LSTM-based Flight Trajectory Prediction

Zhiyuan Shi*†, Min Xu*†, Quan Pan*, Bing Yan* and Haimin Zhang†

* Key Laboratory of Information Fusion Technology, Ministry of Education
Northwestern Polytechnical University, Xi'an, China 710072
Email: Zhiyuan.Shi@student.uts.edu.au
†School of Electrical and Data Engineering
University of Technology Sydney, Ultimo, NSW, Australia 2007
Email: Min.Xu@uts.edu.au

Abstract—Safety ranks the first in Air Traffic Management (ATM). Accurate trajectory prediction can help ATM to forecast potential dangers and effectively provide instructions for safely traveling. Most trajectory prediction algorithms work for land traffic, which rely on points of interest (POIs) and are only suitable for stationary road condition. Compared with land traffic prediction, flight trajectory prediction is very difficult because way-points are sparse and the flight envelopes are heavily affected by external factors. In this paper, we propose a flight trajectory prediction model based on a Long Short-Term Memory (LSTM) network. The four interacting layers of a repeating module in an LSTM enables it to connect the long-term dependencies to present predicting task. Applying sliding windows in LSTM maintains the continuity and avoids compromising the dynamic dependencies of adjacent states in the long-term sequences, which helps to improve accuracy of trajectory prediction. Taking time dimension into consideration, both 3-D (time stamp, latitude and longitude) and 4-D (time stamp, latitude, longitude and altitude) trajectories are predicted to prove the efficiency of our approach. The dataset we use was collected by ADS-B ground stations. We evaluate our model by widely used measurements, such as the mean absolute error (MAE), the mean relative error (MRE), the root mean square error (RMSE) and the dynamic warping time (DWT) methods. As Markov Model is the most popular in time series processing, comparisons among Markov Model (MM), weighted Markov Model (wMM) and our model are presented. Our model outperforms the existing models (MM and wMM) and provides a strong basis for abnormal detection and decision-making.

I. INTRODUCTION

Over the past few years, with the rapid growth of traffic data, transportation management and control becomes data driven and requests significant increasing computing resources. As one important means of modern traffic, civil aviation enjoys a rapid development both in scale and technology. Up to the year 2014, the total number of global civil flights was about 36.5 million, with an average of about 100,000 per day, and has a sustained growth trend. In addition, chartered plane, low altitude unmanned aerial vehicle, helicopters are very active in the low altitude airspace in recent years. Though it leads to a convenient and fast travel all around the world, civil aviation brings huge burden to the Air Traffic Management (ATM) systems. According to the Eurocontrol Base of Aircraft Data (BADA) [1], the trajectories of aircrafts should be predictable in the sense of ATM.

Forecasting is vital to avoid mistakes or reduce errors when making decisions. Recently, forecasting has been widely used
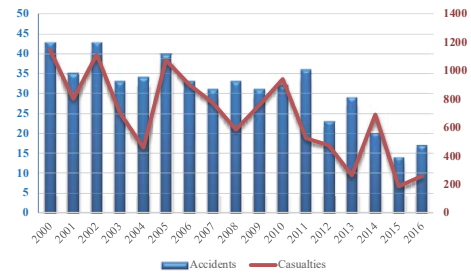

Fig. 1. Air accidents and casualties all over the world since 2000.

for traffic management, such as aircraft trajectory prediction, maritime traffic forecasting, vehicle traffic predicting, traffic flow prediction and pedestrian trajectory prediction. Accordingly, there are a lot of long-term, medium-term, short-term forecasting techniques presented. The most popular model may be the Markov Model (MM) and its variants [2–4], which model trajectory sequentially. Besides MM, other methods used for prediction include Autoregressive model [5], Clustering and Artificial Neural Networks [6, 7], and Deep Neural Networks [8–10].

Trajectory prediction methods for land traffic or pedestrian are strictly constrained by the already-known and stationary road and points of interest (POIs) in the horizontal plane. While in the flight trajectory prediction, way-points are too sparse to label a trajectory in the ascending, cruising and descending phases. The flight envelopes are heavily affected by the external factors, such as geographic factor, bad weather or manual operations. Delays or collisions are mainly resulted from the above mentioned problems. Prediction of flight trajectory becomes more and more vital in ATM. Related prediction works on flight trajectory are reviewed in the following. Delay prediction [11–13] was widely studied, which is related to commercial interests of airlines and choice of passengers. In [14], five machine learning methods including MLR-FS, Ridge, PCR, NNet and GBM were applied for mass estimation for aircrafts at climbing phase. Despite the application of advanced control and monitoring systems on aircrafts, aviation safety still has to be very concerned. Fig. 1 shows the airspace accidents and casualties from the year 2000 to 2016. Potential risks and dangers can be warned in advance
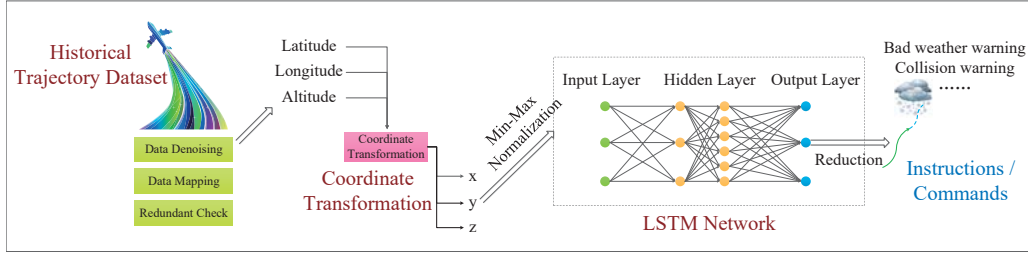
Fig. 2. The proposed LSTM-based flight trajectory prediction.

