Machine learning with overabundant unlabeled samples in data-driven pressure estimation

Junwei Chen[1*], Marco Raiola[1] and Stefano Discetti[1]

1. Dept. of Aerospace Engineering, Universidad Carlos III de Madrid, Leganés, Spain
* Corresponding author: junwei.chen@uc3m.es

Abstract
Data-driven techniques demonstrated to be able to enrich snapshot PIV (Particle Image Velocimetry) with the time resolution from fast probe data, enabling pressure estimation within acceptable accuracy [1]. Acquiring large datasets for proper training is often challenging. Noticing that long sequences of probe data can be captured with significantly less effort than the PIV snapshots, two methods are introduced in this paper, including enforcing Taylor's hypothesis and leveraging a semi-supervised training. The last concept also exploits probe-only measurement without labels, in addition to the probe data directly labeled with PIV snapshots. The semi-supervised training algorithm is compared with the supervised one, showing superior performances in accuracy and reduced storage and computational cost.

Keywords: Pressure Estimation; Machine Learning; Data Driven; POD; Semi-Supervised
Theme/Topic: Pressure and Temperature measurements

1. Introduction
Pressure fields can provide insightful information for locating the source of aeroacoustic noise, improving the design of vehicles and other fields. Traditionally, it is measured pointwise or on solid surfaces using probes or pressure sensitive paint. Pressure fields can be calculated from time-resolved velocity field via Navier-Stokes equation. This requires expensive high-repetition-rate equipment. Recently, we introduced a data-driven method to estimate pressure from time-resolved flow fields reconstructed using snapshot PIV (i.e., with very low temporal resolution) and high-repetition-rate probe data [1]. This principle, and in general flow estimation techniques from sensors, exploit only partially probe data, requiring them to be labelled with a PIV snapshot in order to be used during the training process. Given the overabundance of probe data with respect to PIV data, this means that a significant portion of the probe data acquired are not effectively exploited. The objective of this paper is to exploit overabundant probe data by properly including in the training labelled and unlabelled data. The higher data efficiency of this method would allow to reduce the cost (in time, complexity and money) of the reconstruction experiment without compromising the pressure reconstruction quality.

Two approaches are applied in this work. Firstly, the velocity field can be extrapolated in time nearby the PIV snapshots using Taylor's hypothesis [2], thus extending the database with additional pseudo-labelled samples. This approach is valid only for short-time backtracking or advancement from PIV snapshots, leaving most of the probe data still unlabeled. A second approach based on semi-supervised machine learning is applied. In detail, the flow field can be temporal-spatial decoupled via POD,

$$\boldsymbol{U}(\boldsymbol{x},t)=\sum_i \sigma_i a_i(t)\,\boldsymbol{\phi}_i(\boldsymbol{x}) \tag{1}$$

where the $\boldsymbol{x}$ and t stands for position and time, σ_i is the weight of the ith mode, and a_i is the coefficient for the normalized orthogonal spatial basis $\boldsymbol{\phi}_i(\boldsymbol{x})$. Neural networks regress the relationship $\boldsymbol{f}$ between the probe data $\boldsymbol{p}$ and the coefficients $\boldsymbol{a}$,

$$\widehat{\boldsymbol{a}}(t)=\boldsymbol{f}(\boldsymbol{p}(t)), \tag{2}$$

where $\widehat{\boldsymbol{a}}$ is the coefficient predicted via the model $\boldsymbol{f}$, and the task is to find the optimal model,

$$\underset{\boldsymbol{f}}{\operatorname{argmin}}\,|\boldsymbol{a}-\boldsymbol{f}(\boldsymbol{p})|. \tag{3}$$

Owing to the complexity of the mapping, generally a large number of samples is needed for the training. However, if the model $\boldsymbol{f}$ confirms the assumption of translation invariance in time, which means when the model is trained well at time t_j, $\left|\boldsymbol{a}(t_j)-\boldsymbol{f}\left(\boldsymbol{p}(t_j)\right)\right|\to 0$, it should work well also at time t_{j+k}, with $k=0,\pm 1,\pm 2,\ldots$ This can be fulfilled by introducing another model $\boldsymbol{g}$ on the time derivative of $\boldsymbol{a}$,

$$\hat{\boldsymbol{a}}_t(t) = \boldsymbol{g}(\boldsymbol{p}(t)), \tag{4}$$

and the optimization of $\boldsymbol{f}$ and $\boldsymbol{g}$ does not require to be supervised with the coefficient $\boldsymbol{a}_j$,

$$\underset{\boldsymbol{f},\boldsymbol{g}}{\operatorname{argmin}} \left| \frac{\boldsymbol{f}\left(\boldsymbol{p}(t_{j+1})\right) - \boldsymbol{f}\left(\boldsymbol{p}(t_{j-1})\right)}{t_{j+1} - t_{j-1}} - \boldsymbol{g}\left(\boldsymbol{p}(t_j)\right) \right|_j . \tag{5}$$

This unsupervised part allows enriching the training set by utilizing probe data unlabelled by PIV snapshots. In addition, the supervised machine learning is only trained in several PIV snapshots without knowing the evolution of the flow, while the unsupervised part can be trained in series of probe snapshots, thus covering more conditions of the flow and suppressing unphysical temporal fluctuations.

