Overview of conformal predictors applications in experimental nuclear fusion environments

J. Vega[1], A. Murari[2], S. González[1], A. Pereira[1], I. Pastor[1] and JET-EFDA Contributors[3]
JET-EFDA, Culham Science Center, OX14, 3DB, Abingdon, UK
[1]Asociación EURATOM/CIEMAT para Fusión
[2]Associazione EURATOM/ENEA per la Fusione
[3]See the Appendix of F. Romanelli et al. Proceedings of the 23rd IAEA Fusion Energy Conference 2010, Daejeon, Korea

2nd International Conference 'Frontiers in Diagnostic Technologies' November 28-30 2011

Outline

- Introduction
  - Why do we need reliability?
  - What methodologies can provide reliability?
- Conformal predictors
  - Only the randomness assumption is required
- Results on classifications
  - Image recognition
  - L/H transitions
  - Recognition of local perturbations in plasma emissivity
- Results on regressions
  - Non parametric models

Introduction

- Machine Learning Methods (MLM) are used to make predictions

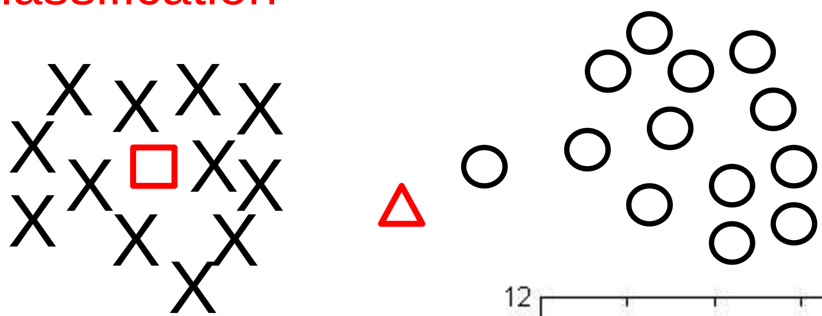
- In machine learning, any object (or sample) is represented by an ordered pair ($\mathbf{x}_i$, y_i)
  - $\mathbf{x}_i \in \mathbb{R}^m$ is the feature vector (the set of m features that characterize the object *i*).
  - y_i is the label of sample *i*. Labels can be
    - *A small finite set: classification* ($y_i \in \{L_1, L_2, ..., L_M\}$)
    - *Any real number: regression* ($y_i \in \mathbb{R}$)

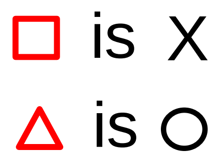
- Training dataset: ($\mathbf{x}_i$, y_i), i = 1,...N
  - A model is created to make predictions: given $\mathbf{x}_i$, <u>the model predicts the label</u>

- Test dataset: ($\mathbf{x}_j$, y_j), j = 1,...M
  - Model validation: a level of confidence can be determined and *it is assumed to be the same for all future samples*

Introduction: why do we need reliability?

- Predictions corresponding to different samples can have different levels of confidence
- Objective: <u>*to qualify*</u> *each particular prediction with a measure of its reliability*
  - *Prediction + reliability*
  - Classification

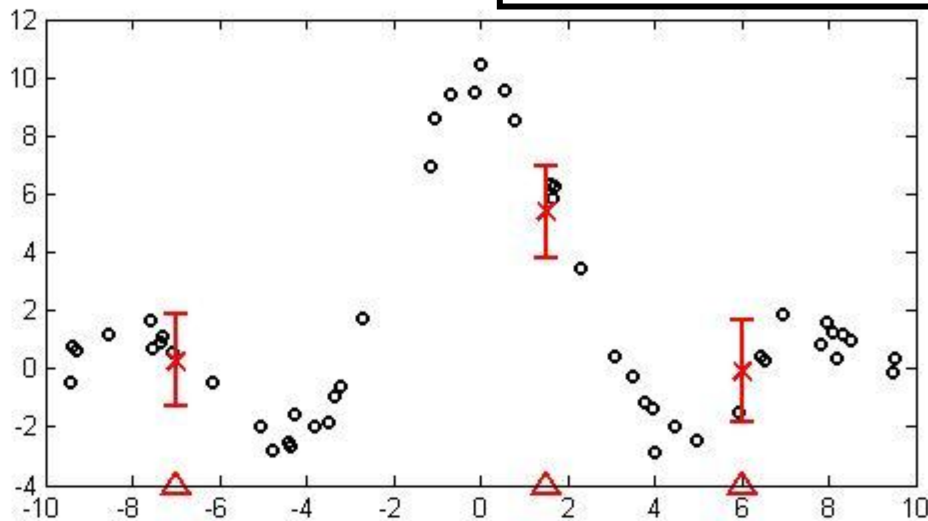




How accurate and reliable are these predictions?

  - Regression

What is the prediction region of the estimations?



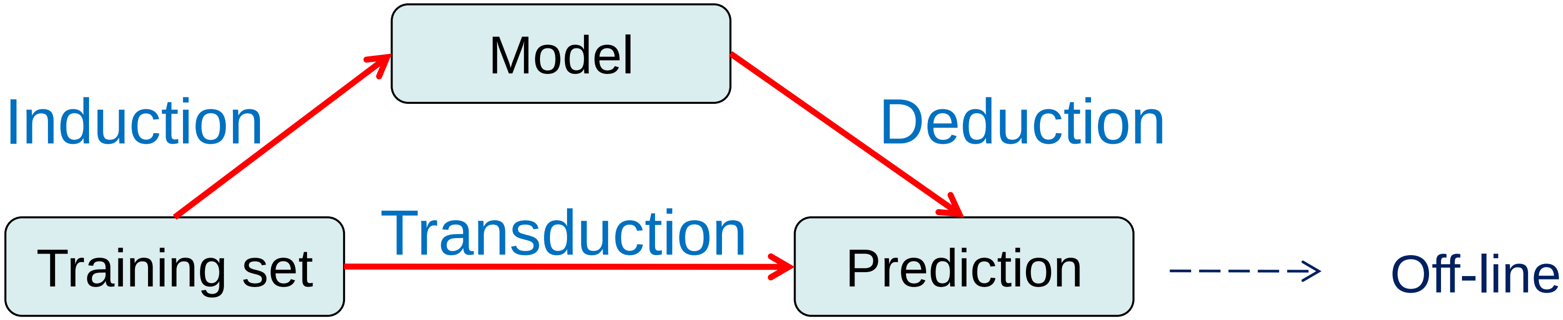
Introduction: alternative methods with reliability

- Aim: to show results in fusion about classification and regression systems with an estimation of the accuracy and reliability of the predictions
- Alternatives under the randomness assumption (independent and identically distributed samples, **iid**)
  - Reliable classification
    - Conformal predictors: *only iid*
    - Bayesian classifiers: *prior probabilities must be known or assumed*
    - Logistic regression: *parametric model whose parameter* ω *has to be determined in an empirical way*:

$$\left\{1+\exp\left[-f\left(x,\omega\right)\right]\right\}^{-1}$$

  - Reliable regression
    - Conformal predictors: *only iid*
    - Bayesian regressors: *prior probabilities*

Conformal predictors (CP)



