MEASURING IMAGE NAVIGATION AND REGISTRATION PERFORMANCE AT THE 3-σ LEVEL USING PLATINUM QUALITY LANDMARKS*

James L. Carr, Ph.D. and Houria Madani, Ph.D.
Carr Astronautics Corp.
1725 Eye St. NW, #300
Washington, D.C. 20006 U.S.A.

jcarr@carrastro.com, hmadani@carrastro.com

ABSTRACT

Geostationary Operational Environmental Satellite (GOES) Image Navigation and Registration (INR) performance is specified at the 3-σ level, meaning that 99.7% of a collection of individual measurements must comply with specification thresholds. Landmarks are measured by the Replacement Product Monitor (RPM), part of the operational GOES ground system, to assess INR performance and to close the INR loop. The RPM automatically discriminates between valid and invalid measurements enabling it to run without human supervision. In general, this screening is reliable, but a small population of invalid measurements will be falsely identified as valid. Even a small population of invalid measurements can create problems when assessing performance at the 3-σ level. This paper describes an additional layer of quality control whereby landmarks of the highest quality ("platinum") are identified by their self-consistency. The platinum screening criteria are not simple statistical outlier tests against sigma values in populations of INR errors. In-orbit INR performance metrics for GOES-12 and GOES-13 are presented using the platinum landmark methodology.

Key words: GOES I-M, GOES-N, Image Navigation and Registration (INR), landmarks, validation

1. INTRODUCTION

The Image Navigation and Registration (INR) systems of the GOES I-M[1] and GOES N-P[2] geostationary weather satellites calibrate and control the geometry of imagery and sounding products from these satellites. The INR system estimates a daily Orbit and Attitude (O&A) solution. Each daily O&A solution is propagated forward to predict image motion for the next day and an Image Motion Compensation (IMC) coefficient set is created and uploaded to the spacecraft. The IMC set is used to compute an IMC signal that adjusts the Imager scan pattern to compensate for the predicted image motion, with the objective of providing a standard and fixed relationship between pixel location and geographic coordinates. Inputs to the O&A generation process include landmark measurements. Landmarks are also used to assess INR performance. Since INR performance requirements are stated at the 3-σ level, it necessary to have a highly robust and reliable method for the validation of landmark measurements.

Landmarks are measured by the Replacement Product Monitor (RPM), one of the components of the INR system. The RPM measures absolute landmark position by correlating a landmark neighborhood against a map. It also makes a relative measurement by correlating neighborhoods from one image against corresponding

---

*Work performed with Carr Astronautics Internal Research and Development (IR&D) funds.

neighborhoods from the previous image. The RPM automatically screens for cloud coverage, poor illumination (visible landmarks), and poor thermal contrast (IR landmarks), enabling it to discriminate valid measurements from invalid measurements and to operate without human supervision. In general, this screening is reliable, but there will always be a small population of invalid measurements falsely identified as valid. Even a population of invalid measurements as small as a tenth of a percent can create problems when assessing performance at the 3-σ level. This paper describes an additional layer of quality control added to the validation process to select landmark measurements of the highest quality ("platinum"). Platinum landmarks are identified by the mutual consistency between absolute, relative, and multi-spectral measurements within the larger set of valid landmarks.

The INR performance metrics and the INR specifications for the GOES I-M and GOES N-P series are provided in Section 2. RPM landmark measurement capabilities are described in Section 3. Section 4 describes the platinum landmark criteria. The platinum landmark methodology is illustrated in Section 5 with flight data from GOES-12 and GOES-13.

2. INR REQUIREMENTS FOR GOES WEATHER SATELLITES

Image navigation is the process of determining the location (latitude and longitude) on the Earth of each pixel or sounding. Image registration is the process of maintaining each pixel or sounding at the same Earth location from one frame to the next. Good INR performance is crucial for accurate location of severe weather events and determination of Atmospheric Motion Vectors (AMVs). INR performance can be expressed using the following metrics:

1. Navigation error (NAV) is the error in pixel or sounding location.
2. Frame-to-Frame Registration (FFR) is the change in NAV from one frame to another.
3. Within-Frame Registration (WIFR) is the relative alignment of pixels or soundings within a frame (related to distortion).

The INR requirements for GOES I-M and GOES N-P are shown in Table 2-1. The data and analyses described in this paper refer only to the Imager; therefore, only Imager requirements are given.

Table 2-1. INR Requirements for The GOES I-M and GOES N-P Imager

INR Metric	GOES I-M Imager 3-σ Requirements		GOES N-P Imager 3-σ Requirements
	Noon ± 8 hrs	Midnight ± 4 hrs	Measured* Non-Eclipse
NAV	112 μrad	168 μrad	65 μrad
15-min FFR	50 μrad	70 μrad	36 μrad
90-min FFR	84 μrad	105 μrad	49 μrad
24-hr FFR	168 μrad	168 μrad	114 μrad
48-hr FFR	210 μrad	210 μrad	N/A
WIFR (25 min)	50 μrad	50 μrad	54 μrad

***Includes an allocation for the inaccuracy of measuring INR with landmarks**

These requirements are stated at the 3-σ level. The meaning of "3σ" is defined by NASA as the error at the 99.7-th percentile, which is the population inside ±3σ on a one-dimensional normal distribution. The interpretation of these requirements is that at least 99.7% (rounded from $erf(3/\sqrt{2}) = 0.9973002...$) of the measurement population should lie within specification thresholds $\pm\theta$.

