Guidance, Navigation, and Control and Co-located Conferences
August 19-22, 2013, Boston, MA
AIAA Guidance, Navigation, and Control (GNC) Conference

AIAA 2013-5030

An Enhancement to the Linear Dynamic System Model for Air Traffic Forecasting

Yi Cao*
School of Aeronautics and Astronautics, Purdue University. West Lafayette, IN 47907, USA

Lingsong Zhang†
Department of Statistics, Purdue University. West Lafayette, IN 47907, USA

Dengfeng Sun‡
School of Aeronautics and Astronautics, Purdue University. West Lafayette, IN 47907, USA

Air traffic forecast is among the most important functionalities of air traffic controls. The Linear Dynamic System Model that predicts the traffic demand within the Air Route Traffic Control Centers serves well for this purpose. This model formulates inflows and outflows between Centers' boundaries by assuming that the boundary crossings between Centers are in conformance with Poisson distribution and that the traffic patterns do not change too much over multiple days. As a result, the traffic in the near future can be predicted based on knowledge of historical traffic patterns and anticipated departures. As a predictive model, its prediction accuracy relies heavily on parameter estimation. In the earlier implementations of this model, the traffic patterns are obtained by averaging estimations of multiple days. However, given the uncertainties in the traffic system and the deficiencies inherent in the radar track data, using observed traffic data of a few days to train the parameters is likely to result in bias due to limited samples. A large training set on the other hand contains a lot of noise, to which the mean is susceptible. This paper introduces the Kernel Density Estimation into the Linear Dynamic System Model. This non-parametric method serves as an enhancement to the model in that it is able to capture the major transition patterns from a large data set in the presence of outliers and data deficiencies regardless of the actual distribution of data. Therefore, this statistical approach is useful in extracting normal traffic patterns that are representative of major behavior of the traffic flows. An one-month traffic simulation shows that, by incorporating with the Kernel Density Estimation, the Linear Dynamic System Model reduces the estimation errors by 20% on average.

Nomenclature

N	Total number of Centers
M	Total number of data samples
$s_{i,j}(t)$	Accumulative number of aircraft transitioning from Center i to Center j at time t
T	Time horizon
ΔT	Time interval for system model evolution
$x_i[k]$	Aircraft count in Center i at time step k
$u_i[k]$	Departures in Center i at time step k
$\beta_{i,j}[t]$	Transition probability describing outflow from Center i to Center j at time step k
$\lambda_{i,j}(t)$	Rate of transition intensity from Center i to Center j at time t
τ	Time interval for counting the number of boundary-crossings
Superscript	
d	Date index

*Ph.D Student, School of Aeronautics and Astronautics, Purdue University. West Lafayette, IN 47907, USA. Student member, AIAA. cao20@purdue.edu.
†Assistant Professor, Department of Statistics, Purdue University. West Lafayette, IN 47907, USA. lingsong@purdue.edu.
‡Assistant Professor, School of Aeronautics and Astronautics, Purdue University. West Lafayette, IN 47907, USA. Member, AIAA. dsun@purdue.edu.

Subscript

i, j	Center index
k	Discrete time step
t	Continuous time

I. Introduction

The National Airspace System (NAS) in the United States is monitored by the Air Traffic Control System Command Center, which is responsible for balancing the demand with system capacity in a safe and efficient manner. Due to the ever-increasing demand in air transportation system in the past decade, air traffic management (ATM) calls for development of more accurate and robust tools for traffic demand forecasting. In ATM, the aircraft count in a defined region is usually used as an indicator of busyness in that airspace control volume. For example, the maximum aircraft count that the Air Traffic Controllers can safely manage within a *sector* is referred to as the *Monitor Alert Parameter* (MAP),[1] which varies over time to reflect the change of control complexity or change of weather throughout the day. Comparing the aircraft count forecast with MAP values improves the situational awareness of facility personnel. Similarly, the aircraft count in the *Air Route Traffic Control Center* (ARTCC, or simply *Centers*), generally termed Center count, is also very important in that the numbers provide a high level view of traffic and trigger appropriate management actions in response to the looming imbalance between demand and capacity. Therefore, the projected Center count in a strategical timeframe is of particular interest to the control authority.

Tools being in use for air traffic forecasting include the Enhanced Traffic Management System (ETMS), the Future ATM Concept Evaluation Tool (FACET), and the Center-TRACON Automation System (CTAS),[2–4] all of which employ trajectory-based approaches for traffic forecasting. Bayen et al. developed a sector-based traffic flow model based on hybrid automata theory.[5] The model establishes a kinematic equation for each aircraft, so it is essentially trajectory-based. Given the computational workload, this predictive model is typically used for modeling traffic within a small region like sector or the Terminal Radar Approach Control (TRACON). A NAS-wide model, the Cell Transmission Model, was later developed by Sun et al, which aggregates en route traffic into flows according to the origin and destination pairs.[6] The travel times across sectors were assumed to be the mean values calculated based on historical data. This model was further refined and extended to a lower order version.[7] The main drawback associated with the trajectory-based models is the increased cumulative errors along the length of look-ahead time. Generally, the prediction accuracy is guaranteed only for a short timeframe, such as twenty minutes. To avoid sophisticated formulation and computational workload, researchers have been actively examining the statistical properties of the traffic system at different aggregation levels. Menon et al. were among the first to model the air traffic in a framework of longitude-latitude tessellation using the Eulerian approach,[8,9] which inspired great interest in the ATM research community. This model established flow dynamics between longitude-latitude grids based on flow conservation principle. A lower order Eulerian approach, the Linear Dynamic System Model (LDSM), was proposed by Sridhar et al,[10] which is a time-variant version of a dynamic stochastic model developed in Ref. [11]. The aircraft count is aggregated at the Center level. With fewer state variables, Eulerian models are more computationally efficient. In addition, Eulerian models focus on flow property instead of individual flight dynamics. As a result, the predictions are more accurate at a macroscopic level in statistical terms.

