Online LSTM Framework for Hurricane Trajectory Prediction

Ding Wang [1] **Pang-Ning Tan** [1]

Abstract

Hurricanes are high-intensity tropical cyclones that can cause severe damages when the storms make landfall. Accurate long-range prediction of hurricane trajectories is an important but challenging problem due to the complex interactions between the ocean and atmosphere systems. In this paper, we present a deep learning framework for hurricane trajectory forecasting by leveraging the outputs from an ensemble of dynamical (physical) models. The proposed framework employs a temporal decay memory unit for imputing missing values in the ensemble member outputs, coupled with an LSTM architecture for dynamic path prediction. The framework is extended to an online learning setting to capture concept drift present in the data. Empirical results suggest that the proposed framework significantly outperforms various baselines including the official forecasts from U.S. National Hurricane Center (NHC).

1. Introduction

Our climate system is now changing more rapidly than ever in the past due to increasing human activities. Recent reports have suggested that the warming of sea surface due to anthropogenic influences will likely trigger more intense and destructive hurricanes (Knutson et al., 2021). Hurricanes are rapidly rotating tropical storm systems with a maximum sustained wind speed of at least 64 knots (119 km/hr). Due to their potentially severe impact, accurate long-range forecasting of the hurricane tracks is essential to provide enough time for emergency management and response teams to issue warnings and organize evacuation efforts. However, long-range hurricane track forecasting is a challenging problem due to the complex, non-linear interactions among various factors in our atmospheric system.

In recent years, there have been growing interests in applying deep learning methods to hurricane prediction tasks (Lee & Liu, 2000; Moradi Kordmahalleh et al., 2016; Alemany et al., 2019; Cox et al., 2018; Gao et al., 2018). However, there are several limitations to these approaches. Many of the methods, particularly those based on recurrent neural network (RNN) and its variants, are mostly designed for short-range forecasts (24 hour or less) only. These methods often utilize only historical data, which are insufficient to capture the current and future environmental conditions that affect the hurricane trajectory path. While some recent studies have applied deep learning techniques, such as generalized adversarial networks (GAN) (Rüttgers et al., 2019) and convolutional LSTM (ConvLSTM) (Mudigonda et al., 2017; Kim et al., 2019), to meteorological data, these models are often trained on coarse-resolution data (e.g., $0.5^\circ \times 0.5^\circ$), and thus, their forecast errors can still be relatively large. Extending these methods to longer lead time prediction is also a challenge due to the inherent error propagation problem (Cheng et al., 2006). Finally, previous works are mostly designed for batch learning algorithms, which are not ideal given the non-stationary nature of the domain.

To address these issues, we propose a novel online LSTM-based framework for long-range hurricane trajectory forecasting. Instead of using historical data, the framework utilizes the outputs generated from a multi-model ensemble of dynamical (physical) models to generate its predictions. The advantage of using a multi-model ensemble to generate the forecasts is that the ensemble members would simulate future state of the atmospheric system based on current environmental conditions. Since not all models generate their forecasts at every time step, this leads to considerable amount of missing values in the ensemble forecast data, a challenge that must be addressed by the deep learning framework. The proposed architecture consists of two stages. The first stage consists of a set of LSTM based layers called model performance layers to learn the performance of the individual ensemble members. The second stage is the prediction layer, which uses the output from the previous stage to generate the final multi-lead time predictions. The proposed framework allows us to learn the nonlinear relationships among the ensemble member forecasts as well as the temporal autocorrelations of the predictions. It also alleviates the missing value problem using a Temporal Decay Memory (TDM) with a masked softmax function for

[1]Department of Computer Science and Engineering, Michigan State University, Michigan, USA. Correspondence to: Ding Wang <wangdin1@msu.edu>.

Tackling Climate Change with Machine Learning Workshop at ICML 2021.

weighting the ensemble member forecasts.

The proposed framework can be trained either in a batch or an online learning setting. The batch learning implementation is known as DTP (**D**eep **T**rajectory **P**rediction). To enable the framework to capture concept drift present in the data, we extended its formulation to ODTP (**O**nline **D**eep **T**rajectory **P**rediction), which is an online learning implementation to handle the non-stationary nature of the domain. The proposed frameworks were applied to real-world data to predict future trajectory paths of hurricanes up to 48 hours lead time. Experimental results showed that ODTP can achieve better performance than DTP, and generally outperforms other baseline approaches, including the official forecasts from the U.S. National Hurricane Center (NHC).