- Inductive conformal predictors
  - Training set: proper training set + calibration set
  - Proper training set -> MODEL
  - MODEL + calibration set: conformal predictor

Real-time

- Conformal predictors are always valid
  - The probability to make an error with a prediction set at a confidence level **1 – ε** is not greater than **ε**
  - It is possible to control the number of wrong predictions by choosing a proper confidence level
    - 80% - 20%, 95% - 5%, 99% - 1%
- The reliability of the predictions is determined through the estimation of two values (confidence and credibility) in the range [0, 1]
  - A large confidence in one prediction means that all labels except the predicted one are unlikely
  - The credibility of a prediction represents how good the training dataset is to predict the label of the new sample

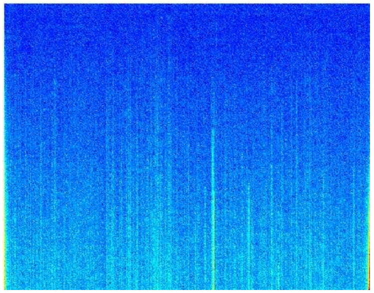
For a complete description, see A. Gammerman talk

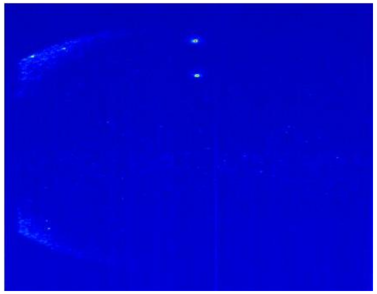
Classifiers

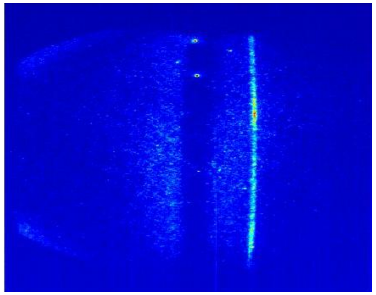
- Image recognition (TJ-II stellarator, off-line & real-time)
- L/H transitions (JET, off-line)
- Recognition and location of local perturbations in the plasma emissivity (off-line & RT simulations)

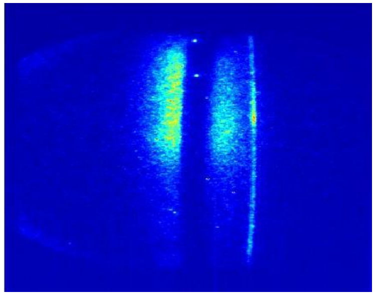
Automatic data analysis in the TJ-II Thomson Scattering diagnostic

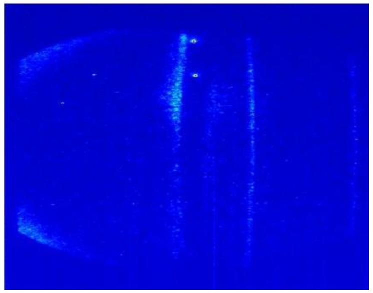
- The Thomson Scattering determines the temperature and density radial profiles of the plasma
- The data acquisition program of the TJ-II Thomson scattering was synchronised to operate and process data in an unattended manner
- The *automatic data processing system* depends exclusively on information collected by the TS diagnostic
  - The specific data processing is dependent on the collected image with the CCD camera
- 5 different types of images = 5 different types of data processing codes
- A multi-class (5) predictor classifier is needed after an image capture to know the specific processing required


CCD camera background


Stray light

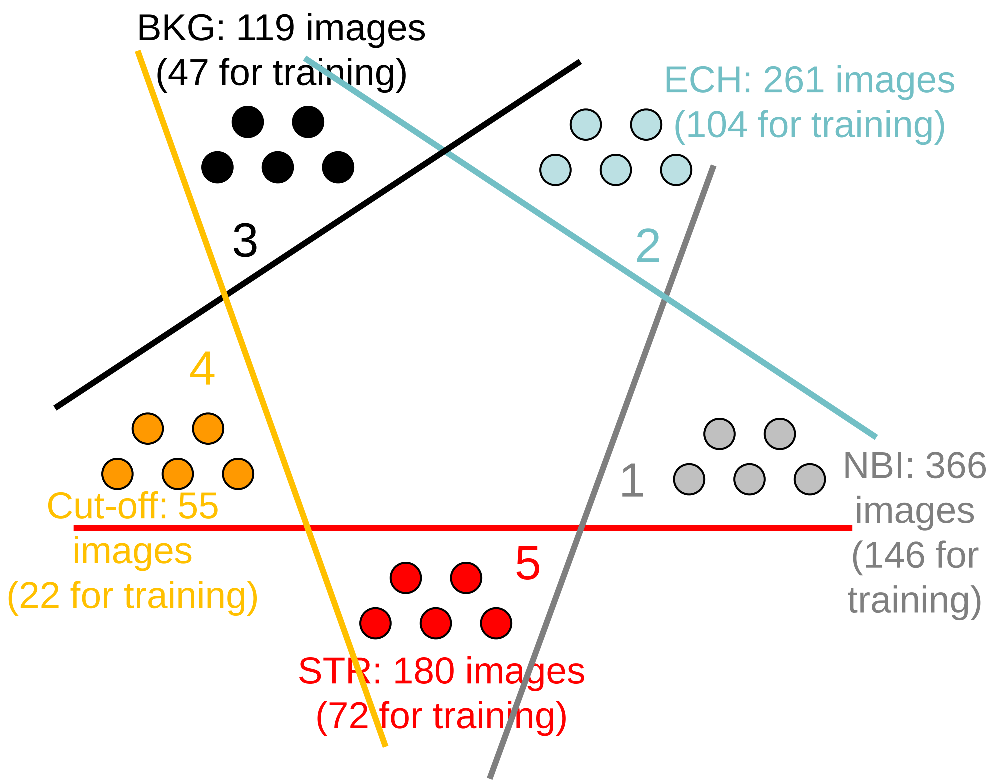

ECH phase


NBI phase


Cut off density during ECH

SVM classifier: one-versus-rest

- Objective: *off-line* and *RT* predictions with the corresponding level of significance

