Deep Learning for Hurricane Track Forecasting from Aligned Spatio-temporal Climate Datasets

Sophie Giffard-Roisin[*]
University of Colorado
Boulder, USA

Mo Yang*
Linear Accelerator Laboratory
Université Paris-Sud, CNRS

Guillaume Charpiat
Inria Saclay–Ile-de-France
LRI, Université Paris-Sud

Balázs Kégl
Linear Accelerator Laboratory
Université Paris-Sud, CNRS

Claire Monteleoni
University of Colorado
Boulder, USA

Abstract

The forecast of hurricane trajectories is crucial for the protection of people and property, but machine learning techniques have been scarce for this so far. We propose a neural network fusing past trajectory data and reanalysis atmospheric images (wind and pressure 3D fields). We used a moving frame of reference that follows the storm center for the 24h tracking forecast. The network is trained to estimate the longitude and latitude displacement of hurricanes and depressions from a large database from both hemispheres (more than 3000 storms since 1979, sampled at a 6 hour frequency). The advantage of the fusion network is demonstrated and a comparison with current forecast models shows that deep methods could provide a valuable and complementary prediction.

1 Introduction

Cyclones, hurricanes or typhoons are words designating the same phenomena: rare and complex events characterized by strong winds surrounding a low pressure area. Their trajectory and intensity forecasts are crucial for the protection of people and property. However, their evolution depends on many factors at different scales, altitudes and time, which leads to difficulties in their modelling. Today, current national forecasts are typically driven by consensus methods able to combine different dynamical models[2]. Statistical forecasting models, on the other hand, still perform poorly with respect to dynamical models, even though the database of past hurricanes is constantly growing. Moreover, a large number of physical variables (pressure, wind fields..) are now available on gridded earth maps from the reanalysis[3], and could be integrated in a statistical or learning method.

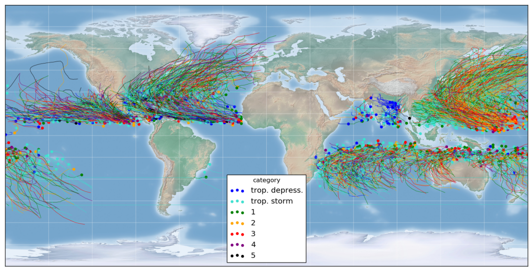
However, only few machine learning methods are tackling the tracking forecast problem. One of them uses a sparse recurrent neural network from only trajectory data (Moradi Kordmahalleh et al. (2016)) and was tested on 6h- and 12h-forecast on only 4 hurricanes. Another study uses storm tracks and reanalysis maps as input for a hybrid ConvNet - LSTM network in order to learn the (x,y) tracking coordinates (Mudigonda et al. (2017)) and showed their 6h-forecast results. The regional map (for image-like physical inputs) was fixed and of size 160 x 80 deg (longitude/latitude). However, a fixed region has three major limitations. Firstly, the tracked storm must stay in the region (while tracks

[*]S. Giffard-Roisin and M. Yang contributed equally to the work. Corresponding author: sophie.giffard-roisin@mines-saint-etienne.org

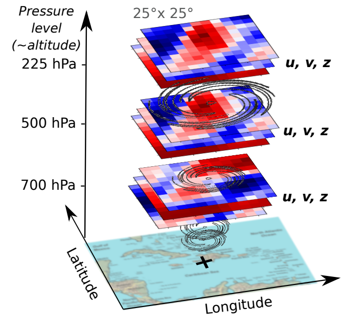
[2]NHC track and intensity models, `www.nhc.noaa.gov/modelsummary.shtml`, Accessed: 2018-07-04.

[3]Reanalysis of past weather data presents a clear picture of past weather, independent of the varieties of instruments used over the years.

Preprint. Accepted to NIPS Modeling and decision-making in the spatiotemporal domain workshop 2018.



(a) Tracking database: more than 3000 tropical/extra-tropical storm tracks since 1979. Dots = initial position, colors = maximal strength (Saffir-Simpson scale).



(b) Global atmospheric grids centered on the storm location: wind fields (u and v) and geopotential height (z).