The Eulerian models must be parameterized using historical data. For example, the key parameter of the LDSM is the transition probability, which characterizes the flow rates between Centers during a given time interval. It is assumed to be the mean flow rate found in the radar records over past two or three days. This approach is easy but not reliable, since the underlying assumption is that the traffic patterns of consecutive days are similar. However, this is not necessarily true given the unpredictable nature of the traffic system. If the reference traffic pattern happens to be an anomaly, using abnormal pattern of the other day to predict the current traffic would potentially result in unacceptable errors. Generally, the influence of anomalies can be minimized by averaging data samples with sufficiently large size, but the side effect is that large data set also increases the chance of wide spread of data. When the data distribution is multimodal, the mean maybe not a representative value to the quantity under estimation. This is very likely in the context of traffic flows. Inspired by Sridhar's work, this paper introduces a more reliable approach, the Kernel Density Estimation (KDE), to enhance the LDSM. With the enhancement, the traffic patterns are estimated using traffic data of multiple days. Unlike the mean approach that equally weighs each data sample, the KDE takes the value corresponding to the mode of the underlying distribution of data. As such, this approach is resistant to outliers. When a large data set is used, the traffic pattern estimated by the KDE is the most typical one observed in the historical data, functioning as the backbone of the LDSM. Although the KDE does not always outperform the mean approach for each day of traffic, especially when the traffic deviate from the normal pattern, statistically, the KDE is still more accurate and reliable in a long timeframe, since abnormal patterns are only transient states and the system has a greater chance of being normal.

This paper is structured as follows. The Linear Dynamic System Model is first reviewed in Section II. Then the Kernel Density Estimation is introduced in Section III. Simulations for a full month traffic are presented in Section IV, where traffic predictions yielded by the LDSM based on the KDE and the mean approach respectively are compared. Concluding remarks and future works are provided in Section V.

II. Review of the Linear Dynamic System Model

The NAS is a structural airspace system with low, high, and super-high altitude layers. In horizontal directions, each layer is divided into Centers. The Linear Dynamic System Model is based on the sectorized airspace. But for simplicity, the low and super-high altitude Centers are often ignored, and traffic of all altitudes are mapped to the high altitude Centers. However, the modeling approach introduced in this paper is also applicable to airspace with more sophisticated structural details. Moreover, only airspace in the continental Unites States is of interest in this paper, thus the system consists of twenty high altitude Centers in the United States and an International airspace that covers the rest airspace outside of the continental United States, as shown in Figure 1. The number of Centers determines the dimension of the system model. In this case, it is $N = 21$.

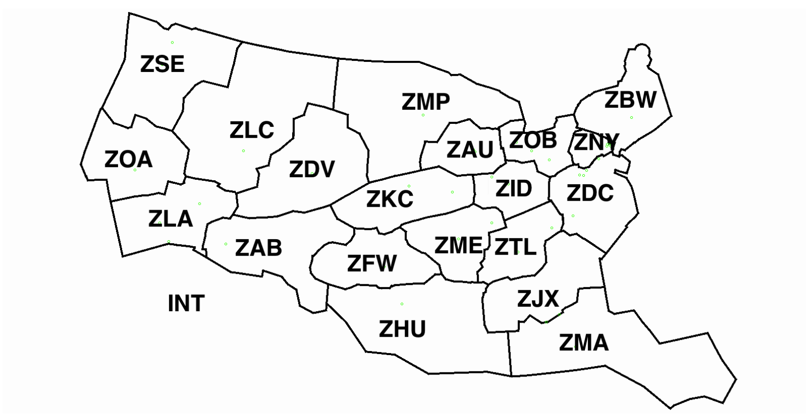


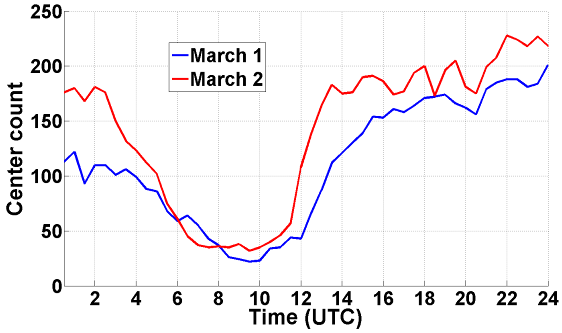
Figure 1. High altitude ARTCCs in the continental United States airspace, the twenty Centers and an International airspace.

The LDSM is an aggregate model that formulates traffic flows between Centers. Suppose T is the prediction time horizon with ΔT being the system evolution time interval, typically $\Delta T = 10$ *minutes*. The aircraft count $x_i[k]$ in Center i is the state variable that the model is trying to predict. Aircraft traveling through Centers forming inflows and outflows at the Center boundaries. The dynamics of the LDSM stems from the flow conservation principle, described by the following discrete-time linear system:

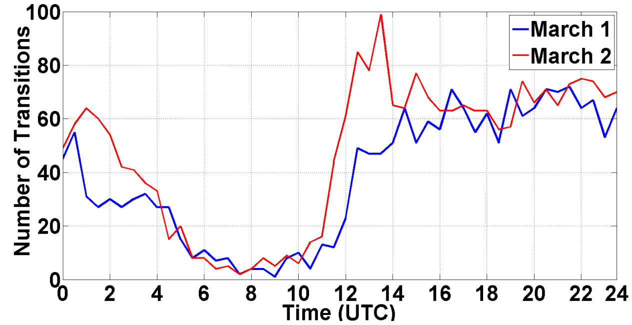
$$x_i[k+1] = x_i[k] - \sum_{j=1}^{N} \beta_{i,j}[k]x_i[k] + \sum_{j=1,j\neq i}^{N} \beta_{j,i}[k]x_j[k] + u_i[k], \quad i \in \{1,2,\cdots,N\} \tag{1}$$

where $u_i[k]$ is the departure into the airspace of Center i at time t. $\beta_{i,j}[k]$ is the transition probability describing the fraction of aircraft in Center i flying to the neighboring Center j from time step k to time step $k+1$. The arrivals landing in Center i is formulated as a special outflow $\beta_{i,i}[k]x_i[k]$. Equation (1) is quite self-explanatory, the aircraft count in the next time step is the sum of the current state, new departures into the Center, and the difference between inflows and outflows in ΔT. As civil flights usually file their flight plans with the Civil Aviation Authority two or three hours prior to departure, thus the departure $u_i[k]$ is a known input to the model. Once the transition probability $\beta_{i,j}[k]$ is determined and an initial state is given, the system model will evolve to generate a time history of the Center counts in an iterative manner. Sridhar et al. proposed to incorporate an additive Gaussian variable in the formulation to account for the uncertainties caused by unregistered departures or flight plan changes.[10] As the focus of this study is to examine the effectiveness of KDE in parameter estimation, we leave the modeling of uncertainties to a future study.

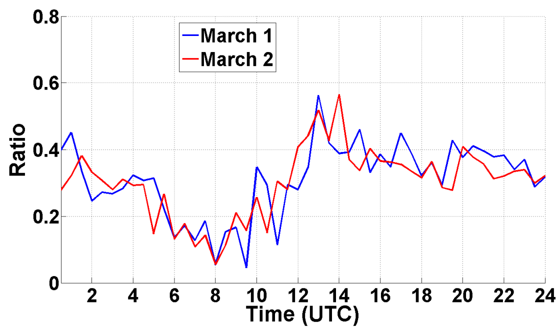
The transition probability $\beta_{i,j}[k]$ is essential to the prediction accuracy of the model. Its physical meaning is that, at any time step, the number of aircraft flowing in or out a Center is proportional to the Center count of the origin



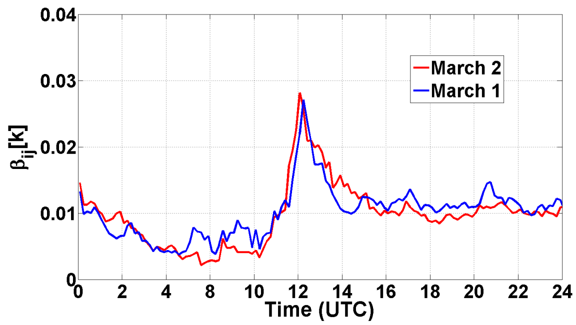
(a) Center count in ZNY on March 1 and 2 in 2010. Data are sampled every one hour.



(b) Number of boundary-crossing in each 1-hour interval from ZNY to ZDC on March 1 and 2 in 2010. Data are sampled every one hour.



(c) Ratio of outflow to the Center count on March 1 and 2 in 2010. Data are sampled every one hour.



(d) Transition probability computed based on data on March 1 and 2 in 2010.

Figure 2. New York Center statistic extracted from ASDI data.

Center. Such assumption is based on observed patterns drawn from historical traffic data. Figure 2(a) shows the daily Center count history in the New York Center (ZNY) on March 1 and 2 in 2010 (data are sampled every one hour). Although the magnitude of Center count varies over days, the patterns of the curves are similar. The saddle shape of the curves reflects the daytime pattern and nighttime pattern inherent in the air traffic. Figure 2(b) shows the history of outflows from ZNY to the Washington Center (ZDC) of the same days. The outflows present a similar pattern as the Center count's. Comparing Figure 2(b) with Figure 2(a), one may find that the magnitude of the outflows are commensurate with the corresponding Center count. Intuitively, the ratio of the outflow to the Center count should be more consistent than the Center count or outflow at its scale, which is manifest in Figure 2(c). The same observation is also found in inflows and outflows between any contiguous Centers. As a result, a time history of the ratio can be used to characterize the transition pattern at Center boundaries.

The number of aircraft passing through a Center boundary during a time interval τ is a *Poisson random variable*,[11] since it can be counted as the event of "arrivals at the Center boundaries" occurring between t and $t+\tau$. When the time interval τ is relatively short, typically one minute, the Poisson process is homogeneous whose probability is given by:

$$P[(s_{i,j}(t+\tau)-s_{i,j}(t))=p]=\frac{(\lambda_{i,j}\tau)^p e^{-\lambda_{i,j}\tau}}{p!}, \quad p=0,1,\cdots \tag{2}$$

where $s_{i,j}(t)$ is the cumulative number of boundary-crossings at time t, and $\lambda_{i,j}$ is the transition rate. Although $\lambda_{i,j}$ is a constant during τ, it would vary slowly throughout the day to reflect change of traffic volume. As a result, it is a time-variant parameter, $\lambda_{i,j}[k]$, representing the average rate of boundary-crossing between time step k and $k+1$. To estimate $\lambda_{i,j}[k]$, a sufficiently large data set must be available for data fitting. But ΔT is too short to host sufficient data samples. For statistical significance, the data are collected in a three-hour period such that there are $60\times 3=180$ data samples in total. The sampling period is a rolling time window centering around the current time step k, i.e., $[k\Delta T-90\ minute,\cdots,k\Delta T,\cdots\Delta T,k+90\ minute]$, then the $\lambda_{i,j}[k]$ derived from data fitting represents the average number of boundary-crossings per minute during the sampling period. Figure 3 shows the normalized distribution of number of boundary-crossings per minute from ZNY to ZDC in different periods on March 1, 2010. The observed distribution is approximated by the Poisson distribution which is impressively close to the observed distribution.