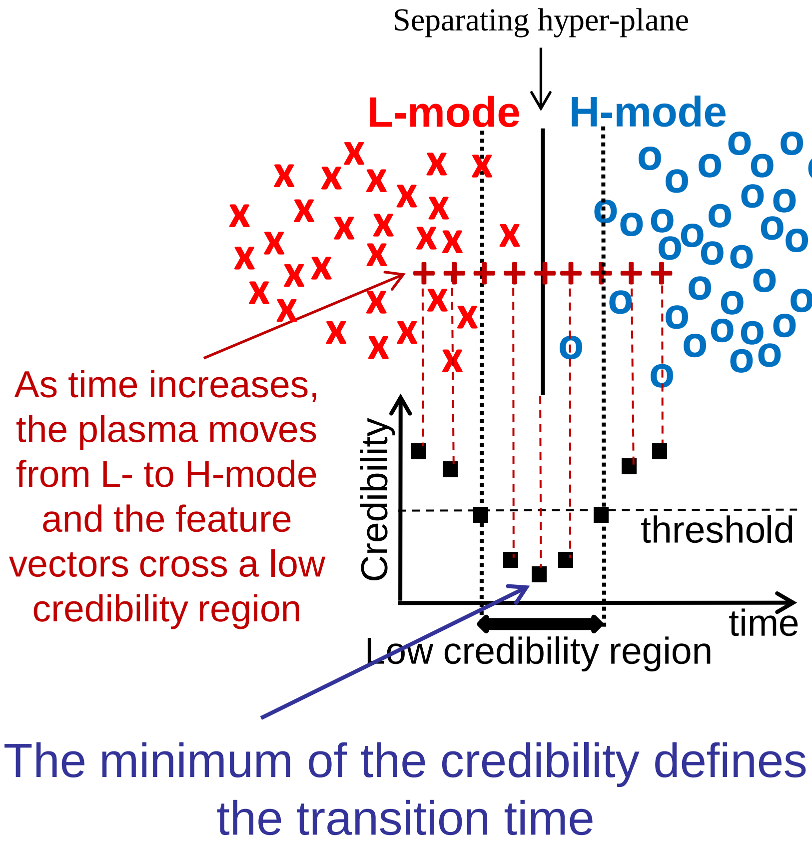


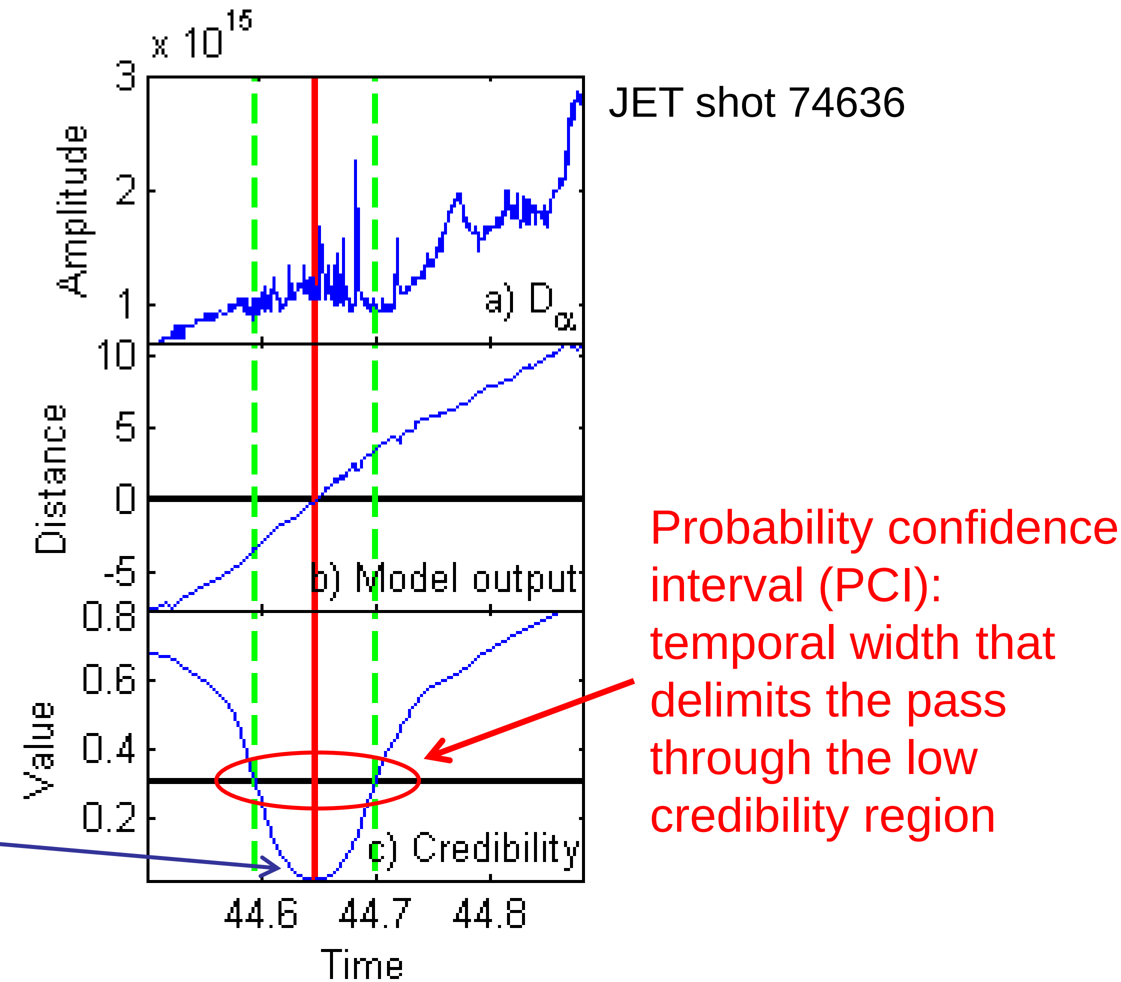
- 981 images (576 x 385 pixels)
- Features: the Haar Wavelet Transform of the images (decomposition at level 3)
  - Elimination of spatial redundancy
  - Dimensionality reduction (72 x 48)
- RBF kernel: $\sigma = 10^5$, $C = 10^3$
- Initial supervised dataset of 391 images
- The new images are added to classify future samples if the credibility is above a certain threshold
- *Off-line classifier*
  - $t_{CPU} = 15.023 \cdot 10^{-3}n + 4.523$ (s)
  - If $n = 600$, $t_{TOT} = 13.54$ s
- *Real-time classifier*: 89.7 ± 14.1 ms (after 600 images)

Predictor	Success rate (%)	Error rate (%)	Ambiguities (%)	Low cred. (%)	Credibility threshold	<conf>	σ_{conf}	<cred>	σ_{cred}
Off-line	**98.30**	0.51	0	1.19	0.05	**0.997**	0.005	**0.552**	0.302
Real-time	**95.70**	0.50	0.50	3.29	0.05	**0.997**	0.071	**0.553**	0.285

L/H transitions in JET

- Automatic prediction of L/H transition times
- Frontier problem





This is a novel interpretation of the credibility in conformal classifiers

L/H transitions in JET

- Steps for the development of a reliable classifier to determine JET L/H transition times with a probability confidence interval through the credibility of a conformal predictor
  - Determination of the best quantities to detect both confinement modes
  - Generation of a model (training + validation) from a dataset of **551** discharges with transition times determined by experts
  - Application of a conformal classifier to a dataset of **1451** discharges between campaigns C15 and C21 (66001-78157) to determine transition times and probability confidence intervals

L/H transitions in JET: step 1

- Automatic determination of best quantities

BNDIAM: Beta normalised with respect to the diamagnetic energy	**LSPRO:** R coordinate outer lower strike point
BT: Toroidal magnetic field	**LSPZO:** Z coordinate outer lower strike point
ELO: Elongation boundary	**RIG:** Radial inner gap
FDWDT: Time derivative of diamagnetic energy	**ROG:** Radial outer gap
IPLA: Plasma current	**AD36:** D_α inner view
LI: Plasma inductance	**TOG:** Top Outer Gap
PTOT: Total heating power	**RAD:** Radiated power
Q95: Safety factor	**TE02:** Temperature at psi = 0.2
TRIL: Lower triangularity	**CR0:** Minor radius
TRIU: Upper triangularity	**RGEO:** Major radius
XPRL: R coordinate lower XP	**LAD3:** Electron density line averaged – core
XPZL: Z coordinate lower XP	**LAD4:** Electron density line averaged – edge
LSPRI: R coordinate inner lower strike point	**WDIA:** Diamagnetic energy
LSPZI: Z coordinate inner lower strike point	**TE08:** Temperature at psi = 0.8

