Machine Learning techniques applied to FDM

Hélder Mendes

FDM Expert

Nicolas Chrysanthos

Ph.D. Student

EOFDM Conference 2013

/ 23 January 2013 /

Contents

Machine Learning (ML)

➔ « development of algorithms which take **data as input** in order to **detect patterns**, **make predictions**, or more generally **gain insights** and knowledge about the mechanism generating the data ».

Machine Learning (ML)

- Recent field of study,

- Has seen tremendous development these recent years,

- Applied to many fields of application:
  - Computer Vision,
  - Finance,
  - Economy,
  - Biology,
  - Marketing.

Machine Learning (ML)

- Essentially two types of Machine Learning:

	SUPERVISED	UNSUPERVISED
Training phase?	YES	NO
Labeled data?	YES	NO
Goal	Make predictions	Discover structure

Added value of using Machine Learning for FDM?

1. Contrary to classical analysis, finds **unexpected** problems!
2. Extract more **value** from the data,
3. Gain **better understanding** of the mechanisms behind the data.

SAFRAN Sagem

The No Free Lunch theorem

- You *can't just throw all your data* into a ML algorithm and expect it to give you **relevant** results…field experts must be involved.
- The goal is to find a **compromise…**

Not enough knowledge!

Too much knowledge!

- Less input from field expert,
- More data,
- Finds more unexpected results,
- Needs more computational ressources
- Higher false positive ratio,
- In the limit, amounts to measuring noise…

- More input from field expert,
- Less data,
- Finds less surprising results,
- Needs less computational ressources
- **Lower false positive ratio**,
- In the limit, amounts to classical deterministic event detection…

SAFRAN Sagem

Machine Learning: good practices

- Machine Learning is **teamwork**:
  - One **Machine Learning practitioner**, with Maths/Computer Science background,
  - One or more **Field Experts**, in this case pilots, flight data analysts and FDM experts.

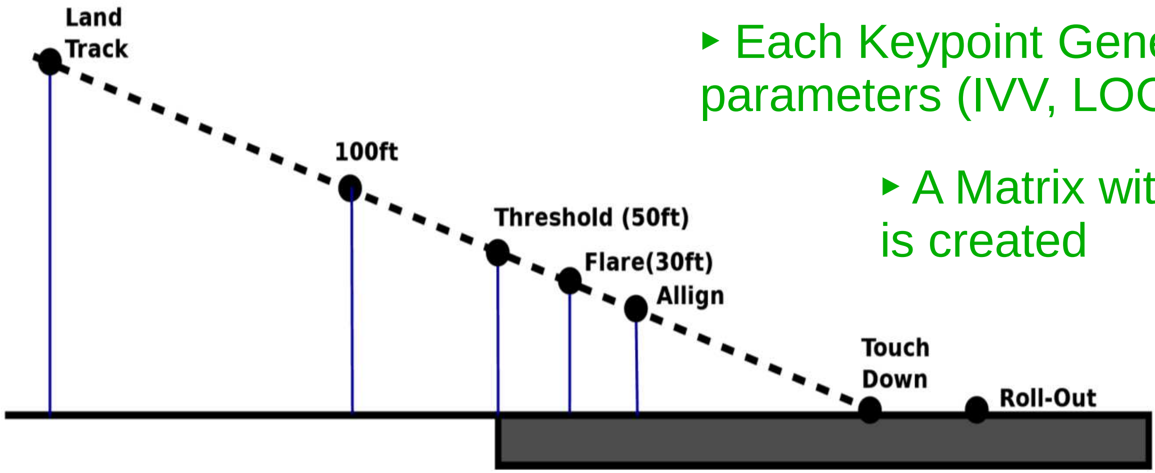
- The Machine Learning practitioner…
  - Stands at the interface between field experts and mathematical methods,
  - Has responsibility to provide insights to the field experts about
    - How algorithms work,
    - What they can do,
    - What they **can't** do
  - Receive feedback from field experts to gain a good comprehension of the data.

- A good way to develop algorithms is to work in an **iterative** way.

Contents

Autoland Flight File Construction

- Procedure based on keypoints data extraction during landing
  - Each Keypoint Generates around 30 snapshot parameters (IVV, LOCDEV, PITCH,VRTG, etc...)
    - A Matrix with 279 dimensions per flight is created



(Sample)

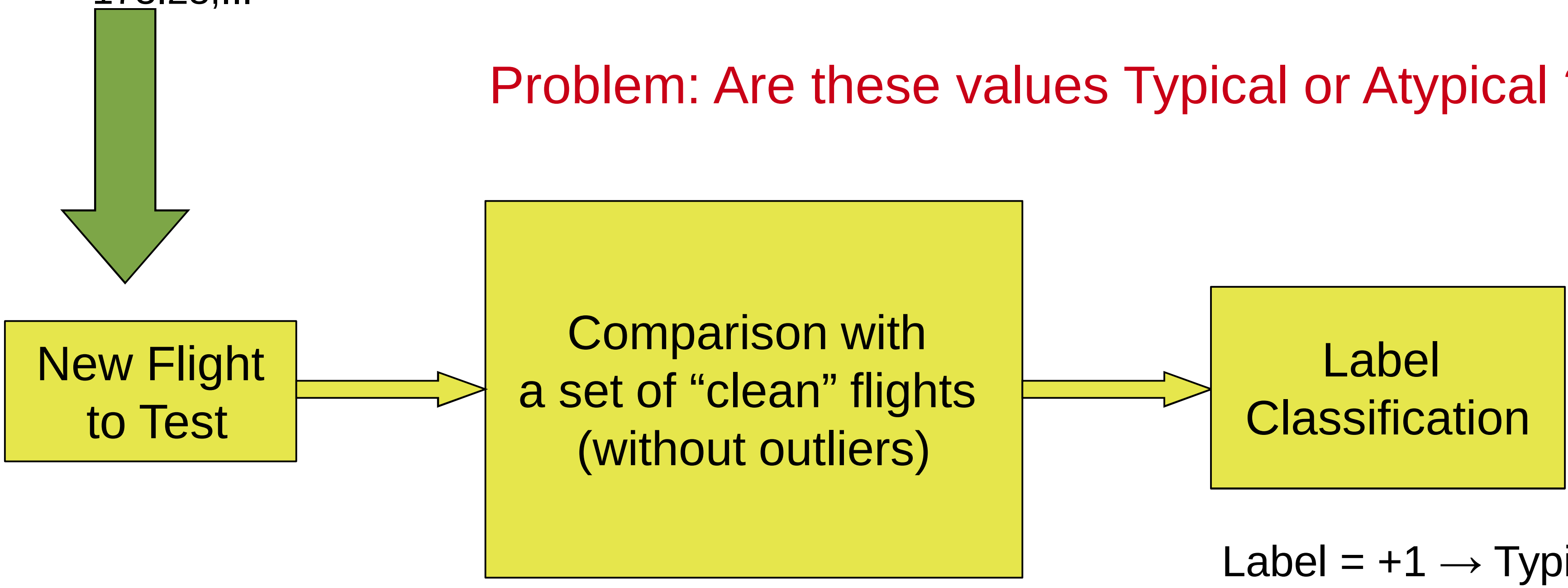
	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	DESTINATION	RUNWAY_LD	DATE_TO	HEIGHTD	VAPR	ROD_LT	ROLL_LT	LOC1_LT	LOC2_LT	GS1_LT	GS2_LT	HDG_LT	PITCH_LT	IAS_LT	ALT_LT	TLA1_LT	TLA2_LT	N11_LT	N12_LT	N21_LT
2	LIS	21	11/02/2011	0	144.75	-12.8	0.35	1.56	0.98	5.86	5.08	205.05	2.46	143.25	524	3	3	62.38	62.5	88.88
3	ORY	6	16/02/2011	20	130.5	-12.27	1.41	-0.39	0	-8.98	-12.11	64.42	2.46	131.88	1156	3	3	46.5	46.38	82.13
4	MXP	35L	01/02/2011	0	139	-13.87	0.7	-1.37	-0.98	-1.95	-1.17	347.43	2.11	140.25	872	3	3	47.5	47.13	83.63
5	LIS	21	07/02/2011	0	125.25	-12.53	-0.35	1.17	0.98	-2.73	-3.52	206.19	2.46	125.38	364	3	3	47.5	47.38	84.75
6	ORY	6	16/02/2011	20	124.25	-11.73	-0.35	-0.59	-0.2	-13.67	-14.06	65.21	3.16	123.63	1144	3	3	46.25	46.88	82.75
7	HEL	04L	24/02/2011	134	126.25	-12.8	0	-1.17	-1.56	-11.33	-12.11	44.21	1.76	130.38	-96	3	3	44.13	44.13	80.75
8	BLQ	12	01/02/2011	0	125.5	-12.8	0	-0.59	-0.39	16.41	16.41	114.96	2.46	127.5	252	3	3	41.88	42.38	82
9	BLQ	12	01/02/2011	0	122	-11.47	0	0.39	0.39	3.91	3.91	114.79	2.81	122.88	292	3	3	44	43.88	82.5
10	VCE	04R	08/02/2011	0	121.5	-11.47	1.05	0	0.39	5.08	5.08	39.02	1.41	127.63	180	3	3	44.25	43.88	82.13
11	MUC	08L	09/02/2011	0	124	-10.93	0.7	0.39	0.78	-2.73	-3.13	81.91	2.46	125.88	1624	3	3	49.63	48.63	82.75
12	LIS	21	11/02/2011	0	124	-10.67	-0.35	-0.59	-0.59	-3.13	-3.52	204.79	2.81	122.88	552	3	3	54.5	55.13	87
13	OPO	17	12/02/2011	108	130	-10.67	-0.7	-0.2	-0.2	-2.34	-3.13	171.04	2.11	130.13	236	3	3	49.5	49.63	84.25