by estimating intentions through trajectory prediction. Future dynamic features can be estimated or predicted by learning the historical observations combined with static information such as airways, airports and no-flight zones (NFZ).

Conventional methods relying on kinematics and aerodynamic models can accurately describe the status of an aircraft [15]. However, related to the types of aircrafts, the coefficients of the 6-DOF equation of motion (see details in III-B) are required for ground stations to predict flight trajectory. However, these coefficients are unobservable from the view of surveillance, which makes it extremely difficult for ground stations to predict the trajectory without knowledge of these coefficients. Moreover, most of the existing methods predict trajectory in 2-D space (i.e. in the horizontal plane). Errors in modelling and sensing the complexity and difficulty in trajectory processing when considering extra dimensional data.

In this paper, we are the first to propose a flight trajectory prediction method based on Long Short-Term Memory (LSTM) network, which can accurately predict 3-D and 4-D flight trajectories without using the physical model of aircraft. As shown in Fig. 2, our method has three major steps, Historical dataset collecting, Coordinate transforming and LSTM network building. Instructions or commands are made by ATM based on the output of our LSTM-based model.

The major contributions of this research can be summarized as follows.

- To the best of our knowledge, we are the first to introduce LSTM network for flight trajectory prediction. Instead of having a single layer, a common LSTM unit consists of a cell, an input gate, an output gate and a forget gate. The cell is responsible for keeping the characteristics of sequence over arbitrary time intervals, which enables the LSTM to process long-term sequences.
- Sliding windows is applied to our LSTM network. In long-term sequences, dynamic features such as pitch or acceleration may be weakened or averaged, which results in low accuracy when predicting. Sliding windows not only maintain the continuity of sequence, but also fully excavate the dynamic dependencies of adjacent points in each window, which helps to improve the prediction precision.
- We are able to predict the trajectory in 4-D space. Mutual constraints and couplings among position, speed and heading are mined in the hidden layers of LSTM network,

which makes it possible for our model to output a smooth and reasonable prediction trajectory in 4-D space.

II. RELATED WORKS

Advanced and densely distributed sensors make it possible to characterize the transportation trajectories into time series. Though the concept of Intelligent Transportation System (ITS) has been proposed and seeking applications to the land, maritime and aviation transport, accidents, breakdowns or emergencies happen unpredictably, which poses a security threat to the public. Detect these threatening events at an early stage, i.e. before they happen will empower government agencies and individual persons for effective and immediate responses and help to minimize impact.

In land traffic, congestion is the primary issue need to be solved in recent period. Traffic flow shows stochastic and nonlinear characteristics due to the practical cases, which makes Kalman filters not able to fully reflect traffic features, though it worked adaptability when traffic is volatile [16]. Clustering methods such as k-NN, performed with the equivalent performance as the linear time-series approach according to [17]. As early as 1970s, the autoregressive integrated moving average (ARIMA) model [18] was applied to short-term traffic flow prediction. And then according to [19], ARIMA (0,1,1) model was the most outstanding forecast method in terms of statistical significance. Recurrent neural network (RNN) showed its advantage in time-series preprocessing, and was applied to traffic prediction in [20]. However, disadvantages of RNNs will be exposed in face of long sequences. Recently, as one of the advanced RNNs, Long Short-Term Memory (LSTM) network [10] was proposed for long-term sequence prediction. In [9], an LSTM network with attention model was utilized for encoding and decoding purpose of pedestrians, combined soft and hardware the motion patterns were also considered. Besides the traffic data, some contextual information can provide significant help for forecasting. User's preferences, traffic police, roads and transportation engineers were considered in [21]. What's more, periodic characteristics can also contribute, and a fuzzy neural network was proposed [22] for forecasting travel speed. A dynamic traffic simulator was first proposed in [5] to generate flows based on historic traffic information. The original-to-destination (OD) matrices were adjusted timely by an optimization methodology. Temporal, spatial and multi-mode characteristics contributed to a

novel prediction approach, which was based on dynamic tensor completion for short-term traffic flow forecast [23].

