# Development and Validation of a Hybrid Remote Sensing and Machine Learning Framework for High-Resolution Terrain Ruggedness (K-Coefficient) Mapping Across Heterogeneous Geomorphological Regions

| | | |
|---|---|---|
| شماره کتابشناسی ملی | : | ۱۰۵۳۸۹۴۱ |
| عنوان و نام پدیدآور | : | Development and Validation of a Hybrid Remote Sensing and Machine Learning Framework for High-Resolution Terrain Ruggedness (K-Coefficient) Mapping Across Heterogeneous ; Research Institute for Geomorphological Regions [et al.]. Earth Sciences.; with cooperation UNESCO Chair on Coastal Geo-Hazard Analysis. |
| شابک | : | ۹۷۸-٦۲۲-۱۵۸۲-۰۷-۵ |
| وضعیت اجرای طرح | : | تهران: نشر خزه، ۱۴۰۵. |
| مشخصات ظاهری | : | ۹٤ ص.: مصور، جدول، نقشه. |
| یادداشت | : | زبان: انگلیسی.<br>کد طرح ۱۰٦-ط-پ-۰۴. |
| یادداشت | : | Nazari Hamid, Behrad AkbarPour, Golchehreh Farahani, Jalal Karami, Saeed Arefipour, Ramin Abbasi. |
| یادداشت | : | کتابنامه: ص. ٦۹-۷۰. |
| توصیفگر | : | تعیین ضریب (K) حفر در تفکیک طیفی زمین های پوشیده (زمین های دستکاری شده): مطالعه موردی برنامه SUTGMM<br>Server-Based Unified Thematic Geological Mappingلاتین |
| توصیفگر | : | پیش‌بینی ژئوفیزیکی<br>Geophysical prediction<br>فرسایش‌پذیری<br>erodibility<br>ناهمواری‌های زمین<br>Landforms<br>تحقیق کمی<br>Quantitative research |
| توصیفگر | : | زاگرس، منطقه<br>zagros Mountains Region<br>دشت مغان<br>Moghan Plain (Iran) |
| دسترسی و محل الکترونیکی | : | ۳۸۹f۲d۳۳۱fb۲-۸e۸۵-٤۵۳٤-۲۳۵-a۰e٤bf۱۲a<br>a۱۱٦۱۷۳۵۵٦۹c-f۳۰-a۳e٤۱a-۳٤۵-۹۹۰۱۲۱۳۵ |

# Development and Validation of a Hybrid Remote Sensing and Machine Learning Framework for High-Resolution Terrain Ruggedness (K-Coefficient) Mapping Across Heterogeneous Geomorphological Regions

**Authors:**

**Hamid Nazari**
**Behrad AkbarPour, Golchehreh Farahani, Jalal Karami, Saeed Arefipour, Ramin Abbasi**

## اطلاعات گزارش

**عنوان:** تعیین ضریب (K) حفر در تفکیک طیفی زمین های پوشیده (زمین های دستکاری شده): مطالعه موردی برنامه_SUTGMM

**مجری:** پژوهشکده علوم زمین

**زبان مرجع:** انگلیسی

**خروجی:** تحلیل و مقاله

**نویسندگان:** حمید نظری، بهراد اکبرپور، گلچهره فراهانی، جلال کرمی، سعید عارفی‌پور، رامین عباسی

**رئیس کرسی یونسکو در مخاطرات زمین شناختی ساحلی:** حمید نظری

**مسئول شورای اجرایی:** مرتضی طالبیان

**ناشر:** نشر خزه

با همکاری کرسی یونسکو در مخاطرات زمین شناختی ساحلی

**چاپ اول:** ۱۴۰۴

**شمارگان:** ۵۰ نسخه

**صفحات:** ۹۴

**شابک:** ۹۷۸-۶۲۲-۱۵۸۲-۰۷-۵

khazepub@gmail.com

**Report Information**

**Title:** Development and Validation of a Hybrid Remote Sensing and Machine Learning Framework for High-Resolution Terrain Ruggedness (K-Coefficient) Mapping Across Heterogeneous Geomorphological Regions

**Employer:** Research Institute for Earth Sciences

**Original language:** English

**Output:** Report, Scientific paper

**Authors:** Hamid Nazari, Behrad AkbarPour, Golchehreh Farahani, Jalal Karami, Saeed Arefipour, Ramin Abbasi

**Chairholder in the UNESCO Chair on Coastal Geo-Hazard Analysis:** Hamid Nazari

**Head of the Executive Council:** Morteza Talebian

**Publisher:** Khazeh Publication

with cooperation UNESCO Chair on Coastal Geo-Hazard Analysis

**First Edition:** 2026

**Edition number:** 50

**Page:** 94

**Shabak:** 978-622-1582-07-5

khazepub@gmail.com

**Scientific Council**

| Name | Affiliation |
|---|---|
| Iraj Abdollahie Fard | Exploration Directorate of NIOC |
| Behrouz Abtahi | Shahid Beheshti University (SBU) |
| Philippe Agard | University of Sorbonne |
| Hamid Alizadeh Lahijani | Iranian National Institute for Oceanography and Atmospheric Science |
| Ryo Anma | Tokushima University |
| Constantin D. Athanassas | School of Mining and Metallurgical Engineering, National Technical University of Athens (NTUA), Greece |
| Franck A. Audemard | Department of Geology, Central University of Venezuela |
| Alice Aurelie | UNESCO Water Sciences Division |
| Ara Avagyan | IGS: Institute Geological Sciences |
| Yeong Bae Seong | Korea University |
| Rick J Bailey | IOC-UNESCO Indian Ocean Tsunami Warning and Mitigation System/ UNESCO |
| Abbas Banj Shafiei | Urmia University |
| Eric Barrier | University of Sorbonne |
| Jean Chery | University of Montpellier |

| | |
|---|---|
| Yahya Djamour | Shahid Beheshti University (SBU) |
| Dariush Esmaili | University of Tehran |
| Issa El-Hussain | Sultan Qaboos University |
| Aram Fathian Baneh | University of Calgary |
| Hassan Fazeli Nashli | University of Tehran |
| Magdi Guirguis | Institut français d'archéologie orientale du Caire |
| Ekkehard Holzbecher | German University of Technology in Oman |
| Egor Krasinskiy | Underwater research center Russian Geographical Society |
| Razyeh Lak | Research Institute for Earth Sciences |
| John Lambert | Deltares, UNESCO |
| Mohammad Mokhtari | International Institute of Earthquake Engineering and Seismology |
| Hamid Nazari | Research Institute for Earth Sciences |
| Jafar Omrani | CGMW- Geological Survey of Iran |
| Klaus Reicherter | Aachen University |
| Jean-François Ritz | University of Montpellier |
| Stefano Salvi | National Institute of Geophysics and Volcanology (INGV) |
| Morteza Talebian | Research Institute for Earth Sciences |

| | |
|---|---|
| Mohammad Tatar | International Institute of Earthquake Engineering and Seismology |
| Katerina Theodorakopoulou | Hellenic Open University-School of Applied Arts and Sustainable Design, Greece |
| Judith Thomalsky | German Archaeological Institute Tehran Branch |
| Richard Walker | University of Oxford |
| Wenjiao Xiao | Chinese Academy of Sciences |
| Alireza Zarasvandi | Iranian Research Organization for Science & Technology (IROST) – Ministry of Science, Research & Technology |
| Mahdi Zare | International Institute of Earthquake Engineering and Seismology |

**Executive Council-Abroad**

| **Name** | **Affiliation** |
|---|---|
| Nasir Ahmadi | Environmental Protection Organization of Mazandaran Province |
| Arash Amini | Golestan University |
| Mir Ali Asghar Mokhtari | University of Zanjan |
| Alireza Amrikazemi | Scientific Coordinator, Qeshm Island UNESCO Global Geopark |

| | |
|---|---|
| Somayeh Arab | University of Tehran |
| Parviz Armani | Imam Khomeini International University |
| Asghar Dolati | Kharazmi University |
| Fahimeh Foroghi | Iranian National Institute for Oceanography and Atmospheric Science |
| Abdolazaim Ghanghormeh | Golestan University |
| Habibolah Ghasemi | Shahrood University of Technology |
| Ataollah Kavian | Environmental Protection Organization of Mazandaran Province |
| Hossein Kouhestani | University of Zanjan |
| Mahdi Moghimi | University of Zanjan |
| Ali Akbar Momeni | Shahrood University of Technology |
| Babak Moradi | Iranian National Institute for Oceanography and Atmospheric Science |
| Seyed Mohsen Mortazavi | Hormozgan University |
| Hasan Nasrollah Zadeh | Caspian Sea Ecological Research Center |
| Ehsan Pegah | Kharazmi University |
| Abdolwahed Pehpouri | Qeshm Island UNESCO Global Geopark |

| | |
|---|---|
| Ahmadreza Rabani | University of Science and Technology of Mazandaran |
| Mahdi Rahmanian | Shargh Daily newspaper |
| Masoud Sadri Nasab | University of Tehran |
| Mojtaba Safari | University of Mazandaran |
| Mohammad Tatar | International Institute of Earthquake Engineering and Seismology |
| Mojtaba Yamani | University of Tehran |

**Executive Council-RIES and GSI (Board of Directors)**

| **Name** | **Affiliation** |
|---|---|
| Seyed Mahmoud Fatemi | Deputy Minister and General Director |
| Morteza Talebian | Director in RIES and President of the Executive Board |
| Mahdi Badpa | Vice President |
| Mahmoudreza Majidifard | Deputy in Research |
| Majid Pourkerman | AI-GEOARCH Specialist |
| Mehdi Basavand | Head of Youth Section |
| Zaynab Kazeminia | Head of the Secretariat |

**Executive members (RIES):**

| | |
|---|---|
| Hourieh Alibeygi | Secretary |
| Keivan Ajdari | Educational expert |
| Afshin Akbarpour | Academic Advisor |
| Mohammad Reza Ghasemi | Academic Advisor |

| | |
|---|---|
| Elaheh Ghayoumi | Web Designer |
| Hamoon Memarian | IT and Information |
| Mehrnoosh Poursaeid | Graphist |
| Manouchehr Qorashi | Advisor |
| Shirin Safavi | Communication Affair |
| Majid Taherverdi | Financial Affair |
| Alireza Vaezi | Academic Advisor |

**Executive members (GSI):**

| | |
|---|---|
| Saeed Arefipour | Research Expert |
| Hanieh Bakhshaei | International Affair |
| Reza Behbahani | Research Expert |
| Javad Darvishi khatooni | Research Expert |
| Mohammadreza Ensani | Research Expert |
| Marziyeh Estrabi | CGMW |
| Narges Ghamami | Executive Director of the Geosciences Quarterly |
| Gholamreza Hoseinyar | Research Expert |
| Somayeh Habibi | Research Expert |
| Mojtaba Kavianpour | Research Expert |
| Maryam Miraftab | Public Relation Affair |
| Elham Moeini | Public Relation Affair |
| Mohammad Ali Shokri | Research Expert |
| Faramarz Taheri Dadgar | Public Relation Affair |

## CONTENTS

## Table of Figures

## Abstract

Quantitative characterization of terrain ruggedness is essential for diverse geoscientific applications, including erosion modeling and hydrological prediction. This research developed and validated a robust, 10-meter resolution K-coefficient framework designed as a continuous, transferable measure of geomorphological complexity. The project progressed from a morphometric baseline ($K_0$) to a Hybrid Machine Learning model that integrated Digital Elevation Model (DEM) derivatives with multi-sensor remote sensing data, specifically Sentinel-1 SAR (Radar VV) and Sentinel-2 NDVI. A central scientific focus was the cross-regional generalization across three distinct Iranian regimes: the complex mountainous Zagros, the arid Ravar, and the lowland agricultural Moghan. To achieve this, a standardized spatial architecture was implemented using metric Universal Transverse Mercator UTM projections and SNAP GPT for radar terrain correction to ensure strict pixel-wise correspondence. The research identified a significant multi-regime behavior in terrain modeling: while Moghan and Ravar are strongly topo-driven ($\text{correlations} \approx 0.88 - 0.90$), the Zagros region exhibits a "generalization problem" where surface morphology is decoupled from ruggedness due to missing lithological drivers such as karstification and carbonate rock fractures. To address this heterogeneity, the study developed a regime-aware mixture framework utilizing an interpretable Ridge Regression formula for topographic anchors and a ReliabilityGate mechanism. This gate uses a disagreement metric ($\Delta$) to prevent "mixture collapse" by selectively activating a non-parametric baseline only when it is consistent with the topographic model. Validation using Leave-

One-Region-Out (LOO) experiments and a $\tau_q = 0.80$ threshold demonstrated a 5% reduction in mean cross-region error. Final outputs include standardized 10m K-coefficient maps mapped to a physical range of [0.01, 0.65], providing a stable and operationally ready tool for national-scale geomorphological assessment. Despite these gains, persistent failure in the Zagros domain suggests that future improvements require explicit integration of geological and structural proxies.

