Improved Estimates of Spitzer Space Telescope Data Volumes with Error Bars

Marc A. Sarrel[1] and Michael Turmon[2]
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109 USA

The number of observations that may be made in a given period of time by the Spitzer Space Telescope is restricted by the amount of available on-board storage. The data volumes of the observations are estimated during observation sequencing. Overestimation may result in unnecessary idle time on the observatory, and Spitzer has a high-level project requirement to be 90% efficient in scheduling observations. On the other hand, underestimation may result in violation of a mission operations constraint to tolerate the loss of a single downlink session without filling the on-board storage. In this paper, we describe statistical methods to increase downlink utilization on Spitzer and enable more observations to be safely taken while taking on a planner-specified level of risk. By comparing predicted science data volumes to the historical record of actual, observed data volumes, we construct a model of the bias (tendency for over prediction) of the predictions for each of the three Spitzer instruments. We also find and validate a variance estimate for the data volume predictor. The variance allows us to quantify the probability of a mis-estimation of any given size. The form of our model allows us to precisely "roll up" the risks associated with three separate downlink sessions into a single high-level probability of buffer under-run which corresponds directly to a planner-enforced requirement.

Nomenclature

PAO_n	=	Period of Autonomous Operation *n* during which x_n is recorded to on-board storage
$Pass_n$	=	Downlink session *n*: data set *n* is transmitted and data set *n*–1 is deleted from on-board storage
x_n	=	Actual data volume recorded during PAO_n (measured in storage units)
$\hat{x}_n$	=	Predicted data volume recorded during PAO_n (measured in storage units)
$\tilde{x}_n$	=	Corrected data volume for PAO_n as determined from our procedure (measured in storage units)
σ_n^2	=	Variance of $\tilde{x}_n$ relative to x_n (measured in storage units squared)
y_n	=	Actual Missed-Pass Minimum (free space in the on-board storage) at the start of $Pass_n$ in case of a single fault (measured in storage units)
$\hat{y}_n$	=	Predicted Missed-Pass Minimum (free space in the on-board storage) at the start of $Pass_n$ in case of a single fault (measured in storage units)
$\tilde{y}_n$	=	Corrected Missed-Pass Minimum as determined from our procedure (measured in storage units)
v_n^2	=	Variance of $\tilde{y}_n$ relative to y_n (measured in storage units squared)
s_n	=	The instrument that is recording data during PAO_n
$P(A \mid B,C)$	=	The probability that A is true, given that B and C are true.
p_{nofill}	=	The minimum acceptable probability that the on-board storage will not fill in case of a single fault
h	=	Bandwidth of the smoother window (measured in storage units).
a_r	=	Coefficients of the *r*-degree polynomial to convert from predicted to corrected data volumes.
$E(s,r)$	=	Cumulative residual function
$K(x,\cdot)$	=	Localized weighting function about a given predicted data volume
ε	=	Error residual between predicted and actual PAO data volumes.

[1] Spitzer Mission Operations Systems Engineer, Systems Engineering Section, 4800 Oak Grove Dr. M/S 264-767, Member.
[2] Principal Engineer, Instrument Software and Science Data Systems Section, 4800 Oak Grove Dr. M/S 306-463.

TP	=	True Positives
FP	=	False Positives
TN	=	True Negatives
FN	=	False Negatives
α	=	False Positive Rate
β	=	True Positive Rate
m	=	Mitigation Rate

I. Introduction

THE Spitzer Space Telescope is the fourth of NASA's Great Observatories.[1-3] It takes astronomical images in infrared light. It was launched in August of 2003, and is in an Earth-trailing, heliocentric orbit. It is cooled to within a few Kelvins of absolute zero by a supply of liquid helium. Spitzer has three science instruments, the Infrared Array Camera (IRAC), the Multi-band Image Photometer (MIPS), and the Infrared Spectrometer (IRS). The cryogen is predicted to run out between January and July of 2009. After that, it will still be possible to operate the IRAC instrument. But, only two of its four detectors will continue to produce valid data at the higher, passively cooled temperature.

Only one of the three instruments is powered on at a time. Each stays on for a campaign of from one to three weeks. There typically one or two times per day that Spitzer downlinks data to Earth. Sometimes, there can be as many as forty hours between telecommunication sessions.

Spitzer has a high-level requirement to operate at or above 90% efficiency. Astronomical observations, science instrument calibrations and spacecraft slews to science targets are considered to be efficient. All other activities, including slews to Earth (for the purpose of downlinking data) and time spent down linking data are considered to be inefficient. The largest single source of inefficiency is the time taken to downlink data from the spacecraft. During a typical week, the spacecraft spends between 3% and 11% of its time downlinking data. The project therefore wants to minimize both the number of telecommunication sessions, and the duration of each session. On the other hand, the project does not wish to fill the on-board storage. This places an upper limit on the time allowed between telecommunication sessions.

The major source of uncertainty in sizing and spacing the downlinks is the on-board lossless data compression algorithm. The project produces an estimated compressed data volume for each downlink. Before the analysis described in this paper was performed, the project had used a fixed limit on the size of downlinks to prevent overfilling the on-board storage. That is, the project assumed that the data volume predictions had no uncertainty in them at all.

Because there is uncertainty in the data volume predictions, and their accuracy is not constant, this fixed limit necessarily had a considerable amount of conservatism built in, reducing operational efficiency. The goal of this paper is to characterize the mean and error bars of actual observed data volumes as a function of predicted data volumes, and thereby allow the project to quantitatively assess the probability of filling the on-board storage (XXX alternate wording: …probability of transgressing customary margins for on-board storage). The project can then operate the spacecraft as efficiently as possible while not being overly conservative or aggressive with the predicted data volumes.

II. Management of the Spitzer On-board Storage

A. Single-Fault Tolerance

Spitzer has a two-phase data transmission and acknowledgment system.[4] That is, a set of data is downlinked on one telecommunications pass and then deleted via ground command on the next. Data that were not received on the ground are retransmitted rather than deleted. In the nominal case, at the start of each downlink the spacecraft has the two most recent sets of data in the on-board storage. The most recent set will be downlinked, and the set before that will be deleted. At the end of each downlink, only one set of data remains on board. A single fault in this system will result in the need to store three sets of data at once, rather than the nominal two. Therefore, to be single-fault tolerant, the project must limit the data volumes such that any three consecutive sets of data will fit in the on-board storage.

A single fault is any problem that prevents either sending the commands to delete the previous set of data, or that prevents receiving the current set of data on the ground. For example, if the project is unable to send the data-acknowledgment commands during $Pass_n$, there will still be two sets of data at the end of $Pass_n$. Consequently, there will be three sets of data at the start of $Pass_{n+1}$. In this case, the project typically is able to send commands during $Pass_{n+1}$ to delete both sets of data, and thereby leave only one set of data on board at the end of $Pass_{n+1}$. A similar sequence occurs when the project fails to receive data. If the project receives no data during $Pass_n$, it will have no acknowledgment command to send during $Pass_{n+1}$. At the end of $Pass_n$ there will be one set of data, because the project sent the acknowledgment command during $Pass_n$. But, at the end $Pass_{n+1}$ there will be two sets of data. It will take several passes to retransmit the missed data and thereby clear the backlog. If both problems, failure to send the command and failure to receive the data, occur at the same time, during $Pass_n$, the result is two independent single failures. The spacecraft would have two sets of data at the end of $Pass_n$ and at the end of $Pass_{n+1}$, but would never have four sets of data at once.