Maritime trajectories are more complex than the land traffics with two-dimensional flexibility, as seaways are not strictly fixed and sparser than roads. Navigation and surveillance information cannot be updated as frequently as that in land traffic. It is essential to extract vessels' waterway pattern to identify regular seaways. Without incorporating motion patterns, directly applying statistical or machine learning approaches and neural networks may lead to large error. The pattern of vessels can be extracted by clustering the trajectories, Lattice-based DBSCAN and fuzzy c-means algorithms were proposed in [7, 24]. With the assistance of automatic identification system (AIS), the authors in [25] were able to predict future trajectories of surrounding vessels in a recursively way.

Delay, 4-D trajectory prediction and collision risks have attracted much research attentions in ATM research community. Statistical methods analyzed air traffic delays in Long-term and Short-term patterns [26]. Bayesian network [11] was proposed to estimate delay propagation. Machine learning [12] was also used for predicting air traffic delays. [13] modeled the delayed flight of an individual airport by the LSTM architecture. Turning points may provide rich information, which were clustered and then used to build a stochastic model [27]. Clustering models were used to identify the hot-spot [28] where airspace collision may occur. Historic flight data was trained on collision risk. Several key factors were also taken into consideration, such as navigation accuracy, airway structure, crossing tracks, aircraft size and the voyage. For 4-D trajectory prediction, K-means was carried out on radar data for aircraft monitoring in actual operation rate. Fuzzy clustering method was used to analyze flight data in arrival phase. In [29], a clustering method combining time warp edit distance (TWED) with K-means algorithm was proposed to improve the accuracy of nominal flight profile.

The rest of the paper is organized as follows. We introduce the assumptions and formulate the problem of flight trajectory prediction in section III. In section IV, we introduce the LSTM network for flight trajectory prediction. Section V presents the experimental results and comparative tests. Concluding remarks are described in section VI.

III. Problem Formulation

Aircrafts must obey the traffic rules just like the vehicles running on the land. The scheduled rules in air traffic are usually called "flight plans", which mainly include route or flight paths, flight levels, special use airspace, alternate airports, time-line and so on. It may be delayed for an aircraft to take off when the ground runway or airways are busy or beyond their capacity according to the timely traffic flow. Though way-points and airways divide the airspace into grid-like fixed routes, heavy rains, thunders, hails or mis-operation still throw a threat on aircrafts when they start the cruising phase. What's more, pilots should keep constant attention on the minimum separation in case of crash. Hidden dangers exist in all phases. The primary task of ATM is to avoid colliding

among aircrafts or obstacles, so as to ensure the unimpeded and orderly operate in the air traffic. Newly technologies are imported to ATM according to the growing traffic flow, such as radio, radar, navigation, communication, computer technology and so on. In terms of the airlines, it is also primary to ensure the safety of passengers and take the economic efficiency (mainly the range and fuel consumption) into consideration. In the Next-Gen air traffic architecture [30], trajectories will be more flexible based on their performance rather than the navigation-based strategy. It becomes more and more difficult and important to predict trajectories precisely.

A. Assumptions

Aircraft model is defined by the main eight terms: aircraft type, mass, flight envelope, aerodynamics, engine thrust, reduced power, fuel consumption and ground movement. The performance will also be influenced by the atmospheric properties: pressure, temperature, density and speed of sound, all these factors are functions of the altitude. In our paper, we make assumptions as follows.

Assumption 1. In this paper, we consider the aircraft as a rigid body for convenience of force analysis and the curvature of the earth is not considered. The mass will affect the thrust applied by the engines especially when climbing. Here, we assume the mass of an aircraft be constant and within the limitation throughout the entire flight mission.

Assumption 2. The speed amplitude is assumed to be constant between the two report time. The ADS-B transmitter messages in a very high frequency, and we record them every 10-30 seconds by ADS-B ground station. We consider the speed to be constant during each period.

Assumption 3. The changes on temperature and pressure are ignored when an aircraft climbs and descends. The effects of earth curvature and wind are also not considered. The intention we mentioned here refers to the changes on heading, speed and altitude.

Assumption 4. Way-points are considered as a circle perpendicular to the horizontal surface, positioned in certain latitude, longitude and altitude. The radius is assumed to be equal with that of the airways (flight envelop). Airways are assumed to be regular cylinders, with standard radius (20km) and certain altitudes.