Keywords: K coefficient, Terrain Ruggedness, Geomorphological Complexity, Slope, NDVI, Hybrid Machine Learning, Random Forest, Hist Gradient Boosting Reliability Gate

# 1. Problem Statement

## *1.1 The Need for Advanced Morphometric Characterization*

Quantitative characterization of terrain ruggedness and geomorphological complexity is a fundamental requirement in environmental and geoscientific applications, including erosion modeling, hydrological prediction, and large-scale geospatial decision-making. In many practical contexts, decision-makers require a robust numerical indicator "the K-coefficient" that can consistently differentiate between smooth lowland plains and highly dissected mountainous landscapes. Unlike categorical labels, continuous spatial variables like the K-coefficient provide higher analytical flexibility and enable integration into advanced machine learning (ML) frameworks [2,3].

## *1.2 Limitations of Conventional Morphometric Descriptors*

The core technical problem lies in the inadequacy of conventional descriptors to fully capture terrain behavior across heterogeneous regions. Current approaches rely on individual features such as:

**Slope:** Captures gradient magnitude but ignores multi-directional variability and local elevation oscillations.

**Curvature:** Describes local convexity/concavity but is highly sensitive to noise and scale.

**Terrain Ruggedness Index (TRI):** While a strong predictor of local variability, it often exhibits high multicollinearity with slope, leading to unstable coefficients in traditional linear models.

Furthermore, these indices are fundamentally geometric; they represent elevation-driven morphology but fail to capture physical surface properties such as scattering roughness, moisture response, and vegetation canopy structure. Without incorporating multi-sensor remote sensing data, such as Sentinel-1 SAR[1] (radar backscatter) and Sentinel-2 NDVI[2] (vegetation), morphometric models remain physically incomplete [1,3].

### *1.3 The "Multi-Regime" Challenge and Generalization Failure*

A critical scientific gap addressed by this research is the lack of cross-regional generalization in existing models. Standard geomorphological models are often locally calibrated and fail when exposed to "distribution shifts" in new geographic domains.

The sources identify a "Zagros Generalization Problem," where models trained on one region (e.g., Moghan or Ravar) fail catastrophically when applied to another (e.g., Zagros). This suggests that the relationship between topography and terrain complexity is not universal but regime-dependent:

**Topo-Driven Regimes:** In regions like Moghan (lowland plains) and Ravar (arid sedimentary/volcanic terrain), the K-

[1] Synthetic Aperture Radar
[2] Normalized Difference Vegetation Index

coefficient is strongly correlated $(0.88 - 0.90)$ with morphometric structure.

**Lithology-Controlled Regimes:** In regions like Zagros, correlations drop significantly (approx. 0.33), likely due to missing drivers such as lithology, karstification, and fracture density which decouple surface morphology from susceptibility patterns [12,13].

### *1.4 Technical Barriers in Data Harmonization*

The development of a unified framework is further hindered by significant technical obstacles in multi-source data integration. Raw satellite products frequently lack consistent Coordinate Reference Systems (CRS) and affine transforms. For instance, Sentinel-1 GRD[3] measurement files often rely on Ground Control Points (GCPs) rather than explicit geocoding, leading to "NaN-dominated" outputs and spatial misalignment in standard GIS pipelines. Additionally, subtle discrepancies such as one-pixel offsets during resampling can break the pixel-wise correspondence required for machine learning inference [4,7].

### *1.5 The Gap: Absence of a Stable, Mixture-Based Framework*

Current research lacks a methodology that can:

**Detect Regime Shifts:** Automatically identify whether a pixel belongs to a stable topo-driven regime or a complex, lithology-controlled one.

[3] Ground Range Detected

**Mitigate Mixture Collapse:** Prevent "naive" mixture models from degrading performance in regions where different predictors (e.g., human-layer water proxies) provide inconsistent signals.

**Standardize Outputs:** Map abstract model scores into a physically interpretable range (e.g.,$[0.01, 0.65]$) that remains stable across national scales [9].

### *1.6 Summary of Research Objectives*

This research addresses these gaps by developing a Hybrid Remote Sensing and Machine Learning Framework. The proposed study moves beyond local calibration to implement a Regime-Aware approach, utilizing a ReliabilityGate mechanism to stabilize cross-region predictions. By integrating topographic derivatives with radar backscatter, optical indices, and anthropogenic "human layers," this study seeks to provide a definitive, transferable solution for high-resolution (10m) K-coefficient mapping across the diverse geomorphological landscapes of Iran.

## Significance and Justification

### *2.1 Scientific Significance*

This research addresses several fundamental gaps in geomorphometry and remote sensing:

**From Categorical to Continuous Modeling:** Traditional geomorphological studies often rely on categorical labels; however, this research promotes continuous spatial variables (the K-coefficient), which provide higher analytical flexibility, enable probabilistic interpretations, and allow for direct integration into machine learning frameworks.

**Overcoming Descriptive Limitations:** No single morphometric feature (slope, curvature, etc.) is sufficient to describe terrain behavior across heterogeneous regions. This study justifies a hybrid approach, integrating morphometric variables with physical surface properties—such as scattering roughness from Sentinel-1 radar and vegetation density from Sentinel-2 NDVI—to capture land surface characteristics that Digital Elevation Models (DEMs) alone cannot express.

**Regime-Aware Discovery:** The research identifies the "multi-regime" nature of landscapes, proving that relationships between topography and terrain complexity are not universal. By uncovering the Zagros Generalization Problem, the study shifts the scientific focus from "global" models to specialized mixture frameworks that account for lithological and structural drivers like karstification.

**Methodological Innovation:** The introduction of the ReliabilityGate mechanism is a significant technical contribution, providing a way to detect and prevent model collapse during cross-regional transfer.

### *2.2 Practical Importance and Applications*

The K-coefficient serves as a robust numerical indicator that differentiates smooth lowlands from highly dissected mountainous landscapes, making it essential for:

**Soil Erosion and Hydrological Modeling:** High-resolution mapping of ruggedness is a prerequisite for predicting erosion patterns and hydrological responses.

**Watershed Prioritization:** The framework enables researchers to conduct watershed susceptibility analysis and prioritize regions for conservation or infrastructure development.

**Environmental and Land Management:** Standardized physical ranges (e.g., K-physical $[0.01, 0.65]$) allow for consistent thresholding and decision-making across diverse geographic domains, from the agricultural plains of Moghan to the arid terrains of Ravar.

### *2.3 Beneficiaries and Policy Relevance*

**Geospatial Decision-Makers:** Government agencies and environmental planners benefit from a reproducible methodology that can be operationally deployed at a national scale with minimal adaptation.

**Resource Managers:** By incorporating "human layers" (e.g., irrigation canals and water frequency), the research provides insights into how anthropogenic transformations alter surface processes, which is vital for agricultural and water infrastructure planning.

**Standardization for Policy:** The project provides a unified 10-meter grid framework, ensuring that multi-source data (topographic, optical, and radar) are harmonized, which is critical for maintaining data integrity in large-scale national production settings.

### *2.4 Timeliness of the Research*

This research is particularly timely due to the current availability of high-resolution satellite constellations and advanced computational tools:

**Data Availability:** The maturity of the Sentinel-1 and Sentinel-2 programs provides the multi-sensor data necessary to improve the physical representativeness of terrain indices.

**Technological Readiness:** The development of sophisticated algorithms like HistGradientBoosting and Ridge Regression, combined with automated processing engines like Sentinel Application Platform (SNAP) on Graph Processing Tool (GPT), now makes it possible to handle the heavy computational burden of processing national-scale, high-resolution datasets.

**Urgency for Transferable Models:** As environmental changes affect diverse regions differently, there is an urgent need for models that do not require local re-calibration but can generalize across different geomorphological regimes.

## Research Objectives

### *3.1 General Objective*

The primary goal of this research is to develop and validate a robust and transferable 10-meter resolution framework for estimating the K-coefficient, a continuous scalar field that serves as a normalized measure of terrain ruggedness and geomorphological complexity. The framework aims to maintain predictive stability across geographically and climatically independent domains specifically the Zagros, Ravar, and Moghan regions by integrating topographic derivatives with multi-sensor remote sensing data.

### *3.2 Specific Objectives*

**Extract and Harmonize Morphometric Indices:** Utilize DEMs to derive a suite of geometric descriptors, including slope, TRI, aspect, and curvature. This objective includes projecting data into region-specific Universal Transverse Mercator (UTM) zones and resampling all topographic layers to a unified 10-meter grid to ensure physically meaningful gradient computations.

**Process Sentinel-1 Radar Backscatter:** Capture physical surface properties such as scattering roughness and micro-texture using Sentinel-1 SAR Vertical-Vertical (VV) polarization. A key sub-objective is implementing a stable preprocessing pipeline using SNAP GPT to perform Range-Doppler terrain correction, thereby overcoming the lack of explicit CRS in raw GRD products.

**Derive and Project Sentinel-2 NDVI:** Incorporate vegetation density and surface greenness by computing the Normalized Difference Vegetation Index (NDVI) from Sentinel-2 or HLS[4] optical products. This objective requires strict CRS management, reprojecting geographic coordinates to metric UTM zones prior to resampling to avoid grid collapse.

**Develop a Hybrid Machine Learning Framework:** Transition from a baseline morphometric index (K0) to a high-capacity Hybrid Machine Learning model. This includes

[4] Harmonized Landsat Sentinel

evaluating algorithms such as Random Forest and HistGradientBoosting to integrate topographic, optical, and radar features into a single predictive pipeline.

**Implement Cross-Region Validation Strategies:** Rigorous evaluation of model transferability using Leave-One-Region-Out (LOO) validation and spatial block cross-validation. This objective aims to test the model's performance under "distribution shifts" when moving between geomorphologically distinct regimes.

**Design a ReliabilityGate-Based Mixture Model:** Develop a stabilized mixture framework that combines an interpretable topo-driven Ridge Regression formula with an empirical baseline. A critical component is the ReliabilityGate, which measures the disagreement ($\Delta$) between components to prevent "mixture collapse" in unstable regions.

**Evaluate Generalization Performance and Regime Drivers:** Analyze the spatial and statistical patterns of model residuals to identify regional limitations, such as the "Zagros Generalization Problem". This involves investigating the role of missing physical drivers—such as lithology, karstification, and fracture density—that decouple surface morphology from terrain complexity in certain regimes.

**Standardize Outputs for Operational Use:** Map abstract model scores into a fixed physical range ($K_{physical} \in [0.01, 0.65]$) to ensure consistent interpretation, thresholding, and decision-making across national-scale applications.

## Research Questions

### *4.1 Primary Research Questions*

To what extent is the K-coefficient governed by purely morphometric derivatives versus multi-sensor physical indicators?

Does the integration of Sentinel-1 radar backscatter and Sentinel-2 NDVI significantly improve the physical representativeness and predictive performance of the K-coefficient compared to models relying solely on DEM-derived geometry?

Can a machine learning model train on a specific geomorphological region effectively generalize to independent domains?

How does the model maintain predictive stability when exposed to "distribution shifts" between geographically and climatically distinct regions such as the arid Ravar mountains and the agricultural Moghan plains?

Is the Zagros region a distinct geomorphological regime that requires non-topographic predictors?

What factors contribute to the "Zagros Generalization Problem," where standard morphometric models fail (producing negative values), and to what extent do missing drivers like lithology, karstification, and fracture density decouple surface morphology from terrain complexity in this domain?

### *4.2 Methodological & Technical Questions*

Can a regime-aware mixture framework mitigate "mixture collapse" in cross-regional applications?

How effective is the ReliabilityGate mechanism—which uses a disagreement metric (Δ) between topo-driven and baseline models—at preventing systematic errors and stabilizing performance across heterogeneous landscapes?

Do anthropogenic surface modifications provide significant explanatory power for K-coefficient patterns?

When modeled as "human residuals" (using NDWI[5] and canal density proxies), can anthropogenic data improve K-estimation in lowland, human-impacted regimes like Moghan where purely geomorphometric signals are weak?

How does a harmonized 10-meter multi-source grid architecture impact the numerical stability of the framework?

What are the technical requirements for ensuring strict pixel-wise correspondence and CRS alignment between SAR, optical, and topographic data to avoid "silent failures" during machine learning inference?

[5] Normalized Difference Water Index

### *4.3 Summary Table: Research Focus Areas*

| Question Type | Core Concept | Relevant Source Evidence |
|---|---|---|
| Scientific | Multi-regime behavior | Zagros, Ravar, and Moghan exhibit distinct correlation structures (e.g., $corr(K, TRI)$ is ~0.90 in Moghan but only ~ 0.33 in Zagros). |
| Methodological | ReliabilityGate | A $T_q = 0.80$ threshold was found to optimize the tradeoff between baseline utilization and stability. |
| Physical | Missing Drivers | Lithological/structural controls in carbonate terrains (karst) are hypothesized to be the primary cause of topo-driven model failure. |
| Technical | Data Integration | Standardizing nodata to -9999 and using SNAP GPT for SAR terrain correction were essential for pipeline reproducibility. |

## Research Hypotheses

Low-slope and low-TRI areas correspond to lower K values. The sources state that the K-coefficient is designed as a normalized measure where values close to zero represent smooth lowland environments. Statistical analysis in regions like Moghan and Ravar shows a very high correlation (0.88 – 0.90) between the K-coefficient and morphometric features such as the TRI and slope. Furthermore, the extracted topo-driven formula identifies slope as the dominant positive driver (coefficient of ~ 0.309), while a composite index ($S$) combining DEM, slope, and TRI is used to effectively gate the ruggedness regime.