2. Validation

The method was tested on different data sets including synthetic and experimental ones. In this abstract the experiment on the wake of a wing already tested in [1], is included. The snapshot PIV is down sampled from the time-resolved one, and pointwise velocity are extract at several positions to simulate probe data. Finally, the pressure fields integrated from the velocity estimated by the simple neural networks trained on the original and the extended database, as well as by the semi-supervised machine learning method, are compared to that from time-resolved PIV directly. An example of pressure field prediction is reported below.

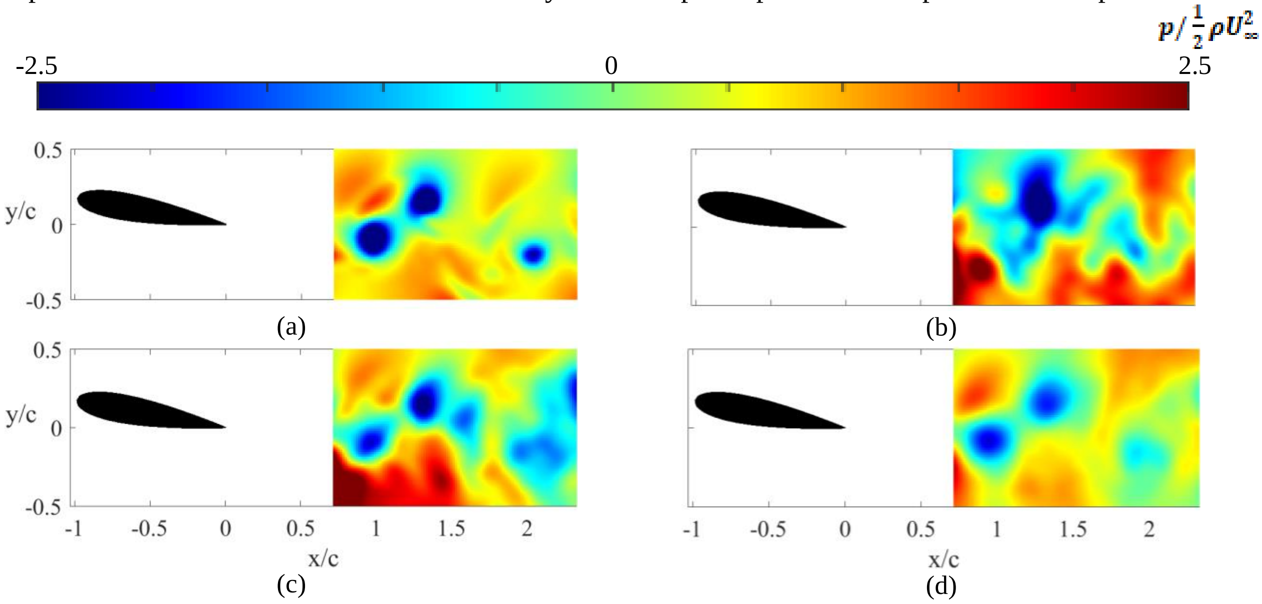

Figure 1 The pressure integrated from (a) time-resolved PIV field, (b) neural networks estimation on original database, (c) neural networks on extended database and (d) semi-supervised machine learning on extended database.

Figure 1 shows the neural networks on extended database with extrapolation using Taylor's hypothesis yields improvements comparing to the one on original database in the prediction of high- and low-pressure areas, and the semi-supervised machine learning on extended database has further improvement. The neural networks on original database reduced the root-mean-square error on the domain shown in the figure and over 981 frames by 50% and 73% comparing to the EPOD in the velocity and pressure prediction, while that on extended database and semi-supervised machine learning on extended database give further reductions by 10% and 13% in velocity, 16% and 36% in pressure prediction comparing to the original neural networks.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 949085).

References

[1] Chen J, Raiola M, Discetti S. Pressure from data-driven estimation of velocity fields using snapshot PIV and fast probes. Experimental Thermal and Fluid Science, 2022, 136: 110647

[2] Taylor GI. The spectrum of turbulence. Proceedings of the Royal Society A: Mathematical, Physical and engineering Sciences, 1938, 164: 476-490