# LATE REQUEST FOR A SPECIAL PROJECT 2025–2027

**MEMBER STATE:** ……Germany……..…

**Principal Investigator[1]:** ……Yijuan Zhang ……..……

**Affiliation:** …Institute of Environmental Physics (IUP), University of Bremen……

**Address:** Otto-Hahn-Allee 1, 28359 Bremen, Germany……

……

**Other researchers:** Dr. Nikos Daskalakis (University of Bremen) Prof. Guy Brasseur (Max Planck Institute for Meteorology) Prof. Mihalis Vrekoussis (University of Bremen) Dr. Claire Granier (NOAA/CIRES/University of Colorado) Dr. Marius Dan (University of Bremen)

**Project Title:** …High-resolution computational modeling of integrated anthropogenic emission inventories for atmospheric transport simulations and numerical weather prediction validation……

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** N/A - New project….. | |
| Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.) | 2025 | |
| Would you accept support for 1 year only, if necessary? | YES ☒ | NO ☐ |

| **Computer resources required for project year:** | | **2025** | **2026** | **2027** |
|---|---|---|---|---|
| High Performance Computing Facility | [SBU] | 6,500,000 | 10,000,000 | 12,000,000 |
| Accumulated data storage (total archive volume)[2] | [GB] | 15000 | 35000 | 50000 |

| **EWC resources required for project year:** | | **2025** | **2026** | **2027** |
|---|---|---|---|---|
| Number of vCPUs | [#] | 200 | 400 | 600 |
| Total memory | [GB] | 1600 | 3200 | 4800 |
| Storage | [GB] | 15000 | 35000 | 50000 |
| Number of vGPUs[3] | [#] | 25 | 50 | 70 |

*Continue overleaf.*

[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.

**Principal Investigator:** ……Yijuan Zhang……

**Project Title:** High-resolution computational modeling of integrated anthropogenic emission inventories for atmospheric transport simulations and numerical weather prediction validation

# Extended abstract

*All Special 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 Special 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 Special 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).

**Project Summary:** This project aims to develop a next-generation integrated anthropogenic emission inventory (CINEI version 2) to support atmospheric transport simulations and datasets within ECMWF's operational framework. The development of CINEI v2.1 will incorporate observational datasets from both satellite and in-situ measurements, implementing advanced emission estimation methods including deep learning algorithms. CINEI v2.1 will achieve global coverage while focusing on highly polluted regions—Eastern Asia (particularly China), South America, and Africa—where anthropogenic emissions are experiencing rapid changes due to industrial growth and mitigation policies. The enhanced 0.1-degree resolution of CINEI v2.1 will enable city-scale modeling and forecasting capabilities. This project will apply CINEI v2.1 in regional Chemistry Transport Models (CTM) and validate its performance against observational datasets. The primary objective is to provide high-resolution emission data and ensure the seamless integration of CINEI v2.1 into ECMWF's operational systems, specifically the Integrated Forecasting System (IFS) and the Artificial Intelligence Forecasting System (AIFS), thereby enhancing atmospheric chemistry modeling within numerical weather prediction frameworks.

**Team Expertise and Computational Requirements:** Our research team brings extensive expertise in HPC utilization and atmospheric numerical modeling. We have successfully published methodological advances in Geoscientific Model Development (GMD), demonstrating our capability in model development and validation. The first version of our emission inventory (CINEI v1.1) has been published in the PANGAEA repository (DOI: 10.1594/PANGAEA.974347) and has gained significant traction in the atmospheric modeling community. Our collaborators possess comprehensive experience in WRF-Chem modeling, satellite data processing, and machine learning applications for atmospheric sciences, ensuring the technical competence required for successful project completion.

This project requires High Performance Computing (HPC) resources to process multi-terabyte observational datasets, train deep learning models, and conduct CTM experiments at unprecedented resolution. Our institutional HPC cluster (Aether) cannot adequately support this project due to limited computational capacity (65 Tflop/s), intense competition among 30+ users, and near-exhausted storage capacity, preventing long-running global simulations essential for this work.

