Trajectory Prediction by Functional Regression in Sobolev Space

K. Tastambekov[a], S. Puechmorel[a], D. Delahaye[a], C. Rabut[b]

[a]Ecole Nationale de L'Aviation Civile, Dept. Mathematiques et Informatique LMA,
7, Avenue Edouard Belin, 31055 Toulouse France
email: kairat@recherche.enac.fr, puechmor@recherche.enac.fr, delahaye@recherche.enac.fr

[b] Institut National Des Sciences Appliquees de Toulouse, Dept. Mathematiques,
135, Avenue de Rangueil - 31077 Toulouse Cedex 4 France,
email: christophe.rabut@insa-toulouse.fr

Abstract—**In this paper we consider the problem of short to mid-term aircraft trajectory prediction. That problem aims to predict collisions between aircraft in airspace. Our approach is based on local functional regression which consists in the three following stages : data pre-processing, localizing and regression. This algorithm has been successfully applied on aircraft trajectories between Toulouse and Paris.**

Keywords : Trajectory prediction, wavelet, functional regression

I. Introduction

A. Trajectory Prediction Metrics

Air traffic management research and development has developped substantial collection of decision support tools (DST) that provide automated conflict detection and resolution, trial planning, controller advisories for metering and sequencing, traffic load forecasting, weather impact assessment. Aircraft trajectory prediction algorithms([25]) are significant components of decision support tools (DST) in order to avoid collisions with others aircraft, arrival metering and other applications in air traffic management. A 4-dimensional (4D) trajectory prediction contains data specifying the predicted horizontal and vertical position of an aircraft over some time span into the future. The ability to accurately predict trajectories for different types of aircraft and under different flight conditions, that involves external actions (pilot, ATC) and atmospheric factors (wind, temperature), is an important factor in determining the accuracy and effectiveness of an air traffic management. Everyday, about 8000 aircrafts fly in the French airspace, inducing a huge amount of control workload (see [27]). Such workload, is then spread by the mean of the airspace sectoring. The airspace is divided into geometrical sectors, each of them being assigned to a controller team. When a conflict between two (or more) aircraft is detected, the controller changes their routes (heading, speed, altitude) in order to keep a minimum distance between them during the crossing. All flying aircrafts are then monitored during their navigation and so from the departure till the destination. When a controller observes its traffic on the radar screen, he tries to identify convergent aircraft which may be in conflict in a near future, in order to apply maneuvers that will separate them. The problem is to estimate where the aircraft will be located in a near future ($10-30$ minutes).

One of the issues in trajectory prediction is to measure how accurately a model will fit to a target trajectory. Unfortunately, many different metrics can be proposed each of them focusing on a specific aspect of accuracy. Most of the time, the proposed metrics fall into one of these categories [26], [27]:

- Time coincidence. The time difference between a predicted event and a real event is used as a measure of TP accuracy. Time coincidence is relevant in applications where synchronizing is important, like sequencing traffic, or when the DST uses time information to inform controller about the actions that have to be taken.
- Spatial coincidence. Similar to the previous one except that spatial distance at specified time (or more generally at events that can be predicted with the knowledge of aircraft positions up to a given time) between the model and the real aircraft is computed. Spatial coincidence can be refined by further splitting into altitude and horizontal error. Furthermore, for some applications, mainly conflict predictors and/or solvers, spatial difference is projected onto a vector normal to the real trajectory (cross-track error) and onto a vector tangent to the real trajectory (along-track error).
- 4D coincidence. Trajectories are considered as 4D curves, and distance between such curves is computed. Most of the metrics derived for spatial coincidence can be extended to the 4D setting, with the benefit of including a kind of time coincidence, thus generalizing in some sense the previous two aspects.
- Morphological similarity. Different in nature from the previous metrics, an intrinsic distance between trajecto-

ries considered as curves in a 3D space can be derived from Riemannian geometry. Since only the shape of the trajectory is taken into account, this metric is relevant mainly for trajectory design tools.

Except for the last one, all those basic metrics can be integrated along trajectories to produce a mean value indicator (the classical L^2 distance is for example obtained by integrating the standard spatial coincidence metric over time).

