TOWARDS A FAIR CENTRIFUGAL FAN DATABASE FOR OFF-DESIGN VALIDATION OF CFD SIMULATIONS

Matthias PROBST, Balazs PRITZ,
Hans-Jörg BAUER

*Karlsruhe Institute of Technology, Institute of Thermal Turbomachinery,
Kaiserstraße 12, 76131 Karlsruhe, Germany*

SUMMARY

Computational Fluid Dynamics simulations are becoming more and more an integral part in the design and optimization process both in academia and industry. The need for reliable results is great and simulations must be validated with experiments. Especially in off-design conditions of turbomachinery strong whirling and unsteady flow fields are challenging CFD codes. Nowadays, not only the raw validation data and its availability is required but also sustainable data management as an integral part of research data. In the present paper the concept of a validation database for a low-pressure centrifugal fan with a spiral casing for off-design conditions fulfilling the so-called FAIR principles of data management is outlined.

INTRODUCTION

Turbomachineries such as fans are energy conversion machines which can be found in many daily technical devices. The bandwidth of power consumption ranges from a few watts to megawatts and energy-saving requirements are getting increasingly strict with every year. Therefore, a continuous study of the flow phenomena inside such machines is necessary to identify and to understand the sources of energy losses. To analyze the flow phenomena beyond empirical findings, Computational Fluid Dynamic (CFD) simulations are nowadays being used to unlock further potentials for design optimizations and energy savings without the need for expensive experiments.

For turbomachineries the optimal operation point typically can be described well by such simulations while the off-design conditions are very challenging to predict due to unsteady effects for instance. CFD is well-established as an engineering tool both in academia and industry to make design decisions [1]. Due to the numerical nature of the simulation together with various model assumptions validation is required. However, as experiments generally are both time- and costintensive and require much know-how in conducting them [2], the success of such studies is not only based on high quality data but the ability to share them [3]. For this reason, sustainable data management is a key aspect of scientific research and should especially respected in the conception phase of databases.

Very recently, the national data infrastructure association (NFDI) [4] has been founded to support systematic and sustainable data management in Germany. One of the key aspects of achieving this, is the implementation of the so-called FAIR principles – Findability, Accessibility, Interoperability, Reusability, which serve as guidelines during data generation but also for software tools, algorithms and workflows [5]. By following these principles not only the value of data alongside with a simpler data discovery shall be achieved but it could also safe significant investment costs. The European Union estimates, that “not having FAIR research data” could cost about $10.2bn per year [6].

A great variety of databases exist for the many different scientific disciplines. It is beyond the scope of this paper to review all of them, however, to the knowledge of the author, none is publicly available for centrifugal fan validation. The most similar case found in literature was presented in the scope of the ERCOFTAC Workshop [7]. Here, a centrifugal pump with a vaned diffuser is available through the OpenFOAM wiki website [8]. This case as well as other popular basic flow cases provided through ERCOFTAC cases are available via websites that describe the case and allow to download comma separated files (CSV) that contain the data. A general problem with such files is that they do not allow to store additional information along with the raw data in a general way. This results in differently organized file contents (number of columns, variable description length in header, …). Consequently, individual software code to explore and use the data for comparison is required.

Another prominent example of open-access online database is the Johns Hopkins Turbulence Databases [9,10]. Formerly, data was only accessible via the web-browser by querying for specific datasets which then could be downloaded. Meanwhile a web service interface for popular programming languages like Python and Matlab is implemented. Such services especially improve the accessibility, interoperability and reusability of the provided data significantly.

GOALS AND DESIGN PRINCIPLES OF THE VALIDATION DATABASE

The primary goal of the centrifugal fan validation database for off-design conditions is the generation of experimental data and publication of that data without access restrictions. Simultaneously, the FAIR principles are incorporated as an integral part. With the appropriate design principle chosen, a transfer to other projects and thus a universal usage of the presented database concepts is possible. This includes the reusability principle on a meta level for the design concept as a secondary goal.

On the pathway towards such a database multiple common challenges must be solved outlined in the following. A general problem of experiments often is the creation of highly heterogeneous data but also the usage of different software. This leads to dependency issues and individual solutions to write, read and explore datasets. As test rigs generally evolve over time [2], limited or no expansibility is given [11] and long-term accessibility is not insured. This however is strongly needed, otherwise the risk exists that the acquired data and knowledge decays quickly [2,12]. Expansibility in this context does not only mean the generation of additional and new data which is intended to be published. It is rather the manipulation of code and software evolutions, the exchange of hardware and acquisition tools or the rearrangement and adaptation of the test setup that cause heterogeneous data and information respectively.