3. MEASURING INR ERROR

The only way to directly measure INR performance is through landmark registration which is referred to as "landmarking". Automatic landmarking is performed in real-time in the RPM, which is one of the components of the GOES ground system. A brief description of the RPM landmarking capabilities follows. A more detailed description of the RPM automatic landmarking process can be found in Madani, Carr, and Schoeser[3].

3.1 RPM LANDMARKING FUNCTIONS

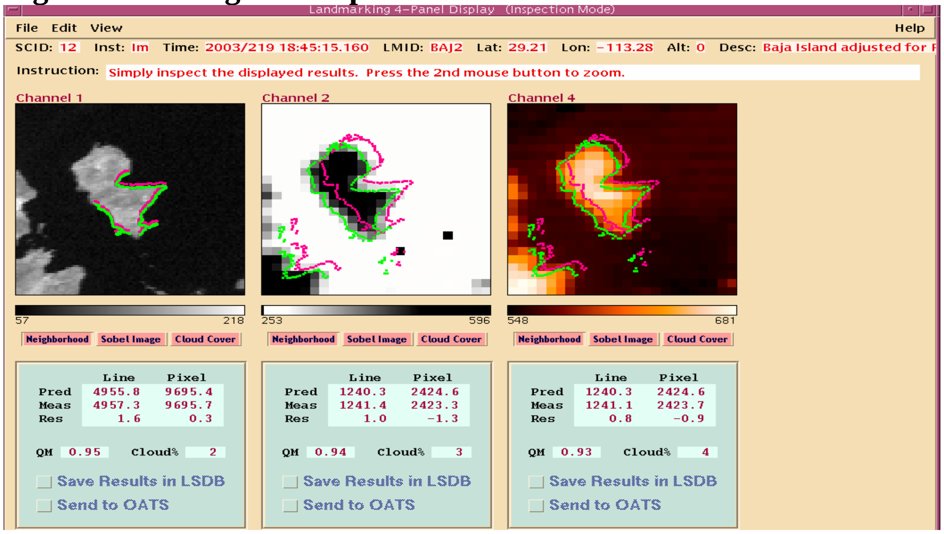
Two independent landmark registration techniques are included in the RPM automatic landmarking function: image-to-map registration, which determines the navigation error between the known geographic position of a given landmark and its position in the acquired image, and image-to-image registration, which determines the change in the apparent position between successive frames for a given landmark. The image-to-map registration algorithm measures an "absolute" error that is a NAV error datum. The image-to-image registration algorithm measures a "relative" error that is a FFR error datum.

The image-to-map registration process is based on a similarity measure that evaluates the similarity between the edges in a neighborhood surrounding a given landmark and a "chip" containing a set of land-water boundaries extracted from a digital map. The digital map used in the RPM is the Global Self-consistent Hierarchical High-resolution Shoreline (GSHHS) database[4]. The shoreline points extracted from the GSHHS are remapped into GOES image coordinates using the GOES Earth Location User's Guide (ELUG)[5]. The search area center-point and dimensions used to correlate each landmark are determined based on the navigation error and its uncertainty estimated by a Kalman Filter embedded within the RPM. The shifts in North-South (NS) and East-West (EW) coordinates needed to match the neighborhood and the chip are the absolute error components, which are also referred to as navigation residuals. An example of a neighborhood-chip correlation is shown in Figure 3-1 for the VIS, IR2 and IR4 channels of GOES-12. The magenta and green dotted curves in Figure 3-1 represent the predicted and measured shoreline points, respectively.

In the image-to-image registration process, a "chip" sub-image extracted from a previous landmark neighborhood is correlated against the current neighborhood representing the same landmark. The NS and EW shifts that maximize the correlation surface value are the relative error components.

The search for the maximum of the correlation surface in both the absolute and relative error processes is conducted at the sub-pixel level to obtain sub-pixel measurement precision.

Figure 3-1. Image-to-Map Correlation for a GOES-12 Landmark



SCID: 12 Inst: Im Time: 2003/219 18:45:15.160 LMID: BAJ2 Lat: 29.21 Lon: -113.28 Alt: 0 Desc: Baja Island adjusted for

Instruction: Simply inspect the displayed results. Press the 2nd mouse button to zoom.

Channel 1

	Line	Pixel
Pred	4955.8	9695.4
Meas	4957.3	9695.7
Res	1.6	0.3

QM 0.95 Cloud% 2

☐ Save Results in LSDB
☐ Send to OATS

Channel 2

	Line	Pixel
Pred	1240.3	2424.6
Meas	1241.4	2423.3
Res	1.0	-1.3

QM 0.94 Cloud% 3

☐ Save Results in LSDB
☐ Send to OATS

Channel 4

	Line	Pixel
Pred	1240.3	2424.6
Meas	1241.1	2423.7
Res	0.8	-0.9

QM 0.93 Cloud% 4

☐ Save Results in LSDB
☐ Send to OATS

3.2 CLOUD MASKING

An important step in the RPM landmarking process is automatic cloud masking. The registration processes described in Subsection 3.1 are performed only on cloud-free pixels. The cloudiness status of each pixel is obtained by combining the results of a series of cloud detection tests on the multi-spectral bands of the GOES images. Currently, GOES neighborhoods are generated in four (for GOES I-L) or three (for GOES M-N) spectral channels at two spatial resolutions as shown in Table 3-1 to cover the VIS and IR window channels. A more detailed description of the RPM cloud detection algorithm can be found in Madani, Carr, and Schoeser[6]. Accurate VIS and IR radiometric calibration is important for accurate cloud detection.