B. What is Functional Data Analysis?

Functional data analysis is an active branch of statistics in which relevant objects are mappings belonging to a well defined space, most of the time a Hilbert space. It has been proved very efficient for problems where preserving the functional nature of data is of great importance: curves classification, functional dependence learning and similar problems. The fundamental aims of functional data analysis are the same as those of conventional statistics ([12], [13], [14], [15], [16]) :

- to formulate the problem at hand in a way amenable to statistical thinking and analysis
- to develop ways of presenting the data that highlight interesting and important features,
- to investigate the variability as well as mean characteristics,
- to build a model for observed data , including those that allow for dependence of one observation or variable of another, etc.

In some cases, original observations are interpolated longitudinal data which are quantities observed as they evolve through time. In other situations, people prefer to use panel data, which are data from a number of observations over time of cross-sectional units like individuals, households, in our case aircraft trajectories.

We encounter functional data in many applications. In our problem, data consist of large number of independent numerical observations coming from ATC radars. Such data represent aircraft trajectories in the French airspace. In recent papers an increasing attention has been paid to linear functional regression, and some of its generalizations. In this setting, either a scalar value or a mapping (the response), possibly contaminated by an independent measure noise is modeled as being linearly dependent on a mapping (the predictor).

In this paper we will present an innovative approach based on functional regression for solving short to mid-term trajectory prediction (TP) problem. The first part of the paper presents our approach and give the associated mathematical modeling. The second part presents initial results on real data for the flight from Toulouse to Paris.

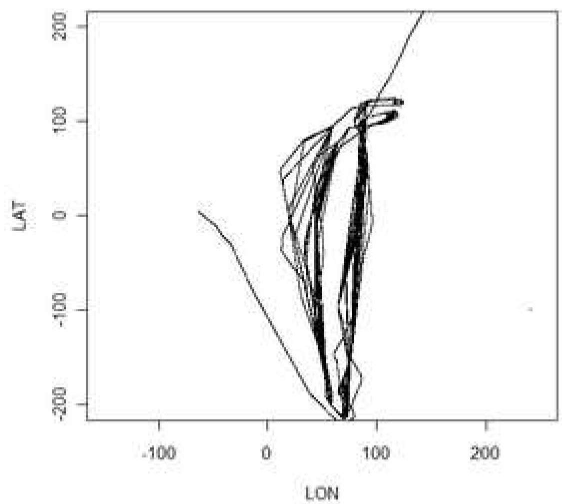

Figure 1. Example of aircraft trajectories data (Toulouse-Paris).

II. LOCAL FUNCTIONAL LINEAR REGRESSION BASED ON WEIGHTED DISTANCE-BASED REGRESSION

A. Problem Statement

The main idea of this paper is to solve the linear functional regression problem using data coming from radar tracker in order to build an enhanced trajectory prediction.An aircraft trajectory is by definition a mapping from a time interval $[a, b]$ to $\mathbb{R}^3$.

Radar data are disturbed by noise measurement and are not regularly sampled in time. These data are then processed by smoothing, approximation, and resampling. This is the first step of the algorithm.

A flight path is controlled by flight dynamics equations and knowing that the pilot's actions are simple, we may assume that aircraft trajectories are C^1, piecewise C^2 functions. Thus, we may assume that observed trajectories are samples of an Hilbert stochastic process (in fact it is even a Sobolev space valued process).

Let $\{X_n, Y_n\}_{n=1}^N$ be a sample of observations identically distributed coming from a Hilbert random processes X, Y defined on intervals τ_X, τ_Y. For our application, X_n represents all past trajectories connecting the same origin destination and Y_n the associated "future" reference trajectories. τ_X is the past time horizon on which we gather trajectory samples and τ_Y is the time horizon on which we do the prediction.

In order adjust our model, real trajectories will be used (from Toulouse to Paris; an example of radar tracks picture of such trajectories can be found on figure 1) for which a reference time will be considered. This time will artificially

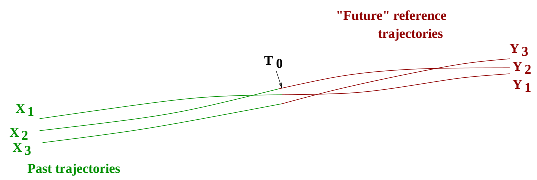

Figure 2. Framework used to adust the model. Our model is trained by the mean of a set of trajectories executed between two airport (in our case Toulouse-Paris). A reference time T_0 is then considered to separate the past and the future. The model uses both parts in order to adjust its parameters. After this training phase, the algorithm is tested by using only the samples coming from the past to predict the future position of the aircraft. Having the reference trajectory, it is very easy to measure the accuracy of the prediction.

separate the past and the "future" (which is known in this framework). Based on the previous position ($\{X_n\}$, our model will produce future predicted positions $\widehat{Y}(t)$ in order to minimize the errors between Y_n and $\widehat{Y}(t_n)$ see figure 2.

