

## Original Article

<https://doi.org/10.12985/ksaa.2019.27.4.027>

ISSN 1225-9705(print) ISSN 2466-1791(online)

## 생성 모델을 사용한 순항 항공기 항후 속도 예측 및 추론

백현진\*, 이금진\*\*

En-route Ground Speed Prediction and  
Posterior Inference Using Generative Model

Hyunjin Paek\*, Keumjin Lee\*\*

## ABSTRACT

An accurate trajectory prediction is a key to the safe and efficient operations of aircraft. One way to improve trajectory prediction accuracy is to develop a model for aircraft ground speed prediction. This paper proposes a generative model for posterior aircraft ground speed prediction. The proposed method fits the Gaussian Mixture Model(GMM) to historical data of aircraft speed, and then the model is used to generate probabilistic speed profile of the aircraft. The performances of the proposed method are demonstrated with real traffic data in Incheon Flight Information Region(FIR).

**Key Words** : Ground Speed Prediction(대지 속도 예측), Departure Manager(출발관리기법), Gaussian Mixture Model(가우시안 혼합 모형), Generative Model(생성 모형), Unsupervised Learning(비지도 학습)

## 1. 서 론

지속적으로 증가하는 국내 및 국제 항공교통수요에 대응하기 위해 출발관리기법(DMAN, Departure Manager)과 같은 의사결정지원도구가 개발되고 있다. 출발관리기법은 항공기의 출발시간과 출발순서를 최적화한 후 이를 관제사들에게 제공함으로써 항공기가 공항 이동지역 및 공역에서 불필요하게 지연되는 것을 막아준다. 항공기의

출발 스케줄을 최적화하기 위해서는 보다 정확한 궤적 예측이 필요하며, 그동안 궤적 예측의 정확도를 향상시키기 위한 다양한 연구가 이루어져 왔다.

기존에는 PMM(Point Mass Model)과 Eurocontrol에서 개발한 BADA(Base of Aircraft Data) 등 항공기 성능 모형(Aircraft Performance Model)을 기반으로 하는 궤적 예측 연구가 주로 수행되었다[1-6]. 하지만 항공기 성능 모형 기반의 궤적 예측은 벡터링(vectoring)과 같이 실제 항공기가 비행할 때 발생하는 다양한 요인들을 고려하지 못하며, 입력 변수로 사용되는 데이터의 정확도 및 접근성이 좋지 않기 때문에 예측 정확도가 떨어진다[7]. 이와 같은 문제를 보완하고 예측 정확도를 향상시키기 위해서 항공기의 의도

Received : 25. Sep. 2019. Revised : 03. Dec. 2019.

Accepted : 24. Dec. 2019

\* 한국항공대학교 항공교통물류학과 석사과정

\*\* 한국항공대학교 항공교통물류학과 교수

연락처자 E-mail : rwyclear@gmail.com

연락처자 주소 : 경기도 고양시 덕양구 항공대학교로 76

(intent)를 추론하여 이를 궤적예측에 반영하는 연구가 이루어졌으며[8, 9], 또한 항공기의 실제 비행 궤적이 기록된 과거 항적 데이터를 학습(Machine Learning)하여 궤적을 예측하는 데이터 기반의 연구가 다양하게 이루어졌다.

관련 연구로, 과거 항적 데이터에 대해 일반화선형모형(generalized linear model)을 사용하여 연속강하운영(CDO, Continuous Descent Operations) 시 필요한 항공기간 분리 간격을 계산한 연구[10]가 있으며, 다른 연구에서는 회귀분석(regression analysis)을 통해 항공기 궤적을 예측하였다[11, 12]. 회귀 분석을 통한 궤적 예측의 정확도를 향상시키고자 가중선형회귀(Weighted Linear Regression)를 사용하여 궤적 예측을 수행한 연구가 있었으며[13, 14], 또 다른 연구에서는 가우시안 혼합모형을 사용하여 공항지역 내 항공기 궤적을 예측하는 방법을 제안하였다[15].

이와 같은 다양한 과거 연구에도 불구하고, 여전히 궤적 예측 모형의 정확도를 향상시키는 것은 풀어야 할 과제이다. 많은 사전 연구 및 문헌들에서 항공기 궤적 예측 정확도에 영향을 주는 요인들을 분석하였다. 그중 대부분의 연구에서는 항공기 속도 오차를 궤적 예측 오차의 가장 큰 요인으로 지목하였으며[16-19], 한 연구에서는 항공기 대지 속도 (Ground Speed)에 대한 예측이 궤적 예측의 정확도를 높이는 핵심 요소라고 지목하였다[20].

본 논문에서는 과거 항적 데이터를 기반으로 만든 가우시안 혼합 모형(Gaussian Mixture Model)을 사용하여 순항 항공기의 대지 속도 (Ground Speed)를 예측하는 방법을 제시하였다. 해당 모형은 속도 예측뿐만 아니라, 사용자가 원하는 만큼 항공기 속도 데이터의 샘플을 생성할 수 있다. 이는 성능 모형에서의 항공기 속도보다 실제 항공기 속도 프로파일과 유사하므로 항공 교통 시뮬레이션이나 다른 모형의 검증 등에 활용될 수 있을 것이다.

