LATE REQUEST FOR A SPECIAL PROJECT 2026–2028

MEMBER STATE: Switzerland

Principal Investigator[1]: Dr. Muhammed Irfan

Affiliation: École Polytechnique Fédérale de Lausanne (EPFL)

Address: Laboratory of Atmospheric Processes and their Impacts (L)
EPFL, ENAC, IIE, LAPI, GR C2 534 (Bâtiment GR), Station 2
1015, Lausanne, Switzerland

Other researchers: Prof. Athanasios Nenes

Project Title: Improving Cloud Droplet Activation in OpenIFS–HAMM7 through Physically-Based and Deep-Learning-Derived Vertical-Velocity Variability

To make changes to an existing project please submit an amended version of the original form.)

If this is a continuation of an existing project, please state the computer project account assigned previously.	**SP** ………….	
Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.)	2026	
Would you accept support for 1 year only, if necessary?	YES ☐ X	NO ☐

Computer resources required for project year:		**2026**	**2027**	**2028**
High Performance Computing Facility	[SBU]	25 000 000		
Accumulated data storage (total archive volume)[2]	[GB]	100000		

EWC resources required for project year:		**2026**	**2027**	**2028**
Number of vCPUs	[#]			
Total memory	[GB]			
Storage	[GB]			
Number of vGPUs[3]	[#]			

[1] The Principal Investigator will act as contact person for this Special Project and, in particular, will be asked to register the project, provide annual progress reports of the project's activities, etc.
[2] These figures refer to data archived in ECFS and MARS. If e.g. you archive x GB in year one and y GB in year two and don't delete anything you need to request x + y GB for the second project year etc.
[3]The number of vGPU is referred to the equivalent number of virtualized vGPUs with 8GB memory.

Continue overleaf.

Principal Investigator: Dr. Muhammed Irfan

Project Title: Improving Cloud Droplet Activation in OpenIFS–HAMM7 through Physically-Based and Deep-Learning-Derived Vertical-Velocity Variability

Extended abstract

Cloud processes play a central role in regulating the Earth's energy balance and hydrological cycle, yet they remain one of the largest sources of uncertainty in climate projections. A dominant contributor to this uncertainty is the representation of aerosol–cloud interactions, particularly the activation of cloud droplets from atmospheric aerosol populations. The number concentration of cloud droplets (CDNC) directly controls cloud albedo, lifetime, and precipitation efficiency, and thus strongly influences effective radiative forcing and climate sensitivity (Boucher et al., 2013; IPCC, 2021).

Within the framework of the Horizon Europe CleanCloud project (Grant Agreement 101137639), this Special Project aims to improve the physical realism of cloud droplet activation in OpenIFS–HAMM7 (cycle 48r1), the atmospheric component of EC-Earth4. The focus is on addressing a long-standing simplification in global climate models: the treatment of subgrid-scale vertical velocity variability (σw).

Cloud droplet activation is fundamentally controlled by the competition between aerosol properties and supersaturation, which is itself governed by vertical velocity fluctuations in cloud-forming regions (Abdul-Razzak & Ghan, 2000; Nenes et al., 2001). In most Earth-system models, including EC-Earth3, σw is prescribed as a globally constant value (typically 0.6–0.8 m s^{-1}), independent of turbulence intensity, stratification, or convective regime. This assumption neglects well-established physical controls on vertical velocity variability and introduces systematic regional and seasonal biases in CDNC and cloud radiative effects (Duran et al., 2018; Bougiatioti et al., 2020).

Recent advances offer two complementary pathways to overcome this limitation. First, physically based formulations diagnose σw from prognostic turbulent kinetic energy (TKE), allowing σw to respond dynamically to atmospheric stability and turbulence (Morales & Nenes, 2014; Duran et al., 2018). Second, data-driven approaches using deep neural networks have demonstrated skill in predicting σw variability across diverse atmospheric regimes when trained on high-resolution simulations and observations (Barahona et al., 2024). However, the robustness and generalization of such machine-learning (ML) parameterizations in extreme or out-of-distribution regimes remains largely unexplored.

This project will implement and evaluate both physically derived and deep-learning-derived σw formulations within OpenIFS–HAMM7 and assess their impacts on CDNC, cloud microphysical properties, and aerosol indirect effects at the global scale. In addition, it will explicitly test the generalization and failure modes of a pretrained ML σw parameterization across physically extreme environments, from the ultra-stable Arctic boundary layer to strongly forced Mediterranean orographic regimes.

Scientific Objectives

The main scientific objectives are:

1. **Evaluate droplet activation physics**: Evaluate the recently implemented Morales & Nenes (2014) activation scheme alongwith the existing Abdul-Razzak & Ghan (2002) formulation in OpenIFS–HAMM7, enabling systematic evaluation of activation sensitivity to σw.

2. **Introduce physically based σw diagnostics**: Diagnose σw dynamically from model-predicted TKE following Duran et al. (2018.

3. **Integrate a deep-learning σw parameterization (Wnet)**: Couple a pretrained neural-network predictor of σw (Barahona et al., 2024) into OpenIFS–HAMM7 in inference-only mode, ensuring computational efficiency and numerical stability.