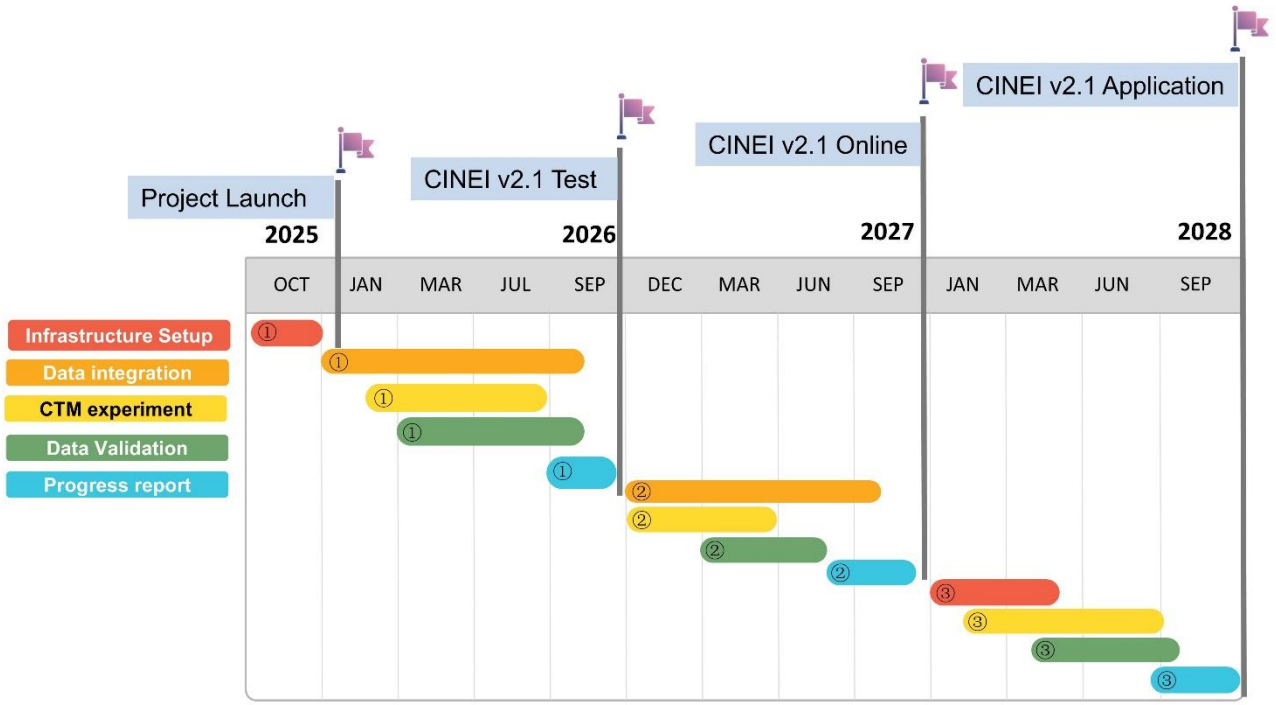


Figure 1. Three-year project timeline for CINEI v2.1 development and application

Table 1. Three-Year Task Lists and Computing Resource Requirements for Each Task

| 1st Year | Tasks | Task Lists | HPC (SBu) 6.5M | vCPU 200 | Memory (GB) 1600 |
|---|---|---|---|---|---|
| Oct. 2025 To Sep. 2026<br><br>**CINEI v2.1 Test Dataset** | **Infrastructure Setup** | • Enable WRF-Chem, Jupyter environments<br>• Data migration, including CINEIv1.1, observation | | 50 | 400 |
| | **Data integration** | • Incorporating Satellite, In-situ measurement<br>• Implement deep learning models using XGBoost for emission estimation (East Asia--China, South America, and Africa) | 3.8M | 100 | 800 |
| | **CTM experiment** | • Conducting WRF-Chem air quality simulation<br>• Three targeted regions in 12 km spatial resolution | 2.2M | 200 | 1600 |
| | **Data validation** | • Validate simulated air pollutant concentrations against regional observations<br>• Quantify uncertainties in CINEI v2.1 emission estimates | 0.5M | 150 | 1200 |
| | **Progress report** | • Completion and submission of CINEI v2.1 test dataset to ECMWF<br>• Deliver comprehensive validation report to ECMWF | | 50 | 400 |

| 2nd Year | Tasks | Task Lists | HPC (SBu) 10M | vCPU 400 | Memory (GB) 3200 |
|---|---|---|---|---|---|
| Oct. 2026 To Sep. 2027<br><br>**CINEI v2.1 Online** | **Data integration** | • Advance emission estimation methodologies<br>• Merge regional datasets into a unified global inventory<br>• Develop and release Python package for user accessibility | 6M | 200 | 1600 |
| | **CTM experiment** | • Conduct global CTM air quality simulation<br>• Implement finer spatial resolution (0.1° × 0.1°) | 3.5M | 400 | 3200 |
| | **Data validation** | • Validate CINEI v2.1 performance in global CTM<br>• Assess CINEI v2.1 global impact on air quality predictions | 0.5M | 300 | 2400 |
| | **Progress report** | • Publish and submit of CINEI v2.1 dataset to ECMWF<br>• Reporting the validation to ECWMF | | 100 | 800 |

| 3rd Year | Tasks | Task Lists | HPC (SBu) 12M | vCPU 600 | Memory (GB) 4800 |
|---|---|---|---|---|---|
