SAOCOM InSAR DEM Validation

A Land Cover-Stratified Analysis

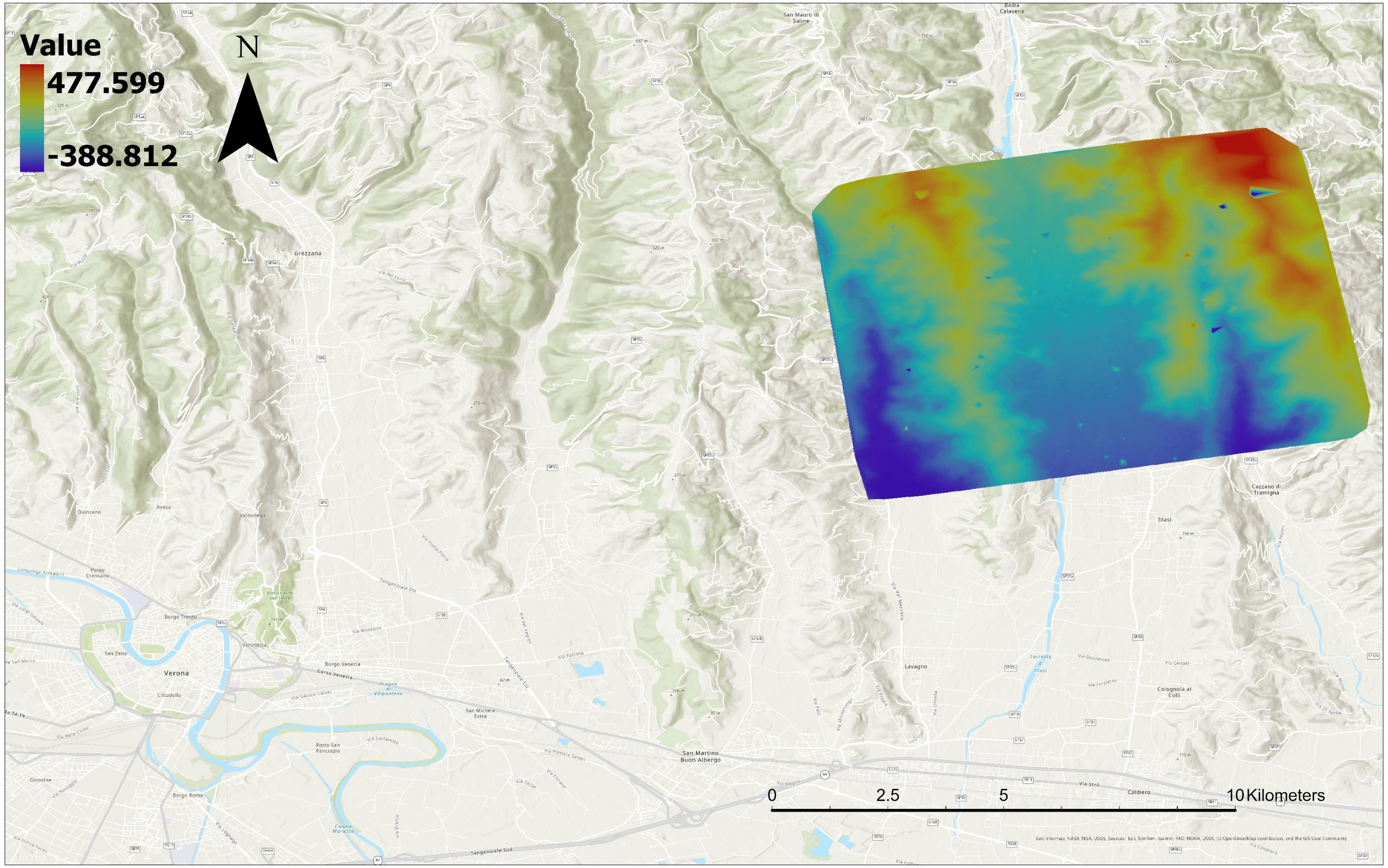
Key Pillars:

- Vertical Accuracy Assessment
- Land Cover Performance
- Spatial Coverage & Void Analysis

Datasets & Study Area

Study Area: Verona, Italy

- Constrained to SAOCOM data hull
- All data projected to UTM Zone 32N



Datasets & Study Area

SAOCOM: L-band InSAR point cloud

Reference Data:

- TINITALY 10m DEM
- Copernicus 30m DEM
- CORINE 2018 Land Cover

```python
# Automatically locate data files in subdirectories
saocom_files = list((DATA_DIR / "saocom_csv").glob("*.csv"))
tinitaly_files = list((DATA_DIR / "tinitaly").glob("*.tif"))
copernicus_files = list((DATA_DIR / "copernicus").glob("*.tif"))
corine_files = list((DATA_DIR / "ground_cover").glob("*.tif"))
sentinel_files = list((DATA_DIR / "sentinel_data").glob("*.tif"))

# Select first match for each dataset
saocom_path = saocom_files[0] if saocom_files else None
tinitaly_path = tinitaly_files[0] if tinitaly_files else None
copernicus_path = copernicus_files[0] if copernicus_files else None
corine_path = corine_files[0] if corine_files else None
sentinel_path = sentinel_files[0] if sentinel_files else None

# Find the corresponding .vat.dbf file for the CORINE raster
corine_dbf_path = None
if corine_path:
    corine_dbf_candidates = list((DATA_DIR / "ground_cover").glob(f"{corine_path.name}.vat.dbf"))
    corine_dbf_path = corine_dbf_candidates[0] if corine_dbf_candidates else None

```

```python
# =============================================================================
# COORDINATE REFERENCE SYSTEM
# =============================================================================
TARGET_CRS = 'EPSG:32632'  # UTM 32N
```

Core Processing Parameters

Coherence Filter: γ ≥ 0.3

• Purpose: Removes noisy, unstable points to ensure data quality.

Grid Resolution: 10 meters

• Purpose: Matches the highest-resolution reference (TINITALY).

Resampling:

• Cubic Convolution for elevation (smooths surfaces).

• Nearest Neighbor for land cover (preserves classes).

```python
# ============================================================================
# RESAMPLE TINITALY TO 10M
# ============================================================================
tinitaly_10m = np.full((grid_height, grid_width), NODATA, dtype=np.float32)

with rasterio.open(tinitaly_path) as src:
    reproject(
        source=rasterio.band(src, 1),
        destination=tinitaly_10m,
        src_transform=src.transform,
        src_crs=src.crs,
        dst_transform=target_transform,
        dst_crs=TARGET_CRS,
        resampling=Resampling.cubic,
        src_nodata=src.nodata,
        dst_nodata=NODATA
    )

# Save resampled TINITALY
tinitaly_10m_path = RESULTS_DIR / "tinitaly_10m.tif"
profile = {
    'driver': 'GTiff', 'dtype': 'float32', 'width': grid_width, 'height': grid_height,
    'count': 1, 'crs': TARGET_CRS, 'transform': target_transform,
    'nodata': NODATA, 'compress': 'lzw'
}
with rasterio.open(tinitaly_10m_path, 'w', **profile) as dst:
    dst.write(tinitaly_10m, 1)
```

Data Preprocessing Workflow

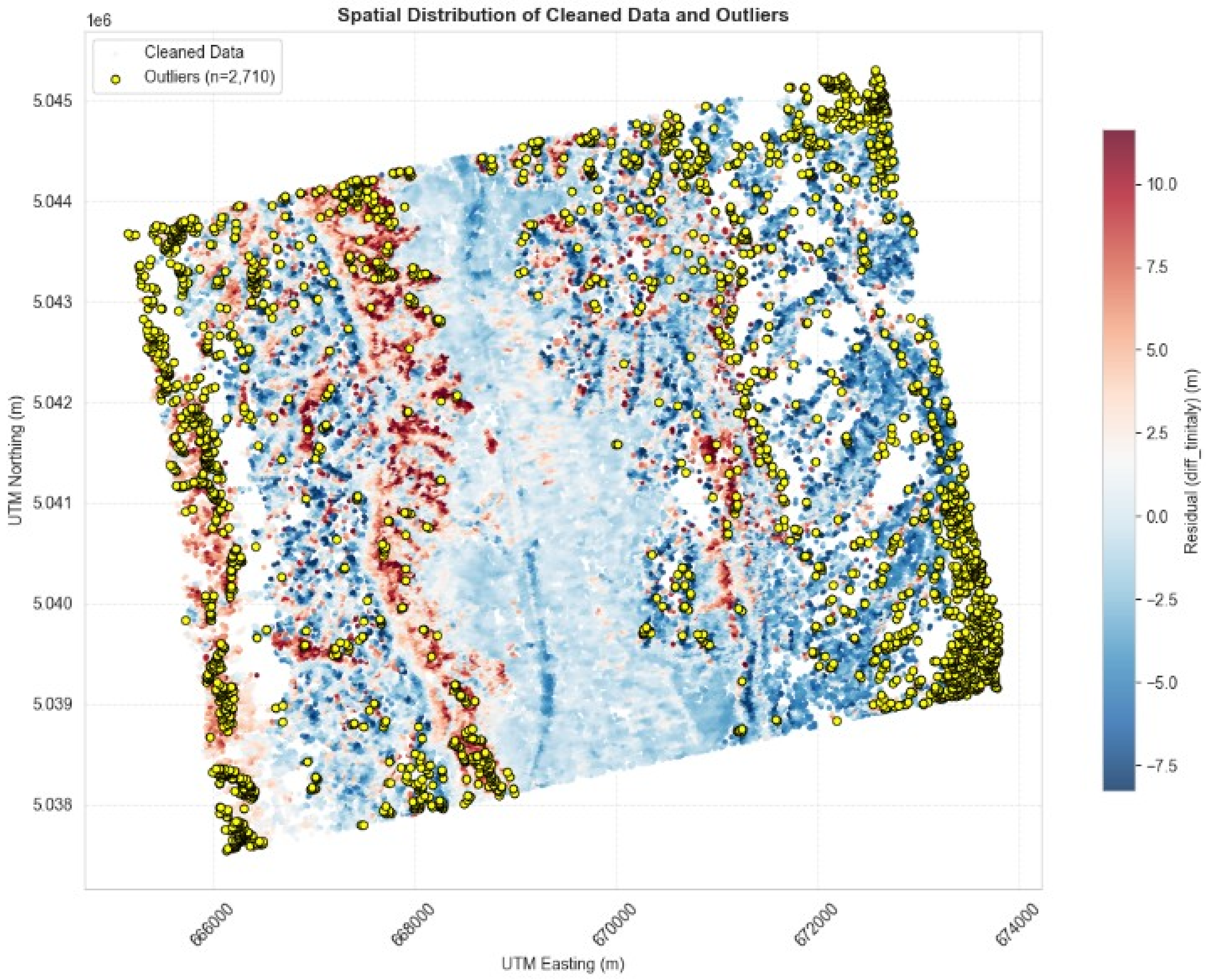
SAOCOM:

1. Filter by coherence (γ ≥ 0.3).
2. Remove invalid points.
3. Apply k-NN algorithm to remove spatial outliers.

Reference DEMs:

1. Reproject to UTM 32N.
2. Resample to 10m grid.
3. Clip to SAOCOM data boundary.

3d



Statistical Metrics Used

Classical Metrics:

- Bias (ME): Systematic error.
- RMSE: Overall error magnitude (sensitive to outliers).