Table 3-1. Spectral and Spatial resolution of GOES I-N Landmark Neighborhoods

Channel Number	Wavelength (μm)	Spatial Resolution (km)
1 (VIS)	0.55-0.75	1
2 (IR2)	3.8-4.0	4
4 (IR4)	10.20-11.2	4
5 (IR5) GOES I-L	11.25-12.5	4

3.3 MEASUREMENT VALIDITY

RPM determines the validity of an absolute error measurement through its Quality Metric (QM). The QM takes into account the following factors: goodness-of-fit of the neighborhood edges to the GSHHS chip, cloud contamination fraction, illumination for the VIS channel, and thermal contrast for the IR channels. Each quality factor is assigned a value between 0 and 1 and a fuzzy-logic combination of these factors, Equation (3-1), results in a QM with a value between 0 and 1.

$$QM = \sqrt[N]{(factor_1)^{\gamma_1}(factor_2)^{\gamma_2}...(factor_N)^{\gamma_N}} \qquad (3\text{-}1)$$

The values of $\gamma_{1...N}$ are such that QM values higher than a specific threshold (typically 0.90) describe a valid measurement. A robust validation of the landmark measurements is important because the INR performance assessments can be significantly affected by a few invalid measurements due to the strict 3-σ requirements described in Section 2. Only measurements that pass the QM test are sent to the Orbit and Attitude Tracking System (OATS) for navigation purposes. An additional filtering process based on the Kalman Filter can also be used to prevent measurements falsely labeled as valid from being sent to OATS. This process, termed the Kalman Filter "veto", is described in Section 4.

Although the RPM relative error measurements are not sent to OATS and are not reported as part of the INR performance regularly published in the NOAA/NESDIS INR website[7], validity criteria have been defined for these measurements with the goal of improving the ability to measure INR performance. The validity criteria for relative error measurements consist in setting a threshold on the normalized value of the peak of the image-to-image correlation surface (ρ) and setting a limit on the total amount of cloudiness of the neighborhood. In addition, a limit is imposed on the time difference between the current and previous frames.

The measurement validity criteria applied to the data presented in this paper are summarized in Table 3-2.

Table 3-2. Valid Measurement Criteria for GOES 12 and GOES 13

Metric	**GOES 12**	**GOES 13**
Quality Metric	QM ≥ 0.90 for VIS QM ≥ 0.87 for IR2 & IR4	QM ≥ 0.90 for VIS,IR2,IR4
Kalman Filter Veto	4σ for VIS 3σ for IR2 & IR4	Not Applied
Relative Error Correlation (ρ)	ρ ≥ 0.9 for VIS ρ ≥ 0.96 for IR2 & IR4	ρ ≥ 0.9 for VIS ρ ≥ 0.96 for IR2 & IR4
Cloudiness	< 5%	< 5%
Maximum Time difference between 2 frames for FFR	120 minutes	120 minutes

4. DISCRIMINATING VALID AND INVALID MEASUREMENTS

Any automatic landmark measurement system – supervised or unsupervised – will produce a population of invalid data. That population, in turn, places verification of specification compliance at risk. To sharpen our meaning, let us state that a measurement is "valid" when it relates to the performance of the system under test. Conversely, a measurement is "invalid" when it does not relate to the performance of the system under test. Unmasked cloudy pixels and poor IR thermal contrast would be the most common reasons that a landmark measurement from RPM would be invalid. Bad landmark morphology (straight lines, repeating shapes) that permits a coastline to match the image edges in multiple ways is another reason. The problem of recognizing when scenes are cloud contaminated, poorly illuminated in a VIS channel, or when thermal contrast is too low to discriminate land from sea in an IR channel is solved in the human mind by a process that one might call "judgment". The RPM implements its judgment through the rule-based QM assigned to each absolute error measurement as described in Subsection 3.3. The RPM is rather successful in identifying invalid data according to the test that a landmark is valid if and only if $QM \ge QM_{Valid}$ (a threshold typically ~0.9). Nonetheless, its judgment is not perfect and neither is that of a human. Two types of errors are possible according to the diagram in Figure 4-1.

Figure 4-1. Type I and Type II Errors in the Validity Hypothesis Test

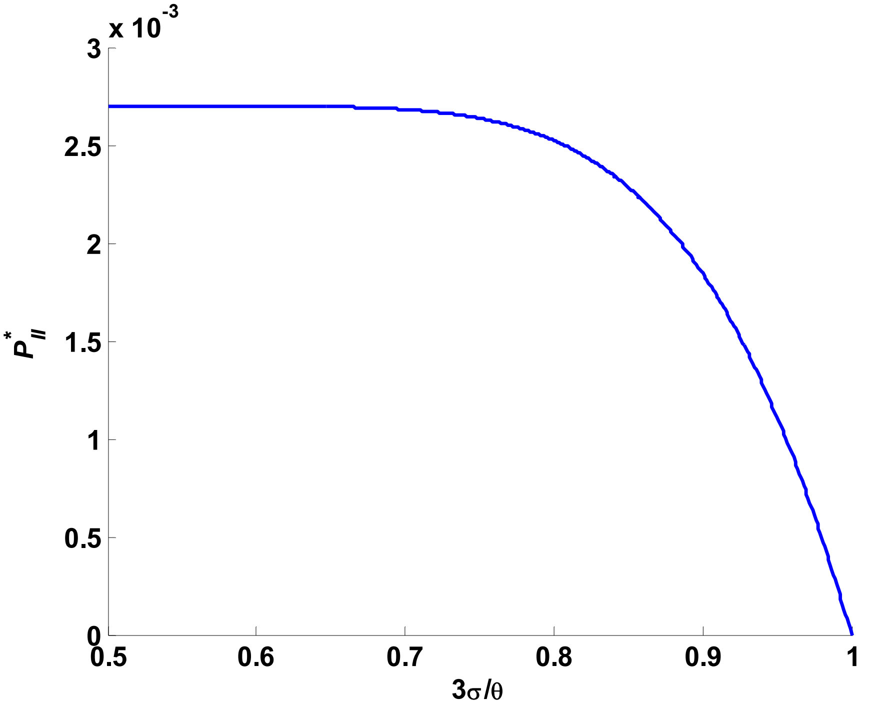
		Decision	
		$QM < QM_{VALID}$ **Reject H_0**	**$QM \ge QM_{VALID}$** **Accept H_0**
Truth	**H_0: Landmark is Valid.**	**Type I Error**	**Correct**
	H_1: Landmark is Invalid.	**Correct**	**Type II Error**

