PREDICTIVE MODELS OF THE NATIONAL AIRSPACE SYSTEM USING BAYESIAN NETWORKS

Todd Lauderdale
NASA Ames Research Center

Keywords: *Bayesian Network, National Airspace System, Statistical Model*

Abstract

This paper focuses on a technique to model the National Airspace System which does not require selection of reference data and provides a framework to discover the complicated effects of congestion in one region of airspace on the operation of another. Since the interactions at this level are complex, only historical data are used in the analysis. From these data, a Bayesian Network model of the system is learned that is capable of predicting the number of aircraft in certain regions of the airspace at a given time with greater accuracy than similar linear regression models. The model is also used to automatically determine which regions of the system are the most critical to overall performance. The model is also used to automatically rank the regions of the system based on how important congestion in each region is to the overall performance of the system.

1 Introduction

It is important to be able to model the National Airspace System (NAS) and predict the effects of congestion. An increase or decrease in an airport capacity, or a change in the number of flights scheduled in a certain region, can affect the congestion in the system as a whole. If the results of these changes in the NAS can be predicted, steps such as reconfiguring the airspace or implementing traffic flow initiatives can be taken to ameliorate the adverse effects.

Many different modeling techniques have been developed in the past for this purpose. Some models are physically explicit trajectory-based models, [1] and [2], which calculate the trajectories of each aircraft in the system. These models can simulate congestion effects, but they require significant computational time and are not guaranteed to represent the behavior of the system. Other models are created from network-flow data and combine physical equations and analysis of historical data, [3], [4] and [5]. These models often do not take capacity and demand into account explicitly, instead relying upon selection of a reference day similar to the day being simulated. Still other models are empirical and purely data-driven, [6], [7], and [8]. These models are mostly used to predict delay instead of aircraft count, and some still rely on selection of an appropriate reference day.

This paper presents a graphical probabilistic model, known as a Bayesian Network, which is derived from historical data and provides estimates of the number of aircraft in regions of the NAS. Other researchers ([9] and [10]) used Bayesian networks in airspace system modeling to study detailed effects such as gate delay. Unlike some other modeling approaches, this model does not require the careful selection of a reference state. Capacity and demand are treated as fully independent variables, and this allows for more modeling flexibility to simulate congestion situations that have not yet been experienced by the system. The model also provides a framework to automatically discover from the data the effects of congestion in one region on the operation of the entire system.

This model and the data used to produce it

are discussed in Sections 2 and 3. The two steps required to learn the model from the data are discussed in Section 4. Results of the model showing predictions for certain days and a comparison of the model with similar linear regression models are presented in Section 5. Finally, conclusions are presented in Section 6.

2　Model of a Center

The NAS is divided into 20 separate regions known as Air Route Traffic Control Centers. Here, they will be referred to as "centers." The purpose of the model described in this paper is to predict the number of aircraft that will be in each center at a given time. A brief introduction to Bayesian Networks will first be presented, and then a Bayesian Network model to predict the number of aircraft in a center will be discussed.

2.1　Bayesian Networks

Bayesian networks are probabilistic graphical models where a node in the graph represents a variable (such as the number of aircraft or the capacity of the area) and an edge between nodes represents an interaction between variables (like the fact that the number of aircraft is dependent on the capacity). Associated with each node are parameters representing the probability of the states of the node given the states of its parents. The Bayesian network represents a joint probability distribution of all variables.

In general, there are two parts of a Bayesian network which must either be determined from the data or assigned: the structure of the graph and, given a particular graph, some parametric representation of the probability of the state of a node given the states of the parent nodes. Advantages of the methodology used to create Bayesian Networks are that it allows for the graph structure to be discovered automatically from the data, instead of assigned *a priori*, and it allows for flexibility to add other variables if desired.

2.2　Network Center Model

The number of aircraft in a center is a function of the time histories of the arrival demand at airports in the center, the departure demand at airports in the center, the number of over-flights and the airport capacities in the center. From this basic premise a simplified model of the center is assumed by only taking into account the states at the current time. In this simplified model, the number of aircraft is given by the scheduled arrivals, scheduled departures, and capacities of all the airports in the center. Because of the different operating states of the system, another important factor to be included is the time of day. A graphical representation of the Bayesian Network associated with this simplified model is shown in Figure 1, and this assumed model can be extended by adding links to other centers.
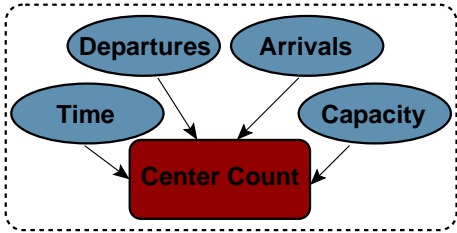

Fig. 1 Representation of the basic Bayesian network used to model a center.