B. Problem description

The behavior of an aircraft is determined by its classical kinematics and aerodynamic models, ie. the 6-DOF equation of motion [15] based on Assumption 1.

$$
\begin{cases}
D = C_D \rho V^2 S/2 \\
L = C_L \rho V^2 S/2 \\
Y = C_Y \rho V^2 S/2 \\
l = C_l \rho V^2 Sb/2 \\
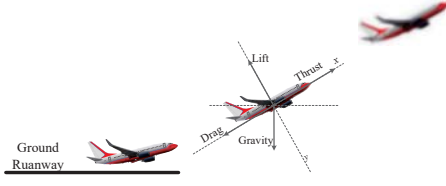M = C_m \rho V^2 S\bar{c}/2 \\
N = C_n \rho V^2 Sb/2
\end{cases}
\tag{1}
$$

Fig. 3. Illustration of the taking-off phase of an aircraft.

Fig. 4. An ADS-B system.

where, ρ is the density of air, V is the true airspeed (TAS), D, L, Y, l, M, N represent the aerodynamic drag, lift and lateral force, roll, pitch and yaw moment of the aircraft, respectively. C_D, C_L, C_Y, C_l, C_m, C_n represent the coefficients of drag, lift and lateral force, roll, pitch and yaw moment of the aircraft accordingly. S, b, $\overline{c}$ represent the wing area, wing span and the average aerodynamic chord length respectively. The performance of an aircraft can be specified into TAS, rate of climb/descent and fuel flow for conditions of climb, cruise and descent at various flight levels.

Flight trajectory is a sequence that the later state of the aircraft is closely related to the former state, $State_{k+1} \propto State_k$. So, the current flying state of an aircraft is mainly determined by the new instruction based on previous status. Then we can probably predict the next near future state of an aircraft by learning its historical trajectory within the physical performance limits.

$$\begin{cases} Pos_{k+1} = Pos_k + Vel_k \Delta t + 1/2 Acc_k (\Delta t)^2 \\ Vel_{k+1} = Vel_k + Acc_k \Delta t \end{cases} \quad (2)$$

The dynamic performance of an aircraft can be predicted through changes of these state parameters in the experimental environment. We take the climbing stage for an example, as illustrated in Fig. 3, at any time the intended movement and altitude of the aircraft in the air are determined by the aerodynamic and aerodynamic moments it receives, as well as the engine thrust and engine torque.

However, in the view of ATM, the kinematics and aerodynamic parameters are unobservable, we can hardly know the real-time attitude information (pitch, yaw and roll) and it will take a long time to estimate them with great uncertainty and inaccuracy. Fortunately, characteristics on trajectory level can be recorded and detected, such as 4-D position, speed, heading, call sign, origin-destination (OD). Aided with Airport Surveillance Radar, the Secondary Surveillance Radar (SSR) and the Automatic Dependent Surveillance-Broadcast (ADS-B), the surveillance system is able to monitor the trajectories of objects in the air. Process and measurement noises exist during the surveillance period, which aggravates the uncertainty and difficulty on trajectory predicting.

IV. METHODOLOGY

Hidden relationship between internal parameters and trajectories can be discovered by analyzing the behavior of trajectory. Also the con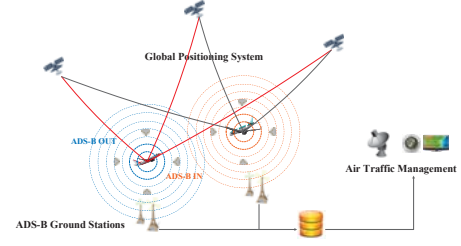textual and geographic environment provide us rich auxiliary information. In this section, we will introduce the key methods as shown in Fig. 2 and the evaluation criterion.

A. Data Collection

Automatic Dependent Surveillance-Broadcast (ADS-B) is gradually introduced to the ATM system due to its high positioning accuracy (GPS level) and high broadcast frequency. It broadcasts position, heading, call-sign, departing and landing airports three to four times per second. Messages are transmitted via ground-to-air and air-to-air data links in certain time period, as shown in Fig. 4.

Multiple ground stations were built to expand the surveillance area, which leads to conflict or redundant data. In addition, measurement and system errors can be different even among the sensors with same type. Raw data recorded by different stations need to be de-noised. The format should be re-defined to map the data sheet. Redundant data is generated in the intersecting surveillance area of sensors, when an aircraft appears in this area. Abnormal data are retained and marked.

B. Coordinate Transformation

As we know, Global Positioning System (GPS) uses the Geocentric Coordinate System, which is one of the WGS-84 coordinate (World Geodetic System-1984 Coordinate System). Elements will have different units of measurement. Intuitively, the value of altitude can be hundreds or thousands times of latitude or longitude, as they are limited to $[0, \pm 90°]$ and $[0, \pm 180°]$ respectively. The unit used in North East Down (NED) coordinate system is kilometer, which guarantee the small difference in numerical values among different dimensions, thereby decrease error in data processing.