If there is more than a single fault, some observations may be skipped to prevent filling the on-board storage. Before each astronomical observation begins, the spacecraft checks to make sure that there is enough free space in the on-board storage to accommodate the predicted data volume of that individual observation plus 7500 SU of margin. Filling the on-board storage has a large chance of putting the spacecraft into standby or safe mode. An entry into either mode would halt science observations, and would take several days to recover from. Under normal circumstances there is always enough space in the on-board storage, the check succeeds, and the observation is run. However, if there has been more than a single fault in the data-transmission system, the check may fail and, if so, the observation is skipped. In this case, the spacecraft performs the initial slew to the target, but sits idle for the duration of the observation. It then continues with the rest of the observations as before. This fill-avoidance check was implemented to minimize the consequences of a failure. But, the project of course wishes to avoid even this minimized consequence whenever possible.

B. Existing Method to Manage Data Volumes

It takes thirty-seven days to develop a sequence of commands for the Spitzer spacecraft to perform science observations. Each science sequence takes approximately seven days to execute on the spacecraft. At any given time, there are five or six science sequences in development and one executing on the spacecraft. The project manages data volumes in two different ways. Before the science sequence begins execution, the project checks that each series of three consecutive sets of data will fit in the on-board storage. During the execution of the science sequence, the project observes the actual data volumes and state of the on-board storage. The project uses that information to predict the future state of the on-board storage, taking into account any faults that have occurred. In the remainder of this paper, we will focus on the pre-execution analysis.

About three weeks before the science sequence is set to begin execution on the spacecraft, the project assesses the predicted data volumes. The fundamental calculation used to analyze the predicted data volumes is shown in Eq. (1).

$$\hat{y}_n = 121\,732 - (\hat{x}_{n-2} + \hat{x}_{n-1} + \hat{x}_n) \tag{1}$$

The predicted missed-pass minimum for $Pass_n$, $\hat{y}_n$, gives the predicted amount of free on-board storage that would remain in case of a single fault. That is, in case the spacecraft had to store three sets of data: $\hat{x}_{n-2}$, $\hat{x}_{n-1}$, and $\hat{x}_n$ at the same time. All data volumes are measured in Storage Units (SU). The on-board storage is 121 732 SU in size. This calculation for predicted missed-pass minimum is performed for each telecommunications session. The criterion for deciding if the $\hat{y}_n$ is acceptable is shown in Eq. (2).

$$\hat{y}_n > 7500 \tag{2}$$

The value of 7500 SU was chosen to allow margin for twenty-four hours of engineering telemetry, and for underestimation of science data volumes. If the on-board storage were actually to fill, there is a large probability that the spacecraft would halt operations and enter standby mode. It typically takes two or three days to resume normal operations following an entry into standby mode. At the beginning of each science observation, the spacecraft checks to see if there is enough free space in the on-board storage for the predicted data volume of that observation plus the 7500 SU margin. If there is not, the observatory does not execute the observation and sits idle for that observation's predicted duration. This protective mechanism prevents a science observation from filling the on-board storage. It has been triggered five to ten times since it was implemented about two years ago. However, no such protective measure can be put in place for engineering telemetry. Engineering telemetry is written at a rate of about 210 SU/h.

If $\hat{y}_n < 7500$ SU, we say that $Pass_n$ is not single-fault tolerant. The data volumes $\hat{x}_{n-2}$, $\hat{x}_{n-1}$ and $\hat{x}_n$ are collectively too large and some mitigation measures must be taken, such as moving science observations from one PAO to another, getting backup ground antennas, or removing science observations altogether; Ref. 1 details these and other strategies.

If $\hat{y}_n \geq 7500$ SU, we say that $Pass_n$ is single-fault tolerant. If $\hat{y}_n$ is as large as 60 000 SU or 70 000 SU, then it may be possible to increase efficiency by removing one of the telecommunication sessions all together. That time, perhaps up to an hour, could then be used for new science observations. This method is somewhat limited because the times of the telecommunication sessions are fixed before the science observations are schedule. Therefore, it is possible to forgo a telecommunications session, but not to move it earlier or later.

C. New Method to Manage Data Volumes

The existing method described in the previous section assumes that the predicted data volumes have no uncertainty. That is, that the actual data volumes will exactly match the predicted. In reality there is quite a bit of uncertainty in the predicted data volumes. As discussed before, this uncertainty is due to the performance of the on-board lossless data-compression algorithm. So, in this paper, we wish to understand and account for this uncertainty.

This uncertainty leads to observational inefficiency. We must skew the predictions systematically in the direction of under-prediction. That is, so that the actual data volumes tend to come out smaller than predicted. This will be shown clearly in Section III. The most direct way to eliminate this inefficiency would be to improve the data volume prediction algorithms on the ground to reduce the variance directly. For several reasons, this is not practical for the Spitzer project at this phase of the mission. So, the approach we take is to better characterize the bias and variance of the existing predictions, and base our decisions whether to approve a set of data volumes on a probabilistic basis taking the variance into account.

First, we define y_n similarly to $\hat{y}_n$, as shown in Eq. (3).

$$y_n = 121\,732 - (x_{n-2} + x_{n-1} + x_n) \tag{3}$$

The actual missed-pass minimum for $Pass_n$, y_n, gives the actual amount of free on-board storage that would remain in case of a single fault. That is, in case the spacecraft had to store three sets of data: x_{n-2}, x_{n-1} and x_n at the same time. Instead of the criterion in Eq. (2), we would like to use as the criterion the probability that y_n will be greater than 7500 given the predicted data volumes and test that number against a given lower limit.

$$P(y_n > 7500 \mid \hat{y}_n) \geq p_{nofill} \tag{4}$$

In Eq. (4), p_{nofill} is the given lower limit of risk that the project has chosen to accept. Its value can range between zero and one. But, for example, p_{nofill} would most likely be in the range of 0.95 to 0.99. Unfortunately, Eq. (4) is too simple for this problem. The variance of the predicted data volumes, and hence of the predicted missed-pass minimum via Eq. (1), is a function of both the predicted data volumes, and which science instrument is on during those PAOs. For a given constant y_n, the associated variance can change significantly depending on the sizes of the three data volumes and which instruments produced which data volumes. So, Eq. (4) would yield unnecessarily large variances in the y_n in a large number of the cases. Instead of the predicted missed-pass minimum, we must use the individual predicted data volumes and instruments for each of x_{n-2}, x_{n-1} and x_n as shown in Eq. (5). This method lets us reduce the variances of the actual missed minimum for those cases that allow it.

$$P(y_n > 7500 \mid s_{n-2},\ \hat{x}_{n-2},\ s_{n-1},\ \hat{x}_{n-1},\ s_n,\ \hat{x}_n) \geq p_{nofill} \tag{5}$$

We assume that there is no uncertainty about which of the three science instruments is on for any given PAO. So, we make no distinction between predicted and actual for it. In the next section, we describe our data set and how we approach the solution of Eq. (5).

III. Data Sets and Statistical Approach

A. The Training Data Set

The data set analyzed in this paper is composed of predicted data volumes, actual data volumes, predicted missed-pass minimums and actual missed-pass minimums for 550 PAOs. The first PAO in the data set is from 2006-10-19 and the last from 2007-10-31, a span of 377 days or almost 54 weeks.

Table 1. Sample of the data set.