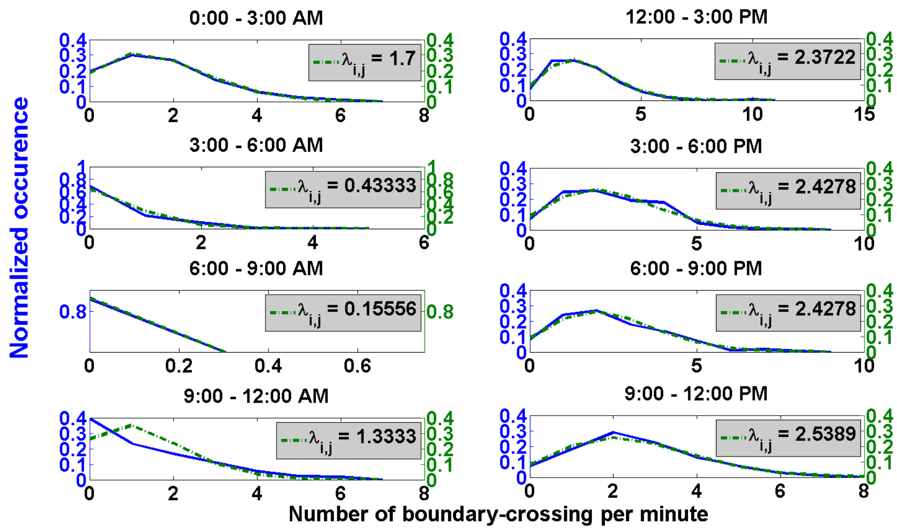


Figure 3. Observed distribution of outflows from ZNY to ZDC (in solid blue) and its Poisson distribution (in green dash) by data fitting, data collected on March 1 in 2010.

With the transition rate, the transition probability $\beta_{i,j}[k]$ can be written as:

$$\beta_{i,j}[k] = \frac{\lambda_{i,j}[k]\Delta T/\tau}{\bar{x}_i[k]}, \quad k \in [0,T] \tag{3}$$

where $\bar{x}_i[k]$ is the mean Center count between time step k and $k+1$ derived from historical data. Figure 2(d) shows the transition probability calculated by Equation (3). Note that the scale of $\beta_{i,j}[k]$ is different from the scale of ratio shown in Figure 2(c), which is due to the different time interval ΔT used. Figure 2(c) is for illustration purpose, thus the time interval is chosen to be one hour. The transition probability $\beta_{i,j}[k]$ in Figure 2(d) is computed for traffic prediction, therefore a smaller interval of 10 minutes is used for better resolution.

In its earlier implementations,[10,11] the LDSM predicts the current traffic using $\beta_{i,j}[k]$ computed from traffic of the previous day. In a sense, the recorded traffic serves as a "template" that the LDSM uses to replicate the traffic with new input $u_i[k]$ and initial state $x_i[0]$. However, relying on traffic pattern of a single day may result in unexpected bias. Figure 4 shows the transition probability history of different days in March 2010. It can be seen that there is no boundary-crossing from ZNY to ZDC between 7 AM and 8 AM on March 5 for unknown reasons. Possible causes involve inclement weather, traffic flow control directives, or data deficiency. No matter whichever is true, this transient state is not representative compared to $\beta_{i,j}[k]$ of the same period on the other days. Another data set on March 7 is found to be discontinued. This might be due to data transfer failure on the vendor side when the computer was trying to connect to the ETMS hubsite. Therefore, use of a single day's traffic record is not reliable. Data of multiple days should be used instead to minimize the chance of data deficiency.

The conventional approach to estimate the traffic patterns based on a large amount of data is to average the $\beta_{i,j}[k]$ over multiple days. The underlying assumption of use of mean is that the data is symmetrically distributed and its probability density function (pdf) is unimodal, such as Gaussian distribution. However, such assumption does not always hold given the variety of sources of uncertainties in the traffic system. Moreover, the distribution of $\beta_{i,j}[k]$ is often irregular, thus no presumptive distributional property can be made. Therefore, the mean is often a biased estimate in the LDSM context. In contrast, the KDE does not have any a priori assumption on the distributional property of the data, and is able to capture the mode of the real distribution. Therefore, the estimated $\beta_{i,j}[k]$ is the most typical value observed in the historical data.

III. The Kernel Density Estimation

This section introduces the Kernel Density Estimation in the context of LDSM. Let subscript $d \in \{0, 1, \cdots, M\}$ denotes the day and M is the total number of day. Using the model introduced in Section II, one can compute a transition probability set $\{\beta^d_{i,j}[k] | k \in [0,T]\}$ for each day. With data of multiple days, the problem can be posted

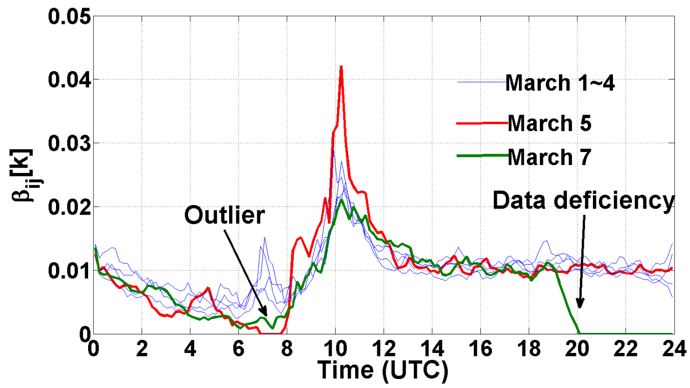


Figure 4. Daily transition probability of boundary-crossing from ZNY to ZDC computed based on historical data.

