

Stochastic conformal anomaly detection and resolution for air traffic control

Hong-Cheol Choi ^{*}, Chuhao Deng, Hyunsang Park, Inseok Hwang

School of Aeronautics and Astronautics, Purdue University, West Lafayette, 47907, IN, USA



ARTICLE INFO

Keywords:

Anomaly detection
Anomaly resolution
Uncertain data
Conformal prediction
Aviation safety
Decision support tool

ABSTRACT

Safety is of the utmost importance in the air traffic system. In recent years, data-driven algorithms have emerged to identify anomalous and potentially unsafe operations based on machine learning techniques. Although many algorithms have shown notable progress in anomaly detection, they have hardly considered the fact that data can be corrupted by noise and uncertainty (e.g., navigation system error) can lead to frequent misdetection and false alarms, which could disturb air traffic controllers and result in system performance degradation. Therefore, an accurate and reliable assessment of emerging safety risks that accounts for and alleviates the effect of uncertainty in data is required for safe and efficient airspace operations. To achieve this goal, this paper proposes a conformal prediction-based framework that explicitly examines uncertainty for reliable anomaly detection and learns online from new streaming data. In addition to supporting the monitoring task of air traffic controllers, the proposed method takes one step forward and provides support for the control task, by offering a resolution strategy when anomaly probability violates the predefined threshold. The proposed method is demonstrated with real air traffic data, called automatic dependent surveillance-broadcast data.

1. Introduction

The volume of operations in aviation has steadily increased, and the Federal Aviation Administration (FAA) forecasts a 2.0% annual growth rate for domestic passengers over the following 20 years ([Federal Aviation Administration, 2020a](#)). This places paramount importance on air traffic control (a service offered by ground-based Air Traffic Controllers (ATCs) to direct the aircraft safely and expedite air traffic flow) and diagnostics related to air traffic and aircraft. However, the current system is unable to follow the growing demand as it is heavily reliant on manual support including manual inspection and communication between pilots and airport operators ([International Air Transport Airport, 2018](#)). ATCs also are burdened with prohibitive responsibility which increases the risk of human error that leads to unsafe operations. Therefore, diagnostics related to air traffic and aircraft have become more important than ever before. In the past, accidents have been the primary sources for identifying problems and developing mitigation plans ([Logan, 2008](#)). However, the industry has shifted towards a more proactive method where potentially unsafe occasions, i.e., anomalies, are identified in advance, and mitigation strategies are established to prevent real accidents. In this regard, it is essential to develop automated diagnostics tools to analyze and detect unusual events that could pose operational or safety-related risks to ensure high efficiency and safety of airspace operations.

Physics-based approach to anomaly detection has been first used in aviation by generating a governing physics-based model which specifies nominal behaviors so that an algorithm can detect specific occurrences that deviate from the model ([Narasimhan](#)

* Corresponding author.

E-mail address: choi642@purdue.edu (H.-C. Choi).

and Brownston, 2007; Zhu et al., 2016). For instance, Kalman filters and observer-based estimation have been employed to estimate the dynamical system behavior based on an explicit mathematical model (Valasek and Chen, 2003; De Loza et al., 2015). However, the limitations of this approach are that the complete modeling of an entire aircraft system, including all the subsystems, is virtually impossible, and generalization of these models is limited due to strict system dependency (Yoon and MacGregor, 2000; Hou and Wang, 2013). Meanwhile, the data-driven approach has emerged as a suitable alternative with the advent of new data technologies. Aviation modernization efforts have led to better accessibility of aviation operational data and an increase in data sources and volume (Federal Aviation Administration, 2020b), driving the recent increase in data-driven anomaly detection algorithms based on various data mining and machine learning techniques (Das et al., 2010; Kong et al., 2014; Puranik and Mavris, 2018; Deshmukh and Hwang, 2019; Choi et al., 2022). In contrast to the physics-based approach, the data-driven approach can benefit from historical air traffic data and flexibly incorporate environmental uncertainty. However, one of the limitations of those algorithms is that they can only implicitly consider uncertainty inherent in deterministic data rather than explicitly considering the uncertainty from measurement and processing noise, which can hinder the accuracy of anomaly detection.

To tackle this limitation, we propose a stochastic anomaly detection and resolution framework based on Conformal Anomaly Detection (CAD) (Laxhammar and Falkman, 2013) by extending CAD to aviation data and incorporating uncertainty explicitly. The contributions of the proposed framework can be summarized in three. Firstly, by designing a new measure that takes account the uncertainty, uncertainty is explicitly incorporated and examined for anomaly detection in the form of probability mass function, thereby enhancing the situational awareness of ATCs. Secondly, the proposed method allows learning online from streaming air traffic data to reflect continuously evolving behaviors in airspace operations and thus can provide the updated anomaly detection criterion in real-time. Lastly, the proposed framework can automatically recommend a resolution strategy when the anomaly probability exceeds a threshold, which supports both ATCs' monitoring and control tasks and thus significantly contributes to alleviating their workload.

The remainder of the paper is organized as follows: Section 2 reviews the existing literature regarding techniques for detecting anomalies and discusses its limitations. Section 3 presents the technical details of the proposed framework for anomaly detection and resolution under uncertainty. Section 4 presents experimental results and analysis with the real air traffic data, called the automatic dependent surveillance-broadcast data, to evaluate performance of the proposed method and provides a discussion on the insights gained. Finally, Section 5 concludes the paper and offers avenues for future works.

2. Literature review

In this section, we describe current anomaly detection algorithms and summarize research gaps we find in the literature, i.e., the uncertainty in data samples from measurement and processing noise, online learning for the air traffic domain, and anomaly resolution.

2.1. Anomaly detection in aviation

Identifying cases that are nonconforming to expected behavior in data is known as anomaly detection (Pimentel et al., 2014), and an anomaly in the air traffic domain is defined as a rare event with a potentially safety-related impact (Deshmukh and Hwang, 2019). Many models that can segregate abnormal data from normal ones have been actively proposed to identify risky behaviors in aviation data with complex characteristics such as high dimensionality, heterogeneity (continuous and discrete variables), and temporality. These models have been successfully implemented to find meaningful events such as runway excursions, conflict situations, and go-around. If labels that distinguish between normal and abnormal data are available, a supervised learning approach can be taken. Accordingly, some anomaly detection models have been developed based on supervised machine learning techniques such as k-nearest neighbors, support vector machines, Bayesian networks, neural networks, and decision trees (Kong et al., 2014; Ghorbani et al., 2009).

Unsupervised learning, however, has been more extensively studied to detect anomalies in aviation data since extensive aviation data typically lack labels indicating whether a flight is normal or abnormal. A distance-based approach that utilizes the distances or the densities of regions from the data have been taken for unsupervised algorithms (Ester et al., 1996; Dani et al., 2015; Shin and Hwang, 2017). Another widely used approach, the One-Class Support Vector Machine (OCSVM), aims to design a hyperplane so that all abnormal cases are on one side and normal cases are on the other (Schwabacher et al., 2009; Ma and Perkins, 2003). Moreover, the Multiple Kernel Anomaly Detection (MKAD) algorithm that can take sequences of both discrete and continuous variables has been proposed to effectively identify operationally significant situations, such as abnormal flight approaches, based on kernel functions and OCSVM (Das et al., 2010). Clustering-based methods have also been studied, such as Cluster-based Anomaly Detection (ClusterAD) (Li et al., 2011, 2015) based on DBSCAN (Ester et al., 1996), which can detect anomalies in an airline dataset for takeoff and approach operations. This group has subsequently proposed ClusterAD-DataSample, a Gaussian Mixture Model (GMM)-based method that can detect anomalous temporal samples during a particular flight phase (Li et al., 2016). The DBSCAN and OCSVM-based algorithm by adopting energy metrics has been proposed for anomaly detection and quantitative analysis (Puranik and Mavris, 2018; Corrado et al., 2021). The authors in Deshmukh and Hwang (2019), Deshmukh et al. (2021) have developed a Temporal logic-based Anomaly Detection algorithm (TempAD) to generate human-readable mathematical expressions, i.e., anomaly detection bounds, from surveillance data in metropolex terminal airspace operations. Deep learning-based techniques such as autoencoder and artificial neural networks have recently drawn more attention in the air traffic domain (Nanduri and Sherry, 2016; Olive and Basora, 2020; Memarzadeh et al., 2020).

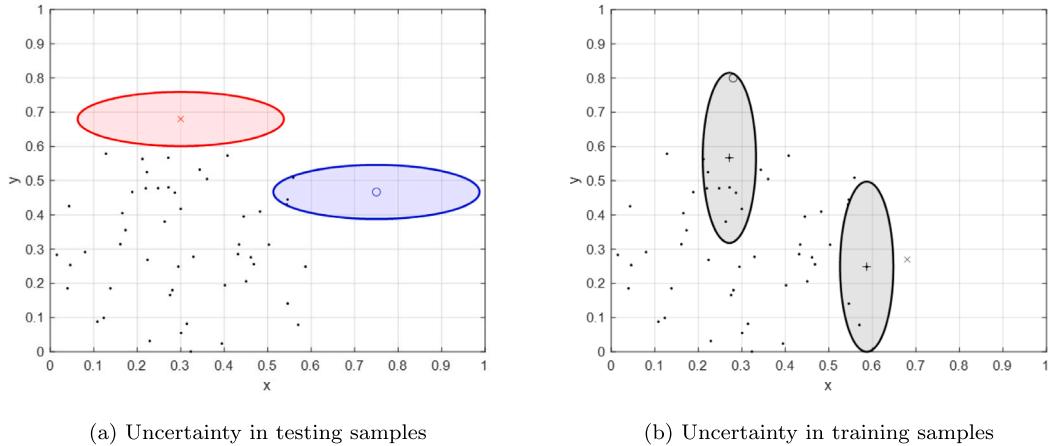


Fig. 1. Anomaly detection issues with uncertain data.

Although many efforts have achieved significant progress, there are still some limitations. Existing anomaly detection algorithms in aviation do not take the uncertainty from measurement and processing noise in the data explicitly, and the online model update is rarely considered except for the study in [Deshmukh and Hwang \(2019\)](#). Even so, only recursive updates are performed using mini-batch data rather than real-time updates corresponding to streaming data. Finally, most studies focus only on monitoring tasks (anomaly detection) and do not extend to control tasks (anomaly resolution).

2.2. Anomaly detection under uncertainty

In recent years, many advanced technologies have been developed to store and record large quantities of data. However, in many cases, the data may still contain errors or may only be partially complete, which is called uncertain data. For example, sensor networks typically create large amounts of uncertain datasets, and the errors may be a result of the imperfections in the hardware used in the data collection process such as the noise in sensor inputs or errors in wireless transmission. The results of data mining applications are known to be negatively affected by the uncertainty in the data ([Aggarwal and Philip, 2008](#)). Due to the possibility that the uncertainty added to a data point could cause it to behave in an outlier-like manner, the outlier detection problem is more difficult for uncertain cases. To illustrate the uncertainty effect in anomaly detection, let two samples be in the dataset along with the corresponding uncertainty areas, as shown in [Fig. 1\(a\)](#). If we did not consider the uncertainty, we would be more likely to label the blue point (marked as a circle) as an anomaly since it seems further away from the given dataset (in black). However, the red point (marked as a cross) is much more likely to be an anomaly because the corresponding uncertainty area does not overlap with the dense regions of the data. Many efforts have been put into handling this type of data. For instance, to quantify the probability that a given uncertain data point is drawn from a dense region, the authors in [Aggarwal and Yu \(2008\)](#) propose the concept of η -probability, which is defined as the probability that the uncertain data point lies in a region with (overall data) density at least η . Although it can remove the uncertainty effect added to testing samples, extending this algorithm to more complex data, such as aviation data, is not trivial.