| Oct. 2027 To Sep. 2028<br><br>**CINEI v2.1 Application** | **Infrastructure Integration** | • Adapt CINEI v2.1 for integration with IFS and AIFS | 3M | 300 | 2400 |
| | **CTM experiment** | • Collaborate with ECMWF teams on IFS and AIFS model integration | 4.5M | 600 | 4800 |
| | **Data validation** | • Assess CINEI v2.1 impact on ECMWF modeling performance | 3.5M | 400 | 3200 |
| | **Progress report** | • Provide operational services for simulations and datasets to ECMWF<br>• Submit comprehensive three-year progress report to ECMWF | 1.0M | 200 | 1600 |

## Research Innovation and Technical Approach

Building upon our published CINEI v1.1 inventory (Zhang et al., 2025; https://doi.pangaea.de/10.1594/PANGAEA.974347) and python code for project framework (Zhang, 2025; https://doi.org/10.5281/zenodo.15000795), this project develops CINEI v2.0 with three major computational advances specifically relevant to ECMWF's modeling framework:

1. Enhanced spatial resolution (0.25° to 0.1°) requiring intensive spatial interpolation algorithms compatible with ECMWF's grid structures
2. Machine learning integration using eXtreme Gradient Boosting (XGBoost) to assimilate TROPOMI satellite observations with ground-based measurements, providing accurate emission updates for operational ECMWF systems
3. Expanded geographical coverage from East Asia to include South America and Africa's most polluted regions, supporting global atmospheric modeling initiatives

## Computational Methodology and ECMWF Data Integration

The workflow integrates multiple high-resolution datasets through computationally intensive processes optimized for atmospheric modeling applications:

- Satellite data preprocessing using TROPOMI observations compatible with ECMWF's observational data assimilation systems
- Deep learning model training for emission adjustment using meteorological fields consistent with ECMWF's atmospheric model formulations
- Spatial disaggregation algorithms that maintain mass conservation properties essential for atmospheric transport modeling
- Validation against regional Chemical Transport Model (WRF-Chem) simulations using meteorological forcing compatible with ECMWF's numerical weather prediction outputs