Type II errors ($QM \ge QM_{Valid}$ when measurement is really invalid) are particularly problematic for a 99.7-percentile "3-σ" test. The reliability of the QM must be held to progressively higher standards as system performance approaches the specification 3σ-threshold θ. Suppose that the underlying system performance is $N(0,\sigma)$; *i.e.*, normally distributed with mean zero and variance σ^2 and that this performance is measured in the presence of an invalid population. If such measurements were truly invalid, one would expect that the invalid population would be $U(-D,+D)$; *i.e.*, uniformly distributed between ±*D*. The probability of making a type-II error (P_{II}) that puts the test at the threshold of expected failure (expected noncompliant population > 0.3%) can be estimated from the underlying distributions when $D \gg \theta$:

$$P_{II}^{*} = 1 - \frac{erf\left(3/\sqrt{2}\right)}{erf\left(\dfrac{\theta}{\sqrt{2}\sigma}\right)} \quad . \qquad (4\text{-}1)$$

Equation (4-1) is plotted in Figure 4-2 as a function of $3\sigma/\theta$ (system performance in relation to the specification threshold). When system performance is better than the specification ($3\sigma/\theta < 0.7$) the validity test needs to be at least 3σ reliable; however, as the system performance approaches the specification threshold, the validity test needs to become asymptotically infallible to avoid an expectation of failure.

Figure 4-2. Threshold Type-II Error Probability to Avoid Expected Failure



Even with proper VIS and IR radiometric calibration and proper QM tuning, it is admitted that the RPM will pass some invalid landmarks as valid ones (a type-II error). To gain margin in a 99.7-percentile test, some additional screening is necessary. Statistical outlier testing against a model is one obvious choice. The OATS implements such screening when it ingests landmarks for orbit and attitude estimation. Landmarks with implausible residuals with respect to the model are rejected. The RPM also has similar capabilities provided by its internal Kalman Filter. The RPM Kalman Filter tracks navigation error in real time so that land-sea masks can be accurately registered to radiometric neighborhoods for cloud detection purposes. In tracking navigation error, the RPM tests the Kalman Filter innovation against a threshold. If the innovation is too large then it is considered to be physically implausible and the Kalman Filter state is not updated. The RPM can also be configured to not pass such "vetoed" landmarks to OATS. When this feature was being tested with the GOES I-M series, it was noticed that some of the vetoed landmarks possessed large QM, appeared perfectly cloud free and perfectly well matched. This led us to develop the notion of "gold" landmarks: those landmark measurements satisfying the super-quality criterion $QM \geq QM_{Gold} > QM_{Valid}$. Gold landmarks cannot be vetoed by the RPM Kalman Filter. Gold landmarks are apparently valid measurements although they fall beyond the tails of a normal distribution.

Kalman Filter innovation testing and other statistical outlier testing at the 3σ-level are appropriate for parameter estimation, but seem akin to cheating in a 3-σ specification compliance test. For this reason, we sought to develop an additional screening step to identify the best of the best (“platinum”) landmark measurements using a test that is not a statistical outlier test in the domain of the specification. Platinum status is conferred on a landmark according to a criterion of consistency between consecutive absolute error measurements and their corresponding relative error measurement. The absolute and relative error algorithms are completely independent. An absolute error measurement that is an outlier with respect to its own population but is consistent with its relative error partner can earn platinum status. The datum of inconsistency between a consecutive pair of absolute measurements (A_n, A_{n-1}) and its corresponding relative measurement ($R_{n,n-1}$) is

$$i_n = A_n - A_{n-1} - R_{n,n-1} . \qquad (4\text{-}2)$$

We form the inconsistency dataset by taking all consecutive absolute error measurements passing the QM validity test that also have high quality relative error measurement partners. Figure 4-3 shows a population of inconsistency metrics for a GOES-12 dataset. Those that are concentrated in the core at (0,0) are the most consistent. The core region can be defined by an *n*σ-error ellipse. Let $M = E[ii^T]$ be the sample covariance about the origin (0,0). The equation $n^2 = x^T M^{-1} x$ defines the contour *x* of the *n*σ-error ellipse about (0,0). A practical choice is $n = 2$, for which the platinum test would be $x^T M^{-1} x < 4$. In the example of Figure 4-3, 88.4% of the points plotted are within the 2σ-error ellipse (close to expectations for a 2-dimensional normal distribution).

