Measuring Ground Delay Program Effectiveness Using the Rate Control Index

Robert L. Hoffman
Metron Scientific Consultants
11911 Freedom Drive
Reston VA 20190
hoff@metsci.com
703-787-8700

Michael O. Ball
Robert H. Smith School of Business
University of Maryland
College Park, MD 20742
mball@rhsmith.umd.edu
301-405-2227

March 29, 2000

1. The Need for Improved Metrics

The increase in air traffic in the United States over the last 20 years has necessitated more frequent use of traffic flow management initiatives. In 1998, there were 187 ground delay programs (GDPs) run at San Francisco airport alone. As a result, there has been considerable interest both in the aviation and research communities regarding traffic flow management and its analysis.

Since the primary purpose of a GDP is to control the rate of flow of aircraft into an airport, the typical metric for evaluating the performance of a GDP is the comparison of actual landings per hour with the planned landings per hour. Often times, upon execution of a ground delay program, the airport acceptance rate (AAR) is unexpectedly low due to weather or other conditions at the airport and not all of the aircraft delivered in a given hour can be accepted. In this case, the standard metric shows poor performance when, in reality, the program was executed as planned. Since better weather forecasting is generally not the aim of improvements in traffic flow management, one can see the need for metrics that assess the execution of initiatives independently of the quality of their design. Our interest in the analysis of GDPs is motivated by the substantial changes GDP planning has undergone through the application of Collaborative Decision Making (CDM) (see Ball et al 1999, Hoffman et al 1999, Wambsganss, M., 1997).

For the case of a ground delay program, one solution is to meter the traffic as it enters the terminal space of the airport, rather than on the runways. We define a radius of geographical distance (or time) about the airport and declare that once a flight has passed over this boundary, it has entered the terminal airspace of the airport. This should roughly correspond to the point at which flights are ordinarily put into a state of airborne holding. (Special care must be taken at airport that hold aircraft at great distances from the terminal space.) The idea is to meter the traffic flow at this boundary rather than at the runways and compare it to the desired flow. Our comparison is based on the Rate Control Index (RCI), which is introduced and analyzed in Hoffman and Ball, 2000.

We assume the existence of a model that will estimate, post facto, for each flight f the amount of airborne holding that it incurred (see Ball et al, 1998 for one such model). From this, we can deduce whether or not f was in a state of airborne holding at a given time.

For contiguous time intervals $t = 1, 2,..., T$, let D_t be the number of flights delivered to the airport during time interval t, meaning, arrived at the border of the terminal space (but not necessarily landed). Let H_t be the number of flights being held in the air at the end of interval t. Let L_t be the number of flights that land during time period t. If we view the airport as a closed system, as in Figure 1, then we have the following elementary relationships:

$$D_t = (H_t + H_{t-1}) + L_t$$

Once the distribution $D = [d_1, d_2, ..., d_T]$ is formed, the question is how to weigh it against the planned distribution, $P = [p_1, p_2, ..., p_T]$. An obvious idea here is use vector subtraction to form a distribution of errors

$$E = P - D = \{ p_1 - d_1, p_2 - d_2,..., p_T - d_T \}$$

then to apply a standard variance technique, such as averaging over the time periods or summing the squares of the deviations.

There are two reasons why this might not be the best approach. First, it develops unintuitive or meaningless units that are difficult to translate into tangible quantities or cost assessments. Secondly, consider the scenario in which 30 flights are planned to arrive in each of two consecutive hours. Suppose that one flight scheduled for the first hour arrives late and spills over into the second hour arrival count, i.e., $P = [30, 30]$ and $D = [29, 31]$. Standard variance techniques would report an average error of one flight per hour, for each of the two hours. While this is statistically correct, the traffic flow "error" was that one flight migrated one hour, and yet this error is recorded twice, once in the first hour and once in the second.

2. The Core of the RCI Metric

The alternative measure of variance that we present here, the rate control index (RCI), records the minimum amount of aggregate flight movement that lead to the deviation of D from P. This is the same as the minimum amount of flight movement that would be necessary to revert D to P. The metric is normalized by dividing this quantity by cost of the worst case scenario. In this paper, we use examples to describe how the metric is calculated. Hoffman and Ball, 2000 gives the underlying mathematics.