Integrating NDVI and radar backscatter improves the physical representativeness and predictive performance of the model. The research is predicated on the limitation that morphometric layers only represent elevation-driven geometry and miss physical surface properties like scattering roughness and vegetation canopy structure. Permutation importance analysis confirms that while topography is dominant, radar (Sentinel-1 VV) and NDVI (Sentinel-2) are significant and moderate contributors to model performance, respectively. Integrating these sensors in a hybrid framework led to a substantial explanatory power with $R^2$ values reaching approximately 0.71 to 0.73 in multi-feature models.

Topography-only models fail to generalize in lithology-controlled or karstified terrains. Evidence from the sources identifies a "Zagros Generalization Problem," where models trained on other regions fail catastrophically when applied to Zagros (yielding $R^2$ values of -0.34 to -0.35). This failure persists even in topo-only models because Zagros behaves as an out-of-distribution domain where surface morphology is decoupled from terrain complexity. Expert consultation suggests this is due to missing drivers such as carbonate rock distribution, fracture density, and karstification, which are not captured by standard morphometric variables.

A reliability-based gating mechanism reduces mixture collapse and stabilizes cross-regional predictions. The sources describe how a "naive" mixture of topo-driven and baseline models can lead to catastrophic failures and mixture collapse in certain domains. To mitigate this, the ReliabilityGate was developed to measure the disagreement (Δ) between model components; it "shuts off" the baseline branch when it is deemed untrustworthy. Experimental results prove that this mechanism completely eliminates mixture collapse, yielding a 5% reduction in mean cross-region error and significantly improving stability in regions like Ravar and Moghan.

## Literature Review

### *6.1 Quantitative Morphometric Analysis and TRI Studies*

Quantitative characterization of terrain is a fundamental requirement in environmental and geoscientific applications, including erosion modeling, hydrological response prediction, and watershed susceptibility analysis. Traditionally, geomorphologists have relied on a suite of conventional descriptors to characterize landforms:

**Slope:** Captures the first-order magnitude of terrain gradient but is limited as it does not represent multi-directional variability or local oscillations.

**Curvature:** Describes local convexity and concavity but is notably sensitive to noise and the scale of the input data.

**Aspect:** Provides directional information but presents challenges in statistical modeling due to its circular nature.

A key metric in modern geomorphometry is the TRI, which describes local elevation oscillation and neighborhood-scale variability. In ML frameworks, TRI has been identified as a dominant feature; permutation importance analysis consistently ranks TRI as the most influential predictor for determining terrain complexity. However, research highlights a significant challenge regarding multicollinearity between TRI and slope. In linear models such as Ridge Regression, slope often absorbs the primary ruggedness signal, leading to negative coefficients for TRI as a statistical balancing mechanism rather than a physical indicator. Thus, the literature suggests that no single feature is sufficient, requiring the development of composite indices like the K-coefficient [1,2,3].

### *6.2 Remote Sensing Integration in Erosion Modeling*

While DEMs provide a structural foundation, they are inherently limited because they represent only geometric variability. For robust erosion and susceptibility modeling, the literature emphasizes incorporating physical surface properties.

**Sentinel-1 (SAR):** Radar backscatter is sensitive to structural roughness and dielectric properties, providing information on surface scattering and micro-texture that DEMs cannot capture. Using VV polarization is particularly effective for general surface scattering sensitivity [4].

**Sentinel-2 (NDVI):** Vegetation indices provide a dimension of surface greenness and canopy structure, which modulate the relationship between terrain and susceptibility [6].

**Anthropogenic Modifiers ("Human Layers"):** Modern landscapes are often shaped by human transformation. Proxies for water management such as the Normalized Difference Water Index (NDWI) and canal area fraction represent anthropogenic modifiers that may influence patterns in lowland agricultural regimes like Moghan.

### *6.3 Machine Learning in Geomorphology*

The field is transitioning from categorical landform labels to continuous spatial variables, which provide higher analytical flexibility and enable integration into advanced ML frameworks [14].

**Algorithmic Selection**: High-capacity learners like Random Forest and HistGradientBoostingRegressor are preferred for their ability to handle non-linear interactions and missing data across large-scale geospatial datasets.

**Interpretability:** While non-linear models offer high performance, there is a growing trend toward using Ridge Regression on robust-normalized variables to extract interpretable, closed-form mathematical equations [8].

**Data Harmonization:** A critical technical barrier identified in the literature is the need for strict pixel-wise correspondence and CRS management. Multi-sensor workflows require reprojecting all data to a metric CRS (e.g., UTM) prior to resampling to avoid "silent failures" like grid collapse [10,11].

### *6.4 Cross-Region Generalization and Distribution Shift*

A significant scientific challenge is the lack of transferability of models across independent geographic domains with distinct geomorphological conditions.

**The Zagros Generalization Problem:** Models trained on one region often fail when applied to another due to distribution shift. For instance, models trained on Moghan and Ravar fail catastrophically in Zagros, yielding negative $R^2$ values.

**Validation Rigor:** Traditional random pixel splitting can overestimate performance due to spatial autocorrelation. The literature recommends LOO and Spatial Block Cross-Validation to assess true predictive power under distribution shift [13].

### *6.5 Regime-Based Modeling and Mixture Frameworks*

Recent evidence suggests that the relationship between topography and terrain complexity is not universal but regime-dependent.

**Topo-Driven vs. Lithology-Controlled:** Regions like Moghan and Ravar are "topo-driven," where $K$ is highly correlated with morphometric indices (corr ≈ 0.90). Conversely, regions like Zagros are likely lithology-controlled, where carbonate rock distribution, fracture density, and karstification decouple surface morphology from ruggedness.

**Gating Mechanisms:** To handle this heterogeneity, the literature proposes Logistic Gating based on morphometric signal strength ($S = zDEM + zSlope + zTRI$), which can achieve high accuracy (AUC[6] ≈ 0.94) in separating rugged topo-driven regimes.

**Stability and Reliability:** A major risk in multi-regime modeling is "mixture collapse," where an unreliable baseline component degrades the final output. The introduction of the ReliabilityGate which measures the disagreement (Δ) between topo-driven and baseline models prevents this collapse by restricting the influence of the baseline to stable areas. Experiments show this mechanism can achieve a 5% reduction in mean cross-region error [9].

## Study Area Description

The research is conducted across three geographically and geomorphologically distinct regions in Iran. These areas were deliberately selected to represent contrasting environmental regimes, ranging from high-altitude mountainous systems to lowland agricultural plains. This diversity is critical for evaluating the multi-regime behavior of the K-coefficient.

---

[6] Area Under the Curve

### *7.1 Zagros Region: The Lithology-Controlled Mountainous Regime*

**Geographic Location:** Located in western Iran, the study focuses on the Kermanshah domain. For spatial processing, it is analyzed within UTM Zone 38N.

**Topographic Characteristics:** This region represents a complex mountainous system characterized by strong elevation gradients and significant structural heterogeneity.

**Geomorphological Features:** The Zagros is identified as a distinct geomorphic regime where surface morphology is often decoupled from terrain complexity. It contains a high proportion of carbonate and karstified terrain.

**Scientific Significance:** In this region, geomorphic processes are likely lithology-controlled. Features such as subsurface dissolution, fracture networks, and karst morphology create a landscape where traditional topographic metrics (like slope) do not fully explain terrain behavior [12].

### *7.2 Ravar Region: The Arid Rugged Regime*

**Geographic Location:** Situated in central/eastern Iran, the Ravar study area is bounded by longitudes 56.4686°E to 57.0314°E and latitudes 30.7229°N to 31.0271°N. It is processed under UTM Zone 40N.

**Climatic Conditions:** Ravar is characterized by an arid to semi-arid climate. Statistical analysis of NDVI (mean ≈ 0.00094) confirms the region is largely barren and extremely dry, with effectively zero vegetation across the 95th percentile of the area.

**Topographic Characteristics:** The landscape exhibits significant ruggedness with desert-dominated patterns.

Unlike the Zagros, Ravar is primarily topo-driven, meaning its K-coefficient is highly correlated($\approx$ 0.90) with morphometric structure. Its geology consists of a mixture of sedimentary and volcanic lithologies [14].

### *7.3 Moghan Region: The Lowland Agricultural Regime*

**Geographic Location:** Located in northwestern Iran, this region is processed under UTM Zone 39N.

**Climatic Conditions**: Moghan is significantly greener and more vegetated than the other study areas. It has a mean NDVI of approximately 0.22, with maximum values reaching 0.90, reflecting its moderate-to-strong vegetation coverage.

**Topographic Characteristics:** Moghan is a low-relief region dominated by flat plains.

**Geomorphological Features:** It is characterized as a younger and more homogeneous sedimentary plain. The geomorphic structure is agriculturally dominated, making it a representative domain for testing the impact of anthropogenic "human layers" (e.g., irrigation canals) on terrain indicators.

### *7.4 Summary of Multi-Regional Differences*

The following table summarizes the contrasting characteristics of the study regions:

| Feature | Zagros | Ravar | Moghan |
|---|---|---|---|
| **Primary Landscape** | **Complex Mountains** | **Arid Rugged Terrain** | **Lowland Plains** |
| Dominant Regime | Lithology-Controlled | Topo-Driven | Topo-Driven |
| Key Driver | Karst/Fracture Density | Elevation Oscillation | Agricultural/Human Layer |
| Vegetation (NDVI) | Moderate | Extremely Low (Barren) | High (Agricultural) |
| Correlation (K, TRI) | Weak (≈ 0.33) | Strong (≈ 0.90) | Strong (≈ 0.88) |

## Datasets

### *8.1 Digital Elevation Model and Morphometric Derivatives*

The DEM serves as the foundational layer for the entire modeling framework, providing the structural basis for all geometric terrain analysis [17].

**Source & Resolution:** The DEM was acquired at an initial resolution of 30 meters and subsequently resampled to 10 meters using a harmonization procedure to match the hybrid feature stack.

**Projection:** Data were projected into region-appropriate UTM coordinate systems to ensure physically meaningful gradient computations:

- Zagros: UTM Zone 38N (EPSG[7]:32638).
- Moghan: UTM Zone 39N (EPSG:32639).
- Ravar: UTM Zone 40N (EPSG:32640).

**Derived Morphometric Layers:** From the 10m DEM, a suite of morphometric indices was extracted, including Slope, TRI, Aspect, and Curvature.

**Preprocessing Steps:**

1. Projecting from geographic to UTM coordinates.
2. Resampling to a 10m grid to improve the fidelity of ruggedness derivatives.
3. Standardizing Nodata values to -9999 to ensure compatibility with processing engines like SNAP.
4. Robust Normalization: Applying z-normalization based on the median and Interquartile Range (IQR) to allow cross-regional comparison.

[7] European Petroleum Survey Group

### *8.2 Sentinel-1 Radar (SAR)*

Radar data were integrated to capture surface scattering properties and textures (e.g., micro-texture and roughness) that elevation models alone cannot express.

**Specifications:** Data were acquired as Sentinel-1 GRD (Ground Range Detected) products in Interferometric Wide (IW) mode with VV polarization.

**Spatial Resolution:** Final processed output is 10 meters.

Preprocessing Steps (SNAP GPT Workflow): Raw Sentinel-1 products lack explicit CRS and rely on GCPs, necessitating a specialized pipeline:

1. **Orbit Correction:** Applying precise orbit files.
2. **Terrain Correction:** Utilizing the Range-Doppler terrain correction in SNAP GPT to geocode the rasters into the metric UTM grid.
3. **Aggregation:** Mosaicking and compositing multiple scenes using robust statistics (mean or median) to reduce speckle noise.
4. **Alignment:** Warping the final RadarVV layer to match the DEM-derived slope template to resolve one-pixel offsets [15].

### *8.3 Sentinel-2 Optical (NDVI)*

NDVI was utilized as an environmental modifier to incorporate vegetation density and surface greenness into the hybrid model.

**Source:** Sentinel-2 or HLS (Harmonized Landsat Sentinel) products.

**Processing Level:** Derived using the Normalized Difference Vegetation Index formula:

$$NDVI = (B08 - B04)/(B08 + B04)$$

**Spatial Resolution:** Resampled to 10 meters.

**Critical Preprocessing Steps:**

1. **Metric Reprojection:** Because raw NDVI is often in geographic coordinates (EPSG:4326), it must be reprojected to a UTM zone first before resampling to 10m to avoid raster collapse (e.g., 1×1-pixel errors).

2. **Grid Alignment:** Using an *align_to_template* script to match the exact transform and dimensions of the slope raster [16].

