Thirteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2019)

Data-Driven Precursor Detection Algorithm for Terminal Airspace Operations

Raj Deshmukh, Dawei Sun and Inseok Hwang
School of Aeronautics and Astronautics, Purdue University
West Lafayette, IN 47906, USA
{rdeshmu, sun289, ihwang}@purdue.edu

***Abstract*—The air traffic management system is one of the most complex man-made systems, with stringent standards for safety and operational performance. Modern surveillance systems make available detailed flight and airport information, through on-board and ground recording systems. These recorded datasets can be used for detecting and/or predicting anomalies which hinder safe and efficient operations. The prediction of an anomaly is performed by identifying events that precede the occurrence of an anomaly, which are called precursors. In this paper, we propose a detection algorithm that can identify precursors for flight anomalies through data-driven models designed with surveillance data recorded in the terminal airspace. The proposed algorithm is demonstrated to detect precursors of flight anomalies in the terminal airspace around LaGuardia (LGA) airport in New York City using real traffic data obtained from the Airport Surface Detection Equipment - Model X (ASDE-X) and the Terminal Automation Information Service (TAIS) surveillance datasets.**

Keywords-air traffic management; terminal airspace; precursor detection; machine learning; anomaly detection; prognosis

I. Introduction

To assist air traffic controllers (ATCs) with airspace operations while ensuring high efficiency and safety, it is important to analyze operational anomalies. Recently, there have been extensive efforts to develop anomaly detection algorithms for unlabeled aviation datasets using machine learning techniques [1]–[6]. Once the anomalies are detected, the next step is to find the *causes* for these anomalies and to aid ATCs and pilots in mitigating such anomalous behavior. The approach to finding the causes is called prognosis and requires the identification of events or conditions that precede the anomaly and have some correlation to the occurrence of the anomaly, called *precursors*. If detected, the precursor could be used to initiate actions to avoid the anomaly from ever occurring or mitigate it.

There have been extensive research efforts to develop precursor detection algorithms to predict anomalies. One approach to prognosis is to use a physics-based or a rule-based model of the system, and this approach relies on models designed by domain experts [7]. Common examples of such an approach is system health monitoring [8][9] and conformance monitoring [10][11]. Another approach is the data-driven approach, also known as the data mining or machine learning approach. This approach uses large records of historical data to learn a model of system behavior. In the aviation domain, data-driven approaches have been applied to structural, gas turbine, battery prognostics, etc. [12]–[14]. Considering that air traffic management is an increasingly complex system with a continuously evolving behavior, we propose a data-driven precursor detection algorithm for anomalies in the terminal airspace.

The main objective of this paper is to develop a supervised precursor detection algorithm to augment the unsupervised anomaly detection algorithm developed in our earlier work [6], called TempAD. The proposed precursor detection algorithm, called *reactive TempAD*, detects precursors using the surveillance data that correlate to specific anomalies in terminal airspace operations. We demonstrate its performance by summarizing and analyzing the results of the proposed algorithm applied to terminal airspace surveillance data, such as Airport Surface Detection Equipment - Model X (ASDE-X) and Terminal Automation Information Service (TAIS) datasets. The recorded datasets are obtained from aircraft during their final approach to the LaGuardia (LGA) airport in New York City.

The outline of the paper is as follows: Section II describes the precursor detection problem and the preliminaries regarding data pre-processing and anomaly detection. In Section III, the precursor detection algorithm is developed using supervised machine learning techniques. In Section IV, we demonstrate the working of the proposed precursor detection algorithm for predicting go-around and S-turn anomalies detected for arrivals to LGA using the ASDE-X and TAIS datasets. Finally, concluding remarks are made in Section V.

II. Preliminaries for Precursor Detection

In this section, we first describe the data which is used as input to the precursor detection algorithm after pre-processing, followed by the anomaly detection algorithm to which we augment our precursor detection algorithm. These help us set up the problem for precursor detection, which we aim to solve.

