En-route Arrival Time Prediction using Gaussian Mixture Model

Hyunjin Paek and Keumjin Lee
Department of Air transportation & Logistics
Korea Aerospace University
Goyang, 10540, Republic of Korea
rwyclear@gmail.com

Adan Ernesto Vela
Department of Industrial engineering & Management systems
University of Central Florida
Orlando, FL, 32816, USA

***Abstract*—Accurate trajectory prediction is required to realize safe and efficient aircraft operations. In this paper, a new framework for predicting arrival time of en-route aircraft using Gaussian Mixture Model (GMM) is proposed. The proposed method fits the historical trajectory data with GMM whose variable is a set of arrival times at the significant points along a specific air route. The flight times to the defined points along the air route are computed conditioned on the observed flight times for the previous points that the aircraft has already passed by. The form of prediction output from the proposed model is the probability distribution which would increase its applicability to various fields due to its probabilistic nature. The performance of the proposed method is demonstrated by applying it to real flight data in Incheon Flight Information Region (FIR).**

Keywords-component; Trajectory Prediction; Departure Manager; Gaussian Mixture Model; Probabilistic Model; Unsupervised Learning

I. INTRODUCTION

Decision support tools such as Enhanced Traffic Management System, Departure Manager (DMAN), and Traffic Management Advisor are being developed and utilized to reduce and manage air traffic congestion both in airspaces and at airports by optimizing the aircraft's departure sequence and times. To improve the performances of those tools, various studies have been made to develop the techniques for more accurate trajectory prediction.

The classical approach for trajectory prediction is based on aircraft dynamics and performance data such as Base of Aircraft Data (BADA) [1–6]. Since trajectory prediction is a complex process which requires various information including meteorological conditions, aircraft performance and flight intent, there is still a significant room for improving its accuracy [7-9].

More recently, air traffic practitioners and researchers have identified and applied data-driven techniques to address problems related to trajectory prediction. In such studies, researchers propose using regression analysis methods to predict aircraft trajectories [10–11]. Other studies [12–13] performed trajectory prediction using weighted linear regression. Still, another study [14] proposed a method for predicting aircraft trajectories in terminal airspace using a Gaussian Mixture Model (GMM). As a related study, the separation interval between aircraft required for continuous descent operation was calculated using a generalized linear model [15].

In this study, a new framework for predicting fix-point arrival times of en-route aircraft using a GMM is proposed. The proposed method fits the historical trajectory data with a GMM whose output prediction variables is a set of arrival times at the way-points along a specific air route. Accordingly, the arrival times at future way-points over which aircraft will fly are computed conditioned information gained from earlier segments of the trajectory. As time progresses, and new data is updated, the predicted arrival times at subsequent way-points are automatically updated.

Typically, data-driven trajectory models seek to predict arrival times at a specific way-point. Thus, if the goal is to predict an aircraft's location at multiple way-point along its route (especially, at merge points), then multiple independent prediction models are required, one for each way-point. The proposed GMM model is able to overcome this limitation by concurrently learning the correlation between arrival times at multiple way-points. As such, the proposed GMM model is able to produce multiple predictions for all way-points, thereby removing the need to multiple models. The form of prediction output from the proposed method is a probability distribution, unlike traditional regression techniques. The capability of GMMs to produce probabilistic measures increases its applicability and value to various fields such as stochastic arrival scheduling or risk management.

Research supported by the Ministry of Land, Infrastructure, and Transport, South Korea, under contract 19ATRP-C088155-06