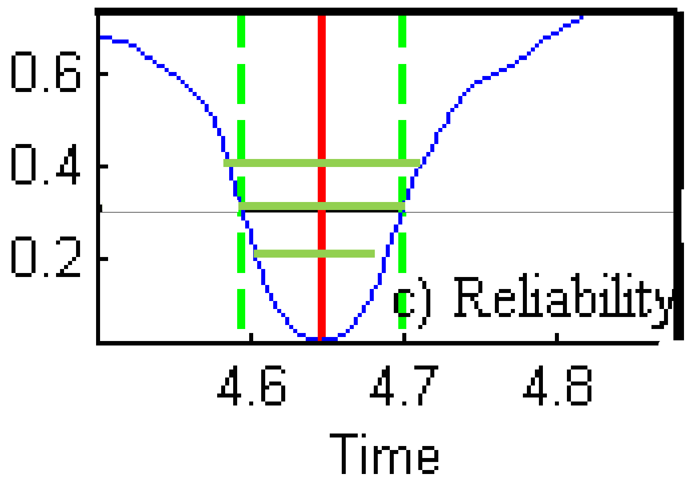
- SVM based method that eliminates one by one the least important quantities without increasing the model complexity

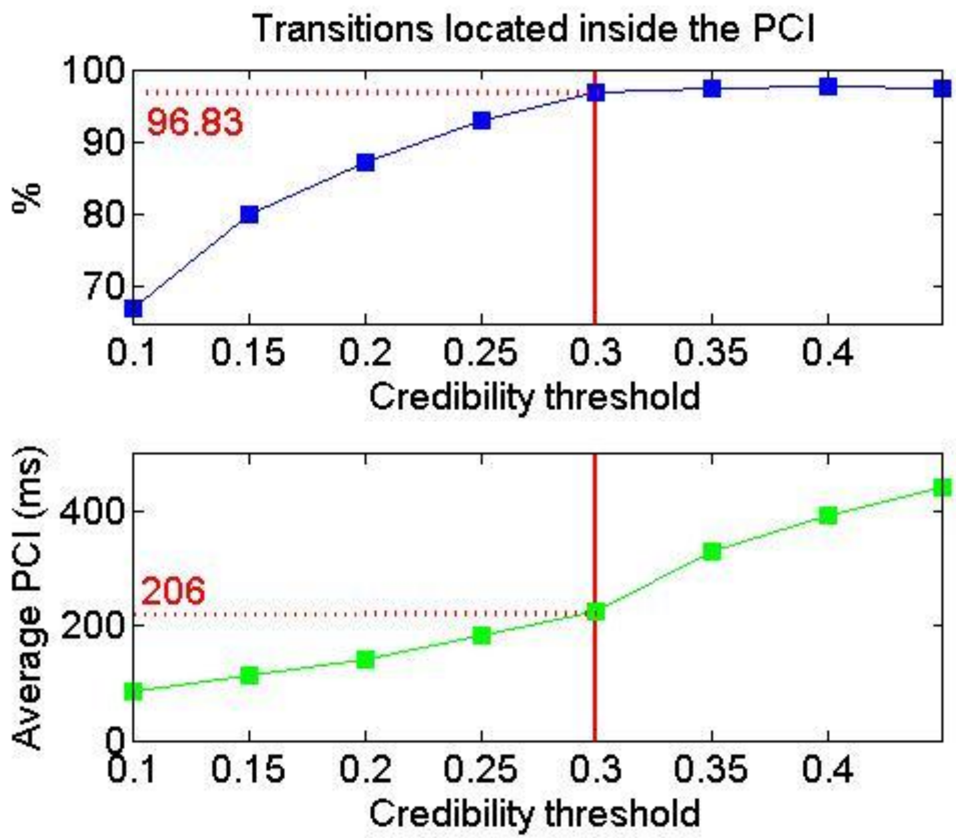
S. González et al. Rev. Sci. Ins. 81, 10E123 (2010)

L/H transitions in JET: step 2

- Model creation
- Probability confidence interval with known transition times
  - 551 discharges: 141 training + 410 validation

As mentioned, the time instant of minimum reliability defines the transition time

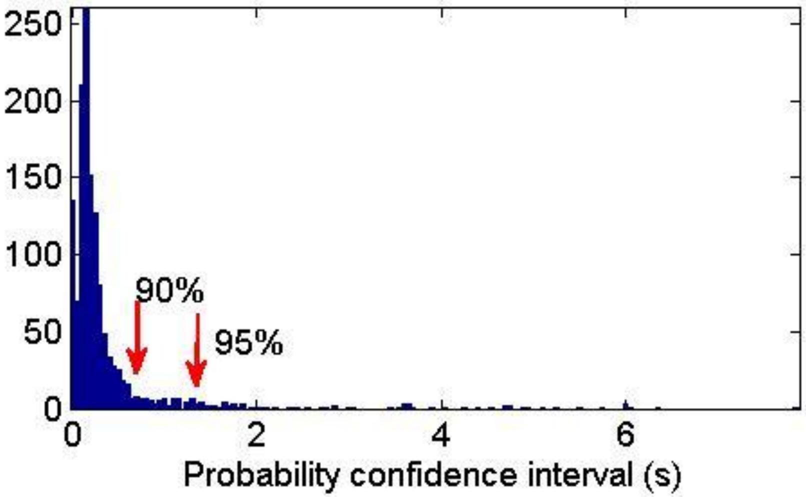




Different thresholds generate different average widths of the PCI

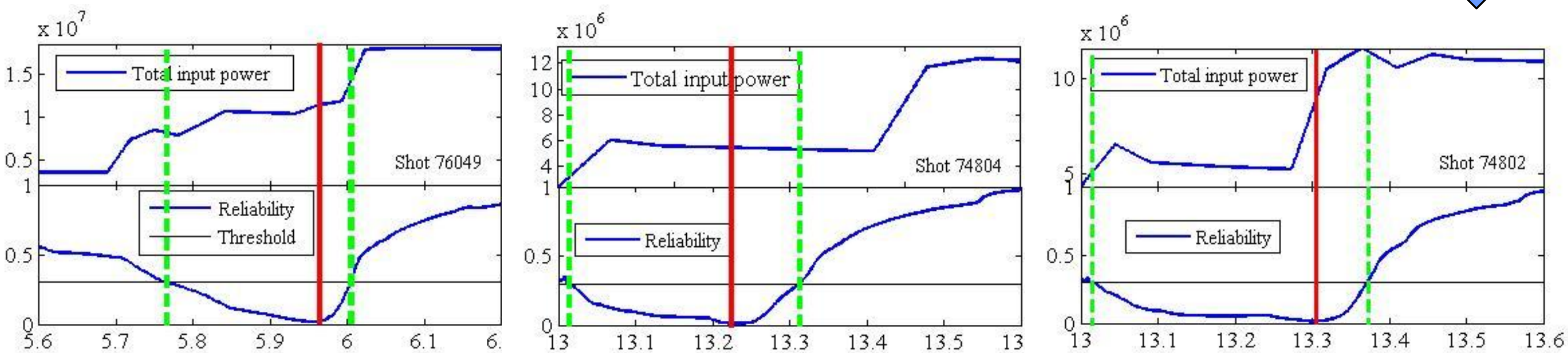
L/H transitions in JET: step 3

- Conformal classifier applied to **<u>1451</u>** discharges to determine transition times and the PCI