Robust Metrics:

- Median: Central tendency (resists outliers).
- NMAD: Robust standard deviation.

Why robust? Our data has outliers; these metrics give a truer picture of performance.

Height Statistics Summary

This table shows the fundamental statistical distributions for all elevation datasets. Note the large range of the raw SAOCOM data and the differences in mean/median values, underscoring the need for calibration.

==

HEIGHT STATISTICS SUMMARY (all values in meters)

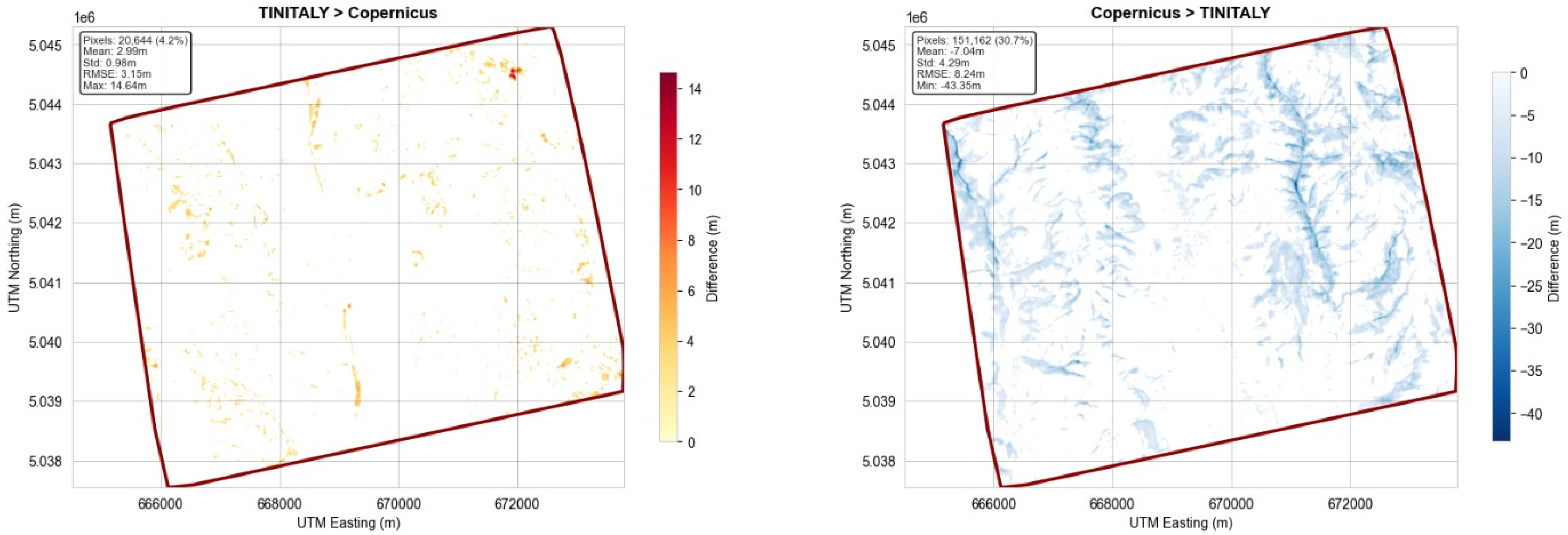
==

Dataset	Count	Min	Max	Mean	Median	Std Dev	Range	Q25	Q75	IQR
SAOCOM (Relative)	66765	-562.00	1163.70	340.09	327.00	116.87	1725.70	252.00	418.80	166.80
TINITALY (at SAOCOM pts)	66690	99.32	825.80	343.79	330.04	116.54	726.48	255.95	421.12	165.17
Copernicus (at SAOCOM pts)	66765	100.09	826.59	345.34	331.56	116.86	726.50	256.84	423.97	167.13
TINITALY (Full Grid)	493124	98.84	826.76	354.23	341.34	129.15	727.92	256.91	442.43	185.52
Copernicus (Full Grid)	495376	99.60	827.02	356.89	343.88	129.64	727.42	258.68	446.53	187.85

==

Reference DEM Comparison

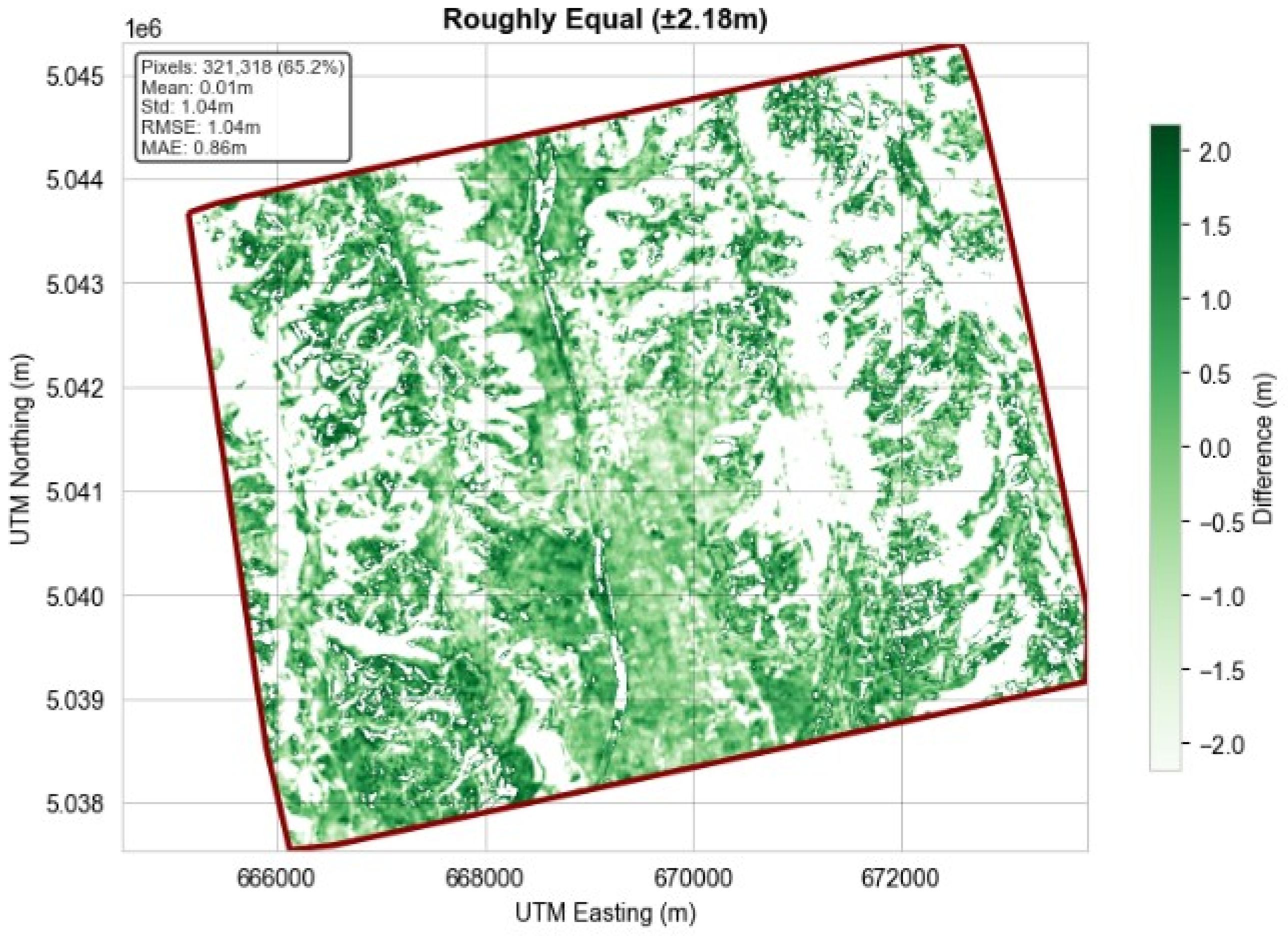
This figure visually breaks down the differences between the two reference DEMs. The maps highlight spatial patterns of disagreement, while the histogram shows the statistical distribution.

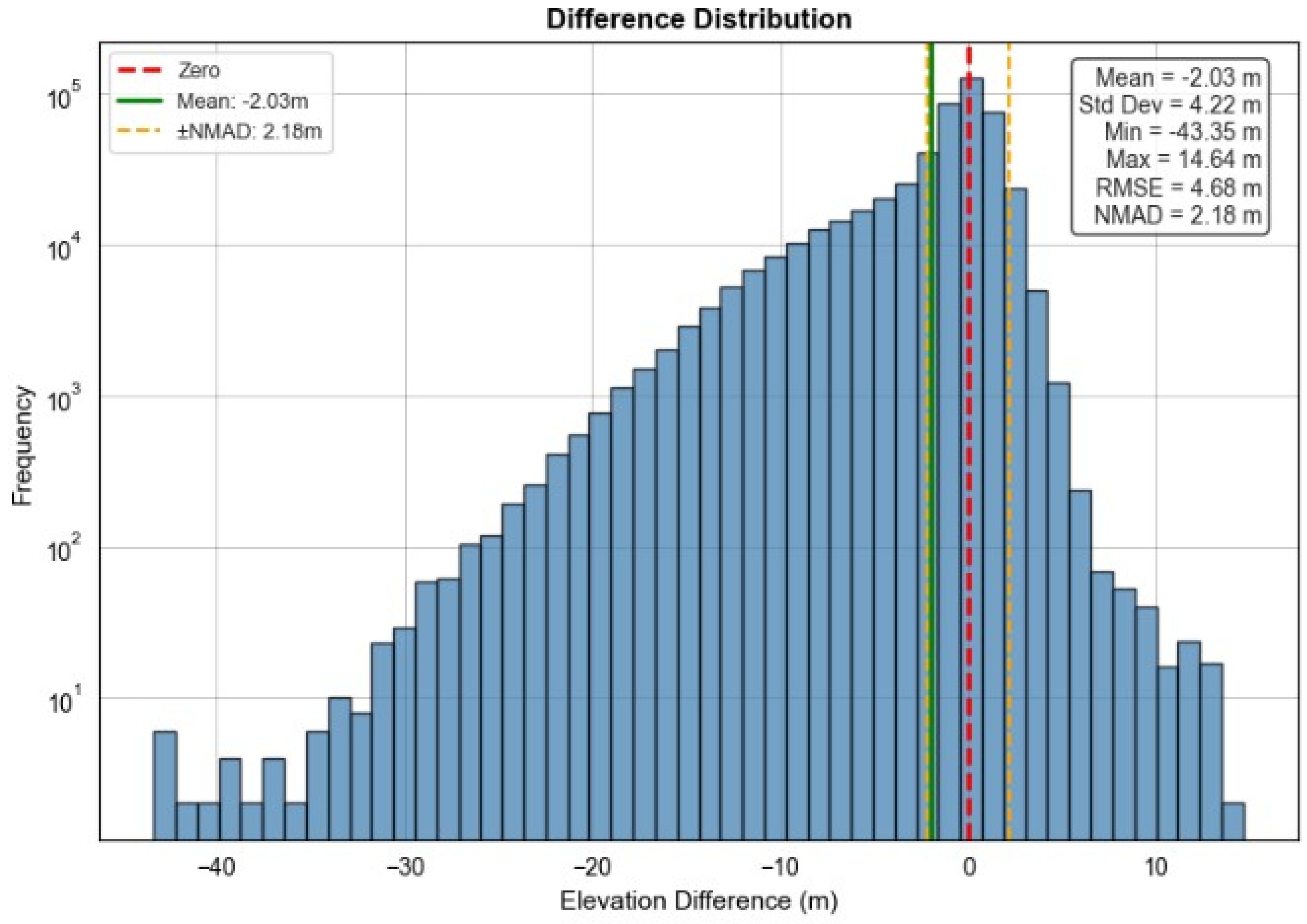


Reference DEM Comparison

This figure visually breaks down the differences between the two reference DEMs. The maps highlight spatial patterns of disagreement, while the histogram shows the statistical distribution.

