AAS 11-123

SEQUENTIAL ORBIT-ESTIMATION WITH SPARSE TRACKING

John H. Seago,* Jacob Griesbach,†
James W. Woodburn‡ and David A. Vallado§

A properly initialized sequential orbit estimator will converge to less certain state estimates whenever the density of the spacecraft tracking is relatively low; nevertheless, these estimates should remain viable as long the estimator's working assumptions are met. However, when tracking data are sparse, violations of certain working assumptions may pose special problems, such as when outliers exist. In this paper, some mitigation strategies are explored for such cases and these are contrasted with typical batch-least-squares techniques, with the observation that a sequential orbit estimator presents it own benefits whenever tracking data are sparsely distributed.

INTRODUCTION

A sequential estimator processes observations one at a time and provides an updated state estimate immediately after each observation is processed. A batch-least-squares estimator employing traditional normal equations processes an entire sequence of observations in a single procedure, and afterwards provides a singly updated estimate at a user-specified epoch which may be propagated forward or backward across time.[1, 2, 3] Each method presents certain strengths and weaknesses in the presence of "sparse" spacecraft tracking. *Sparseness* is a relative term depending upon the accuracy requirements of problem to be solved, but in the context of this discussion "sparse tracking" might be the frequency and quality of observations that results from lowered tasking of the US Space Surveillance Network (SSN) for the purposes of routinely maintaining a space-object catalog. The two areas of consideration for this discussion are outlier rejection and accurate prediction in the presence of sparse data.

OUTLYING MEASUREMENTS

Generally speaking, a discordant measurement exists whenever its measurement error is highly inconsistent with the assumptions about its expected (null) distribution. Because "error" is an offset relative to "truth", and truth is not usually known outside of a simulation, measurement residuals relative to some "best" estimate serve as the testable statistic.

---

* Astrodynamics Engineer, Analytical Graphics Inc., 220 Valley Dr., Exton, Pennsylvania, 19341-2380.
† Senior Advisory Scientist, Analytical Graphics Inc., 7150 Campus Dr., Suite 260, Colorado Springs, Colorado, 80920-6522.
‡ Chief Orbital Scientist, Analytical Graphics, Inc., 220 Valley Creek Blvd., Exton, PA, 19341-2380.
§ Senior Research Astrodynamicist, Center for Space Standards and Innovation, 7150 Campus Dr., Suite 260, Colorado Springs, Colorado, 80920-6522.

Batch least-squares estimators process all observational data using a common reference trajectory that is smoother than that from a sequential filter. It is a (perhaps common) perception that this specific property aids the batch-least-squares technique in the detection of outliers whose residuals are "considerably larger than the average value" and therefore outlier detection is relatively easier compared to a Kalman filter.[4] Yet, when identifying outliers by residual examination, "considerably larger" becomes somewhat subjective, so in practice this approach is made more objective by performing a *statistical test of hypothesis* on suspected residuals.[5, 6] In these situations, one tests whether

$$| r(t_i) | / \sigma_r > C \,, \tag{1}$$

where $| r(t_i) |$ is the magnitude of an individual measurement residual at time (t_i), σ_r is some indicator of residual uncertainty (sample or theoretical), and C is some user-specified threshold that corresponds to an improbable value for $r(t_i) / \sigma_r$, the *measurement residuals ratio*. A rejection threshold of $C = 3$ is a commonly used, this critical value intending to test a single outlier from an otherwise normally distributed population.[7]

Sequential Filter

For the sequential filter case, the *measurement residuals ratios* are the predicted measurement residuals divided by the measurement-error root-variance as computed from the filter. The following notation (due to Maybeck, 1979) specifies the definition of predicted filter residuals, or *innovations*.[8] For a linear sequential estimator such as the Kalman filter, let the measurement update equations be expressed as:

$$\mathbf{K}(t_i) = \mathbf{P}(t_i^-)\mathbf{H}^{\mathrm{T}}(t_i)\left[\mathbf{H}(t_i)\mathbf{P}(t_i^-)\mathbf{H}^{\mathrm{T}}(t_i) + \mathbf{R}(t_i)\right]^{-1} \tag{2}$$

$$\hat{\mathbf{x}}(t_i^+) = \hat{\mathbf{x}}(t_i^-) + \mathbf{K}(t_i)\left[\mathbf{z}_i - \mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-)\right] \tag{3}$$

$$\mathbf{P}(t_i^+) = \mathbf{P}(t_i^-) - \mathbf{K}(t_i)\mathbf{H}(t_i)\mathbf{P}(t_i^-) \;. \tag{4}$$

where $\hat{\mathbf{x}}(t_i^-)$ is the *a priori* state (or, state-correction) estimate array and $\mathbf{P}(t_i^-)$ is the *a priori* state-error covariance matrix, $\mathbf{K}(t_i)$ is the Kalman filter gain, $\hat{\mathbf{x}}(t_i^+)$ is the *posteriori* state estimate array and $\mathbf{P}(t_i^+)$ is the *posteriori* state-error covariance matrix updated by measurement vector $\mathbf{z}_i$ at time (t_i). The matrix $\mathbf{H}(t_i)$ is realized according to the analytical observation-state relationship at time (t_i), *e.g.*,

$$\mathbf{z}(t_i) = \mathbf{H}(t_i)\mathbf{x}(t_i) + \mathbf{v}(t_i) \;\; ; \;\; E\{\mathbf{v}(t_i)\} = \mathbf{0} \;\; , \; \mathrm{Cov}(\mathbf{v}(t_i)) = E\left\{\mathbf{v}(t_i)\mathbf{v}^{\mathrm{T}}(t_j)\right\} = \begin{cases} \mathbf{R}(t_i); & t_i = t_j \\ \mathbf{0}; & t_i \neq t_j \end{cases} \tag{5}$$

where $\mathbf{v}(t_i)$ is an array of white measurement noise having an expected value of zero and variance $\mathbf{R}(t_i)$. The array of predicted filter residuals is defined as the difference between the actual measurement and the best prediction of the measurement just before it is actually taken:

$$\mathbf{r}(t_i^-) = \mathbf{z}_i - \mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-) \;\; , \tag{6}$$

which has mean and covariance:

$$E\left\{\mathbf{r}(t_i^-)\right\} = \mathbf{0} \;\;,\;\; E\left\{\mathbf{r}(t_i^-)\mathbf{r}^{\mathrm{T}}(t_i^-)\right\} = \mathrm{Cov}\left(\mathbf{r}(t_i^-)\right) = \mathbf{H}(t_i)\mathbf{P}(t_i^-)\mathbf{H}^{\mathrm{T}}(t_i) + \mathbf{R}(t_i) \tag{7}$$

The inverse of the variance expression in Eq. (7) is identified as part of the Kalman gain **K** in Eq. (2). The measurement update equation Eq. (3) is an explicit function residual at time (t_i) and its variance, such that Eq. (3) can be rewritten as:

$$\hat{\mathbf{x}}(t_i^+) = \hat{\mathbf{x}}(t_i^-) + \mathbf{P}(t_i^-)\mathbf{H}^{\mathrm{T}}(t_i)\left[\mathrm{Cov}\left(\mathbf{r}(t_i^-)\right)\right]^{-1}\mathbf{r}(t_i^-) \tag{8}$$

If $\mathbf{R}(t_i)$ is assumed diagonal such that all observation errors at time (t_i) are independent, then $\mathbf{z}_i$ and $\mathbf{r}(t_i^-)$ are scalar quantities, as is the covariance of $\mathbf{r}(t_i^-)$, such that Eq. (3) is a function of the sequence of scalar residual ratios.

For the system described by Eq. (3), Eq. (5), and Eq. (7), the residual sequence $\mathbf{r}(t_i^-)$ will be *white* and Gaussian.[9, 10] Therefore, the sequence of *residual ratios* (the elements of $\mathbf{r}(t_i^-)$ divided by the diagonal elements the residual covariance of Eq. (7)) for t_i, $i = 1, 2, \ldots n$ is distributed according to a Gaussian distribution with zero mean and unit variance, such that a statistical test can be performed on each member of the sequence $\mathbf{r}(t_i^-)$, $i = 1, 2, \ldots n$ to identify outlying measurements.