		Predicted		Actual	
Num	Instrument	Missed Pass Minimum	PAO Volume	Missed Pass Minimum	PAO Volume
127	IRS	47,105	23,768	47,752	23,397
128	IRS	29,445	41,588	30,077	41,537
129	IRS	16,518	39,858	17,104	39,694
130	IRAC	14,060	26,226	16,027	24,474
131	IRAC	20,598	35,050	23,206	34,358
132	IRAC			24,569	38,331
133	IRAC		24,761	24,981	24,062
134	IRAC		18,122	42,004	17,335
135	IRAC	56,010	22,839	59,621	20,714
136	IRAC	52,043	28,728	57,631	26,052
137	IRAC	28,617	41,548	37,704	37,262
138	IRAC	23,347	28,109	33,988	24,430
139	IRAC			24,622	35,418
140	IRAC			31,785	30,099
141	IRAC		29,155	31,014	25,201
142	MIPS		35,234	33,448	32,984
143	MIPS	29,173	28,170	36,790	26,757
144	MIPS	20,779	37,549	27,927	34,064

We chose this data set because the ground configuration related to producing estimates of data volumes was stable during this period. And, the spacecraft and ground stations performed well during this period, so that virtually all data were received on the ground. Nonetheless, there were several instances where we had to exclude data from the analysis.

Table 1 shows a few sample rows from the data set. We have numbered the five hundred and fifty rows from 85 to 634 to correspond to our project-internal tracking tool. Each row represents a single telecommunications session. Empty cells in the table represent invalid or excluded data. Three predicted missed-pass minimums, numbers 132 to 134, are excluded because of the one excluded predicted PAO volume number 132. Table 2 shows how the terms in Eqs. (1-5) correspond to the data set as shown in Table 1. In this scheme x_n is recorded to the on-board storage before $Pass_n$.

Table 2. Correspondence of the data set to the equations.

		Predicted		Actual	
Num	Instrument	Missed Pass Minimum	PAO Volume	Missed Pass Minimum	PAO Volume
n-2	s_{n-2}	$\hat{y}_{n-2}$	$\hat{x}_{n-2}$	y_{n-2}	x_{n-2}
n-1	s_{n-1}	$\hat{y}_{n-1}$	$\hat{x}_{n-1}$	y_{n-1}	x_{n-1}
n	s_n	$\hat{y}_n$	$\hat{x}_n$	y_n	x_n

The two adjacent and excluded PAO volumes in rows 139 and 140 combine to exclude four missed-pass minimums, numbers 139 to 142. See Eq. (1) for both cases. Also note that every excluded actual PAO volume resulted in three excluded actual missed-pass minimums. That is, none of the actual PAO volumes were close

Table 3. Summary of excluded (invalid) and included (valid) PAO data volumes.

	Number of PAO Volumes						
	Predicted		Actual		Pairs		
Instrument	Invalid	Valid	Invalid	Valid	Invalid	Valid	Total
IRAC	15	118	0	133	15	118	133
IRS	0	145	1	144	1	144	145
MIPS	5	267	3	269	8	264	272
Total	20	530	4	546	24	526	550

enough to overlap. A total of sixty out of a possible seventy-two pairs of missed-pass minimums were excluded. The net result is that we have 526 valid pairs of PAO data volumes and 490 valid pairs of missed-pass minimums, each out of a possible 550. See Tables 3 and 4.

Table 3 shows a summary of how many PAO volumes were excluded and how many were included. We need a valid pair of predicted and actual data volumes to include a given PAO in the analysis. The two columns under the

Table 4. Summary of excluded (invalid) and included (valid) missed-pass minimums.

	Number of Missed Pass Minimums						
	Predicted		Actual		Pairs		
	Invalid	Valid	Invalid	Valid	Invalid	Valid	Total
Total	50	500	12	538	60	490	550

Pairs heading show that there were no cases where both the actual and predicted volumes were invalid for the same PAO. That is, 20+4-24 = 0.

We excluded predicted PAO volumes as invalid for one of the following two reasons: 1) the predicted data volume was larger than could be transmitted in the corresponding telecommunications session – a bug in the prediction software causes a known error (11 exclusions), or 2) the PAO contained a certain type of observation with the IRAC instrument that produces a particularly large volume of data – the prediction algorithm does not work well for this class of observations (9 exclusions).

We excluded actual PAO volumes as invalid for one of the following two reasons: 1) science observations were autonomously skipped on board the spacecraft – this action is taken to prevent the on-board storage from filling when there is insufficient free space (2 exclusions), or 2) science observations were removed by ground intervention during sequence execution – this action allows the project to extend a telecommunications session and thereby clear a backlog of retransmitted data more quickly (2 exclusions).

Table 4 shows a summary of how many missed-pass minimums were excluded and how many were included. In order to count a missed-pass minimum as valid, the corresponding PAO volume and the two preceding PAO volumes must all be valid. See Eqs. (1, 3). That is, if one or more of these three PAO volumes were excluded, then the missed-pass minimum is excluded as well.

Unlike in Table 3, we do not show the numbers for each instrument. We cannot in all cases assign a missed-pass minimum to a single instrument. After each instrument transition the next two missed-pass minimums are mixtures of PAO data volumes from the incoming and outgoing instruments. For example, see rows 130, 131, 142, and 143 in Table 1. There are two missed-pass minimums for which both the actual and predicted values were invalid. That is, 50+12-60 = 2. These two are in fact just the first two missed-pass minimums in the data set. We expect this because it is never possible to calculate the missed-pass minimums for the first two elements of the data set. By comparing Tables 2 and 3, we can see that there only were fifty predicted missed-pass minimums that were excluded rather than the maximum of sixty. Sixty is the maximum because it is three times the number of excluded PAO volumes, twenty. However, in several cases, the excluded PAO volumes were close enough to each other so as to overlap, saving a net of ten missed-pass minimums from exclusion.

B. Characteristics of the Data Set

Figure 1 shows a scatter plot of actual versus predicted PAO data volumes for all 526 valid pairs. Data volumes for IRAC are colored red, those for IRS and green and those for MIPS are blue. The diagonal line shown in black has a slope of one and a *y*-intercept of zero. We can observe several patterns from this plot. First, the predicted PAO volumes for each of the three instruments show different characteristics of variance. IRS has the least variance; its data points are closely clustered. IRAC has the most variance. And, MIPS is in between the two. Second, the variances for IRAC and MIPS tend to increase as predicted data volume increases. Third, there is a certain amount of conservatism built in to the estimates. Most of the data points are below the diagonal line. That is, the actual PAO volume turned out to be less than predicted. Since the IRS data volumes show the most consistent variance, there is very little if any conservatism built in for that instrument. It is this conservatism that is a source of observational inefficiency for Spitzer, and which we want to eliminate using the analysis described in this paper. Since the predictions for IRS are already well behaved, the most benefit will be gained for the IRAC and MIPS instruments.

