REQUEST FOR AN EMI R&D PROJECT 2027–2029

Country/Organisation: GREECE

Principal Investigator[1]: KONSTANTINOS KOLOMVATOS

Affiliation: UNIVERSITY OF THESSALY

Address: 3o km Lamia – Athens. Nea Ampliani area, Lamia P.C. 35100

Other Researchers: TERRA Consortium

Project title: An intelligent platform for integrating climate services (TERRA)

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.)	2027	
Would you accept support for 1 year only, if necessary?	**YES** ☒	~~NO~~ ☐

Computer resources required for project year:		**2027**	**2028**	**2029**
High Performance Computing Facility	[SBU]	6.5M	6.5M	6.5M
Graphics Processing Unit Cluster-A	[GBU]	4.3M	4.3M	4.3M
Graphics Processing Unit Cluster-B	[GBU]	-	-	-
Accumulated data storage (total archive volume)[2]	[GB]	2283	2283	2283

[1] The Principal Investigator is the contact person for this EMI R&D Project

EWC resources required for project year:		**2027**	**2028**	**2029**
Number of vCPUs	[#]	88	88	88
Total memory	[GB]	192	192	192
Storage	[GB]	710	710	710
Number of vGPUs[3]	[#]	4	4	4

Continue overleaf.

Principal Investigator: KONSTANTINOS KOLOMVATSOS

Project Title: An intelligent platform for integrating climate services (TERRA)

Extended abstract

A. Project Description

The TERRA project develops an intelligent platform for integrating climate services by combining Copernicus Earth Observation data, advanced AI methods, in-situ measurements and Digital Twin technologies. It builds on Copernicus Sentinel observations together with operational Copernicus services such as the Copernicus Emergency Management Service (CEMS) and the Climate Change Service (C3S), aiming to improve environmental monitoring, forecasting, and decision support. Instead of producing isolated AI results, TERRA follows a modular Digital Twin approach that creates persistent and continuously updated digital representations of environmental systems. The project focuses on three main areas: assessment of water contamination in coastal areas and in the water cycle, illegal human activity detection in harbour areas for environmental risk assessment, and coastal change monitoring from satellite remote sensing. Each output will be visualized into the platform which is designed to provide reusable product chains and interoperable services that can support emergency response, climate adaptation, and long-term environmental resilience across different regions.

B. Scientific Plan

The requested ECMWF HPC resources directly support the implementation and scaling of existing TERRA work packages, mainly within WP2 (Use Cases and Demonstrators) and WP3 (Intelligent Platform Development). WP2 defines the operational use cases of the project and provides the scientific context for the proposed workloads. Demonstrator 1 focuses on water contamination assessment and water-cycle monitoring, combining hydrological analysis, Earth Observation data, and predictive modelling for pollution forecasting. Demonstrator 2 supports the overall framework through harbour monitoring and environmental risk assessment using vessel detection and maritime situational awareness. Demonstrator 3 focuses on Digital Twins for coastal change monitoring, including shoreline evolution, vegetation edge stability, and coastal-state forecasting. These demonstrators define the real environmental problems that require large-scale processing and repeated forecasting cycles. WP3 provides the computational and technical backbone that justifies the use of ECMWF HPC resources. In particular, Task 3.3 develops the AI models for water quality prediction using multi-temporal Sentinel-2/3 data and meteorological variables, requiring repeated model training, fine-tuning, transfer learning, inference, and long-term historical reprocessing. Task 3.4 focuses on advanced satellite image processing, including high-resolution segmentation, object detection, super-resolution, and EO-based feature extraction across large spatial and temporal scales. Task 3.5 implements the Digital Twin framework, where hydrological states, observation states, and prediction states are continuously updated to support persistent monitoring and forecasting services. In parallel, Task 3.1 supports data management and orchestration through scalable storage and workflow execution, while Task 3.2 ensures the interoperability of Copernicus services, APIs, and external data sources. WP4 (Platform integration and validation) and WP5 (Demonstration activities) ensure the integration, validation and operational assessment of the forecasting outputs generated by WP2 and WP3.

