DE LA RECHERCHE À L'INDUSTRIE

cea

First WEST experimental data analysis using machine learning algorithms

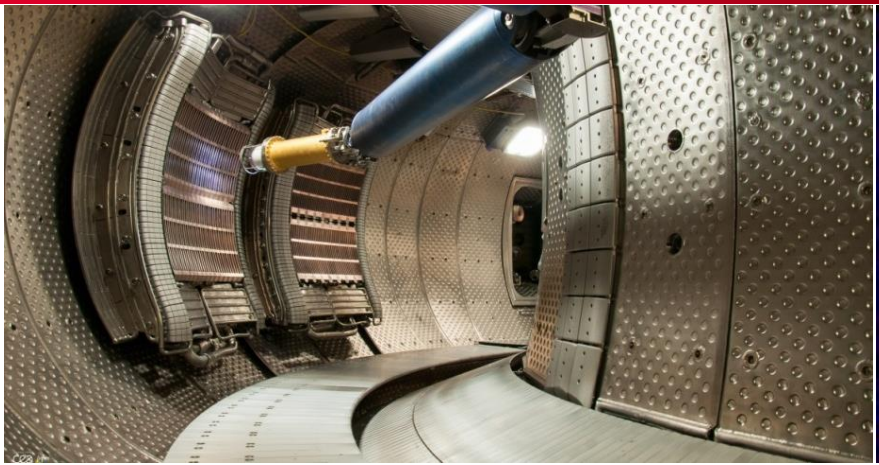
J. Morales, L. Fleury, P. Maini, J.-F. Artaud, H. Ancher, B. Faugeras, E. Nardon, F. Imbeaux and the WEST team

13 May 2019

IRFM

www.cea.fr

WEST
Tungsten (W) Environment Steady-state Tokamak



I_p (q_{95}~2.5)	1 MA
B_T	3.7 T
R	2.5 m
a	0.5 m
A (aspect ratio)	**5 – 6**
n_{GW} (1MA)	1.5 $10^{20}m^{-3}$
P_{ICRH}	**9 MW**
P_{LHCD}	**7 MW**
$time_{flattop}$ (0.8 MA)	**1000 s**

WEST diagnostics

Diagnostics organized into groups related to their measurement principle

Magnetics

Neutrons

X-Ray measurements

- Soft X-ray (impurities)
- Hard X-ray (fast electron)

Optical

- Interferometry (density)
- Polarimetry
- Thomson scattering
- Visible cameras
- Infrared cameras

Spectroscopy

- Visible (plasma edge)
- UV (plasma edge-core)
- 2D X-ray (T ions at plasma core)

Bolometry (plasma radiation)

Microwave

- Reflectometers (density)
- Electron cyclotron emission (T electrons)

Plasma Wall Interaction

- Reciprocating probes
- Pecker probes
- Langmuir. probes
- Calorimetry
- Thermocouples
- Barometry
- Mass spectroscopy

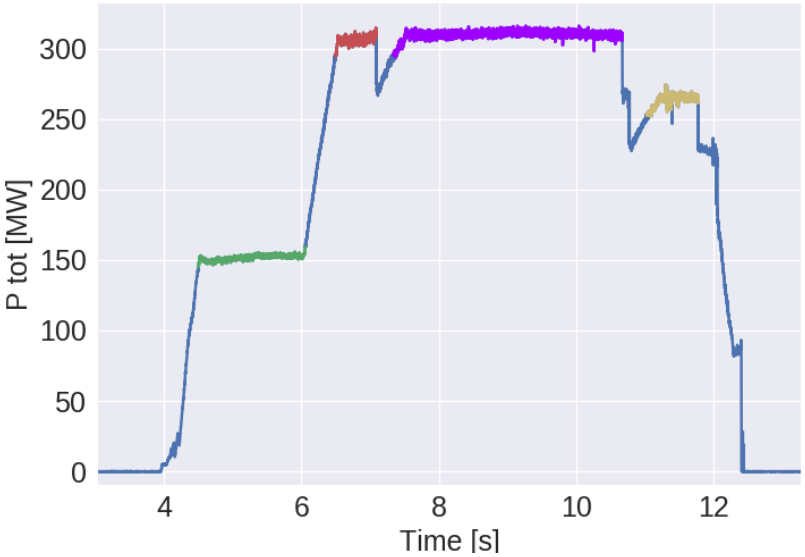
Total **50 diagnostics**

WEST data signal example

First use of a pattern recognition tool

- We are interested on **steady-states** reached by the plasma ► need for a **plateau recognition algorithm**

