



SESAR Innovation Days 2017, Belgrade, Serbia

28 – 30 November 2017

Short-term 4D Trajectory Prediction Using Machine Learning Methods

Zhengyi WANG

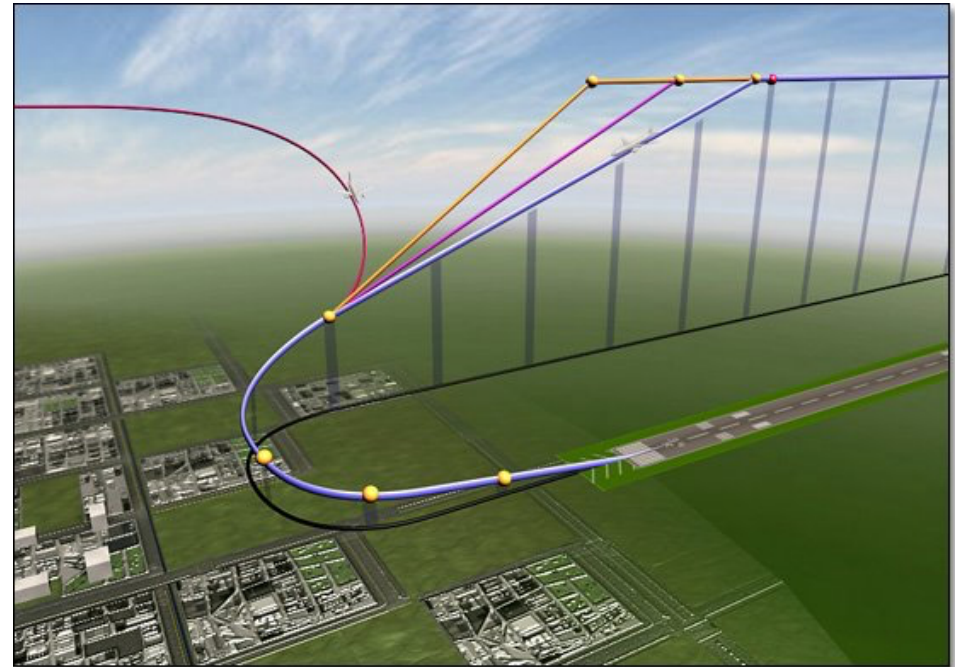
Man LIANG

Daniel DELAHAYE

ENAC, 31055, Toulouse, France

4D trajectory prediction

- ❖ Influence factors:
 - Aircraft weight
 - Pilot actions
 - Wind and temperature
 - etc.
- ❖ Two categories of prediction:
 - short-term: in minutes
 - long-term: in hours



uncertainty

accuracy

efficiency

Previous works

4D trajectory prediction model

```
graph TD; A[4D trajectory prediction model] --> B[Aircraft performance model]; A --> C[Machine learning model];
```

Aircraft performance model

Physics-based approach

Deterministic model

- aircraft motion: BADA
- planned flight routes
- predicted atmosphere condition
- Aircraft Intent

Machine learning model

Data-driven approach

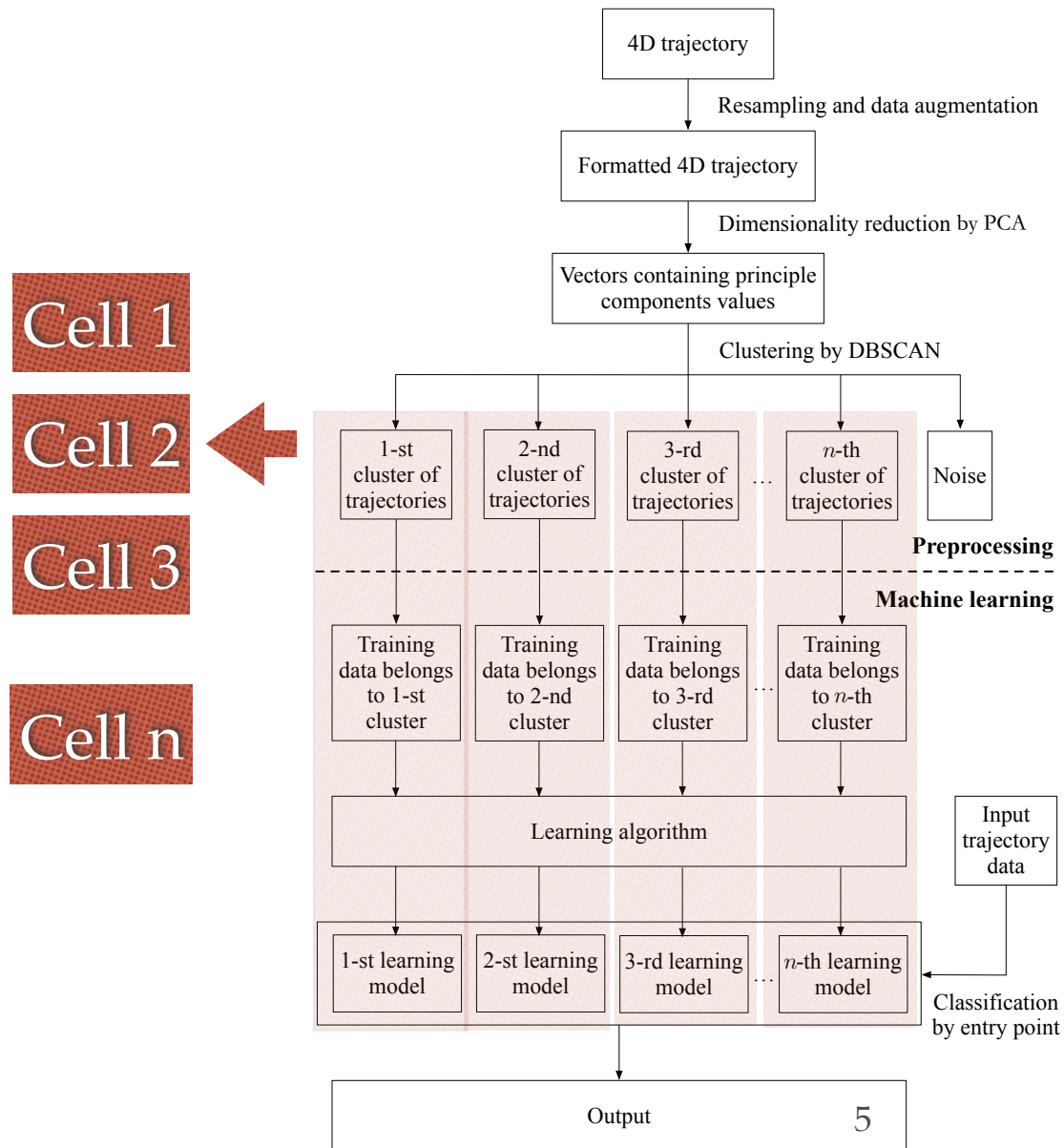
Learning model

- input data: surveillance data and meteorological data
- learning algorithm
- data-mining