4. **Quantify impacts on CDNC and cloud properties**: Compare prescribed, TKE-derived, and ML-derived σw formulations and assess their effects on CDNC, cloud optical depth, and aerosol–cloud interactions at global and regional scales.

5. **Assess ML generalization and robustness**: Evaluate the ability of a globally trained ML turbulence parameterization to generalize to extreme atmospheric regimes and develop physically consistent fallback and hybrid strategies to ensure safe deployment in Earth-system models.

Simulation Design

We will conduct **six global 2-year simulations** with **OpenIFS–HAMM7 (cycle 48r1)** to systematically evaluate the impact of σw formulation and droplet activation scheme on cloud droplet number concentration (CDNC).

Each simulation includes:

- **1 year spin-up** (model equilibration of aerosol and cloud fields)
- **1 year analysis period** for diagnostics and intercomparison.

Simulation	Activation	σw formulation	Purpose
ARGs08	Abdul-Razzak & Ghan (2002)	Constant σw = 0.8 m s^{-1}	Baseline
ARGtke	Abdul-Razzak & Ghan (2002)	TKE-derived σw	Physics-based
ARGwnet	Abdul-Razzak & Ghan (2002)	Wnet σw	ML-based
MNs08	Morales & Nenes (2014)	Constant σw = 0.8 m s^{-1}	Alternative physics
MNtke	Morales & Nenes (2014)	TKE-derived σw	Fully physical
MNwnet	Morales & Nenes (2014)	Wnet σw	AI-enhanced

Total simulations: 6 × 2 years = **12 model-years**

ML Generalization Across Extreme Regimes

The robustness of the ML-based σw parameterization will be evaluated across contrasting environments:

- **Villum Research Station (Greenland):**
  Ultra-stable Arctic boundary layers with weak turbulence and strong inversions.
- **Mt. Helmos / CHOPIN (Mediterranean):**
  Strongly forced orographic flows with intense shear- and convection-driven turbulence.

Where training-distribution information is available, out-of-distribution (OOD) conditions will be diagnosed using standardized distance and Mahalanobis-type metrics. Failure modes such as σw underestimation in strong shear or overestimation in stable stratification will be linked to CDNC and cloud-field biases.

Low-cost stabilization strategies will be tested, including OOD-aware fallback to TKE-derived σw, variance limiting, and hybrid ML–physics blending. These approaches will demonstrate that ML parameterizations can be safely deployed in global climate models without retraining.

Computational Resource Justification

High-Performance Computing

Requested: 25,000,000 SBU (2026)

OpenIFS–HAMM7 global simulations with interactive aerosols, prognostic TKE, and online ML inference are computationally demanding. Based on prior experience, one model year requires approximately 2 million SBU at the selected resolution. The total requirement of 12 model-years corresponds to ~24 million SBU, with the remaining allocation covering sensitivity tests, restarts, and diagnostics.

Data Storage

The project requires an estimated **80–100 TB** of accumulated archive storage over its duration. Each OpenIFS–HAMM7 global simulation produces approximately **8–12 TB** of output due to the need to archive three-dimensional aerosol, cloud microphysical, and turbulence-related diagnostics in addition to standard meteorological fields.

Key archived variables include cloud droplet number concentration (CDNC), subgrid-scale vertical velocity variability (σw), aerosol size distributions, and cloud microphysical properties, stored at daily resolution with higher-frequency output for selected diagnostics and regions. Additional storage is required for diagnostics associated with the evaluation of physically based and machine-learning-derived σw formulations.

With six two-year simulations conducted over the project lifetime, the requested storage is commensurate with the scientific objectives and necessary to support regime-dependent analysis, reproducibility, and model development.

References

Abdul-Razzak, H., & Ghan, S. J. (2000). A parameterization of aerosol activation: 1. Single aerosol type. *Journal of Geophysical Research*, 105(D5), 6837–6844.

Abdul-Razzak, H., & Ghan, S. J. (2002). A parameterization of aerosol activation: 2. Multiple aerosol types. *Journal of Geophysical Research*, 107(D3), 4026.

Barahona, D., et al. (2024). A deep-learning parameterization of vertical velocity variability for aerosol–cloud interactions. *Journal of Advances in Modeling Earth Systems*.

Bougiatioti, A., et al. (2020). Aerosol–cloud interactions and their representation in climate models. *Atmospheric Chemistry and Physics*, 20, 915–933.

Boucher, O., et al. (2013). Clouds and aerosols. In *IPCC AR5 WG1*.

Duran, E., et al. (2018). Turbulence controls on cloud droplet number concentration. *Journal of the Atmospheric Sciences*, 75, 3059–3080.

IPCC (2021). *Sixth Assessment Report, Working Group I*. Cambridge University Press.

Morales, R., & Nenes, A. (2014). Characteristic updrafts for computing aerosol activation. *Atmospheric Chemistry and Physics*, 14, 11485–11502.

Nenes, A., et al. (2001). Cloud droplet activation theory revisited. *Journal of Geophysical Research*, 106(D18), 22947–22964