Let μ_X,μ_Y and B_X,B_Y be means and covariance kernels respectively (μ_X represents the mean trajectory on the set $\{X_n\}$ on τ_X and B_X the associated standard deviation.

The functional linear model has the general form [26]:

$$\widehat{Y}(t) = \widehat{f}(t) + \int_{\tau_X} \widehat{K}(t,s)X(s)ds$$

where $\widehat{f}(t)$ a smooth square integrable mapping which represents the mean of learning trajectories data set and $\widehat{K}(t,s)$ is a smooth square integrable matrix valued kernel. Next integral:

$$\int_{\tau_X} \widehat{K}(t,s)X(s)ds$$

is the deviation of the predicted trajectory from the mean $\widehat{f}(t)$.

The solution of the functional regression problem is the optimal couple $(\widehat{f},\widehat{K})$ that minimize the mean square error between Y and $\widehat{Y}$. Several expansions has been tried to model trajectories, but after some experiments we decide to use wevelet decomposition it is the which miniminize the prediction error. In our algorithm we use such wavelet in Sobolev space instead of the regular L_2([19], [20], [21]).

B. Wavelets in Sobolev space

At the beginning of 1980s, many scientists were already using "wavelets" as an alternative to traditional Fourier analysis. The word "wavelet" is used in mathematics to denote a kind of orthonormal bases in L_2 with remarkable approximation properties. The theory of wavelets was developed by Y.Meyer, I.Daubechies, S.Mallat and others in the end of 1980s, [1], [2], [3], [4], [5].

Definition 1: **Sobolev space.** Let $s \in \mathbb{N}$. The function $f \in L_2(\mathbb{R})$ belongs to the Sobolev space $W^s(\mathbb{R})$ [17], [18], if it is s-times weakly differentiable, and if $f^{(j)} \in L_2(\mathbb{R}), j = 1, 2, \ldots, s$. In a Sobolev space the norm is given by :

$$\|f\|^2_{W^s(\mathbb{R})} = \|f\|^2_{L_2(\mathbb{R})} + \|f^{(s)}\|^2_{L_2(\mathbb{R})}$$

Any $f \in L_2(\mathbb{R})$ can be represented as a series (convergent in $L_2(\mathbb{R})$) which is the definition of standard wavelet decomposition [6], [7], [8], [9] :

$$f(t) = \sum_{k\in\mathbb{Z}} c_k\varphi_k(t) + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} c_{jk}\psi_{jk}(t) = \sum_{i\in\mathbb{Z}} c_i\varphi_i(t)$$

where c_k, c_{jk} are some coefficients, and

$$\|f\|^2_{L_2(\mathbb{R})} = \sum_{k\in\mathbb{Z}} c_k^2 + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} c_{jk}^2.$$

It was shown in [1],[4], that a function f lies in $W^s(\mathbb{R})$ if and only if

$$\sum_{k\in\mathbb{Z}} c_k^2 + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} 2^{sj}c_{jk}^2 < +\infty.$$

Moreover, the discrete equivalent norm in Sobolev space $W^s(\mathbb{R})$ is

$$\|f\|^2_{W^s(\mathbb{R})} \approx \sum_{k\in\mathbb{Z}} c_k^2 + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} 2^{sj}c_{jk}^2$$

where s is the smoothness order of the Sobolev space.

For our application, trajectories will be modeled by such decomposition (wavelet in Sobolev space)[23].

III. Application to Trajectory Prediction

This section presents the application of the previous medeling to the trajectory prediction problem. The wavelet decomposition has been extended to the 3-dimensional case in order to fit the aircraft trajectories. Several kind of wavelets have been tried, but Daubechies 4 has produced more better results. The first part presents the solving of the function regression and the second one gives the results of our algorithm on real data set.