Objective

- ❖ We propose a novel **short-term** trajectory prediction model in Terminal Maneuvering Area (TMA) to predict **Estimated Time of Arrival (ETA)**.
- ❖ This model can be divided into **two** main parts:
 1. Preprocessing part
 - ❖ to identify the 4D trajectories into different clusters and remove the noises in an efficient way.
 2. Machine learning part
 - ❖ to apply Multi-Cells Neural Networks (MCNN) technique to process different traffic data prediction.

Methodology



Preprocessing

- data clearing
- data clustering

Machine learning:

- parallel processing
- Neural Network

Cell 1

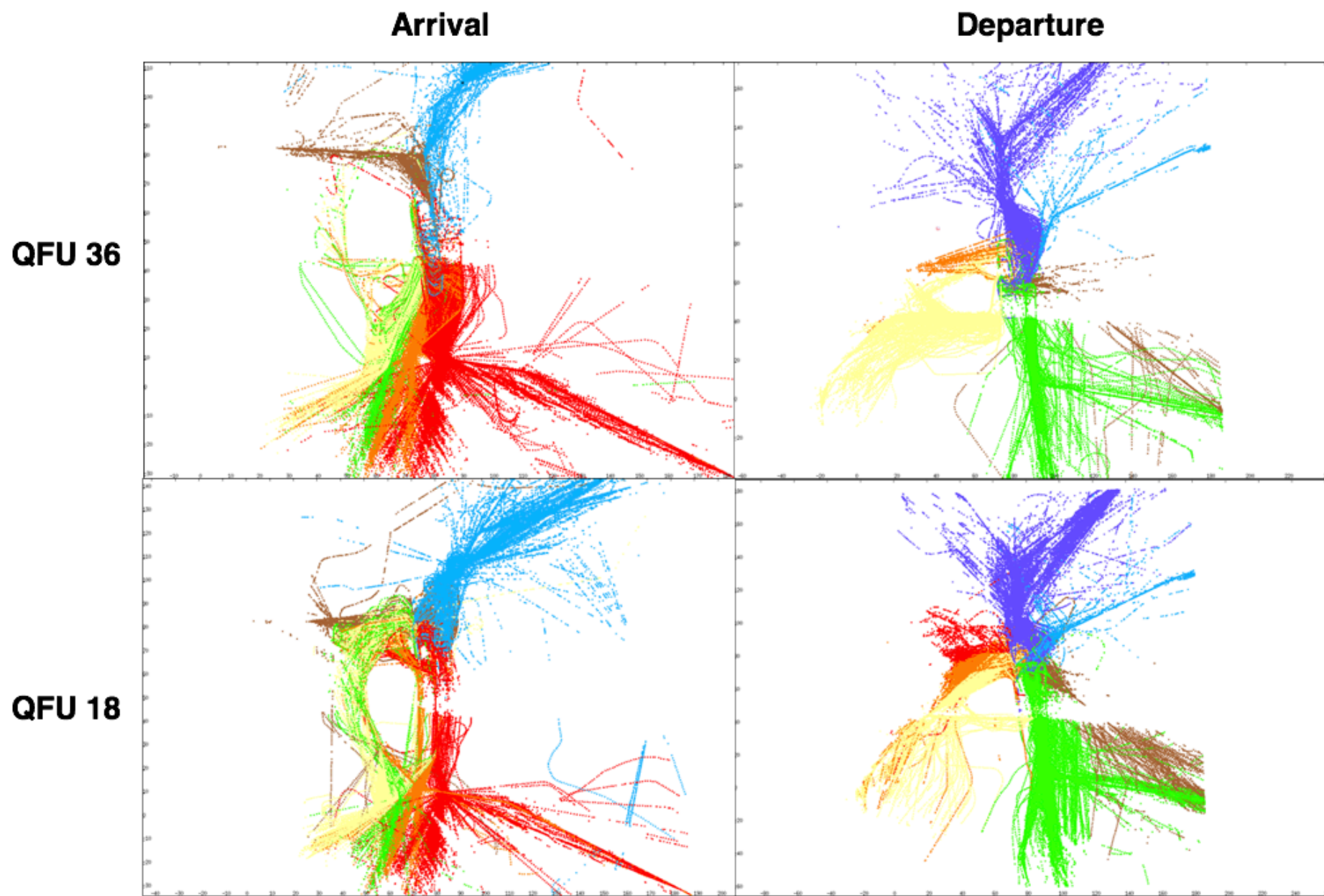
Cell 2

Cell 3

Cell n

Input data

- ❖ The available dataset includes **ADS-B records in July, 2017** over the TMA of Beijing Capital International Airport (**BCIA**).
- ❖ **Each record** of data contains:
 1. Type of operation (departure / arrival)
 2. Record beginning time
 3. Aircraft number
 4. Position(X,Y,Z)
 5. Heading
 6. Horizontal velocity
 7. Vertical velocity
- ❖ Each record with the same aircraft number belongs to an aircraft i , and the collection of all records for that aircraft forms the trajectory.
- ❖ We select the dataset with **36288 flights and 3242384 trajectory points**.