Once this model is learned, it will be capable of predicting the number of aircraft in each center at a certain time given the airports' scheduled arrivals and departures and the expected capacities for the airports. The methods used to learn the model will be discussed in Section 4.

3　Data

The model requires historical data which describe four quantities in epochs of 15 minutes for the entire NAS: the scheduled arrivals at an airport, the scheduled departures at an airport, the arrival and departure capacity of an airport, and the observed number of aircraft in a center. These data can be gathered from three different sources. The airport schedule data for commer-

cial aircraft can be found in the Bureau of Transportation Statistics (BTS) Airline On-Time data. This dataset provides the scheduled and actual arrivals and departures for flights of the commercial airlines which report their data to the BTS. The airport capacity data are collected by the Federal Aviation Administration (FAA) for the top 75 airports in the country and made available through the Aviation System Performance Metrics (ASPM) database. The number that is used as the capacity for each airport is the sum of the airport acceptance rate and the airport departure rate found in that database. Finally, the number of aircraft in the center is determined from the Aviation Situation Display to Industry (ASDI) dataset. The ASDI data are processed using Future ATM Concepts Evaluation Tool (FACET), [2], to count the number of aircraft in each center every 15 minutes.

Since much of the data are collected at the airport level, these data must be aggregated by center. Thus, the scheduled departures for all airports in each center are summed to create a center-wide number referred to as the scheduled center departures. This aggregation is also used to create the scheduled center arrivals, and the capacities of all airports in the center are summed to create the center capacity.

All data were collected and processed for January 1, 2005 to June 30, 2007. Three separate groups of data were then formed. Approximately half of the data were grouped in one database for model parameter learning. A quarter of the data were grouped in a second database for model scoring, and the remaining quarter of the data were reserved for model evaluation and comparison.

4   Model Learning

The procedure used to create the model from the data will now be discussed. An outline of the procedure discussed in the following sections is: (1) a graph is assumed; (2) partitions to discretize the data are assumed; (3) the data are divided into groups based on the graph and which partitions they lie in; (4) points are added to the data to

smooth the data and reduce noise; (5) a probability density function is fit to each group of data; (6) the model is scored using reserved data; (7) changes are made to the graph or the partition; and (8) start at step three and iterate until satisfied with the results.

This procedure can be logically divided into two parts. The first part of the process, known as parameter search, is to learn the parameters of each probability density function given a particular graph. This corresponds to steps three to six. The second part of the process, known as structure search, is change the structure of the graph to better capture the interactions in the system. This corresponds to steps seven and eight.

The model of a single center shown in Figure 1 is simple in the sense that the number of aircraft in a center is only a function of the variables of that center. In the NAS, this is not the case. The effects of a capacity constraint in one center can propagate to other centers by various mechanisms. To capture these effects graphically, links from other centers to the current center can be included (see Figure 2) using structure search.
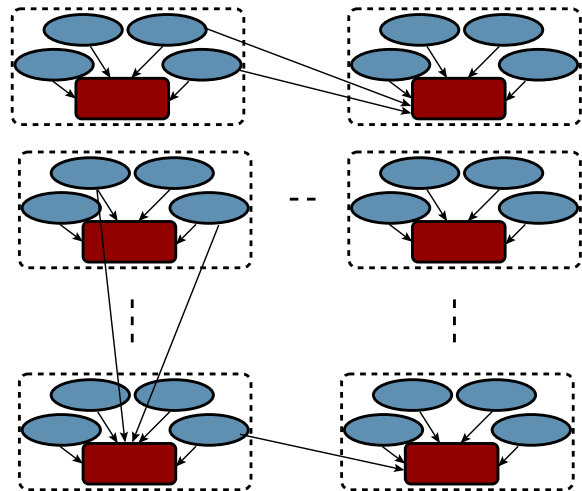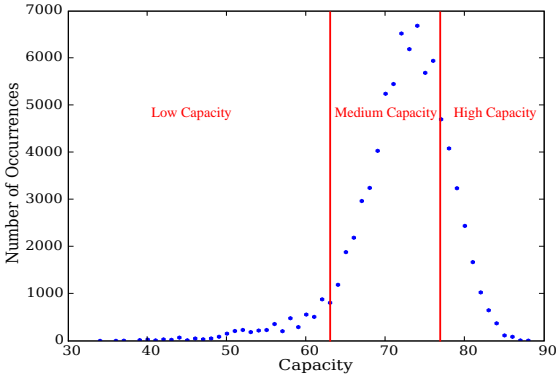

Fig. 2 A complex Bayesian Network which includes dependencies between centers.