More recently, stochastic anomaly detection algorithms have been developed to handle uncertainty, based on the belief that data-driven techniques can intrinsically address uncertainty, as the data is inherently corrupted with them. Existing stochastic algorithms first estimate the probability density function based on orthogonal series approximation and probability distribution estimation (e.g., kernel density estimation) ([Jansson et al., 2015](#)) or the energy consumption modeling of each system mode in the hybrid system ([Windmann et al., 2013](#)). And then, the algorithms detect if the new incoming data point is located outside the predefined confidence interval from the estimated distribution. Similarly, by combining Recurrent Neural Network (RNN) and GMM, a stochastic RNN has been proposed to detect anomalies based on the reconstruction probability ([Su et al., 2019](#)). However, note that these stochastic anomaly detection algorithms only focused on the estimation or inference of the resultant distribution of stochastic variables corrupted by noise and cannot discern the two sources of uncertainty in data: the underlying distribution versus the additive noise. The data sample is still assumed to be deterministic, which can adversely affect the performance of anomaly detection.

2.3. Limitations of existing literature

As mentioned in previous sections, existing anomaly detection algorithms in aviation mainly stick to a deterministic approach, disregarding uncertainty attributed to the noise in sensor inputs. The treatment of uncertainty remains important even in future air transportation systems, which are characterized by the transition from radar-based operations to satellite-based navigation and surveillance. Satellite-based operations with the Automatic Dependent Surveillance-Broadcast (ADS-B) system could enhance aircraft

Table 1
Measurement uncertainty in ADS-B and SBAS.

Typical operation	ADS-B		SBAS	
	Horizontal accuracy	Vertical accuracy	Horizontal accuracy	Vertical accuracy
En-route	4 nmi	N/A	2 nmi	N/A
Terminal	1 nmi	N/A	0.4 nmi	N/A
Approach/Departure	0.3 nmi	N/A	220 m	N/A
Precision approach	30 m	45 m	16 m	20 m

navigation, surveillance, and position-tracking accuracy (Sankararaman et al., 2017; Zhang and Mahadevan, 2019). One of these efforts is adopting a Satellite-Based Augmentation System (SBAS) for performance-based navigation (ICAO, 2016) to provide a more precise navigation service to meet the future increasing demands. For this technology, International Civil Aviation Organization (ICAO) defines the required performance (ICAO, 2006). For instance, the required accuracy in the horizontal dimension is 0.4 nmi. The measurement uncertainty in ADS-B and SBAS is shown in Table 1. Therefore, since state-of-the-art technology will still have uncertainty though it could be smaller, possible errors should be considered explicitly in safety-related assistant tools to properly assess the safety risks of the current flight and avoid incorrect decisions. On the other hand, the outlier detection algorithm (Aggarwal and Yu, 2008) presented in Section 2.2 is not directly applicable to air traffic surveillance data characterized by high dimensionality, heterogeneity, and temporality. In addition, uncertainty in training samples is not incorporated into the existing method, although it could have much effect on the performance. For instance, in Fig. 1(b), the blue circle appears more abnormal than the red cross in terms of the distance to the high-density area, but when the uncertainty of training samples is taken into account (shaded in gray), the red sample is more likely to be abnormal because it is outside the corresponding uncertainty area. This is particularly important in the sense that the uncertainty added to other training samples could alter the overall data distribution.

Meanwhile, existing anomaly detection models are typically trained by batch historical datasets, making their models static under a situation where streaming aviation data continues to evolve in real-time. In other words, they rarely consider using newly incoming data to incrementally update the previous model for the next prediction of anomalies, which makes it challenging to reflect continuously evolving behaviors of air traffic in the airspace. For iterative mini-batch data learning (Deshmukh and Hwang, 2019), it is unclear how to determine the appropriate time for such an update. Hence, it gives rise to the need for online learning algorithms to sequentially update the model parameters as each new training sample is observed.

Additionally, to reduce the workload on ATCs and pilots, automation or decision support tools can be developed, which aid the decision-making process for anomaly resolution. If a system can automatically detect and correct anomalies, it can go a long way in reducing human involvement in keeping the system safe. In this regard, anomaly detection and resolution have been widely studied in other areas such as cloud, log, and firewall (Ahad et al., 2015; Saâdaoui et al., 2017; Mahindru et al., 2021), while existing algorithms in aviation mainly focus on anomaly detection. Similar to conflict detection and resolution (Matsuno et al., 2015; Hernández-Romero et al., 2020) in the air traffic domain, anomaly detection should be extended to anomaly resolution to provide an advisory strategy to rectify potential safety-related issues.

To fill the above mentioned research gap, a stochastic conformal anomaly detection and resolution framework is proposed in this paper. Firstly, to quantify uncertainty in training and testing samples, hybrid estimation is employed as it can represent the probability distributions of the state estimate. Trajectory pattern classification based on Long Short-Term Memory (LSTM) is adopted to compare a new sample with samples with similar properties, thereby alleviating computational costs. Secondly, to explicitly take uncertainty into account explicitly and achieve online learning for anomaly detection, we design a new nonconformity measure based on conformal prediction. Lastly, we obtain a probability mass function (pmf) from which an anomaly probability can be computed at the current timestep. If an anomaly probability exceeds a predefined threshold, the algorithm automatically provides an anomaly resolution strategy generated by hybrid trajectory prediction. The technical details of the proposed framework will be presented in the next section.

3. Algorithm development

The overall architecture of the proposed algorithm is shown in Fig. 2, which consists of offline training and online operation. In the training stage, a set of time-series data (i.e., air traffic surveillance data) is preprocessed and then used for the training of (i) classification model and (ii) hybrid trajectory prediction. In the online operation stage, streaming data is sequentially fed into the processing, monitoring, and control steps. The data to a trajectory pattern classification model is a four-Dimensional (4D) trajectory which can be expressed as:

$$X_t = [\chi_1^1, \chi_1^2, \dots, \chi_1^f, \dots, \chi_t^1, \chi_t^2, \dots, \chi_t^f] \quad (1)$$

where t represents the current timestep, f represents the feature dimension (latitude, longitude, and altitude in this paper), and χ represents the corresponding feature value. This 4D trajectory is split into the horizontal and vertical dimensions (X_t^H, X_t^V), and two dimensions are processed separately for the following steps: hybrid estimation, stochastic conformal anomaly detection, and hybrid trajectory prediction. As an output of the proposed framework, an anomaly probability and an anomaly resolution strategy (if needed) are provided.

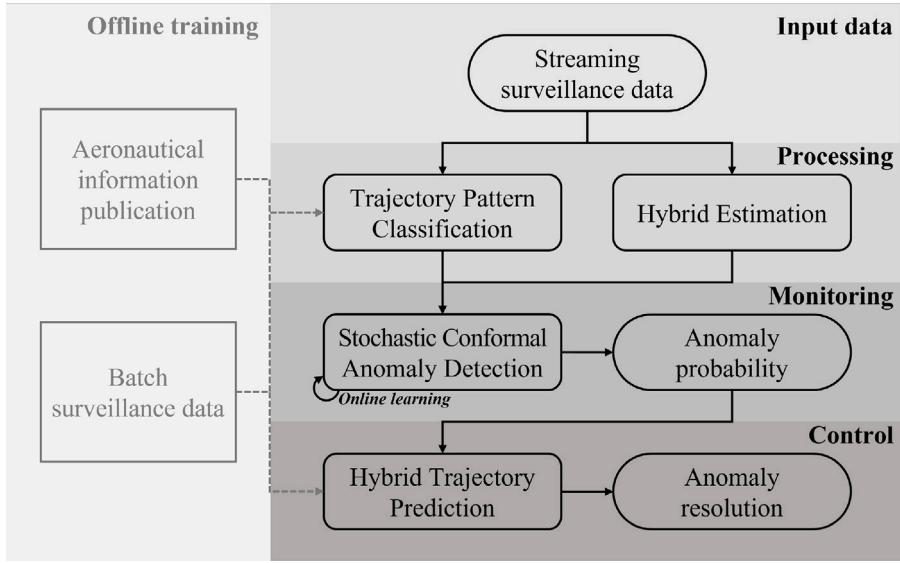


Fig. 2. The schematic diagram of the proposed framework.

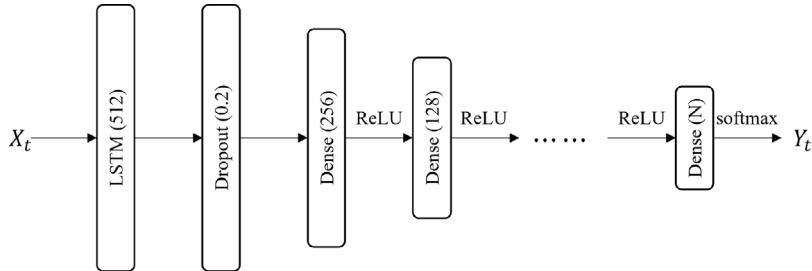


Fig. 3. The neural network architecture for trajectory pattern classification.

3.1. Trajectory pattern classification and hybrid estimation

Given that multiple patterns of trajectories have already been identified by clustering methods (Gariel et al., 2011; Deng et al., 2022), a real-time trajectory pattern classification model is required to classify a new incoming flight into one of the identified patterns. The trajectory pattern classification problem can be formulated as a supervised multi-class classification problem, and the solution is provided in the form of a vector (Y_t) with one element equal to 1 and the rest equal to 0. The position of the non-zero element represents the pattern number that the current flight belongs to. To solve this problem, we use LSTM (Deng et al., 2022; Hochreiter and Schmidhuber, 1997) to utilize its capability in memorizing the information in the preceding timesteps. This capability can play an important role in the classification problem since it can address temporal correlation in time-series data. More importantly, LSTM-based methods can effectively deal with an incomplete trajectory in real-time. As shown in Fig. 3, the classification model architecture consists of multiple layers: a layer of LSTM, a dropout layer, several fully-connected layers with the Rectified Linear Unit (ReLU) activation function, and a fully-connected layer with the softmax function. Following a standard procedure for hyperparameter tuning, the hyperparameters such as the dropout rate and size of each layer are determined based on the categorical cross-entropy loss value across multiple tests. Finally, by using the trained model of the aforementioned architecture, the probabilities for trajectory patterns can be provided as an output. Based on the probabilities, the pattern that the current flight belongs to can be determined as the one with the highest probability in real-time.

By utilizing hybrid estimation, we can filter out a considerable portion of inaccuracies and errors from the measurement data based on the aircraft's dynamics. To accurately account for the aircraft's motion in the airspace, which involves frequent changes of flight modes, e.g., from straight flight to turning or from altitude hold to altitude change, the aircraft's dynamics is modeled as a Stochastic Linear Hybrid System (SLHS) (Seah and Hwang, 2009; Liu and Hwang, 2011), in which the discrete dynamics describes the transitions in the flight modes and the continuous dynamics represents how the aircraft's continuous states (such as position and speed) evolve over time in each mode. For each discrete state, or mode, $q(t) \in \mathbb{Q} := \{1, \dots, n_d\}$ where n_d is the total number of modes, the SLHS model is given by:

$$x(t+1) = A_{q(t)}x(t) + E_{q(t)}\omega_{q(t)}(t) \quad (2)$$

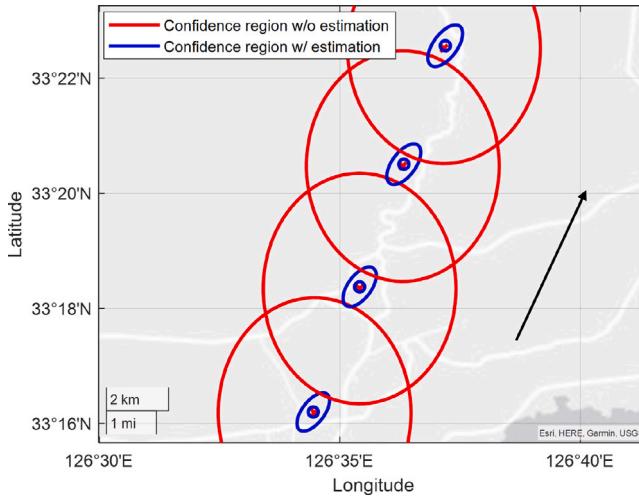


Fig. 4. Comparison of the confidence regions according to whether or not estimation was performed.

$$y(t) = Cx(t) + v(t) \quad (3)$$

where at given timestep t , $x(t) \in \mathbb{R}^l$ is the continuous state, $y(t) \in \mathbb{R}^p$ is the measurement, and $\omega_{q(t)}(t)$ and $v(t)$ are the process noise and the measurement noise, assumed to be zero-mean white Gaussian noises with the covariances $Q_{q(t)}(t)$ and $R(t)$, respectively. The system matrices $A_{q(t)}$, $E_{q(t)}$ and $C_{q(t)}$ are with proper dimensions for each mode $q(t)$. The discrete dynamics is represented by a constant mode transition probability matrix

$$\Pi = [\pi_{ij}]_{i,j \in \mathbb{Q}} \quad (4)$$

where π_{ij} is the constant mode transition probability from mode i to mode j and $\sum_{i \in \mathbb{Q}} \pi_{ij} = 1$ for $j \in \mathbb{Q}$.