### *8.4 Human Layer Dataset Google Earth Engine (GEE)*

These layers represent anthropogenic modifiers and surface water infrastructure proxies, which were exported from Google Earth Engine (GEE).

**Composition (3-Band Stack):**

1. ***NDWI_median*:** Stable indicator of surface moisture.

2. ***NDWI_waterfreq***: Persistence metric for water bodies.

3. ***Canal_area_fraction*:** Proxy for irrigation canal density.

Resolution & Projection: 10 meters, aligned to the regional UTM framework [5].

### *8.5 Vector Shapefiles*

Vector data were used as the primary spatial boundary definition for the project.

**Functions:**

Defining the Region of Interest (ROI) for each study area.

Generating bounding boxes (WGS84) for data acquisition from services like ASF Alaska and NASA AppEEARS.

Constraining all raster clipping and masking operations to ensure fully overlapping domains for multi-source data.

### *8.6 Summary of Dataset Harmonization*

| Feature | DEM / Morphometrics | Sentinel-1 (Radar) | Sentinel-2 (NDVI) | Human Layers |
|---|---|---|---|---|
| Final Resolution | 10 Meters | 10 Meters | 10 Meters | 10 Meters |
| Projection | Regional UTM | Regional UTM | Regional UTM | Regional UTM |
| Nodata Encoding | -9999 | -9999 | -9999 | -9999 |
| Primary Tool | GIS/Python | SNAP GPT | Python/Warp | GEE / Python |

**Methodology**

The research methodology is implemented in a multi-stage workflow, progressing from spatial harmonization and hybrid feature extraction to the development of a Regime-Aware ReliabilityGate framework.

### *9.1 DEM Preprocessing*

The foundation of the geospatial framework is the DEM, which serves as the reference for all morphometric analysis.

**Projection to UTM:** Because morphometric computations assume metric units, all DEMs are projected into region-appropriate UTM zones: UTM 38N for Zagros, 39N for Moghan, and 40N for Ravar.

**Resampling to 10 m:** To establish a unified spatial grid, DEMs are resampled to a common 10-meter resolution. This harmonization improves the fidelity of ruggedness derivatives, which can be underestimated on coarser grids.

**Nodata Handling:** To ensure compatibility with processing engines like SNAP and Rasterio, all nodata values are standardized to a numerical sentinel of -9999.

**Grid Harmonization:** The 10 m resampled DEM is used as the reference template grid for the reprojection and alignment of all multi-source raster layers.

### *9.2 Morphometric Derivatives*

A suite of geometric descriptors is extracted to characterize terrain complexity [2].

**Feature Extraction:** Standard derivatives including Slope, TRI, Aspect, and Curvature are computed from the 10 m DEM.

**Robust Normalization:** To allow cross-regional comparison, all variables are transformed using Robust Z-Normalization:

$$z(x) = \frac{x - median(x)}{IQR(x)}$$

This reduces sensitivity to outliers across diverse landscapes.

Logistic Transformation for$(K_0)$: In the initial stage, a baseline index $(K_0)$ is constructed using a logistic transformation of weighted robust-normalized TRI and slope. This produces a continuous score bounded within [1].

### *9.3 Sentinel-1 Radar Processing*

Sentinel-1 SAR data are integrated to capture surface scattering roughness and micro-texture not expressed by DEM derivatives.

**GRD Data Acquisition:** Sentinel-1 GRD products are acquired in IW mode with VV polarization [4,7].

**SNAP GPT Workflow:** Raw products rely on GCPs and lack explicit CRS, leading to mosaic failures in standard pipelines. Consequently, a definitive SNAP GPT pipeline is implemented, applying Orbit Correction and Range-Doppler Terrain Correction.

**Median Compositing:** Multiple scenes are aggregated using median compositing to reduce speckle noise and scene-specific variability.

**Alignment to DEM Grid:** Radar composites are warped using a dedicated script (*align_to_template.py*) to match the exact dimensions and transform of the DEM-derived slope template.

### *9.4 NDVI Processing*

Vegetation density is incorporated as an environmental modifier to provide contextual surface information.

**Band Selection & Calculation:** NDVI is calculated from Sentinel-2 or HLS Red (B04) and NIR (B08) bands using the formula:

$$NDVI = (B08 - B04)/(B08 + B04)$$

**Reprojection and Resampling:** To avoid "raster collapse" (corrupted 1×1-pixel outputs), NDVI is reprojected to a metric UTM zone first and then resampled to 10 m.

**Alignment:** The final NDVI layer is aligned to the regional slope template to resolve one-pixel offsets caused by reprojection rounding.

### *9.5 Hybrid Machine Learning Model*

The framework expands from a morphometric baseline into a high-capacity Hybrid Remote Sensing and Geomorphometric model.

**Feature Selection:** The model utilizes 11 predictors, including five topographic layers (DEM, slope, TRI, curvature, aspect), radar backscatter, NDVI, and three "human layers" (*NDWI_median*, *NDWI_waterfreq*, *canal_fraction*).

**Sampling Strategy:** Approximately 200,000 random pixels are extracted from each region (600,000 total). A revised valid mask ensures that samples include partial data (e.g., radar gaps) to reduce sampling bias [9].

**Train/Test Split:** An 80/20 random split is utilized for internal validation, stratified by region.

**Algorithm**: HistGradientBoostingRegressor is selected for its efficiency with large datasets and support for non-linear interactions, configured with $\text{max\_depth} = 10$ and $\text{max\_iter} = 800$ [10].

**Feature Importance:** Permutation Importance is used for interpretability, identifying TRI as the dominant predictor, followed by DEM and radar [8].

### *9.6 Cross-Region Validation*

To assess true model transferability under distribution shift, the study implements rigorous validation strategies.

**Leave-One-Region-Out:** The model is trained on two regions and tested on the third, testing whether learned relationships hold in independent domains.

**Performance Metrics:** Validation is conducted using $R^2$, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

**Distribution Shift Analysis:** Experiments identified a "Zagros Generalization Problem," where models trained on Moghan/Ravar yield negative $R^2$ in Zagros due to missing lithological drivers.

### 9.7 Regime-Based Modeling

To reconcile regional differences, the project adopted a Regime-Aware Mixture Model.

**Topo Ridge Formulation:** A stable, interpretable formula is extracted using Ridge Regression on robust-normalized variables for the topo-driven regime:

$$\widehat{K}_{topo} = 0.4088 + 0.0654z(DEM) + 0.3091z(slope) - 0.1084z(TRI) + \cdots$$

**Composite Index S:** A morphometric signal index is defined as:

$$S = z(DEM) + z(slope) + z(TRI)$$

**Baseline Construction:** A non-parametric baseline ($K_{base}$) is constructed using S-binned medians.

**ReliabilityGate Definition:** A disagreement metric $\Delta = |K_{base} - \widehat{K}_{topo}|$ is used to detect baseline instability.

**Tau Quantile Sweep:** A threshold is learned from the distribution; a sweep of $\tau_q$ values identified 0.80 as the optimal setting to balance stability and performance.

**Mixture Formulation:** The final K estimate is a weighted combination:

$$\widehat{K} = w\widehat{K}_{topo} + (1 - w)(K_{base} + \hat{r}_{human})$$

where is controlled by the ReliabilityGate and represents a secondary residual correction derived from human layer proxies.

## Innovation

### *10.1 Integration of Physical and Machine Learning Modeling*

The innovation of this research lies in its transition from conventional, locally calibrated morphometric indices to a regime-aware, hybrid modeling framework that ensures national-scale transferability. Based on the sources, the following five pillars define the scientific and technical novelty of this project:

Unlike traditional approaches that rely solely on geometric descriptors (slope, curvature), this framework integrates elevation-driven morphology with physical surface properties captured by multi-sensor remote sensing.

**Multi-Sensor Fusion:** It combines Sentinel-1 SAR (capturing scattering roughness and micro-texture) and Sentinel-2 NDVI (capturing vegetation canopy) to represent land surface properties that DEMs cannot express.

**Anthropogenic Context:** It innovatively incorporates "human layers" (proxies for water management like NDWI and canal density) as a residual correction term to account for human-induced transformations in lowland regimes.

### *10.2 The ReliabilityGate Mechanism*

The project introduces the ReliabilityGate, a novel diagnostic and control component designed to ensure model stability during cross-regional transfer.

**Disagreement Metric (Δ):** It measures the absolute difference between a non-parametric baseline ($K_{base}$) and a stable topo-driven ridge model ($\widehat{K}_{topo}$).

**Failure Prevention:** If the disagreement (Δ) exceeds a learned threshold ($\tau$), the mechanism "shuts off" the potentially unstable baseline and reverts the model to the reliable topographic anchor. This explicitly prevents "mixture collapse," where an unreliable component would otherwise degrade overall predictive performance.

### *10.3 Stabilized Mixture Modeling Approach*

The final production model is not a single "black-box" algorithm but a stabilized mixture framework that reconciles interpretability with predictive power.

**Ridge-Stabilized Formula:** The model extracts an interpretable, closed-form linear equation for topographic ruggedness using Ridge Regression, which mitigates the instability caused by high multicollinearity between slope and TRI.

**Empirical Anchoring:** This is combined with an S-binned median baseline (robust to outliers) and a human residual pathway, allowing the model to adapt to different landscapes without losing physical grounding.

### *10.4 Cross-Region Evaluation Framework*

The research moves beyond standard validation techniques to implement a stringent cross-region evaluation framework.

**Leave-One-Region-Out:** By training on two regions and testing on an independent third region (e.g., Zagros, Ravar, or Moghan), the study rigorously evaluates performance under "distribution shift".

**Spatial Rigor:** The framework utilizes spatial block cross-validation (20 km blocks) to eliminate leakage from local spatial autocorrelation, ensuring that the model is learning generalized physical relationships rather than memorizing spatial coordinates.

### *10.5 Multi-Regime Modeling Interpretation*

A significant scientific innovation is the shift from a "global model" philosophy to a multi-regime interpretation of geomorphology.

**Regime Identification:** The study identifies that terrain behavior is governed by distinct regimes: "topo-driven" regimes (like Moghan and Ravar) where morphology directly explains ruggedness, and "lithology-controlled" regimes (like Zagros) where factors like karstification and carbonate rock fracture density decouple surface morphology from the target metric.

**Logistic Gating:** It uses a logistic gating function based on morphometric signal strength ($S = zDEM + zslope + zTRI$) to automatically classify pixels into the appropriate predictive regime with an AUC of approximately 0.95.

## Tools and Software

The implementation of the K-coefficient framework utilizes a robust computational environment designed for memory-safe raster operations, automated SAR processing, and reproducible machine learning workflows.

### *11.1 Python Ecosystem (Core Analysis & Machine Learning)*

Python serves as the primary language for data orchestration, feature engineering, and modeling.

**Scikit-learn:** This library is central to the project's modeling phase. It is used to implement the Hist Gradient Boosting Regressor (configured with $\text{max\_depth} = 10$ and $\text{max\_iter} = 800$), the SimpleImputer for handling missing values (NaNs), and the Ridge Regression model used to extract the interpretable topo-driven formula. It also facilitates stratified sampling and permutation importance analysis.

**Rasterio:** Used for standard raster I/O and spatial operations. While raw Sentinel-1 products were not initially compatible with Rasterio due to GCP-based referencing, it is used throughout the pipeline for mosaicking, warping, and geocoding checks once data are geocoded.

**NumPy:** Utilized for high-performance array manipulations, specifically for calculating Robust Z-Normalization (based on median and IQR) and the disagreement metric ($\Delta$) between model components.

**Custom Pipeline Scripts:** Several dedicated Python scripts were developed to automate the **workflow, including:**

***align_to_template.py*:** Resolves one-pixel discrepancies by warping NDVI and RadarVV rasters to the slope template.

***run_hybrid_k.py*:** Executes the final model inference across regional data directories.

**postcheck_hybrid.py:** Conducts statistical validation of ruggedness thresholds.

### *11.2 SNAP GPT (Sentinel-1 SAR Processing)*

The ESA GPT is a mandatory preprocessing component for radar data.

**Terrain Correction:** Used to perform Range-Doppler terrain correction and orbit file application **to geocode raw GRD measurement TIFFs into standard projected rasters.**

**Error Mitigation:** SNAP was critical for overcoming the lack of explicit CRS and affine transforms in raw Sentinel-1 products. The workflow was specifically configured to run in offline mode to ensure stability without network-dependent auxiliary data downloads.

### *11.3 QGIS and GDAL (Visualization and Spatial Utilities)*

**QGIS:** Employed for the initial definition of the Region of Interest (ROI) using vector shapefiles and for visual quality assurance (QA) of the final prediction maps.

**GDAL:** Serves as the backend for many Python-based transformations. It was specifically utilized for reprojecting the Zagros K rasters from geographic (EPSG:4326) to UTM metric coordinates to resolve sampling alignment issues.

### *11.4 Google Earth Engine*

GEE was integrated into the pipeline to handle the acquisition of large-scale "Human Layers".