Drop Image/Figure Here





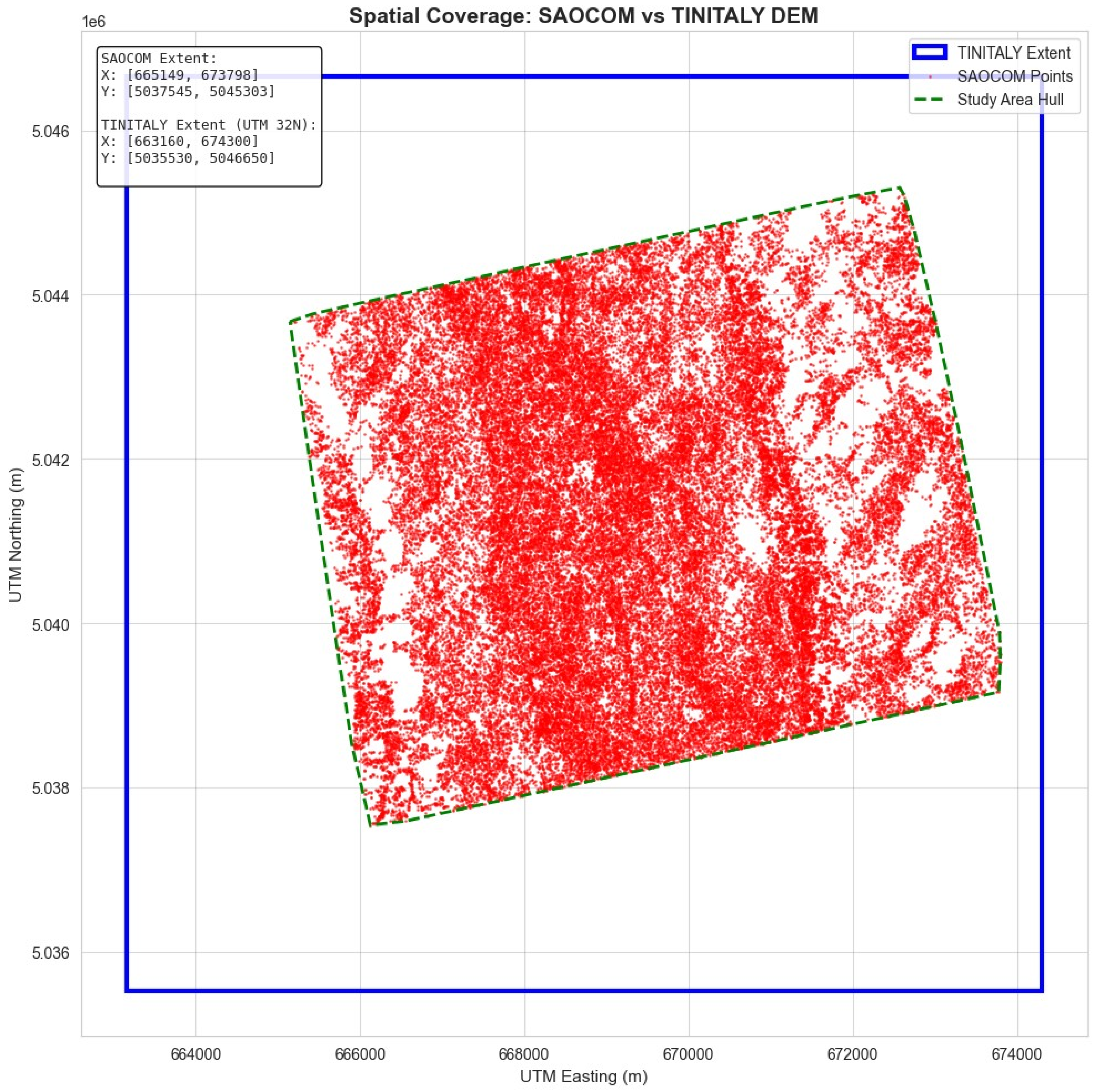
SAOCOM Spatial Coverage Analysis

A grid-based analysis reveals the extent of data gaps.

Void Percentage: 87.0%

This means for every 10 pixels in the study area, nearly 9 have no SAOCOM data.

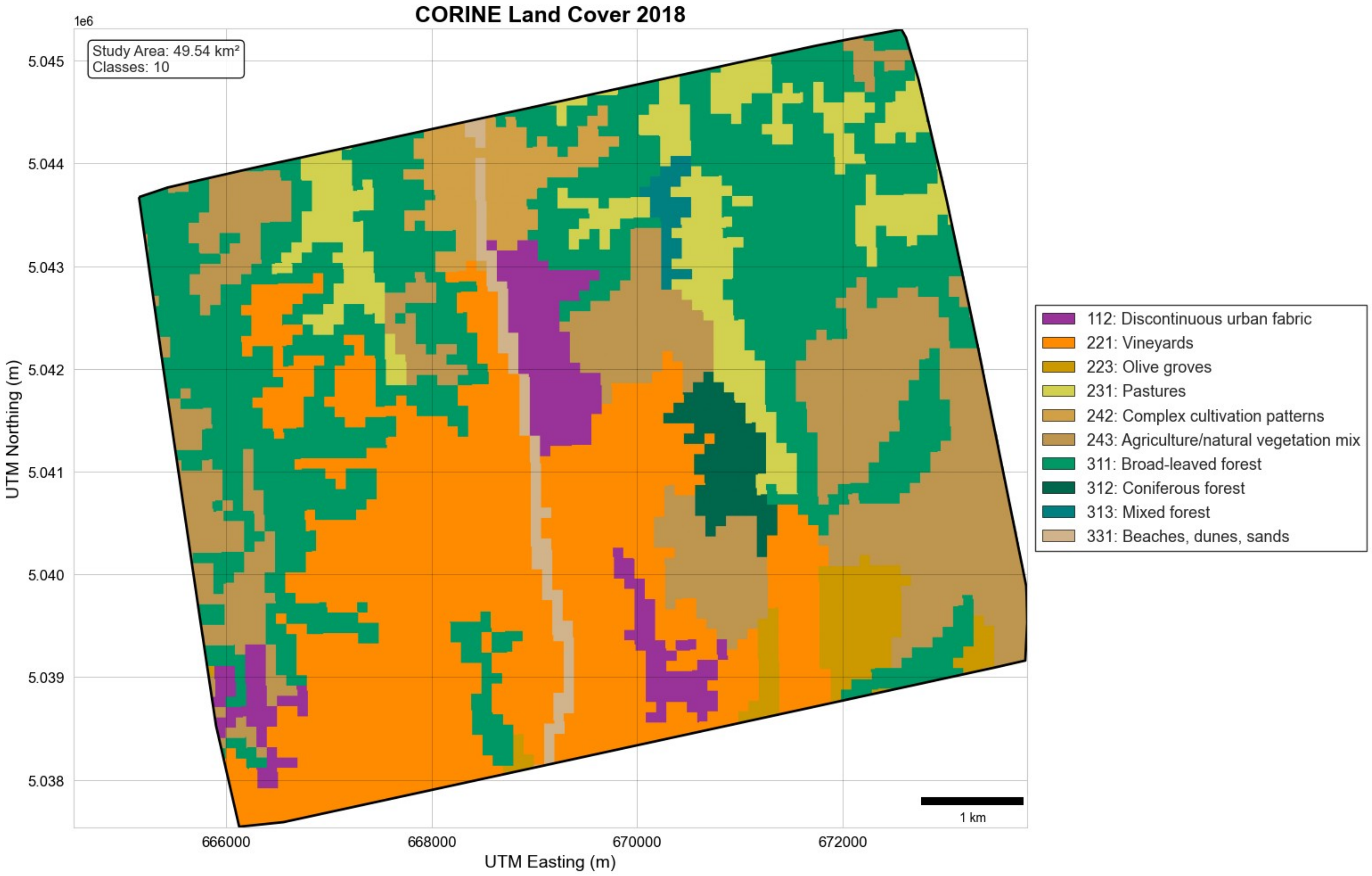
What land cover types are in these voids?



CORINE Land Cover

The CORINE 2018 dataset was processed to classify the terrain.

• 10 unique classes were found in the study area.
• A custom color palette was used for clarity in all subsequent maps and charts.

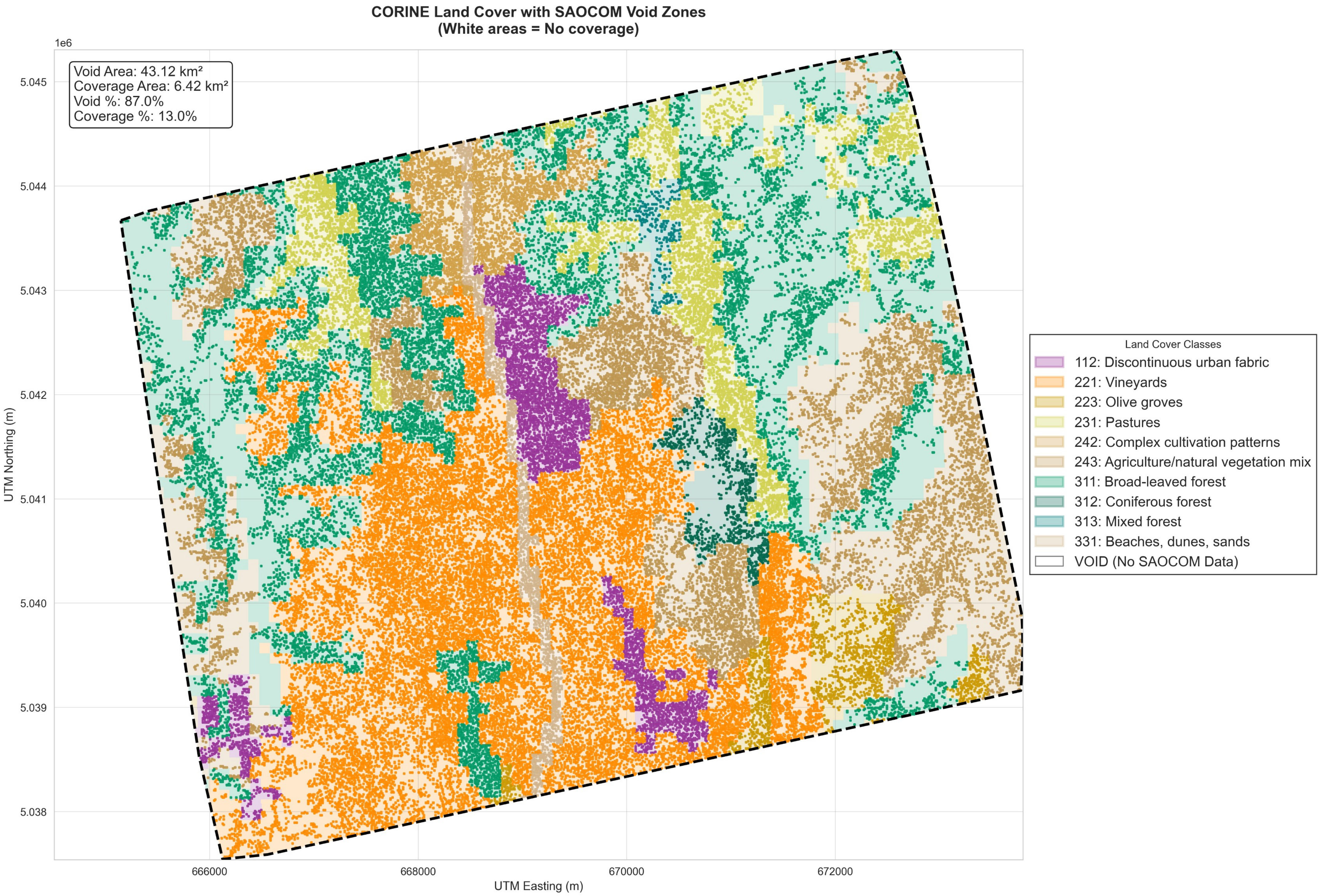


Linking Elevation to Land Cover

Each SAOCOM point was tagged with the specific land cover class it falls on.

