Kim, Amy, Hansen, Mark.

Validation of runway capacity models.

AUTHOR POST PRINT VERSION

Kim, A., & Hansen, M. (2010). Validation of runway capacity models. Transportation Research Record, 2177(1), 69-77. https://doi.org/10.3141/2177-09

VALIDATION OF RUNWAY CAPACITY MODELS

Amy Kim
Department of Civil and Environmental Engineering
107 McLaughlin Hall
University of California, Berkeley
Berkeley, USA
amy_kim@berkeley.edu

Mark Hansen
Department of Civil and Environmental Engineering
106 McLaughlin Hall
University of California, Berkeley
Berkeley, USA
mhansen@ce.berkeley.edu

November 15, 2009

5,511 words + 3 figures + 3 tables = 7,011 words

ABSTRACT

There are many runway capacity estimation models currently available today, and developers usually claim that their models have been validated. However, information about the validation process is often limited, and different models are validated at differing levels of complexity. This paper proposes two validation methodologies that can be used to test model predictions against reality. We demonstrate the two methodologies on two models—the Airfield Capacity Model (ACM) and Runway Simulator (rS)—and two airports—San Francisco International (SFO) and Los Angeles International (LAX). The results indicate that when arrivals and departures are considered separately, both ACM and rS tend to over-predict capacities under good visibility conditions, and predict larger ranges of capacity values than are seen empirically. However, when considering total operations (arrivals and departures together), both models results failed to provide good estimates of total throughput at both airports. Overall, arrival and departure capacity estimates from rS typically better reflect empirical capacities than those from ACM.

1. Introduction

There are many runway capacity estimation models commercially available and in use today. These models span a wide range of types, scope, output, and capabilities. Model developers usually claim that their model has been validated, but there are several issues that arise with these validation claims. Firstly, information on the calibration and validation processes used is often vague or unclear. Secondly, model validations are performed to various standards and differing levels of complexity. Finally, the main validation exercises were often carried out by the developers themselves as opposed to an unaffiliated third party.

In this paper, we attempt to address the three issues mentioned above. We propose and demonstrate two validation methods that can be used to compare the estimates from a runway capacity model against empirical counts of arrival and departure throughput. These two methods account for the fact that capacity may not be directly observable, since it represents an upper limit rather than the actual number of operations. These validation methods are demonstrated on results from two models—the Airfield Capacity Model (ACM) and Runway Simulator (rS). This paper will provide a description of the models and the validation methodology, describe the data used, present validation results and suggested directions for future work.

2. BACKGROUND

2.1 Defining Capacity

The definition of runway capacity is a topic that has been discussed and debated extensively for many years. The ACM defines runway capacity to be the average maximum sustainable throughput *(1)*; capacity estimates from rS are based on the same definition. Throughput (or count; these words will be used interchangeably in this paper) refers to the number of aircraft that use the runway system at a given airport over a given unit of time. Sustainability is the idea that the airport can maintain this average throughput for long periods of time under sufficient demands *(2)*. However, because the main factors that affect throughput change over time, throughput will also vary accordingly.

This analysis attempts to validate model capacity estimates against empirical throughput data. To choose the appropriate throughput data, we filter it through carefully defined criteria that help ensure we are choosing the data that best reflects capacity. The criteria that we use, as well as the entire data selection process, is discussed in Section 4.1.

2.2 Factors Affecting Runway Capacity

There are many phenomena that can affect the number of aircraft able to land and depart at an airport. Those that typically have the greatest effects are listed below *(3)*:

- Weather characteristics (visibility, cloud ceiling, precipitation, location and duration of adverse weather front, position of sun, etc.), and subsequent meteorological condition designation;
- Air traffic control separation requirements;
- State and performance of the Air Traffic Management (ATM) system;
- Number of runways in use and their geometric layout;
- Layout of the airfield (including all components such as location of gates and taxiways);
- Aircraft fleet mix and performance characteristics;
- Mix and sequencing of arrival and departure aircraft;
- Runway occupancy times;
- Overall arrival/departure split;

- Airline policies regarding landing, takeoff, taxiing;
- Pilot familiarity with airport, experience and skill, and
- Controller environment, workload, experience and skill.

Depending on their purpose, design and type, capacity models attempt to account for some combination of the above factors. Many factors can be explicitly included in a model's inputs, although simpler models will account for less factors and each to a more limited degree. Some factors, namely the last three, are difficult to account for as they may be subjective measures. In addition information may not be available, or if it is, too labor intensive to collect and include. However, there have been efforts to include controller workload factors.

2.3 Description of Capacity Models

2.3.1 Overview

The runway capacity models commercially available today span a wide range of scopes, capabilities, and complexities *(4)*. Models can be categorized in several ways; here we categorize them by three important aspects: calculation method, stochastic capability, and model scope. The first two are independent of one another; however, they serve to isolate key differences between models.

Runway capacity models calculate capacity analytically or through simulation. Analytical models are mathematical representations of operations, and can be implemented using a calculator or spreadsheet. They rely on a set of key inputs and variables to quickly, simply, and efficiently estimate the average behavior of entities (in this case, aircraft). Simulation models attempt to characterize changing conditions over time. They can be further categorized as macroscopic, mesoscopic, and microscopic. Macroscopic models are like analytical models in that they rely on key inputs to represent the average behavior of entities over time. However, they are updated with changing information in discrete time steps. In microscopic simulation, aircraft (for instance) are represented individually, and the model creates and records their interactions with one another and their environment. Microscopic models tend to be more comprehensive in accounting for more factors that affect capacity. Mesoscopic models combine elements of both macro and microscopic models. All the above models can be placed on a sliding scale of computational complexity, from simple (analytical) to highly complex (microscopic simulation).

Models can be deterministic, or stochastic to varying degrees. A model's level of stochasticity depends on how many parameters (and which ones) are treated as random variables.