TAP PORTUGAL

NEW Autoland Flight Values created:

TJF,321,1988,LIS,03,OPO,17,20/08/12,108,130.00,-10.93,0.00,-0.78,-0.39,3.52,5.08,173.23,...

Problem: Are these values Typical or Atypical ?



Label = +1 → Typical
Label = -1 → Atypical

Possible Solutions

Input Space classification procedure

- Creation of a dataset without outliers (clean all cases with values over ± 3 σ) = Training Dataset

- Determination of the average (μ) and Standard deviation (σ) of each dimension (spreadsheet column) of the Training Dataset

- Each new value is Normalised → remove the average and the scaling of each parameter

$$z_i = \frac{x_i - \mu}{\sigma}$$

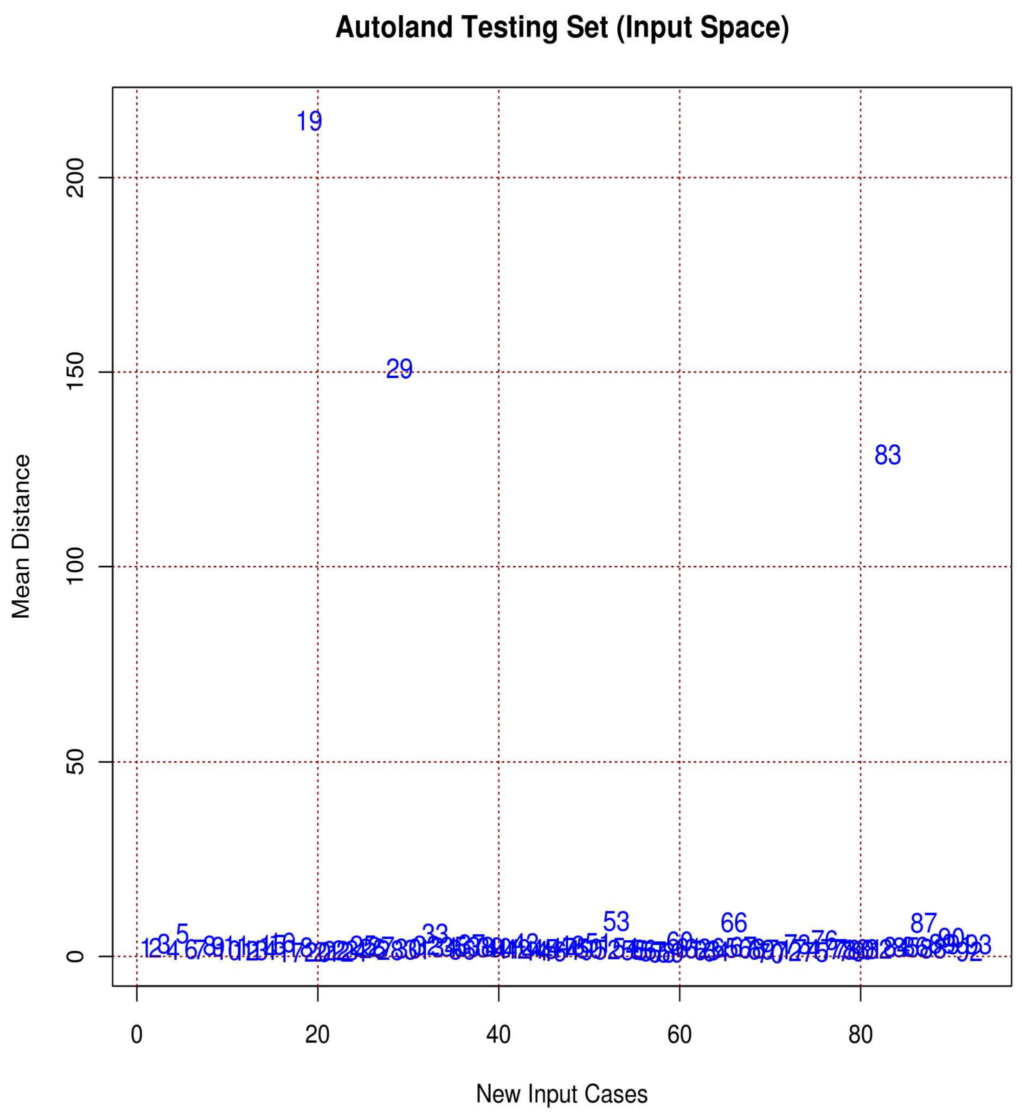
- Calculation of the average distance

$$d_{med} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} z_i^2}$$

- Each flight produces one average distance, which is plotted for a fast comparison of several flights....

Average distances for 93 new flights (Input Space)

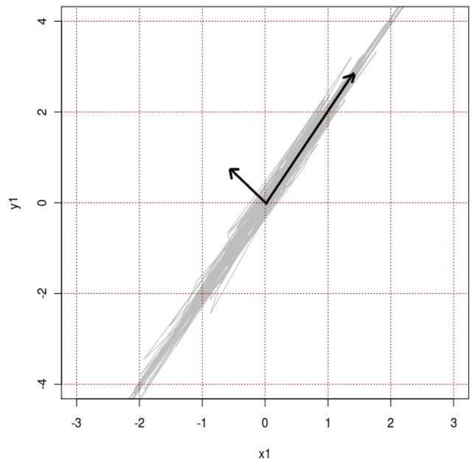
TAP PORTUGAL



A STAR ALLIANCE MEMBER

TAP PORTUGAL

PCA – Principal Component Analysis



- Decomposition of the Data Matrix by its most important directions (Principal Components).

- Each new dimension is obtained by a Linear Combination of the original dimensions (ALT,CAS,PITCH,...).

- The new dimensions are ordered by decreasing order of variance and are uncorrelated with each other.