Sparse Data Considerations

A test of statistical hypothesis on the residual ratios tends to correctly identify outlying measurements *providing that the noise modeling is appropriate for the situation*. To better accomplish this, the Orbit Determination Toolkit (ODTK) uses a gravity process noise model that allows the user to account for errors of commission (uncertainty in the gravity field coefficients) and/or errors of omission (uncertainty due to truncating the gravity field).[11] Relative corrections to the satellite's nominal solar pressure coefficient and ballistic coefficient can also be modeled as an exponentially fading process, allowing for more realistic representations of the dynamical uncertainty of these parameters. A test of consistency between filter and smoother solutions may be used to verify that dynamical uncertainties are plausibly modeled for both simulated and actual tracking data.[12] Statistical tests of residual whiteness can also be used to assay goodness-of-fit.[13]

Consideration of Eq. (7) reveals that a large residual variance can be almost entirely due to the uncertainty of the state rather than the measurement uncertainty $\mathbf{R}(t_i)$, creating a situation where a measurement can have very large error but its residual ratio will not appear outlying. To illustrate, an example of a simulated scenario is considered: a low-Earth-orbiting spacecraft in a nearly circular orbit at 35° inclination and at 350 km altitude. The spacecraft is tracked over a two-week interval by 95 direction-cosine pairs from the Air Force Space Surveillance System (AFSSS), and 127 range and angle pairs from the Eglin radar. Figure 1 illustrates the residual ratios of the simulated observation data and reveals the relative sparseness of the tracking data, showing how a sequential filter can maintain the orbit estimate given very low tracking levels for an object having large dynamical uncertainties. Figure 2 illustrates how this dearth of tracking data affects the tangential uncertainty of the estimated spacecraft location, where the 99% uncertainty grows to 40 km to 50 km when large gaps occur in the sparse tracking span. Figure 3 illustrates how the Eglin range-residual uncertainties are largest at the start of the pass due to large satellite position uncertainty, but once the satellite position uncertainty is consequently reduced from measurement processing, the residual uncertainty is reduced per Eq. (7).

Studies with SSN tracking data have observed discordant observations occasionally at the beginning of a radar tracking sequence.[14] The hypothesis for their origin is that the radar's tracking

filter prematurely reports while still in an acquisition mode that is much noisier than typical. For a filter experiencing high process noise with sparse tracking, it is possible that this type of discordant observation could go undetected due to the large state uncertainty at the end of lengthy tracking outage. One might assign higher uncertainties to the first observation of radar tracks to mitigate this.

Filter Divergence

Filter divergence is the persistent condition where the state is no longer meaningfully updated despite the availability of observations. Processing a discordant measurement can drive the state to an incorrect solution and simultaneously reduce the uncertainty of the state, causing the filter to become overconfident about an incorrect state estimate. When undue authority is given to the prior state, measurement that were not "sufficiently close" to the incorrect state may be unnecessarily rejected. Or, future measurements may be processed, but their uncertainty is large relative to the supposedly high confidence of the estimate, resulting in small and ineffective gain. Sometimes called *smugness*, this situation may be symptomatic of highly optimistic process noise modeling.

Filter divergence tends to be readily identified by a lack of measurement processing. In ODTK for example, *divergence* is practically defined by the maximum number of measurements sequentially rejected before some corrective action needs to take place. Once the established divergence criterion is reached, the divergent condition might be remedied in various possible ways.

- The rejection threshold for residual ratios could be temporarily relaxed to allow the acceptance of measurements. (This is the option exercised in ODTK via its *dynamical editing* criterion.)
- The uncertainty of the state could be sufficiently inflated to allow the acceptance of new data.
- Sequential estimates could be saved over time; the state prior to any evidence of divergence could be used to restart the recursive process, with expert analyst involvement to discover what went wrong.
- Sequential processing could be restarted by a short-arc solution from a batch-least-squares estimation (discussed in the sequel) or possibly an initial orbit determination.

In practice, SSN tracks tend to be short, typically three observations spaced about 10 seconds apart. If a divergence criterion is considering more than a few observations, it will span multiple trackers. In sparse-data environments, the operational definition of *divergence* criterion might consider a sequence of tracks rather than a sequence of measurements. The probability that a single sensor is producing discordant data is much greater than two sensors independently producing discordant data.

Nonlinearity Effects. The classic behavior of a filter in the presence of non-linearity effects is to *over-correct*, meaning that the filter reduces the state error covariance too much during the measurement update, which then can lead to filter divergence.[15] *Over-correction* can be addressed by reducing the influence of the measurements during the period of time when non-linearity effects are significant. A number of so-called "de-weighting" methods may be utilized to limit the Kalman filter gain $\mathbf{K}(t_i)$, including both *ad-hoc* techniques and higher order filters. In a second order filter, for example, a term is added to the measurement error variance which effectively lowers the filter gain thus reducing the state correction and the reduction of the state error covariance.[15] A possible *ad-hoc* approach would temporarily add to the measurement uncertainties $\mathbf{R}(t_i)$ in Eq. (2) to limit the gain. The indicator for such replacement would likely be governed by the

magnitude of $\mathbf{r}(t_i^-)$ in Eq. (6) or the resultant state correction in Eq. (3) and Eq. (8). However, in the routine maintenance of Earth-orbiting space-objects, non-linearity effects tend to be ignored for detecting outliers; the evidence for this partly being that the vast majority of current orbit determination programs which have been used successfully for decade are based on linearized estimation theory.

Filter Initialization

The true error in the initial state estimate must be consistent with the initial state-error covariance.[16] When there is limited information *a priori*, often a "Kalman filter may be initialized using a very large initial covariance and any convenient initial condition, with the hope that the filter will eventually converge to a statistically correct behavior."[17] An initial state-error covariance populated with relatively large values minimizes the influence of a highly approximate initial state as observations are processed. An additional constraint is that the initial state-error covariance matrix must be positive definite; this requirement is met by populating the off-diagonal elements of initial state-error covariance with zeroes to create a diagonal matrix.

Often with sparse-data, the first observations are far from the initializing epoch. Thus, a relatively lengthy propagation from uncertain initial conditions might cause the state uncertainties can grow to very large levels. The uncertainty will reduce to more realistic levels after sufficient observations have been processed; however, with sparse observations this may take a relatively long time. The time-interval over which the state-error covariance may be considered unrealistic is known as the *filter-initialization* period. This interval varies depending on the accuracy of the initial state and state-error covariance, and the observability of the state parameters over time based on the processing of measurement data across the filter-initialization time interval.[18] Figure 4 illustrates simulated growth in tangential position uncertainty caused in part by an unrealistic diagonal state-error covariance.

A potential alternative to diagonal initialization is to start with the covariance from a batch-least-squares solution from a relatively short sequence of observations. If the time interval is relative short, the accumulated dynamical modeling error will not be significant enough to adversely influence the realism of the covariance. However, extending the state space beyond spacecraft position and velocity will almost certainly result in additional observability problems. Therefore, initializing the covariance with a short least-squares fit will tend to be problematic with short arcs of sparse data.

Using a short-arc least-squares covariance may hold more advantage when re-initializing a divergent filter. In this case, one likely has access to reasonably accurate force-model parameter estimates from prior catalogued solutions. Combining such information with the improved realism of a least-squares covariance is an interesting area for future study, although experience suggests that using nominal initial covariance matrices based the spacecraft orbital regime oftentimes works satisfactorily.

