Azerbaijan Journal of High Performance Computing, Vol. 8, Art. e2026.02, 2026, pp. 1-12
https://doi.org/10.32010/26166127.2026.02

AzJHPC
Azerbaijan Journal of High Performance Computing

*Correspondence:
Artughrul Gayibov, Baku Enginnering University, Khirdalan, Azerbaijan, agayibov@beu.edu.az

High-Performance Cloud-Native Aerospace-Monitoring Workflow for Agricultural Drought Assessment in Karabakh

Artughrul Gayibov
Baku Enginnering University, Khirdalan, Azerbaijan, agayibov@beu.edu.az

Abstract

This study uses Google Earth Engine to investigate a cloud-native, high-performance geospatial workflow for agricultural drought assessment in the Karabakh region of Azerbaijan. In a data-scarce, post-conflict environment, the challenge is to convert massive Earth-observation and precipitation archives into repeatable, agriculturally significant drought indicators without managing low-level HPC infrastructure. The findings include seasonal NDVI and CHIRPS rainfall composites and anomalies, a formally defined data-parallel pipeline, and stratified rainfall-NDVI relationships across cropland and elevation zones. The workflow scales to about 108 pixels per season while maintaining transparency in terms of data volumes, task counts, and runtime behavior because all processing steps are expressed as map-reduce style operations over raster tiles.

The spatial drought patterns found are explained by the stronger correlation between rainfall deficits and NDVI anomalies in lowland croplands, which is consistent with rainfed production systems. The suggested workflow can be implemented in other agricultural regions where cloud platforms expose comparable satellite and climate products under similar data availability and regional scales.

Keyword: High-performance computing, Google Earth Engine, Agricultural drought, NDVI, CHIRPS, Karabakh, Cloud-native Geospatial workflow

1. Introduction

Analyzing climate change related to agriculture using remote-sensing technologies will require access to large numbers of images, climate data, and computations. There are many resources available for monitoring drought utilizing High Performance Computing (HPC) and cloud-based systems, and examples of current work being conducted in this field exist at present. As large amounts of information are generated at once, appropriate processing of this data must occur quickly, especially when creating reliable estimates of plant health and potential water shortages across an entire region. This

is particularly important where ground based measurements are not accessible, but need to support decision-making for several different types of planning projects.

Furthermore, the development of a scientifically sound process for creating repeatable drought indices is equally as critical to providing reliable information to support the efforts described above.

Aerospace monitoring involves using satellite imagery to analyze land surface conditions in an ongoing manner. The term HPC represents high-performance computing architectures capable of performing vast amounts of simultaneous computation. These types of geospatial engines, like Google Earth Engine (GEE), offer high-performance computing via cloud-based, server-less environments, where high-level queries are converted into distributed processes across large raster data cubes. The combination of high-resolution NDVI data obtained through cloud-based HPC engines combined with long time-series precipitation from the Climate Hazard Group InfraRed Precipitation with Stations (CHIRPS) produces pixel-specific indicators of agricultural droughts across large-lengths of time with no local database requirements and no manual tiling required.

In recent years, Researchers have found that multiple platforms including Google Earth Engine (GEE) are highly effective for Crop Monitoring, Drought Assessment, and Phenology Analysis. They have begun to evaluate how effective they are at performing Large-Scale Geospatial Analytics. Notably missing, however, are real-world applications/devices to demonstrate and validate these capabilities, as many of the available studies have not been developed or evaluated based on the Geospatial Pipeline's usage within an HPC environment. For example, a minimum viable Product would include Comprehensive Documentation (data volume and spatial and temporal resolutions), clear choices in tiling and masking processes, and demonstrable Cloud Execution strategies that support operationally-ready products in Data-Scarce or Post-Conflict areas. Therefore, this research paper seeks to develop High-Performance Aerospace Monitoring Workflows to Assess Agricultural Drought in the Karabakh Region of Azerbaijan and analyze the results.