With the SLHS model of an aircraft, we use an estimation algorithm called Residual-Mean Interacting Multiple Models (RM-IMM) to estimate the aircraft states (Hwang et al., 2006). The RM-IMM algorithm can reduce false mode estimation by increasing the difference between the likelihood of the correct mode and those of the other modes, thereby giving more distinct mode probabilities that can reduce false mode estimation. Based on the aircraft's dynamics corresponding to the correctly identified mode, the RM-IMM can produce a state estimate with a covariance which is significantly smaller than that from the measurement alone, and an illustrative example is shown in Fig. 4.

3.2. Stochastic conformal anomaly detection

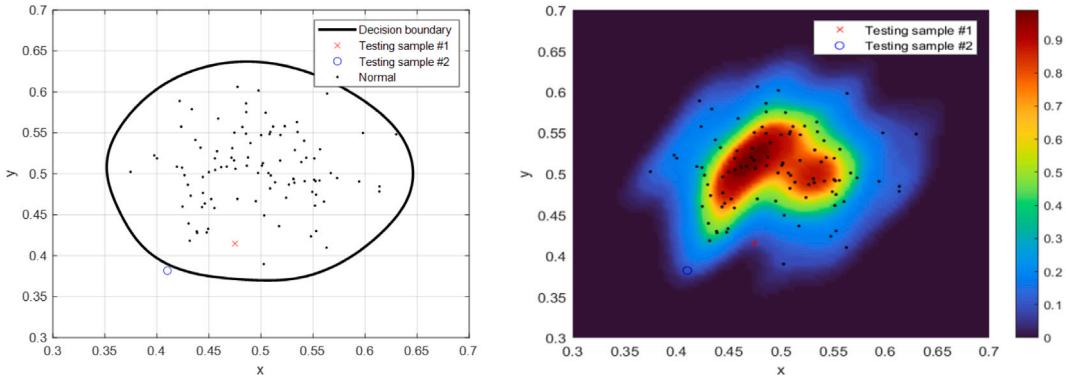
Conformal prediction is a technique for providing valid measures of confidence for individual predictions made by machine learning algorithms. Suppose we have a set of similar trajectories, $\{Z_i\}_{i=1}^n$ where $Z_i = \{z_i(t)\}_{t=0}^T$ is an aircraft trajectory, $z_i(t)$ is a track point at timestep t , and T is the final timestep. Based on conformal prediction (Shafer and Vovk, 2008), the p -value for a new incoming trajectory Z_{n+1} with respect to $\{Z_i\}_{i=1}^n$ is estimated by

$$p_{n+1} = \frac{|\{j = 1, \dots, n : \alpha_j \geq \alpha_{n+1}\}|}{n+1} \quad (5)$$

where $|A|$ is the cardinality of the set A and α_j is a suitable NonConformity Measure (NCM). Conformal Anomaly Detection (CAD) (Laxhammar and Falkman, 2010) is an extension of the framework of conformal prediction. Analogous to statistical hypothesis testing, ϵ corresponds to an upper bound of the probability of erroneously rejecting the null hypothesis that Z_{n+1} and $\{Z_i\}_{i=1}^n$ are from the same probability distribution. In other words, if $p_{n+1} < \epsilon$, Z_{n+1} is detected as an anomaly. The only design component of CAD is the NCM. In principle, any real-valued function that accurately discriminates between examples from the normal and abnormal classes would be appropriate. It is shown in Laxhammar and Falkman (2013) that CAD can address conventional issues of previous anomaly detection algorithms, e.g., invalid statistical assumptions and ad-hoc anomaly thresholds, and proposes Sequential Hausdorff Nearest Neighbours Conformal Anomaly Detector (SHNN-CAD) by choosing the summation of directed Hausdorff distances to k -nearest neighbors as the NCM.

Although SHNN-CAD has shown good performance in trajectory data, it is not appropriate for aviation data in that temporality is neglected, and online learning cannot be achieved due to the high dimensionality of the air traffic data. To address these issues along with uncertainty in real air traffic data, we redesigned the NCM from our previous work (Choi and Hwang, 2022). In this paper, a new NCM of trajectory Z_i is given as

$$\alpha_i = \max_{t \in [0,T]} \left[\sum_{j \in \mathbb{J}, j \neq i} -\mathcal{N}(\mu_i(t); \mu_j(t), \Sigma_j(t)) \right] \quad (6)$$



(a) Anomaly detection boundary using support vector machine (b) Anomaly detection boundary using proposed method

Fig. 5. Anomaly detection comparison under uncertainty in training samples.

where $\mathbb{J} = \{1, \dots, n+1\}$, and $\mu_j(t)$ and $\Sigma_j(t)$ are the mean and covariance of the state estimate of data point at timestep t in trajectory Z_j . We incorporate Σ_j , the known uncertainty of the training set, into the NCM, to explicitly consider the uncertainty in CAD. A summation of the PDF of other data points evaluated at the new data point can be interpreted as a measure of conformity, which implies the possibility of new data that comes from the distribution of other data at a specific time. Note that we multiply it by minus one to convert it to an NCM for consistency. This NCM can reflect uncertainty in training samples as shown in Fig. 1(b).

In order to numerically compute the distribution of the p -value under uncertainty in testing samples, the logistic function can first replace the inequality in Eq. (5).

$$p_{n+1} = \frac{1}{n+1} \sum_j g(\alpha_j - \alpha_{n+1}) \text{ where } g(x) = \frac{l}{1 + e^{-kx}} \quad (7)$$

where l is the curve's maximum value and k is the logistic growth rate or steepness of the curve. This paper sets l to 1 and k to an arbitrary large number, e.g., 10^8 . However, when α_j equals α_i , the logistic function would give 0.5, which is different from the result calculated using Eq. (5). To tackle this problem, by slightly modifying Eq. (5), we employ the smoothed p -value (Laxhammar and Falkman, 2013) given as

$$p_{n+1} = \frac{|\{j : \alpha_j > \alpha_{n+1}\}| + \tau |\{j : \alpha_j = \alpha_{n+1}\}|}{n+1} \quad (8)$$

where $\tau \in [0, 1]$. If τ is chosen as 0.5, Eq. (7) reaches Eq. (8) as k goes to infinity. Then, by combining the current data point's distribution and the p -value in Eq. (7), we can compute the expected p -value over the given dimensions. Hence, for a sequentially updated data point with covariance $(\mu_{n+1}(t)$ and $\Sigma_{n+1}(t))$ at the current timestep, the expected p -value over a small closed set \mathbb{Z} can be computed as

$$\mathbb{E}[p_{n+1,t}(z)] = \frac{1}{Pr(\mathbb{Z})} \int_{\mathbb{Z}} p_{n+1,t}(z) \cdot \mathcal{N}(z; \mu_{n+1}(t), \Sigma_{n+1}(t)) dz \quad (9)$$

with a probability of $Pr(\mathbb{Z}) = \int_{\mathbb{Z}} \mathcal{N}(z; \mu_{n+1}(t), \Sigma_{n+1}(t)) dz$. Finally, the pmf of the p -value at the current timestep can be computed by iterating the numerical integration with respect to discretized \mathbb{Z} over the current data distribution, $\mathcal{N}(z; \mu_{n+1}(t), \Sigma_{n+1}(t))$. Note that any other probability distribution function for the current data point can be used in Eq. (9) as long as it is integrable.

To see how much uncertainty in training samples can affect the anomaly detection result, we compare the proposed method with a conventional method, OCSVM. As shown in Fig. 5(a), the decision boundary is created using deterministic samples. This boundary labels sample #1 as normal and sample #2 as abnormal. However, by considering uncertainty, the proposed method creates a p -value distribution obtained by the expected p -value on the horizontal plane. Contrary to the previous result, sample #1 is labeled as abnormal, while sample #2 is labeled as normal, as shown in Fig. 5(b). Therefore, it is concluded that the uncertainty added to training samples can result in the overall distribution in such a way that the decision criterion is different from the deterministic result.

The pmf can be computed by integrating Eq. (9) numerically over the current data distribution. However, this can be computationally challenging from a numerical perspective. Hence, the pmf may be estimated with the use of numerical sampling for an online application. Note that the computation cost increases proportional to the number of samples used. Hence, the sampling size for integration is carefully selected so that information loss is minimized. Hence, to select the number of numerical samples for integration, we use the Kullback–Leibler (KL) divergence by computing the difference between two distributions. The KL divergence, i.e., relative entropy, is expressed as

$$D_{KL}(P \parallel Q) = \sum_{x \in \lambda} P(x) \log \left(\frac{P(x)}{Q(x)} \right) \quad (10)$$

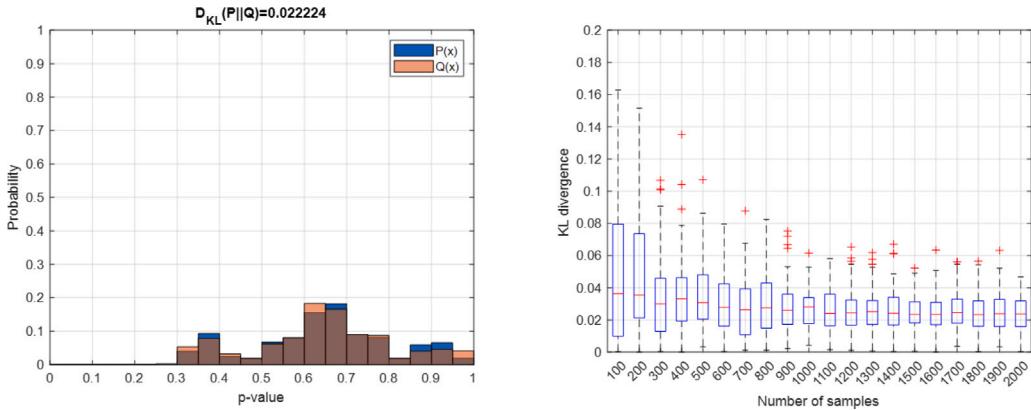


Fig. 6. The comparison example of probability distribution (left) and the boxplot of Kullback–Leibler divergence (right).

Table 2

Comparison of the computation time for online learning on sequentially updated point (unit: second).