2. Preliminaries

Consider a set of hurricanes, $\{h_1, h_2, \ldots, h_C\}$, ordered by their start times. Assume there are n_i data points (time steps) associated with hurricane h_i and $N = \sum_{i=1}^{C} n_i$. Let T be the forecast horizon, i.e., maximum lead-time, and M be the number of ensemble members. Let $\mathcal{X} \in \mathbb{R}^{2\times M\times T\times N}$ be the set of trajectory forecasts generated by the ensemble of dynamical (physical) models, where each $\mathcal{X}^t \in \mathbb{R}^{2\times M\times T}$ corresponds to the hurricane trajectory forecasts generated at time step t. Let $\tilde{n}_i$ be the cumulative number of time steps from hurricane h_1 to h_i, i.e. $\tilde{n}_i = \sum_{j=1}^{i} n_j$. Thus, $\{\mathcal{X}^j \mid \tilde{n}_{i-1} < j \leq \tilde{n}_i\}$ denote the set of ensemble forecast data associated with hurricane h_i.

Let $\mathcal{Y} \in \mathbb{R}^{2\times T\times N}$ be the ground truth locations for $\mathcal{X}$, where $\mathbf{Y}^t \in \mathbb{R}^{2\times T}$ is the corresponding ground truth locations, from $t+1$ to $t+T$, for the forecasts generated at time step t for the T lead times. At each time step t, our goal is to learn a function f that maps the ensemble member forecasts $\mathcal{X}^t$ to the multi-lead time forecasts $\mathbf{Y}^t$ with minimal error. The trajectory forecasts of the ensemble members for lead time τ at time t is denoted as $\mathbf{X}^{t,\tau} \in \mathbb{R}^{2\times M}$, with the corresponding ground truth location $\mathbf{y}^{t,\tau} \in \mathbb{R}^2$. Suppose $\mathcal{E} \in \mathbb{R}^{M\times T\times N}$ is the distance errors corresponding to trajectory forecasts $\mathcal{X}$. At each time step t, $\mathbf{E}^t \in \mathbb{R}^{M\times T}$ is the distance errors for all the ensemble members at all lead times computed based on their ground truth locations. The geographic distance error for ensemble member m at time t with lead time τ is $e^{t,\tau,m} = R_e \Delta\theta(\mathbf{x}^{t,\tau,m}, \mathbf{y}^{t,\tau})$, where R_e is the earth radius and $\Delta\theta(\cdot)$ is great circle central angle between the forecasted and ground truth locations (Williams, 2013). Let $\tilde{\mathbf{e}}^{t,m} = [e^{t-T,T,m}, e^{t-T+1,T-1,m}, \ldots, e^{t-1,1,m}]$ be the distance errors at time t associated with the multi-lead time forecasts generated by the ensemble member m. Let $\mathcal{K} \in \{0,1\}^{M\times T\times N}$ be the mask values corresponding to trajectory forecasts set $\mathcal{X}$. If the mask value is equal to 1, then the corresponding ensemble member forecast is available; otherwise, the corresponding forecast is missing. At each time step t, let $\mathbf{K}^t \in \{0,1\}^{M\times T}$ be the mask values associated with all the ensemble members for all lead times. The mask value of an ensemble member m for forecasts generated at time t with lead time τ is denoted as $k^{t,\tau,m} \in \{0,1\}$.