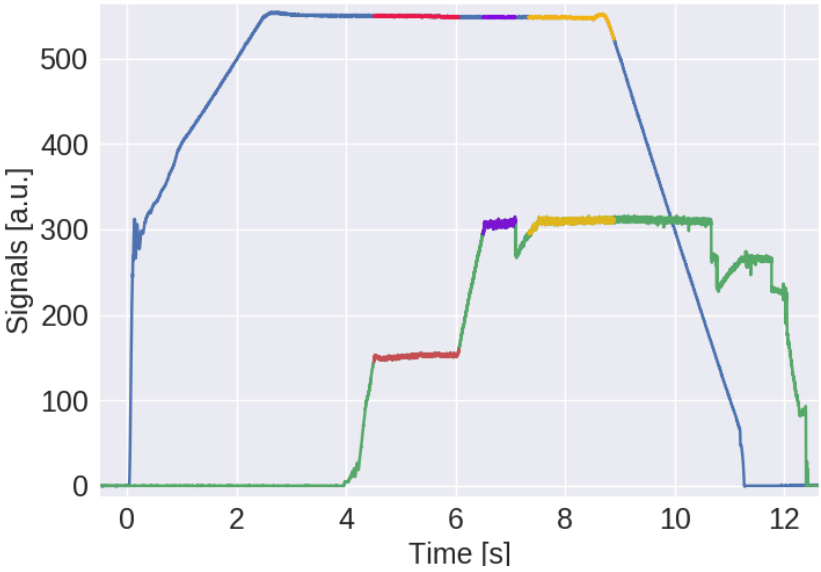
- In this case we use a Kernel density linear estimator provided by `scikit-learn`



WEST data signal example

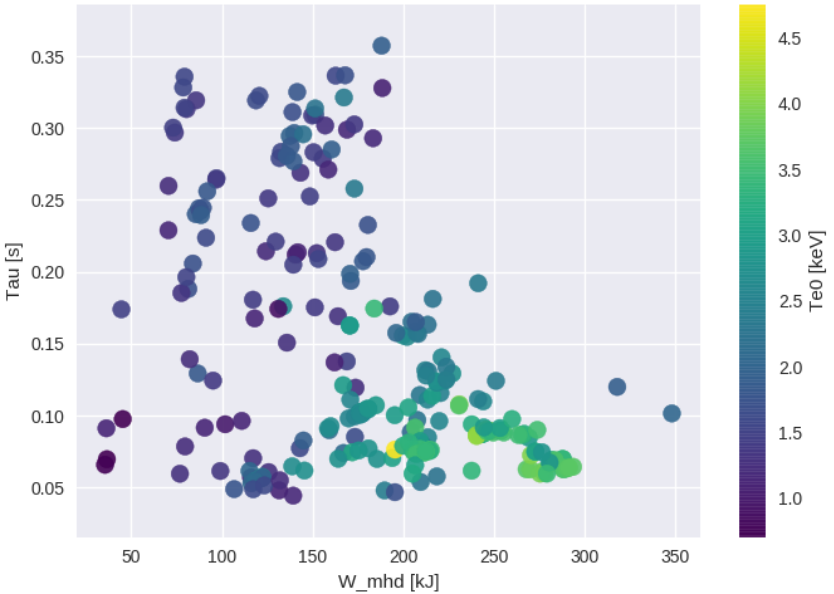
First use of a pattern recognition tool

- We are interested on **steady-states** reached by the plasma ► need for a **plateau recognition algorithm**
- In this case we use a Kernel density linear estimator provided by `scikit-learn`
- Also we need to find steady-states taking into account more than one single signal