Figure 1: Tracking data and registered reanalysis data.

often cross oceans, see Fig. 1a), forcing the selection of a large region, even if it is constrained by memory issues (Mudigonda et al. (2017)). Moreover, learning local phenomena on a large and not centered image can be difficult. Finally, it prevents information transfer between storms coming from different basins or regions, while ground truth data is scarce. In a recent work (Giffard-Roisin et al. (2018)), we showed the advantage of using a moving reference CNN model for forecasting hurricanes tracks 6 hours into the future with respect to the other learning methods (30km error with respect to more than 60km). However, a 6h-forecast is of no use for catastrophe planning and it is not possible to compare to current forecasts as the smallest standard is 24 hours.

In this work, we propose to use a moving frame of reference that follows the storm center for a 24h-forecast tracking task. We pose the tracking problem as the estimation of the displacement vector $\vec{d}$ between current and future locations. Moreover, we propose to use the reanalysis data as cropped images (25 x 25 degrees) centered on the storm location. That way, the computation is reduced and we can learn from storms coming from a large number of hurricane basins from both hemispheres. We include past temporal information by adding the reanalysis maps from previous time steps. We propose a fusion convolutional neural network taking into account past trajectories and reanalysis images (wind fields and pressure), and we treat each time step of a storm as a training data point.

2 Tracking Data and Reanalysis Data Processing

Tracking Data from Both Hemispheres. The raw storm track data is composed of more than 3000 extra-tropical and tropical storm tracks since 1979 extracted from the NOAA database IBTrACS Knapp et al. (2010), see Fig. 1a. The tracks are defined by the 6-hourly center locations (latitude and longitude). They come from both hemispheres and the number of records per storm varies from 2 to 120 (total: more than 90,000 time steps). A storm's future displacement (here in 24h) can be predicted from its historical displacement. We define a displacement as the vector $\vec{d} = (\delta lon_t, \delta lat_t)$ between two successive locations of one storm, t being a multiple of 6 hours. We used as features the two past displacements of the storm. We added also some "0D-features" from the IBTrACS database: the current latitude, longitude, and max. sustained windspeed, the Jday predictor (DeMaria et al. (2005)), and the current distance to land. In total, 9 features per time step are extracted.

Reanalysis Data. The trajectory of a storm depends on large scale atmospheric physical phenomena. We applied a sparse feature selection technique (automatic relevance determination, based on linear regression) over 10 available reanalysis fields on pressure levels from the ERA-interim database (Dee et al. (2011)). It highlighted the usefulness of wind fields and geopotential heights (that can be seen as pressure maps). Thus, we extracted them on the neighborhood of the storm at every time step t. Specifically, we extracted the u-wind, v-wind and z fields on a 25x25 degree grid centered on the current storm location, at 3 atmospheric pressure levels ($700/500/225hPa$). The choice of the 3 pressure levels was inspired by related work in the literature on statistics forecast models (DeMaria

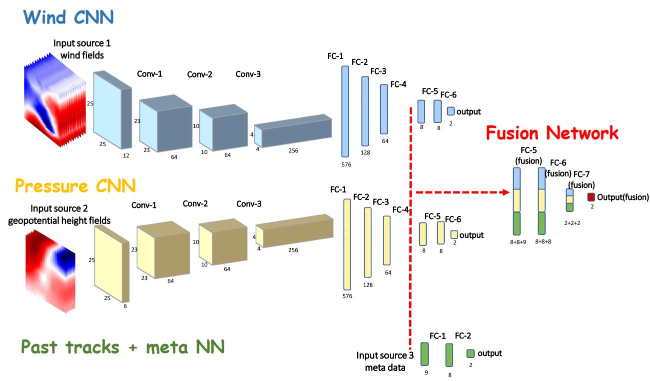


Figure 2: General architecture: the three types of data are feeding three neural networks trained separately. The final fused network is re-trained before predicting the 24h-forecast displacement.

et al. (2005)). In order to capture the dynamics, we extracted the wind fields measured at $t-6h$ at the same locations: the data can be seen as 9 small videos of 2 frames each.

