FAA - 81-18

DOT/FAA/EM-83-13
DOT-TSC-FAA-81-18

Statistical Analysis and Time Series Modelling of Air Traffic Operations Data from Flight Service Stations and Terminal Radar Approach Control Facilities:

Two Case Studies

Norman J. Meyerhoff
George H. Wang

Transportation Systems Center
Cambridge MA 02142

October 1981
Final Report

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

U.S. Department of Transportation
Federal Aviation Administration

Office of Systems Engineering Management
Washington DC 20591

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/EM-82-13		

4. Title and Subtitle	5. Report Date
STATISTICAL ANALYSIS AND TIME SERIES MODELING OF AIR TRAFFIC OPERATIONS DATA FROM FLIGHT SERVICE STATIONS AND TERMINAL RADAR APPROACH CONTROL FACILITIES: TWO CASE STUDIES	October 1981
	6. Performing Organization Code: DTS-522
7. Author(s): Norman J. Meyerhoff and George H. Wang	8. Performing Organization Report No.: DOT-TSC-FAA-81-18
9. Performing Organization Name and Address: U.S. Department of Transportation, Research and Special Programs Administration, Transportation Systems Center, Cambridge MA 02142	10. Work Unit No. (TRAIS): FA259/R2119
	11. Contract or Grant No.
12. Sponsoring Agency Name and Address: U.S. Department of Transportation, Federal Aviation Administration, Office of Systems Engineering Management, Washington DC 20591	13. Type of Report and Period Covered: Final Report, Oct 1979-Sep. 1980
	14. Sponsoring Agency Code: FAA/AEM-200

15. Supplementary Notes

16. Abstract

Two statistical procedures have been developed to estimate hourly or daily aircraft counts. These counts can then be transformed into estimates of instantaneous air counts. The first procedure estimates the stable (deterministic) mean level of hourly or day of the week patterns by statistical models. The second procedure estimates both deterministic and stochastic periodic (hourly or day of the week) patterns by stochastic time series models. Both statistical procedures have been used to analyze traffic at the St. Louis TRACON and Los Angeles Flight Service Station.

This report analyzes hourly variations in operations at the St. Louis TRACON, for each day in four representative months in 1979. It also analyzes daily variations for three months in 1979 of flight plan activity at the Los Angeles FSS.

The results of these analyses are given preliminary interpretations, and are available for possible application to other facilities. They are also available for other applications, such as estimation of instantaneous air counts in hubs.

17. Key Words	18. Distribution Statement
Terminal Radar Approach Control, Flight Service Station, Aircraft Counts, Aircraft Operations	DOCUMENT IS AVAILABLE TO THE PUBLIC THROUGH THE NATIONAL TECHNICAL INFORMATION SERVICE, SPRINGFIELD, VIRGINIA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	59	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

PREFACE

At the request of the FAA's Office of Systems Engineering Management, the Transportation Systems Center implemented a program to estimate instantaneous counts of airborne aircraft over the continental United States. Counts over Air Route Traffic Control Centers (ARTCC) and Terminal Radar Approach Control Facilities (TRACONS) are of special importance, since these counts help system designers develop system tradeoffs and technical specifications for new equipment and computers to update or replace those currently in the field (i.e., IBM 9020's in the ARTCCs and Univac ARTS-3 in the TRACONS).

In the past, there has been limited effort to systematically estimate counts. This report describes several important phases in a new structured program to systematically estimate air counts.

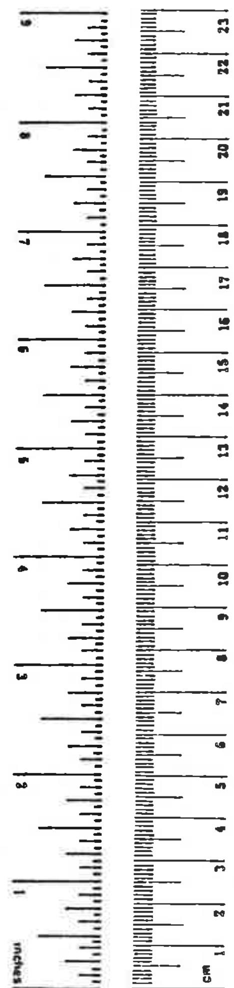
A crucial step in the estimation of air counts is the estimation of daily or hourly operations at FAA facilities. The principal outputs from this study are statistical methods for predicting hourly or daily operations with application to certain facilities. Concurrent with this research, models to convert operations to air counts have been developed, although these models are not described in this report. The prediction of facility operations is sufficiently important to FAA program planning to warrant a separate, comprehensive report. This work was performed in FY'80.

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price $2.25, SD Catalog No. C13.10:286.



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F

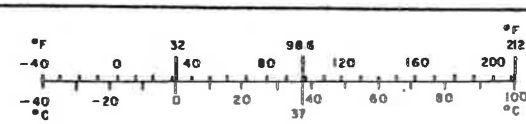


TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF TABLES (CONT.)

EXECUTIVE SUMMARY

Two statistical procedures have been developed to estimate hourly or daily aircraft counts. These counts can then be transformed into estimates of instantaneous air counts. The first procedure estimates the stable (deterministic) mean level of hourly or day of the week patterns by statistical models. The second procedure estimates both deterministic and stochastic periodic (hourly or day of the week) patterns by stochastic time series models. The foregoing statistical procedures have been used to analyze traffic at the St. Louis TRACON and Los Angeles Flight Service Station.

This study analyzes hourly variations in operations at the St. Louis TRACON, for each day in four representative months in 1979. It also analyzes daily variations for three months in 1979 of flight plan activity at the Los Angeles FSS. The results of these analyses are given preliminary interpretations, and are available for possible application to other facilities. They are also available for other applications, such as estimation of instantaneous air counts.

1. INTRODUCTION

The systematic estimation and prediction of characteristic variations in air traffic operations at FAA facilities is needed in order to help plan for future air traffic control systems, programs and possible facility modification. Two important examples of these characteristic variations, or representative patterns, are the changes in operations with day of the week and hour of the day.

However, hourly or daily patterns, while important to many applications, do not satisfy all requirements. Estimates of instantaneous aircraft counts are required for many applications. In the past, the methods for counting in-flight aircraft have consisted mainly of reducing radar tapes (when available), counting aircraft images on air traffic controller plan view displays, or on some occasions, requiring the assistance of the Civil Air Patrol to count in-flight aircraft. While there is merit in these approaches, they also have their limitations. For instance, they are costly. Therefore, mathematical models to convert FAA operations data to estimates of instantaneous counts have been developed at the Transportation Systems Center. This report will concentrate only on the analysis and modeling of daily and hourly operations data at certain FAA facilities. Their conversion to counts of in-flight aircraft will be discussed in a separate publication. But the major impetus for undertaking the analysis of hourly operations data is to provide input to models which estimate instantaneous aircraft counts.

Useful information can be gleaned from analysis of daily or hourly operations, and this information is relevant to interpretation of instantaneous aircounts. The first part of this report extracts weekly systematic patterns in daily instrument flight rules (IFR) and visual flight rules (VFR) flight plan activity of general aviation aircraft at the Los Angeles Flight Service Station. This information is important because it identifies preferred days. For example, which weekdays tend to be busy and which tend to be quiet, and which weekend days are busiest. This information is needed by the statistical survey designs used to obtain data from flight service stations (discussed in another report). Los Angeles FSS data is a necessary requirement to analyze aviation activity (with and without flight plans) in the crowded Los Angeles area. Since VFR activity data from different FSSs in a limited region tend to be correlated with each other, the daily variations of flight plan activity at Los Angeles must be typical of neighboring flight service stations (and also similar to temporal variations of itinerant GA activity at neighboring airfields).

The second part of the report develops models for systematic patterns and peak activity in a TRACON. A model and related software were developed and tested on data obtained from the St. Louis TRACON. Hourly operations data, for each day in four representative months in 1979, was the input. The model predicts hub loading on an hourly basis, and as indicated above, can be converted to peak and average instantaneous aircount within each hourly interval.

All models permit direct forecasting of their output. For flight plan forecasts, the level of forecasting is daily, compared with the FAA's annual forecasts of operations. The sensitivity of the model is thereby improved.

This study accomplishes two main purposes. First, special statistical models and procedures for the estimation of hourly and day of the week patterns for air traffic activity were developed. Second, these statistical procedures were applied by estimating hourly and day of the week patterns with selected air traffic data.

A common practice in previous studies of temporal variations in air traffic activity has been simply to plot changes in sample data as a function of day of week or time of day. This approach cannot provide reliable estimates of representative patterns in air activity. This is becuase these estimates are adversely affected by sampling variations. In this study, two statistical approaches are proposed to estimate the average level and variances of these temporal patterns.

In the first approach, a simple model, similar to a one-way analysis of variance model, is adopted for hourly or daily air traffic data. One standard and two robust statistical procedures are suggested to estimate the parameters of this model. The second approach applies a time series modeling technique suggested by Box and Jenkins (1970) to model temporal variation patterns for air traffic data. This approach tailors a specific model for each series. In general, it will provide a more suitable model to explain the temporal variations of a particular series than the model obtained by the first approach. However, the second approach is not economical if used to analyze a large set of air traffic data.

Both approaches have been applied to St. Louis TRACON data. These empirical results indicate that there are diurnal variations in the hourly data and weekly and seasonal variations in the daily data collected from different seasons in the year of 1979. Their estimated hourly and day of the week patterns provide valuable information for the estimation of instantaneous air counts (see Meyerhoff (1980)) which can be applied to the planning of air traffic control facilities and the airport capacity problem.

This study is organized into three sections. Section 2 presents our first statistical approach to estimate temporal variation patterns for air traffic activity. Section 3 presents Box-Jenkins time series approach to model the stochastic behavior of temporal variation patterns for air traffic activity. This approach is also applied to the estimation of daily models for flight plan data collected from the Los Angeles Flight Service Station. Summary and conclusions are contained in Section 4.

2. ESTIMATION OF HOURLY AND DAY OF THE WEEK PATTERNS FOR SAMPLED AIR TRAFFIC DATA

In Section 2.1, we present some standard and robust statistical procedures to be used in this study and the reasons why both of these two types of procedures will be used simultaneously in the analysis of sampled air traffic data. In Section 2.2, these statistical procedures are illustrated with estimation of hourly and day of the week patterns with sampled hourly and daily air traffic data.

2.1 STATISTICAL PROCEDURES

In applied statistics, statisticians usually face a problem of choosing alternative statistical procedures in their analysis. The general principle in guiding our selection of alternative techniques are: (1) statistical properties of alternative techniques; (2) the primary purpose of the study undertaken and; (3) computational costs and requirements associated with alternative procedures. In this study, the primary goal is to estimate the representative (average) hourly or day of the week patterns for sampled air traffic data. Therefore, those statistical techniques which will give reliable location and scale estimates (average patterns and variations around them for data under a wide variety of conditions) will be adopted.

One of the difficulties in analyzing air traffic data is that maverick observations invariably find their way into the data set. These observations may be due to meteorological IFR weather conditions, facility outages or other special events. These outliers are values which in a dramatic way do not conform to the average (typical) behavior of most observations. Planning a new facility on the basis of an occasional outlier is, in general, not cost-effective.

The most commonly calculated statistics are the mean and variance of a batch of numbers. These two statistics are not robust to a small fraction of the observations. For example, suppose we have six numbers: 4, 5, 3, 2, 4, 3. In this case, the arithmetic mean of 3.5 seems to be a good description of the average number. Suppose one more observation of 45 is added, the mean now jumps to 9.5. This value does not represent the major bulk of this data set. This example illustrates that mean and variance are not robust in the sense that they are seriously affected by a single large fluctuating observation.

One natural solution to this problem is to use robust statistical methods. The statistical procedures whose results are not dramatically affected by a small fraction of observations are usually called robust statistical procedures. This example will intuitively explain the reasons why robust statistical procedures will be adopted in our analysis of sampled air traffic data.

The standard and robust statistical procedure adopted in this study will be presented in the rest of this section.

Suppose sample operations data obtained at equal intervals are designated by X_{ij}, where i = 1, 2, ...7, and j = 1, 2, ... 52. In our case, i is the ith day of the week and j is the jth week of the year.

The statistical model in general form is:

$$X_{ij} = U_i + e_{ij} \tag{1}$$

where U_i represent the deterministic effect of the ith day of the week and e_{ij} is assumed to be synmetrically distributed with zero mean and variance σ^2_{ei}.

The estimator for the U_i is defined as:

$$\hat{U}_i = \sum_{j=1}^{N_i} X_{ij}/N_i \qquad i = 1, 2, \ldots 7 \tag{2}$$

and the variance estimator of $\hat{U}_i$ is defined as:

$$\hat{\sigma}^2_{e_i} = \sum_{j=1}^{N_i} (X_{ij} - \hat{U}_i)^2 / (N_i - 1) \tag{3}$$

where N_i is the number of observations of the ith day of the week.