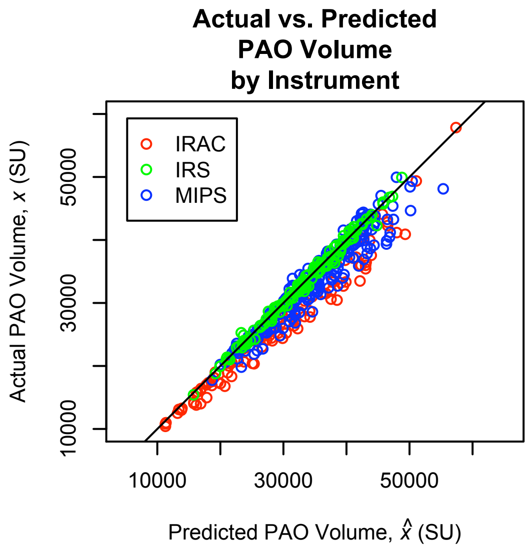

Figure 1. Actual vs. Predicted PAO Volume by Instrument

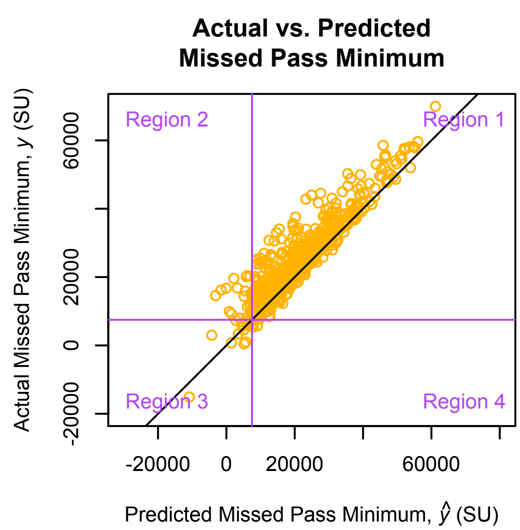

Figure 2. Actual vs. Predicted Missed-pass Minimum

Figure 2 shows a scatter plot of actual versus predicted missed-pass minimums for all 490 valid pairs. The diagonal line shown in black has a slope of one and a *y*-intercept of zero. In Fig. 2, the conservatism of data volume prediction is shown by most of the data points being above the diagonal line. That is, the actual missed-pass minimum turned out to be larger than predicted. This is the reverse sense compared to the data volumes in Fig. 1. This reversal is caused by the minus signs in Eqs. (1 and 3). The graph in Fig. 2 is additionally divided into four regions, which are separated by the vertical and horizontal purple lines. These lines are each located at 7500 SU on their respective axes. This is the same 7500 SU limit from Eqs. (2, 4 and 5). The 7500 SU limit is important because of the automated on-board fill-avoidance check described in Section II. A.

Data points in region 1 represent times when both the predicted and actual missed-pass minimums were greater than 7500 SU. They represent little risk of filling the on-board storage, although to the extent that the actual missed-pass minimum was greater than the predicted, they are a potential source of inefficiency. That is, there might have been needless idle time in those PAOs when the predicted missed-pass minimum was close to 7500 SU, near the left edge of region 1, because the project incorrectly thought that it was constrained by the on-board storage. Data points in region 2 represent times when the predicted missed-pass minimum was less than or equal to 7500 SU, but the actual was greater than 7500 SU. All the points in this region represent potential sources of inefficiency as above since the project incorrectly thought that we were constrained by on-board storage. In addition, when the predicted missed-pass minimum is less than 7500 SU, the project takes some mitigation measure to reduce the risk of filling the on-board storage. See Ref. 4 for details. These mitigation measures add some operational risk and complexity of their own, and the project would like to minimize their use. The project sometimes chooses to take mitigation measures when the predicted missed-pass minimum is above, but still close to the 7500 SU limit. This is true especially when the IRAC instrument is on. Data points in region 3 represent times when the both the predicted and actual missed-pass minimums were less than or equal to 7500 SU. They represent times when the risk of filling the on-board storage is real and the project correctly knew this and took the appropriate mitigation measures. As before, the project would like to minimize the need for mitigation. And, there is some limit to the magnitude of excess data volume for which the mitigation measures can compensate. For planning purposes, this limit is generally around 12 000 SU. So, any missed-pass minimum below about -4500 SU is too low and cannot be completely mitigated. Data points in region 4 represent times when the predicted missed-pass minimum is greater than 7500 SU, but the actual is less than or equal to 7500 SU. Points in region 4 are the most dangerous. The represent times when the project incorrectly believed that no mitigation measures were needed. In general, the goals are to keep the data points in Fig. 2 as close to the diagonal line as possible while minimizing the number of points in regions 4, 2 and 3 in that priority order. As we discussed before, the project is not able at this time to improve the data-volume prediction algorithm and thereby reduce the variance in Fig. 2 directly. Table 5 shows the number of missed-pass minimum data points in each region of Fig. 2.

IV. Application of Statistical Analysis to Spitzer Mission Operations

The buffer overflow constraint is expressed directly in terms of MPM: we require that with high probability, $y_n > 7500$ SU. To quantify this probability, first, we show how to improve the biased estimate of y_n. Second, we

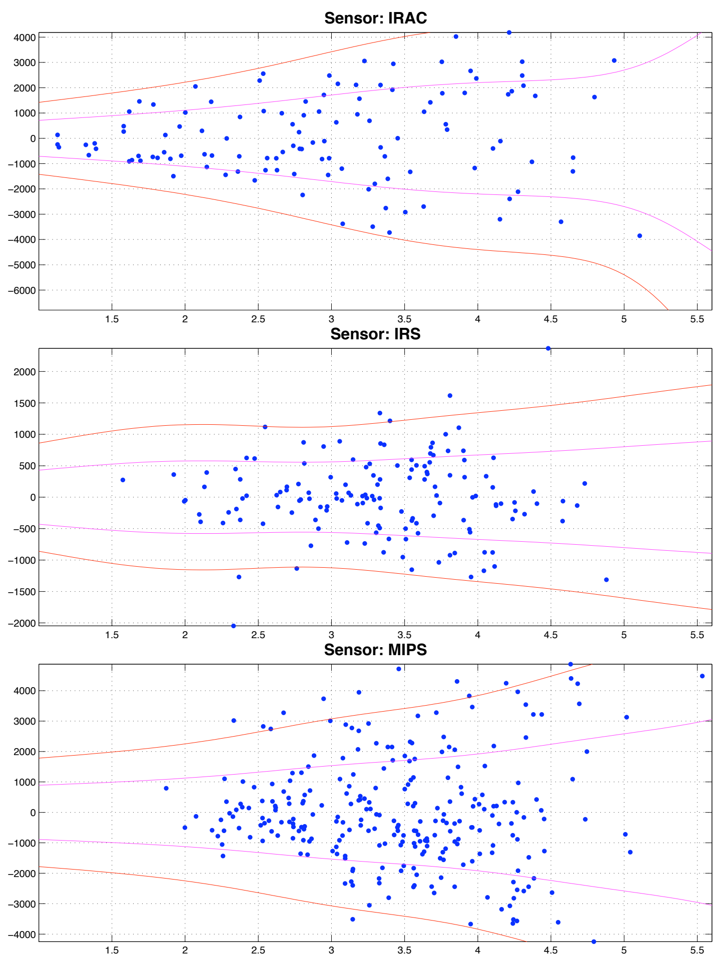

Figure 3: Error residuals versus prediction, for each instrument.

find a variance estimate for the error of our corrected prediction. In this way, the probability of buffer overrun can be estimated from historical data, and then used for future decisions. The simple relationship, shown in Eqs. (1, 3), of MPM to the three preceding PAO values makes it easy to proceed, bottom-up, from corrected PAO values to corrected MPM values. Accordingly, we now describe the PAO correction and accompanying variance estimate.

The tools we have to correct the bias in the estimate are the estimate itself, $\hat{x}_n$, and the instrument tag s_n. As seen above in Fig. 1, the instrument tag is an important factor in determining the prediction error. Note that, unlike MPM, the PAO values are not linked in time, so the samples ($\hat{x}_n$) can be viewed as independent in time. We adopt an r-degree polynomial regression model for $\tilde{x}_n$ as an estimate of x_n:

Table 5. Polynomial fits and residuals for each instrument type.

Instrument	a_0 (SU)	a_1 (SU)	E_0 (SU)	E (r = 1) (SU)	E (r = 2) (SU)	E (r = 3) (SU)
IRAC	638	0.8788	3680	1823	1787	1707
IRS	1.4	0.9963	635	622	621	618
MIPS	1135	0.9298	2243	1777	1763	1746

$$\tilde{x}_n = a_0 + a_1\hat{x}_n + \cdots + a_r(\hat{x}_n)^r \tag{6}$$

where we select the polynomial degree r in response to the data. The coefficients a are a function of s_n, but the dependence has been suppressed for compactness.