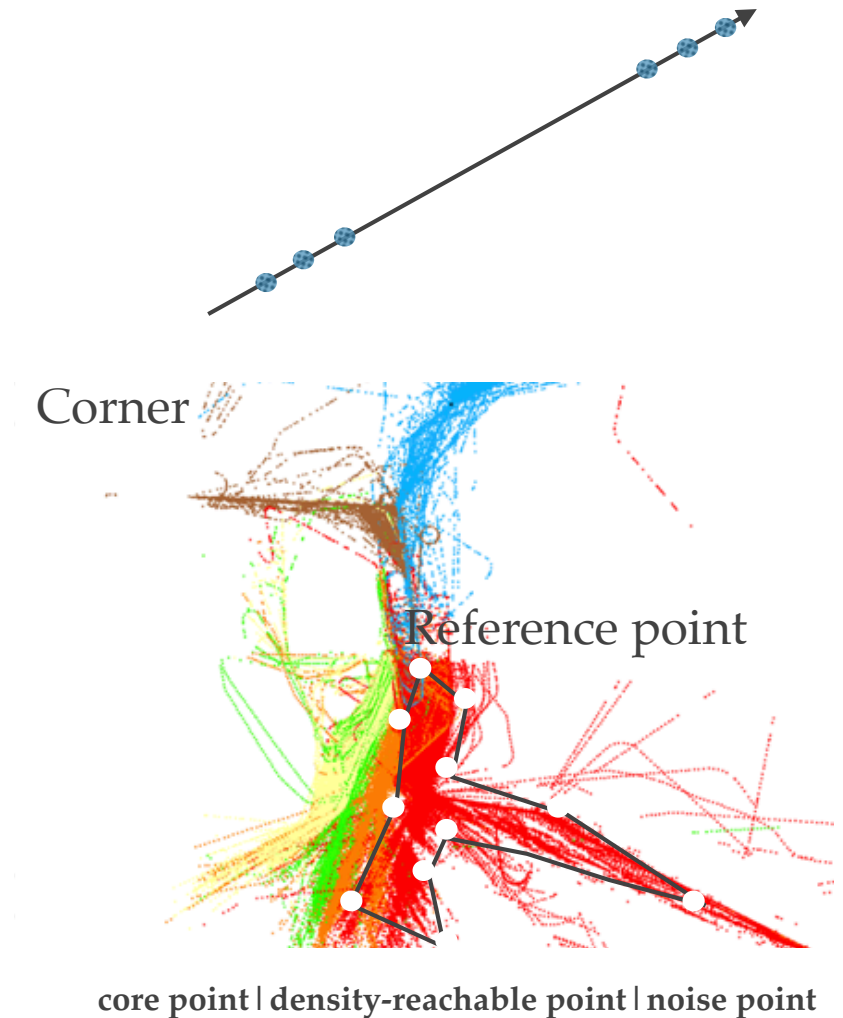
Runways 18R/36L and 18L/36R traffic samples in Beijing Capital International Airport

Clustering-based preprocessing

❖ Four steps:

- 1) **Data cleaning and formatting:** trajectories less than 50 points are eliminated. All trajectories are resampled to same length for clustering.
- 2) **Dimensionality augmentation:** to improve the clustering performance. The final features are: position (X , Y , Z), heading Ψ , distance from the reference point R, distance from the corner point D, angular position from the reference point Θ .
- 3) **Principal component analysis:** to reduce redundant elements which may decrease computational efficiency, even lead to larger errors.
- 4) **Clustering via DBSCAN*:** to divide the dataset into several clusters and noises.

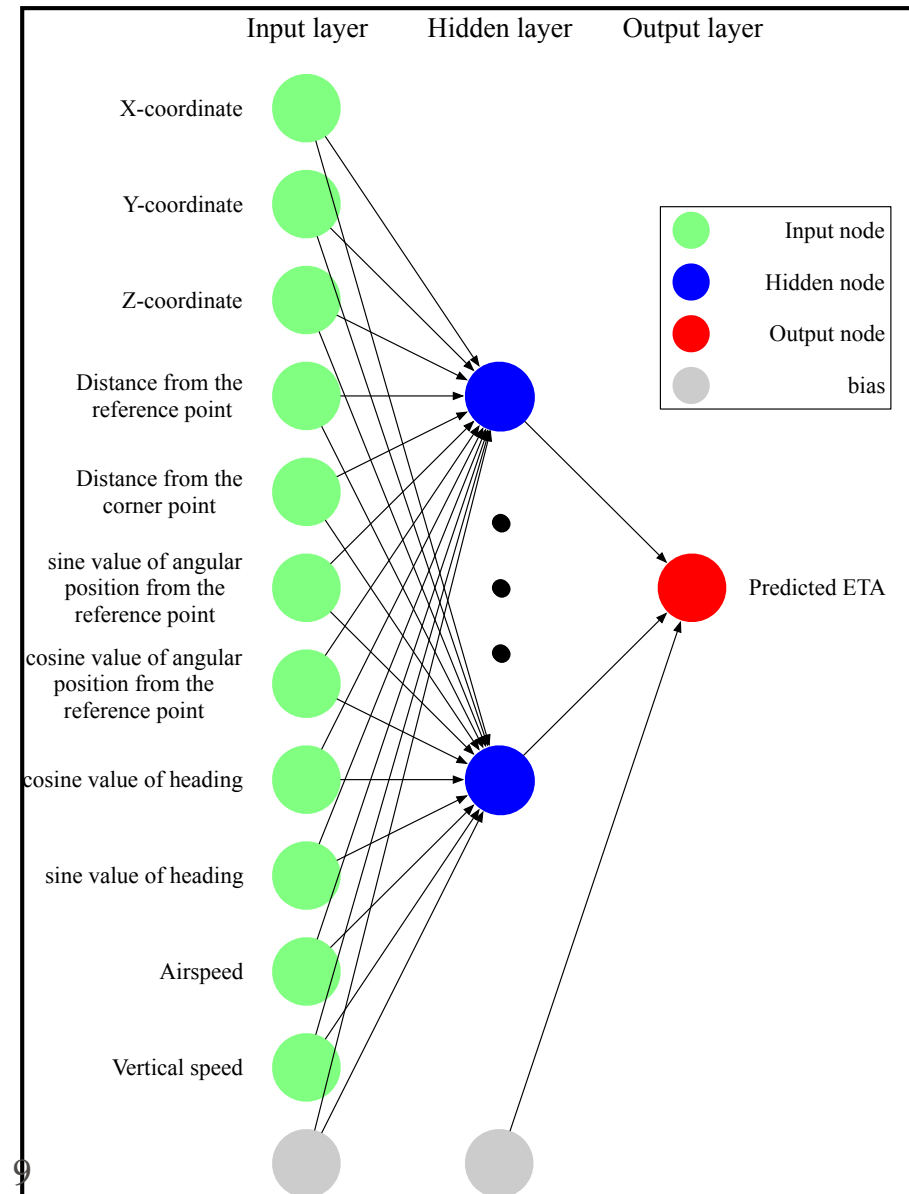
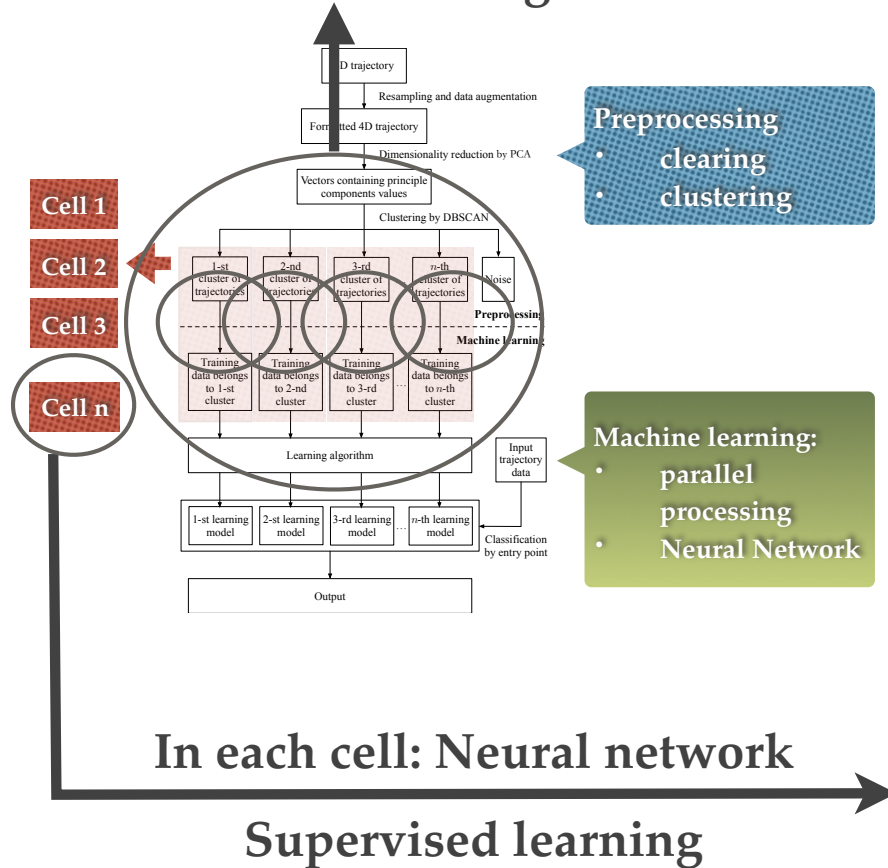
*DBSCAN: Density-Based Spatial Clustering of Applications with Noise



