

# **Report of the Rolling Spike Task Force**

**April 1, 1998**

Robert Hoffman, Univ. MD  
Lara Shisler, Metron  
Ken Howard, Volpe  
Mark Klopfenstein, Metron  
Michael Ball, Univ MD

This is a report by the members of the rolling Spike Task Force of the Collaborative Decision Making (CDM) Analysis subgroup on the phenomenon known as the "rolling spike" that has been observed on the Flight Schedule Monitor (FSM) during ground delay programs (GDP's).

We will begin with a general description of the phenomenon and some conclusions, then present the details of the statistical analyses that have lead us to those conclusions, and close with a list of recommendations.

## **1. What is the rolling spike phenomenon?**

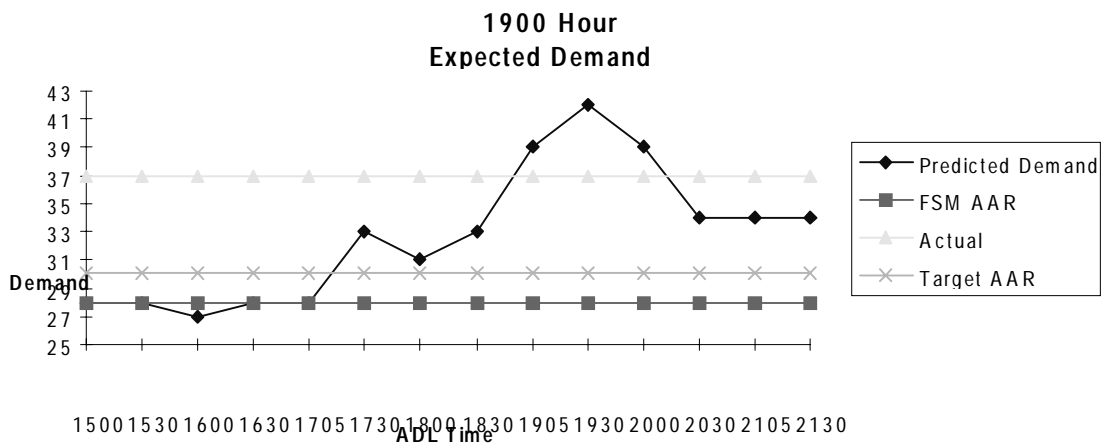
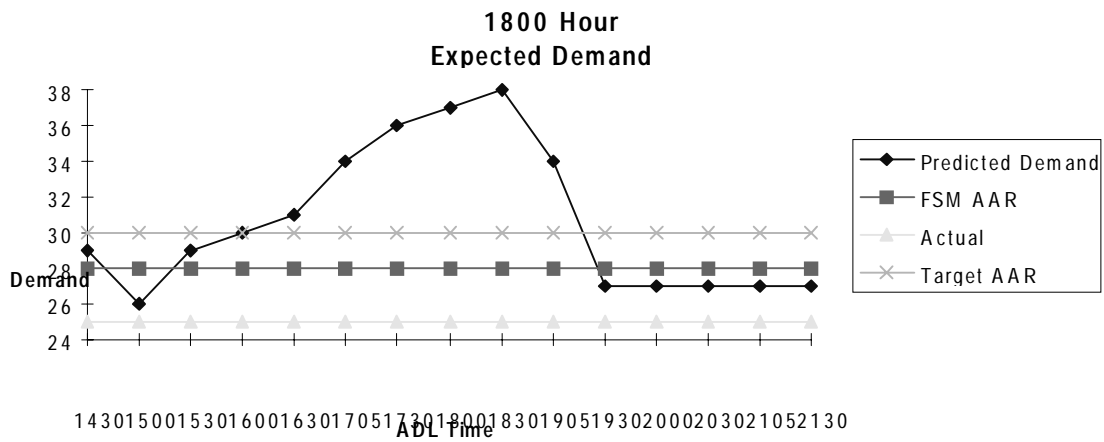
The rolling spike phenomenon within the CDM system is characterized by a persistent redistribution of predicted demand from time intervals at a constant distance out from the current time, into later time intervals. It is most noticeable and of greatest concern when the redistribution forms "spikes" in predicted demand of sizes greater than the capacity of the airport in question. A key component of this phenomenon is that when the period, or periods, in which the spike occurs comes to pass, the actual number of flights that arrive in that period is significantly less than expected. Since the primary objective of a GDP is to take control of the flow rate of flights into an afflicted airport, this phenomenon has become a major concern to both the specialists at the Air Traffic Control Systems Command Center (ATCSCC) and the CDM community.