	90%	**95%**
L/H	191 ms	233 ms

Interpretation

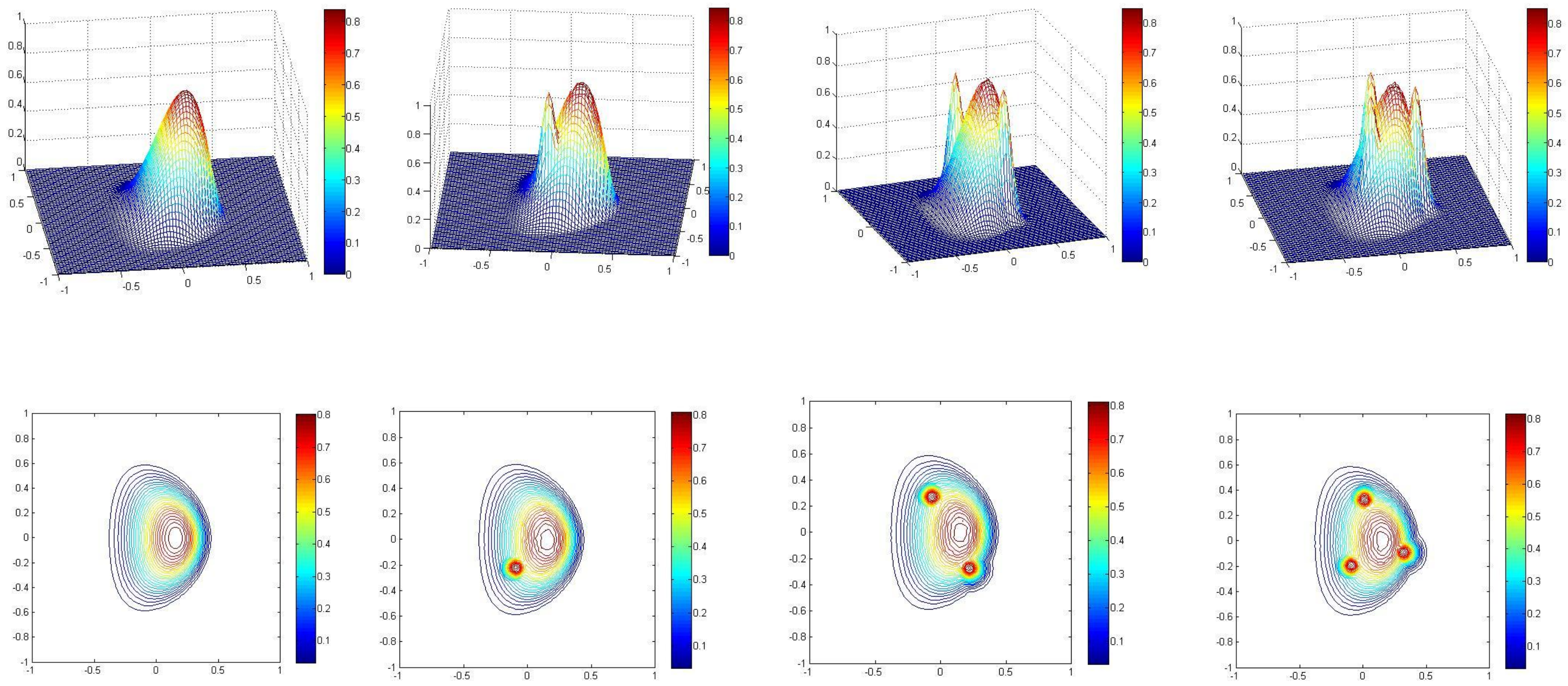


Long intervals of low credibility regions are associated to discharges in which the input power is not introduced in an abrupt way. The plasma remains in the L/H frontier during a long time and **the credibility can be used to detect this physical behaviour**

If the most extreme cases are not considered, the width of the probability confidence interval is about 100 ms

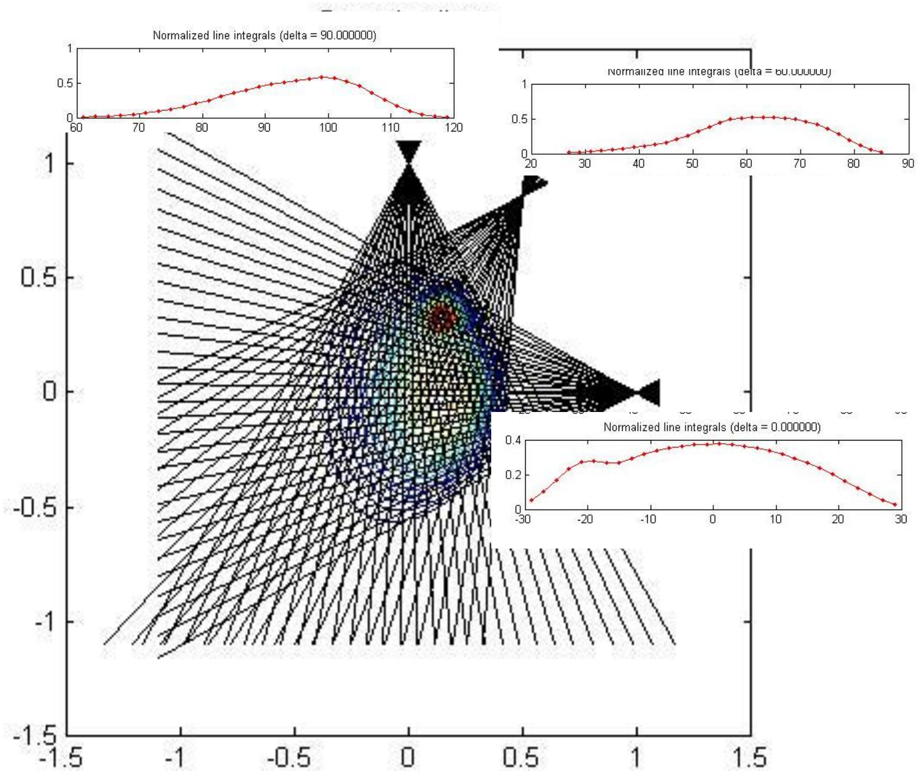
Plasma emissivity and local perturbations

- Simulation to detect and locate a number of local perturbations in the plasma
  - Soft X-rays or bolometry



