Identification of edge-localized instabilities in nuclear fusion plasmas using pattern recognition techniques

A. Shabbir[1,2] , G. Hornung[1] and G. Verdoolaege[1,3]

[1]Department of Applied Physics, Ghent University, Belgium
[2]MPI for Plasma Physics, Garching, Germany
[3]LPP–ERM/KMS, Brussels, Belgium

General Scientific Meeting 2016

Belgian Physical Society

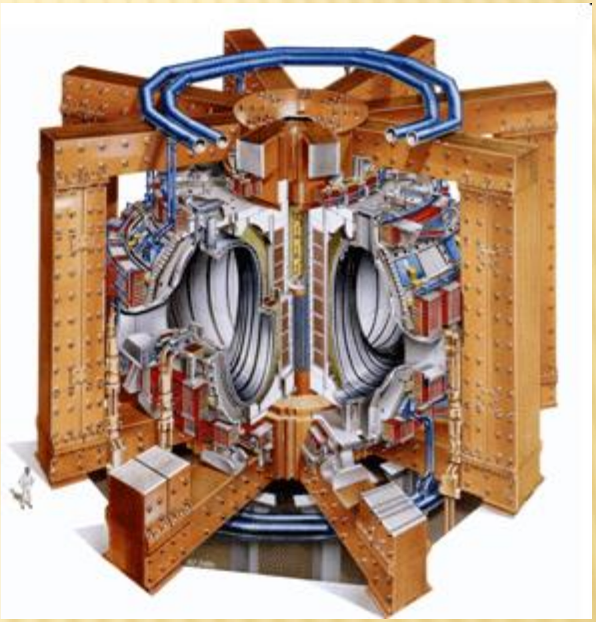
Magnetic confinement fusion

On of the most promising routes towards nuclear fusion on earth ⟶ magnetic confinement of a hot hydrogen plasma in a tokamak device

High confinement (H-mode) regime in tokamaks is accompanied by an edge instability:

Edge-localized modes (ELMs)

Type I ELMy H-mode is the reference operating scenario for next-step fusion device ITER



JET tokamak, situated in Oxfordshire UK, is currently the world’s largest operational tokamak device

Edge-localized modes (ELMs)

ELMs → intense, short duration, repetitive events that result in sudden expulsion of energy and particles from the plasma edge

➕ Beneficial for impurity control

➖ Degrade confinement + large ‘uncontrolled’ ELMs will cause intolerable transient heat loads on the plasma-facing components in ITER

In this work, a classification scheme for ELM types is presented which:

- Effectively incorporates inherent stochasticity of ELMs and the substantial measurement uncertainties
- Provides an automated, fast, high-accuracy and standardized classification for ELM types which can considerably reduce the effort of ELM experts in identifying ELM types
- Demonstrates that the distributions of ELM properties contain more information than the mean values alone

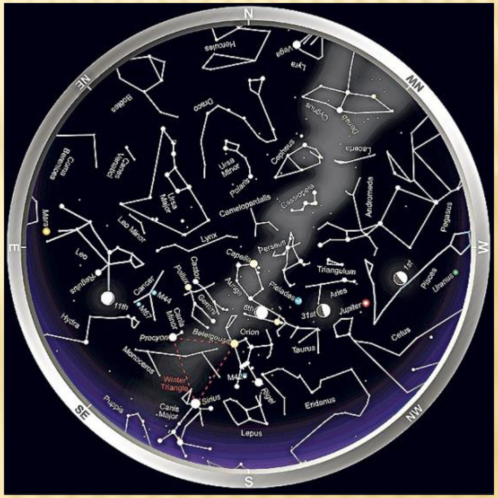
METHODOLOGY

Pattern recognition

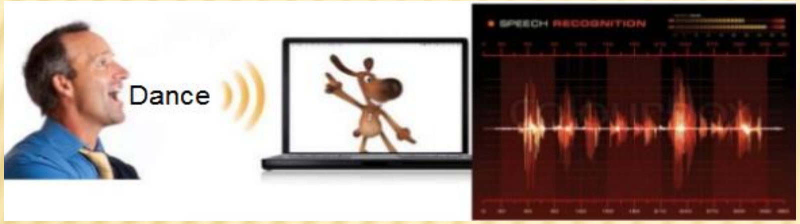
"The assignment of a physical event to one of several pre-specified categories" --- Duda & Hart

