https://ntrs.nasa.gov/search.jsp?R=19950011756 2020-06-16T09:37:54+00:00Z

N95- 18171

1994 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

111756
55-82
33965
p- 25

JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

ASSESSMENT OF THE SFC DATABASE FOR ANALYSIS AND MODELING

PREPARED BY:	Dr. Martha A. Centeno
ACADEMIC RANK:	Assistant Professor
UNIVERSITY AND DEPARTMENT:	Florida International University Industrial & Systems Engineering Department
NASA/KSC	
DIVISION:	TP-OAO
BRANCH:	
NASA COLLEAGUE:	Amanda Mitskevich
DATE:	August 18, 1994
CONTRACT NUMBER:	University of Central Florida NASA-NGT-60002 Supplement: 17

ACKNOWLEDGMENTS

I would like to acknowledge and thank you the following person for making this summer experience a reality: Kari Stiles, Amanda Mitskevich, Timothy Barth, William Bastedo, and anyone else who contributed to make my appointment reinstatement possible.

To the former TP-OAO staff: Amanda, Bill, JoAnn, Kathy, Laura, and Tim ***GRACIAS!!*** Thanks are specially given to Amanda for keeping me on my toes with her questions, and for re-directing me with her insightful comments. Thank you, Boss!☺

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This effort is a continuation of the one initiated during the summer of 1993, concerning the utilization of the SFC data. During the summer of 1993, we discovered the actual configuration of the SFC database and found out the several aspects of the data entry process; i.e. the actual ***form*** of the SFC database. This summer we set out to do some actual analysis with the SFC contents. In order to do that, however, we had to know the ***actual values*** that are being stored in the SFC database.

SFC is one of the four clusters that make up the Integrated Work Control System (IWCS), which will integrate the shuttle processing databases at Kennedy Space Center (KSC). The IWCS framework will enable communication among the four clusters and add new data collection protocols. The Shop Floor Control (SFC) module has been operational for two and a half years; however, at this stage, automatic links to the other 3 modules have not been implemented yet; except for a partial link to IOS (CASPR). SFC revolves around a DB/2 database with PFORMS acting as the database management system (DBMS). PFORMS is an off-the-shelf DB/2 application that provides a set of data entry screens and query forms. The main dynamic entity in the SFC and IOS database is a ***task***; thus, the physical storage location and update privileges are driven by the status of the WAD. Complete discussion of the 1993 effort is found in the report *"Issues Regarding Data Collection, Data Extraction, and Data Analysis"* by Centeno and Colucci (1993).

As we explored the SFC values, we realized that there was much to do before actually engaging in continuous analysis of the SFC data. Half way into this effort, it was realized that full scale analysis would have to be a future third phase of this effort. So, we concentrated in ***getting to know the contents*** of the database, and in ***establishing an initial set of tools*** to start the continuous analysis process. Specifically, we set out to

1. Provide specific procedures for statistical models, so as to enhance the TP-OAO office analysis and modeling capabilities
2. Design a data exchange interface
3. Prototype the interface to provide inputs to SCRAM
4. Design a modeling database

These objectives were set with the expectation that, if met, they would provide former TP-OAO engineers with tools that would help them demonstrate the importance of process-based analyses. The latter, in return, with help them obtain the cooperation of various organizations in charting out their individual processes.

Sections 2 and 3 address most of the issues that raised new questions regarding the contents of SFC's database, and their impact on analysis. Sections 4, 5, and 6 describe the initial set of tools developed. Section 8 summarizes results and recommendations.

2. SFC RECORDS THAT NEED TO BE UPDATED

As part of the data retrieval process, it was found that many records do not have complete information. Although this situation is relatively normal in a software system of the magnitude of SFC, it must be corrected in as much as possible. It has been found, for instance, that there are approximately 111,000 ! ☹ *tasks worked* (ACTTRNID = '31') records which have either a null, a non-printable character, a 0, or a blank space in the STS_NO field of the ACTVEMPL table. In the early stages of SFC implementation, there was no STS_NO field in the table; it was added later on. A similar situation was found for *delays* records. Furthermore, since some of the analyses will be done on a *"per wad type"* basis, the completeness of ACTVEMPL on the WAD_TYPE field was checked. It was found that 26% of the *tasks worked* records and 41% of the *delays* records do not have a value in this field. Identifying the wad type is a feasible, yet cumbersome task that, at this time, may not be worth pursuing because losing those *wad_type*-less records will not have an adverse effect on the various analyses (Figures #3 and #4).

Table #1 gives a tally of the *tasks worked* and *delays* records in ACTVEMPL (as of July 6, 1994) for each one of the flows, including those unidentified flows. It can be seen from this table that about 111,000 (≈42%) *tasks worked* records belong to *unknown* flows (Figure #1). Similarly, only 887 records were found to belong to STS-52 and STS-53 combined, which is an abnormally low value for completed flows. Similarly, 53% of *delays* records (Figure #2) belong to *unknown* flows.

Table #1: TASKS WORKED and DELAYS records in ACTVEMPL per STS_NO

STS_NO	COUNT() for tasks worked	COUNT() for delays	STS_NO	COUNT() for tasks worked	COUNT() for delays
	159	9216	56	9956	1119
weird	1		57	11455	1010
—	2		58	15581	1208
	71066		59	12953	580
weird	1		60	16471	954
TBD	2270	206	61	19321	893
0	41870	2051	62	13278	635
5	5	1	63	132	2
15	7		64	11499	525
16	3		65	14734	656
17	4		66	3605	225
18	1		67	4	
19	6		68	8717	407
47	1093	3	69	4	
51	12284	987	73	34	1
52	739	4			
53	166	3			
54	2359	50			
55	7989	895	**Grand Total**	277,769	21,631