**Anthropogenic Proxies:** Used to calculate and export 10-meter stacks of *NDWI_median*, *NDWI_waterfreq*, and $canal\_area\_fraction$. These layers were exported as multi-band GeoTIFFs, simplifying feature loading and ensuring consistent band ordering for the human residual pathway.

### *11.5 Jupyter Notebooks*

Jupyter Notebooks are used throughout the development lifecycle for exploratory data analysis (EDA), diagnostic tile-based experiments (such as the Zagros subregion analysis), and iterative testing of the ReliabilityGate quantiles ($\tau_q$).

## Expected Results

### *12.1 High-Resolution (10 m) K-Coefficient Maps*

The primary tangible output of this research is a set of continuous scalar field maps representing terrain ruggedness and geomorphological complexity.

**National-Scale 10 m Resolution**: The pipeline produces 10 m resolution maps that ensure high fidelity for slope and ruggedness derivatives, which are often underestimated on coarser grids.

**Dual-Scale Mapping:** The results include both abstract coefficient rasters and standardized physical-scale rasters ($K_{Physical} \in [0.01, 0.65]$).

**Consistent Output Distribution:** Statistical validation across the pilot regions (Zagros, Ravar, and Moghan) confirms that the maps appropriately reflect geomorphological context—for example, Ravar and Zagros exhibit high mean values($\sim 0.52 - 0.54$), while the Moghan plain shows a significantly lower mean($\sim 0.35$)

### *12.2 Cross-Region Validated Model*

The research provides a model that has been rigorously tested for transferability and stability across independent geographic domains.

**Leave-One-Region-OutValidation:** The model is validated by training on two geomorphologically distinct regions and testing on a third, ensuring it can handle "distribution shifts".

**Spatial Rigor:** To prevent performance inflation caused by spatial autocorrelation, the model's predictive power is verified using spatial block cross-validation (20 km blocks).

**Diagnostic residual mapping:** The framework generates residual rasters ($K_{True} - K_{Pred}$) to quantify spatial error patterns and identify regions where missing physical drivers (like lithology) may be influencing results.

### *12.3 Regime-Aware Mixture Framework*

A key innovation of the research is the development of a framework that adapts its logic based on the identified geomorphological regime.

**Automated Regime Classification:** The framework utilizes a Logistic Gating function (achieving an AUC of $\sim 0.9466$) to determine if a pixel belongs to a "topo-driven" or "lithology-controlled" regime based on its morphometric signal strength ($S$).

**ReliabilityGate Mechanism:** This novel component measures the disagreement ($\Delta$) between the topo-driven ridge model and a non-parametric baseline. It prevents "mixture collapse" by shutting off the baseline branch when it is deemed untrustworthy in out-of-distribution domains.

### *12.4 Quantified Generalization Performance*

The effectiveness of the proposed framework is supported by specific numerical performance metrics derived from the pilot studies:

**Error Reduction:** The ReliabilityGate mixture framework achieved an absolute improvement in mean cross-region RMSE, resulting in a 5% reduction in mean error compared to a standard topo-driven baseline.

**Performance Stability:** In the Ravar region, the mixture framework increased the $R^2$ value from 0.689 to 0.776, demonstrating a significant gain in cross-regional predictive accuracy.

**Regime Optimization:** The framework identified $\tau_q = 0.80$ as the optimal reliability threshold, balancing the need for stability with the benefits of baseline integration across diverse landscapes.

### *12.5 Production-Ready Mathematical Formulation*

The research culminates in a fully transparent and reproducible mathematical architecture.

**Interpretable Topo-Ridge Formula:** A closed-form linear equation is provided for the topo-driven regime, identifying slope as the dominant positive driver (coefficient $\approx 0.31$).

**The Mixture Equation:** The final prediction is governed by the ReliabilityGate mixture formula:

$$\widehat{K} = w\widehat{K}_{topo} + (1-w)(K_{base}(S) + \hat{r}_{human})$$

where is the dynamic reliability weight $\hat{r}_{human}$ and is an optional correction term for anthropogenic "human layers" (NDWI/canal proxies).

**Standardized Normalization:** All inputs are processed through Robust Z-Normalization (using median and IQR), ensuring that the production formula is applicable to regions with vastly different numeric scales.

## Timeline

### *13.1 Research Timeline*

The project is structured into five key phases:

Data Foundation, Multi-Sensor Engineering, Algorithmic Development, Rigorous Validation, and Final Production.

| Month | Phase | Activity |
|---|---|---|
| 1 | Data Acquisition | Download regional vector shapefiles, 30m DEMs, Sentinel-1 GRD (IW mode), and Sentinel-2/HLS optical imagery for Zagros, Ravar, and Moghan. |
| 2 | Spatial Harmonization | Project DEMs to regional UTM zones (38N, 39N, 40N); resample to a unified 10m grid; and standardize all nodata values to -9999. |
| 3 | Radar Processing | Implement the SNAP GPT pipeline to perform Range-Doppler terrain correction and orbit file application for Sentinel-1; generate median-composited RadarVV layers. |
| 4 | Feature Alignment | Compute NDVI (reprojected to UTM prior to 10m resampling) and extract Human Layers (NDWI/canal proxies) from GEE; execute template-based grid alignment to resolve one-pixel offsets. |
| 5 | Baseline Extraction | Apply Robust Z-Normalization (median/IQR) to all morphometric features and extract the interpretable Ridge Regression formula for the topo-driven regime. |
| 6 | Regime Gating | Define the composite signal index ($S$) and train the Logistic Gating function to classify pixels into topo-driven or non-topo-driven regimes. |
| 7 | Mixture Modeling | Develop the ReliabilityGate mechanism to measure baseline disagreement ($\Delta$) and integrate the Human Residual pathway for lowland correction. |

| | | |
|---|---|---|
| 8 | Hyperparameter Sweep | Conduct the $\tau_q$ quantile sweep (evaluating the 0.30–0.80 range) to optimize the ReliabilityGate threshold for cross-regional stability. |
| 9 | Cross-Region Validation | Execute LOO experiments; perform Spatial Block Cross-Validation (20km blocks) to eliminate autocorrelation leakage. |
| 10 | Failure Analysis | Conduct tile-based diagnostic experiments (50km tiles) to investigate the Zagros Generalization Problem and the impact of missing lithological drivers. |
| 11 | Final Production | Generate definitive 10m K-coefficient maps; standardize outputs to the physical range ($K \in [0.01, 0.65]$) to enable operational thresholding. |
| 12 | Synthesis & Reporting | Consolidate mathematical formulations and final performance metrics ($R^2$, RMSE, MAE) into the comprehensive research report. |

## Deliverables

### *14.1 Research Deliverables*

**Final K-Coefficient Maps (GeoTIFF):** High-resolution (10 m) continuous scalar field maps representing terrain ruggedness for the Zagros, Ravar, and Moghan regions. These files (e.g., *FINAL_RAVAR_K_COEFFICIENT.tif*) are provided as georeferenced GeoTIFFs in region-specific UTM projections (Zones 38N, 39N, and 40N) with standardized nodata encoding of -9999.

**Physical-Range Standardized K Maps:** A set of standardized rasters where the abstract coefficient is mapped into a fixed, operationally interpretable physical range of [0.01, 0.65]. These products (e.g., FINAL_ZAGROS_K_PHYSICAL_0p01_0p65.tif) enable consistent thresholding and cross-regional decision-making.

**Saved Machine Learning Models *(.pkl)*:** Serialized model artifacts, including the *final_hybrid_k_model.pkl* (Random Forest / HistGradientBoosting) and the finalized ReliabilityGate mixture model. These include the trained weights for the topographic ridge formula, the logistic gating coefficients, and the human residual regressor.

**Python Processing & Inference Scripts:** A reproducible library of scripts used for the end-to-end pipeline, including:

***align_to_template.py*:** Ensures strict pixel-wise correspondence between multi-source layers.

***run_hybrid_k.py*:** Executes model inference across regional datasets.

***make_k_physical_all.py*:** Standardizes outputs to the physical parameter domain.

***postcheck_hybrid.py*:** Conducts statistical validation and threshold analysis.

**SNAP XML Workflows:** XML-based graph processing files for SNAP GPT, detailing the mandatory Sentinel-1 SAR preprocessing steps. These workflows automate orbit correction, Range-Doppler terrain correction, and geocoding to ensure radar data is perfectly aligned with the 10 m DEM grid.

**Full Technical Report:** A comprehensive document detailing the multi-regime modeling philosophy, the identification of the "Zagros Generalization Problem," and the final Regime-Aware ReliabilityGate framework. The report synthesizes the scientific findings regarding the role of lithology and human anthropogenic modifiers.

**Numerical Validation Appendix:** A detailed appendix containing final performance metrics ($R^2$, RMSE, MAE) derived from LOO validation and the quantile sweep. This includes the optimized results for the MixRel + Human configuration at the selected threshold of $\tau_q = 0.80$

## Figures

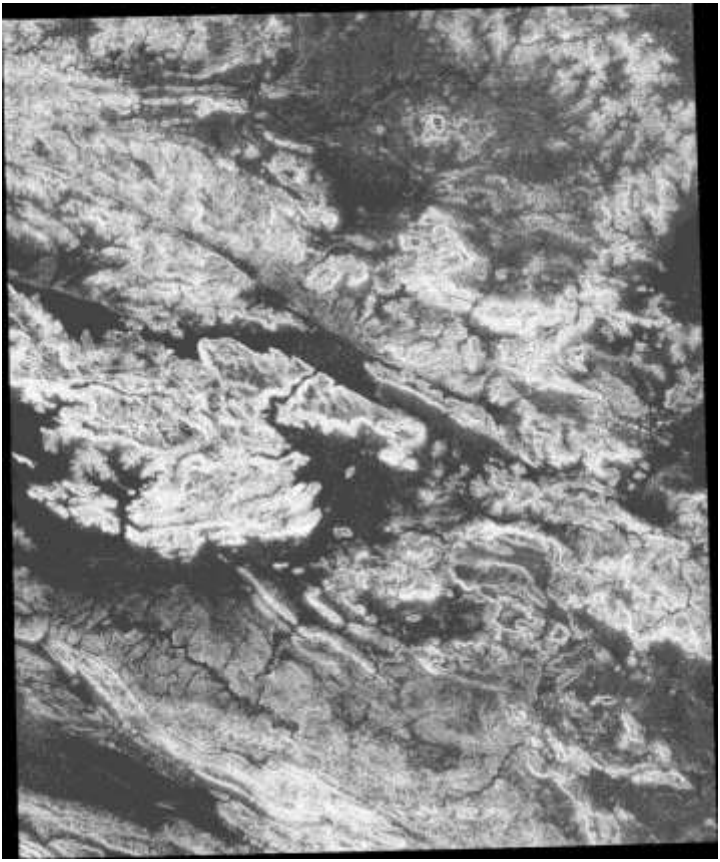

Figure 1. Zagros Baseline $K_0$ Map (Geomorphometric Index): Spatial distribution of the baseline geomorphometric $K_0$ index over the Zagros study area, computed from robust-normalized terrain derivatives (slope and TRI) and transformed into a continuous [0–1] range using a sigmoid function. This

layer represents the initial topography-driven susceptibility component prior to introducing radar and vegetation modifiers.

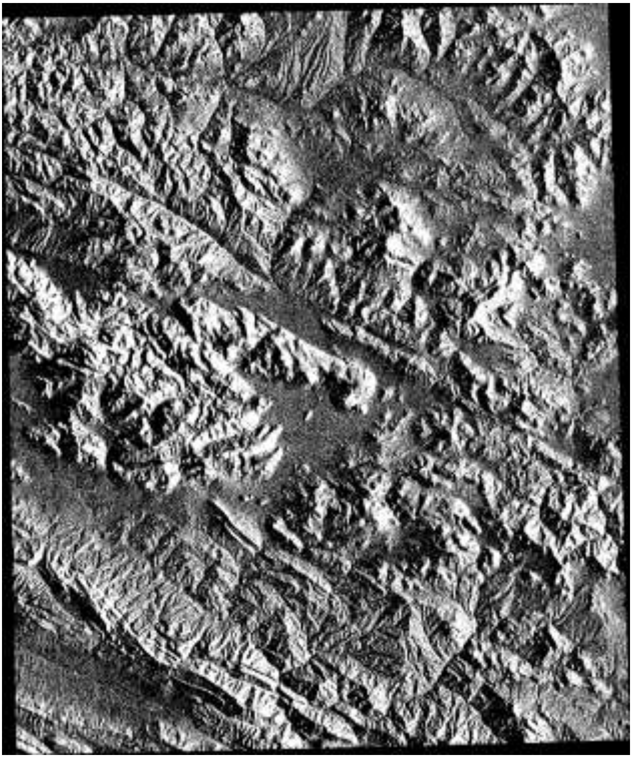

Figure 2. Zagros Terrain Aspect Map: Aspect (slope direction) map derived from the 10 m DEM for the Zagros region. Aspect provides directional information about terrain orientation, which can influence surface energy balance, moisture retention, and erosion-related processes, and was included as a supplementary morphometric variable in the modeling framework.

