A short presentation of my work experience and current projects

Emmanuele Peluso

WIP 04.03.24

Outline

- Personal introduction
- Key-words: data analysis, diagnostics.
    - Symbolic Regression and SVM
    - Causality detection
    - Bolometry and Tomography
- Work in progress

Personal Introduction. Highligts from my CV

- **Background**
  - Bachelor ('Calibrazione di una sorgente di neutroni da 14 MeV per misure di sezioni d'urto per la fusione nucleare', *FNG, ENEA, data analysis*) and Master degrees ('Scattering collettivo di onde millimetriche su instabilità magnetoidrodinamiche sul TOKAMAK FTU', *FTU, ENEA, data anlaysis*) in Physics, La Sapienza, Roma.
  - PhD in Industrial Engineering ('Genetic Programming for Symbolic Regression in Nuclear Fusion', *JET, CCFE, data analysis – European label*), Tor Vergata.

- **Work Experiences in Fusion** (Fellowship in 2011 (JET restart), Research fellowships (2014-2020) and RTDa at Tor Vergata, (2022-2023):
  - JET (Mobility and missions):
    - From 2013 to 2019 during different missions up to around 1y and a half on site, working quite often in control room as VSO.
  - AUG (Missions in 2019 almost three months on site, later, due to covid, only remotely in 2020-2022)
  - Deputy coordinator of the EUROfusion task on 'EUROfusion database on Profile and Confinement within the Work-Packages JET1 and MST1 (now WPPrIO)' in 2019, then the coordinator left and I stood as coordinator.
  - Conferences, at least one per year either attending them or joining remotely during covid.

Outline

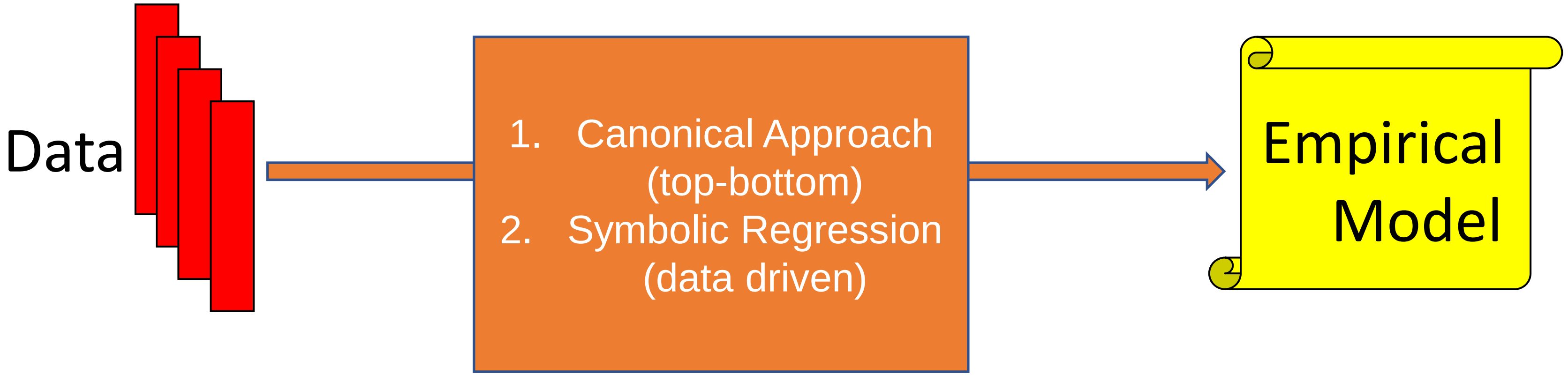
- Personal introduction

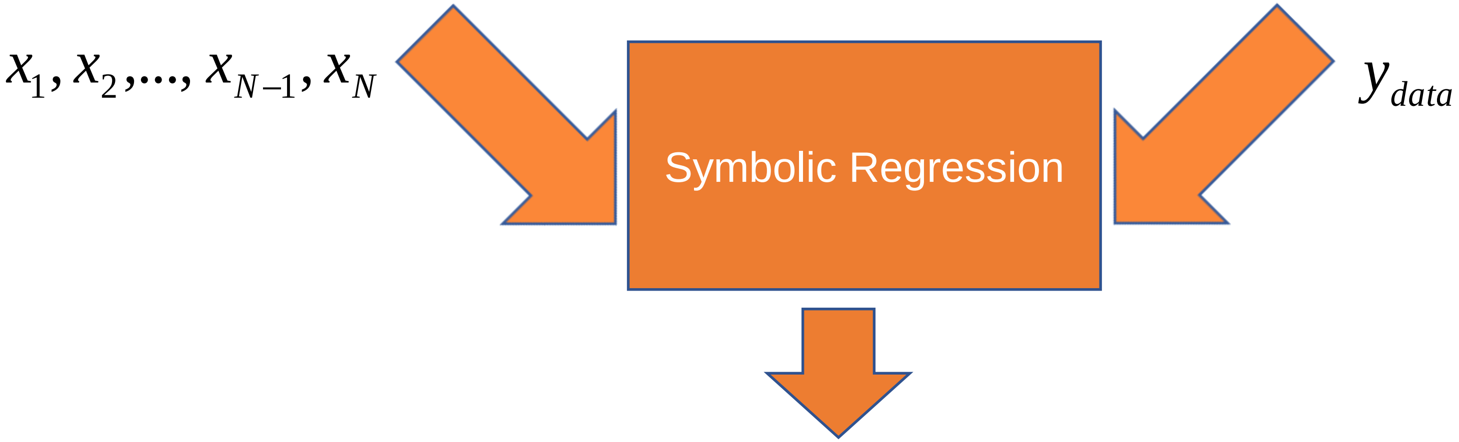
- Key-words: data analysis, diagnostics.
  - Symbolic Regression and SVM
  - Causality detection
  - Bolometry and Tomography

- Work in progress

Data Analysis: Symbolic Regression

- It is a Machine Learning, data driven technique.