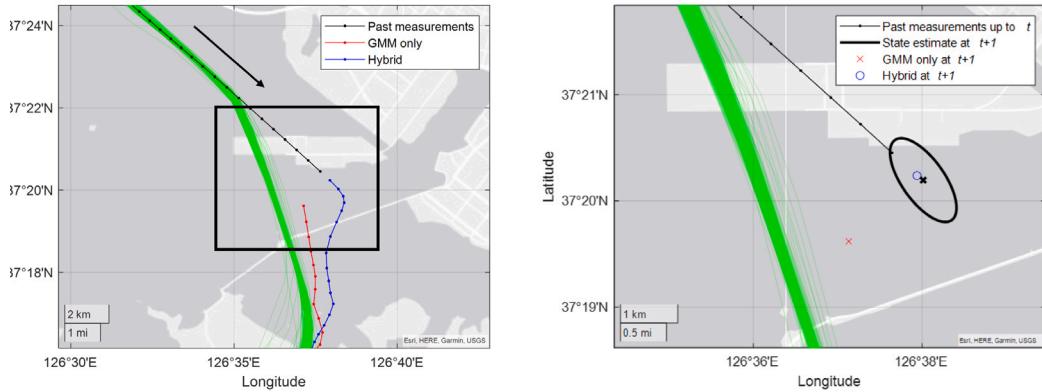
Size	SHNN-CAD	Proposed
$n = 500$	5.7124 ± 0.1119	0.0085 ± 0.0008
$n = 1000$	11.4326 ± 0.1546	0.0158 ± 0.0006
$n = 1500$	17.1893 ± 0.1997	0.0235 ± 0.0007
$n = 2000$	22.7703 ± 0.1437	0.0313 ± 0.0007
$n = 2500$	28.7610 ± 0.3539	0.0390 ± 0.0008
$n = 3000$	34.9428 ± 1.0616	0.0480 ± 0.0025

This is a measure of how one probability distribution Q is different from a reference probability distribution P . In this paper, P is obtained by numerical integration and considered as the ground truth. Q is obtained by random sampling from the current data distribution, and the result depends on the number of samples. For illustration, in Fig. 6, two distributions are plotted, showing slight differences between them, and the KL divergence value is computed as 0.02224. A small KL divergence value means the sampling method matches the reference distribution well, i.e., very little information loss. Based on the Monte Carlo simulation using real ADS-B data, the KL divergence is tested with respect to the number of samples. As shown in Fig. 6, when the number of samples becomes 1000, it seems no significant change is observed in the KL divergence, especially in terms of variance. Hence, in this paper, we select the number of sampling as 1000.

As an output of stochastic conformal anomaly detection, the anomaly probability is provided, and the current dataset is updated with recomputed NCMs and new data points. Since aviation data are continuously received, online learning on sequentially updated points could be desirable in order to maintain a more timely and precise detection model which can more accurately represent dynamically changing air traffic. For online learning, the proposed algorithm uses the NCM expressed as a summation of the probability densities, unlike SHNN-CAD. For a sequentially updated point in a new incoming trajectory Z_{n+1} , the algorithm updates the NCM, α_i , by adding the probability density given the state estimate ($\mu_{n+1}(t)$ and $\Sigma_{n+1}(t)$) of the updated point. Compared to the directed Hausdorff k -nearest neighbors in SHNN-CAD, the proposed method can significantly decrease computational cost. The overall time complexity of SHNN-CAD is given as $\mathcal{O}(n \cdot L_{max} \log(k) + n \cdot L_{max}^2 \log(L_{max}) + n \cdot k)$ in Laxhammar and Falkman (2013), while that of the proposed method is $\mathcal{O}(3n \cdot m_t + m_t)$, where L_{max} is the maximum number of line segments among a total of n trajectories and m_t is the number of track points at timestep t . Note that m_t is less than or equal to n since the length of each trajectory is generally different. To compare the computation cost of the proposed method and the SHNN-CAD, given n trajectories with the same final timestep $T = 200$, we record the computation time measured using a computer with an Intel Core i7-9750H CPU and 16 GB RAM. As shown in Table 2, the proposed method takes much less time than SHNN-CAD. This result indicates that the proposed method can work well even when multiple trajectories come in simultaneously. This is especially important because tens of aircraft arrive and depart simultaneously in terminal airspace. The importance of online learning in anomaly detection will be further discussed with in Section 4.3.

3.3. Anomaly resolution

The existing algorithms described in Section 2.1 have focused only on anomaly detection. Hence, when the system detects abnormal behaviors, resolving anomalies should strongly depend on a manual way by human operators such as ATCs and pilots, thereby increasing the workload. However, if advisory tools could suggest a strategy to resolve anomalies, it would reduce the workload of human operators to a reasonable level. In this regard, we propose an anomaly resolution method by employing a trajectory prediction algorithm, which is based on the fact that data-driven trajectory prediction algorithms leverage historical data



(a) Trajectory prediction example made by two methods when an anomaly occurs (b) The first prediction points and state estimate distribution before measurement update

Fig. 7. Comparison of trajectory prediction methods for anomaly resolution.

and are mainly derived from frequently observed trajectories. Trajectory prediction algorithms produce a maximum *a posteriori* estimate of the future trajectory, meaning that the predicted trajectory is the one with the highest probability, given the available information (from model, data, and past trajectory). Naturally, since abnormal trajectories with large deviations are rarely observed, the predicted trajectory will generally return to the normal, despite the temporary deviation from the nominal.

To assure anomaly resolution, the trajectory prediction algorithm is trained by the dataset consisting of normal trajectories whose p -value is larger than the predefined alarm rate, ϵ . The prediction example is shown in Fig. 7(a), where the historical trajectories are depicted in green. When an anomaly occurs, and past measurements are fed into trajectory prediction algorithms, a future trajectory is predicted by two methods (in red and blue), GMM only and hybrid methods (Choi et al., 2021). In this example, both cases return to the set of normal trajectories, and it seems that the red trajectory gets close to the normal range faster than the blue case. However, the problem is that the red trajectory has a significant gap from the current flight (in black). One potential reason is that the data-driven only method could result in infeasible maneuvers because anomalous trajectories were rarely observed in the historical data. Contrastively, the blue case is sensible in terms of the aircraft's dynamics because the aircraft's current motion described by its dynamics is directly incorporated into hybrid trajectory prediction. In the hybrid method, the data-driven and physics-based methods are integrated in such a way that RM-IMM infers the current behavior of the aircraft and estimates its future trajectory based on the future pseudo measurements provided by GMM.

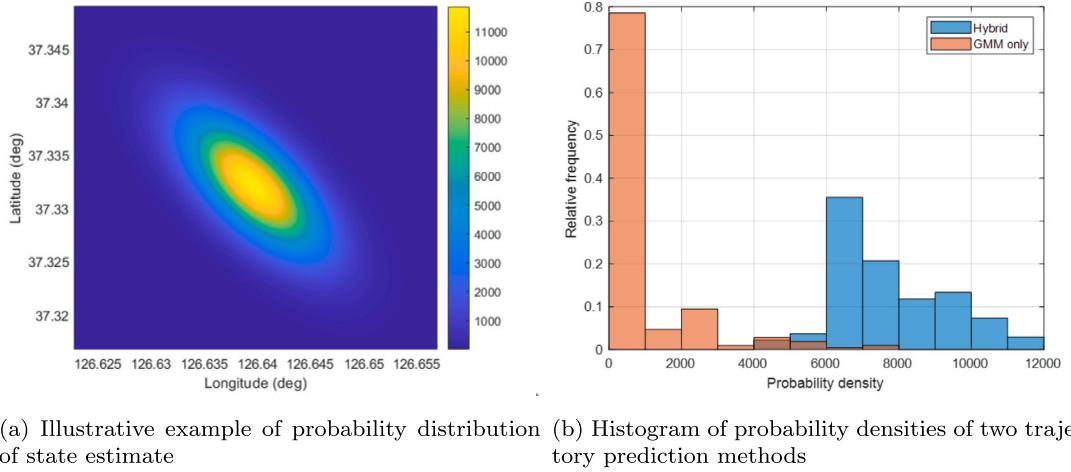
To quantitatively measure how much predicted trajectories conform to the aircraft's dynamics, we analyze and compare the probability density of the first predicted point according to the distribution given by the SLHS model in Eqs. (2)–(4), i.e., $\mathcal{N}(\hat{x}(t+1); \hat{x}(t+1|t), P(t+1|t))$, where $\hat{x}(t+1)$ is the predicted point and $\hat{x}(t+1|t)$ and $P(t+1|t)$ are the mean and covariance of $x(t+1)$ given $x(t)$ and Eqs. (2)–(4). For instance, the first predicted point by two methods and the state estimate are plotted together in Fig. 7(b). The actual probability distribution of the state estimate is obtained in the form of Gaussian by hybrid estimation as shown in Fig. 8(a). We only focus on the first point because the infeasible first predicted point could make the rest of the predictions pointless. Therefore, predictions are performed only when a trajectory significantly deviates from the normal ones, i.e., anomaly in the horizontal dimension, and then the first point's probability density is recorded. The results are presented in Fig. 8(b). A large probability density is interpreted as a good fit of the first prediction point to the aircraft's dynamics. The histogram of GMM only is skewed to the left and the case of the hybrid method is concentrated on around 8,000. The means of the probability density are 737.36 and 8579.4, respectively. Therefore, we can conclude that the hybrid method can provide a reasonable anomaly resolution strategy conforming to the aircraft's dynamics.

4. Test and analysis

In this section, we first prepare historical ADS-B data in various airspace. To evaluate the performance of the proposed framework described in Section 3, we study unstable approaches labeled by subject matter experts, online learning in en route, and anomaly resolution in terminal airspace.

4.1. Data preparation and trajectory pattern analysis

ADS-B technology has been deployed to enhance aviation safety and efficiency by enabling aircraft to determine their position with respect to other similarly-equipped aircraft, using satellite, inertial, and radio navigation (Olive and Morio, 2019). ADS-B Out periodically emits (at approximately 1 Hz) the aircraft's position (latitude and longitude), altitude, heading, horizontal and vertical speed, along with other relevant parameters, to ground stations and other equipped aircraft. Accordingly, the recorded ADS-B data



(a) Illustrative example of probability distribution of state estimate (b) Histogram of probability densities of two trajectory prediction methods

Fig. 8. Probability density comparison of two trajectory prediction methods as a measure of conformance to dynamics.

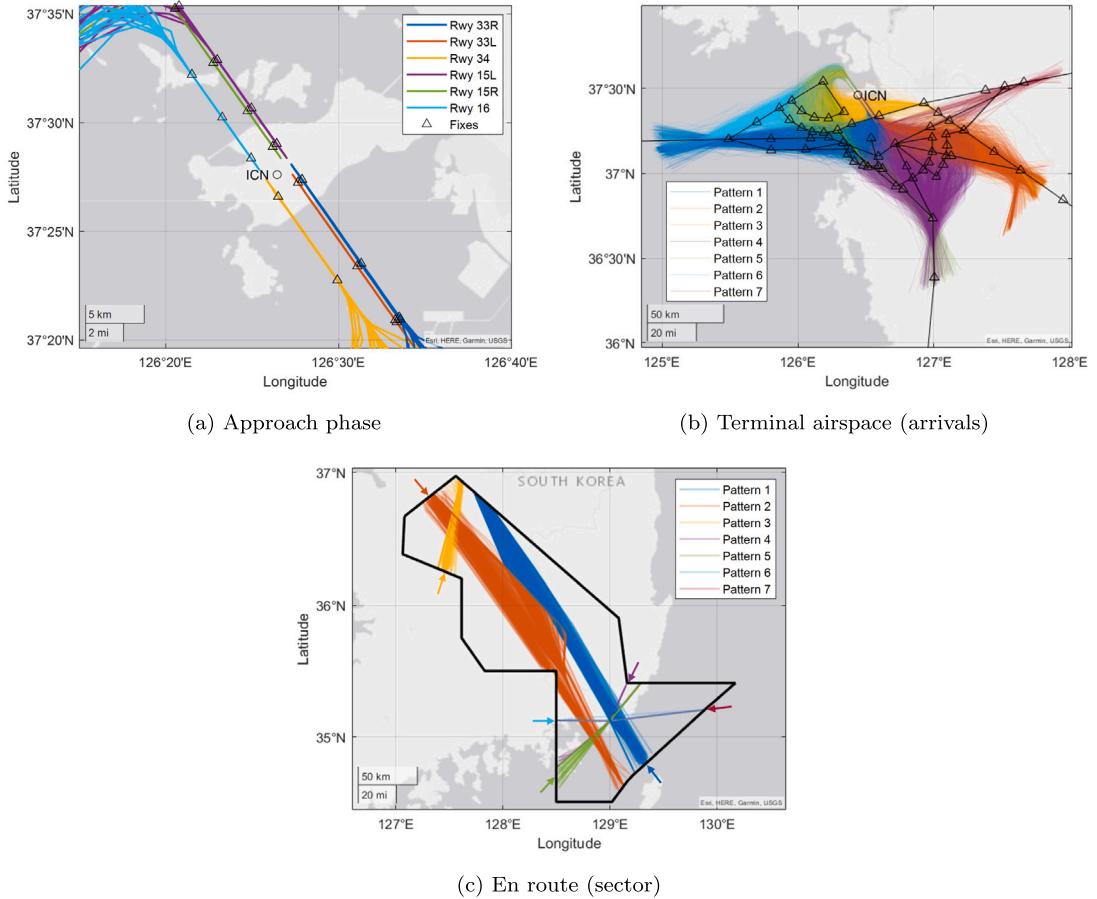


Fig. 9. Data preparation results of one-month trajectories.

contains general information about the aircraft (flight ID, origin, and destination) and the aircraft's state information (time, position, and speed). In this paper, we use the ADS-B data recorded around Incheon international airport (ICN), South Korea, in 2019.