Suppose (L, B, H) represents the 3-D position in GPS-84 coordinate system, accordingly, the position in NED (x, y, z) can be transformed by

$$\begin{cases} x = \left[N_R \left(1 - e_1^2\right) + H\right] \sin B \\ y = (N_R + H) \cos B \cos L \\ z = (N_R + H) \cos B \sin L \end{cases} \quad (3)$$

where $e_1^2 = \left(a^2 - b^2\right)/a^2$ represents the first eccentricity of ellipsoid. $N_R = a \left/ \sqrt{1 - e_1^2 \sin^2 B}\right.$, a is the semi-major axe of the earth and b is the semi-minor axe of the
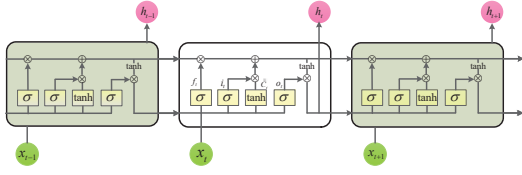
Fig. 5. Typical structure of LSTM network.

earth. Corresponding to WGS-84, $a = 1,378,137m$ and $b = 6,356,752m$.

C. LSTM network

Recurrent Neural Network (RNN) shows its own advantages on sequence prediction in recent years, however, gradients exploding or vanishing when dealing with long-term series. Essentially, LSTM is one of RNN, it has the same inputs and outputs as RNN, but more complex network structure with more parameters. Information passes optionally through the gates and knowledge is encoded up to the observed step. The gates make the cell of an LSTM network act as a memory unit that keeps the features of the input and updates the cell state according to the input of every time stamp. The structure of an LSTM network is shown in Fig. 5.

$$\begin{cases} f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right) \\ i_t = \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right) \\ \tilde{C}_t = \tanh\left(W_C\left[h_{t-1}, x_t\right] + b_C\right) \\ C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \\ o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \\ h_t = o_t * \tanh\left(C_t\right) \end{cases} \quad (4)$$

where, W_f, W_i, W_C and W_o are the coefficient matrices, b_f, b_i, b_C and b_o are the bias matrices. f_t decides what information is going to cast away from the cell state, we often call this layer as the "forget gate layer". The output of this layer can be a random value between 0 and 1. i_t and C_t decide the newly added information to the cell state, this layer is known as the "input gate layer". The output is based on the filtered cell state.

Normalization is necessary before folding the data as training set to LSTM network. Min-Max normalization is a linear strategy, it transforms x to $y = (x - \min)/(\max - \min)$, min and max are the minimum and maximum values of the observed set X. The features of data are scaled between 0 and 1.

D. Index of Performance

Euclidean distance is the most direct way to measure the similarity between two sequences. The mean absolute error (MAE), the mean relative error (MRE) and the root mean square error (RMSE) are most common indicators in many practical applications. We also use the dynamic warping time (DWT) to measure the similarity between predicted trajectory and the ground truth.

Euclidean Distance

$$\begin{cases} MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - \hat{f}_i| \\ MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{|f_i - \hat{f}_i|}{f_i} \\ RMSE = [\frac{1}{n}\sum_{i=1}^{n}(|f_i - \hat{f}_i|)^2]^{\frac{1}{2}} \end{cases} \quad (5)$$

where f_i is the observation and $\hat{f}_i$ is the prediction.

Dynamic Time Warping (DTW)

DTW is known for the successful application in speech recognition. It can be seen as an optimal match between two given sequences with certain restrictions.

Let $Tra_1 = \left(tr_1^1, tr_1^2, \cdots, tr_1^m\right)$ and $Tra_2 = \left(tr_2^1, tr_2^2, \cdots, tr_2^n\right)$ be any two trajectories with different length. A $m \times n$ distance matrix $Dis_{m \times n}$ is constructed by calculating the Euclidean distance among the points of trajectories. The element in this distance matrix can be represented by $d_{ij} = \left|tr_1^i - tr_2^j\right|$, constrained by $1 \leq i \leq m$ and $1 \leq j \leq n$. $W = (w_1, w_2, \cdots w_k)$ is referred to the warping path on the grid. The DTW distance between Tra_1, Tra_2 is calculated as:

$$d_{dtw}\left(Tra_1, Tra_2\right) = \min\left\{\frac{1}{M}\sum_{i=1}^{M} w_i\right\} \quad (6)$$

An optimal path is generated by minimizing cumulative distance by dynamic programming. Defining a cumulative matrix $R_{m \times n} = [r(i,j)]_{m \times n}$ to measure the minimum distance between two sequences. It can be calculated as:

$$r(i,j) = d_{ij} + \min\begin{cases} r(i, j-1) \\ r(i-1, j-1) \\ r(i-1, j) \end{cases} \quad (7)$$

Finally, the distance of two sequences can be denoted by the cumulative distance, namely $d_{dtw}(L_1, L_2) = r(m, n)$.