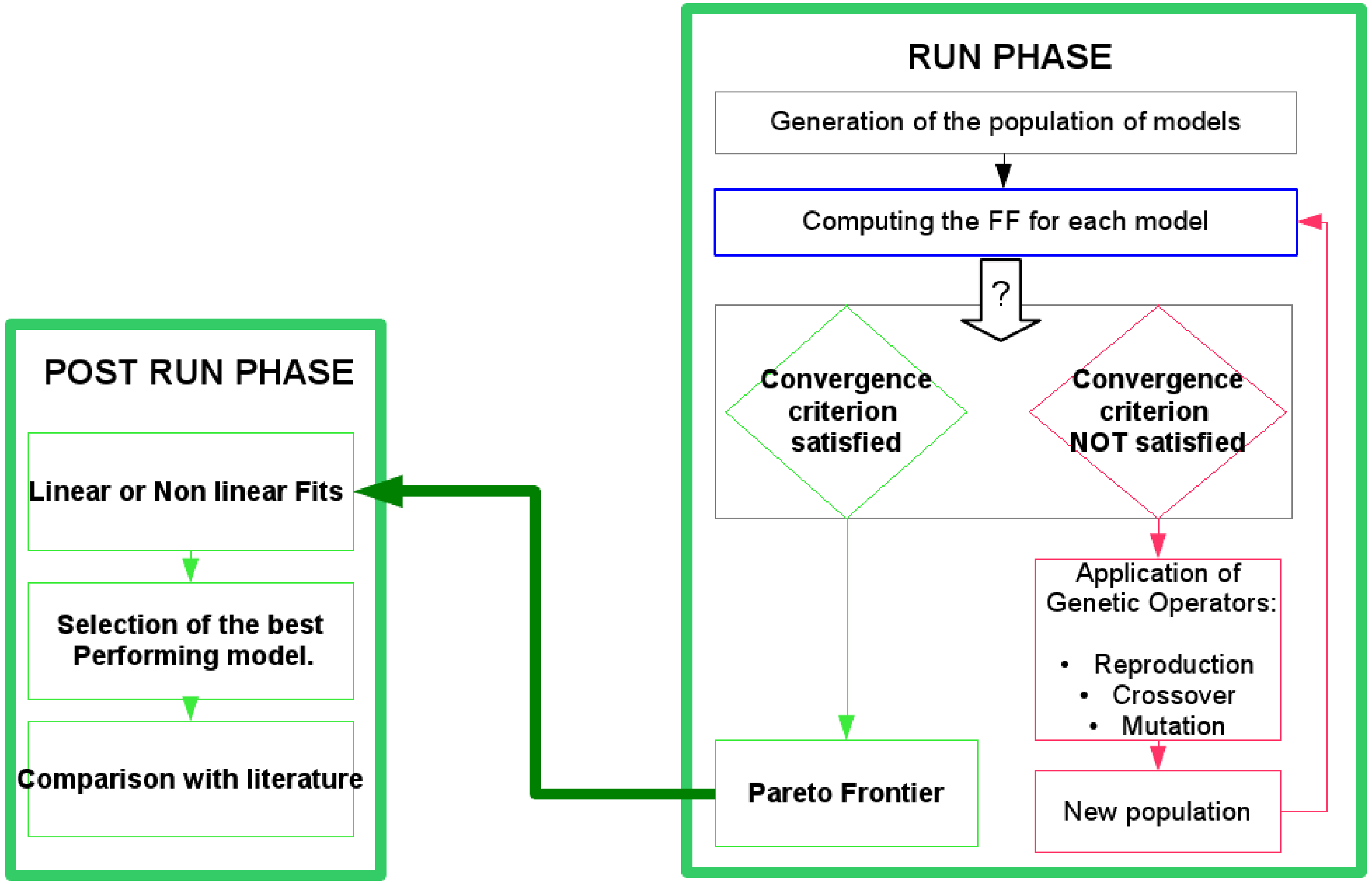


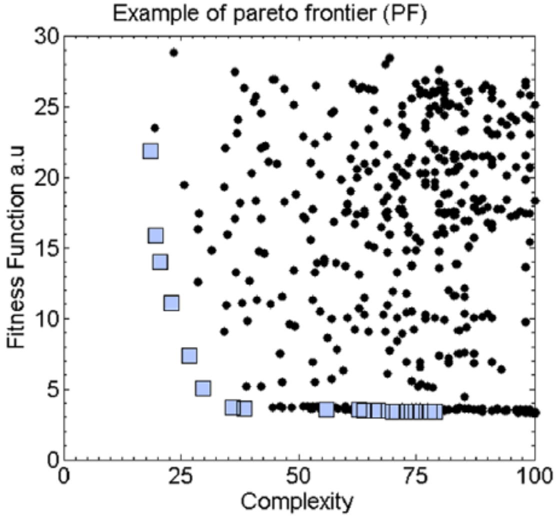


$$y_{\mathrm{mod}\ by\ SR} = \alpha_1 \cdot \sin(\alpha_2 \cdot x_1^{\alpha_3}) + \alpha_4 \cos(\alpha_5 \cdot x_1^{\alpha_6} \cdot x_2^{\alpha_7})$$