A pattern reflects an object, process or event

During recognition (or classification), categories (classes) are assigned to the events (or objects)



Patterns of constellations



Speech recognition

Signature recognition

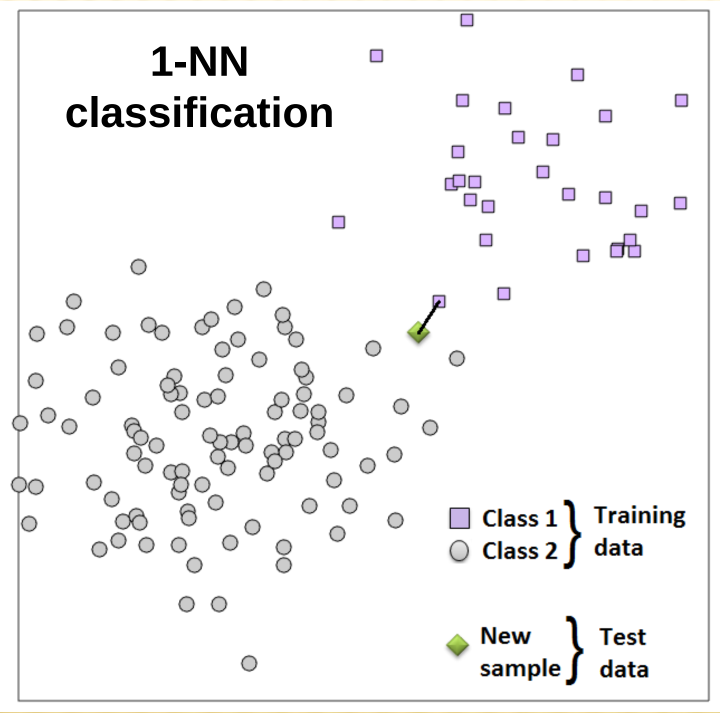
Facial recognition

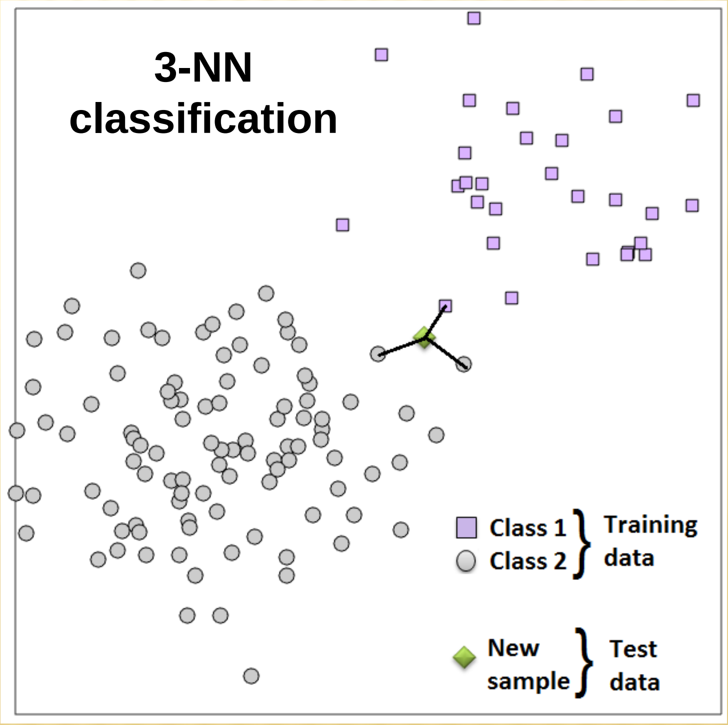
Texture patterns

K-nearest neighbor (kNN) classifier

To classify an unknown event:

- Compute distance to events in training data
- Identify *k* nearest neighbors
- Use class labels of the nearest neighbors to determine class label of the test event (majority vote)





Information geometry

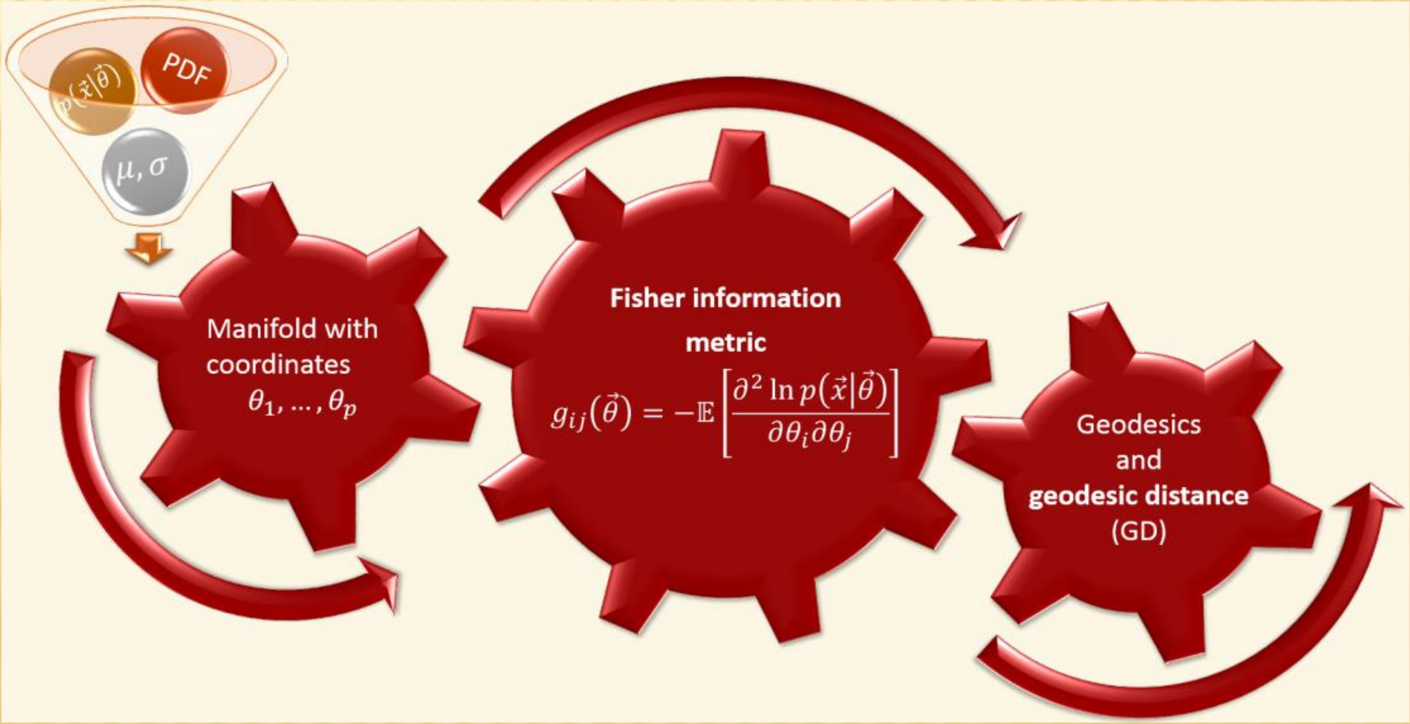
Fusion plasmas $\longrightarrow$ **Measurement uncertainties + fluctuations**