The open source and general-purpose file format HDF5 is identified to be used in a data storage strategy. It combines raw data (stored as “datasets”) and meta data (stored as “attributes”) and is independent of the programming environment as well as the operating system while experiencing great acceptance of a large community in many disciplines [13,14].

Modern databases follow management principles very similar to those described in the FAIR principles. The aerodynamic open benchmark database described in [11] summarizes requirements for such databases. Most of them are also adopted for the presented database.

The presented database is intended to become a publicly available database on an open-access basis using the hierarchical file format HDF5 as the means of data storage and organization. Unlike other database concepts, the provided data will not be limited to experimental and descriptive data, which commonly can be downloaded via a website as standard text or comma-separated files.

Besides experimental data like the standard measures (volume flow rate, pressure, …) and more specific measures in form of detailed velocity field using Particle Image Velocimetry (PIV), different reference computational meshes for CFD simulations will be provided. In addition to this, validation metrics will be provided to compare CFD results against the available PIV results. These three key building blocks illustrated in Figure 1 are the services offered to the user.

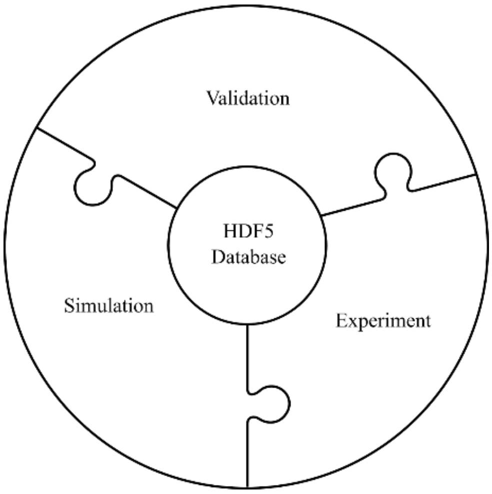

Figure 1: Illustration of the three key building blocks of the centrifugal fan validation database. They can be interpreted as services which are the integral parts of the database and interact with HDF5 files as the core data storage strategy.

The main goals and concepts for the presented fan validation database outlined in this section can be summarized as follows and is partly motivated by [2,11]:

(i) Publicly accessible velocity field data in multiple planes by means of PIV inside the fan rotor and in the spiral casing relative to the volute. In addition, performance curve data should be provided also for operation points where no PIV data exists.

(ii) Publicly accessible CFD meshes as reference using a common file format allowing to be used in every CFD software. The General Notation System (CGNS) [15] as the well-excepted open source storage standard is used to provide the meshes as HDF5 files.

(iii) Rich dataset information through a self-descriptive file format which stores raw data alongside with meta information. This is achievable through the HDF5 file format.

(iv) Assurance of continuous development and adoption of all hardware, software and datasets through object-oriented programming using a widely used language such as python. Codes should be made available and maintained in public software repositories.

(v) Simple data discovery and state-of-the art interoperable open-access is required to achieve acceptance and consequently usage by the community.

EXPERIMENTAL AND NUMERICAL SETUP

The experimental and numerical setup has already been presented in [16], however, a short overview is given here for the sake of completeness.

The investigated fan is designed to be comparable to industrial fans. The nine-blade, backward-curved rapid-prototyping design is rather simple but therefore allows to build a high-quality block-structured mesh for CFD. The supporting disk and the spiral casing are made from acrylic glass to provide optical access. The specific fan properties at design point can be read in [16].

The total pressure increase is measure before and after the fan ("In", "Out" in Figure 2). The pressure drop $p_{\dot{V}}$ in the across the orifice in the straight pipe allows to compute the volume flow rate according to DIN EN ISO 5167.

The velocity fields in the blade channel and the spiral casing is resolved by means of 2D2C-PIV. This allows to measure the two in-plane velocity components at the region of interest. For the PIV setup a 532 nm wavelength Nd:YAG dual cavity laser and a 14 bit pco.pixelfly camera with 1392x1040 pixels is used to record the Di-Ethyl-Hexyl-Sebacic-Acid-Ester (DEHS)-seeded flow. Statistical significance is checked for each recording to determine a reliable phase-averaged velocity field. Laser and camera are both mechanically connected and mounted on a traverse parallel to the fan axis. This allows to measure multiple plane consecutively. Thus, calibration only must be performed once and particles stay in focus even when moving the PIV setup.

