Life Conclusion

We show that by analysing a sliding window of the Kalman Filter performance the anomaly metric can be used for detecting small persistent non-modelled forces that would not be visible in an impulse analysis. We have shown that of the techniques assessed the Bootstrap variance technique is:

- More sensitive to smaller changes,
- More robust to sub-optimal process noise,
- Increases in sensitivity if changes are assessed over longer periods,
- Able to detect changes from a baseline with non-modelled continuous thrust.

We found the Bootstrap variance test performs better than tests on the Bootstrap mean. Although not explicitly shown here, we found bootstrapping to be absolutely key to the extraction of statistical properties on which we could generate hypotheses.

Bootstrapping does incur a computational cost, if this is a concern, then the empirical distribution function tests may be more appropriate.

6. CONCLUSION AND FUTURE WORK

This report has shown that it is possible to detect both impulsive manoeuvres and continuous thrust by applying sample analysis techniques on information available while tracking an object using a Kalman Filter. This Anomaly Metric measures the square Mahalanobis distance of a predicted from an observed sensor measurement accounting for both predicted state and observational anomaly which is information that is readily available at the Kalman Update step.

We have shown that a simple χ^2 provides a reliable, simple and efficient means to detect impulses, but that it is less effective at detecting pattern of life changes due to poor stability to maladjusted process noise and being limited to detecting the onset of thrust.

We have shown that by using the Bootstrapping technique to evaluate and compare the variance of the Anomaly Metric over a sliding window and the prototypical baseline, we can detect small but persistent changes in non-modelled continuous thrusts, and that this approach is robust to maladjustment and can detect both onset and change of thrust.

We have shown that while both effects can be extracted by analysing conformance of observed anomaly metric to and empirical distribution function established from the prototypical baseline, it not the best technique in either case. Of

the hypothesis tests we tried, Anderson-Darling performed marginally better than two-sample Cramer-von-Mises and Kolmogorov-Smirnov. We have not been able to consider whether the technique was hampered due to the limited size of the prototypical baseline.

The analysis in [7] showed that their "Algorithm 1" was unable to detect impulses of comparable size to our approach. It is not clear if this is due to less favourable process noise conditions in that case, or that the algorithm washed out the result. Our approach benefited from the extra data available by not combining the observations into an attributable, and for impulse detection, the distribution conformance hypothesis tests excelled over analysis of gross distribution properties such as the mean.

In order to progress this work prior to integrating any of these techniques into an operational catalogue management system, the same analysis should to be performed on real data.

[1] E. Grun, H.A. Zook, H. Fechtig, and R.H. Giese. *Collisional balance of the meteoritic complex*, volume 62, pages 244–272. Elsevier, 1985.
[2] D. Vallado. *Fundamentals of Astrodynamics and Applications*. 2001.
[3] M. Duncan and A. Long. *Realistic Covariance Prediction for the Earth Science Constellation*. 2007.
[4] R. Van der Merwe and E.A. Wan. *The square-root unscented Kalman filter for state and parameter-estimation*, volume 6, pages 3461–3464 vol.6. 2001.
[5] W. Zaidi and M.D. Hejduk. *Earth Observing System Covariance Realism*. 2016.
[6] J. Pawloski, M. Duncan, S. Hesar, T. Lechtenberg, and J. Wysack. *Realistic Covariance Generation for the GPM Spacecraft*. 2018.
[7] R.V. Valenzuela, J.C.S. Merino, J.M.G. Garcia, J.G. Vioque, and F.R.G. Jimenez. *Manoeuvre detection for near-orbiting objects*. 2021.
[8] D. Chicco and G. Jurman. *The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation*. 2020.
[9] R. Davidson and J. G. MacKinnon. *The Size Distortion of Bootstrap Tests*, volume 15, pages 361–376. Cambridge University Press, 1999.
[10] R. Beran. *Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements*, volume 83, pages 687–697. American Statistical Association, Taylor Francis, Ltd., 1988.
[11] D. Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. 1994.
[12] K.R. Malkareddy and R. Yerukala. *Bootstrap Graphical Test for Equality of Variances*. 2011.