Measurement $\longrightarrow$ **Probability distribution**

Information geometry $\longrightarrow$ **Family of probability distributions**

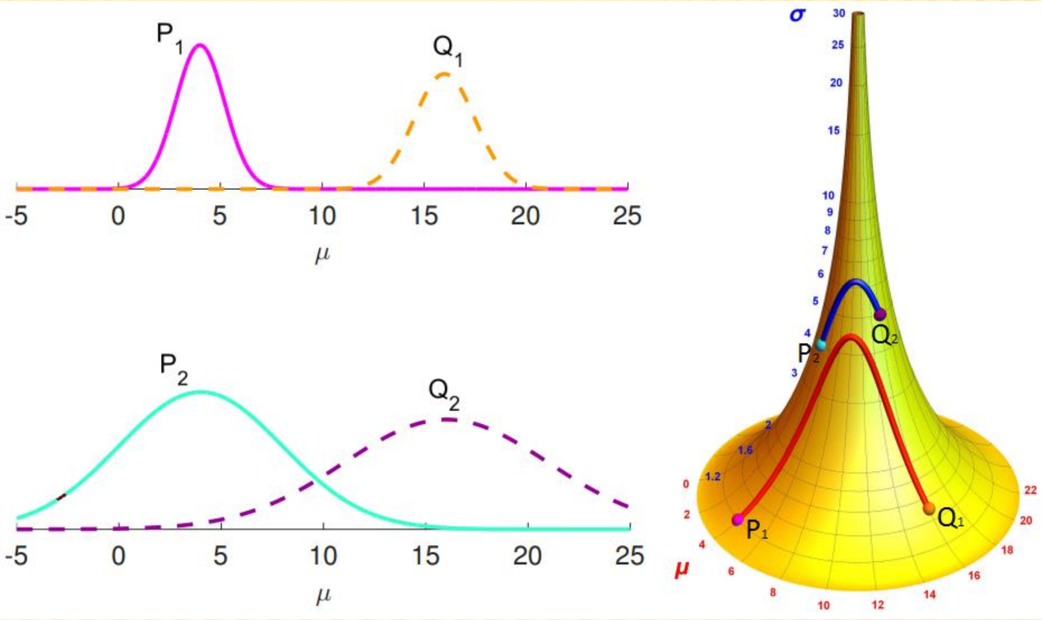
$\downarrow$

Riemannian manifold



Rao geodesic distance (GD)

A *geodesic curve* on a manifold *M* is locally the *shortest path* between points (probability distributions)

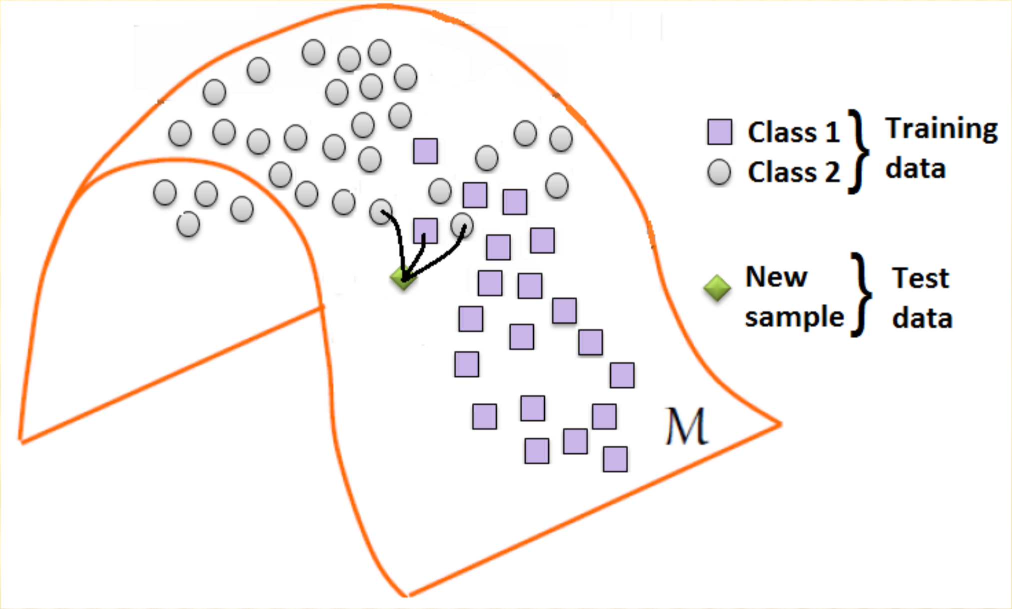


$$Eucl.dist\,(P_1, Q_1) < Eucl.dist\,(P_2, Q_2)$$

$$\mathrm{GD}(P_1, Q_1) > GD(P_2, Q_2)$$

GD is a natural, intrinsic distance measure on the manifold of probability distributions