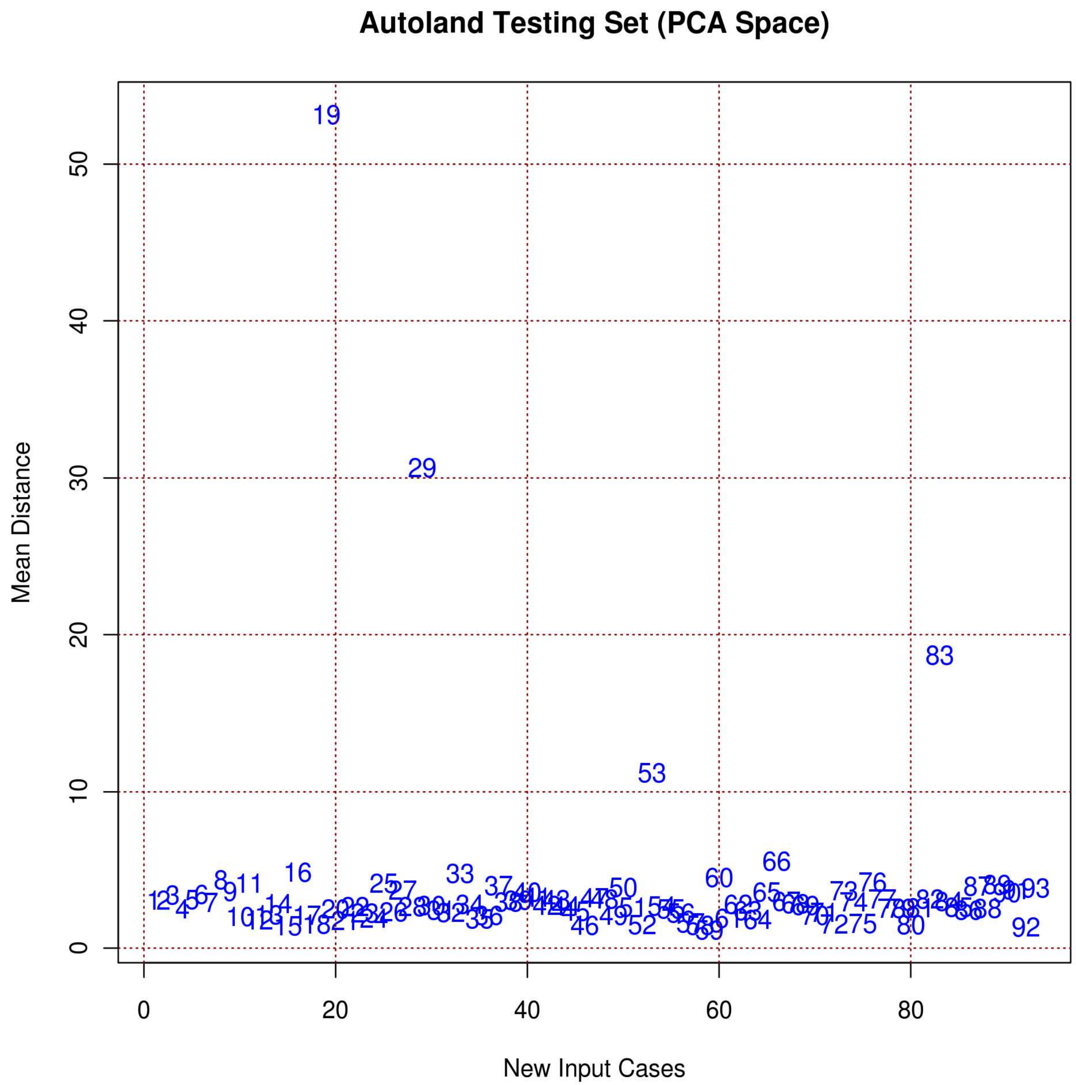
- A new flight is projected into the Principal Components space

- The Mean Distance is calculated as in the previous case.

- Plotting the Mean Distance for the same 93 flights using PCA....

Average distances for 93 new flights (PCA Space)

TAP PORTUGAL



A STAR ALLIANCE MEMBER

PCA and Input Space Summary

TAP PORTUGAL

- Very useful for a fast evaluation of a set of flights

- Both methods apply to numerical values with Gaussian distributions

- Still there is data that cannot be tested with the two methods presented before:

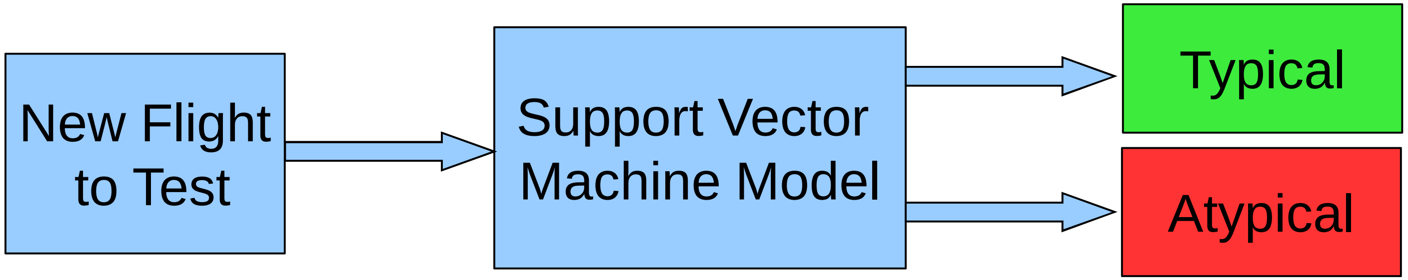
- Binary data (Landing Gear Sqwat Switches, Engine Valves, etc...)
- Text data (Tail Number, Departure Airport, etc...)
- Bit sequences (Autopilot Modes, etc...)

- New techniques based on “Statistical Learning Theory” can be used to overcome these limitations

Support Vector Machine (SVM)

TAP PORTUGAL

▸New approach for classification → Machine Learning Methods



▸Support Vector Machine – It's one algorithm that can "learn" from previous data generated and classify new inputs (flights)

▸This algorithm is capable of dealing with binary and text data

SVM Model Construction

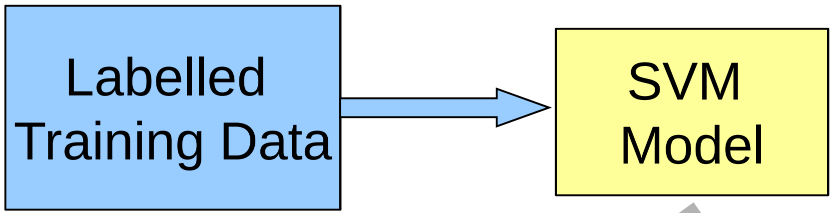
TAP PORTUGAL

- The data structure is composed by data (x_{ij}) and its corresponding label (yi)

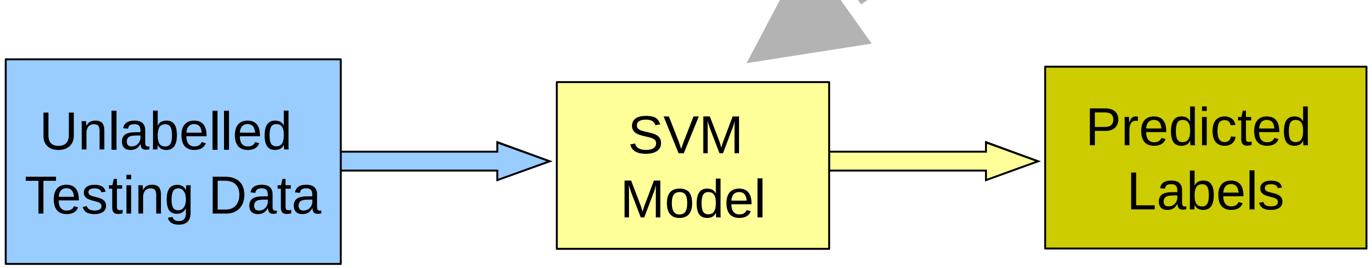
$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} & , y_1 \\ x_{21} & x_{22} & \dots & x_{2m} & , y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} & , y_n \end{bmatrix}$$

- Supervised Learning Method → It relies on a Training Database with data and its corresponding known label