The seven charts in Appendix A give an example of a rolling spike that occurred at Boston's Logan Airport on March 3, 1998. Each chart is a demand prediction for each of the hours 1700 - 2300. For instance, the first chart gives the demand predictions for 1700 - 2300 hours, based on predictions made at 1700 (the "view" time). Note that there is a bulking up of demand in the hours 1700, 1800, 1900. In the next chart (1800 view), the bulk has shifted slightly to the right. Note that the "predicted" demand for the 1700 hour has become the actual number of arrivals and that this number (39) is slightly less than the 42 predicted at 1700. Move quickly through the subsequent charts and you will see the bulk continually shift to the right (forward in time) and become more concentrated, leading to the large spike in the 2200 hour of the 2200 view.

The rolling spike phenomenon is analogous to the "wave" that is created by the spectators at a football stadium when columns of spectators rise and fall in sequence. In the case of demand predictions, the columns of demand correspond to the columns of spectators. For

instance, scroll through the charts in Appendix A in sequence, keeping your eye on just the 2000 hour and you will see it gently rise from chart to chart and then drop once the view time reaches 2100. Just as there is no perfect wave in a football stadium, there is no perfect rolling spike in demand predictions. The effect is more pronounced at certain days at certain airports and, at other times, barely noticeable. There is a great amount of subjectivity as to how long a spike must “roll” in order for it to constitute a rolling spike.

The following two charts show a more detailed analysis of the pattern of growth of predicted demand for the hours 1800z and 1900z, respectively, at SFO during a January 31, 1998 GDP. The physical time (times of observation) are plotted on the horizontal axis in each chart. As you can see, the expected demand is initially very close to the AAR which was used to run the program. But as time moves forward, the demand starts to rise and finally reaches a peak during the arrival hour being observed. There is then a drop in the predicted demand to the number of aircraft which actually had AZ messages during that hour. In both graphs, the difference between the peak expected demand and the actual arrivals is all accounted for by aircraft which were delayed and arrived in a later hour.



## **2. Is the rolling spike phenomenon a problem to be concerned about?**

Yes. Recall that (1) one of the ultimate goals of CDM in general and FSM in particular is to provide accurate demand forecasts of arrivals into a given airport and that (2) the rolling spike phenomenon is sustained by unfulfilled demand predictions. Both the specialists at the ATCSCC and the operational centers of the airlines are becoming increasingly dependent on the forecasts provided through FSM so it is crucial that these forecasts be as accurate as possible. In particular, it has been reported that a rolling spike can prevent a specialist from running compression for fear of compounding the undesirable effects of heavy spike in demand. Also, widespread ground delay during a GDP is considered unacceptable because flights are supposed to be departing close to their issued controlled times of departures (EDCT's).

There is some sentiment amongst the CDM members that, since the demand of a rolling spike "never materializes" it is in some sense fictitious and, therefore, not of great concern. However, our analyses show that the majority of the demand of a rolling spike is quite real; it's just that the flights are arriving later than expected. We should add that part of the reason that spikes in demand never materialize is because each airport has a limited arrival capacity and the rest are put into an airborne queue, hence, not counted as arrivals until a later time period. In other words, if arrival count is the measure of materialization of predicted demand, then materialization itself is limited, especially during a time of reduced airport capacity.

We believe the rolling spike phenomenon to be a general trend that is specific to neither CDM nor FSM. The rolling spike phenomenon has been observed by the specialists at the ATCSCC long before CDM. It has received much more attention during CDM, however, since a wide body of CDM participants have access to demand forecasts through FSM and GDP's in general have fallen under heavy scrutiny.

### 3. What causes the rolling spike?

There are many factors that contribute to the rolling spike. In theory, any one of them could cause the phenomenon. After a review of demand predictions throughout the day for several airports (BOS, SFO, EWR), however, we have concluded that, in practice, the primary cause is *the net redistribution of predicted demand to later time intervals as the (CDM) system is notified of arrival delays incurred both on the ground and in the air, ground delay being more of a cause than air delay. Flight cancellations is a secondary but significant cause.* Moreover, in order for there to be a true rolling spike, there must be a continued pattern of delay (or cancellations) throughout an extended period of observation.

Let us consider the effect that the ground delay of a single flight can have on demand predictions. Suppose that at time 1500, flight AIR1280 is predicted to depart at 1550 and arrive at in BOS at 1750, and that it is just one of 60 flights predicted to arrive in the 1700-1759 hour. If AL1280 departs 70 minutes late (at 1700) but this is not known to the system (CDM) until its departure, then the demand prediction for 1700-1759 will remain at 60 flights (all else being equal) until 1700, at which time the demand for 1700-1759 will be dropped from 60 to 59 and the demand for the 1800-1859 hour will be increased by one flight. In this manner, some of the predicted demand for the 1700-1759 hour is redistributed to the next hour. If this pattern of delay occurs for a significant number of flights and on a regular basis (say, more than one hour), then demand is continually being redistributed to later time intervals and there is a “spike” of demand rolling through time but hovering very close to the current (physical) time. The movement of the spike is a natural consequence of the CDM system revising (increasing) estimated times of departure (ETD’s).