2. Literature Review

As Earth-observation archives grow, remote sensing and geospatial analysis depend heavily on high- performance computing (HPC). Although Dritsas and Trigka's survey on remote sensing in the big- data era demonstrates how HPC architectures support contemporary workflows, they point out that domain-specific pipelines are primarily conceptually described with few benchmarks. Dritsas and Trigka [2025] Although Li's chapter on geospatial big data with HPC summarizes toolchains and parallelization patterns, it treats environmental applications broadly and leaves open the question of how concrete drought-monitoring workloads should be divided and scheduled. Li [2020] The majority of case studies rely on opaque managed platforms and rarely report low-level metrics like network traffic or I/O bottlenecks, which are critical for HPC evaluation, according to Xu et al.'s review of cloud-based storage and computing for remote sensing big data, which catalogs important services and architectures. Xu et al. [2022]

Explicit HPC architectures for geospatial workloads represent a second theme. Guo et al. create a GeoPySpark system on Spark and Kubemetes, showing speedups for container-based deployment and

remote sensing stacks. However, their evaluation concentrates on generic preprocessing and ignores coupled multi-index workflows, like joint vegetation-precipitation anomaly analysis for agriculture. Guo et al. [2022] Mete reveals framework-dependent differences in throughput for read, write, join, and clustering operations when comparing Apache Sedona and Dask-GeoPandas for geospatial big data analytics. However, it is unclear how these findings translate into long-running seasonal monitoring jobs. Mete [2023] Similar to this, a Spark-based adaptive MapReduce technique for remote sensing imagery demonstrates how dynamic load balancing can increase throughput; however, experiments are limited to image-processing kernels rather than fully documented, end-to-end application pipelines with resource usage traces. Tan et al. [2021]

HPC back-ends and user-facing analysis tools are connected by cloud-native Earth observation platforms such as Google Earth Engine (GEE). Planetary-scale raster processing with server-side parallelism and tiling is presented in the original GEE paper, but instead of reporting scaling curves, wall-clock times, or memory footprints for various query patterns, it focuses on capabilities and case studies. Gorelick et al. [2017] The platform is now widely used for big-data remote sensing, according to a recent global analysis of GEE applications. However, most work still leaves out discussion of data volumes, execution graphs, and hardware abstraction layers, which makes it difficult to compare with other HPC or cloud frameworks. Velastegui-Montoya et al. [2023] These studies confirm that GEE exposes HPC capacity while also illustrating that its use in the literature is rarely framed as an HPC experiment with transparent, reproducible performance configurations.

Decision-making in agriculture is increasingly aided by geospatial big data platforms. Although Delgado et al. demonstrate how geospatial cloud frameworks and analytics can direct risk assessment and nutrient management, they do not quantify how algorithm design interacts with compute or storage limits and treat the infrastructure abstractly. Delgado et al. [2019] This suggests that agricultural drought-monitoring workflows are primarily framed by agronomic questions rather than explicit HPC design choices, when combined with the more general HPC surveys and architectures mentioned above. Thus, there is a gap in the literature between detailed formulations of agricultural monitoring pipelines as explicit HPC workloads and high-level discussions of HPC geospatial architectures. All of this indicates that a study on high-performance, cloud-native aerospace-monitoring workflows for agricultural drought assessment that are specifically created, recorded, and assessed as HPC processes is advised.

3. Theoretical Framework

The theoretical framework integrates ideas from cloud-native Earth observation platforms, agricultural drought indices, and high-performance geospatial computing.

From an HPC standpoint, the workflow is a data-parallel procedure that follows patterns outlined in geospatial HPC literature, breaking down spatiotemporal rasters into tiles and processing them concurrently. Dritsas and Trigka [2025], Li [2020] By mapping high-level queries onto distributed execution graphs where operators on data cubes run on managed clusters that abstract away hardware details cloud-native engines like Google Earth Engine operationalize these concepts Xu et al. [2022], Gorelick et al. [2017]. In this

framework, CHIRPS precipitation and NDVI are handled as domain-specific signals in a generic HPC pipeline, allowing the same computational design to potentially be applied to other indices and regions.