Learning model

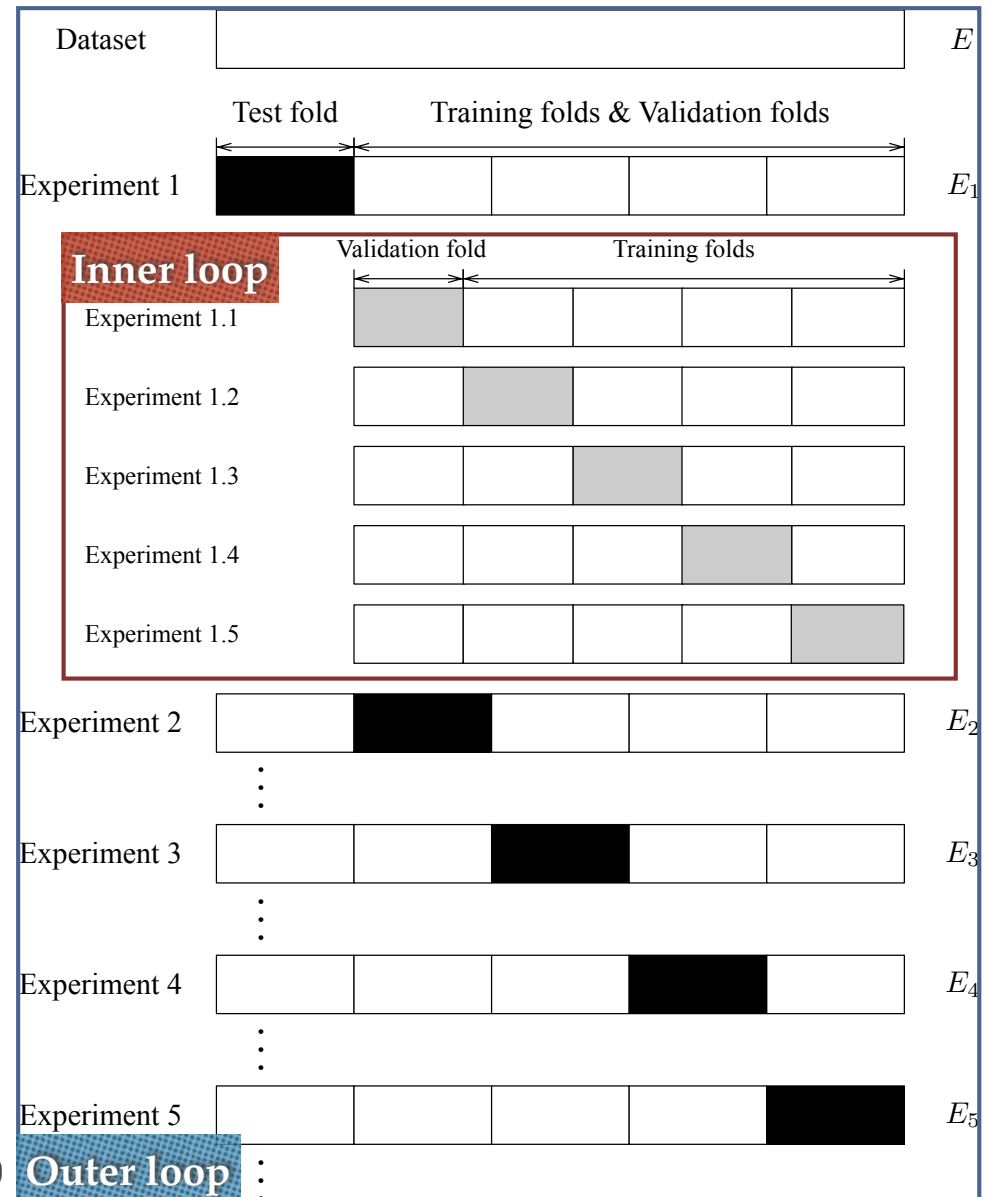
Objective: Estimated Time of Arrival (ETA) prediction

