Probabilistic Aircraft Trajectory Prediction Considering Weather Uncertainties Using Dropout As Bayesian Approximate Variational Inference

Yutian Pang, Yongming Liu
Arizona State University, Tempe, Arizona, 85287, USA

AIAA SciTech Forum 2020
Orlando, Florida

Outline

1. Introduction
2. Preliminaries
3. Data Source
4. Network Architecture
5. Experimental Results
6. Conclusion

Introduction – Motivations

- Weather-related safety concerns and huge uncertainties exist during daily aviation operations.
  - Unforeseeable: The convective weather conditions can develop rapidly [1].
  - Frequent: 30% ~ 40% of the delayed flights are caused by weather related reasons [2].
- The workload of controllers can increase significantly during certain extreme climatic events.
- The development of flight trajectory planning tools help reduce these concerns.
- NextGen requires an accurate and automatic flight trajectory planning tool to update the flight route.
- To achieve this, researchers focus on,
  - System-level trajectory planning and dynamic weather reroutes.
  - Deterministic prediction and probabilistic prediction.

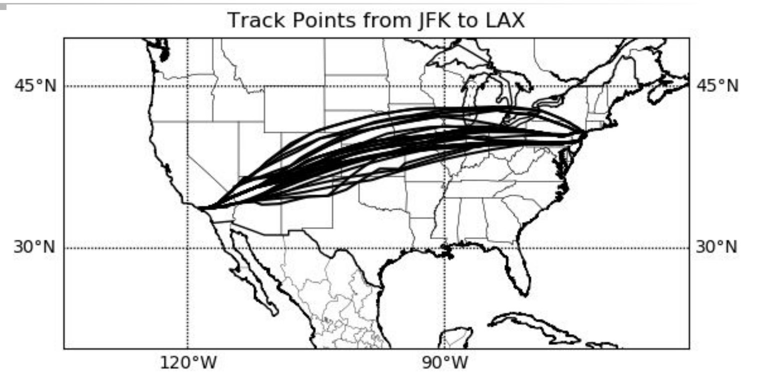


Fig.1 Flight Tracks from JFK to LAX on Dec 19, 2018

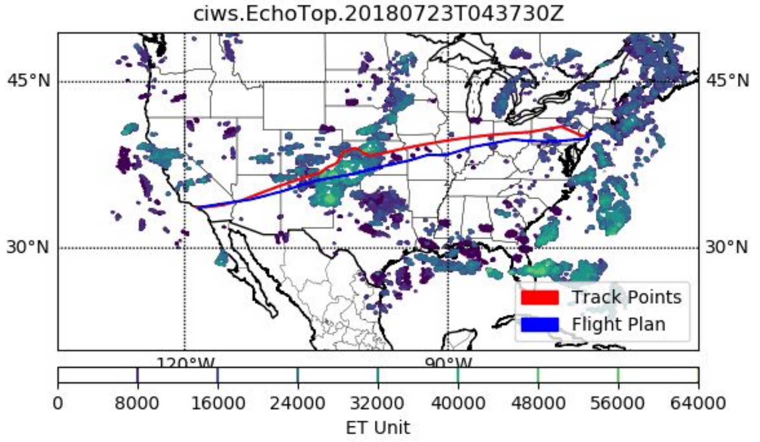


Fig.2 Convective Weather Conditions during the Flight Period from 04:45:00 to 11:47:30 on July 23, 2018

[1] Erzberger, H., Lauderdale, T., and Chu, Y., "Automated conflict resolution, arrival management, and weather avoidance for air traffic management," Journal of aerospace engineering, Vol.226, No. 8, 2012, pp. 930–949.

[2] Belcastro, L., Marozzo, F., Talia, D., & Trunfio, P. (2016). Using scalable data mining for predicting flight delays. ACM Transactions on Intelligent Systems and Technology (TIST), 8(1), 5.