These estimators can be extended to estimate the location (mean) and scale (variance) parameters of the distribution on the ith hour of the day using hourly rather than daily samples. The estimators described in (2) and (3) are commonly used and are referred to as a standard method.

The first robust statistical technique we use is the Box plot and schematic plot. These graphical display techniques were suggested by Tukey (1977). A Box plot displays a batch of data. Nine values to represent ranges in a set of data are conventionally used: the upper and lower detached points, the upper and lower outside points; the upper and lower side values; the upper and lower hinges (these are the same as upper and lower quartile) and the median. The configuration of a Box plot is shown in Figure 1. The definitions of these summary measures are:

(1) Upper Hinge: The definition of Upper Hinge is identical to that of the upper quartile, which is widely used in statistics. The Upper Quartile is the value such that 75 percent of the observations are less than or equal to the value denoted by Q_3; while 25 percent of the observations are larger than Q_3.

(2) Lower Hinge: It is identical to the definition of Lower Quartile. Lower Quartile is the value such that 25 percent of observations are less than or equal to this value, which is denoted as Q_1.

(3) Upper Side Value: It is the largest data value less than the upper hinge (i.e., Upper Quartile) plus Midspread [Upper Quartile - Lower Quartile]. But if that value is less than upper quartile, the upper side value is equal to upper quartile.

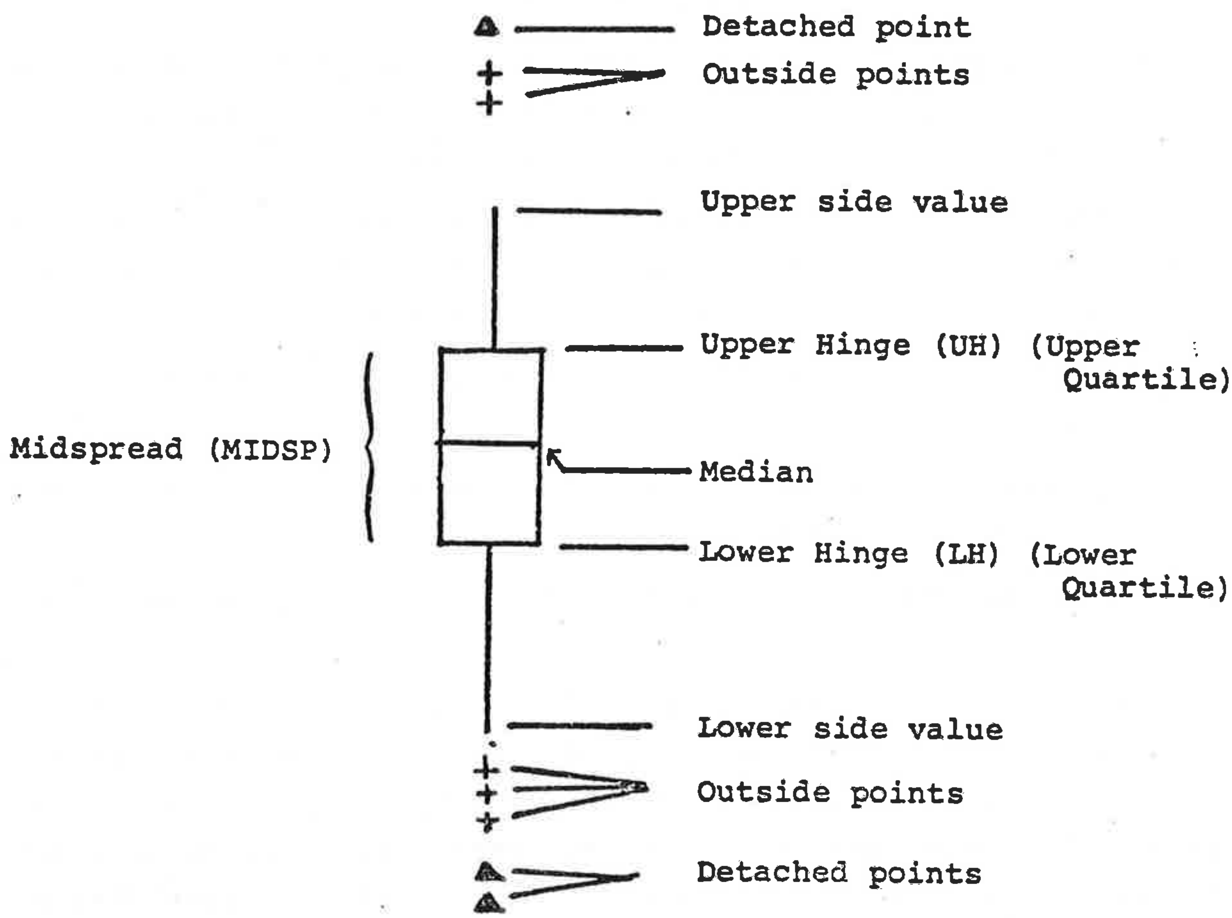


FIGURE 1. THE CONFIGURATION OF A BOX PLOT

(4) Outside Points: The definition of Outside Points are data values between UH + 1.5 MIDSP and UH + MIDSP or LH - 1.5 MIDSP and LH - MIDSP.

(5) Detached Points: The definition of Detached Points are data values greater than UH + 1.5 MIDSP or less than LH - 1.5 MIDSP.

(6) Median: It is the value such that 50 percent of the observations are less than it and 50 percent greater than it.

In this study, the Box plot technique is used to estimate the empirical distribution of the ith day of the week or the ith hour of the day. The merits of Box plots are as follows:

(1) They display the levels and variability of a batch of data.

(2) They are resistant to bad data points and indicate the numbers of bad data points in the plot.

(3) In the process of constructing a Box plot, outliers are identified.

For comparison purposes, the Box plot for each day of the week or for each hour of the day is displayed in the same figure and this figure is called a schematic plot by Tukey (1977). This displays the center and spreads of the empirical distributions for each day of the week or each hour of the day, thereby providing a preliminary model for the data.

The second robust statistical estimator for the mean U_i is the trimean estimator. It is defined as:

$$\hat{\hat{U}}_i = \frac{X_{(1/4)} + 2X_{(1/2)} + X_{(3/4)}}{4} \qquad (4)$$

where, $X_{(\frac{1}{4})}$ denotes the lower 25 percentile of the data batch

$X_{(\frac{1}{2})}$ denotes the median of the data batch

$X_{(3/4)}$ denotes the upper 75 percentile of the data batch.

The variance formula for the trimean is defined as scaled midspread (SMIDSP),

$$\text{Var}\ (\hat{\bar{U}}_i) = \frac{X_{(3/4)} - X_{(1/4)}}{1.349}$$

The robust estimators considered here work well for symmetrical distributions. If the empirical distribution is very skewed, then a power transformation [Box and Cox (1964)] of the original data should be used, and this robust estimator will be applied to the transformed data. The location and scale estimators for the original data can be obtained by transforming back from the location and scale estimates obtained from transformed data.

Other types of robust estimators such as a trimmed mean can be also used in this study. For computional purposes, this estimator is not adopted in this study.

Finally, our strategy in the estimation of representative patterns is as follows: (1) both standard and robust statistical methods are applied to the representative patterns of interest. If the results obtained by both procedures are similar, then the standard method is recommended; (2) On the other hand, if the results obtained from the standard method are quite different from the results obtained by those robust procedures, then there are two paths we can follow: (a) the results obtained by robust statistical methods is recommended. This is because these procedures have a built-in property to be resistent to bad data points and robust to the deviation from the assumed model of probability distribution; (b) further research should be taken to understand those reasons which may cause the occurrence of these bad data points. Then an alternative model should be specified for this data set. This approach is not economical in analyzing a large amount of data because it is tailor-made model for a limited number of <u>important</u> data sets.

2.2 EMPIRICAL RESULTS

2.2.1 Flight Plans Filed at Los Angeles Flight Service Station

The daily number of (IFR and VFR) flight plans filed at the Los Angeles Flight Service Station were processed for the period June to August 1979. They were supplied on Form 7230-13 by the Los Angeles FSS. These data are plotted in Figure 2 and Figure 3, which summarize, respectively, the variations in VFR and IFR flight plan activity for each day in the cited interval. There exist weekly cycles in the data over this sample period. The standard statistical method and the robust procedures described in the previous section are employed to estimate deterministic (fixed) day of the week patterns for VFR and IFR flight plan data. The estimated patterns and other related summary statistics of the data are presented in Tables 1 and 2. The graphical displays of schematic plots for the VFR and IFR day of the week patterns are shown in Figures 4 and 5, respectively.

From these empirical results, several interesting findings can be summarized as:

(1) On the average, the peak day of the week for VFR is Saturday while the peak day of the week for IFR is Friday.

(2) The number of IFR flight plans filed on a week-day is larger in general than those filed on a weekend. However, on the average, the number of flight plans filed in the weekend and Friday for VFR is larger than those in the rest of the week.

These observations verify our expectations. IFR activity is primarily business-related, so the number of IFR flight plans should decrease on weekends. Many VFR flight plans are filed by private individuals, so VFR activity should increase as the weekend approaches. (Many individuals leave the Los Angeles area on weekends.)