AI-Driven Solution

Once we have our high-resolution maps ready we will perform several specialized tasks to ensure accurate and reliable forecasts. For water quality, we use Bidirectional LSTM layers and Temporal Attention Mechanisms to learn from complex historical patterns and produce 3-to-7-day forecasts. To handle small objects like maritime vessels in medium-resolution imagery, we employ YOLO-family detectors featuring Space-to-Depth (SPD) convolutions and deformable backbones that adapt to specific ship shapes. We also use the Segment Anything Model (SAM) for "zero-shot" detection, allowing the system to identify unknown vessels without requiring prior training. For coastline monitoring, we use RobustUNet for pixel-level segmentation, followed by a Mamba-inspired temporal mixer that works with LSTM memory to predict future shoreline positions with high efficiency. Finally, to make our system trustworthy, we use Gaussian Process Regression (GPR) with Matern kernels and Monte Carlo Dropout to measure uncertainty, ensuring every prediction comes with a clear confidence level. The technical framework for Task 2 is supported by D3.3 (Sections 5.1, 3.6, and 4.2) for the LSTM and weather fusion logic, and D3.5 (Section 5.2) for its integration into the Digital Twin's "Prediction State".

C.1. Justification of Computer Resources Requested (training in HPC)

The requested HPC resources are necessary to support the computational intensity of high-resolution regional pre-processing, the iterative deep-learning workflows, and finally to support validation and testing. In Table 1, for each job we justify the required computational resources, including both SBU usage and storage needs.

Table 1: Identified AI-Driven Jobs

AI-Driven Solutions	*SBU / Year*	*GBU-A / Year*	*Storage / Year*
Job 1: Forecasting of water contamination in coastal areas and in the water cycle with Digital Twin.	*2M*	*600K*	*25GB*
Job 2: Vessels detection on high resolution satellite imagery.	*2.1M*	*1.6M*	*250GB*
Job3:Illegal vessels activity detection in harbour areas leveraging Copernicus services. Job 4: Prediction of pollution in port areas.	*1M*	*500K*	*6 GB*
Job 5: Detection of Coastal Changes (erosion/flood) from Satellite Remote Sensing with Digital Twins.	*1.4M*	*1.6M*	*2TB*
Total resource requirements	***6.5M***	***4.3M***	***2.3TB***

Note: GPU resources were estimated using ECMWF's GPU accounting approach, where GBU corresponds to GPU-seconds of usage. Therefore, the use of 1 GPU for 1 hour corresponds to 3,600 GBU. The estimated GBU requirement for each job is calculated as:

Estimated GBU = Number of GPUs × Runtime per experiment in hours × 3,600 × Number of experiments

Justification for Job 1:

The proposed workflow generates water quality predictions through a hybrid pipeline that combines Earth Observation data processing, environmental modelling, and machine learning. For the primary Area of Interest (AOI), covering the Sperchios River and Malian Gulf, the region is divided into 10 spatial tiles, with the possibility to scale up to 30 tiles depending on resolution requirements. Each tile follows the same processing pipeline, including data collection, cleansing and interpolation, multi-source data fusion, feature engineering, and training or inference using a Bi-directional LSTM-based forecasting model. Benchmarking experiments for the 10-year historical period 2016–2026, with approximately 3,600 timesteps and around 20 features per timestep, show that one tile requires about 200–240 seconds on a single CPU core. Additional hydrological preprocessing using high-resolution DEM data requires approximately 120 seconds per AOI.