Introduction – Previous Work

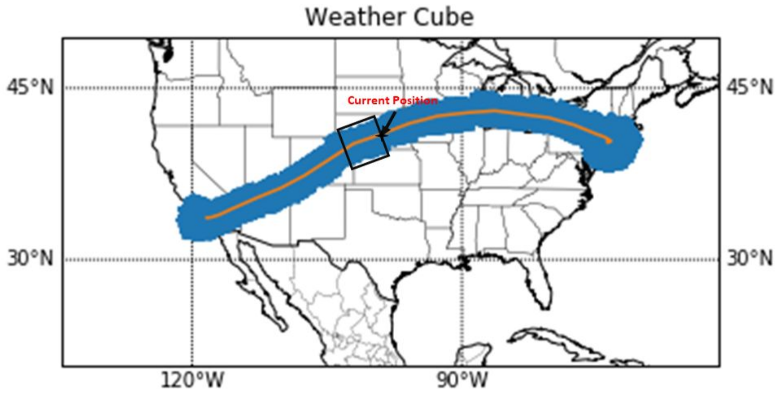


Fig.3 Weather cube generating algorithm

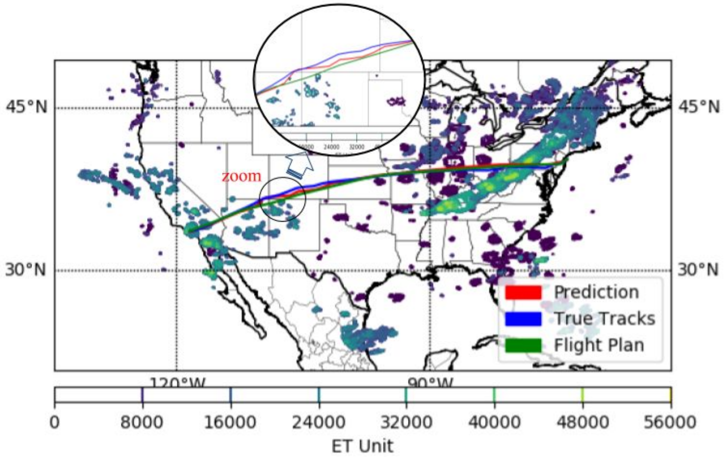


Fig.4 A prediction result

- Previous work [3] use recurrent neural network to predict the flight trajectory from JFK to LAX.
  - Extract weather data along each track point.
  - Modify the recurrence of RNN cell to incorporate weather features into the model.
- Issues,
  - A long prediction (6 hours flight) is untrustworthy.
  - Needs better prediction accuracy.
  - No prediction uncertainties.

[3] Pang, Y., Yao, H., Hu, J., & Liu, Y. (2019). A Recurrent Neural Network Approach for Aircraft Trajectory Prediction with Weather Features From Sherlock. In AIAA Aviation 2019 Forum (p. 3413).

Introduction – Solutions

- To accommodate these, we
  - perform sector-specific prediction rather than prediction across multiple control sectors.
  - incorporate Bayesian framework to introduce uncertainties into the prediction.