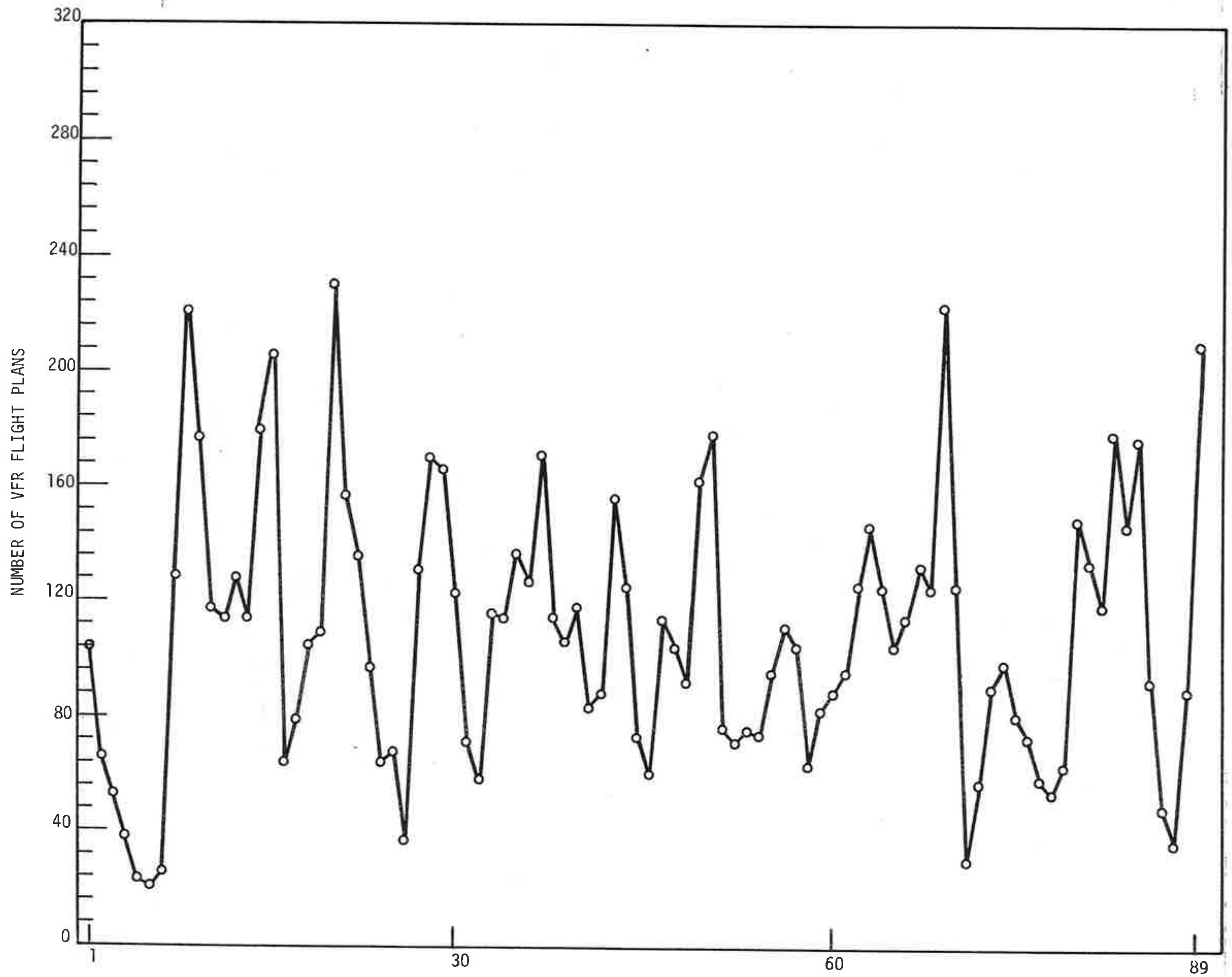


FIGURE 2. DAILY VFR FIGHT PLAN ACTIVITY AT THE LOS ANGELES FLIGHT SERVICE STATION, JUNE 1 - AUGUST 31, 1979

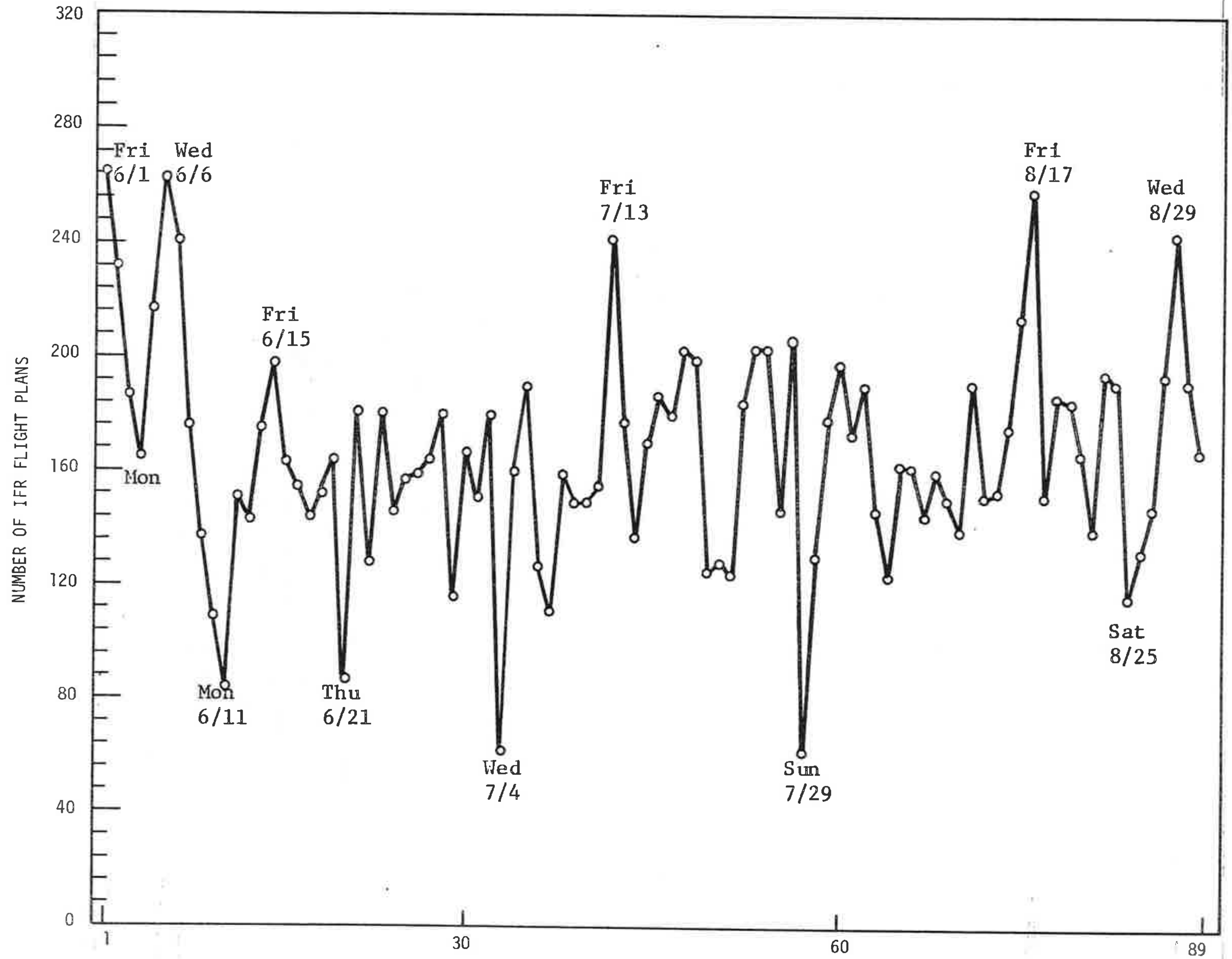


FIGURE 3. DAILY IFR FLIGHT PLAN ACTIVITY AT THE LOS ANGELES FLIGHT SERVICE STATION, JUNE 1 - AUGUST 31, 1979

TABLE 1. DAY OF THE WEEK PATTERNS (IN LEVELS) FOR DAILY NUMBER OF VFR FLIGHT PLANS FILED AT LOS ANGELES FLIGHT SERVICE STATION, JUNE, JULY, AUGUST 1979

	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.	Sun.
Mean	77.77	83.62	92.77	106.31	139.93	140.46	113.46
S.D.	23.19	33.61	38.96	46.22	47.14	46.42	52.32
Median	73.00	82.00	109.	104.	133.	146.0	123.
MIDSP	21.50	35.58	34.09	31.13	58.57	31.13	80.01
Tri-Mean	75.30	82.5	102.75	102.25	134.75	144.75	120.25
SMIDSP	21.50	34.09	31.88	25.95	61.53	27.43	79.31
Max.	117	148	133	230	222	221	178
Min.	38	23	21	26	73	58	30
Range	79	125	112	204	149	163	148

Note: S.D. = Standard Deviation

SMAD = scaled median absolute deviation,

i.e., $MED|X - Med(X)|/0.6745$

SMIDSP = scaled midspread i.e., $\left(\frac{X_{(3/4)} - X_{(1/4)}}{1.349}\right)$

TABLE 2. DAY OF THE WEEK PATTERNS (IN LEVELS) FOR DAILY NUMBER OF IFR FLIGHT PLANS FILED AT LOS ANGELES FLIGHT SERVICE STATION, JUNE, JULY, AUGUST 1979

	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.	Sun.
Mean	147.46	171.46	170.92	178.46	195.2	151.	143.69
S.D.	24.86	20.58	50.49	37.22	36.33	35.36	38.38
Median	151.	166.	164.0	175.25	190.	139.	137.
SMAD	20.76	22.24	37.1	37.87	17.79	34.09	44.48
Tri-Mean	152.	167.25	167.5	178.25	188.75	142.	144.5
SMIDSP	13.34	22.98	40.2	31.87	17.05	26.69	42.99
Max.	184	217	263	241	265	232	190
Min.	84	149	62	87	146	116	62
Range	100	68	201	154	119	116	128

Note: S.D = Standard Deviation

SMAD and SMIDSP are defined in the footnote of Table 1.