Figure 3. Zagros Slope Map (Degrees): Slope map generated from the 10 m DEM of the Zagros study area, expressed in degrees. This layer quantifies terrain steepness and is a primary morphometric driver controlling geomorphic instability, runoff potential, and erosion susceptibility.

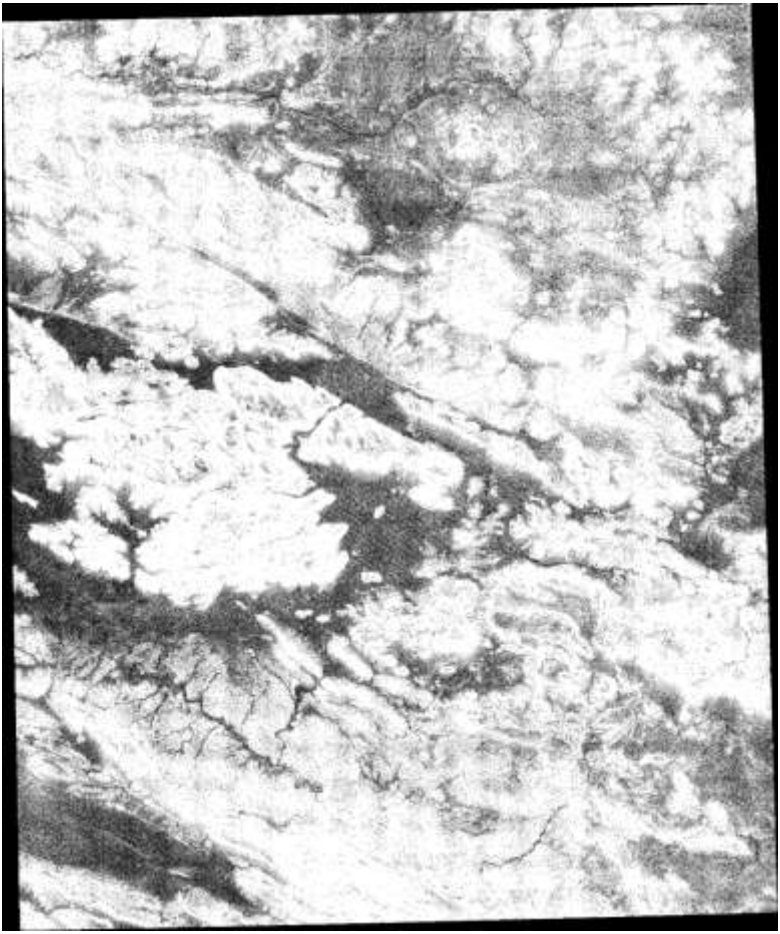

Figure 4. Zagros Terrain Ruggedness Index (TRI) Map: TRI map computed from the DEM, representing local elevation variability and surface roughness at pixel scale. TRI is one of the dominant predictors in the proposed framework and serves as a robust proxy for rugged terrain and geomorphological complexity.

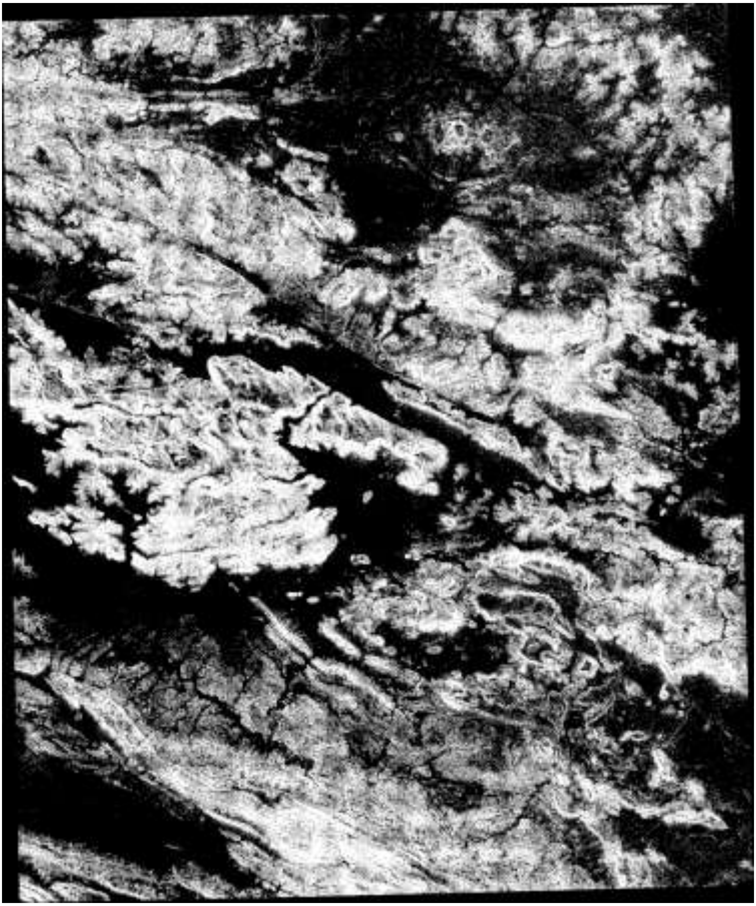

Figure 5. Zagros AI-Based K Prediction Map (K_AI): Spatial output of the geomorphometric AI model (K_AI) for the Zagros region. This map represents the probability-based machine learning prediction of ruggedness-related K behavior using morphometric predictors, prior to integration with radar and vegetation variables in the hybrid framework.

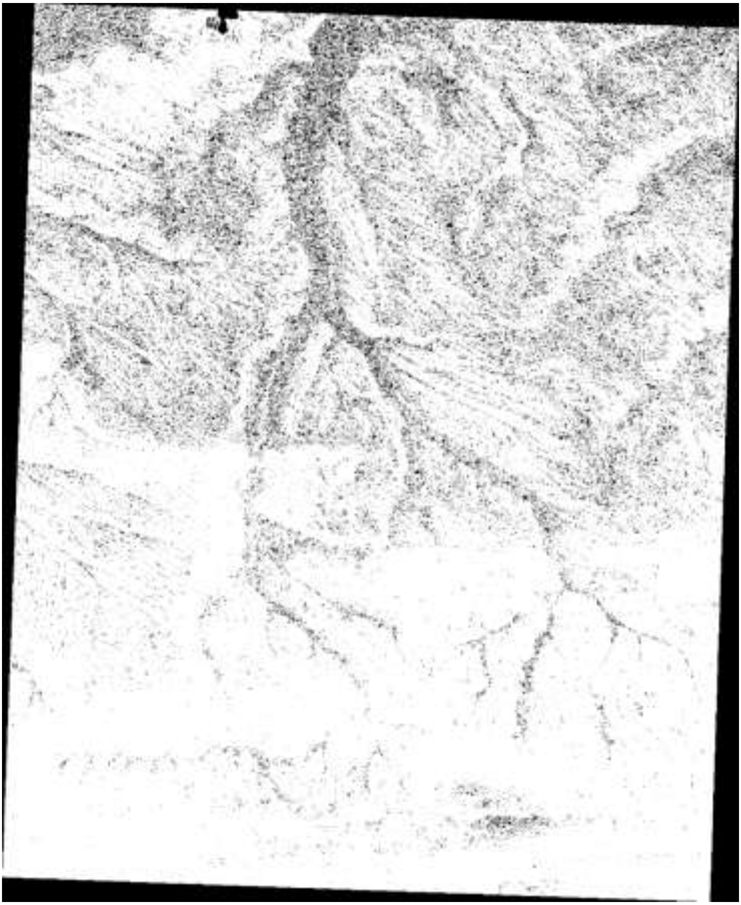

Figure 6. Moghan Slope Map (Degrees): Slope map derived from the 10 m DEM for the Moghan region. The generally low slope distribution reflects the predominantly lowland agricultural character of Moghan, explaining its overall lower predicted K values compared to rugged mountainous regions.

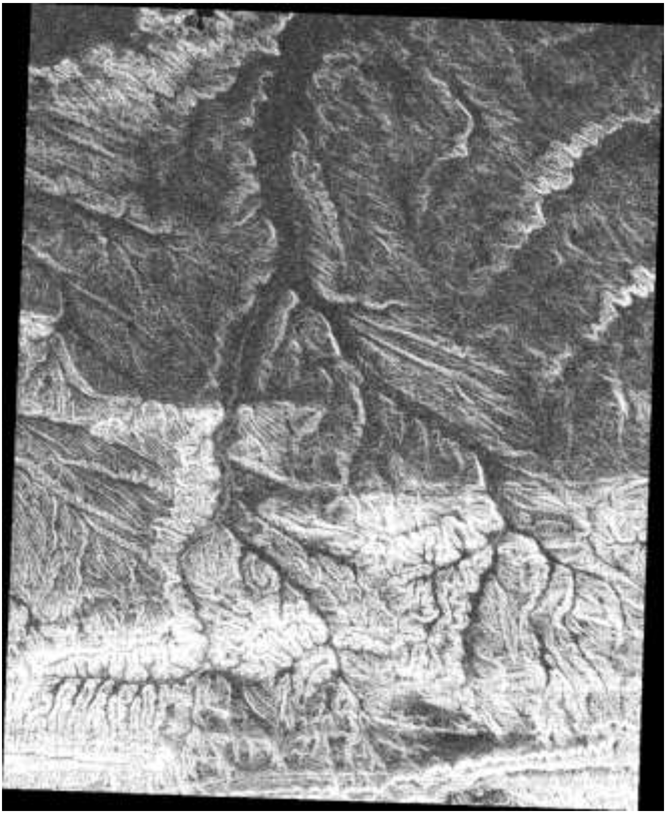

Figure 7. Moghan Baseline $K_0$ Map (Topo-Driven Index): Baseline $K_0$ index map for Moghan computed using robust-normalized slope and TRI features and sigmoid transformation. This layer provides the initial morphometric susceptibility estimate before integration with NDVI and Sentinel-1 radar backscatter.

Figure 8. Moghan Terrain Aspect Map: Aspect map extracted from the 10 m DEM for the Moghan region. This layer describes the directional orientation of terrain slopes and was used as an auxiliary morphometric descriptor within the expanded modeling experiments.

Figure 9. Moghan Sentinel-1 Radar Backscatter Map (VV): Processed Sentinel-1 VV polarization radar layer for the Moghan study area, produced through SNAP GPT terrain correction and geocoding, followed by mosaicking and compositing. This raster captures surface scattering characteristics related to roughness, soil moisture variability, and land cover structure.

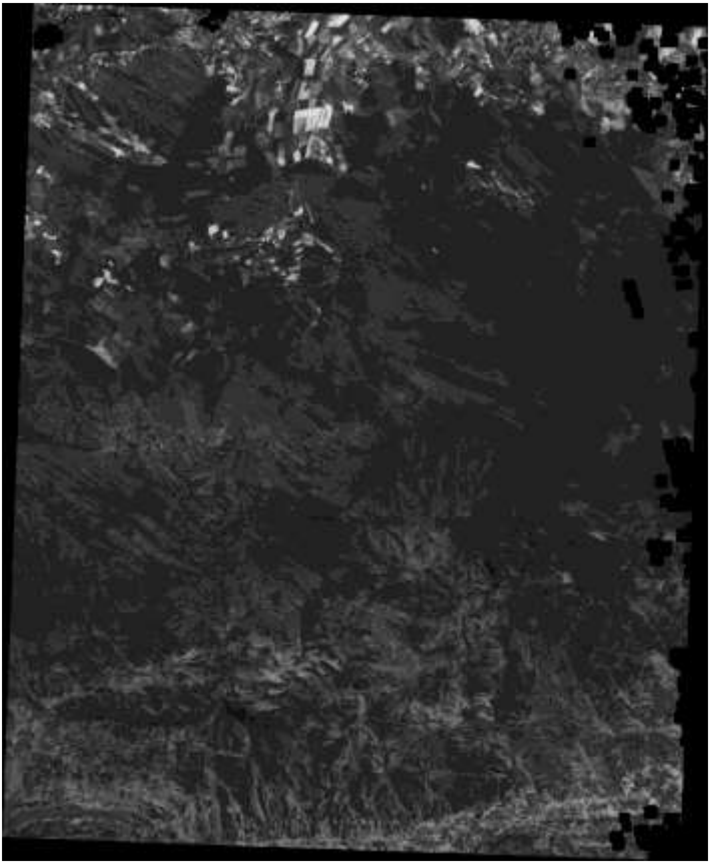

Figure 10. Moghan NDVI Map (10 m): Normalized Difference Vegetation Index (NDVI) map for Moghan, derived from Sentinel-2 red and near-infrared bands and resampled to 10 m resolution. NDVI serves as a vegetation proxy influencing erosion sensitivity and soil protection mechanisms.

Figure 11. Ravar Terrain Aspect Map: Aspect map generated from the 10 m DEM of the Ravar study area. This morphometric layer captures slope direction patterns and provides supporting terrain characterization for geomorphological interpretation.