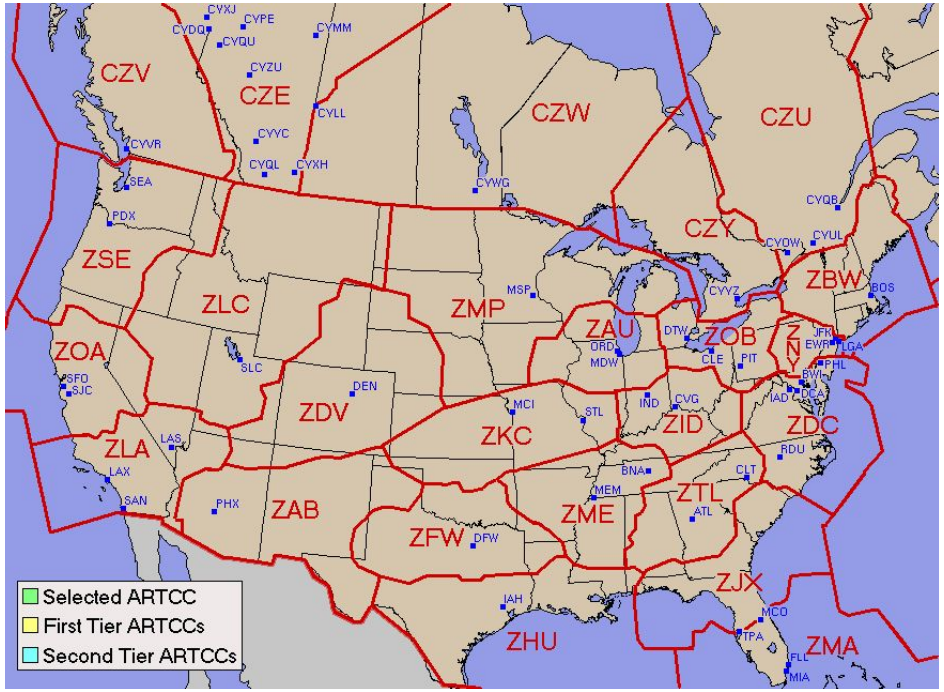
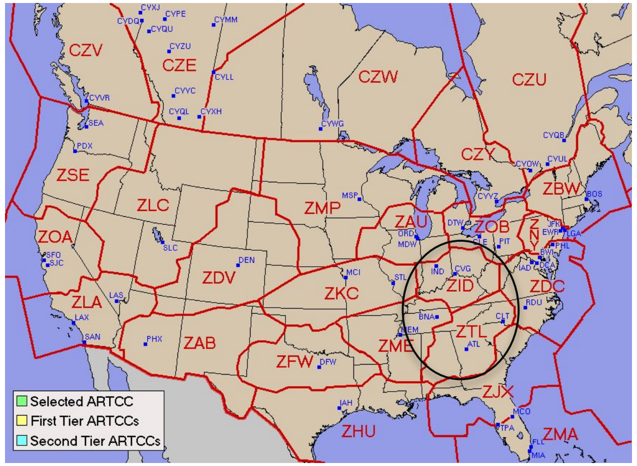


Fig.5 United States Air Route Traffic Control Center (ARTCC)
Source: https://www.fly.faa.gov/ois/tier/themap.html

Introduction – Case Study

- Start from data analysis.
  - It's reported to have severe weather condition on 06/24/2019* within the United States airspace.
  - Tornado and high wind are reported at numerous locations in the east.
- Sector ZID (Indianapolis) and ZTL (Atlanta) roughly cover the area of interest.