A further consistency check is added when simultaneous multi-spectral measurement pairs are present. In this case, one may suppose that the channel-to-channel alignment cannot change too rapidly and outliers in the coregistration datum $C^{N,M} = A^N - A^M$ for channels *N* and *M* from the same frame and the same site can be cast out. We call those cast out “tarnished” platinum landmarks. A population of coregistration statistics $C^{IR4,IR2} = A^{IR4} - A^{IR2}$ is plotted in Figure 4-4. It is of course arguable whether such editing is appropriate for verification of a 3σ channel-to-channel coregistration requirement.

5. FLIGHT DATA ANALYSIS

A platinum landmark analysis is implemented in Matlab. The first pass identifies the high-quality absolute and relative error measurements, computes their inconsistency statistics, and models the coregistration error between channels. Those IR landmarks attaining platinum status are realigned to the VIS channel using the coregistration model. NAV error and FFR errors are plotted, WIFR statistics are computed, and the 99.7-th percentile points of the cumulative error distributions are estimated.

Figure 4-3. Inconsistency Data for GOES-12 IR2 Absolute and Relative Errors. Recall that IR2 is sampled at 64 μrad in the EW direction and 112 μrad in the NS direction. Those landmarks outside the 2σ-error ellipse are shown in red. Landmark data with inconsistencies inside the ellipse (black) are conferred platinum status.

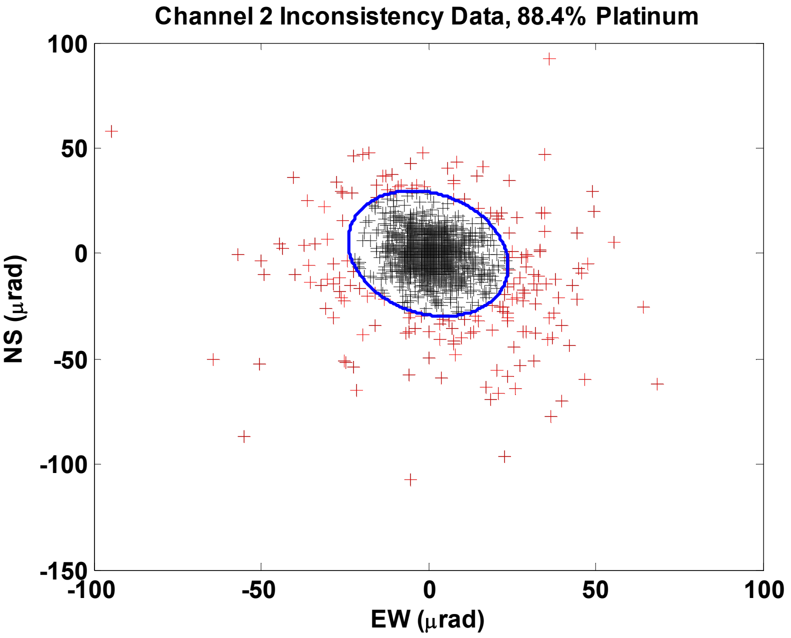
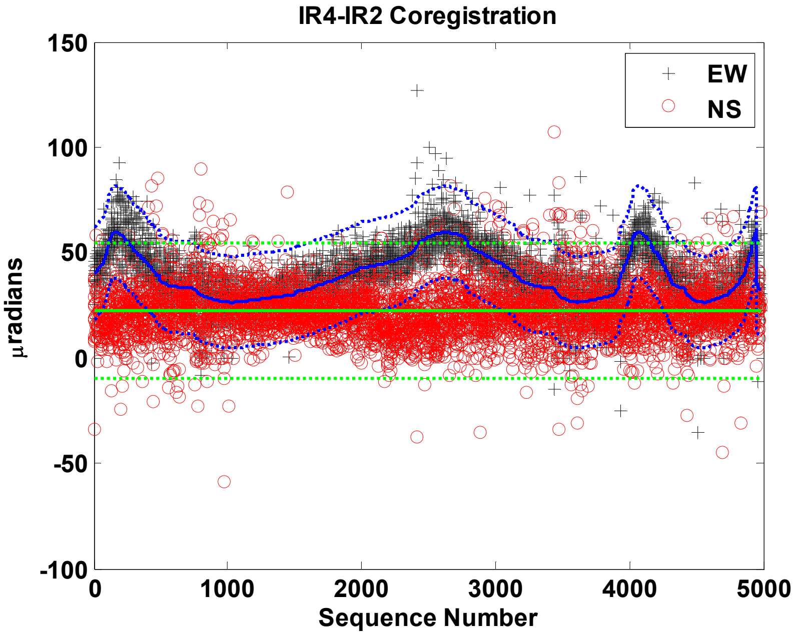


Figure 4-4. IR4-IR2 for GOES-12 Simultaneous Landmarks Plotted over 3 Days. Landmark absolute errors are plotted by chronological sequence number, not time, and fit with a harmonic model (EW) or a constant bias (NS). The ±3σ bands about the mean are plotted. Those falling outside the band are tarnished if they have otherwise earned platinum status.