II. Methodology

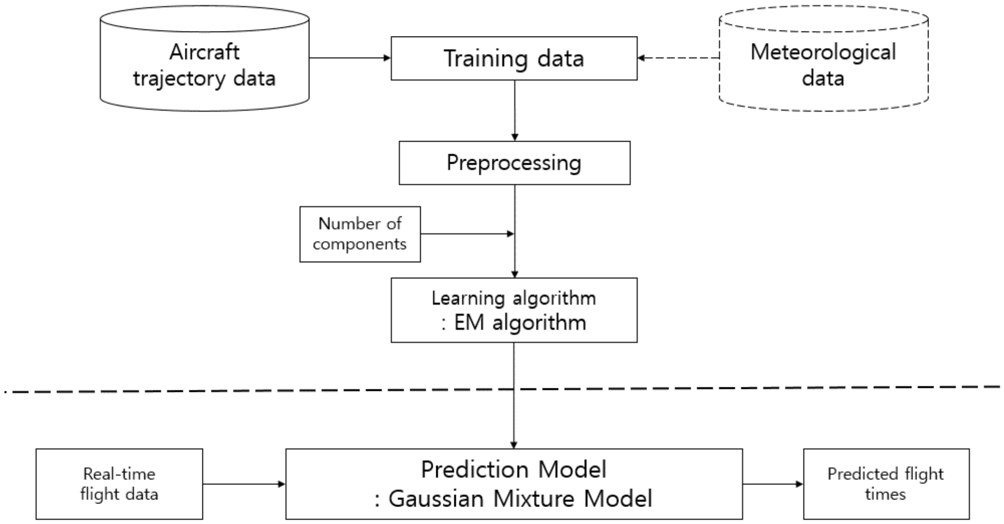

Figure 1. Flow chart of the overall process of the proposed method

A. Overview

In this study, a model for predicting arrival times at the defined points along a specific air route is proposed. Fig.1 shows the overall framework of the proposed method. The method consists of 2 stages: (1) learning stage and (2) prediction stage. In the learning stage, the GMM is generated based on historical data using the EM (Expectation-Maximization) algorithm. The random variable X of GMM represents the flight times to the defined points along the air route from its first way-point. The defined points are set to evenly divide the whole air route, and the number of defined points is d.

In the prediction stage, the flight times to the defined points along the air route are predicted. As shown in Fig. 2, the distributions of the predicted flight times are computed conditioned on the observed flight times for the previous points that the aircraft has already passed by.

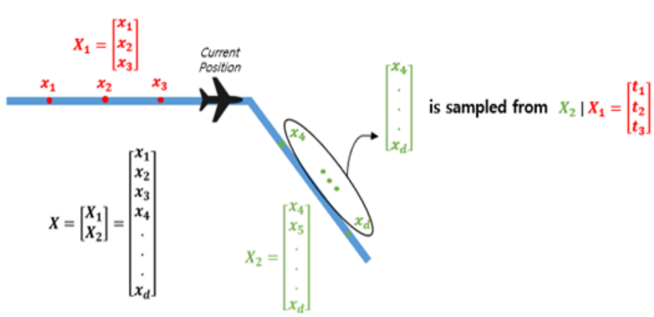

Figure 2. Application example of proposed prediction model

B. Data preprocessing

The radar data used in this study includes aircraft state such as longitude, latitude, altitude, ground speed, call sign by flight time. To generate GMM, resampling process using linear extrapolation and interpolation is performed for each trajectory to obtain the flight times to the defined points. The incomplete trajectories whose data points either starts after the first way-point or ends 12.5 NM before the final way-point are excluded.

An aircraft may fly off the air route due to various reasons including controller instructions or bad weather. For those flights, the flight time associated with the original trajectory point is considered as the flight time of the foot of the perpendicular line to the air route, as shown in Fig. 3.

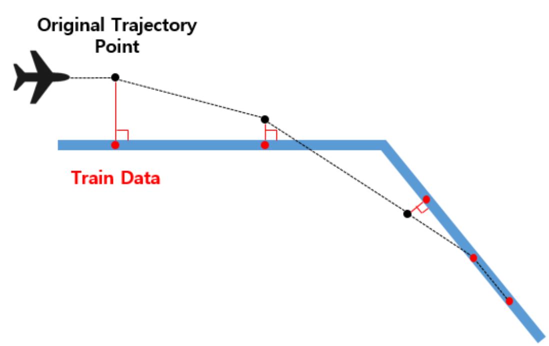

Figure 3. Data preprocessing for the flight off the air route

C. Learning with Gaussian Mixture Model

GMM is a mixture model consisting of multiple Gaussian distributions along with the mixture weights for each distribution. In this study, the parameters (i.e., mean, covariance matrix, and weight) for each distribution are estimated from historical data by using EM algorithm.