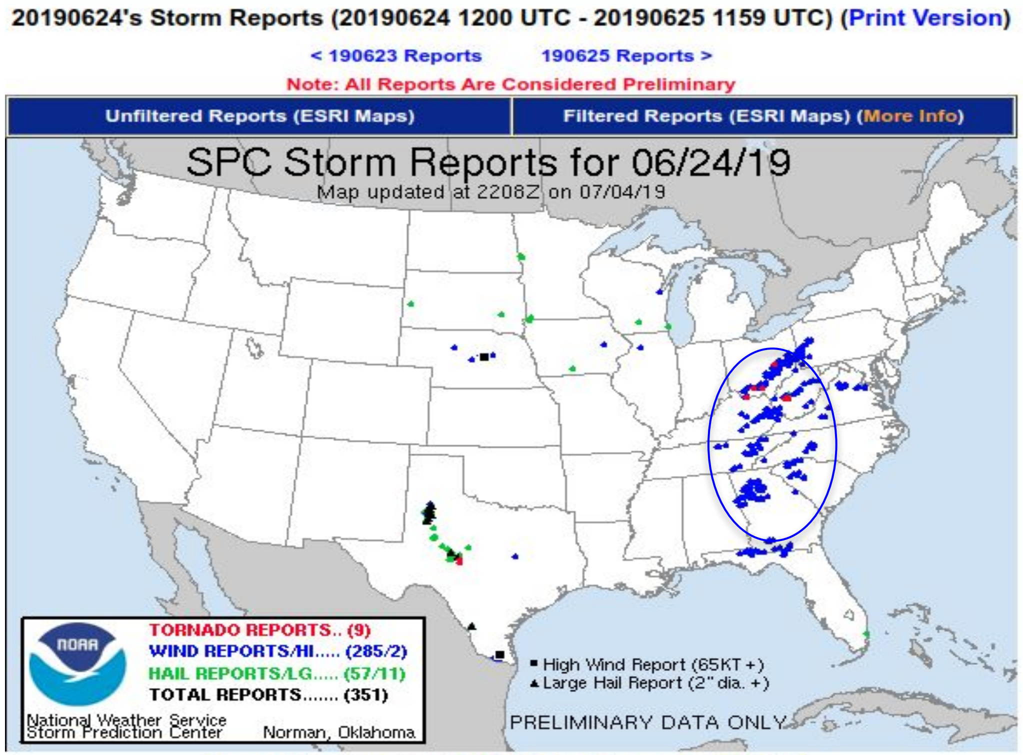


Fig.6 SPC Storm Reports for 06/24/19

*NOAA Storm Prediction Center: *https://www.spc.noaa.gov/exper/archive/events/*

Preliminaries – Variational Inference

- Bayesian theorem,
$$p(\omega|X,Y) = \frac{p(Y|X,\omega)p(\omega)}{p(Y|X)}$$
- Bayesian inference,
$$p(y^*|x^*,X,Y) = \int p(y^*|x^*,\omega)p(\omega|X,Y)d\omega$$
- Approximating variational posterior,
$$KL(q_\theta(\omega)||p(\omega|X,Y)) = \int q_\theta(\omega)log\frac{q_\theta(\omega)}{p(\omega|X,Y)}d\omega$$
- Minimizing KL divergence is equivalent to maximizing the evidence lower bound (ELBO) w.r.t. $q_\theta(\omega)$
$$\mathcal{L}_{VI}(\theta) := \int q_\theta(\omega)\,log(p(Y|X,\omega))\,d\omega - KL(q_\theta(\omega)||p(\omega))$$

Notes,
X – training inputs
Y – training labels
x* – testing inputs
y* – testing labels
ω – model parameters

Preliminaries – Bayesian Neural Nets

- The gradient descent of BNN can be achieved by minimizing the Negative ELBO,

$$\mathcal{L}_{VI}(\theta) := KL(q_\theta(\omega)||p(\omega)) - \mathbb{E}_{q_\theta(\omega)}[log(X,Y|\omega)]$$

- Bayes By Backprop uses the chain rule to calculate the gradient,

$$\frac{\partial}{\partial\theta}\mathbb{E}_{q_\theta(\omega)}[f(\omega,\theta)] = \mathbb{E}_{q(\epsilon)}[\frac{\partial f(\omega,\theta)}{\partial\omega}\frac{\partial\omega}{\partial\theta} + \frac{\partial f(\omega,\theta)}{\partial\theta}]$$

- A neural network with arbitrary depth and non-linearities, with Bernoulli dropout applied before every weight layer, is mathematically equivalent to an approximate variational inference which results in uncertainty estimates [4].

[4] Gal, Y., and Ghahramani, Z., "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," international conference on machine learning, 2016, pp. 1050–1059.

Preliminaries – Bayesian Convolutional Neural Nets

- To perform approximate VI to convolutional layers,
  - Place a prior distribution over each kernel and approximately integrate them with Bernoulli variational distributions.
  - Sample the Bernoulli random variables and multiply with the weight matrix.
- This approximating distribution is also equivalent to applying the dropout after every convolution layer as well as inner-product layers [5].

- <u>The implementation of the Bayesian CNN is as simple as using dropout after every convolution layer before pooling [6].</u>

[5] Gal, Y., "Uncertainty in deep learning," Ph.D. thesis, PhD thesis, University of Cambridge, 2016.

[6] Gal, Y., and Ghahramani, Z., "Bayesian convolutional neural networks with Bernoulli approximate variational inference," arXiv:1506.02158, 2015.

