USE OF THE FAST NEAREST-NEIGBHOR ALGORITHM FOR THE TROPICAL CYCLONE TRAJECTORY FORECAST

Jean Marc RABEHARISOA, Miloud BESSAFI, A. Adolphe RATIARISON, Olivier RAKOTOVAZAHA

ABSTRACT

Tropical cyclone tracks are predicted by an analog ensemble forecast model. The analogs for which we wish to make a forecast are selected among all the situations met in the past in a basin. A new analogs selection technique known as ' Partial Axis Tree' (PAT) was inserted in the self adapting model. This algorithm makes it possible to carry out research on the point neighbourhood given. The quality of prevision is evaluated by calculating the distance between the real position and the calculated position for a time lag of 6 hours. Compared to the former selection method, which is used as reference method, the new method of selection improves, on average, the quality of forecast.

KEYWORDS:

Predictors, PAT, track, situation, basin, analog, statistics, self adapting

1 - INTRODUCTION:

The used methods for the forecasting of cyclone trajectory divide generally in two big classes: the statistical models, the numeric models. Since these last thirty years, some considerable progress have been gotten with the meteorological models with the advent of the computers more and more effective in term of memory capacity and calculation speed, the development of new observations means (ex: the meteorological satellites, the radars, the embarked flights, the drifting buoys), the integration of sophisticated preliminary treatment model to establish an initial state of the atmosphere (data assimilation). The trajectory forecasting majority is of CLIPER model (Climatology and Persistence) type introduced by Neumann (Neumann, 1981). An analog ensemble forecast model is used by Oliver Sievers (O. Sievers, 2000) to predict cyclone trajectory in the Atlantic and east Pacific basins. It is self-adapting model. K. Fraedrich (K. Fraedrich, 2003) used the analog ensemble forecast model in the Australian basin. Compared to his model of reference (CLIPER), they found the good performance in forecast error.

The purpose of this paper is to integrate a new analogs selection technique in the self-adapting model in the worry to improve the tropical cyclone tracks forecast. For verification purposes these forecasts are compared with the former analogs selection, which is used as a reference selection method. The outline of this note is as follows: In section 2, the new analogs selection technique is explained. Section 3 applies the method in the austral basins. Section 4 shows the results and discussion. Finally, section 5 summarizes the results.

2- Fast nearest neighbor algorithm: Partial Axis Tree (PAT)

Partial Axis Tree (PAT) permits to divide the space of points in an efficient manner in term of speed for the "nearest neighbor" determination (James Mc Names, 1999).

This research algorithm is based on a very fast tree pruning, thanks to its elimination criteria power, by limiting the necessary storage space in memory. PAT uses principal component analysis (PCA) by constructing the principal axis of a given points with like objective to divide the dataset in ***nc*** region in the same point number. This process repeats itself for every subgroup until the number of point in a subgroup is lower to ***nc***.

The fast nearest neighbor algorithm consists of two parts:

- Principal Axis Tree Construction,
- Principal Axis Tree Search.

* **Principal Axis Tree Construction**

Schematically, it takes place in 4 stages:

1. If the number of points assigned to the node in progress is lower to ***nc***, the node is said terminal node and its treatment is finished, otherwise go to the second stage.
2. Construct the principal axis for the points in progress and project these points on this axis,
3. Share the set of the projected points in ***nc*** distinct regions so that every region contains the same number of points. Each region contains the entire part of point number divided by ***nc***.
4. Assign to every created region a node label. There are ***nc*** distinct nodes.

A tree, for which the set of the points is assigned to the root, is then obtained. These points are separated in different regions.

* **Principal Axis Tree Search**

Via the principal axis that has been saved to every stage, the region that contains the given point can be determined and therefore the associated child node is known. This determination is done by projecting the aforesaid point on the associated principal axis and doing a binary search among the ***nc*** regions limits. From there, the algorithm tempts to eliminate the sibling nodes via the elimination criteria.

1. If the criterion is satisfied, the sibling nodes are eliminated and the analysis goes back up to the related node;
2. If the criterion is not satisfied, the algorithm takes down in the nearest sibling node for an analysis more deepened and we take to the first stage of the first part.