In this study, K-means++ algorithm [16] was used to set the initial parameters efficiently. K-means++ algorithm is a heuristic algorithm for determining the centroids in K-means clustering. The statistics such as the mean and covariance of clusters are used as initial parameters for EM algorithm. This process helps to speed up the computation process and increase the chance that the estimated values will converge to the true ones.

D. Prediction method

The conditional distributions of the predicted flight times are computed based on the aircraft's current position (longitude, latitude) and the flight times observed so far for the previous points.

Consider the Gaussian mixture model consist of n (K=n) of d-dimensional multivariate normal distributions $N_1(\mu_1,\Sigma_1)$, $N_2(\mu_2,\Sigma_2)$,...,$N_n(\mu_n,\Sigma_n)$, each with mixture weights $\omega_1,\omega_2,...,\omega_n$. Suppose that an aircraft has already passed by m of defined points (m<d) and a set of observed flight times is given as $A=[a_1,a_2,...,a_m]$. The probability that the aircraft falls within the j [th] distribution can be computed by Bayesian inference as follows.

$$\widehat{\omega}_j=P(z=j|A)=\frac{\omega_j p(A|z=j)}{\omega_1 p(A|z=1)+\omega_2 p(A|z=2)+...+\omega_n p(A|z=n)} \quad (1)$$

Where p(A|z=j) represents the probability of observing A given j [th] multivariate normal distribution (i.e, likelihood). Given that mean and covariance matrix of the j^{th} multivariate normal distribution as in (2), mean ($\widehat{\mu}_j$) and covariance matrix ($\widehat{\Sigma}_j$) of the conditional distribution given A can be calculated as (3).

$$\mu_j=\begin{bmatrix}\mu_{j1}\\ \mu_{j2}\end{bmatrix} \text{ with sizes } \begin{bmatrix}m\times 1\\ (d-m)\times 1\end{bmatrix}$$

$$\Sigma_j=\begin{bmatrix}\Sigma_{j1} & \Sigma_{j2}\\ \Sigma_{j3} & \Sigma_{j4}\end{bmatrix} \text{ with sizes } \begin{bmatrix}m\times m & m\times(d-m)\\ (d-m)\times m & (d-m)\times(d-m)\end{bmatrix} \quad (2)$$

$$\widehat{\mu}_j=\mu_{j2}+\Sigma_{j3}\Sigma_{j1}^{-1}(X-\mu_{j1})$$

$$\widehat{\Sigma}_j=\Sigma_{j4}-\Sigma_{j3}\Sigma_{j1}^{-1}\Sigma_{j2} \quad (3)$$

The conditional distribution of Gaussian distribution follows Gaussian distribution as well. Therefore, the most likely flight time for the remaining points can be computed as weighted sum of mean flight times of the distributions as follows.

$$\text{Predicted Flight Times} = \sum_{j=1}^{n} \widehat{\omega}_j\, \widehat{\mu}_j \quad (4)$$

Finally, the arrival times for the defined points that the aircraft will pass in the future can be computed as:

$$\text{Predicted Arrival Times} = T_0 + \sum_{j=1}^{n} \widehat{\omega}_j\, \widehat{\mu}_j \quad (5)$$

where T_0 represents the current time.

As noted previously, one of the strengths of the proposed GMM model is its ability to not only provide a set of arrival-time predictions for multiple way-points, but also its ability to provide additional distribution measures of the predicted times. Specifically, the variance of predicted flight times is computed as follows:

$$\text{Variance of Predicted Flight Times} = \sum_{j=1}^{n} \widehat{\omega}_j^{\,2}\, \widehat{\Sigma}_j. \quad (6)$$

