

Aircraft trajectory recognition via statistical analysis clustering for Suvarnabhumi International Airport

Patcharin Kamsing^{†*}, Peerapong Torteeka[‡], Soemsak Yooyen[†], Siriporn Yenpiem[⊥], Daniel Delahaye[⊞], Philippe Notry[⊞],
Thaweerath Phisannupawong[†], Sittiporn Channumsin[±]

[†]Department of Aeronautical Engineering and Commercial Pilot, International Academy of Aviation Industry

King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

[‡]National Astronomical Research Institute of Thailand, ChiangMai 50180, Thailand.

[⊥]President Office, Aeronautical Radio of Thailand LTD, Bangkok 10120, Thailand.

[⊞]École Nationale de l'Aviation Civile(ENAC), Toulouse, France.

[±]Geo-Informatics and Space Technology Development Agency (GISTDA), Chonburi 20230, Thailand.

*Email: insom.patcharin@gmail.com, patcharin.ka@kmitl.ac.th

Abstract—Since Suvarnabhumi International Airport is considered to be the biggest airport in Thailand, a big travelling-hub of southeast Asia and plays a significant part to the economy of Thailand relying on the tourism industry, an aircraft trajectory recognition is essential to support the high traffic management system from around the world. The first and essential stage of airport capacity enhancement is descriptive-analytic in several sections of the airport, including flight trajectory behaviors in order to plan an improvement procedure in the future. This experiment deploys K-mean and Gaussian Mixture clustering to compare results by using available automatic dependent surveillance-broadcast (ADS-B) dataset provided by the bigdata system from various websites. The test varies the number of clustering from three to ten and measures how similar an object is to its cluster by using the Silhouette score. Gaussian Mixture clustering produces at least three unique flight trajectories when setting the number of clustering equal to four, giving the Silhouette score of 0.43. K-mean clustering with the number of clustering equal to ten gives the highest Silhouette score of 0.45. However, its routes are not clearly recognized when compared with the Gaussian Mixture clustering. Although the overall results are not clearly shown in the pattern, it is enough to describe the trajectory patterns of the aircrafts taking off or landing over Suvarnabhumi International Airport.

Index Terms—Trajectory, Clustering, Airport Capacity, Gaussian Mixture Model(GMM), K-mean

I. INTRODUCTION

Air Traffic Management(ATM) is a trend towards increasing interest in the growth of numerous flights. Suvarnabhumi International Airport is considered to be the biggest airport in Thailand and a big travelling-hub of Southeast Asia. Based from the data collected from <https://opensky-network.org> [1] on September 12, 2017, compared the data collected on September 12, 2019, over Suvarnabhumi International Airport, the callsign value which reflects the airline and the flight number has been growing up to 16% from 1,562 callsigns to 1,822 callsigns in the year 2019. Therefore, Air Traffic Management should be recognized to enhance the capacity of the airport.

This work was supported by King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand.

According to the standard separation requirement [2], aircrafts that operate in the Terminal Maneuvering Area(TMA) are required to be separated vertically at least 1,000 feet and separated horizontally at least three nautical miles(NM). This requirement is one significant rule for Air Traffic Control(ATM) to manage all traffic in pre-setting routes, such as arranging and managing the queue for an airplane to enter a crossing point that is well located and does not change with time, before landing. If the amount of route is small, the air traffic controller can manually handle with the available air traffic tools. However, a research on planning and management air traffic is on-demand and tends to apply many optimization algorithms. For example, researchers deploy a dynamical system by implementing a radar observation(position and speed) [3]. Direct Route Optimization for air traffic presented to solve a problem about deciding to present the direct route of controller [4]. The controller does not offer a direct route in the day time because high traffic volume may be impacted to interfere with other traffic, whereas, during the night time controller often offers the direct route because of the short distance, and they make sure that it will not affect other traffics. Some papers implement an optimization method such as optimal control [5] and mixed-integer optimization [6] applying speed to detect and solve aircraft conflicts. A*algorithm is well-known for route finding and has been used for separating air traffic flows by allocating 3D-trajectories [7]. Also, A*algorithm has been modified to improve Flight Management System(FMS) by solving the optimal control problem in the vertical plane [8]. The popular route from trajectories can be define to optimize the route in proper way [9] [10]. The method retrieves a popular route such as developing a Coherence Expanding algorithm together with the absorbing Markov Chain and a Maximum Probability Product [9].

Artificial Intelligence(AI) also can work as an optimizing tool in several areas such as remote sensing [11] [12] in classifying the concentrated land cover. The global approach using a genetic algorithm widespread apply with trajectories dataset for various purposes. An example of using a genetic

algorithm is to separate air traffic flows by 3D-trajectories [7]. A clustering algorithm is a kind of AI categorized as unsupervised learning. The clustering algorithm for recognizing aircraft trajectory is not a highly sophisticated method to obtain route patterns. For instance, [13] implements Density-based spatial clustering of applications with noise (DBSCAN) to cluster air traffic; the paper demonstrates the sensitivity of these algorithms to different hyper-parameters. This paper presents a K-mean clustering and Gaussian Mixture Clustering for grouping unstructured trajectories over Suvarnabhumi International Airport to be beneficial for airport capacity enhancement.