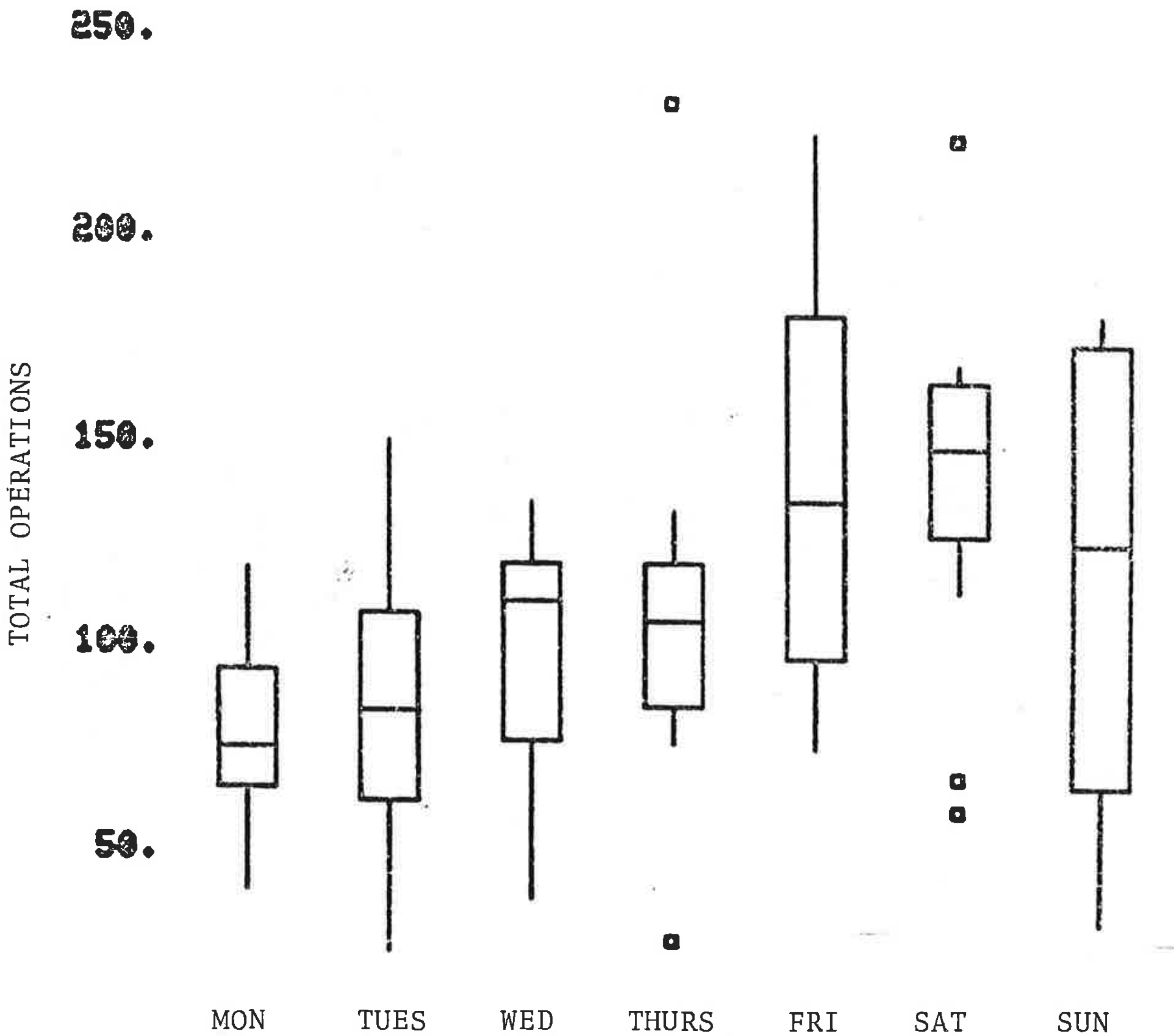


FIGURE 4. SCHEMATIC PLOT OF DAY OF THE WEEK PATTERN FOR VFR FLIGHT PLANS FILED FROM LOS ANGELES FSS

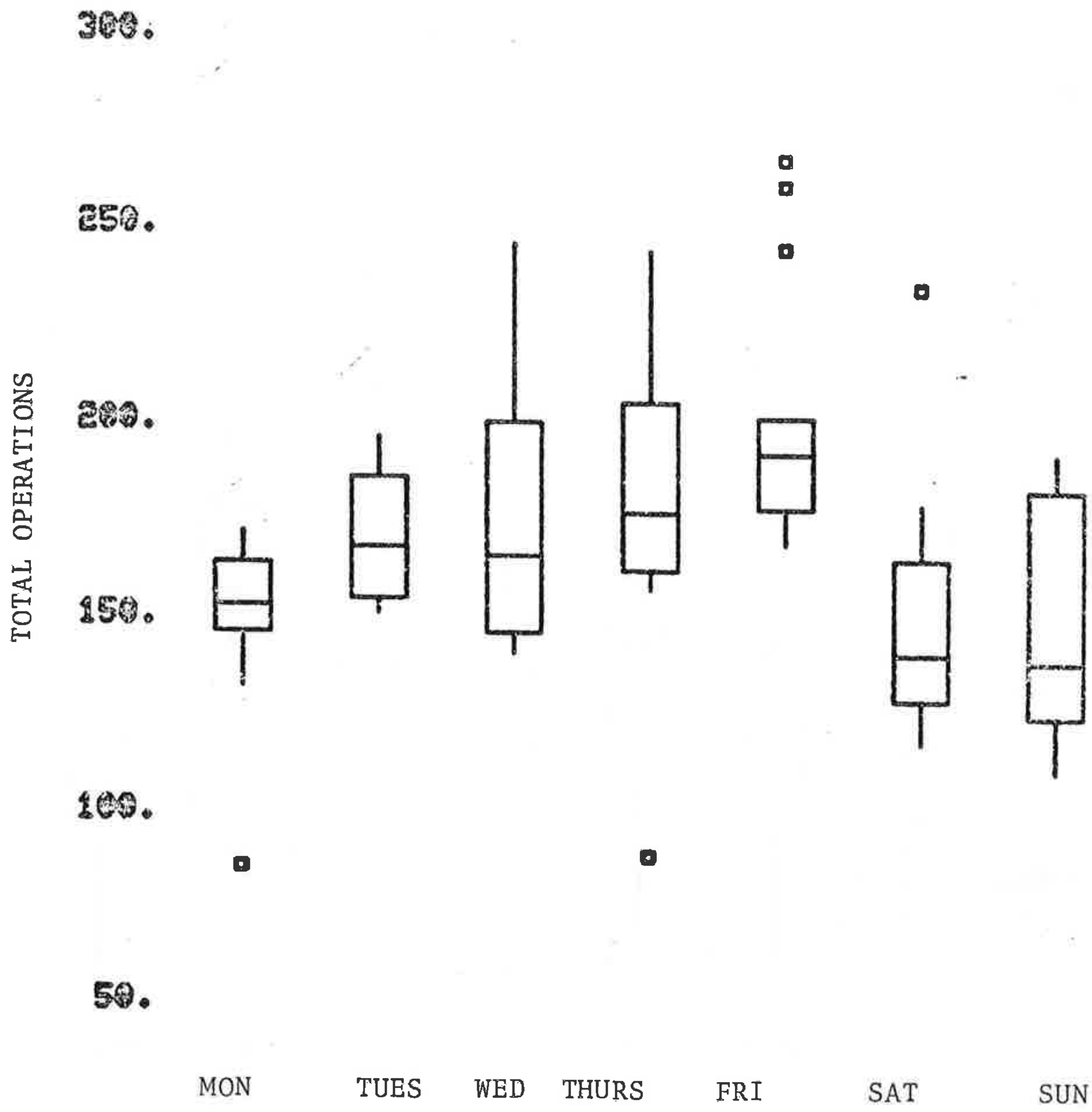


FIGURE 5. SCHEMATIC PLOT OF DAY OF THE WEEK PATTERN FOR IFR FLIGHT PLANS FILED FROM LOS ANGELES FSS

(3) From Tables 1 and 2, observe that there are some positive associations between the peak (i.e., max) flight plan counts for each day of the week and their corresponding location estimates (i.e., mean levels).

(4) To measure the variability of the data around their mean level, the coefficients of variations for each day of the week are computed and presented in Table 3. The coefficient of variations (C.V.) of IFR for each day of the week are uniformly smaller than those for VFR. Operationally, daily variations in VFR activity tend to be larger than IFR because the former are more sensitive to weather.

2.2.2 Hourly Operations at St. Louis TRACON

This section summarizes the result of statistical analyses on operations data at the St. Louis TRACON. The St. Louis TRACON records hourly operations for each day of the year in manual form. These operations are disaggregated in various ways, such as by type of aircraft. The analyses described herein were applied only to the total aircraft count. The procedures are equally applicable to disaggregated counts. They can also be applied to any TRACON tabulating its operations in the manner described.

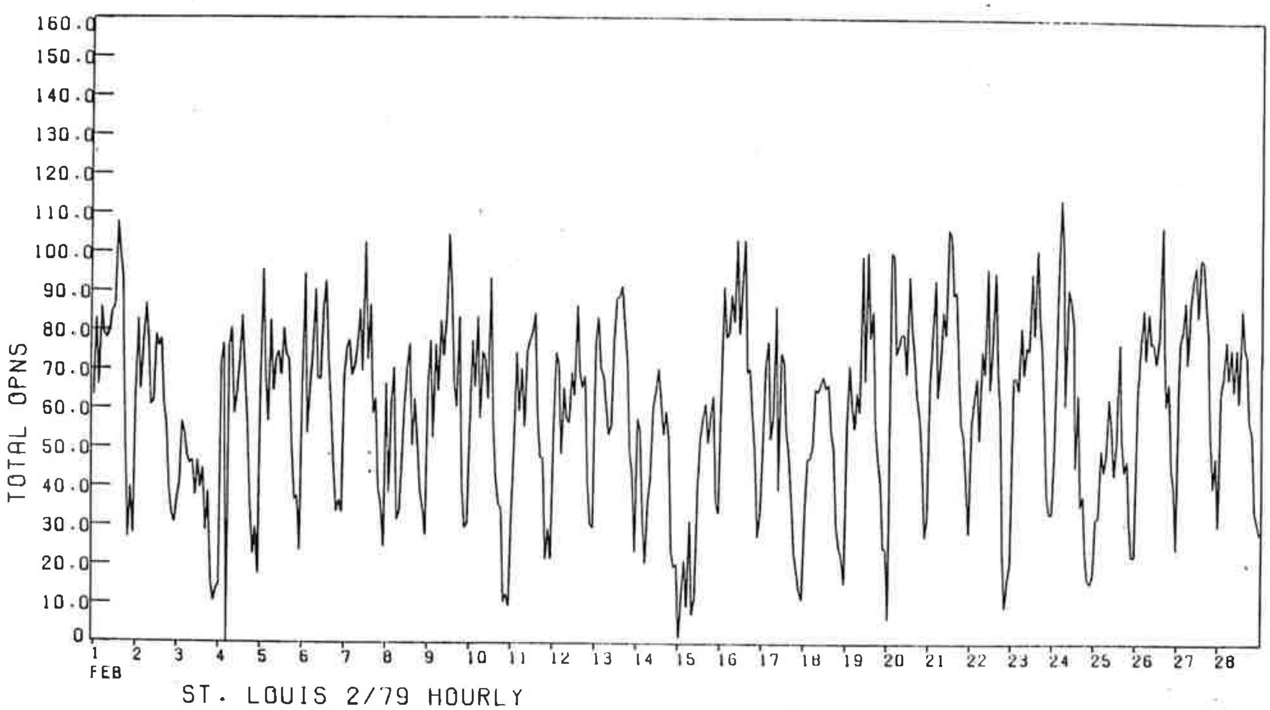
Manual data supplied by the St. Louis TRACON were automated at TSC, then analyzed using specially developed software. Daily data from the middle month of each season of the year was supplied (i.e., February, May, August, November). The assumption is that the middle month is representative of its season, e.g., February is representative of January and March. Experience indicates that the middle month of three consecutive months is representative of the two adjacent months. To have performed the same analyses for 12 months would have been costly.

Hourly operations data at the St. Louis TRACON for each day in February, May, August and November, 1979, are displayed in Figures 6 and 7, respectively. It is hoped that a study of hourly

TABLE 3. COEFFICIENTS OF VARIATION IN FLIGHT PLANS FILED AT LOS ANGELES FLIGHT SERVICE STATION FOR DAY OF THE WEEK (JUNE, JULY, AUGUST, 1979)*

	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.	Sun.
IFR	0.08	0.14	0.23	0.18	0.09	0.18	0.30
VFR	0.28	0.41	0.31	0.25	0.45	0.19	0.65

*The coefficient of variation used here is defined as $C.V. = \frac{SMIDSP}{Tri\text{-}Mean}$. Clearly the lower bound of C.V. is zero. Small values of C.V. indicate less variation around the mean level.



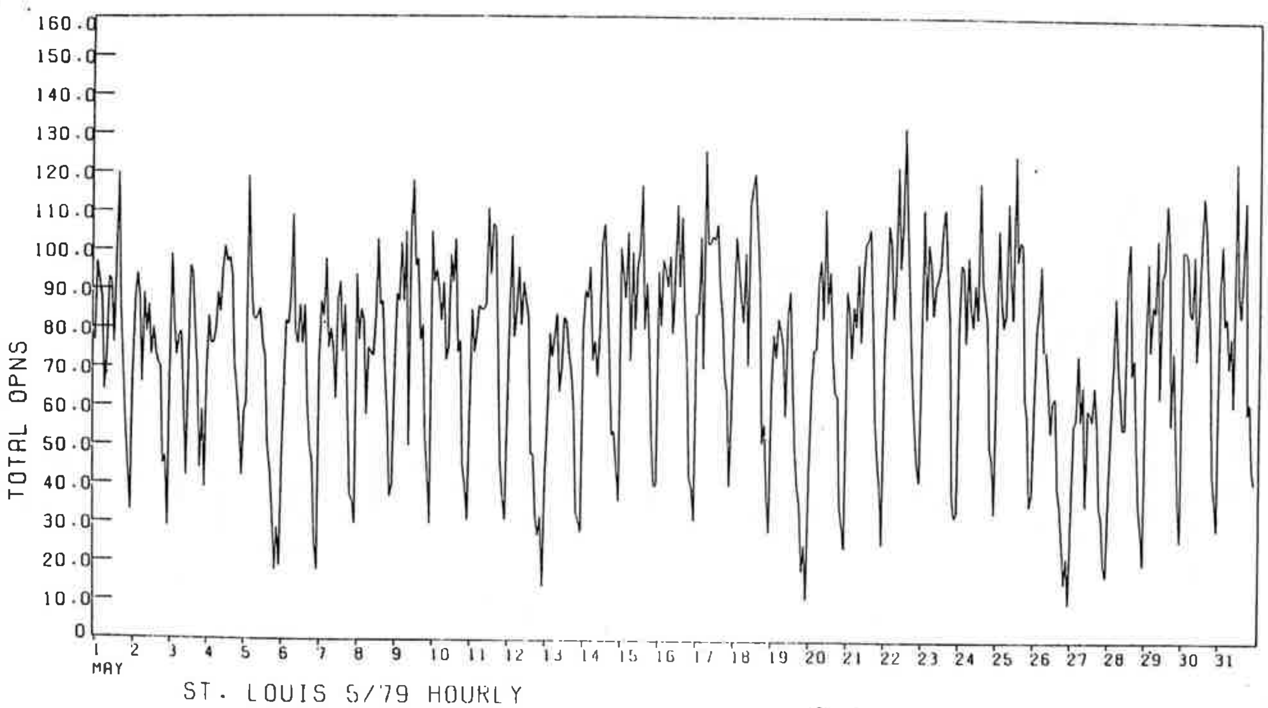
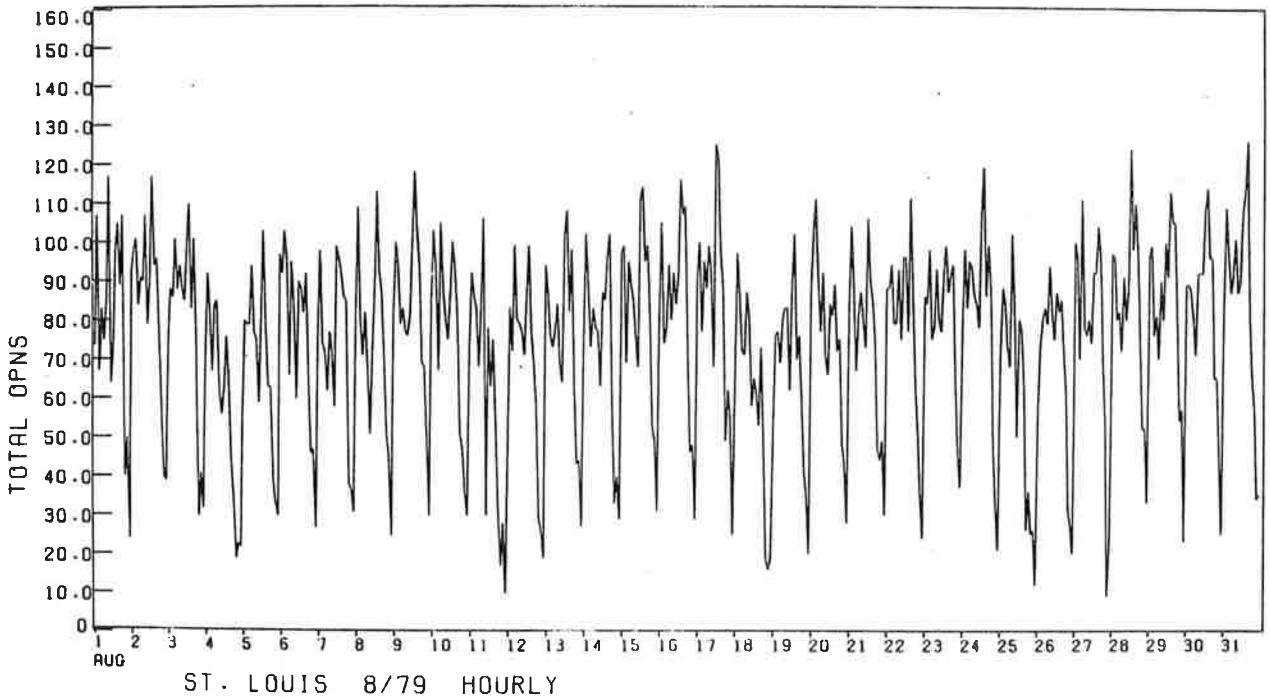