The specific form of structure search used in this paper is described in Section 4.2 and is important for improving the predictions of the model. It can also allow for relative ranking of

links. For instance, this ranking can be used to determine whether a reduction of the capacity of New York Center (ZNY) has a larger effect on the results of Boston Center (ZBW) than a reduction of capacity of Minneapolis Center (ZMP).

4.1 Parameter Search

The steps necessary to learn the probability relationships between the variables for any given graph such as the one shown in Figure 1 will now be discussed. The first step is to discretize the data by grouping them into bins. In subsequent steps, the bin boundaries will be determined from the data, but the initial boundaries can be picked arbitrarily. For example, as shown in Figure 3, the capacity of the center can be divided into three separate bins. Capacity values below 62 are grouped together in the first bin. Capacity values between 62 and 76 are grouped together into the second bin, and capacity values above 76 are grouped in the third bin. Bins for the scheduled arrivals, the scheduled departures, and the time of day are also created. The number of partitions and the location of these partitions are important parameters in the model. How these variables are refined will be discussed later.

sumption of what the probability of the center count will be, and it is used for Bayesian learning of the parameters, which smoothes the data and minimizes noise. To gain insight into a reasonable form of this prior, the total distribution of center counts for all data points aggregated across all bins and all times is plotted in Figure 4. This distribution is bimodal. The two peaks correspond roughly to nighttime operations and daytime operations. To create a reasonable prior for the data, these two states must be differentiated between. This is accomplished by using the scheduled arrivals, because the number of scheduled arrivals during nighttime operations is usually much lower than daytime operations. The distribution for low scheduled arrivals is shown in Figure 5(a) and the distribution for high scheduled arrivals is shown in Figure 5(b). For the nighttime prior, a log-normal probability distribution is fit to the data, and for the daytime prior a Gaussian distribution is fit. Since these prior distributions are only used to condition the data, it is not necessary that they fit the data perfectly.

Fig. 3 The distribution of capacities for New York Center and an example of bin partitions.
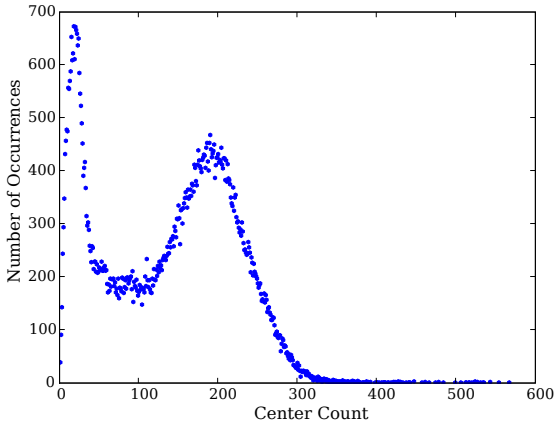
Fig. 4 The distribution of center counts for all data points in the parameter learning dataset for New York Center.

Another important step in the process of learning the model is to determine a distribution that gives a prior probability for the data. This prior probability density function is an as-

With the data partitions and priors in place, a probability density function (pdf) for the number of aircraft in the center for each possible combination of bins is learned. For example, imagine that there are 2 bins (high and low) for each of the

(a)   (b)
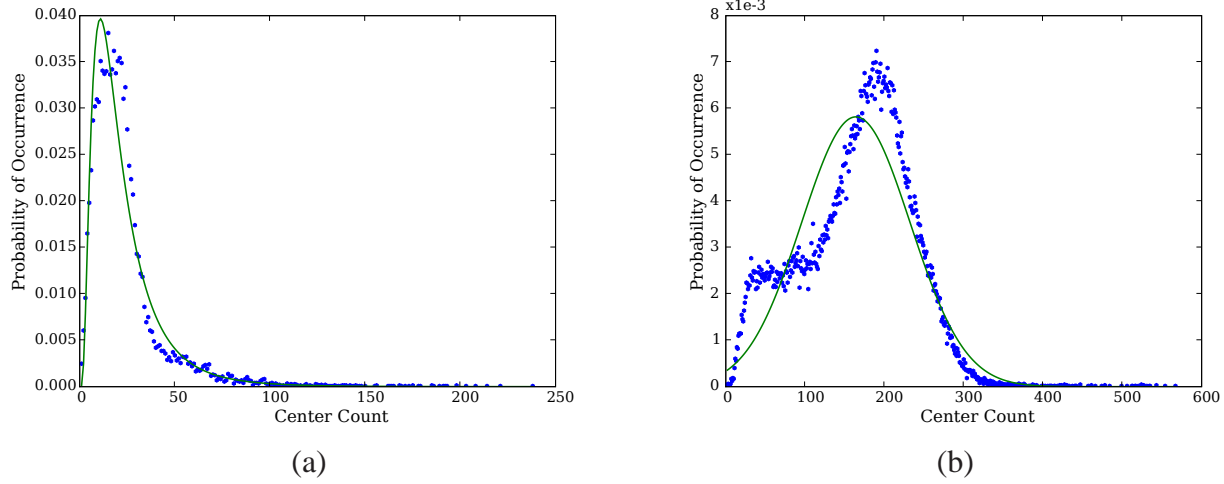
Fig. 5 The New York Center counts and fitted distributions for (a) low arrival demand and (b) high arrival demand.

four input variables. Then a pdf for each combination of bins must be learned, for a total of 16. Each pdf is known as a conditional probability distribution, since it is dependent on the states of the input data.