The CFD domain is bounded to the marks "In" and "Out" as illustrated in Figure 2. A block-structured mesh is chosen in order to achieve a high-quality discretization. CFD results are phase-averaged to be comparable to the PIV measurements.

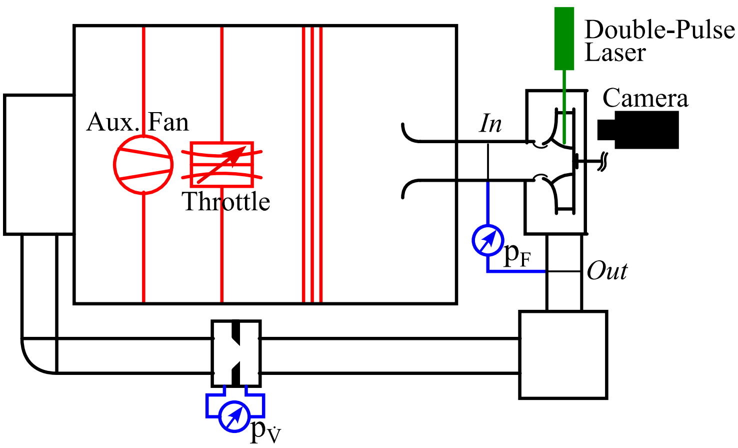

Figure 2: Experimental set-up of the closed-loop centrifugal fan validation test rig.

DATABASE DATAMANAGEMENT AND WORKFLOW

In the development towards the FAIR centrifugal fan database, three conceptual layers were identified to achieve the before mentioned goals and respect the current guiding principles for sustainable research data management. The layout is given in Figure 3 and is summarized in the following:

(i) The first layer incorporates the test bench infrastructure as well as the data storage (first row in Figure 3). It is therefore spatially separated from the user and solely the generated data will be accessible internally or via a web-service. The storage is on a local server only containing HDF5 files and is referred to as the *HDF5 database* in the following.

(ii) The second layer can be interpreted as an interface providing all tools necessary to generate datasets respecting the FAIR principles, access the website and explore as well as process the available data. These tools are available through software repositories for both experimentalists and users.

(iii) The third layer represents the user targeting to validate CFD results computed with own codes for example. The user may be an on-site scientist or from outside the network. In both cases the connection to the database and consecutive processing steps are performed in a Jupyter Notebook [17] environment.

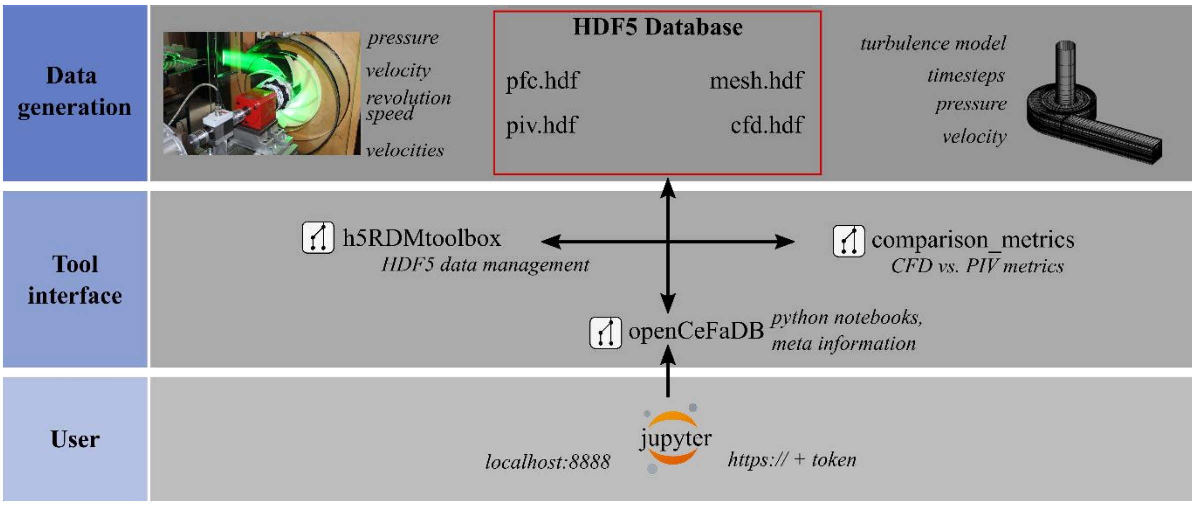

Figure 3: Three-layer concept to manage the database from data creation to exploration by basing everything around HDF5 as well as python and Jupyter Notebook respectively.