Table 5 shows the $r = 1$ correction coefficients for each sensor. The strong linear effect ($\tilde{x}_n \approx \hat{x}_n$) is evident, as is the conservatism (the linear coefficient in each case is less than unity). Of course, we expect a strong linear relation between observed data volumes and $\hat{x}_n$, but we fitted up to third-degree polynomials in case other effects were present. The benefit to including higher-order terms can be measured by the cumulative residual E, which we define as the per-sample RMS error remaining after correction, as a function of sensor s and polynomial degree r:

$$E(s,r) = \Big[\frac{1}{N(s)} \sum_{\substack{n=1 \\ s(n)=s}}^{N} (x_n - \tilde{x}_n)^2\Big]^{1/2} \tag{6a}$$

The sum extends only over the training data for a given sensor, and $N(s)$ is the number of samples for sensor s. With the scaling shown, the residual is in units of SU. For calibration, we also find E_0, the residual with no correction, that is, where $\tilde{x}_n = \hat{x}_n$. To justify higher-order terms, the drop in residual would have to be significant in relation to E_0. For IRS and MIPS, there is clearly no benefit to higher-order corrections. For IRAC, there seems to be some benefit to cubic terms. But on inspection, over half of the drop in IRAC residual is due to a single point at the edge of the domain, so we have used only a linear correction. The resulting residuals are shown in the left-hand panels of Fig. 3 as a function of $\hat{x}_n$.

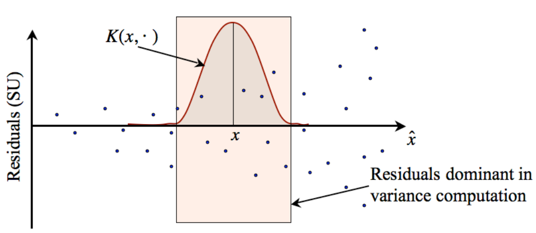

Figure 4: Schematic of local variance estimation.

This correction removes the bias, or conservative estimation, that was built in to the PAO predictions. The next step is to derive the variance estimate σ_n^2. As above, the variance is a function of both $\hat{x}_n$ and s_n. The best starting point is the residual plots in Fig. 3. The variance typically increases as the prediction size increases. However, it is unclear that there is a simply parameterized expression for the variance – e.g., it does not appear to be linear with $\hat{x}_n$. We preferred not to make any *a priori* assumptions about the variance structure, and chose to estimate it non-parametrically, as a weighted average of nearby residuals.

That is, for each instrument type, s

$$\sigma^2(s,x) = \sum_{\substack{i=1 \\ s(i)=s}}^{N} K(x,\hat{x}_i)\ (x_i - \hat{x}_i)^2 \tag{7}$$

where the $N(s)$ weights $K(x,\cdot)$ are proportional to a Gaussian window centered at x with given bandwidth h, i.e.

$$K(x,\hat{x}_i) \propto \exp -\tfrac{1}{2}\left[(x-\hat{x}_i)/h\right]^2, \text{ and } \sum_i K(x,\hat{x}_i)=1 \tag{7a}$$

This is notationally cumbersome, but conceptually simple. Figure 4 shows the idea: a localized weighting function (or "kernel") K is placed about a point of interest x and the points falling within the core of the weights contribute most strongly to the variance at x. The set of all residuals in the training data is plotted as points. The variance estimate at some point x is defined as the weighted average, across all points, of the squared residuals. All residuals appear in the summation, but they are weighted by the kernel centered at x, so the points within the marked square contribute most significantly to the weighed average. The bandwidth is chosen as $h = 5000$ SU, a compromise between smoothing and sensitivity to local variance fluctuations. The resulting variance curves are overlaid on the residuals in the right-hand panels of Fig. 3. The results are not sensitive to our particular choice of smoothing kernel and bandwidth.

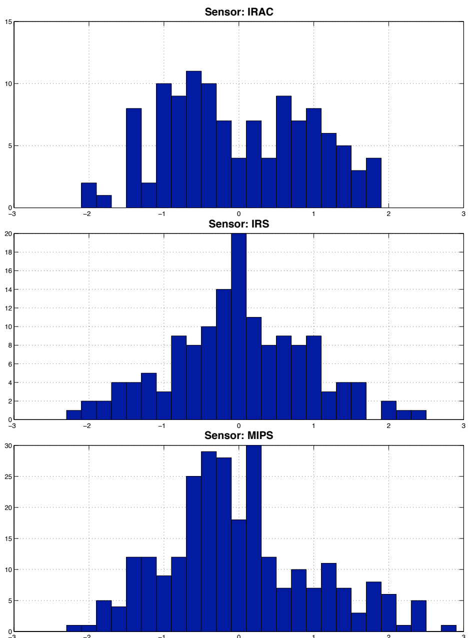

Figure 5: Histograms of normalized estimation-error residuals for each sensor.

This method of determining a localized variance was examined by Carroll (1982), in the context of a heteroscedastic linear regression model. In our notation, this would model the observed PAO volume x as

$$x \;=\; (a_0 + a_1\hat{x}) \;+\; \sigma(s,\hat{x}$$

where σ is as above, and the errors ε are symmetric about zero and have unit variance. The first term is deterministic, and is the center of the scatter of possible observed PAO data volumes. The second term is a scale factor which influences spread of data volume errors. The scale factor depends on both the sensor type and the size of the estimated data volume. Earlier work (e.g., Hildreth and Houck 1968) considered parametric expressions for the scale factor σ, and subsequent work has considered the interplay between the deterministic and random components (Hall and Carroll 1989, Wang, Brown, Cai and Levine 2008), a differencing method which partly breaks this interplay (Levine and Brown 2007), and the case when $\hat{x}$ is multivariate (Spokoiny 2002; Cai, Levine, and Wang 2006).

After selecting the correction coefficients of Table 5, and establishing the method of computing variances, the error model for x is complete. The next step is to examine the standardized residuals to determine if the error distribution is nearly-Gaussian. The standardized residuals correspond to the ε term above. By dividing by the local standard deviation, we put them on the same scale so they may be compared directly. The most straightforward method is to compute histograms, which are shown in Fig. 5. IRS and MIPS are reasonably well-fit by a Gaussian distribution, but IRAC seems to depart from Gaussianity. Modeling this departure could be computationally prohibitive in our context (we prefer a spreadsheet implementation), so we have chosen to invoke the Gaussian assumption for IRAC as well.

The corrected PAO model is easy to roll up into a MPM model via this analog to Eqs. (1, 3):

$$\tilde{y}_n \;=\; 121732 \;-\; (\tilde{x}_{n-2} + \tilde{x}_{n-1} + \tilde{x}_n) \tag{8}$$

$$v_n^2 = \sigma_{n-2}^2 + \sigma_{n-1}^2 + \sigma_n^2 \tag{9}$$

where we have used that means add, and that variances are additive for independent prediction errors. Here, the three terms contributing to $\tilde{y}_n$ are determined via (6) and depend on $(\hat{x}_{n-2}, s_{n-2})$, $(\hat{x}_{n-1}, s_{n-1})$, and $(\hat{x}_n, s_n)$ respectively. Similarly, the three variance terms depend on $(\hat{x}_{n-2}, s_{n-2})$, etc., and each is computed using (7). In many cases, s_n is different among the three constituent PAOs. By the Gaussian assumption, we predict that the MPM y is Gaussian distributed about the mean value $\tilde{y}_n$, and with variance v_n^2, which allows us to compute confidence intervals of any type, including (6).