## II. 항공기 속도

### 2.1 항공기 속도와 항로의 상관성

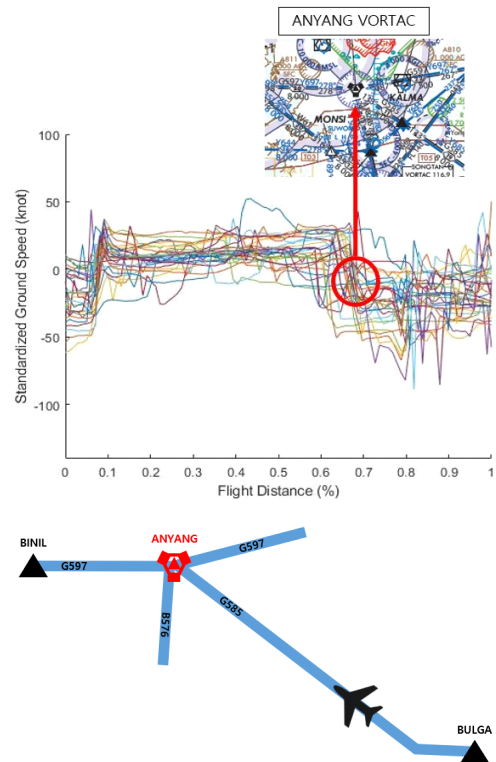


Fig. 1. Standardized aircraft ground speed & air-route

Fig. 1은 2015년 1월, 2월, 10월, 11월에 BULGA~BINIL Way Point(WP)구간을 순항한, 2989대의 항공기의 항적 데이터 중 임의로 추출한 30대의 항공기의 표준화된 대지 속도(Standardized Ground Speed)를 비행 거리에 따라 나타낸 그림이다. 항공기별로 전체 구간에서 속도의 평균을 0으로 표준화시키고 색깔을 다르게 표시하였다. Fig. 1에서 항공기 속도 변화를 보면 x축(Flight Distance)이 0~0.1인 구간에서 항공기의 속도가 전체적으로 증가하는 경향성을 보이고, 0.1~0.6 구간에서는 일정한 경향성을 보이며, 0.6~0.7 구간에서는 감소하는 경향성을 보인다. 그 중에서도 항공기 속도가 급격하게 감소하는 구간인 빨간색으로 표시한 구간(0.6~0.7 구간)은 ANYANG VORTAC이 위치한 지점으로, 교통량이 많고 여러 항로가 합류하는 지점이다. 해당 WP 구간 이외의 구간에 대해 추가적인 분석을 해본 결과, ANYANG VORTAC과 같이 여러 항로의 합류점이 되는

WP 구간, 주변에 공항이 존재하는 WP 구간 혹은 교통량이 많은 접근관계구역에 위치한 WP 구간에서는 항공기의 속도가 주로 감소하는 경향성을 보였고, 반면 이와 같은 제약 조건이 없고, 주로 길이가 긴 구간에서는 항공기의 속도가 증가하는 경향성이 나타났다.

이와 비슷한 실험의 일환으로 인천 FIR 내 특정 WP 구간을 비행한 항공기들 간 속도 데이터의 상관성을 분석해 보았다. 순항 단계의 항공기만을 대상으로 실험을 진행하였으며, 각 WP 구간을 비행한 항공기 100대를 임의로 추출한 후, 항공기간 대지 속도의 상관계수를 계산했다. 그 결과, 동일 항로를 비행한 항공기를 대상으로 구한 상관계수의 평균은 약 0.3에서 0.4 정도로 유의한 양의 상관관계를 보인 반면, 서로 다른 항로를 비행한 항공기간 상관계수의 평균은 0에 근접하거나, 음수로 상관관계가 거의 없었다. 나아가 각 항공기의 대지 속도 데이터에서 바람 성분을 제거하여 진대기 속도(True Air Speed)를 구하고 상관계수를 계산해본 결과, 동일 항로를 비행한 항공기간 상관계수의 평균은 대지 속도를 기준으로 계산한 상관계수의 평균보다 더 높게 나왔고, 서로 다른 항로를 비행한 항공기간 상관계수는 마찬가지로 0에 근접하거나 음수가 나왔다.

순항 단계에서 항공기는 각 항공사의 Cost Index에 제시된 지시 속도를 따라 비행한다. Airbus(1998)에서 발간한 Cost Index 매뉴얼에 따르면, 항공기가 비행할 때 공역별로 소모되는 연료비용이 매우 상이하기 때문에 항공사에서는 이를 고려하여 항로별로 다른 Cost Index를 채택해야 한다[21]고 명시되어 있으며, 이 또한 항로와 항공기 속도 변화가 밀접한 상관성이 있음을 보여준다.

본 논문에서 제시한 가우시안 혼합 모형은 비지도 학습(Unsupervised Learning) 알고리즘을 사용하는 모형으로, 사용자가 성분의 개수만 정해 주면 자동적으로 특정 항로를 비행한 과거 항적 데이터에 포함된 항공기 위치와 속도 데이터의 연관성에 대해 기계학습을 하여 해당 항로에서 항공기 속도 변화의 경향성을 찾아내며, 이를 바탕으로 항공기 항후 속도를 예측할 수 있

다.

### III. 항후 속도 예측 모형

#### 3.1 예측 모형 생성 방법

Fig. 2는 본 연구에서 제안하는 데이터 분석 기반 항공기 항후 속도 예측 모형의 알고리즘을 나타낸다. 해당 알고리즘은 크게 선형 외삽법(Linear Extrapolation)과 선형 보간법(Linear Interpolation)을 사용한 데이터 전처리 단계와 EM (Expectation-Maximization) 알고리즘을 사용한 가우시안 혼합 모형 구축 단계로 이루어진다.