Lastly, runway capacity models' scopes can range from being able to represent aircraft operations on runways only to aircraft operations at gates, on aprons, taxiways, and in airspace. Very sophisticated models can incorporate numerous complex factors and operations that affect capacity, even beyond those listed in the previous section.

2.3.2 Airfield Capacity Model (ACM)

The ACM was initially developed by a consortium in the late 1970s and then modified by the FAA and MITRE CAASD, with the last modification made in 1981. It is an analytic model that calculates the hourly capacity of runway systems given continuous demand *(5,6)*. It asks the user for basic operating and geometric characteristics, which it then converts to numerical inputs for its calculations. The ACM can estimate capacities for 15 simple runway configurations, from a single runway to 4 runways in varying configurations. The model's default assumption is that there is a 5% probability of violating separation standards, and this is used to determine aircraft spacing on runways.

The ACM was validated in the early 1980s by the FAA; capacity estimates for certain runway configurations were deemed to be reasonably accurate. We were not able to obtain more information on the validation work. The ACM is mainly used by the FAA and their consultants *(6)*.

2.3.3 Runway Simulator (rS)

rS was developed by MITRE CAASD, and is an intermediate effort between a simple analytical model and a complex discrete event simulation model. rS simulates individual aircraft movements on runways and airspace in the immediate vicinity of the airport, under continuous demand. Like many simulation models, rS is based on “blocking” rules, meaning that it is built on a link-node system where each link can only hold a pre-specified maximum number of aircraft at any given time. It is a dynamic model that incorporates some stochasticity in its inputs, including runway occupancy time separation buffers, arrivals, etc. rS is capable of estimating both capacity and delay (which requires input of a schedule). rS requires a basic set of operational inputs (not very different from ACM) although it does require more physical parameter inputs. Users can set up an analysis in rS relatively quickly in comparison to other more complex simulation models.

rS was validated by MITRE by comparing capacity results from rS to those of ACM for a number of simple scenarios *(2)*. As basic calculations are found to be correct, they were assumed to remain so for more complex scenarios. In addition, the animation can be viewed to insure that all ATC rules specified are followed correctly. The program is mainly used for in-house studies, although the Federal Aviation Administration (FAA) has begun using it as well.

3. DATA

The data for this analysis was obtained from the Aviation System Performance Metrics (ASPM) database, which is part of the FAA’s Operations and Performance Data system. Data from the “Download/Airport” section of the ASPM database was used in particular. This data includes hourly and quarter-hourly arrival and departure counts, demands, various weather conditions, and visibility conditions (either visual (VFR) or instrument (IFR) flight rules). The data does not contain individual flight information. The counts are based on individual aircraft landing and take-off times as supplied through Airline Service Quality Performance (ASQP) data or Enhanced Traffic Management System (ETMS) messages. The data is available for 77 major airports in the United States.

To understand our methodology and results, it is necessary to understand the demand data in our data set. Conceptually, demand is the number of flights that “want” to perform an arrival or departure movement within a particular time period. It is based on the flight plan that is filed just before a flight takes off at the origin airport. In most cases, a flight counts toward demand beginning in the time period it is filed to land or take-off, continuing through all time periods until it actually does so. The only exception is when a flight arrives or takes off in a time period earlier than planned, in which case it is counted toward the demand in this earlier time period. This procedure ensures that the throughput never exceeds the demand. When the throughput and demand are equal, all flights are able to make their desired movement; none are forced to wait until the next time period to perform their desired operation. However, when demand exceeds throughput, flights are delayed. A shortcoming of this method for determining demand is that demand is not updated based on delays that are incurred 1) between the time the flight plan is filed and the aircraft is taxiing for take-off (departure demand) or 2) en route to the destination airport (arrival demand). The implication is that the actual airfield demand may in reality be lower than the ASPM demand data reports. This can lead to incorrectly attributing a difference between count and demand to a capacity constraint. This is not taken into consideration in the ensuing analysis, but has been done so previously by Hansen *(7)*.

Quarter-hour and individual flight data from 2006 was obtained for both SFO and LAX. However, the runway configurations identified in the LAX ASPM data were found to be incorrect, so it was replaced with runway configuration data from the Performance Data Analysis and Reporting System (PDARS). PDARS is a joint NASA-FAA effort developed by ATAC Corporation. The database is fed by radar track and flight plan information directly from Automatic Radar Terminal System (ARTS) computers at Terminal Radar Approach Control (TRACON) facilities, and from the host computers at Air Route Traffic Control Centers (ARTCCs), which provide precise state information for each aircraft every 2 seconds. As

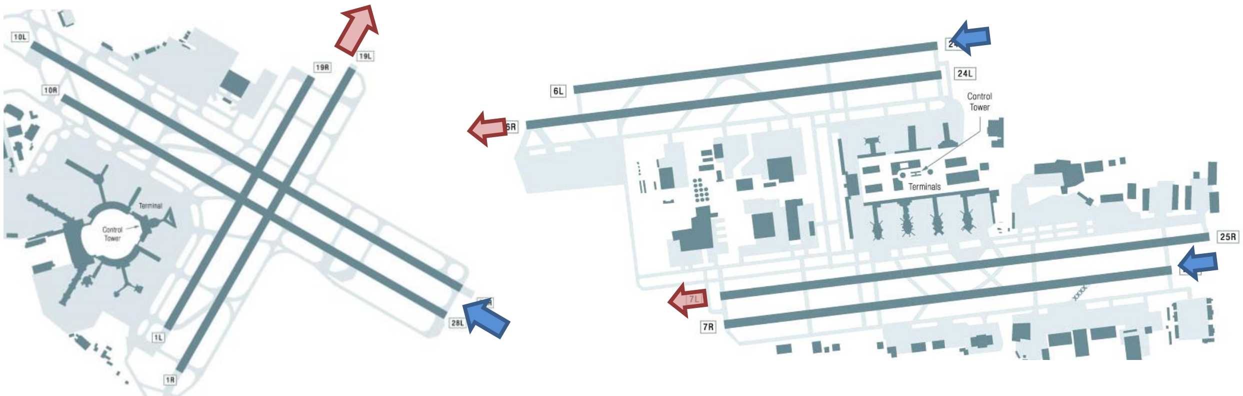
PDARS data was readily available for January through March 2005, ASPM data for the corresponding time period was used instead of 2006 data.

