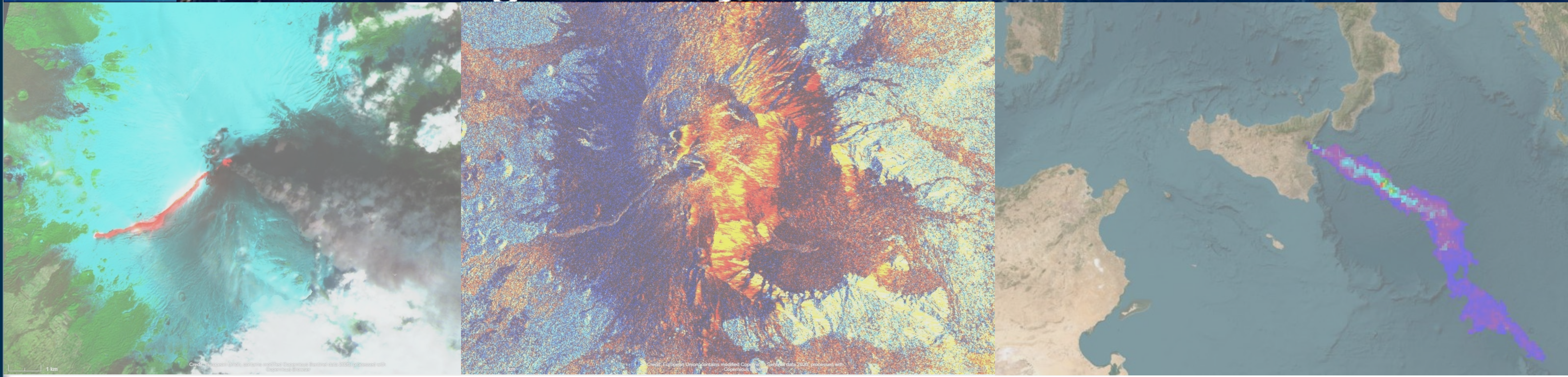
INGV Istituto Nazionale di Geofisica e Vulcanologia



Filling the Multisensor Gap: Multimodal Generative Models for Volcano Monitoring

Simona Cariello [1], Claudia Corradino [1], Giovanni S. Di Bella [1, 2] & Ciro Del Negro [1]

[1] Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Etneo – Sezione di Catania, Piazza Roma 2, 95125 Catania, Italy
[2] Dipartimento di Ingegneria Elettrica, Elettronica e Informatica, Università di Catania, Via Santa Sofia 64, 95123 Catania, Italy

TechnoLab

simona.cariello@ingv.it

INGV Istituto Nazionale di Geofisica e Vulcanologia

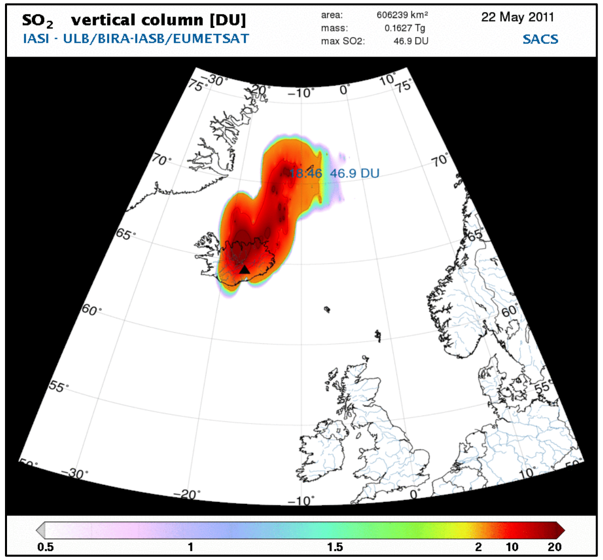
Volcano Monitoring

Volcanic phenomena evolve rapidly; continuous monitoring is essential to understand activity and forecast eruptions.

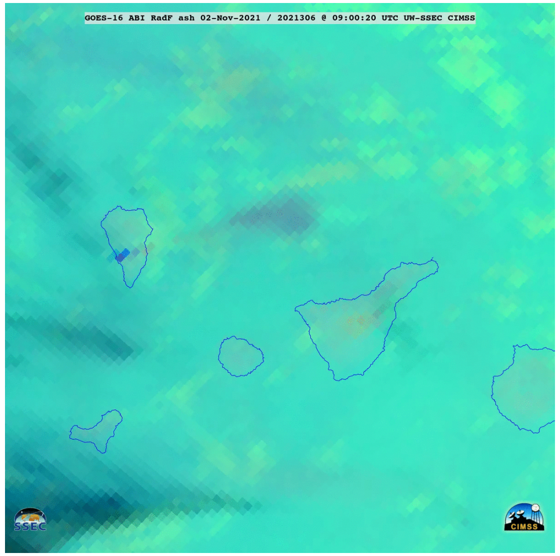
Volcano monitoring aims to detect early signals of activity by observing:

- Thermal anomalies
- Ground deformation
- Surface and morphological changes
- Gas and ash emissions