This paper is presented in the following section. Section II describes related works about trajectory over Suvarnabhumi International Airport and Clustering Method, namely K-mean Clustering and Gaussian Mixture Clustering. Section III begins with dataset explanation and data query filtering and followed by implementation of two clustering methods. Section IV is about comparative results from the two clustering algorithms and discussion. The conclusion is to draw in the final section of this paper.

II. RELATED WORK

A. K-mean Clustering

K-mean Clustering, it is also referred to as Lloyd's algorithm, mainly deploys iteration technique and proceeds by alternating between two steps: assignment step and update step [14]–[18]. Given the initial set of k mean $m_1^{(1)}, \dots, m_k^{(1)}$, and in the first step assign each observation to the cluster whose to mean has the least squared Euclidean distance as (1), this is intuitively the “nearest” mean.

$$d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + \dots + (q_n - p_n)^2} \quad (1)$$

where $\mathbf{p} = (p_1, p_2, \dots, p_n)$ and $\mathbf{q} = (q_1, q_2, \dots, q_n)$ are two points in Euclidean n -space. Then, the observation $\mathbf{S}_i^{(t)}$ compute by:

$$\mathbf{S}_i^{(t)} = \{x_p : \|x_p - m_i^{(t)}\|^2 \leq \|x_p - m_j^{(t)}\|^2 \forall j, 1 \leq j \leq k\} \quad (2)$$

where each x_p is assigned to exactly one $\mathbf{S}_i^{(t)}$, even if it could be assigned to two or more of them. The next step is update step by calculating the new means or centroids of the observations in the new clusters by (3).

$$m_i^{(t+1)} = \frac{1}{|\mathbf{S}_i^{(t)}|} \sum_{x_j \in \mathbf{S}_i^{(t)}} x_j \quad (3)$$

The iteration processes will converged when the assignments no longer change.

B. Gaussian Mixture Model

Gaussian Mixture Model(GMM) [19]–[22] is a kind of mixture density models, which assumes that each component belongs to Gaussian distribution. GMM assumes a mixture model compose of c Gaussian density components with parameters $\theta_k = \{u_k, \Sigma_k\}$ in the k th component. The probability density of x_i is formulated in GMM by:

$$p(x_i | \pi, \theta) = \sum_{k=1}^c \pi_k p(x_i | \theta_k) \quad (4)$$

where $\theta = \{\theta_1, \theta_2, \dots, \theta_c\}$ is the parameters of all the components and π_k is the mixing weight of the k th component, satisfying $\pi_k \geq 0$ and $\sum_{k=1}^c \pi_k = 1$. The k th Gaussian is denoted by:

$$p(x_i | \theta_k) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_k|}} \exp\left(-\frac{(x_i - u_k)^T \Sigma_k^{-1} (x_i - u_k)}{2}\right) \quad (5)$$

where u_k and Σ_k are the mean and the covariance matrix, respectively. The parameters $\{\theta, \pi\}$ can be estimated by iteration method like K-mean clustering algorithm by using maximizing the likelihood function using the Expectation-Maximization(EM) algorithm in terms of the following updates:

$$\omega_i^{k(t)} = \frac{\pi_k^{(t)} p(x_i | \theta_k^{(t)})}{\sum_{j=1}^c \pi_j^{(t)} p(x_i | \theta_j^{(t)})} \quad (6)$$

$$\pi_k^{(t+1)} = \frac{\sum_{i=1}^N \omega_i^{k(t)}}{N} \quad (7)$$

$$u_k^{(t+1)} = \frac{\sum_{i=1}^N (\omega_i^{k(t)} x_i)}{\sum_{i=1}^N \omega_i^{k(t)}} \quad (8)$$

$$\Sigma_k^{(t+1)} = \frac{\sum_{i=1}^N \omega_i^{k(t)} (x_i - u_k^{(t+1)})(x_i - u_k^{(t+1)})^T}{\sum_{i=1}^N \omega_i^{k(t)}} \quad (9)$$

III. METHOD

A. Dataset

According to the Federal Aviation Administration (FAA)'s policy issues, air traffic control (ATC) authorizations to persons seeking to operate aircraft that are not equipped with Automatic Dependent Surveillance-Broadcast(ADS-B) Out equipment in ADS-B airspace after January 1, 2020 [23]. ADS-B is a surveillance technology that allow an aircraft to determine its position via satellite navigation and periodically broadcasts it, enabling it to be tracked. ADS-B has two different services: “ADS-B Out” and “ADS-B In.” There are two types of ADS-B:

- Transmits at 1090 MHz using the mode-S extended squitter of the SSR transponder, with about 50 kHz of bandwidth.

- Transmits at 978 MHz(UAT, Universal Access Transceiver), using a larger bandwidth of about 1.3 MHz and also transmitting weather data.