Plasma emissivity and local perturbations

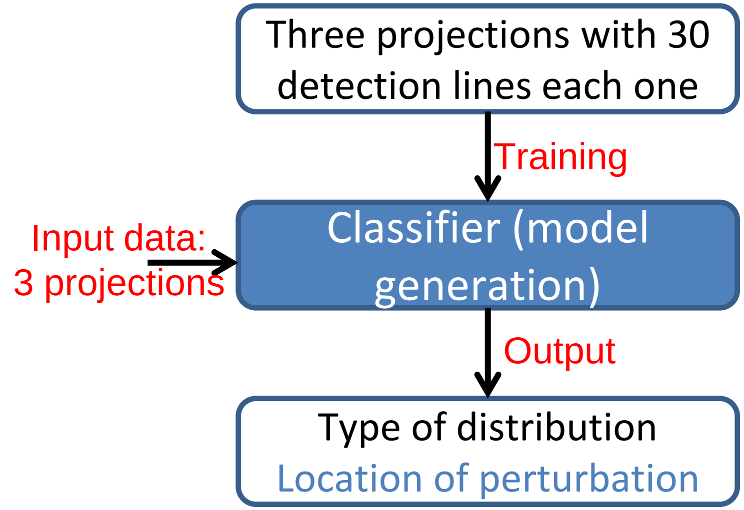
- Plasma can be "*seen*" through a very limited number of projections
  - Projections are made up of line integrals
- Arrays with fan-like geometry are typical in fusion
  - Detectors occupy a reduced space but the line integrals cover all the plasma
- Depending on the problem to solve and inherent constraints (initial distribution, spatial resolution, ill-posed problems), tomography can be unfeasible
- Can we determine the number of local perturbations at a given time instant?

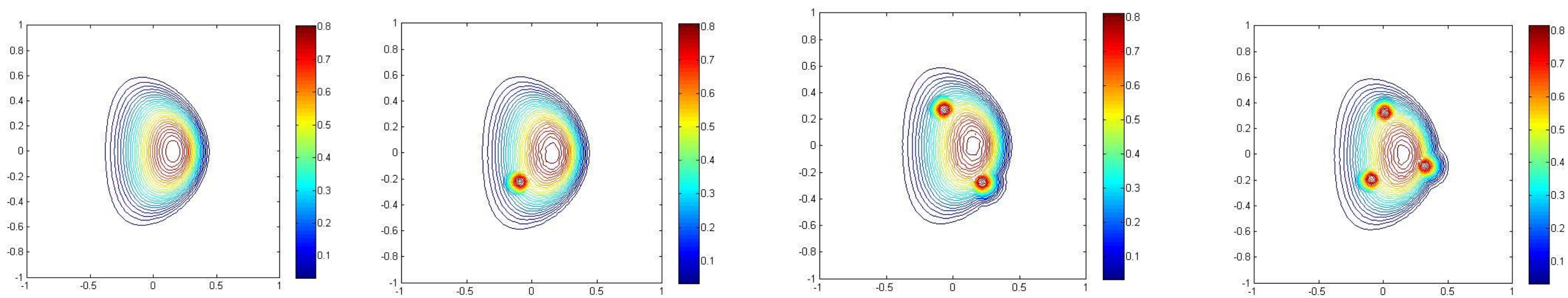


Plasma emissivity and local perturbations

- Machine learning methods can be applied to determine the number of local perturbations from projections

- Training datasets (SVM & one-vs-rest)





60 distributions (Gaussian noise).

60 distributions: local perturbation at 60 different positions (Gaussian noise)

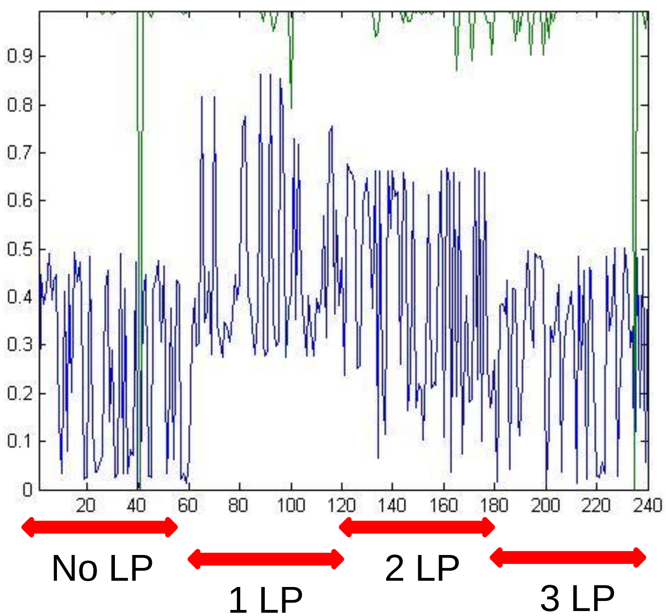
60 distributions: local perturbations at 60 different positions (Gaussian noise)

60 distributions: local perturbations at 60 different positions (Gaussian noise)

Plasma emissivity and local perturbations

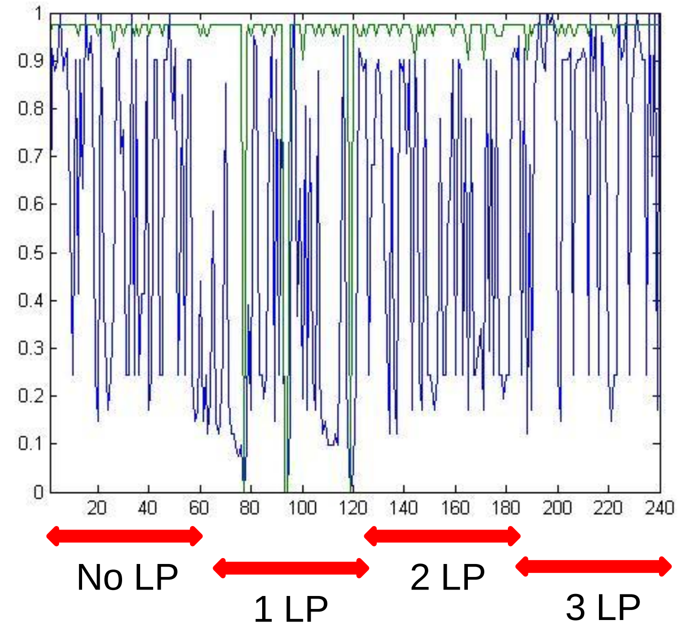
- Test datasets: 60 sets of 3 projections per class

Conformal prediction
Confidence level +98%



Success rate: 95.42% (229/240)
Ambiguities: 0.83% (2/240)
Failure rate: 3.75% (9/240)

Inductive conformal prediction
PTS (70%) + CS (30%)
Confidence level +95%



Success rate: 97.92% (235/240)
Ambiguities: 1.67% (4/240)
Failure rate: 0.42% (1/240)

Plasma emissivity and local perturbations

- By subtracting the projections without local perturbations from the projections measured, plasma chords with enhanced emission are determined in each array
- The barycentre of the resulting triangle is assumed to be the centre of the local perturbations