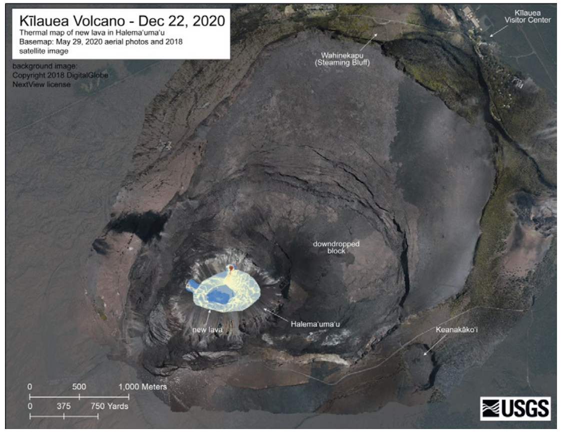
GAS EMISSIONS



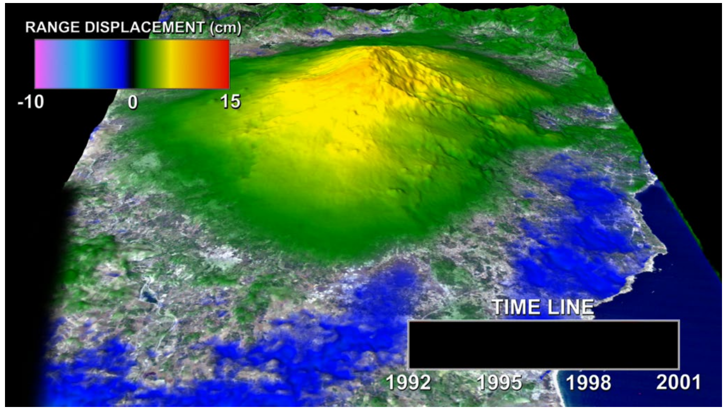
ASH EMISSIONS



THERMAL EMISSIONS



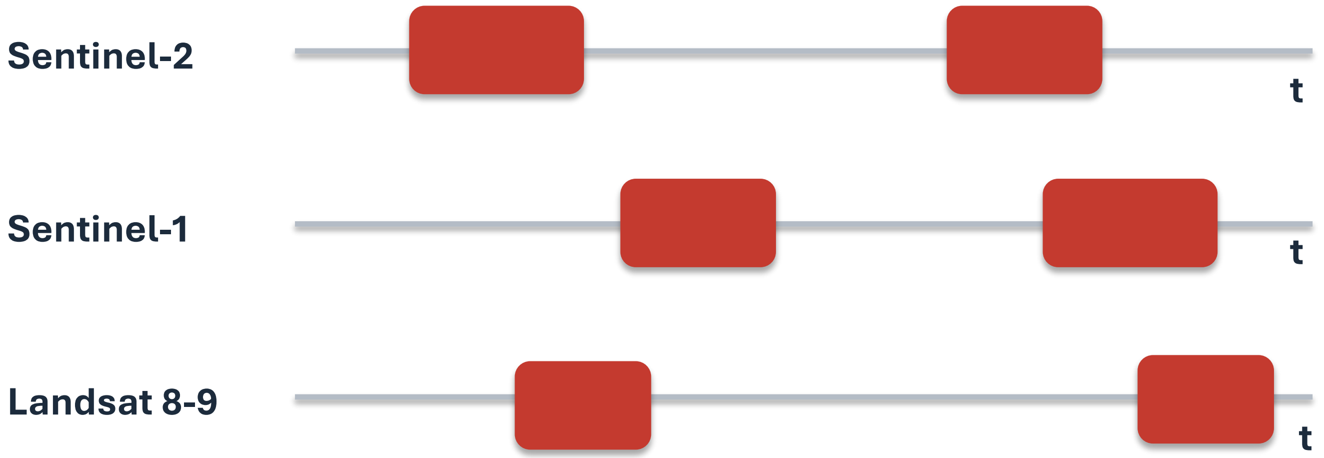
GROUND DEFORMATION



The Multisensor Temporal Gap

Complementary satellite missions rarely observe the volcano at the same time.

- Sentinel-2 and Landsat 8/9 provide optical and thermal observations
- Sentinel-1 provides SAR observations on a different acquisition schedule

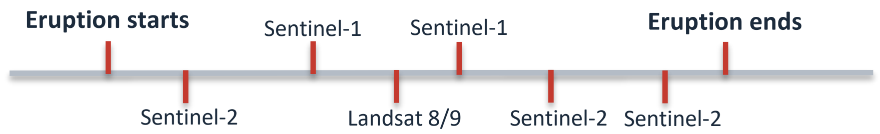


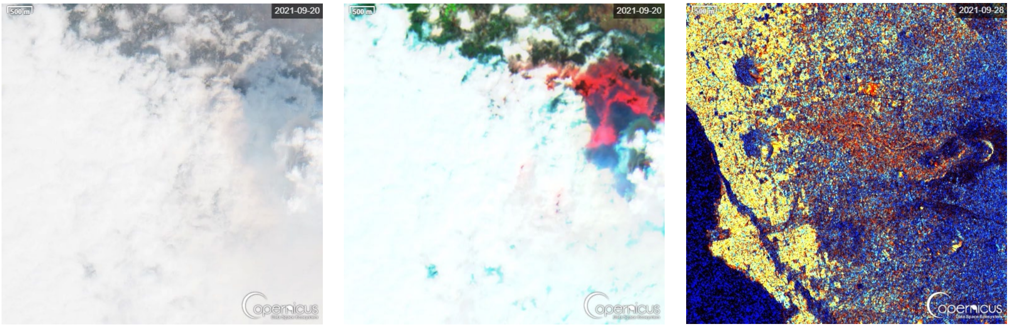
Asynchronous acquisitions create critical information gaps

Why the Gap Matters in Operational Volcanology

Volcanic activity evolves rapidly and may occur between satellite overpasses. Temporal mismatches between sensors complicate interpretation during critical phases of an eruption.

- Sentinel-2 captures the eruption and provides optical information on the volcanic surface.
- Sentinel-1 acquisitions may not be available at the same time
- Without SAR data, lava deposits cannot be reliably mapped using radar change detection.
- By generating synthetic Sentinel-1 from Sentinel-2 imagery, we can compute post–pre differences and recover key information for lava deposit mapping.





Recovering missing SAR observations could significantly improve volcanic change detection.

Objective

Can we reconstruct missing sensor observations from the modalities that are available?

Available modalities → Multimodal model → Synthetic missing observation

Multimodal Models for Earth Observation

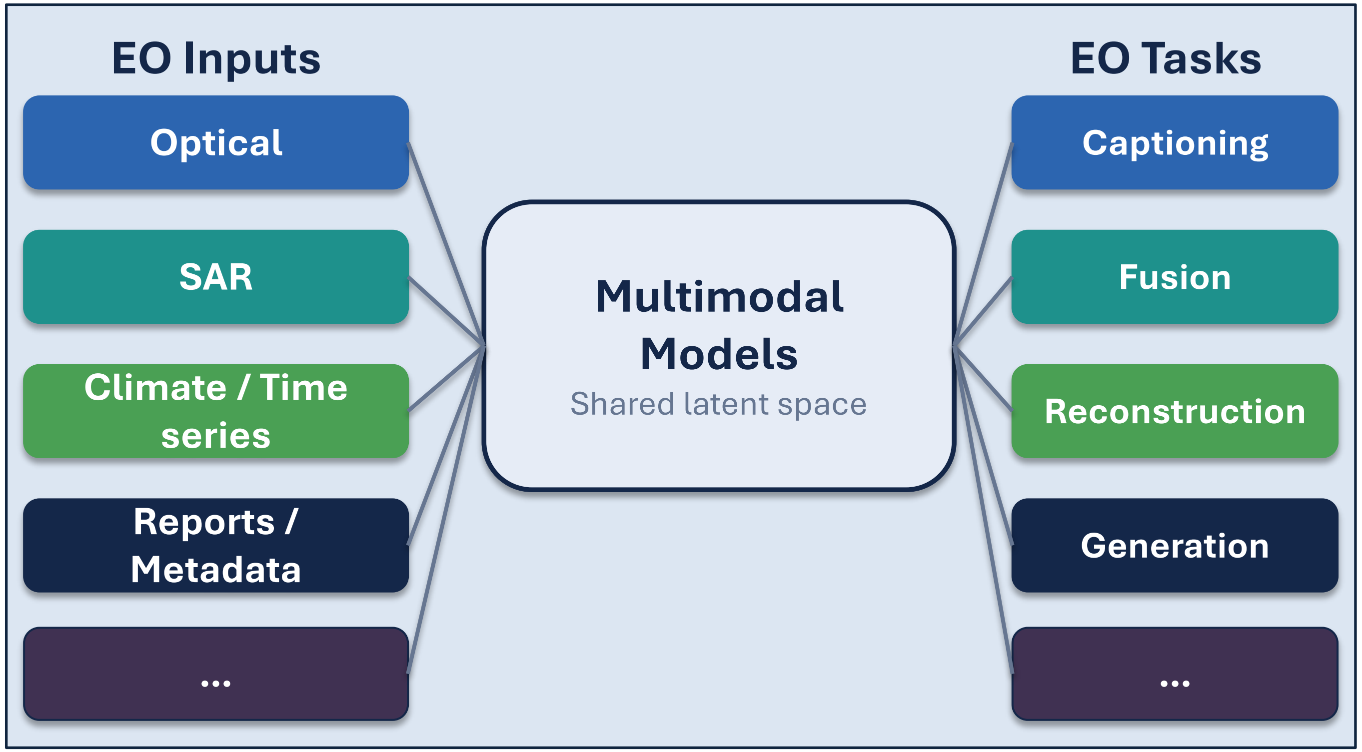
Definition — Models that learn a shared representation across heterogeneous modalities and can reason or generate across them.