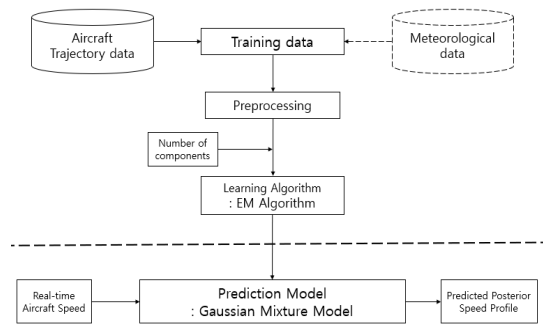


Fig. 2. Flow chart of proposed method

##### 3.1.1 항적 데이터 전처리 과정

본 연구에서 사용한 항적 데이터에는 항공기 호출부호, 비행시간별 항공기 상태 정보(경도, 위도, 고도, 대지 속도) 등의 정보가 포함되어 있다. 특정 항로 구간에 대해 기록된 항적 데이터의 개수는 항공기마다 다를 수 있으며, 보통은 5초에 1번씩 기록된다. 항적 데이터 중 어떤 항공기의 경우에는 기록된 항적의 개수가 너무 적어서 예측 모형의 학습 데이터로 활용될 수 없다. 본 연구에서는 기록된 항적 데이터의 개수가 평균보다 현저히 적은(평균의 70% 미만) 항공기나 대상 항로 구간의 전체 길이를 100%라고 하였을 때, 출발점으로부터 5% 뒤의 지점 이후부터 항적 데이터가 기록된 항공기 또는 도착점으로부터 5% 앞의 지점 이전에 항적 데이터가 기록되지 않은 항공기는 학습 데이터에서 제외시켰다.

항적 데이터 전처리 과정은 가우시안 혼합 모

형 구축을 위해 각 항공기에 해당하는 항적 데이터의 차원을 동일하게 맞춰주는 과정이다. 먼저, 항적 데이터가 출발점보다 뒤에서부터 기록되었거나, 도착점보다 앞에서 끊긴 경우에는 선형 외삽법을 사용하여 출발점부터 도착점에 해당하는 항적 데이터를 계산한다. 그 이후, 전체 항적 데이터에 대해 선형 보간법을 사용하여 해당 항로 구간내의 출발점과 도착점을 포함한  $X_c$  개의 동일한 지점에서의 항공기 속도를 계산할 수 있다.

대상 항로를 비행한 항공기들 중에서는 항로를 정확하게 잘 따라서 비행한 항공기들도 있지만, 악기상이나 관제사의 지시 등의 이유로 벡터링을 하여 항로를 벗어나 비행한 항공기들도 있다. 이러한 항공기들의 경우, Fig. 3과 같이 항공기의 현재 위치에서 항로상에 수선을 내린 후, 수선의 발에 해당하는 지점에 항공기가 위치한다고 지정하였다. 항로를 벗어나 비행한 항공기들의 항로상의 위치를 지정해줌으로써 항로를 잘 따라 가지 않은 항적들도 예측 모형의 학습 데이터로 사용할 수 있었으며, 항로를 벗어난 항공기들에 대해서도 같은 방법으로 항후 속도 예측을 수행할 수 있었다.

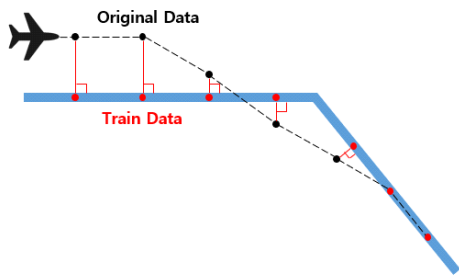


Fig. 3. Data preprocessing procedure

### 3.1.2 가우시안 혼합 모형

기계학습법을 이용한 대부분의 예측 모형은 분별 모형(Discriminative Model)이다. 분별 모형은 학습된 데이터를 바탕으로 주어진 입력 변수를 계산하여 가장 알맞은 하나의 답을 출력한다. 반면, 가우시안 혼합 모형은 생성 모형(Generative Model)으로, 분별 모형과 같이 입력 변수에 대

응하는 가장 알맞은 출력 변수를 내놓을 수 있지만, 이와 더불어 가우시안 분포가 가지는 확률적 특성을 사용하여 조건에 맞는 출력 변수를 원하는 양만큼 확률적으로 생성할 수 있다.

가우시안 혼합 모형은 여러 개의 가우시안 분포(정규분포)와 각 분포에 대한 가중치(weight)로 이루어진 혼합 모형이며, 주어진 관측 값에 대해 EM 알고리즘을 사용하여 만들어진다. EM 알고리즘은 확률 모형(Probabilistic Model)에서 관측 값에 대한 가능도(Likelihood)를 최대화하여 모수를 추정하는 반복 알고리즘으로, Expectation 단계와 Maximization 단계로 구성된다. Expectation 단계에서는 관측 값에 대한 추정 값의 가능도 함수(Likelihood Function)를 구하여 관측 값이 특정 군집(Cluster)에 속할 사후 확률(Posterior Probability)을 계산한다.