A. Input Data and Pre-processing

The input surveillance data comprises of the aircraft track data which contains the position (latitude and longitude), speed, altitude, heading and a unique flight identifier for each flight in a time-series format, recorded for arrival flights to the LaGuardia (LGA) airport. We source the surveillance data from the Airport Surface Detection Equipment – Model X (ASDE-X) [15] and the Terminal Automation Information Service (TAIS) [16] datasets. The ASDE-X dataset has a detection range of about 20 nautical miles from the airport and records data at every second. On the other hand, the TAIS dataset has a detection range of about 141 nautical miles from the airports and records data at a rate of 5 seconds. The recording used in this paper from ASDE-X is between April 6th to 24th, 2016, which has records of 9,634 arrivals at LGA, and for TAIS is during the months of September, October and November 2016, which has records of 36,243 arrivals at LGA.

The setup of the proposed algorithm requires extracting specific aircraft states which are combined and identified as precursors. We use features corresponding to the following dimensions in this paper:

- Horizontal (H): Obtained from the positional (latitude and longitude) time-series data
- Vertical (V): Obtained from the altitude (h) time-series data
- Speed (S): Obtained from the ground speed (v) time-series data
- Specific Total Energy (STE): $h + v^2/2g$
- Specific Potential Energy Rate (SPER): $\dot{h}$
- Specific Kinetic Energy (SKE): $v^2/2g$

where g is the gravitational acceleration constant. Note that the first three dimensions are directly available from the ASDE-X and TAIS datasets, but the features in the energy dimensions need to be derived from the others.

B. Anomaly Detection

To detect anomalies in unlabeled terminal airspace surveillance data, we use the temporal logic-based anomaly detection algorithm called TempAD presented in our earlier study [6]. This algorithm is capable of generating mathematical expressions called predicates for the bounds of normal operations in terms of time and physical parameters in a human-readable format. These bounds comprise the anomaly detection model for that specific dataset. An example of such a model for arrivals to the LaGuardia airport runway 22 is shown in Figure 1. The figure presents the anomaly detection models in blue. Any violation of these models is labeled as an anomalous trajectory and is plotted in red, while the normal trajectories are plotted in green.

The corresponding anomaly detection model is expressed as:

$$G_{[20,0]}(-1.1429x + y - 123.1929 > 0) \\ \wedge G_{[20,0]}(-1.3409x + y - 139.8514 < 0)$$

Here, x and y are the longitude and latitude, respectively, in degrees. The expression can be understood as a requirement of all normal aircraft arriving to runway 22 from the northeast to lie within the given bounds during the final 20 nautical miles before touchdown at all points of time (G).

This anomaly detection algorithm has been shown to be capable of detecting anomalies in all five dimensions. Furthermore, the detected anomalies can be segregated and characterized automatically, based on the specific relevance they have to air traffic operations. For example, some identified anomalies are go-around, S-turn (path stretch), overspeed/underspeed, late interception of glideslope, energy excess/deficit, and change of runway anomalies.

Once the flight trajectories are labeled as normal or anomalous, we now find the precursors to these anomalies. In the next section, we describe the precursor detection algorithm, and present specific approaches to apply it to a few of the anomaly types described above.

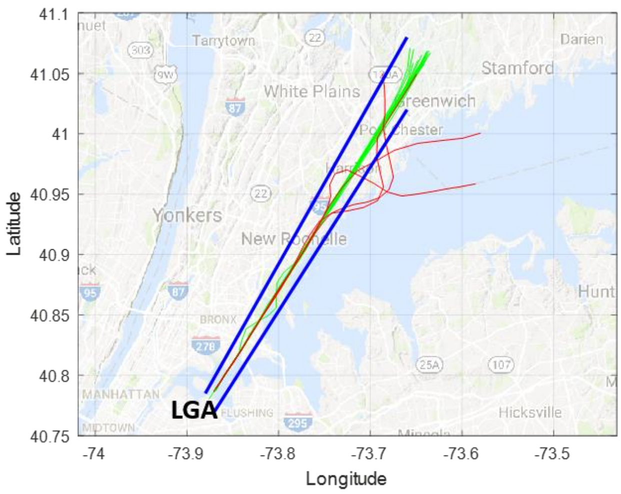

Figure 1. Horizontal anomaly detection for arrivals to LGA RWY22

III. Reactive TempAD Algorithm Development

For the detection of precursors, we investigate the causal relations of the anomalies identified by TempAD and present the concept of a reactive temporal logic algorithm, called *reactive TempAD* [17] to detect precursors (causes) to these anomalies (effects).