Figure #1: Distribution of *tasks worked* records

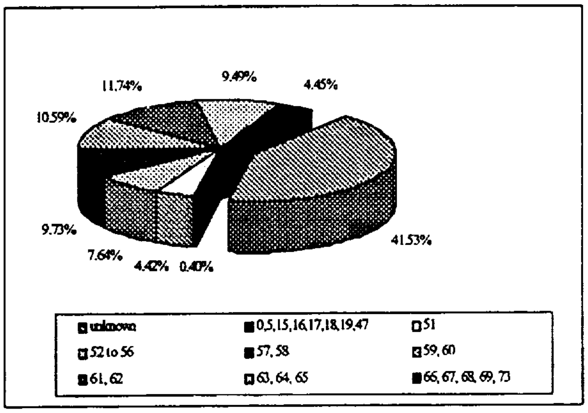


Figure #2: Distribution of *delays* records

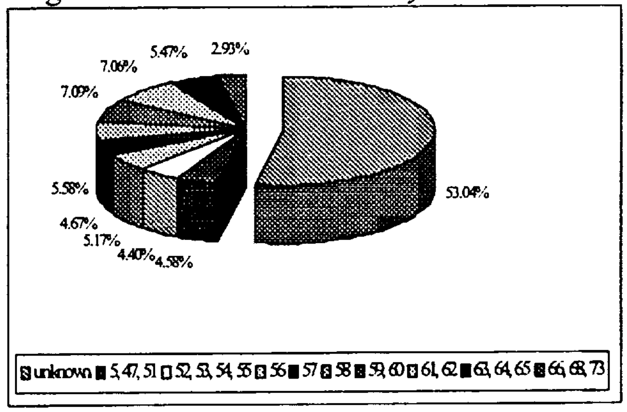


Figure #3: Distribution of *wad_types* in *tasks worked* records

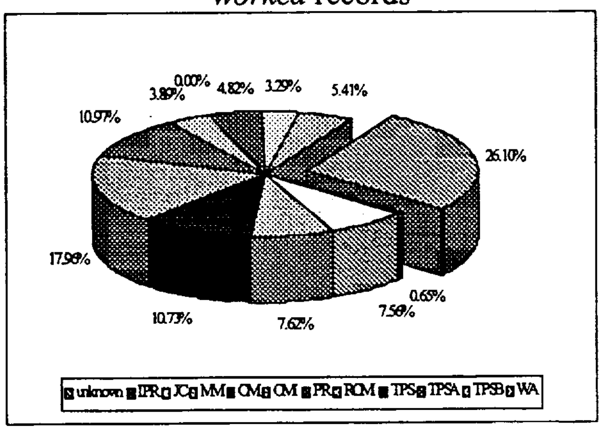
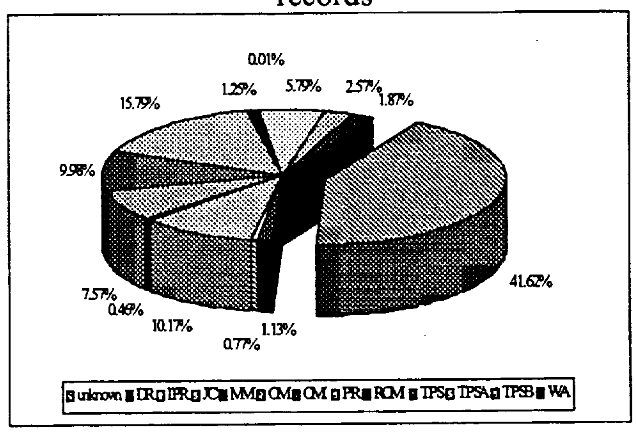


Figure #4: Distribution of *wad_type* in *delays* records



To identify the ***flow-less*** records, a series of queries have been run against ACTVEMPL (KSC3) using the dates of each flow at the OPF (Table #3). The results of these queries have been summarized in Table #2. About 70,000 of these flow-less records have one blank space or a '000 ' in STS_NO. However, there are records from 1993 and 1994 that have a '000' in the STS_NO field. The latter situation should not be occurring especially since the software has been upgraded to automatically download the *sts_no* from IOS.

Sometimes an orbiter is processed at two OPF facilities. To keep track of the data downloaded for each facility, Table #3 assigns a sequential key to each *flow/OPF* pair to be used in Table #2.

A counting query was issued to check how many of the records with '000%' in STS_NO were *notes or remarks*. This was done to assess whether the extra mainframe processing time was worthwhile, or if these records could be easily removed using a PC-based tool (e.g. Excel 5.0 or a Visual Basic (or C) program). A comment on the subject of notes is that their entry in the database does not seem to be consistent with the overall design of SFC. ACTVEMPL contains various types of records: tasks worked, delays, and so forth. Each transaction has a different value for *acttrnid*, yet when it comes to *notes*, they get the same *acttrnid* as *tasks worked*.

Table #2: Number of orphan records per flow

		Tasks Worked		*Delays*	
Key	*flow*	*records with sts_no = ' '*	*records with sts_no = '000'*	*records with sts_no = ' '*	*records with sts_no = '000'*
4	sts-49	---	---	---	---
2	sts-46	2112	---	188	---
7	sts-52	7751	---	918	---
10	sts-54	5773	---	563	---
1	sts-45	2	---	---	---
8	sts-53	10728	---	1964	---
11	sts-55	2158	13	360	3
5	sts-50	2386	---	904	---
3	sts-47	7988	---	1034	---
9	sts-53	6607	---	1071	---
13	sts-57	---	10	---	
6	sts-51	---	19	---	1
16	sts-61	---	59	---	5
	Total	**45,505**	**101**	**7,002**	**9**

Table #3: Key to Table #2[1]

				Dates at OPFs		Dates at VAB & Pad	
Key	flow	orbiter	OPF	In	Out	In	Out
1	sts-45	OV-104	2	1-Dec-91	13-Feb-92		
2	sts-46	OV-104	1	3-Apr-92	5-Jun-92		
3	sts-47	OV-105	3	1-Jun-92	16-Aug-92	17-Aug-92	12-Sep-92
4	sts-49	OV-105	1	1-Dec-91	7-Mar-92	7-Mar-92	9-May-92
5	sts-50	OV-102	3	9-Feb-92	30-May-92		
6	sts-51	OV-103	3	16-Apr-93	24-Jun-93	24-Jun-93	12-Sep-93
7	sts-52	OV-102	1	10-Jul-92	20-Sep-92		
8	sts-53	OV-103	2	17-Feb-92	8-Aug-92		
9	sts-53	OV-103	3	17-Aug-92	3-Nov-92		
10	sts-54	OV-105	1	21-Sep-92	23-Nov-92	23-Nov-92	13-Jan-93
11	sts-55	OV-102	2	12-Nov-92	27-Jan-93		
12	sts-56	OV-103	3	9-Dec-92	3-Mar-93	3-Mar-93	8-Apr-93
13	sts-57	OV-105	1	19-Jan-93	24-Mar-93	24-Mar-93	21-Jun-93
14	sts-58	OV-102	2	6-May-93	12-Aug-93	12-Aug-93	7-Oct-93
15	sts-59	OV-105	1	13-Dec-93	15-Mar-94	15-Mar-94	9-Apr-94
16	sts-61	OV-105	1	1-Jul-93	21-Oct-93		
17	sts-62	OV-102	2	1-Nov-93	27-Jan-94	27-Jan-94	24-Feb-94
18	sts-65	OV-102	2	10-Mar-94	21-Jun-94	21-Jun-94	7/8/94

[1] Dates were taken from Volume II of Schedule and Status Summary Enhancement Analysis KSC Processing Summary Data, May 18, 1993. This table should be updated as the flows are processed through the OPF facilities.

In reviewing extracted data for *tasks worked* and *delays*, it was found that some wads have a *blank* space in their name (PARTN), or they have a double hyphen ('--'). The rule of the majority seems to indicate that these cases are not supposed to exist. A possible reason for this situation is a bar coding error since the error is consistent across the same wad. Specific examples of this situation are given below. In these examples the '^' symbol represents a *blank* space.

V1262.002-C-R0^1	V30-14343-B-R0^1	V9002.10E/2-01^18
V41-10017-B-R0^1	V63-50006-H-R0^1	V5C06.001-B01-R^0
V02-50002-H-R0^1	V1008.001-Q-R0^1	V1165.013-S-R0^1
V9023.001/5-111692-15	APU-4-12-^293	RMS-201-^202-018
V9001^VL^1	V9028/5-092492-02	V9023.001/3-0614
V9045C/3-042693-^^		

The importance of knowing if the names of the wads (*partn*) in SFC are correct is critical to automatically group them for various types of analyses. A wad that differs just by one character in its *partn* field will be considered a different wad. To alleviate this problem, either the contents of SFC must be corrected, or the grouping routines have to be built with pseudo smart grouping capabilities, using a cross referencing table. Since the wads are mostly downloaded from IOS, it seems reasonable that *partn* be corrected directly into the database, so that future occurrences of the wad do not exhibit the same problem. Furthermore, by correcting these discrepancies at the source (database), future software applications will not have to take care of it over and over again.

The high number of wads with inconsistencies in *partn*, led us to run a query to identify all '31' and '37' entries in SFC which contain a "/" or a "\" in *partn* for the OPFs and the VAB/PAD. The results of this query show that there are 7,625 records (as of 8/8/94) under this situation. 97.1% of them are type '31', with the rest being type '37'. Most of these records were posted by the OPFs (94.57%) (Figure #5), with the VAB/PAD posting the other ones. Figure #6 shows the incidence of this situation over time, and Figure #7 shows it per OPF.