Firstly, we perform data preprocessing because some of the recorded trajectories are incomplete and have noisy track points, which could degrade the performance of data-driven algorithms. In the given dataset, two major issues are considered: (i) heavily

Table 3
Trajectory pattern classification performance with different flight times.

Flight time	AUC	Accuracy	F1 score	Recall	Precision
100%	1.0000	1.0000	1.0000	1.0000	1.0000
75%	1.0000	0.9992	0.9982	0.9994	0.9970
50%	0.9999	0.9657	0.9417	0.9763	0.9276
25%	0.9982	0.7333	0.7019	0.8483	0.7276

missing data points, and (ii) partially noisy data points (e.g., physically infeasible spikes in the final approach phase). Since those recordings do not conform to the aircraft dynamics and/or flight operations procedures, both issues should be addressed. As a result of data preprocessing, trajectories of the first case are excluded from the dataset, but in trajectories of the second case, only the corresponding points are removed. Secondly, as the time interval of recorded data is different for each trajectory, the state information is resampled to have the same time interval (10 s). Trajectories in the ADS-B data typically cover all the flight phases from departure at the origin to the approach at a destination airport, which can be divided into three phases: (i) approach/departure, (ii) terminal airspace, and (iii) en route. In terminal airspace, trajectories are cut at 70 nmi away from ICN because that is roughly where most flights are heading towards each entry fix. In addition, trajectories are cropped from the sector entry point to the exit point in en route airspace. Lastly, the dataset in each airspace is divided into a training dataset and a testing dataset with the ratio 8:2. Measurement uncertainty is set based on [Table 1](#) and processed in the form of Gaussian by hybrid estimation mentioned in Section 3.1.

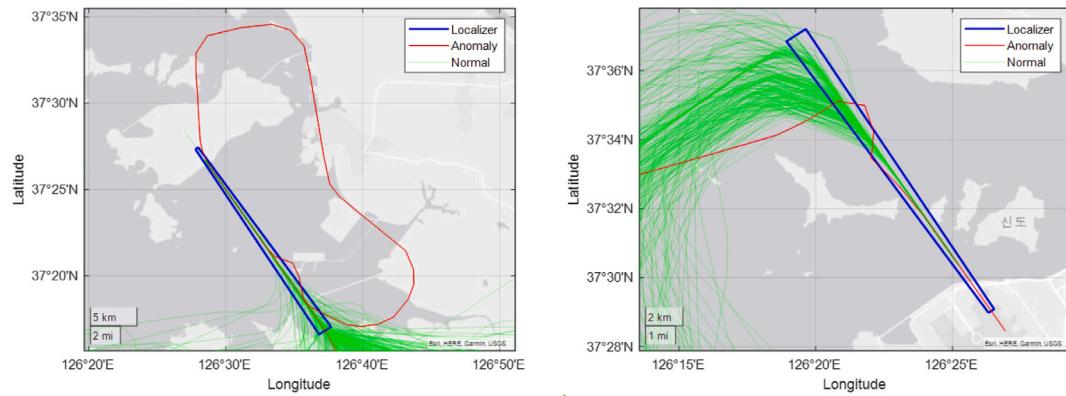
Since the initial and last stages of trajectories are aligned with the runway, trajectories can be easily distinguished by the runway, as shown in [Fig. 9\(a\)](#). However, in terminal airspace and en route, airspace operations are complex, resulting in diverse aircraft maneuvers. To capture complex aircraft operations, multiple clusters, i.e., trajectory patterns, can be identified by applying existing clustering algorithms to each airspace. In this paper, we employ a density-based method, called DBSCAN (Density-Based Spatial Clustering of Applications with Noise) ([Ester et al., 1996](#)) because it works well with time-series data, and there is no need to predefine the number of patterns. For our dataset, we apply DBSCAN with a distance threshold of 0.3 and a minimum sample size of 30 in [Fig. 9\(b\)](#), while in [Fig. 9\(c\)](#), we employ a distance threshold of 0.3 and a minimum sample size of 20. The trajectory pattern identification is particularly important in that anomaly detection on individual clusters (patterns) with similar properties can be more effective and efficient. For instance, in [Fig. 9\(c\)](#), the flights in the red pattern enter from the northwest, and the flights in the blue pattern enter from the southeast. Hence, by dividing one from the other, there is no need to compare trajectories in opposite directions, and a more accurate model can be obtained using only similar trajectories.

Based on the trajectory patterns identified in [Fig. 9](#), the classification of trajectory patterns serves as a fundamental component in this paper for real-time anomaly detection and resolution. To evaluate the performance of the proposed trajectory pattern classification method, we conduct classification experiments using trajectories with different flight times. As performance measures, we compute accuracy, F1 score, recall, and precision, which are widely used for a supervised classification model ([Deshmukh et al., 2020](#)). These measures are computed using the number of True Positive cases (TP), the number of True Negative cases (TN), the number of False Negative cases (FN), and the number of False Positive cases (FP). Additionally, we include the Area Under the receiver operating characteristic Curve (AUC), as the aforementioned measures highly rely on what threshold is chosen. In other words, AUC is a threshold-invariant measure of the ability to distinguish between classes.

The classification results are presented in [Table 3](#). The range for all the measures in the table is [0,1], and closer to 1 means better performance. Since this problem is multi-class classification, except for accuracy, all the metrics are the averaged values of multiple classes. The results demonstrate that our model achieves high accuracy in classifying trajectories, which can help detect and resolve anomalies in real-time scenarios. Note that the classification performance significantly drops when using only 25% of the trajectory compared to 50%. This finding suggests the possibility of incorrect classification in real-time applications where only partial trajectory is available, especially at the beginning. However, we acknowledge that initial misclassification can be resolved as more data becomes available over time.

4.2. Anomaly detection performance evaluation

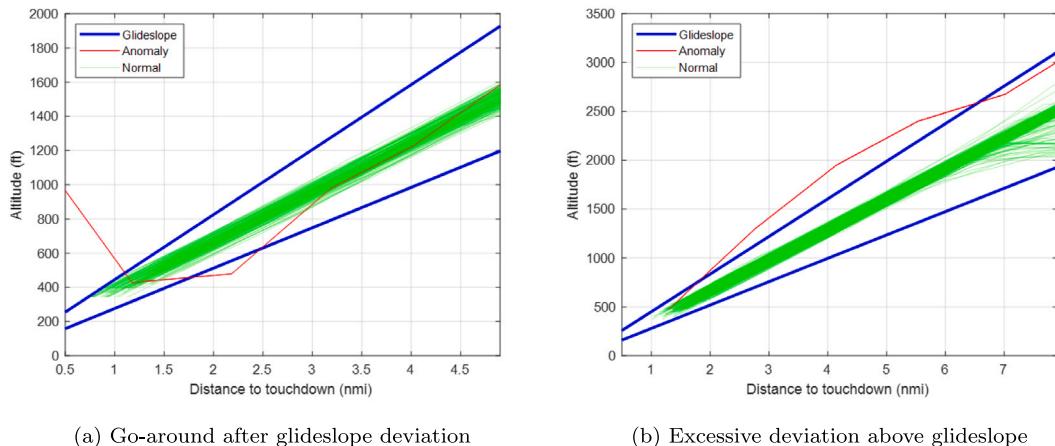
In order to evaluate the performance of the proposed framework, we use the unstable approach data labeled by subject matter experts. In this dataset, two types of unstable approaches are defined. One is horizontal deviation, which is called localizer deviation, and the other is vertical deviation called glideslope deviation. Around airports, localizer and glideslope transmitters are installed to assist the safe approach and landing ([Hall and Soares, 2008](#)). If flights deviate significantly from these guidance, it could lead to adverse events. For instance, as shown in [Figs. 10\(a\)–11\(a\)](#), serious deviation may increase the probability of an unstable landing, followed by a go-around. Therefore, subject matter experts define a criterion to detect deviations in each dimension. For a localizer, the range of 1° to the left and right of its course is defined as a normal area, and for a glideslope, a range of 0.7° above and below the 3° slope is defined as a normal area. [Fig. 10\(b\)](#) illustrates an S-turn maneuver that repeatedly deviates from a localizer bound in order to increase the distance or reduce airspeed in congested airspace ([Timar et al., 2012](#)). Note that deviations that occur only after flights fully intercept a localizer are considered anomalies. Hence, in [Fig. 10](#), some trajectories (in green) intercept a localizer late or passing through the blue bound, but it is not labeled as anomalies. [Fig. 11\(b\)](#) illustrates an anomalous flight in the vertical dimension, which deviates to a higher angle than normal.



(a) Go-around after localizer deviation

(b) S-turn deviation

Fig. 10. Illustrative examples of localizer deviation.



(a) Go-around after glideslope deviation

(b) Excessive deviation above glideslope

Fig. 11. Illustrative examples of glideslope deviation.

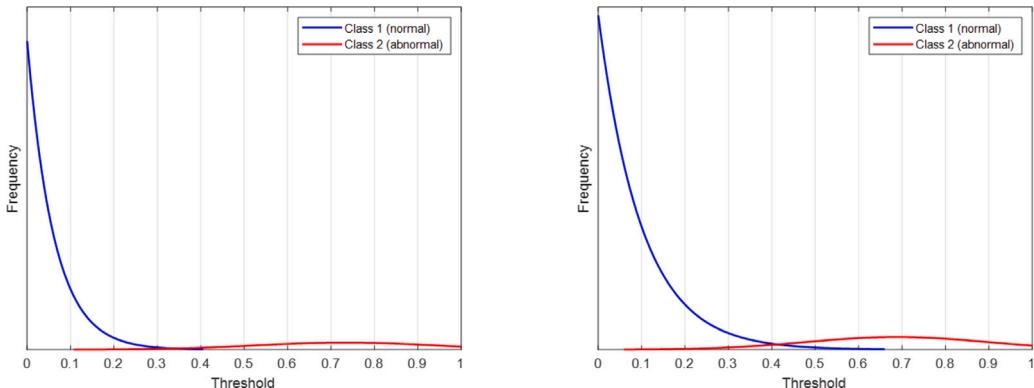


Fig. 12. Class distributions in the imbalanced datasets (left: localizer, right: glideslope).

Table 4
Performance results of unstable approaches.