4. METHODOLOGY

4.1 Experimental Procedure

Several steps were taken to perform the validation exercise. The first step involves choosing the hours to be analyzed, by grouping the quarter-hourly ASPM data into hourly bundles starting on the hour. The purpose, again, is to maximize our likelihood of obtaining data that best reflects capacity-constrained conditions. Complete candidate hours for analysis were identified by filtering the hours through several criteria:

- The predominant runway configuration was in use for the entire hour (28L,28R | 1L,1R at SFO, 24R,25L | 24L,25R at LAX), see Figure 1 *(8)*;
- The weather designation was VMC or IMC for all quarter-hour periods of the total hour, and
- The hour falls within the period of the day with the highest average demands (which, based on the data, was found to be between 9 am and 2 pm at both airports).



(a) SFO **(b) LAX**

Source: *Airport Capacity Benchmark Report 2004*. Federal Aviation Administration *(8)*.

FIGURE 1 Layout of Airport Runways.

After filtering, the next step involves randomly drawing 50 hours (approximately half VMC and half IMC) from each filtered set. After filtering, only 20 IMC hours were available for analysis at both SFO and LAX, but about 30 VMC hours were available. A sample of 50 hours was deemed sufficient for this analysis, although it would be preferable to have more in future studies if possible.

The next step is to obtain capacity estimates from both ACM and rS for each of the 50 hours. Each hour is distinguished in the data by meteorological condition, fleet mix, and arrival/departure split (%). Runway configurations are held fixed at each airport's predominant configuration. As the purpose of this work was to assess model performance using minimal to no calibration, no additional edits were made after the above features were input into the models.

The resulting set contains predicted ACM and rS capacities, observed counts, and other information (from the data) for each of 50 hours at LAX, and likewise for SFO. This data serves as the basis for our two validation methods, which we now discuss.

4.2 Comparison of Predicted Capacities with Demand-Unconstrained Counts

The first set of validation metrics are those developed by Theil in the 1960s for comparing predicted values to realizations *(9)*. The method is based on a simple comparison of the realized counts and the capacities predicted by the models. Recognizing that counts may be reflecting demand constraints rather than capacity constraints, we selected observations (from our filtered data set of 50) where demand exceeded the capacity for that time interval. Since our data set consists of hours from the busy period of the day, this turned out to be true for the majority of our observations. In addition to plotting demand-unconstrained counts against capacity, we calculated the Theil inequality coefficient and its components for each model. The inequality coefficient is a measure of the seriousness of a prediction error.

$$U = \frac{\sum_i (P_i - A_i)^2}{\sum_i A_i^2} \quad (1)$$

Where

U is the inequality coefficient for operation type o (arrivals, departures, or both combined),

P_i is the predicted value for o in observation i, and

A_i is the realized value for o in observation i.

If a given model yielded capacity predictions that perfectly fit the empirical evidence, we would expect that U approaches *0*. The inequality coefficient may be decomposed into three parts that isolate the differences between predicted and actual values: bias or error in central tendency, U_m; unequal variation, U_s; and incomplete covariation, U_c. These components are normalized so that they sum to 1; as such they indicate what proportion of the total prediction error can be attributed to each of these three effects. The inequality proportions are given by:

$$U_m = \frac{(\bar{P} - \bar{A})^2}{\frac{1}{n}\sum (P_i - A_i)^2} \quad (2)$$

$$U_s = \frac{(s_P - s_A)^2}{\frac{1}{n}\sum (P_i - A_i)^2} \quad (3)$$

$$U_c = \frac{2 \cdot (1-r) \cdot s_P \cdot s_A}{\frac{1}{n}\sum (P_i - A_i)^2} \quad (4)$$

In these expressions s_P and s_A are the standard deviations of the predicted and actual values, r is the correlation coefficient between P and A, and n is the sample size.

4.3 Censored Regression Model

We also used censored regression to evaluate the two models. A censored regression model is equivalent to an ordinary least squares (OLS) regression model in that it relates a dependent random variable Y to a set of independent variables $X_1, X_2, \ldots, X_n$ *(10)*. However, in censored regression it is assumed that Y cannot be observed beyond some minimum or maximum threshold value (or both). For instance, if a value of Y is larger than the maximum threshold value, Y_{max} then only Y_{max}, is observed. The true value of Y - the latent variable Y^* - cannot always be observed due to this censoring effect, although $X_1, X_2, \ldots, X_n$ are always observable. Tobit regression accounts for this by ensuring that the regression model parameters estimate the effects of $X_1, X_2, \ldots, X_n$ on the latent variable Y^* and not on the censored (observed) variable Y.

In this analysis the dependent observed variable Y is the airport's arrival or departure throughput Q, while the latent variable Y^* is the airport's arrival or departure capacity C. When aircraft demand exceeds count in a given time period, capacity can be equated to the count. However, counts cannot exceed the demand in any given time period, although the airport's capacity may in fact be greater than the demand in that time period. As a result we are limited in our ability to use counts to measure capacity. The observed throughput in time period t is censored from above, and there are two situations that can arise *(7)*.

$$Q_o(t) = \begin{cases} C_o(t), & if \quad 0 < C_o(t) < D_o(t) \\ D_o(t), & if \quad C_o(t) \ge D_o(t) \end{cases} \quad (5)$$

Where

$Q_o(t)$ is the throughput, or the "observed" capacity for operation type o in time interval t $(=Y)$,

$C_o(t)$ is the true (or latent) capacity for operation type o in time t $(=Y^*)$, and

$D_o(t)$ is the demand (or upper bound of observable capacity) for operation type o in time t $(=Y_{max})$.