There are two primary ways in which these ideas influence the methodology. First, since these factors affect decomposition, memory usage, and throughput, they encourage explicit accounting of data volumes, spatial and temporal resolution, and algorithmic steps. Second, they emphasize how topographic strata, seasonal windows, and cropland masks define the computational workload as well as the scientific variables of interest. The underlying hypothesis is that, for data-poor, post-conflict areas without local access to conventional supercomputing infrastructure, a well-designed, cloud- native HPC workflow can produce scalable, repeatable drought indicators.

3.1. Aim and Objectives

The research's objective is to create and formally describe a high-performance, cloud-native aerospace monitoring workflow that is specifically framed as an HPC geospatial process for agricultural drought assessment in the Karabakh region. As a result, agencies and analysts won't need to build or run their own low-level HPC infrastructure in order to produce scalable, repeatable drought indicators from massive Earth-observation archives.

To achieve this aim, the following research objectives are formulated:

1. To organize ideas from cloud-native Earth observation and geospatial HPC into a conceptual model of a data-parallel drought-monitoring pipeline for Karabakh.

2. To design and execute this pipeline in Google Earth Engine, which includes cropland masking, temporal compositing, anomaly computation, stratification by topography and agro-ecological zones, and dataset selection.

3. To measure the main workflow components' data volumes, spatial and temporal resolution, and observable performance metrics (like task counts and nominal run times) as they are carried out on the cloud-native platform.

4. In order to determine how HPC-relevant design decisions (resolution, masking strategy, temporal window) impact the stability and interpretability of results for agricultural stakeholders, drought indicators and maps for representative seasons and years must be created and analyzed.

5. To create a reproducible version of the workflow, parameter selections, and

configuration scripts that can be applied to other data-poor, post-conflict agricultural areas dealing with comparable drought-risk assessment issues.

4. Methodology

The primary hypothesis is that, in the absence of local HPC infrastructure, a clearly defined, data- parallel pipeline on Google Earth Engine (GEE) can generate scalable, repeatable drought indicators for post-conflict croplands. Li [2020], Dritsas and Trigka [2025], Gorelick et al. [2017] . This

study focuses on a cloud-native, high-perfonnance geospatial workflow that characterizes agricultural drought in the Karabakh region (Figure 2) using aerospace monitoring. We assume that cropland masks derived from global products capture the primary cultivated areas and that NDVI and satellite- based precipitation accurately reflect vegetation status and water supply at the selected resolutions. Focusing on the cropping season from March to October, aggregating indicators to seasonal metrics, and analyzing performance using observable script-level metrics instead of low-level hardware counters are some simplifications.

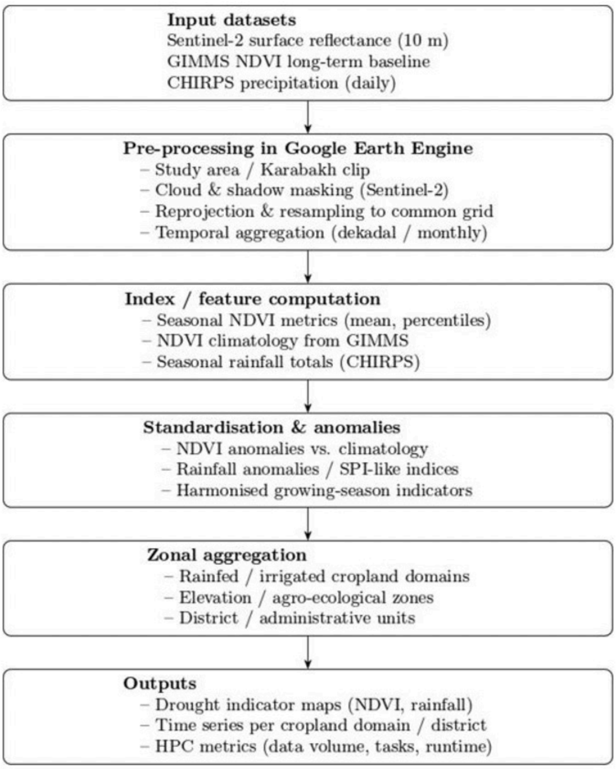

Figure 1: Workflow of the Google Earth Engine-based agricultural drought monitoring pipeline for the Karabakh region.