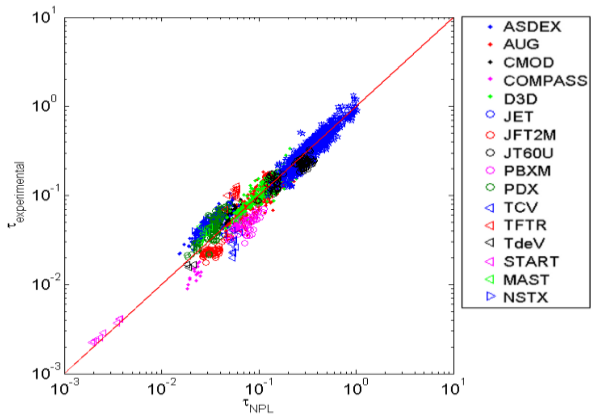
Data Analysis: Symbolic Regression

- SR can be shortly described by this Flowchart





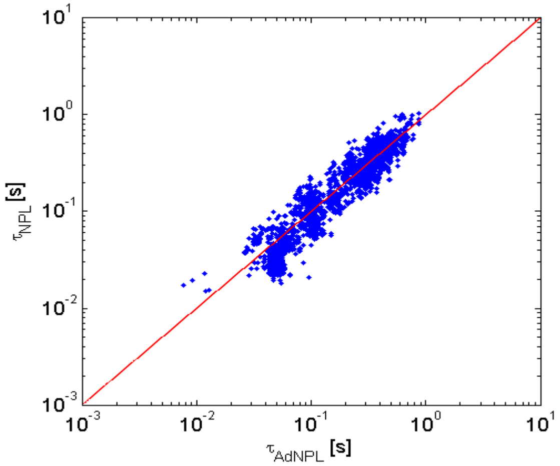
Data Analysis: Symbolic Regression, why and an application



ipb98y2[1]	$\tau_E = 5.62 \cdot 10^{-2} I^{0.93} B^{0.15} n^{0.41} M^{0.19} R^{1.97} \varepsilon^{0.58} k_a^{0.78} P^{-0.69}$
NPL	$\tau_E = 3.67_{3.66}^{3.69} \cdot 10^{-2} I^{1.01_{0.99}^{1.02}} R^{1.73_{1.71}^{1.75}} k_a^{1.45_{1.41}^{1.49}} P^{-0.74_{-0.75}^{-0.72}} h(n,B)$ $h(n,B) = n^{0.45_{0.44}^{0.46}} \cdot \left(1 + e^{-9.40_{-9.69}^{-9.11} \cdot (n/B)^{-1.37_{-1.41}^{-1.32}}}\right)^{-1}$

	AIC	BIC	MSE [10^{-3} s^2]	KLD
ipb98y2	-19416.86	-19362.86	1.866	0.0337
NPL	-19660.03	-19599.04	1.724	0.0254

•*The AIC and BIC are two statistical estiamtors rewarding accuracy and precision respectively in terms of average residuals and standard deviation. The KLD is then a measure of dissimilarity (between the pdfs of data and the model)*