as: given a set of data samples $\{\beta^d_{i,j}[k] | d \in \{0, 1, \cdots, M\}\}$, estimate the mode of the probability density function $f(\beta_{i,j}[k])$. The observed distribution of $\beta^d_{i,j}[k]$ is irregular (will be shown in Section IV A). In other words, one cannot parameterize $f(\beta_{i,j}[k])$ with any presumptive distribution. Therefore, an optimal way is to approximate $f(\beta_{i,j}[k])$ nonparametrically. The KDE is a popular approach to estimate the underlying probability function without assuming its statistical properties.[13] The estimator is expressed in the following form:

$$\hat{f}_{KDE}(\beta_{i,j}[k]) = \frac{1}{M}\sum_{d=1}^{M} K_h(\beta_{i,j}[k] - \beta^d_{i,j}[k]) = \frac{1}{Mh}\sum_{d=1}^{M} K(\frac{\beta_{i,j}[k] - \beta^d_{i,j}[k]}{h}), \quad k \in [0, T] \tag{4}$$

where $K_h(\cdot)$ is the *kernel* function, defined as $K_h(y) = \frac{1}{h}K(y/h)$. It is usually a symmetric unimodal probability density function. In this paper, the standard Gaussian kernel function $N(0, 1)$ is used for easy implementation and nice statistical properties:[14]

$$K(\frac{\beta_{i,j}[k] - \beta^d_{i,j}[k]}{h}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\beta_{i,j}[k] - \beta^d_{i,j}[k]}{h})^2} \tag{5}$$

Equation (4) sums up the envelope of the kernel function centered at the data samples such that the shape of the distribution to be estimated is approximated by the sum.

h is a smoothing parameter that determines the width of the kernel function. The selection of h is crucial since an inappropriate value will either oversmooth the density function or make it spiky.[14] A comprehensive survey of bandwidth selection can be found in Ref. [15]. The Direct Plug-In (DPI) bandwidth selector proposed in Ref. [16] is used here because of two important reasons: 1) the corresponding KDE has a reasonable small asymptotic mean square error, 2) estimating h is computational fast. Let $K(\cdot)$ be the kernel function, and $f(y)$ be the underlying probability function. $R(K) = \int_{-\infty}^{+\infty} K(z)^2 dz$ is the "roughness" for the kernel function, and $\sigma^2(K) = \int_{-\infty}^{+\infty} z^2 K(z) dz$ is the kernel variance. An asymptotic optimal bandwidth selector (given $f(y)$ is known) is given by:

$$h_A = \left[\frac{R(K)}{\sigma^2(K)^2 R(f'')M}\right]^{\frac{1}{5}}. \tag{6}$$

Note that $R(f'')$ is unknown. If we replace $R(f'')$ by an estimate, then it is called the Direct Plugin (DPI) bandwidth selector:

$$h_{DPI} = \left[\frac{R(K)}{\sigma^2(K)^2 \widehat{R(f'')} M}\right]^{\frac{1}{5}}. \tag{7}$$

Note that $\widehat{R(f'')}$ may also depend on the unknown density function $f(y)$. Iteratively estimating f and h is commonly used to achieve the optimal estimation. In practice, it turns out that 2 iterations are sufficient, which dramatically reduces the computational complexity (See section 3.6.1 in Ref. [14]).

The location of the peak in the estimated kernel density estimate corresponds to the mode from the underlying distribution, i.e., the most likely value that the random variable may take:

$$\hat{\beta}_{i,j}[k] = \arg\max_{\beta_{i,j}[k]} \hat{f}_{KDE}(\beta_{i,j}[k]) \tag{8}$$

The mode $\hat{\beta}_{i,j}[k]$ is a more robust estimate than the mean $\bar{\beta}_{i,j}[k]$, especially when there is a large sampling pool. Because the outliers usually make up the minor cluster of the data set while the mode corresponds to the major cluster, taking the mode consequently avoids the influence of outliers. As the size of the sampling pool increases, $\hat{\beta}_{i,j}(t)$ captures the most typical behavior of the traffic.

IV. Simulation Results

To set up the LDSM, the ETMS data which provide 4-dimentional trajectory of each flight are used.[2] The boundary-crossing and departure information that is essential to the parameterization of the model can be easily extracted from the aircraft trajectories. Aircraft from outside of the continental United States airspace are classified as departures in the International Center. The data set contains NAS-wide air traffic from March through June in 2010, among which traffic from March 1 through May 31 (92 days in total) is used to calculate the transition probability, and traffic in June (31 days in total) is used for model validation. Considerable amount of deficiencies are found in the traffic data of June 14, thus this date is excluded from traffic simulation. $\bar{\beta}_{i,j}[k]$ and $\hat{\beta}_{i,j}[k]$ are used in the LDSM respectively, and their associated traffic predictions are compared.