GD-based kNN classifier



k-nearest neighbors of a ‘test’ probability distribution function (PDF) are the ‘training data’ PDFs that have the *k* smallest GDs to the test PDF

EXPERIMENTS & RESULTS

Dataset

100 plasma discharges from the JET tokamak (operating with the carbon wall)

Type I ELMs = 69 discharges

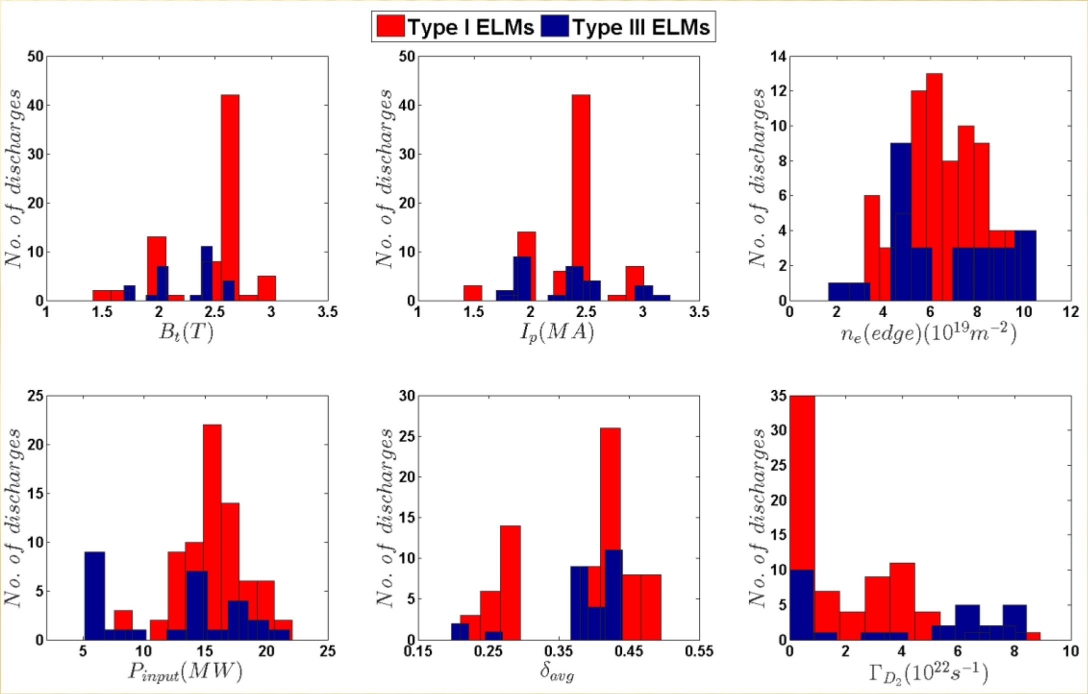
Type I ‘high frequency’ (HF) ELMs = 5 discharges

Type III ELMs = 26 discharges

Analysis restricted to time intervals → stationary plasma conditions

The global plasma parameters considered for each discharge are:

- Toroidal field = $B_t\ (T)$,
- Plasma current = $I_p\ (MA)$,
- Line-integrated edge density = $n_e\ (10^{19} m^{-2})$,
- Input power = $P_{input}\ (MW)$,
- Average triangularity = δ_{avg}
- Gas fuelling = $\Gamma_{D_2}\ (10^{22} s^{-1})$