V. EXPERIMENTS

The Markov Models are widely used in nonlinear systems as heuristic methods. Based on probability transfer matrix, these models have simple form. Taking fully advantage of historical data, they can be applied to practical systems. As also popular in time series processing, we compare LSTM network and Markov Models in this part.

A. Data Description

Our data is collected by multiple ADS-B ground stations every 15 seconds. The record period lasted from June 2017 to November 2017 for five months. Each trajectory traveled for around two and half hours. Raw data may contain redundant data, mistakes, noised data and conflict features. Data compression, simplification, interpolation and abnormal detection are needed before we use it. We selected an abnormal trajectory as test data, others are input as the training set.

B. Network Training

We build our LSTM network based on Keras [31]. The dynamic states of aircraft acts as the input to our LSTM based network. Position, speed and heading are deeply coupled and mutually constrained. The ideal length of input sequences should not be too long or too short to mine these hidden constrains reasonably. Sliding windows are used in the training process, the length of window and step size can be set experimentally according to practical applications.

According to the regional regulations and agreements, the time interval between two aircrafts must be more than 10 minutes, and the vertical spacing should be over 300m. In addition, an aircraft performs differently in each phase. Sliding windows can avoid compromising the dynamic characteristics in long-term sequences. Taking the sampling period into consideration, the length of window is set to be 10, and shift by one each time. This guarantees a constant overlap with prior windows. Our LSTM-based network is structured in four layers with two hidden layers, and predicts by 10 points (time interval be 5-10 minutes). The first hidden layer is designed with 30 neurons, then feeds into the other layer with 60 neurons. Prediction results will be outputted from the fully connected layer of one neuron with a linear activation function.

Avoiding model over-fitting, dropout technique is used, with the ratio of 0.2. When compiling the model, we choose the mean squared error (mse) as the loss function, RMSprop as the optimizer.

C. Experimental Results

The trajectories recorded in the past five months are used for training the network. As the dynamic models in each phase are partially known priorly, 5% of the training data are selected as the validation set to train the hyper parameters.

For $i > 1$ and $j > 1$, the recursion in DTW is initialized as,

$$\begin{cases} r(0,0) = 0 \\ r(i,0) = +\infty \\ r(0,j) = +\infty \end{cases} \tag{8}$$

We compare our model with widely used Markov Model and weighted Markov Model used in [2] [3]. Limitations of Markov Models appear when the probabilities are close to each other, only the maximum one is selected, which results in inaccurate estimation of the stochastic process. Besides, prediction by Markov Models relies on pre-defined classification, while the range and number of this classification are determined by the practical problems and may be decided by artificial experience. While LSTM performs more adaptively than Markov Models.

The indicators are listed in Table I. Difficulty increases when we enlarge the dimensions of training data. Additional errors will be generated by the coordinate transformation and network prediction. We decompose the errors into three dimensions in order to make a clear demonstration. From the numerical analysis shown in Table I and the prediction of the trend shown in Fig. 6a, we can see that, LSTM network can predict a smoother trajectory than that in Markov Models. One of the possible reasons is that LSTM keeps long-term features of the sequence. Sharp and serrated trajectories are generated by the Markov Models. The weighted Markov Model performs better than the unweighted one. Since the adjacent states of weighted MM are assigned with large weights, which weaken the error between prediction and ground truth. In Fig. 6b, all these models encounter large errors when considering the altitude dimension. As the normalization and classification in LSTM and Markov Models will be affected by the large ranged amplitude. Fig. 7 shows detailed errors along the three directions. Random maneuvering and bad weather create fluctuating or sharp turns, lacking of such training data makes us unable to capture the dynamic features. So spikes exist in the output, but our model converges quickly.

We apply our LSTM model to a practical scene. An aircraft takes off after taxiing on the runaway as usual. However, it climbs with an abnormal heading but normal climb rate. Along with its heading direction, a mountain exists whose altitude is slightly higher than that of the aircraft. The aircraft will hit the mountain if it continues to follow the normal operation. Fortunately, ATM gave an urgent order timely, which successfully avoided the accident. Our model can tell the potential risk with predicted trajectory and environmental factors.

VI. CONCLUSION

We formulated a trajectory prediction method based on an LSTM network. This method was tested on the flight trajectory recorded by ADS-B ground stations. Different from conventional model-based methods used for flight trajectory prediction, our method avoids the complicated coefficients estimation process. Applying sliding windows makes the LSTM network able to track every phase of the trajectory and converge quickly. Moreover, our method can accurately predict flight trajectory in both 3-D and 4-D space. Experiments showed that our method outperforms the widely used methods, such as MM and wMM.