One can see from this hypothetical example that it is not possible for the ground delay of a flight departing from an origin airport at a great distance (several hours) to contribute to a rolling spike close to the physical time because, once it is airborne, its ETD is updated in the system and the demand forecasts are adjusted accordingly . This could, however, contribute to a rolling spike that lingers several hours out in the demand forecast.

Departure time accuracy plays a key role in the rolling spike phenomenon. Next, we outline a method for producing a reasonable estimated time against which to compare the actual time of departure.

1. Take the OETD at the time of departure
2. Remove any “time out” delay (the automatic ETMS pushback).
3. Remove any “last minute” updates from the airlines (e.g., if the OETD was 1827, then at 1830 the airline told us the time would be 1900, use 1827 as the prediction).

The two data sets we studied were for SFO during a GDP (3/12/98), and BOS without a GDP, but with bad weather (there was almost a GDP). What we found was that for the SFO GDP case, the OETD did not require any adjustments. That’s because the OETD is

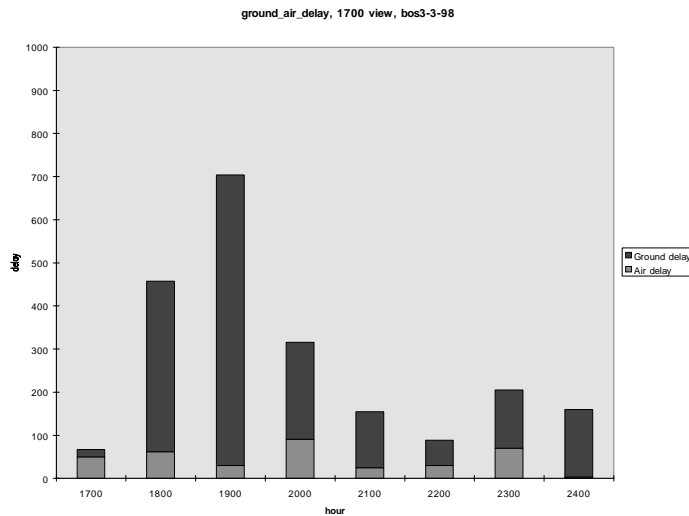
frozen once an EDCT comes in. For the BOS data, we did the manual adjustments. We also broke out the flights for which the airline was sending runway departure times to see if they varied. The table below summarizes the findings. Note that while the averages are not all that bad, the standard deviations are pretty high.

| <b>Departure Time Performance (Actual – Estimated)</b> |                      |                      |            |            |
|--|----------------------|----------------------|------------|------------|
| <b>Data Set</b>  | <b>Average Error</b> | <b>Standard Dev.</b> | <b>Min</b> | <b>Max</b> |
| SFO – Controlled                                       | -11                  | 53                   | -460       | 225        |
| BOS – All  | 10                   | 21                   | -38        | 149        |
| BOS - Airline  | 8                    | 16                   | -12        | 64         |

Airborne delay contributes to the rolling spike phenomenon in two ways. The first way is from airborne delay incurred before arriving at the terminal space of the airport in question. For instance, suppose that throughout the day, a significant number of flights increase their ETA's while they are enroute. (This could be caused by unexpected head winds or overly optimistic estimated enroute times.) Then the system will exhibit a steady pattern of redistribution of arrival predictions to later time intervals and a rolling spike is observed. The second way for airborne delay to contribute to the rolling spike is for there to be an arrival queue at the terminal space of the airport. In this case, the ETA's of the flights are not updated until they are close to the destination airport, thus contributing to a spike close to the current physical time.

In the flight records that we examined, it was quite common for a flight to have incurred *both* ground delay and airborne delay. For this reason, most rolling spikes are not attributable to strictly airborne delay or ground delay. Later in this report, we will present the relative roles that these two effects have played for the airports we studied. For now, we offer as a general rule of thumb, that a rolling spike close to the current hour is most likely due to ground delay of short-haul flights and/or airborne delay of general flights, while a rolling spike far away from the current hour is most likely due to ground delay of long-haul flights. The former type of close-in spike is more common than the latter, particularly at an airport with a lot of short to medium haul flights. In fact, we strongly suspect that the bad weather conditions on March 3, 1998 led to massive ground delays and caused the rolling spike at Logan airport for that day.

Below is a chart of the amounts of ground delay and airborne delay that directly contributed to unfulfilled demand predictions (made at 1700z) at Boston's Logan airport on March 3, 1998. The numbers were arrived at by computing for each future hour  $h$  the total amount of ground delay (and air delay) from all the flights that were expected to arrive in  $h$  but arrived in an hour later than  $h$ . This chart is typical of the many airports and days that we studied. Note that ground delay dominates airborne delay in all hours except the 1700 hour. This shows that the unfulfilled demand in the 1700 hour was caused mostly by airborne delays while in all other hours, it was caused mostly by ground delay.