2개( $K=2$ )의  $d$ 차원 다변수 정규분포(Multivariate Gaussian Distribution)가 혼합되어 있는 경우를 예시로 들어보겠다. 각 분포에 대한 혼합 가중치(Mixture Weight)를  $w_1, w_2 (w_1 + w_2 = 1)$ 라 하고, 관측 값( $x_i$ ) 이 두 개의 분포 (각각  $N_1(\mu_1, \Sigma_1), N_2(\mu_2, \Sigma_2)$ 로 표기) 중 어느 분포에서 수집되었는지 가리키는 잠재변수(Latent Value)를  $z_i$ 라고 하며,  $n$ 개의 관측 값( $x_1, x_2, \dots, x_n$ )이 수집되었을 때, 매개변수(parameter)는  $\theta = (\mu_1, \mu_2, \Sigma_1, \Sigma_2, w_1, w_2)$ 에 대한 가능도 함수는 식 (1)과 같이 계산할 수 있다. 식 (1)에서  $f$ 는 다변수 정규분포의 확률밀도함수(Probability Density Function)이며,  $I(z_i = j)$ 는  $z_j = j$ 일 경우 1, 그렇지 않을 경우 0을 나타내는 Indicator Function이다.

$$L(\theta; x, z) = P(x, z | \theta) = \prod_{i=1}^n \prod_{j=1}^2 [f(x_i; \mu_j, \Sigma_j) w_j]^{I(z_i = j)} \quad (1)$$

현재의 모수  $\theta^t$ 에 대한  $z_i$ 의 조건부 분포  $C_{j,i}^t = P(Z_i = j | X_i = x_i; \theta^t)$ 는 베이즈 정리(Bayes theorem)에 의해 식 (2)와 같이 계산한다.

$$C_{j,i}^t = \frac{w_j^t f(x_i; \mu_j^t, \Sigma_j^t)}{w_1^t f(x_i; \mu_1^t, \Sigma_1^t) + w_2^t f(x_i; \mu_2^t, \Sigma_2^t)} \quad (2)$$

다음 단계인 Maximization 단계에서는 로그가능도 함수 값을 최대화 하는 최우추정치(Maximum log-Likelihood Estimate)에 해당하는 매개변수를 찾아내어 현재의 다변수 정규분포보다 관측 값에 대한 가능도(likelihood)가 더 높은 새로운 다변수 정규분포들을 찾는다. 먼저,  $w^t = (w_1^t, w_2^t)$ 라 하였을 때, 가중치( $w^t$ )에 대한 최우추정치( $w^{t+1}$ )는 식 (3)과 같이 유도되고, 평균과 공분산( $\mu_1^t, \Sigma_1^t$ )에 대한 최우추정치( $\mu_1^{t+1}, \Sigma_1^{t+1}$ )는 식 (4)와 같이 유도된다.

$$w^{t+1} = \underset{w^t}{\operatorname{argmax}} ([\sum_{i=1}^n C_{1,i}^t] \log w_1^t + [\sum_{i=1}^n C_{2,i}^t] \log w_2^t) \quad (3)$$

$$(\mu_1^{t+1}, \Sigma_1^{t+1}) = \underset{\mu_1^t, \Sigma_1^t}{\operatorname{argmax}} \sum_{i=1}^n C_{1,i}^t (-\frac{1}{2} \log |\Sigma_1^t| - \frac{1}{2} (x_i - \mu_1^t)^T \Sigma_1^{t-1} (x_i - \mu_1^t)) \quad (4)$$

따라서 현재 가중치( $w_j^t$ )에 대한 최우추정치( $w_j^{t+1}$ )는 식 (5)와 같고, 현재 평균과 공분산( $\mu_1^t, \Sigma_1^t$ )에 대한 최우추정치( $\mu_1^{t+1}, \Sigma_1^{t+1}$ )는 식 (6)과 같으며, ( $\mu_2^{t+1}, \Sigma_2^{t+1}$ )도 마찬가지로 방법으로 계산할 수 있다.

$$w_j^{t+1} = \frac{1}{n} \sum_{i=1}^n C_{j,i}^t \quad (5)$$

$$\mu_1^{t+1} = \frac{\sum_{i=1}^n C_{1,i}^t x_i}{\sum_{i=1}^n C_{1,i}^t}, \Sigma_1^{t+1} = \frac{\sum_{i=1}^n C_{1,i}^t (x_i - \mu_1^{t+1})(x_i - \mu_1^{t+1})^T}{\sum_{i=1}^n C_{1,i}^t} \quad (6)$$

EM 과정을 반복하여 관측 값에 대한 매개변수의 가능도를 점차 높여주고, 가능도의 증가폭이 정해진 기준점( $\epsilon=1e-8$ ) 이하가 되면 알고리즘이 종료된다.

본 연구에서는 초기 매개변수를 효율적으로 설정하기 위해서 학습 데이터에 대해 K-평균++ 알고리즘(K-means++ algorithm)[22]을 사용하였다. 초기 매개변수를 잘 설정하면 계산 속도가 더 빨라지고, 결과물이 올바른 값에 수렴할 가

능성이 더 높아진다.