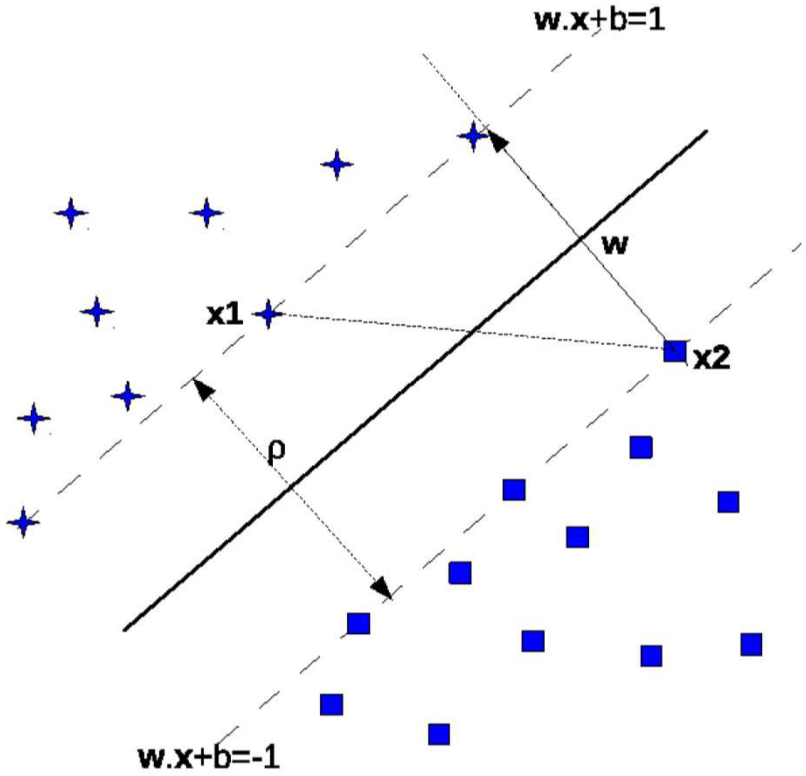
- Step #1 – Training Process



- Step #2 – Prediction for new inputs



The key ideas for SVM classification are:



- Determination of the best separation line between the two clusters of data

The best line is the one that maximizes the margin (ρ)

- The separation becomes possible using the vectors that lie on the lines...

$$\mathbf{w}^{\mathbf{T}}\mathbf{x_i} + b = +1 \qquad \mathbf{w}^{\mathbf{T}}\mathbf{x_i} + b = -1$$

- ... which are called the support vectors

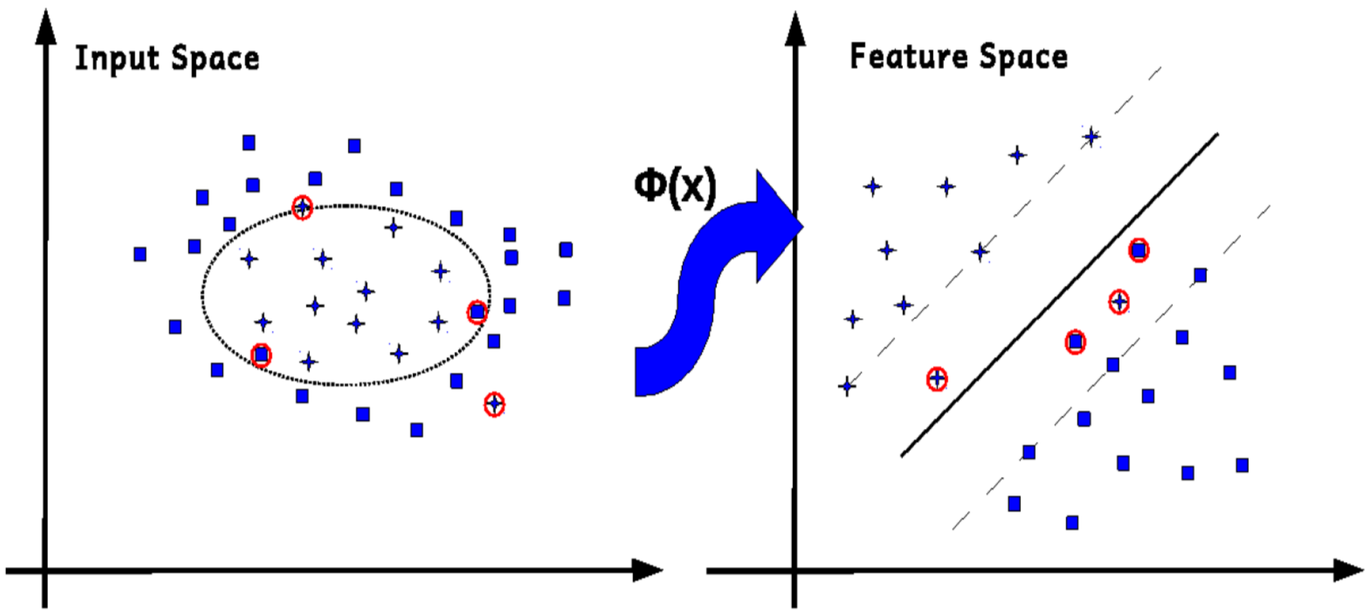
- Each new entry (x) is evaluated according to a decision function....

$$f(\mathbf{x}) = sign(\mathbf{w}^{\mathbf{T}}\mathbf{x} + b)$$

- ... That provides the corresponding Label (+1 / -1)

Kernel Machine

- Real Data is affected by noise, outliers and is not linearly separable normally
- A mapping function Φ(x) is used to transpose the data to a space where the SVM can be used directly (linear separation)
- The SVM classification is operated in the Feature Space



Automatic Classification Process

TAP PORTUGAL

Inputs for Reliable Training Database construction	Use of the appropriate software to build the model (LIBSVM)	New entries (flights) are classified according to the previous model
Input Space PCA Space Pilot Reports Other relevant input	Construction of a Robust SVM Model	Classification Of new flights

Sofware packages used for this process:

- AGS (Analysis Ground Station) from SAGEM Défense et Sécurité - Safran Group
- LIBSVM – SVM integrated software from National Taiwan University (GNU License)
- R statistics Environment from University of Auckland (GNU License)

Contents

Flight data provided courtesy of

Detecting atypical flights

- « When considering a population of flights, an **atypical flight** is a flight which is in a sense different from the majority of the other flights ».

- Atypical flights may present **operational or safety issues** and thus need to be studied by an FDM expert!

Difference with classical procedures

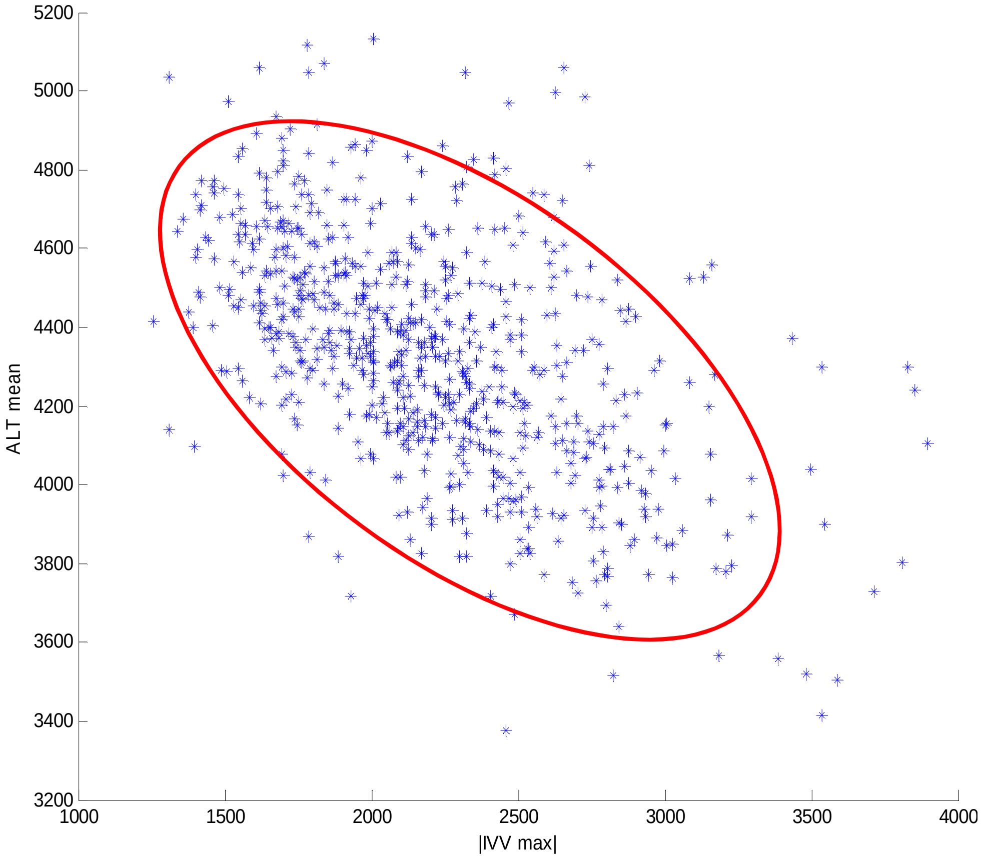
- Main difference with classical procedures :
  - In classical procedures, flights are studied **one at a time**, isolated.
  - In ML methods, flights are studied **with respect** to a population of other flights.