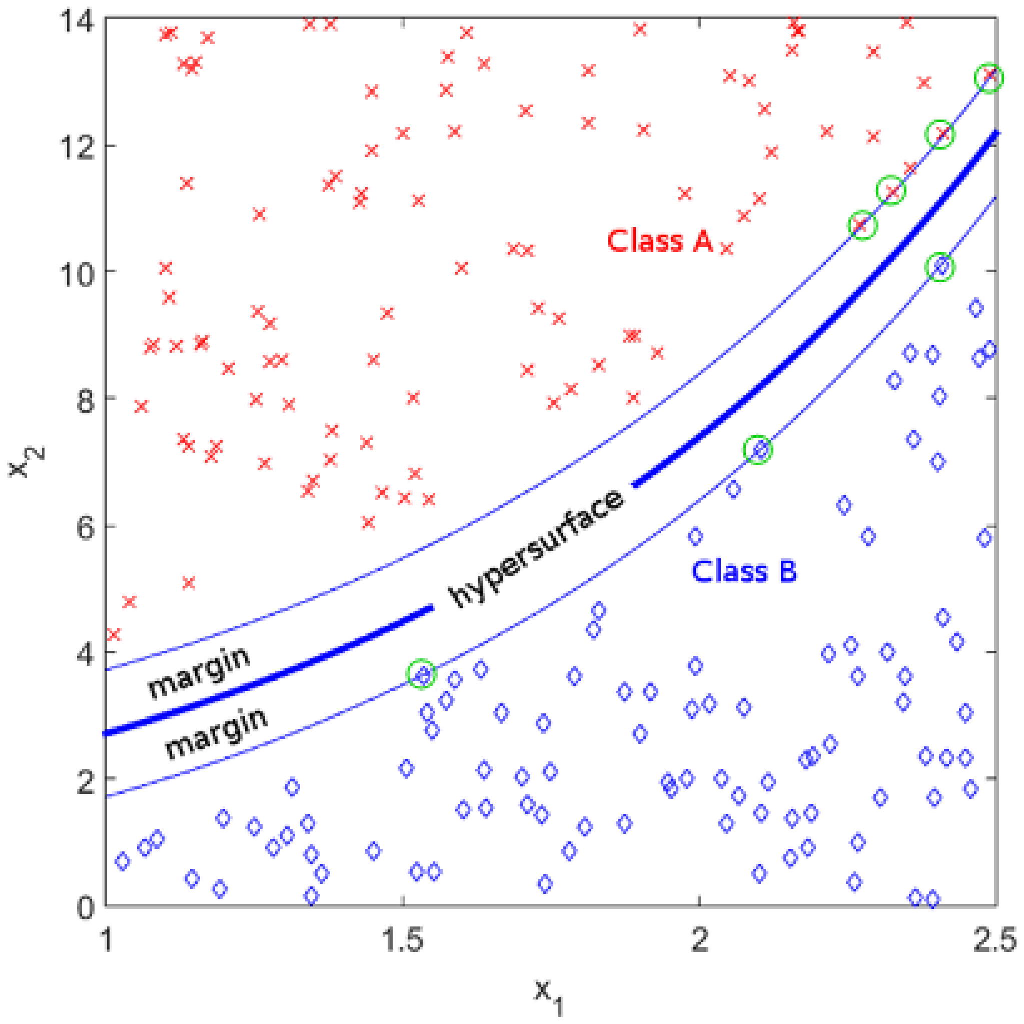


equation	τ [s]
Dimensionless models	$2.97_{2.78}^{3.16}$
Dimensional model	$2.83_{2.42}^{3.31}$
ipb98y2	3.66

- ~20% reduction of the estimate of the confinement time to ITER using data not with ILW[1]
- Recently the ITPA-20 scalings confirmed a reduction of the prediction to ITER [2] .

[1] A. Murari *et al* 2015 *Nucl. Fusion* **55** 073009
[2] Verdoolaege G. et al, "*The updated ITPA global H-mode confinement database: description and analysis*" 2021 Nucl. Fusion 61 076006,

- SVM is a Machine Learning technique.



•SVMs can be used here for classification.

•SVMs find the multidimensional hypersurface separating classes of data.

•For linearly separable problems this hypersurface is identified as the one having the maximum distance (margins) between the closest points of the two different classes (Support Vectors).

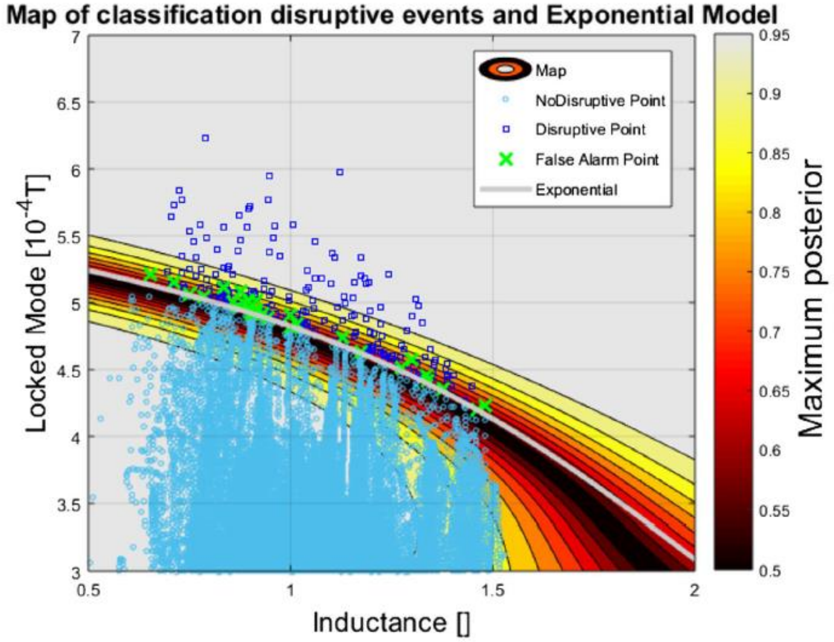
•An example on the left. The data belonging to the two classes are reported in red and blue in the 2D feature space of the coordinates (x1,x2).

•Also non linearly separable problems can be studied using a kernel trick

•In this case the hypersurface function f(x) separating classes, is described as a sum of non linear (kernel) functions, each centered on a SV.

Data Analysis: Disruptions, SVM and SR

- Once the points of the hypersurface are numerically estimated, it is possible to run a SR trying to find a simpler analytical expression for the boundary[3]



$$LM(l_i) = a_0 e^{a_1 l_1^{a_2}}$$

$$a_0 = 5.4128 \pm 0.0031;$$
$$a_1 = -0.11614 \pm 0.00085;$$
$$a_2 = 2.21 \pm 0.011;$$

C29-C30

Method	Success rate	Missed	Early	Tardy	False	Mean (ms)
SVM	97.9% (183/187)	0% (0/187)	0% (0/187)	2.1% (4/187)	2.8% (29/1020)	335
Equation	97.9% (183/187)	0% (0/187)	0% (0/187)	2.1% (4/187)	2.8% (29/1020)	336

[3] Murari, A.,. *et al.* Data driven theory for knowledge discovery in the exact sciences with applications to thermonuclear fusion. *Sci Rep* **10**, 19858 (2020). https://doi.org/10.1038/s41598-020-76826-4

Outline

- Personal introduction

- Key-words: data analysis, diagnostics.
  - Symbolic Regression and SVM
  - Causality detection
  - Bolometry and Tomography

- Work in progress

Data Analysis: Causality detection

- Many experiments are performed to either/both asses the causal relationship of one physical quantity evolving in time to another one or/and the possible synchorinization between them.

- In fact, even in the case of time series, quantifying causality remains a difficult issue.