This research is our initial step of applying LSTM for flight trajectory prediction. In future, multi-modal data including images, audios and videos will be considered. Our LSTM-based trajectory prediction model will be modified to fuse different modality data.

REFERENCES

[1] P. Note and E. Note, "European organisation for the safety of air navigation," *GEN*, vol. 4, p. E2.

[2] C. Yuan, D. Li, and Y. Xi, "Campus trajectory forecast based on human activity cycle and markov method," in *Cyber Technology in Automation, Control, and Intelligent*

Methods	MAE			MRE			RMSE			DTW	
	lat*	lon*	hgt*	lat	lon	hgt	lat	lon	hgt	3-D	4-D
LSTM	0.0725	0.0552	77.9472	0.0029	5.09e-04	0.0197	0.2295	0.1337	134.512	0.0216	26.3263
wMM**	0.0788	0.0910	141.3293	0.0032	8.31e-04	0.0339	0.1058	0.1481	193.4384	0.0690	70.6659
MM**	0.1065	0.1045	159.9611	0.0043	9.56e-04	0.0396	0.1376	0.1633	194.8655	0.0992	104.8770

* lat, lon, hgt stand for $latitude, longitude$ and $altitude$, respectively.
** wMM indicates weighted Markov Model, while MM means Markov Model.
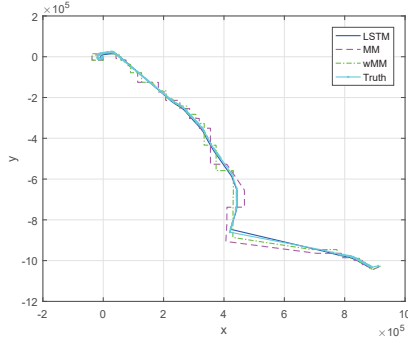
(a) 3-D trajectory prediction
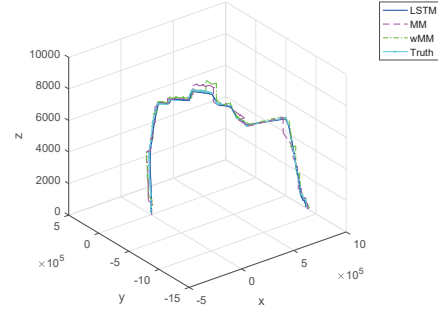
(b) 4-D trajectory prediction

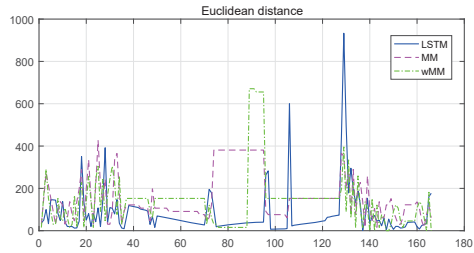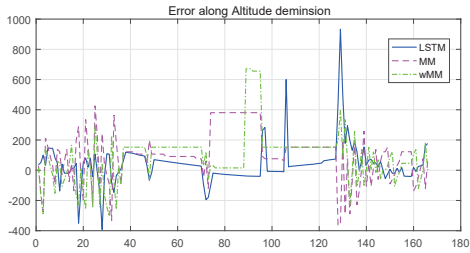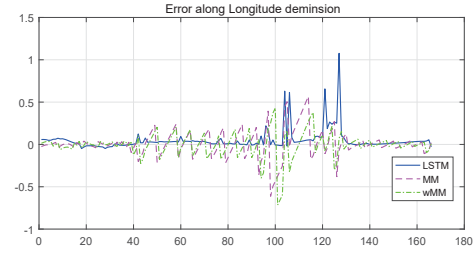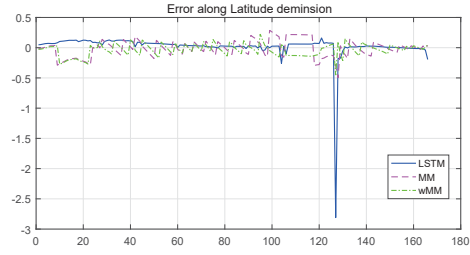Fig. 6. Predictions of trajectory by LSTM, Markov Model and weighted Markov Model.

Fig. 7. Euclidean errors of the three models.

Systems (CYBER), 2015 IEEE International Conference on, pp. 941–946, IEEE, 2015.

[3] B. Wang, Y. Hu, G. Shou, and Z. Guo, "Trajectory prediction in campus based on markov chains," in *International Conference on Big Data Computing and Communications*, pp. 145–154, Springer, 2016.

[4] M. A. Awad and I. Khalil, "Prediction of user's web-browsing behavior: Application of markov model," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 4, pp. 1131–1142, 2012.

[5] A. Abadi, T. Rajabioun, and P. A. Ioannou, "Traffic flow prediction for road transportation networks with

limited traffic data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 653–662, 2015.