Parallel learning



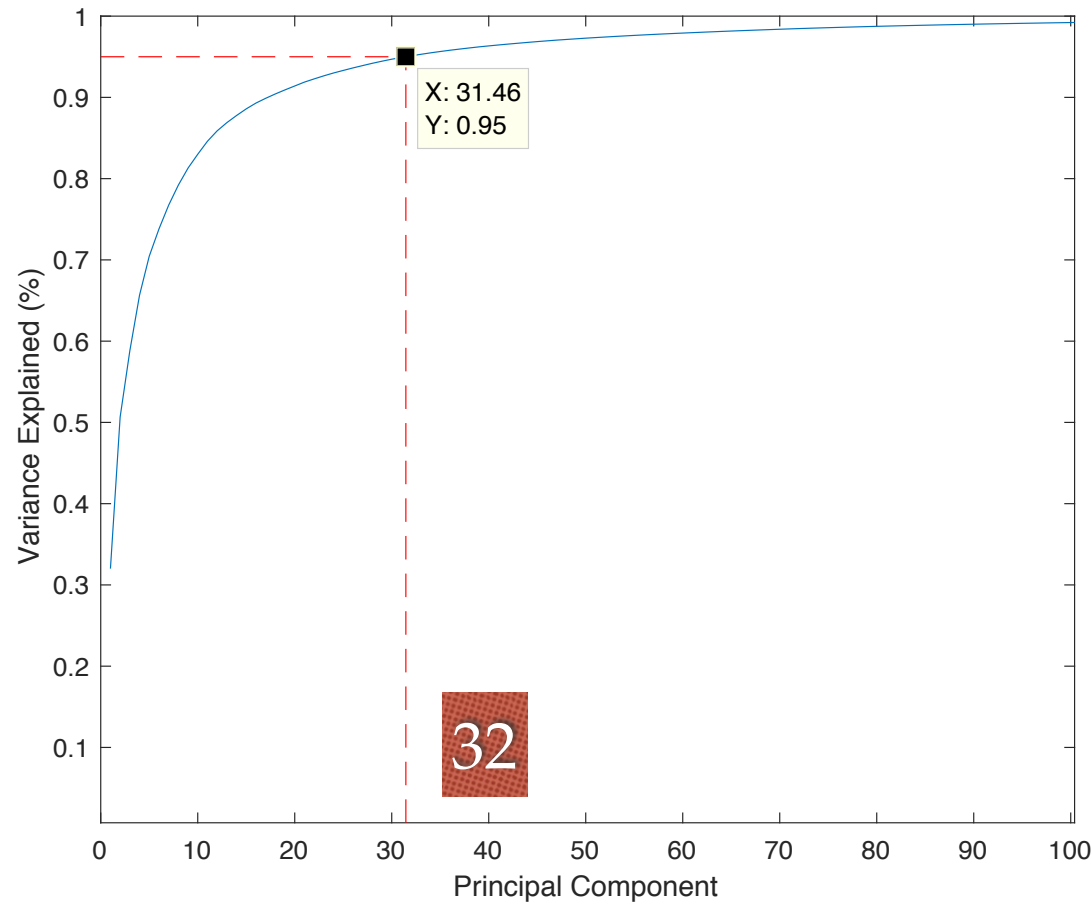
Validation of learning algorithm

- ❖ Nested cross validation
 - The **inner** loop, for parameters selection, such as learning rate, number of hidden nodes;
 - The **outer** loop, to validate the robustness of our prediction model.
- ❖ The proportion of training sets, validation sets and test sets is 64% / 16% / 20%.



