Davide Valsecchi

Italian - Born 09/09/1994 in Bergamo (IT) | Geneva, Switzerland (Permit B)
dvalsecchi@ethz.ch | +41-767250344 | LinkedIn | GitHub

RESEARCH PROFILE

Postdoctoral Researcher at ETH Zurich based at **CERN** for the **CMS Experiment**, with a dual focus on **precision Higgs boson measurements** and **advanced Machine Learning** applications. Currently serving as the coordinator of the **CMS E/Gamma Physics Object Group**, leading the calibration and reconstruction strategy for electrons and photons. Expert in **Generative Deep Learning**, pioneering applications to **unbinned likelihood fits**, **data-simulation correction**, and **phase-space integration** (Matrix Element Method). Strong track record in **detector reconstruction algorithms** and novel data analysis techniques.

PROFESSIONAL EXPERIENCE

Postdoctoral Researcher

2022 – Present

ETH Zurich, Prof. Wallny Group (CMS Experiment)

- **Higgs Physics Analysis:** Leading the analysis of the Higgs boson in the $t\bar{t}H$ production mode, focusing on the $H \to b\bar{b}$ decay channel. Developed a novel **background estimation** strategy based on data-driven corrections using machine learning techniques, improving the modeling of the dominant $t\bar{t}$+jets background. Currently developing the full LHC Run 3 analysis, with the goal of observing for the first time the $t\bar{t}H$ process in this challenging final state ($H \to b\bar{b}$).
- **Object Reconstruction R&D**: Designed new machine learning (ML) algorithms to reconstruct electron/photon deposits in the CMS electromagnetic calorimeter (ECAL) and improve the performance of the CMS E/Gamma physics objects. Developed novel strategies for the calibration of the heavy-flavour jet identification algorithms in Lorentz-boosted topologies, crucial for many physics analyses at the LHC.
- **Novel analysis methods**: Developing novel strategies to apply the Matrix Element Method in the complex CMS experiment data using **Transformers coupled to generative models** for multi-dimensional phase-space integration.
- **Statistical Methods & Calibration**: Pioneering **unbinned likelihood fit** based on generative ML models (Normalizing Flows) for high-dimensional parameter estimation and **data-driven corrections**. Building CMS-wide applications in the context of detector calibration.
- **Mentorship & Teaching:** Supervised 3 PhD and 10+ BSc/MSc students at ETH Zurich. Teaching assistant for the Master's course "Statistical Methods and Analysis Techniques in Experimental Physics". Invited seminar speaker at multiple institutions, including University of Warsaw, University of Bristol, Purdue University, and Milano-Bicocca. Lecturer for PhD courses on machine learning applications in high-energy physics.
- **Common tools development:** Lead developer of `PocketCoffea`, a columnar analysis framework based on Scikit-HEP, adopted by multiple analysis groups. Contributing to `Zuko`, for Normalizing Flows in PyTorch.

PhD Student

Oct. 2018 – Jan. 2022

Università degli Studi di Milano-Bicocca & CERN Doctoral Student Program

- **Standard Model Measurement:** Led the analysis for the Vector Boson Scattering (VBS) search in the semileptonic channel ($WV \to \ell\nu qq$). Developed a novel signal extraction strategy based on ML classifiers, leading to the first evidence of this process at the LHC.
- **ECAL Reconstruction:** Designed and implemented the first Graph Neural Network (GNN) based superclustering algorithm for the CMS Electromagnetic Calorimeter (ECAL). Replaced legacy heuristic algorithms, improving energy resolution in high-pileup conditions.
- **ECAL Trigger Optimization:** Optimized low level hardware mechanisms for the ECAL trigger primitives generation, studying new approaches in legacy hardware to reduce the contamination from spurious electronic noise.

CERN Technical Student *2016 – 2017*

CERN (CMS Experiment)

- **ECAL DAQ Optimization:** Optimized the data acquisition system for the CMS Electromagnetic Calorimeter (ECAL) during the LHC Run 2. Developed new monitoring tools and implemented improvements in the DAQ software, leading to increased efficiency and reduced downtime of the ECAL during data-taking.