III. NUMERICAL STUDY

A. Target air route and flight data

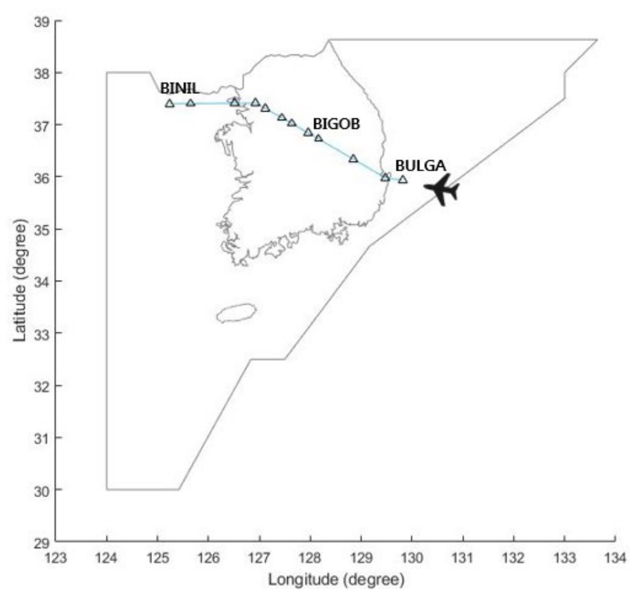

Figure 4. Target air route (G585-G597)

We applied the proposed method to the air route G585-G597 in Incheon Flight Information Region (FIR). As shown in Fig.4, G585-G597 consists of 12 way-points from BULGA to BINIL.

This air route is often used by aircraft cruising from Fukuoka FIR towards Shanghai FIR. The distance from BULGA to BINIL is 253NM and the total flight time is about 42 minutes.

The trajectory data set is made of 2980 flights that have flown the target air route in three months of 2015. 90% of the data (i.e., 2685 flights) were used for training GMM, and the rest of them (i.e., 295 flights) were used as test data. The number of defined points along the air route was set to be 56 where the interval between two consecutive points is approximately 5NM.

B. Generating Gaussian mixture model

In this study, we set the number of components distributions in GMM to be three (K=3). Fig.5 shows the histograms of flight times from BULGA to each way-point of G585-G597. As shown in the figure, the variabilities in flight time are relatively high, and the means and variances for each way-point are shown in Table I. Fig. 6 shows the observed and fitted distributions of flight time to GUKDO. As shown in the figure, the one component is dominant but other components are still significant. Note that the fitted GMM contains the flight time distributions for all way-points, and the one in Fig. 6 was obtained by marginalizing the GMM for GUKDO. The parameters of each component distributions are shown in Table II.

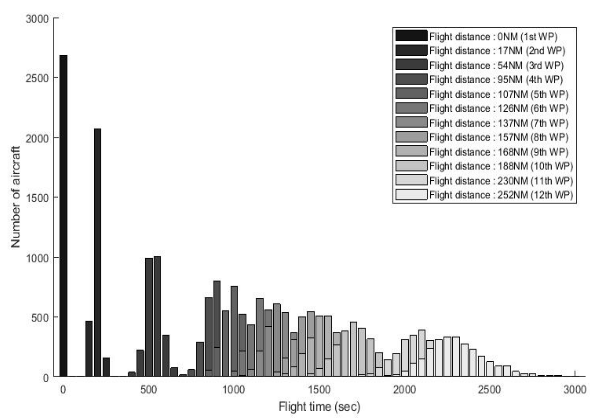

Figure 5. Histograms of flight times from BULGA to each way-point

TABLE I. MEANS AND VARIANCES OF FLIGHT TIME FOR THE WAY-POINTS

Way-point	Mean (sec)	Variance (sec)
BULGA (1st WP)	0	0
POHANG (2nd WP)	166.46	446.06
ELAPI (3rd WP)	506.37	2210.16
BIGOB (4th WP)	877.23	5224.79
BASEM (5th WP)	991.62	6455.04
GUKDO (6th WP)	1160.05	8520.51
KAKSO (7th WP)	1264.96	9955.19
KALMA (8th WP)	1443.38	12627.22
ANYANG (9th WP)	1533.88	14404.26
GOGET (10th WP)	1713.48	18393.18
NOPIK (11th WP)	2098.50	26914.39
BINIL (12th WP)	2287.56	31769.05