For a specific combination of bins, all of the data points in the parameter learning database associated with those bins are selected. To incorporate the data from the prior distributions, points sampled from the appropriate prior are added to the data. Figure 6 shows an example of a pdf learned from a specific combination of bins in the actual data with the points sampled from the prior included. The pdf function fit is a mixture of a log-normal distribution and a Gaussian distribution:

$$X = \alpha N(\mu_1, \sigma_1^2) + (1 - \alpha) Log - N(\mu_2, \sigma_2^2). \quad (1)$$

In this equation $N(\mu_1, \sigma_1)$ is a Gaussian distribution with mean μ_1 and standard deviation σ_1, and $Log - N(\mu_2, \sigma_2^2)$ is a log-normal distribution where μ_2 and σ_2 are the mean and the standard deviation of the logarithm of the points in the distribution. Also, the parameter α provides a relative weight between the Gaussian and log-normal portions of the mixture. The five parameters of the mixture, σ_1, μ_1, σ_2, μ_2, and α, are determined by maximizing the likelihood of the data given the mixture parameters.

It is important to note that it was determined that the center count is not distributed in a Gaussian fashion. This is especially true when the demand on the system is low and thus the total number of aircraft in the system is low, but this effect can be seen even when the center count is higher. The effect of the mixture in the pdf is illustrated by the peaked and asymmetrical nature of the curve in Figure 6. Similarly, the log-normal nature of the data at low demand is shown in Figure 5(a).
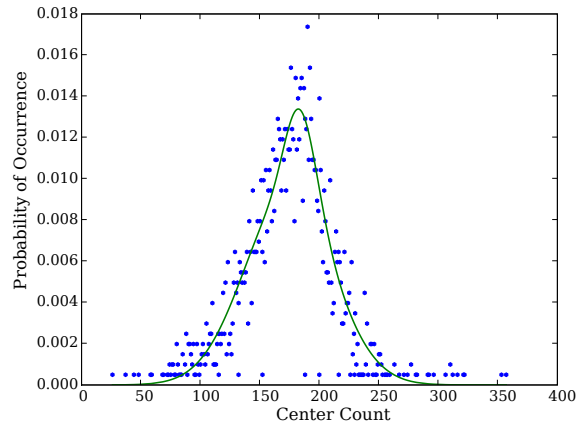
Fig. 6 An example of the distribution learned for New York Center for evening with medium capacity and high arrival and departure demand.

Once the pdf for each combination of bins is learned, the model can be scored using the data from the model scoring database. The score is the log likelihood of the data in the model scoring database given the Bayesian network learned in the parameter learning step. This is given by the equation

$$LL = \sum_{i=1}^{n} \ln P\left(x_i.c | x_i.i, \mathcal{M}\right), \qquad (2)$$

where x_i is a point in the scoring database which center count $x_i.c$ and other input data $x_i.i$. Also, $\mathcal{M}$ represents the current Bayesian Network, and n is the number of points in the scoring database. The parameters of the model, such as the location of a bin partition or the number of partitions, can be varied and the resulting log-likelihood indicates whether the change is good or bad for the model.

In practice, learning the locations of the partitions is accomplished by optimizing each partition type independently from the other partitions. So, to determine the number and locations of the time partitions, the partitions of all other variables are held fixed. The time partitions are then varied and scored using the structure learning dataset. The partition with the best score is then selected for the model. Using this optimization, most centers have three time bins, five arrival and departure bins, and three capacity bins. A more complete optimization obtained by varying the partitions of each variable together may result in a better model.

4.2 Structure Search

The second part of the learning process is known as structure search, and it involves determining which nodes in the graph should be connected. In this paper, the structure search will be used to determine from the data where congestion has the most effect on the operation of the NAS. For this purpose, the basic graph for Center B, shown in Figure 1, is used as a starting point and a link to one node of another center model is added to create a graph such as the one shown in Figure 7. The parameters of the graph are then

learned using the parameter learning dataset and the procedure described in the previous section. This graph is then scored using the model scoring dataset, and this score can be compared against the original score to determine whether the link was beneficial to the model. If the link improves the score, there is a statistical relationship between the newly connected nodes.
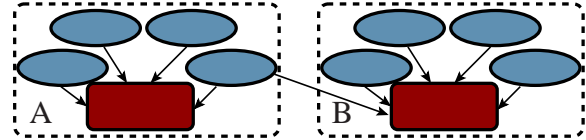
Fig. 7 Including a single link from an input of Center A to the output of Center B.