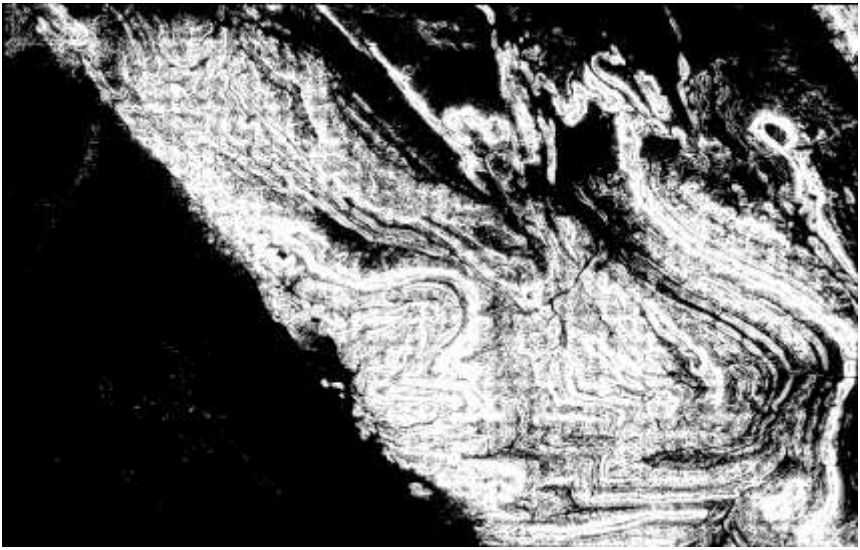

Figure 12. Ravar AI-Based K Prediction Map (K_AI): Spatial prediction of K_AI for the Ravar region based on the geomorphometric machine learning model. The output reflects terrain-driven ruggedness susceptibility patterns prior to hybrid integration with NDVI and radar features.

Figure 13. Ravar Hybrid K Coefficient Map (K_Hybrid): Final hybrid K coefficient prediction for the Ravar region, generated using the trained machine learning model integrating slope, TRI, NDVI, and Sentinel-1 VV radar backscatter. The resulting raster represents a continuous K score capturing both geomorphometric and environmental controls.

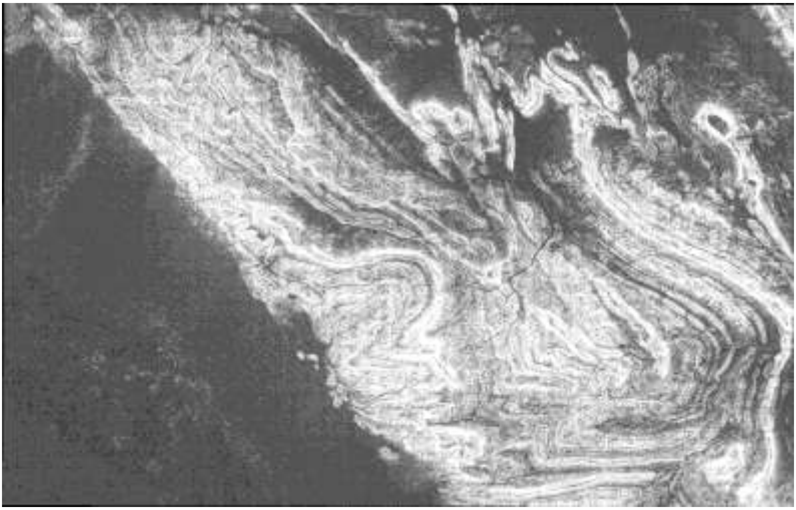

Figure 14. Ravar Baseline $K_0$ Map: Baseline $K_0$ index map for the Ravar region computed from robust-normalized slope and TRI and transformed via sigmoid scaling. This map provides a reference morphometric-only susceptibility estimate for comparison with the hybrid model output.

Figure 15. Ravar Slope Map (Degrees): Slope map derived from the 10 m DEM of the Ravar study area. This layer quantifies terrain steepness and acts as a key input variable controlling the topographic component of K variability.

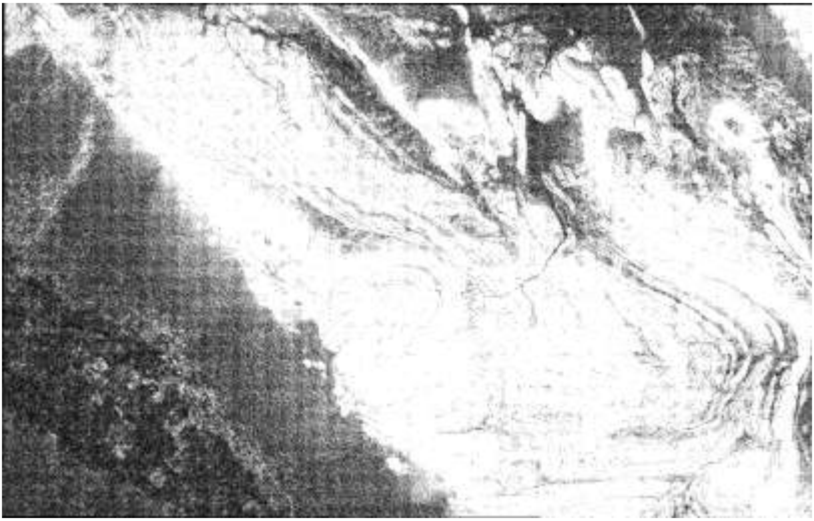

Figure 16. Ravar TRI Map: TRI raster for the Ravar region, representing local elevation heterogeneity and micro-topographic roughness. This index provides strong explanatory power in both baseline and machine learning K prediction stages.

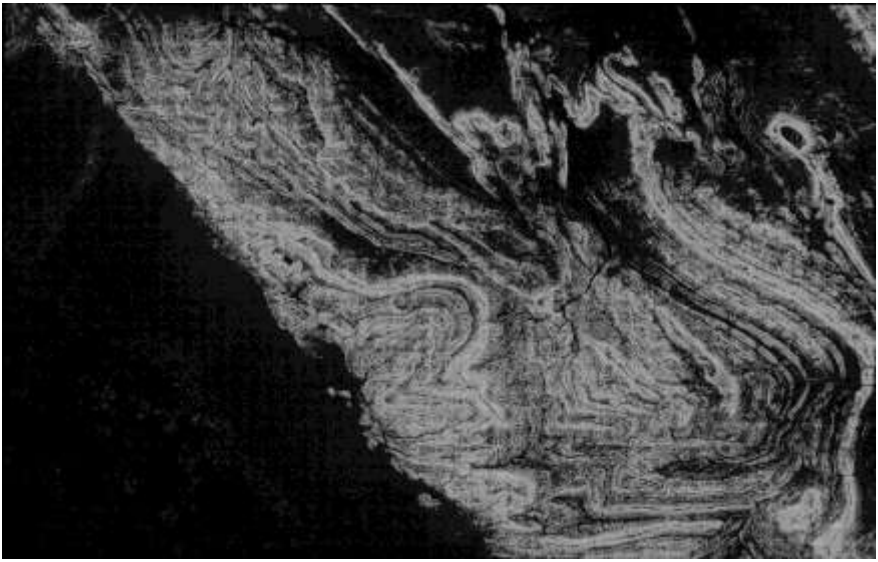

Figure 17. Ravar Physical-Range Standardized K Map [0.01–0.65]: Standardized physical-range K map for Ravar, produced by transforming the continuous hybrid K coefficient output into the fixed interval [0.01, 0.65]. This normalization ensures consistent interpretability and comparability across different regions.

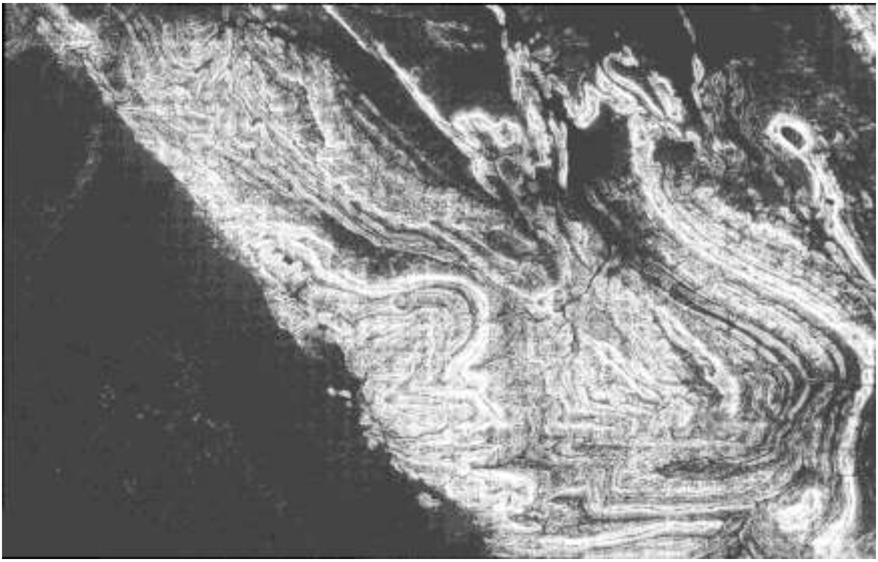

Figure 18. Full-Model K Prediction Map (All Features Integrated): Spatial prediction raster generated by the full multi-feature machine learning model, integrating morphometric terrain layers, radar backscatter, NDVI, and human-related proxy variables. This map represents the most comprehensive predictive output of the framework.

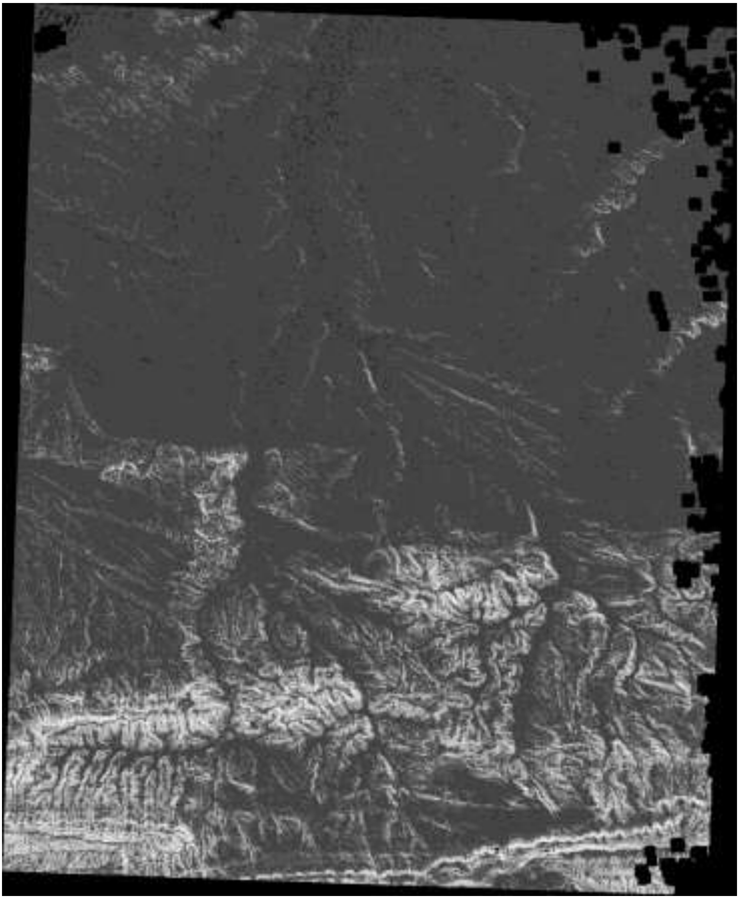

Figure 19.Moghan Final K Coefficient Map: Final continuous K coefficient prediction map for Moghan, produced using the hybrid modeling framework. The spatial distribution highlights the dominance of lowland behavior, with limited high-K zones corresponding to localized roughness or terrain transitions.

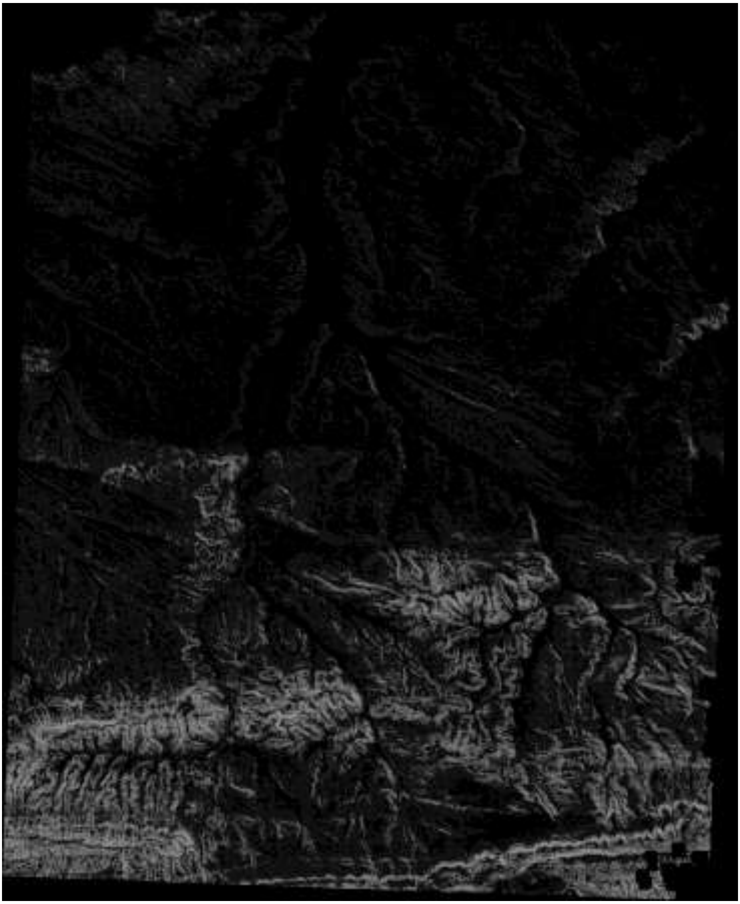

Figure 20. Moghan Physical-Range Standardized K Map [0.01–0.65]: Physical-range standardized K output for Moghan, scaled into the interval [0.01, 0.65]. This product enables direct quantitative comparison with the standardized outputs of Zagros and Ravar.