As an example of the computation, consider a four-hour GDP in which $P = [30, 30, 30, 30]$ is the distribution of planned arrivals (to the terminal space), and $D = [27, 32, 35, 24]$ is the distribution

of flights delivered to the terminal space. Then the rate control index for D, relative to P, denoted by RCI(P, D), is computed as follows.

Step 1: Compute $P - D$, the minimum amount of flight movement that is necessary to turn D into P. We sweep left to right through D (increasing t), moving however many flights (call it f_{t+1}) are necessary from d_t to d_{t+1} to achieve $d_t = d_t - f_{t+1} = p_t$. Counting left-hand movements as negative and right-hand movements as positive, we must move 3 flights from hour 1 to hour 0 ($f_1 = -3$), move 1 flight from hour 2 to hour 1 ($f_2 = -1$), and move 4 flights from hour 2 to hour 3 ($f_3 = 4$). See Figure 2.

So far, D has been transformed into the distribution [30, 30, 30, 28]. Note that this is two flights short of the desired distribution, because

$$\sum_t p_t - \sum_t r_t = 120 - 118 = 2$$

We create a slush fund of 2 flights at the end of D to compensate for the lack of conservation of flights.[1] Equivalently, we could have started with an augmented distribution $D' = [27,32,35,24,2]$ and an augmented $P' = [30,30,30,30,0]$. To complete the example, we move 2 flights from hour 4 to hour 3 ($f_4 = 2$). The summary of flight-movements is given in Table 1.

Now we can compute the cost of transforming D into P. Let c^- be the average cost (say, in dollars per hour) of delaying a flight for one unit of time and let c^+ be the average cost (say, in dollars per hour) of a flight arriving early by one unit of time. Then we break the flight-movements into left-hand movements
$$M^- = |-3| + |-1| + |-2| = 6$$
and right-hand movements
$$M^+ = 4$$
corresponding to tardiness and earliness of arrival, respectively. The final cost is given by

$$c^- M^- + c^+ M^+ = 6c^- + 4c^+$$

In this example, we opt to set $c^- = c^+ = 1.0$ to obtain pure units of $4 + 6 = 10$ flight-hours. In other words, D was off from P by 10 flight-hours. Intuitively, this means that in order to turn D into P, one would have to do the work equivalent to moving 1 flight 10 hours, or 2 flights 5 hours, etc. One

---

[1] Similarly, one can compensate for unanticipated arrivals by adding the surplus to the other distribution. However, there is less intuitive justification for this.

variation on the metric is to retain the positive and negative sums to show the breakdown of this total.

Step 2: To normalize the metric, we divide the distribution error by the cost of the worst-case scenario. In this case, this is the five-hour redistribution W of the 120 flights with the highest cost difference, $P - W$. In general, this involves solving a max-min problem, whose solution leads to some interesting mathematics, too lengthy to present here. For now, we take as a given that W corresponds to the scenario in which all the flights land in the final hour, that is, $W = [0,0,0,0,120]$ with corresponding cost

$$P' - W = 30 \times \left(|-4| + |-3| + |-2| + |-1| \right) = 30 \times 10 = 300 \text{ flight-hours}$$

Thus, the rate control index we assign to the distribution D is

$$\frac{10 \text{ flight-hours}}{300 \text{ flight-hours}} = 0.03\overline{3}$$

In order to make the index more palatable, we subtract this ratio from 1.0 and multiply by 100%:

$$RCI(P', R) = \left(1.0 - 0.3\overline{3} \right) \times 100\% \; \square \; 96.67\%$$

96.67%. The interpretation of the index is that the realized distribution achieved 96.67% of the intended (planned) distribution.

The rate control index can accommodate revisions to planned acceptance rates by setting p_t equal to the largest target value for time period t that was set prior to time period t. This way, the target value p_t can be changed as many times as one wishes but the program will ultimately be evaluated on the final target value.

3. Sample Analyses

Figures 3 and 4 shows the trend of the RCI metric for SFO and EWR, respectively. The results were recorded over the 30-month period from January 1997 to March 1999. This captures a full year of data prior to the implementation of CDM prototype operations in January of 1998. Note that there is much greater variance in the monthly RCI value at EWR than at SFO. (See the lines with round icons.) This is likely due to the unpredictable nature of east coast traffic and the complexity of the airspace surrounding the New York City area (it lies on the boundary of more than one Center).