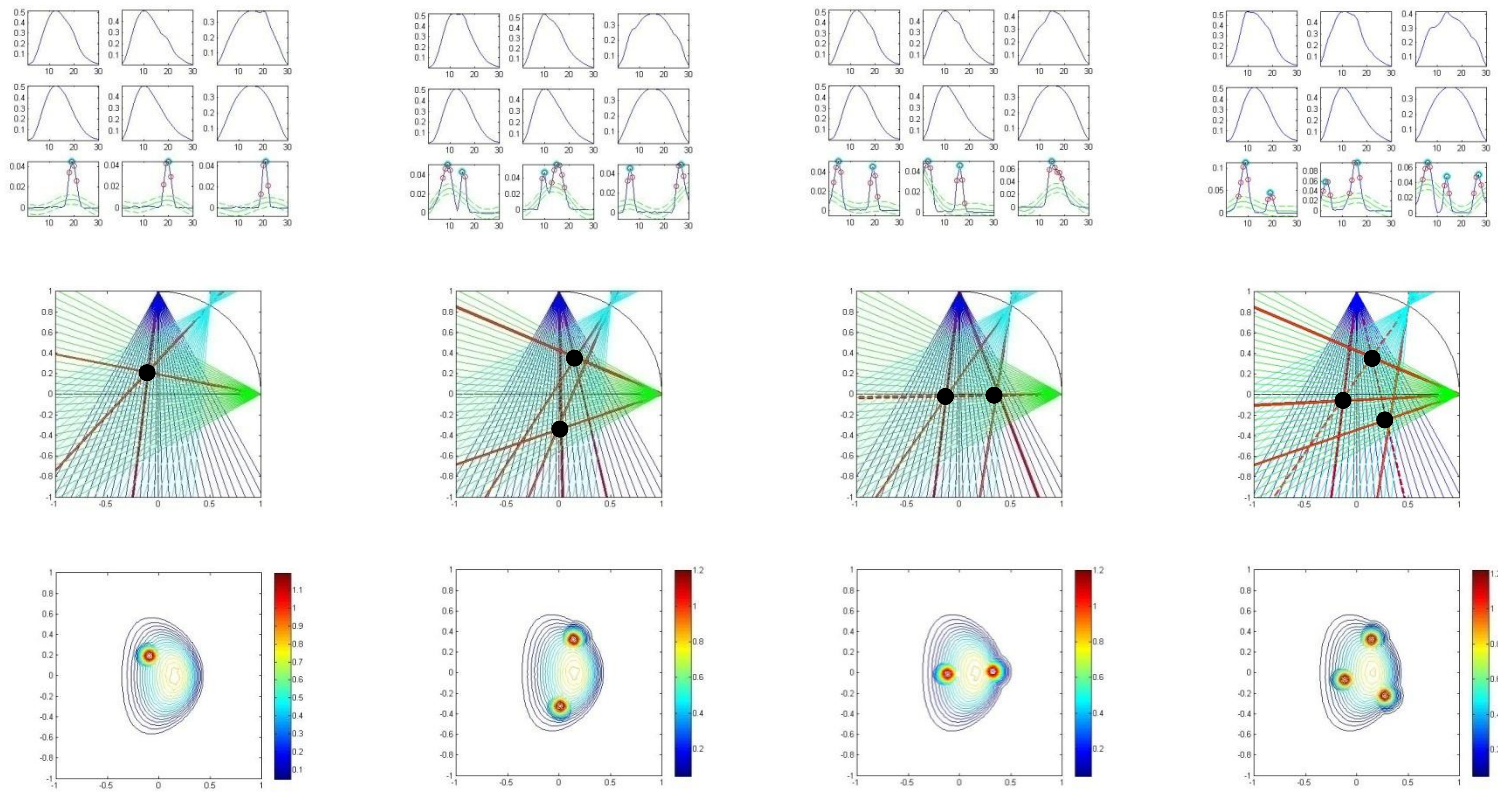
Line integrals with perturbation

Line integrals without perturbation

Difference

Predicted local perturbations and spatial location

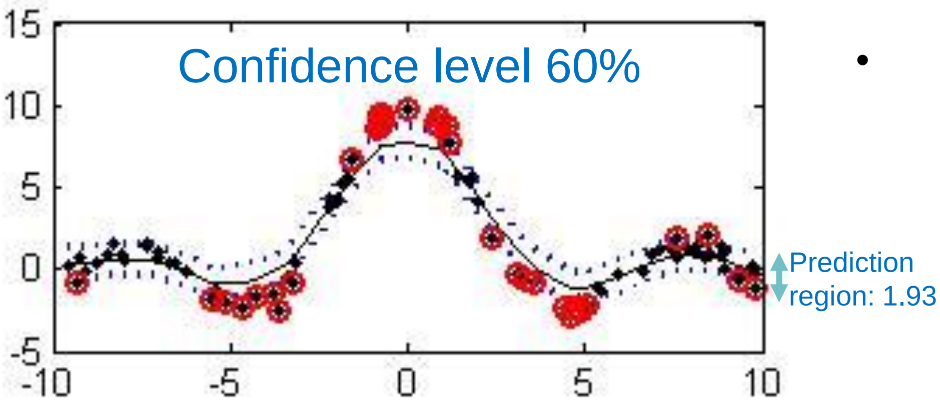
Initial emissivity distribution

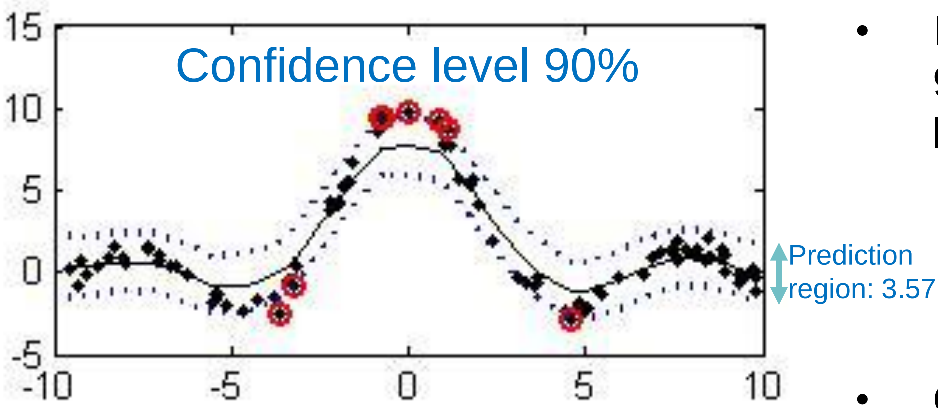


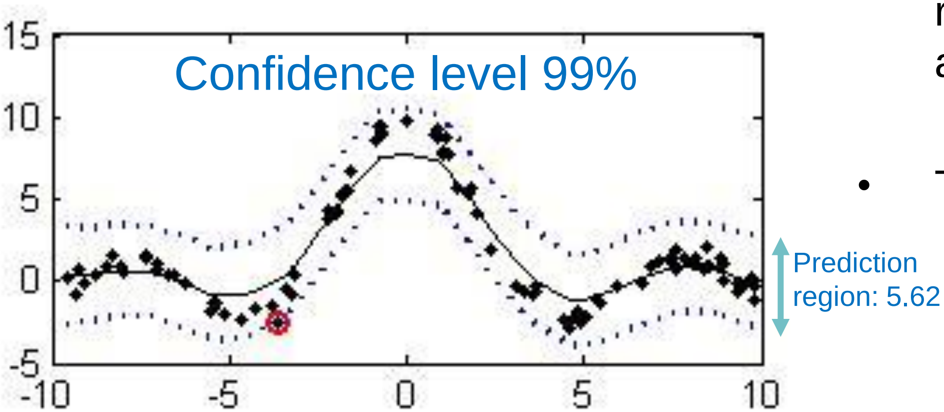
Conformal regressors

- Non parametric models (L/H transitions in JET, off-line)