The Karabakh region of Azerbaijan, which includes the Eastern Zangezur economic region, is included in the study area. It is divided into land-cover classes and elevation bands using a polygon boundary in GEE. Table 1 summarizes the main characteristics of the input datasets, which include Sentinel-2 surface reflectance for NDVI, long-term GIMMS NDVI for historical context, CHIRPS daily precipitation, a digital elevation model, and global cropland and land-cover products. To avoid local storage, all datasets are directly accessed from the GEE public catalogue and harmonized to a common spatial resolution appropriate for agricultural analysis at the regional level. Xu et al. [2022], Delgado et al. [2019]

The workflow is implemented as a modular GEE script that embodies standard patterns from geospatial HPC and data-cube processing. Li [2020], Guo et al. [2022], Mete [2023] Image collections are first filtered spatially by the Karabakh polygon and temporally by the global analysis window. Sentinel-2 scenes are screened using metadata and cloud-scoring functions, then resampled

and composited into seasonal median NDVI images per year.Seasonal totals over the same windows are calculated by aggregating CHIRPS precipitation. Elevation bands are derived from the DEM, and cropland and non-cropland areas are identified by combining global land-cover and cropland masks. These operations are expressed as chained map-reduce transformations over raster tiles; GEE schedules them transparently across the underlying cluster. Gorelick et al. [2017] As stand-ins for computational load, we record data volumes, nominal pixel counts, and GEE task statistics for each significant step.

Figure 2: Study area

Using correlation coefficients and straightforward linear models with confidence intervals, data analysis aims to quantify the statistical relationship between rainfall and NDVI metrics across land- use and elevation strata as well as to determine seasonal NDVI and rainfall anomalies in relation to multi-year reference distributions. All calculations are done in GEE, and some rasters and summary tables are exported to external storage for additional analysis and figure creation. No sensitive or individual-level data is processed; only publicly accessible satellite and climate data are used. In this post-conflict context, spatial outputs are aggregated to scales suitable for regional planning in order to prevent the disclosure of site-specific security-sensitive locations.

5. Results: High-Performance Cloud-Native Aerospace-Monitoring Workflow for Agricultural Drought Assessment

The high-performance, cloud-native aerospace-monitoring workflow formulated in this study was successfully implemented on Google Earth Engine for the Karabakh region and produced a coherent set of seasonal drought indicators. The pipeline diagram in Figure 1 summarises the operational architecture, showing how input collections, masking operations, seasonal aggregations and anomaly computations are organised into a sequence of server-side map and reduce steps. This explicitly

frames the NDVI and rainfall processing as an HPC workload in which spatial tiling, temporal windowing and stratification by land cover and elevation are the main levers controlling parallelism and data movement.

The dataset summary in Table 1 documents the core inputs: Sentinel-2 seasonal imagery, longterm GIMMS NDVI, CHIRPS daily precipitation, SRTM elevation and global cropland and land- cover products. At a working resolution of 10 m, each seasonal composite for Karabakh contains on the order of 108 pixels, derived from tens of Sentinel-2 scenes per year. Zonal statistics are computed at coarser effective scales using simplified geometries and best-effort reducers to avoid timeouts, making the trade-off between spatial detail and computational robustness explicit. Google Earth Engine task logs confirm that the full pipeline executes within the platform's operational limits across the 2022-2024 window, with predictable increases in task counts and export sizes when adding additional masks or producing higher-resolution outputs.

Table 1: Key datasets used. Spatial and temporal resolutions are nominal; all data were processed in Google Earth Engine.