There appears to be a seasonal drop in the RCI value at each airport: for SFO, the drop occurs in the winter months, for EWR, the drop occurs in the summer months. These may be due to weather conditions. The drops were most pronounced in 1998, when the effects of El Nino were at their greatest. The RCI metric screens out effects of weather conditions immediately at the airport but traffic flow into the airport could be affected by weather conditions elsewhere in the NAS (e.g., this could delay flow into the airport and lower the RCI value).

We smoothed out the variance of the monthly points by computing a moving average over four months (see the lines with the square icons). Further smoothing was obtained by computing a cumulative average over all months since January of 1997 (see the lines with the triangular icons). Three "checkpoints" are worth noting: the first month for which a four-month cumulative average had been computed (April 1997), the start of CDM (January 1998) and the last point (March 1999). For SFO, these points were 91.42, 92.33 and 92.75. For EWR they were 89.13, 86.56 and 87.91. This shows a slight rise in the (cumulative average) RCI value for each of the airports since the start of CDM[2]. However, this trend is probably not statistically significant. We feel that a longer history is required to judge the significance of CDM's impact.

Figure 5 shows a graphical analysis of the performance of a ground delay program (GDP) conducted at San Francisco airport (SFO) on March 5, 1998. The planned acceptance rate of flights (the *PAAR* distribution) was set at 32 flights per hour for each of the six hours of the GDP. The RCI value assigned to this GDP was 88.24%. This is below the 1998 RCI average for SFO, approximately 92.0%, it would be a likely candidate for further analysis. The RCI value of 88.24% was computed based on *PAAR* versus *DEL*, the distribution of flights delivered to the terminal airspace of the airport (but not necessarily landed). Also shown are the distributions *LAND*, flights landed at the airport and *ABH*, the size of the airborne holding queue at the end of each hour.

A quick glance at Figure 5 shows why the RCI(*PAAR*, *DEL*) was not closer to the optimal value, 100%. There were too many flights delivered to the airport early in the program: in the first hour, 41flights were delivered when only 32 were intended. One possible explanation for this is that the GDP was implemented too late and some of these 41 flights were already airborne when the program was planned, hence, they could not have been held on the ground. Some of these flights may have been assigned ground holds but departed too early, but this is a less likely explanation. Moreover, there were far too few flights delivered in the 2100z hour (16 flights versus 32). Some of the flights intended to arrive in hour 2100z may have arrived in the 2000z hour. Note that it took the remainder of the program for this airborne holding queue to dissipate that developed from the glut of flight delivered in the 2000z hour.

---

[2] Since the number of ground delay programs varies with the month, these results do not give equal weight to ground delay programs. Different results are obtained when equal weight is given to each program, i.e, averaging over all prior ground delay programs. The method we have adopted screens out some of the seasonal effects of weather. The legitimacy of this method is confirmed by averaging over each season of the year (results similar to those above are obtained)

4. Aggregate Vs. Nominal

If one is only concerned with the *aggregate* flow of traffic during a GDP, then the rate control index that we have presented is sufficient. However, since a ground delay program assigns specific flights to specific time intervals, a more complete assessment of GDP execution can be made by employing a *nominal* version RCI, which we will denote RCI_{NOM}. This version of the metric measures the total deviation of individual flights from their intended arrival slots. That is, for each flight f, we compute the amount of arrival delay,

$$M_f = |\text{actual arrival time} - \text{planned arrival time}|$$

to arrive at the unnormalized RCI score is $\sum_f M_f$. This represents the amount of flight movement through time that would be necessary to restore all flights to their planned arrival periods. This is normalized by dividing by $\sum_f W_f$, where W_f is the most arrival delay that could have occurred for flight f (i.e., the farthest time period from its planned period of arrival). As with the aggregate version, this ratio is subtracted from 1.0 and multiplied by 100% to arrive at the final rate control index.

To demonstrate how RCI_{NOM} and RCI_{AGG} are used in conjunction, we have plotted in Figure 6 one ordered pair, (RCI_{NOM}, RCI_{AGG}), for each GDP run at SFO during the period January 1 to October 28, 1998. Consider the point corresponding to the July 10th GDP, (64, 92). The relatively high value of RCI_{AGG} (92%) indicates that, in the aggregate, the program was very successful. However, the low value of RCI_{NOM} (64%) reveals that, too often, flights arrived outside of their intended time interval. This discrepancy reveals that for a typical time interval t, the number of flights intended to arrive in t that arrived in some other interval was generally close to the number of flights that arrived in interval t that were not intended to. In other words, the aggregate success was the result of a great deal of balancing of early and late arrivals.

