7th Asia-Pacific Conference on Plasma Physics, 12-17 Nov, 2023 at Port Messe Nagoya

An intrinsically probabilistic approach to analyzing stochasticity and uncertainty in fusion plasmas using information geometry

G. Verdoolaege
Department of Applied Physics, Ghent University
e-mail (speaker): geert.verdoolaege@ugent.be

Various phenomena and processes in plasmas are of a strongly stochastic nature or are affected by considerable uncertainty. Some examples of stochastic phenomena are plasma instabilities and turbulent fluctuations, while measurement of key plasma parameters occurs with instruments that can be quite complex and that may introduce considerable uncertainty on the estimated quantities of interest. In such cases, a probabilistic description of the corresponding data often provides significant added value. In this work, we leverage the powerful methods of information geometry (IG).

In IG, a parametric family of probability densities is interpreted as a Riemannian differentiable manifold [1]. Each point on the manifold corresponds to a specific probability density function (PDF) within the family, and the family parameters represent a coordinate system on the manifold. The Fisher information (covariance of the score) provides a unique metric tensor. This allows a mathematically well-founded distance measure between PDFs: the Rao geodesic distance (GD). In turn, similarity measurement between PDFs in fusion science is essential for quantifying changes in stochastic processes or measurement characteristics under varying plasma conditions. Furthermore, similarity forms the basis of many pattern recognition methods, such as classification and regression.

In this work, we first illustrate our methods by a study of waiting time distributions of a class of MHD instabilities occurring in high-confinement (H-mode) tokamak plasmas, called edge-localized modes (ELMs). The variability of the distribution shape across ELM types is highlighted and modeled with Gaussian and Weibull distributions [2]. Next, a nearest neighbor algorithm based on the GD between the PDF models is shown to successfully identify ELM types [3].

We then proceed to regression analysis based on the Rao GD. This has led to geodesic least squares (GLS) as a new regression technique operating in spaces of probability distributions [4]. Owing to its solid foundations rooted in IG, GLS is robust against outliers and model misspecification, yet it is easily implemented without expert statistics knowledge. We demonstrate the robustness properties of GLS using regression of ELM size (energy drop) vs. waiting time [5]. Additional interpretation is gained from visualizations using several models for the manifold of Gaussian probability distributions. We then concentrate on the application of GLS in a recent update of the key scaling law for the global energy confinement time $\tau_{\mathrm{E,th}}$ in H-mode tokamak plasmas (Figure 1) [6]. GLS is shown to compare favorably to other robust regression methods, including minimum distance estimation techniques based on alternative similarity measures.

With the applications showed in this work, we aim to demonstrate the suitability of the powerful framework of IG for quantifying stochasticity in the complex and challenging environment posed by fusion plasmas.

References

[1] S.-I. Amari and H. Nagaoka, “Methods of Information Geometry”, AMS, New York, NY (2000)

[2] G. Verdoolaege *et al*, Proc. 45th EPS Conference on Plasma Physics, P2.1078 (2018)

[3] A. Shabbir *et al*, Rev. Sci. Instrum. **87**, 11D404 (2016)

[4] G. Verdoolaege, Entropy, **17**, 4602 (2015)

[5] G. Verdoolaege *et al*, AIP Conf. Proc. **1853**, 100002 (2016)

[6] G. Verdoolaege *et al*, Nucl. Fusion **61**, 076006 (2021)

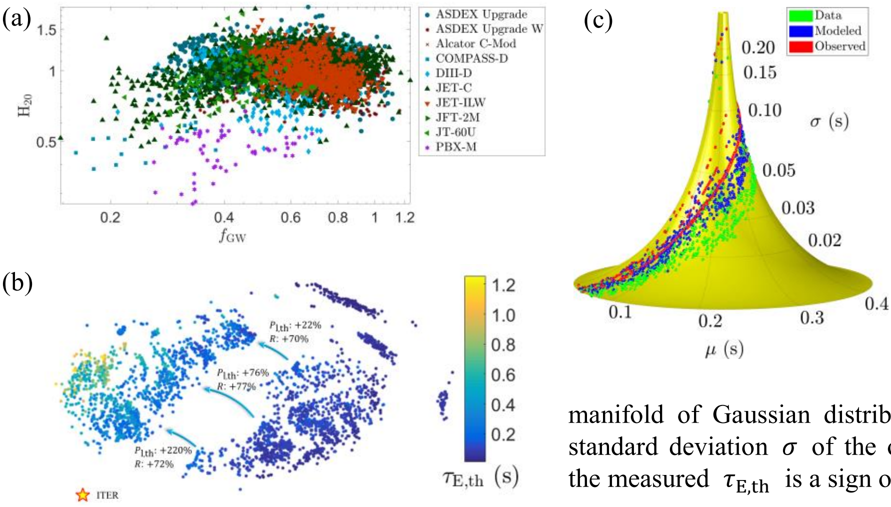

Figure 1. Visualization of regression results obtained with GLS from the ITPA20-IL global H-mode confinement scaling [6], showing (a) confinement enhancement factor H_{20} vs. Greenwald fraction f_{GW} (note the outliers from PBX-M), (b) a projection of the data indicating the gap in power $P_{\mathrm{l,th}}$ and machine size R, and (c) the pseudosphere model of the manifold of Gaussian distributions, where the generally higher standard deviation σ of the observed distributions, compared to the measured $\tau_{\mathrm{E,th}}$ is a sign of additional sources of uncertainty.