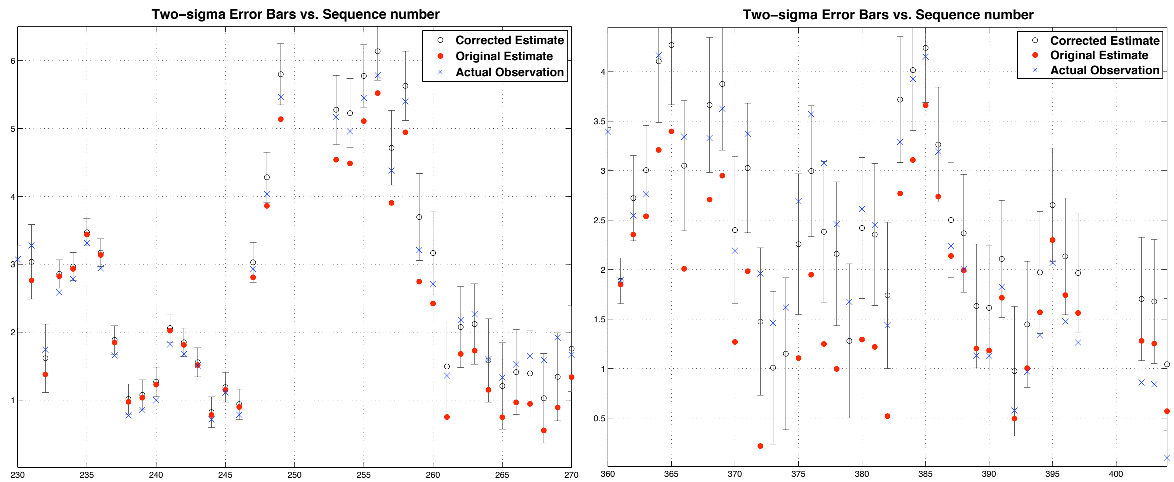

Figure 6. Probable Intervals

Figure 6 shows the effectiveness of this error model for y by using the training data. (Our model has so few parameters that over-fitting seems unlikely, making it sensible to evaluate performance on training data. See section V below for results on a separate holdout set.) The horizontal axes show MPM value in two segments of the 487-value training sequence. The error bars are centered on the black circles – we have plotted the two-sigma error bars, which should contain 95% of the samples. The original, uncorrected MPM estimate is shown in red, and the actual MPM value is the blue cross, almost all of which lie in the expected range. The pervasive undershoot of the original MPM estimates is also clear, so the value of the correction is clear. As for the error bars, of the 487 MPM values, we would expect 22 to fall outside the two-sigma error bars; in fact, 59 values do. The non-Gaussian errors referred to above cause part of this. Another cause is that the errors ε of (7b) are not independent as a function of pass number n, so there are additional cross-terms in the variance (9).

V. Model Validation

A. The Holdout Data Set

The data set used to validate the statistical model is identical in form to the training data set described in section III A. It contains data for 325 PAOs, starting 2007-11-02 to 2008-07-10, a span of 241 days or almost 35 weeks. This data set is compatible with the training data set because the ground configuration is the same as for the training data set. There were a few more operational problems with ground equipment, so proportionately more data points

Table 6. Summary of excluded (invalid) and included (valid) PAO data volumes from the holdout data set.

	Number of PAO Volumes -- Holdout						
	Predicted		Actual		Pairs		
Instrument	Invalid	Valid	Invalid	Valid	Invalid	Valid	Total
IRAC	9	92	0	101	9	92	101
IRS	8	101	2	107	9	100	109
MIPS	14	101	0	115	14	101	115
Total	31	294	2	323	32	293	325

were excluded. Data were excluded for the same reasons as before. Table 6 shows a summary of how many PAO volumes were excluded from and how many were included in the validation data set. Table 7 shows a summary of

Table 7. Summary of excluded (invalid) and included (valid) missed-pass minimums from the holdout data set..

	Number of Missed Pass Minimums						
	Predicted		Actual		Pairs		
	Invalid	Valid	Invalid	Valid	Invalid	Valid	Total
Total	50	500	12	538	60	490	550

how many missed-pass minimums were excluded and included.

B. Characteristics of the Holdout Data Set

Figure 7 shows a scatter plot of actual versus predicted PAO data volumes for all 293 valid pairs. Figure 8 shows

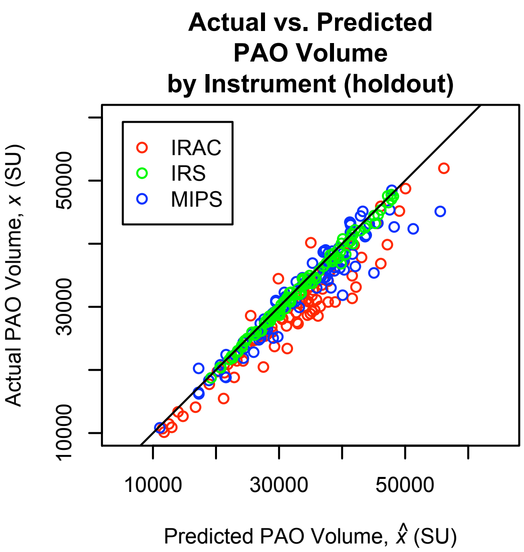

Figure 7. Actual vs. Predicted PAO Volume by Instrument -- holdout

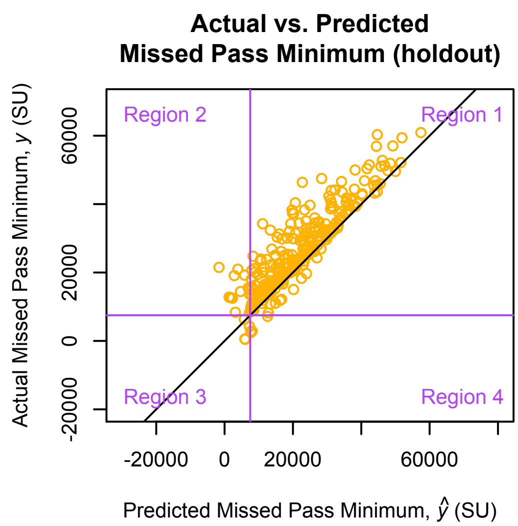

Figure 8. Actual vs. Predicted Missed-pass Minimum -- holdout

a scatter plot of actual versus predicted missed-pass minimums for all 276 valid pairs. The data in both figures are similar to the data in Figs. 1 and 2 in terms of their relation to the black diagonal lines and their variance.

C. Results of Validation

All the modeling described in section IV was completed before examining the validation set. The key plots for validation are the analogs of Fig. 6 showing highly probable intervals. Two representative segments are shown in Fig. 8. The color-coding is as above: the interval is two-sigma wide with a black circle at its center, the original estimate is a red dot, and the actual observation is a blue cross. Again, many of the original estimates are far from the actual value and well outside the confidence interval, so the procedure here is clearly improving the estimates as well as providing reasonable error bounds.

Regarding the quantitative accuracy of the error bounds, over 90% of the observed points are inside the two-sigma interval. Of the 273 MPM values, we would expect 13 to land outside the two-sigma band, but in fact 23 do. As indicated above, we believe this is caused by some nongaussianity in the residuals, as well as the temporally-dependent errors.