Typical EO modalities — Satellite imagery (optical, SAR), time series, geospatial metadata, climate variables, textual reports.

Core architecture — Modality-specific encoders → fusion/shared latent space → generative decoder.

What they enable — Cross-modal understanding, sensor fusion, missing-data reconstruction, synthetic generation, and grounded analysis.

Why EO is a natural fit — Earth data are inherently multimodal, multi-sensor, and multi-scale, so joint modeling improves completeness and robustness.

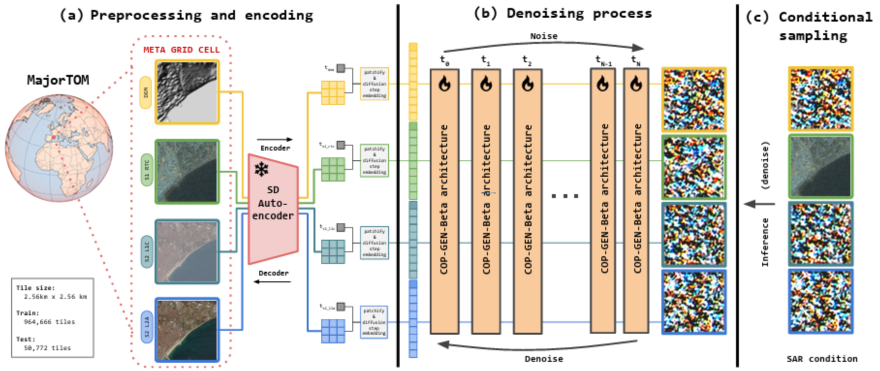


Multimodal EO Models used

Two recent approaches illustrate different directions in multimodal Earth observation modeling.

COP-GEN Beta

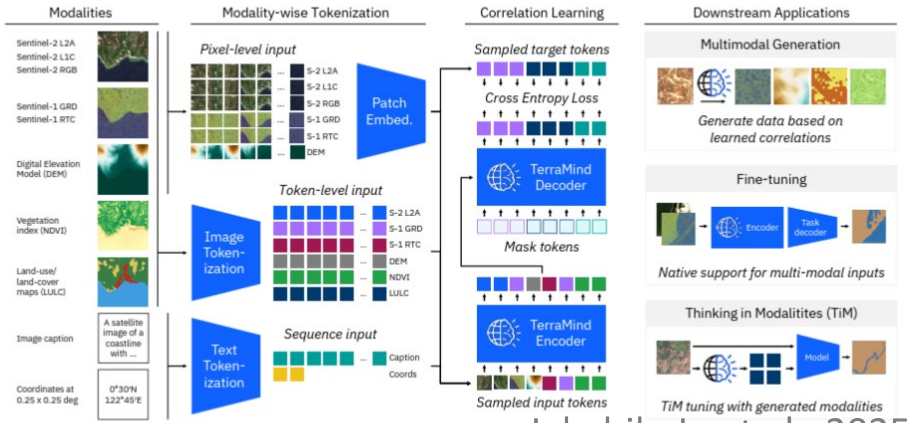
- Diffusion-based multimodal generative model for EO
- Cross-modal satellite image generation
- Training dataset: Major TOM



Espinosa, M., et al., 2025

TerraMind

- Multimodal geospatial foundation model
- Shared multimodal EO representation
- Any-to-any multimodal EO generation and reasoning
- Training dataset: TerraMesh

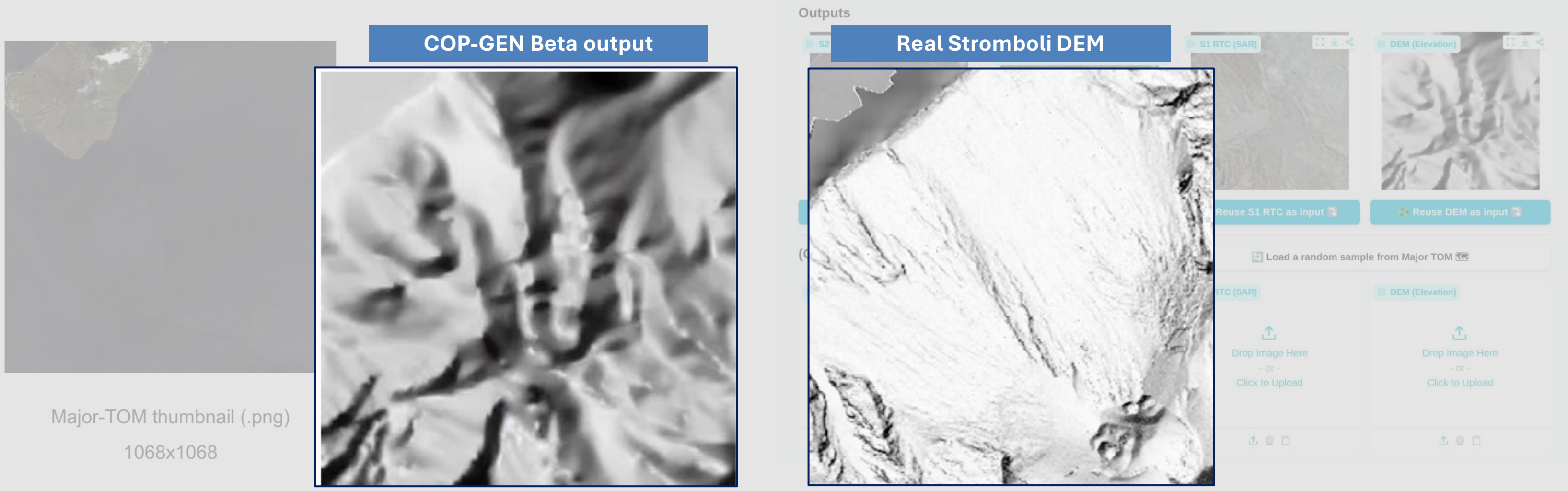


Jakubik, J., et al., 2025

Key difference: COP-GEN focuses on multimodal image generation, while TerraMind is a multimodal foundation model for EO representation learning

COP-GEN Beta

COP-GEN Beta tested using the official online tool, the input images prepared following Major-TOM format thumbnail and then cropped to 256 × 256 pixels.