EDUCATION

Master's in Particle Physics (110 cum laude) *2016 – 2018*
Università degli Studi di Milano-Bicocca

Bachelor's in Physics (110 cum laude) *2013 – 2016*
Università degli Studi di Milano-Bicocca

COORDINATION ROLES

E/Gamma Physics Object Group Coordinator *2025 – Now*

CMS Collaboration (Physics Coordination)

- Co-lead of the CMS E/Gamma Physics Object Group (POG) coordinating the calibration and reconstruction activities for electron/photon objects in the CMS experiment.
- Responsible for the overall strategy and performance of E/Gamma reconstruction, including the precise energy measurement of photons and electrons and the development of new ML-based algorithms for object identification, reconstruction, and calibration.
- Liaison between POG and CMS Physics Analysis groups to ensure optimal object performance for all CMS physics measurements.

ML Software Coordinator *2022 – 2025*

CMS Collaboration (Offline & Computing)

- **Coordination (L3):** Led the Machine Learning sub-group within the CMS Offline Software area. Responsible for the integration and validation of ML models in the official reconstruction software (`CMSSW`).
- **Software Integration:** Managed the integration of ONNXRuntime and PyTorch C++ inference engines into the multi-million line `CMSSW` codebase, enabling the deployment of Deep Learning models in the High Level Trigger (HLT) and Offline reconstruction.
- **Performance Optimization:** Optimized ML inference on heterogeneous hardware (CPU/GPU) using GPU compatibility libraries.
- **Algorithm Deployment:** Bridge between Physics Object Groups (POGs) and Offline Software, ensuring new architectures (GNNs, DeepSets) were successfully deployed in production campaigns.

SELECTED PUBLICATIONS & REPORTS

Full list available on Inspire HEP.

- **From bins to flows: unbinned and multivariate scale factors**. D. Valsecchi on behalf of the CMS Collaboration, CMS Performance Note (presented at ACAT 2025).
  - Developed a novel unbinned and multivariate approach for data-simulation corrections using Normalizing Flows. Demonstrated the method on the calibration of electron/photon identification scale factors, achieving improved precision compared to traditional binned approaches.
- **Computing the Matrix Element Method with Generative Machine Learning**. D. Valsecchi on behalf of the CMS Collaboration, CMS Performance Note (presented at EuCAIF 2025).
  - Developed a generative framework for Matrix Element Method integration. Leveraged Normalizing Flows coupled with Transformers to efficiently sample the **high-dimensional phase space** of multi-body final states, achieving significant speedups of this theoretically well-motivated approach.
- **Performance of heavy-flavour jet identification in Lorentz-boosted topologies in proton-proton collisions at $\sqrt{s} = 13$ TeV**. CMS Collaboration, JINST 20 (2025) P11006.

- Developed novel strategies for the calibration of the heavy-flavour jet identification algorithms in Lorentz-boosted topologies, crucial for many physics analyses at the LHC. Demonstrated the performance of these algorithms on simulated events, and introduced novel calibration techniques using proton-proton collision data collected at $\sqrt{s} = 13$ TeV during the 2016-2018 LHC data-taking period.