This method is useful because it not only determines the best links for the graph, but it can also provide a relative ranking between the links. As discussed in Section 5.3, by ranking the relative importance of a link between one center and each other center in the system the model can be used to predict where in the system that capacity reductions or demand increases have the most effect on the rest of the system.

5 Results

After learning the model, there are many different ways to interpret the output. The model can be run on large sets of data for evaluation. The model can also be used to predict what will happen for given inputs at a center level or system wide. The results of changes in demand profiles or capacity reductions can also be predicted. Also, the model can be used to rank the relevance of centers.

5.1 Center and System Wide Predictions

The prediction capabilities of the model at the system and center level will be discussed first. The Bayesian network model used for these results is the base model shown in Figure 1. The quality of the results can be analyzed using multiple metrics such as the root mean squared error

of the predicted flight count or the probability of the test data given the model. These metrics will be compared against two linear regression models.

The model can be used to predict two different aspects of the number of aircraft in the center for any given input state. First, it can predict the number of aircraft in the center, and second, it can predict the probability of a certain number of aircraft in a center. The count prediction is performed by taking the expected value of the pdf associated with the input bins that the data point lies in. Since the pdfs that were fit to the data are mixtures of log-normal and Gaussian distributions as given in Equation 1, the expected value is

$$x_{exp} = \alpha\mu_1 + (1-\alpha)e^{\left(\mu_2 + \sigma_2^2/2\right)} \qquad (3)$$

where α, μ_1, μ_2, and σ_2 are the parameters of the mixture for that bin.

Once the estimate for each point in the evaluation set is created, the data can be plotted on a chart showing the predicted value versus the actual value. These plots can be created for all points in the evaluation dataset for each center or for all centers together. If the predictions were perfect, all data would lie on a line with unit slope and zero y-intercept. Figure 8 shows the results for all centers combined. As seen in the figure, the data are somewhat tightly distributed about the line which indicates that the model provides a good estimate for the actual system. A measure of the variation of the data from the line is the root mean squared (RMS) error. The data in Figure 8 has a RMS error of 37. This error was better than the RMS error for the linear regression models discussed subsequently and given in Equations 4 and 5. For all of the evaluation data the average flight count is 160 aircraft. So, the RMS error divided by the average flight counts is approximately 23%.

Figure 9 shows the RMS error for each center divided by the average number of aircraft in that center. The prediction errors are fairly consistent around 23% with the maximum error being 29% for Boston Center (ZBW) and the minimum error being 16% for Los Angeles Center (ZLA).
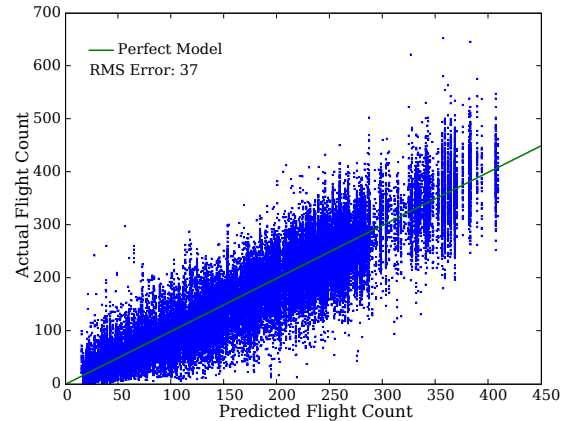

Fig. 8 The actual values compared with the predicted values for all of the evaluation data.
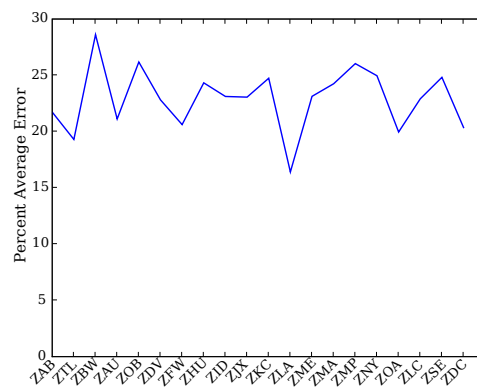

Fig. 9 The RMS error of the prediction divided by the average flight count for each center.