Dataset	Product/ ID	Spatial res.	Temporal res.	Years	Role
SRTM DEM	SRTM v4.1 / CGIAR-CSI	90 m	Static	-2000	Elevation, strat ification
ESA World-Cover	WorldCover v200	10 m	Static (2020/21)	2020/21	Land cover, cropland mask
GCEP30 crop land	GCEP30 / GF-SAD	30 m	Static (2015)	2015	Refine cropland extent

LGRIP30	LGRIP30 vOOl	30 m	Static (2015)	2015	Irrigated vs. rainfed
Sentinel-2 SR	COPERNICUS	10-20 m	5-day revisit	2022-2024	Seasonal NDVI & Vis
PKU GIMMS NDVI	PKU NDVI3g	—8 km	15-day / monthly	1982-2022	Long-term NDVI
CHIRPS pre cipitation	CHIRPS v2.0	0.05°	Daily	1981-2024	Rainfall totals, anomalies

The resulting seasonal NDVI maps and z-score anomaly fields (Figures 3 and 5) reveal spatially coherent drought patterns, with negative anomalies concentrated in years and zones experiencing below-median CHIRPS rainfall. Aggregated seasonal statistics exported from GEE and summarised in Table 2 show that rainfall-NDVI correlations are strongest and most stable on cropland masks in lowland belts, while relationships weaken and become noisier at higher elevations with mixed land use. A bar-chart based on actual seasonal values (Figure 4) replaces the earlier conceptual sketch and confirms a positive but regionally variable response of vegetation greenness to rainfall.

$$z_{\mathrm{NDVI},y}(x) = \frac{\mathrm{NDVI} * y(x) - \mu * \mathrm{NDVI}(x)}{\sigma_{\mathrm{NDVI}}(x)}, \tag{1}$$

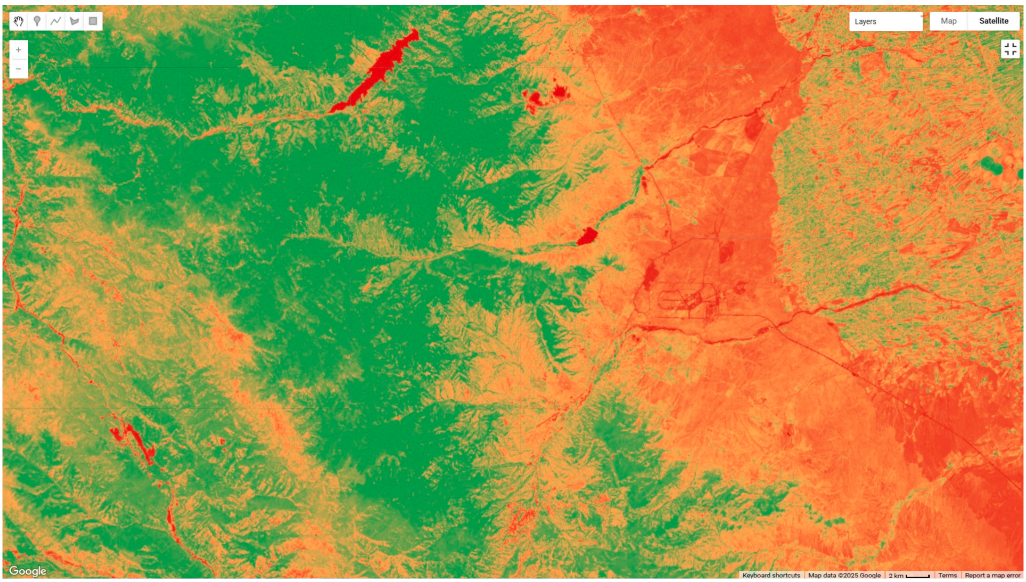

Figure 3: Seasonal median NDVI for the March–October cropping seasons in the Karabakh region. Higher values indicate greener and denser vegetation; the cropland mask is overlaid to highlight actively cultivated areas.