The utility of the platinum landmark analysis is illustrated with VIS landmark data obtained from GOES-13 when it was taken out of storage in late summer of 2007. GOES-13 attitude determination and control should be very stable. One would generally not expect that valid landmarks would suddenly deviate from the main NAV error population on account of attitude determination and control errors. Such outliers could be edited with a statistical outlier test, but such a test presupposes what is reasonable behavior for the system. Indeed, rapid deviations could occur with anomalous Imager servo performance. Figure 5-1 shows the GOES-13 NAV error computed for all putatively valid VIS landmarks (QM ≥0.9). Those attaining platinum status are plotted in blue and those failing to attain platinum status are plotted in magenta. Figures 5-2 and 5-3 present similar measurements for 15-minute FFR and 90-minute FFR respectively. Figure 5-1 has five noncompliant landmarks from a total of 1781 "valid" landmarks, a percentage of 99.719%, which is at the threshold of failure. However, all five noncompliant landmarks are suspect as apparent statistical outliers. In fact, none of these five landmarks attain platinum status. Compliance with the specification is 100% when evaluated with the platinum standard. Similarly, two of 706 15-minute FFR data points and two of 922 90-minute data points are noncompliant; but, none are platinum quality.

Complete multi-spectral results for GOES-12 are shown in Figures 5-4 to 5-7. Figure 5-4 shows the derived coregistration model for GOES-12. Figure 5-5 is the NAV error with IR landmarks corrected so that they align with the VIS channel and Figures 5-6 and 5-7 show FFR metrics. IMC set transitions (frequently needed with the GOES I-M series) are marked on Figures 5-6 and 5-7. FFR error is not assessed across such transitions for GOES I-M; however, FFR error is assessed across daily IMC set transitions for GOES-N.