A. Solving the Functional Regression

One of the goal of our approach is to select the Hilbert random process from the set of trajectories connecting different points (airports in our application) [26]. As we said above, for French airspace only, there are about 8000 aircrafts every day connecting many Origins-Destination pairs. Without the knowledge of the origins and destination airports, trajectory prediction problem is much harder to address. It means that we have to extract a subset of trajectories connecting the same origins-destination pair. For our application, we have decided to keep the tracks of the flights from Toulouse to Paris for a given day (May 8 2009).

Let X_n (response Y_n) be the realization of predictor process X (response for process Y) corresponding to observation n

in the data set. Let M_n (response L_n) be the number of samples available for this observation (number of radar plots of a given trajectory) and let $X_{n,j}$, $j = 1, ..., M_n$ (response $Y_{n,j}$, $j = 1, ..., L_i$) be the actual samples along trajectories X_n (response Y_n) with corresponding sample times $\tau_{n,j}$ (response $\nu_{n,j}$). The number of samples M_n, L_n and the sampling times are assumed to be random variables independent from the processes X and Y. The first step towards solving the problem is to resample time intervals and to compute the missing data. To make an expansion of the predictor and response on respective basis and to compute coordinates at a given fixed time interval we can use several basis representations, such as wavelets, cubic splines, karhunen-loeve expansion, etc. An important step in the design of a linear smoother is the choice of weighted kernel and bandwidth. The problem has been addressed in the field of non parametric statistics and it is known that the kernel has less influence than the bandwidth. Let us now introduce some examples of kernels. The Epanechnikov kernel is defined by the following equation :

$$K_e(t) = \frac{3}{4}(1 - t^2)1_{[-1,1]}(t)$$

This kernel has some interesting optimality properties and is easy to compute. Another choice is the Gaussian kernel :

$$K_g(t) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2})$$

For very fast computation, it is still possible to use a uniform kernel :

$$K_u(t) = \frac{1}{2}1_{[-1,1]}(t)$$

Since the data set is usually large (around 1000 trajectories sampled at 10s), a compactly supported kernel in the local linear smoother allows a reduced computational load and a complexity mostly independent of the number of samples of the trajectory. The Gaussian kernel is not compactly supported, but decreases very fast at infinity so that practically it can be set to 0 outside a compact interval. And as a simple particular case we can define the simple window function:

$$W_d(X_i, X_j) = \begin{cases} 1, & \text{if } d(X_i, X_j) \leq d \\ 0, & \text{if } d(X_i, X_j) > d \end{cases}$$

Where $d(X_i, X_j) = ||X_i - X_j||$ is the distance between two trajectories, X_i and X_j respectively.

Finding the right predictor is a critical task in applying functional regression. For trajectory prediction purpose, it is natural to consider a part of the observed trajectory as the learning set, and a part of the future trajectory as target. The learning database has thus been chosen by selecting homogeneous 256 radar plots (per trajectory) from a day of traffic. Those learning trajectories have been divided into two parts with 128 plots each, corresponding to the past and the “future” .

One simple approach to estimate $f(t)$ is to center the observed Y_n and the given X_n by subtracting their sample average functions $\overline{X}$ and $\overline{Y}$. Here and later we consider centralized X_n, Y_n and the model becomes :

$$\widehat{Y}(t) = \int_{\tau_x} \widehat{K}(s,t)X(s)ds$$

Then, the regression problem becomes to find an optimal $\widehat{K}(t,s)$ minimizing following expression:

$$\sum_{n=1}^{N} U_d ||Y_n(t) - \int K(t,s)X_n(s)ds||^2_{(W^1)^3}$$

where U_d is one of the weighted window kernel function described above([24]).

The kernel $\widehat{K}(t,s)$, $X_k(s)$ and $Y_k(t)$ can be expressed using the wavelet basis $(\phi_i)_{i\in\mathbb{N}}$, $(\psi_i)_{i\in\mathbb{N}}$ as:

$$X_n(s) = \sum_j a_j^n \phi_j(s), \quad Y_n(t) = \sum_i b_i^n \psi_i(t)$$

$$K(t,s) = \sum_i \sum_j K_{ij}\phi_j(s)\psi_i(t)$$

Where ϕ_i and ψ_j are wavelet basis functions, respectively to the τ_X and τ_Y time intervals. Using the orthonormality of basis, the regression problem becomes to find the minimum of the sum:

$$\min_{K_{ij}} \sum_{n=1}^{N} U_d ||Y_n(t) - \int K(t,s)X_n(s)ds||^2_{(W^1)^3} =$$

$$\min_{K_{ij}} \sum_{n=1}^{N} U_d \sum_{i=1}^{P} (b_{in} - \sum_{j=1}^{Q} a_{jn}K_{ij})^2$$