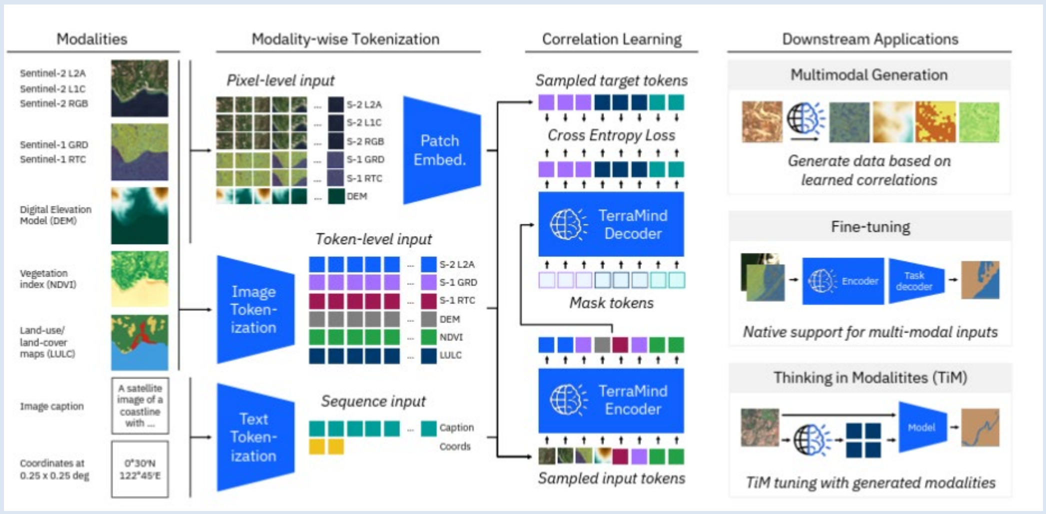
Input Major-TOM-like crop 256×256 → COP-GEN Beta → Generated Outputs (S2 / S1 / DEM)

Key observation: Sentinel-1 and DEM outputs contained little or no meaningful volcanic information.

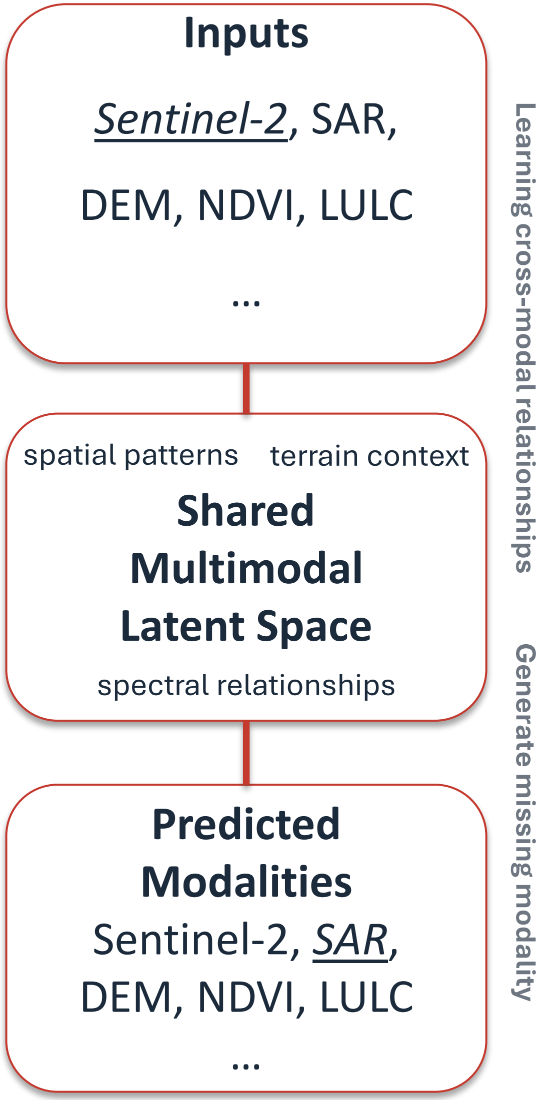
The model remained strongly tied to the Major-TOM training domain.

TerraMind

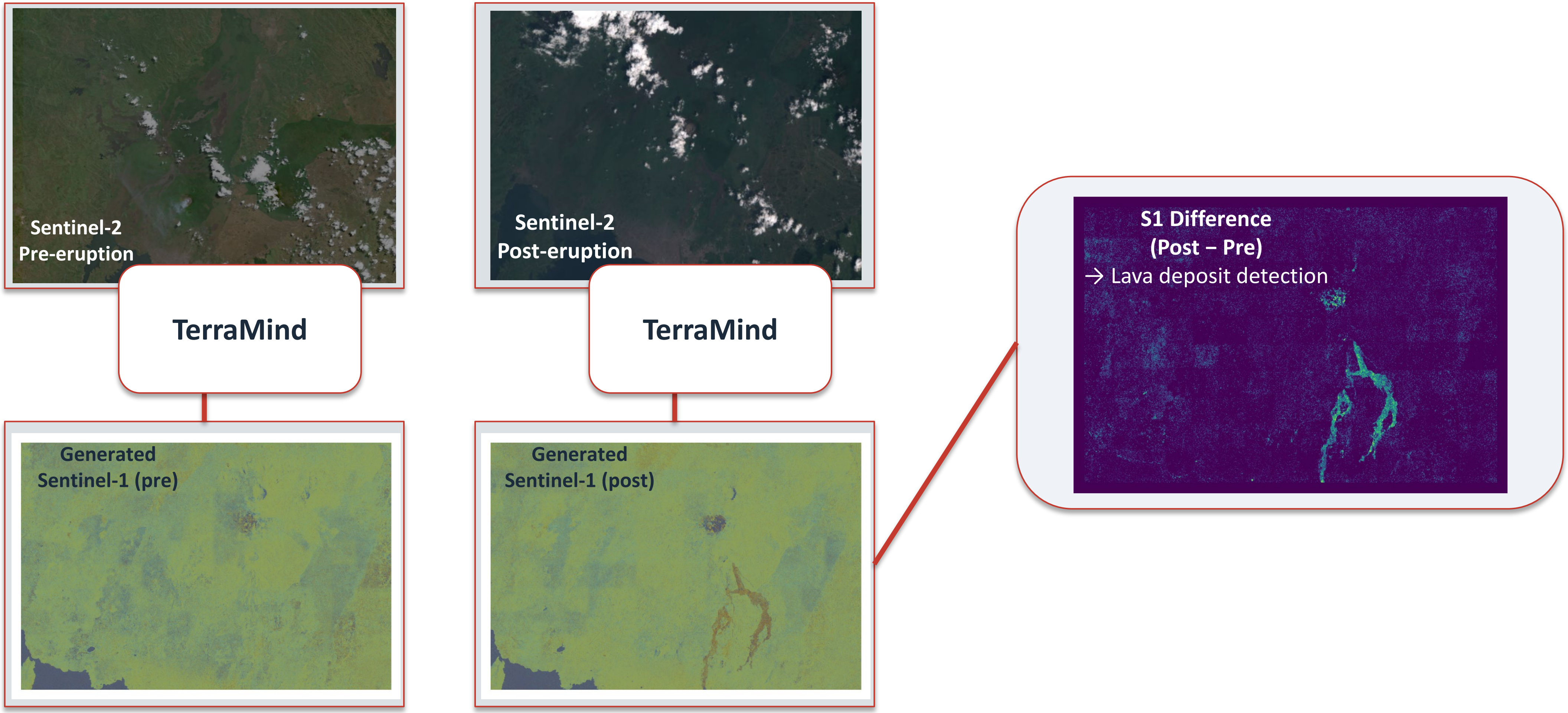
- Multimodal geospatial foundation model
- **Shared multimodal EO representation**
- **Any-to-any multimodal EO generation and reasoning**
- Training dataset: TerraMesh
- **Captures both spatial detail (pixel-level) and semantic relationships (token-level)**



