<u>LATE</u> REQUEST FOR A EMI R&D PROJECT 2026–2028

Country/Organisation: …Sweden…..…

Principal Investigator[1]**:** …Yi-Chi Wang……………….…………………………………..……

Affiliation: …Swedish Meteorological and Hydrological Institute…………..……

Address: …Folkborgsvägen 17, 603 82 Norrköping, Sweden …………………..……

…………………………………………………………..……………………………………………………..……

Other researchers: Fuxing Wang (SMHI, fuxing.wang@smhi.se)

Toni Fernandez Andrade (AEMET, afernandeza@aemet.es)

Project Title: HCLIM-AI Benchmark and Tutorial: Evaluating ML Emulators and AI Forecasting Workflows for Regional Climate Modelling with HCLIM

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 ☒	NO ☐

Late EMI R&D Projects are normally approved for 1 year, but exceptionally can be approved for 3 years.

Computer resources required for project year:		**2026**	**2027**	**2028**
High Performance Computing Facility	[SBU]	0		
Graphics Processing Unit Cluster-A	[GBU]	1,050,000		
Graphics Processing Unit Cluster-B	[GBU]	990,000		
Accumulated data storage (total archive volume)[2]	[GB]	4000		

[1] The Principal Investigator will act as contact person for this EMI R&D 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.

EWC resources required for project year:		2026	2027	2028
Number of vCPUs	[#]	0		
Total memory	[GB]	0		
Storage	[GB]	0		
Number of vGPUs[3]	[#]	0		

Continue overleaf.

Principal Investigator: Yi-Chi Wang

Project Title: HCLIM-AI Benchmark and Tutorial: Evaluating ML Emulators and AI Forecasting Workflows for Regional Climate Modelling with HCLIM

Extended abstract

All EMI R&D Project requests should provide an abstract/project description including a scientific plan, a justification of the computer resources requested and the technical characteristics of the code to be used. The completed form should be submitted/uploaded at https://www.ecmwf.int/en/research/special-projects/special-project-application/special-project-request-submission.

Following submission by the relevant Member State the EMI R&D Project requests will be published on the ECMWF website and evaluated by ECMWF and its Scientific Advisory Committee. The requests are evaluated based on their scientific and technical quality, and the justification of the resources requested. Previous EMI R&D Project reports and the use of ECMWF software and data infrastructure will also be considered in the evaluation process.

Requests exceeding 10,000,000 SBU should be more detailed (3-5 pages).

Scientific Motivation

The HARMONIE-Climate (HCLIM; Belušić et al., 2020) model is developed and maintained by a multi-institutional consortium and is widely used for regional climate research and climate-service applications across Europe and beyond. A major focus of HCLIM development has been the production of kilometre-scale, convection-permitting climate simulations. These simulations provide important added value for precipitation extremes, heatwaves, wind, and other local-scale processes, but their high computational cost limits their use in large ensembles, uncertainty quantification, and operational climate services.

Recent advances in machine learning (ML) offer opportunities to emulate selected aspects of convection-permitting simulations and accelerate the generation of high-resolution climate information. However, robust progress requires standardized datasets, transparent evaluation protocols, reproducible baselines, and users who understand both the capabilities and failure modes of ML models. The HCLIM community currently lacks a shared framework that joins these scientific benchmarking and capacity-building needs.

The HCLIM consortium has produced a unique collection of convection-permitting simulations, including the Nordic Convection-Permitting Climate Projection (NorCP; Lind et al., 2020, 2023) and other kilometre-scale experiments covering different regions and applications. These physically consistent datasets provide an exceptional testbed for evaluating AI-based emulation and downscaling methods.