- For the time series, Norbert Wiener proposed a concept of causality based on prediction:

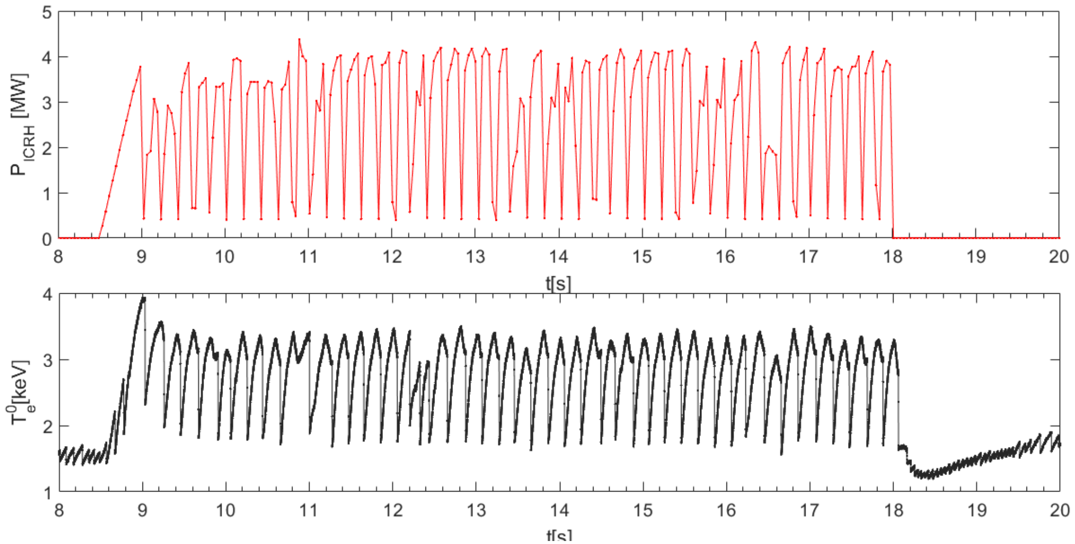
IF the knowledge of the past of signal “J” helps predicting signal “I”, THEN “J” is considered causal to “I”.

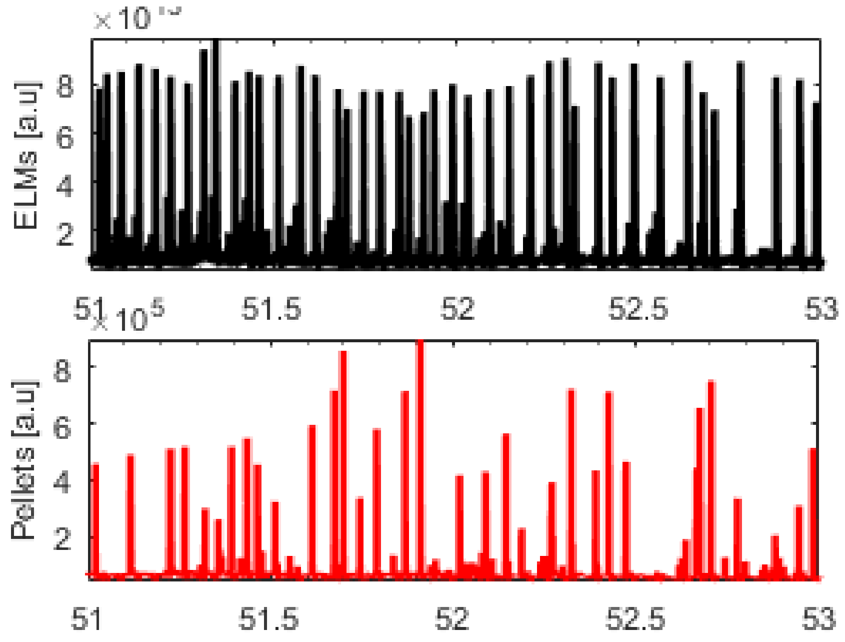
- This approach was formalised by Clive Granger and is called Granger causality.

- Studying the behaviour of two or more dynamical systems, the concept of “synchronization” occurs when a relationship of coupling can be defined, specifically, the **generalized synchronization** refers to completely different systems where the dynamical variables of one system **(the response system)** are determined by the other system **(the drive system)**

Data Analysis: Causality detection

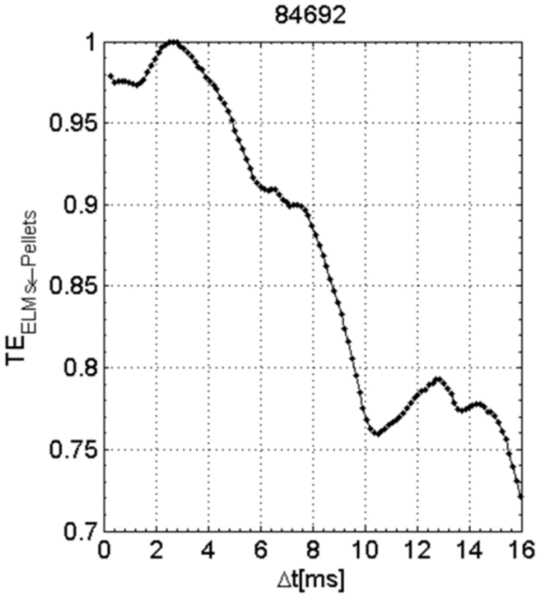
- Two examples are reported: either the sawteeth pacing and ELMs pacing[3].



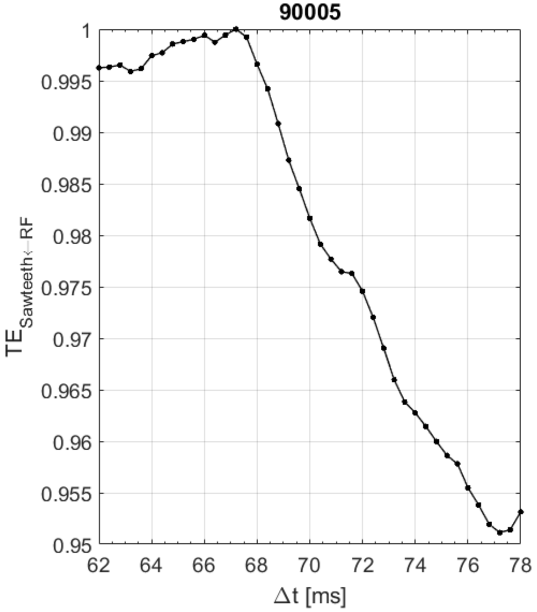


Data Analysis: Causality detection

- In this case the Transfer Entropy $TE_{i\to j} = I\left(i_n^{(l)}; j_{n+1} | j_n^{(k)}\right)$ has been applied to estimate the the maximum time interval into which two physical quantities are coupled (i.e syncronized) and in which one observable can be thougth as the «drive mechanism» of the second observable.
- Considering the sawteeth pacing, I also applied a further refinement[4] trying to decouple the effect of natural, not paced, sawteeth[5].