Using the newest ECMWF HPC2020 accounting model, with SBU = P × N × T and P ≈ 0.00473, one tile execution on one CPU core requires approximately: SBUtile ≈ 0.00473 × 1 × 220 ≈ 1.0428 SBUs

This corresponds to approximately 11.5 SBUs for one AOI with 10 tiles, or 34.5 SBUs for a higher-resolution setup with 30 tiles. For repeated Digital Twin experiments and scenario runs, assuming 100 executions, the expected requirement is approximately 1,0428–3,128 SBUs, depending on the number of tiles. In addition, an auxiliary model will be developed to include in-situ sensor data, such as turbidity, conductivity, and temperature from USV measurements. Benchmarking for this component is still ongoing, but its computational cost is expected to be comparable to or lower than the main forecasting model. A future extension will move from tile-level forecasting to finer per-pixel processing within 512 × 512 satellite imagery. This will significantly increase the computational demand and is the main reason for requesting additional HPC capacity. Based on current estimates, this higher-resolution scenario may require up to approximately 20,000 SBUs per AOI for one execution, and up to 2,000,000 SBUs for 100 experimental or Digital Twin simulation runs.

For this job we are going to need around 25GB to store the corresponding ML models and the datasets needed for trainings, including satellite data from Sentinel-2 and Sentinel-2 Copernicus Satellites and ECMWF meteorological datasets.

GPU resources are required for accelerating selected training, fine-tuning, and evaluation of the Bi-LSTM and attention-based forecasting models used for water contamination prediction. Since this is mainly a time-series forecasting workflow, it does not require large multi-GPU execution. A moderate allocation of 600,000 GBU-A per year is requested for this job.

The estimate assumes 1 GPU, approximately 10 hours per selected training or evaluation run, and 17 selected GPU-accelerated runs per year: 1 GPU × 10 hours × 17 runs × 3,600 = 612,000 GBU-A/year

This is rounded to 600,000 GBU-A per year. The remaining tile-based preprocessing, hydrological processing, data fusion, and Digital Twin execution will remain covered by the CPU-based SBU allocation.

Justification for Job 2:

VTG contributes the vessel-detection workflow within the TERRA platform, focusing on the detection of small maritime vessels from high-resolution and super-resolved satellite imagery. The workflow supports the harbour and coastal monitoring demonstrators by transforming Copernicus-derived imagery into AI-ready tiles, training a YOLO-based vessel detector, validating the model, and generating geospatial vessel-detection products for further analysis and visualization. The computational workload is dominated by model training. Current benchmarking is based on CPU-only execution using 16 CPU cores, 150 training epochs, and an average execution time of approximately 1,800 seconds per epoch. Using the newest ECMWF HPC2020 accounting model: where P = 0.00473, N = 16 CPU cores, and T = 1,800 × 150 seconds, the estimated requirement is: SBU = 0.00473 × 16 × (1,800 × 150) = 20,476 SBUs

This corresponds to approximately 75 hours for a complete training campaign under the current benchmark configuration. The requested HPC allocation is required for model training, benchmarking, validation, and reproducible experimentation. While operational inference is expected to be significantly less demanding and may be executed within the cloud infrastructure, the training phase requires substantial computational resources due to repeated loading and augmentation of satellite imagery, model optimization, validation, and checkpoint generation. The estimated requirement of 20,476 SBUs therefore represents the minimum computational allocation needed to complete one full vessel-detection training cycle and up to 2,047,680 SBUs for performing 100 training cycles that allows for hyperparameter configuration and conducting experimental validation per year to meet the project objectives.

GPU resources are required for accelerating selected YOLO-based vessel-detection training, super-resolution experiments, segmentation-assisted analysis, validation, and benchmarking on high-resolution satellite imagery. This is one of the most GPU-dependent tasks in the project, but not all experimental cycles need to be executed fully on GPU resources. A moderate allocation of 1,600,000 GBU-A per year is requested for this job.

The estimate assumes 2 GPUs, approximately 20 hours per selected training or validation run, and 11 selected GPU-accelerated runs per year: 2 GPUs × 20 hours × 11 runs × 3,600 = 1,584,000 GBU-A/year

This is rounded to 1,600,000 GBU-A per year. CPU-based preprocessing, image tiling, geospatial handling, and large-scale data preparation remain covered by the SBU allocation.