This creates the final analysis dataset, enabling us to ask:

- "How does accuracy change between forests and urban areas?"
- "Which land cover types have the most data voids?"



Height Residuals by Land Cover

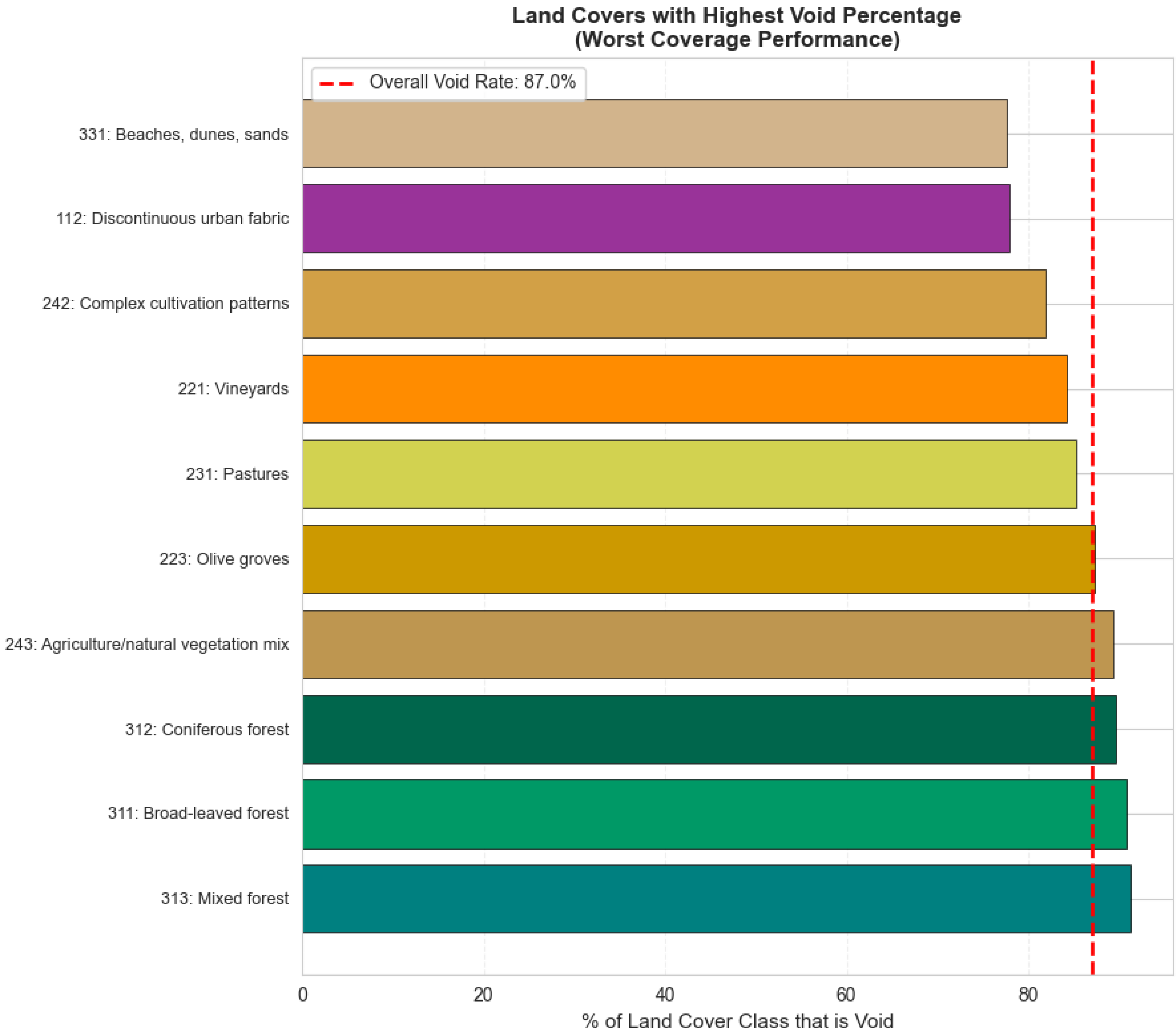
This table shows the error statistics for each land cover class. Performance varies significantly. Urban areas show lower error (NMAD < 3.5m), while forests show much higher error (NMAD > 5m).

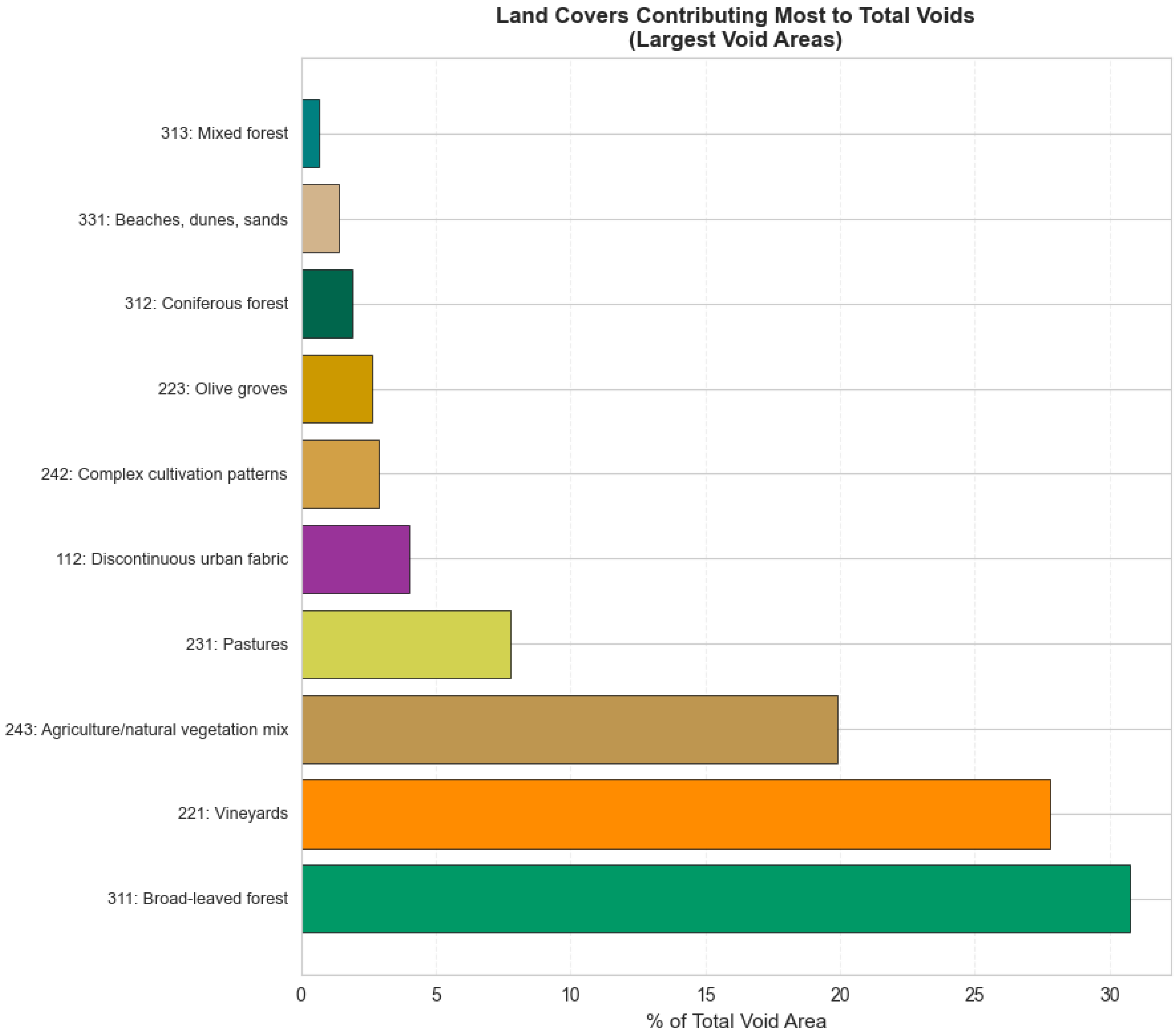
HEIGHT RESIDUAL STATISTICS by CORINE Land Cover (N > 50)
(Residual = Calibrated SAOCOM Height - TINITALY Reference DEM)

corine_code	LC_Label	N_Points	Median_Diff_m	NMAD_m	Mean_Diff_m	Std_Dev_m
243	Agriculture/natural vegetation mix	9,796	-0.98 m	NaN	-0.70 m	4.68 m
331	Beaches, dunes, sands	1,804	-1.03 m	1.34 m	-1.13 m	1.67 m
311	Broad-leaved forest	12,631	+1.98 m	NaN	+2.30 m	5.85 m
242	Complex cultivation patterns	2,807	-0.89 m	1.35 m	-0.62 m	2.38 m
312	Coniferous forest	978	+2.79 m	6.22 m	+2.70 m	6.53 m
112	Discontinuous urban fabric	5,107	+0.55 m	1.59 m	+0.83 m	1.99 m
313	Mixed forest	263	+2.09 m	5.40 m	+2.05 m	5.38 m
223	Olive groves	1,552	-3.41 m	2.72 m	-2.91 m	3.61 m
231	Pastures	5,872	-0.82 m	NaN	-0.49 m	4.51 m
221	Vineyards	23,142	-0.12 m	1.92 m	+0.29 m	3.34 m