TE example for Pellets pacing. The highest time interval for the coupling is (2.7 ± 2)ms



TE example for sawteeth pacing. The highest time interval for the coupling is (67 ± 9)ms

[4] Entropy 2020, 22(8), 865;
[5]E. Peluso *et al* 2022 *Plasma Phys. Control. Fusion* **64** 084002

Outline

- Personal introduction

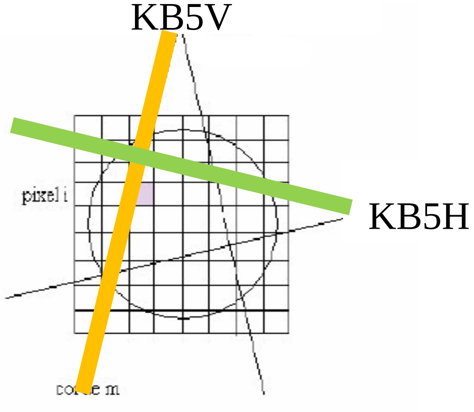
- Key-words: data analysis, diagnostics.
  - Symbolic Regression and SVM
  - Causality detection
  - Bolometry and Tomography

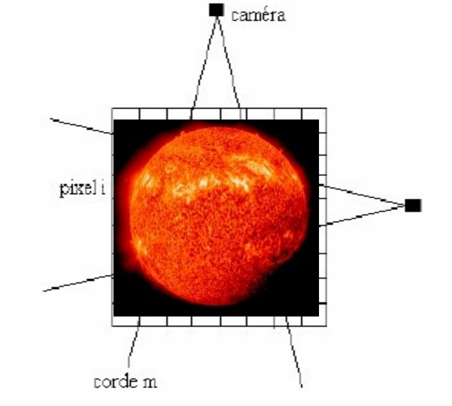
- Work in progress

Diagnostics: Bolometry and Tomography

- Tomography is a technique used to perform the inversion of bolometric measurements in order to infer indirect quantities (emissivity profiles, radiated powers,…) from obtained images (tomograms)

$\overbrace{KB5H\ \&\ KB5V\ meas.}^{M}$ → $\overbrace{Tomographic\ code}^{A^{-1}}$ → $\overbrace{Tomogram}^{S}$





Diagnostics: Bolometry and Tomography

- Maximum Likelihood (ML) assumptions: the emission and the measures are ruled by Poisson processes;
- The probability of obtaining the measurement $g = \{g_m | m = 1, \dots, N_d\}$

if the image is $f = \{f_n | n = 1, \dots, N_p\}$ is given by the likelihood function:

$$L(g/f) = \prod_k \frac{1}{g_k!} (\bar{g})^{g_k} \times exp(-\bar{g}) \quad \Rightarrow f_{ML} = argmax_f L(g/f)$$

- Practically then:

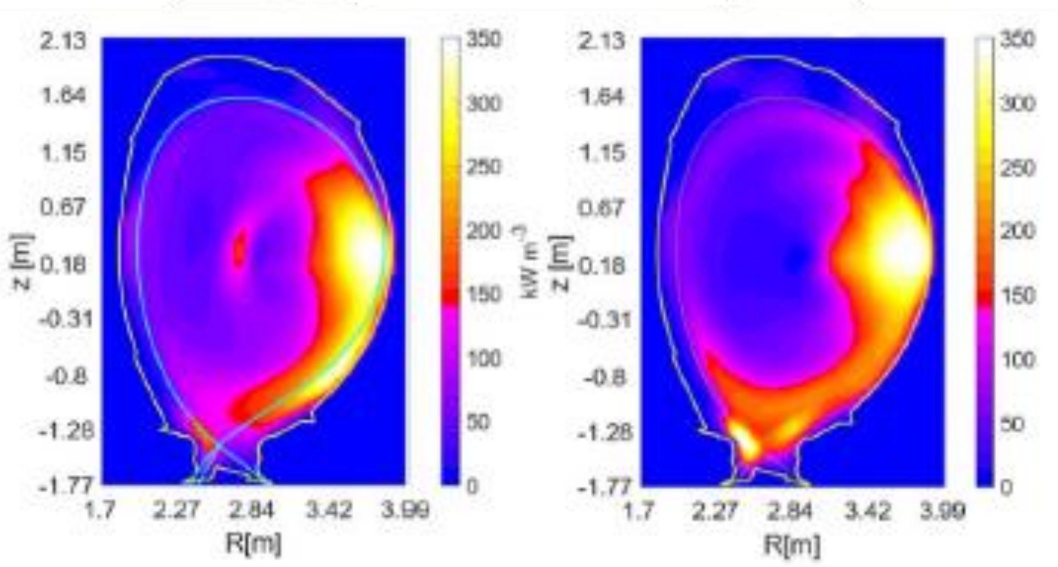
$$f_n^{(k+1)} = \frac{f_n^{(k)}}{\sum_m H_{mn}} \sum_m \left(g_m / \sum_j H_{mj} f_j^{(k)} \right) H_{mn}$$

$$\begin{aligned} \varepsilon^{(k+1)} &= \\ &= diag\left[\hat{f}^{(k)}\right] diag[s^{-1}] H^T diag\left[H\hat{f}^{(k)}\right]^{-1} n \\ &+ \left[I - diag\left[\hat{f}^{(k)}\right] diag[s^{-1}] H^T diag\left[H\hat{f}^{(k)}\right]^{-1} H \right] \varepsilon^{(k)} \end{aligned}$$