Batch Least Squares

Consider the situation where all observations are considered together as a batch, such that $\mathbf{z}_i$, $i = 1, 2, \ldots n = \mathbf{z}$. Then, the traditional linear weighted batch-least-squares *normal equations*,

$$\hat{\mathbf{x}} = \left[\mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{H}\right]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{z}\,, \tag{9}$$

result from a minimization of the scalar cost function

$$J = \tfrac{1}{2}\left[\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\right]^{\mathrm{T}} \mathbf{W}\left[\mathbf{z}_i - \mathbf{H}\hat{\mathbf{x}}\right]. \tag{10}$$

where the matrix **H** is realized according to the analytical observation-state relationship

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}, \tag{11}$$

v being an array of white measurement noise having zero expected value and weighted according to $\mathbf{W}^{-1}$. By analogy to the sequential estimator, the weight matrix **W** may be recognized as being equivalent to $\mathbf{R}^{-1}$. The residual model for the batch case is also analogous to that of the sequential case, such that a computed measurement is subtracted from the observed measurement, *i.e.*,

$$\mathbf{r} = \mathbf{z} - \mathbf{H}\hat{\mathbf{x}} \tag{12}$$

where each element of **r** is a residual with respect to the batch solution. Substituting Eq. (9) into Eq. (12), it can be shown that for an array of batch residuals **r**,

$$\mathbf{r} = \mathbf{z} - \mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{H}\right]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{z} = \left(\mathbf{I} - \mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{H}\right]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{W}\right)\mathbf{z} = (\mathbf{I} - \mathbf{\Lambda})\mathbf{z} \tag{13}$$

where

$$\mathbf{\Lambda} = \mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{H}\right]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{W}. \tag{14}$$

is sometimes known as the (weighted) *hat matrix*, or *projection matrix*, and which might be thought of as a batch-least-squares analog to the Kalman gain **K**. For the linear model of Eq. (11), $\mathrm{E}\{\mathbf{z}\mathbf{z}^{\mathrm{T}}\} = \mathrm{E}\{\mathbf{v}\mathbf{v}^{\mathrm{T}}\} = \mathbf{W}^{-1}$, such that the $n \times n$ covariance matrix of the batch residuals may be expressed as

$$\mathrm{Cov}(\mathbf{r}) = \mathrm{E}\left\{(\mathbf{I} - \mathbf{\Lambda})\mathbf{z}\left((\mathbf{I} - \mathbf{\Lambda})\mathbf{z}\right)^{\mathrm{T}}\right\} = (\mathbf{I} - \mathbf{\Lambda})\mathrm{E}\left\{\mathbf{z}\mathbf{z}^{\mathrm{T}}\right\}(\mathbf{I} - \mathbf{\Lambda})^{\mathrm{T}} = (\mathbf{I} - \mathbf{\Lambda})\mathbf{W}^{-1}(\mathbf{I} - \mathbf{\Lambda})^{\mathrm{T}}. \tag{15}$$

This can be further reduced to

$$\mathrm{Cov}(\mathbf{r}) = \mathbf{W}^{-1} - \mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{H}\right]^{-1}\mathbf{H}^{\mathrm{T}} = (\mathbf{I} - \mathbf{\Lambda})\mathbf{W}^{-1}. \tag{16}$$

If *a priori* information is available, the prior state and state uncertainty can be included as an observation, with the observation-state relationship being identity.[19] Residual ratios of unit variance may be created by dividing the elements of **r** by the diagonal elements the batch-residual covariance Cov(**r**), similar to what was previously suggested for the sequential case.

Per Eq. (14), the diagonal elements of the batch-residual covariance Cov(**r**) are entirely functions of the $n \times n$ hat matrix **Λ**, the content of which is dictated by **H**. Even if $\mathbf{W}^{-1}$ were diagonal, the (symmetric and idempotent) hat matrix **Λ** is generally non-diagonal; therefore, Cov(**r**) is non-diagonal and the batch-fit residual ratios are not independently distributed.

The lack of statistical independence in the elements of **r** complicates the application of statistical tests to identify outlying measurements in batch-least-squares problems. An assumption of $\mathrm{Cov}(\mathbf{r}) \approx \mathbf{W}^{-1}$ (or $\mathbf{\Lambda} = \mathbf{0}$) would result in incorrect residual uncertainties. Even if **Λ** were computed to improve the realism of Cov(**r**),* its contribution is still likely to under-represent the *true* level

---

* **Λ** is a not an ordinarily computed byproduct of the batch-least-squares process.

of residual uncertainty, due to no dynamical process noise within the batch-least-squares formulation. Therefore, the use of theoretical residual variance tends to be impractical for testing batch-least-squares residuals.

Mitigation Strategies. Particularly during the earliest iterations of the batch-least-squares orbit-determination procedure, the residuals are affected by slowly-varying systematic errors caused by the errors in the state parameters and uncompensated sensor biases.[20] The root-mean-square (RMS) of the batch-least-squares residuals is sometimes used *post facto* to establish the scale factor σ_r in Eq. (1), RMS being computed as:

$$\text{RMS} = \sqrt{\frac{\sum_{i}^{n} (r_i)^2}{n}} \ , \tag{17}$$

While RMS editing is commonly practiced, the following are notable disadvantages.

- Because batch-least-squares residuals are not independently and identically distributed, and are heteroskedastic in practice,* the RMS is not a reliable estimate of residual variance, such that a $C \times$ RMS criterion may not reject data at the rate presumed.
- There is no RMS value before the first iteration through the data. To reject data on the first pass usually requires an initial guess or an initial pass through the observations to compute residual RMS without any state update.
- The RMS is estimated directly from the sample *post facto*; therefore, it can already be contaminated by outlying observations not previously detected. This weakens the test and potentially masks the presence of discordant observations.
- Because the convergence of non-linear batch-least-squares estimator is often gauged by changes in the (weighted) RMS, inefficient outlier editing can substantially affect changes in the RMS and can result in slow convergence of the batch-least-squares estimator.

Unfortunately, when tracking data are sparse, a criterion based on $C \times$ RMS can be completely ineffective. If the RMS is estimated from a sample of size n, then

$$\max\{|r_1|, |r_2|, \ldots |r_n|\} \le (n)^{1/2} \times \text{RMS} \ , \tag{18}$$

which is to say, outlier detection based on $C \times$ RMS is impossible whenever $C > (n)^{1/2}$.† To illustrate by example, $10 \times$ RMS is an example of an operationally practiced criterion that initially screens for "impossible" values; however, the sample size n must first be greater than 100 (that is, C^2) to detect an outlier.[5]

Also, RMS calculations for data-rejection purposes are typically unweighted and may be based on observations only of a specific measurement *type*. For example, while twenty-four tracking measurements of azimuth, elevation, and range may be available for maintaining an orbit, only eight of those measurements may be ranging measurements. With only eight observations per measurement type, a $3 \times$ RMS editing criterion could never reject a single ranging measurement because $3 > (8)^{1/2}$. Of course more robust estimators of scale beyond RMS could be used, although their reliability tends to presume that the data are independently and identically

---

* Heteroskedasticity is when the variance functionally varies over time.
† This inequality is proven as an Appendix.

distributed with known distributional characteristics, and their reliability also decreases with very small sample sizes.[5]

Another operationally practiced alternative to RMS editing includes *variate-difference methods* that look for outliers within a sequence of residual *differences*.[20] Editors based on first- or second-differences can be successful at finding isolated outliers or outlier pairs, although their usage adds complexity to the estimation process. There are also certain drawbacks to the outlier methods based on residual differences:[14]

- Two or more successive outliers may mask each other when differenced, thus requiring higher-order differences.
- An outlying variate difference often consists of an acceptably small deviate subtracted from an unacceptable large one. Identifying which of these two is becomes increasingly cumbersome with higher order differences.
- Residual differences relative to the un-converged solution may exhibit large amounts of nonrandom behavior, which may not appear very Gaussian. Preliminary outlier identification based on a Gaussian assumption may be too aggressive.
- Differencing of successive residuals is uninformative on short sequences typical of some types of US space-surveillance tasking, *e.g.*, three (3) observations.

Non-Statistical Rejection Schemes

Certain rejection schemes, which are not based on the distribution of measurement residuals, can be applied to batch and sequential estimators. A hard limit may be imposed on what is an acceptable magnitude for a measurement. Geometry constraints may be imposed, such as rejection of very low elevation tracking, or tracking data that appear to have originated outside the field of regard of the sensor. As previously mentioned, the probability of a faulty measurement at the start of a pass is greater than subsequent measurements, and these tend to occur only in particular classes of sensors (*e.g.*, radars); therefore, one mitigation strategy might down-weight these specific measurements as a rule.

Cross-Tagged Measurements

Another common source of outlying data is the association of an observation with the wrong spacecraft, known as *cross-tagging*. This is a particular issue for objects within the geosynchronous belt, where multiple spacecraft can be in close spatial proximity, and where optically-based angular measurements are the dominant means of tracking. The observation association processing that assigns the observations to a certain spacecraft can be inaccurate due to the inherent limitations of angles-only techniques, and the proximity of the observational data values.