Void Analysis by Land Cover

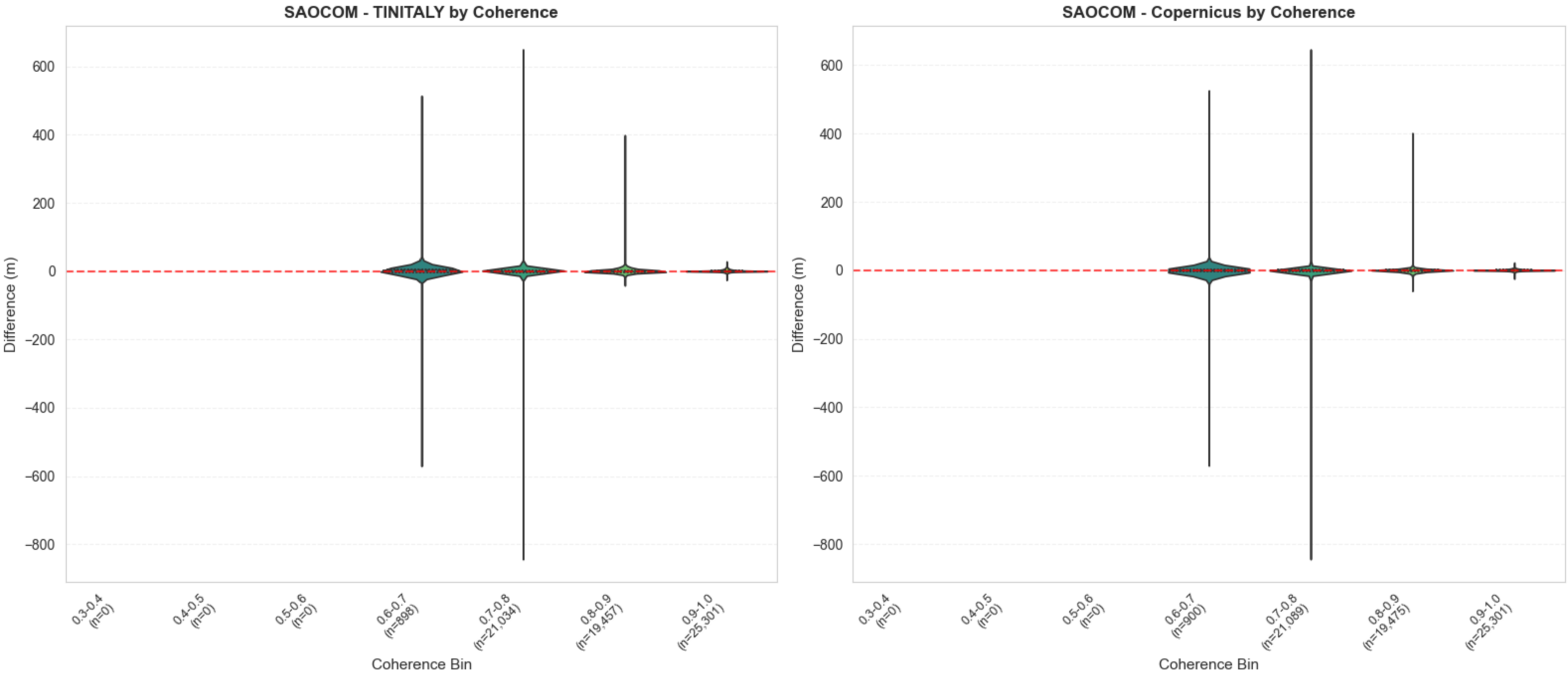
This table quantifies which land cover classes are most affected by data gaps. Water bodies and forests are almost entirely voids, while vineyards and forests are the largest contributors to the total void area.





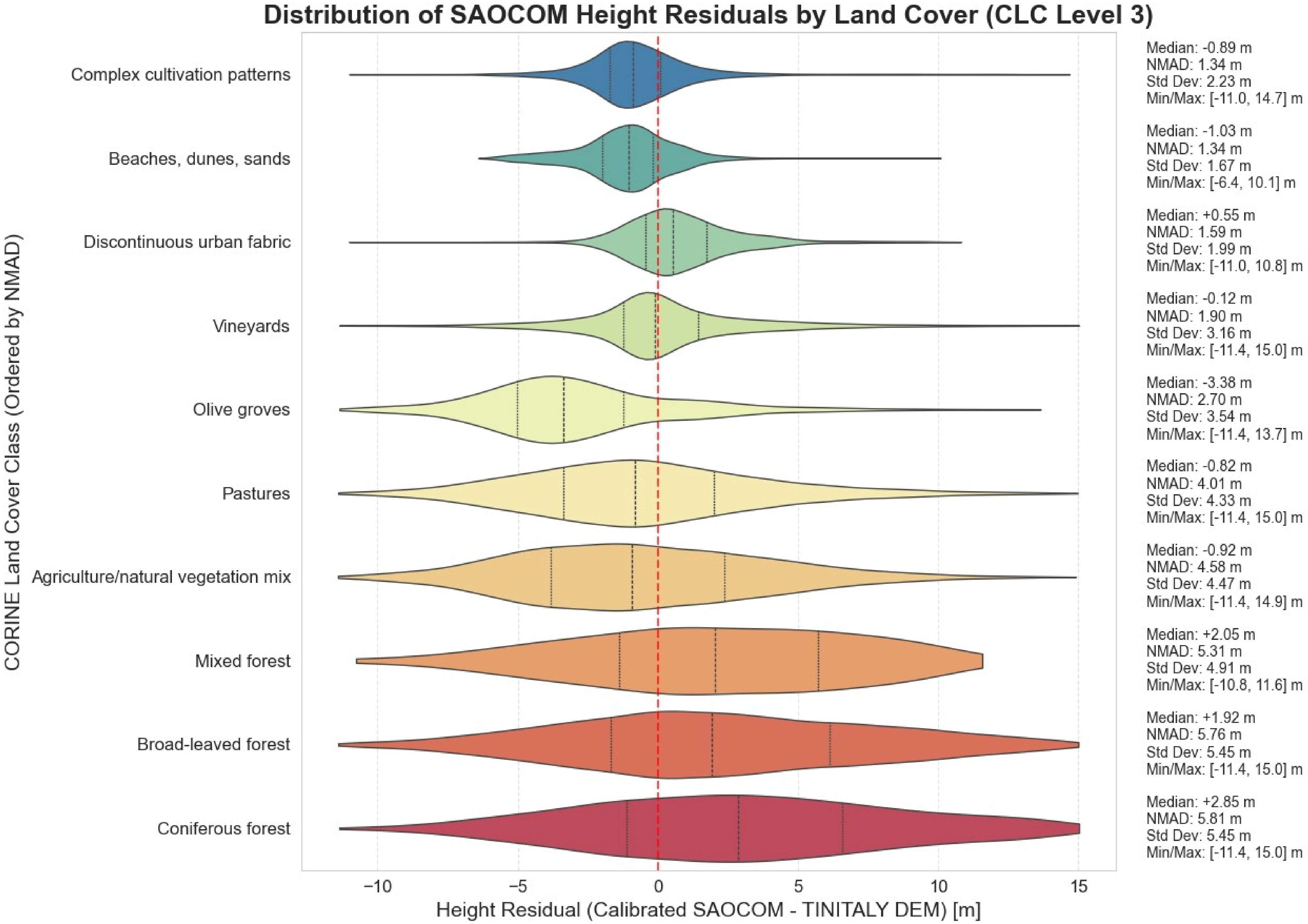
Error vs. Coherence

These plots show the clear relationship between signal quality (coherence) and accuracy. As coherence increases, the error distribution tightens significantly, and the median error approaches zero.



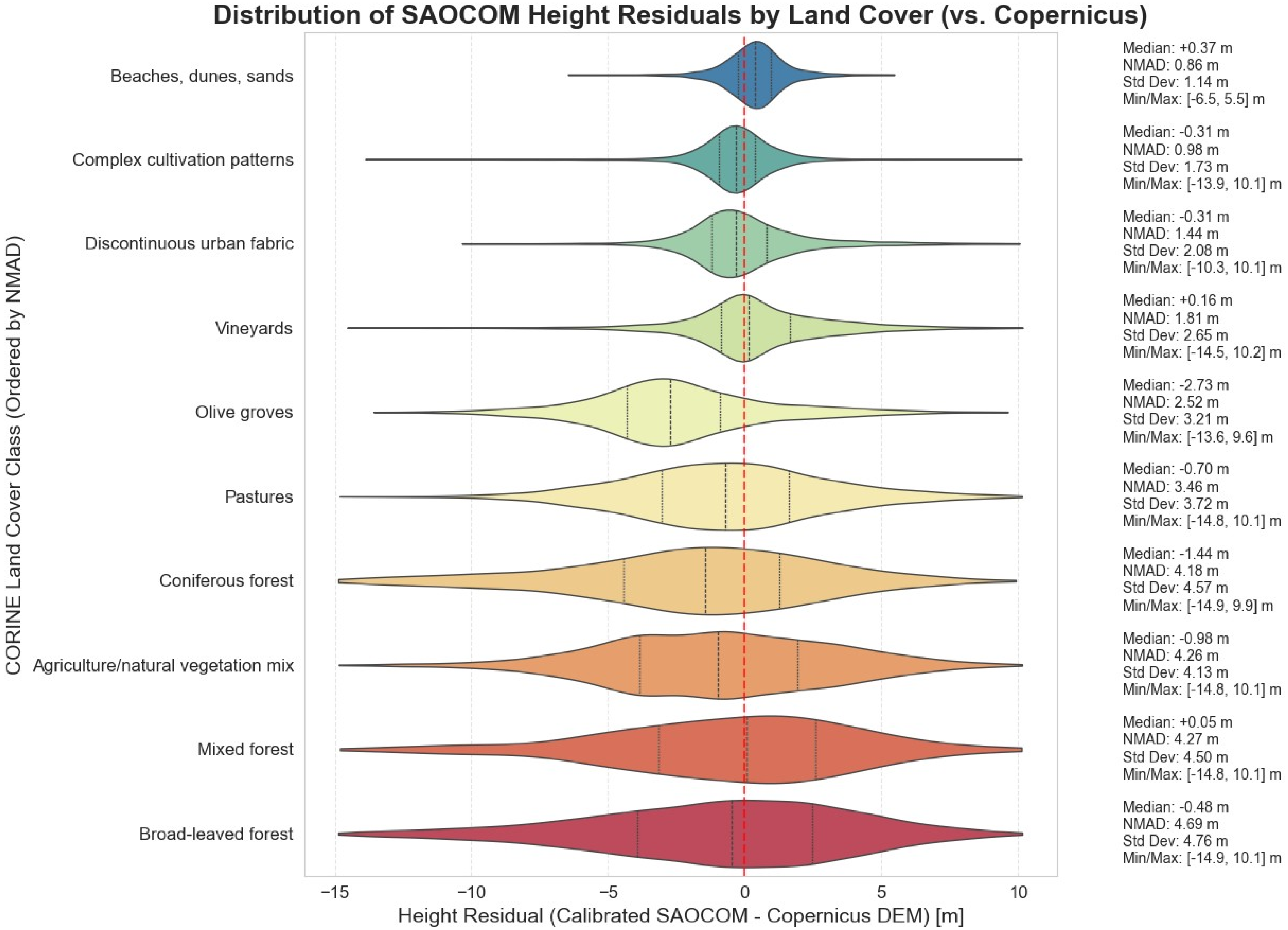
Error by Land Cover

These plots visualize the error distributions for each major land cover class. Note the wide shape for forests (high uncertainty) compared to the narrow shape for urban fabric (more reliable).