The computations needed to estimate the probability of MPM exceeding an arbitrary threshold are simple enough to allow implementation as part of the existing spreadsheet-based workflow. The model consists of the correction coefficients and a variance lookup table which tabulates Eq. (7) for each instrument *s* over a list of

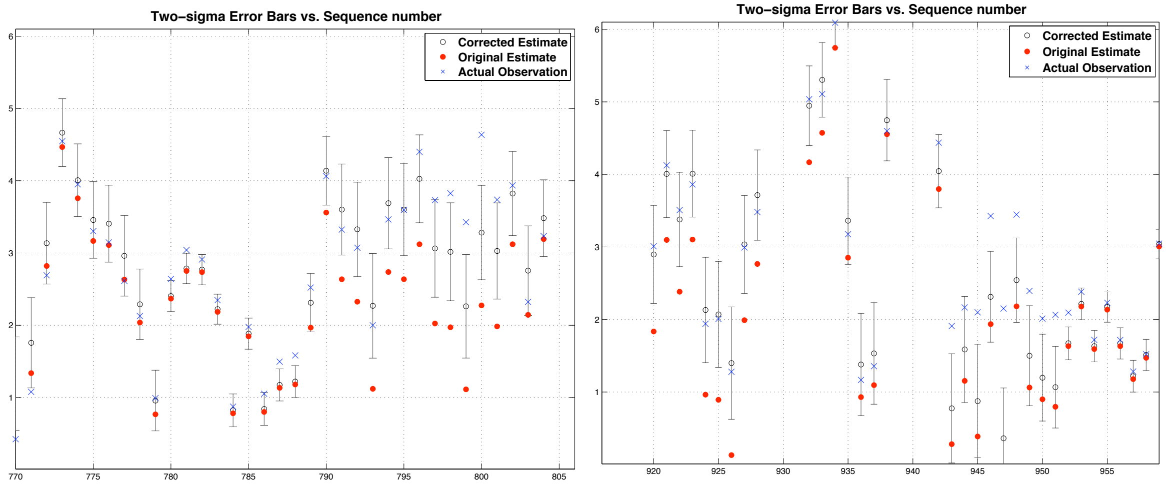

Figure 8. Probable Intervals for the holdout set.

regularly-spaced values *x*. In an operational mode, the constituent PAO data estimates are known, as well as the instrument types, from which the corrected MPM value Eq. (8) and its error bars Eq. (9) would be computed.

VI. Conclusion

The intent of this analysis was to improve Spitzer's operational efficiency by better understanding the chance of filling the on-board storage in case of a single fault in the data transmission system. To demonstrate this, we

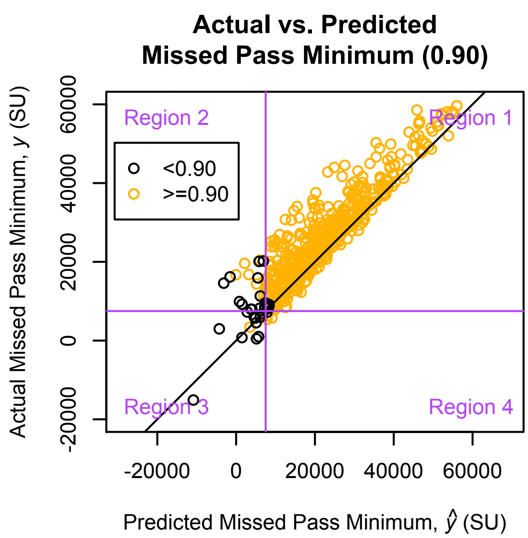

Figure 10. Actual vs. Predicted Missed-pass Minimum with p_{nofill} set at 0.90.

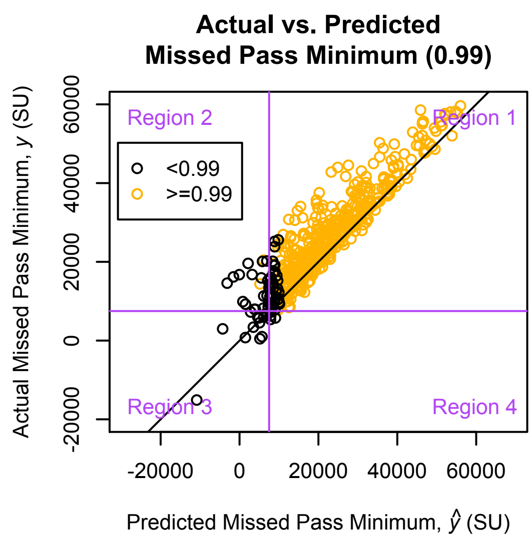

Figure 11. Actual vs. Predicted Missed-pass Minimum with p_{nofill} set at 0.99.

compare the False Negative Rate, False Positive Rate and Mitigation Rate of the original system to those of the proposed system. We show the rates of the proposed system as functions of p_{nofill}.

To begin with, we illustrate the effect that varying p_{nofill} has on the training data set. Figure 10 is a modified version of Fig. 2. We chose p_{nofill} to be 0.90 and have colored all the points whose probability of not filling is less than 0.90 black. Figure 11 is the same as Fig. 10 but we have set p_{nofill} to be 0.99 to illustrate the effect of varying

p_{nofill}. For the purposes of hypothesis testing, in Figs. 10 and 11, black points below the horizontal purple (i.e. in Regions 3 and 4) line represent True Positives. Orange points above the line (i.e. in Regions 1 and 2) represent True Negatives. Black points above the line represent False Positives. And, orange points below the line represent False Negatives. For comparison, in Fig. 2, points in Region 3 represent True Positives. Points in Region 1 represent True Negatives. Points in Region 2 represent False Positives. And, points in Region 4 represent False Negatives.

$$\alpha = \frac{FP}{FP + TN} \tag{10}$$

$$\beta = \frac{FN}{TP + FN} \tag{11}$$

Table 8. False Positive and False Negative Counts Rates corresponding to the original method shown in Fig. 2.

	Count
Region 1, *TN*	452
Region 4, *FN*	4
Region 3, *TP*	12
Region 2, *FP*	22
Total	490
False Positive Rate, *α*	0.046
False Negative Rate, *β*	0.250
Mitigation Rate, *m*	0.069

$$m = \frac{TP + FP}{TP + FP + TN + FN}$$

The false positive rate in Table 8 is defined according to Eq. (10). Similarly, the false negative rate is defined according to Eq. (11). We also define the Mitigation Rate, m. This is the fraction of time that the operations team will have to take a mitigation measure. This will also be a factor in the project's choice of p_{nofill}. It is defined in Eq. (12).

In both the original and the new methods, we wish to lower the false positive rate, α, without raising the false negative rate, β, too much. Lowering α allows us to be more operationally efficient because it makes us more certain that we will not fill the on-board storage in case of a single fault and potentially allows us to collect more data per PAO. But, when we lower α, we consequently

Table 8. False Positive and False Negative Counts Rates corresponding to the new method shown in Figs. 10 and 11.

p_{nofill}	0.80	0.81	0.82	0.83	0.84	0.85	0.86	0.87	0.88	0.89
orange above, *TN*	466	465	463	463	462	460	459	457	457	457
orange below, *FN*	6	6	5	5	5	5	5	5	5	4
black below, *TP*	10	10	11	11	11	11	11	11	11	12
black above, *FP*	8	9	11	11	12	14	15	17	17	17
Total	490	490	490	490	490	490	490	490	490	490
False Positive Rate, *α*	0.017	0.019	0.023	0.023	0.025	0.030	0.032	0.036	0.036	0.036
False Negative Rate, *β*	0.375	0.375	0.313	0.313	0.313	0.313	0.313	0.313	0.313	0.250
Mitigation Rate, *m*	0.037	0.039	0.045	0.045	0.047	0.051	0.053	0.057	0.057	0.059

p_{nofill}	0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
orange above, *TN*	456	454	453	451	446	440	437	433	429	420
orange below, *FN*	4	4	4	4	2	2	1	0	0	0
black below, *TP*	12	12	12	12	14	14	15	16	16	16
black above, *FP*	18	20	21	23	28	34	37	41	45	54
Total	490	490	490	490	490	490	490	490	490	490
False Positive Rate, *α*	0.038	0.042	0.044	0.049	0.059	0.072	0.078	0.086	0.095	0.114
False Negative Rate, *β*	0.250	0.250	0.250	0.250	0.125	0.125	0.063	0.000	0.000	0.000
Mitigation Rate, *m*	0.061	0.065	0.067	0.071	0.086	0.098	0.106	0.116	0.124	0.143

raise β. Raising β means that we will needlessly take mitigation measures more often. Lowering α also raises m. There is an operational cost and risk associated with taking a mitigation and the project would wish to know this directly. Figures 12 and 13 show this relationship for the training data set and the hold out data set respectively. The solid lines show the results for the new method, while the dashed lines show the results for the original method. Roughly speaking the new method is equivalent to the original method if we choose p_{nofill} to be about 0.93. There seems to be a sweet spot for p_{nofill} of about 0.96 that significantly decreases α, by about a factor of about four, without increasing β too much, only by a factor of about two.