This pilot project will establish an HCLIM-AI community benchmarking framework as the scientific foundation for a practical tutorial on ML emulators and regional AI forecasting workflows, delivered alongside the HCLIM Working Week in October 2026. Compact convolutional neural network (CNN; Rampal et al. 2022) and generative adversarial network (GAN; Wang et al. 2026) baselines will be trained and evaluated against selected HCLIM simulations. The tutorial will connect basic ML concepts and CNN emulators to generative emulation, before progressing to notebook-based inference with ECMWF's Artificial Intelligence Forecasting System (AIFS), with emphasis on AIFS-CRPS v1 and v2 (Lang et al., 2026; Lang et al., 2025). The advanced module will also examine whether AIFS output can provide initial and boundary conditions for HCLIM, test selected regional ML weather models developed within AEMET-NWP, and compare ML and physics-based forecasts. AIFS itself will be used for inference only and will not be treated as an HCLIM emulator benchmark.

The resulting datasets, metrics, baseline results, and tutorial materials will provide a common reference for future HCLIM-AI developments, support ECMWF and European Meteorological Infrastructure activities on machine learning and climate modelling.

Scientific Plan

WP1: Benchmark Dataset Development

Benchmark datasets will be derived from existing HCLIM kilometre-scale simulations, initially focusing on the selected HCLIM convection-permitting experiments. Predictor variables will consist of large-scale atmospheric and surface fields from the driving simulations. Target variables will include precipitation, temperature, wind, and selected climate-extreme indicators. The pilot datasets will use compact spatial and temporal subsets that are large enough to demonstrate scientific workflows but small enough for reproducible tutorial use.

WP2: Baseline CNN and GAN Benchmarking

Reproducible baseline models will be developed rather than extensively optimized. A compact CNN emulator will establish a deterministic reference. A conditional GAN will then demonstrate generative emulation of fine-scale spatial variability and extremes. Evaluation will use the existing climate-metrics package and will address mean climate statistics, extremes, spatial variability, temporal consistency, deterministic skill, and distributional behaviour. Training, validation, and independent testing will be separated explicitly.

WP3: HCLIM ML-Emulator Tutorial

A tutorial activity will be held alongside the HCLIM Working Week in October 2026 for approximately 15 participants from the HCLIM community. For this activity, the benchmark workflow will be organized into three progressive hands-on modules. Module 1 will introduce supervised learning, normalization, convolutions, loss functions, training diagnostics, validation, and a compact CNN emulator. Module 2 will introduce generator-discriminator training, adversarial and content losses, stochastic sampling, mode collapse, and appropriate spatial and distributional evaluation of GAN output. Module 3 will demonstrate AIFS and regional ML forecasting workflows. Participants will run notebook-based inference experiments with pre-trained AIFS models, emphasizing AIFS-CRPS v1 and v2, and compare their outputs, execution times, and resource use. A feasibility exercise will assess the variables, temporal frequency, domain treatment, and data conversion required to use AIFS forecasts as initial and lateral boundary conditions for HCLIM. Subject to software and model availability, selected regional ML weather models developed within AEMET-NWP will also be demonstrated. Their forecast quality, computational speed, and resource consumption will be compared with suitable physics-based reference forecasts.

WP4: Reproducibility, Delivery, and Reporting

Instructor environments and exercises will be tested on both GPU clusters and rehearsed before the October 2026 Working Week. Pinned software environments, compact datasets, validated checkpoints, and precomputed fallback outputs will reduce technical risk and prevent uncontrolled resource consumption during the tutorial. The AIFS and regional-model experiments will use short forecast periods and a small set of representative cases. Each experiment will record model version, input data, initialization, forecast length, hardware, numerical precision, wall-clock time, peak GPU memory, GBU consumption, and verification results. After delivery, the materials will be corrected and consolidated into a reusable community package, and all measured resource use will be documented in the final report.

Expected Outcomes & Educational Activities

The project will explore and develop:

- A prototype HCLIM-AI benchmark dataset drawn from existing convection-permitting runs.
- Baseline CNN and GAN setups with preliminary results and evaluation metrics.
- **A practical tutorial package** featuring three modular hands-on tracks (CNN emulators, GAN emulators, and AIFS/regional ML workflows) to build technical capacity across participating teams.
- An initial feasibility summary evaluating AIFS-derived driving conditions to guide future scaling and resource planning.