	AUC	Accuracy	F1 score	Recall	Precision
Localizer	0.9856	0.9641	0.7143	0.9091	0.5882
Glideslope	0.9617	0.9442	0.5882	0.7672	0.4762

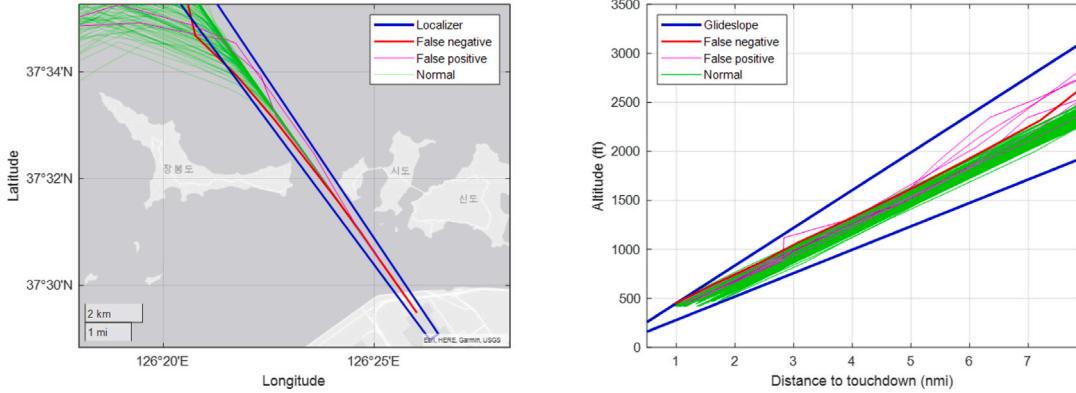


Fig. 13. False alarm examples of localizer (left) and glideslope (right) deviations.

The proposed stochastic conformal anomaly detection framework described in the previous section is applied to detect localizer and glideslope deviations. For instance, if the output, i.e., anomaly probability given as $Pr(p_i \leq \epsilon)$, is greater than or equal to a predefined threshold (δ), the trajectory is labeled as an anomaly by the proposed framework. However, if the true label for this trajectory is actually negative, this is an FP case. To compute the performance measures, the decision threshold needs to be determined in advance. For binary classification, the default threshold is generally set to 0.5. However, this may not be ideal for some cases, especially for imbalanced datasets. Fig. 12 illustrates the importance of threshold selection, which is obtained by fitting the datasets to the distribution. For finding the best threshold, the authors in Zou et al. (2016) have proposed a threshold selection method that maximizes F-score given as

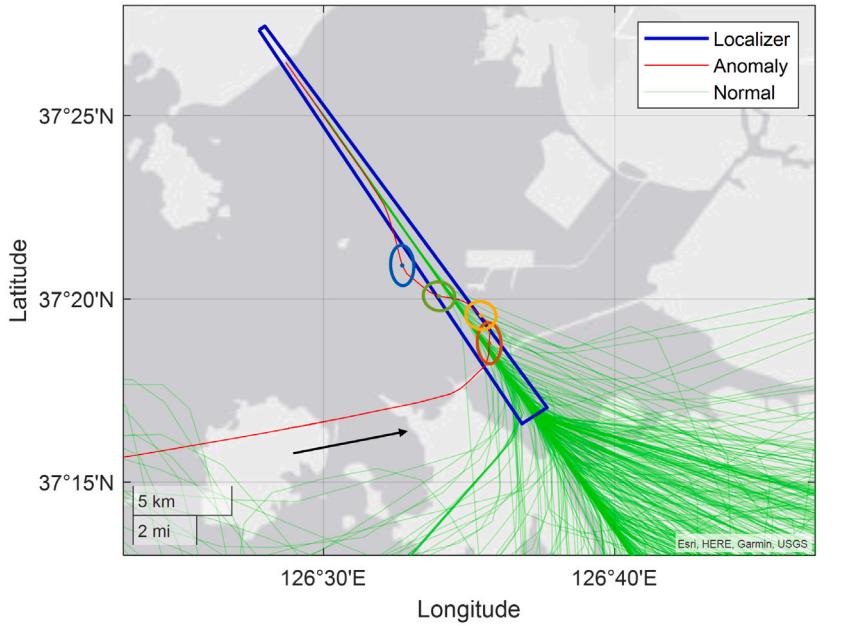
$$F_\beta = \frac{(\beta^2 + 1) \cdot \text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}} \quad (11)$$

where β is a parameter for a weight between precision ($\frac{TP}{TP+FP}$) and recall ($\frac{TP}{TP+FN}$). Using this method, the thresholds of the localizer and glideslope are determined to be 0.382 and 0.414, respectively.

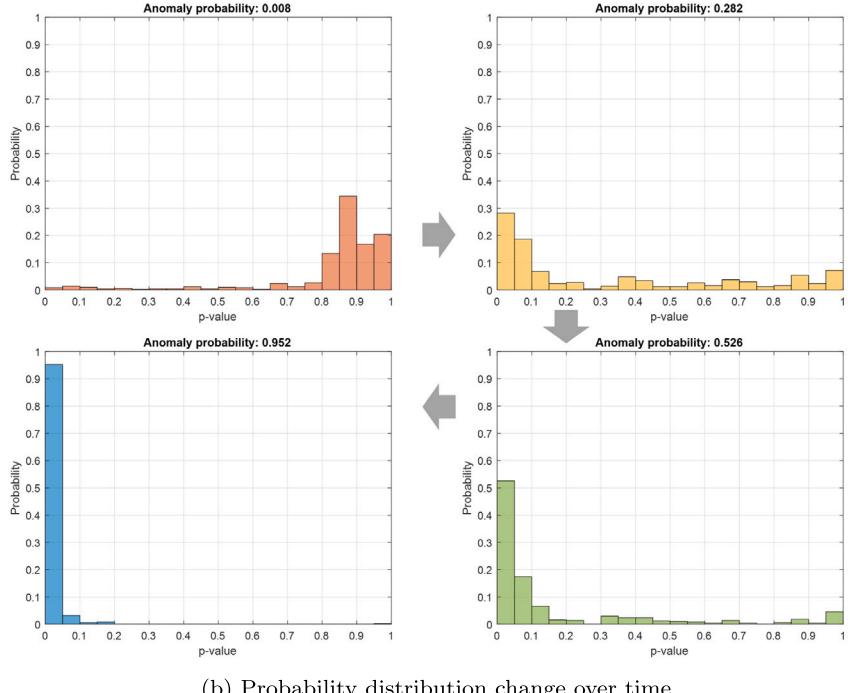
The overall results are summarized in Table 4. AUC and accuracy show the outstanding performance of the proposed framework, while others are acceptable. Note that F1 score and Precision in both cases are relatively low, which can be explained by false alarms in Fig. 13. Since the data labels are determined by the current practice (in blue), the data distribution is not considered at all. Hence, the red and magenta trajectories are determined as FN and FP, respectively, although both cases are reasonable in terms of the data distribution and especially FP (in magenta) could be related to potential safety risks. In addition, the performance of glideslope deviation is much worse than that of localizer deviation, which may be attributed to unreliable altitude distribution between 1 and 2 nmi. As shown in Figs. 11 and 13, the last altitudes of trajectories show significant variations and tightly fit the blue bound, which could cause false alarms frequently. Therefore, this performance is more likely to originate from the inherent characteristics of the data than the proposed method.

In Fig. 14, by using the proposed algorithm, anomaly detection under uncertainty is illustrated. The anomalous flight approaching runway 33R is depicted in red, and the black arrow indicates the moving direction. The mean and covariance of the estimated states are represented by dot and ellipse, respectively, and color coded by timestep. The corresponding pmf with an anomaly probability is shown in Fig. 14(b) clockwise. Since we set the predefined alarm rate ($\epsilon = 0.05$) by considering the class ratio in the given dataset, an anomaly probability, $Pr(p_i \leq \epsilon)$, is the sum of all the probabilities that the p-value is less than or equal to ϵ . After intercepting the localizer at the red point, this flight shows the S-turn maneuver, and then the probability distribution becomes skewed to the left. Correspondingly, the anomaly probability increases from 0.8% to 28.2%, 52.6%, and 95.2%, which means we can conclude this flight is an anomaly with high confidence.

Note that although the yellow point (mean) is located outside the localizer bound, it is determined as normal due to uncertainty, which can be explained by the green normal trajectories intercepting from the right. On the other hand, although the dark green point (mean) is located inside the localizer bound, it is highly probable to be abnormal. One possible explanation is that the normal trajectories and their uncertainty are well-aligned with the localizer; thus, the left half of the green ellipse is unprecedented. These observations can be missed by the current (rule-based) practice and deterministic methods. In addition, the pmf information could be beneficial to enhance the situational awareness of ATCs in practice.



(a) Localizer deviation case showing S-turn maneuver



(b) Probability distribution change over time

Fig. 14. Stochastic conformal anomaly detection during the approach phase.

4.3. Online learning for anomaly detection

Aircraft travel routinely in each airspace, and accordingly, new data becomes available in sequential order. To respond to this situation, the best predictor, i.e., the trained anomaly detection model, needs to be updated at each timestep, as opposed to batch learning update which generates a model by using the entire training dataset altogether. Therefore, online learning is particularly important in that, by taking new streaming data in a sequential manner, the bounds represented by the learned model could evolve over time. For instance, once a group of normal flights taking straight flights comes in, the bound could be shrunken, as shown

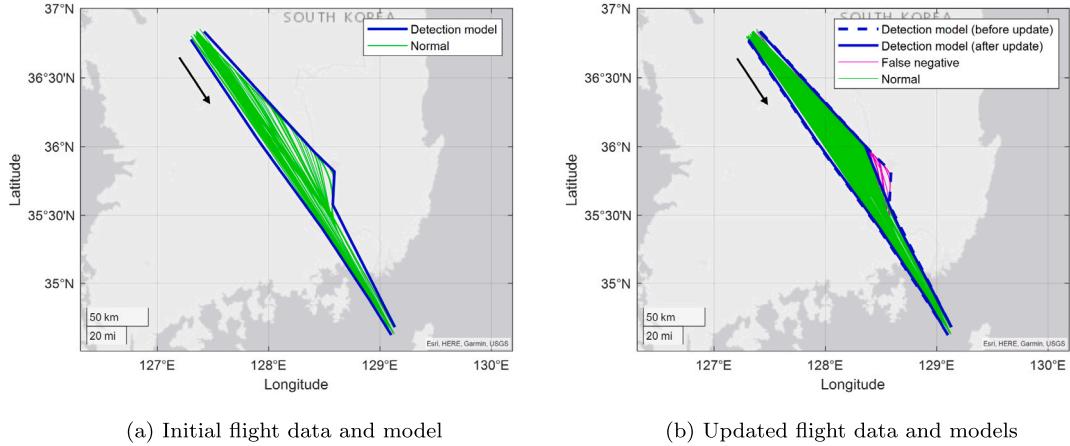


Fig. 15. Illustration showing the requirement of online learning in the horizontal dimension.

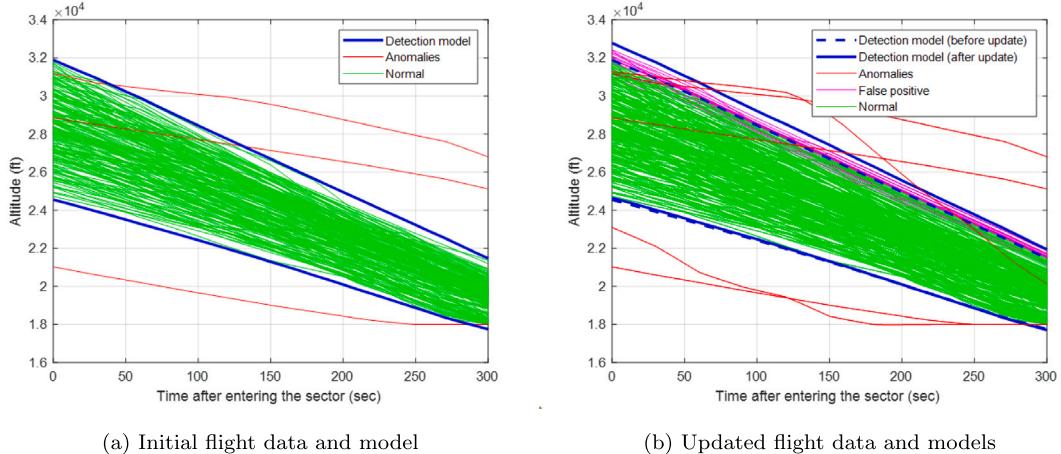


Fig. 16. Illustration showing the requirement of online learning in the vertical dimension.

in Fig. 15. On the other hand, if trajectories slightly out of the current bound are fed into the model, the bound could be updated by extending the existing area as shown in Fig. 16. In both cases, false alarms (in magenta) occur because anomaly detection is performed based on the initial model without an update. In particular, since false negative cases in Fig. 15(b) overlap the blue pattern in Fig. 9(c), those flights are definitely abnormal.