In the first scenario described by Equation 5 counts are less than the demand; in this case capacity can be equated to the count. The second scenario is the upper censor where counts equal demand, and therefore capacity is measured to be this demand (although it could in reality be higher, therefore the censoring effect).

The basic Tobit regression model specification is introduced here.

$$C_o(t) = \beta_0 + \beta_1 \cdot Mod_{x,o}(t) + \varepsilon \quad (6)$$

$$Q_o(t) = \min[D_o(t), C_o(t)] \quad (7)$$

Where

β_0, β_1, and σ_0^2 are estimated parameters,

$Mod_{x,o}(t)$ is the capacity estimate from model x (ACM or rS), for operation o in t, and

ε is the iid error term, distributed Normal with mean 0 and variance σ_0^2.

The model parameters are estimated from the data using maximum likelihood estimation (MLE).

If a given model yielded perfect capacity predictions, we would expect $\beta_0 \rightarrow 0$, $\beta_1 \rightarrow 1$, and $\sigma_0^2 \rightarrow 0$. Thus the coefficients yielded by estimation provide a basis for scoring the validity of the models. The ACM and rS model estimates were obtained and then used as explanatory variables in the regression model. To avoid confusion, the capacity regression model will be referred to as the empirical or regression model, while the ACM and rS models will be called the test models (if not referred to by their names).

4.4 Results

4.4.1 Predicted Capacities versus Unconstrained Counts

Figs. 2 and 3 compare predicted capacities and realized counts for each model and each airport. For arrival and departure counts, in the case of SFO it appears that the rS model results yield better agreement with observed values (Fig. 2). At LAX, aside from the rS results in IMC, neither model does very well for arrivals or departures (Fig. 3). The VMC estimated capacities appear to be greater than the actual capacities in all cases except rS at SFO. With regard to total operations, ACM consistently overestimates VMC capacities, while rS does so for LAX. Neither model predicts capacity variation within an airport or visibility condition very well.

Table 1 presents the Theil analysis results. At SFO, the inequality coefficients for the rS capacity estimates are much less than those of the ACM model estimates, confirming rS' better predictive capability. The primary sources of inequality are also different, with bias (U_m) the major source in the ACM model compared to incomplete covariation (U_c) in the rS model. However, as Table 1 displays the proportions rather than the actual error values, the results might lead us to believe that rS' error due to covariance was as great as the ACM bias. It was in fact observed that the ACM bias was very high whereas the rS covariance error was typically lower. In the case of LAX, the inequality coefficients for the two models are comparable, as are the inequality proportions. In general, unequal variation (U_s) is the smallest contributor to the inequality of the predicted and actual data sets. Aggregating across the two airports, the rS model emerges as the better predictor, primarily because it exhibits less bias.

When the results from both airports are combined in one analysis, it can again be seen that the rS results are better than the ACM results. As seen in Figures 2 and 3, the VMC capacity estimates are much higher than the observed throughput, and this is reflect in the fact that bias (U_m) is the major source of inequality in both ACM and rS' VMC results (this is also true for the individual airports results above). The main source of inequality for IMC observations is incomplete covariation (U_c).

Arrival and departure (and subsequently, total) capacity predictive performance appears to be highly correlated. Inequality coefficients for arrivals and departures are generally of very similar magnitudes, as are the inequality proportions.

4.4.2 Regression Model I

The results of the basic model (Equations 6 & 7) for ACM and rS are reported in Table 2. The first section of the table contains results for SFO, the middle section for LAX, and the bottom for the combined observations from both airports. Recall that the empirical model results are based on about 30 VMC and 20 IMC observations at each airport.

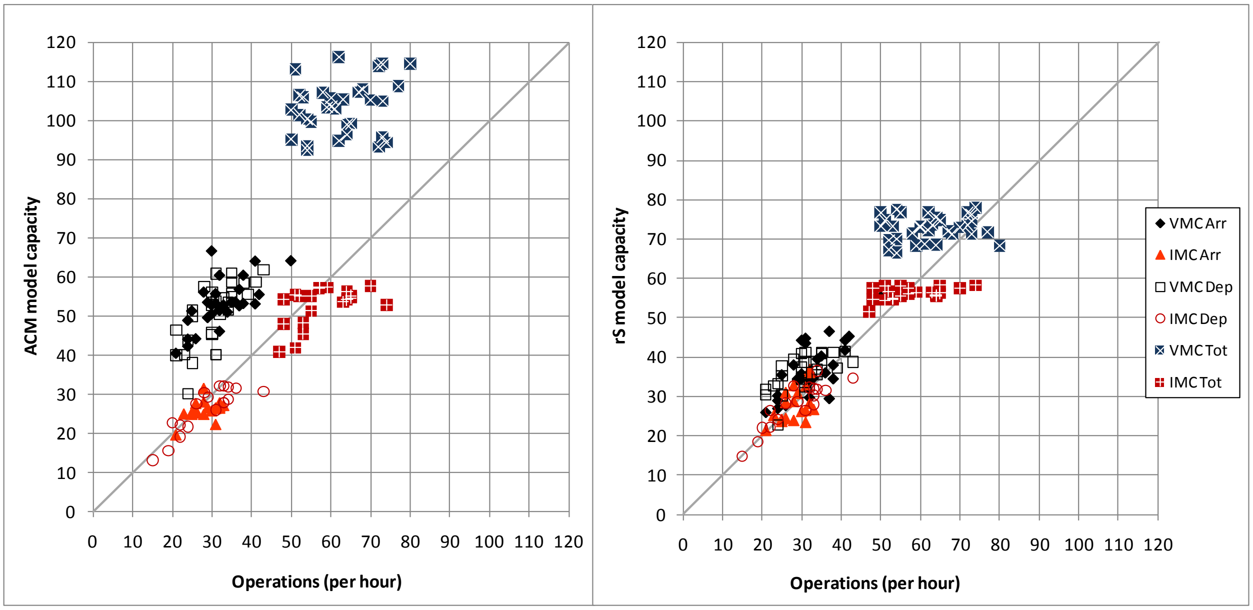


FIGURE 2 Model Capacity Estimates versus Unconstrained Counts, SFO.

