LATE REQUEST FOR A EMI R&D PROJECT 2026–2028

Country/Organisation: Austria / University of Vienna

Principal Investigator[1]: Sarah Hauser, Dr.-Ing. (B.Eng., M.Sc.)

Affiliation: University of Vienna, Department of Geography and Regional Research, Geoecology

Address: Geoecology
Josef-Holaubek-Platz 2 (UZA II)
1090 Vienna, Austria

Other researchers: none

Project Title: Scalable and Semantically Enriched Earth Observation Processing for Environmental Monitoring

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.	**N/A**	
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	0	0
Graphics Processing Unit Cluster-A	[GBU]	0	0	0
Graphics Processing Unit Cluster-B	[GBU]	0	0	0
Accumulated data storage (total archive volume)[2]	[GB]	0	0	0

[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	[#]	8	8	8
Total memory	[GB]	64	64	64
Storage	[GB]	2000	2000	2000
Number of vGPUs[3]	[#]	1	1	1

Continue overleaf.

Principal Investigator: Sarah Hauser, Dr.-Ing. (B.Eng., M.Sc.)

Project Title: Scalable and Semantically Enriched Earth Observation Processing for Environmental Monitoring

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).

Project summary and scientific objectives

This project investigates scalable and semantically enriched methods for processing, combining and interpreting multi-source Earth observation time series. It forms a methodologically connected component of a broader habilitation research programme in geoecology and environmental Earth observation. The central scientific objective is to explore how incomplete and heterogeneous observations can be reconstructed adaptively and transparently while retaining information about data quality, uncertainty and provenance. The project connects several methodological components, including Earth observation data preparation and representation, adaptive handling of temporal gaps, multi-sensor integration, semantic enrichment and environmental model development. Sentinel-1 synthetic aperture radar, Sentinel-2 multispectral imagery and climate and land-surface data will provide a shared experimental basis. Selected environmental systems will be used to investigate how well the methods adapt to different surface characteristics, temporal dynamics, data availability and monitoring objectives.

The project builds on previous methodological work by Schmitt and colleagues and on the principal investigator's research concerning multisensor and multitemporal Earth observation, environmental reference data, feature and label enhancement, and transparent assessment of spatio-temporal predictions. The Kennaugh element representation provides a consistent basis for preparing and comparing multi-polarised, multi-temporal and multi-scale SAR information (Schmitt et al., 2015). Its extension towards multi-source and multi-temporal representations provides a methodological basis for combining radar and optical information (Schmitt et al., 2020).

One component of the proposed work will establish reproducible Sentinel-1 processing based on the Kennaugh representation. A dedicated processing implementation will make the relevant processing steps controllable and scalable and will generate radar features for subsequent time-series, fusion and modelling experiments. The processor is therefore a supporting methodological component rather than the principal scientific outcome. Existing software such as ESA SNAP may be used for reference processing and comparison where appropriate. The potential of Kennaugh elements for environmental monitoring has already been demonstrated in forest applications, including early detection of bark beetle infestation from Sentinel-1 observations (Hechtl et al., 2025). The resulting Sentinel-1 representations will be investigated together with Sentinel-2 observations, ECMWF ERA5 and ERA5-Land variables, and ancillary terrain, land-cover and ecosystem information. This component will explore how radar, optical and environmental data can be harmonised and combined while retaining relevant spatial, temporal and contextual information. Previous work on an open benchmark data set for forest characterisation from Sentinel-1 and Sentinel-2 time series provides an existing basis for evaluating multi-sensor processing and machine-learning approaches in forest systems (Hauser et al., 2024). A central component will explore adaptive approaches for detecting and reconstructing gaps in Earth observation time series. Gaps may result from cloud cover, acquisition schedules, quality filtering, missing observations or inconsistent temporal sampling. Different reconstruction strategies will be compared with respect to accuracy, robustness and suitability for different environmental conditions. Beyond producing a filled time series, the work will investigate mechanisms for transparent feedback on the observations, temporal relationships and assumptions supporting each reconstructed value.

Semantic enrichment will connect observations and derived features with information about acquisition conditions, environmental context, data quality, processing decisions, labels and model outputs. Previous research on label optimisation in dynamic Earth observation demonstrates the importance of considering label quality, temporal validity and refinement as integral parts of Earth observation modelling (Hauser et al., 2025). The proposed work will extend this perspective to reconstructed observations and multi-source time series. Methods for model-agnostic assessment of trust in spatio-temporal geospatial predictions provide a further conceptual basis for communicating data support, uncertainty and potential limitations to users and downstream models (Hauser, 2026).