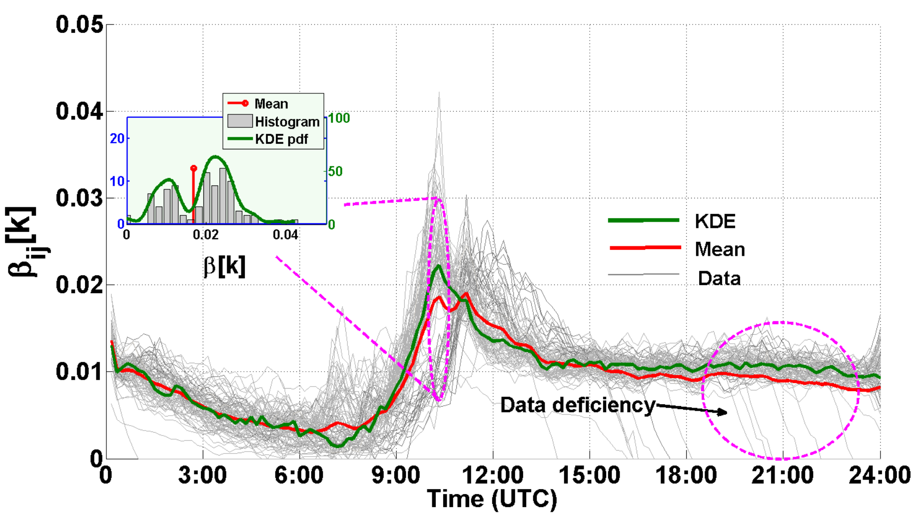
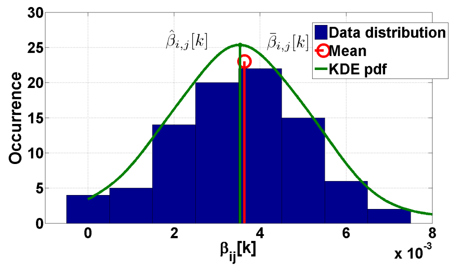


Figure 5. Outflow transition probability history for ZNY-ZDC from March 1 through May 31, 2010.

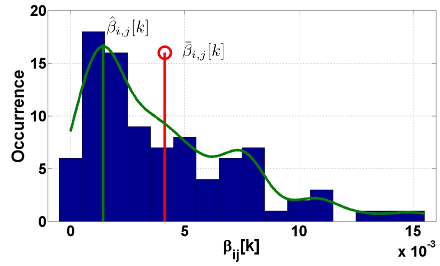
A. Transition probability estimation

Figure 5 shows the daily transition probability of outflow from ZNY to ZDC in the three-month period. In a 24-hour timeframe, the transition probability presents a similar pattern with certain degree of variations on each day. Both KDE and the mean are effective in capturing the pattern. But the mean is prone to deviating from the pattern when data deficiencies are present. The subplot in Figure 5 is a snapshot of the distribution of $\beta^d_{i,j}[k]$ at a specified time step. It can be seen that the distribution is bimodal. The mean is right between the two peaks representing none of the two data clusters. In contrast, the KDE well approximates the shape of the distribution. The mode resulting from the estimated distribution captures the major data cluster. In addition, $\beta^d_{i,j}[k]$ of some days diminish to zero after 15:00 due to data deficiency. The mean is "dragged" down by the zeros, resulting in underestimation of the outflow rate. As the zeros account for only a few of the data samples, the KDE treats these zeros as outliers. Then, $\hat{\beta}_{i,j}[k]$ is not impacted.

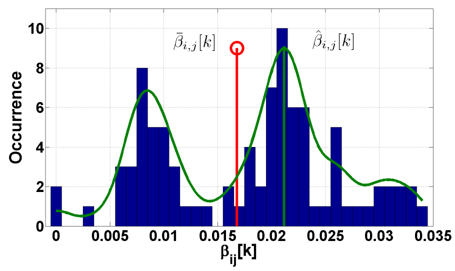
A more detailed examination of the results are shown in Figure 6. The subfigures corresponds to four representative scenarios found in the distribution of $\beta^d_{i,j}[k]$ at different time steps. The samples are presented in histogram with resolution equals to 0.001. In Figure 6(a), $\hat{\beta}_{i,j}[k]$ and $\bar{\beta}_{i,j}[k]$ agree with each other when the distribution is close to Gaussian and unimodal. In Figure 6(c), the distribution is bimodal, hence the mean situates between the two peaks. Figure 6(b) and Figure 6(d) show scenarios where the mean deviates from the peaks of the distribution due to outliers. These scenarios drawn from the observed $\beta^d_{i,j}[k]$ demonstrates the irregularity inherent in the distribution of transition probability. The KDE is able to identify the major cluster of the data regardless of the actual distribution of the data.



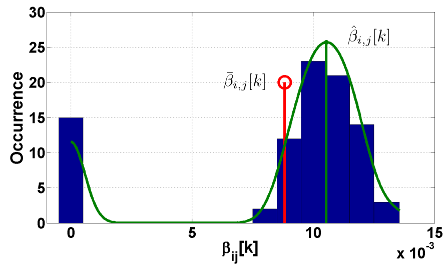
(a) 6:00 UTC



(b) 7:00 UTC



(c) 10:00 UTC

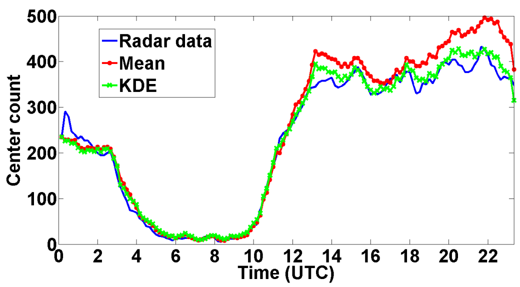


(d) 21:00 UTC

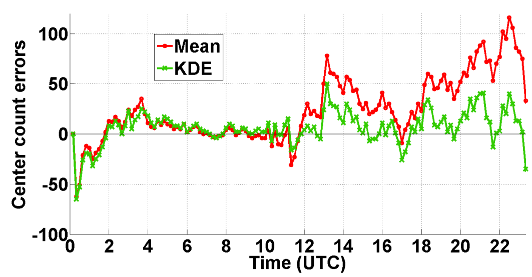
Figure 6. Snapshots of data distribution over 90 days' traffic at different time step of the day. Histogram bin resolution = 0.001.