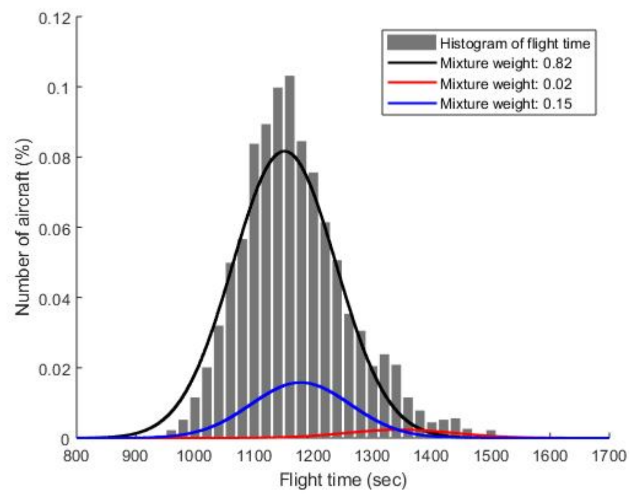

Figure 6. Observed and fitted distributions of flight time to GUKDO

TABLE II. PARAMETERS OF EACH COMPONENT IN FIG.6

Component	Mean	Variance	Mixture weight
1st component	1151	7613	0.82
2nd component	1346	8160	0.02
3rd component	1178	7053	0.15