### 3.2 항후 속도 예측 및 추론 방법

본 연구에서 예측 모형의 입력변수는 현재까지 기록된 항공기의 위치(위도, 경도)와 속도이고, 출력변수는 현재 위치부터 항로를 따라 목적지까지 항후 속도의 분포이다. 입력변수가 주어지면, 먼저 데이터 전처리 과정을 통해 입력된 항적 데이터의 차원을 모형과 동일하게 맞춰주고, 다음 단계로 예측 모형에서 베イズ 추론(Bayesian Inference)과 조건부 분포(Conditional Distribution)의 특성을 활용하여 현재까지 기록된 항공기의 속도 분포를 고려한 항후 속도 분포가 결과값으로 출력된다.

결과값을 계산하는 과정을, 혼합 가중치가 각각  $w_1, w_2$ 인 2개의 d차원 다변수 정규분포  $N_1(\mu_1, \Sigma_1), N_2(\mu_2, \Sigma_2)$ 가 혼합된 가우시안 혼합 모형을 예시로 들어보겠다. 어떤 항공기의 속도 데이터  $X = [x_1, x_2, \dots, x_m]$ 가  $m(m < d)$ 개 주어졌을 때, 해당 항공기가 첫 번째 다변수 정규 분포에 해당할 확률은 베イズ 추론에 의해 식 (7)과 같이 표현할 수 있다.

$$\Pr(z=1|X) = \frac{w_1 p(X|z=1)}{w_1 p(X|z=1) + w_2 p(X|z=2)} \quad (7)$$

$\Pr(z=1|X)$ 은 결국 사후 분포(posterior distribution)의 혼합 가중치( $\hat{w}_1$ )와 같으며, 마찬가지로 방법으로  $\hat{w}_2$ 를 계산할 수 있다. 첫 번째 다변수 정규분포의 평균과 공분산 행렬을 식 (8)과 같다고 할 때, 사후 분포의 평균  $\hat{\mu}_1$ 과 사후 분포의 공분산  $\hat{\Sigma}_1$ 은 조건부 분포의 특성에 의해 식 (9)와 같은 방법으로 계산할 수 있다.  $\hat{\mu}_2, \hat{\Sigma}_2$ 도 마찬가지로 계산한다.

$$\mu_1 = \begin{bmatrix} \mu_{11} \\ \mu_{12} \end{bmatrix} \text{ with sizes } \begin{bmatrix} m \times 1 \\ (d-m) \times 1 \end{bmatrix} \\ \Sigma_1 = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{13} & \Sigma_{14} \end{bmatrix} \text{ with sizes } \begin{bmatrix} m \times m & m \times (d-m) \\ (d-m) \times m & (d-m) \times (d-m) \end{bmatrix} \quad (8)$$

$$\begin{aligned}\hat{\mu}_1 &= \mu_{12} + \Sigma_{13} \Sigma_{11}^{-1} (X - \mu_{11}) \\ \hat{\Sigma}_1 &= \Sigma_{14} - \Sigma_{13} \Sigma_{11}^{-1} \Sigma_{12}\end{aligned}\quad (9)$$

가우시안 분포의 조건부 분포는 모분포와 마찬가지로 가우시안 분포를 따른다. 이러한 특성은 가우시안 혼합 모델로 하여금 데이터를 쉽게 샘플링할 수 있게 해준다. 위의 방법들을 통해 주어진 항적데이터에 대한 향후 속도 분포(조건부 분포)의 평균과 공분산을 계산할 수 있다. 본 연구에서 향후 속도 예측은 가우시안 혼합 모형 각 성분의 혼합 가중치와 그 성분에 해당하는 평균을 사용하여 식 (10)과 같은 방법으로 계산하였다.

$$\text{Predicted Posterior Speed} = \sum_{i=1}^2 \hat{w}_i \hat{\mu}_i \quad (10)$$

항후 속도에 대한 추론(샘플링)의 경우, 각 성분의 혼합 가중치와 평균, 공분산을 사용하였으며,  $k \sim N(0, I_{d-m})$ ,  $\hat{\Sigma}_j = U_j S_j U_j^T$ 에 대해 식 (11)과 같은 방법으로 계산하였다. 이는 표준정규분포를 따르는 확률변수( $k$ )를 사용하여 다변량 정규 분포에서 임의의 표본을 추출하는 방법과 동일하며,  $U_j$ 와  $S_j$ 는 사후 분포에 대한 공분산 행렬의 특이값(Singular Value)이다. 하지만 특이값을 그대로 사용할 경우, 공분산 행렬에 많은 noise가 포함되어 있기 때문에 샘플링 결과에 좋지 않은 영향을 미친다. 따라서 샘플링시에는 특이값을 일부만 남기고(truncate), 가장 큰  $r$ 개의 값들만 추출하여 사용하였다.

$$\text{Posterior Speed Sample} = \hat{\mu}_j + U_j S_j^{1/2} k \quad (11)$$

## IV. 향후 속도 예측 사례 연구

### 4.1 대상 공역 및 항적 데이터

본 연구에서는 인천 FIR 내의 항공로 G585-G597을 대상으로 제안된 예측 모형을 만들고 성능을 측정하였다. Fig. 4는 해당 항로를 구성하는 WP중 BULGA부터 BINIL까지의 구간을

나타낸다. 해당 항로는 보통 후쿠오카 FIR에서 상하이 FIR 방향으로 순항하는 항공기들이 사용한다. BULGA부터 항로를 따라 BINIL까지의 거리는 약 253NM로 비행시간은 약 45분가량 소요된다.