GD-based kNN classification using global plasma parameters

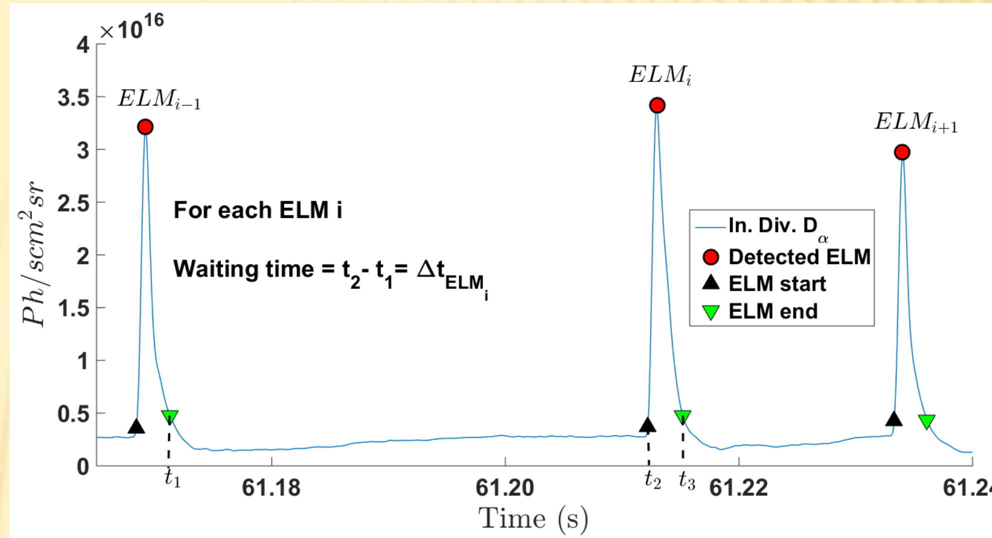
We assume that the error bars associated with each plasma parameter pertain to a statistical uncertainty in the data, specifically that it represents a single standard deviation

Underlying distribution of the global plasma parameters is Gaussian with the measurement = mean (μ) and its error bar = standard deviation (σ)

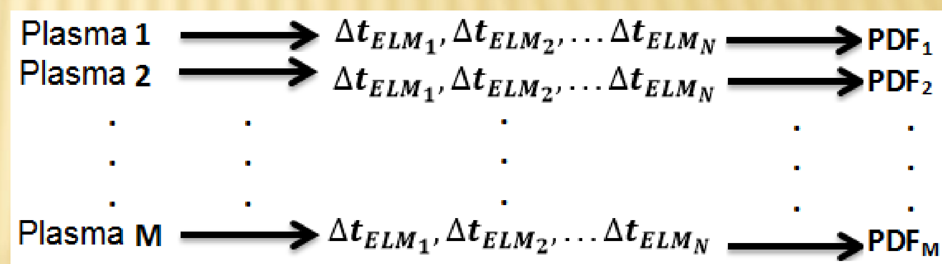
Predictors	Distance measure	k	Classification success rate (%)		
			I	III	Avg
μ	Euclidean	1	89.2	69.2	84.0
μ, σ	Euclidean	1	89.2	69.2	84.0
$\boldsymbol{\mu, \sigma}$	**GD**	**1**	**95.9**	**84.6**	**93.0**

Extraction of ELM waiting times

A threshold-based ELM detection algorithm is used for extracting the ELM waiting times