This sparse-tracking situation is one for which the outlier-rejection capabilities of the filter is thought to provide significant benefit. The force-model uncertainties tend to be much less for those orbits; the probability of accepting cross-tagged observations is lowered and their rejection by the filter aids in their identification. Figure 5 provides an illustration of how sparsely distributed residual ratios respond to the insertion of mis-tagged observations from a nearby spacecraft, after which original tracking data are restored about a day later. In contrast, the inability of an RMS-based editor to reject cross-tagged measurements might coax a batch-least-squares estimator away from the object of interest and onto the nearby orbiting spacecraft.

Maneuvering Spacecraft

Maneuvering spacecraft are another special class of sparse-data problem to which the sequential filter can provide benefit. Observations after the time of the maneuver will tend to be rejected

by a sequential estimator, or, the solution will adjust to the new orbit if the maneuver is small enough. For a batch-least-square estimator, the fit span is usually segmented at the time of the maneuver if the maneuver time is detectable or otherwise known in advance. In situations where the maneuver is small or is not detected, the solution over the entire fit span can effectively degrade.

PREDICTION

In principle, the physical modeling can be made the same for sequential and batch-least-squares estimators. The question becomes whether one estimator provides an estimate that forecasts more accurately than the other given that model. For both types of estimators, the solution at the end of the data span will be conditioned on all available data. If the dynamic force modeling is the same, the difference in solutions is generally due to how the observational data influence the final state in the presence of process noise. In a sequential filter, the influence of an observation lessens as the observation time gets farther from the epoch of solution due to the influence of process noise. With a batch-least-squares estimator, the lack of process noise allows observations to remain more influential over long time spans.

This effect of process noise could lead one to conjecture that a sequentially-based estimate at the end of the fit span may better reflect the local tendencies of the time-varying estimate, whereas the batch least-square estimate may better reflect the global tendencies of the time-varying estimate. For orbit determination applications, it may be reasonable to hypothesize that global averages will forecast in the very long term better than more localized solutions. Hence, a least-squares estimate may forecast more accurately in the very long term whereas a sequential estimate may forecast better in the very near term. For example, an extended Kalman filter/smoother was demonstrated in at least one situation* to predict less accurately during extended data gaps compared to a batch-least-squares estimator.[21]

Two Parameter Gauss-Markov Sequence

There are different approaches for modeling stochastic (random) process noise for state, acceleration, and measurement errors, which may improve the predictive tendencies of a sequential estimator. The stochastic process-noise model historically used in ODTK is a two-parameter *Gauss-Markov* sequence having the scalar representation:

$$x(t_k) = e^{-\alpha|t_k - t_{k-1}|}\, x(t_{k-1}) + \sqrt{1 - e^{-2\alpha|t_k - t_{k-1}|}}\, \sigma_Z\, Z(t_{k-1}) \;;\;\; k \in \{1,2,\cdots\}; Z(t_k) \sim N(0,1) \qquad (19)$$

where σ_Z is the root-variance scale factor applied to independent (white) draws $Z(t_k)$ of a Gaussian distribution of zero mean and unit variance, and α is a positive constant prescribing the degree of sequential autocorrelation. In practice, the user defines α through the exponential half-life $\tau_{½} = -\ln(½) / \alpha$.† These two parameters are generically known in ODTK as the *white-noise sigma* and the process *half-life*.

The model expressed in Eq. (19) is *Gaussian* in the sense that a population of independent realizations of $x(t_k)$ will be distributed multivariate normal at time (t_k). It is *Markov* in the sense that

---

* The example cited used a reduced-dynamic orbit determination technique involving estimated empirical accelerations fit *densely distributed* GPS-receiver observations, which is not really a sparse-data application.

† For a time interval equal to the half-life $\tau_{½}$, an estimated bias will decay by a factor of two (2) during estimation to its *a priori* value, in the absence of measurements.

the relationship of $x(t_k)$ does not extend beyond the preceding observation.[22] The mean and variance as a function of time interval can be expressed as:

$$\begin{aligned} E\{x(t_k)\} &= e^{-\alpha|t_k - t_{k-1}|} E\{x(t_{k-1})\} \ ; \\ \mathrm{Var}(x(t_k)) &= E\left\{\left(x(t_k) - E[x(t_k)]\right)^2\right\} = \sigma_Z^2 \ . \end{aligned} \tag{20}$$

The expected value of the sequence $x(t_k)$, $k = \{1, 2, \dots\}$ asymptotically approaches zero as $(t_k - t_{k-1}) \rightarrow \infty$.

The recursive *Gauss-Markov* sequence $x(t_k)$ requires two additional values for initialization at time (t_0). For many estimation and modeling purposes within ODTK, $x(t_k)$ in Eq. (19) represents a *difference* or offset from some *a priori* constant value of the state parameter being estimated or modeled; therefore, an *a priori* constant value is needed for initialization. The second additional parameter is known as the *bias sigma*, which the user prescribes as the square root of the error variance of the *a priori* constant value.

Three-Parameter Vasicek Sequence

A stochastic model related to the *Gauss-Markov* sequence is the three-parameter *Vasicek* (Wah-SI-check) sequence, its scalar representation being:

$$\begin{gathered} V(t_k) = e^{-a|t_k - t_{k-1}|} V(t_{k-1}) + \left[1 - e^{-a|t_k - t_{k-1}|}\right] b + \sqrt{1 - e^{-2a|t_k - t_{k-1}|}} \, \frac{\sigma}{\sqrt{2a}} Z(t_{k-1}) \ ; \\ k \in \{1, 2, \cdots\}; Z(t_k) \sim N(0,1) \end{gathered} \tag{21}$$

The *Vasicek* stochastic sequence is a more general, mean-reverting sequence that separates both short-run and long-run time-varying bias estimation with a single stochastic sequence.[23] The *Vasicek* sequence also has the same exponentially decaying correlation function as the *Gauss-Markov* sequence. (Sequences with exponentially decaying autocorrelation are sometimes described in literature as originating from, or as specialized solutions for, an Ornstein-Uhlenbeck process.[24]) The original application of the *Vasicek* sequence was for financial applications, *e.g.*, to simultaneously model a short-run interest rate level together with a long-run interest rate level. The mean and variance as a function of time interval can be expressed as:

$$\begin{aligned} E\{V(t_k)\} &= e^{-a|t_{k1} - t_{k-1}|} \, E\{V(t_{k-1})\} + \left(1 - e^{-a|t_{k1} - t_{k-1}|}\right) b \ ; \\ \mathrm{Var}(V(t_k)) &= E\left\{\left(V(t_k) - E[V(t_k)]\right)^2\right\} = \frac{\sigma^2}{2a}\left(1 - e^{-2a|t_k - t_{k-1}|}\right) \ ; \end{aligned} \tag{22}$$

As $(t_k - t_{k-1}) \rightarrow \infty$, the expected value of the sequence $V(t_k)$, $k = \{1, 2, \dots\}$ asymptotically approaches b. The variance of the sequence asymptotically approaches $\sigma^2/2a$.

Implementation in ODTK

The three-parameter *Vasicek* sequence is an added feature to ODTK Version 6.1. The *Vasicek* sequence of Eq. (21) can be viewed as an extension to the two-parameter *Gauss-Markov* sequence of Eq. (19) as it produces a numerically equivalent result providing that $b \equiv 0$, $a \equiv \alpha$, and $\sigma/\sqrt{2a} \equiv \sigma_Z$. For simplicity of implementation and to leverage user familiarity with ODTK's two-parameter *Gauss-Markov* model, ODTK's three-parameter *Vasicek* model defines a through the exponential half-life $\tau_{½} = -\ln(½) / a$. The user also specifies the asymptotic root-variance $(\sigma/\sqrt{2a})$ as the so-called *short-term sigma*, just as if it were the *white-noise sigma* σ_Z in the *Gauss-Markov* model.