Jakubik, J., et al., 2025

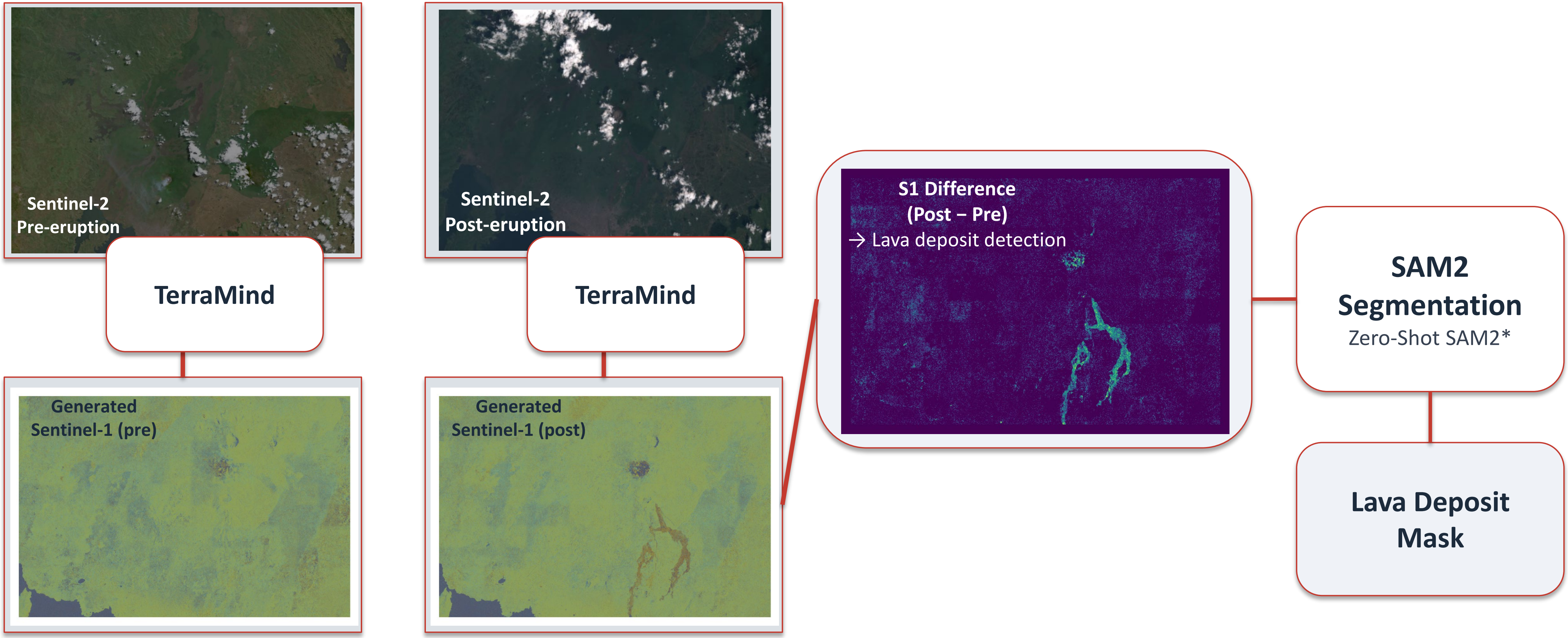


TerraMind Workflow for Volcanic Change



Synthetic SAR differences highlight newly emplaced lava deposits.

TerraMind Workflow for Volcanic Change

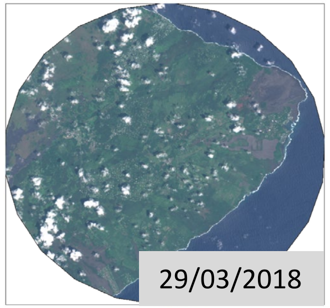


Synthetic SAR differences highlight newly emplaced lava deposits.

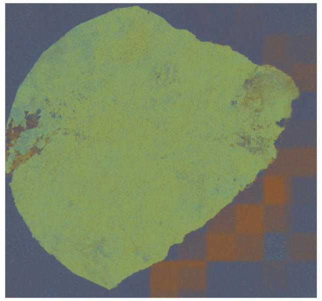
*[*Cariello et al., 2026 (Under review)]*