Preliminaries – Bayesian Recurrent Neural Nets

- Variational dropout,

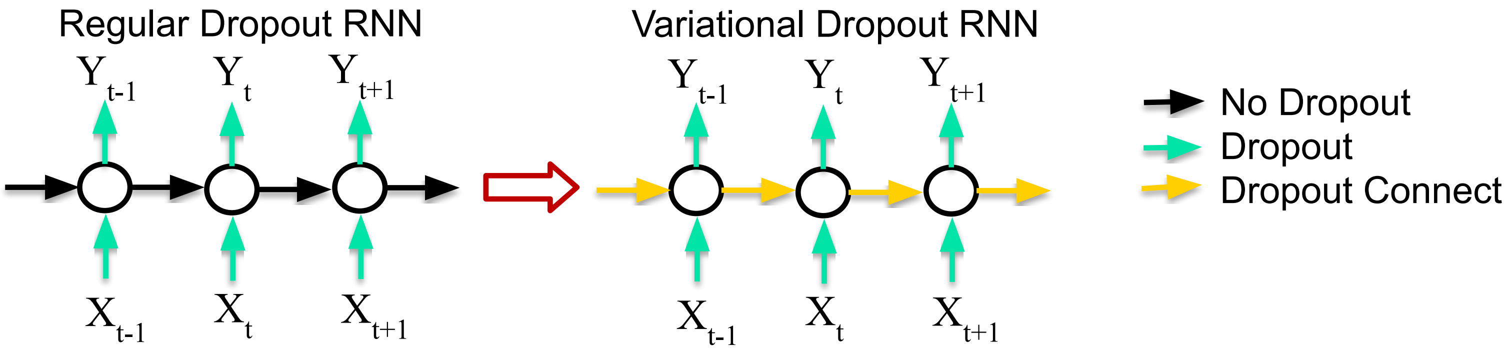


Fig.7 Variational Dropout

- Variational dropout can drop the connection between each state of RNN folds by randomly mask each row of the approximating distribution to zero as,

$$q(\omega_k) = pN(\omega_k; 0, \sigma^2 I) + (1-p)N(\omega_k; m_k, \sigma^2 I)$$

  - where k is the row index, p is the dropout probability and σ^2 is the small noise.
- We optimize over m, the variational parameters of the random weight matrices, or RNN weight matrices in standard view.

Preliminaries – Bayesian Recurrent Neural Nets

- Implementing the approximate inference to RNN is identical to implementing dropout in RNNs with the same network units dropped at each time step, randomly dropping inputs, outputs, and recurrent connections [7].

- <u>The final Bayesian framework is simply applying regular dropout after each convolutional layers and variational dropout to recurrent cells.</u>

- Uncertainty estimates: MC dropout

$$p(y^*|x^*, X, Y) \approx \int p(y^*|x^*, \omega)q(\omega)d\omega \approx \frac{1}{K}\sum_{k=1}^{K} p(y^*|x^*, \hat{\omega_k})$$

[7] Gal, Y., and Ghahramani, Z., “A theoretically grounded application of dropout in recurrent neural networks,” Advances in neural information processing systems, 2016, pp. 1019–1027.

Data Source – Database

- Sherlock Data Warehouse (SDW)
  - SDW [8] is a big data system for data visualization to support air traffic management (ATM) research, which includes a database, a web-based graphical user interface (GUI) and other services.
  - Data of SDW primarily comes from the FAA and the National Oceanic Atmospheric Administration (NOAA).
  - We use sector Integrate Flight Format (IFF) data for flight plans and real tracks, and Convective Integrated Weather Service (CIWS) for EchoTop (ET) weather data.
  - ZTL on June 24th, 2019.

[8] Arneson, H. M., Hegde, P., La Scola, M. E., Evans, A. D., Keller, R. M., & Schade, J. E. (2019). Sherlock Data Warehouse.

Data Source – Flight Data

- Sector IFF data for each day is around 200 MB with over 10 million rows.
- Flight Records includes flight tracks, flight plan and flight information.
  - Flight tracks includes spatial and temporal coordinates.
  - Flight Plan comes as a string of waypoints. We create a web-based mining tool to parse it into WGS84 coordinates using online database (*opennav.com*).