Suppose one makes the two notational substitutions $\alpha \rightarrow a$, and $\sigma_Z \rightarrow (\sigma/\sqrt{2a})$ in Eq. (19) as previously mentioned. Considering that $x(t_k)$ in Eq. (19) is a modeled offset relative to some *a priori* constant value, suppose one also replaces $x(t_k)$ with an explicit expression representing the offset, *i.e.*, $x(t_k) = V(t_k) - b$, where b is a constant. One then finds Eq. (19) becoming:

$$\left(V\left(t_k\right)-b\right)=e^{-a\left|t_k-t_{k-1}\right|}\left(V\left(t_{k-1}\right)-b\right)+\sqrt{1-e^{-2a\left|t_k-t_{k-1}\right|}}\,\frac{\sigma}{2a}Z\left(t_{k-1}\right) \tag{23}$$

which reduces to Eq. (22) by inspection. Therefore, the three-parameter *Vasicek* model of ODTK extends the original two-parameter *Gauss-Markov* sequence by explicitly including the constant offset b within the *Gauss-Markov* sequence. Doing this allows b to become another parameter that can be estimated, together with $V(t)$ within the state space. Therefore, given enough tracking data, the constant parameter (or, more correctly, a correction to the user's specified constant) can be directly estimated. ODTK refers to the estimated correction to the user's specified constant the *long-term bias*.

Prediction Using a *Long-Term Bias*

Because the *Gauss-Markov* sequence $x(t_k)$ is a modeled offset to some user-specified constant, in the absence of measurements its expected value reverts to its asymptotic value of zero at a rate of decay that is dependent on the exponential half-life $\tau_{½} = -\ln(½) / \alpha$. Thus, a forecast of a bias-state estimate modeled as a *Gauss-Markov* sequence will revert back to the *a priori* constant value specified by the user sooner or later. In this case, the accuracy of very long-term predictions is governed by the correctness of the *a priori* constant value specified by the user.

The *Vasicek* sequence addresses this issue by estimating an additional correction b to the *a priori* constant value which does not decay with time. Rather, at the end of the fit interval, the estimated short-term component of the sequence (*i.e.*, $V(t)$) will fade at a rate of decay that is dependent on the exponential half-life $\tau_{½} = -\ln(½) / a$, reverting to the estimated long-term component b. Thereby, the forecast will be influenced more by the long-term estimate as the prediction extends farther from the end of the fit interval.

The *Vasicek* long-term parameter is therefore entirely analogous to modeling an estimated constant in batch-least-squares, such that one might expect the prediction with the long-term parameter to behave like estimated constants propagated in batch-least-squares. This seems relevant considering that some research and development with the US space-object catalog oftentimes presumes that a long-term average is a suitable operational proxy for a spacecraft parameter's "true" value. For example, in the development of the Air Force HASDM atmospheric model, a batch-least-squares algorithm was used to estimate approximately 3200 ballistic-coefficient (B) estimates per spacecraft, all of which "were used to obtain an average value, the 'true' B value, over the 31-year time period."[25] When data are sparse, estimation of such constant parameters is less practical with least-squares; however, a recursive approach using the *Vasicek* long-term parameter will asymptotically approach the long-term mean with a conveniently single fit span that can be indefinitely long.

Examples

An example of a simulated observation set is now considered: a low-Earth-orbiting (LEO) spacecraft in a nearly circular orbit at 98° inclination and at 750 km altitude, tracked over a one-month interval by 274 direction-cosine pairs from the Air Force Space Surveillance System (AFSSS) and 477 range and angle pairs from various SSN radar sites. The simulated measurements were fit by the sequential filter using a *Vasicek* model for the ballistic coefficient and solar radiation pressure coefficient; however, the filter was initialized with coefficient values too low

by 10%. Figure 6 shows that ballistic coefficient estimate over time increases to approximately 10%, that is, back to the correct value used to simulate the measurements. The short-term and long-term components of the solar-radiation-pressure coefficient estimate did not separate significantly during the two-week time interval under consideration, which assumed a solar-radiation-pressure coefficient half-life of two days. It is conjectured that the long-term behavior of the solar-radiation-pressure coefficient was not likely observable for the scenario being considered.

Another scenario was considered; a geosynchronous spacecraft in a nearly circular orbit at 7.5° inclination, tracked over a four-month interval by 276 range and angle pairs from a simulated radar, and 1352 angle-pairs from four SSN optical sites. In the same way, the filter was initialized with a solar-radiation-pressure coefficient value too low by 10%. Using a coefficient half-life of seven days, Figure 7 shows that ballistic coefficient estimate over time increases back to approximately 10% over a period of several months. The experiment was repeated using a coefficient half-life of one-half day, with the outcome illustrated in Figure 8. In this case the long-term bias component appeared to regress faster toward the simulated value.

Using the LEO scenario, Figure 9 illustrates that improved prediction accuracy may be expected when using the *Vasicek* sequence in lieu of the *Gauss-Markov* sequence. Here, simulated data were generated using a *Gauss-Markov* sequence for the ballistic-coefficient correction of with sigma of 20% and half-life of 90 minutes. The *a priori* ballistic coefficient was decreased by 10% and the coefficient was filtered assuming a *Gauss-Markov* sequence. The estimate was repeated using a *Vasicek* sequence for the ballistic coefficient with a long-term sigma of 10% and a short-term sigma of 20% and half-life of 90 minutes. In both cases the solution was allowed to predict an additional nine days, and then each result was differenced against the original simulation case. The idea was to simulate a scenario with the spacecraft having variable area throughout the orbit. Both filtered solutions looked comparable to each other and reasonably matched the simulation during the data-processing interval. In the prediction interval, the *Vasicek* produced a better absolute prediction than the *Gauss-Markov*, which is expected. Also, the *Gauss-Markov* prediction error is not strongly consistent with its covariance in this example.

CONCLUDING SUMMARY

Sparse data present special challenges to the orbit-determination problem. Limited observations result in larger parameter uncertainties and degraded prediction accuracy on average. If outlier contamination exists, sparse data make it hard or impossible to identify faults. This is true for both batch and sequential estimation, but the problem will manifest itself in different ways. With the sequential filter, limited data can result in larger residual uncertainties which might increase the probability of accepting bad data using a statistical hypothesis test. With a batch estimator, outliers can contaminate the iterated global estimate as well as the scale estimate (such as RMS) used to identify outliers, making *ad hoc* detection schemes less efficient or insufficient, and potentially slowing or prevent solution convergence. It is therefore reasonable to conclude that neither a batch-least-squares approach nor a filtering approach is ideal at rejecting outliers from sparse data.

Under certain circumstances, the two methods can complement each other. For example, a least-squares solution can be used to initialize a sequential filter, and the filter's autonomous editing of mis-tagged data and inherent maneuver detection capability can be used for these situations which can be more problematic for batch-least-squares. Also, outlier mitigation schemes which are not statistically based can be useful with either approach.

For prediction, a capability has been added to ODTK Version 6.1 to separate a stochastic model parameter into a varying short-term estimate and an asymptotic long-term estimate, assum-

ing that there is a long-term component to the behavior that is approximately constant. This so-called *long-term bias* should have long-term predictive behavior more like batch-least-squares estimates, which allows for an asymptotic constant parameter to be estimated with sparse data over an indefinitely long fit span.

APPENDIX A: MAXIMUM LIMIT FOR RMS EDITING

The following derivation is an extension of a result by Shiffler (1988) which accommodates RMS as a scale estimate when testing for outliers within a sample.[26] Let $\{r_1, r_2, \ldots r_{(n-1)}\}$ denote a random sample of size (n-1), reordered from smallest value to highest value. Let $\mathrm{RMS}_{(n-1)}$ denote the root-mean-square computed from this sample according to Eq. (17).

Let us propose to add to this set an even larger value $r_{(n)} > r_{(n-1)}$, which is possibly outlying. From Eq. (17), it is straightforward to show that

$$\left(\mathrm{RMS}_{(n)}\right)^2 = \left(\mathrm{RMS}_{(n-1)}\right)^2 \frac{(n-1)}{n} + \frac{r_{(n)}^{\ 2}}{n} \ , \tag{24}$$

where $\mathrm{RMS}_{(n)}$ is the RMS computed via Eq. (17) after $r_{(n)}$ is added to the ordered set. The question of interest is: what is the maximum possible value of the ratio $R_{(n)} = r_{(n)} / \mathrm{RMS}_{(n)}$?