Consider a set of N signals (e.g., flights) $\{s_i\}_{i=1}^N$, each of which has the length of T time steps. By truncating s_i to its final $\tilde{t}$ time steps, let $s_{i,e}$ denote the latter part of s_i (i.e., for $[T-\tilde{t}, T]$) and $s_{i,c}$ denote the earlier part of s_i (i.e., for $[0, T-\tilde{t})$), as shown in Figure 2. We assume that an anomaly occurs in the latter part of a signal (called *effect signal*, $s_{i,e}$) and hence its corresponding precursor can be detected in the earlier part of the signal (called *cause signal*, $s_{i,c}$). Note that $\tilde{t}$ is a design parameter which can be determined by analyzing the results of the anomaly detection algorithm (e.g., if all the anomalies occur during the final 5 time steps, $\tilde{t}$ is set as 5).

Then, based on TempAD, we propose a framework for precursor detection, as presented in Figure 3. For the set of unlabeled signals $\{s_{i,e}\}_{i=1}^{N}$, we first apply TempAD to identify whether a signal is normal or abnormal, by which each signal can be labeled ($p_i = 1$ for normal and $p_i = 0$ for abnormal) and an effect model (or anomaly detection model) φ_e is obtained. Then, the set of labeled signals $\{s_{i,c}, p_i\}_{i=1}^{N}$ is fed into a supervised learning algorithm for precursor detection, called reactive TempAD, to learn a cause model (or precursor detection model) φ_c such that *if and only if $s_{i,c}$ violates φ_c in $[0, T-\tilde{t})$, it is guaranteed that $s_{i,e}$ violates φ_e within $[T-\tilde{t}, T]$.*

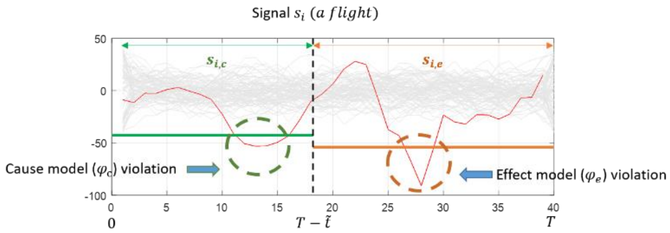

Figure 2. Cause and effect in a signal

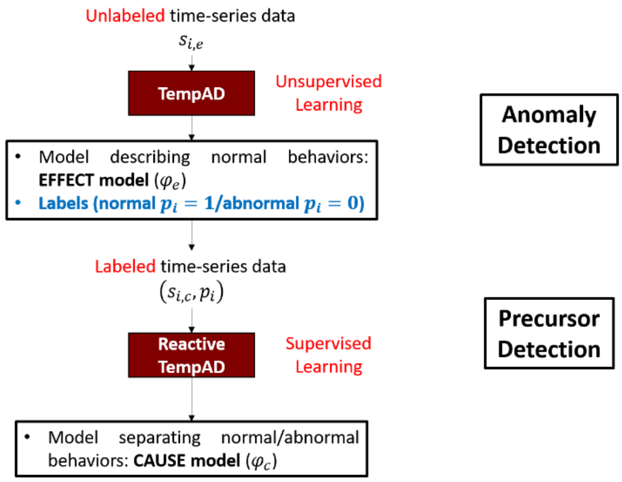

Figure 3. Architecture of proposed precursor detection algorithm with TempAD

Note that the main difference between reactive TempAD and the original TempAD algorithm is that reactive TempAD involves supervised learning which incorporates the labeled signal information. We use the support vector machine (SVM) and artificial neural network (ANN) machine learning algorithms for generating the separating hyperplane for classification.