To determine how well the method works, it is compared to two linear regression models. Linear regression is also a Bayesian network, but in the model the center count is assumed to be a Gaussian distribution with a continuous linear dependence on the input parameters. The two linear regression models used as benchmarks are one using only first order dependence on the four input parameters:

$$cc = \gamma_1 n_a + \gamma_2 n_d + \gamma_3 n_c + \gamma_4 t + \gamma_5, \quad (4)$$

and one with dependence on the input parameters up to the fourth order:

$$cc = \sum_{i=1}^{4} \left(\gamma_{1i} n_a^i + \gamma_{2i} n_d^i + \gamma_{3i} n_c^i + \gamma_{4i} t^i \right) + \gamma_5, \quad (5)$$

where cc is the center count, n_a is the number of arrivals, n_d is the number of departures, n_c is the capacity, t is the time of day (in coordinated universal time), and the γ constants are determined by performing regression on the data in the parameter learning database. The model including up to 4th order terms is the most accurate linear regression model found.

The comparison of the models is accomplished by comparing the log of the probability of the evaluation data given the model which is given in Equation 2, where in this case the data are from the evaluation database and the models can be either Bayesian Networks or linear regression models. Basically this metric says how well the model fits the evaluation data, with the ideal model having a probability of one giving a log-likelihood of zero.

Using the log-likelihood of the Bayesian model as a baseline, the percent increase of the Bayesian model over the other two models is computed using the formula:

$$\%_{increase} = \frac{LL_{reg} - LL_{Bayes}}{LL_{reg}}. \quad (6)$$

Figure 10(a) shows the average over all centers of the percent increase in log-likelihood of the Bayesian model over the log-likelihood of the two regression models. The Bayesian model has

approximately a 9% increase in log-likelihood over the 1st order linear regression model and it has approximately a 4% increase over the best linear regression model which included up to 4th order dependence on the input variables. Figure 10(b) shows this same comparison at the center level.
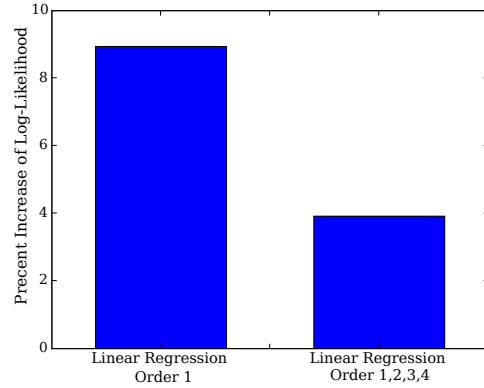

(a)
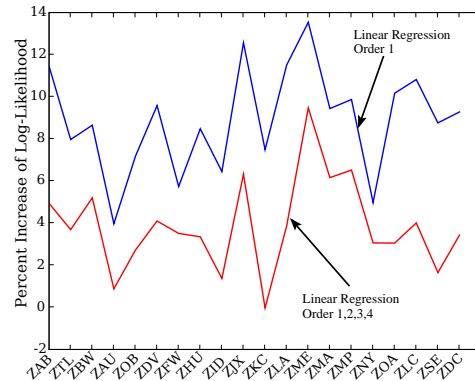
(b)

Fig. 10 A comparison of the percent increase of the log-likelihood of the data given the Bayesian model with two linear regression models for: (a) the average log-likelihood increases for all centers and (b) for each center.

5.2 Daily Predictions

In addition to looking at the results in aggregate it is illustrative to look at a prediction for a continuous period of time. Figure 11 shows a comparison of system-wide traffic counts for two 24-hour time periods and the predictions for those

periods. The solid lines show a Tuesday in the NAS, with the red line showing the actual flight count and the blue line showing the prediction. The dashed lines correspond to a Sunday. The variations in the predictions for the two different days show that the model is responding to the changes in demand between the two days.
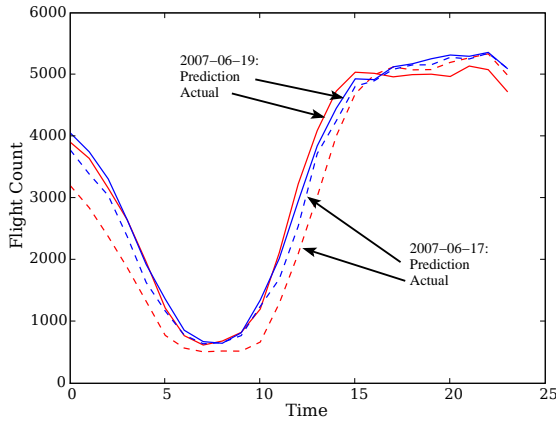
Fig. 11 System-Wide Prediction for a weekday (6/19/07) and a Sunday (6/17/07).