3. Proposed DTP Framework

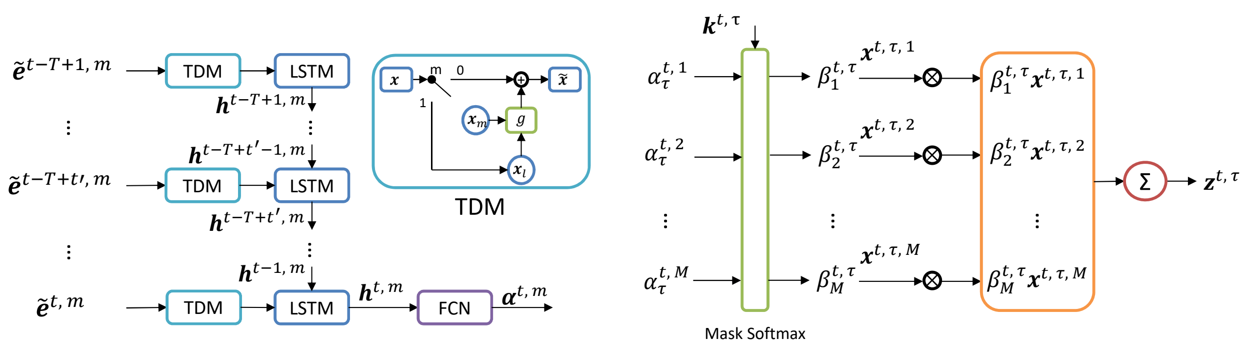
An overview of the proposed DTP framework is shown in Figure 1. In the first stage, as shown in Figure 1(a), a set of LSTMs were trained to learn an embedding of the ensemble members based on their model performance. Specifically, the distance errors associated with each ensemble member were used as input to the LSTM instead of their forecast locations for two reasons. First, learning an embedding of the model output location is harder as the latitude and longitude tend to vary significantly from one hurricane to another, unlike the distance error of the model output, which has a limited range of variability. Second, data imputation is required since the ensemble member forecasts contain missing values for certain lead times. While the missing values in distance error can be effectively imputed using historical distance errors, it is not optimal to directly impute the trajectory forecasts since the current location can be very different than the previous locations. We implemented an approach known as Temporal Decay Memory (TDM) to impute the missing values before providing the data as input to the LSTMs. In the second stage shown in Figure 1(b), the outputs of model performance layer were combined to generate attention-like weights for each ensemble member forecast. The final multi-lead time predictions are computed based on the attention weights and multi-model ensemble forecasts. The details of the proposed DTP framework are discussed below.

3.1. Temporal Decay Memory

Inspired by Temporal Belief Memory (TBM) (Kim & Chi, 2018), we designed a method called Temporal Decay Memory (TDM) to impute the distance errors in hurricane trajectory forecasts as shown in Figure 1(a). It contains a gating unit, corresponds to the missing gate $m \in \{0,1\}$ to indicate whether a value is missing. If the value is missing, then $m = 1$, otherwise $m = 0$. If the value is observed ($m = 0$), the observed value is directly passed to the output. If the value is missing ($m = 1$), the value passed to the output is a combination of the last observation value $\mathbf{x}_l$, the mean value of all observations $\mathbf{x}_m$ and the time interval Δt between the current time and the last observation time. In TDM, we use a function g to calculate the imputed missing value based on $\mathbf{x}_l$ and $\mathbf{x}_m$ as follows:

$$\tilde{\mathbf{x}} = g(\Delta t, \mathbf{x}_l, \mathbf{x}_m) = e^{-\Delta t/\lambda}\mathbf{x}_l + (1 - e^{-\Delta t/\lambda})\mathbf{x}_m \quad (1)$$

where λ is the hyperparameter for the decay rate.



(a) Model performance layer (b) Prediction layer

Figure 1. The figure illustrate the DTP framework. Figure (a) shows model performance layer. Figure (b) shows prediction layer.

3.2. Model Performance Layer

The model performance layer learns the performance of the individual ensemble members, as well as the temporal dependencies between different lead times. Let $\tilde{\mathbf{e}}^{t,m} \in \mathbb{R}^T$ denotes the distance errors of the forecasts generated by ensemble member m with respect to the predicting location at time t. The model performance layer takes the sequence $\{\tilde{\mathbf{e}}^{t-T+1,m}, \tilde{\mathbf{e}}^{t-T+2,m}, \ldots, \tilde{\mathbf{e}}^{t,m}\}$ of an ensemble member as input and output a final hidden state $\mathbf{h}^{t,m}$ using a series of LSTM models. The LSTM output is thus a feature embedding of the ensemble model m at time t. The hidden state $\mathbf{h}^{t,m}$ is calculated as a function of the previous hidden state $\mathbf{h}^{t-1,m}$ and $\tilde{\mathbf{e}}^{t,m}$, i.e., $\mathbf{h}^{t,m} = LSTM\left(\tilde{\mathbf{e}}^{t,m}, \mathbf{h}^{t-1,m}\right)$. A fully connected network (FCN) takes the hidden state $\mathbf{h}^{t,m}$ of the ensemble member as input to generate a multi-lead time performance vector as follows:

$$\boldsymbol{\alpha}^{t,m} = FCN(\mathbf{h}^{t,m}) \in \mathbb{R}^T \tag{2}$$

3.3. Prediction Layer

In this layer, a temporal attention-like mechanism is used to automatically generate the weight for each ensemble member across all lead times. Based on the outputs of the model performance layer, a multi-lead time performance vector $\alpha^{t,m}$ is obtained for each model m at time t. The attention weights for all ensemble members across all lead times are computed using a masked softmax layer. The masked softmax function sets the weights of the missing forecasts to zero. The attention weight $\boldsymbol{\beta}^{t,\tau} \in \mathbb{R}^M$ for all ensemble members at time t with lead time τ can be calculated based on the following equation:

$$\beta_m^{t,\tau} = \frac{\exp(a_\tau^{t,m})}{\sum_{i=1}^M \exp(a_\tau^{t,i})}$$
$$\text{where } a_\tau^{t,m} = \begin{cases} \alpha_\tau^{t,m}, \text{if } k^{t,\tau,m} = 1 \\ -\infty, \text{otherwise} \end{cases} \tag{3}$$

Finally, the multi-lead time predictions at time step t can be computed as a linear combination of the ensemble member forecasts using the following equation:

$$\mathbf{z}^{t,\tau} = \sum_{i=1}^M \beta_i^{t,\tau} \mathbf{x}^{t,\tau,i} \tag{4}$$

3.4. ODTP Framework

Since the DTP framework is trained in a batch mode, its model is susceptible to concept drift and becomes outdated when applied to the hurricane prediction task. To overcome this limitation, the DTP framework is extended to an online learning approach called ODTP, which allows the model to be continuously updated as new observation data become available. Unlike the model performance layer in DTP that employs a fixed sequence of length T to train the LSTM model, ODTP considers only a sequence of length 1 to update the model incrementally. Furthermore, for each given sequence of length T, the hidden state in DTP is initialized to zero. In contrast, the hidden state of the LSTM cell is inherited from its previous time step. This strategy allows the hidden state for each model to be continuously updated to fit new observations in an online fashion. Following the strategy described in (Xu et al., 2014), to alleviate the error propagation problem, the algorithm will backtrack to its previous T time steps and restart the update from time step $t - T$ and incrementally update the model until the current time step t. The online learning with backtrack and restart strategy adopted by ODTP helps to adapt the model to concept drift and overcome the error propagation problem. The pseudocode for ODTP framework is given in Algorithm 1.

4. Experiments

The hurricane best track (ground truth) data and NHC official forecasts are available from the NHC website[1], while

[1] https://www.nhc.noaa.gov

Input: Hyperparameter $\mathbf{\Theta}$ for ODTP model M
Output: Forecasts $\mathbf{z}$
Initialize: Pre-train model $M^{(0)}$ using training dataset;
for *t = 1,2, . . . , N* **do**
    Observe the trajectory location at time t
    /* Backtracking and restart step */
    for $t' = t - T, t - T + 1, \ldots, t - 1$ **do**
        Update model $M^{(t')}$ using backpropagation with all observed trajectory locations till current time t
    end
    $M^{(t)} \leftarrow M^{(t-1)}$
    /* Prediction step */
    for $\tau = 1, 2, \cdots, T$ **do**
        Generate trajectory predictions $\mathbf{z}^{t,\tau}$ using model $M^{(t)}$
    end
end

Algorithm 1: Proposed ODTP framework

the ensemble member forecasts from the years 2012 to 2020 were downloaded from the Hurricane Forecast Model Output website at University of Wisconsin-Milwaukee[2]. According to NHC, 46 models were used in the preparation of their official forecasts. However, only 27 of them have data available from the year 2012 to 2020 at the UWM website. We use these 27 models as ensemble members in our experiments. The final dataset contains 336 tropical cyclones with a total of 7364 observations at 6 hourly intervals. Each tropical cyclone has an average length of 21.9 time steps. For ensemble members with 12-hourly intervals, we performed linear interpolation to impute the missing values so that every ensemble member has 6-hourly forecasts. The hurricane data from 2012 to 2017 (208 tropical cyclones) were used for training and validation while those from 2018 to 2020 (128 tropical cyclones) were used for testing.