- Some flights may be overlooked by classical procedures but detected by ML methods…
  - And vice-versa !

- **ML techniques are complementary with classical detection procedures**.

2 parameters Example

- Example study with 2 « parameters »:
  - Data from **721 flights from Porto to Orly**,
  - Plotted the mean ALT and max IVV for each flight during **approach phase**…
  - **Each point represents a flight.**
  - In this simple example, anyone could easily spot which are the atypical flights.



- ML algorithms do the same, but…
  - In **any number of dimensions**,
  - For any number of flights,
  - And provide **quantitative results**

SAFRAN Sagem

Choosing a dataset

- Choosing the dataset is the most important step!
  - Garbage in garbage out…

- There are two decisions:
  1. What population?
     - All flights that land on Orly? Or all Porto-Orly flights ?
  2. What parameters?
     - Study trajectory, mechanical parameters? Or engine parameters? Or both ?

Choosing a dataset : what population?

- Some *rules of thumb*: for « kernel methods » (widely used techniques such as SVM)

  - Increasing number of flights...
    - Greatly increases computation time,
    - Mildly increases precision of results.
    - Example: with 10 times the number of flights, the computation time is multiplied by 100, whereas the precision is at best multiplied by ≈ 3.16

  - Good care must be taken when « sampling » the dataset from different sources:
    - With 5 Porto-Orly flights in the middle of 700 Berlin-Orly flights, these 5 flights will be in minority and thus are likely to be considered atypical.

Choosing a dataset : what parameters?

➔ It is best to study *groups of parameters* rather than all parameters at once.

- Make several studies:
  - One for engine parameters,
  - One for trajectory parameters,
  - Etc.
- Each study gives a different *point of view.*

➔ ***Divide and conquer!***

Choosing a dataset : what parameters?

➔ Some other advices:

- In the field of FDM there are many types of parameters, which must not be treated the same way:
  1. Continuous parameters, such as Altitude, N1,
  2. Angular parameters, such as Heading, Roll, Pitch,
  3. Unordered discrete parameters, such as Autopilot mode (OFF/ FlightDirector/ Command)
  4. Ordered discrete parameters, such as FLAP (0, 5, 10, 15, 25, 30)

- Parameters evolve in time,
  - We can take *features* such as mean, max, min etc.
  - Or we can take *snapshots*.

Example study

➔ We have studied 721 flights from Porto to Orly 26, from **Transavia France**, same aircraft, **from approach till touch down** (10000 feet to 0).

➔ 14 parameters, in first pass we studied the **mean** of all parameters

- Position (latitude, longitude)
- Altitudes, heading,
- Roll, Pitch
- Accelerations (angular and along axes)
- Speeds (vertical and longitudinal)
- N1.

A word on the mathematics

➔ We have used our own detection method:

- Based on *Kernel Entropy Component Analysis*, a recent (2010) dimensionality reduction technique,
- Strong theoretical guarantees from nonparametric statistics,
- Better results than state of the art One-Class SVM,
- Very robust even with highly « polluted » dataset.

Results

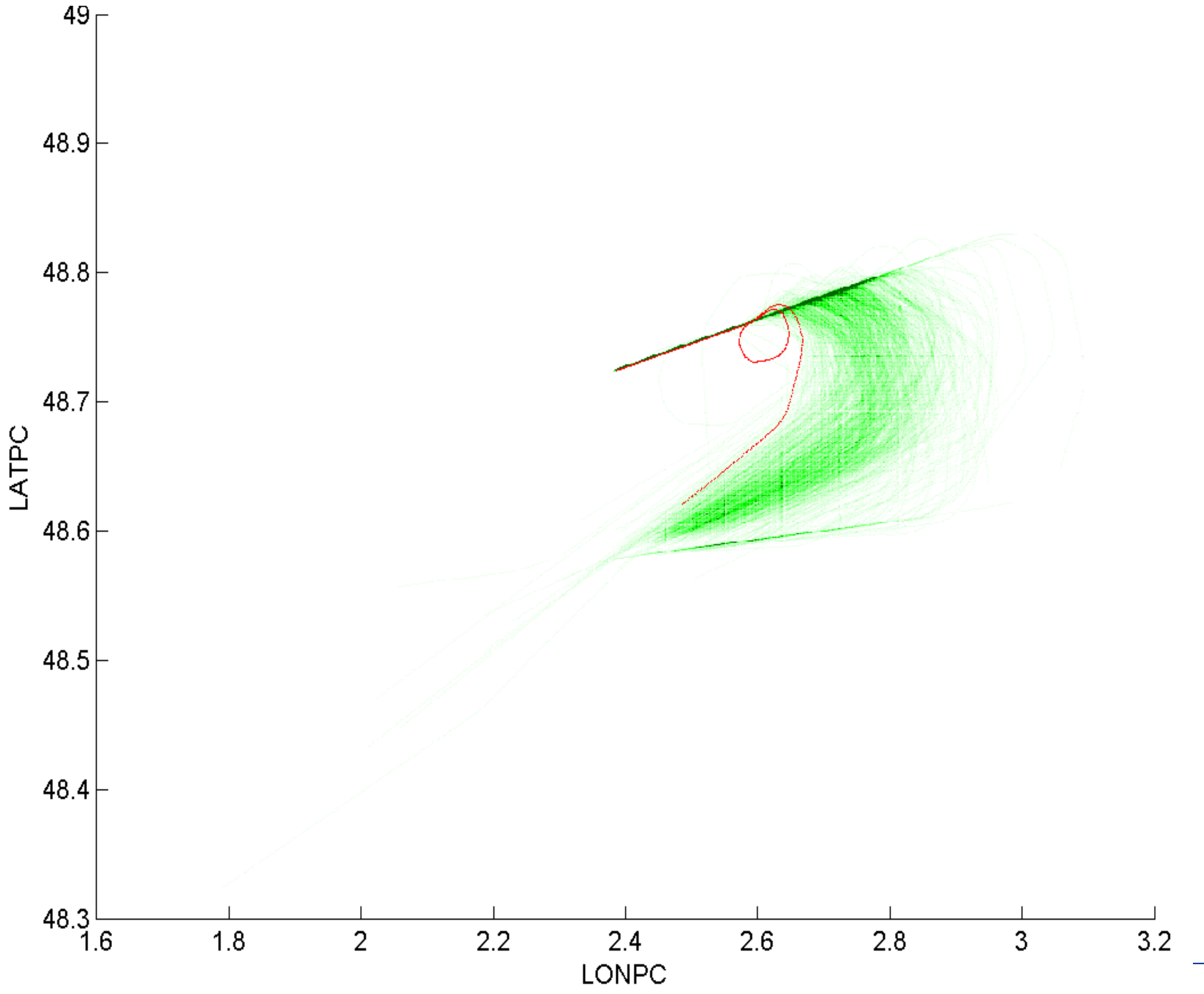
➔ Of the 721 flights, **35 are detected**: each flight is given a **pvalue:**

- « A pvalue is the probability that, **under normal conditions**, a flight at least as extreme could occur by chance alone »

- A flight with a pvalue<0.01 is considered very likely to be an atypical flight,
- A flight with a pvalue<0.001 is considered extremely likely to be an atypical flight.