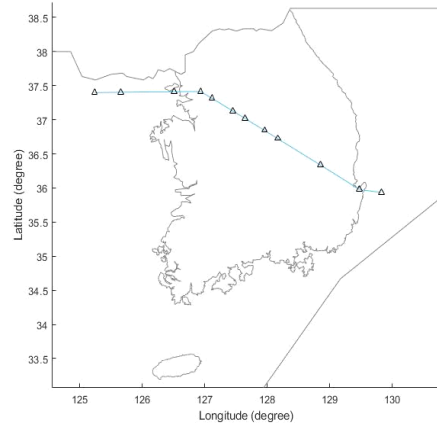


Fig. 4. Target air-route (G585-G597)

항적 데이터의 경우, 2015년 3개월 동안 해당 구간을 비행한 항공기 2,989대의 레이다 관측 자료를 대상으로 하였다. 전체 항공기의 약 90%에 해당하는 2,690대의 항공기는 학습 데이터로 사용하였고, 나머지 약 10%에 해당하는 항공기는 예측 모형의 성능을 측정하는데 사용하였다.

학습 데이터에 사용된 항적 데이터 길이(차원)는 254( $X_c = 254$ )로 동일하게 맞추어 주었다. 데이터의 길이를 254로 설정해줌으로써, 대상 WP 구간(약 253NM)에 대해 약 1NM 간격으로 항공기 속도를 예측 및 샘플링할 수 있었다.

### 4.2 가우시안 혼합 모형 생성

가우시안 혼합 모형을 생성할 때에는 성분의 개수(가우시안 분포의 개수, K)를 설정해 주어야 한다. 금번 사례 연구에서는 K=5로 설정하였다. 성분의 개수를 5개로 설정하고 EM 알고리즘을 통해 모형을 생성하면, 5개의 가우시안 분포( $N_j, j=1, 2, \dots, 5$ )와 각 분포에 대한 혼합 가중치( $w_j, \sum_{j=1}^5 w_j = 1$ )로 구성되는 가우시안 혼합 모형이 만

들어진다. 혼합 가중치는 각 분포가 전체 모형에서 가지는 비율(proportion)을 의미하며, 각 가우시안 분포는 크기가  $1 \times 254$ 인 평균( $\mu_j$ )과 크기가  $254 \times 254$ 인 공분산( $\Sigma_j$ ) 행렬을 가진다.

이 때, 각 성분의 평균은 해당 항로에서 비행한 항공기들이 나타낸 5개의 전형적인 속도 프로파일(Archetypal Speed Profile)을 의미한다고 할 수 있으며, 특정 성분의 혼합 가중치가 높을수록 항공기가 해당 성분이 나타내는 속도 프로파일을 따라 비행할 확률이 높다. Fig. 5는 생성된 가우시안 혼합 모형 각 성분의 평균과 혼합 가중치를 나타낸 그림이다.

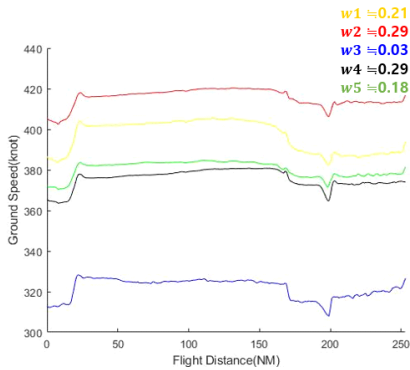


Fig. 5. Archetypal speed profiles

#### 4.3 향후 속도 예측 및 추론

학습 데이터 이외의 항적 데이터(검증 데이터)를 사용하여 항공기의 향후 속도를 예측, 샘플링을 하였다. Fig. 6은 검증 데이터에서 임의로 추출한 4대의 항공기의 속도 데이터가 출발지부터 50개 지점에 대해 주어졌을 때 향후 속도를 예측, 샘플링한 결과이다. 그림에 주어진 속도 데이터는 '\*'로 표시하였으며, 실제 속도분포는 빨간색, 예측한 향후 속도는 파란색, 샘플링 결과는 회색 실선으로 나타났다. 속도를 샘플링할 때 사용되는 특이값은 가장 큰 5개( $r=5$ )의 값을 사용하였는데, 이는  $r$ 값을 너무 작게 설정하면 샘플링 결과에 실제 항공기의 속도 변화가 충분히 반영되지 못하고  $r$ 값을 너무 크게 설정하면 지나치게 많은 noise가 발생하기 때문이다.

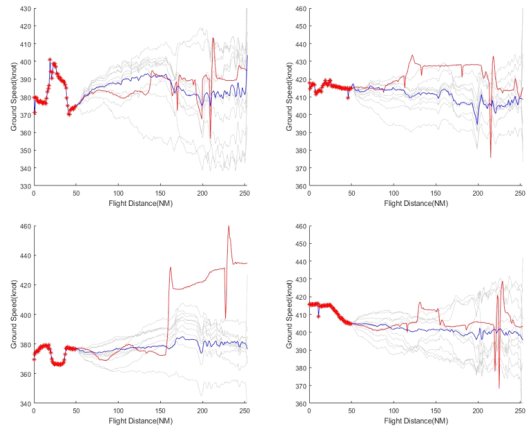


Fig. 6. Prediction & inference examples (50 data given)

Fig. 7은 검증 데이터에서 임의로 추출한 1대의 항공기의 속도 데이터가 1개, 10개, 20개, 30개 주어졌을 때 예측, 샘플링한 결과를 나타낸다. 데이터가 1개 주어진 경우에 비해 나머지 경우에서 향후 항공기 속도에 대한 증가, 감소 여부를 비교적 잘 예측한 것으로 나타났다.