Following an argument posed by Shiffler (1988), it is obvious that the ratio $R_{(n)}$ is maximized when the denominator $\mathrm{RMS}_{(n)}$ is minimized; this minimization occurs when $\mathrm{RMS}_{(n-1)} = 0$ (or alternately, when all members of the set $\{r_1, r_2, \ldots r_{(n-1)}\}$ are exactly zero). In this minimizing situation,

$$R_{(n)}^{\ 2} \equiv \frac{r_{(n)}^{\ 2}}{\left(\mathrm{RMS}_{(n)}\right)^2} = \frac{r_{(n)}^{\ 2}}{\left(\left(\mathrm{RMS}_{(n-1)}\right)^2 \frac{(n-1)}{n} + \frac{r_{(n)}^{\ 2}}{n}\right)} = \frac{r_{(n)}^{\ 2}}{\left(\frac{r_{(n)}^{\ 2}}{n}\right)} = n \ . \tag{25}$$

Identical reasoning can be used to show that the maximum possible value for $R_{(0)}$ as caused by a large negative value $r_{(0)}$ is also n.

Such arguments prove that $| r_{(n)} / \mathrm{RMS}_{(n)} | \le (n)^{1/2}$ always, regardless of the magnitude of the outlier $r_{(n)}$, and an equality only holds under the highly implausible condition that every other member of the set $\{r_1, r_2, \ldots r_{(n-1)}\}$ is exactly zero. Neglecting the implausible circumstance then,

$$C < (n)^{1/2} \ . \tag{26}$$

is a necessary condition for identifying $r_{(n)}$ as an outlier if using the criterion $| r_{(n)} | > C \times \mathrm{RMS}$.

APPENDIX: FIGURES

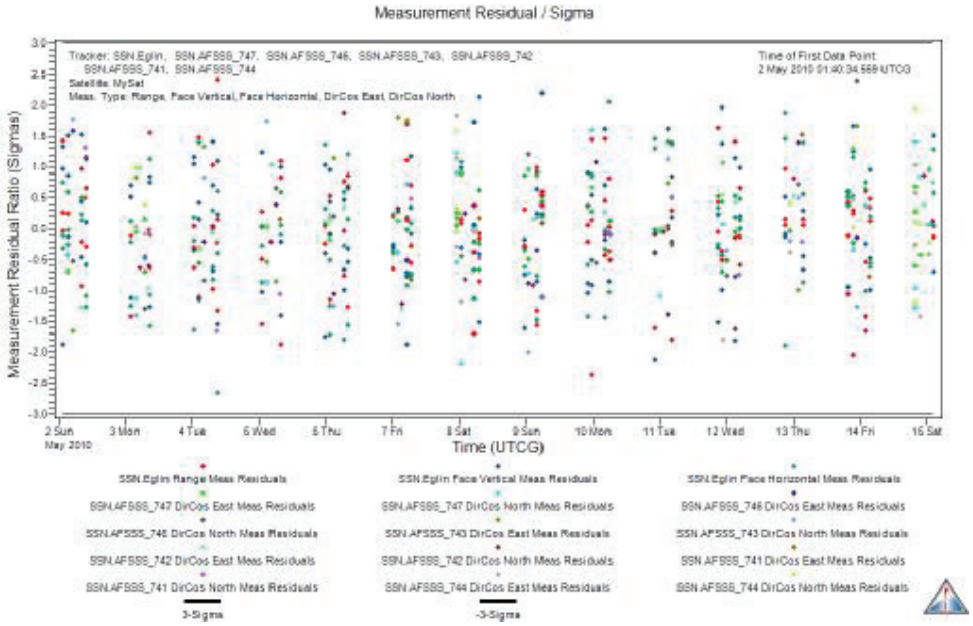

Figure 1. Sparse Data Case 7A – Normalized Tracking Residuals

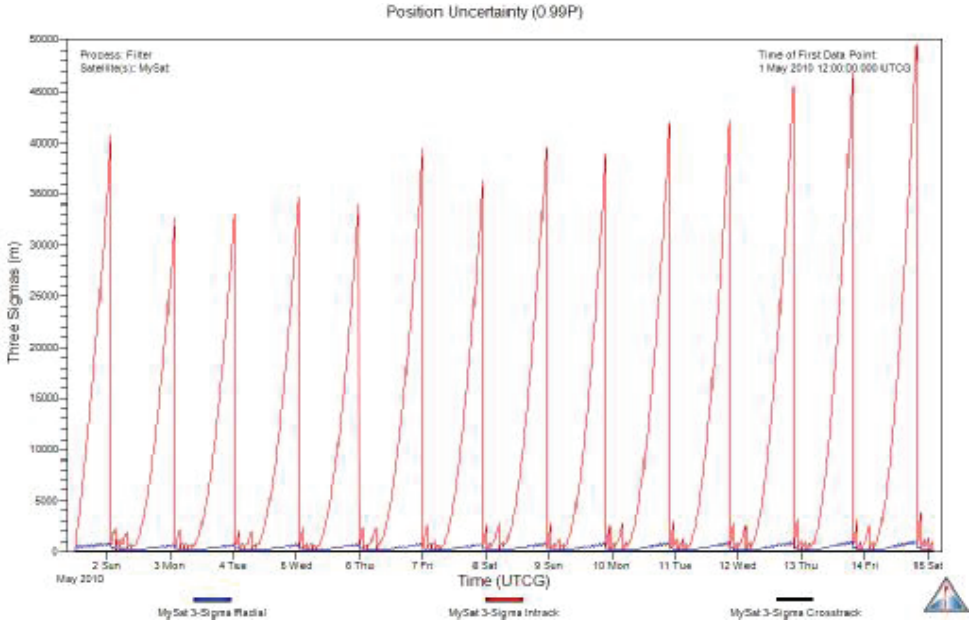

Figure 2. Sparse Data Case 7A – Position Uncertainty

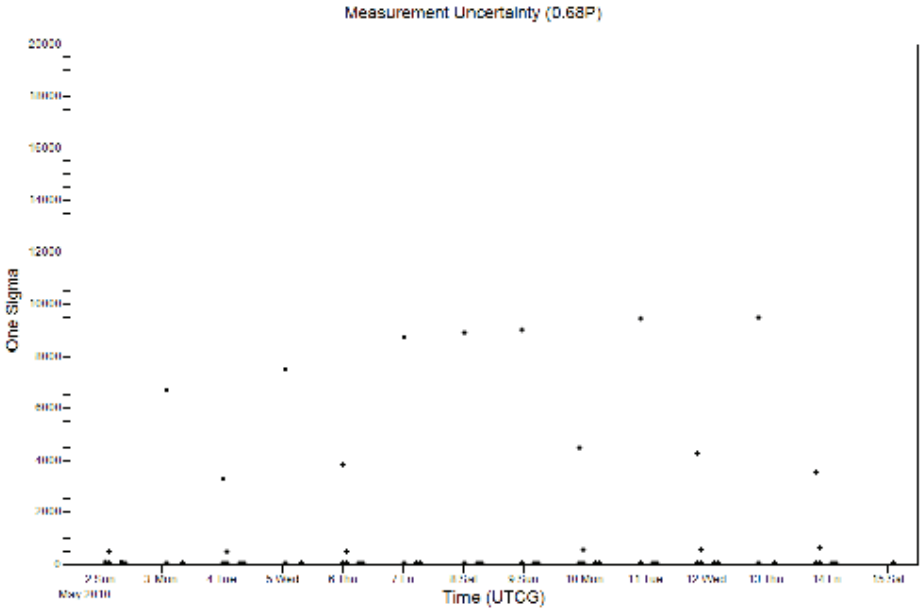

Figure 3. Range Residual Uncertainty (m) for Sparse Data Case 7A

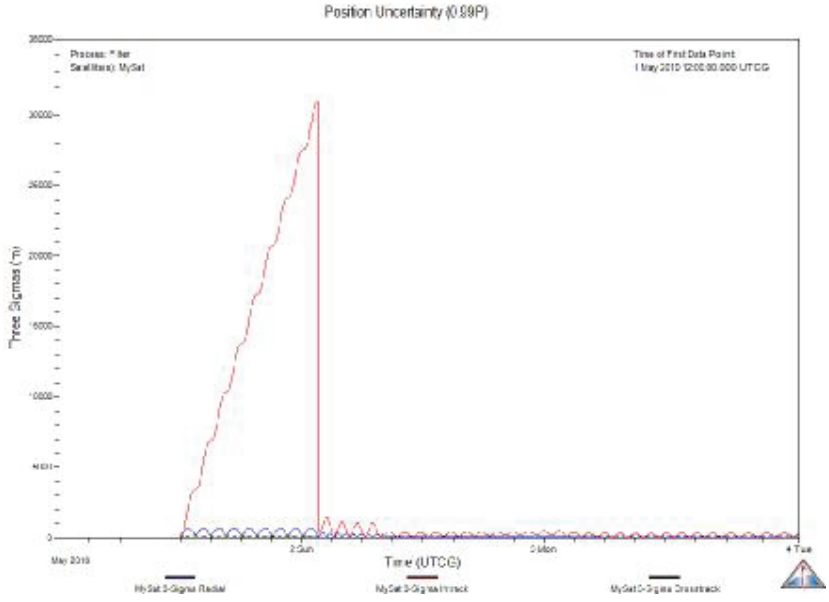

Figure 4. Tangential Position Error Uncertainty Growth From Initialization Using a Diagonal Covariance Matrix.