WEST reduced database

- Using the plateau recognition algorithm a reduced WEST database is constructed from statistical moments calculated at each detected plateau
- We are interested in studying the confinement quality using the plasma **stored energy W_{mhd}** and the **confinement time Tau** $= \frac{W_{mhd}}{P_{tot}}$ (main target features)
- First constructed reduced database: **36 features** (diagnostic measurements) and **254 samples** (detected plateaus)



Plot example of mean quantities at detected plateaus

First data analysis

Main feature detection for W_{mhd} target

Three different feature selection algorithms are tested:

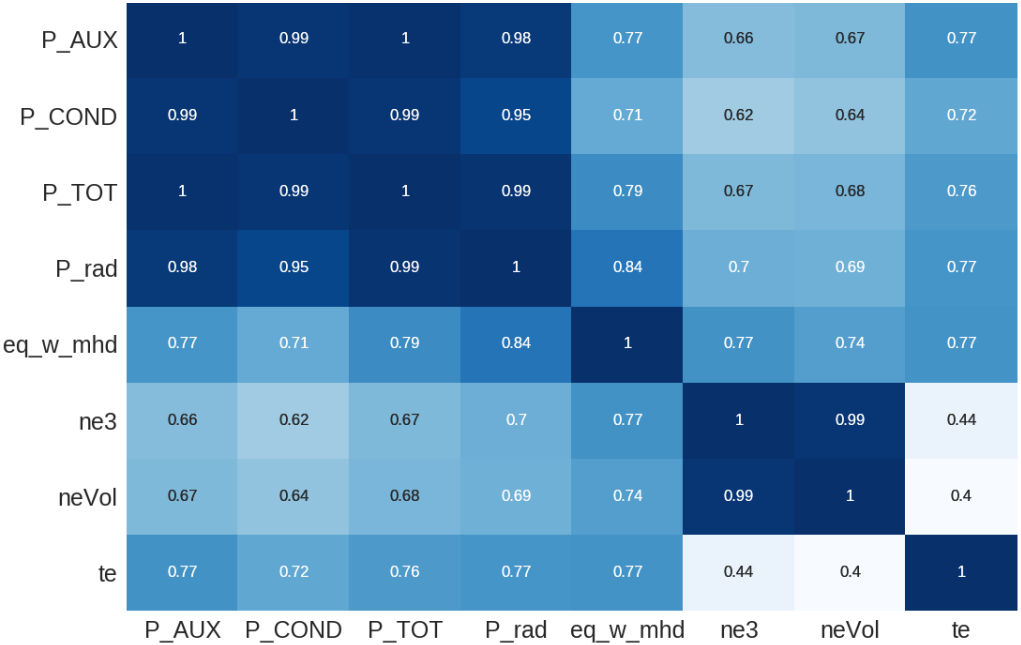
- **Pearson correlation**: measures the linear correlation between variables
- **Mutual information**: measures linear and non-linear dependences between variables
- **Random forest**: feature importance estimated by how much the tree node that use that feature reduce impurity on average in all decision trees

First data analysis

Main feature detection for W_{mhd} target

Three different feature selection algorithms are tested:

- **Pearson correlation**: measures the linear correlation between variables
- **Mutual information**: measures linear and non-linear dependences between variables
- **Random forest**: feature importance estimated by how much the tree node that use that feature reduce impurity on average in all decision trees