As the arrows in Figure 3 suggest the layers are not strictly separated but allow interaction which mainly is driven by the second layer housing the software repositories. In the following the layers and the associated workflow as described in more detail.

On the data generation layer (i) two main sources of data exist: One being the operational point data such as pressures, temperatures and revolution speed and the PIV velocity fields. The integral values origins from many different acquisition devices which are captured using LabView. By storing the data stream in a single HDF file ("performance curve" (PFC) data file *pfc.hdf* in Figure 3) the number of files of each individual acquisition is reduced from many to one.

Information about the devices involved in the measurement as well as technical drawings are registered in the *openCeFaDB* repository [18] together with all required meta data. In a post-processing step, the performance curve HDF5 file is updated in this respect. Additionally, all units and descriptions of the datasets are checked and updated if needed. Only then a full self-descriptive dataset exists which can be published to the HDF5 database. This measurement file can exist exclusively while this is not the case for the PIV recordings.

The velocity field data is directly linked to the current operation mode of the fan. Therefore, the final *piv.hdf* file either contains a group linking data from a PFC-file or simply copies the data into it. The raw images are processed using PIVview, which generates thousands of files per measurement. For each snapshot two images exist and one netCDF4 (similar to HDF5) is written, containing the PIV data. Moreover, as multiple planes are recorded for a single operation point, the overall amount of file grows accordingly. Although the used PIV software provides already the netCDF4 file format per snapshot, which is very similar to the HDF5 file format, it is not yet fully populated with meta data. Therefore, each snapshot is updated and converted to the HDF5 format. To reduce the many thousands of snapshot velocity data files of multiple planes, a single HDF5 case file is generated containing four-dimensional datasets. The dimensions indicate plane (z-location), time and y- and x-location in the field of view. Those dimensions are linked to coordinate datasets using HDF5 dimension scales. Like this, a full and distinct description of the file content is achieved. After adding the final meta data information (technical drawings, device data, ….) the PIV measurement case is published in the HDF5 database.

Simulation-wise the workflow looks slightly different and follows a different motivation. Although the main purpose and surely the larger part of the presented validation database is the making available of experimental data, the simulation side is also supported by providing CFD reference meshes. The common standard for CFD input and output files which contain both the grid and the solution and associated auxiliary information is the General Notation System (CGNS) [15]. Originally CGNS was built using the Advanced Data Format (ADF) but is no switched to using HDF5 to gain parallel I/O and compression capabilities [15]. This agrees with the design choice to only use HDF5 files within the presented database. Any CFD simulation that is given as reference or example in the context of this database will not be the original result file of the used software. Only extracted data according to the field of views in the PIV recordings will be provided. Again, the four-dimensional dataset layout as already described for PIV data will be used.

The data management strategy for the PIV and CFD cases is implemented in python and available through the software repository *h5RDMtoolbox* [19]. This repository provides three python packages: The package *x2hdf* provides classes and methods for CFD and PIV files to convert from the respective output file(s) to a single HDF file while preserving and enriching the final file with meta data. The second package is called *h5wrapperpy* and implements classes that are wrappers around an HDF file. Besides many assisting functions to facilitate input und visualization the core functionality is given by enforcing the user to use meta information for HDF groups and datasets and to use physical units for the HDF datasets. This ensures consistency and the self-descriptiveness and reusability of the files.

The tool is mainly needed in the generation of the database files; however, users may also benefit from the other additional features built around the HDF file. The third package of the *h5RDMtoolbox* is called *h5database* and implements filtering methods to a (posix) HDF file databases. In the current state, it does not use any existing database package like e.g. SQLite. Instead a so-called flat table of content HDF5 file is generated for each data type (PIV, PFC and CFD). In such a file all HDF5 files of the respective data type are stored as so-called external links. Database requests are processed by scanning the table of content file and checking the filter request for each registered file.