Community Benefit and Relevance to NWP

The project will establish the first shared HCLIM benchmark and tutorial framework based on kilometre-scale climate simulations. The work is relevant to numerical weather prediction because it develops practical understanding of data-driven mappings between atmospheric states and resolutions, spatial verification, generative representation of unresolved variability, and inference with operationally relevant global AI forecast models. Testing AIFS-CRPS v1 and v2, examining the technical pathway from AIFS output to HCLIM initial and boundary conditions, and comparing regional ML models with physics-based forecasts will connect the tutorial directly to forecast-system design. These activities will strengthen links between the HCLIM community and ECMWF's AI software and data ecosystem. The project is intended to complement, not replace, physics-based NWP and regional climate modelling.

Benchmark datasets and evaluation protocols will support future work on AI emulation of regional climate simulations, AI-driven regional-climate ensembles, and hybrid AI-physics modelling. Tutorial materials will be made available to the HCLIM community subject to data licences and ECMWF access conditions.

Estimation of Computing Resources

The request is intentionally bounded and supports a one-year pilot combining benchmark preparation, compact baseline training, reproducibility testing, and supervised tutorial delivery. GBU is defined as GPU-seconds of use; one GPU-hour therefore consumes 3,600 GBU. We request a maximum of 1.05 million GBU on GPU Cluster-A (291.67 GPU-hours) and 0.99 million GBU on GPU Cluster-B (275 GPU-hours), for a combined ceiling of 2.04 million GBU (566.67 GPU-hours).

GPU Resource Breakdown

- CNN module and deterministic baseline: **540,000 GBU (150 GPU-hours)**, including compact training runs, validation, pre-course testing, and participant exercises.
- GAN module and generative baseline: **720,000 GBU (200 GPU-hours)**, reflecting generator-discriminator updates, sampling, stability tests, and tutorial exercises with bounded configurations.
- AIFS and regional ML forecasting module: **780,000 GBU (216.67 GPU-hours)** for AIFS-CRPS v1/v2 inference, HCLIM initial/boundary-condition feasibility tests, selected AEMET-NWP regional-model experiments, physics-ML comparison, rehearsal, and one bounded training or fine-tuning experiment.

Jobs will be sized and monitored separately on each cluster. Short forecast periods, strict stopping criteria, checkpoints, and precomputed outputs will prevent uncontrolled repeated training. Any substantial surplus will be released in accordance with ECMWF resource-management procedures.

CPU, EWC, and Storage Requirements

No dedicated HPC CPU or EWC resources are requested for this pilot project. To support data extraction, preprocessing, benchmark generation, statistical analysis, and hands-on tutorial activities for approximately 15 participants, an **ECFS/MARS archive allocation of 4 TB** is requested. The estimated storage requirements are as follows:

- AIFS and regional ML activities: up to **100 GB** per participant, including model environments, initial conditions, and outputs from selected inference experiments.
- SRGAN activities: up to **30 GB** of outputs per participant.
- CNN activities: up to **20 GB** per participant.
- Shared files: approximately 1000 GB for datasets, model files, source code, environments, and consolidated tutorial materials.

The expected working requirement is approximately **3.5 TB**. The requested 4 TB provides contingency for ensemble forecasts, longer inference runs, duplicated working files, and temporary outputs.

Technical Characteristics and Risk Management

The tutorial and baseline models will use Python-based accelerator workflows with standard scientific Python tools. CNN and GAN configurations will specify software versions, random seeds, input and output dimensions, batch size, numerical precision, checkpoint frequency, and evaluation metrics. Conservative batch sizes and compact subsets will control GPU memory.

The advanced module will use ECMWF-supported inference routes and pre-trained AIFS-CRPS v1 and v2 models (Lang et al., 2026; ECMWF, 2026). Notebook workflows will standardize initial-condition retrieval, model execution, forecast output, diagnostics, timing, and resource logging. The HCLIM coupling-feasibility assessment will identify required atmospheric and surface variables, vertical-level mappings, grid and spectral transformations, temporal frequency, file formats, metadata, and lateral-boundary update intervals; it will determine technical readiness rather than undertake a production HCLIM integration. Selected AEMET-NWP regional ML models will be run only where code, weights, dependencies, and test data are available, with BRIS treated as an optional candidate. Access-controlled data and model assets will not be redistributed. The small training or fine-tuning experiment will use a compact dataset, fixed epoch or step limit, predefined stopping criteria, and a documented resource ceiling.