Figure 5-1. Navigation Error for GOES-13

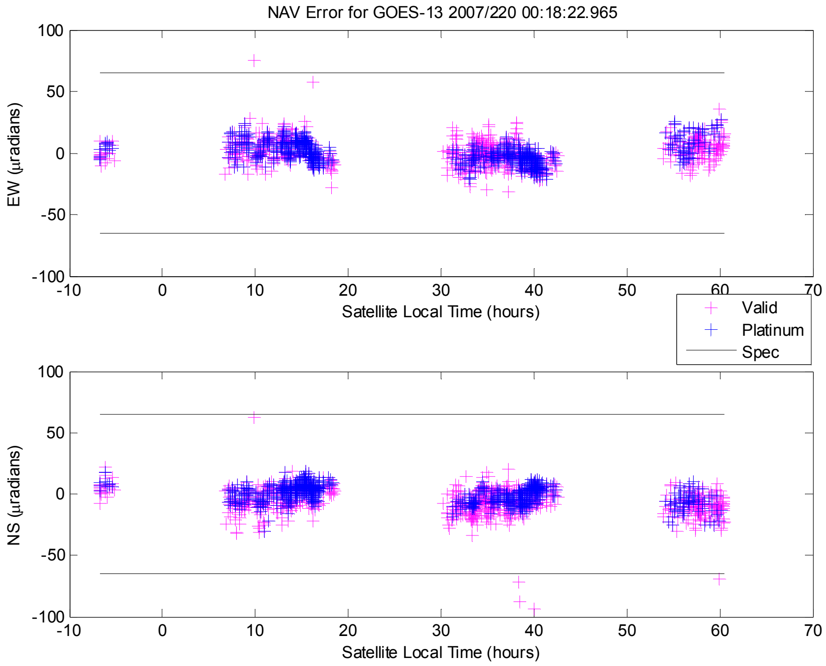


Figure 5-2. 15-Minute FFR Error for GOES-13

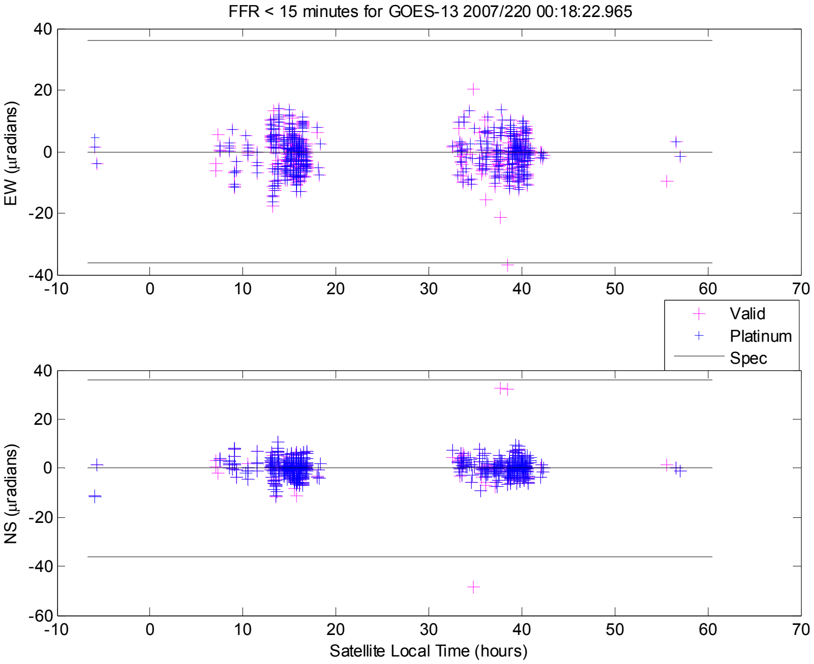


Figure 5-3. 90-Minute FFR Error for GOES-13

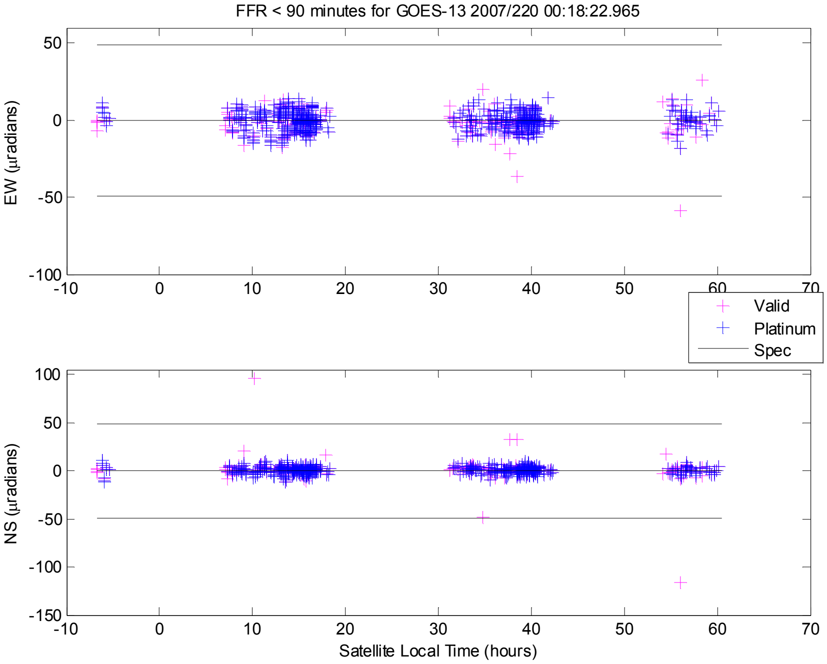


Figure 5-4. Coregistration Model for GOES-12

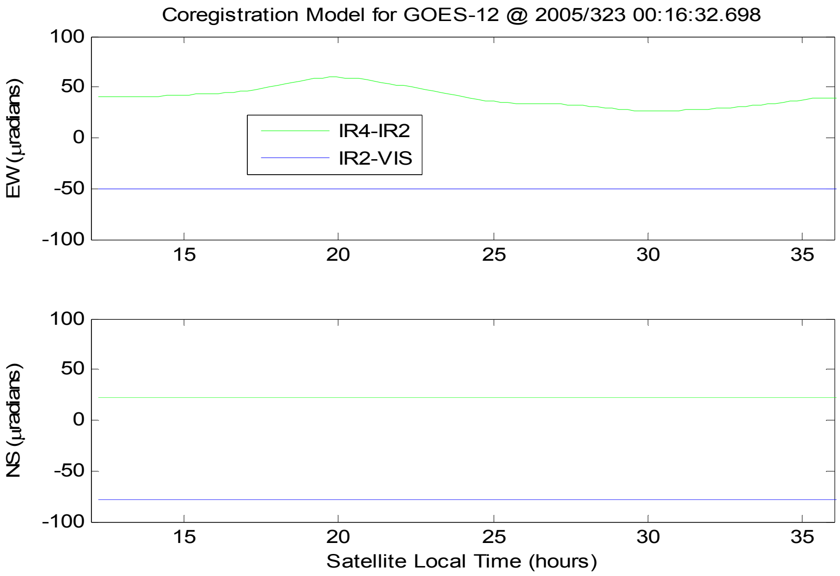


Figure 5-5. NAV Error for GOES-12

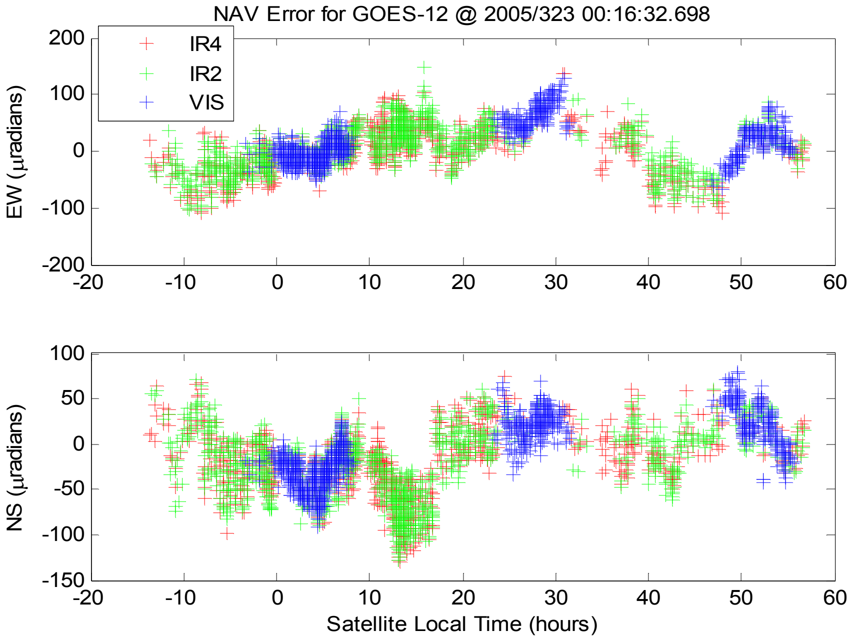


Figure 5-6. 15-Minute FFR Error for GOES-12

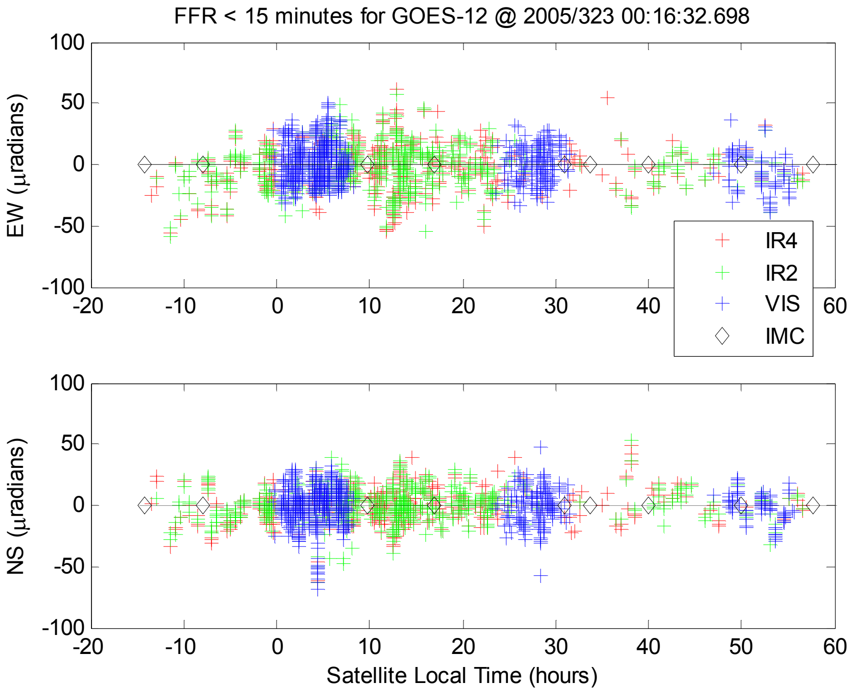
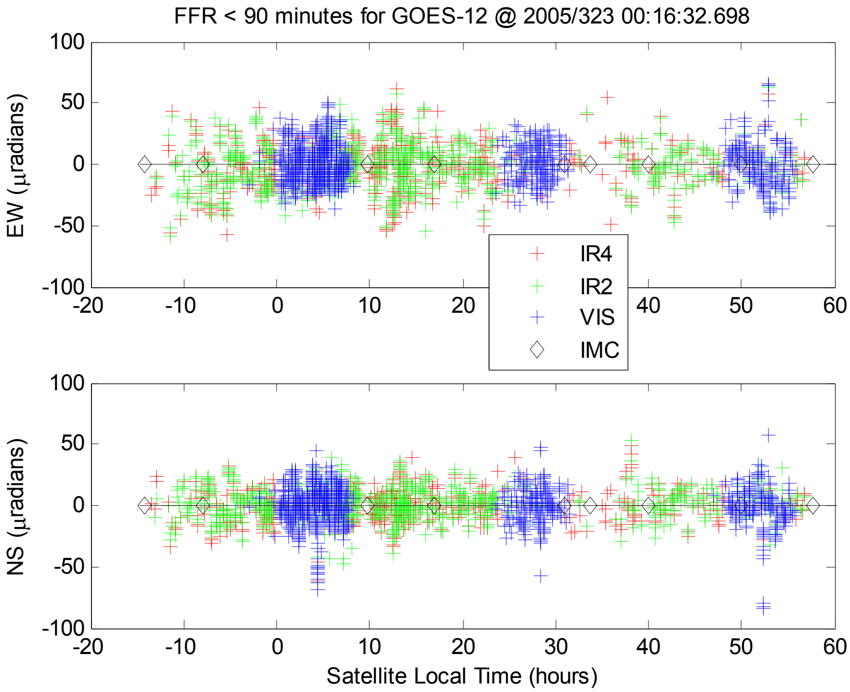


Figure 5-7. 90-Minute FFR Error for GOES-12



6. CONCLUSION

Platinum landmarks are distinguished by their self-consistency and are more suitable for verification of 3-σ performance requirements than the larger set of "valid" landmarks produced by the RPM. This larger set may contain invalid landmarks that have been falsely designated as valid. Invalid data must be edited from performance datasets unless the RPM quality screening is more reliable than the system under test. As the underlying system performance approaches the specification threshold, the RPM must approach infallibility. Statistical outlier editing in the domain of the specification presupposes what constitutes reasonable behavior for the system. The platinum landmark methodology avoids such assumptions within the domain of the test.