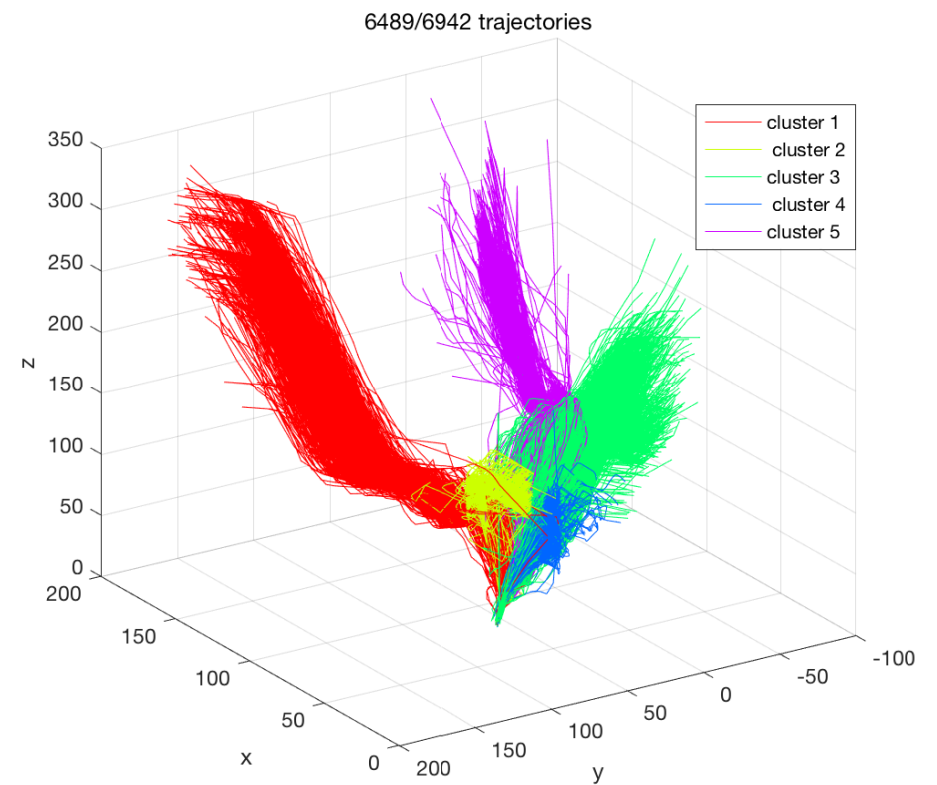
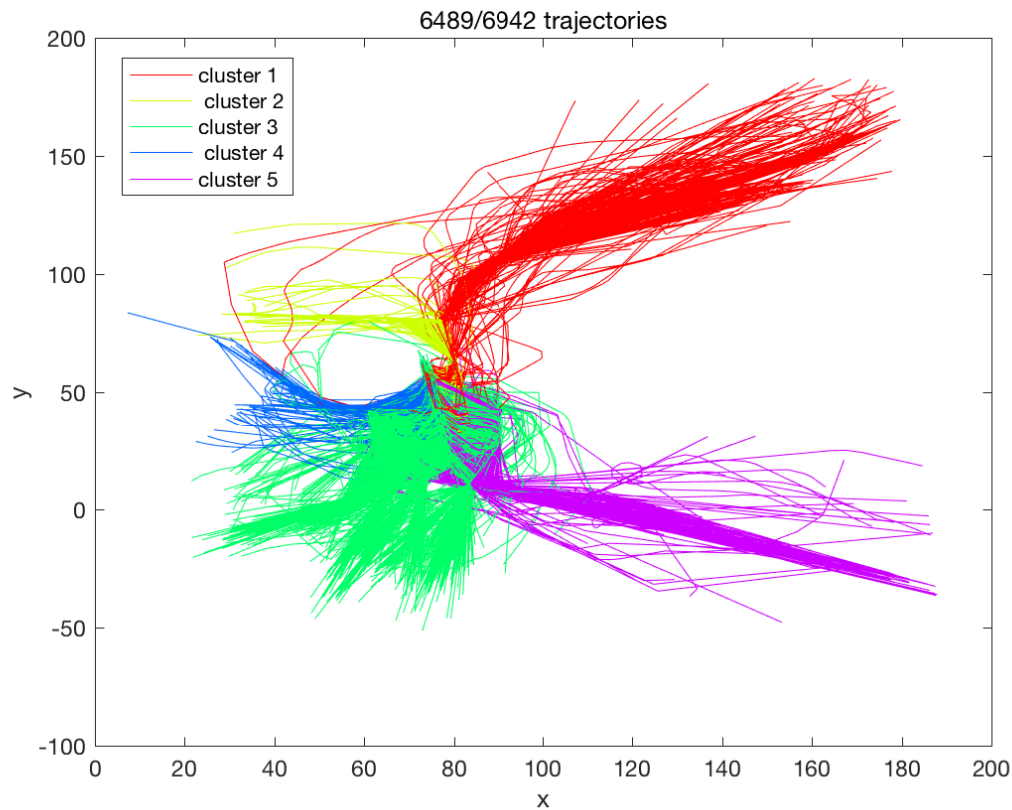
Experiment and result

We use dataset containing 8677 arrival flights of QFU 36.



Result 1: Principal component

Experiment and result

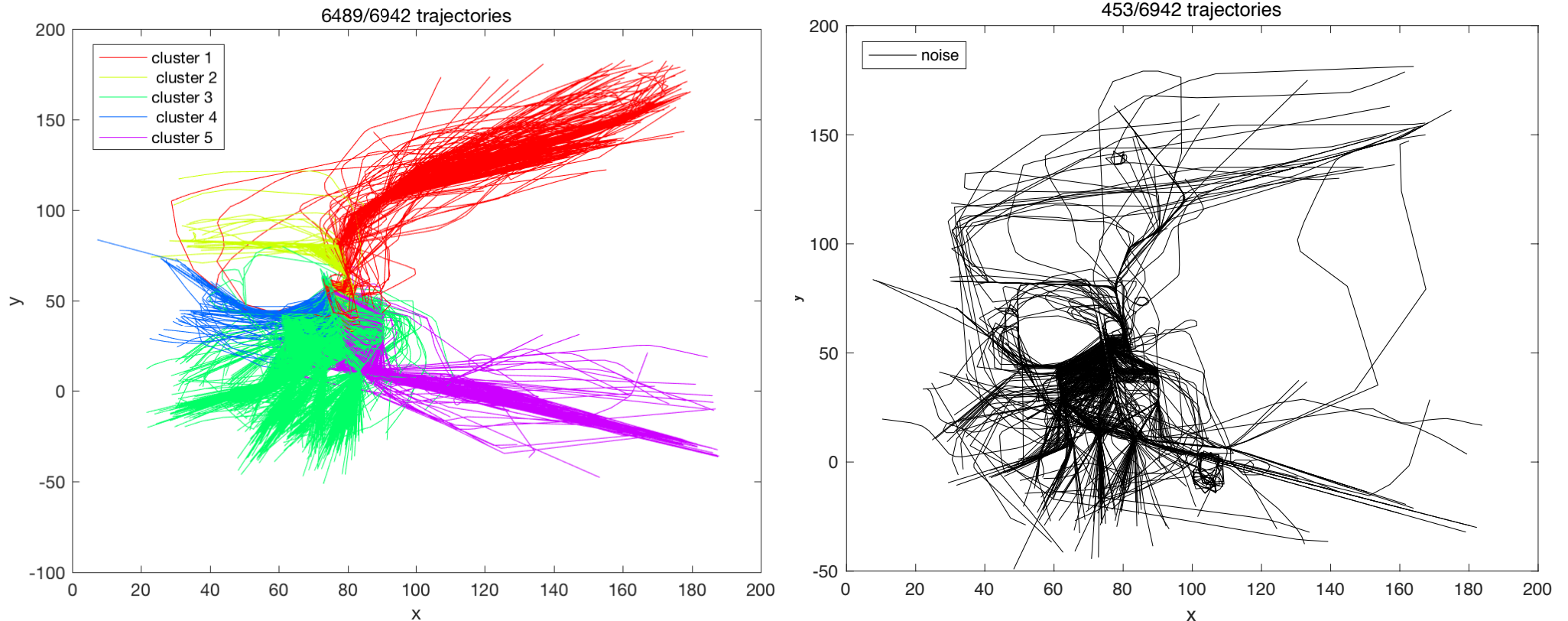


Result 2: Clustering with $\varepsilon = 1.8$ and MinPts = 200

(ε affects the size of cluster, MinPts affects the noise identification)