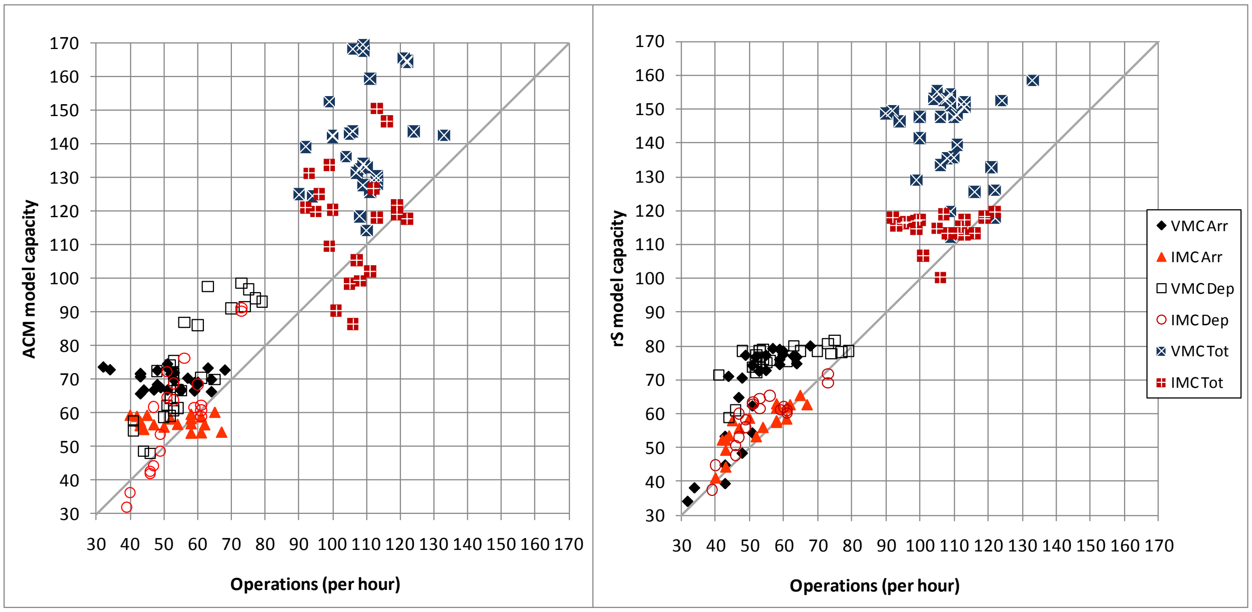


FIGURE 3 Model Capacity Estimates versus Unconstrained Counts, LAX.

TABLE 1 Prediction-Realization Analysis

SFO		RMS error			Inequality Coefficient (U)			Inequality Proportions								
								Arrival			Departure			Total		
		Arr	Dep	Tot	Arr	Dep	Tot	U_m	U_s	U_c	U_m	U_s	U_c	U_m	U_s	U_c
ACM	VMC	21.3	20.6	42.0	0.65	0.66	0.66	0.94	0.00	0.06	0.92	0.01	0.06	0.95	0.00	0.05
	IMC	3.9	4.3	8.2	0.14	0.15	0.14	0.36	0.06	0.58	0.32	0.09	0.59	0.37	0.10	0.54
ACM Total		16.8	16.4	33.1	0.54	0.54	0.54	0.48	0.25	0.27	0.47	0.22	0.31	0.48	0.27	0.26
rS	VMC	6.6	6.9	13.5	0.20	0.22	0.21	0.46	0.00	0.54	0.58	0.04	0.38	0.56	0.17	0.27
	IMC	3.8	3.4	7.0	0.13	0.12	0.12	0.01	0.06	0.93	0.01	0.06	0.93	0.04	0.71	0.26
rS Total		5.6	5.8	11.4	0.18	0.19	0.19	0.21	0.08	0.71	0.26	0.00	0.74	0.24	0.00	0.76
LAX		RMS error			Inequality Coefficient (U)			Inequality Proportions								
								Arrival			Departure			Total		
		Arr	Dep	Tot	Arr	Dep	Tot	U_m	U_s	U_c	U_m	U_s	U_c	U_m	U_s	U_c
ACM	VMC	19.9	17.8	37.2	0.38	0.30	0.34	0.76	0.09	0.15	0.80	0.06	0.14	0.80	0.04	0.16
	IMC	10.4	11.1	21.4	0.20	0.20	0.20	0.24	0.41	0.35	0.28	0.40	0.32	0.26	0.13	0.61
ACM Total		16.8	15.4	31.9	0.32	0.27	0.29	0.56	0.02	0.43	0.59	0.19	0.22	0.59	0.13	0.28
rS	VMC	16.8	19.6	36.3	0.32	0.33	0.33	0.74	0.10	0.15	0.81	0.07	0.11	0.80	0.01	0.19
	IMC	6.0	6.4	12.5	0.11	0.12	0.12	0.37	0.14	0.48	0.40	0.02	0.58	0.41	0.13	0.46
rS Total		13.6	15.6	29.3	0.26	0.27	0.27	0.56	0.10	0.34	0.60	0.00	0.40	0.60	0.07	0.33
BOTH Airports		RMS error			Inequality Coefficient (U)			Inequality Proportions								
								Arrival			Departure			Total		
		Arr	Dep	Tot	Arr	Dep	Tot	U_m	U_s	U_c	U_m	U_s	U_c	U_m	U_s	U_c
ACM	VMC	20.7	19.3	39.7	0.47	0.42	0.45	0.86	0.02	0.12	0.86	0.00	0.14	0.88	0.00	0.12
	IMC	7.8	8.4	16.1	0.18	0.19	0.19	0.03	0.08	0.89	0.04	0.46	0.50	0.03	0.30	0.67
ACM Total		16.8	15.9	32.5	0.39	0.35	0.37	0.51	0.03	0.46	0.52	0.09	0.38	0.53	0.07	0.40
rS	VMC	12.6	14.4	27.1	0.29	0.31	0.30	0.55	0.23	0.23	0.60	0.09	0.31	0.60	0.16	0.24
	IMC	4.9	5.1	10.1	0.12	0.12	0.12	0.11	0.07	0.82	0.11	0.10	0.80	0.10	0.12	0.78
rS Total		10.3	11.6	22.0	0.24	0.26	0.25	0.37	0.22	0.41	0.40	0.13	0.47	0.40	0.19	0.41