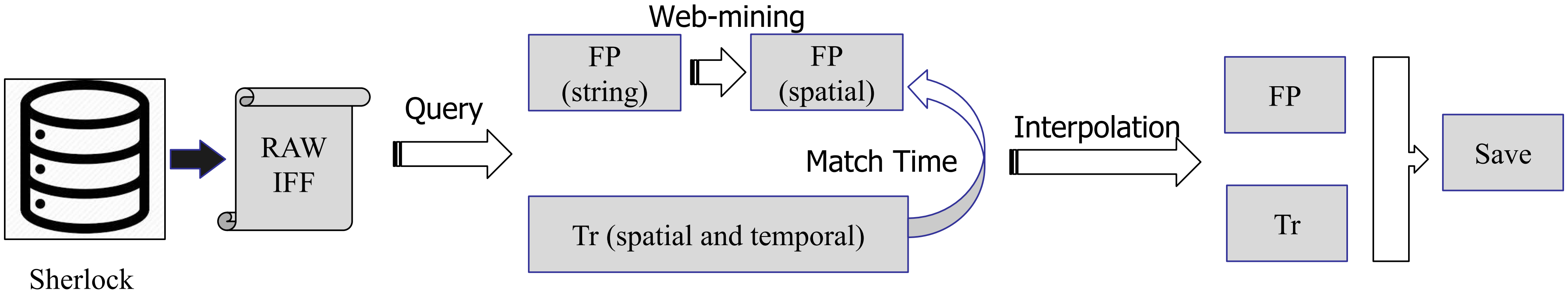


Fig.8 Flowchart: Raw IFF data processing

Data Source – Flight Data

- The processing of sector IFF data follows a similar procedure as the previous work and can be conclude as,
  - <u>Query → Match → Interpolate → Equalize</u>
- The dimension of the processed flight data is 6255×50×2.

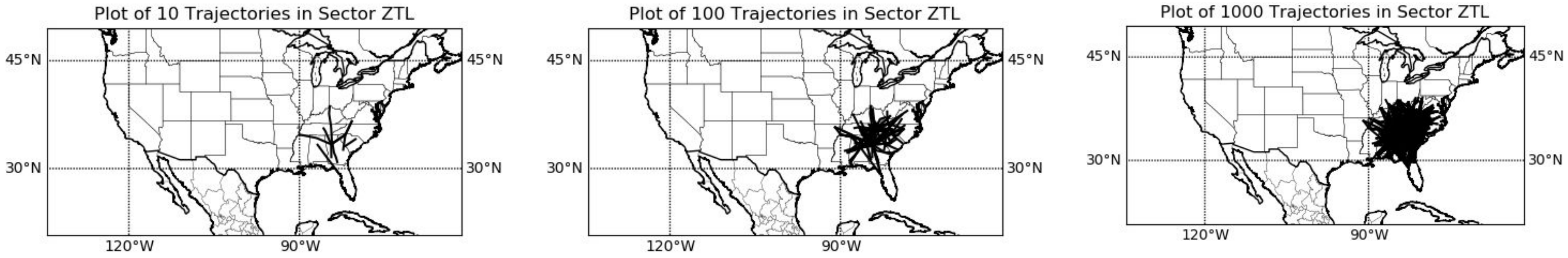


Fig.9 Flights in sector ZTL on 06/24/2019

Data Source – Weather Data

- The two key features of CIWS are EchoTop (ET) and Vertically Integrated Liquid (VIL), both come with current and forecast dataset in Sherlock. We only use current ET for demonstration.
- At each track point, we take out the weather tensor of size 32×32×1 except for the starting point. The orientation of the weather tensor is rotated with the aircraft heading angle.
- The processed weather tensor has size 6255×49×32×32×1.

Table 1 EchoTop Key Features

Parameters	Current	Forecast
Dimension	1x1x3520x5120	24x1x3520x5120
Range Latitude / °	[19.36, 48.90]	
Range Longitude / °	[-134.35, -61.65]	
Update Frequency / s	150	300