The harmonised and semantically enriched time series will subsequently be used in statistical and machine-learning experiments. Models will be evaluated not only according to predictive performance but also with respect to robustness, transferability and sensitivity to input-data quality. Depending on the selected application, this may include environmental classification, estimation of ecosystem properties, analysis of seasonal dynamics or detection of environmental change. Potential test domains include High Arctic landscapes such as Axel Heiberg Island, peatland and wetland systems, and forest systems. These domains represent complementary evaluation environments rather than independent projects. The High Arctic experiments can build on an existing monthly harmonised Sentinel-1 and Sentinel-2 data set for Axel Heiberg Island (Cantatore et al., 2026). Forest applications can draw on existing time-series and reference data for forest characterisation. Peatlands and wetlands provide an additional domain in which surface moisture, vegetation dynamics, climate variability and seasonal data availability are closely connected. Not every domain will require the same processing or modelling approach; their role is to test adaptability and reveal where application-specific modifications remain necessary.

The initial project phase will establish the European Weather Cloud environment, organise data access, assemble representative data sets and implement the core Sentinel-1 processing components. Work during 2027 will concentrate on adaptive gap-handling experiments, transparent feedback mechanisms, semantic descriptions and selected approaches for combining Sentinel-1, Sentinel-2 and environmental data. During 2028, the methodological components will be evaluated in the selected environmental domains and combined with model-training experiments. The final phase will assess uncertainty, transparency and transferability and consolidate reusable processing components and documentation. If only one year of support is initially available, the Sentinel-1 processing component and one transparent time-series reconstruction demonstrator will provide a self-contained first result.

Processing will primarily use Python packages developed or maintained in version-controlled Git repositories. The anticipated software stack includes xarray, Dask, rasterio, GDAL, GeoPandas, scikit-learn and, where required, PyTorch or comparable machine-learning libraries. QGIS and ESA SNAP may be used for visual inspection, cartography, reference processing and quality control. Reproducible software environments will support recurring interactive development as well as unattended raster-processing and model-training tasks lasting from several hours to several days.

Eight vCPUs and 64 GB of memory are requested for parallel raster preprocessing, multi-temporal data stacks and combined Python and GIS workflows. Two terabytes of storage are required for active Sentinel-1 and Sentinel-2 scenes, environmental data, intermediate products and reproducible outputs. One vGPU is requested for selected machine-learning experiments and will not be required continuously.

By the end of 2028, the project is expected to provide evaluated Earth observation processing components, evidence concerning adaptive and transparent reconstruction of incomplete time series, tested approaches for combining radar, optical and environmental information, and modelling experiments across contrasting environmental systems. Results are expected to contribute to peer-reviewed publications and the broader habilitation research programme. Reusable code and processing documentation will be shared where licensing, data protection and publication considerations permit.

References:

Schmitt, A., Wendleder, A. and Hinz, S. (2015). The Kennaugh element framework for multi-scale, multi-polarized, multi-temporal and multi-frequency SAR image preparation. ISPRS Journal of Photogrammetry and Remote Sensing, 102, 122–139. https://doi.org/10.1016/j.isprsjprs.2015.01.007

Schmitt, A., Wendleder, A., Kleynmans, R., Hell, M., Roth, A. and Hinz, S. (2020). Multi-Source and Multi-Temporal Image Fusion on Hypercomplex Bases. Remote Sensing, 12(6), 943. https://doi.org/10.3390/rs12060943

Hauser, S., Ruhhammer, M., Schmitt, A. and Krzystek, P. (2024). An Open Benchmark Dataset for Forest Characterization from Sentinel-1 and -2 Time Series. Remote Sensing, 16, 488. https://doi.org/10.3390/rs16030488

Hechtl, C., Hauser, S., Schmitt, A., Heurich, M. and Wendleder, A. (2025). Kennaugh Elements Allow Early Detection of Bark Beetle Infestation in Temperate Forests Using Sentinel-1 Data. Forests, 16, 1272. https://doi.org/10.3390/f16081272

Hauser, S., Augner, L. and Schmitt, A. (2025). Perfect Labelling: A Review and Outlook of Label Optimization Techniques in Dynamic Earth Observation. Remote Sensing, 17, 1246. https://doi.org/10.3390/rs17071246

Cantatore, F., Hauser, S. and Schmitt, A. (2026). Monthly Harmonized Sentinel-1 and Sentinel-2 Dataset over Axel Heiberg Island. Zenodo. https://doi.org/10.5281/zenodo.20509321

Hauser, S. (2026). HELIX-Trust: A Model-Agnostic Trust Framework for Spatio-Temporal Geospatial Predictions. gis.Science. https://doi.org/10.14627/gis.science.2026.2.3