Therefore, the KDE yields an estimate that is most representative of the transition pattern.

B. Nationwide Air Traffic Prediction



(a) Center count



(b) Prediction errors

Figure 7. Center count predictions for the ZNY Center on June 1, 2010.

Figure 7 compares the Center counts in the ZNY Center on June 1 predicted by using the KDE and the mean respectively. Obviously, the prediction based on the KDE is closer to the actual Center count tracked by radar. The improvement of accuracy is especially evident during the high traffic period, i.e. 12:00-24:00 UTC,. The prediction errors associated with the KDE are less than 50, in contrast to 100 associated with the mean approach. To quantify the deviation from the observed Center count, the normalized model error is used. Define vectors $X_i^d = [x_i^d[0], \cdots, x_i^d[T]]$ and $\hat{X}_i^d = [\hat{x}_i^d[0], \cdots, \hat{x}_i^d[T]]$, where $x_i^d[k]$ is the average Center count observed during the interval $[k\Delta T, (k+1)\Delta T)$ on date d, and $\hat{x}_i^d(t)$ is the corresponding prediction associated with the KDE, then the normalized model errors is defined as:

$$\mu(\hat{X}_i^d) = \frac{1}{T}\sum_{k=0}^{T}\left(\frac{\hat{x}_i^d[k] - x_i^d[k]}{x_i^d[k]}\right) \times 100\%, \qquad i \in N, d \in days\ in\ June \tag{9}$$

For the ZNY Center on June 1, $\mu(\hat{X}_i^d) = 6.19\%$ and $\mu(\bar{X}_i^d) = 10.52\%$. The percentages suggest better performance of the KDE within the 24-hour time horizon. Figure 8 shows the normalized model errors of the twenty Centers for the whole month (June 14 is exluded). The normalized model errors are presented in box plot to show the variation of the

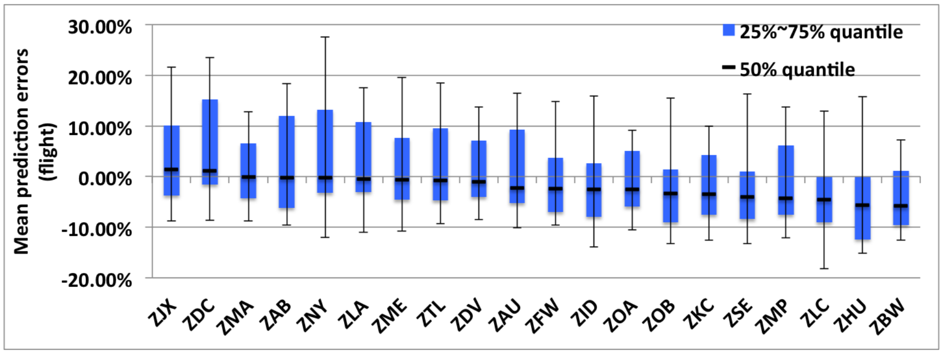


Figure 8. Normalized model errors by LDSM using KDE, traffic in June 2010.

KDE performance. For the majority of Centers, the normalized model errors are bounded by $\pm 20\%$, and data within $25\% - 75\%$ quantile are bounded by $\pm 10\%$.

To measure the accuracy improvement in prediction, the following metrics are defined:

$$L_2(\hat{X}_i^d) = \sqrt{(\hat{X}_i^d - X_i^d)(\hat{X}_i^d - X_i^d)'} \tag{10}$$

$$Ratio_i^d = \frac{L_2(\bar{X}_i^d) - L_2(\hat{X}_i^d)}{L_2(\bar{X}_i^d)} \times 100\%, \qquad d \in days\ in\ June \tag{11}$$

where $L_2(\cdot)$ is the L_2 *distance* that measures the deviation from the observation. The results are shown in Table 1 and Figure 9. $Ratio_i^d$ is positive, indicating larger deviation associated with the mean approach. Table 1 shows that the KDE achieves smaller deviation from the actual Center count for most of days in the month. In Figure 9, 25%-75% quantile of the data are above zero for the twenty Centers. The medians range from 9% to 50%, with a mean of 25.43%, indicating that the KDE outperforms the mean approach for most of the days. But there are also days where the KDE leads to larger deviations (see rows with $Ratio_i^d < 0$ in Table 1). This is due to the highly dynamic nature of traffic patterns. If the traffic to be estimated deviates from the normal patterns, such as those transition probabilities locates in the minor cluster in Figure 2(c), then the mode is not representative to such "deviations." In contrast, the mean is a better estimate in this case. However, as Table 1 suggests, LDSM using the KDE is more accurate and reliable in statistical terms.

Table 1. Days of prediction improvement & deterioration in June 2010 (30 days in total).

Center	ZAB	ZAU	ZBW	ZDC	ZDV	ZFW	ZHU	ZID	ZJX	ZKC
$Ratio_i^d > 0$	29	28	24	28	30	26	21	27	29	27
$Ratio_i^d < 0$	1	2	6	2	0	4	9	3	1	3
Center	ZLA	ZLC	ZMA	ZME	ZMP	ZNY	ZOA	ZOB	ZSE	ZTL
$Ratio_i^d > 0$	29	27	29	25	26	28	28	28	28	25
$Ratio_i^d < 0$	1	3	1	5	4	2	2	3	2	5

V. Conclusion and Future Work

This paper introduces a statistical approach to the Linear Dynamic System Model, which is able to capture the mode of distribution of boundary-crossings between Centers. The Kernel Density Estimation enables the model to make full use of historical data and to extract useful information from a large data set with deficiency, not only enhancing the robustness of the model but also improving the prediction accuracy of the aggregate air traffic model at the system level. Simulation results show that the LDSM based on KDE produces an estimate of Center count with errors less than 20%, and reduces 50% of the prediction errors that are associated with LDSM based on conventional mean approach.