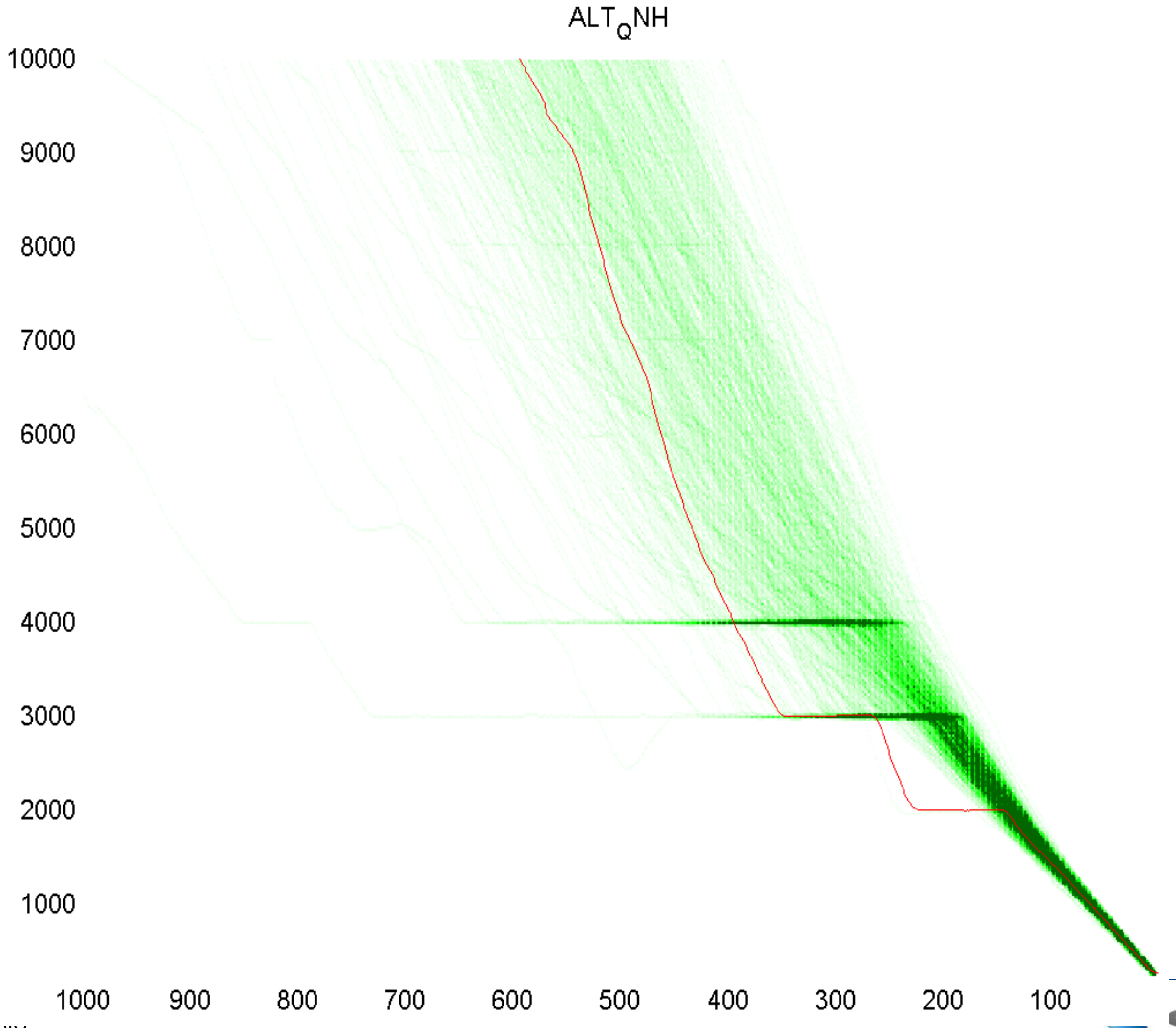
Example of atypical flights

➔ Atypical flight 1: pvalue = 0.0000000001, **trajectory plot**



Example of atypical flights

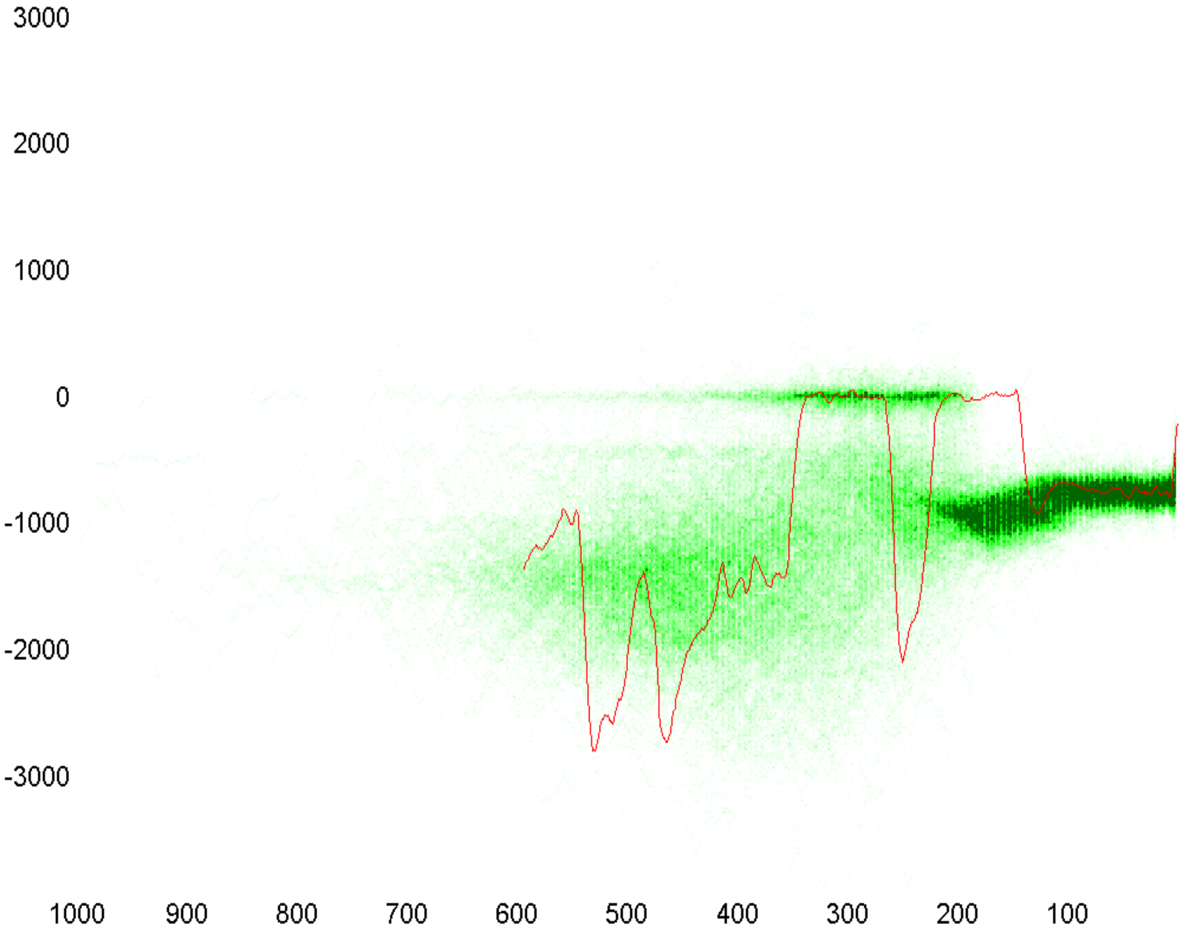
➔ Atypical flight 1: pvalue = 0.0000000001, **altitude plot**