We suspected that airborne delay would be more of a factor in the rolling spike phenomenon at SFO than at an airport like BOS, which has many more short and medium haul flights. Indeed, this is the case. On average, over the data we studied, SFO has twice the ratio of airborne holding to ground holding than does BOS that directly contributes to flights arriving in a time period later than expected. See Appendix C for details.

We should point out that in any given hour during a rolling spike observation period, there are two ways for demand to be unfulfilled. One, as we have mentioned, is for some flights to be moved to later time intervals but the other is for some of the flights to be moved to earlier time intervals. In order for there to be a rolling spike, however, the redistribution must be skewed in favor of distribution to later time intervals.

Next, we list each of the factors that can contribute to a rolling spike, discuss the nature of its contribution and suggest ways to alleviate its effects.

### **1. Ground delay (departure after ETD).**

*Contribution:* Already discussed.

*Solutions:* More frequent, more accurate ground delay predictions from the airlines.

### **2. Airborne delay (arrival after ETA), underestimated ETE's.**

*Contribution:* Both enroute delays and delays at the terminal space of the destination airport have already been discussed. We comment that we have no formal method at this time for distinguishing these two through analysis of flight records. Clearly, underestimated ETE's (estimated enroute time) can contribute to inaccurate demand predictions.

*Solutions:* More frequent, more accurate enroute time predictions from the airlines.

### **3. Flights missing their EDCT's.**

*Contribution:* This falls under the more general category of ground delays because during a GDP, the EDCT is the same as the ETD (at least, initially).

*Solutions:* More frequent, more accurate enroute time predictions from the airlines.

#### **4. Cancellations.**

*Contributions:* We have found that canceled flights contribute to a small but significant amount of unfulfilled demand predictions. See Appendix B for more details.

*Solutions:* While a certain amount of cancellation (particularly during a GDP) is inevitable, it is important that the system be made aware of this as soon as possible in order to correct demand predictions.

#### **5. Pop-up flights.**

*Contribution:* If a flight is suddenly added to the CDM system then it will increase the demand prediction for its respective arrival hour. A steady flow of pop-up flights can contribute to the build up of demand in future time periods. We have found this to be a relatively minor factor.

*Solutions:* The effect of pop-up flights can be nullified only by increasing the lead time on the introduction of these flights to the system.

#### **5. Time out cancellations/delays.**

*Contributions:* If no departure is reported for a scheduled flight, then ETMS continually pushes back its ETD (and ETA) until the flight is “timed out”. This contributes to unfulfilled demand.

*Solutions:* Better data from the airlines and the elimination of “dead” flights from the OAG.

#### **6. Diversions (non-reported).**

*Contributions:* Unreported diversions lead to unfulfilled demand but not in a significant amount.

*Solutions:* Elimination of “dead” flights from the OAG and better notification from the airlines on diversions.