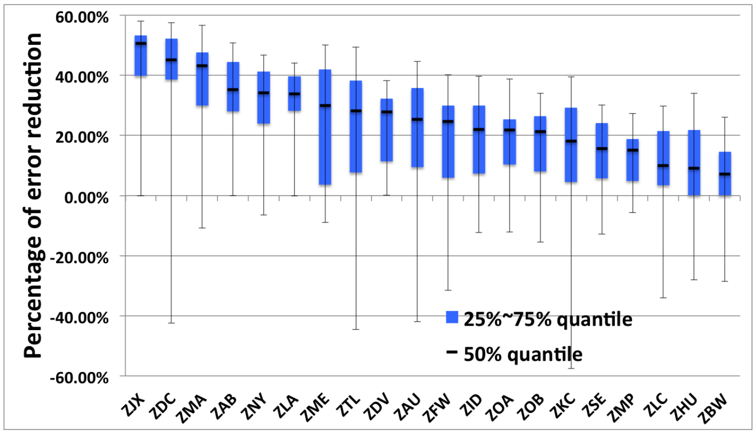


Figure 9. Prediction improvement with respect to the mean approach.

For evaluation purpose, the look ahead time used in this paper is 24 hours, a realistic configuration does not predict traffic for that long time horizon though. Instead, the initial state in the LDSM can be updated periodically using the current observation. As such, the prediction errors should be smaller than that reported in this paper. Another potential improvement that can be made to this model is to classify the traffic patterns in terms of weather, and estimate the transition probabilities for each weather category such that an appropriate traffic pattern can be selected when the weather forecasting for the intermediate future is ready. But this effort requires much more traffic data along with weather information. One or two years ETMS data are desirable so as to accommodate sufficiently large samples for each weather category.

References

[1]Federal Aviation Administration, Air Traffic Organization Policy , Section 8, Chapter 17, ORDER JO 7210.3X, February 9, 2012.

[2]"Enhanced Traffic Management System (ETMS) Functional Description," Volpe National Transportation System Center, VNTSC-DTS56-TMS-002, Cambridge, MA, January, 2000.

[3]Bilimoria, K. D., Sridhar, B., Chatterji, G. B., Sheth, K. S., and Grabbe, S. R.,"FACET: Future ATM Concepts Evaluation Tool," Air Traffic Control Quarterly, Vol. 9, No. 1, pp. 1-20, 2001,

[4]Denery, D. and Erzberger, H., "The Center/TRACON Automation System: Simulation and Field Testing," Proceeding of the Advanced Workshop on Air Traffic Management, Capri, Italy, October 1995.

[5]A. M. Bayen, P. Grieder, G. Meyer, C. J. Tomlin, "Lagrangian Delay Predictive Model for Sector-Based Air Traffic Flow," Journal of Guidance, Control, and Dynamics, Vol. 28, No. 5, September-October 2010.

[6]D. Sun,and A. M. Bayen, "Multicommodity Eulerian-Lagrangian Large-Capacity Cell Transmission Model for En Route Traffic," Journal of Guidance, Control and Dynamics. Vol. 31. No.3. May-Jun 2008.

[7]Y. Cao, D. Sun, "A Link Transmission Model for Air Traffic Management," AIAA Journal of Guidance, Control and Dynamics. Vol.34, No. 5, pp 1342-1351, 2011.

[8]P. K. Menon, G. D. Sweriduk, and K. D. Bilimoria, "A New Approach for Modeling, Analysis, and Control of Air Traffic Flow," AIAA Journal of Guidance, Control, and Dynamics, 2002-5012, August 2002.

[9]P. K. Menon, G. D. Sweriduk, T. Lam, and G. M. Diaz, "Computer-aided Eulerian air traffic flow modeling and predictive control," *AIAA Conference on Guidance, Navigation, and Control* , Vol. 29 No.1, 2006, pp. 12-19.

[10]B. Sridhar, T. Soni, K. Sheth, and G.B. Chatterji, "An Aggregate Flow Model for Air Traffic Management," Journal of Guidance, Controland Dynamic, pp.992-997, Jul-Aug. 2006.

[11]S. Roy, B. Sridhar, G. C. Verghese, "An Aggregate Dynamic Stochastic Model for an Air Traffic System," 5^{th} Eurocontrol/Federal Aviation Agency Airtraffic Management Research and Development Seminar, Budapest, Hungary, Jun. 2003.

[12]B. Sridhar, N. Y. Chen, H. K.Ng, "An Aggregate Sector Flow Model for Air Traffic Demand Forecasting," 9^{th} AIAA Aviation Technology, Integration, and Operations Converence, Sep 21-23, 2009, Hilton head, South Carolina.

[13]B.W. Silverman, Density Estimation for Statistics and Data Analysis. London: Chapman & Hall/CRC, 1998.

[14]M. P. Wand, M. C. Jones. Kernel Smoothing. London: Chapman & Hall/CRC, 1995.

[15]M. C. Jones, J. S. Marron, S. J. Sheather, "A brief survey of bandwidth selection for density estimation." Journal of the American Statistical Association 91 (433): 401407, 1996.

[16]S. J. Sheather, and M. C. Jones, "A reliable data-based bandwidth selection method for kernel density estimation," Journal of the royal Statistical Society, Series B, Vol. 53, pp. 683-690, 1991.