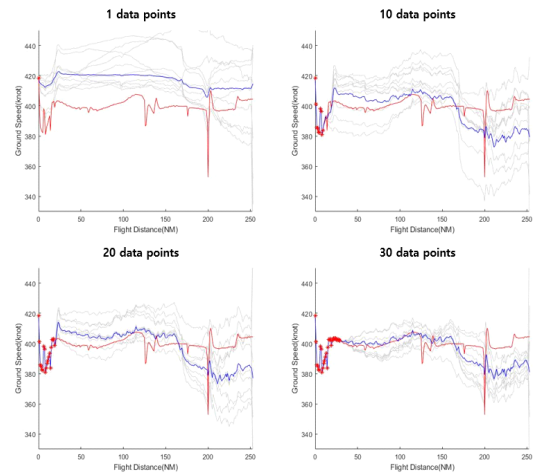


Fig. 7. Prediction & inference examples (1, 10, 20, 30 data given)

가우시안 혼합 모형을 만들 때에는 적절한 성분의 개수를 채택하여야 한다. 성분의 개수가 너무 적으면 모형의 유연성이 떨어지기 때문에 예측 정확도가 낮아지고, 성분의 개수가 너무 많을 경우에는 과적합(Overfitting) 문제가 발생한다. 기



본 연구에서는 AIC(Akaike Information Criterion) 혹은 BIC(Bayesian Information Criterion)를 사용하여 적합도를 측정하거나[23], 모형의 성분 간 거리를 측정하는 방법을 사용하여 적절한 성분의 개수를 도출하였다[24]. 하지만 본 연구에서 제안한 모형의 목적은 데이터를 클러스터링하는 것이 아니라, 예측하는 것이 목적이므로, 검증 데이터에 대한 예측 값의 평균제곱근오차(RMSE, Root Mean Square Error)를 측정하여 오차가 가장 작게 나오는 혼합 모형을 채택하였다. Fig. 8은 가우시안 혼합 모형 성분 개수에 따른 항공기 향후 속도 예측 오차를 나타낸다. 이 때, 검증 데이터로 사용된 항공기는 총 299대이며, 항공기 별로 속도 데이터가 10개 주어졌을 때 향후 속도 값(244개)에 대한 RMSE를 계산하고, 299대의 항공기에 대해 평균을 내었다. 결과적으로, 성분의 개수가 5개일 때 RMSE의 평균이 가장 작게 계산되었으며, 성분의 개수가 6개 이상 되는 시점부터 오차가 증가하였다.

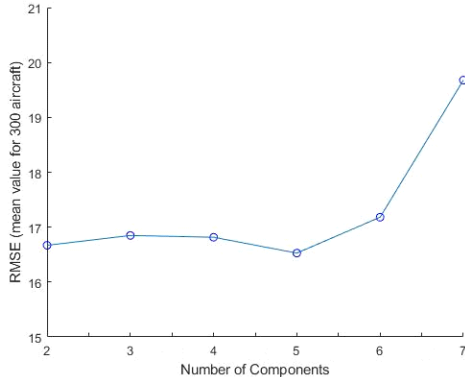


Fig. 8. Mean RMSE by number of components

본 연구에서는 이와 같은 이유로 예측 모형의 성분을 5개로 설정하였으며, 해당 모형을 사용하여 항공기 향후 속도를 예측해본 결과, 속도 데이터가 1개 주어진 경우 항공기 299대에 대한 RMSE의 평균은 약 18.1258, 10개 주어진 경우는 16.5274, 20개 주어진 경우는 16.0599로 계산되었고, 예측 모형이 속도 데이터가 1개 주어진 경우에도 향후속도를 평균적으로 20knot 오차범위 안에서 예측할 수 있었다.

## V. 결 론

본 연구에서는 가우시안 혼합 모형을 사용하여 순항 항공기의 속도를 예측하는 새로운 방법을 제시하였다. 항공기의 속도를 예측하는 것은 궤적 예측의 정확도를 향상시키는 데 필수적인 요소이며, 본 연구에서 제안한 기법은 다양한 궤적 예측 모형의 정확도를 개선시키는데 활용될 수 있다. 또한 가우시안 혼합 모형은 생성 모형으로, 학습된 실제 항적 데이터와 유사한 항공기 속도의 표본을 사용자가 원하는 만큼 샘플링할 수 있고, 샘플링한 결과물은 다양한 항공 교통 시뮬레이션 모형에 활용될 수 있을 것이다.

향후 연구로는 순항 항공기 속도 예측 모형의 정확도를 향상시켜야 하며, 다른 예측 모형과 비교하여 본 연구에서 제안한 모형의 성능을 검증할 필요가 있다.

## 후 기

본 연구는 국토교통과학기술진흥원의 “항공기 출발 및 도착 통합관리기술연구” (19ATRP-C088 155-06)의 일환으로 수행되었으며, 지원에 감사드립니다.