## Specific HPC Resource Justification:

- **Oct. 2025-Sep. 2026 (6.5M SBU):** Development of CINEI v2.1 test dataset covering three targeted regions (East Asia/China, South America, and Africa) with 12 km resolution CTM validation simulations
- **Oct. 2026-Sep. 2027 (10M SBU):** Global implementation of CINEI v2.1 at 0.1° resolution, merging regional datasets into unified global inventory, and conducting global CTM validation
- **Oct. 2027-Sep. 2028 (12M SBU):** Adaptation of CINEI v2.1 dataset for IFS and AIFS usage within ECMWF, integration with operational systems, and establishment of real-time emission services with 6-hourly update cycles

## Technical Characteristics and Software Requirements

The computational framework utilizes:

- WRF-Chem atmospheric chemistry model with MOZART-4 mechanism (200+ gas-phase species)
- Python-based processing tools for multi-dimensional data fusion across 8 emission sectors
- XGBoost machine learning implementation requiring GPU acceleration
- Conservative regridding algorithms using xESMF library for mass-conserving grid transformations
- NetCDF data formats compatible with ECMWF's data infrastructure

# Expected Contributions and Deliverable Datasets to ECMWF

Results will provide ECMWF with unprecedented high-resolution emission datasets and operational capabilities:

- Enhanced numerical weather prediction accuracy through improved aerosol-meteorology interactions in IFS and AIFS
- Validated emission datasets compatible with ECMWF's global modeling framework for atmospheric transport simulations
- Operational emission services integrated with weather prediction systems for real-time forecasting applications
- Direct support for CAMS (Copernicus Atmosphere Monitoring Service) through quality-assured emission inputs with uncertainty quantification

Specific Deliverables:

- CINEI v2.1 global dataset at 0.1° resolution in GRIB2 format, fully adapted for IFS and AIFS modeling within ECMWF
- Python package v2 for data access, processing, and operational integration
- Comprehensive validation reports documenting performance metrics, uncertainty quantification, and model improvements for ECMWF evaluation

The methodology advances computational approaches for Earth system modeling and directly contributes to ECMWF's mission of improving numerical weather prediction through better representation of atmospheric composition.

## Previous Work and Validation

Our team has demonstrated computational expertise through:

- Successfully published CINEI v1.1 emission inventory with comprehensive validation
- Proven WRF-Chem modeling capabilities with 49-process MPI configurations
- International collaboration experience through atmospheric chemistry working groups
- Validated methodology achieving 12% normalized bias for ozone predictions

**Justification for the requirement of External HPC Resources:** Our institutional HPC cluster (Aether) at the University of Bremen provides limited computational capacity that is insufficient for this large-scale project. The Aether system offers a theoretical peak performance of approximately 65 Tflop/s across 56 compute nodes (2× Intel Xeon E5-2690v4 per node), with 128-256 GB RAM per node, 1.1 PB BeeGFS project storage, and 44 TB home storage with Intel OmniPath interconnect.

However, this system currently supports over 30 active users, resulting in intense competition for computational resources. Additionally, the available disk space is nearly exhausted, severely limiting our ability to process large-scale atmospheric datasets or conduct long-running CTM simulations. While our research group has scheduling priority, the cluster's capacity is fundamentally insufficient for the computational demands of developing a global high-resolution emission inventory. Therefore, access to ECMWF's HPC resources is essential for successful project completion.

References:


- Zhang, Y., Brasseur, G., Kanakidou, M., et al.: Towards an integrated inventory of anthropogenic emissions for China, EGUsphere, https://doi.org/10.5194/egusphere-2025-268, 2025.
- Published CINEI v1.1 dataset: https://doi.org/10.1594/PANGAEA.974347
- Public code for current project (CINEIv1.1): https://doi.org/10.5281/zenodo.15000795


This research directly supports ECMWF's objectives by providing essential emission data for atmospheric modeling applications while advancing computational methodologies for numerical weather prediction and atmospheric composition monitoring.