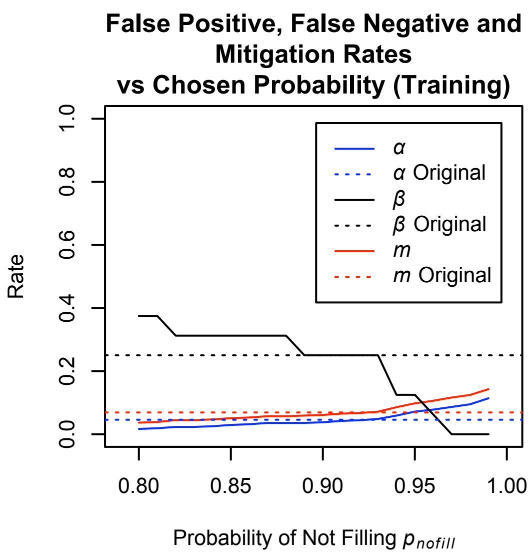

Figure 12. False Positive, False Negative and Mitigation Rates as a function of p_{nofill} for the training data.

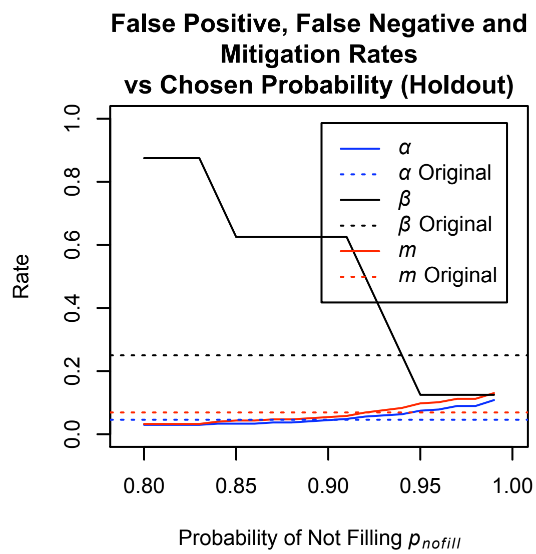

Figure 13. False Positive, False Negative and Mitigation Rates as a function of p_{nofill} for the holdout data.

There are still some limitations of this new method. First, we must manually identify those large IRAC observations described in section III. A. Since we excluded those outliers from the training data set, this statistical model does not apply to them. These cases are few, but when they occur we must add some extra margin by hand. Second, sometimes we intentionally do not leave enough time during a telecommunications pass to downlink all the data collected during the previous PAO. We call this carryover data and the model was not designed to handle this case. In this case, the missed pass minimum is calculated the same way as in Eq. (1), but with an extra term added in for the carryover data at the end of $Pass_{n-3}$. As remarked in section IV, the error bars could be better calibrated by accounting for between-PAO correlations in the residuals, and by accounting for the slightly non-Gaussian error distributions.

Finally, after Spitzer's liquid helium runs out in about April 2009, we will have to retrain and revalidate the model. After that time, only two of the IRAC instrument's four sensor arrays will continue to operate. The MIPS and IRS instruments will be permanently turned off since they cannot return valid science data at the higher temperatures. A first approximation might be to cut the IRAC science data volumes in half. But, we have no reason to assume that the two remaining sensor arrays currently account for half the IRAC data volume. In fact, our best indication is that they account for about 60% of the current volume. The fact that only one instrument will continue to operate will remove one source of variability in the model. But, the other source of variability, the predicted size of the PAO data volume, $\hat{x}_n$, will remain. See Fig. (3).

The exact payoff of this change in terms of increased operational efficiency is still to be determined. And, it may be difficult to isolate since so many other factors outside the scope of this paper govern the operational efficiency. It is not always possible to schedule additional observations even if more time is made available. But, choosing a p_{nofill} greater than the current equivalent of 0.93 will reduce risk of filling the on board storage regardless of any increase in operational efficiency, but at the cost of increased operational complexity due to increased numbers of mitigations. The decision of whether to switch to the new system and, if so, what value for p_{nofill} to choose remains a question for Spitzer mission management.

Acknowledgments

The research and development described here was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1]Sarrel, M. A., Hunt, J. C., Jr., "Evaluating Requirements on the Spitzer Mission Operations System Based on Flight Operations Experience," *Proceedings of the SPIE – Astronomical Instrumentation Conference,* Vol. 7016, SPIE, Bellingham, WA, USA, 2006, paper number 7016-61, pp. 701661-1 – 701661-12.

[2]Dodd, S. R., Levine, D. A., "The Spitzer Space Telescope's Performance: Getting the most out of a Great Observatory," *Proceedings of the SPIE – Astronomical Instrumentation,* Vol. 6270, SPIE, Bellingham, WA, USA, 2006, paper number 6270-02, pp. 627002-1 – 627002-12.

[3]Bennett, L., Comeau, S., Levine, D. A., Dodd, S. R., "The Spitzer Science Center System Description and Lessons Learned Due to Two Years of Operations," *Proceedings of the SPIE – Astronomical Instrumentation,* Vol. 6270, SPIE, Bellingham, WA, USA, 2006, paper number 62704-14, pp. 627014-1 – 627014-9.

[4]Sarrel, M. A., Carrion, C., Hunt, J. C., Jr., "Managing the On-Board Data Storage, Acknowledgement and Retransmission System for Spitzer," *SpaceOps 2006 Conference,* AIAA, Reston, VA, USA, 2006, paper number AIAA 2006-5564, pp. 1-24.

[5]Cai, T., Levine, M. and Wang, L. "Variance function estimation in multivariate nonparametric regression," technical report to appear in *J. Multivariate Analysis*, 2006.

[6]Carroll, R. J., "Adapting for heteroscedasticity in linear models," *Annals of Statistics*, Vol. 10(4), 1982, pp. 1224-33.

[7]Hall, P. and Carroll, R. "Variance function estimation in regression: The effect of estimating the mean." *Jour. Royal Statist. Soc*. (Ser. B), 1989, Vol. 51, pp. 3–14.

[8]C. Hildreth and J.P. Houck, "Some estimators for a linear model with random coefficients," *Jour. Amer. Statist. Assoc.*, 1968, Vol. 68, pp. 584-595.

[9]M. Levine and L. Brown, "Variance estimation in nonparametric regression via the difference sequence method" *Annals of Statistics*, 2007, Vol. 35, No. 5, pp. 2219–2232.

[10]Spokoiny, V. "Variance estimation for high-dimensional regression models," *J. Multivariate Anal.,* 2002, Vol. 82, 111–133.

[11]Wang, L., Brown, L. D., Cai, T. & Levine, M., "Effect of mean on variance function estimation in nonparametric regression," *Annals of Statistics,* 2008, Vol. 36, pp. 646-664.