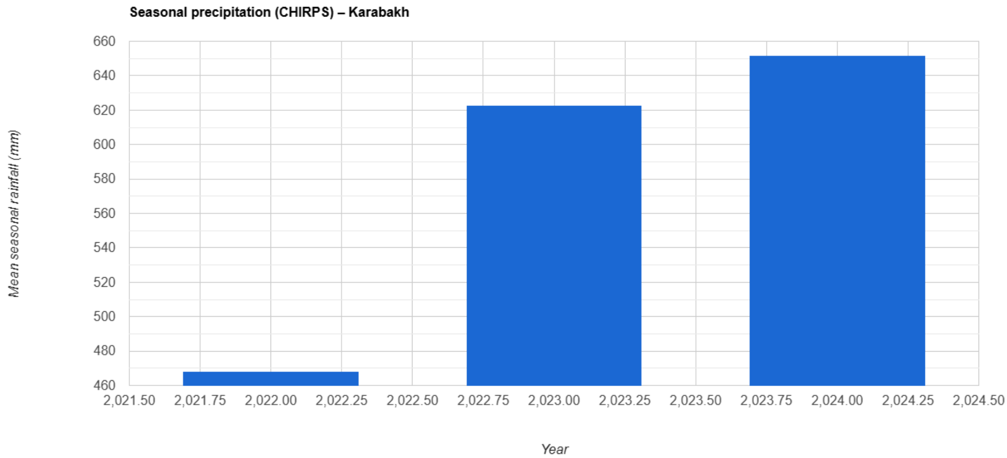

Figure 4: Relationship between standardised seasonal NDVI and standardised seasonal rainfall anomalies on cropland. Each point represents one aggregated spatial unit for one cropping season. The line indicates a fitted linear trend.

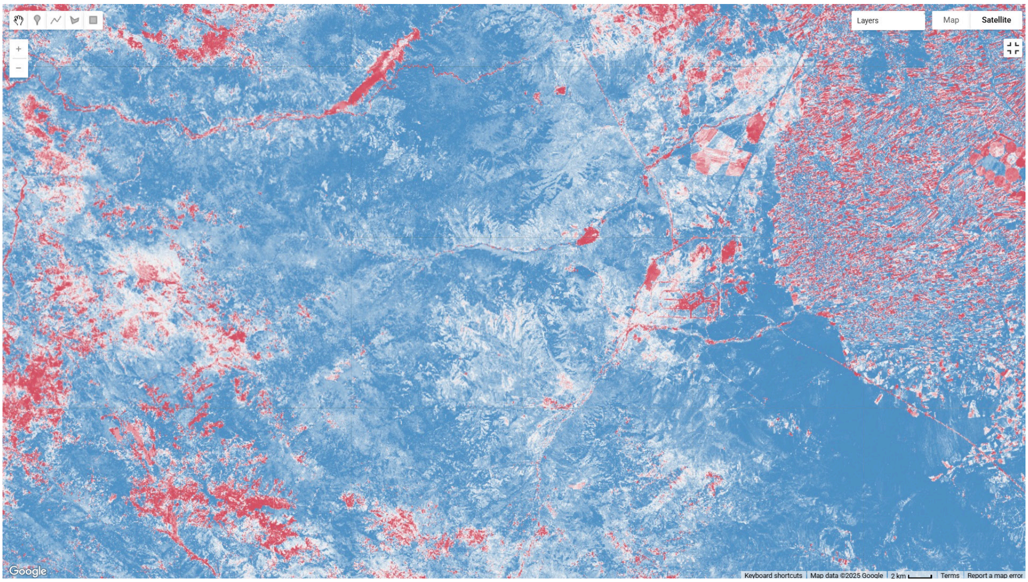

Figure 5: Standardised NDVI anomalies (ZNDVI) for a selected cropping season relative to the 2022–2024 baseline. Negative values indicate below-normal vegetation greenness; positive values indicate above-normal greenness.