Here the expansions were truncated to a fixed rank. Then $\widehat{f}(t)$ can be founded by the next formula:

$$\boxed{\widehat{f}(t) = \overline{Y}(t) - \int_{\tau_X} \widehat{K}(t,s)\overline{X}(s)ds}$$

$$\boxed{= \sum_{i=1}^{P}(\bar{b}_i - \sum_{j=1}^{Q} \widehat{K}_{ij}\overline{a}_j)\psi_i(t)}$$

which is nothing but a linear mean square problem that can be solved with the help of normal equations or using SVD.

B. Application to real data

For the first test we use one day air traffic between Toulouse and Paris airports. There are 52 aircraft trajectories flying in both directions and one trajectory will be used as “real” trajectory for which prediction accuracy will be evaluated. Each of the 52 trajectory will be selected as “real” to build a cross validation procedure in order to improve robustness. The least mean square problem was solved using window kernel function. Figure 3 shows the first 42 minutes of real aircraft trajectories started from Toulouse-Blagnac airport. And the second figure 4 consists of two parts. The first part shows 21 minutes of real trajectories and the second part is the predicted

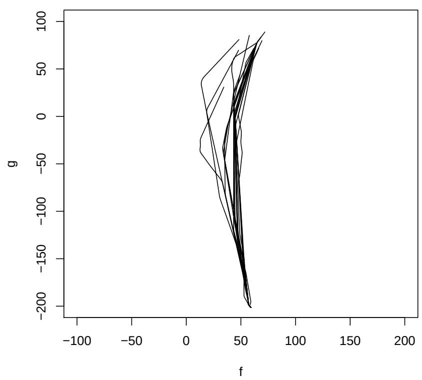

Figure 3. Real trajectories

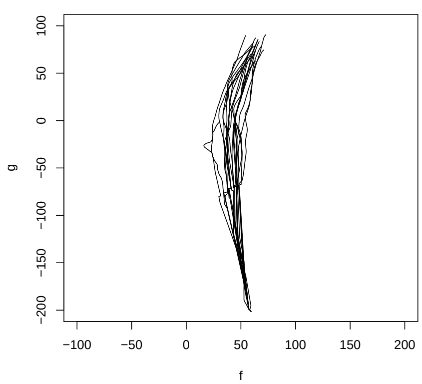

Figure 4. Predicted trajectories

trajectories. The shift is located at the middle of trajectories (50 on the Y axis).

A cross validation procedure was designed as follows:

- Pick a trajectory and remove it from the leaning database.
- Compute prediction error on this trajectory using the others as learning set.
- Do the same with couples of trajectories removed.
- The relative prediction error computed for each trajectory is given by:

$$\frac{||\widehat{X}_0 - X_0||^2}{\frac{1}{N}\sum_{n=1}^{N}||X_i - \overline{X}||^2}$$

where X_0 is a real trajectory, $\widehat{X}_0$ is predicted and $\{X_n\}_1^N$ is the set of learning trajectories.

Results are summarized in the following table for which each cell give the prediction relative error for a given "real" trajectory :

Relative errors			
0.1622	0.2269	0.2020	0.1848
0.1719	0.1994	0.1935	0.1940
0.2176	0.1653	0.1812	0.1636
0.1940	0.1553	0.2977	0.2090
0.2168	0.1686	0.1691	0.9003
0.2214	0.2393	0.3565	0.2039
0.2020	0.1681	0.1498	0.1841
0.3023	0.2183	0.1734	0.2410
0.1679	0.1622	0.2502	0.2880
0.1620	0.1847	0.2920	0.1911
0.2901	0.1968	0.1960	0.1788
0.2345	0.1969	0.1861	0.2205
0.2458	0.2873	0.1636	0.1767

As we can see, the results produced by this new approach are very good for a prediction horizon of 20 minutes.