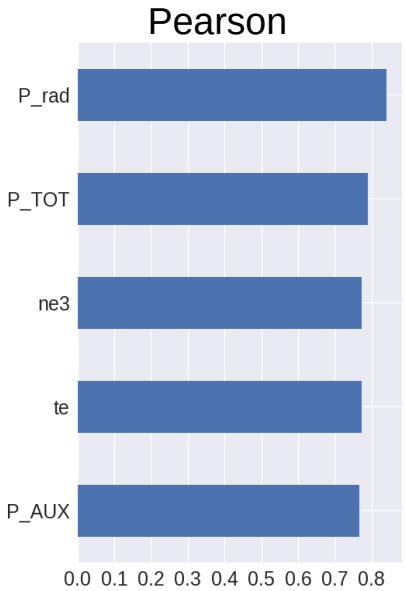


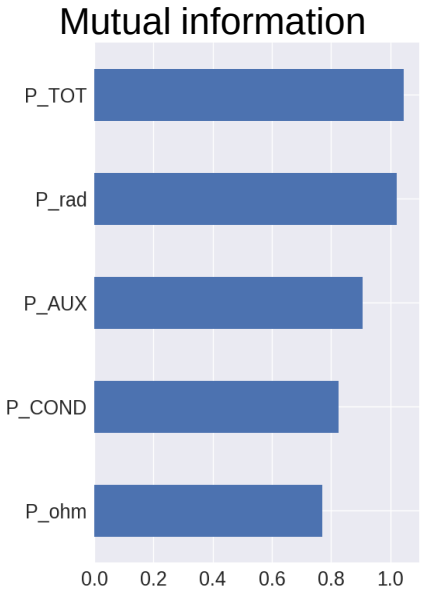
First data analysis

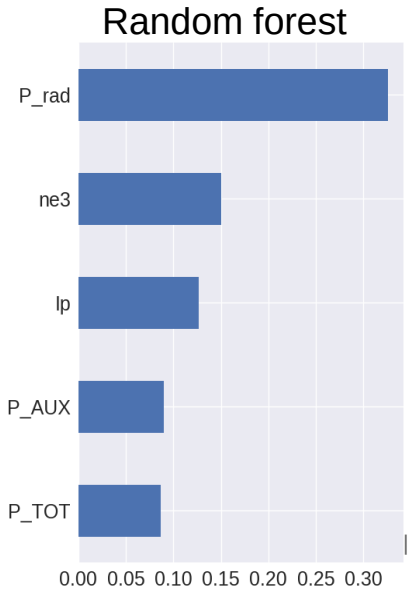
Main feature detection for W_{mhd} target

Three different feature selection algorithms are tested:

- **Pearson correlation**: measures the linear correlation between variables
- **Mutual information**: measures linear and non-linear dependences between variables
- **Random forest**: feature importance estimated by how much the tree node that use that feature reduce impurity on average in all decision trees (cross validation mean error of 14%)