Figure #5: Distribution of wads with inconsistencies

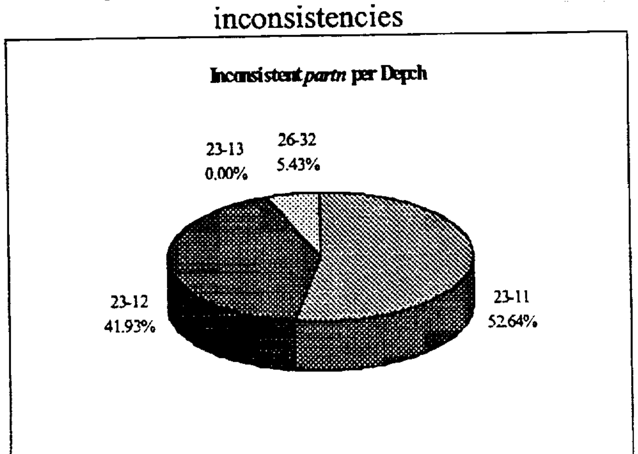


Figure #6: Frequency of wads w/inconsistencies over time

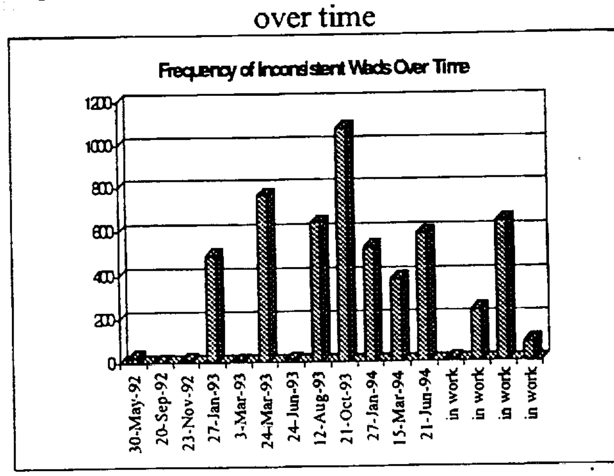
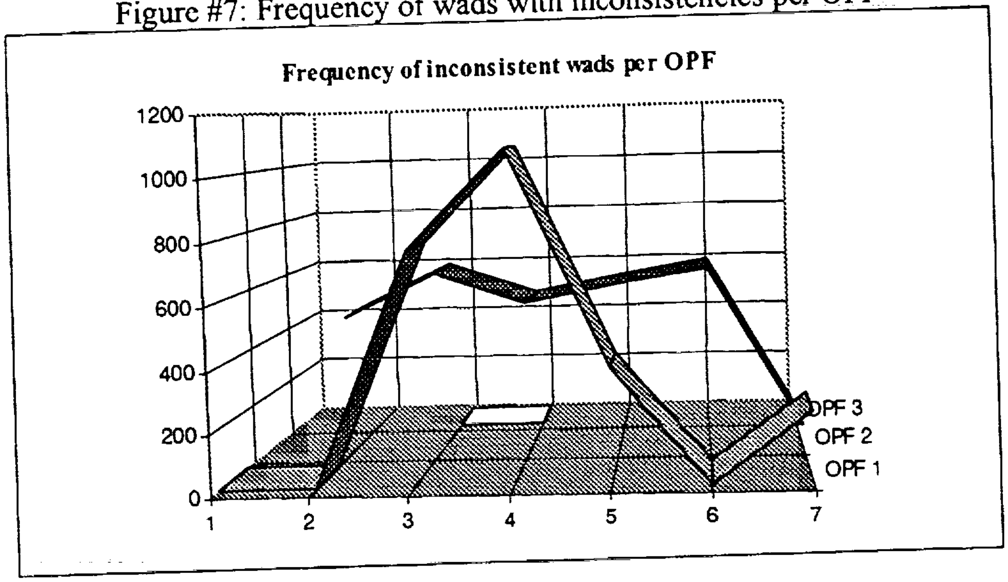


Figure #7: Frequency of wads with inconsistencies per OPF



It is clear from Figure #7 that for some unknown reason, OPFs 1 and 2 have a higher incidence of discrepancies with the value for *partn*. This needs further investigation.

4. SETTING UP HISTORICAL SUMMARIES

This section describes an initial set of historical summaries for each one of the shuttle flights for which there was data. Summaries are for both *delays* as well as *tasks worked* records.

Setting up these initial set of summaries, and enabling the mechanisms to make it a continuous process, required a thorough exploration of the SFC database contents. This exploration helped us to better understand how to manipulate the SFC data, but at the same time, like any other exploration, it raised new interesting questions.

It was learned that, confidence aside, many of the records in the SFC database cannot be used for analysis. Specifically, it was found that many '31' and '37' type entries

1. were done as *trial* records. During the early stages of implementation, engineers at various facilities needed to practice with the system, so they entered records which have a *non-wad* value in *partn*.

2. were not logged off until months later or were never logged off. In the beginning the technicians were not given the appropriate training to deal with the system. This resulted in very *long* or *negative* delay and work duration.

3. were not updated appropriately when converted from a type '31' record to a type '37' record. Some *delay* records show a "CD", "SQ" or other invalid delay code in *p_sub_stat,* which should represent the delay category in a type '37' record.

4. seem to have been entered accidentally. Their delay or work duration is less than one minute.

5. are *notes* or *remarks*.

Some of these problems can be readily overcome by conditioning the query (e.g. **where** *partn* **not like "%NOTE%"**.) The other ones have to be taken care of once the data has been imported into Excel 5.0. The following criteria has been implemented in the Excel 5.0 templates to get rid of non-useful records:

tasks worked (type '31')

1.	TIME ELAPSED BETWEEN TECHNICIANS *CLOCKING OUT* OF THE TASK IS MORE THAN 7 HOURS.	**{max(*sdate+stime*) - min(*sdate+stime*) } > 7 hours**
2.	TIME ELAPSED BETWEEN TECHNICIANS *CLOCKING IN* OF THE TASK IS MORE THAN 7 HOURS.	**{max(*actcdate+actctime*) - min(*actcdate+actctime*) } > 7 hours**
3.	WORKED TIME IS LESS THAN 10 MINUTES (INCLUDING NEGATIVE).	**{max(*sdate+stime*) - min(*actcdate+actctime*) } < 0.167 hours**
4.	WORKED TIME IS MORE THAN 60 DAYS	**{max(*sdate+stime*) - min(*actcdate+actctime*) } > 1440 hours**
5.	RECORD IS A TRIAL RECORD. TRIAL RECORDS HAVE A NUMBER AS THE FIRST CHARACTER AND A "-" AS THE SECOND CHARACTER OF *PARTN*	**examples:** 2-111692-5 3-011293-6
6.	THE *CLOCK OUT* DATE IS 2 OR MORE DAYS AFTER THE ROLL OVER DATE.	**max(*sdate+stime*) > roll over date + 2**

delays (type '37')

1.	DELAY CODE IS INVALID.	**examples:** null, one blank space, CD, SQ, C24, ACT, SQ, PA, ST, NW
2.	DELAY TIME IS LESS THAN 5 MINUTES (INCLUDING NEGATIVE).	**(*sdate+stime*) - (*actcdate+actctime*) < 0.083 hours**
3.	DELAY TIME IS MORE THAN 60 DAYS	**(*sdate+stime*) - (*actcdate+actctime*) > 1440 hours**
4.	THE *CLOCK IN* DATE IS AFTER THE ROLL OVER DATE.	***actcdate+actctime* > roll over date**

5. RECORD IS A TRIAL RECORD. (See explanation given for ***tasks worked***)

Data was downloaded for as many flows as possible. There were only two records for *sts-45*, so no further processing was done for this flow. All queries used to extract data have been stored in ***sf0110***. All extracted files are saved on floppy diskettes.

Each extracted file was cleaned up using the criteria given before. It can be stated that the data entry process (overall) has been improving as SFC reaches steady state conditions. This can be seen in Figure #8 (*tasks worked*) and Figure #9 (*delays*). The same thing can be said for each OPF although their individual learning curve differs greatly as shown in Figure #10 and Figure #11.