Based on a specific CFD simulation, that a user conducted, a typical request would be to find a corresponding PIV recording. The filter request respond is again a HDF5 file but now only containing external links to those files matching the filter. After processing this file, it can be either deleted or used further in the processing workflow, e.g. shared with others. While this seems to be a brute-force approach to find specific HDF files, three aspects should be noted: Firstly, the main properties which the filtering-algorithm is looking for are attributes (e.g. mean value of the dataset "pressure difference"). Thus, no large datasets must be read into RAM to be analyzed. Secondly, one of the big advantages of HDF files is the efficient reading of data. Thirdly, the amount of data generated in the context of this fan database is by far not as large as astronomical data, hence sophisticated query algorithms like [20] are not needed.

The filtering tasks are applied in the third layer, when users actively search for matching validation data to their CFD results. Before doing so, the external user needs to establish a connect to the database. This will be achieved using the highly scalable data service (HSDS [21]) provided by the HDF group. Although this remote service for HDF file exploration is designed and optimized for performance regarding very large databases, the posix-version of it can perfectly be used also for this relatively small database concept.

All interaction with the fan database is optimally performed using Jupyter Notebooks. Jupyter Notebooks combine programming code alongside with front-end web page text. Code output, plots, images and even videos together with explanatory text in a single document make websites like for the ERCOFTAC test cases, for instance, needless. Everything can be explained in the Jupyter Notebook in a tutorial style and the user can directly start the analysis, which is provided through the project's repository *openCeFaDB*. As python is one of the most used programming languages and Jupyter Notebooks are arising more and more to a standard way of exploring and processing data, this is the efficient way of exploring the database and the obvious choice [22].

The final repository in the "tool interface layer" (ii), the python package *comparison_metrics* [23], is the final puzzle piece remaining from Figure 1 in this outline. The user may perform the validation with personal methods and approaches. However, as there is a wide range of comparison methods from qualitative to quantitative ones in different degrees of detail [24,25], the project accompanies with a collection of popular comparison metrics for scalar and vector fields through the above mentioned package.

At the current state the database is at testing mode. All repositories will be publicly available with extensive Jupyter Notebook tutorials once the database goes online. A first release of the database is planned for mid 2022.

It is worth noting that the outlined concept is not restricted to this particular project of the centrifugal fan validation database but rather can be understood as a generic approach to organize fluid investigations. Even if the access outside an institutional network, for instance, is not needed, all other concepts can be adopted by using the python codes available through the repositories.

CONCLUSION

With increasing use of computational fluid dynamics in many engineering disciplines as a design tool and to understand flow phenomena there is also the need for validation data from experiments. For fundamental flow types benchmark data exists. However, most are not following adequate, state-of-the-art data management principles like the FAIR principles supported by the national data infrastructure association for instance. In addition, for the field of centrifugal fans no publicly available database fulfilling the standards exists to the knowledge of the authors. Hence, an open-access public centrifugal fan test bench with a generic rotor is built to provide such needed experimental data to test CFD codes and further study flow phenomena inside the fan. In the genesis of the database the FAIR principles are considered which will allow a comprehensible (self-explaining), long-term data availability that is expandable and comes with open-source, object-oriented software to assist in the exploring process but is no prerequisite. This is mainly achieved by using python as the programming language and by storing all data in the widely used open-source general purpose file format HDF5.

The centrifugal fan database is under current development and therefore only locally accessible at the moment, but a first release is expected soon. HDF5 files are accessible remotely and readable by various programming languages. However, the provided codes suggest using the popular Jupyter Notebooks as in many other successful projects.

The presented database concept is not restricted to the fan validation test bench but can be applied to any research data acquisition or generation of any type. The provided software repositories especially open-up the possibility to use the concept in other CFD and/or PIV projects. It even is possible to apply the principle on already generated data to enhance accessibility, long-term usage and interoperability for instance.

BIBLIOGRAPHY

[1] R. V. Wilson, F. Stern, H. W. Coleman et al., *Comprehensive approach to verification and validation of CFD simulations-Part 2: Application for RANS simulation of a cargo/container ship.*, J. Fluids Eng., vol. 123, no. 4, pp. 803–810, **2001**.

[2] N. Preuß and P. Pelz, *Integrated management of experimental research- and meta-data for fan test rigs. in* International Conference on Fan Noise, Aerodynamics, Applications and Systems, **2018**.

[3] J. Georgieva, V. Gancheva, and M. Goranova, *Scientific data formats. in* International Conference on Applied Informatics and Communications, **2009**.