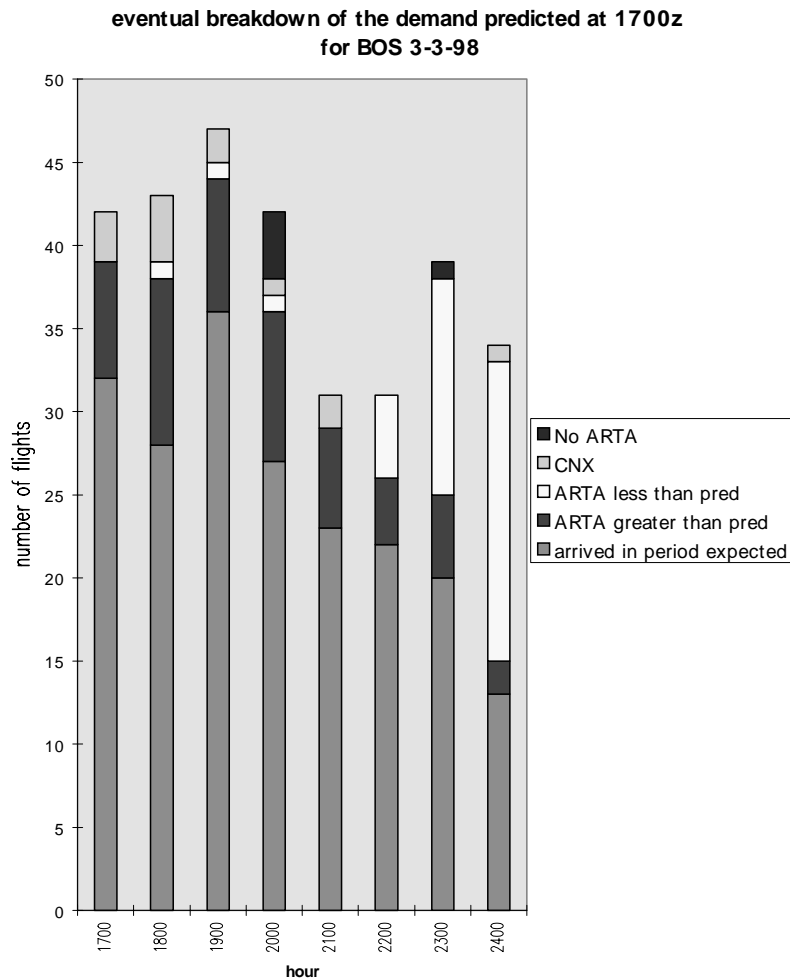
#### **7. White hats.**

*Contribution:* Flights that are given special exemption from ground delay through personal communication between the ATCSCC specialist and an airline can contribute to unfulfilled demand. However, since these flights are being moved forward in time, they are not a direct cause of a rolling spike. White hats are very hard to identify in retrospect but do not appear to be aggravating the rolling spike in any way.

*Solutions:* None needed.

## 5. The Need For Stochastic Modeling

In addition to each of the solutions mentioned above, a stochastic model should be developed that would help to revise FSM demand forecasts based on historical analyses similar to the chart below. The overall height of each column in the chart below represents the number of flights that were predicted to arrive (as of 1700z) in the respective hour. The column is broken down according to the number of flights that fell into certain categories (e.g., the number of flights that were eventually canceled). The flights in the “no ARTA” category and “CNX” category comprise the fictitious portion of the predicted demand while the flights in the “ARTA less than predicted” and “ARTA greater than predicted” category will be distributed to earlier and later time periods, respectively.



Work has begun by members of NEXTOR (National Center for Excellence in Aviation Operations Research) on such a stochastic model. Software has been developed that will input the flight records for an entire day at a single airport and generate for each hour  $h$  of the day (or any time period) a statistical analysis of the demand predictions that were made for the remainder of the day. The analysis reviews the accuracy of the predictions

(in hindsight) and keeps careful track of the numbers of flights that fell into the categories such as (1) predicted to arrive but arrived in a later time interval (2) predicted to arrive but arrived in an earlier time interval (3) canceled at a later time (4) pop-up flight (5) never received an ARTA for that flight, and so on. Using such data for a wide variety of days (both GDP's and non-GDP's), distributions can be constructed that will be used to guide a stochastic model that will revise demand predictions.

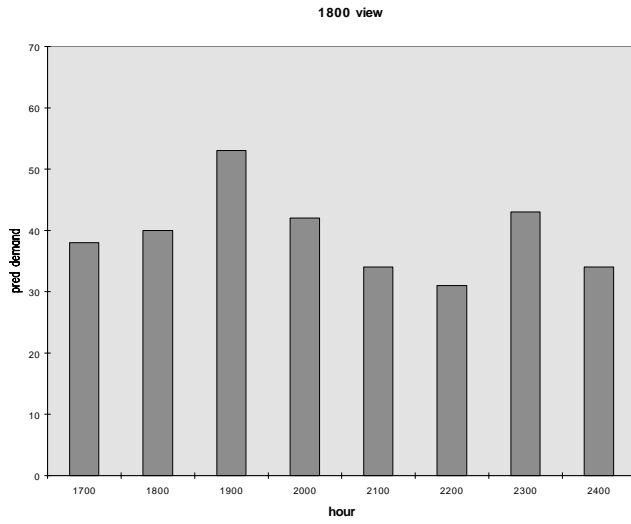
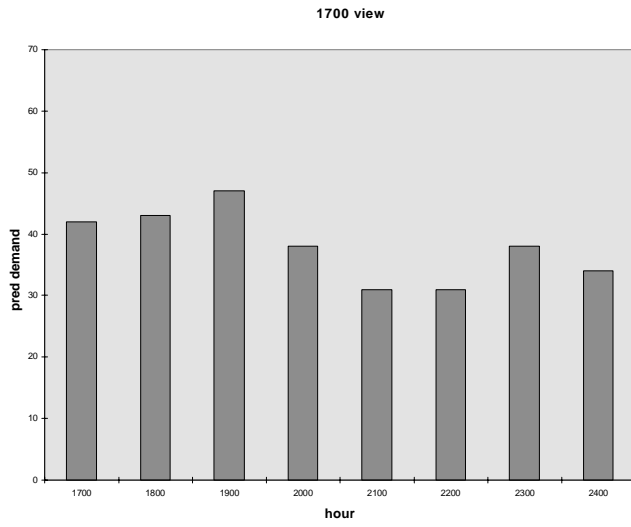
## **7. Conclusions and Recommendations**

The rolling spike phenomenon is a very real phenomenon observed in the demand forecast displayed by FSM. The primary cause is the persistent redistribution of predicted demand to later time intervals as the (CDM) system is notified of arrival delays incurred both on the ground and in the air. For the airports we studied, ground delay was more of a cause than air delay. Other contributing factors are flight cancellations and underestimated enroute times. This phenomenon is peculiar to neither CDM nor FSM but it is to be considered a hindrance to one of the primary objectives of CDM and FSM, which is to improve demand forecasting at the airports.