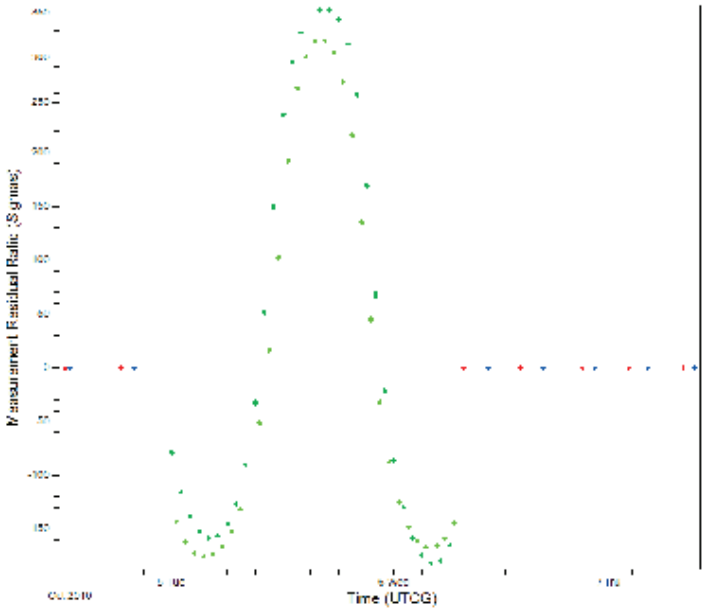

Figure 5. Filter Residuals of Cross-Tagged Ranging Measurements.