Results Pu’u ‘O’o Rift Eruption, Hawaii, May 2018

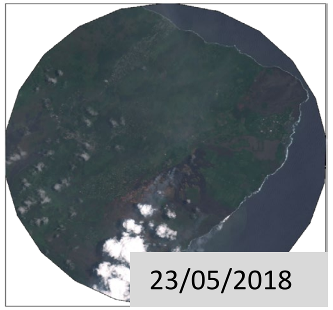
Sentinel 2 – Pre Eruption



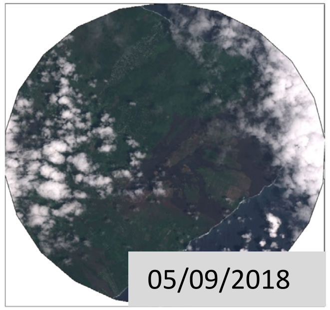
TerraMind



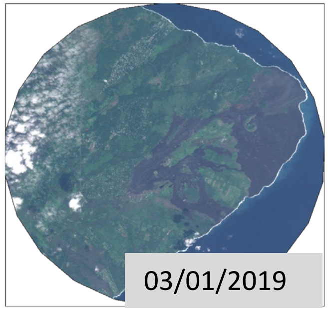
Sentinel 2 – Post Eruption



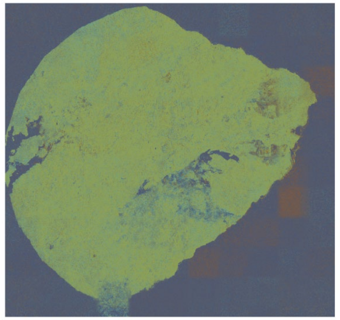
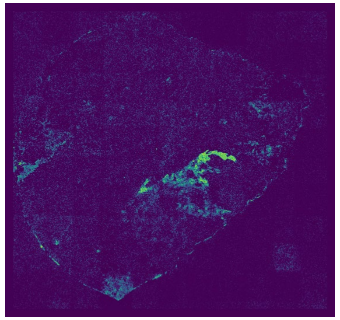
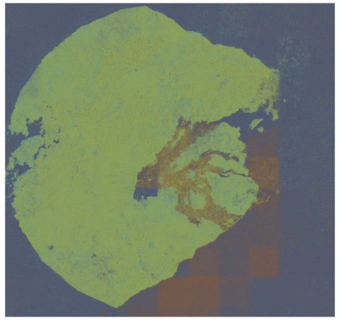
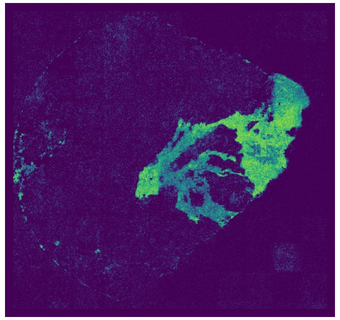
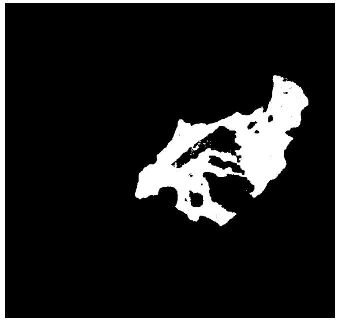
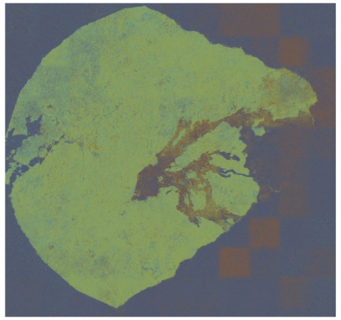
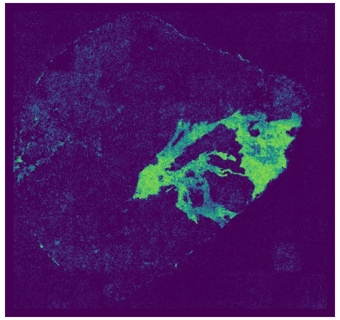
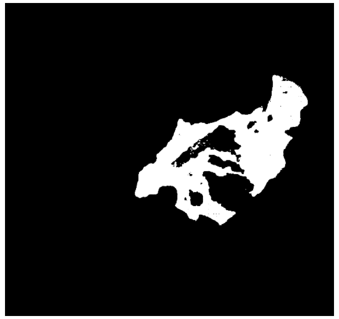
TerraMind



TerraMind

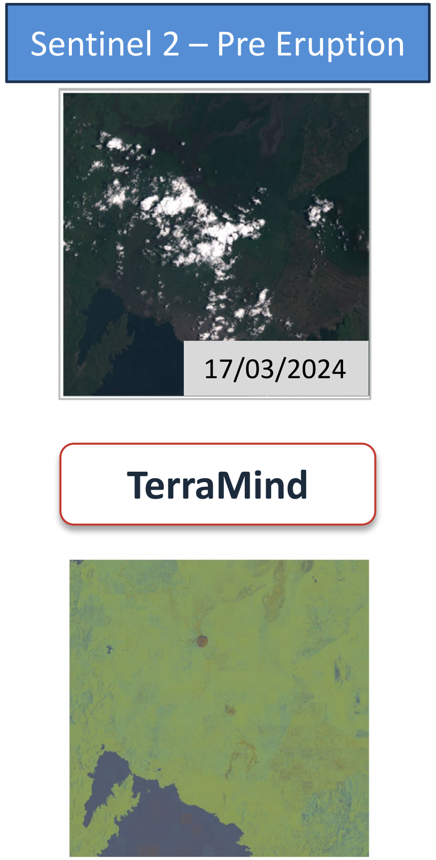


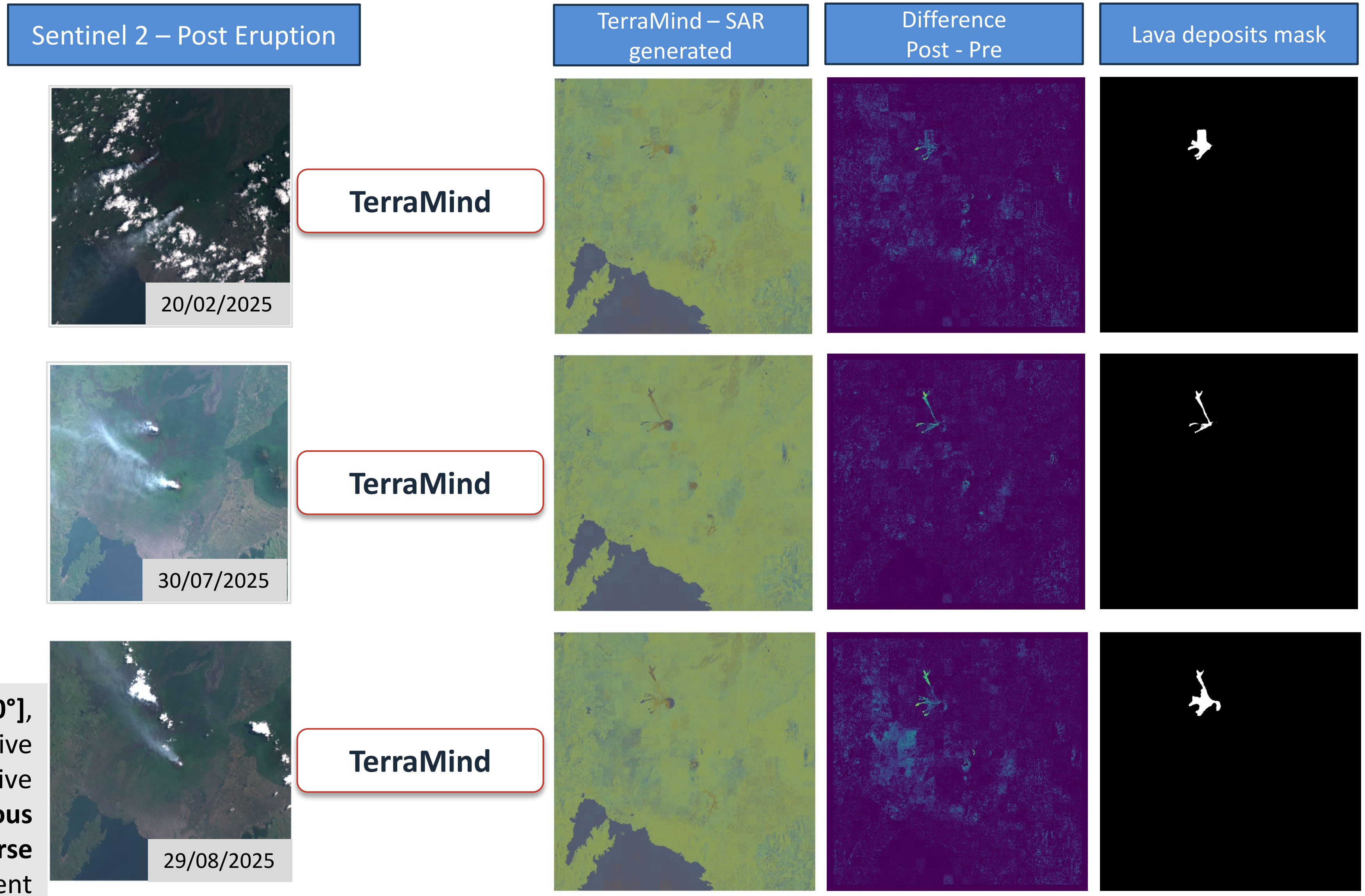
TerraMind

TerraMind – SAR generated	Difference Post - Pre	Lava deposits mask
		
		
		

The slope ranges from **[~2°–6°]**, continuous effusive activity producing **spatially extensive and morphologically homogeneous** basaltic lava fields, with **high vegetation cover** in areas not recently resurfaced due to rapid tropical ecological succession.