C. *Arrival time prediction*

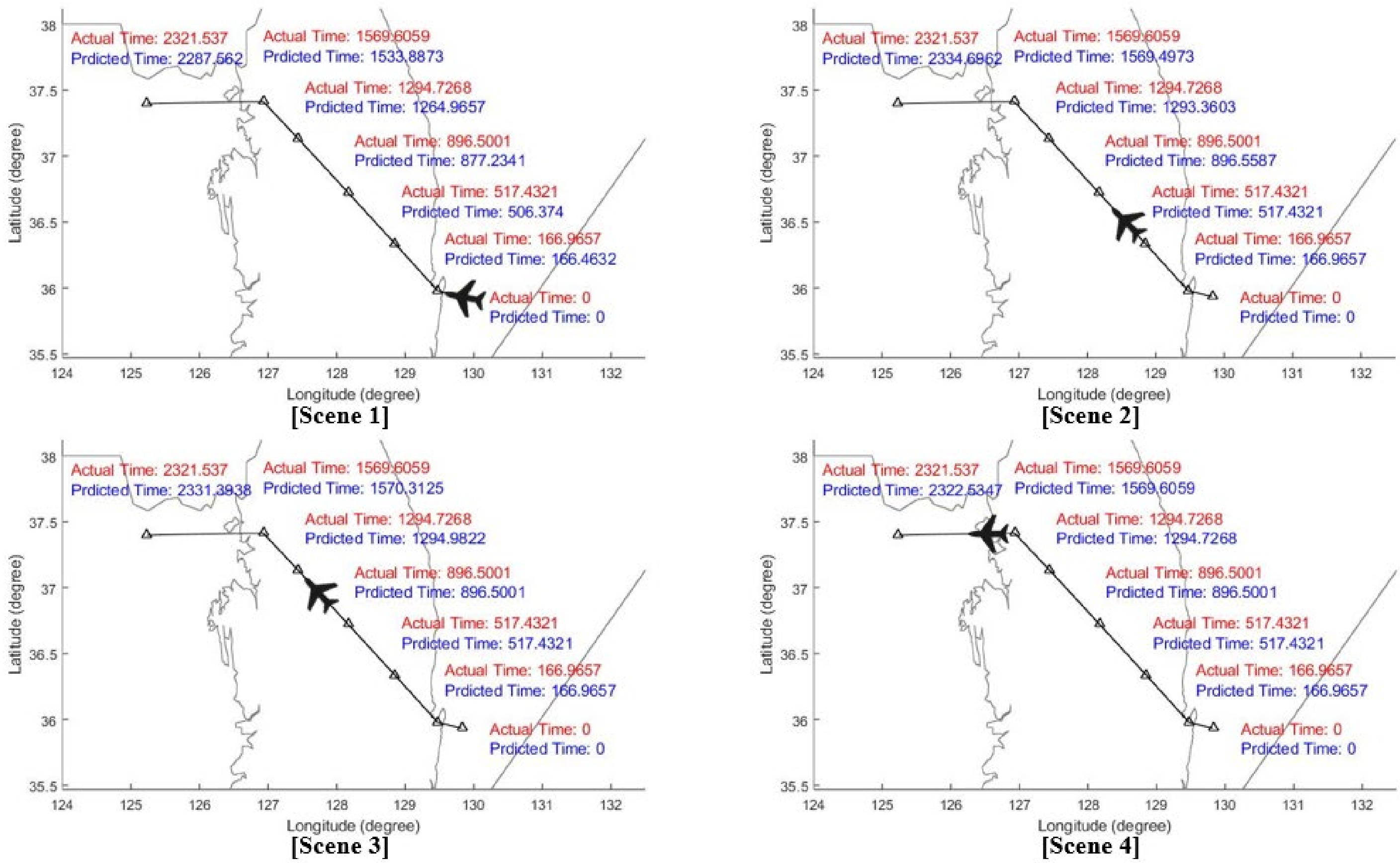

Figure 7. Proposed model applied to G585-G597

Fig. 7 shows a series of snapshots of an aircraft flying along G585-G597. The actual arrival times to the way-points are shown in red and predicted arrival times are shown in blue. As shown in Scene 1, the proposed model initially underestimated the flight times to the way-points (i.e, the predicted arrival times are less than actual times). However, as shown in Scene 2, once the actual arrival time to the third way-point (ELAPI) turned out to be greater than the predicted value, the model could adjust predicted arrival times to reflect the past observations. Although the model overestimated the arrival times in [Scene2], it could continuously adjust its predictions as more observations were collected as shown in Scenes 3 and 4.

Fig. 8 shows the actual and the predicted flight times of an aircraft at the final way-point (BINIL). The red line in the figure represents the actual flight time and the blue line represents the predicted flight time as the aircraft flies. As shown in the figure, the prediction accuracy is getting better as the aircraft continues toward BINIL. Fig. 9 shows the histogram of prediction errors for BINIL using the test data when the aircraft locates at BIGOB. Table III shows the means and Root Mean Square Error (RMSE) of the prediction errors for all way-points.

Fig.10 shows how the RMSE changes as aircraft approach the final fix BINIL. Each red circle in the figure represents the RMSE when the aircraft has just passed each way-point (i.e., the corresponding actual arrival time is observed). As shown in the figure, RMSE was about 186 seconds initially, but once actual arrival times started being observed, the prediction error decreases rapidly. The prediction RMSE at the second way-point was about 75.78 seconds and becomes lower than one minute after the third way-point.

We compared the prediction error of proposed model with a Linear Regression (LR) prediction model. Table III shows the means and RMSE of the prediction errors for each prediction model. As shown in the table, the performance of proposed model does not significantly outperform the LR model in terms of prediction accuracy. However, the primary distinction of the proposed model from other prediction models is that the prediction output is a probability distribution of arrival times (i.e., prediction interval). This probabilistic feature can be used in many ways.

In Fig.11, the blue bar graph represents the histogram of the GMM estimated variances of the flight times for BINIL using the test data when the aircraft locates at GUKDO; the red dashed line in the figure represents the sampled variance of all the actual flight times of the aircraft using the same data. The estimated variances of flight times were computed using (6), while the sample variance of actual flight times from GUKDO to BINIL was 5408.74. As shown in the figure, the estimated variances of predicted flight times are much smaller than the sample variance of flight times. This result reaches a conclusion that the proposed model decreases the uncertainty of predicting flight times.

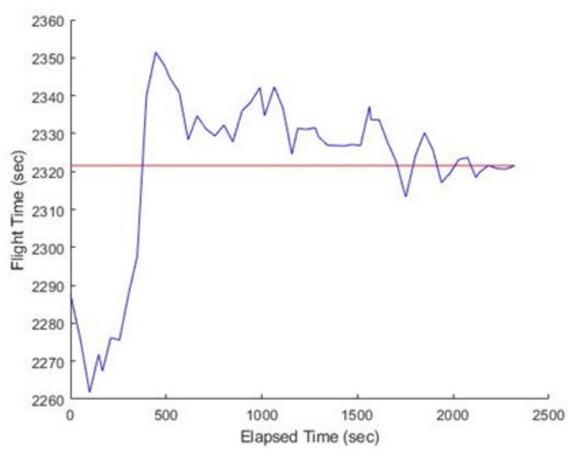

Figure 8. Actual flight time and predicted flight time at BINIL

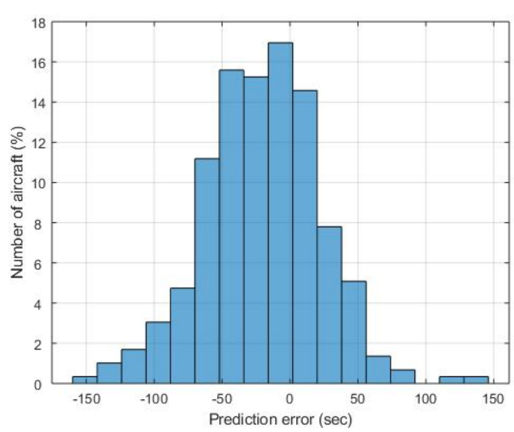

Figure 9. Prediction errors for BINIL from BIGOB (157 NM to go)