Diagnostics: Bolometry and Tomography

- The application of the ML method to bolometric data has also been extensively studied. Methodologies have been developed to take into account the effect of noise, systematic errors, uncertainties on the position of the magnetic surfaces, the failure of channels, as well as to minimize the occurrence of artefacts[6-9] and more recently, also to deal with symmetry breaking events.

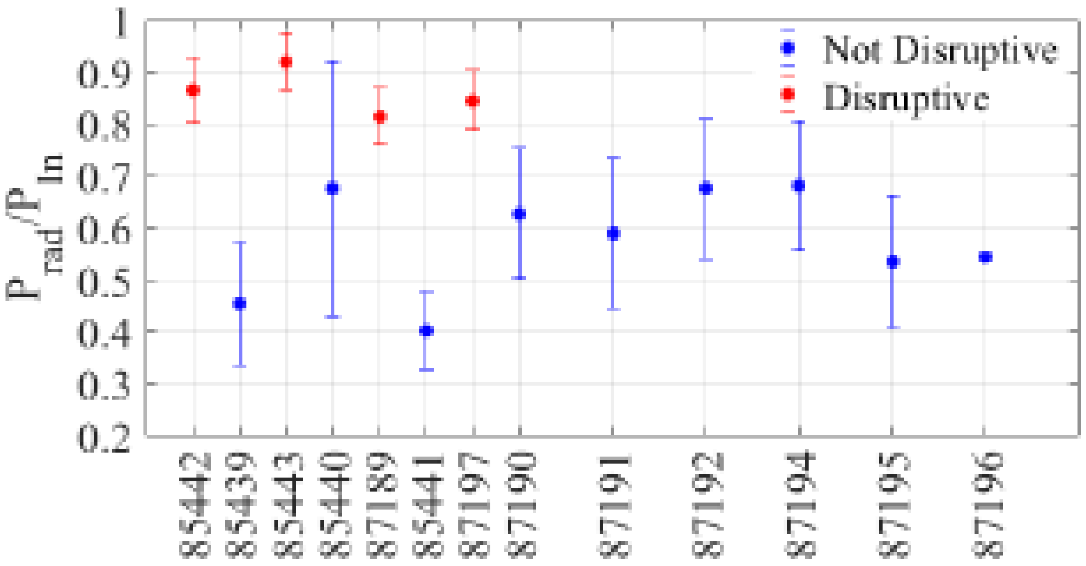
- The code has been transfered to AUG[10]



- Pulse 92398 at 7.5 s. Left: not optimized tomograms effected by a core artefact which occurrence can influence transport studies; right: optimized tomogram [9].

[6] T. Craciunescu, et al. , Rev. Sci. Instrum. 89, 053504 (2018)
[7] E. Peluso , et al. , Rev. Sci. Instrum. 90, 123502 (2019)
[8] E. Peluso, et al., Fusion Engineering and Design, 146, 2124–2129, (2019)
[9] E.Peluso, et al., Plasma Physics. Controlled. Fusion 64 (2022) 045013.
[10] Teddy Craciunescu *et al* 2023 *Phys. Scr.* **98** 125603

Diagnostics: Bolometry and Tomography



- An example related to the thermal stability studies performed by using the ML method on Neon seeded pulses from C32-C33 JET campaigns.
- Radiated fraction estimated 50 ms before the beginning of the current quench (disr) or before the end of the external heating (not disr). Estimated uncertainties are also reported[11]

[11] A. Murari et al 2020 Nucl. Fusion 60 046030

Outline

- Personal introduction

- Key-words: data analysis, diagnostics.
  - Symbolic Regression and SVM
  - Causality detection
  - Bolometry and Tomography