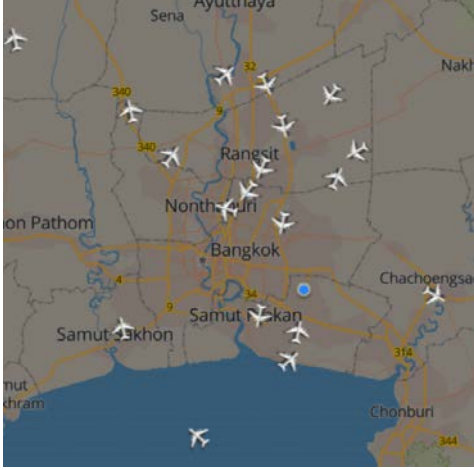


Fig. 1: Trajectory over Suvarnabhumi Airport from <https://opensky-network.org> [1]

ADS-B data used in the experiment belong to the first type of ADS-B. The data query from website “opensky-network.org” [1], from selected date on September 12, 2019, having 671 distinct icao24 (the 24-bit ICAO transponder ID) and 1822 callsign, which most airlines indicate the airline and the flight number in the callsign, but there is no unified system. The total record is 1,141,900 rows.

B. Procedures

Processes for trajectory clustering are as follows:

- Selecting the grid area that cover Suvarnabhumi International Airport, which has a roughly estimated latitude location between 13.401178 and 14.237951, and a longitude location between 100.354682 and 100.946156.
- Querying all trajectory data which flies over the grid area on September 12, 2019.
- Filtering the only trajectory from its origin(in case take off) or destination(in landing case) in the area of Suvarnabhumi Airport.
- Finding the 3-dimensional Euclidean distance of each trajectory and to assign one feature for both(K-mean and GMM) clustering algorithm denoted as “3d_distance”.
- Finding the time difference once the airplane enters the focus area till the last update position or landing completed denote as “timediff”.
- Selecting features for eight features, namely 1) first_latitude, 2) last_latitude, 3) first_longitude, 4) last_longitude, 5) first_timeupdate, 6) last_timeupdate, 7) 3d_distance, 8) timediff
- Clustering the features by K-mean and Gaussian Mixture clustering to compare results.

IV. RESULTS AND DISCUSSION

A landing case is selected to test the experiment. Fig. 2 demonstrates 1,822 flights over Suvarnabhumi International Airport on September 12, 2019. There are 419 callsigns for landing case. The majority of flights are entering to the crossing point from the northeast direction. Only small flights come from the southwest direction. From Fig. 2 is also roughly estimated that the crossing point is located in the northeast direction of the image, which has a high density of trajectories.

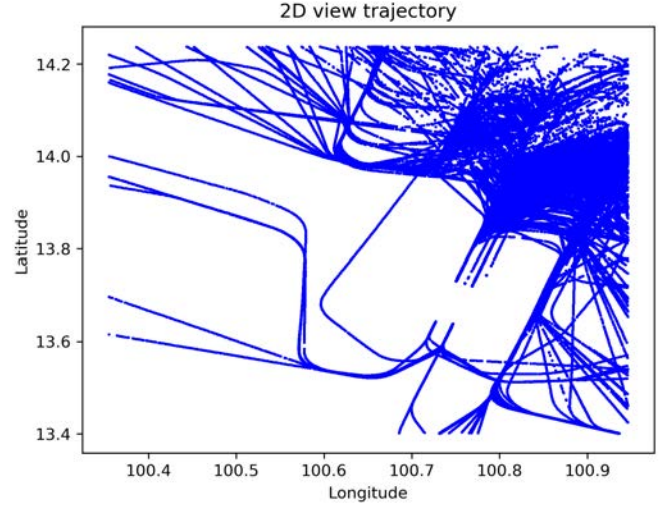


Fig. 2: Trajectory over Suvarnabhumi international Airport

To identify the number of routes, K-mean and GMM are implemented by using the Silhouette score [24], which measures how similar an object is to its cluster (cohesion) compared to other clusters (separation), as a validating tool. The Silhouette ranges from -1 to +1, where a high value indicates that the object is well matched to its cluster and poorly matched to neighboring clusters. The results show that the Silhouette score of K-mean reaches the highest score at k (number of clusters) equal to ten, while GMM has the highest score at k equal to four. K-mean and GMM produce the Silhouette score at k ranking from 3 to 10 as TABLE I.

TABLE I: Silhouette score [24]

k_cluster	K-mean	Gaussian Mixture
3	0.40143	0.40143
4	0.42977	0.43087
5	0.42452	0.37414
6	0.42318	0.38094
7	0.41967	0.41064
8	0.43717	0.26432
9	0.41784	0.38541
10	0.45274	0.36031

Fig. 3 and Fig. 4 expresses the clustering result of K-mean at k equal ten and GMM at k equal four, respectively. Moreover, all paths of those clustering demonstrate as Fig. 5 and Fig. 6.