7. APPENDIX - BOOTSTRAPPING METHODS

In sample statistics, bootstrapping methodologies have been shown to yield improvements in discrimination over the empirical distribution methods [9] [10].

The Bootstrap method is to randomly re-sample the test samples with replacement. This allows for many new samples to be produced, and such that statistical properties of those samples converge on a normal distribution due to the central limit theorem. The two Bootstrap routines will be used in this report are the Bootstrap t-test [11] on the sample mean, and the a Bootstrap test on the sample variance due to Malkareddy and Yerukala [12].

It is important to note that the minimum p-value a Bootstrap can indicate is given by the number of Bootstrap samples used. This means that for low probability thresholds, a large number of B-Bootstrap estimates are required. For this report we have used B=10,000 throughout, which gives a minimum significance value of 1e-4.

7.1 Variance Bootstrap

For i samples, each of size j we can draw b Bootstrap samples to obtain our Bootstrap distribution:

$$\bar{y}_{ib} = \frac{1}{n}\sum_{j=1}^{n} Y_{ijb} \tag{4}$$

$$s^2_{ib} = \frac{1}{n-1}\sum_{j=1}^{n} (Y_{ijb} - \bar{y}_{ib})^2 \tag{5}$$

Which gives the variance Bootstrap distribution:

$$s_b^2 = \frac{1}{k}\sum_{i=1}^{k} s_{ib}^2 \tag{6}$$

For our B variances we can now obtain a Bootstrap variance distribution and get the percentile of our significance level (α):

$$m_1 = max(1, [\frac{\alpha}{2}(B+1)]) \tag{7}$$

$$m_2 = min([(1-\frac{\alpha}{2})(B+1)], B) \tag{8}$$

This variance confidence bounds are then the m_1^{th} and m_2^{th} values in the ordered Bootstrap distribution. If any of the k sample variances fall outside of these confidence values, they are deemed heterogenous to the significance level α. In the case of an increase in variance, a significance value can be obtained by looking at the fraction of variance points greater than the sample variance in terms of the Heaviside $\Theta()$ (step) function:

$$p_1 = \frac{1}{B}\sum_{b=1}^{B} \Theta(s_b^2 - s_{test}^2) \tag{9}$$

$$p_2 = \frac{1}{B}\sum_{b=1}^{B} \Theta(s_{clean}^2 - s_b^2) \tag{10}$$

$$p = min(p_1, p_2) \tag{11}$$

If equation 11 exceeds a significance threshold an increase in variance (with time) has been found to that confidence interval. For decreases in variance with time the same test is performed on:

$$p = min(1-p_1, 1-p_2) \tag{12}$$

7.2 Bootstrap T-test

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\sigma_x^2/N + \sigma_y^2/M}} \tag{13}$$

For each sample, produce a new set by adding the combined sample mean $\bar{z}$ such that:

$$x_i' = x_i - \bar{x_i} + \bar{z}, \qquad y_i' = y_i - \bar{y_i} + \bar{z} \tag{14}$$

Re-sampling (with replacement) x_i' and y_i' to give x_i^* and y_i^* respectively, recalculate the t statistic B times where B is the number of Bootstrap trials:

$$t^* = \frac{\bar{x^*} - \bar{y^*}}{\sqrt{\sigma_x^* 2/N + \sigma_y^* 2/M}} \tag{15}$$

From this distribution you can evaluate a significance level (p) by obtaining the number of bootstrapped t values that are greater than the baseline in equation 13 in terms of the Heaviside $\Theta()$ (step) function:

$$p = \frac{1}{B}\sum_{i=1}^{B} \Theta(t_i^* - t) \tag{16}$$