Experiment and result

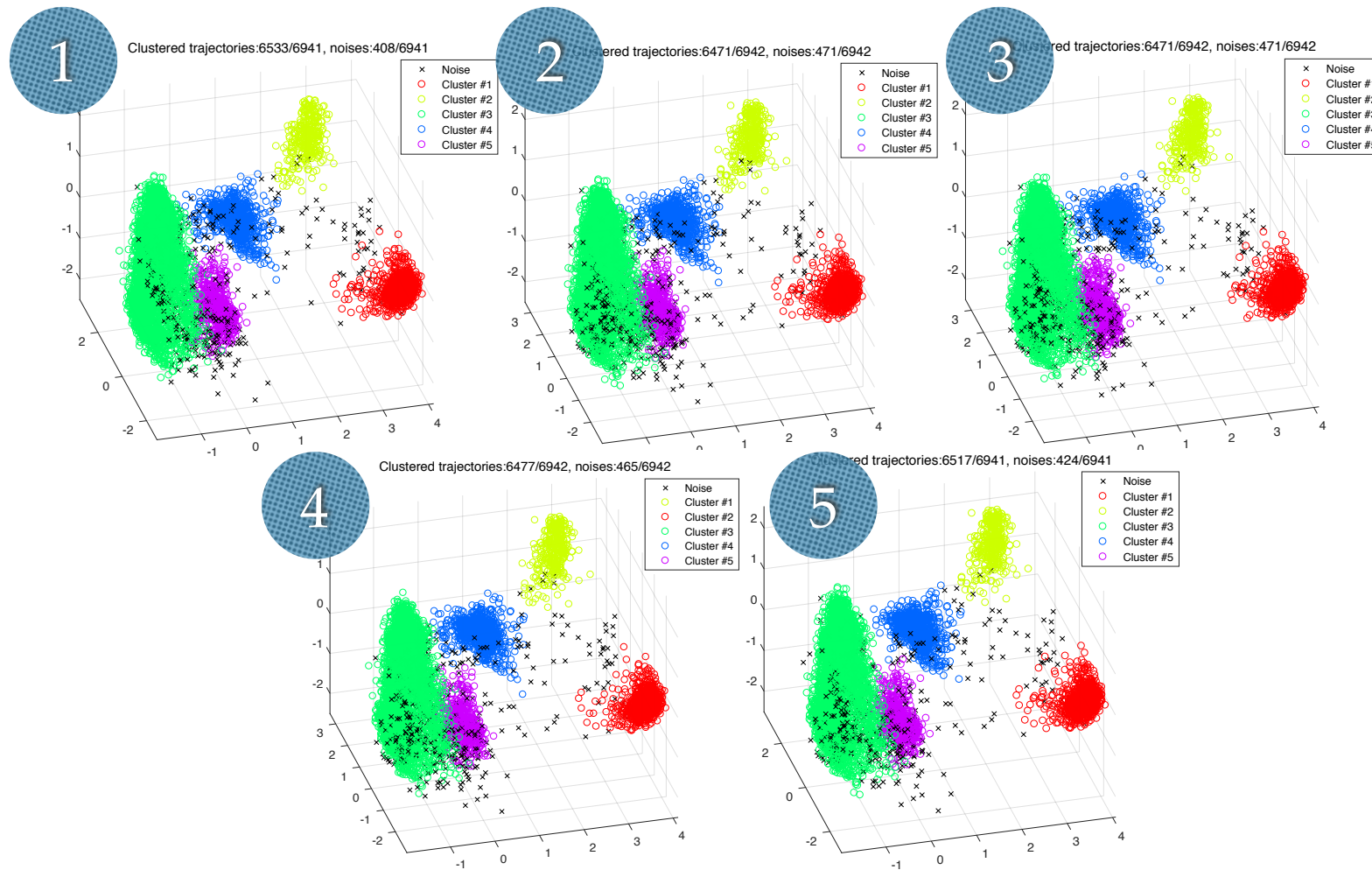
Clustered trajectories account for 93.47%, Noises represent 6.53% of total trajectories.



Result 3: Noise results in clustering with $\varepsilon = 1.8$ and MinPts = 200

(ε affects the size of cluster, MinPts affects the noise identification)

Experiment and result



Result 4: Clustered partitions for each iteration in outer loop

Experiment and result

We use **Mean Absolute Error (MAE)** and **Root Mean Squared Error (RMSE)** to evaluate our trajectory prediction model performance. We compare our **NN** prediction model with **Multiple Linear Regression (MLR)** model which is the most common form of regression analysis, frequently applied to prediction.

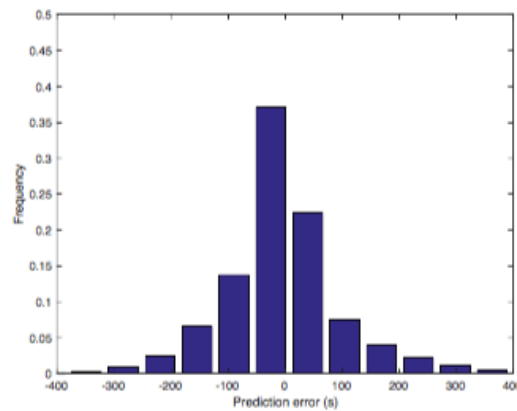
Result 5: Performance on ETA prediction of NN and MLR with preprocessing step

Partition number	percentage	MAE for NN+P. (s)	RMSE for NN+P. (s)	MAE for MLR+P. (s)	RMSE for MLR+P. (s)
Cluster 1	13.85%	106.08	141.51	113.67	150.20
Cluster 2	5.62%	82.91	108.08	92.99	118.59
Cluster 3	58.39%	61.68	97.81	82.48	117.14
Cluster 4	13.64%	46.00	69.39	51.09	75.12
Cluster 5	8.51%	88.76	124.31	97.42	132.62
Total	100%	69.19	104.82	84.37	119.13

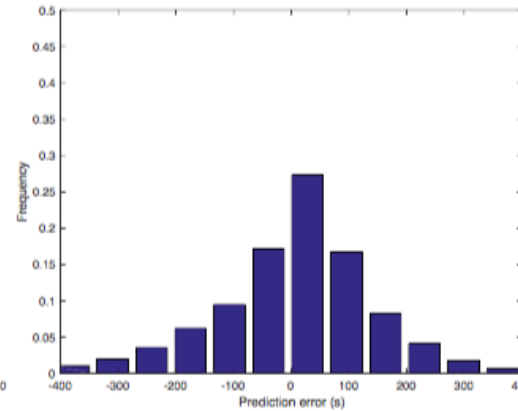
Result 6: Performance on ETA prediction of NN and MLR without preprocessing step