## References

- [1] Gallo, E., Lopez-Leones, J., Vilaplana, M. A., Navarro, F. A., and Nuic, A., “Trajectory computation infrastructure based on BADA aircraft performance model”, 2007 IEEE/AIAA 26th Digital Avionics Systems Conference, Dallas, Texas, Oct 21, 2007.
- [2] Dupuy, M. D., and Porretta, M., “Preliminary results for a robust trajectory prediction method using advanced flight data”, 2007 IEEE/AIAA 26th Digital Avionics Systems Conference, Dallas, Texas, Oct 21, 2007.
- [3] Lymperopoulos, I., Lygeros, J., “Model based aircraft trajectory prediction during takeoff”,



- AIAA Guidance, Navigation, and Control Conference and Exhibit, Colorado, Keystone, Aug 21-24, 2006.
- [4] Roy, K., Levy, B., and Tomlin, C., "Target tracking and estimated time of arrival (eta) prediction for arrival aircraft", AIAA Guidance, Navigation, and Control Conference and Exhibit, Colorado, Keystone, Aug 21-24, 2006.
  - [5] Raghunathan, A. Gopal, V., Subramania, D., Beigler, L., and Samad, T., "3D conflict resolution of multiple aircraft via dynamic optimization", AIAA Guidance, Navigation, and Control Conference and Exhibit, Austin, Texas, Aug 14, 2003.
  - [6] Crisostomi, E., Lecchini-Visintini, A., and Maciejowski, J., "Combining Monte Carlo and worst-case methods for trajectory prediction in air traffic control: a case study", 6th EUROCONTROL Innovative Research Workshop and Exhibition, Bretigny sur Orge, France, 2007, pp.259-269.
  - [7] Vourous, G., "Data-driven aircraft trajectory prediction exploratory research", The DART project by EU and EUROCONTROL, 2017, pp.9-10.
  - [8] Liu, Y., and Li, X. R., "Intent based trajectory prediction by multiple model prediction and smoothing", AIAA Guidance, Navigation, and Control Conference, Kissimmee, Florida, January 5-9, 2015.
  - [9] Yepes, J. L., Hwang, I., and Rotea, M., "New algorithms for aircraft intent inference and trajectory prediction", Journal of Guidance, Control and Dynamics, 30(2), 2007, pp.370-382.
  - [10] de Leege, A. M. P., van Paassen, M. M., and Mulder, M., "A machine learning approach to trajectory prediction", AIAA Guidance, Navigation, and Control Conferences, Boston, MA, August 9-22, 2013.
  - [11] Tastambekov, K., Puechmorel, S., Delahaye, D., and Rabut, C., "Aircraft trajectory forecasting using local functional regression in Sobolev Space", Transportation Research Part C, 39(1): Emerging Technologies, 2014, pp. 1-22.
  - [12] Hong, S., and Lee, K., "Modeling the air traffic controller's direct-to operation using logistic regression", 15th AIAA Aviation Technology, Integration, and Operations Conference, Dallas, Texas, Jun 22-26, 2015.
  - [13] Kim, S., and Lee, K., "En-route trajectory prediction via weighted linear regression", Journal of the Korean Society for Aviation and Aeronautics, 24(3), 2016, pp.44-52.
  - [14] Chae, H., "En-route arrival time prediction via locally weighted linear regression with barycentric interpolation", M.D. Thesis, Korea Aerospace University, Goyang, Korea, Jul, 2019.
  - [15] Barratt, S., T., Kochenderfer, M., J., and Boyd, S., P., "Learning probabilistic trajectory models of aircraft in terminal airspace from position data", IEEE Transactions on Intelligent Transportation Systems, Nov 28, 2018.
  - [16] Mondoloni, S., and Bayraktutar, I., "Impact of factors, conditions and metrics on trajectory prediction accuracy", 6th USA/ Europe Air Traffic Management Research and Development Seminar, Baltimore, Maryland, Jun 27-30, 2005.
  - [17] Vivona, R., Cate, K., and Green, S., "Comparison of aircraft trajectory predictor capabilities and impacts on automation interoperability", 11th AIAA Aviation Technology, Integration, and Operations Conference, Virginia Beach, Virginia, Sep 20-22, 2011.
  - [18] Fukuda, Y., Shirakawa, M., and Senoguchi, A., "Study on trajectory prediction model", ENRI International Workshop on ATM/CNS, Tokyo, Japan, 2009.
  - [19] Fukuda, Y., Shirakawa, M., and Senoguchi, A., "Development and evaluation of tra-

- jectory prediction model", 27th International Congress of the Aeronautical Sciences, Nice, France, Sep 19-24, 2010.
- [20] Schuster, W., and Ochieng, W., "Performance requirements of future trajectory prediction and conflict detection and resolution tools within SESAR and NextGen: Framework for the derivation and discussion", *Journal of Air Transport Management*, Vol. 35, pp.92-101, Mar, 2014.
- [21] Airbus, "Flight operations support & line assistance-getting to grips with the cost index issue 2", Airbus Customer Service, May, 1998, pp.7-9, 14-17, 57-60.
- [22] Arthur, D., and Vassilvitskii, S., "k-means++: The advantages of careful seeding", *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, New Orleans, Louisiana, Jan 7-9, 2007.
- [23] Leroux, B., G., "Consistent estimation of a mixing distribution", *The Annals of Statistics*, 20(3), 1992, pp.1350-1360.
- [24] Chen, J., and Kalbfleisch, J., D., "Penalized minimum-distance estimates in finite mixture models", *The Canadian Journal of Statistics*, 24(2), 1996, pp.167-175.