Primary risks are environment incompatibility, data or model-access delays, differences between AIFS versions, incomplete availability of regional ML code or weights, HCLIM–AIFS variable and grid incompatibilities, GPU-memory limitations, adversarial or fine-tuning instability, and differing participant experience. Mitigation measures include pinned environments, version-specific notebooks, smoke tests on both clusters, an early inventory of required HCLIM boundary variables, compact data, validated checkpoints, bounded run configurations, optional treatment of BRIS, precomputed fallback results, and instructor monitoring of GBU use.

Reporting and Acknowledgement

A final report will document the development and tutorial activities benchmark datasets, model configurations, AIFS-CRPS v1/v2 and regional-model experiments, the HCLIM initial/boundary-condition feasibility assessment, forecast-quality comparisons, tutorial delivery, execution times, GPU-memory use, and actual GBU consumption. Public materials resulting from the activity will acknowledge ECMWF computing and archive facilities using the wording requested in the EMI R&D project guidelines or an appropriate modification.

References

Belušić, D., de Vries, H., Dobler, A., et al. (2020). HCLIM38: a flexible regional climate model applicable for different climate zones from coarse to convection-permitting scales. Geoscientific Model Development, 13, 1311–1333. https://doi.org/10.5194/gmd-13-1311-2020

Lind, P., Belušić, D., Christensen, O. B., et al. (2020). Benefits and added value of convection-permitting climate modeling over Fenno-Scandinavia. Climate Dynamics, 55, 1893–1912. https://doi.org/10.1007/s00382-020-05359-3

Lind, P., Belušić, D., Médus, E., et al. (2023). Climate change information over Fenno-Scandinavia produced with a convection-permitting climate model. Climate Dynamics, 61, 519–541. https://doi.org/10.1007/s00382-022-06589-3

Moldovan, G., Pinnington, E., Prieto Nemesio, A., et al. (2026). AIFS Single 1.1.0: an update to ECMWF's machine-learned weather forecast model AIFS. Geoscientific Model Development, 19, 4703–4724. https://doi.org/10.5194/gmd-19-4703-2026

Lang, S., Alexe, M., Clare, M. C. A., et al. (2026). AIFS-CRPS: ensemble forecasting using a model trained with a loss function based on the Continuous Ranked Probability Score. npj Artificial Intelligence. https://doi.org/10.1038/s44387-026-00073-7

Lang, S., Leutbecher, M., and Maciel, P. (2025). A multi-scale loss formulation for learning a probabilistic model with proper score optimisation. arXiv:2506.10868. https://doi.org/10.48550/arXiv.2506.10868

ECMWF (2026). AIFS ENS v2 model card and inference notebook. ECMWF model repository. https://doi.org/10.57967/hf/8763

Rampal, N., Gibson, P. B., Sood, A., Stuart, S., Fauchereau, N. C., Brandolino, C., Noll, B., and Meyers, T. (2022). High-resolution downscaling with interpretable deep learning: Rainfall extremes over New Zealand. Weather and Climate Extremes, 38, 100525. https://doi.org/10.1016/j.wace.2022.100525

Wang, F., Wang, Y.-C., Kourabbaslou, H. E., Krus, K., Aldama-Campino, A., Nikulin, G., Döscher, R., Lind, P., Mirjalili, S., Lennard, C., and Schenk, F. (2026). Emulating land-atmosphere feedbacks in convection-permitting regional climate models using machine learning. In revision with Artificial Intelligence for the Earth Systems.

Anemoi Contributors (2026). Anemoi: an open-source framework for developing, training, and deploying machine-learning weather forecasting models. https://github.com/ecmwf/anemoi (accessed 21 August 2026).