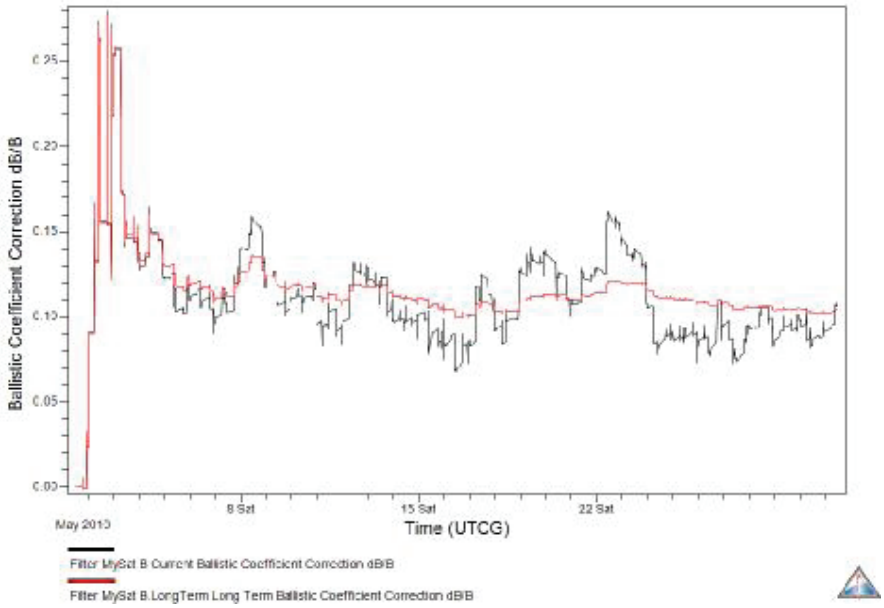

Figure 6. 10% Positively Biased Drag Coefficient, 1-Day Half-Life.+10.4%

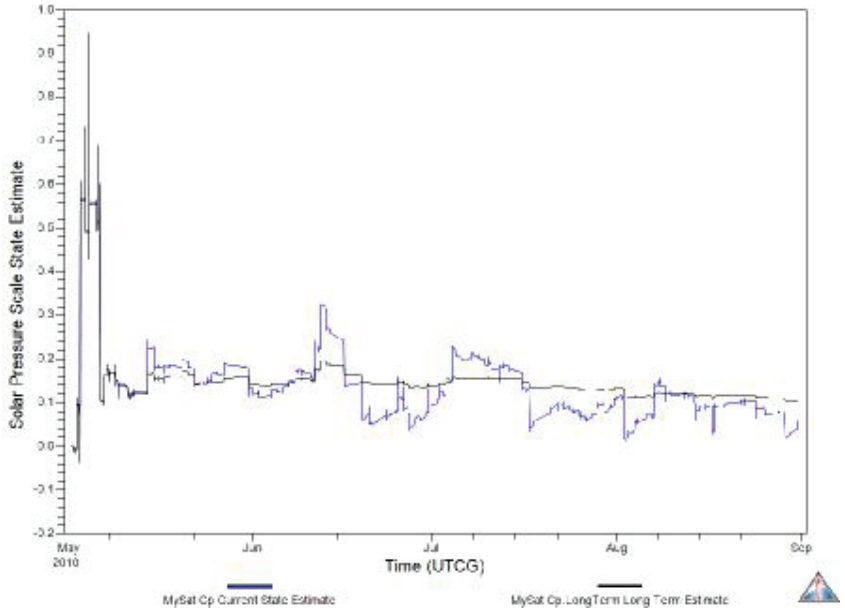

Figure 7. 10% Positively Biased Solar Pressure Scale, 7-Day Half-Life. +10.6%

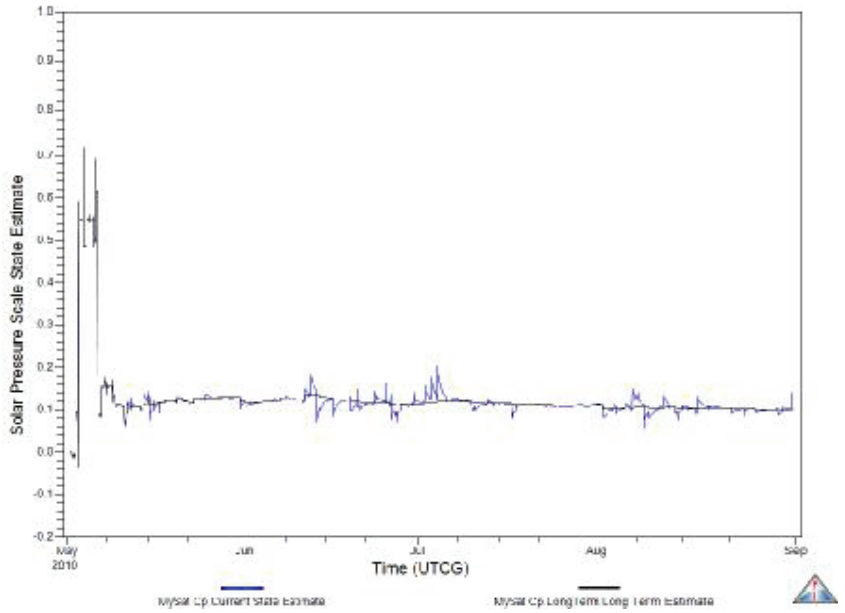

Figure 8. 10% Positively Biased Solar Pressure Scale, 0.5 Day Half-Life, +10.2%

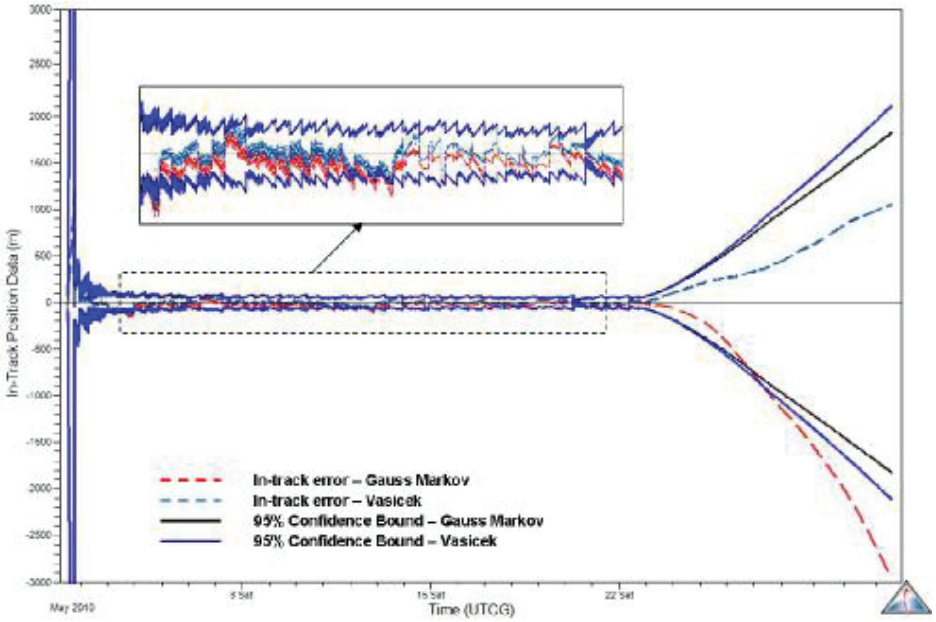

Figure 9. Propagation of *Gauss-Markov* and *Vasicek* Estimates.

REFERENCES

[1] Swerling, P. (1959), "First order error propagation in a stagewise smoothing procedure for satellite observations." Journal of the Astronautical Sciences, Vol. 6, pp. 46-52.

[2] Tapley, B.D., B.E. Schutz, G.H. Born (2004), Statistical Orbit Determination, Elsevier Academic Press, Burlington, MA. pp. 199-208.

[3] Wright, J.R. (2010), *Orbit Determination Tool Kit Theorems & Algorithms*. Appendix B., Analytical Graphics, Inc.

[4] Montenbruck, O., E. Gill, *Satellite Orbits*. *Models, Methods and Applications*. Springer, 2000, p. 287.

[5] Seago, J.H., M.A. Davis, W.R. Smith (2005), "Estimating the Error Variance of Space Surveillance Sensors." Paper AAS 05-127, from Vallado *et al*. (eds.), *Spaceflight Mechanics 2005 - Advances in the Astronautical Sciences. Vol. 120, Part I*, Proceedings of the AAS/AIAA Space Flight Mechanics Meeting, Copper Mountain, Colorado, January 23-27, 2005, pp. 367-386.

[6] Vallado, D.A., J.H. Seago (2009), "Covariance Realism." Paper AAS 09-304, from Rao, *et al*. (eds.), *Astrodynamics 2009 - Advances in the Astronautical Sciences, Vol. 112, Part I,* Proceedings of the 19th AAS/AIAA Astrodynamics Specialist Conference, Pittsburgh, Pennsylvania, August 9-13, 2009, p. 51.

[7] Wright, T.W. (1884), *A Treatise on the Adjustment of Observations by the Method of Least Squares, With Applications To Geodetic Work and Other Measures Of Precision.* Van Nostrand, New York, NY.

[8] Maybeck, P.S. (1979) *Stochastic Models, Estimation, and Control - Vol. 1*. Academic Press, New York, pp. 231-32.

[9] Anderson, B.D.O., J.B. Moore (1979), *Optimal Filtering*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey. p. 231.

[10] Maybeck, P.S. (1982) *Stochastic Models, Estimation, and Control - Vol. 2*. Academic Press, New York, p. 136.

[11] Wright, J.R. (1981), "Sequential Orbit Determination with Auto-Correlated Gravity Model Errors", *Journal of Guidance and Control*, Vol. 4, No. 3, May 1981, pp. 304-309.

[12] McReynolds, S.R. (1984), "Editing Data Using Sequential Smoothing Techniques for Discrete Systems." Paper AIAA-1984-2053, Proceedings of the AIAA/AAS Astrodynamics Conference, Seattle, WA, August 20-22, 1984.

[13] Seago, J.H., D.A. Vallado (2010), "Goodness-of-Fit Tests for Sequential Orbit Determination." Paper AAS 10-149, from Mortari, D. *et al*. (eds), *Space Flight Mechanics 2010 - Advances in the Astronautical Sciences, Vol. 136, Part I*, Proceedings of the 13th AAS/AIAA Space Flight Mechanics Meeting, San Diego, California, February 14-17, 2010, pp. 749-768.

[14] Seago, J.H., M.A. Davis, A.E. Reed (2003), "Characterization of Space Surveillance Sensors Using Normal Places." Paper AAS 03-248, from Scheeres, D.J., *et al*. (eds.), *Space Flight Mechanics 2003 - Advances in the Astronautical Sciences, Vol. 114, Part III*, Proceedings of the 13th AAS/AIAA Space Flight Mechanics Meeting, Ponce, Puerto Rico, February 9-13, 2003, p. 2155.

[15] Woodburn, J.W., S. Tanygin (2010), "Detection of Non-Linearity Effects during Orbit Estimation." Paper AAS 10-239, from Mortari, D. *et al*. (eds), *Space Flight Mechanics 2010 - Advances in the Astronautical Sciences, Vol. 136, Part I*, Proceedings of the 20th AAS/AIAA Space Flight Mechanics Meeting, San Diego, California, February 14-17, 2010, p. 2041.

[16] Bar-Shalom, Y., X.R. Li, T. Kirubarajan (2001), *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, New York, p. 248.

[17] Fisher, H.L., M.D. Shuster, T.E. Strikwerda (1989), "Attitude Determination for the Star Tracker Mission." Paper AAS 89-365, from Thornton *et al*. (eds.) *Astrodynamics 1989 - Advances in the Astronautical Sciences. Vol. 71, Part I*, Proceedings of the AAS/AIAA Astrodynamics Specialist Conference, August 7-10,1989, Stowe, Vermont, p. 140.

[18] Tanygin, S., J.R. Wright (2004), "Removal of Arbitrary Discontinuities in Atmospheric Density Modeling." Paper AAS 04-176, from Coffey et al. (2004), *Spaceflight Mechanics 2004 - Advances in the Astronautical Sciences. Vol. 119, Part II,* Proceedings of the 14th AAS/AIAA Space Flight Mechanics Meeting, Maui, Hawaii, February 8-12, 2004, p. 1186.

[19] Tapley, B.D., B.E. Schutz, G.H. Born (2004), *Statistical Orbit Determination*, Elsevier Academic Press, Burlington, MA. p. 189.

[20] Koskela, P.E., Astrodynamic Analysis for the Advanced Orbit/Ephemeris Subsystem (AOES). Aeronutronic Publication No. U-4180, PHILCO-Ford Corporation, September 1, 1967, p. 111.

[21] Montenbruck, O., T. van Helleputte, R. Kroes, E. Gill (2005), "Reduced dynamic orbit determination using GPS code and carrier measurements." *Aerospace Science and Technology*, Vol. 9, pp. 261–71.

[22] Bendat, J.S., A.G. Piersol (2000), Random Data: Analysis & Measurement Procedures. Wiley-Interscience. p. 159-61.

[23] Glasserman, P. (2004), Monte Carlo Methods in Financial Engineer. Springer., p. 110.

[24] Bar-Shalom, Y., X.R. Li, T. Kirubarajan (2001), *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, New York, pp. 232-3

[25] Bowman, B. (2002), "True Satellite Ballistic Coefficient Determination for HASDM." Paper AIAA 2002-4887, Proceedings of the AIAA/AAS Astrodynamics Specialist Conference and Exhibit, August 5-8, 2002, Monterey, California.

[26] Shiffler, R.E. (1988), "Maximum Z Scores and Outliers," *The American Statistician*, Vol. 42, No. 1. pp. 79-80.