Gaussian and 2-parameter Weibull distributions are fit to the ELM waiting times

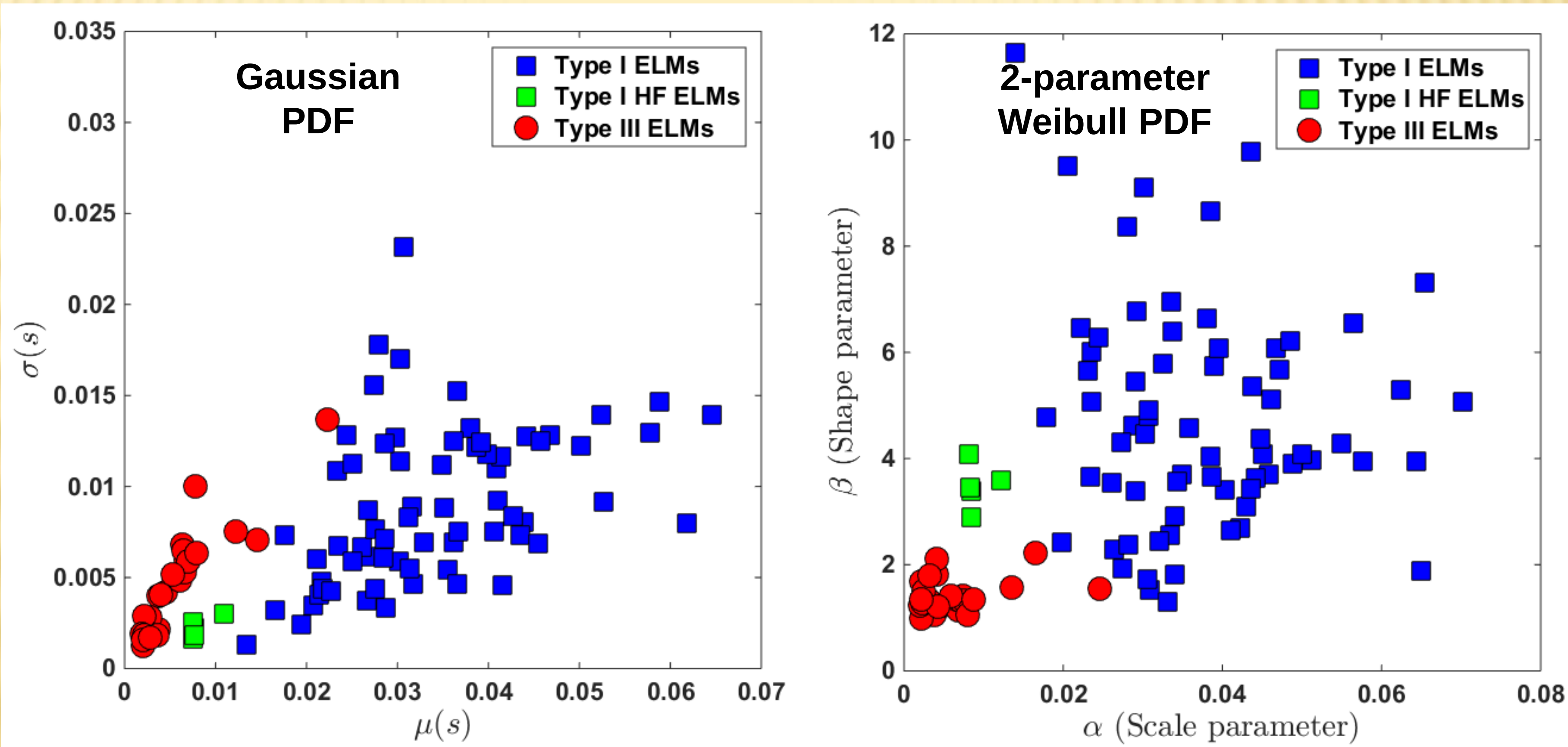


Free parameters of the distributions

↓

Maximum likelihood estimation (MLE)

ML estimates for distribution fits to ELM waiting times



GD-based kNN classification using ELM waiting times

Predictors	Distance measure	k	Classification success rate (%)		
			I	III	Avg
μ	Euclidean	1	95.9	84.6	93.0
μ, σ	Euclidean	1	95.9	84.6	93.0
$\boldsymbol{\mu}, \boldsymbol{\sigma}$	**GD**	**1**	**97.3**	**96.2**	**97.0**
β, α	Euclidean	1	94.6	80.8	91.0
β, α	GD	1	97.3	92.3	96.0

CONCLUSIONS

- **Complete distributions of plasma quantities contain more information than mere averages**

- **GD is an apt similarity measure for comparing probability distributions**

- **A fast, standardized classification scheme for ELM types is presented which:**
  - ❑ **reduces the effort of ELM experts in identifying ELM types**
  - ❑ **can complement phenomenological approaches**

- **The presented method is generic and can also be applied to other classification problems in fusion and astronomy**