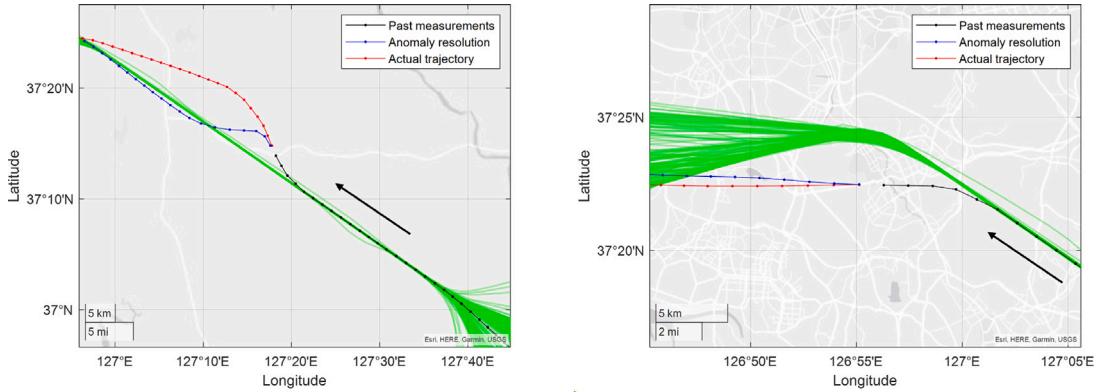
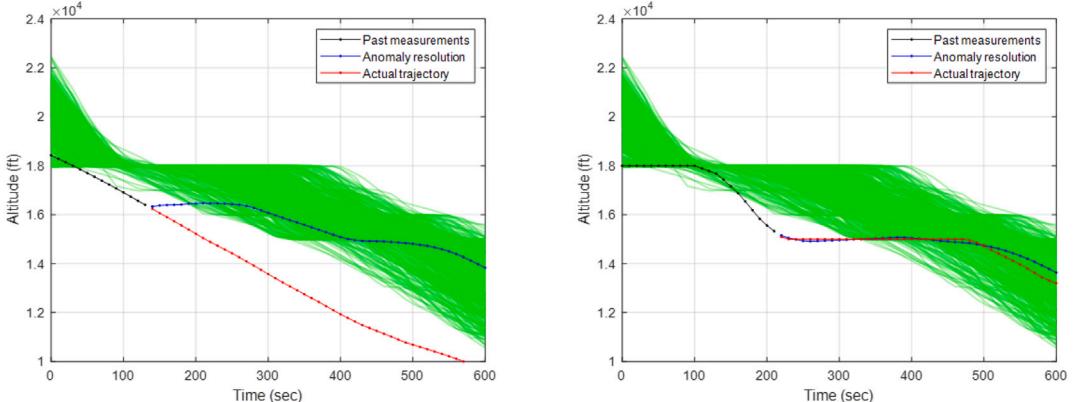
To evaluate online learning performance in anomaly detection, we design an experiment using two patterns (red and yellow) in Fig. 9(c). For comparison, we choose TempAD (Deshmukh and Hwang, 2019) because it can generate human-readable anomaly detection bounds where we can easily identify false alarms by visual inspection. Since both TempAD and the proposed algorithm are unsupervised learning, there are no true labels for anomalies. Hence, we first generate two TempAD models using 80% and 100% of the data, respectively. Then, we label the testing data (20%) with the second TempAD model (iteratively trained by the entire data), and these labels are assumed as the true labels. Lastly, the proposed online/offline learning and the first TempAD model are compared using the testing data. The performance measures are computed using the assumed true labels, and AUC is omitted since TempAD provides a single detection bound (no threshold), i.e., predicates.

The comparison results are summarized in Table 5. In both dimensions, the proposed method (online) significantly outperforms TempAD and the proposed method (offline) in terms of accuracy and F1 score. Compared to the proposed method (offline), TempAD shows slightly better performance by taking advantage of OCSVM-like optimization. The considerable accuracy difference is observed between horizontal and vertical dimensions, which is mainly due to the data imbalance. In other words, the number of actual positive samples is much smaller than that of actual negative samples, causing higher false alarms in the vertical case. On the other hand, recall (vertical case) and precision (horizontal case) of TempAD are computed as 1, as we assume labels generated by TempAD trained by whole data are true labels. To put it in another perspective, in the horizontal dimension, the updated model gets narrower than the initial one, and thus, if a sample is labeled as abnormal by the initial model, it must be abnormal from the perspective of the updated model, thereby resulting in zero false positives. Similarly, in the vertical dimension, the detection model is updated to

Table 5

Performance comparison between offline (TempAD and proposed method) and online learning (proposed method).

Dimension	Method	Accuracy	F1 score	Recall	Precision
Horizontal	TempAD	0.9730	0.6286	0.4583	1.0000
	Proposed(off)	0.9688	0.5714	0.4167	0.9091
	Proposed(on)	0.9896	0.8936	0.8750	0.9130
Vertical	TempAD	0.8739	0.4340	1.0000	0.2771
	Proposed(off)	0.8676	0.4112	0.9565	0.2619
	Proposed(on)	0.9706	0.7308	0.7917	0.6786

**Fig. 17.** Illustrative examples of anomaly resolution in the horizontal dimension.**Fig. 18.** Illustrative examples of anomaly resolution in the vertical dimension.

get wider, and hence, there are no false negatives. Due to this issue, the recall or precision of TempAD is larger than that of the proposed method. Still, the proposed method clearly shows better performance in terms of the F1 score, defined as the harmonic mean of precision and recall. Therefore, it is concluded that online learning can play a critical role in anomaly detection.

4.4. Effectiveness of anomaly resolution

If the anomaly probability of a new incoming flight violates the predefined threshold, the proposed framework automatically generates and suggests a resolution strategy to a human operator (ATC and/or pilots), who then can follow the provided strategy to rectify the anomaly. Anomaly resolution is created based on the hybrid trajectory prediction algorithm (Choi et al., 2021), as the data-driven only method shows poor conformance to the aircraft's dynamics as discussed in Section 3.3.

To measure effectiveness of our anomaly resolution strategy, the proposed framework is tested with anomalous flights in terminal airspace. For better illustration, we only investigate anomalies that can be resolved within a typical travel time range, which means trajectories with extremely large deviations, e.g., holding and diverting, are excluded. As an example, anomalous flights belonging to the yellow pattern of Fig. 9(b) are illustrated in Fig. 17. If an anomaly is detected at the last black point, the proposed framework provides an anomaly resolution strategy, while an actual flight follows a red trajectory. The left figure shows a case where our

Table 6
Anomaly resolution performance.

Dimension	Elapsed time (<i>resolution</i>)	Elapsed time (<i>actual</i>)
Horizontal	86.43 ± 58.93 s	281.19 ± 95.54 s
Vertical	64.88 ± 36.13 s	217.81 ± 127.79 s

strategy is much quicker than the real flight, whereas the right figure shows a case where ours is similar to the actual flight. Similarly, Fig. 18 illustrates two cases in the vertical dimension.

As shown in these examples, not only our anomaly resolution strategies but also actual trajectories generally maneuver in the direction that can resolve the current deviation. Note that all the cases in the given dataset are successfully settled by the proposed method within the 3 min prediction horizon. Table 6 summarizes the results of elapsed time for resolution. The elapsed time of the proposed method in the horizontal and vertical dimension are reduced by 69.26% and 70.21% compared to the elapsed time of actual trajectories on average. Therefore, the detected anomalies can be effectively resolved by the proposed anomaly resolution method.

5. Conclusions

In this paper, a stochastic conformal anomaly detection and resolution framework is proposed to detect flight anomalies from noisy aviation data and recommend a dynamics-based strategy to resolve the detected anomalies for real-time application. The proposed framework is developed based on conformal prediction so that it can achieve online learning and explicitly considers uncertainty in both training and testing data. In addition, by incorporating hybrid trajectory prediction, the proposed framework can generate a resolution suggestion that conforms to the aircraft's dynamics. Our framework has been demonstrated with real air traffic data for performance evaluation. A set of performance measures for machine learning techniques are used to show anomaly detection performance under uncertainty during the approach phase. The test results show that the proposed method not only achieves competitive performance but also captures information that deterministic anomaly detection algorithms or current practice could miss. Also, the benefit of online learning is demonstrated through experiments, showing a significant performance improvement over offline learning. Lastly, it is shown that the proposed method can resolve anomalous situations more effectively compared to the actual trajectories that resolve them. Therefore, the proposed framework can contribute to reducing the workload of human operators and enhancing their situational awareness. Furthermore, note that our proposed framework is general enough to be applied to unmanned aircraft system traffic management where uncertainty becomes more critical.

We note that this study has three limitations that need to be addressed by future research. Firstly, although online learning of the proposed method showed much reduced computational cost compared to the existing algorithm, a large dataset consisting of more than 1–2 years of aviation data is still difficult to be processed in a real-time setting. As a practical solution, taking note of the fact that the Aeronautical Information Publication (AIP) defines airspace operations and is frequently updated, pruning the training dataset according to the AIP update may be applied but the effectiveness of this technique remains hindered since it relies heavily on operational updates. Therefore, the computational time complexity that grows with the size of the dataset should be addressed from an algorithmic aspect. Secondly, the proposed framework can be extended to include features such as speed and energy, which are critical states of an aircraft, thereby enabling anomaly resolution strategies that are congruent with the speed and energy management based on flight dynamics. Moreover, by designing a new nonconformity measure for high-dimensional data, future research can include investigating a combination of multiple features for anomaly detection. Lastly, the proposed framework in this paper is developed based on the conventional conformal prediction under the assumption that samples are independent and identically distributed. Recent extensions of conformal prediction that relax the exchangeability condition can be beneficial for anomaly detection in air traffic control, as it needs to be implemented on real-time streaming data, which may not satisfy the exchangeability condition due to the fact that the patterns of trajectories can change depending on the weather, season, or even ATCs' preference. In this regard, relating our proposed framework to the extensions of conformal prediction will be studied as one of our future works. Therefore, our future works will study these limitations in the above-mentioned directions for more practical applications.

CRediT authorship contribution statement

Hong-Cheol Choi: Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Chuhao Deng:** Software, Writing – review & editing. **Hyunsang Park:** Data curation, Writing – review & editing. **Inseok Hwang:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2020-KA158275). The authors are grateful to Dr. Hak-Tae Lee at Inha University for his valuable comments and support.