TABLE 2 Model I Results

SFO		Departure			Arrival			Total (Dep & Arr)		
		Estimate	*Error*	*t-stat*	*Estimate*	*Error*	*t-stat*	*Estimate*	*Error*	*t-stat*
ACM	β_o	21.0	2.604	8.04	22.2	2.004	11.06	50.8	3.585	14.16
	β_1	0.2	0.055	4.22	0.2	0.052	4.03	0.1	0.043	2.79
	σ_o	1.7	0.1	17.12	1.6	0.105	15.08	2.1	0.073	28.61
rS	β_o	7.8	2.579	3.02	12.4	3.023	4.1	39.2	8.329	4.70
	β_1	0.7	0.08	8.61	0.6	0.093	5.98	0.3	0.127	2.56
	σ_o	1.6	0.096	16.2	1.4	0.078	17.8	2.1	0.077	27.23
LAX		**Departure**			**Arrival**			**Total (Dep & Arr)**		
		Estimate	*Error*	*t-stat*	*Estimate*	*Error*	*t-stat*	*Estimate*	*Error*	*t-stat*
ACM	β_o	22.7	2.712	8.35	60.7	13.709	4.43	94.6	5.759	16.42
	β_1	0.5	0.043	11.58	-0.1	0.212	-0.56*	0.1	0.044	2.40
	σ_o	1.7	0.082	20.52	2.2	0.082	26.66	2.2	0.099	21.94
rS	β_o	21.0	5.565	3.77	22.3	3.558	6.27	102.6	10.318	9.95
	β_1	0.5	0.09	5.82	0.5	0.057	8.53	0.05	0.079	0.57*
	σ_o	2.1	0.085	24.84	1.8	0.091	20.1	2.2	0.098	22.50
BOTH Airports		**Departure**			**Arrival**			**Total (Dep & Arr)**		
		Estimate	*Error*	*t-stat*	*Estimate*	*Error*	*t-stat*	*Estimate*	*Error*	*t-stat*
ACM	β_o	8.6	1.933	4.44	11.7	2.247	5.22	23.4	3.390	6.91
	β_1	0.6	0.031	20.48	0.6	0.046	12.30	0.6	0.028	20.37
	σ_o	2.1	0.056	38.24	2.3	0.067	33.84	2.8	0.058	49.07
rS	β_o	7.9	1.714	4.59	10.0	1.358	7.40	20.8	2.873	7.23
	β_1	0.7	0.040	17.69	0.7	0.032	20.82	0.6	0.031	20.54
	σ_o	1.9	0.070	27.53	1.7	0.074	23.29	2.5	0.061	40.29

** Results are not significant at the 95% confidence level.*

We first discuss the regression results for departure and arrival operations observed separately. All β_0 estimates are much greater than 0, and all β_1 estimates are smaller than 1. The magnitudes of the β_1 estimates indicate that the ranges of capacity estimates from ACM and rS are larger than the corresponding ranges of empirical throughputs. The β_0 estimates are relatively large for the same reason. These results imply that when predicted capacities are low, actual capacities are likely to exceed the predictions. However, when predicted capacities are high, the actual capacities are likely to be lower. At SFO, the rS regressions' β_0 estimates are lower and β_1 estimates are higher than those of the ACM regressions, suggesting that rS performs better in predicting actual throughputs for a given configuration. The standard deviations in the ACM regression model are also higher than those of the rS regression model. At LAX it appears that the parameter estimates from the ACM and rS regressions are comparable, except those of the ACM arrivals regression. The ACM arrival capacity predictions proved very insensitive to the different input conditions of the 50 hourly samples in IMC and VMC. As a result, the β_1 estimate is insignificant and β_0 simple reflects the average empirical capacity for the entire 50 hour sample.

For total operations (arrivals and departure combined), it appears that the variation in total throughput at each airport is not captured by the model. β_0 values are extremely high while β_1 values are very small and in one case insignificant, indicating that the model results give us little to no information about actual throughput. These results are supported by Figures 2 and 3.

For the combined airports models (last section of Table 2), the β_0 values tend to be smaller and the β_1 values larger than those of the other two models. Moreover, the rS and ACM estimated parameters are quite similar in the combined model, implying that the models do fairly well in predicting the difference in capacity between SFO and LAX in their primary configurations.

The results of Tables 1 and 2, in addition to Figures 2 and 3, suggest that this basic model (Equations 6 and 7) may not be the most appropriate model to use, in assessing total capacity estimates. The results lead to some combination of three conclusions. Firstly, the capacity models are doing a poor job of using the inputs (the various geometric, weather, and operational conditions) to form their estimates. Secondly, the inputs require some adjustment based on better or more detailed field knowledge. Lastly, the phenomena causing the throughput variations observed in the data are not captured by the model inputs. These phenomena could include any combinations of those listed in Section 2.2 including aircraft mix, sequencing of arrival and departure aircraft on the runway, visual obscuration effects that encourage instrument landings in VFR conditions, pilot experience, and air traffic controller workload and culture. These phenomena could have significant effects on capacity such that it may not be reasonable to include them in the error term of Equation 6. This can be investigated as part of a sensitivity analysis of results in future work.