- Work in progress

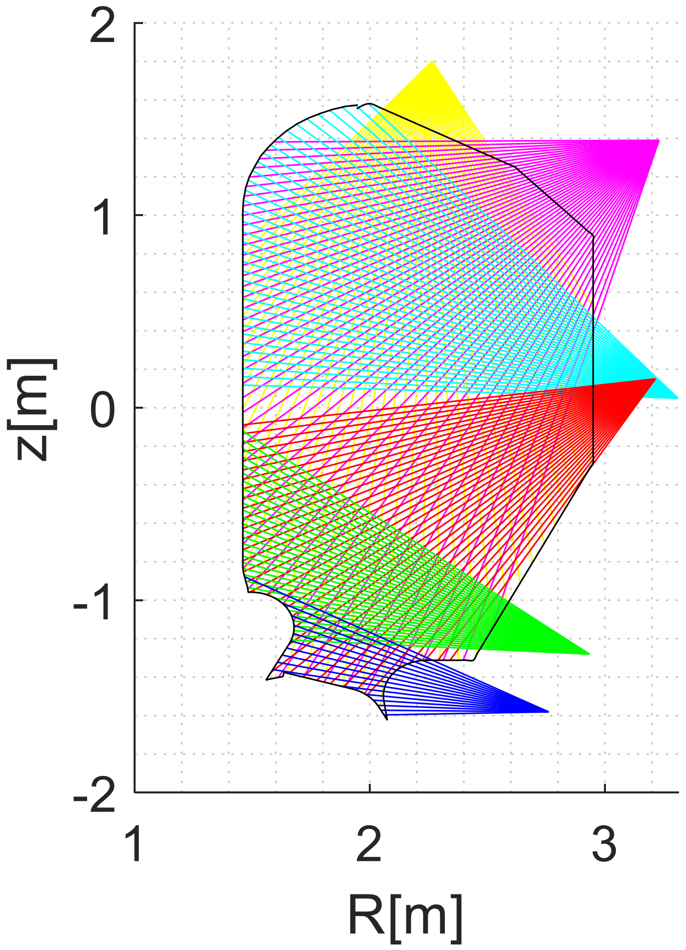
Work In Progress

- I am working on DTT bolometry with G. M. Apruzzese, L.Senni (CNR) V.D’Agostino (Tor Vergata), A.Belpane (RFX), S.Palomba (DTT), M.Gelfusa (Tor Vergata).

- At the moment, we are working on the latest design of the diagnostic and with the 2023 divertor configuration;

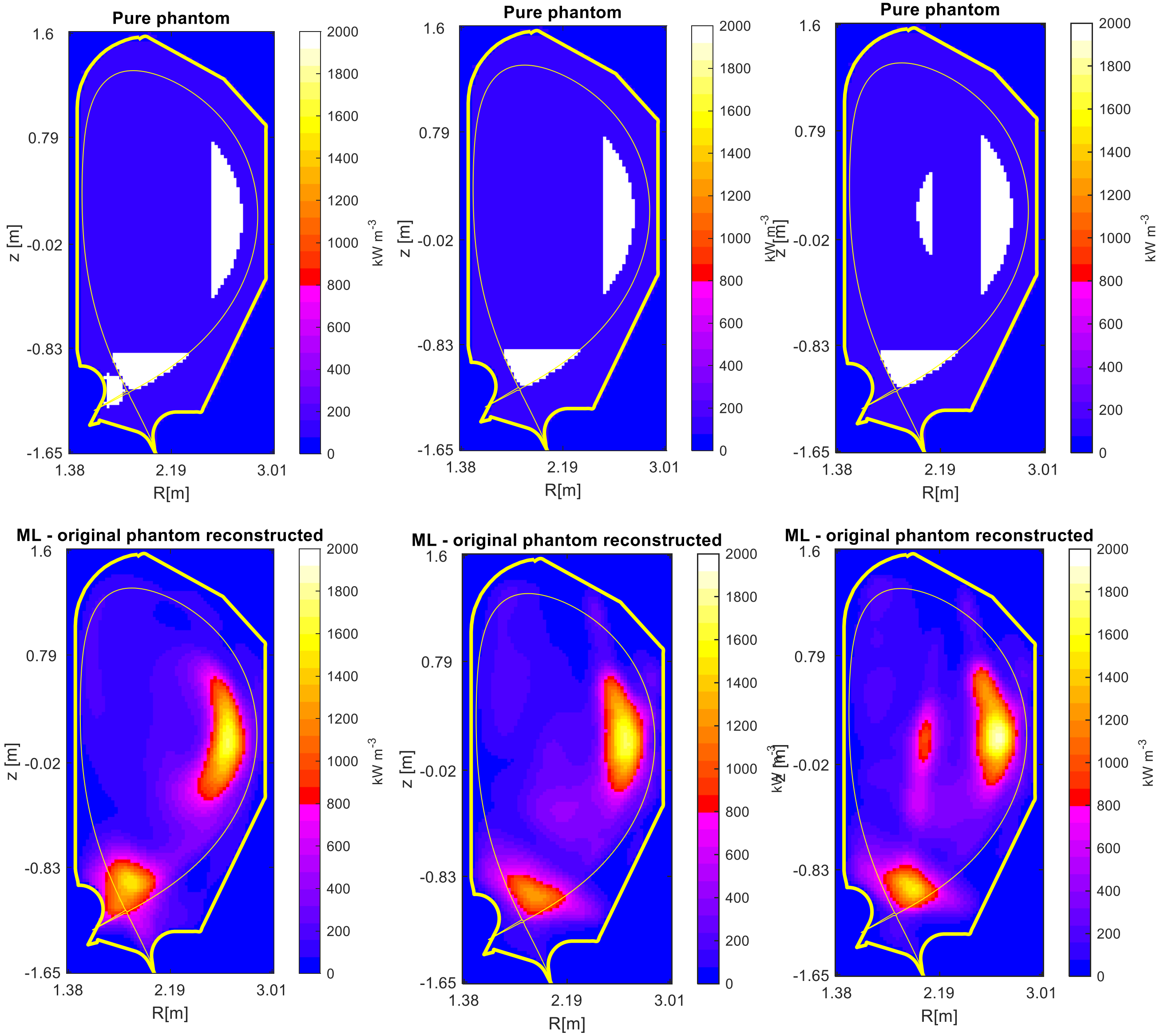
- I am also envolved in the WPPrIO as the EUROfusion confinement database coordinator;

Work In Progress



- 2023 layout on the left (from A.Belpane, V.D’Agostino, and S.Palomba):
  - Yellow,48 [from LFS to HFS] LOSs, P1;
  - Magenta [from top to bottom], 48 LOSs, P2;
  - Cyan, [from top to bottom], 36 LOSs, P3up;
  - Red, [from top to bottom], 36 LOSs, P3down;
  - Green [bottom to top], 32 LOSs, P4;
  - Blue, [from top to bottom], 16 LOSs, DIV1.
- All LOSs have been considered at this stage

Work In Progress



Thank you for your attention !