Since the highest possible score of a program is (100%,100%), the best-run programs can be found in the upper right of the scatter plot. Note that all of the points lie above the 45-degree line. This is to be expected since as RCI_{NOM} increases, more flights are arriving in their planned arrival periods and so the aggregate distribution of flights is more likely to match the planned distribution, which also increases the value of RCI_{AGG}. Note that when $RCI_{NOM} = 100\%$, the only way for RCI_{AGG} to fall below 100% is if there are arrivals not anticipated by the GDP.

The center of mass of the squares lies at about (82, 92), marked by the cross. This indicates that the ground delay performance at SFO is generally quite good. In general, RCI_{NOM} is about 10% less than RCI_{AGG}.

5. Closing Remarks

The development of an aggregate and nominal version of the rate control index captures the two crucial aspects of traffic flow management: *how many* flights flowed through a region of space and *which* flights flowed through the region. We have shown how to apply the metric at the terminal space of an airport to factor out the effects of inaccurate forecasts from performance analysis of a traffic flow initiative, the metric can used in a more traditional way by metering traffic at the runways.

Proper analysis of airport arrival flow can be achieved by computing four versions of the metric: aggregate and nominal versions metered at the terminal space, and aggregate and nominal versions metered at the runways. Since, in its most general form, the RCI metric generalizes to a comparison of two finite distributions, it can be used to meter the flow of traffic through any region of space.

6. References

Ball, Michael, R. Hoffman, W. Hall, and A. Muharremoglu, 1998, "Collaborative Decision Making in Air Traffic Management: A Preliminary Assessment", NEXTOR technical report RR-99-3.

Ball, Michael O., R. Hoffman, C. Chen, and T. Vossen, 1999, "Collaborative Decision Making in Air Traffic Management: Current and Future Research Directions", Proceedings of Capri ATM-99 Workshop, pp 267-276.

Hoffman, Robert and M. Ball, 2000, "The Rate Control Index for Traffic Flow", submitted for publication.

Hoffman, Robert, M. Ball, A. Odoni, W. Hall, and M. Wambsganss, 1999, "Collaborative Decision Making in Air Traffic Flow Management", manuscript in preparation.

Wambsganss, M., 1997, Collaborative decision making through dynamic information transfer, Air Traffic Control Quarterly, 4, 107-123.
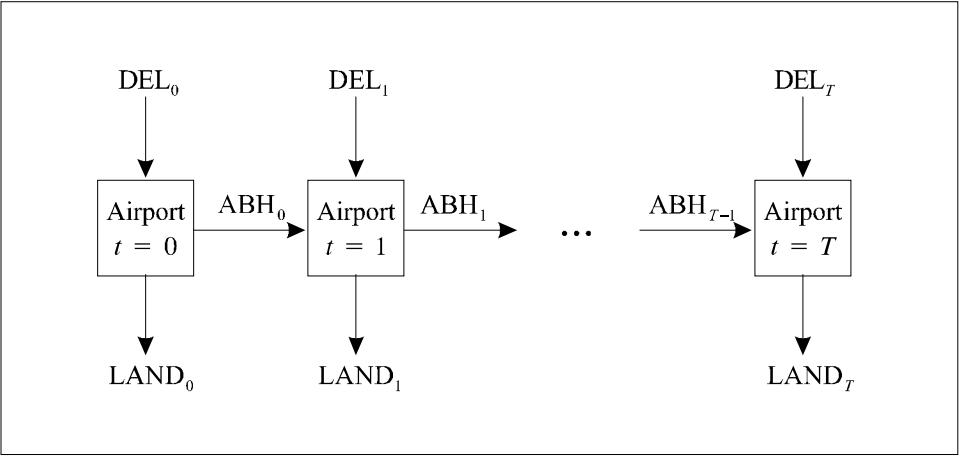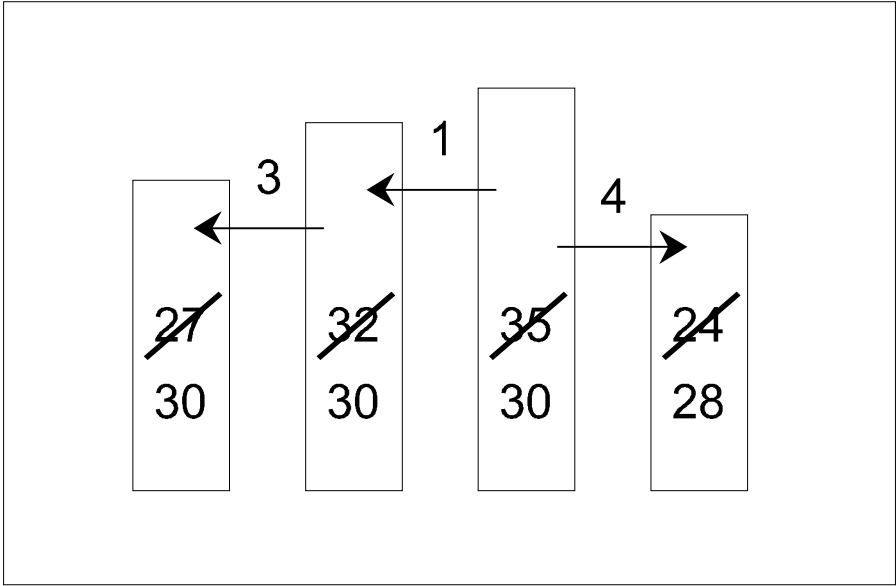
Figure 1: Model of an airport as a closed system