Furthermore, note that the cause model and the effect model may be implemented using different features. This is essential for precursor detection in aircraft operations, since a cause in one feature dimension can have an effect in another feature dimension. For example, a go-around anomaly could be detected using an effect model $\varphi_{e,alt}$ implemented using the altitude feature, but the cause model φ_c for precursor detection does not *necessarily* indicate any physical and operational significance in the altitude feature but can be better described in other features (e.g., energy).

In the next section, we will choose specific anomalies and apply the precursor detection algorithm to predict these anomalies using the aircraft states prior to the occurrence of the anomaly.

IV. PRECURSOR DETECTION USING REACTIVE TEMPAD

In this section, we apply the proposed precursor detection algorithm to the go-around anomaly and the S-turn anomaly and analyze the results.

A. Precursor Detection for Go-around Anomalies

A go-around anomaly is one where the pilot in the anomalous flight aborts the landing during the short final, climbs using a missed approach procedure, circles around and attempts landing again. A visualization of the horizontal trajectory during a go-around is illustrated in Figure 4. A go-around anomaly is a severe anomaly, in the sense that it manifests as an anomaly and is detected in all the dimensions.

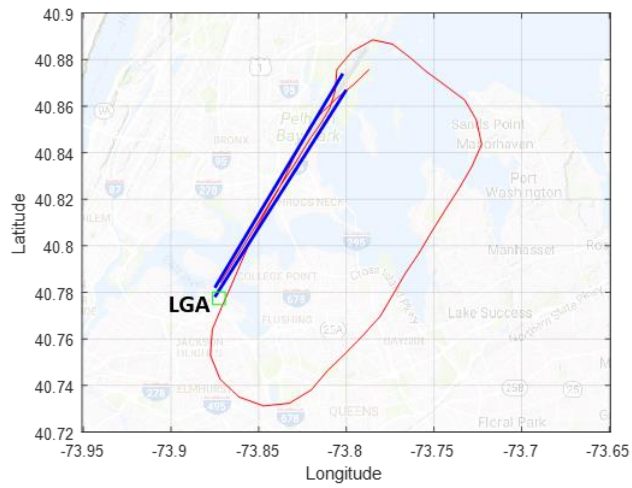

Figure 4. Horizontal view of go-around anomaly

To illustrate the use of reactive TempAD for precursor detection, we implement this algorithm to find a precursor for all the go-around anomalies detected at LGA airport across 19 days of the ASDE-X data (9,634 flights). To determine the feature capable of delivering the best precursor detection performance, we search across several features and choose the feature that grants the highest F1 classification score, thereby implying a good correlation between the occurrence of the precursor and the occurrence of the anomaly. The F1 classification score can be computed as:

Precision = ratio of true anomaly detection of total precursor detections

Recall = ratio of anomalies detected by precursors over total true anomalies

F1 score $= 2 \times \frac{Precision \times Recall}{Precision + Recall}$.

It was observed that a derived feature (f) from other energy features gave the highest F1 score from all tested candidate features:

$$f = \frac{SPER}{SPE} \times SKE = \frac{\dot{h}}{h} \times \frac{v^2}{g}.$$

Here, SPER is the specific potential energy rate, SPE is the specific potential energy (altitude) and SKE is the specific kinetic energy.

The results of reactive TempAD with this feature f are presented in Figure 5, Figure 6 and Table 1: the go-around flights (signals) are in red and the normal flights are in green. Figure 5 demonstrates go-around anomaly detection, while Figure 6 demonstrates go-around precursor detection. Thus, if the go-around flights violate the cause (or precursor) model (φ_c) in Figure 6, it is guaranteed within a margin of error that the effect (or anomaly) model (φ_e) will be violated in Figure 5, where the cause and effect models are respectively given as:

$$\varphi_c = G_{[0,55)}\left(\frac{SPER}{SPE} SKE < 37\right)$$
$$\varphi_e = G_{[55,60)}(1650 \times t + 75 \times \text{altitude} < 123750).$$