- **One Flow to Correct Them all: Improving Simulations in High-Energy Physics with a Single Normalising Flow and a Switch**. D. Valsecchi et al. *Comput Softw Big Sci 8, 15* (2024).
  - Architected a conditional Normalizing Flow framework to correct **Data-Simulation differences** in high-dimensional variables. Achieved precise closure of detector-level distributions by learning the morphing function between simulation and data unbinned distributions.
- **Deep Learning Techniques for Energy Clustering in the CMS Electromagnetic Calorimeter**. D. Valsecchi on behalf of the CMS Collaboration, *2022 IEEE NSS-MIC-RTSD* (IEEE Conference Proceeding).
  - Detailed performance study of the GNN-based superclustering algorithm for the CMS ECAL. Demonstrated significant improvements in energy resolution and pileup robustness compared to legacy algorithms.
- **Deep learning techniques for energy clustering in the CMS ECAL**. D. Valsecchi et al. *J. Phys.: Conf. Ser. 2438 012077* (2021).
  - Designed the first GNN architecture for ECAL energy clustering. Formulated the task as a graph-based linking problem to resolve overlapping showers in high-pileup conditions, outperforming the legacy geometrical algorithm.
- **Evidence for WW/WZ vector boson scattering in the decay channel $l\nu qq$ produced in association with two jets**. CMS Collaboration. *Phys. Lett. B 834 137438* (2022).
  - *PhD Thesis Result.* First evidence of the VBS process in the semileptonic channel at the LHC. Developed the DNN-based signal extraction strategy to isolate the rare VBS signal from large V+jets backgrounds. Combined deep classification networks with data-driven background estimation techniques to achieve a signal significance of 4.4 standard deviations, providing crucial insights into the electroweak symmetry breaking mechanism.

CONFERENCE CONTRIBUTIONS

- **From bins to flows: a neural network approach to unbinned data-simulation corrections**. *ACAT 2025*
- **Efficient data movement for Machine Learning inference in heterogeneous CMS software**. *FastML 2025*
- **Conference Track Convener**, *CHEP 2024*. Member of the scientific committee for the "Offline Computing" track, responsible for the review and selection of contributed papers and talks.
- **ML in Higgs Physics**. Plenary talk at *Higgs 2024*
- **Higgs / Top / EWK and combined EFT results at CMS** Parallel talk at *LHCP 2023*

TEACHING EXPERIENCE

- **Student Supervision:** Supervised 3 PhD students and 10+ BSc/MSc students at ETH Zurich, providing guidance on research projects, data analysis techniques, and scientific communication.
- **Teaching Assistant: "Statistical Methods and Analysis Techniques in Experimental Physics"**, *ETH Zurich (2025-2026).* Master's course covering advanced statistical methods in High Energy Physics.
- **Invited Seminar: "Machine Learning in CMS: New Approaches for High-Energy Physics Challenges"**, *University of Warsaw (November 2025).* Seminar for the High-Energy Physics group, covering the latest developments in machine learning applications within the CMS experiment, including novel approaches for data-simulation corrections and generative modeling for physics analyses.
- **PhD Course Lecturer: "Normalizing Flows in High-Energy Physics Applications"** *Purdue University (November 2025).* Invited lecture for the PhD course "Machine Learning for High-Energy Physics", covering the theory and applications of Normalizing Flows in particle physics.
- **Invited Seminar: "Machine Learning in CMS: New Approaches for High-Energy Physics Challenges"**, *University of Bristol (November 2024).* Seminar for the High-Energy Physics group, covering the latest developments in machine learning applications within the CMS experiment.
- **PhD Course Lecturer: "Machine Learning tools for precision physics"**, *Milano-Bicocca (October 2024).* PhD course about ML applications in high-energy physics, with a focus on the Transformer architecture and probability modeling.
- **Teaching Assistant: "Physics Lab 1"**, *ETH Zurich (2022-2025).* Supervision of laboratory experiments for first-year physics students.

TECHNICAL SKILLS

- **Statistics:** Hypothesis Testing (CLs), Maximum Likelihood Estimation, Unfolding, Systematic Uncertainty profiling.
- **Machine Learning:** Geometric Deep Learning (GNNs), Transformers, Generative Models (Normalizing Flows, Flow-matching), Differentiable Programming.
- **Languages:** Python, C++, CUDA, ROOT (RooFit/RooStats).
- **Libraries:** PyTorch, TensorFlow, Scikit-HEP, Coffea, Pandas, Dask, SciPy, ONNXRuntime.
- **Computing:** Git, Docker/Apptainer, Linux, Grid Computing (HTCondor), CI/CD (GitLab, GitHub Actions).