Figure 2: Flight movements necessary to transform *D* into *P*

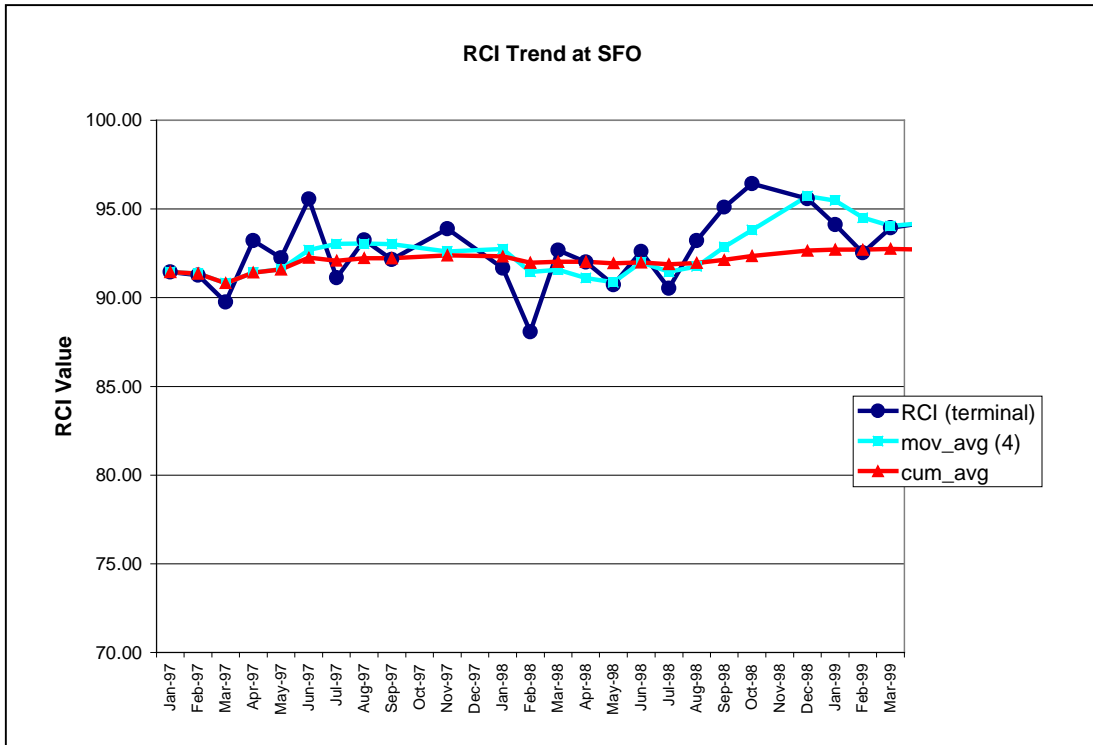| Movement	Notation	Flight-hours
Hr 0 → Hr 1	f_1	−3
Hr 1 → Hr 2	f_2	−1
Hr 2 → Hr 3	f_3	4
Hr 3 → Hr 4	f_4	−2