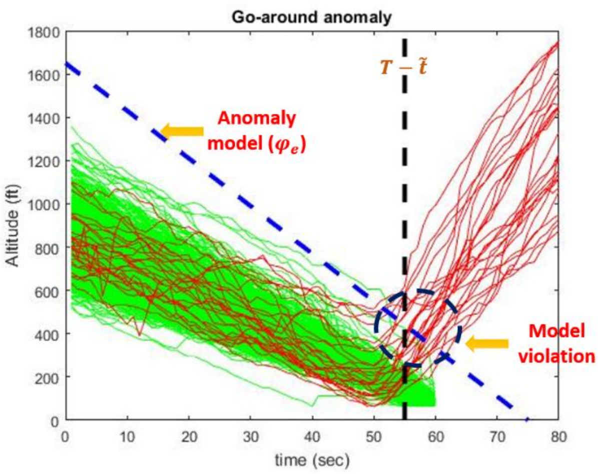

Figure 5. Go-around anomaly detection model

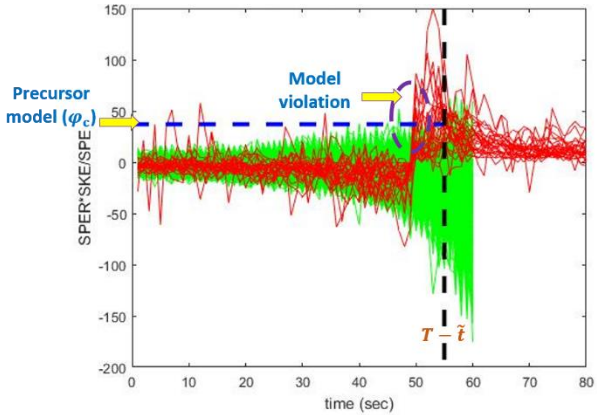

Figure 6. Precursor detection model

Using Figure 6, we can detect violations of the precursor model by determining signals which go above the precursor model (in blue) at any point of time through their trajectory in the first 55 time steps. Thus, using Figure 5 and Figure 6, we can determine true positives as flights that violate the precursor model in the first 55 time steps and the anomaly model in the last 5 time steps. Similarly, true negatives, false negatives (missed detections) and false positives (false alarms) can be determined to form a confusion matrix. Table 1 summarizes the results of reactive TempAD for the go-around anomaly in such a confusion matrix. The accuracy of the algorithm is 99.81% when 5-fold cross validation is used. The precision value is 87.50%, the recall value is 72.92% and thus, the F1 score is 79.55%.

TABLE 1. CONFUSION MATRIX FOR GO-AROUND PRECURSOR DETECTION USING REACTIVE TEMPAD-SVM

True \ **Predicted**	*Anomalous flights*	*Normal flights*	*Total*
Anomalous flights	True positive 35	False negative 13	48
Normal flights	False positive 5	True negative 9,581	9,596
Total	40	9,594	9,634

To characterize the operational usability of the precursor detection algorithm, we introduce the concept of look-ahead time, which is the time difference between the detection of the precursor and the time of occurrence of the anomaly. It is desirable that this look-ahead time be as long as possible, since it allows more time for the ATC and pilot to make decisions and better accommodates pilot input lag.

The above feature f resulted in an average look-ahead time of 7 seconds, which is a decent number for precursor detection. However, if it were to be applied to real-time implementation, it might not be sufficient enough since it gives the pilot and ATC only 7 seconds to react to the anomaly. From aircraft design and missed approach studies, 7-8 seconds are always required for engine spool-up during a go-around [18] and thus the look-ahead time achieved here is just feasible to be used.

To allow for a longer reaction time and make the precursor detection algorithm more effective, we modified the framework of the reactive TempAD algorithm to use the artificial neural network (ANN) learning algorithm (called *reactive TempAD-ANN*) instead of SVM (called *reactive TempAD-SVM*). ANN automatically weighs the best combination of diverse features in a manner better than SVM and gives better test results. Further, ANN can merge information contained in multiple features. Thus, it explores more complex causal relations than SVM and is expected to give a longer average look-ahead time. Note that due to the complex nature of the ANN algorithm, it is not always possible to physically interpret or to present the mathematical expression for the classification model succinctly.