Conformal predictors: regression







- $f(x) = 10\dfrac{\sin|x|}{|x|}$, *x* has been drawn randomly in [-10, 10]
  - Training set: $\{(x_i, f(x_i)), i = 1,\ldots, M\}$
  - Plain line: regression line
  - Dotted lines: prediction region
  - Black points are inside the prediction region
  - Red points are outside the prediction region
- In terms of CP, a prediction region of 60%, (respectively 90% and 99%) covers each prediction with probability at least 0.6, 0.9 and 0.99
  - At most, 40%, 10% and 1% of the initial dataset will be outside the prediction region (35%, 8% and 1% respectively in the plots)
- Given an initial dataset and a confidence level for the regression, the prediction region for each *x* can be seen as an error bar of the prediction
- The larger confidence level the greater prediction region
  - If the confidence level is 100%, the error bar is infinite (probability 1 of having any value for the prediction)

Non-parametric model for L/H transitions in JET

- Alternative approach to parametric models (for example, scaling laws)

- Model $P_L = f(n_e, B_t, S)$
  - P_L: loss power (input power – d(total plasma energy)/dt) ($1.8 \leq P_L \leq 18.5$ MW)
  - n_e: line average electron density ($0.67 \cdot 10^{19} \leq n_e \leq 3.46 \cdot 10^{19}$ m^{-3})
  - B_t: magnetic field ($1.59 \leq B_t \leq 3.43$ T)
  - S: plasma surface ($3.10 \leq S \leq 4.67$ m^2)

- On-line protocol: the goal is to predict each consecutive response given the corresponding feature vectors and all the previous observations
  - Each prediction is qualified with its own prediction region

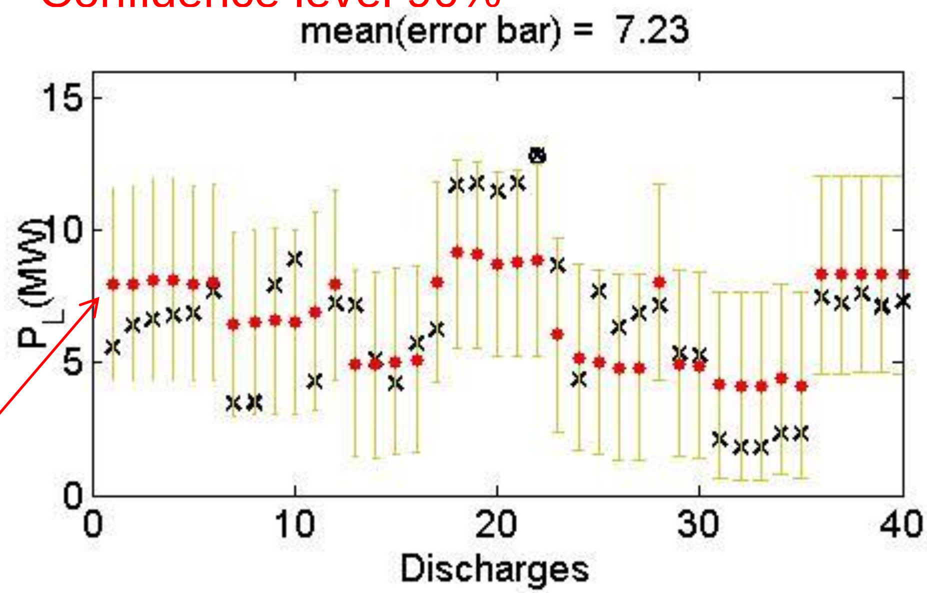
- Dataset: 558 discharges between 73337 (C21) and 78156 (C26)

Non-parametric model for L/H transitions in JET

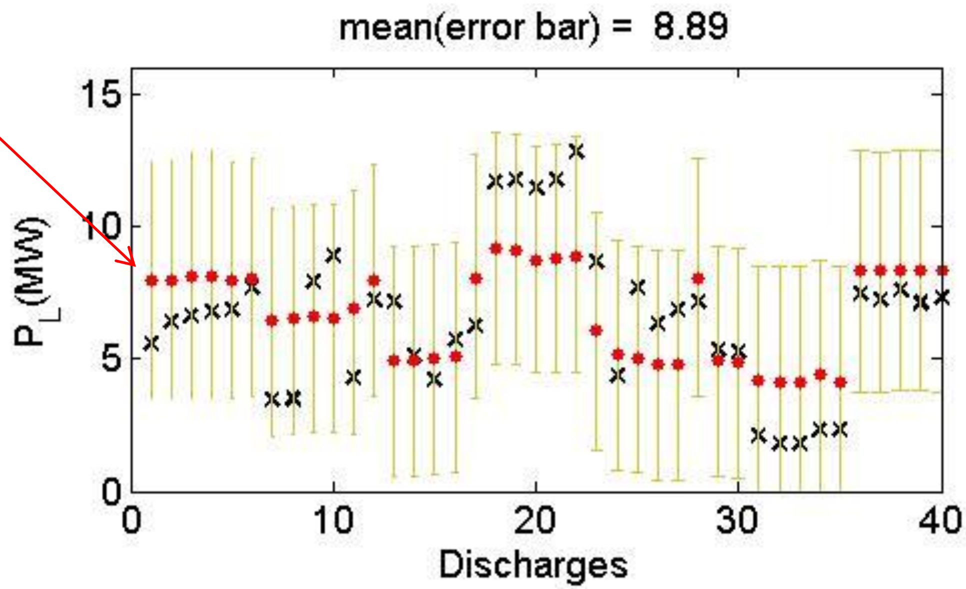
- Model estimation
  - Computation of the centroid of the discharges with the objects $(P_L, n_e, B_t, S)_i$, i = 1,...,558
  - Training dataset: the closest 286 discharges to the centroid

Estimations

- Confidence level 90%



- Confidence level 95%



The larger confidence level the greater prediction region

Summary

- CP have shown a high reliability in *ad hoc* classifiers for
  - Image recognition (TJ-II TS diagnostic)
  - Frontier problem (L/H transition times in JET)
    - A new interpretation of the credibility
  - Simulations to determine both number and spatial location of local perturbations in the plasma

- CP have been used under real-time requirements and also show a high reliability of the classifier
  - TJ-II TS diagnostic
  - Simulations to determine both number and spatial location of local perturbations in the plasma

- CP have been used to determine error bars in regressions with a non-parametric model
  - $P_L = f(n_e, B_t, S)$ in L/H transitions in JET