Table 1: Summary of flight movements through time

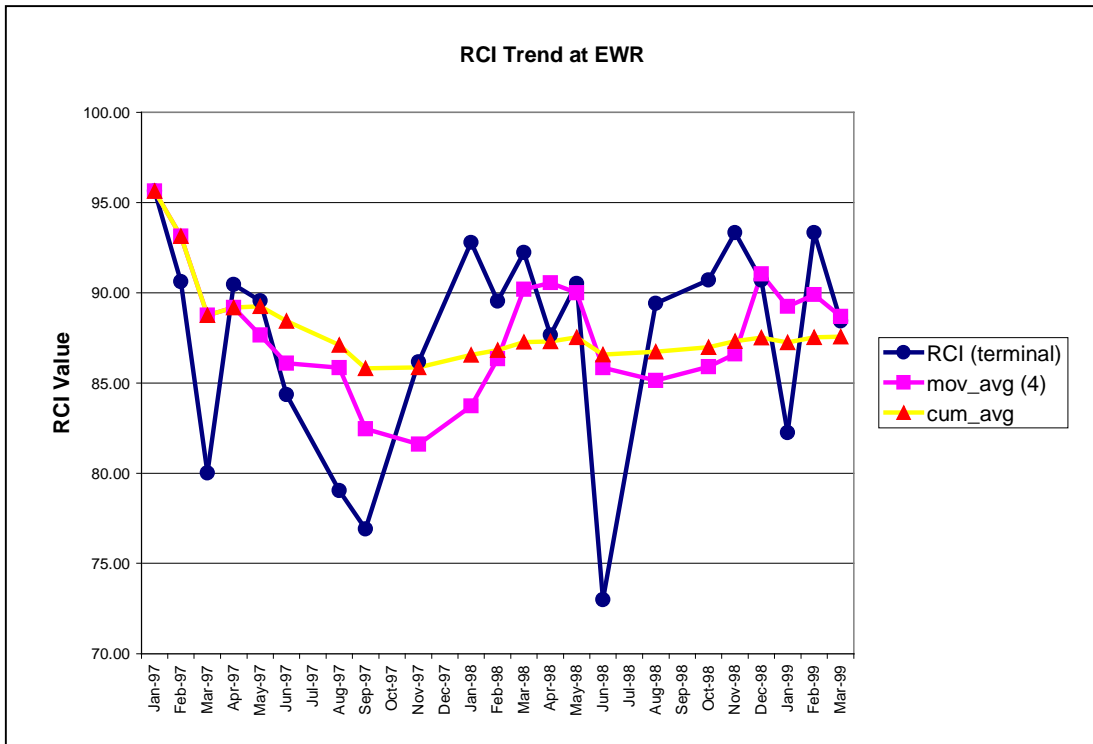Figure 3: Cumulative Average of the RCI Metric recorded at SFO

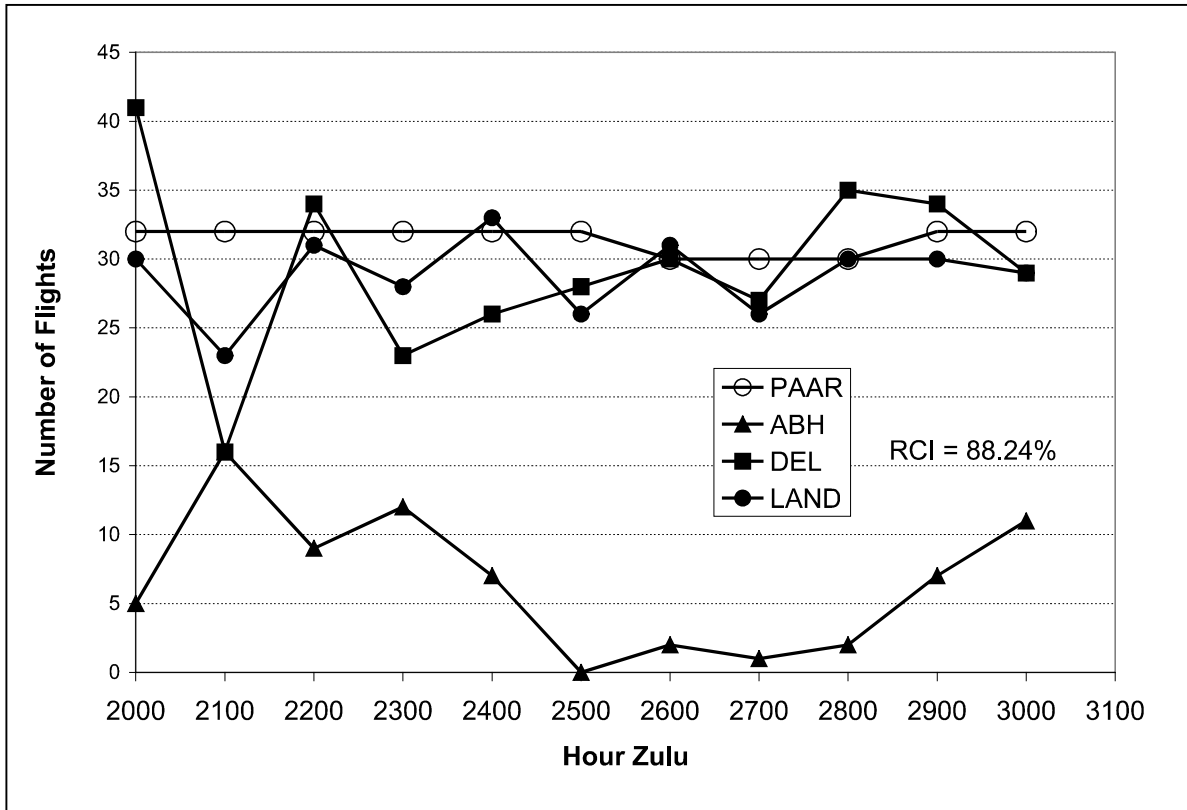Figure 4: Cumulative Average of the RCI Metric recorded at EWR