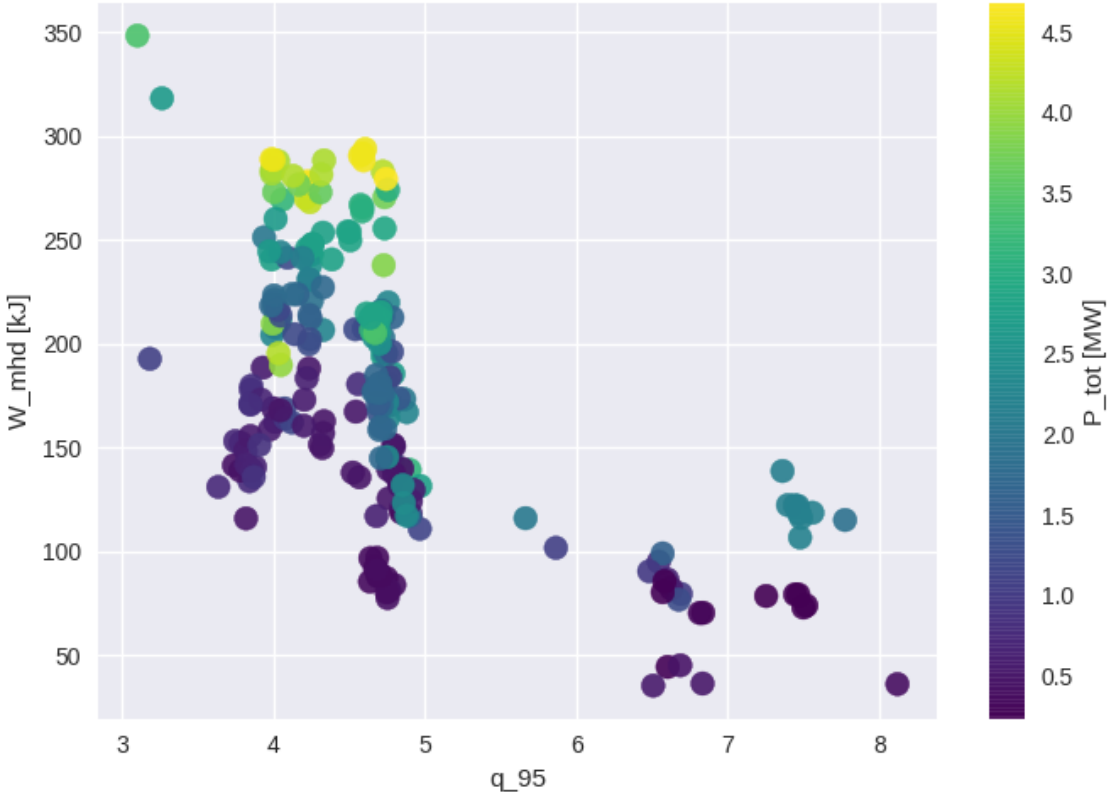






First data analysis

W_{mhd} evolution as a function of main detected features

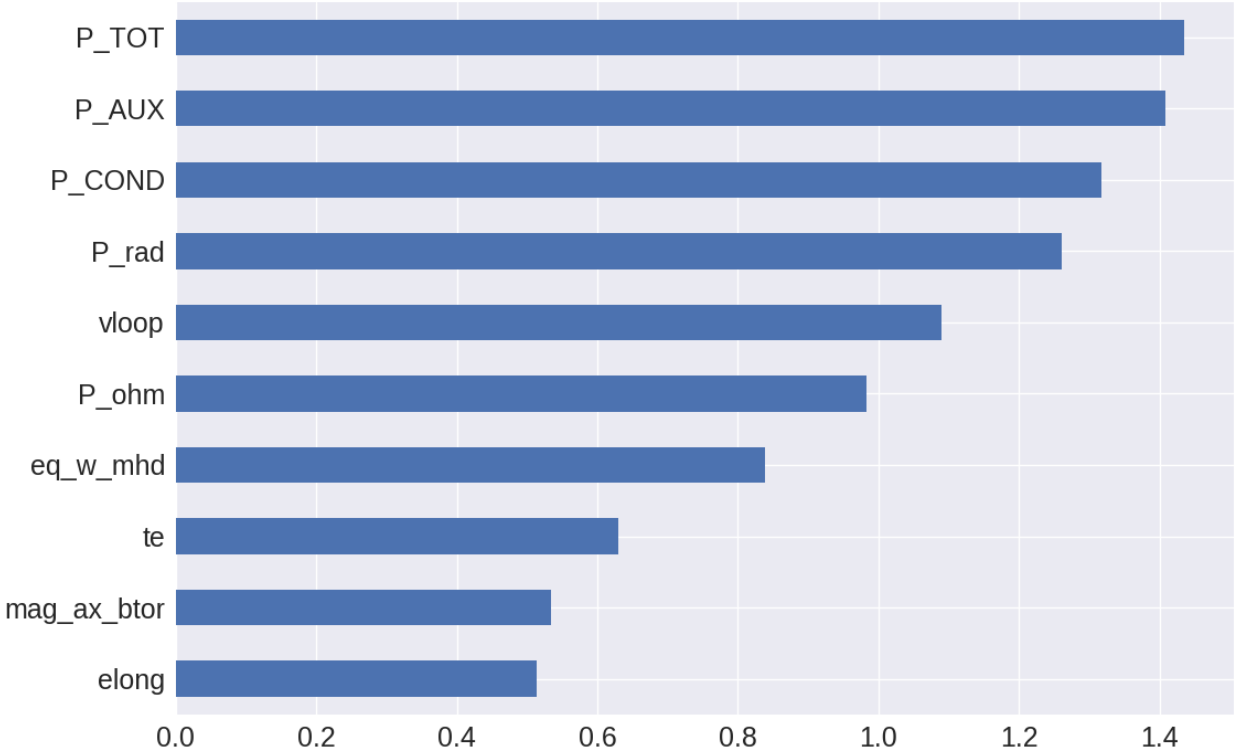


Stored plasma energy decreases as a function of q_95 (edge safety factor) for a given total power level