IV. Conclusion and future work

The functional data analysis has been applied in order to build a new algorithm for aircraft trajectory prediction problem. This approach uses only previous radar tracks for a given origins destination pair. A learning process enable the adjustment of parameters. This model is based on localization of functional linear regression model using wavelets in Sobolev space. This method produces efficient results with hight robustness.

In a next step, a larger data base will be used for the same origins destination pair (Toulouse-Paris), by taking several years of data. Then, we will try it on some other origins destination pair (midge range and long range). Finally, we will determine the limit of this approach by increasing the prediction time interval.

References

[1] I. Daubechies, Ten lecture on wavelets, *SIAM*, 1992.

[2] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, *IEEE Trans. Inf. Th.*, 1990.

[3] C.K. Chui, An Introuduction to wavelets, *Xian Jiaotong Univ. Press*, 1994.

[4] Y. Meyer, Wavelets and Operators, *Cambridge Univ. Press*, 1992

[5] Y. Meyer, Wavelets: Algorithms and Applications, *SIAM*, 1993.

[6] W. Hardle, D. Picard, G. Kerkyacharian and A. Tsybakov, Wavelets, Approximation and Statistical Applications, *Un premier resultat du seminaire Paris-Berlin*, 1997.

[7] S.A. Dianat, R. Rao, Wavelet transforms: Theory and applications, *Proc. SPIE 12th Annu. Int. Symp. Aerosense*, 1998.

[8] M.A. Cody, The fast wavelet transform, *Dr. Dobb's J.*, 1992.

[9] Rong-Qing Jia, Jianzhong Wang and Ding-Xuan Zhou, Compactly supported wavelet bases for Sobolev spaces, *Appl. Comput. Harmon. Anal. 15*, 2003.

[10] I.T. Joliffe, Principal component analysis, *Springer-Verlag*, 1986.

[11] J.E. Jackson, A user's guide to principal components, *John Wiley*, 1991.

[12] J. Ramsay, B. Silverman, Functional Data Analysis, *Springer series in statistics*, 1997.

[13] J. Ramsay, B. Silverman, Functional Data Analysis (second edition), *Springer series in statistics*, 2005.

[14] F. Ferraty, Ph. Vieu, Nonparametric functional data analysis, *Springer series in statistics*, 2006.

[15] J. Ramsay, G. Hooker, S. Graves, Functional data analysis with R and MATLAB, *Springer series in statistics*, 2009.

[16] J. Ramsay, B. Silverman, Applied functional data analysis: Methods and case studies, *Springer series in statistics*, 2002.

[17] R. Adams, Sobolev spaces, *Academic press*, 1975.

[18] I. Gihman and A. Skorohod, Introduction to the theory of random processes, *"Nauka"*, 1965

[19] Maya R. Gupta, Nathaniel P. Jacobson, Wavelet principal component analysis and its application to hyperspectral images, *IEEE*, 2007.

[20] Feng Lu, Zhaoxia Yang and Yuesheng Li, Wavelets approach in choosing adaptive regularization parameter, *Lecture Notes in Computer Science*, 2007.

[21] R. A. Devore, Fast wavelet techniques for near-optimal image processing, *IEEE Military Communications Conference Record*, 1992.

[22] Bhavik R. Bakshi, Multiscale PCA with application to multivariate statistical process monitoring, *AIChE Journal*, 1998.

[23] R. Suyundykov, S. Puechmorel, L. Ferre, Multivariate Functional Data Classification using wavelets in Sobolev space, *Ecole Nationale de l'Aviation Civile*, 2010

[24] Eva Boj, Pedro Delicado, Josep Fortiana, Local Linear Functional Regression based on Weighted Distance-Based Regression, *Universitat de Barselona, Universitat Politecnica de Catalunya*, 2008

[25] Chester Gong, Dave McNally, A methodology for automated trajectory prediction analysis, *San Jose State University, NASA Ames Research Center, Moffett Field, CA, 94035, NASA Ames Research Center, Moffett Field, CA, 94035*, 2004

[26] D. Delahaye, S. Puechmorel, L. Boussouf, Trajectory prediction: the functional regression approach, *Toulouse, France*

[27] S. Mondoloni, S. Swiertstra and M. Paglione, Assessing trajectory prediction performance metrics definition, 24-th Digital and Avionics Systems Conference, 2005.