An algorithm of partial distance is applied on the set of the points that belong to these remaining terminal nodes, to have the given point nearest neighbors. Let's recall that there is, by construction, less ***nc*** point in the terminal nodes.

In summary, the process begins with the root, tempts to eliminate whole sections of the tree via the elimination criteria and takes down toward the terminal node that is correspondent to the point of which we look for the neighbourhood.

The partial distance algorithm requires defining a distance formula.

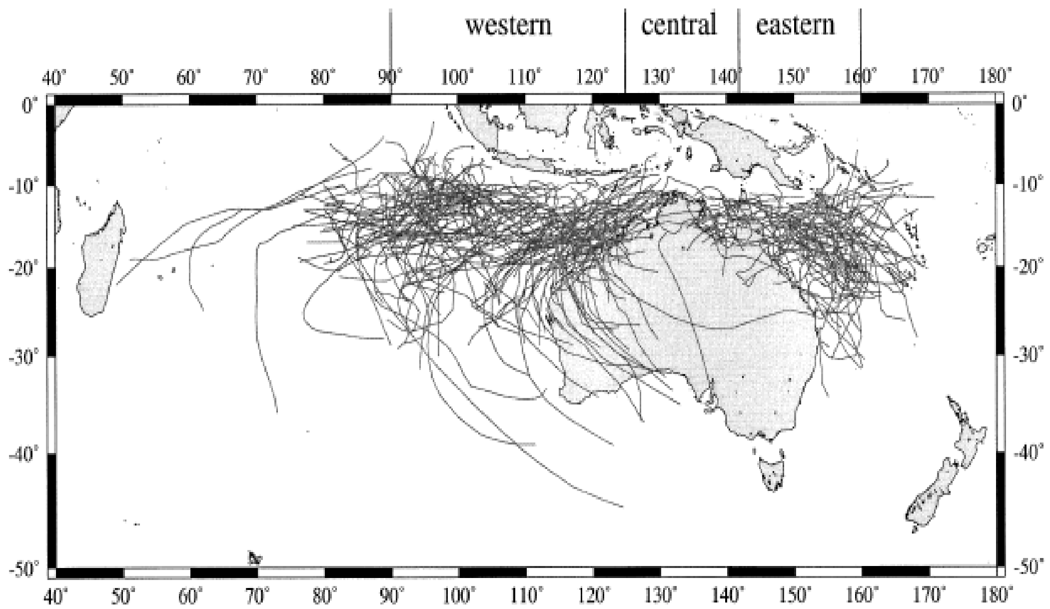

Fig.1: Tropical cyclone tracks in the Australian region (1958-2000)

3 - Cyclone trajectory forecast:

We work in the australian region (Fig.1). The used dataset provides the trajectories of past cyclones in the Australian region. Every cyclone of the basin is subdivided in several situations. Every situation provides the following parameters: zonal and meridional displacements, positions and time. Displacements for a time lag of 6h are used up to 24h in the past, which together with the positions and the year day, characterize region and season. Data are divided into:

- a dependent set for building model and
- an independent set for verification.

We tried to recover the situations continuation of the independent set by using elements of dependent set. The number of nearest neighbors that defines the analog has been fixed to 20 (K. Fraedrich, 2003). The pondered euclidean metric is used (James Mc Names, 1999):

$$D_{WE}^2(q, x_i) = \sum_{j=1}^{n_d} W_j^2 (q_j - x_{i,j})^2$$

Where

- $x_{i,j}$ is the *j*-th parameter of *i*-th tropical cyclone historic basis situation
- q_j the *j*-th considered situation parameter,
- W_j the metric weight,
- n_d the space dimension that is 19 in our case.

The used weights metrics are the metric weights obtained by Fraedrich (K. Fraedrich, 2003). We used the analogs selection method of Fraedrich et al as reference method.
The meridional and zonal cyclone positions are calculated by:
1. Arithmetic mean (Oliver Sievers, 2000);
2. Regression method with principal component analysis (M. Bessafi, 2002).