Justification for Job 3 and 4:

The proposed workflow provides continuous monitoring and intelligent analysis of maritime traffic through the integration of real-time and historical Automatic Identification System (AIS) data with AI-based trajectory prediction and anomaly detection models. The study focuses on a monitored maritime area containing approximately 500 vessels, with operational inference applied to a filtered subset of 170–250 vessels per prediction cycle.

The AI model is trained on approximately 8 GB of historical AIS and SAR data collected over a one-year period. Three machine learning architectures are developed and evaluated, Random Forest, LSTM, and Transformer-based models, to identify the configuration that maximises prediction accuracy while maintaining computational efficiency suitable for near-real-time deployment. Once trained, the selected model generates a 30-minute trajectory forecast of 360 predicted positions at 5-second intervals for each qualifying vessel, updated every 30 seconds. Predicted trajectories are continuously compared against incoming AIS observations to compute per-vessel deviation metrics, with a quantitative risk score assigned to each monitored vessel based on trajectory prediction error, navigational consistency, proximity to critical maritime assets, and historical behavioural patterns.

The computational workload encompasses AI model training, multi-architecture evaluation, hyperparameter optimisation, continuous operational inference, anomaly detection, GNSS spoofing identification, and periodic model retraining against continuously acquired AIS and SAR observations. Based on the expected workload, Job 3 is estimated to require approximately 30,000 SBUs per year × 33 tests ≈ 1milion SBUs per year, covering both the research and operational phases of the system throughout the project lifetime.

Regarding Job4 the proposed workflow monitors water quality and vegetation condition in the vicinity of the Port of Gdańsk using Sentinel-2 imagery. Historical results are available for the period 2016–2026, while new products are generated on a monthly basis as new satellite observations become available. The initial deployment requires the processing of the complete historical archive, corresponding to approximately 11 years of monthly observations across 24 spatial tiles. After the historical baseline has been established, the system performs one operational update per month to incorporate newly available Sentinel-2 data. Based on current benchmarking, the processing of a single tile requires approximately 60 seconds on one CPU core. Using the newest ECMWF HPC2020 accounting model this corresponds to approximately 0.284 SBUs per tile. Monthly processing of the full Area of Interest (24 tiles) requires approximately 6.83 SBUs. The one-time preparation of the historical archive requires approximately 1,000 SBUs, while subsequent monthly

updates introduce only a small additional computational cost. The requested resources therefore support both the initial generation of the historical baseline and the continuous operational monitoring of water quality and vegetation conditions throughout the project lifetime.

GPU resources are required only for accelerating the deep-learning components of the harbour-monitoring and port-pollution workflows, including selected LSTM and Transformer-based trajectory-prediction, anomaly-detection, and pollution-prediction experiments. A significant part of this workload remains CPU-based, including Random Forest models, AIS/SAR data handling, geospatial preprocessing, and routine Sentinel-2 processing. Therefore, a moderate combined allocation of 500,000 GBU-A per year is requested for Jobs 3 and 4.

The estimate assumes 1 GPU, approximately 10 hours per selected deep-learning experiment, and 14 selected GPU-accelerated runs per year: 1 GPU × 10 hours × 14 runs × 3,600 = 504,000 GBU-A/year. This is rounded to 500,000 GBU-A per year.

Justification for Job 5:

The proposed workflow supports the detection and forecasting of coastal change through a Digital Twin pipeline combining Sentinel-2 image processing, automated vegetation-edge extraction, shoreline state construction, and uncertainty-aware forecasting. The workflow processes 5-band Sentinel-2 imagery, extracts vegetation-edge geometry using a RobustUNet model, converts the results into transect-based time series, and generates forecast and uncertainty products for Digital Twin operation.

The next development stage includes: (i) multi-AOI historical reprocessing and vegetation-edge dataset generation, (ii) repeated retraining of the RobustUNet extraction model as new labelled data become available, (iii) retraining and calibration of the forecasting models for 6- and 12-month prediction horizons, and (iv) benchmarking alternative forecasting approaches to improve prediction accuracy and robustness.