FIGURE 6. HOURLY OPERATIONS AT ST LOUIS TRACON, EACH DAY IN FEBRUARY AND MAY 1979



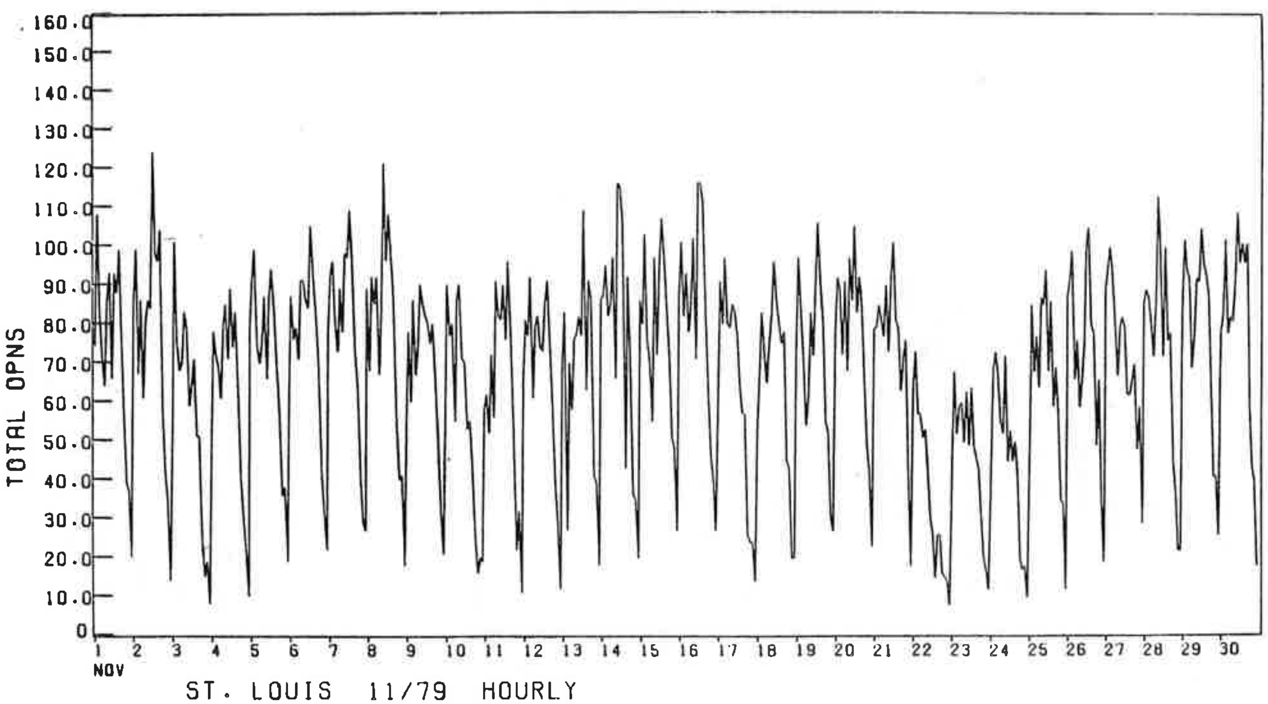


FIGURE 7. HOURLY OPERATIONS AT ST. LOUIS TRACON, EACH DAY IN AUGUST AND NOVEMBER 1979

data collected for one month from each of the four seasons will provide insight into the changes in levels and variability of hourly patterns of the day due to seasonal variations. Four hourly patterns (from 8 to 24 hours) of a typical day estimated from four different sampled months data are reported in Tables 4 through 7, respectively. Figures 8 through 12 display representative hourly patterns for February, May, August and November and all four months combined. From these empirical results, the major characteristics of hourly patterns in different seasons can be summarized as follows:

(1) The peak hour of the day is from 17 to 18 hour, i.e., 5 to 6 p.m. This is true for all sampled monthly data considered here.

(2) The peak hour of the morning period (defined by the range 8 a.m. to 12 a.m.) is in the interval from 9 to 10 a.m. Again, this morning peak pattern remains the same for all of the monthly data examined in this study.

(3) In general, the shape of the hourly patterns for different sampled months are similar but the levels of hourly patterns vary with different seasons. For example, the levels of hourly patterns in August is the highest and the hourly pattern in February is the lowest among these four sampled monthly data. This implies that there is some seasonal variation in the levels of hourly patterns for operations in the St. Louis TRACON. These regular hourly patterns for the different sampled months has its origin in the observation that pilots' habits are fairly regular. This regular pattern is an important prerequisite for forecasting and also for inferring operations at other TRACONS.

(4) From Tables 7 to 10, the variability of the hourly pattern in August is less than those estimated from the rest of three sampled months. This is because the weather at this time of the year is relatively good, i.e., VFR.

TABLE 4. THE HOURLY PATTERNS OF A TYPICAL DAY IN FEBRUARY 1979, ST LOUIS TRACON

Hours	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Mean	53.3	70.4	69.2	61.6	67.1	68.9	70.3	70.71	79.4	80.8	69.3	59.7	54.1	34.6	29.9	25.1
S.D.	20.1	21.0	19.2	22.3	16.0	19.1	20.1	14.9	17.2	16.9	16.5	14.6	14.9	13.5	9.9	6.9
Median	60.	75.5	69.5	67.5	70.5	75.5	73.5	70.0	79.5	81.5	70.5	60.5	54.0	38.	30.5	25.0
SMAD	11.1	11.8	13.3	17.8	12.6	15.5	18.0	11.1	17.1	14.8	22.2	13.3	9.6	16.3	10.4	8.2
Tri-mean	60.0	73.3	69.3	65.8	69.6	73.	71.3	70.6	80.5	81.5	69.3	60.3	52.4	35.3	30.3	25.5
SMIDSP	16.6	17.8	14.1	16.1	12.2	17.1	16.3	11.5	17.04	15.9	20.0	12.6	12.3	13.7	10.4	7.4
Max.	78.	101	115	100	90	94	104	107	105	108	100	94	84	64	49	34
Min.	2	13	22	0	32	8	12	38	46	40	36	29	24	10	11	10
Range	76	88	93	100	58	86	92	69	59	68	64	65	60	54	38	24

TABLE 5. THE HOURLY PATTERNS OF A TYPICAL DAY IN MAY, 1979, ST. LOUIS TRACON

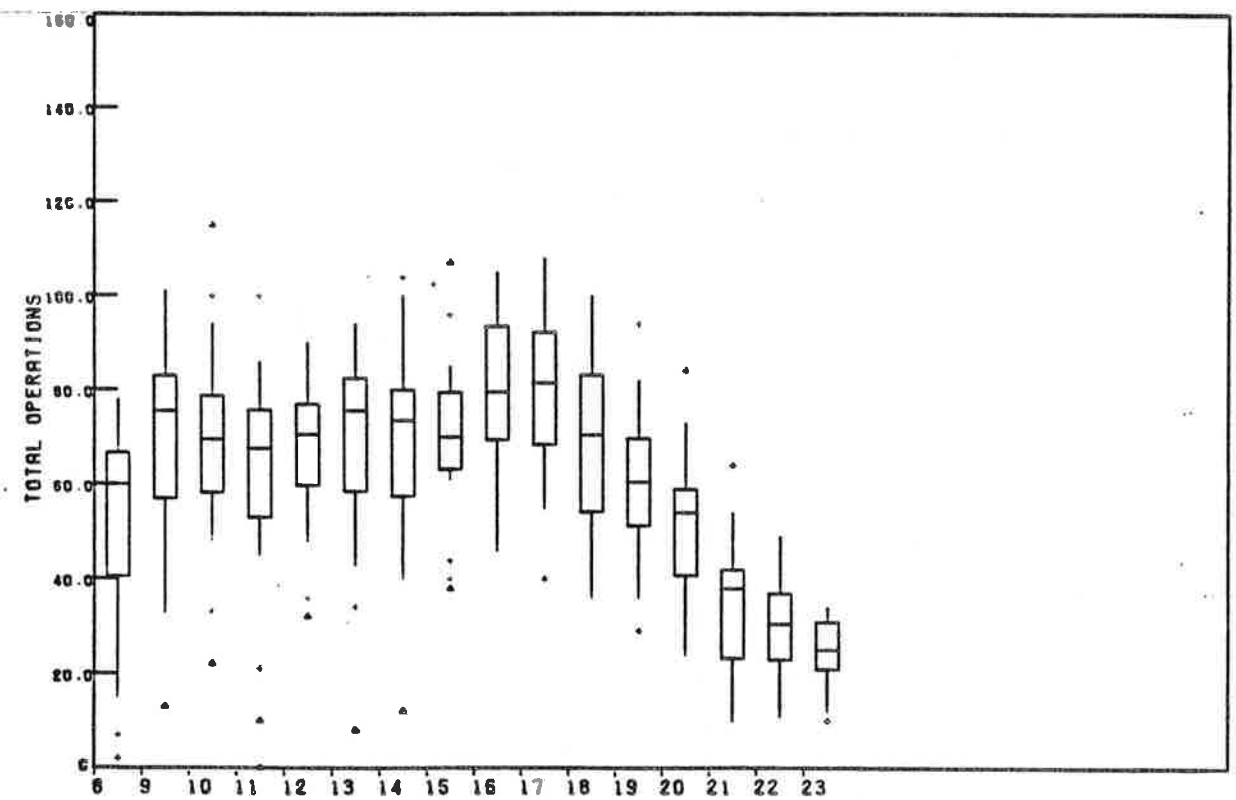
Hours	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Mean	72.4	85.3	86.7	86.4	81.0	88.8	75.5	85.4	92.7	95.4	87.5	77.2	61.	44.6	37.5	29.9
S.D.	18.8	15.2	12.7	9.8	12.9	12.6	16.3	18.2	16.	17.4	21.1	19.3	16.0	13.3	9.0	10.6
Median	74	87	85	86	82	89	77	86	94	96	93	83	66.0	46.0	37	30.0
SMAD	22.2	14.8	14.8	13.3	10.4	16.3	13.3	14.8	16.3	17.8	20.7	16.3	14.8	14.8	10.3	11.8
Tri-Mean	73.0	87.5	85.5	86.2	80.5	89.	75.3	85.5	93.5	95.5	91.5	80.0	64.7	45.5	37.5	30.5
SMIDSP	19.3	13.3	14.8	12.6	10.4	16.3	14.1	14.8	14.8	17.8	23.7	19.3	17.1	14.8	10.3	14
Max.	105	111	119	109	126	113	122	124	125	132	120	114	83	64	59	61
Min.	34	56	57	70	57	64	35	42	59	57	40	35	26	15	20	10
Range	71	55	62	35	69	49	87	82	66	75	80	79	57	49	39	51

TABLE 6. THE HOURLY PATTERNS OF A TYPICAL DAY IN AUGUST 1979, ST. LOUIS TRACON

Hours	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Mean	80.0	94.8	84.3	81.9	80.5	86.8	81.8	71.5	92.7	98.2	90	83.9	65.7	44.2	38.7	26.2
S.D.	13.9	9.3	10.7	11.3	9.7	7.9	12.6	14.6	14.4	16.7	15.5	18.4	16.9	13.8	11.4	6.5
Median	83.0	97	83	80	80	87.0	80.0	71	91	99	92	88	65	46	40	27
SMAD	14.8	10.4	10.4	11.9	11.8	8.9	7.4	14.8	13.3	20.7	17.8	16.3	16.3	11.8	11.8	5.9
Tri-Mean	82.3	96	83.7	80.7	80	86.7	80	71.3	92.3	99	92.5	86.5	66	44.5	40.	26.5
SMIDSP	12.6	10.4	12.6	12.6	11.8	9.6	8.1	14.1	11.1	20.7	16.3	17.8	13.3	14.8	11.8	5.9
Max.	100	109	111	111	105	102	117	100	125	121	126	109	107	68	57	39
Min.	45	74	67	67	62	68	51	30	62	61	53	26	31	17	9	10
Range	55	35	44	44	43	34	66	70	63	60	73	83	76	51	48	29