Furthermore, a new feature (distance to preceding aircraft) is introduced into the feature space, which determines the vertical and horizontal distances to the nearest preceding flight (arriving or departing) and attempts to learn from loss of separation in terminal airspace. With the new feature set, using reactive TempAD-ANN, the results are improved with a test accuracy of 99.87%, precision of 89.13%, recall of 85.41% and an F1 classification score of 87.23%, as shown in Table 2. More importantly, the average look-ahead time is increased to 11 seconds, an improvement of nearly 57%.

TABLE 2. CONFUSION MATRIX FOR GO-AROUND PRECURSOR DETECTION USING REACTIVE TEMPAD-ANN

Predicted / **True**	***Anomalous flights***	***Normal flights***	***Total***
Anomalous flights	True positive 41	False negative 7	48
Normal flights	False positive 5	True negative 9,581	9,596
Total	46	9,588	9,634

It is important to note that the number of false negatives using the reformulated reactive TempAD-ANN has nearly halved. This is important, as minimizing instances of false negatives (anomalies not predicted by the precursor, i.e., missed detections) is vital in safety-critical applications such as air traffic management (ATM).

B. Precursor Detection for S-turn Anomalies

To further demonstrate the applicability and performance of the reactive TempAD algorithm, we demonstrate precursor detection for the S-turn anomaly. An S-turn anomaly is a special anomaly in the horizontal dimension which involves an extension of flight path, called a path stretch [19]. They are generally caused due to significant safety and operations related incidents and are critical anomalies in the terminal airspace [20]. Examples of S-turn trajectories detected by the TempAD algorithm for arrivals to LGA runway 22 and runway 4 are presented in red in Figure 7 and Figure 8, respectively.

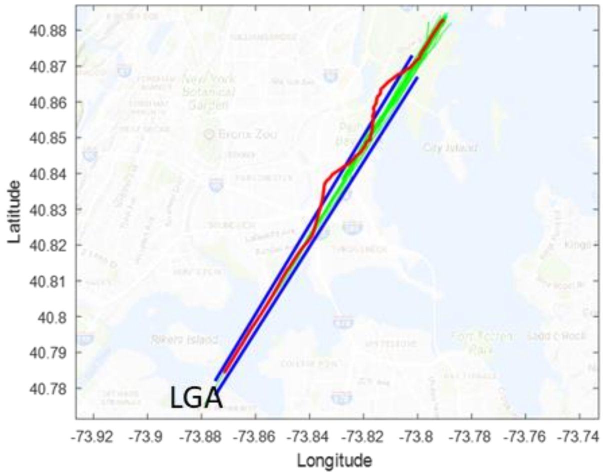

Figure 7. Horizontal view of S-turn anomaly in arrivals to RWY22 at LGA

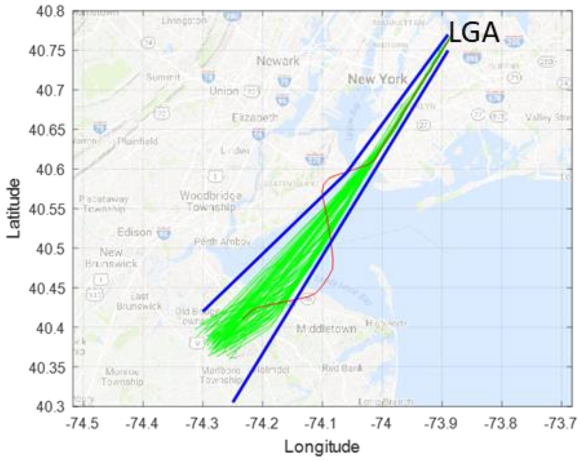

Figure 8. Horizontal view of S-turn anomaly in arrivals to RWY4 at LGA

To illustrate the use of reactive TempAD for precursor detection for S-turn anomalies, we implement this algorithm to find a precursor for all the S-turn anomalies detected at LGA airport across 88 days (September – November 2016) of the TAIS data (36,243 flights). Considering that the TAIS dataset (nearly 3 months) is much larger in size than the ASDE-X dataset (19 days), a higher training data size is expected to result in more effective models using supervised learning. The precursors for the S-turn anomaly were searched from features comprising of separation from preceding flight (both arriving and departing flights); excess approach speed; and altitude. Here, the first feature is related to loss of separation, while the last two features relate to unstable approach of the arriving flight. The results of testing reactive TempAD-ANN over 88 days of the TAIS dataset for arrivals to LGA are presented in Table 3.