TABLE III. MEANS AND RMSES OF PREDICTION ERRORS FOR BINIL

Way-point	GMM		LR	
	Mean	RMSE	Mean	RMSE
BULGA	-125.81	186.26	-125.81	186.26
POHANG	-36.05	75.78	-37.40	74.49
ELAPI	-25.36	57.39	-27.18	58.21
BIGOB	-20.76	47.73	-22.28	48.34
BASEM	-16.99	44.71	-18.82	44.81
GUKDO	-16.86	42.90	-18.21	43.42
KAKSO	-17.06	42.63	-18.26	43.04
KALMA	-16.90	42.37	-18.47	43.12
ANYANG	-13.24	32.37	-14.67	32.29
GOGET	-3.49	16.08	-3.62	15.01
NOPIK	-0.76	6.40	-0.70	5.95

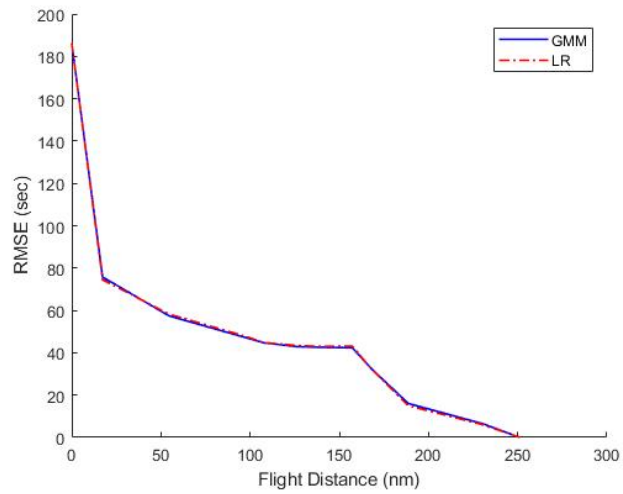

Figure 10. Prediction RMSE by flight distance

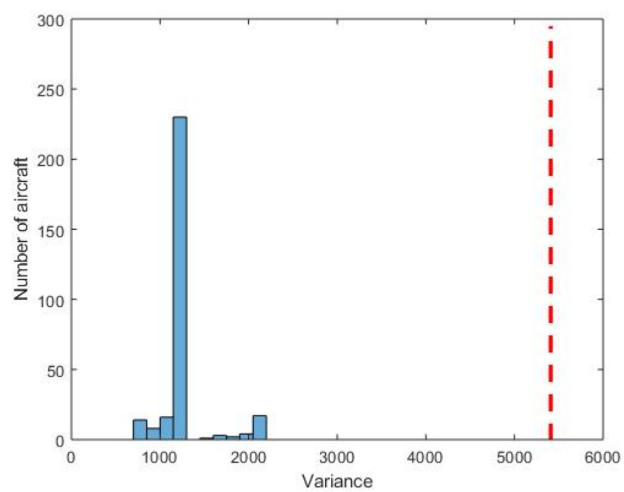

Figure 11. Histogram of variance of predicted flight times from GUKDO

IV. CONCLUSION

In this study, we proposed a new method for predicting the flight time of aircraft using Gaussian mixture model. The proposed model was able to predict arrival time within one minute of RMSE over about 30 minutes look-ahead time.

Only the arrival times gained from earlier segments of the trajectory was considered to compute the arrival times at future way-points over which aircraft will fly. For the future work, the other features which might have a big potential impact on the arrival time such as overall traffic density or cruise speed of the leading aircraft would be included. Inclusion of other features can increase the prediction accuracy of the model.