REFERENCES

---

[1] **GOES I-M Data Book**, NASA, DRL 101-08, 31 August 1996. Available for download at http://goes.gsfc.nasa.gov/text/goes.databook.html (current on 8/15/07).

[2] **GOES-N Data Book**, NASA, Available for download at http://goes.gsfc.nasa.gov/text/goes.databookn.html (current on 8/15/07).

[3] Madani, H., J. Carr, and Cathya Schoeser, "Image Registration using AutoLandmark", *Proceedings of the 2004 IEEE International Geoscience and Remote Sensing Symposium*, September 20-24, 2004.

[4] Wessel, P., and W. H. F. Smith, A Global Self-consistent, Hierarchical, High-resolution Shoreline Database, *J. Geophys. Res., 101*, 8741-8743, 1996.

[5] U.S. Department of Commerce NOAA/NESDIS, "Earth Location User's Guide (ELUG)", March 1998, URL current on August 12, 2007, Available for download at http://rsd.gsfc.nasa.gov/goes/text/ELUG0398.pdf (current on 8/15/07).

[6] Madani, H., J. Carr, and Cathya Schoeser, *op.cit.*, 2004.

[7] U.S. Department of Commerce NOAA/NESDIS/OSO, http://www.oso.noaa.gov/goes/inrstat/ (current on 8/15/07).