4.1. Baseline and Evaluation Metrics

We compared DTP against the following baseline methods:
Ensemble mean: This method uses the mean of the ensemble member outputs at each lead time as its predictions.
Persistence: This method assumes the moving speed at each time step is the same as the previous time step.
Passive-Aggressive(PA) (Crammer et al., 2006): This is a well-known online learning algorithm that updates the weights based on newly observed data points.
ORION (Xu et al., 2014): This is an online multi-task learning algorithm for multi-lead time forecasting.
OMuLeT (Wang et al., 2020): This is a recently developed online learning algorithm for trajectory prediction.

[2]http://derecho.math.uwm.edu/models

Method	Trajectory error (in n mi)			
Lead Time	12	24	36	48
Ensemble Mean	23.30	36.34	50.22	65.03
Persistance	34.84	88.89	155.87	229.63
LSTM	41.64	94.50	160.35	232.80
PA (online)	23.30	36.34	50.23	64.80
ORION (online)	23.37	36.36	50.21	65.00
OMuLeT (online)	**22.33**	**35.33**	48.97	63.77
DTP	23.20	36.08	49.72	64.40
ODTP (online)	22.90	35.50	**48.85**	**63.27**
NHC (gold standard)	24.59	38.49	52.17	65.74

Table 1. Trajectory forecast errors for different methods at varying lead times from 12 to 48 hours.

NHC: This is the gold standard, corresponding to the official forecasts generated by NHC.
LSTM: This is the vanilla LSTM architecture trained on the historical trajectories with a window size of 48 hours.

For a fair comparison, the models are trained on the same training set. The forecasts are evaluated using the Mean Distance Error (MDE) function as follows:

$$\text{MDE}_\tau = \frac{1}{\sum_t k^{t,\tau}} \sum_{t,k^{t,\tau}=1} distance(\mathbf{z}^{t,\tau}, \mathbf{y}^{t,\tau}) \qquad (5)$$

4.2. Experimental Results

The results comparing the hurricane trajectory prediction errors for various methods from 12 hour to 48 hour forecast lead times are shown in Table 1. First, observe that the persistence and vanilla LSTM methods have the worst performance as both methods rely only historical trajectory data only. Second, the performance of the batch DTP framework is slightly worse than OMuLeT, an online learning framework, though they both outperform other baselines including the NHC official forecasts. Finally, ODTP improves the performance of DTP for all lead times as it can adapt to changes in the distribution of the data. It outperforms all other online models such as PA, ORION, and OMuLeT for 36 hour or more forecast lead times. This suggests the benefits of using a nonlinear model to capture the relationships of the multi-lead time forecasts.

To verify that the model performance layer learns an embedding of the ensemble member performance, we analyze the relationship between the input and output of the model performance layer. For each ensemble member m, the LSTM input in the model performance layer is the distance error $\tilde{\mathbf{e}}^{t,m}$ while its output corresponds to the embedding vector, $\boldsymbol{\alpha}^{t,m}$. Figure 2 shows the scatter plots of mean residual distance error $e^{t-\tau,m,\tau}$ against the mean vector $\alpha_\tau^{t,m}$ for all the time steps in a given hurricane for the AVNO model. The correlation between the mean distance errors and mean vector $\alpha_\tau^{t,m}$ are -0.7372, -0.5440, -0.3450, -0.2128 for 12-hour,

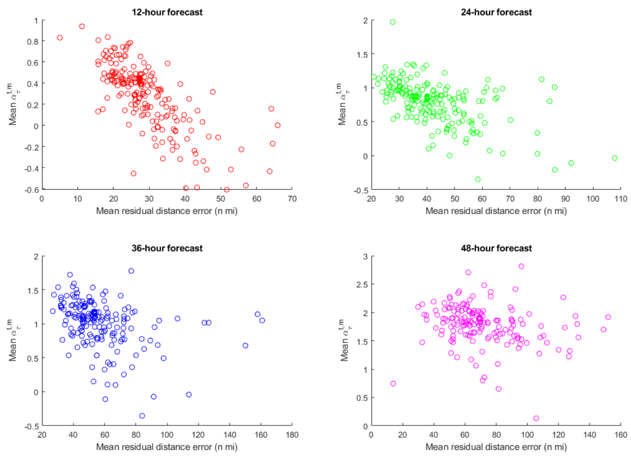


Figure 2. Mean residual distance error $e^{t-\tau,m,\tau}$ vs. mean $\alpha_\tau^{t,m}$ of all the time steps within one hurricane for physical model AVNO.