Figure #8: Deleted records (Tasks worked)

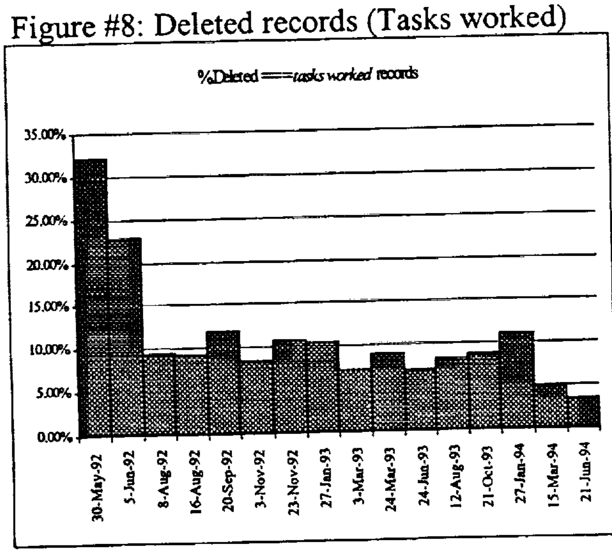


Figure #9: Deleted records (Delays)

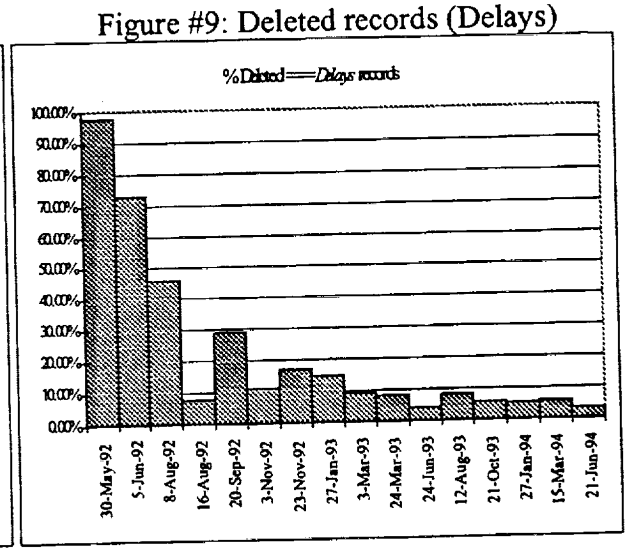


Figure #10: Deleted records (per OPF)

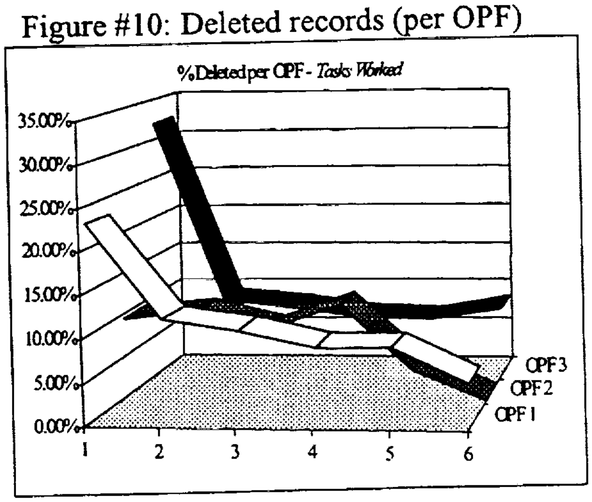
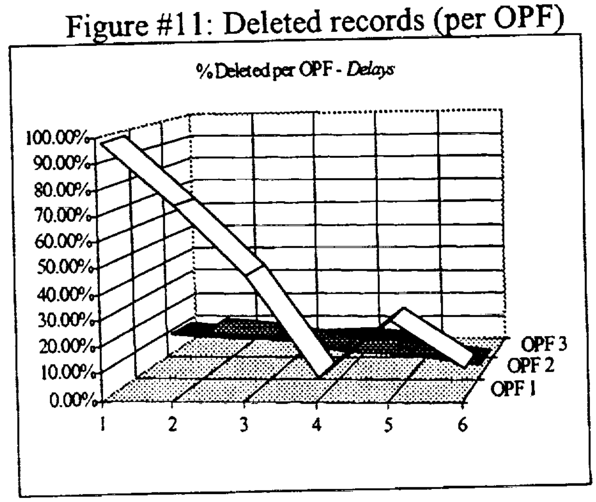


Figure #11: Deleted records (per OPF)

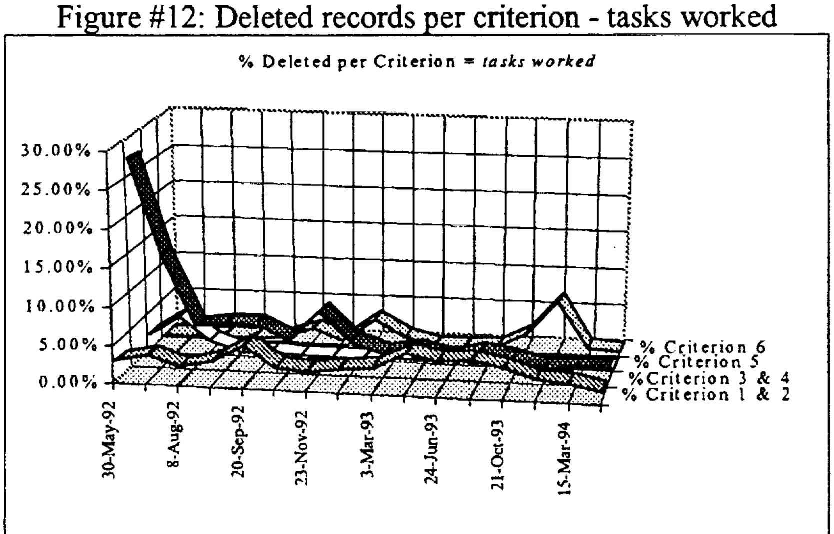


It is interesting to see from Table #4 and Figure #12 how the relevance of each criterion has changed over time. Entries with too small or too large work duration have steadily decreased, whereas technicians clocking in/out at different times for the same task has maintained the same level. The latter may be an indicator for further investigation (*why are technicians clocking in/out at significantly different times for the same task? Are they still using the "assigned" shift of the technicians to update actscode?*). Most of the improvements seen with regards to criterion 1 & 2 and 3 & 4 are mostly due to better training and software improvements respectively.

Table #4: Cleaning results - tasks worked

sts_no	Date Out of OPF	Records left	Criterion 1 & 2	Criterion 3 & 4	Criterion 5	Criterion 6	Total Deleted	% Deleted
sts-50	30-May-92	1223	46	492	38	0	576	32.02%
sts-46	5-Jun-92	1145	51	218	70	0	339	22.84%
sts-53(a)	8-Aug-92	6607	156	405	112	16	689	9.44%
sts-47	16-Aug-92	4508	150	308	0	0	458	9.22%
sts-52	20-Sep-92	4453	270	302	16	10	598	11.84%
sts-53(b)	3-Nov-92	2097	59	93	0	39	191	8.35%
sts-54	23-Nov-92	3201	85	288	1	8	382	10.66%
sts-55	27-Jan-93	3897	116	188	0	145	449	10.33%
sts-56	3-Mar-93	3421	117	106	2	38	263	7.14%
sts-57	24-Mar-93	3128	176	126	1	0	303	8.83%
sts-51	24-Jun-93	782	37	22	0	0	59	7.02%
sts-58	12-Aug-93	4840	242	183	4	3	432	8.19%
sts-61	21-Oct-93	6077	252	189	0	138	579	8.70%
sts-62	27-Jan-94	3768	105	75	0	290	470	11.09%
sts-59	15-Mar-94	3727	84	89	0	17	190	4.85%
sts-65	21-Jun-94	4036	52	83	1	5	141	3.38%