Results Nyamuragira Eruption, Democratic Republic of Congo



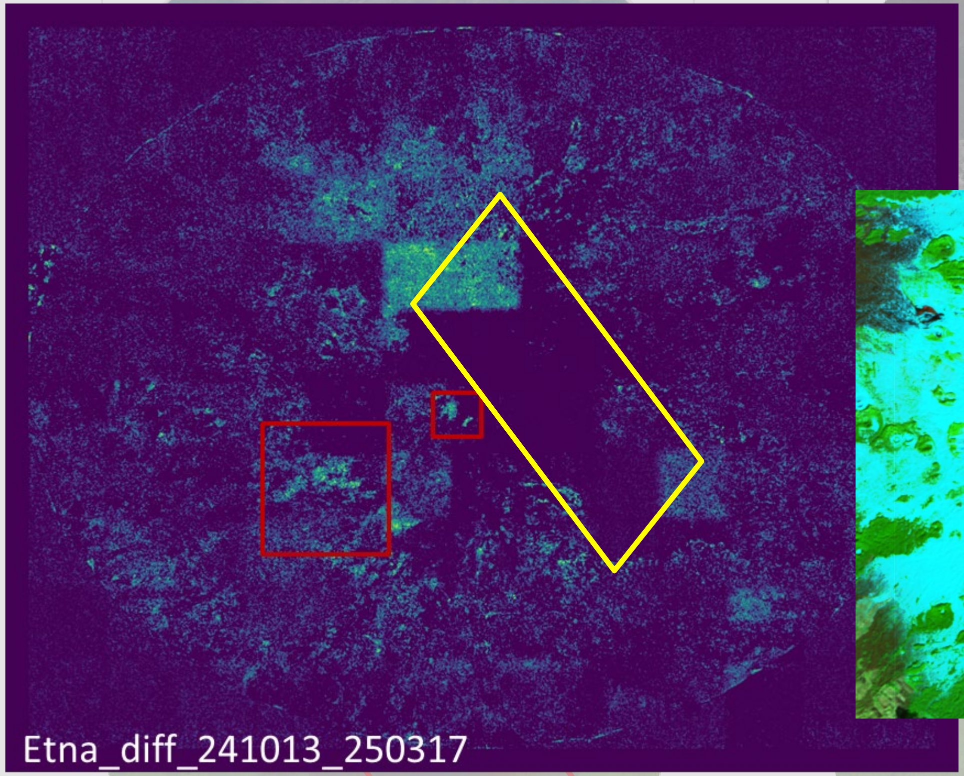


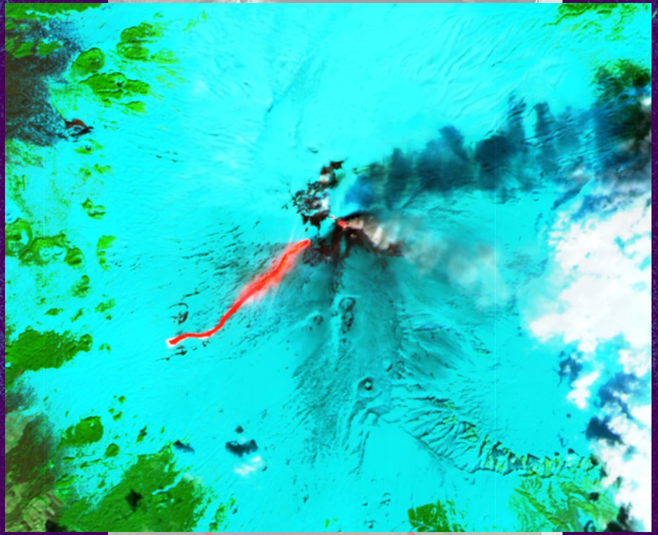
The slope ranges from **[~3°–10°]**, episodic but voluminous effusive activity producing spatially extensive and **morphologically homogeneous** basaltic lava fields, with **sparse vegetation** due to frequent resurfacing.

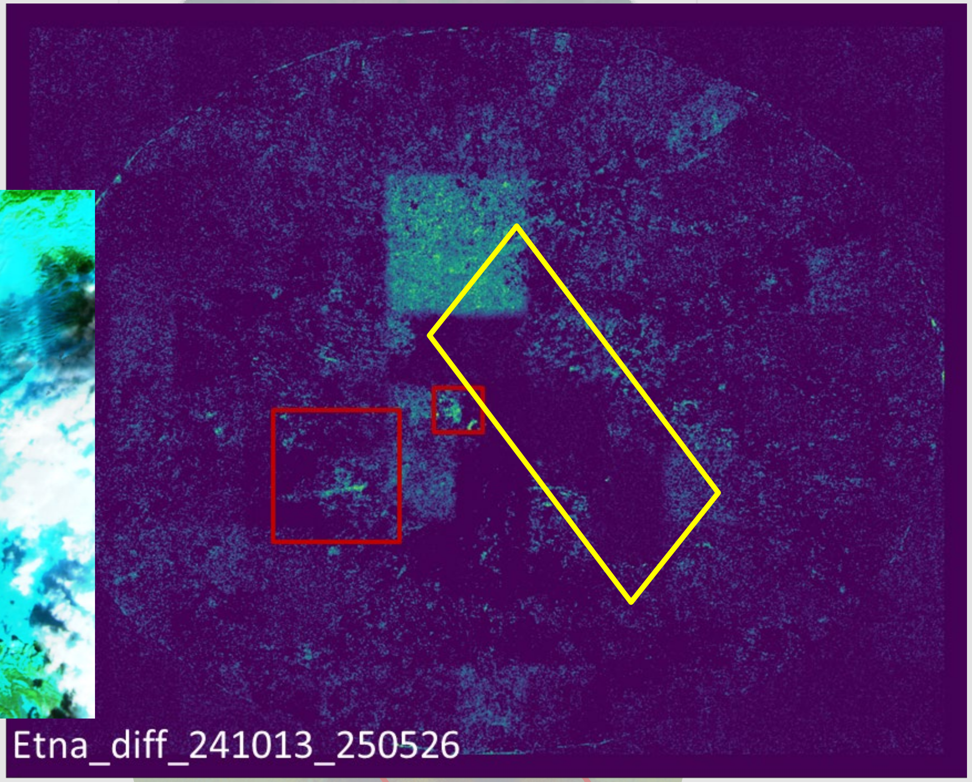
When this approach it does not work?