The distance between the real position and the calculated position has been calculated with the following formula (K. Fraedrich, 2003):

$$E_{\text{mod}\,ele} = 111 \arccos\,[\sin(y_0)\sin(y_f) + \cos(y_0)\cos(y_f)\cos(x_0 - x_f)]$$

Where:

- (x_0, y_0) is the observed position and
- (x_f, y_f) is the calculated position.

This distance defines the forecast error. Average forecast error over all independent data set have been calculated for a time lag of six hours for estimate the performance of analogs selection.

4 - RESULTS

Using two different calculation methods (Arithmetic mean, regression method with principal component analysis) for an analogs selection method, we obtained a same result. Fig.2 show calculated trajectories examples with two different analogs selection methods and the corresponding forecast error. The average forecast error is shown in Fig.3. We notice that, forecast depends on the tropical cyclone trajectory form to forecast; therefore the individual forecast error also depends on the tropical cyclone trajectory form. For the tropical cyclones having straight trajectory, the calculated trajectory brings closer the real cyclone trajectory and the analogs selection with PAT carries away before the reference analogs selection of the point of view forecast error.

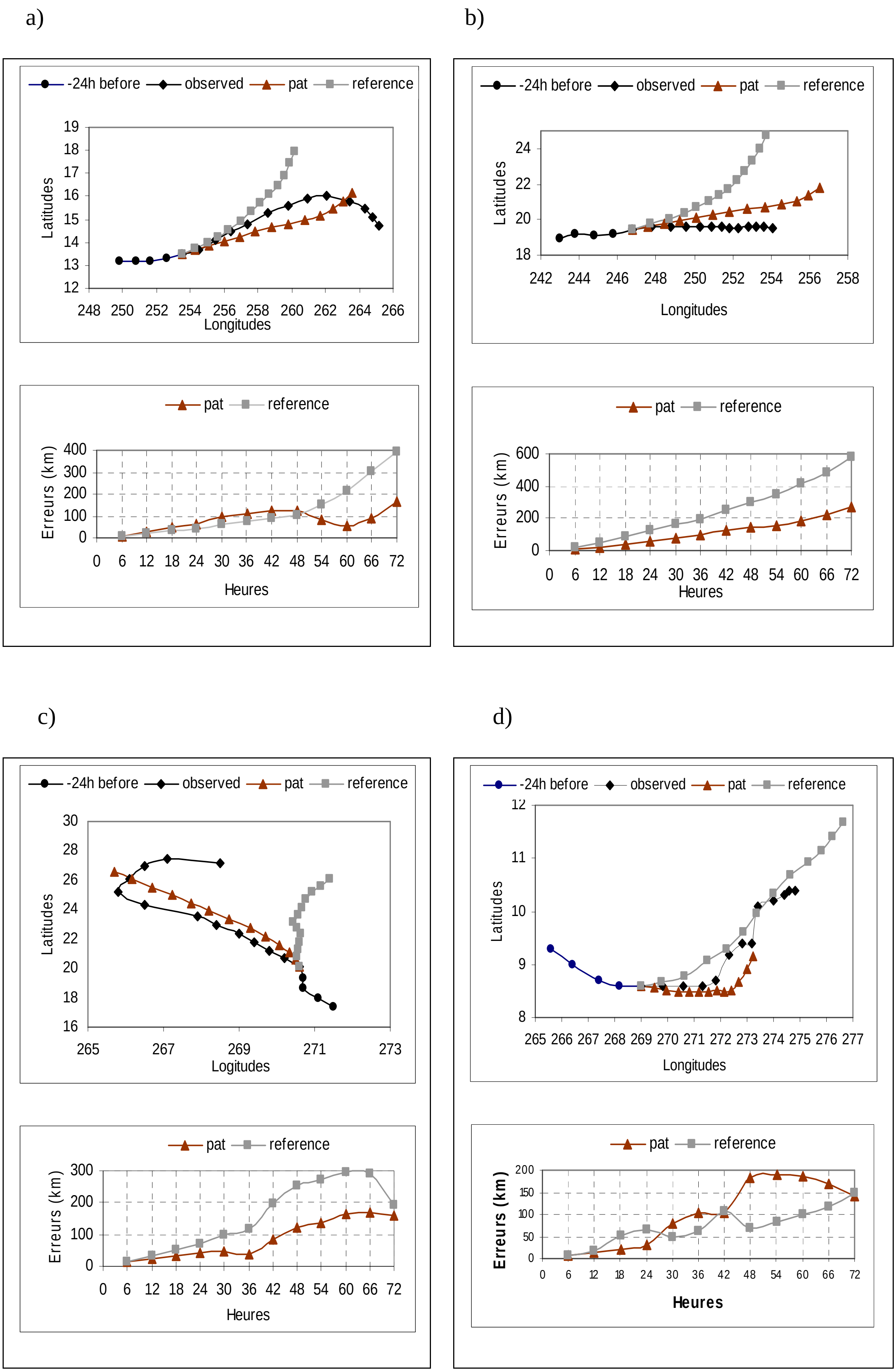

Fig.2. Examples predicted trajectories and the forecast error corresponding

For the tropical cyclones having any trajectory, the calculated trajectory departs of the real trajectory, in this case the reference selection method gives us good forecasting in relation to our method. These qualities of forecasting are explained by the difference between the two analogs selection methods. The selection method P.A.T. begins with a preliminary selection by reducing the data number in the dataset and continues the selection with the measure of distance