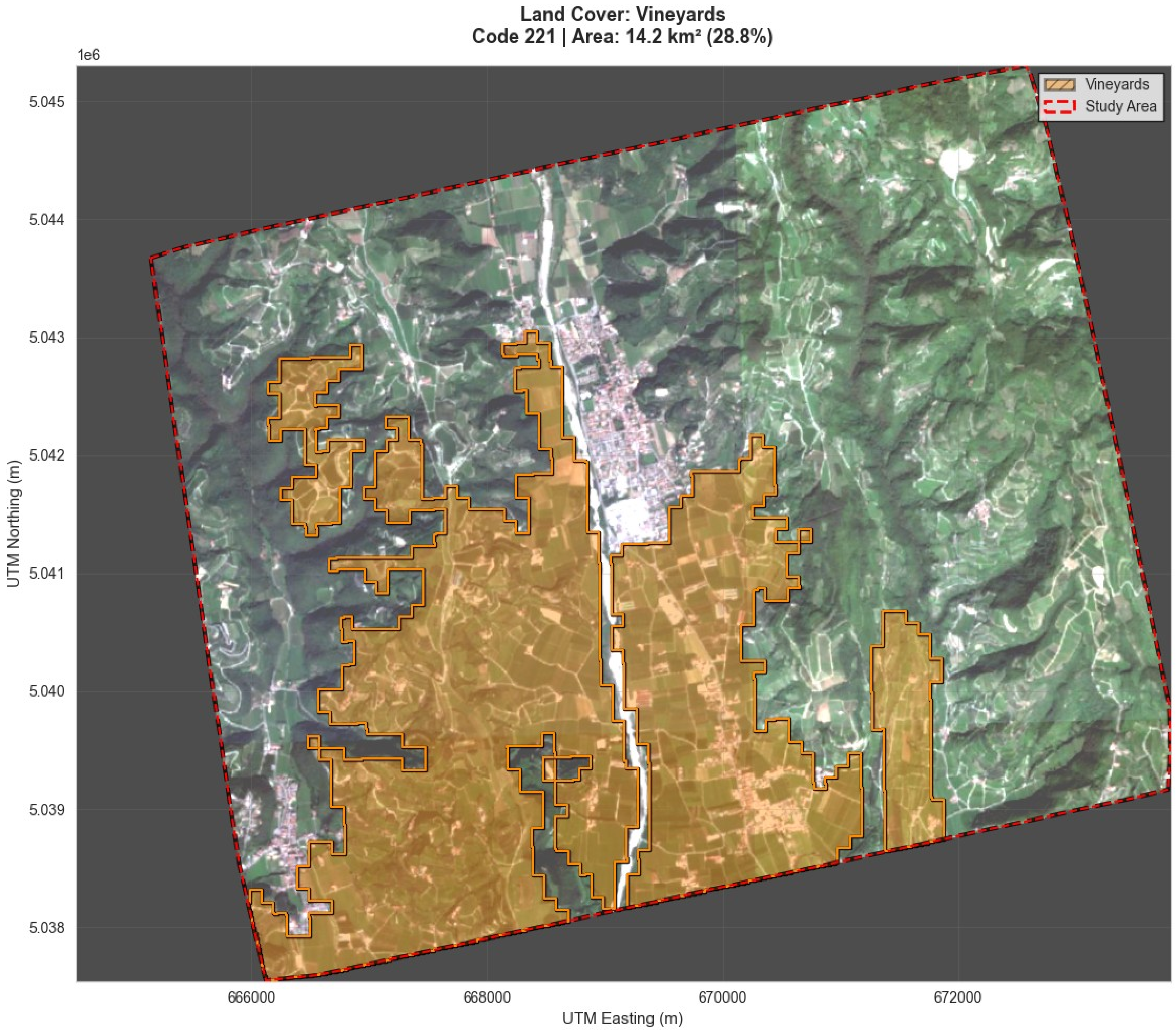
Error by Land Cover

These plots visualize the error distributions for each major land cover class. Note the wide shape for forests (high uncertainty) compared to the narrow shape for urban fabric (more reliable).



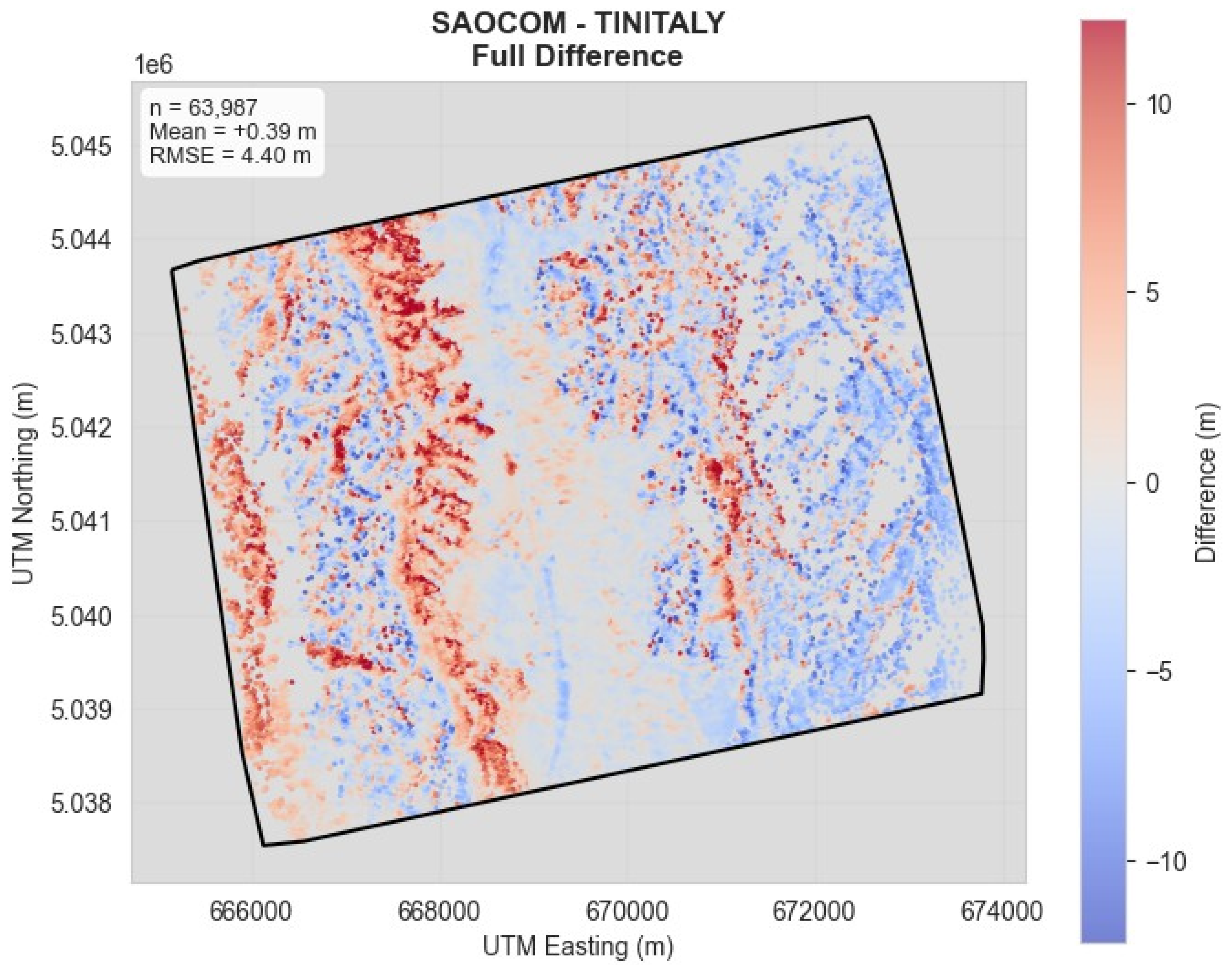
Individual Land Cover Overlay

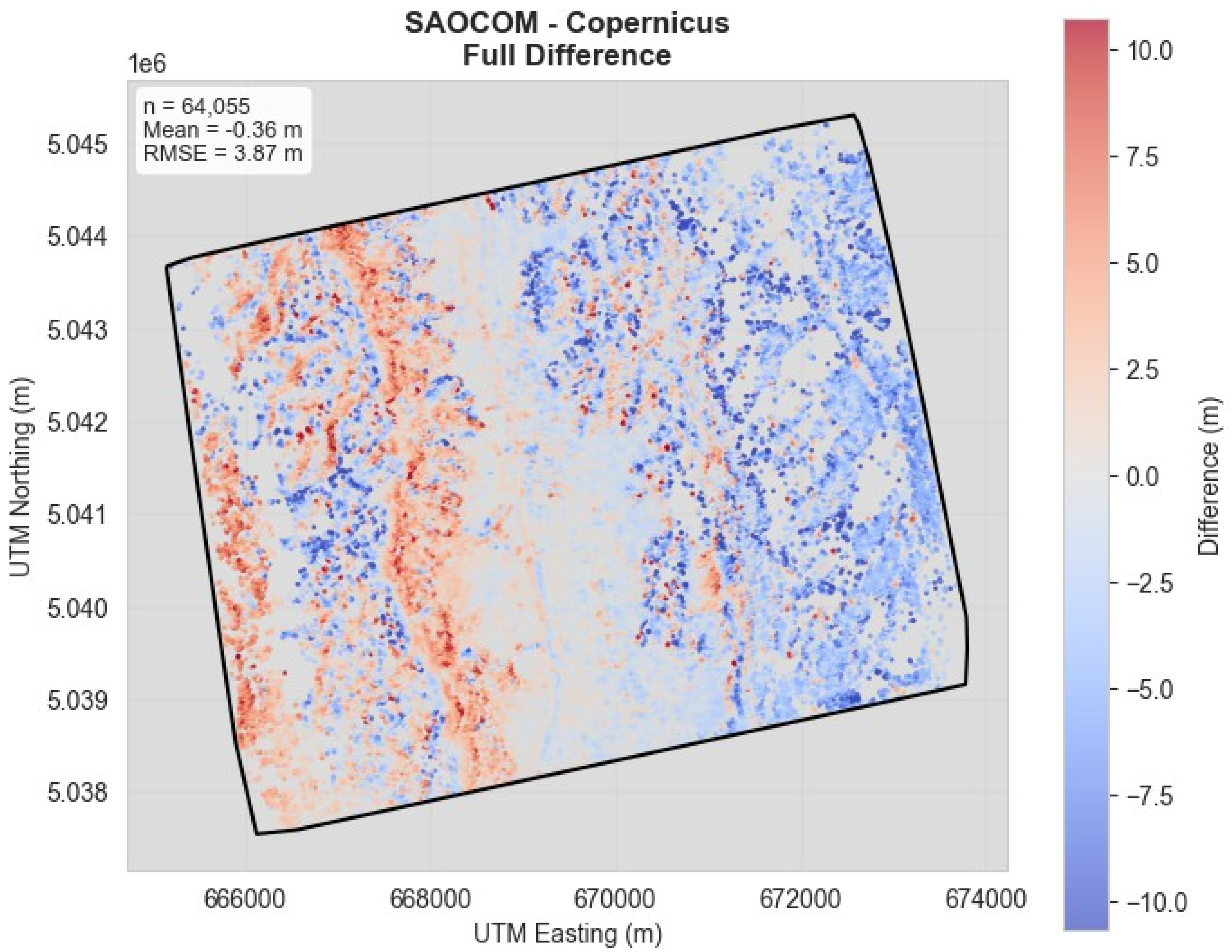
To understand the spatial distribution, each class was map over a satellite image. This example shows the extensive "Vineyards" class, a key agricultural feature in the region.