Targets for which SAR is more challenging and less straightforward to interpret due to topographic and morphological complexity.

Etna





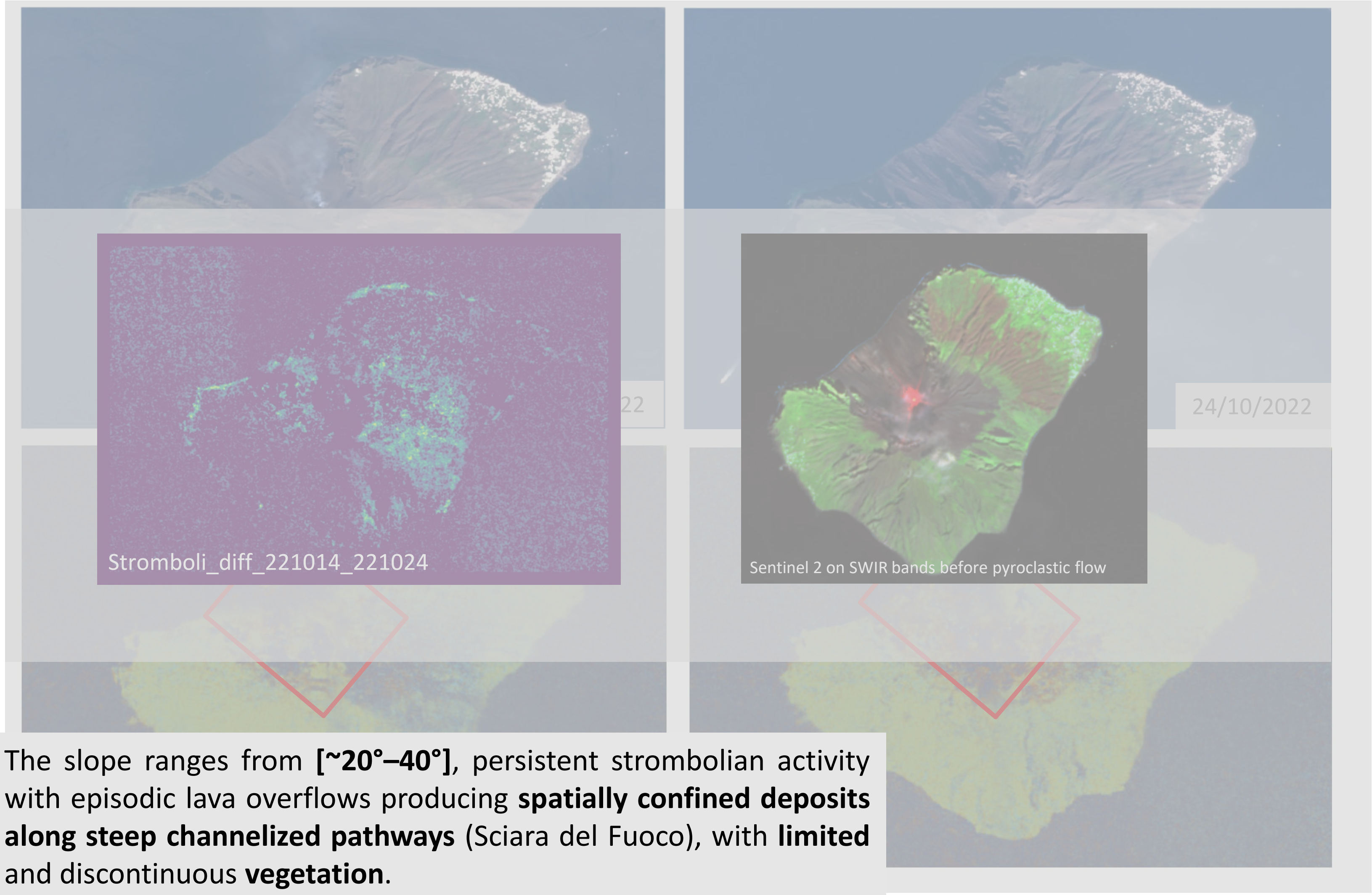


The slope ranges from **[~5°–30°]**, frequent effusive activity producing **spatially heterogeneous lava flows and mixed pyroclastic–lava deposits**, with **vegetation** distribution **strongly controlled by elevation** and eruptive history.

When this approach it does not work?

Targets for which SAR is more challenging and less straightforward to interpret due to topographic and morphological complexity.

Stromboli



The slope ranges from **[~20°–40°]**, persistent strombolian activity with episodic lava overflows producing **spatially confined deposits along steep channelized pathways** (Sciara del Fuoco), with **limited** and discontinuous **vegetation**.

- **Observed limitation:**
  Generated SAR are less reliable in steep volcanic environments.

- **Possible explanation:**
  - Cone-shaped morphology introduces strong slope asymmetry.
  - Ascending and descending SAR acquisitions illuminate different slope sectors.
  - Areas visible in one orbit direction may appear shadowed or geometrically distorted in the other.

To sum up…

Volcano	Slope range	Morphology	Vegetation	Surface roughness (SAR)	Backscatter heterogeneity	Temporal coherence	Geometric distortion risk
Pu′ u ′O′o′, Kīlauea	~2°–6°	Extensive homogeneous basaltic lava fields	High (rapid tropical regrowth)	Low–moderate (pāhoehoe smooth / ‘a‘ā rough)	Low–moderate (large uniform lava units)	High on old flows, low on active rift	Very low
Nyamuragira	~3°–10°	Extensive homogeneous basaltic flows	Sparse (frequent resurfacing)	Moderate (blocky basalt + flow textures)	Moderate (episodic resurfacing patches)	Moderate–high on young flows	Low
Mount Etna	~5°–30°	Heterogeneous lava + pyroclastic deposits	Variable (altitude-controlled)	High (blocky lava, scoria, mixed textures)	High (multi-event overlapping flows)	Low–moderate (rapid surface changes)	Moderate–high
Stromboli	~20°–40°	Confined lava channels (Sciara del Fuoco)	Limited, discontinuous	High (blocky lava + tephra + cliffs)	Moderate (localized but persistent activity)	Low (rapid local changes in confined area)	Very high (shadow + layover dominant)

Conclusions

- Multimodal models can help reconstruct missing EO observations and reduce multisensor temporal gaps.
- TerraMind produced more informative cross-modal predictions than COP-GEN Beta in volcanic monitoring scenarios.
- Synthetic SAR differences can highlight volcanic surface changes and lava emplacement.
- Combining generative models with foundation models (e.g., SAM2) enables automatic extraction of volcanic deposits.
- Model performance is reliable in moderate-slope volcanic environments.
- Steep volcanic topography (e.g., Etna, Stromboli) introduces limitations due to SAR geometric effects and acquisition geometry.
- These results suggest that multimodal generative models could support future operational volcanic monitoring systems.

THANKS

Find out more
about our research:

*ct.ingv.it/**technolab**/*

Simona Cariello

simona.cariello@ingv.it