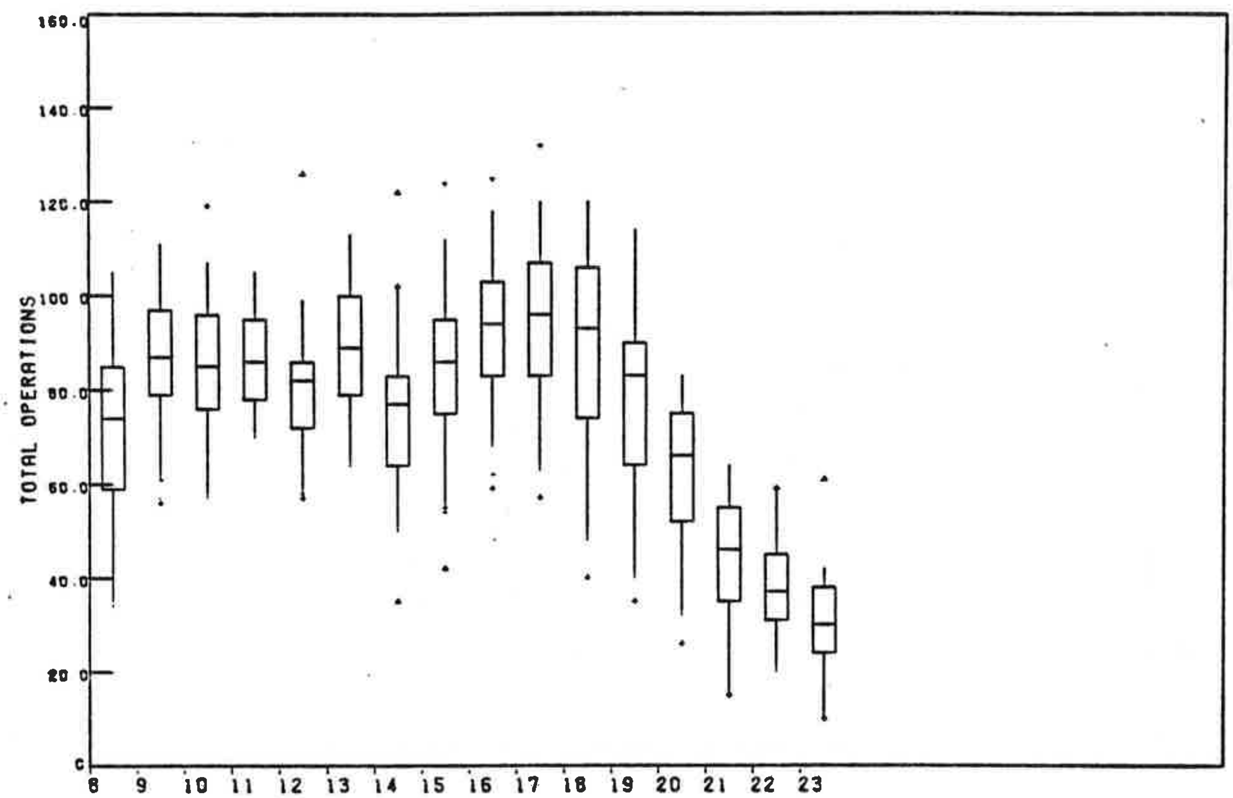
TABLE 7. THE HOURLY PATTERNS OF A TYPICAL DAY IN NOVEMBER 1979, ST. LOUIS TRACON

Hours	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Mean	69.5	85.5	79.5	77.5	68.9	76.0	83.1	80.2	85.7	85.1	77.1	72.9	53.13	38.5	32.2	18.4
S.D.	15.4	11.8	17.5	10.1	10.9	12.8	13.12	19.6	20.1	23.4	21.1	18.2	17.6	14.4	10.3	6.2
Median	73.5	87.0	82.0	77.5	69.5	79.	84.5	82	87.5	87	81.5	78	57	40.5	33	19
SMAD	18.5	11.1	16.3	11.1	12.6	11.1	9.6	16.3	16.3	25.2	21.5	13.3	16.3	9.6	11.1	7.41
Tri-Mean	72.5	86.0	82.5	77.5	69.5	77.8	84.5	81	86.3	88.3	79.5	76	56.3	39.3	32	18
SMIDSP	20.01	10.4	16.3	11.1	11.8	14.1	9.6	15.0	17.7	22.9	27.4	14.8	15.5	11.9	13.3	7.4
Max.	89	108	103	97	92	97	113	121	124	116	111	104	86	76	59	29
Min.	42	62	27	57	51	50	43	31	26	15	26	26	16	15	14	8
Range	47	46	76	40	41	47	70	90	98	101	85	78	70	61	45	21



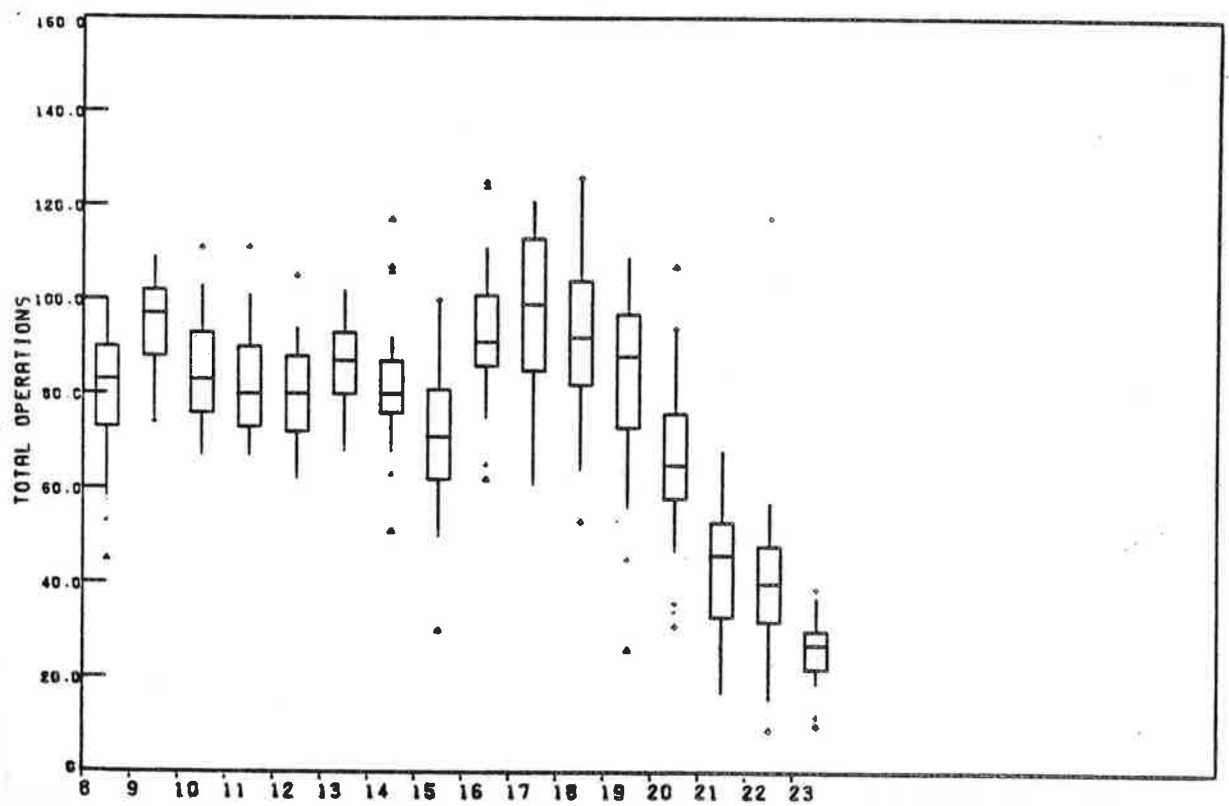
St. Louis Hourly TRACON Operations for Feb. 1979

FIGURE 8. ST. LOUIS TRACON, HOURLY OPERATIONS FOR FEBRUARY 1979



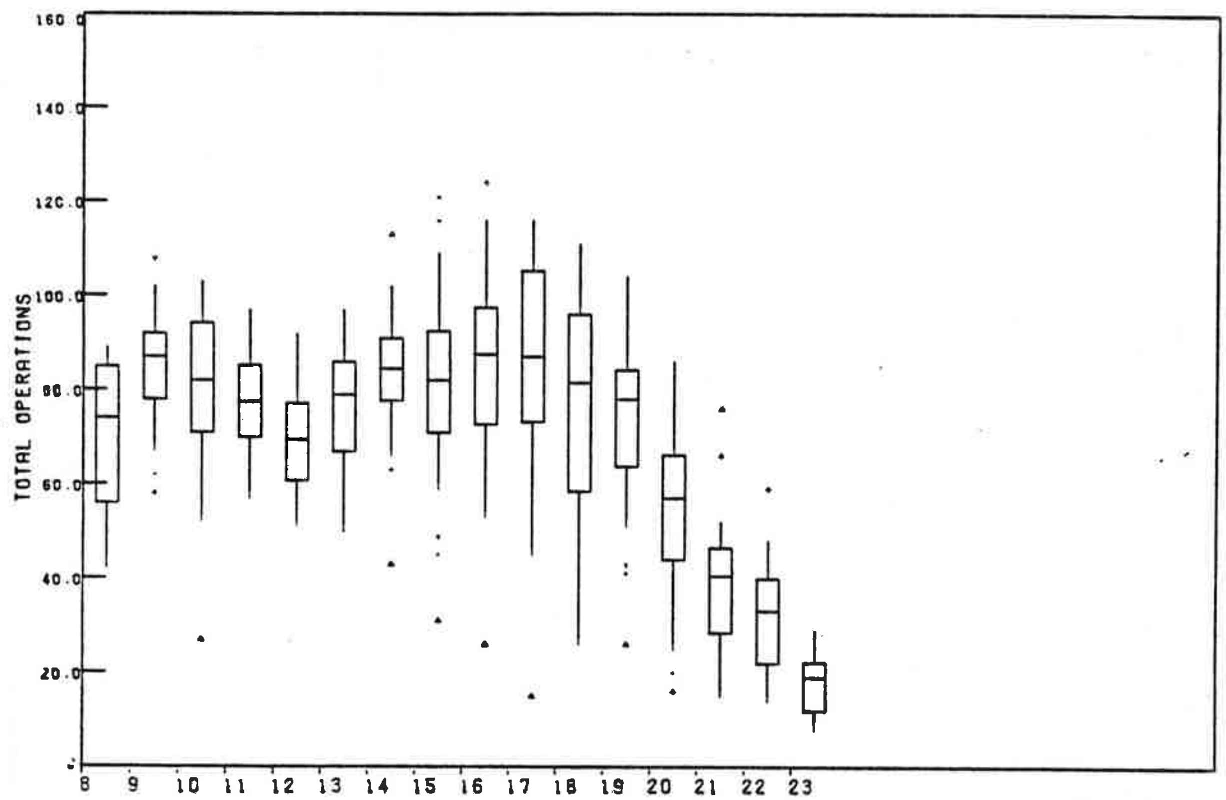
St. Louis Hourly TRACON Operations for May 1979

FIGURE 9. ST. LOUIS TRACON, HOURLY OPERATIONS FOR MAY 1979



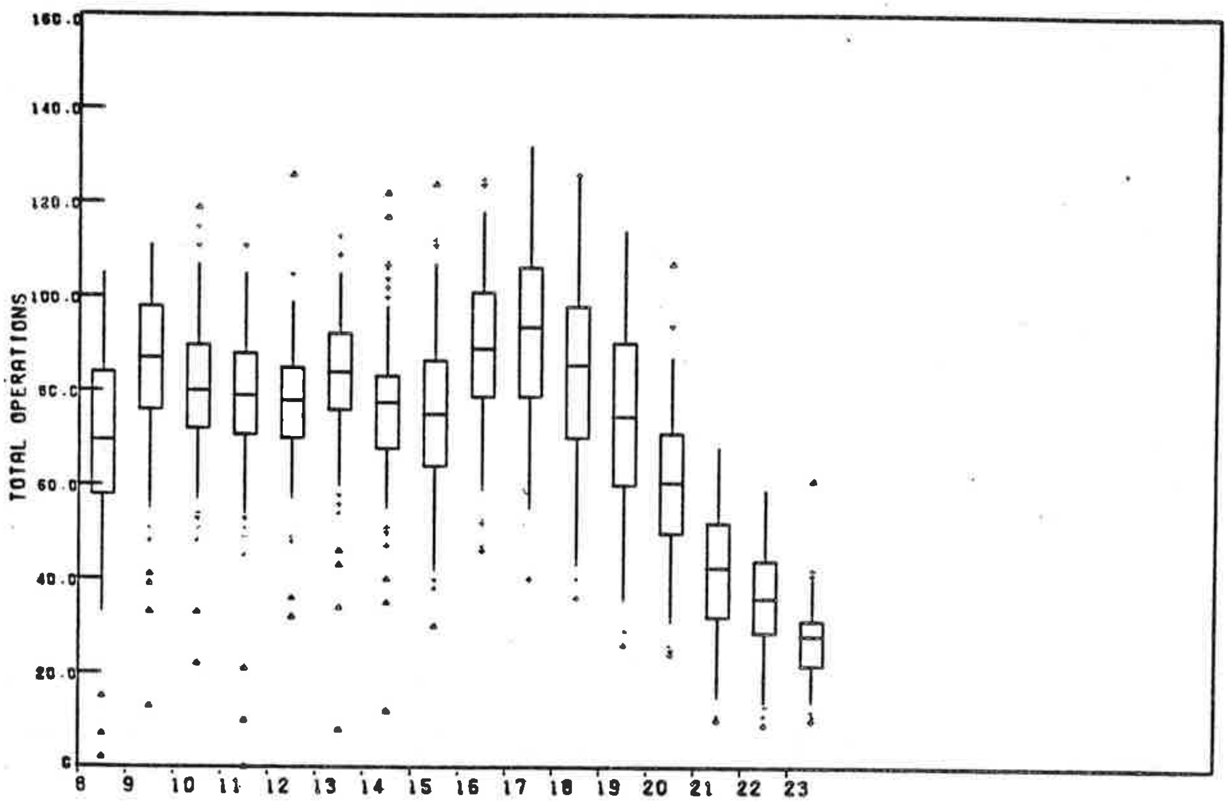
St. Louis Hourly TRACON Operations for August 1979