References

- Aggarwal, C.C., Philip, S.Y., 2008. A survey of uncertain data algorithms and applications. *IEEE Trans. Knowl. Data Eng.* 21 (5), 609–623.
- Aggarwal, C.C., Yu, P.S., 2008. Outlier detection with uncertain data. In: Proceedings of the 2008 SIAM International Conference on Data Mining. SIAM, pp. 483–493.
- Ahad, R., Chan, E., Santos, A., 2015. Toward autonomic cloud: Automatic anomaly detection and resolution. In: 2015 International Conference on Cloud and Autonomic Computing. IEEE, pp. 200–203.
- Choi, H.-C., Deng, C., Hwang, I., 2021. Hybrid machine learning and estimation-based flight trajectory prediction in terminal airspace. *IEEE Access* <http://dx.doi.org/10.1109/ACCESS.2021.3126117>.
- Choi, H.-C., Deng, C., Park, H., Hwang, I., 2022. Gaussian mixture model-based online anomaly detection for vectored area navigation arrivals. *J. Aerosp. Inf. Syst.* 1–16.
- Choi, H.-C., Hwang, I., 2022. Toward real-time stochastic conformal anomaly detection in terminal airspace. In: ICRAT 2022 10th International Conference on Research in Air Transportation. ICRAT Florida, US.
- Corrado, S.J., Puranik, T.G., Fischer, O.P., Mavris, D.N., 2021. A clustering-based quantitative analysis of the interdependent relationship between spatial and energy anomalies in ADS-B trajectory data. *Transp. Res. C* 131, 103331. <http://dx.doi.org/10.1016/j.trc.2021.103331>.
- Dani, M.C., Freixo, C., Jollois, F.-X., Nadif, M., 2015. Unsupervised anomaly detection for aircraft condition monitoring system. In: 2015 IEEE Aerospace Conference. IEEE, pp. 1–7. <http://dx.doi.org/10.1109/AERO.2015.7119138>.
- Das, S., Matthews, B.L., Srivastava, A.N., Ozga, N.C., 2010. Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 47–56. <http://dx.doi.org/10.1145/1835804.1835813>.
- De Loza, A.F., Cieslak, J., Henry, D., Dávila, J., Zolghadri, A., 2015. Sensor fault diagnosis using a non-homogeneous high-order sliding mode observer with application to a transport aircraft. *IET Control Theory Appl.* 9 (4), 598–607.
- Deng, C., Choi, H.-C., Park, H., Hwang, I., 2022. Trajectory pattern identification and classification for real-time air traffic applications in Area Navigation terminal airspace. *Transp. Res. C* 142, 103765.
- Deshmukh, R., Hwang, I., 2019. Incremental-learning-based unsupervised anomaly detection algorithm for terminal airspace operations. *J. Aerosp. Inf. Syst.* 16 (9), 362–384. <http://dx.doi.org/10.2514/1.I010711>.
- Deshmukh, R., Sun, D., Kim, K., Hwang, I., 2020. Reactive temporal logic-based precursor detection algorithm for terminal airspace operations. *J. Air Transp.* 28 (4), 155–163.
- Deshmukh, R., Sun, D., Kim, K., Hwang, I., 2021. Temporal logic learning-based anomaly detection in metroplex terminal airspace operations. *Transp. Res. C* 126, 103036. <http://dx.doi.org/10.1016/j.trc.2021.103036>.
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise.. In: Kdd, Vol. 96, No. 34. pp. 226–231.
- Federal Aviation Administration, 2020a. FAA Aerospace Forecast: Fiscal Years 2020–2040. US Department of Transportation Washington, DC, USA.
- Federal Aviation Administration, 2020b. Modernization of the U.S. Airspace. US Department of Transportation Washington, DC, USA.
- Gariel, M., Srivastava, A.N., Feron, E., 2011. Trajectory clustering and an application to airspace monitoring. *IEEE Trans. Intell. Transp. Syst.* 12 (4), 1511–1524. <http://dx.doi.org/10.1109/TITS.2011.2160628>.
- Ghorbani, A.A., Lu, W., Tavallaei, M., 2009. Network Intrusion Detection and Prevention: Concepts and Techniques, Vol. 47. Springer Science & Business Media.
- Hall, T., Soares, M., 2008. Analysis of localizer and glide slope flight technical error. In: 2008 IEEE/AIAA 27th Digital Avionics Systems Conference. IEEE, pp. 2–D.
- Hernández-Romero, E., Valenzuela, A., Rivas, D., 2020. Probabilistic multi-aircraft conflict detection and resolution considering wind forecast uncertainty. *Aerosp. Sci. Technol.* 105, 105973.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780.
- Hou, Z.-S., Wang, Z., 2013. From model-based control to data-driven control: Survey, classification and perspective. *Inform. Sci.* 235, 3–35.
- Hwang, I., Balakrishnan, H., Tomlin, C., 2006. State estimation for hybrid systems: applications to aircraft tracking. *IEEE Proc. D* 153 (5), 556–566.
- ICAO, 2006. ICAO standards and recommended practices.
- ICAO, 2016. Draft 2016–2030 Global Air Navigation Plan, Tech. Rep., Doc 9750-AN/963, fifth ed. Montreal, Canada.
- International Air Transport Airport, 2018. Future of airline industry 2035.
- Jansson, D., Medvedev, A., Axelson, H., Nyholm, D., 2015. Stochastic anomaly detection in eye-tracking data for quantification of motor symptoms in Parkinson's disease. In: Signal and Image Analysis for Biomedical and Life Sciences. Springer, pp. 63–82.
- Kong, Z., Jones, A., Medina Ayala, A., Aydin Gol, E., Belta, C., 2014. Temporal logic inference for classification and prediction from data. In: Proceedings of the 17th International Conference on Hybrid Systems: Computation and Control. pp. 273–282. <http://dx.doi.org/10.1145/2562059.2562146>.
- Laxhammar, R., Falkman, G., 2010. Conformal prediction for distribution-independent anomaly detection in streaming vessel data. In: Proceedings of the First International Workshop on Novel Data Stream Pattern Mining Techniques. pp. 47–55.
- Laxhammar, R., Falkman, G., 2013. Online learning and sequential anomaly detection in trajectories. *IEEE Trans. Pattern Anal. Mach. Intell.* 36 (6), 1158–1173.
- Li, L., Das, S., John Hansman, R., Palacios, R., Srivastava, A.N., 2015. Analysis of flight data using clustering techniques for detecting abnormal operations. *J. Aerosp. Inf. Syst.* 12 (9), 587–598. <http://dx.doi.org/10.2514/1.I010329>.
- Li, L., Gariel, M., Hansman, R.J., Palacios, R., 2011. Anomaly detection in onboard-recorded flight data using cluster analysis. In: 2011 IEEE/AIAA 30th Digital Avionics Systems Conference. IEEE, <http://dx.doi.org/10.1109/DASC.2011.6096068>, 4A4–1.
- Li, L., Hansman, R.J., Palacios, R., Welsch, R., 2016. Anomaly detection via a Gaussian mixture model for flight operation and safety monitoring. *Transp. Res. C* 64, 45–57. <http://dx.doi.org/10.1109/10.1016/j.trc.2016.01.007>.
- Liu, W., Hwang, I., 2011. Probabilistic trajectory prediction and conflict detection for air traffic control. *J. Guid. Control Dyn.* 34 (6), 1779–1789.
- Logan, T.J., 2008. Error prevention as developed in airlines. *Int. J. Radiat. Oncol. Biol. Phys.* 71 (1), S178–S181.
- Ma, J., Perkins, S., 2003. Time-series novelty detection using one-class support vector machines. In: Proceedings of the International Joint Conference on Neural Networks, 2003, Vol. 3. IEEE, pp. 1741–1745. <http://dx.doi.org/10.1109/IJCNN.2003.1223670>.
- Mahindru, R., Kumar, H., Bansal, S., 2021. Log anomaly to resolution: AI based proactive incident remediation. In: 2021 36th IEEE/ACM International Conference on Automated Software Engineering. ASE, IEEE, pp. 1353–1357.

- Matsuno, Y., Tsuchiya, T., Wei, J., Hwang, I., Matayoshi, N., 2015. Stochastic optimal control for aircraft conflict resolution under wind uncertainty. *Aerospace Sci. Technol.* 43, 77–88.
- Memarzadeh, M., Matthews, B., Avrek, I., 2020. Unsupervised anomaly detection in flight data using convolutional variational auto-encoder. *Aerospace* 7 (8), 115. <http://dx.doi.org/10.3390/aerospace7080115>.
- Nanduri, A., Sherry, L., 2016. Anomaly detection in aircraft data using recurrent neural networks (RNN). In: 2016 Integrated Communications Navigation and Surveillance. ICNS, Ieee, <http://dx.doi.org/10.1109/ICNSURV.2016.7486356>, 5C2–1.
- Narasimhan, S., Brownston, L., 2007. HyDE-a general framework for stochastic and hybrid modelbased diagnosis. *Proc. DX* 7, 162–169.
- Olive, X., Basora, L., 2020. Detection and identification of significant events in historical aircraft trajectory data. *Transp. Res. C* 119, 102737.
- Olive, X., Morio, J., 2019. Trajectory clustering of air traffic flows around airports. *Aerospace Sci. Technol.* 84, 776–781. <http://dx.doi.org/10.1016/j.ast.2018.11.031>.
- Pimentel, M.A., Clifton, D.A., Clifton, L., Tarassenko, L., 2014. A review of novelty detection. *Signal Process.* 99, 215–249.
- Puranik, T.G., Mavris, D.N., 2018. Anomaly detection in general-aviation operations using energy metrics and flight-data records. *J. Aerosp. Inf. Syst.* 15 (1), 22–36. <http://dx.doi.org/10.2514/1.I010582>.
- Saâdaoui, A., Souayah, N.B.Y.B., Bouhoula, A., 2017. FARE: FDD-based firewall anomalies resolution tool. *J. Comput. Sci.* 23, 181–191.
- Sankararaman, S., Roychoudhury, I., Zhang, X., Goebel, K., 2017. Preliminary investigation of impact of technological impairment on trajectory-based operation. In: 17th AIAA Aviation Technology, Integration, and Operations Conference. p. 4488.
- Schwabacher, M., Oza, N., Matthews, B., 2009. Unsupervised anomaly detection for liquid-fueled rocket propulsion health monitoring. *J. Aerosp. Comput. Inf. Commun.* 6 (7), 464–482.
- Seah, C.E., Hwang, I., 2009. Stochastic linear hybrid systems: Modeling, estimation, and application in air traffic control. *IEEE Trans. Control Syst. Technol.* 17 (3), 563–575.
- Shaffer, G., Vovk, V., 2008. A tutorial on conformal prediction. *J. Mach. Learn. Res.* 9 (3).
- Shin, S., Hwang, I., 2017. Data-mining-based computer vision analytics for automated helicopter flight state inference. *J. Aerosp. Inf. Syst.* 14 (12), 652–662. <http://dx.doi.org/10.2514/1.I010517>.
- Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., Pei, D., 2019. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2828–2837.
- Timar, S.D., Griffin, K., Borener, S., Knickerbocker, C., 2012. Analysis of s-turn approaches at john f. Kennedy airport. In: 2012 IEEE/AIAA 31st Digital Avionics Systems Conference. DASC, IEEE, pp. 3C1–1. <http://dx.doi.org/10.1109/DASC.2012.6382310>.
- Valasek, J., Chen, W., 2003. Observer/Kalman filter identification for online system identification of aircraft. *J. Guid. Control Dyn.* 26 (2), 347–353.
- Windmann, S., Jiao, S., Niggemann, O., Borcherding, H., 2013. A stochastic method for the detection of anomalous energy consumption in hybrid industrial systems. In: 2013 11th IEEE International Conference on Industrial Informatics. INDIN, IEEE, pp. 194–199.
- Yoon, S., MacGregor, J.F., 2000. Statistical and causal model-based approaches to fault detection and isolation. *AIChE J.* 46 (9), 1813–1824.
- Zhang, X., Mahadevan, S., 2019. Ensemble machine learning models for aviation incident risk prediction. *Decis. Support Syst.* 116, 48–63.
- Zhu, S.-P., Huang, H.-Z., Peng, W., Wang, H.-K., Mahadevan, S., 2016. Probabilistic physics of failure-based framework for fatigue life prediction of aircraft gas turbine discs under uncertainty. *Reliab. Eng. Syst. Saf.* 146, 1–12.
- Zou, Q., Xie, S., Lin, Z., Wu, M., Ju, Y., 2016. Finding the best classification threshold in imbalanced classification. *Big Data Res.* 5, 2–8.