Figure #12: Deleted records per criterion - tasks worked



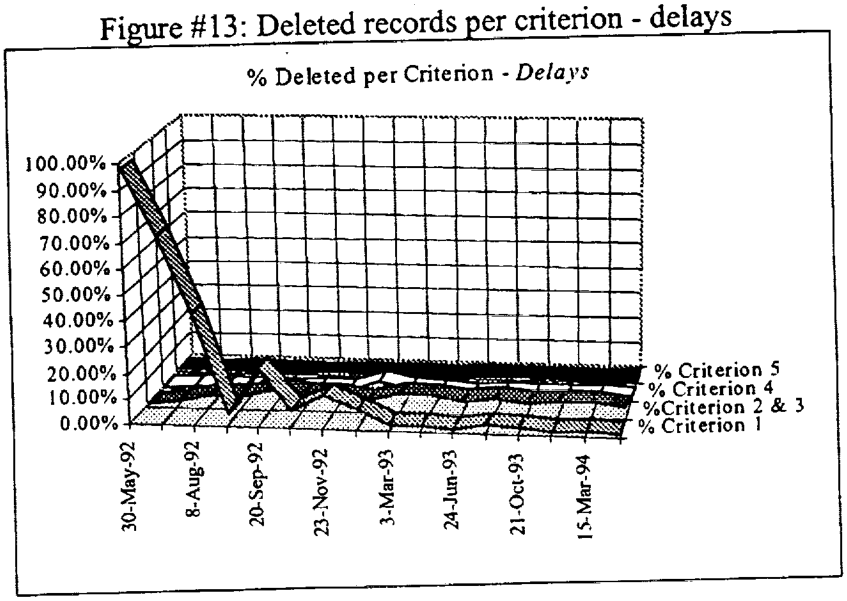
Not surprisingly, criterion #5 (trial records) has maintained a very low profile. What came as a surprise is the rise of the level of criterion 6 (posting entries after the roll over date). Again this must be further investigated.

For delays records, the situation has greatly improved as far as non-usable records are concerned (Table #5 and Figure #13). It most be pointed out, however, that the number of recorded delays seem to be steadily decreasing. This should be a great news if one were confident on the reliability of the data. There are strong reasons to believe that such a decrease is due to willful avoidance of entering delays and not due to an improvement of the shuttle assembly process. This is another issue that needs further investigation.

Table #5: Cleaning results - delays

sts_no	Date Out of OPF	records left	Criterion 1	Criterion 2 & 3	Criterion 4	Criterion 5	Total Deleted	% Deleted
sts-50	30-May-92	22	880	0	0	1	881	97.56%
sts-46	5-Jun-92	51	135	2	0	0	137	72.87%
sts-53(a)	8-Aug-92	1061	831	75	0	0	906	46.06%
sts-47	16-Aug-92	951	36	41	0	0	77	7.49%
sts-52	20-Sep-92	650	207	63	0	1	271	29.42%
sts-53(b)	3-Nov-92	939	64	53	0	0	117	11.08%
sts-54	23-Nov-92	470	68	27	1	0	96	16.96%
sts-55	27-Jan-93	952	78	40	46	0	164	14.70%
sts-56	3-Mar-93	770	13	52	13	1	79	9.31%
sts-57	24-Mar-93	545	9	36	2	0	47	7.94%
sts-51	24-Jun-93	617	2	27	0	0	29	4.49%
sts-58	12-Aug-93	851	21	52	0	0	73	7.90%
sts-61	21-Oct-93	640	13	26	1	0	40	5.88%
sts-62	27-Jan-94	451	1	23	1	0	25	5.25%
sts-59	15-Mar-94	380	3	20	1	0	24	5.94%
sts-65	21-Jun-94	476	0	16	0	0	16	3.25%

Figure #13: Deleted records per criterion - delays



From the cleaning exercise, one must learn whether the data entry process keeps on improving until it reaches a steady state. In a manufacturing setting, the rule of thumb is to accept a batch if it has, statistically speaking, at most π_o percentage of *defective* (non-usable records in this case). The π_o value is never more than 10%, with the preferred value being less than 5%. Establishing whether a batch of *products* is acceptable (good batch) is done by taking a sample of size *n* and using the percentage of defectives (**P**) found in the sample as the *estimator* of the batch's true percent defective (π). In the SFC case, even though one may think of the *records* being the product to *inspect*, one does not need to sample because the capability for a 100% inspection is readily available. Therefore, one only needs to use the Excel 5.0 templates to find the true π value for the given batch. Once the value of π is known, if it is too high, the reasons for the increase must be investigated. At the same time, if the number of records left is too little, no further analyses can be done for that flow. To update the cleaning statistics, see Section 5.1 of this report.

Tables #4 and #5 clearly show that a great improvement has occurred since the inception of SFC. Because a starting point is needed, it is recommended that any flow yielding at most π = 10% be used to set up and revise analyses. As the SFC software, IWCS, and the data entry process settle, the π value should be revised down until it reaches less than 2%. Putting this rationale to work, the paragraphs below present initial assessment of the following flows: **STS-56, STS-57, STS-58**. These flows, although chosen arbitrarily, provided the basis to exemplify some of the problems that inconsistent wad naming brings into analysis. More on this later on.

A point of clarification is that the cleaning process does not assess thoroughly the quality of the data entry process; hence, it does not say the whole story regarding the reliability of the data. The cleaning process deals only with records that were actually *entered*. If records of delays, for instance, are not entered, there is no way that the cleaning process herein described will detect that. This cleaning process is done to remove from the data those records that are an obvious data entry error due to a weak implementation of the data entry process.

Due to time constraints, the assessment is limited to gathering basic summaries for these three initial flows. The varied nature of wad work contents, in conjunction with the fact that many wads are unique to a flow, it was decided that only wads which begin with the letter "V" would be taken into consideration to conduct the multiple flow analysis. However, this is not true for generating inputs for SCRAM. SCRAM input file will contain all the wads that experienced a delay, even if they are IPR or PR or TSPB.

Table #6 gives a summary of the historic processing of the three flows. As it can be seen, each one of these flows was processed at a different OPF (1, 2, and 3), and each involved a different orbiter (Columbia, Discovery, and Endeavour). Time constraints prevented a multiple flow analysis where the orbiter (or the OPF) was the same; however, this kind of summaries can be done by simply choosing the flows for the same OPF.

STS-56 had a total of 2151 *tasks worked* records (for wads starting with a "V") for a total of 737 distinct wads processed. STS-57 had a total 1902 *tasks worked* records for a total of 593 distinct wads processed. STS-58 had a total of 2286 *tasks worked* records for a total of 663 distinct wads. To understand what is meant by distinct wads, keep in mind that a wad may be completed in multiple sessions. Although it is suspected that multiple session may also indicate multiple runs of the same wad (which means the wad suffix should be different), there is no way to know, at the moment, the truth about this situation until the data entry processes is consolidated. Therefore, the work duration for a wad is the sum of the individual records work duration.

Table #6: Sample flows for multiple flow comparison

STS-56	**Orbiter:**	*OV-103*	(querya14.dat)	
	OPF	*3*	**Delays Deleted**	xx
	Left OPF	*3-Mar-93*	**Tasks Work Deleted**	xxx
STS-57	**Orbiter:**	*OV-105*	(querya14.dat)	
	OPF	*1*	**Delays Deleted**	7.94%
	Left OPF	*24-Mar-93*	**Tasks Work Deleted**	8.83%
STS-58	**Orbiter:**	*OV-105*	(querya14.dat)	
	OPF	*2*	**Delays Deleted**	7.90%
	Left OPF	*8-Aug-93*	**Tasks Work Deleted**	8.19%

Given the fact that STS-56 processed 737 wads (set A), STS-57 processed 593 (set B) wads, and STS-58 (set C) processed 663, one might expect to find a great deal of overlapping among set A. b, and C that, when laid out as in Figure #14, the number of rows in that matrix would be no more than a 1,000 (roughly). Unfortunately, this is not the case with these three flows. When the information for the flows was re-arranged as in Figure #14, there were 1516 rows in the matrix. About 1100 of these rows had only one observation; thus, several of the basic statistical summaries (e.g. standard deviation, mode) could not be computed (see Figure #15).