Table 2 has shown that both models' capacity estimates are fairly good on average across two airports, whereas they are not as good for specific airports.

4.4.3 Regression Model II

The basic regression model was modified to include another parameter that distinguishes between VMC and IMC test model capacity estimates. This serves to further isolate the effect of visibility condition on the test models' predictive performance.

$$C_o(t) = \beta_0 + \beta_1 \cdot Mod_{x,o}(t) + \beta_2 \cdot I_0(VMC = 1) + \varepsilon \tag{8}$$

$$Q_o(t) = \min[D_o(t), C_o(t)] \tag{9}$$

Where

β_2 is an estimated parameter, and

$I_o(VMC=1)$ is an indicator variable set to 1 if operation type *o* in time *t* occurs under VMC conditions, and 0 if it occurs in IMC.

The results of the model in Equations 8 and 9 are contained in Table 3. Overall, the inclusion of the VMC indicator variable improves the estimates for β_0 and β_1. β_2 is an estimate of the adjustment that must be made to VMC test model capacity estimates to reflect actual capacities. With its addition, β_0 decreases in most of the regressions, while β_1 increases towards 1. This implies that the test models can more accurately predict variations in capacity given a particular visibility condition, rather than the effect that visibility condition has on capacity.

The β_2 values in Table 3 are each less than zero (except that of the ACM arrivals regression at LAX, which is subject to the same problems that were discussed in the previous section). This implies that both models overestimate the difference between VMC and IMC capacities. Figs. 2 and 3 have shown that this occurs because both models tend to overestimate VMC results, ACM more so than rS.

TABLE 3 Model II Results

SFO		Departure			Arrival			Total (Dep & Arr)		
		Estimate	***Error***	***t-stat***	***Estimate***	***Error***	***t-stat***	***Estimate***	***Error***	***t-stat***
ACM	β_o	11.9	2.692	4.41	12.9	4.022	3.22	38.0	10.362	3.67
	β_1	0.7	0.096	6.87	0.6	0.152	4	0.4	0.197	1.89*
	β_2	-14.2	2.778	-5.12	-12.3	3.78	-3.24	-13.6	10.297	-1.32*
	σ_o	1.5	0.092	15.72	1.4	0.105	13.77	2.1	0.078	26.37
rS	β_o	1.4	3.255	0.42*	11.4	3.172	3.58	36.6	27.618	1.32*
	β_1	1.0	0.121	8.2	0.6	0.109	5.64	0.4	0.493	0.76*
	β_2	-6.2	1.695	1.69*	-1.5	1.319	-1.15*	-0.9	8.610	-0.10*
	σ_o	1.4	0.082	0.08*	1.4	0.081	16.85	2.1	0.079	26.58
LAX		**Departure**			**Arrival**			**Total (Dep & Arr)**		
		Estimate	***Error***	***t-stat***	***Estimate***	***Error***	***t-stat***	***Estimate***	***Error***	***t-stat***
ACM	β_o	22.0	2.479	8.8	95.4	32.732	2.92	94.9	5.826	16.29
	β_1	0.5	0.043	12.64	-0.7	0.572	-1.29*	0.1	0.052	1.94*
	β_2	-3.6	1.559	-2.28	8.9	7.176	1.24*	0.3	3.163	0.09*
	σ_o	1.6	0.087	18.86	2.2	0.08	27.07	2.2	0.099	21.96
rS	β_o	-5.1	6.61	-0.77*	19.6	3.331	5.89	113.6	14.524	7.82
	β_1	1.0	0.118	8.69	0.6	0.056	10.59	-0.06	0.128	-0.46*
	β_2	-13.7	2.634	-5.12	-6.7	1.747	-3.83	4.4	4.168	1.07*
	σ_o	2.0	0.099	19.78	1.7	0.077	22.05	2.2	0.102	21.51
BOTH Airports		**Departure**			**Arrival**			**Total (Dep & Arr)**		
		Estimate	***Error***	***t-stat***	***Estimate***	***Error***	***t-stat***	***Estimate***	***Error***	***t-stat***
ACM	β_o	8.6	1.488	5.80	5.8	2.065	2.82	17.0	3.059	5.57
	β_1	0.8	0.036	21.73	0.8	0.059	14.24	0.8	0.042	18.24
	β_2	-11.9	1.580	-7.53	-14.7	2.064	-7.12	-25.4	3.492	-7.26
	σ_o	1.9	0.061	30.81	2.0	0.075	27.31	2.6	0.061	42.33
rS	β_o	9.4	1.695	5.55	11.4	1.320	8.61	23.0	2.830	8.13
	β_1	0.8	0.037	20.12	0.7	0.031	22.88	0.7	0.030	22.92
	β_2	-6.6	1.140	-5.78	-5.0	1.110	-4.47	-11.1	2.081	-5.31
	σ_o	1.8	0.076	23.97	1.6	0.062	26.17	2.3	0.065	36.26

** Results are not significant at the 95% confidence level.*

ACM appears to overestimate VMC capacities at SFO but underestimate the lower IMC capacities. The rS model appears to do a better job of estimating capacities than ACM, although again it seems to slightly overestimate VMC capacities.

At LAX it again appears that capacity estimates from rS more closely reflect empirical values than do the ACM results. The complete failure of ACM to predict arrival (and subsequently, total arrival + departure) throughput is apparent. For departures ACM overestimates the difference between VMC and IMC capacities and, as in the previous results, exaggerates capacity variability. rS greatly exaggerates the differences in capacity between IMC and VMC at LAX. However, asides from this rS does well in predicting capacity variations for arrivals and departures, as implied by the β_1 coefficient matching the ideal value of 1.