24-hour, 36-hour, 48-hour lead time forecasts, respectively. The results suggest there is a significant negative relationship between the residual distance error $e^{t-\tau,m,\tau}$ and $\alpha_\tau^{t,m}$. The larger the residual error, the smaller the magnitude of the embedding vector $\alpha_\tau^{t,m}$. This suggests that $\alpha_\tau^{t,m}$ indeed captures the performance of the ensemble member m.

5. Conclusions

This paper presents an LSTM based trajectory forecasting framework called DTP and its online counterpart, ODTP. Unlike existing approaches, the proposed frameworks aim to produce accurate long-range forecasts by leveraging the outputs generated from an ensemble of dynamical models. To handle missing values in the ensemble member forecast data, a novel TDM (Temporal Decay Memory) gating mechanism was developed. Experimental results on real-world data showed that ODTP outperforms all the baseline methods for long-range forecasts up to 48 hours lead time.

References

Alemany, S., Beltran, J., Perez, A., and Ganzfried, S. Predicting hurricane trajectories using a recurrent neural network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 468–475, 2019.

Cheng, H., Tan, P.-N., Gao, J., and Scripps, J. Multistep-ahead time series prediction. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 765–774, 2006.

Cox, T. S., Hoi, C. S., Leung, C. K., and Marofke, C. R. An accurate model for hurricane trajectory prediction. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, volume 2, pp. 534–539. IEEE, 2018.

Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., and Singer, Y. Online passive-aggressive algorithms. *Journal of Machine Learning Research*, 7:551–585, 2006.

Gao, S., Zhao, P., Pan, B., Li, Y., Zhou, M., Xu, J., Zhong, S., and Shi, Z. A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network. *Acta Oceanologica Sinica*, 37(5):8–12, 2018.

Kim, S., Kim, H., Lee, J., Yoon, S., Kahou, S. E., Kashinath, K., and Prabhat, M. Deep-hurricane-tracker: Tracking and forecasting extreme climate events. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1761–1769. IEEE, 2019.

Kim, Y.-J. and Chi, M. Temporal Belief Memory: Imputing Missing Data during RNN Training. In *In Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI-2018)*, 2018.

Knutson, T., Chung, M., Vecchi, G., Sun, J., Hsieh, T.-L., and Smith, A. Climate change is probably increasing the intensity of tropical cyclones. *ScienceBrief Review*, 2021.

Lee, R. S. and Liu, J. N. Tropical cyclone identification and tracking system using integrated neural oscillatory elastic graph matching and hybrid RBF network track mining techniques. *IEEE Transactions on Neural Networks*, 11 (3):680–689, 2000.

Moradi Kordmahalleh, M., Gorji Sefidmazgi, M., and Homaifar, A. A Sparse Recurrent Neural Network for Trajectory Prediction of Atlantic Hurricanes. In *Proceedings of the Genetic and Evolutionary Computation Conference 2016*, pp. 957–964. ACM, 2016.

Mudigonda, M., Kim, S., Mahesh, A., Kahou, S., Kashinath, K., Williams, D., Michalski, V., O'Brien, T., and Prabhat, M. Segmenting and tracking extreme climate events using neural networks. In *Deep Learning for Physical Sciences (DLPS) Workshop, held with NIPS Conference*, 2017.

Rüttgers, M., Lee, S., Jeon, S., and You, D. Prediction of a typhoon track using a generative adversarial network and satellite images. *Scientific reports*, 9(1):1–15, 2019.

Wang, D., Liu, B., Tan, P.-N., and Luo, L. OMuLeT: Online Multi-Lead Time Location Prediction for Hurricane Trajectory Forecasting. In *Proceedings of 34th AAAI Conference on Artificial Intelligence*, 2020.

Williams, E. Aviation Formulary V1.47. 2013.

Xu, J., Tan, P.-N., and Luo, L. ORION: Online Regularized multI-task regressiON and its application to ensemble forecasting. In *ICDM*, pp. 1061–1066. IEEE, 2014.