Figure #14: Layout for multiple flow file

partn	*flow 1*	*flow 2*	....	*flow n*
wad_1	$duration_{11}$	$duration_{12}$		$duration_{1n}$
wad_i	$duration_{i1}$	$duration_{i2}$		$duration_{in}$
wad_m	$duration_{m1}$	$duration_{m2}$		$duration_{mn}$

These findings led us to try to include an additional flight, so we included STS-59 (Endeavour, OPF 1). It was found that it had a total of 2219 *tasks worked* records for a total of 703 distinct wad. Yet, despite the fact that the number of wads processed in this flight seems to be a "normal" count, the number of wads in the multiple flow matrix grew from 1516 to 2040, which means that about 75% of the wads in STS-59 **were new wads**. This may be true, but it needs to be further investigated, especially because the naming inconsistencies may be the cause of this situation. An example is given in Figure #16. Wad *V1047* seems to have a date attached to it. What does this mean? Should this be the same wad? Multiple runs of the same wad in the same flow? This situation must be clarified; otherwise, we will keep getting nowhere in our analysis: even with the information from four flows, for only 25% of the records it was possible to compute something as simple as the standard deviation of the work duration.

Figure #15: Sample of multiple flow basic summaries (part 1)

sample size	minimum	maximum	range	standard deviation	arithmetic average	mode	5th percentile	median
1	3.60	3.60	0.00	Can't Compute	3.60	#N/A	3.60	3.60
1	22.98	22.98	0.00	Can't Compute	22.98	#N/A	22.98	22.98
2	8.38	24.30	15.92	11.25	16.34	#N/A	9.18	16.34
1	9.05	9.05	0.00	Can't Compute	9.05	#N/A	9.05	9.05
1	21.98	21.98	0.00	Can't Compute	21.98	#N/A	21.98	21.98
1	0.93	0.93	0.00	Can't Compute	0.93	#N/A	0.93	0.93
1	5.02	5.02	0.00	Can't Compute	5.02	#N/A	5.02	5.02
1	3.67	3.67	0.00	Can't Compute	3.67	#N/A	3.67	3.67
1	18.00	18.00	0.00	Can't Compute	18.00	#N/A	18.00	18.00
1	22.05	22.05	0.00	Can't Compute	22.05	#N/A	22.05	22.05
1	0.92	0.92	0.00	Can't Compute	0.92	#N/A	0.92	0.92

Figure #15: (continued - part 2)

95th percentile	95% C.I. - lower bound	95% C.I. - upper bound	partn	56	57	58
3.60	0.00	0.00	V00-10071-F-R01	3.60		
22.98	0.00	0.00	V00-10072-A-R01			22.98
23.50	0.00	0.00	V00-10072-R01	8.38	24.30	
9.05	0.00	0.00	V02-40002-J-R01	9.05		
21.98	0.00	0.00	V02-50002-H-R01	21.98		
0.93	0.00	0.00	V05-50004-E-R01	0.93		
5.02	0.00	0.00	V070-2-15-153			5.02
3.67	0.00	0.00	V070-2-15-158			3.67
18.00	0.00	0.00	V070-3-16-175	18.00		
22.05	0.00	0.00	V070-5-04-0054		22.05	
0.92	0.00	0.00	V10-00001-B-R01			0.92

Figure #16: Sample of naming problem

V1047 /2-051193-03			4.883	
V1047/3-011993-18	1.783			
V1047/3-021993-12	1.050			
V1047/5-020893-0		6.900		
V1047/5-020894-12				9.750
V1047/5-022293-0		11.167		
V1047/5-022494-01				2.450
V1047/5-022594-11				8.467
V1047/5-031093-0		3.200		
V1047/5-031293-0		8.217		

Delay records for STS-56, STS-57, STS-58 confirm what was long known: the frequency of a delay category does not tells the whole story; rather the accumulated time of such delay category is a better indicator of reality. This can be seen in Figure #17 and #18 where B31 was the delay category with the highest frequency, but it was not the highest contributor to the total stoppage hours in these flows. Figure #19 further confirms this situation, but with another delay category.

Figure #17: Frequency and accumulated time - STS 56

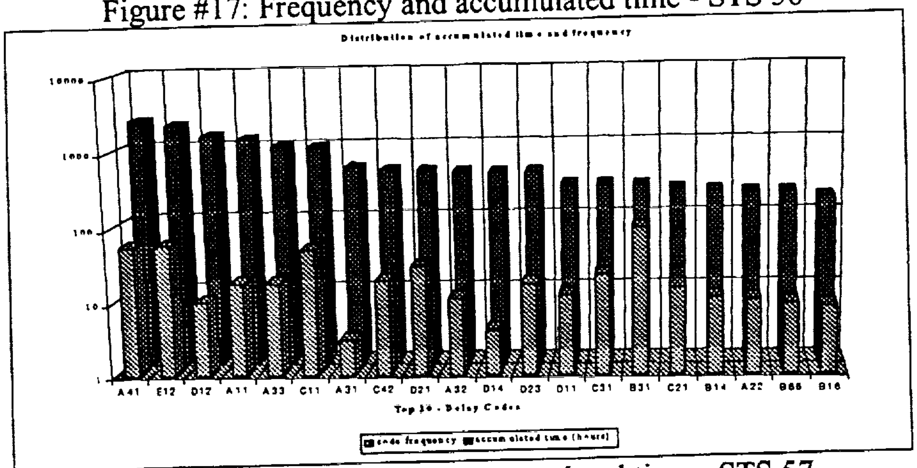


Figure #18: Frequency and accumulated time - STS 57

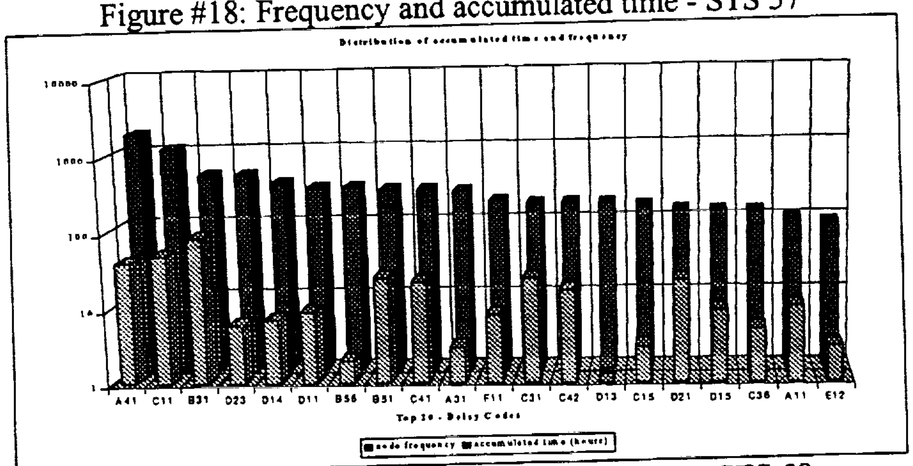
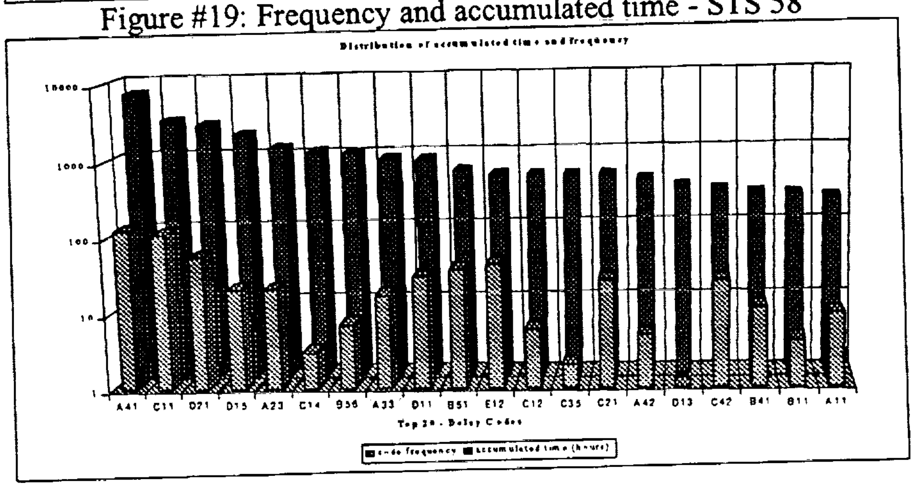