It is evident that both models completely fail to predict total (arrival + departure) throughput, based on the coefficient values and the fact that many values are insignificant at the 95% confidence level. This implies that much of the variability in arrival and departure counts that the models were able to predict arises from differences in arrival-departure mix.

From the combined airports regression (bottom section of Table 3), the performance of ACM and rS are more comparable for arrivals, departures, and total operations. Overall, the rS model does a better job in predicting the capacity difference between VMC and IMC (demonstrated by β_2 estimates that are closer to zero). The models do equally well in predicting capacity variation from other sources, based on the β_1 results. There remains a tendency for the models to exaggerate capacity variation compared to what is actually observed. These results, like those of Table 2, are greatly influenced by the difference in capacity between SFO and LAX, and it is the ability of ACM to accurately predict that difference that makes it appear competitive with rS.

4.4.4. Model Assumptions and Limitations

As discussed previously there are many factors that affect capacity, which were not included as inputs to the two models tested here. Some of the differences between model predictions and empirical values may be attributed to the exclusion of these factors in the regression model. For instance, pilots can choose an instrument landing even during VFR conditions for various reasons such as visual obscuration by haze and unfamiliarity with an airport (the chances of which are higher with international flights, which both SFO and LAX have a large number of). If data on this kind of occurrence could be obtained, regression model II could be modified to account for it.

We have also seen that the test models generally appear to overestimate VMC capacities and underestimate IMC capacities. Despite careful filtering, we cannot guarantee that our count data always reflects maximum throughput conditions, which might explain the overestimation. The underestimation might be caused by the fact that although the airport is designated to be operating under IMC conditions, actual operating conditions might have been somewhat better. Both discrepancies might also be attributed to the fact that weather designation for a 15-minute period reflects conditions at the beginning of the time period, and conditions may have changed sometime thereafter.

ACM and rS also may not be using some inputs in ways that best reflect the way they actually impact operations. This can be tested by including other characteristics as independent variables in the regression model. For instance, the direction and speed of wind is the dominant factor in choosing a runway configuration. A primary configuration at each airport is favored because it allows for maximum aircraft operations. Somewhat unfavorable winds (not unfavorable enough, however, to switch configuration) can certainly be a factor in decreasing the rate of operations at each airport. This was further investigated by including wind speeds and directions in regression model II in various ways; however, this did not yield significant results.

6. CONCLUSIONS & FUTURE WORK

This paper has introduced two methodologies for validating capacity model results against empirical data. The validation results indicate that neither ACM nor rS predict realized throughputs with great accuracy, although rS does somewhat better. The models do best at estimating capacity differences between different runway layouts, and resulting from variations in the mix of arrivals and departures. They both appear to overestimate VMC capacities, although this might be occurring because the count data does not reflect maximum throughput conditions. Other, more minor, capacity differences arising from factors such as fleet mix are not successfully predicted by either model.

The work discussed in this paper can be continued and improve upon in several directions. It would be of interest to test other capacity models, particularly one of the more complex microscopic simulation models that are often used today. Capacities for additional runway configurations at LAX and SFO, as well as configurations at other busy airports, could be estimated and the regression model re-specified to include these. Arrival and departure interaction effects would be worthwhile to investigate. Also, the capacity model estimation results could be compared against empirical capacity estimates based on other data sources such as PDARS.

It would be particularly worthwhile to continue the work of Section 4.4.4 by performing sensitivity tests that investigate which other characteristics may have had a significant effect on capacity. The results of the sensitivity tests could then be used to improve regression model specification. The sensitivity tests would be very worthwhile if additional data could be obtained for both SFO and LAX.

ACKNOWLEDGMENT

The authors would like to thank Joe Post at the FAA for his support of this research study.

REFERENCES

(1) Swedish, William J. *Upgraded FAA Airfield Capacity Model Supplemental User's Guide*. The Mitre Corporation, McLean, Virginia, February 1981.

(2) Barrer, John, and Peter Kuzminski. *Airport Capacity Analysis Through Simulation (ACATS): User's Guide*. Technical Report, Mitre Corporation, 2006.

(3) DeNeufville, Richard, and Amedeo Odoni. *Airport Systems: Planning, Design, and Management*. McGraw-Hill, 2003.

(4) Lucic, Panta, Melissa Ohsfeldt, Mark Rodgers, and Alexander Klein. *Airport runway capacity model review*. Research report by CSSI and Air Traffic Analysis, Inc. for FAA ATO-P Performance Analysis and Strategy, August 2007.

(5) Ball, Carl T. Model Users' Manual for Airfield Capacity and Delay Models. FAA, U.S. Department of Transportation, 1976.

(6) Odoni, Amedeo et al. *Existing and Required Modeling Capabilities for Evaluating ATM Systems and Concepts*. International Center for Air Transportation at MIT, for AATT Program (NASA), March 1997, http://web.mit.edu/aeroastro/www/labs/AATT/index.html. Accessed June 21, 2008.

(7) Hansen, Mark. Post-Deployment Analysis of Capacity and Delay Impacts of an Airport Enhancement: Case of a New Runway at Detroit. *Air Traffic Control Quarterly*, Vol. 12, No. 4, 2004, pp. 339-365.

(8) *Airport Capacity Benchmark Report 2004*. Federal Aviation Administration. http://www.faa.gov/about/office_org/headquarters_offices/ato/publications/bench/2004download.htm. Accessed February 12, 2007.

(9) Theil, Henri. *Applied Econometric Forecasting*. North-Holland Publishing Company, Amsterdam, The Netherlands, 1966.

(10) Greene, William H. *Econometric Analysis*, *4th Ed*. Prentice Hall, New Jersey, 2000.