Fig. 5a,5g,5h of K-mean clustering shows a quite similar routes, which entering to the crossing point and prepare

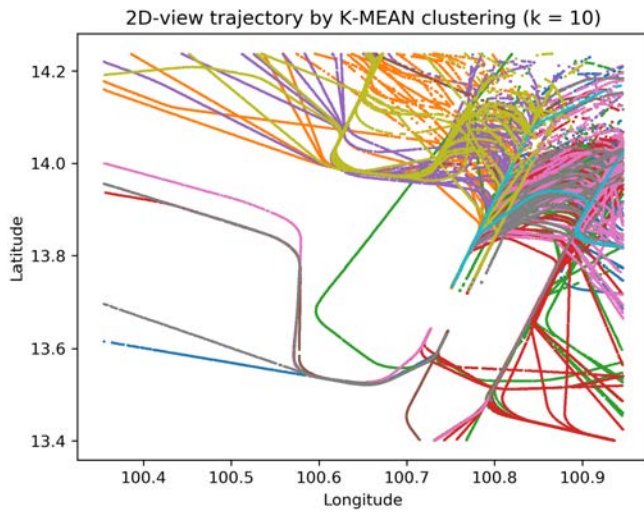


Fig. 3: 2D view trajectory by K- MEAN clustering(k=10)

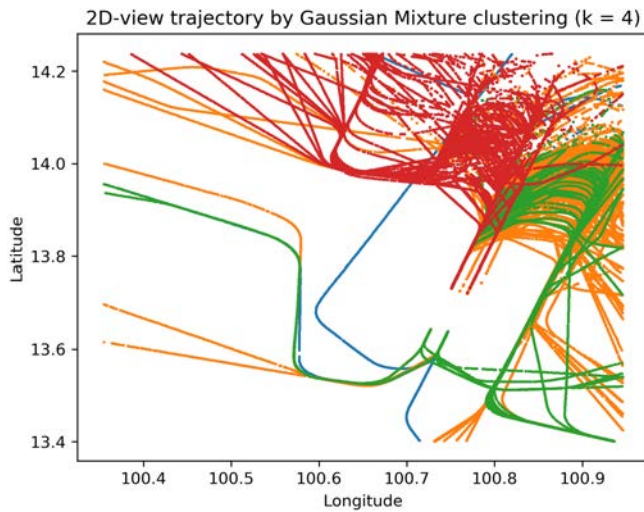


Fig. 4: 2D view trajectory by Gaussian Mixture clustering(k=4)

landing from the right side of images. Fig. 5c and 5d are also from the right side, but it has a difference from the first case that the route has a flight from southwest making it has a longer distance before entering the crossing point. In contrast, Fig. 5b,5e,5i shows another direction to entering the crossing point which it is from right side of image. Some of them are directly entering the crossing point, while others have flight as “U” shape before entering to the point.

Following K-mean, GMM clustering results show that there is a flight entering to the crossing point from both(left and right) sides. However, the flight from the right side could not distinctively separate into two cases as K-mean clustering at $k = 10$. For Fig. 5f,5j,6a could not be recognized as any route pattern as explained before.

The method is also extended and tested on other days, which is October 14, 2019, and October 15, 2019, as demonstrated