TABLE 3. CONFUSION MATRIX FOR S-TURN PRECURSOR DETECTION USING REACTIVE TEMPAD-ANN

Predicted / True	*Anomalous flights*	*Normal flights*	*Total*
Anomalous flights	True positive 714	False negative 195	909
Normal flights	False positive 97	True negative 35,237	35,334
Total	811	35,432	36,243

The classification results have a test accuracy of 99.19%, a precision of 88%, recall of 78.57% and an F1 classification score of 83.01%. The average look-ahead time for precursor detection for the S-turn anomaly is nearly 9 seconds.

V. CONCLUSION

This paper considers the problem of detecting precursors to anomalies in flights using aircraft surveillance datasets. The identified precursors can potentially be used to avoid or mitigate the corresponding anomaly, if detected sufficiently in advance to the anomaly. To address the problem of precursor detection, a temporal logic-based precursor detection algorithm – called reactive TempAD – has been developed which determines causal relations to the occurrence of the anomaly. This algorithm is a supervised learning algorithm that relies on labels generated using an unsupervised anomaly detection algorithm. For the purpose of demonstration, we use real surveillance datasets of Airport Surface Detection Equipment - Model X (ASDE-X) and Terminal Automation Information Service (TAIS) for arrivals to the LaGuardia (LGA) airport. Using these datasets, precursors to go-around and S-turn anomalies are detected and corresponding operational performance results are analyzed to demonstrate the working of the algorithm.

In the future, we plan to extend the applicability of precursor detection to several other types of anomalies detected in the terminal airspace. Furthermore, it was observed that some anomalies were preceded by a sequence of precursors. Facilitating the algorithm to analyze this sequence and detect the precursor that is the most distant from the anomaly will enhance the algorithm by generating a longer look-ahead time, thereby giving the ATC and pilots a longer time to make decisions and mitigate the anomaly.

ACKNOWLEDGMENT

This research is funded by the National Aeronautics and Space Administration (NASA) through the NASA Big Data Analytics project (NNH15ZEA001N). The authors are grateful to Dr. Andrew Churchill at Mosaic ATM, Megan Hawley at Honeywell, and Dr. Nikunj C. Oza at NASA Ames for their valuable comments and continued support.

REFERENCES

[1] L. Li, R.J. Palacios, and R. Welsch, “Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring,” Transportaion Research Part C: Emerging Technolgies, 64, pp. 45-57, March 2016.

[2] G. Dimitry, B. Matthews, and R. Martin, “Aircraft anomaly detection using performance models trained on fleet data,” IEEE Conference on Intelligent Data Understanding (CIDU), pp. 17-23, October 2012.

[3] A.N. Srivastava, “Greener aviation with virtual sensors: a case study,” Data Mining and Knowledge Discovery, 24(2), pp. 443-471, March 2012.

[4] S. Das, B.L. Matthews, A.N. Srivastava, and N.C. Oza, “Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study,” Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data mining, pp. 47-56, July 2010.

[5] L. Li, S. Das, R.J. Hansman, R. Palacios, and A.N. Srivastava, “Analysis of flight data using clustering techniques for detecting abnormal operations,” Journal of Aerospace information systems, 12(9), pp. 587-598, September 2015.

[6] R. Deshmukh, and I. Hwang, “Anomaly Detection Using Temporal Logic Based Learning for Terminal Airspace Operations,” AIAA Scitech 2019 Forum, January 2019.

[7] M. Schwabacher, “A survey of data-driven prognostics,” Infotech@ Aerospace, September 2005.