Figure #19: Frequency and accumulated time - STS 58



Based on only these three flows, nothing reliable can be said about the regularity with which delays occurred; however, these types of charts can help in identifying possible bottleneck organizations. One must be careful in drawing conclusions from these charts because the numeric measurement does not necessarily removes the need to improve an organization's process. There is always room for improvement, but most importantly, perception of being a bottleneck most be taken into consideration. An example of this can be found in the *snnndlyb.xls* templates for these flows, under the ***logistics*** worksheet. Logistics has always been among the organizations with high frequency codes (25% of all the delays count are related to Logistics); yet, Logistics is not the highest contributor to the total number of hold hours (about 10% all delays accumulated time), but Logistics has been perceived as a mjor bottleneck. This findings were presented to the NASA side of Logistics, and A. Mitskevich has began to collaborate with Logistics, so that they can chart out their process.

Although time did not permit any further analysis, the capabilitity to possibly build probability functions for the top 30 delay categories exists. A third Excel 5.0 template (*snnndlyc.xls*) computes basic summaries of the top 30 delay categories, but because of the way it is laid out (Figure #20), some of its information may be exported as a text, and then imported into SIMAN IV's INPUT module.

Figure #20: Layout for delays - Single flow

code 1	***code 2***		....	***code 30***
duration$_1$	duration$_{11}$	duration$_{i2}$		duration$_{1n}$
duration$_i$	duration$_{i1}$	duration$_{i2}$		duration$_{in}$
		duration$_{m2}$		
	duration$_{m1}$			duration$_{mn}$

4. CONTINUOUS GATHERING OF HISTORICAL SUMMARIES

The generic process to gather summaries for one OPF-related flight operations consists of 4 macro steps:

1. <u>EXTRACT DATA</u>: After roll over, the OPF is expected to have closed all pending tasks regarding that particular flight. To ensure that all tasks are closed, allow for a couple of days before data is extracted.

   a) Edit appropriate generic query.

   b) Run query in QMF

c) Export results from query to a PC diskette

2. CLEAN EXPORTED FILE: QMF reporting facility adds about 20 rows of heading and formatting information that is needed only if the file is to be printed from the mainframe. It also places information, at the beginning of each line in the resulting set, which identifies the characteristics of the "record". This extra information needs to be removed before any analysis is done. This cleaning can be done using Excel 5.0 or the *clean* option of SMART.

3. REMOVE NON-USABLE RECORDS: Many records in the SFC database cannot and should not be used in any type of analysis because of data entry problems. Some of the data entry problems can be readily detected from the data itself, so the Excel 5.0 templates *snnnwrka.xls* and *snnndlya.xls* should be used to applied the appropriate criteria. More on this step later in this section.

4. COMPUTE BASIC SUMMARIES: The basic summaries are done by using the various Excel 5.0 templates and the SMART prototype

5. CONDUCT FURTHER ANALYSIS: This may be done by using the SMART interface, if and when fully implemented. The actual analysis will depend on the objective of the modeling activity.

4.1. Updating the Cleaning Statistics

Updating the cleaning statistics requires some manual data transfer. This could be later automated if the SMART concept is further pursued. In the mean time, use the *cleansfc*.xls production Excel 5.0 file. These file has four worksheets named taskswoked, delays, chartswork, and chartsdelay. The name of the worksheets is self explanatory as far as what they contain. This is what needs to be done:

1. As you interact with the template *snnnwrka.xls* and *snnndlya.xls,* write down how many records were deleted with each criterion.

2. Write down how many records were left after the cleaning exercise.

3. Open the *cleansfc.xls* production file. Enter data accordingly based on the data being *delays* or *tasks worked.*

The *cleansfc.xls* file is setup to handle 26 flows. Except for the charts per OPF, everything is setup to pick up the data as soon as the data is entered in the appropriate place. For the "per OPF" charts, enter the data under the appropriate OPF work area. Charts will be updated automatically.

4.2. Cleaning Downloaded Files

At this point in time, the *clean data* option of SMART has not been implemented, yet data can be processed using Excel 5.0. Every extracted file needs to be *"cleaned"* meaning that any header information that QMF places at the top and left of the data must be removed. It also means that records that are suspected of being *"not useful"* records must be eliminated before any analysis is done.

The *snnnwrka.xls* template is designed to clean up the *tasks worked* file downloaded from the mainframe. This template has a series of conditional Excel 5.0 statements to implement the delete criteria (as given in Section 4 of this report) for *tasks worked* records. The first three rows of the template are used for general headings and control data. Beginning column **I** is where the conditional formulas are entered. Data exported from SFC is to be stored beginning on row 4 of columns **A** to **H**. At the same time that this templates cleans the downloaded data, it creates a subset of the data that will, later on, be used to generate inputs for SCRAM.

This template should be used to clean data after a flow, OPF section, has concluded. Detail instructions are in another report submitted to NASA. Once there is a "clean" file of *tasks worked* records. From here, the ScramWorkTime worksheet could be exported (comma delimited) in preparation for the interaction with SMART. However, remember that SCRAM requires a *delays* files too. Cleaning *delays* files is very similar to cleaning *tasks worked.*

The *snnndlya.xls* template is designed to clean up the *delays* file downloaded from the mainframe. This template has a series of conditional Excel 5.0 statements to implement the delete criteria (as given in Section 4 of this report) for *delays* records. The first three rows of the template are used for general headings and control data. Beginning column **I** is where the conditional formulas are entered. Data exported from SFC is to be stored beginning on row 4 of columns **A** to **H**. At the same time that this templates cleans the downloaded data, it creates a subset of the data that will, later on, be used to generate inputs for SCRAM.

This templates is similar in nature to *snnnwrka.xls*. Consequently, the instructions to work with this template are very similar. they have been fully detailed in another report submitted to NASA.

4.3. Multiple Flow Basic Summaries

Work records can be used to estimate how long is actually taken to complete a wad. The varied nature of wads, however, does not allow (at the moment) for such estimation directly from the SFC data. Many wads (e.g. IPR, PR) are unique to a flow; thus, there will always be only ONE observation for these wads, across all the flows. Other wads (e.g. OMI) change in contents from flow to flow, which makes them illegible for across flows comparisons. Taking these facts into account, it was decided that, at the moment, only those wads that begin with a "V" would be used. Other types of wads could be added later on.

To conduct the multiple flow basic summaries calculations, you must interact with the *snnnwrka.xls* (already a clean *task work* records file), *snnnwrkb.xls* (already containing the worksheet with wads that begin with "V" only), SMART (to re-arranged all the flows in a single file), and with *snnnwrkc.xls* (a new template onto which you will paste the multiple flow single file).

The *snnnwrkb.xls* is a template that must be used after the files for all the flows to processed have been cleaned. This template must be given a unique name, making sure that no other file is overwritten. This template has an Excel 5.0 condition to eliminate those records with wads not beginning with a "V". It has 3 worksheets: *basetable, countofwads,* and *multiflowexport*. The *basetable* is the one that has the conditional excel function to identify if the wad begins with a "V" or not. *CountOfWads* has the necessary conditional Excel 5.0 function to found out how many unique wads were processed in the flow. *MultiFlowExport* has the necessary data columns to be used by the SMART interface in building the multiple flow single file.