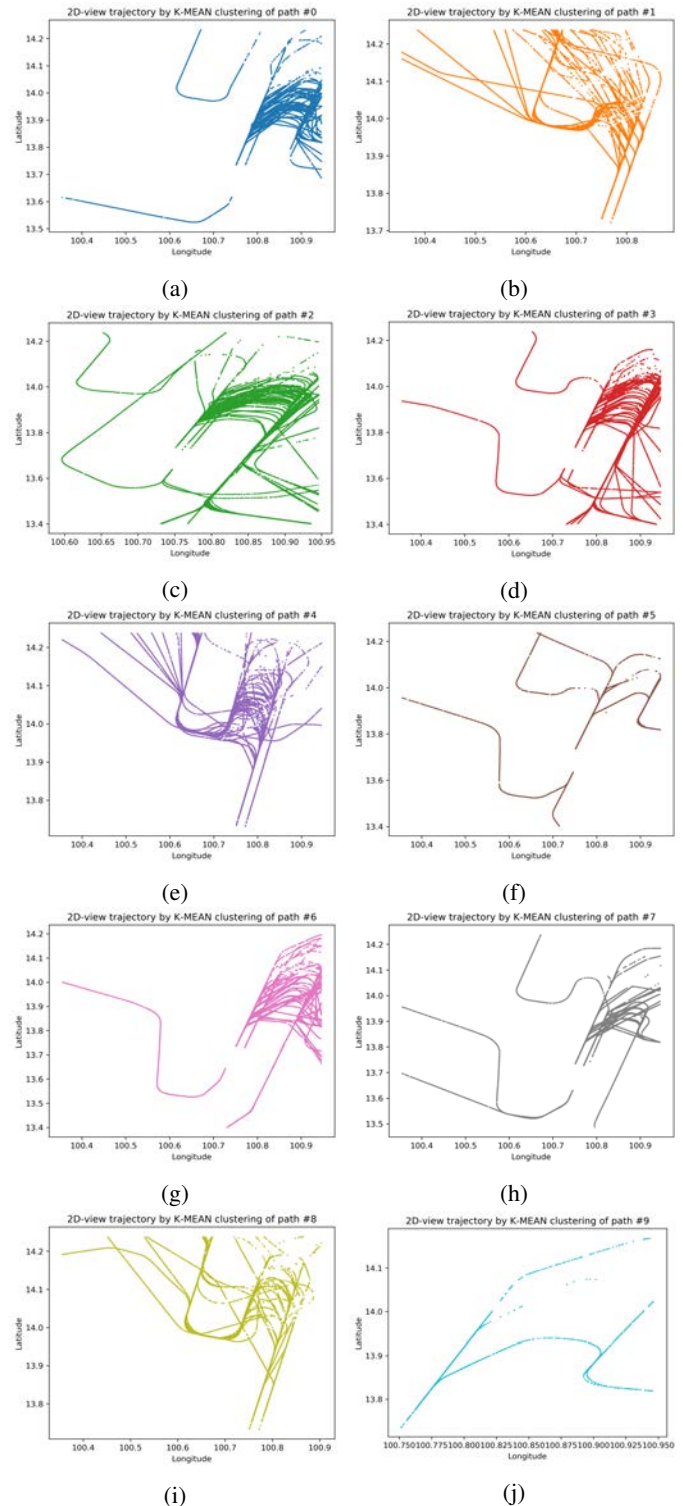


Fig. 5: 2D view each trajectory by K-MEAN clustering (k=10)

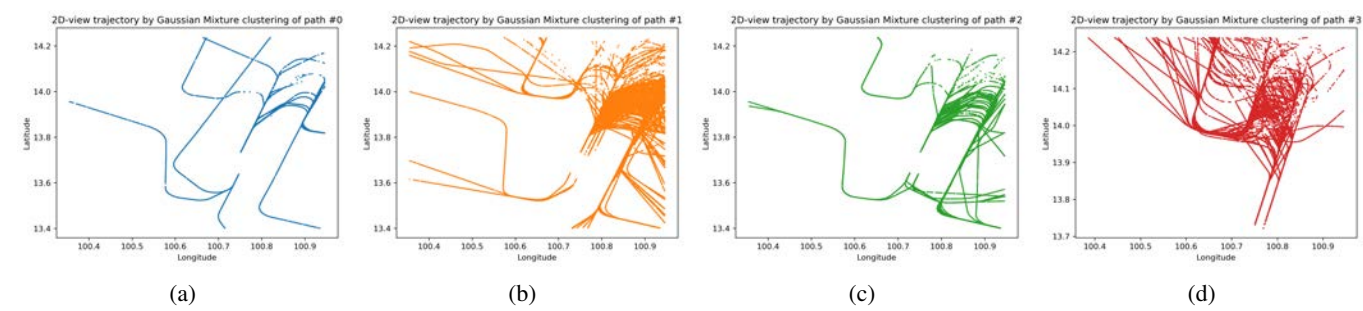


Fig. 6: 2D view each trajectory by Gaussian Mixture clustering (k=4)