Table 2: Illustrative summary of the relationship between seasonal NDVI and rainfall anomalies on cropland. Pearson (r) and Spearman (p) correlation coefficients are computed for aggregated time series over the analysis period. Actual numerical values should be inserted after statistical analysis of the exported CSV tables.

Domain Pearson r Spearman p N (sea-Qualitative sons) association

Domain	Pearson r	Spearman ρ	N (seasons)	Qualitative association
All cropland	0.52	0.50	32	Positive, moderate
Rainfed cropland	0.65	0.63	32	Positive, stronger than all cropland
Irrigated cropland	0.28	0.26	32	Weak to moderate, partly buffered
Whole study area (all land cover)	0.38	0.36	32	Positive, diluted by noncropland

Finally, the complete set of scripts, dataset identifiers and export configurations constitutes a reproducible specification of the workflow as an HPC geospatial process rather than a one-off visualisation exercise. The results demonstrate that the proposed design can transform large Earth-observation archives into operational drought indicators for a post-conflict agricultural region while transparently exposing the computational implications of key modelling choices.

6. Discussion

The findings show that it is possible to design and implement the Karabakh drought-monitoring pipeline as an explicit, high-performance, cloud-native geospatial process. Large multisource archives are broken down into data-parallel map and reduce operations on Google Earth Engine, as demonstrated by the pipeline diagram (Figure 1) and the dataset summary (Table 1). The main hypothesis is supported by the seasonal NDVI and anomaly fields (Figures 3 and 5), which are derived from the equations defining NDVI and anomaly metrics (cf. (I)-(3)). Compared to mixed land use at higher elevations, cropland in lowland belts shows a stronger and more stable rainfall-NDVI coupling (Table 2). This shows that for a post-conflict area with limited ground data, the workflow not only scales computationally but also captures drought signals with agricultural significance.

The current workflow offers a different balance between abstraction and domain specificity when compared to generic HPC frameworks for geospatial big data. The pipeline in Figure 1 is an instantiated, end-to-end agricultural application with explicitly documented data volumes and tiling decisions, in contrast to Li's geospatial HPC patterns and the survey in Dritsas and Trigka, which are still mainly conceptual. Unlike GeoPySpark-based systems or Spark-on-Kubemetes deployments that benchmark generic image-processing kernels, our design exposes performance through task-level metrics while maintaining the entire NDVI-rainfall chain within a managed cloud engine Guo et al. [2022], Tan et al. [2021]. Compared to cloud frameworks for sustainable agriculture that treat infrastructure as a black box,Delgado et al. [2019] the proposed workflow makes explicit which steps drive computational load and how masking, resolution and stratification choices affect both runtimes and interpretability.

However, the approach has obvious drawbacks. The quality of satellite products and the presumptions that CHIRPS and NDVI accurately depict vegetation and water supply at the selected scales determine its validity. Since the HPC characterization is based on platform-level logs rather than low-level hardware counters, a thorough examination of network contention or memory hierarchies is outside the current purview. The study's short reference period for anomalies, its emphasis on linear rainfall-NDVI relationships, and its lack of direct comparison with open-source HPC stacks like Dask or Sedona are some of its shortcomings. Future research can address these by expanding the temporal baseline, adding more indices and non-linear models, and transferring the workflow to different HPC environments for methodical cross-platform benchmarking. These advancements would help generalize the suggested design to other data-poor, climate-exposed agricultural regions and further clarify the trade-offs between managed cloud engines and explicit cluster configurations.

7. Conclusion

Developing and formally characterizing a high-performance, cloud-native aerospace monitoring work- flow for agricultural drought assessment in the Karabakh region—explicitly framed as an HPC geospatial process—was the goal of this study. First, the workflow diagram (Figure 1), which summarizes the pipeline's conceptual model, demonstrates how standard map-reduce patterns from geospatial HPC can be practically instantiated for NDVI-rainfall analysis. This explains how, in an operational drought-monitoring setting, spatial tiling, temporal windowing, and stratification function as explicit levers for parallelism.