The proposed model did not out-perform the prediction accuracy of the linear regression prediction model, but the main advantage of the proposed model is that the form of output from the proposed model is a probability distribution of arrival time. Also, the single of proposed model can predict multiple arrival times at defined points. The predicted arrival times at each point are updated each time when the new flight data are given.

The probabilistic nature of the output would increase its applicability of the proposed method to various fields. To prevent unnecessary delays or collisions corresponding to ever increasing air traffic, new methods for predicting aircraft arrival times need to be developed.

ACKNOWLEDGEMENT

This research was supported by the Ministry of Land, Infrastructure, and Transport, South Korea, under contract 19ATRP-C088155-06.

REFERENCES

[1] E. Gallo, J. Lopez-Leones, M. A. Vilaplana, F. A. Navarro, and A. Nuic, “Trajectory computation infrastructure based on BADA aircraft performance model,” 26th Digital Avionics Systems Conference, October 2007.

[2] M. D. Dupuy, and M. Porretta, “Preliminary results for a robust trajectory prediction method using advanced flight data,” 26th Digital Avionics Systems Conference, October 2007.

[3] I. Lymperopoulos, and J. Lygeros, “Model Based Aircraft Trajectory Prediction during Takeoff,” AIAA Guidance, Navigation, and Control Conference and Exhibit, August 2006.

[4] K. Roy, B. Levy, and C. Tomlin, “Target tracking and estimated time of arrival (ETA) prediction for arrival aircraft,” AIAA Guidance, Navigation, and Control Conference and Exhibit, p. 6324, August 2006.

[5] A. Raghunathan, V. Gopal, D. Subramanian, L. Biegler, and T. Samad, “3D conflict resolution of multiple aircraft via dynamic optimization,” AIAA guidance, navigation, and control conference and exhibit, p. 5675, August 2003.

[6] E. Crisotomi, A. Lecchini-Visintini, and J. Maciejowski, “Combining Monte Carlo and worst-case methods for trajectory prediction in air traffic control: a case study,” 6th EUROCONTROL Innovative Research Workshop and Exhibition, p. 259-269, 2007.

[7] J. Rudnyk, J. Ellerbroek, and J. M. Hoekstra, “Trajectory Prediction Sensitivity Analysis Using Monte Carlo Sumulations,” AIAA Aviation Technology, Integration, and Operations Conference, June 2018.

[8] Y. Liu, and X.R. Li, “Intent Based Trajectory Prediction by Multiple Model Prediction and Smoothing,” AIAA Guidance, Navigation, and Control Conference, January 2015.

[9] J. Lovera Yepes, I. Hwang, and M. Rotea, “New Algorithms for Aircraft Intent Inference and Trajectory Prediction,” Journal of Guidance, Control and Dynamics, Vol.30, No. 2, pp. 370-382, 2007.

[10] K. Tastambekov, S. Puechmorel, D. Delahaye, and C. Rabut, “Aircraft Trajectory Forecasting using Local Functional Regression in Sobolev Space,” Transportation Research Part C, Vol. 39, No. 1: Emerging Technologies, p. 1-22, 2014.

[11] S. Hong, and K. Lee, “Modeling the Air Traffic Controller’s Direct-to Operation using Logistic Regression,” 15th AIAA Avation Technology, Integration, and Operations Conference, June 2015.

[12] S. Kim, and K. Lee, “En-route Trajectory Prediction via Weighted Linear Regression,” Journal of the Korean Society for Aviation and Aeronautics, Vol. 24, No. 3, p. 44-52, 2016.

[13] H. Chae, and K. Lee, “En-route Arrival Time Prediction via Locally Weighted Linear Regression and Interpolation,” 38th Digital Avionics Systems Conference, Sep 2019.

[14] S. T. Barratt, M. J. Kochenderfer, and S. P. Boyd, “Learning Probabilistic Trajectory Models of Aircraft in Terminal Airspace from Position Data,” IEEE Transactions on Intelligent Transportation Systems, November 2018.

[15] A. De Leege, M. van Paassen, and M. Mulder, “A machine learning approach to trajectory prediction,” AIAA Guidance, Navigation, and Control (GNC) Conference, p. 4782, 2013.

[16] D. Arthur, and S. Vassilvitskii,: k-means++: The Advantages of Careful Seeding,” Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms, p. 1027-1035, January 2007.