The SMART interface has one option on the main menu that refers to *tasks worked*. Under such option, you will find another option that refers to multiple flows. Again, the SMART interface is very straight forward to use.

The *snnnwrkc.xls* is a template that has all the statistical functions to compute the basic summaries across the flows, for each wad. These basic summaries include a confidence interval, which will be computed only if there are enough data points for the wad (more than 5). If there are enough data points, the confidence interval will be computed using the *t-student* distribution for sample sizes less than 25 observations, and it will use the normal distribution otherwise.

Steps to follow have been detailed in another report submitted to NASA.

4.4. Single flow basic summaries -delays

The *snnndlyb.xls* is a template that must be used after the delay file for a flow has been cleaned. This template must be given a unique name, making sure that no other file is overwritten. This template has a series of Excel 5.0 conditions and graphs to summarized the behavior of delays. It also has two worksheets to export data, so that the SMART interface can generate a file to gather basic statistical summaries about each one of the top 30 delay code.

The SMART interface has one option on the main menu that refers to *delays*. Under such option, you will find another option that refers to single flow. Again, the SMART interface is very straight forward to use.

The *snnndlyc.xls* is a template that has all the statistical functions to compute the basic summaries across the codes. These basic summaries include a confidence interval, which will be computed only if there are enough data points for the wad (more than 5). If there are enough data points, the confidence interval will be computed using the *t-student* distribution for sample sizes less than 25 observations, and it will use the normal distribution otherwise.

Details on the interaction are part of another report submitted to NASA

5. PREPARING INPUTS FOR SCRAM

SCRAM is a modeling tool that is being developed by Lumina, Inc. through a SBIR contract. The main purpose of SCRAM is to identify and quantify the contributors to overall costs and schedule risk in a shuttle processing flow. Once the initial model is constructed, SCRAM will use Bayes' Theorem to revise the probabilities of wads experiencing delays and delay duration as data is collected in the SFC database. These revised probability functions are then utilized to update the network of shuttle processing activities, including those activities in the critical path.

Inputs for SCRAM must be provided in a "spread-sheet" like format, with data laid out as shown in Figure #21; therefore, it is necessary to download the data from SFC and process it, so that such format is complied with. Necessary Excel 5.0 templates and Visual Basic routines have been set up to carry out this process. The Visual Basic routine has been incorporated into the SMART (Shop Floor Modeling, Analysis, and Reporting, Tool) prototype.

Figure 21: Layout of SCRAM input file

wad_1	$workd_1$	$delcod_{11}$	$deldur_{11}$	...	$delcod_{1k}$	$deldur_{1k}$	...	...
wad_2	$workd_2$	$delcod_{21}$	$deldur_{21}$	...	...		$delcod_{2k+1}$	$deldur_{2k+1}$
:	:	:	:	:	:	:		:
wad_i	$workd_i$	$delcod_{i1}$	$deldur_{i1}$	...	...		$delcod_{in}$	$deldur_{in}$
:	:	:	:	:	:	:		:
wad_m	$workd_m$	$delcod_{m1}$	$deldur_{m1}$	...	$delcod_{mn}$	$deldur_{mn}$		...

There are two possible ways in which the process is initiated: 1) data has just been downloaded from SFC, and 2) data has been downloaded from SFC and it has been cleaned using the Excel 5.0 templates. The inner works of these templates has already been addressed in another section of this report; however, it is necessary to ***emphasize*** that once the records have been cleaned up, the resulting Excel 5.0 file must be **cleared up** in those cells that have no data (*ScramDelayTime* and *ScramWorkTime* sheets of Excel 5.0 files *snnnwrka.xls* and *snnndlya.xls*). To clear cells up, highlight the appropriate cells, click on edit, clear, all in Excel 5.0.

Details on the interaction are part of another report submitted to NASA

A decision to create the *s65wrka.txt* file was made because when testing the Visual Basic procedure, it was found that many records were not being included in the ***s65wrka.out*** file. The reason for delay records not to be included is the lack of at least one matching work record. The initial testing of the Visual Basic routine was done using data for *sts 51* which, for an unknown reason, had a large number of delay records (≈ 350 out of 617) without a matching work record. However, this situation does not seem to be the law of the land because when *sts65* was processed, only 2 delay records (out of 476) were excluded. Appendix C gives samples of the SCRAM input files for *sts-51* and *sts-65*

6. THE SMART PROTOTYPE

The main idea of the S.M.A.R.T. (Shop floor Modeling, Analysis, and Reporting Tool) framework is to have a cohesive and integrated environment that supports analysis and modeling using the SFC data. To avoid re-inventing the wheel, the S.M.A.R.T. framework would use off-the-shelf data processing and analysis tools in as much as possible. Where these tools fail to meet specific requirements, the S.M.A.R.T. framework would integrate customized data processing and analysis routines.

To facilitate various analyses, such as ANOVA test, time series and so forth, the S.M.A.R.T. framework proposes to utilize a database to maintain a history of the analysis results and decisions made. Full implementation of the S.M.A.R.T. framework requires an in-depth study of several issues (such as feasibility of integrating heterogeneous tools in this context, and the development or modification of analysis techniques to better handle the uniqueness of the SFC data), which are beyond the scope of this effort. However, steps toward enabling the ***data exchange*** capabilities of the S.M.A.R.T. framework have been taken. The data exchange interface was pursued because of the large amount of data that need to be re-arranged, once downloaded from the mainframe, before any kind of analysis can be done (e.g. SCRAM, multiple flows). It is expected that the working option of the S.M.A.R.T. framework will facilitate the processing of these large quantities of information.

The current implementation of the S.M.A.R.T. is limited to read in files exported from Excel 5.0 (comma delimited) and re-arranging these files, with some basic computations (e.g. work time per wad), so that they can b used with other tools. Specifically the following options are operational in the S.M.A.R.T. framework:

The documentation of the prototype can be found in another report submitted to NASA

7. RESULTS AND RECOMMENDATIONS

Results of this effort include:

- A thorough consensus of the completeness (or incompleteness) of the SFC data. We learned that a lot of the data in SFC cannot be used for a variety of reasons, including the natural evolution of SFC, and misunderstanding of the data entry process on the part of the technicians. We also learned that things have been improving over time.

- An understanding of the ACCESS database management system (DBMS), and its potential as the DBMS of choice for fully designing and developing the modeling database (if so desired)

- An understanding of the Visual Basic programming language, and its potential as development tool for the S.M.A.R.T. (Shop floor Modeling, Analysis, and Reporting Tool) framework.

- A set of Excel 5.0 templates that, in conjunction with Visual Basic routines, enable the "cleaning" of downloaded data, the generation of inputs for SCRAM, across flows descriptive statistics of *tasks worked*, monitoring improvements in the SFC data entry process, gathering of descriptive statistics for delay categories. Further, various files could be exported into SIMAN IV's Input module to establish probability functions for the delay category.

- Last, but not least, once again, Dr. Centeno goes back with a bag full of great experiences to use in her future research and teaching endeavours.

Among the recommendations of this effort are:

- Pursue the update of as many SFC records as possible.

- Request that the ***notes*** records be given another *acttrnid*, not '31' or '37', and that the existing records be updated.

- Thoroughly investigate the issue of inconsistent *partn*. This is very crucial to accumulate observations.

- Thoroughly investigate why some delays are never put in work.

- Clarify why records are being posted against a flow that has already landed. Take appropriate corrective actions to make this situation disappear.

- Acquire a new computer workstation with at least 24 Mb of RAM, preferably 32 Mb, and with at least 900 Mb of hard disk. This workstation is necessary to maintain a history of the various analyses that will eventually be done.

- Acquire Excel 5.0 as soon as possible. Schedule the acquisition of Visual Basic and ACCESS.