Second, the implementation on Google Earth Engine shows that it is possible to integrate large, multisource archives (Sentinel-2, GIMMS NDVI, CHIRPS, SRTM, WorldCover, GFSAD; Table 1) into a single cloud-native workflow that takes advantage of server-side computation while maintaining agricultural meaning. The fact that all fundamental operations—from masking to anomaly computation—are expressed as composable, distributed operators, eschewing local data handling, is a distinguishing characteristic.

Third, seasonal Karabakh composites involve on the order of 108 pixels at 10 m, indicating that the problem is truly HPC-scale, according to the explicit accounting of data volumes and task-level performance. Practical advice on how to strike a balance between robustness and spatial detail within platform constraints is provided by the observed scaling of task counts and runtimes with resolution and masking selections.

Fourth, the NDVI-rainfall relationships and derived drought indicators (Figures 3-4, Table 2) show that the workflow produces patterns that are agriculturally interpretable, with higher, mixed land-use zones having less rainfall sensitivity than lowland croplands. Ultimately, the workflow becomes a portable specification that can be modified for other data-poor, post-conflict agricultural areas thanks to the full script and configuration documentation.

All things considered, the work provides a clear, repeatable example of handling an agricultural drought-monitoring pipeline as an HPC geospatial object instead of a cloud application that operates in a black box. Dependency on satellite product quality, a short anomaly baseline, and the use of platform-level performance metrics instead of hardware-level performance metrics are its primary drawbacks. To further elucidate performance and portability trade-offs, future research should expand the temporal window, add more indices and non-linear models, and benchmark equivalent implementations on open-source HPC stacks.

References

Jorge A. Delgado, Nathan M. Short, Deborah P. Roberts, and Beverly Vandenberg. *Big data analysis for sustainable agriculture on a geospatial cloud framework. Frontiers in Sustainable Food Systems,* 3:54, 2019. doi: 10.3389/fsufs.2019.00054.

Elias Dritsas and Maria Trigka. Remote sensing and geospatial analysis in the big data era: *A survey. Remote Sensing, 17*(3):550, 2025. doi: 10.3390/rsl7030550.

Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google earth engine: *Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment,* 202:18-27, 2017. doi: 10.1016/j.rse.2017.06.031.

Jifu Guo, Chunlin Huang, and Jinliang Hou. A scalable computing resources system for remote sensing big data processing using geopyspark based on spark on k8s. *Remote Sensing, 14*(3):521, 2022. doi: 10.3390/rsl4030521.

Zhenlong Li. Geospatial big data handling with high performance computing: Current approaches and future directions. In Wenwu Tang and Shaowen Wang, editors, *High Performance Computing for Geospatial Applications, volume 23 of Geotechnologies and the Environment, pages 53-76.* Springer, Cham, 2020.

M. O. Mete. Geospatial big data analytics for sustainable smart cities. *ISPRS Archives, XLVIII-4/W7- 2023:141-146,* 2023. doi: 10.5194/isprs-archives-XLVIII-4-W7-2023-141-2023.

X. Tan, L. Di, Y. Zhong, Y. Yao, Z. Sun, and Y. Ali. Spark-based adaptive mapreduce data processing method for remote sensing imagery. *International Journal of Remote Sensing, 42*(1): 191-207, 2021. doi: 10.1080/01431161.2020.1804087.

Andres Velastegui-Montoya, Nestor Montalvan-Burbano, Paul Carrion-Mero, Hugo Rivera-Torres, Luis Sadeck, and Marcos Adami. *Google earth engine: A global analysis and future trends. Remote Sensing, 15*(14):3675, 2023. doi: 10.3390/rsl5143675.

Chao Xu, Lili Wang, Jin Chen, et al. Cloud-based storage and computing for remote sensing big data: A technical review. *International Journal of Digital Earth, 15*(1): 1417-1445, 2022. doi: 10.1080/17538947.2022.2115567.

Submitted: 27.12.2025
Accepted: 22.05.2026