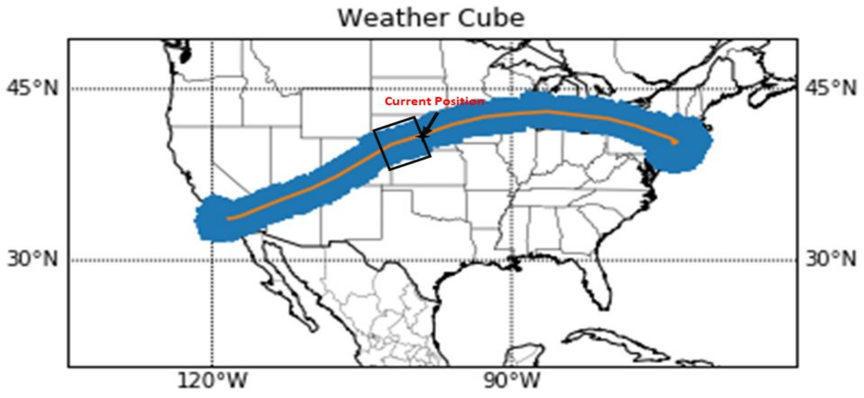


Fig.10 Weather cube generating algorithm

Network Architecture – Details

Table 2 Model Layer Setup

Layers	Input Size	Output Size	Dimension	Others
Conv3d_1	[N, 49, 32, 32, 1]	[N, 49, 10, 10, 3]	[1, 5, 5, 3]	Strides: [1, 3, 3, 1], No Padding
Conv3d_2	[N, 49, 10, 10, 3]	[N, 49, 4, 4, 1]	[1, 3, 3, 1]	Strides: [1, 3, 3, 1], Zero Padding
Flatten	[N, 49, 4, 4, 1]	[N, 49, 16]		
Dense_1	[N, 49, 16]	[N, 49, 4]	4	
Dense_2	[N, 49, 4]	[N, 49, 1]	1	
Concat	[N, 49, 1]	[N, 49, 3]		Concatenate with Flight Plan *p*
LSTM	[N, 49, 3]	[N, 49, 128]	128	Variational Dropout
Dense_3	[N, 49, 128]	[N, 49, 64]	64	
Dense_4	[N, 49, 64]	[N, 49, 32]	32	
Dense_5	[N, 49, 32]	[N, 49, 2]	2	No Dropout

Network Architecture – Details

- Dropout,
  - Regular dropout is applied after each dense layer.
  - Variational dropout is applied to LSTM cell.
  - Dropout ratio is 0.5 for all layers.

- Model,
  - Zero initialization is used to LSTM cell.
  - Root mean squared error.
  - Training with momentum optimizer with value of 0.9.
  - Weight decay is set to the learning rate.

Experimental Results

- We perform *K=100* MC dropout tests and evaluate the mean and variance of the prediction.
- Mean and variance are the key parameters to determine the confidence ellipses at each prediction point.