First data analysis

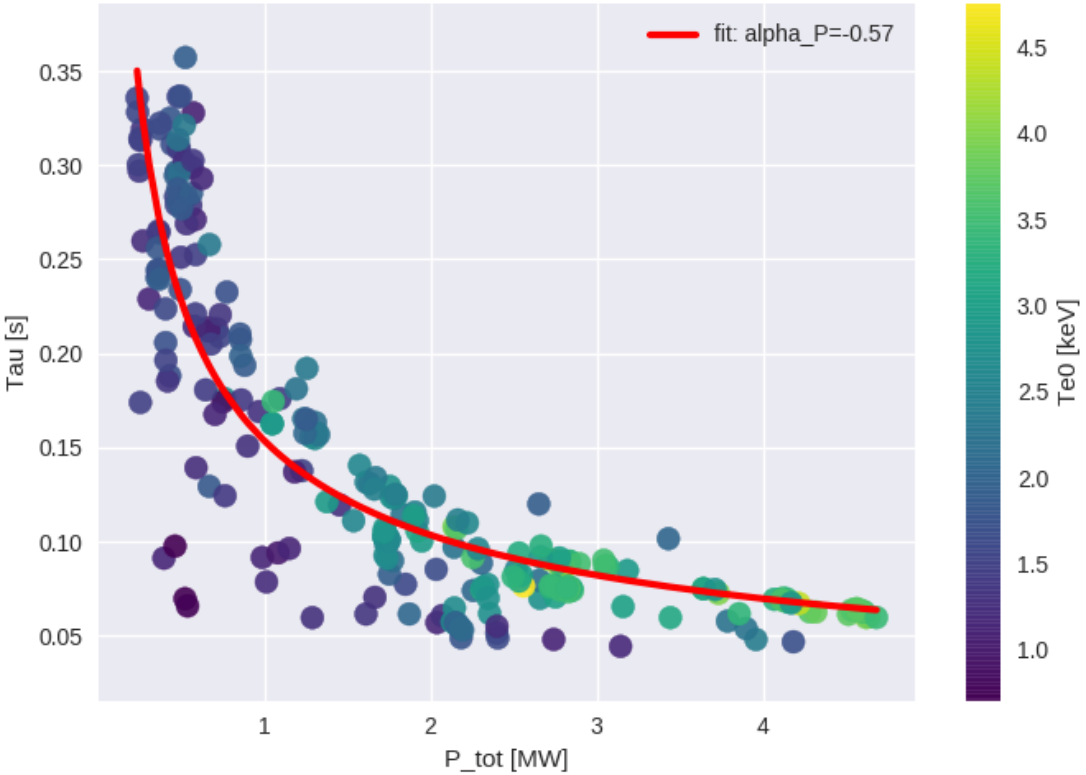
Main feature detection for confinement time (Tau) target

Mutual information is used to detect nonlinear relations between variables



First data analysis

Confinement time scaling with total power



In the literature total power exponents in scaling laws vary between: -0.73 and -0.5 (our value: -0.57)

Conclusion and perspectives

Summary

- To carry out plasma steady-state analyses a WEST reduced database is created. To this aim a plateau recognition algorithm for time signals is developed
- The plasma energy content measured by W_{mhd} and the confinement time (Tau) are the selected targets
- To detect the principal features impacting the targets three feature selection algorithms are tested (Pearson correlation, mutual information and random forest)
- The evolution of the target variables as a function of the main features is analyzed. Comparison with literature shows good agreement in the total power exponent computed for Tau scaling

Conclusion and perspectives

Summary

- To carry out plasma steady-state analyses a WEST reduced database is created. To this aim a plateau recognition algorithm for time signals is developed
- The plasma energy content measured by W_{mhd} and the confinement time (Tau) are the selected targets
- To detect the principal features impacting the targets three feature selection algorithms are tested (Pearson correlation, mutual information and random forest)
- The evolution of the target variables as a function of the main features is analyzed. Comparison with literature shows good agreement in the total power exponent computed for Tau scaling

Perspectives

- More diagnostics data will be added to reduced database
- Failure and error detection algorithms for diagnostics
- Improve main feature selection, regression for targets of interest, WEST pulses classification… More ideas are more than welcomed!