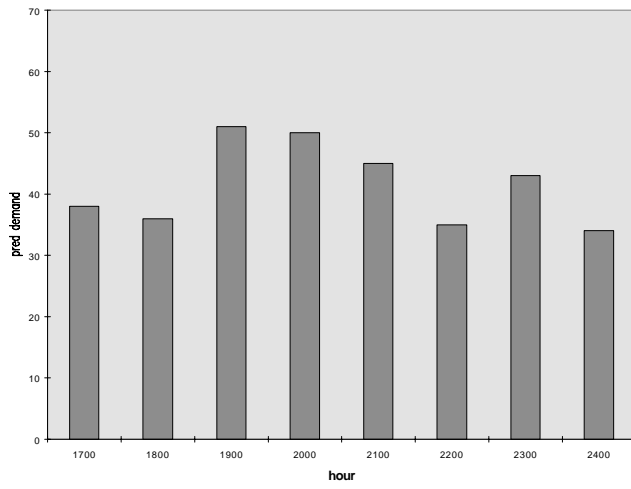
We recommend two courses of action on the part of the CDM community:

1. That the airlines make an effort to submit more timely, more accurate data over the CDM string, especially with respect to departure delays and cancellations.
2. That a statistical model be designed, tested, and implemented into FSM so that FSM users can obtain an alternate view of demand forecasts that has been tempered with historical and probabilistic information, especially regarding flight delays and cancellations. Such a model should be airport-specific, time-specific and dynamic.

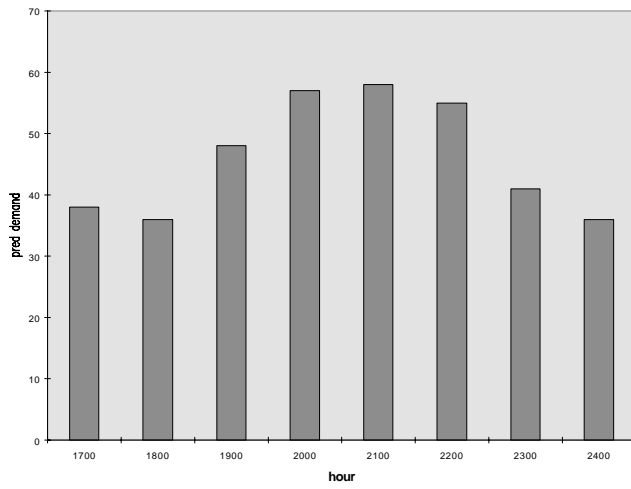
## Appendix A: the rolling spike at BOS 3/3/98