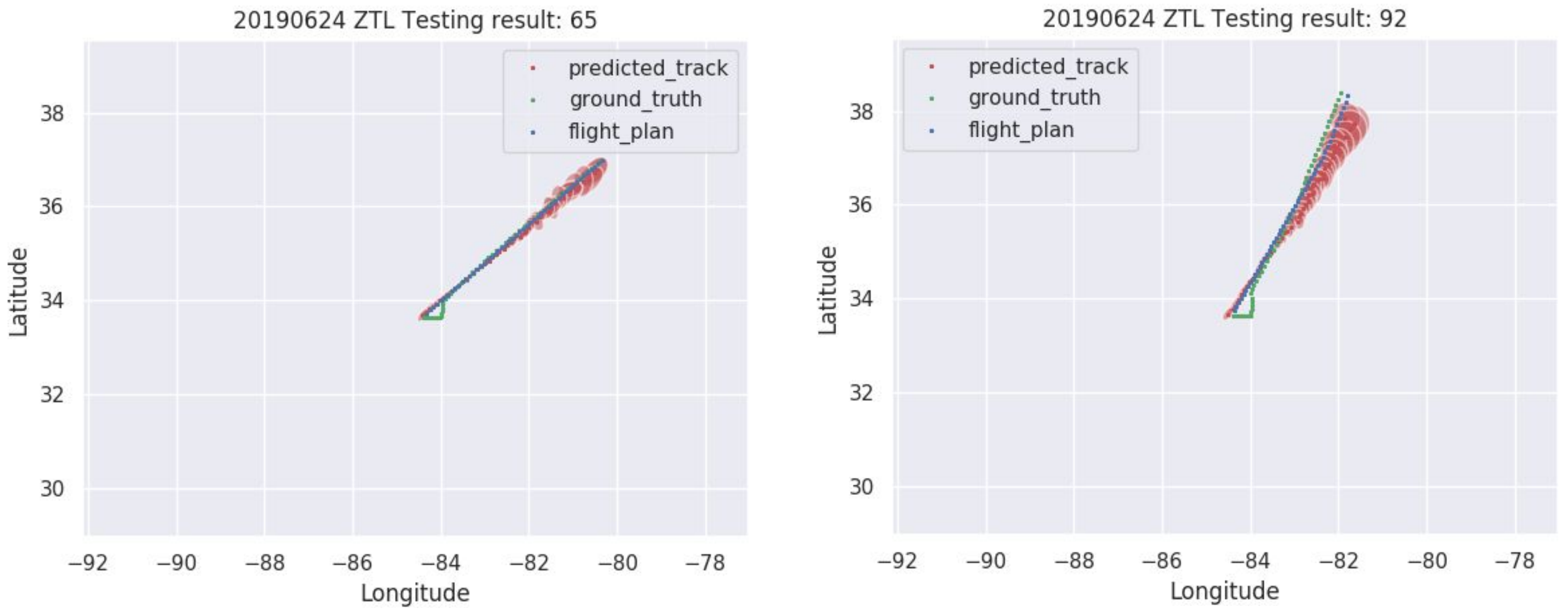


Figure 11. Model Prediction

[9] Wang, Y., Pang, Y., & Liu, Y. (2020). A Voice-Communication Augmented Simulation Framework for Aircraft Trajectory Simulation. In AIAA SciTech 2020 Forum.

Variance Reduction

- ❖ To numerically evaluate the mean model performance, we define the following quantities,

$$L2_k^{ori} = \sum_i^n \sum_j^d (Y_{k,i,j}^{true} - Y_{k,i,j}^{fp})^2$$

$$L2_k^{new} = \sum_i^n \sum_j^d (Y_{k,i,j}^{true} - Y_{k,i,j}^{pred})^2$$

$$reduction = \frac{Var(L2_k^{ori}) - Var(L2_k^{new})}{Var(L2_k^{ori})}$$

- ❖ Dropout is good at deviation variance reduction but less competitive in prediction accuracy.

Table 3 Comparison of Result

Models	Percentage of Flights Reduced	Overall Variance Reduction
Conv-LSTM	47.0%	12.3%
Dropout as Bayesian	26.2%	16.8%

- ❖ Generative Model.

Conclusion

- The formulation of Bayesian Neural Network is as simple as applying regular dropout after each fully-connected layer and performing variational dropout to RNN cells.
- The experimental result shows that the model can predict the aircraft trajectory as well as outputting an uncertainty bound when performing MC dropout during testing.

Future Work

- Exact Variational inference for Bayesian NNs.
- Knowledge discovery of the current dataset.

Acknowledgements

The work related to this research is supported by funds from NASA University Leadership Initiative program (Contract No. NNX17AJ86A, Project Officer: Dr. Anupa Bajwa, Principal Investigator: Dr. Yongming Liu). The support is gratefully acknowledged. We also would like to thank Dr. Hao Yan from Arizona State University for helpful suggestions on the formulation of the model.

References

[1] Erzberger, H., Lauderdale, T., and Chu, Y., "Automated conflict resolution, arrival management, and weather avoidance for air traffic management," Proceedings of the Institution of Mechanical Engineers, Part G: Journal of aerospace engineering, Vol.226, No. 8, 2012, pp. 930–949.

[2] Belcastro, L., Marozzo, F., Talia, D., & Trunfio, P. (2016). Using scalable data mining for predicting flight delays. ACM Transactions on Intelligent Systems and Technology (TIST), 8(1), 5.

[3] Pang, Y., Yao, H., Hu, J., & Liu, Y. (2019). A Recurrent Neural Network Approach for Aircraft Trajectory Prediction with Weather Features From Sherlock. In AIAA Aviation 2019 Forum (p. 3413).

[4] Gal, Y., and Ghahramani, Z., "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," international conference on machine learning, 2016, pp. 1050–1059.

[5] Gal, Y., "Uncertainty in deep learning," Ph.D. thesis, PhD thesis, University of Cambridge, 2016.

[6] Gal, Y., and Ghahramani, Z., "Bayesian convolutional neural networks with Bernoulli approximate variational inference," arXiv preprint arXiv:1506.02158, 2015.

[7] Gal, Y., and Ghahramani, Z., "A theoretically grounded application of dropout in recurrent neural networks," Advances in neural information processing systems, 2016, pp. 1019–1027.

[8] Arneson, H. M., Hegde, P., La Scola, M. E., Evans, A. D., Keller, R. M., & Schade, J. E. (2019). Sherlock Data Warehouse.

[9] Wang, Y., Pang, Y., & Liu, Y. (2020). A Voice-Communication Augmented Simulation Framework for Aircraft Trajectory Simulation. In AIAA SciTech 2020 Forum.

Questions?

Thanks.