3 The Model

General Framework: Fusing Convolutional Networks. Even though the long-short-term memory (LSTM) networks are designed for predicting time-series events, they are difficult to train and simpler CNNs can often outperform LSTMs (Bai et al. (2018)), and encode time frames as different channels already proved its efficiency (de Bezenac et al. (2017)). Because of the different nature of the data sources, it is not straightforward to mix all the data into a neural network (NN). We propose a fusion of three different NN architectures (see Fig. 3). The Wind CNN and Pressure CNN are convolutional NN that take atmospheric fields as input, while the Past tracks + meta NN is a small network which takes 0D features as input. Each stream network first learn its parameters independently for the same task, i.e. predicting the 24h-forecast displacement $\vec{d} = (\delta lon_{24h}, \delta lat_{24h})$. We then integrate the three networks into a fusion network and retrain it (see Fig. 3).

Wind CNN and Pressure CNN. The two CNN networks are very similar, however the type of data is different thus different learning rates were need, that is why we separate them into two networks. The Wind CNN data consists in 12 channels, and 6 channels for the Pressure CNN (concatenation of every dimension). We used a typical CNN architecture alternating convolutional layers (Conv layer) and max-pooling layers with fully connected layers at the end (Simonyan and Zisserman (2014)). All hidden layers are equipped with the rectification (ReLU) non-linearity and batch normalization. We have evaluated different configurations (from one to four Conv layers) of Wind CNN on our validation set before selecting 3 Conv layers.

Past tracks + meta NN. We designed a small neural network (two fully connected layers) represented as the green stream in Fig. 3. It is able to learn the future displacement from its past displacements and other handcrafted data (see section 2). We use two past displacements (from $t-12h$ to $t-6h$ and from $t-6h$ to t) because more past tracks did not improve the performance.

Combining Neural Networks. Once the three individual stream networks are trained, we concatenate their 3 last layers and add a layer at the end of the network as the fused output layer. We initialize to zero the weights of the new connections in these 3 layers (across streams). We then re-train the whole fused network by allowing every weight to be optimized. The number of fused layers (3) was determined by comparing different configurations.

Algorithmic Details. The storms were randomly separated in 3 sets as follows: train (60%) / valid (20%) / test (20%). All time instants were treated independently within each set and the input data was standardized. The loss function was set as the mean square error (MSE) in kilometers between

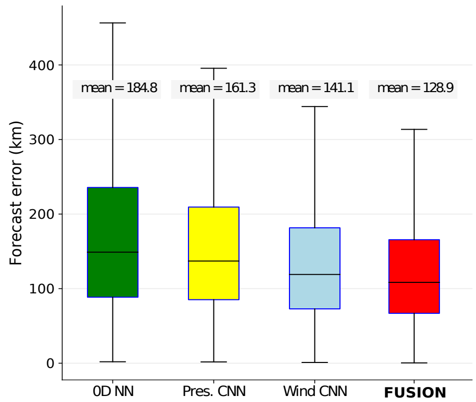


Figure 3: 24h-forecast results on the test set (storms coming from all oceanic basins), in distance between predicted and real location.

Model	Atlantic errors (km)		East Pacific errors (km)	
	mean	std	mean	std
BCD5	125	90	112	78
Fusion	115	67	94	59

Table 1: Mean and standard deviation 24h-forecast errors for the Atlantic and Pacific basins on part of the test set (total = 4349 time steps).

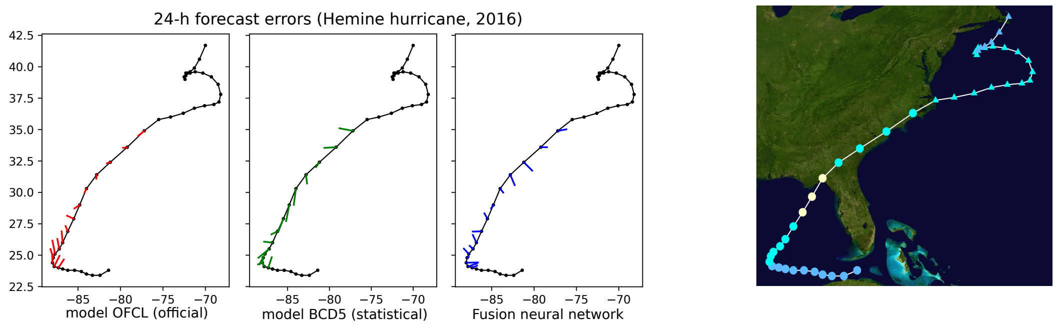


Figure 4: 24-h forecast errors (4 time steps ahead) on Hermine hurricane in 2016. The bars connect each pair of predicted and ground truth location. The larger the length, the larger the error.

the forecast and the true storm location at $t + 24h$. We added an L2 penalty on the weights of the model (*coef.* $= 0.01$). The training was performed by the Adam optimizer, and each model converged within 200 epochs. Every evaluation was repeated three times and an average score was computed. Our implementation uses PyTorch 4.0 on 4 TitanX GPUs with data parallelism (Krizhevsky (2014)).