Gridded SAOCOM Residuals

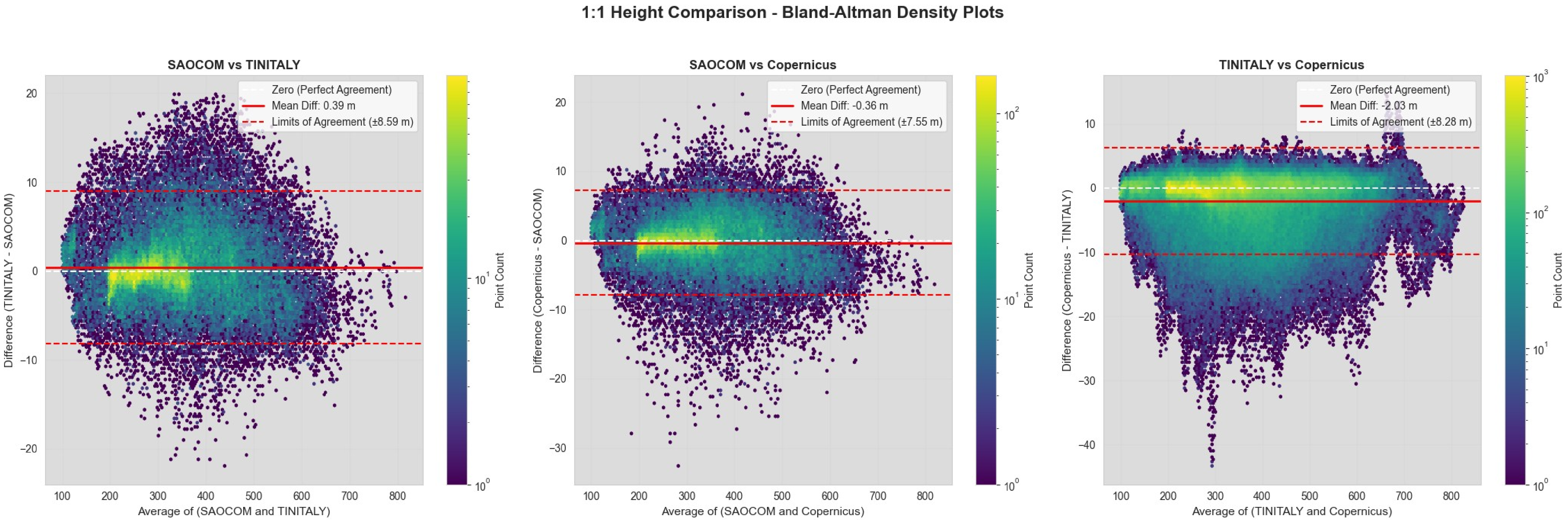
The point-based errors were interpolated into a continuous map. This allows us to visualize the spatial patterns of where the SAOCOM DEM is higher or lower than the reference DEMs.





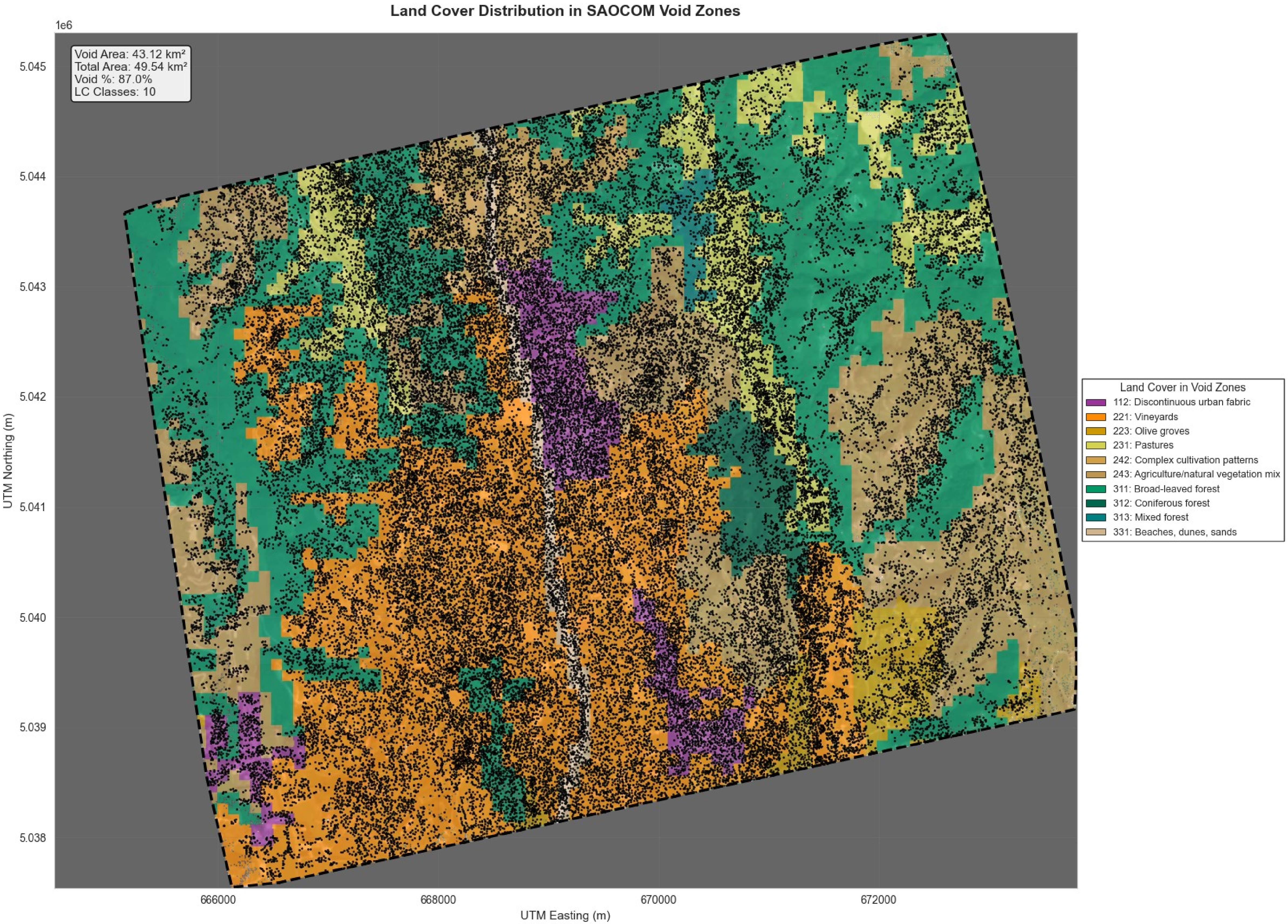
Height Correlation

These plots directly compare elevation values. While there is a strong linear relationship (high correlation), the spread of points around the 1:1 line visually represents the error.



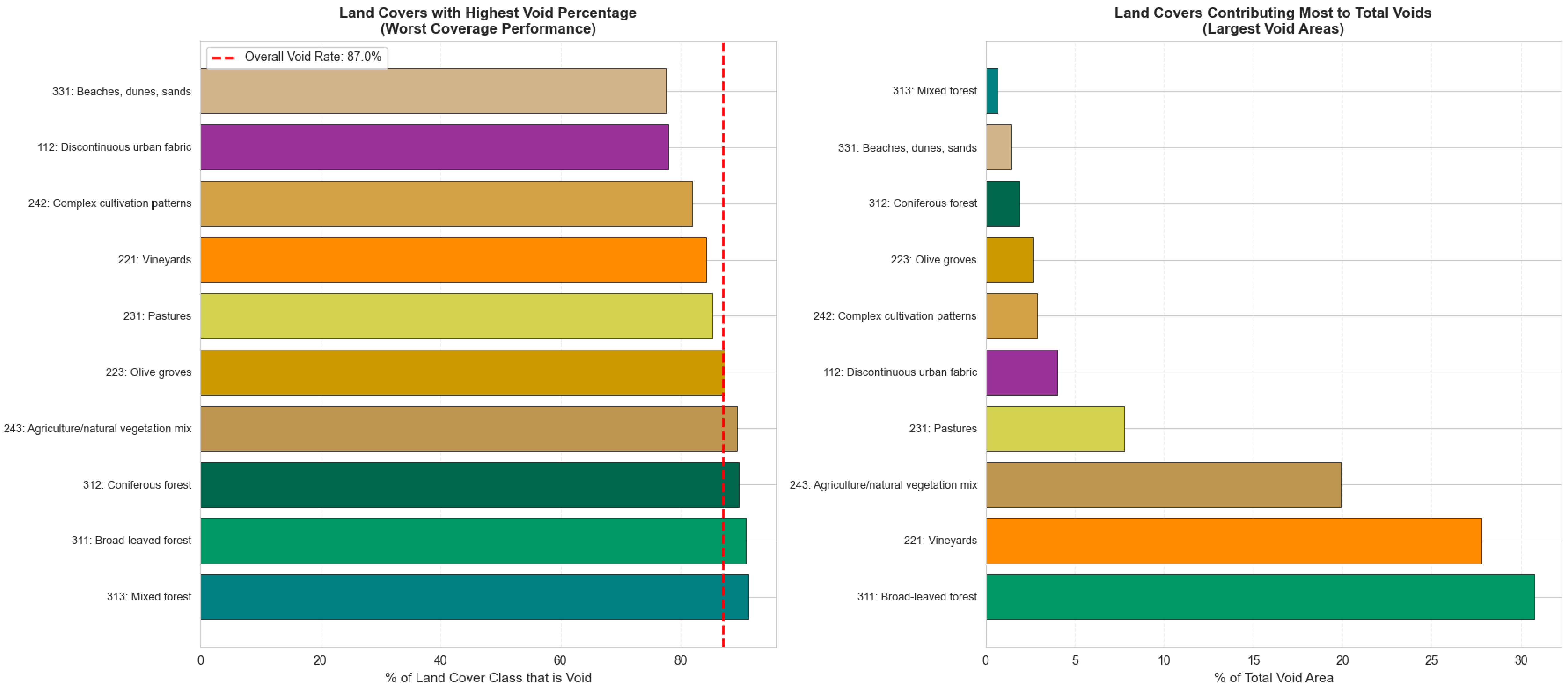
Land Cover Ins

This map shows only the land cover present in areas where SAOCOM data is missing. It clearly demonstrates that forests, vegetation, and water dominate the data gaps, pointing to decorrelation as the cause.