FIGURE 10. ST. LOUIS TRACON, HOURLY OPERATIONS FOR AUGUST 1979



St. Louis Hourly TRACON Operations for Nov. 1979

FIGURE 11. ST. LOUIS TRACON, HOURLY OPERATIONS FOR NOVEMBER 1979



St. Louis Hourly TRACON Operations for Feb., May, Aug. and Nov. 1979 combined

FIGURE 12. ST. LOUIS TRACON, HOURLY OPERATIONS FOR FEBRUARY, MAY, AUGUST, AND NOVEMBER 1979, COMBINED

For planning and comparison purposes, convert the day of the week and hourly patterns (in levels) into an indexed form. The day of the week pattern in index form can be calculated in the following manner:

(1) Compute the overall mean from the means of the ith day, $i = 1, 2, \ldots, 7$, i.e.,

$$U_i = \sum_{i=1}^{7} \frac{\hat{U}_i}{7} .$$

(2) Divide the mean of ith day $\hat{U}_i$ by $\hat{U}$, i.e.,

$$\overline{U}_i = \hat{U}_i / \hat{U}, \quad i = 1, 2, \ldots 7.$$

(3) Multiply $\overline{U}_i$ by 100.0 and the resulting figure is the index number for the ith day of the week. The sum of the index numbers for the seven days of the week is equal to 700.

The index form of hourly patterns can be computed in the same way as the index number for the day of the week patterns except that the notation of the ith day is replaced by the notation of the ith hour and the periodity of 7 is replaced by the periodity of 16.

Following these procedures, the day of the week and hourly patterns in index form are presented in Tables 8 and 9, for the Los Angeles Flight Service Station and St. Louis TRACON operations.

Tabulated operations at any TRACON include only those aircraft in contact with the TRACON and within the boundary of the TRACON (or several miles beyond the boundary). Normally, this includes hub IFR or controlled VFR flights.

The St. Louis TRACON does not exhibit two distinctive operations peaks - one in the morning and another in the afternoon - as are observed in other TRACONS (and ARTCCS). Such a bi-modal profile is typical of a commuter morning and evening "rush hour" pattern. The St. Louis TRACON also has this pattern, but aircraft arriving from other time zones smooth out this pattern. Hence, St. Louis is typical of other TRACONS in Central time zones. More

TABLE 8. INDICES OF THE DAY OF THE WEEK PATTERNS FOR FLIGHT PLANS FILED IN THE LOS ANGELES FLIGHT SERVICE STATION (JUNE - AUGUST 1979)

Type of Flight Plan	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.	Sun.
IFR	93.31	102.67	102.83	109.43	115.86	87.18	88.71
VFR	69.12	75.75	94.32	93.86	123.74	132.87	110.42

TABLE 9. INDICES OF HOURLY PATTERNS OF A TYPICAL DAY FOR OPERATIONS AT ST. LOUIS TRACON, 1979

Hours *	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Feb	97.17	118.70	112.26	106.56	112.76	118.21	115.46	114.33	130.36	131.98*	112.23	97.65	84.85	57.16	49.09	41.30
May	97.23	116.59	113.92	114.85	107.26	118.58	100.33	113.92	124.58	127.25*	121.92	106.60	86.21	60.63	49.97	40.64
August	109.01	127.15	110.86	106.89	105.96	114.83	105.96	94.43	122.65	131.12*	122.5	114.56	87.4	58.94	52.98	35.01
Nov	104.77	124.27	119.22	111.99	100.43	122.11	117.05	124.7	127.6*	114.88	109.82	81.36	56.79	46.79	46.24	26.0

* Denotes the peak hour of the day.

**8 is taken to mean 8-9, 9 is taken to mean 9-10, etc.

analysis would be required to confirm this observation, although similar patterns have been observed in ARTCCs within the Central U.S. (Meyerhoff, 1979).

3. STATISTICAL TIME SERIES MODELS AND FORECASTS FOR SAMPLED AIR TRAFFIC OPERATIONS DATA

In the last section, statistical procedures were proposed to estimate representative (average) hourly and the day of the week patterns for sampled air traffic data. However, the major drawbacks of these procedures are:

(1) They fail to describe stochastic behavior (changing patterns) of the data over the sample period.

(2) In the estimation of the location parameters of our interest, the estimates are not precise because we neglect the information of the serial correlations contained in the sample data.

In order to remedy these problems, a time series modeling technique proposed by Box and Jenkins (1970) will be introduced and illustrated with sampled air traffic data later in this section. The main advantage of time series modeling approach is that it can describe both deterministic (fixed) and stochastic (changing) components in the data. However, the disadvantage of this procedure is:

(3) the implementation of this method generally requires sophistication in statistics.

(4) the construction of a fitted model requires relatively higher computation cost. However, once the model is constructed, the person exercising this mode need not possess special skills in statistics.

3.1 AUTOREGRESSIVE AND INTEGRATED MOVING AVERAGE (ARIMA) MODELING METHODOLOGY

Suppose time series data are observed at equal intervals and are denoted by X_1, X_2, X_3 ... X_{t-1}, X_t, X_{t+1} ..., X_N. B represents a back shift operator such that $BX_t = X_{t-1}$. A class of seasonal models for X_t proposed by Box and Jenkins (1970) is written as:

$$\Phi_1(B)\Phi_2(B^s)(1-B)^d\ (1-B^s)\ X_t(x) = \theta_o + \theta_1(b)\theta_2(B^s)e_t \qquad (5)$$

where:

(1) $\Phi_1(B)$ and $\theta_1(B)$ are polynomials in B of degree p and q, respectively. All of the roots of $\Phi_1(B)$ and $\theta_1(B)$ lie outside the unit circle.

(2) $\Phi_2(B^s)$ and $\theta_2(B^s)$ are polynomials of degree p and Q, respectively, in B^s. The periodicity in the data is denoted by s. For example, s=12 for monthly data and s=7 for weekly cycles. All roots of $\Phi_2(B^s)$ and $\theta_2(B^s)$ lie outside the unit circle.

(3) θ_o is a trend term.

(4) $(1-B)^d$ and $(1-B^s)^D$ represent the degree of difference and periodic difference.

(5) $X_t^{(\lambda)}$ denotes the Box-Cox transformation (see Box and Cox (1964), Box and Jenkins (1973) and $X^{(\lambda)}$ =

$\frac{X_t^\lambda - 1}{\lambda}$. When $\lambda \to o$ then X_t is equal to the log transform of the original observations and when $\lambda=1$, there is no transformation.

(6) e_t is assumed to be independent and normally distributed with zero mean and constant variance σ_e^2.

The Box-Jenkin's iterative modeling procedure for fitting ARIMA models to the data consists of the following steps:

(1) Model Identification

(2) Parameter Estimation

(3) Diagnostic Checking

(4) Forecasting

These steps are executed sequentially, with step 4 conditional on satisfactory results of step 3. If step 3 shows that the fitted model is not adequate in the sense that residuals are not white noise then a modified model must be proposed and then steps 1 to step 3 will be repeated again, as follows:

(1) Model Identification: The function of the model identification is to determine a tentative model which might be appropriate to the series being considered. The sample autocorrelation functions and partial sample autocorrelations are the major tools to be used to determine if the data can be represented by a stationary or nonstationary time series and also to determine which model in the class of the ARIMA process appears to be most appropriate. The transformation parameter λ of the data can be roughly estimated from the range-mean plot of original data. Alternatively, the transformation parameter can be estimated simultaneously with the other parameters of the tentatively identified model in the second step of estimation.

(2) Estimation: The parameters of the tentatively identified model with the transformation parameter λ of the data are estimated by non-linear least squares. These estimates are a symptotically equivalent to the maximum likelihood estimates for parameters of the model.

(3) Diagnostic Checking: The estimated model is considered adequate if the residuals from the estimated model can be considered as white noise. To test the hypothesis that the residuals are the realization of a white noise process, sample autocorrelations of the residuals are calculated and compared with their standard errors. Then a joint test of the serial independence of the residuals suggested by Ljung and Box (1978) is made by calculating:

$$\tilde{Q} = n(n+2) \sum_{K=1}^{J} (n-K)^{-1} \gamma^2_{\hat{e}_t}(K) \qquad (6)$$

where,

n = number of data observations remaining after the application of the difference operators; and $\gamma_{\hat{e}}(K)$ = sample autocorrelation of the residual $\hat{e}_t$

Under the null hypothesis of serial independence, $\tilde{Q}$ is asymptotically distributed as λ^2 with (J-p-q) degrees of freedom. The estimated model is considered adequate if we fail to reject the null hypothesis of independence of the estimated residuals.

(4) Forecasting. Suppose at time t an observation is forecasted ℓ time periods ahead. Following Box and Jenkins, define an origin t for the lead time ℓ. It has been shown that at time t the minimum mean square error of a forecast of a future observation $X_{t+\ell}$ is given by:

$$X_t(\ell) = E(X_{t+\ell}|_t) \tag{7}$$

where, $E(X_{t+e}|_t)$ denotes the conditional expectation of $X_{t+\ell}$, which is conditional on all the information up to and including time t. The procedure to calculate $E(X_{t+\ell}|_t)$ for an ARIMA model is best illustrated by an example. Suppose a model represented by:

$$(1-B)(1-\theta B^7)X_t = (1-\Phi_1 B)e_t \tag{8}$$

The Equation (8) can be written as:

$$X_t = X_{t-1} + \theta X_{t-7} - \theta X_{t-8} + e_{t-1} + e_t$$

Recall that $E(e_t) = 0$. The point forecast for $X_t(t+\ell)$ can be computed as:

$$\hat{X}_t(t+1) = X_t + \theta X_{t-6} - \theta X_{t-7} + {}_1e_t, \quad \ell = 1$$

$$\hat{X}_t(t+\ell) = X_{t+\ell-1} + \theta X_{t+\ell-7} - \theta X_{t+\ell-8}, \quad \ell > 1$$

The derivation of general formula for the variance of the point forecasts for an ARIMA model is found in Box and Jenkins (1970) and Fuller (1976).

3.2 EMPIRICAL RESULTS

In this section, Box-Jenkin's iterative modeling procedure is applied to estimate ARIMA models for daily flight plan data filed at the Los Angeles Flight Service Station. Two daily time series of 92 values each are available to us from June 12 to August 31, 1979. These data are the daily number of VFR and IFR flight plans, respectively, filed at the Los Angeles FSS. The plots of these two series are shown in Figures 2 and 3. These plots suggest that there are stochastic weekly cycles in these daily data. Pilots' habits tend to be repetitive such as prefered times of the day for departures.

For identification purposes, the sample autocorrelations of VFR, X_{1t}, and IFR, X_{2t}, are computed and presented in Tables 10 and 11. In order to examine the effects of removing weekly cycles in the data, the seventh difference of these original data are taken. (Removing weekly cycles helps to study the auto correlation of variations in other components of daily data. These results are reported in Tables 12 and 13, respectively.)

The behavior of these sample autocorrelations of the original and differenced series suggest that we might tentatively identify the model,

$$(1-\Phi_1 B)(1-B^7)\ X_{it}^{(\lambda)} = (1-\theta_7 B^7)\ e_t \quad (i = 1,\ 2.) \qquad (9)$$

as being appropriate for both sets of daily data.

A non-linear least square fit is used to estimate the parameters of the models as well as the transformation parameter λ of the original data. The estimated time series models and associated statistics are summarized in Table 14. The number in parentheses under each estimated parameter is the associated standard error. Since there is no unique representation of ARIMA models for each series, a competing model called (b) for each data set is also fitted and presented in Table 14.