[6] H. Chang, Y. Lee, B. Yoon, and S. Baek, "Dynamic near-term traffic flow prediction: system-oriented approach based on past experiences," *IET Intelligent Transport Systems*, vol. 6, no. 3, pp. 292–305, 2012.

[7] Z. Xiao, L. Ponnambalam, X. Fu, and W. Zhang, "Maritime traffic probabilistic forecasting based on vessels' waterway patterns and motion behaviors," *IEEE Transactions on Intelligent Transportation Systems*, 2017.

[8] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: deep belief networks with multitask learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2191–2201, 2014.

[9] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Soft+ hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection," *arXiv preprint arXiv:1702.05552*, 2017.

[10] Z. Zhao, W. Chen, X. Wu, P. C. Chen, and J. Liu, "Lstm network: a deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.

[11] N. Xu, G. Donohue, K. B. Laskey, and C.-H. Chen, "Estimation of delay propagation in the national aviation system using bayesian networks," in *6th USA/Europe Air Traffic Management Research and Development Seminar*, 2005.

[12] J. J. Rebollo and H. Balakrishnan, "Characterization and prediction of air traffic delays," *Transportation research part C: Emerging technologies*, vol. 44, pp. 231–241, 2014.

[13] Y. J. Kim, S. Choi, S. Briceno, and D. Mavris, "A deep learning approach to flight delay prediction," in *Digital Avionics Systems Conference (DASC), 2016 IEEE/AIAA 35th*, pp. 1–6, IEEE, 2016.

[14] R. Alligier, D. Gianazza, and N. Durand, "Machine learning and mass estimation methods for ground-based aircraft climb prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3138–3149, 2015.

[15] J. Roskam, *Airplane flight dynamics and automatic flight controls*. DARcorporation, 1998.

[16] J. Guo, W. Huang, and B. M. Williams, "Adaptive kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 50–64, 2014.

[17] G. A. Davis and N. L. Nihan, "Nonparametric regression and short-term freeway traffic forecasting," *Journal of Transportation Engineering*, vol. 117, no. 2, pp. 178–188, 1991.

[18] M. S. Ahmed and A. R. Cook, *Analysis of freeway traffic time-series data by using Box-Jenkins techniques*. No. 722, 1979.

[19] M. Levin and Y.-D. Tsao, "On forecasting freeway occupancies and volumes (abridgment)," *Transportation Research Record*, no. 773, 1980.

[20] P. Lingras, S. Sharma, and M. Zhong, "Prediction of recreational travel using genetically designed regression and time-delay neural network models," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1805, pp. 16–24, 2002.

[21] S. H.-A. Hashmienejad and S. M. H. Hasheminejad, "Traffic accident severity prediction using a novel multi-objective genetic algorithm," *International Journal of Crashworthiness*, pp. 1–16, 2017.

[22] J. Tang, F. Liu, Y. Zou, W. Zhang, and Y. Wang, "An improved fuzzy neural network for traffic speed prediction considering periodic characteristic," *IEEE Transactions on Intelligent Transportation Systems*, 2017.

[23] H. Tan, Y. Wu, B. Shen, P. J. Jin, and B. Ran, "Short-term traffic prediction based on dynamic tensor completion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 8, pp. 2123–2133, 2016.

[24] S. Gan, S. Liang, K. Li, J. Deng, and T. Cheng, "Trajectory length prediction for intelligent traffic signaling: A data-driven approach," *IEEE Transactions on Intelligent Transportation Systems*, 2017.

[25] S. Hexeberg, A. L. Flåten, E. F. Brekke, *et al.*, "Ais-based vessel trajectory prediction," in *Information Fusion (Fusion), 2017 20th International Conference on*, pp. 1–8, IEEE, 2017.

[26] Y. Tu, M. O. Ball, and W. S. Jank, "Estimating flight departure delay distributionsła statistical approach with long-term trend and short-term pattern," *Journal of the American Statistical Association*, vol. 103, no. 481, pp. 112–125, 2008.

[27] M. Gariel, A. N. Srivastava, and E. Feron, "Trajectory clustering and an application to airspace monitoring," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1511–1524, 2011.

[28] M.-H. Nguyen and S. Alam, "Airspace collision risk hotspot identification using clustering models," *IEEE Transactions on Intelligent Transportation Systems*, 2017.

[29] X. Tang, J. Gu, Z. Shen, and P. Chen, "A flight profile clustering method combining twed with k-means algorithm for 4d trajectory prediction," in *Integrated Communication, Navigation, and Surveillance Conference (ICNS), 2015*, pp. S3–1, IEEE, 2015.

[30] "The future of the nas," *Federal Aviation Admin*, 2016.

[31] F. Chollet, "Keras (2015)," *URL http://keras. io*, 2017.