between the considered situation and the other remaining situations. For the reference method, the measure of similarity begins with the statistical distance calculation between the considered situation and all situation of the dataset while only taking an analog situation for a tropical cyclone. On average, the obtained forecast error by using the P.A.T. analogs selection is lower to the forecast error with reference analogs selection. The difference increases from +36h.

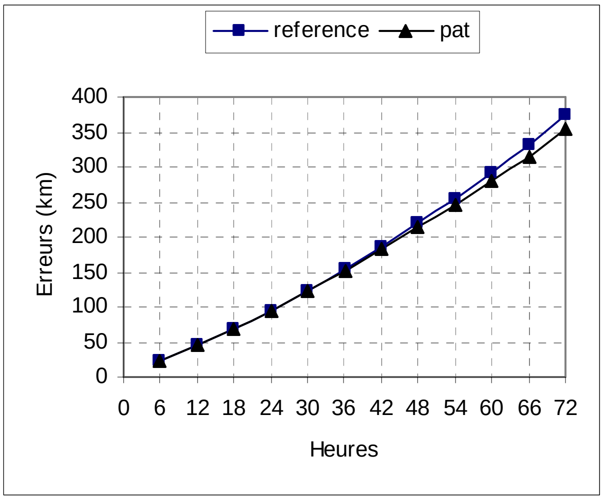

Fig.3: Mean forecast error

5 - CONCLUSION

The analogs selection methods P.A.T., compared with the reference analogs selection method, improve the forecasting of trajectory on average. The improvement is not even considerable. Combined to the calculation model arithmetic mean or to the calculation model regression with the principal component analysis, P.A.T is not sufficient for the forecast improvement. To reach the objective, best is to combine this selection method to another calculation model.

REFERENCES:

K. Fraedrich, C.C. Raible, and F. Sielmann (2003): Analog Ensemble Forecasts of Tropical Cyclone Tracks in the Australian Region. *Bull. Amer. Meteor. Soc. 18:1-11.*

James Mc Names (1999): Innovations in local modelling for time series prediction. *A Dissertation submitted to the Department of Electrical Engineering of the requirements for the degree of Philosophy.*

M.Bessafi, A. Lasserie-Bigorry, C.J.Neumann, F.Pignolet-Tardan, D.Payet and Lee-Ken (2002): Statistical Prediction of Tropical Cyclone Motion: An Analog-CIPER Approach. *Bull. Amer. Meteor. Soc., 17:821-831.*

Neumann (1981): Trends in forecasting the tracks of Atlantic tropical cyclones. *Bull. Amer. Meteor. Soc. 62:1473-1485.*

Olivier Sievers, K.Fraedrich, C.C. Raible (2000): Self-adapting analog ensemble predictions of tropical cyclone tracks. *Weather and Forecasting 15: 623-629.*