To test the adequacy of these estimated models, two diagnostic checking tests are used. First, we computed the sample autocorrelations of residuals and present these results in Tables 15 to 18. Since none of the sample autocorrelations of those residuals are (statistically) significantly different from zero, there is no basis on which to question the adequacy of these fitted

TABLE 10. SAMPLE AUTOCORRELATIONS OF ORIGINAL VFR FLIGHT PLAN DATA, X_{1t}

lags	1- 8	0.52	0.04	-0.16	-0.22	-0.11	0.02	0.10	-0.06
	ST.E*	0.10	0.13	0.13	0.13	0.14	0.14	0.14	0.14
lags	9- 16	-0.19	-0.12	-0.18	-0.10	0.07	0.07	0.03	-0.03
	ST.E.	0.14	0.14	0.14	0.14	0.14	0.14	0.15	0.15
lags	17 - 24	-0.10	-0.14	-0.03	0.21	0.18	0.07	0.01	-0.13
	ST.E.	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15

*ST.E. = standard error of sample autocorrelations.

TABLE 11. SAMPLE AUTOCORRELATIONS OF ORIGINAL IFR FLIGHT PLAN DATA, X_{2t}

lags	1- 8	0.28	0.04	-0.08	-0.00	-0.06	0.06	0.15	-0.01
	ST. E.*	0.10	0.11	0.11	0.11	0.11	0.11	0.11	0.12
lags	9- 16	-0.13	-0.15	-0.00	0.04	0.10	0.07	0.00	-0.11
	ST. E.	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
lags	17-24	-0.01	-0.08	-0.04	-0.05	0.05	-0.02	0.02	-0.10
	ST. E.	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12

*ST.E. = standard error of sample autocorrelations.

TABLE 12. SAMPLE AUTOCORRELATIONS OF SEVENTH DIFFERENCES IN VFR DATA ($(1-B^7)X_{1t}$)

lags	1 - 8	0.59	0.24	0.06	-0.01	0.01	-0.14	-0.30	-0.24
	ST. E.	0.11	0.14	0.15	0.15	0.15	0.15	0.15	0.15
	9 - 16	-0.17	-0.62	-0.03	-0.07	-0.10	-0.09	-0.00	0.04
	ST.E.	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
	17 - 24	0.06	0.07	0.09	0.11	-0.01	-0.03	-0.00	-0.06
	ST. E.	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17

TABLE 13. SAMPLE AUTOCORRELATIONS OF SEVENTH DIFFERENCES IN IFR DATA ($(1-B^7)X_{2t}$)

lags	1- 8	0.26	0.24	0.06	0.08	-0.13	-0.04	-0.30	-0.02
	ST. E.	0.11	0.12	0.12	0.12	0.12	0.12	0.12	0.13
	9 - 16	-0.06	-0.12	0.06	0.13	0.10	-0.06	0.01	-0.10
	ST. E.	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.14
	17- 24	0.12	-0.16	-0.01	-0.09	0.02	-0.08	0.11	-0.01
	ST. E.	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14

TABLE 14. FITTED DAILY TIME SERIES MODELS FOR FLIGHT PLANS FILED AT THE LOS ANGELES FLIGHT SERVICE STATION

TIME SERIES	ARIMA MODELS TIME PERIOD = JUNE TO AUGUST 1979	TESTS OF RESIDUALS $\tilde{Q}$	D.F	ESTIMATED STANDARD ERRORS
(1) VFR FLIGHT PLAN (a)	$\underset{(0.092)}{(1-0.6003B)}(1-B^7)X_{1t}^{(0.5)}$ + $\underset{(0.092)}{(1-0.6310B^7)}e_t$	14.86	22	2.163
(b)	$(1-B)(1-B^7)X_{1t}^{(0.5)}$ = + $\underset{(0.10)}{(1-0.4808B)}\underset{(0.08)}{(1-0.755B^7)}e_t$	20.89	22	2.345
(3) IFR FLIGHT PLAN (a)	$\underset{(0.10)}{(1-0.363B)}(1-B^7)X_{2t} + \underset{(0.09)}{(1-0.6256B^7)}e_t$	13.61	22	40.85
(b)	$(1-B)(1-B^7)X_{2t} + \underset{(0.087)}{(1-0.6399B)}\underset{(0.10)}{(1-0.571B^7)}e_t$	13.96	22	43.65

TABLE 15. SAMPLE AUTOCORRELATIONS OF RESIDUALS FROM IFR FLIGHT PLAN DATA (MODEL (a))

lag	1- 8	-0.12	0.06	-0.03	0.13	-0.12	0.06	0.08	0.09
	ST. E.	0.11	0.11	0.11	0.11	0.11	0.11	0.12	0.12
lag	9- 16	-0.00	-0.04	0.11	0.08	0.04	-0.07	0.06	-0.10
	ST. E.	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
lag	17- 24	0.17	-0.02	0.04	-0.06	-0.01	-0.06	0.11	-0.03
	ST. E.	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.13

TABLE 16. SAMPLE AUTOCORRELATIONS OF RESIDUALS FROM IFR FLIGHT PLAN DATA (MODEL (b))

lag	1- 8	0.03	-0.05	-0.15	-0.01	-0.14	0.06	0.07	0.11
	ST. E.	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
lag	9 - 16	-0.03	-0.12	0.05	0.10	0.06	-0.09	0.05	-0.06
	ST. E.	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
lag	17 - 24	0.15	-0.01	0.03	-0.07	-0.06	-0.06	0.12	-0.00
	ST. E.	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.13

TABLE 17. SAMPLE AUTOCORRELATIONS OF RESIDUALS FROM VFR FLIGHT PLAN DATA (MODEL (a))

lag	1 - 8	0.02	-0.14	-0.01	-0.07	0.12	-0.03	0.10	-0.08
	ST. E.	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
lag	9- 16	-0.12	0.21	0.02	0.01	-0.01	-0.10	-0.02	0.05
	ST. E.	0.11	0.12	0.12	0.12	0.12	0.12	0.15	0.15
lag	17 - 24	0.10	0.01	-0.03	0.13	-0.10	-0.07	0.09	-0.03
	ST. E.	0.11	0.12	0.12	0.12	0.12	0.12	0.12	0.12

TABLE 18. SAMPLE AUTOCORRELATIONS OF RESIDUALS FOR VFR FLIGHT PLAN DATA (MODEL (b))

lag	1 - 8	0.16	-0.23	-0.17	-0.15	-0.00	-0.04	0.10	-0.12
	ST. E.	0.11	0.11	0.12	0.12	0.12	0.12	0.12	0.12
lag	9- 16	-0.12	0.17	0.04	-0.03	-0.06	-0.08	-0.05	0.05
	ST. E.	0.12	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Lag	17- 24	0.11	0.01	-0.02	0.09	-0.05	-0.09	0.04	-0.01
	ST. E.	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13

TABLE 19. DAILY FORECASTS OF VFR FLIGHT PLANS FILED AT LOS ANGELES FLIGHT SERVICE STATION

		Day of the Week	Lower Limits	Point Forecasts	Upper Limits	Actual Values
8-22	1979	Wed.	85.7	137.43	201.3	132.9
23		Thu.	61.4	114.7	184.5	117.9
24		Fri.	71.0	131.0	206.6	178

TABLE 20. DAILY FORECASTS OF IFR FLIGHT PLANS FILED AT LOS ANGELES FLIGHT SERVICE STATION

		Day of the Week	Lower Limits	Point Forecasts	Upper Limits	Actual Values
8-22	1979	Wed.	124.8	191.4	217.9	139
23		Thu.	139.2	188.7	238.4	194
24		Fri.	154.3	204.2	254.0	191

models. However, for further verification, the values of the overall joint λ^2 test, $\tilde{Q}$, are also calculated and presented in the last column of Table 14. These calculated values of $\tilde{Q}$ should be compared with the corresponding tabulated value of the λ^2 distribution for 22 degrees of freedom and 5 percent level of significance.

Again, all calculated $\tilde{Q}$ are less than the tabulated λ^2 value of 33.9 (at $\alpha=0.05$, and degrees of freedom=22). According to these two test results, each of the null hypotheses that these estimated residuals are distributed as white noise are not rejected. Hence, these fitted models are considered as satisfactory models for our data.

For VFR flight plans, the fitted model (a) is preferred over the fitted model (b). The reason is that the value of estimated standard errors and of $\tilde{Q}$ statistic of model (a) are uniformly smaller than those of model (b). For a similar reason, the fitted model (a) is chosen for IFR flight plan data.

A stochastic (changing) day of the week component of the VFR data can be estimated in the following way:

$$\hat{W}_{it} = (1-B^7)(1-0.613B^7)^{-1}\ X_{1t} \qquad (10)$$

and similarly, the stochastic weekly component of the IFR data can be estimated in two steps. First, we estimate

$$\tilde{W}_{2t} = (1-B^7)(1-0.625B^7)^{-1}\ X_{2t}^{(0.5)} \qquad (11)$$

The estimated component $\tilde{W}_{2t}$ is squared in order to transform it back to the original scale.

For example, we rewrite the fitted model (a) for VFR and IFR flight plans as follows:

$$X_{1,t}^{(0.5)} = 0.6003X_{1,t-1}^{(0.5)} + X_{1,t-7}^{(0.5)} - 0.6003X_{1,t-8}^{(0.5)} \qquad (12)$$

$$-0.63e_{1,t-7} + e_{1t}$$

$$X_{2,t}=0.363X_{2,t-1}+X_{2,t-7}-0.363X_{2,t-8}-0.625e_{2,t-7}+e_{2,t} \qquad (13)$$

From these two equations, we can observe the difference in the stochastic nature of IFR and VFR flight plan data. The estimated coefficient of $X_{2,t-1}$ is less than that of $X_{1,t-1}$. This implies that the value of IFR at time t is less dependent on the value at the time t-1. In other words, IFR data has relatively short memory of its past behavior in comparison with the stochastic nature of VFR data. This deduction agrees with the observed operational characteristics of aircraft on IFR and VFR flight plans. IFR flight plans are utilized primarily by business or commercial aircraft and they are not subject to most variations in weather (hence a short memory - see above). On the other hand, VFR flight plans are subject to weather and are filed primarily by private aircraft, many of whom are less likely to be on a schedule.

One-day, two-day and three-day ahead point forecasts at the time origin, August 21, 1979, are computed and summarized in Tables 19 and 20.

These models perform reasonably in forecasting because all the actual values fall within their corresponding 75 percent forecasting intervals. The large forecasting intervals are due to the large fluctuations in the daily operations data itself. Further improvement in the forecasts for daily operations should take acacount of additional factors such as local weather conditions, but this approach is more expensive than the one used here. Finally, it should be remembered that we are modelling trends, not outliers.

To sum up, the construction of an ARIMA model for daily operations data serves two purposes: (1) an estimated ARIMA model can be used to construct a linear filter, which can be used to estimate stochastic (changing) day of the week components in the daily flight plan data, and (2) the fitted model can be used to generate shortrun daily point forccasting.

5. REFERENCES

1. Box, G.E.P. and Cox, D.R. (1964), An Analysis of Transformation, Journal of Royal Statistical Society B, 221-243.
2. Box, G.P. and Jenkins, G.M. (1970). Time Series Forecasting and Control. Holden-Day Inc. San Francisco.
3. Box, G.P. and Jenkins, G.M. (1973). Some Comments on a Paper by Chatfield and Prothers and a Review by Kendall. Journal of Royal Statistical Society, A 136, 337-345.
4. Fuller, W.A. (1976). Introduction to Statistical Time Series. New York: John Wiley and Sons.
5. Ljang, G.M. and Box, G.E.P. (1978). On a Measure of Lack of Fit in Time Series Models. Biometrika 65 (2) 297-303.
6. Meyerhoff, Norman (1980). Estimation of the Instantaneous Air Count of General Aviation Aircraft Flying Under Visual Flight Rules. Paper presented at annual meeting of Transportation Research Board, Washington DC.
7. Meyerhoff, Norman (1979). Air Traffic System Measures and Data Program, U.S. Department of Transportation, Transportation Systems Center.
8. Tukey, J.W. (1977). Explanatory Data Analysis. Reading, Mass: Addison-Wesley.
9. Wang, G.H. and McCarthy, M. (1978). Seasonal Models and Forecasts for Aggregate Aircraft Operations at FAA Towered Airports. Aviation Forecasts 1979. U.S. Department of Transportation, Federal Aviation Administration.
10. Wang, George H.K. (1978). Modeling Weekly Cycles of Aircraft Operations with Daily Time Series Data, A Case Study of O'Hare Airport. Proceedings of Business & Economics Section, American Statistical Association, 655-661.

225 copies

U.S. Department
of Transportation

**Research and
Special Programs
Administration**

Kendall Square
Cambridge, Massachusetts 02142

Official Business
Penalty for Private Use $300

Postage and Fees Paid
Research and Special
Programs Administration
DOT 513

U.S. MAIL