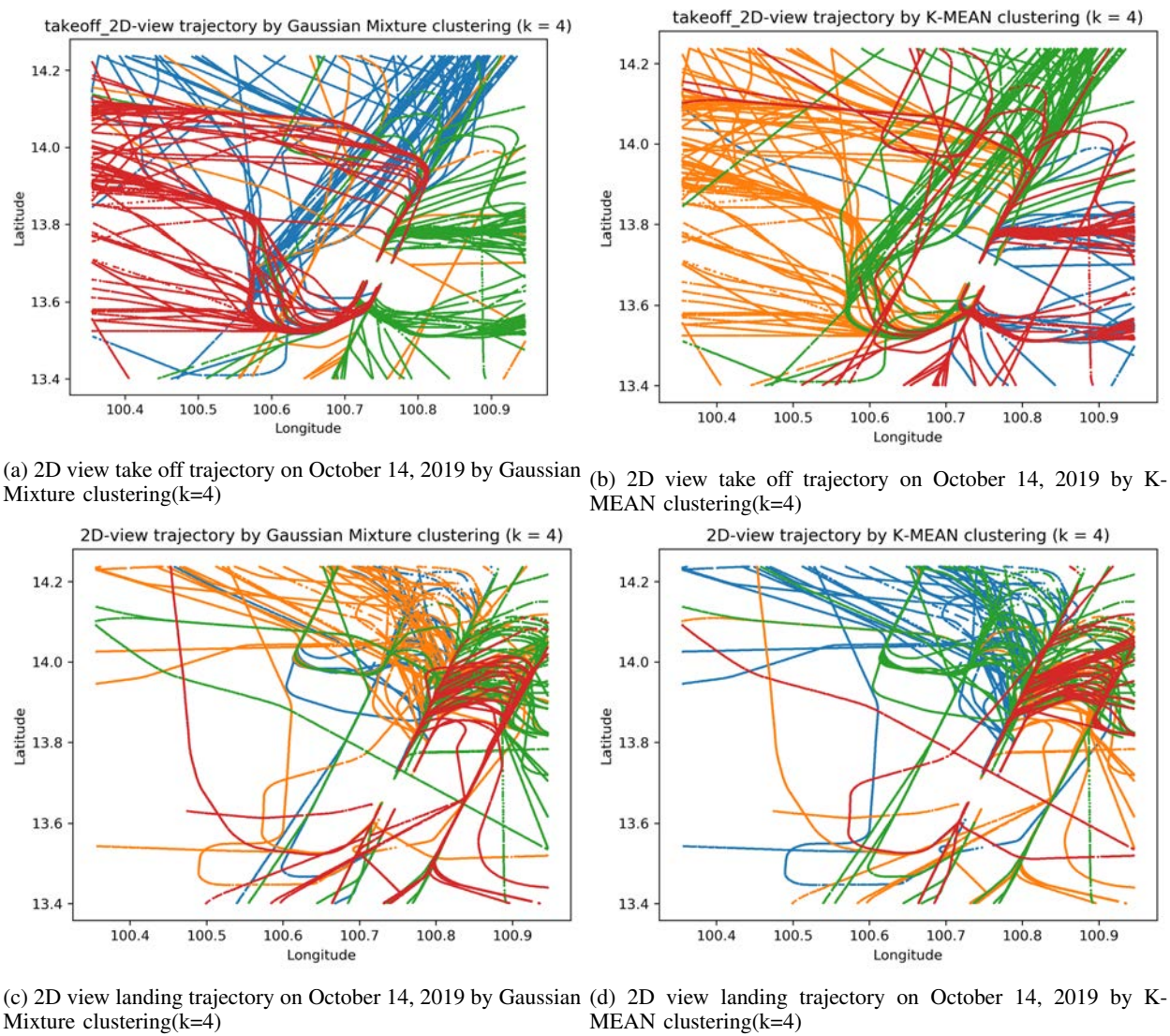
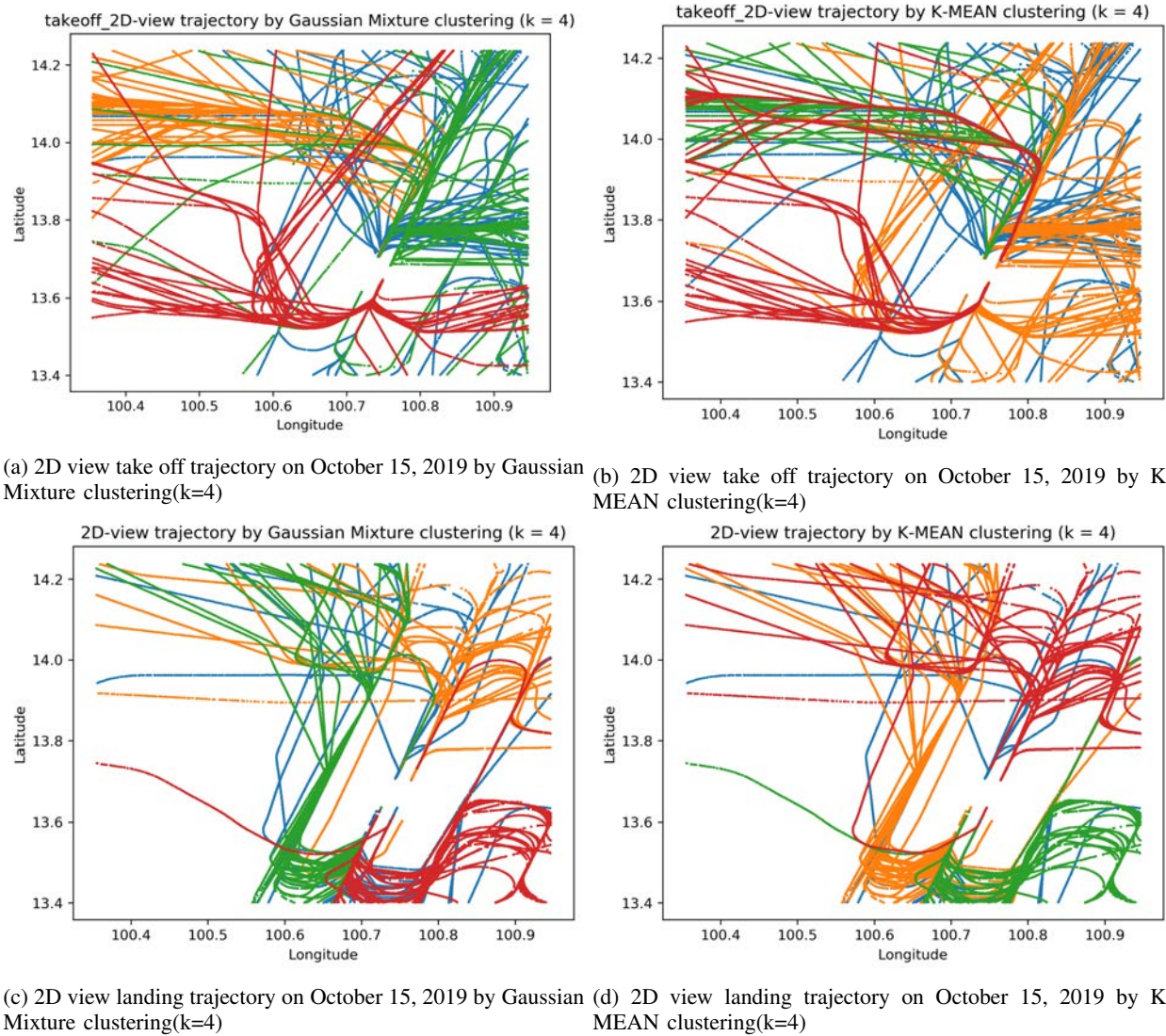