SAFRAN Sagem

Example of atypical flights

➔ Atypical flight 1: pvalue = 0.0000000001, **vertical speed plot**



Example of atypical flights

➔ Atypical flight 1: pvalue = 0.0000000001, **air speed plot**

SAFRAN Sagem

Example of atypical flights

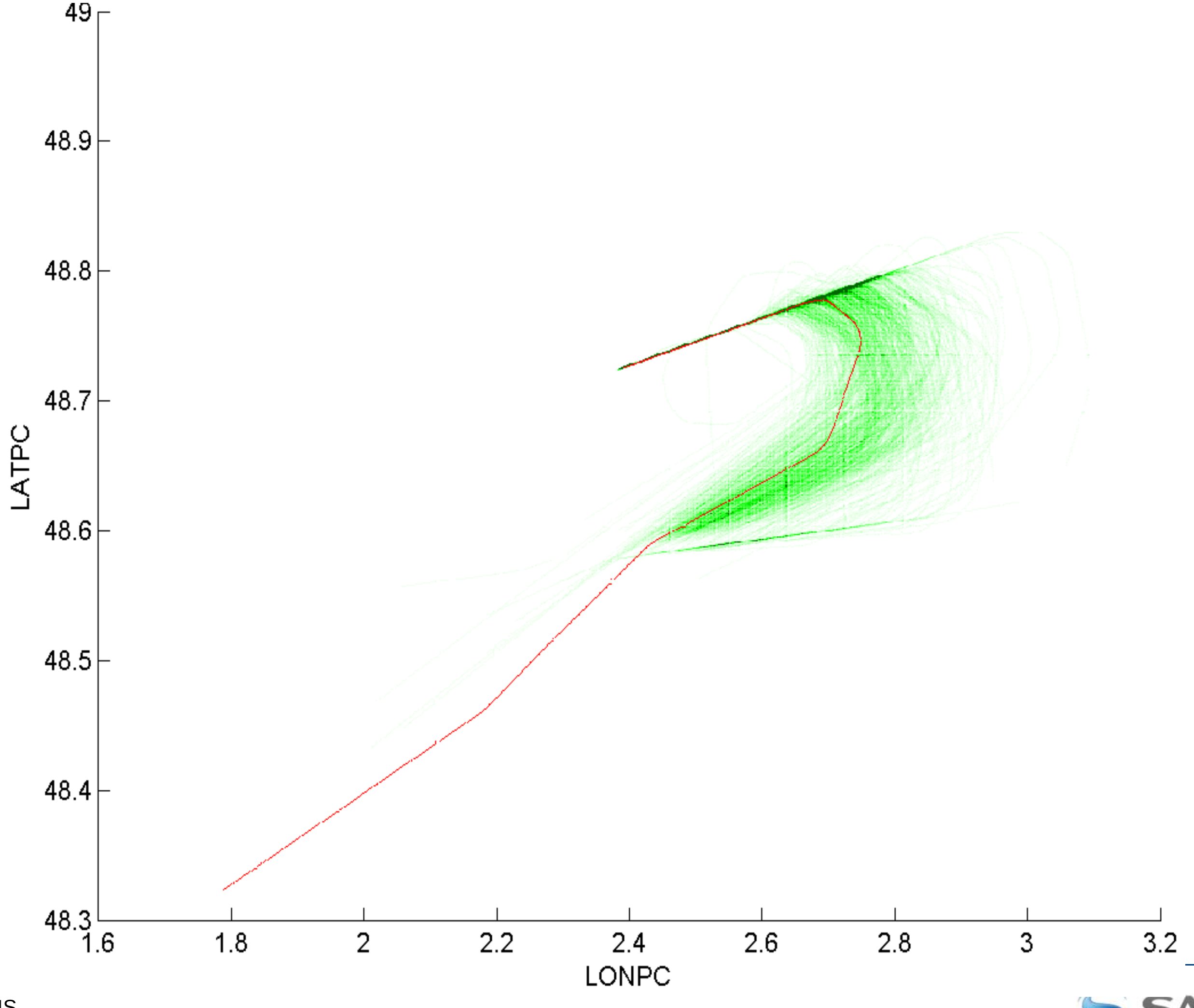
➔ Atypical flight 1: pvalue = 0.0000000001

- Classical analysis:
  - Only one event: High Vertical Speed approach, class 1

- Diagnostic:
  - Meteo is clear,
  - Probably due to operational constraints.

Example of atypical flights

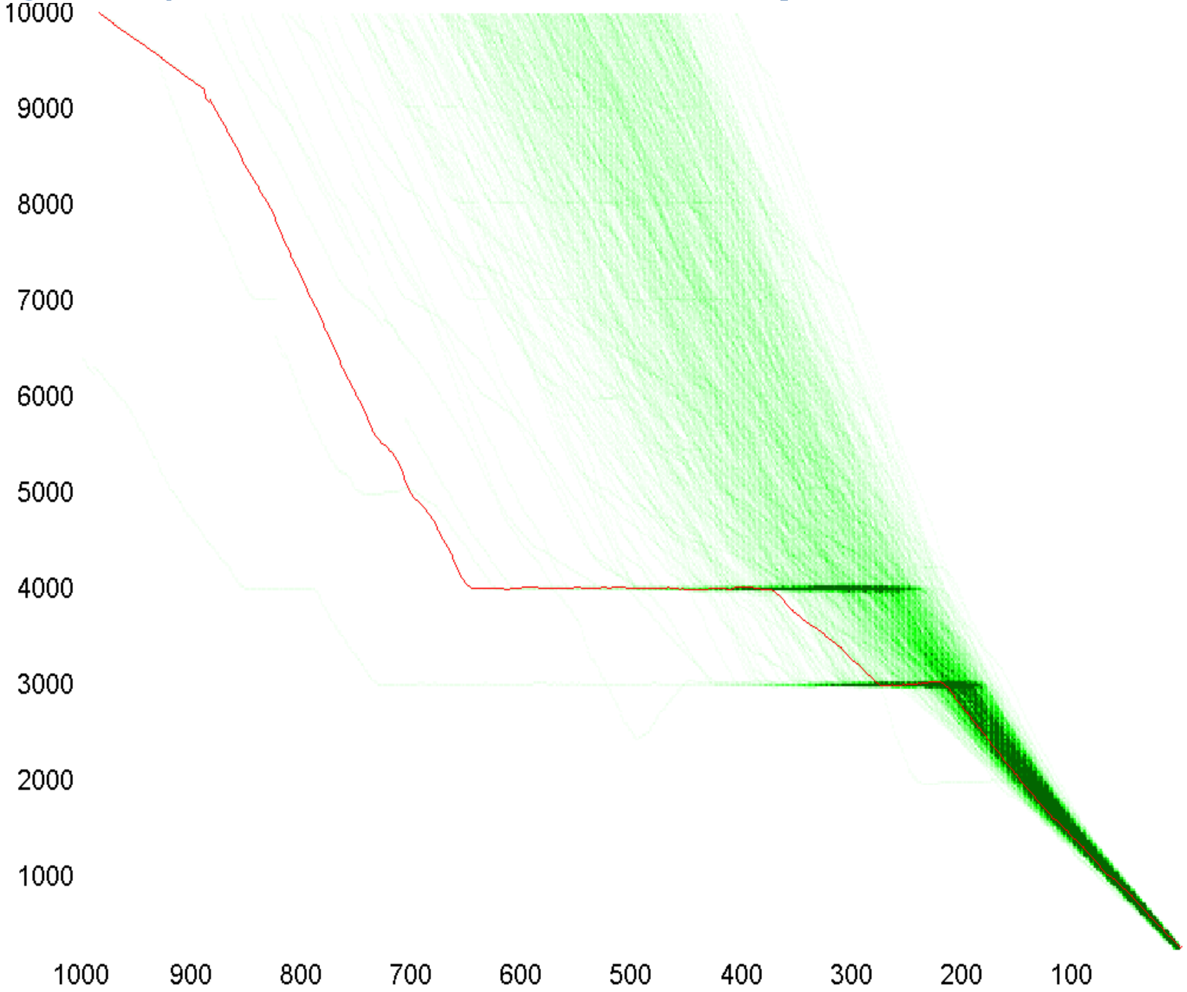
➔ Atypical flight 2: pvalue = 0.0004, **trajectory plot**