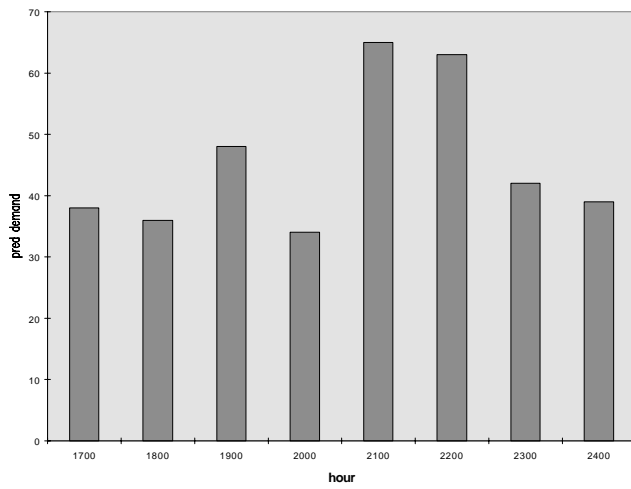
1900 view



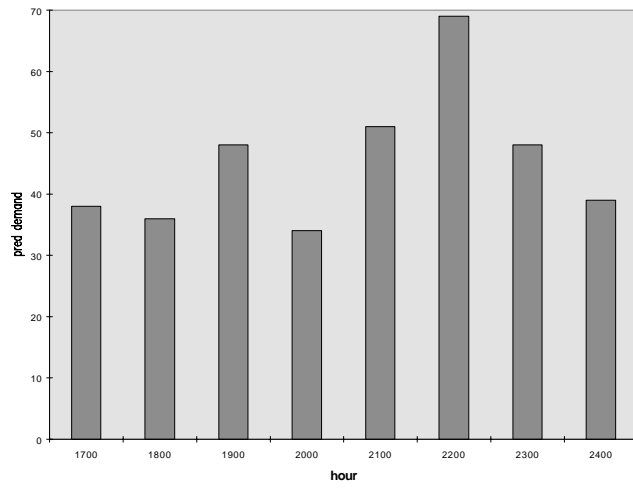
2000 view



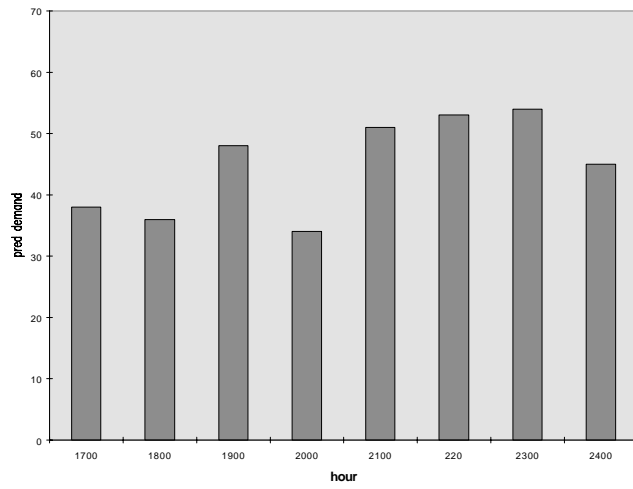
2100 view



2200 view



2300 view



## Appendix B: cancellations at BOS 3/3/98 and SFO 3/12/98

We examined some BOS data during a non-GDP (but bad weather) and some SFO data during a GDP to see who was canceling and for what reasons. In the BOS data, we found that roughly half of the cancellations for the CDM airlines were not sent as FX messages. The SFO data was better. However, two things became evident:

1. The number of timed-out flights, especially at BOS, makes it very difficult to predict accurately.
2. While there are fewer timed-out flights for CDM participants, there are more than one would like.

| <b>Cancellations at BOS</b> |                        |                        |                              |                  |
|-----------------------------|------------------------|------------------------|------------------------------|------------------|
| <b>Data Set</b>             | <b>Total Cancelled</b> | <b>Via CDM (FX+ID)</b> | <b>Other Pre-dep (RZ+DV)</b> | <b>Timed Out</b> |
| BOS – All                   | 68                     | 19                     | 18                           | 31               |
| BOS - CDM Participants      | 37                     | 19                     | 8                            | 10               |
| BOS - Non-participants      | 31                     | 0                      | 10                           | 21               |

| <b>Cancellations at SFO</b> |                        |                        |                                 |                  |
|-----------------------------|------------------------|------------------------|---------------------------------|------------------|
| <b>Data Set</b>             | <b>Total Cancelled</b> | <b>Via CDM (FX+ID)</b> | <b>Other Pre-dep (RZ+DV+SI)</b> | <b>Timed Out</b> |
| SFO – All                   | 79                     | 36                     | 32                              | 11               |
| SFO - CDM Participants      | 49                     | 36                     | 8                               | 5                |
| SFO - Non-participants      | 30                     | 0                      | 24                              | 6                |

One can argue that the timed out flights for CDM participants are due to bad data from OAG, and the CDM data feed should not be blamed. So we looked at who got time-out delays but operated. Obviously, if the flights operated, they should have gotten data updates from the participants. But we see that even in this case there are significant holes in the data that the airlines are sending.

| <b>Time-out Delayed Flights</b> |              |                         |
|---------------------------------|--------------|-------------------------|
| <b>Airport</b>                  | <b>Total</b> | <b>CDM Participants</b> |
| BOS                             | 276          | 194                     |
| SFO                             | 189          | 140                     |

## Appendix C: ratios of ground delay to air delay that directly contribute to unfulfilled demand predictions

From the following charts, we have concluded that at SFO, ground delay is about twice as much of a factor in unfulfilled demand than is airborne delay and that at BOS, ground delay is about four times as much of a factor than is airborne delay. A great deal of averaging went into the construction of the tables, so these conclusions should not be overly interpreted.

In order to put the tables in perspective, we now explain the method by which they were constructed. Consider the entry in the SFO table in the 1700 row and the 2/3/98 column (the entry is 0.37). Here's how this number was arrived at. Set the view time to 1500. List all the flights which were predicted (as of view time 1500) to arrive in the 1700 hour but that actually arrived in a later time interval. Add up the total number of minutes of ground delay for all those flights and the total number of airborne delay for all those flights and take the ratio of the two totals (ground to air). Call this ratio  $R(1500,1700)$ . Repeat this process for view times 1600 and 1700 to form  $R(1600, 1700)$  and  $R(1700, 1700)$ . Now average  $R(1500, 1700)$ ,  $R(1600, 1700)$  and  $R(1700, 1700)$ . This produces the entry 0.37 which is an average over all predictions for the 1700 hour of the relative roles that ground and air delay played in unfulfilled demand for that hour.