Using the newest ECMWF HPC2020 accounting model the annual resource estimate is based on the following major activities:

- *Historical Sentinel-2 reprocessing and vegetation-edge dataset generation: 217,728 SBUs/year*
- *RobustUNet retraining and evaluation: 696,730 SBUs/year*
- *Forecasting model retraining and calibration: 181,440 SBUs/year*
- *Forecasting model benchmarking and comparison: 217,728 SBUs/year*
- *Hindcast validation and uncertainty analysis: 96,768 SBUs/year*

This results in a total estimated requirement of approximately 1.4 million SBUs per year. The associated storage requirement is estimated at approximately 2 TB per year, including Sentinel-2 imagery, vegetation-edge products, transect time-series, model checkpoints, forecasting outputs, uncertainty products, and experiment logs. Access to HPC resources is essential to support large-scale historical reprocessing, repeated model retraining, uncertainty analysis, and the expansion of the coastal Digital Twin across multiple Areas of Interest.

GPU resources are required for accelerating selected RobustUNet-based segmentation, retraining, evaluation, forecasting-model calibration, benchmarking, and uncertainty-aware validation for coastal-change detection. This is the largest GPU allocation because the workflow includes image-based segmentation and repeated model evaluation over Sentinel-2 datasets. A moderate allocation of 2,700,000 GBU-A per year is requested for this job.

The estimate assumes 2 GPUs, approximately 25 hours per selected segmentation, retraining, or validation run, and 15 selected GPU-accelerated runs per year: 2 GPUs × 20 hours × 11 runs × 3,600 = 1.584.000 GBU-A/year

Historical reprocessing, geospatial preparation, transect generation, and routine Digital Twin execution remain primarily covered by the CPU-based SBU allocation.

C.2 Cloud Infrastructure and Operational Platform Requirements

The requested cloud infrastructure resources support the operational deployment of the TERRA platform and its services. Unlike the HPC workloads described previously, which focus on large-scale experimentation, training, and historical reprocessing, the cloud infrastructure is intended to support the continuous execution of the platform, including: frontend and visualization services, backend APIs, database operations, EO data ingestion, scheduled forecasting jobs, user authentication and access management, geospatial processing services, operational monitoring and logging, and near real-time execution of daily workflows.

The table below presents the estimated cloud resource requirements for each module of the TERRA architecture, following a per-module deployment approach. Each module is expected to be developed, packaged, and executed as an

independent Docker container or containerized service, allowing modular deployment, easier maintenance, scalability, and clearer resource allocation. For the ML-related components, only operational inference services are included in the cloud resource estimates. Large-scale model training, benchmarking, and extensive historical reprocessing are considered HPC workloads and are therefore excluded from the cloud infrastructure requirements presented below.

The cloud resource estimates presented above define the minimum operational setup required to deploy the TERRA platform with basic scalability. The minimum values support normal operation and limited concurrent workloads, while the maximum values provide initial scaling capacity for increased data volumes, users, AOIs, and processing frequency. These estimates exclude large-scale model training, historical reprocessing, and other HPC workloads.