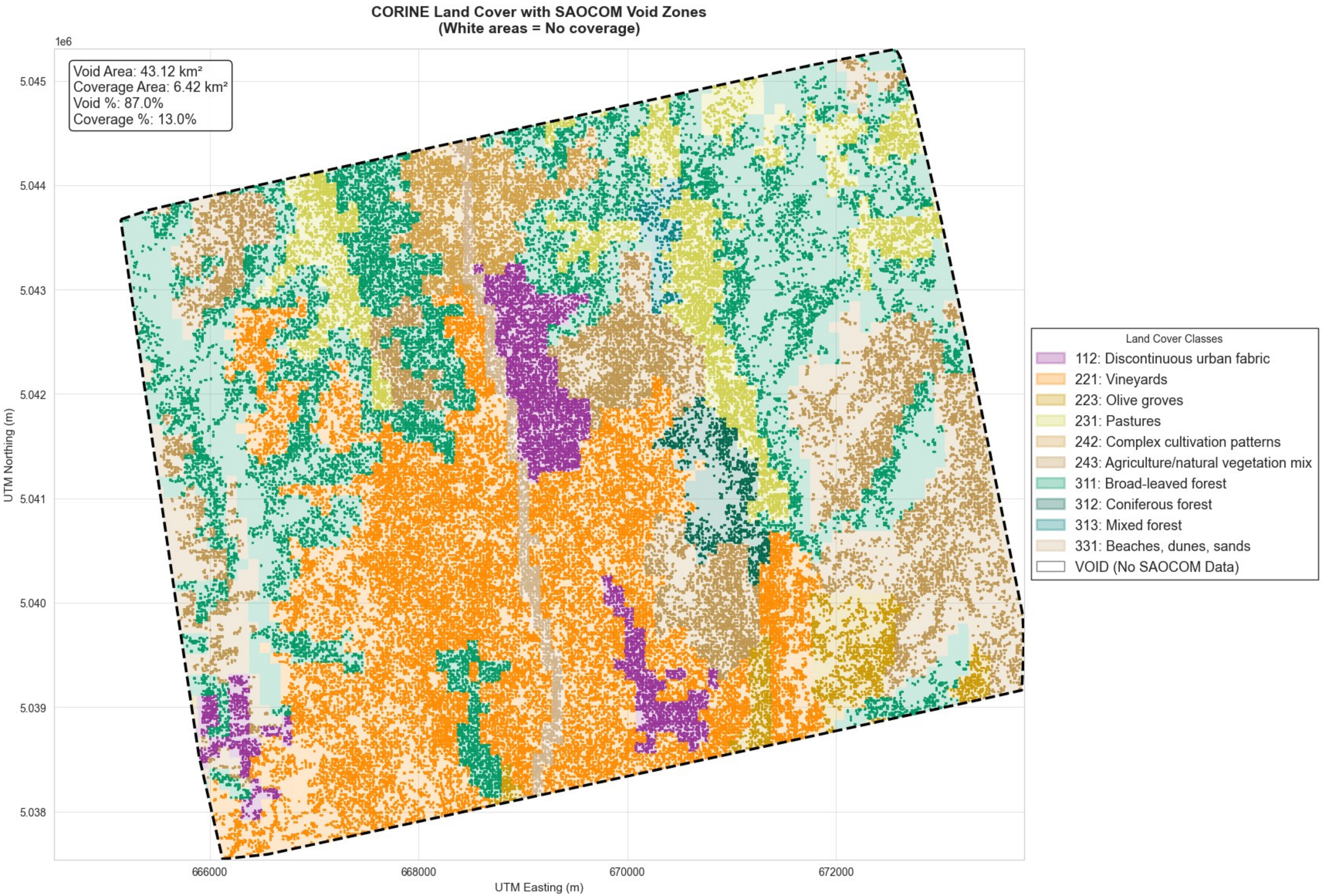
Quantifying Void Contributors

These charts pinpoint the most problematic land cover classes. The left chart shows the worst relative coverage (water), while the right chart shows the largest contributors to total void area (forests).



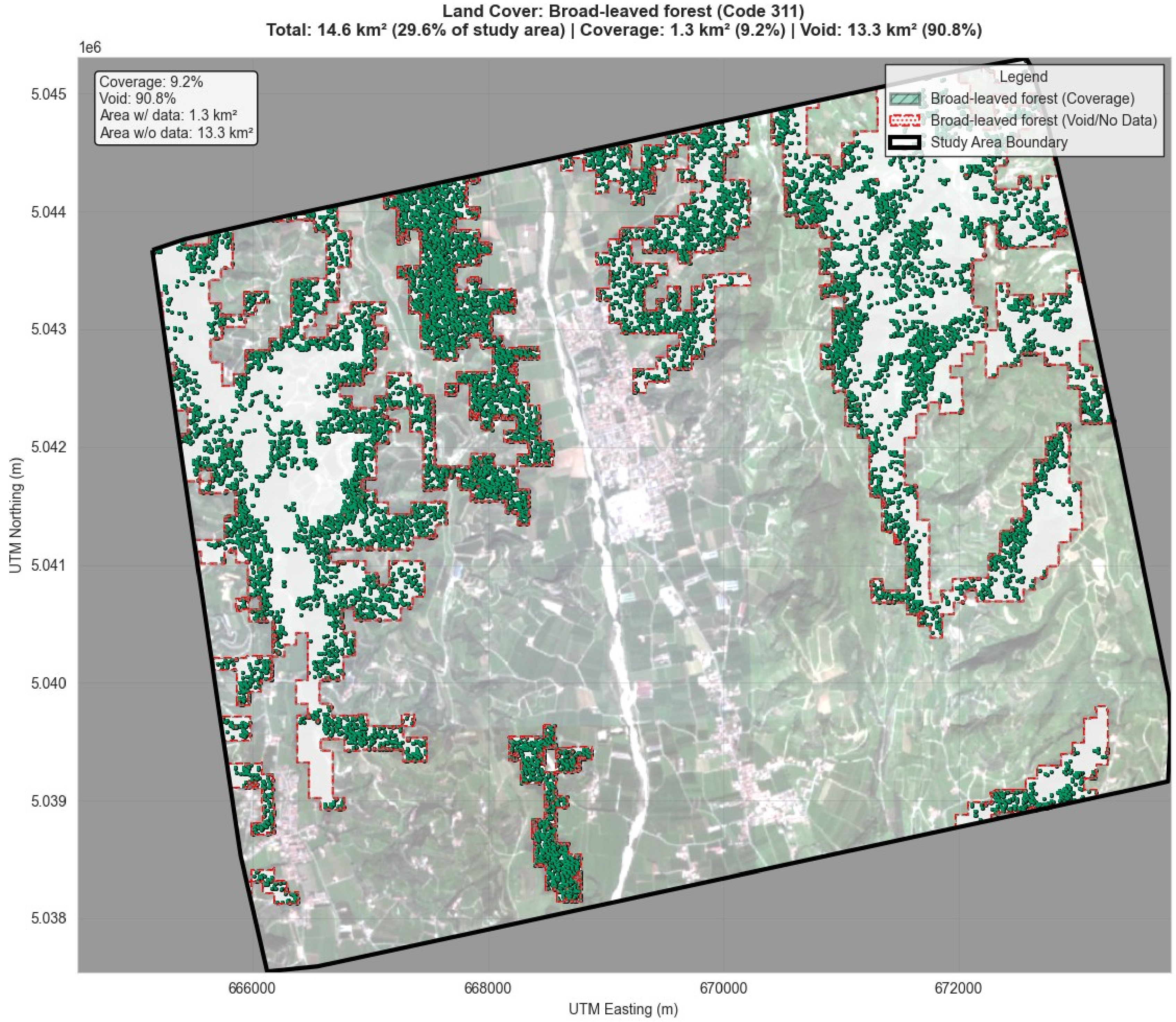
Coverage and Voids Across the Landscape

This map shows all land cover types, with white areas indicating the SAOCOM data voids. This single image provides a powerful summary of the 87% void statistic and the fragmented nature of the coverage.



Coverage vs. Voids: A Closer Look

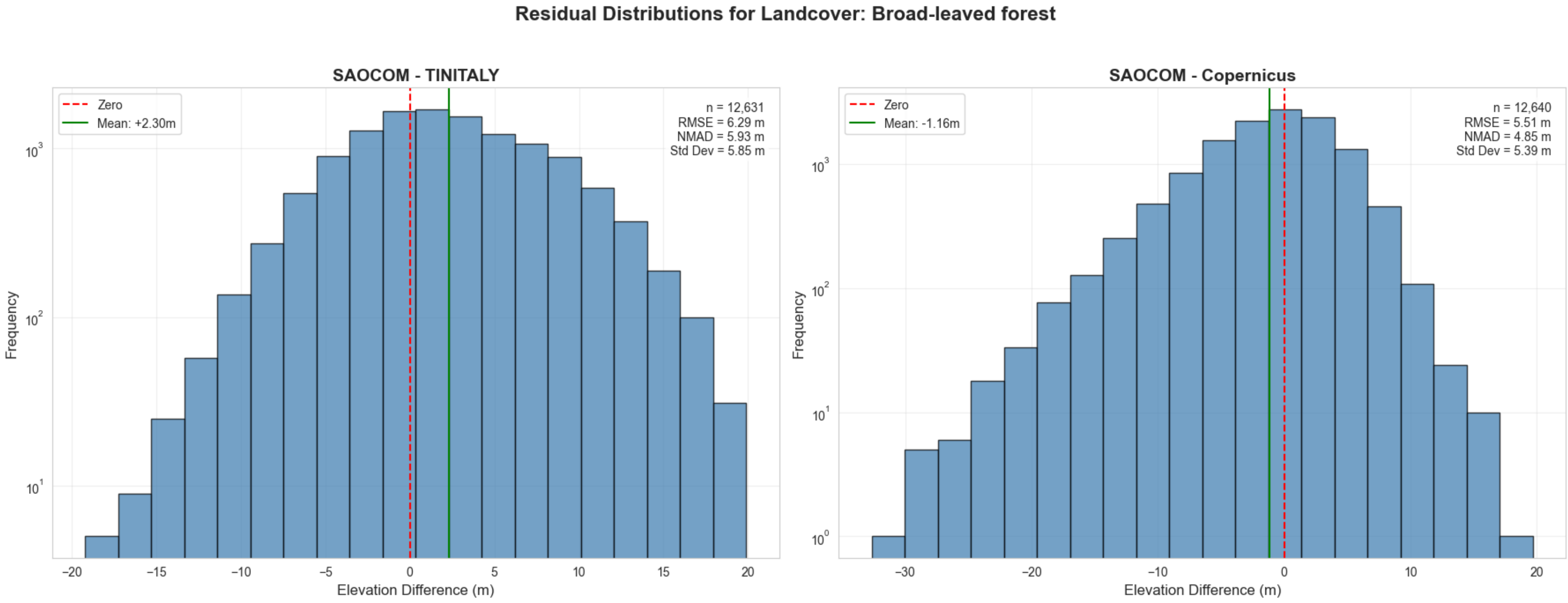
This map zooms in on a single class, partitioning it into areas of successful SAOCOM coverage versus void areas. This is essential for assessing the data's utility for specific applications (e.g., forestry).



Coverage vs. Voids: A Closer Look

This map zooms in on a single class, partitioning it into areas of successful SAOCOM coverage versus void areas. This is essential for assessing the data's utility for specific applications (e.g., forestry).

Drop Image/Figure Here



Conclusions

Accuracy: SAOCOM can produce accurate elevation data, but performance is highly dependent on coherence and land cover type.

Coverage: The primary limitation is the 87% data void, concentrated in forests, complex vegetation, and water.

Utility: Best suited for urban and open agricultural areas; less reliable in vegetated or complex terrain.

Future Work

• Investigate data fusion methods to intelligently fill data voids using other sensors.
• Test the impact of different InSAR processing parameters on improving coverage.
• Analyze the influence of seasonality and temporal changes on L-band coherence.