Fig. 7: 2D view each trajectory on October 14, 2019(k=4)

Fig. 8: 2D view each trajectory on October 15, 2019($k=4$)

in Fig.7 and Fig. 8 by setting the number of clustering(k) equal to 4 since the above experiments show that there is a maximum of four roots that the controllers tend to recommend for the landing of aircraft. On October 14, 2019, 472 callsigns take off and 210 callsigns for landing, while on October 15, 2019, 456 callsigns take off, and 164 callsigns for landing over Suvarnabhumi International Airport. The results corresponds with the output from September 12, 2019, which the pattern is not recognizable for both takeoff and landing cases. Therefore, further study and research should be conducted by cooperating between researchers, controllers, and regulators to improve trajectory by having a maximum capacity, performance, and securities.

V. CONCLUSION

The experiment is set to recognize aircraft trajectory for Suvarnabhumi International Airport based on ADS-B data from an available downloaded website. This experiment implements

a K-mean and Gaussian Mixture Model for the clustering trajectory. The result shows that there are mainly three route patterns, one from the left side and the other two from the right side both entering the crossing point, which located in the northeast of the airport or image. Further study and continuous research should be conducted to reach the highest airport capacity and to support the development of the aviation industry in Thailand.

ACKNOWLEDGMENT

The authors would like to thank all the staff of International Academy of Aviation Industry, KMITL, for all their support. The authors are also grateful for the recommendation from a staff officer of Aeronautical Radio of Thailand Company Limited or AEROTHAI to point out this useful and exciting topic research.