[4] S. Kraft, A. Schmalen, H. Seitz-Moskaliuk et al., *Nationale Forschungsdateninfrastruktur (NFDI) e. V. Aufbau und Ziele.*, **2021**.

[5] M. D. Wilkinson, M. Dumontier, I. J. J. Aalbersberg et al., *The FAIR Guiding Principles for scientific data management and stewardship.*, Scientific data, vol. 3, pp. 1–9, **2016**.

[6] PwC EU Services, *Cost of not having FAIR research data: Cost-Benefit analysis for FAIR research data.*, **2019,** doi: 10.2777/02999.

[7] J. F. Combès, *Test Case U3: Centrifugal Pump with a Vaned Diffuser. ERCOFTAC Seminar and Workshop on Turbomachinery Flow Prediction VII, Aussois, jan 4-7.*, **1999**.

[8] *ERCOFTAC Centrifugal Pump with a vaned diffuser.*, **12/6/2021,** http://www.openfoamwiki.net/index.php/Sig_Turbomachinery_/_ERCOFTAC_centrifugal_pump_with_a_vaned_diffuser.

[9] Y. Li, E. Perlman, M. Wan, Y. Yang, C. Meneveau, R. Burns, S. Chen, A. Szalay & G. Eyink, *A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence.*, **2008**.

[10] Y. Li, E. Perlman, M. Wan et al., *A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence.*, Journal of Turbulence, vol. 9, N31, **2008**.

[11] W. BAI, L. LI, Z. LI et al., *CFD V & V and Open Benchmark Database.*, Chinese Journal of Aeronautics, vol. 19, no. 2, pp. 160–167, **2006**.

[12] W. K. Michener, *Meta-information concepts for ecological data management.*, Ecological informatics, vol. 1, no. 1, pp. 3–7, **2006**.

[13] N. Preuss, G. Staudter, M. Weber et al., *Methods and Technologies for Research- and Metadata Management in Collaborative Experimental Research.*, Applied Mechanics and Materials, vol. 885, pp. 170–183, **2018**.

[14] *The HDF Group.*, **12/6/2021,** https://support.hdfgroup.org/HDF5/users5.html.

[15] C. Rumsey, B. Wedan, T. Hauser et al., *Recent updates to the CFD general notation system (CGNS). in* 50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition, p. 1264, **2012**.

[16] B. Pritz, J. Walter, and M. Gabi, *A centrifugal fan test bench for validation data at off-design conditions. in* International Conference on Fan Noise, Aerodynamics, Applications and Systems, **2018**.

[17] T. Kluyver, B. Ragan-Kelley, F. Pérez et al., *Jupyter Notebooks - a publishing format for reproducible computational workflows. in* Positioning and Power in Academic Publishing:

Players, Agents and Agendas, Fernando Loizides and Birgit Scmidt, Eds., pp. 87–90, IOS Press, **2016**.

[18] M. Probst, *Python package to interface with the open centrifugal fan database.,* **5/3/2022,** https://github.com/MatthiasProbst/openCeFaDB.

[19] M. Probst, *HDF5 Research Data Management Toolbox.,* **5/3/2022,** https://github.com/MatthiasProbst/h5RDMtoolbox.

[20] L. Gosink, J. Shalf, K. Stockinger et al., *HDF5-FastQuery: Accelerating Complex Queries on HDF Datasets using Fast Bitmap Indices. in* 18th International Conference on Scientific and Statistical Database Management (SSDBM'06), pp. 149–158, IEEE, **03-05 July 2006**.

[21] HDF5 Users., *The HDF Group.,* **12/6/2021,** https://www.hdfgroup.org/.

[22] J. M. Perkel, *By jupyter, it all makes sense.,* Nature, vol. 563, no. 7729, pp. 145–146, **2018**.

[23] M. Probst, *Python package providing comparison metrics.,* **5/3/2022,** https://github.com/matthiasprobst/comparison_metrics.

[24] W. L. Oberkampf and T. G. Trucano, *Verification and validation in computational fluid dynamics.*, Progress in aerospace sciences, vol. 38, no. 3, pp. 209–272, **2002**.

[25] C. Willman, B. Scott, R. Stone et al., *Quantitative metrics for comparison of in-cylinder velocity fields using particle image velocimetry.,* Experiments in Fluids, vol. 61, no. 2, **2020**.