Figure 12(a) shows predictions for two days in New York Center. The solid lines show a bad weather day, and the dashed lines show a good weather day a week later. The prediction for the bad weather day is different than the good weather day. This shows that the model is responding to capacity restrictions. However, the prediction is not as low as the actual count. This is because the day is an outlier, so it would be difficult for a statistical model to capture the total effect.

As discussed earlier, another output of the model is the probability of any center count occurring given a particular input. This probability can be used to determine a confidence in the expected value prediction, which is similar to the probabilities of a weather forecast occurring. The confidence, *con*, is given by
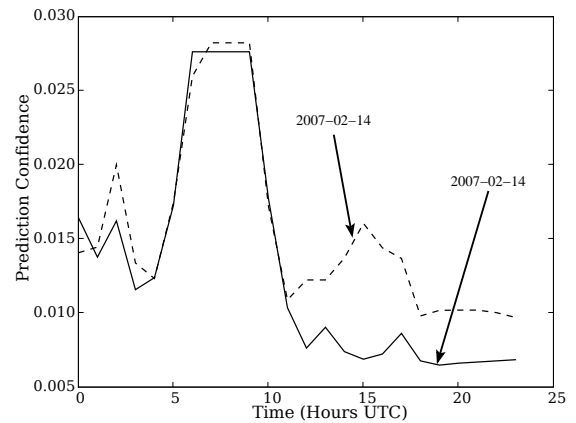
$$con = P(x_{i,exp}|x_i.i, \mathcal{M}), \qquad (7)$$

where $x_{i,exp}$ is the expected value, $x_i.i$ is the input data, and $\mathcal{M}$ is the model. Figure 12(b)

shows that the confidence in the bad weather day prediction is lower than the confidence in the good weather day prediction. Thus, even though the prediction is not as accurate as for the good weather day, the model will indicate that as well.
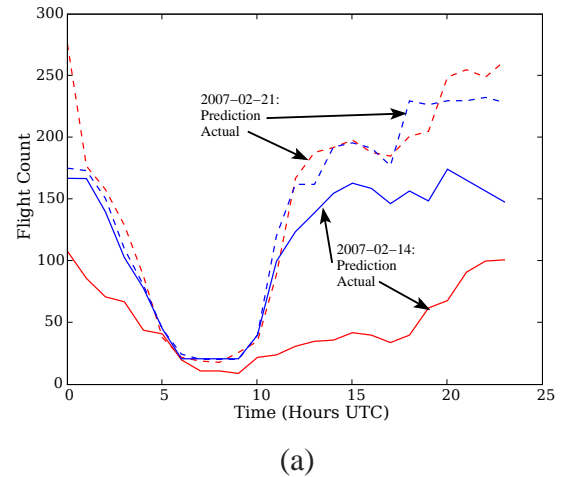
(a)

(b)

Fig. 12 (a) A comparison of the predicted and actual flight counts for a bad weather day and a good weather day in New York Center. (b) The confidence in the prediction for each day.

5.3 Relevance of Links

The previous results dealt with the basic center model shown in Figure 1. Next, structure search will be used to uncover the dependencies in the system. To determine the relevance of inputs of one center on the results of another, the model

for a graph like the one shown in Figure 7 where Center B is held fixed is learned. If adding this link to the graph improves the score for predictions in Center B, then that node of Center A affects the results of Center B. Then all other centers, shown as Center A, are iterated through and the resulting model is scored. After iterating through all centers they can be ranked by how much they improved the model for center B.

The first link analyzed connects the demand node of Center A to the center count of Center B. Demand is defined to be the sum of the scheduled arrivals and the scheduled departures. Every center in the NAS is iterated through as Center A and as Center B, and the score for each graph is recorded. Then, for each center the importance of the demand in every other center can be ranked using these scores with the caveat that, if a link from Center A does not improve the score of Center B, then it is assigned a ranking of 20. These rankings can then be summed for the demand of all centers. For example, if New York Center (ZNY) is ranked 4th for every center and each of those links improves the score of the other center, then the aggregate ranking of ZNY will be 80. If ZNY was ranked 4th in every center, but it only improved the score for 10 of them, then it's aggregate ranking would be 240.

In Figure 13 the result of performing this model learning and ranking for the demand of all 20 centers in the continental United States and summing each center's rank is shown. These scores are then normalized by subtracting the ranking of the lowest ranked center. Using this methodology, centers with lower ranking can be considered to more important to the system with respect to their demand. Interestingly, from the rankings in Figure 13 centers such as Fort Worth Center (ZFW) and New York Center (ZNY) with busy airports such as DFW airport in Dallas and New York's JFK airport rank as important, while centers that are generally considered less connected such as Seattle Center (ZSE) or Miami Center (ZMA) are ranked as less important.