SAFRAN Sagem

Example of atypical flights

➔ Atypical flight 2: pvalue = 0.0004, **altitude plot**



SAFRAN Sagem

Example of atypical flights

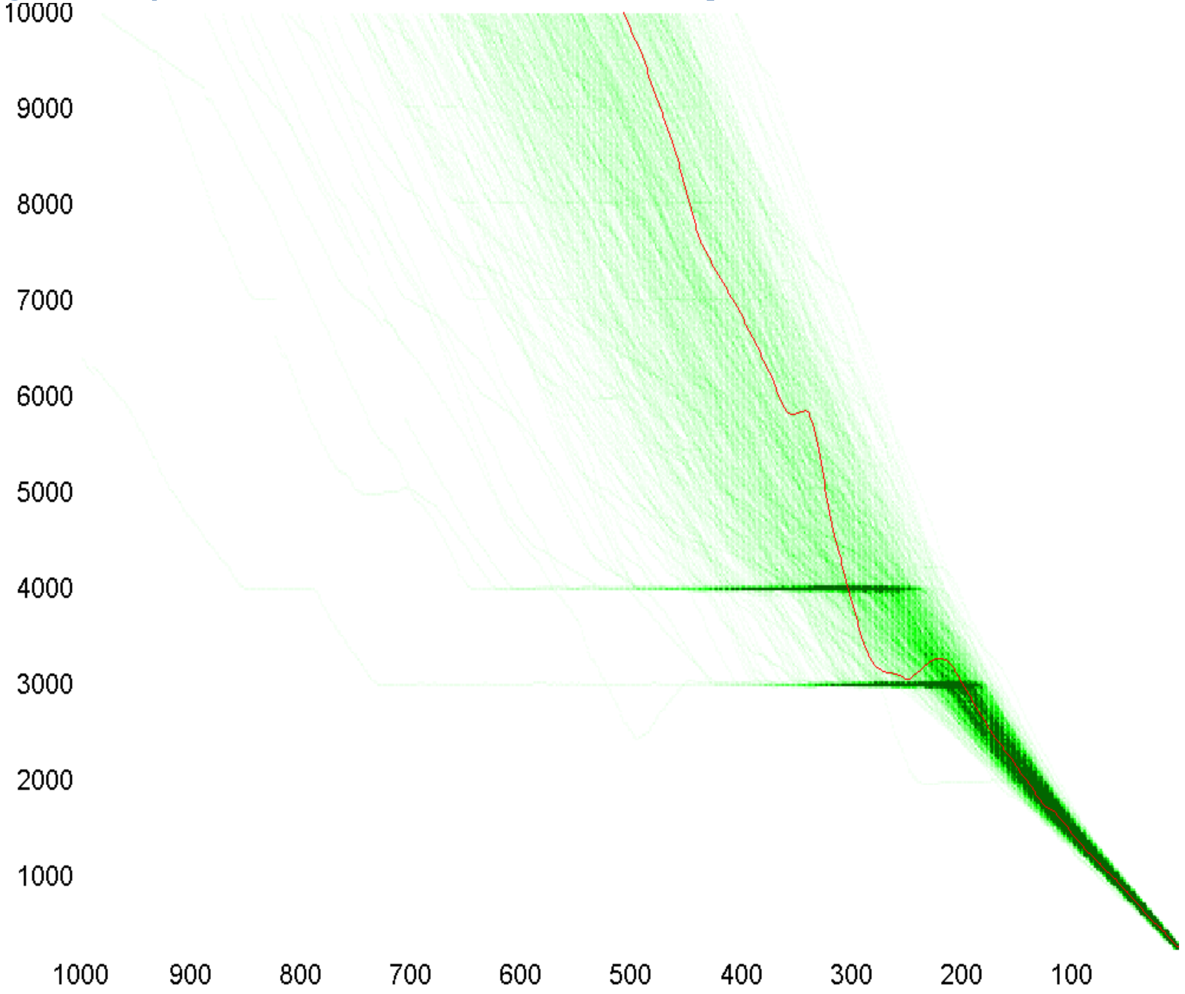
➔ Atypical flight 2: pvalue = 0.0004

- Classical analysis:
  - **No event detected** with the classical analysis.

- Diagnostic:
  - Meteo: thunder; cumulonimbus clouds, towering cumulus clouds observed
  - Meteorological constraints: the pilot had to lower his altitude to avoid the cumulonimbus cloud.

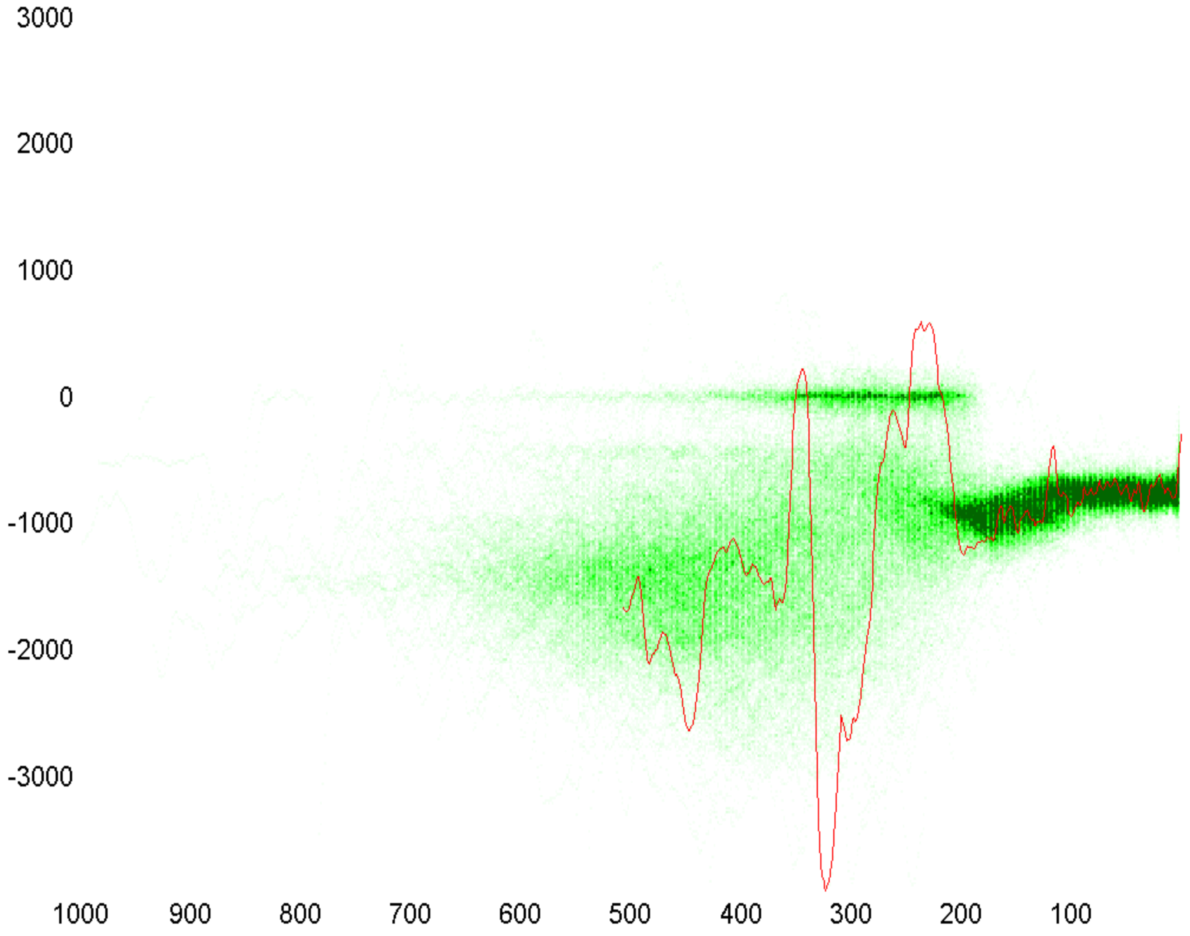
Example of atypical flights

➔ Atypical flight 3: pvalue = 0.02, **altitude plot**



Example of atypical flights

➔ Atypical flight 3: pvalue = 0.02, **vertical speed plot**



SAFRAN
Sagem

Example of atypical flights

➔ Atypical flight 3: pvalue = 0.02

- Classical analysis:
  - One event detected: High Speed - Below 5,000FT, class 1
- Diagnostic:
  - There is a very high vertical speed 300 seconds before touchdown, and irregular altitude profile.
  - This was not detected by classical procedures.

SAFRAN Sagem

Contents

Conclusion

- **A good « feel » of the data is more important than complicated mathematical algorithms**. Thus FDM experts must be involved during the conception of algorithms.
- Machine Learning has proved its usefulness as a powerful tool in complement to classical event detection procedures.
- **Good graphical representations** are key to the comprehension and interpretation of Machine Learning results.
- Up until recently, Machine Learning was very rarely applied to FDM, so there are many exciting things to discover!

Thank you