[8] I . Turner, and A. Bajwa, “A survey of aircraft engine health monitoring systems,” 35th Joint Propulsion Conference and Exhibit, June 1999.

[9] W. Staszewski, C. Boller, C., and G.R. Tomlinson, Health monitoring of aerospace structures: smart sensor technologies and signal processing, John Wiley & Sons, April 2004

[10] C.E. Seah, A. Aligawesa, and I. Hwang, “Algorithm for conformance monitoring in air traffic control,” Journal of Guidance, Control, and Dynamics, 33(2), pp. 500-509, March 2010.

[11] G.W. Mann, and I. Hwang, “Four-Dimensional Aircraft Taxiway Conformance Monitoring with Constrained Stochastic Linear Hybrid Systems,” Journal of Guidance, Control, and Dynamics, 35(5), pp. 1593-1604 September 2012.

[12] I. Tumer, and E.M. Huff, “Analysis of triaxial vibration data for health monitoring of helicopter gearboxes,” Journal of vibration and acoustics, 125(1), pp. 120-128, January 2003.

[13] H.E.G. Powrie, and C.E. Fisher, “Engine health monitoring: Towards total prognostics,” Proceedings of the IEEE Aerospace Conference, vol. 3, pp. 11-20, March 1999.

[14] L. Atlas, et al., “An evolvable tri-reasoner IVHM system,” Proceedings of the IEEE Aerospace Conference, vol. 6, pp. 3023-3037, 2001.

[15] “Airport Surface Detection Equipment, Model X (ASDE-X),” URL: https://www.faa.gov/air_traffic/technology/asde-x/, Augsut 2014.

[16] “SWIM Terminal Data Distribution System (STDDS),” URL: https://www.faa.gov/air_traffic/technology/swim/stdds/media/FIXM_Mediated_STDDS_Data_Overview_v2.pdf, October 2017.

[17] Z. Kong, A. Jones, A., A.M. Ayala, E.A. Gol, and C. Belta, “Temporal logic inference for classification and prediction from data,” Proceedings of the 17th international conference on Hybrid systems: computation and control, pp. 273-282, April 2014.

[18] T. Blajev, and W. Curtis, “Go-Around Decision Making and Execution Project: Final Report to Flight Safety Foundation," Flight Safety Foundation, 2017.

[19] S.D. Timar, K. Griffin, S. Borener, and C.J. Knickerbocker, “Analysis of s-turn approaches at John F. Kennedy airport,” In Digital Avionics Systems Conference (DASC), October 2012.

[20] S. Green, R. Vivona, R., M. Grace, M., and T.C. Fang, “Field evaluation of descent Advisor Trajectory Prediction accuracy for en-route clearance advisories,” Guidance, Navigation, and Control Conference and Exhibit, 1998.

Raj Deshmukh is a Ph.D. student in the School of Aeronautics and Astronautics at Purdue University. He received his Master’s degree from the

same school in May 2017. His research interests are oriented towards machine learning applications to air traffic management and towards distributed consensus estimation & control.

Dawei Sun received his Master's degree in May 2018, and is now a Ph.D. student in the School of Aeronautics and Astronautics at Purdue University. His research interests include resilient control and machine learning application to air traffic management.

Inseok Hwang is a professor in the School of Aeronautics and Astronautics at Purdue University. His research interests lie in modeling, estimation, and control of cyber-physical systems (CPS) and their applications to safety critical systems such as aircraft/spacecraft/unmanned aerial systems (UAS), air traffic management (ATM), and multi-agent systems. For his research, he leads the Flight Dynamics and Control/Hybrid Systems Laboratory at Purdue University. He received the NSF CAREER award in 2008, was selected as one of the nation's brightest young engineers by the National Academy of Engineering (NAE) in 2008, and received the AIAA Special Service Citation in 2010. He is an Associate Fellow of AIAA, and a member of the IEEE Control Systems Society and the IEEE Aerospace and Electronic Systems Society. He is currently an associate editor of the IEEE Transactions on Aerospace and Electronic Systems and the Asian Journal of Control.