The process outlined above for ranking demand links is repeated for ranking capacity links. The results of this analysis are shown in Figure
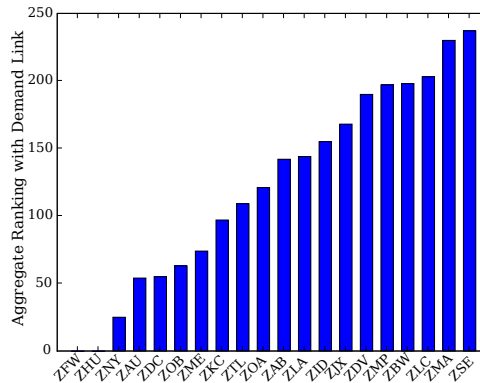

Fig. 13 The normalized sum of the rankings for demand.

14. Again, lower ranked centers are considered to be more relevant to the system. The centers that are generally considered to be more important to the system such as ZNY, Chicago Center (ZAU), and Atlanta (ZTL) are again ranked as important by the model. Thus the model is capable of learning from the data which centers are critical to the operation of the NAS.
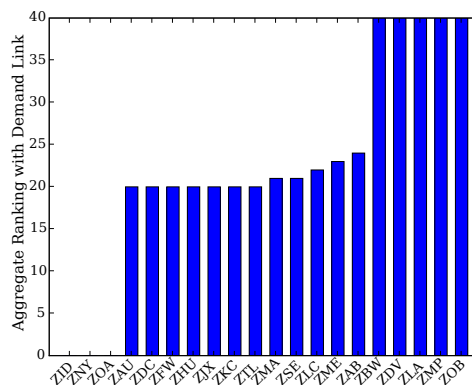

Fig. 14 The normalized sum of the rankings for capacity.

6 Conclusions

In this paper, a Bayesian network model was automatically created from historical data. The model is capable of predicting the number of aircraft in a center with a total RMS error of 37 air-

craft using only scheduled airport demand, airport capacities, and the time of day.

The model performed better than linear regression models using the same inputs. It also was shown to be responsive to demand fluctuations and capacity reductions. So, using the model it is possible to predict what the future number of aircraft will be and to know the relative confidence in that prediction. This model is coarse in that large amounts of data are grouped into large spatial regions, but it shows promise for the method. A model with more detail, such as one using the same techniques but modeling the data at the sector level, might provide more actionable results.

Finally, the model was used to automatically discover from the data a ranking of the importance of the demand and the capacity of the centers in the NAS. This ranking roughly corresponds to the generally accepted importance of centers in that ZNY, ZAU and ZTL appear near the top of the rankings.

7   Copyright Statement

References

[1] L. Meyn, R. Windhorst, K. Roth, D. Van Drei, G. Kubat, V. Manikonda, S. Roney, G. Hunter, and G. Couluris. Build 4 of the Airspace Concept Evaluation System. In *AIAA 4th Aviation Technology, Integration and Operations (ATIO) Forum*, 2001.

[2] K. Bilimoria, B. Sridhar, G. Chatterji, K. Sheth, and S. Grabbe. FACET: Future ATM Concepts Evaluation Tool. In *Proceedings of the 3rd USA/Europe ATM 2001 R&D Seminar*, 2001.

[3] D. Sun, I. Strub, and A. Bayen. Comparison of the performance of four eularian network flow models for strategic air traffic flow management. *Networks and Heterogeneous Media*, 2(4):569–594, 2007.

[4] B. Sridhar, T. Soni, K. Sheth, and G. Chatterji. An aggregate flow model for air traffic managment. In *AIAA Guidance, Navigation and Control Conference and Exhibit*, 2004.

[5] I. Strub and A. Bayen. Optimal control of air traffic networks using continuous flow models. In *AIAA Conference on Guidance, Control and Dynamics*, 2006.

[6] B. Sridhar and S. M. Swei. Relationship between weather, traffic and delay based on empirical methods. In *6th AIAA Aviation Technology, Integration and Operations Conference (ATIO)*, 2006.

[7] N. Xu, L. Sherry, and K. B. Laskey. Multifactor model for predicting delays at U.S. airports. In *87th Transportation Research Board Annual Meeting*, 2008.

[8] D. A. Smith, L. Sherry, and G. Donohue. Decision support tool for predicting aircraft arrival rates, ground delay programs, and airport delays from weather forcasts. In *International Conference on Research in Air Transportation (ICRAT-2008)*, 2008.

[9] K. B. Laskey, N. Xu, and C-H. Chen. Propagation of delays in the national airspace system. In *22nd Conference on Uncertainty in Artificial Intelligence*, 2006.

[10] N. Xu, K. B. Laskey, C-H. Chen, S. C. Williams, and L. Sherry. Bayesian network analysis of flight delays. In *Transportation Research Board 86th Annual Meeting, Washington, D.C.*, 2007.