Figure 5: Analysis of ground delay program performance at SFO, March 3, 1998
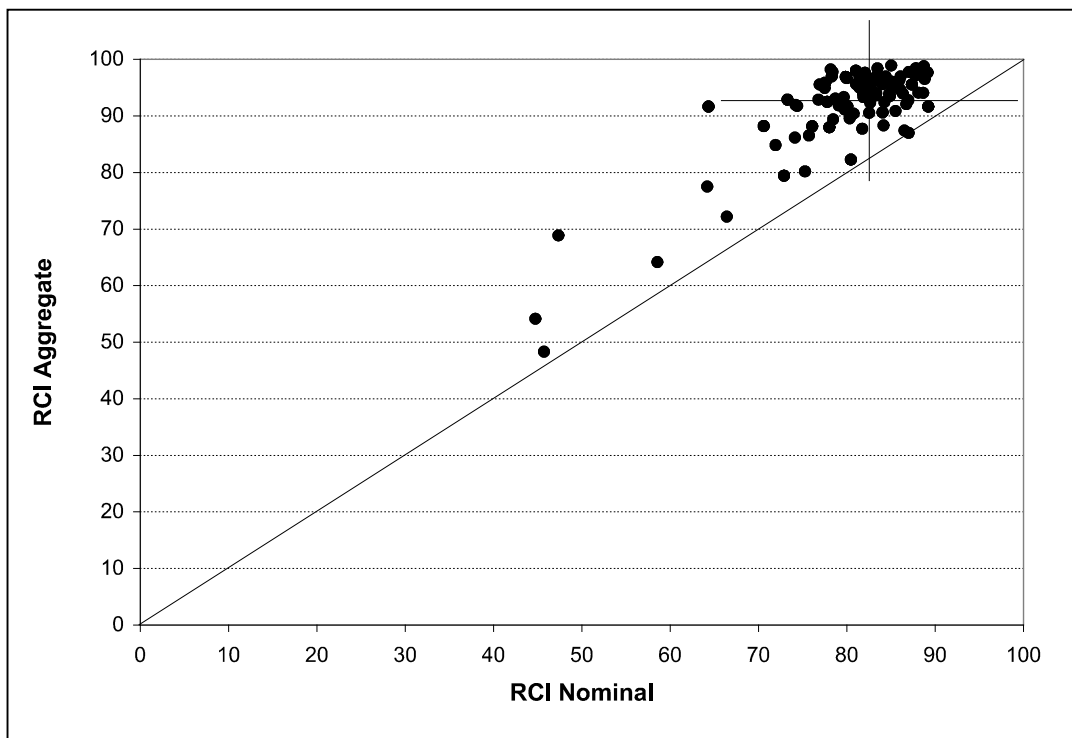

Figure 6: Nominal and aggregate RCI values for ground delay
programs at SFO, Jan – Oct 1998