Module Name	***vCPUs***	***RAM (GB)***	***Storage (GB)***	***vGPUs***	***Justification***
[M1] Data ingestion module	*16*	*64*	*100*	*No*	*M1 acquires and handles large volumes of data from Copernicus and external repositories. The module formulates the multi-dimensional data streams that feed the TERRA backend and requires sufficient CPU and memory resources to avoid delays and increased latency across the platform.*
[M2] ML Models	*16*	*16*	*100*	*2*	*M2 hosts the machine-learning and deep-learning inference services used across the TERRA platform, including Swin2SR super-resolution, SAM vessel segmentation, YOLO vessel detection, and environmental forecasting models. GPU acceleration is required for operational inference, while large-scale training and historical reprocessing are treated as HPC workloads. The requested resources support model loading, image tiling, data processing, and concurrent inference requests, while storage is used for cached models, temporary inputs, and generated outputs.*
[M5] IoT Sensors	*8*	*16*	*50*	*No*	*M5 supports the collection of IoT sensor data through a publish-subscribe interface, where devices upload locally collected information to the TERRA platform. The module manages continuous data streams that propagate throughout the TERRA pipeline and therefore requires sufficient memory resources to efficiently process incoming data and avoid bottlenecks.*
[M6] Data Fusion and data pre-processing	*16*	*32*	*200*	*No*	*M6 collects, validates, transforms, interpolates, and combines multi-dimensional Earth Observation and environmental data. The requested resources support intermediate datasets, processed outputs, and scalable inference workflows across additional AOIs and increased processing frequencies.*
[M7] Digital Twins	*14*	*28*	*150*	*2*	*M7 aggregates the requirements of the three TERRA Digital Twin use cases. The module orchestrates workflows, maintains DT state, supports scenario and what-if analysis, and provides visualization and validation services using outputs from the ML Models and Data Fusion modules. The requested resources support concurrent DT services, geospatial processing, forecast workflows, and parallel what-if scenarios, while storage is used for DT state, cached outputs, and metadata/model artifacts. GPU resources are optional and mainly support accelerated scenario execution and model-based inference.*
[M8] TERRA Interface	*2*	*4*	*10*	*No*	*M8 provides the TERRA web interface and visualization services. The requested resources support the minimum operational deployment and scalability requirements of the TERRA web application.*
Enabling Services	*16*	*32*	*100*	*No*	*For enabling services like Orchestrator, AAI, Accounting, Logging, Gateway API*
Total Requirements	***88***	***192***	***710***		
Object Storage	*-*	*-*	*510*	*No*	*Justification is provided in different section*

Block Storage	*-*	*-*	*200*	*No*	*Storage is needed to run the web application, databases, and processing services*

Justification for Object Storage

Object storage is needed to store large Copernicus datasets, satellite images, GeoTIFF files, model outputs, and other generated products in a scalable and cost-effective way.

Regarding Demonstrator 1: The deployment of Demonstrator 1, requires a minimum of 10 GB and should scale up to at least 100 GB of object storage. The minimum supports several years of typical inference runs for one AOI, while the maximum provides capacity for additional AOIs, more frequent runs, and longer retention of satellite images and plots. The hydrological pipeline requires storage for DEM rasters, derived hydrological rasters (e.g., flow accumulation), ERA5-Land forcing files, river graph products, daily segment-level hydrological features, diagnostic plots, and metadata. For one AOI and multi-year processing (2016-present), an allocation of 100 GB is needed.

Regarding Demonstrator 2: For Vessels anomaly detection and port pollution functionality TERRA requires 400 MB per month and historical data 400 x 12 x 11 = 52800 MB for job 4. For the super resolution and vessels detection functionality TERRA requires a minimum of approximately 10 GB, scaling to approximately 100–300 GB of object storage. VTG's persistent objects are the vessel-detection outputs: per-tile JSON detection records (bounding box, confidence, label, area — segmentation masks are not serialised, so these are only a few KB per tile) together with annotated visualisation rasters (~2–8 MB per 1024-px tile). The minimum (~10 GB) covers the detection vectors and visualisation overlays for a single AOI over typical inference runs, with the 4× super-resolved Sentinel-2 rasters treated as transient and discarded after processing. The maximum (~200 GB) provides capacity for retaining the super-resolved rasters (~3–6 MB per tile), for additional AOIs, multi-year history, and longer retention of detection imagery and plots.

Regarding Demonstrator 3: The Coastal Change and Prediction Digital Twin requires a minimum of approximately 20 GB, scaling up to approximately 50 GB of operational storage. This capacity is used to retain the latest data and intermediate processing outputs for the supported AOIs, including recent Sentinel-2 imagery, fixed geospatial reference data, transect geometries, vegetation-edge products, waterline outputs, temporary raster files, and forecast results. The minimum supports the current set of AOIs and routine Digital Twin execution, while the upper capacity allows for additional AOIs, larger satellite scenes, and temporary retention of intermediate processing products. These estimates refer only to the local operational storage required during Digital Twin execution and exclude long-term historical records stored separately in the TERRA database or object-storage infrastructure.