Model	MAE (s)	RMSE (s)
MLR without P.	108.03	160.40
NN without P.	76.28	127.76

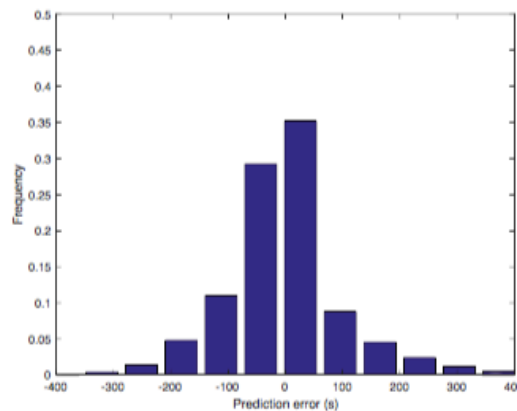
Experiment and result



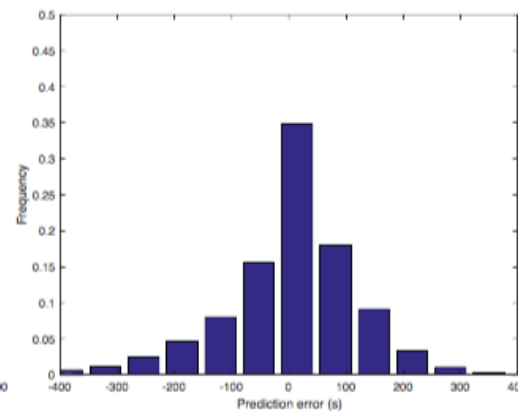
(a) NN without Preprocessing



(b) MLR without Preprocessing



(c) NN with Preprocessing



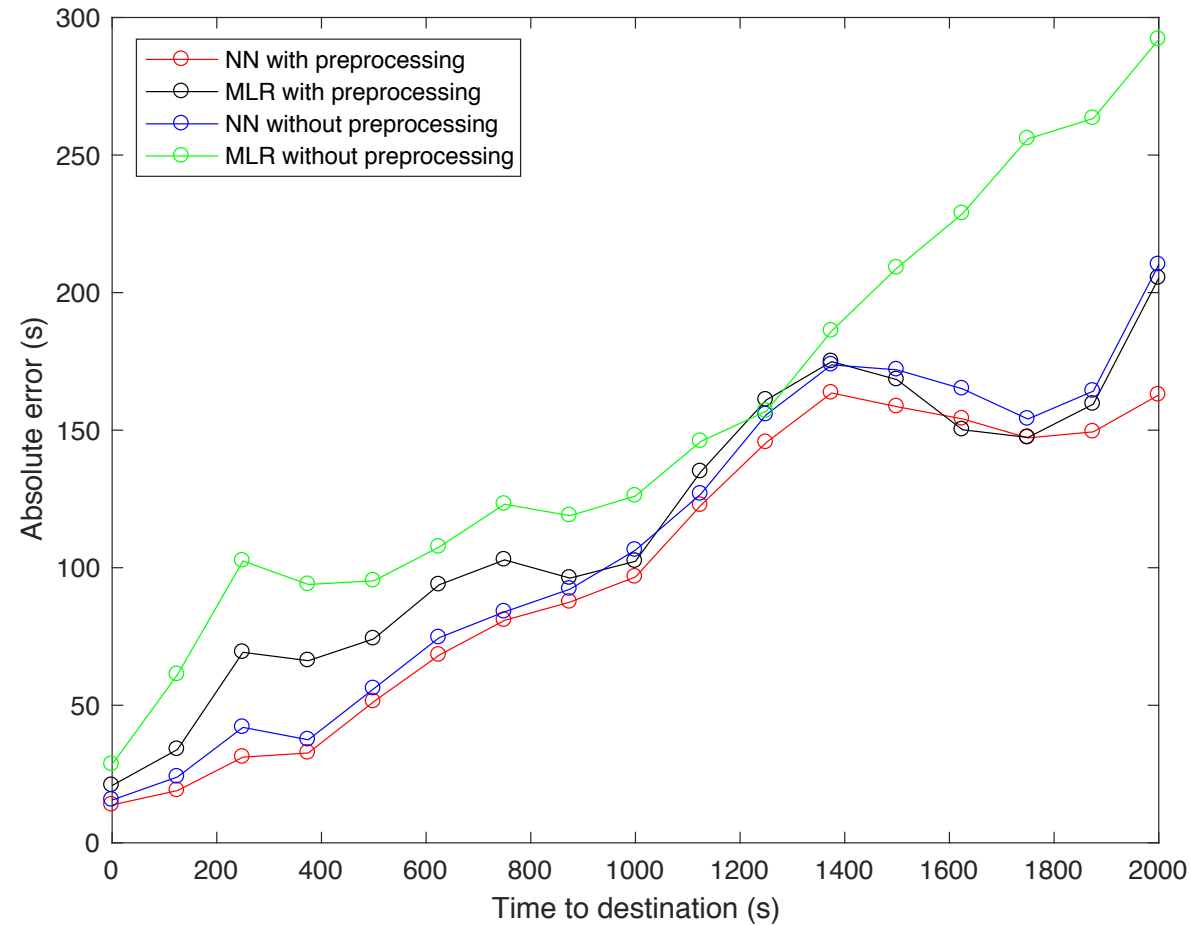
(d) MLR with Preprocessing

Y: Frequency,
which presents the
percentage of
trajectories on the
associated error.

X: Prediction error

Result 7: The distribution of ETA prediction errors with different methods

Experiment and result



Result 8: Mean absolute error of ETA prediction with the time to destination

Conclusion

1. We propose a novel hybrid 4D trajectory prediction model based on clustering and MCNN.
2. Our proposed model is robust. The preprocessing part can provide high-quality inputs to the prediction part.
3. Our proposed model can improve the accuracy of prediction, compared with Multiple Linear Regression (MLR) model.

Future work

- ❖ Future work could be conducted in **different look-ahead times**, on a comparison with results from model-based methods, as well as on studying prediction accuracy for **other trajectory variables** besides ETA. Moreover, more complex prediction model, such as **deep learning** approaches, would be very valuable.

End of presentation

Thanks for your
attentions!

Questions?

cauc_wzy@163.com

man.liang@enac.fr

daniel@recherche.enac.fr