REFERENCES

- [1] M. Schäfer, M. Strohmeier, V. Lenders, I. Martinovic and M. Wilhelm, "Bringing up OpenSky: A large-scale ADS-B sensor network for research," IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks, Berlin, 2014, pp. 83-94.
- [2] Michael S. Nolan. "Fundamentals of Air Traffic Control." Thomson-Brooks/Cole, 2004.
- [3] D. Delahaye and S. Puechmorel, "Air traffic complexity based on dynamical systems," 49th IEEE Conference on Decision and Control (CDC), Atlanta, GA, 2010, pp. 2069-2074. doi: 10.1109/CDC.2010.5718004
- [4] A. Pasini, P. Notry and D. Delahaye, "Direct Route Optimization for Air Traffic Management Improvement," 2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC), London, 2018, pp. 1-9. doi: 10.1109/DASC.2018.8569362
- [5] Sophie Constans, Bastian Fontaine, and Rémy Fondacci "Minimizing Potential Conflict Quantity with Speed Control." in INO Workshop, 2005.
- [6] Sonia Cafieri and Nicolas Durand. 2014. "Aircraft deconfliction with speed regulation: new models from mixed-integer optimization." J. of Global Optimization 58, 4 (April 2014), 613-629. DOI=<http://dx.doi.org/10.1007/s10898-013-0070-1>
- [7] D. Gianazza and N. Durand, "Separating air traffic flows by allocating 3D-trajectories," The 23rd Digital Avionics Systems Conference (IEEE Cat. No.04CH37576), Salt Lake City, UT, USA, 2004, pp. 2.D.4-21. doi: 10.1109/DASC.2004.1391275
- [8] R. A. Altava, J. C. Mere, D. Delahaye and T. Miquel, "Flight Management System Pathfinding Algorithm for Automatic Vertical Trajectory Generation," 2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC), London, 2018, pp. 1-9. doi: 10.1109/DASC.2018.8569254
- [9] Z. Chen, H. T. Shen and X. Zhou, "Discovering popular routes from trajectories," 2011 IEEE 27th International Conference on Data Engineering, Hannover, 2011, pp. 900-911. doi: 10.1109/ICDE.2011.5767890
- [10] Ling-Yin Wei, Yu Zheng, and Wen-Chih Peng. 2012. Constructing popular routes from uncertain trajectories. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '12). ACM, New York, NY, USA, 195-203. DOI: <https://doi.org/10.1145/2339530.2339562>
- [11] P. Insom et al., "A Support Vector Machine-Based Particle Filter Method for Improved Flooding Classification," in IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 9, pp. 1943-1947, Sept. 2015.
- [12] P. Insom, C. Cao, P. Boonsrimuang, S. Bao, W. Chen and X. Ni, "A support vector machine-based particle filter for improved land cover classification applied to MODIS data," 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, 2016, pp. 775-778.
- [13] C. E. V. Gallego, V. F. Gómez Comendador, F. J. Saez Nieto and M. G. Martinez, "Discussion On Density-Based Clustering Methods Applied for Automated Identification of Airspace Flows," 2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC), London, 2018, pp. 1-10. doi: 10.1109/DASC.2018.8569219
- [14] https://en.wikipedia.org/wiki/K-means_clustering
- [15] J. Qi, Y. Yu, L. Wang and J. Liu, "K*-Means: An Effective and Efficient K-Means Clustering Algorithm," 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom), Atlanta, GA, 2016, pp. 242-249. doi: 10.1109/BDCloud-SocialCom-SustainCom.2016.46
- [16] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 881-892, July 2002. doi: 10.1109/TPAMI.2002.1017616
- [17] G. A. Wilkin and X. Huang, "K-Means Clustering Algorithms: Implementation and Comparison," Second International Multi-Symposiums on Computer and Computational Sciences (IMSCCS 2007), Iowa City, IA, 2007, pp. 133-136. doi: 10.1109/IMSCCS.2007.51
- [18] A. Kumar, R. Sinha, V. Bhattacherjee, Daya Shankar Verma and S. Singh, "Modeling using K-means clustering algorithm," 2012 1st International Conference on Recent Advances in Information Technology (RAIT), Dhanbad, 2012, pp. 554-558. doi: 10.1109/RAIT.2012.6194588
- [19] H. Sun, Z. Liu and L. Kong, "A Document Clustering Method Based on Hierarchical Algorithm with Model Clustering," 22nd International Conference on Advanced Information Networking and Applications - Workshops (aina workshops 2008), Okinawa, 2008, pp. 1229-1233. doi: 10.1109/WAINA.2008.45
- [20] Y. Zhao, A. K. Shrivastava and K. L. Tsui, "Regularized Gaussian Mixture Model for High-Dimensional Clustering," in IEEE Transactions on Cybernetics, vol. 49, no. 10, pp. 3677-3688, Oct. 2019. doi: 10.1109/TCYB.2018.2846404
- [21] Y. Li, J. Zhang, Z. Ma and Y. Zhang, "Clustering Analysis in the Wireless Propagation Channel with a variational Gaussian Mixture Model," in IEEE Transactions on Big Data. doi: 10.1109/TBDATA.2018.2840696
- [22] W. Cai, L. Lei and M. Yang, "A Gaussian Mixture Model-based clustering algorithm for image segmentation using dependable spatial constraints," 2010 3rd International Congress on Image and Signal Processing, Yantai, 2010, pp. 1268-1272. doi: 10.1109/CISP.2010.5647653
- [23] DEPARTMENT OF TRANSPORTATION Federal Aviation Administration, "Statement of Policy for Authorizations to Operators of Aircraft That are Not Equipped With Automatic Dependent Surveillance-Broadcast (ADS-B) Out Equipment", Federal Register, Rules and Regulations, April 1, 2019 (<https://www.federalregister.gov/documents/2019/04/01/2019-06184/statement-of-policy-for-authorizations-to-operators-of-aircraft-that-are-not-equipped-with-automatic>)
- [24] Rousseeuw, Peter. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." (1987).



Patcharin Kamsing received doctoral degree from the University of Chinese Academy of Sciences, Beijing, China sponsor by CAS-TWAS president fellowship. She currently is lecturer and assistant dean of International Academy of Aviation Industry, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. Her research is image processing and remote sensing applications.



Peerapong Torteeka received doctoral degree in Celestial Mechanics and Applied Astrometry Engineering at National Astronomical Observatories of Chinese Academy of Sciences (NAOC), University of Chinese Academy of Sciences, Beijing, China. He currently works at National Astronomical Research Institute of Thailand, research interests include passive optical-based small and dim space debris recognition and tracking system, moving object extraction, autonomous detection, robotics telescope for the space observation.



Soemsak Yooyen currently is Dean of International Academy of Aviation Industry, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. He received Doctor of Philosophy from Mie University, Japan. His research is about materials science.



Siriporn Yenpiem currently is strategic planning manager at AEROTHAI - Aeronautical Radio of Thailand Ltd, Bangkok, Thailand. She received doctor of philosophy from Dhurakij Pundit University International College, Thailand.



Daniel Delahaye is a Professor at ENAC (Ecole Nationale de l'Aviation Civile) (France). His research interests lie in optimization methods applied to the problems of air traffic management.



Philippe Notry currently is head of ATC training management unit ENAC (Ecole Nationale de l'Aviation Civile) (France). His interests lie in air traffic management.



Thaweerath Phisannupawong is currently a bachelor degree student, major in Aeronautical Engineering and Commercial Pilot (International Program), from the Department of Aeronautical Engineering, International Academy of Aviation Industry. His research interests are included the image processing and application of deep learning on aeronautical engineering.



Sittiporn Channumsin currently is researcher in Geo-Informatics and Space Technology Development Agency (GISTDA). He received doctoral degree from the university of Glasgow. His research is about space traffic management.