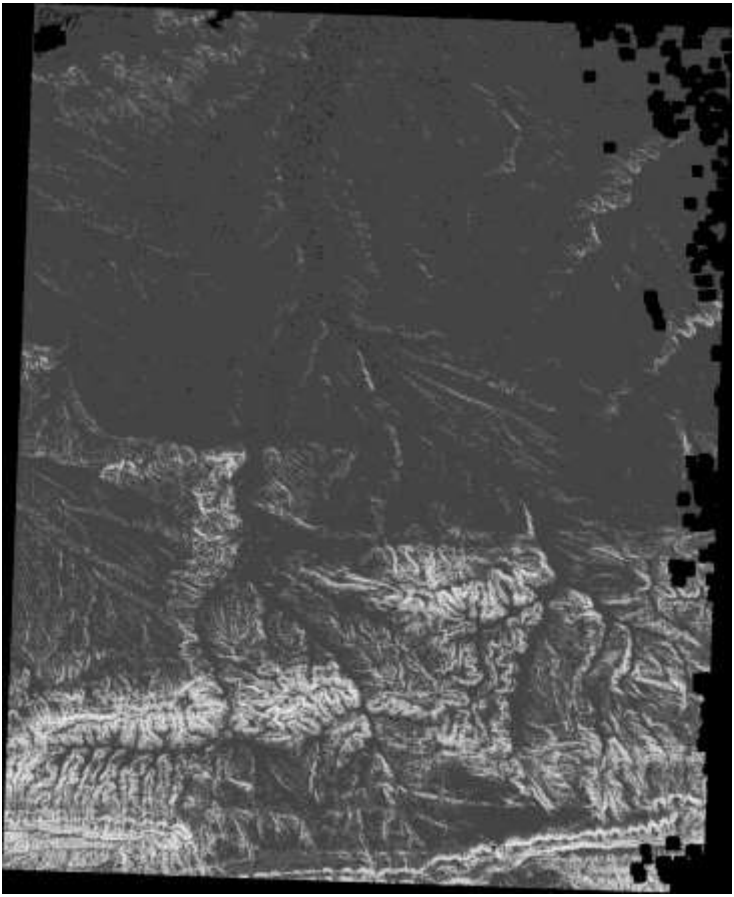

Figure 21. Full-Model K Prediction Map (Duplicate Output / Alternative Run): Full-model predicted K map produced from an alternative execution or repeated modeling run. This output is included to confirm reproducibility and numerical stability of the prediction pipeline under consistent preprocessing conditions.

Figure 22. Zagros Physical-Range Standardized K Map [0.01–0.65]: Physical-range standardized K raster for the Zagros region. The hybrid coefficient output was transformed into a fixed physical range [0.01, 0.65] to provide a comparable metric across regions and support consistent interpretation of high- and low-susceptibility zones.

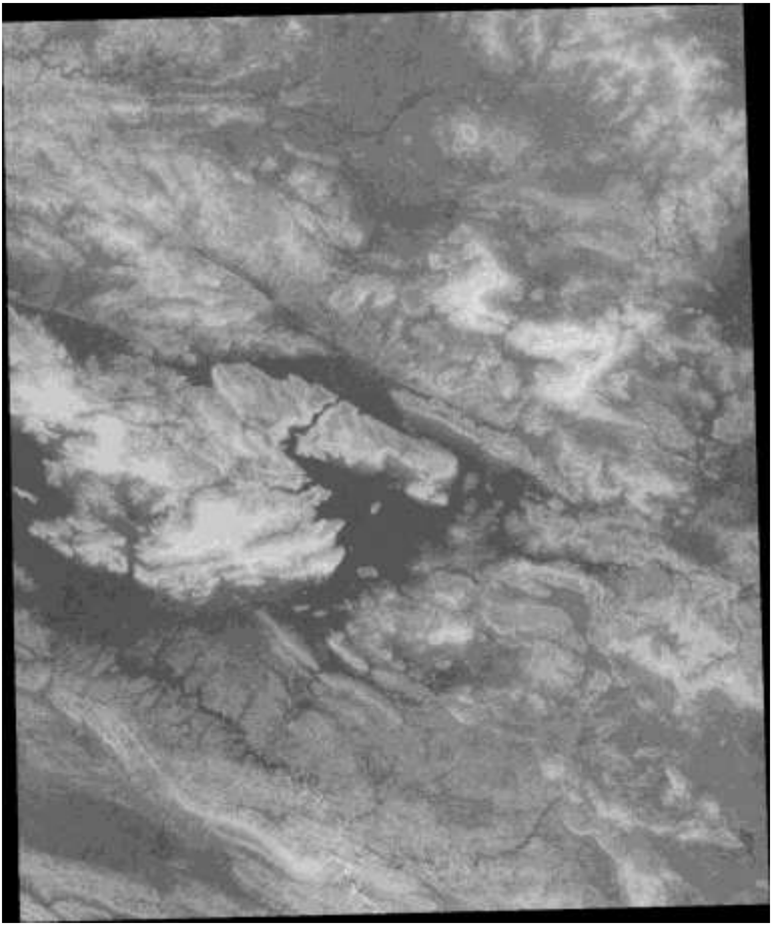

Figure 23. Zagros Full-Model Predicted K Map: Predicted K coefficient map for the Zagros region generated by the full-feature machine learning model. This raster integrates morphometric terrain descriptors with remote sensing variables and represents the most complete modeled K estimate for the Zagros domain.

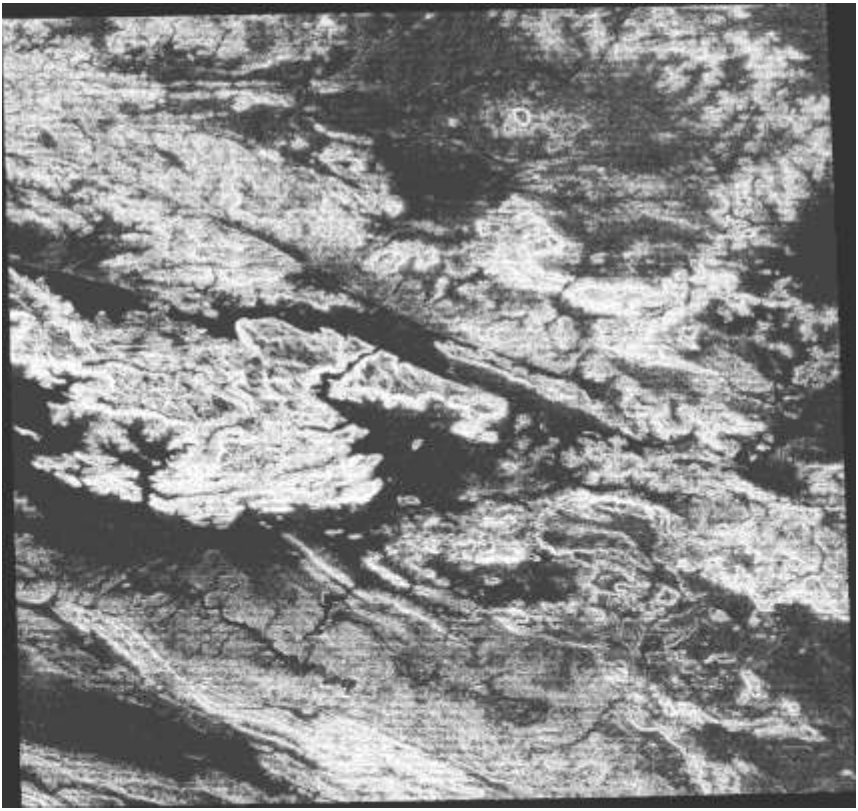

Figure 24. Zagros K Coefficient Reference / Hybrid Output Map: Final K coefficient map for the Zagros study area, representing the continuous modeled output of the framework. This layer serves as a primary reference for analyzing model transferability and assessing the distinct geomorphological behavior of Zagros relative to Moghan and Ravar.

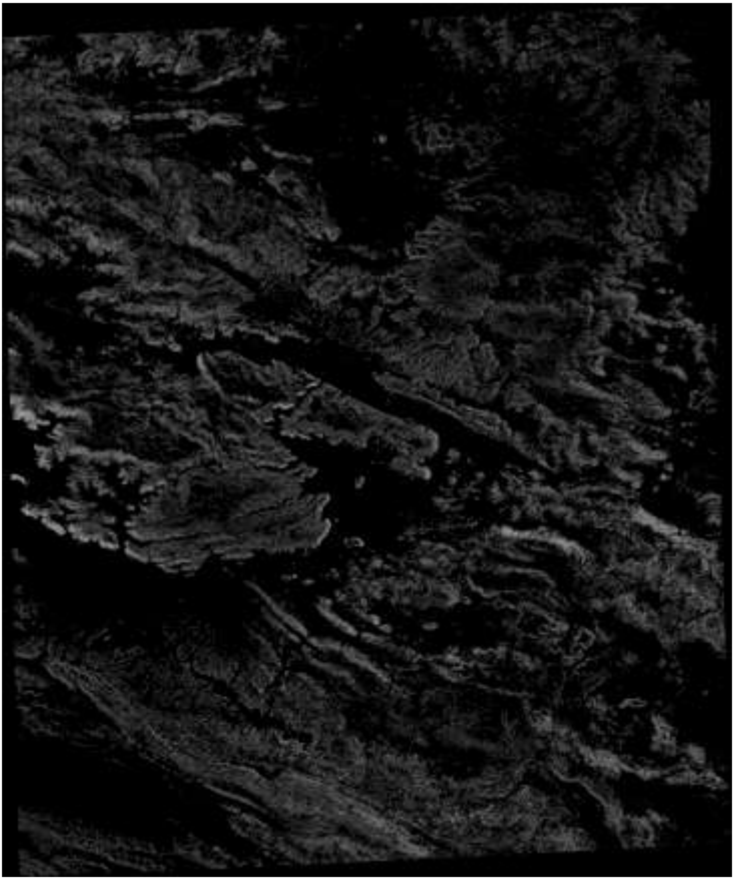

Figure 25. Zagros Residual Map (True K − Predicted K): Residual raster computed as the difference between the reference K map and the predicted K output for the Zagros region (Residual = K_true − K_pred). Positive values indicate underestimation by the model, while negative values indicate overestimation, revealing spatial patterns of model bias potentially linked to missing lithological or karst-related drivers.

## References

[1] Riley, S. J., DeGloria, S. D., & Elliot, R. (1999). A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences, 5*(1-4), 23-27.

[2] Moore, I. D., Grayson, R. B., & Ladson, A. R. (1991). Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrological Processes, 5*(1), 3-30.

[3] Wilson, J. P., & Gallant, J. C. (2000). *Terrain Analysis: Principles and Applications*. John Wiley & Sons. (Context for the limitations of geometric descriptors and the need for compositeindices).

[4] Filipponi, F. (2019). Sentinel-1 GRD Preprocessing Workflow. *Proceedings, 18*(1), 11.

[5] Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18-27.

[6] Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment, 8*(2), 127-150.

[7] European Space Agency (ESA). (n.d.). *SNAP – Sentinel Application Platform*. [Online]. Available: http://step.esa.int.

[8] Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32.

[9] Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics, 12*(1), 55-67.

[10] Ke, G., et al. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*.

[11] Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830.

[12] Agard, P., et al. (2005). Convergence history across Zagros (Iran): Constraints from collisional and earlier deformation. *International Journal of Earth Sciences, 94*, 401-419.

[13] Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography, 40*(8), 913-929.

[14] Zeraatpisheh, M., et al. (2019). Digital mapping of soil properties using machine learning framework in an arid region. *Earth Science Informatics*.

[15] Alaska Satellite Facility (ASF) DAAC. (n.d.). *Sentinel-1 SAR Ground Range Detected (GRD) products*.

[16] NASA AppEEARS. (n.d.). *Extraction of HLS and Sentinel-2 optical products*.

[17] United States Geological Survey (USGS). (n.d.). *Digital Elevation Model (DEM) data sources*.