D. Technical Characteristics of the Code to Be Used

The TERRA platform is designed to process large amounts of Earth Observation data and update Digital Twin states continuously through repeated operational workflows. The code is not based on a single model, but on a combination of data ingestion, image processing, forecasting, and state-update pipelines that work together as one system. Most workflows follow a tile-based approach. Large Areas of Interest, such as river basins, coastal zones, and harbour regions, are divided into smaller spatial tiles so they can be processed independently and in parallel. This allows large-scale execution across many CPU and GPU nodes and makes the workflows suitable for HPC environments.

Regarding Deep Learning Frameworks, PyTorch ***(Paszke, et al., 2019)*** *and TensorFlow* ***(Abadi, et al., 2016)*** *are mainly used for training and running its AI models. For coastline monitoring, TERRA applies RobustUNet* ***(Choi, et al., 2021)*** *to detect the Vegetation Edge from Sentinel-2* ***(Sentinel-2, 2024)*** *images. For environmental forecasting, custom Bidirectional LSTM* ***(Zhou, 2024)*** *models with attention mechanisms* ***(Anonymous, 2025)*** *are used to predict water quality and other environmental indicators for the next 3–7 days. High-resolution analysis is improved through the Swin2SR* ***(Conde, Choi, Burchi, & Timofte, 2022)*** *super-resolution model. In harbour monitoring, YOLO* ***(Redmon, Divvala, Girshick, & Farhadi, 2016)****-based object detection models are used for vessel detection and maritime activity analysis. Hydrological preprocessing for Digital Elevation Model correction, depression filling, D8/DINF flow direction, flow accumulation, and river network extraction is conducted using tools such as WhiteboxTools* ***(WhiteboxTools: Hydrological analysis tools, 2024)****, rasterio* ***(Rasterio: Access to geospatial raster data, 2024)****, and QGIS12* ***(Graser, Sutton, & Bernasocchi, 2025)****-based workflows. The code to be used is developed in Python* ***(Python programming language, 2024)***

E. References

[1] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), (pp. 265–283). Retrieved from https://arxiv.org/abs/1605.08695

[2] Anonymous. (2025). Attention mechanisms in neural networks: A comprehensive mathematical treatment from theory to implementation. Retrieved from https://arxiv.org/abs/2601.03329

[3] Choi, S., Jung, S., Yun, H., Kim, J., Kim, S., & Choo, J. (2021). RobustNet: Improving domain generalization in urban-scene segmentation via instance selective whitening. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Retrieved from https://arxiv.org/abs/2103.15597

[4] Conde, M., Choi, U.-J., Burchi, M., & Timofte, R. (2022). Swin2SR: SwinV2 transformer for compressed image super-resolution and restoration. European Conference on Computer Vision (ECCV) Workshops. Retrieved from https://arxiv.org/abs/2209.11345

[5] Graser, A., Sutton, T., & Bernasocchi, M. (2025). The QGIS project: Spatial without compromise. 6(7), p. 101265. doi:10.1016/j.patter.2025.101265

[6] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., & Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. 32. Retrieved from https://arxiv.org/abs/1912.01703

[7] Python programming language. (2024). Retrieved from https://www.python.org/

[8] Rasterio: Access to geospatial raster data. (2024). Retrieved from https://rasterio.readthedocs.io/en/stable/

[9] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (pp. 779–788). Retrieved from https://arxiv.org/abs/1506.02640

[10] Sentinel-2. (2024). Retrieved from https://sentiwiki.copernicus.eu/web/sentinel-2

[11] WhiteboxGeo geospatial software. (2024). Retrieved from https://www.whiteboxgeo.com/geospatial-software/

[12] WhiteboxTools: Hydrological analysis tools. (2024). Retrieved from https://www.whiteboxgeo.com/manual/wbt_book/available_tools/hydrological_analysis.html

[13] Zhou, J. (2024). Bi-directional LSTM-GRU based time series forecasting approach. 3(2), pp. 222–231. doi:10.62051/ijcsit.v3n2.26