4 Experimental Evaluation

Fig. 3 shows the 24h-forecast results on the test set (14,256 time steps) in absolute distance error. We can see the improvement of fusing networks with respect to the Wind CNN, Pressure CNN and Past tracks + meta NN. We also compared our fusion model CNN with existing forecasting models: BCD5 is a statistical model which is often used to benchmark other storm track forecasting methods, and OFCL is the National Hurricane Center official forecast (consensus of dynamical models).[4] We extracted the BCD5 prediction results of years 1989-2016 in the Atlantic and Eastern Pacific basins. We compare in Table 1 our fusion network with the statistical BCD5 on the test hurricane instants where both methods provided a forecast (4349 time instants from 258 storms). On both basins, our fusion network behaves better than the BCD5 model on average. Such comparison is not possible with the OFCL as this model is modified every year and they only provide forecasts of the version N of the model for the year N. We don't know the performance of the recent models on previous years, and it would be unfair for them to compare with old results (obtained with earlier, less efficient models). Analyzing the mean errors per year, our deep learning model performs better than the OFCL

[4]National Hurricane Center Forecast Verification, `https://www.nhc.noaa.gov/verification/verify6.shtml`, Accessed: 2018-07-31.

forecast until year 2010 for the Pacific basin (2005 for the Atlantic). During the years 2010, the OFCL method improved and its mean errors per year are smaller than ours.

We also compared qualitatively the predictions with both OFCL and BCD5 models for recent storms of the test set, as the Hermine hurricane in 2016 (Fig. 4). The small bars connect each pair of predicted and ground truth location (after 24 hours). The larger the length, the larger the error. Even though the official OFCL model has globally smaller forecast errors, on some time points our model outperforms the OFCL. Moreover, the 3 forecasts have often different directions. If we don't expect to perform better than a current official ensemble of dynamical models, a neural network model can help the current forecast modellers by providing a complementary prediction that could be integrated in a consensus method.

5 Conclusion

We designed a neural network for the storm track 24h-forecasting using a moving frame of reference able to use a common dataset and a common training for every hurricane of both hemispheres. We demonstrated the benefit of coupling past displacements and registered reanalysis images. By comparing results with current forecast models, we think that such a different approach can be beneficial if integrated in a consensus method.

References

Bai, S., Kolter, J. Z., and Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

de Bezenac, M., Pajot, A., and Gallinari, P. (2017). Deep learning for physical processes: Incorporating prior scientific knowledge. *arXiv preprint arXiv:1711.07970*.

Dee, D. P., Uppala, S. M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al. (2011). The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, 137(656):553–597.

DeMaria, M., Mainelli, M., Shay, L. K., Knaff, J. A., and Kaplan, J. (2005). Further improvements to the statistical hurricane intensity prediction scheme (ships). *Weather and Forecasting*, 20(4):531–543.

Giffard-Roisin, S., Yang, M., Charpiat, G., Kégl, B., and Monteleoni, C. (2018). Fused deep learning for hurricane track forecast from reanalysis data. In *Climate Informatics Workshop Proceedings 2018*, Boulder, United States.

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J. (2010). The international best track archive for climate stewardship (ibtracs) unifying tropical cyclone data. *Bulletin of the American Meteorological Society*, 91(3):363–376.

Krizhevsky, A. (2014). One weird trick for parallelizing convolutional neural networks. *arXiv preprint arXiv:1404.5997*.

Moradi Kordmahalleh, M., Gorji Sefidmazgi, M., and Homaifar, A. (2016). A sparse recurrent neural network for trajectory prediction of atlantic hurricanes. In *Proceedings of the Genetic and Evolutionary Computation Conference 2016*, pages 957–964. ACM.

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

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.