In the far right column of the SFO table, the entries in the row have been averaged over all days. The average in the lower right corner of the table is an average of the column above (over all hours). This number is an indicator of the relative roles that ground and airborne delay played in unfulfilled demand, averaged over the hours 1500 - 2900 and over all days.

|      | SFO    | SFO    | SFO    | SFO     | SFO     | SFO     | SFO     |          |
|------|--------|--------|--------|---------|---------|---------|---------|----------|
| hour | 2/3/98 | 2/6/98 | 2/7/98 | 2/12/98 | 2/14/98 | 2/19/98 | 2/26/98 | avg      |
| 1500 | 0.24   | 1.89   | 0.14   | 1.3     | 1.78    | 0.07    | 0.94    | 0.908571 |
| 1600 | 0.03   | 0.17   | 0.32   | 0.73    | 0.63    | 0.23    | 0       | 0.301429 |
| 1700 | 0.37   | 0.37   | 0.13   | 0.39    | 0.22    | 1.17    | 1.38    | 0.575714 |
| 1800 | 11.19  | 0.91   | 0.7    | 0.6     | 1.59    | 1.56    | 3.87    | 2.917143 |
| 1900 | 1.16   | 1.44   | 0.88   | 3.13    | 2.14    | 1.8     | 11.25   | 3.114286 |
| 2000 | 1.91   | 2.94   | 1.41   | 1.12    | 1.98    | 2.23    | 1.83    | 1.917143 |
| 2100 | 13.84  | 4.49   | 1.16   | 1.17    | 0.89    | 1.67    | 0.96    | 3.454286 |
| 2200 | 3.5    | 5.11   | 4.37   | 1.28    | 1.88    | 1.88    | 0.72    | 2.677143 |
| 2300 | 2.06   | 4.13   | 1.03   | 0.71    | 7.8     | 3.86    | 2.28    | 3.124286 |
| 2400 | 2.65   | 1.66   | 3.06   | 1.11    | 3.4     | 3.29    | 1.19    | 2.337143 |
| 2500 | 4.1    | 3.19   | 1.17   | 0.77    | 1.53    | 5.61    | 1.52    | 2.555714 |
| 2600 | 1.16   | 1.37   | 1.49   | 1.55    | 0.31    | 4.68    | 0.16    | 1.531429 |
| 2700 | 1.13   | 1.47   | 1.22   | 1.04    | 0.41    | 1.85    | 1.32    | 1.205714 |
| 2800 | 0.87   | 1.4    | 0.81   | 1.23    | 0.34    | 2.97    | 0.89    | 1.215714 |
| 2900 | 0.29   | 1.41   | 1.4    | 1.02    | 0.21    | 2.05    | 0.39    | 0.967143 |
|      |        |        |        |         |         |         |         |          |
|      |        |        |        |         |         |         | avg:    | 1.92019  |

|      | BOS     | BOS     | BOS     | BOS     | BOS    |          |
|------|---------|---------|---------|---------|--------|----------|
| hour | 2/19/98 | 2/20/98 | 2/21/98 | 2/24/98 | 3/3/98 | avg      |
| 1500 | 0       | 6.15    | 20.15   | 3.84    | 23.29  | 10.686   |
| 1600 | 10      | 11.09   | 0       | 6.27    | 11.74  | 7.82     |
| 1700 | 0       | 2.3     | 0       | 5.66    | 0      | 1.592    |
| 1800 | 2.75    | 0.42    | 3.57    | 5.43    | 1.86   | 2.806    |
| 1900 | 1.08    | 2.68    | 0.72    | 9.65    | 14.73  | 5.772    |
| 2000 | 1.76    | 0       | 0.31    | 3.84    | 2.45   | 1.672    |
| 2100 | 3.24    | 0       | 12.9    | 5.42    | 3.27   | 4.966    |
| 2200 | 9.98    | 1.98    | 3.53    | 3.18    | 1.92   | 4.118    |
| 2300 | 5.53    | 1.9     | 5.73    | 1.28    | 2.36   | 3.36     |
| 2400 | 4.23    | 2.46    | 0.84    | 0.49    | 18.26  | 5.256    |
| 2500 | 3.99    | 4.38    | 0.97    | 0.74    | 7.42   | 3.5      |
| 2600 | 0.94    | 1.85    | 11.87   | 0.96    | 1.1    | 3.344    |
| 2700 | 0.45    | 4.12    | 1.92    | 0.66    | 3.7    | 2.17     |
| 2800 | 0.67    | 3.04    | 0       | 8.09    | 0      | 2.36     |
| 2900 | 0       | 0       | 0       | 0       | 5.27   | 1.054    |
|      |         |         |         |         |        |          |
|      |         |         |         |         | Avg    | 4.031733 |