## Slide 1: SAOCOM InSAR DEM Validation – Key Pillars

* **Vertical Accuracy Assessment:** The study evaluates how accurate the SAOCOM InSAR-derived DEM heights are compared to known references. This involves computing error metrics (bias, RMSE, etc.) to quantify how far off the SAOCOM elevations are from truth data.
* **Land Cover Performance:** It examines error and data availability stratified by land cover type. Different surfaces (urban, forest, water, etc.) affect InSAR signal quality, so the DEM’s accuracy and coverage are assessed for each class (e.g. how well SAOCOM performs in forests versus urban areas).
* **Spatial Coverage & Void Analysis:** The study analyzes where the SAOCOM DEM has valid data and where it has gaps (“voids”). It calculates the percentage of the area with no data and maps those voids. The aim is to identify patterns (are voids concentrated in certain terrain or land cover?) and quantify data completeness.

## Slide 2: Datasets & Study Area

* **Study Area – Verona, Italy:** The analysis is focused on the region around Verona. The spatial extent is defined by the hull of available SAOCOM data. All datasets were reprojected to the same coordinate system (UTM Zone 32N) so they align on a common grid in meters.
* **SAOCOM L-band InSAR Point Cloud:** This is the primary dataset – elevation points derived from SAOCOM satellite’s L-band radar interferometry. It provides a cloud of elevation measurements (in WGS84 lat/long originally, converted to UTM). Each point has attributes like height and coherence (signal quality).
* **Reference DEMs:** Two terrain models are used as ground truth for validation:
  + TINITALY 10 m DEM: A high-resolution (10 m) digital elevation model of Italy (likely based on photogrammetry or LiDAR), used as a precise reference. It’s originally on the WGS84 ellipsoidal vertical datum (the code checked that it doesn’t use a geoid).
  + Copernicus 30 m DEM: The Copernicus GLO-30 DEM (European Space Agency) at 30 m resolution, used as a secondary reference. It uses the EGM2008 geoid for heights. Both DEMs were clipped to the study area and resampled to 10 m for direct comparison.
* **CORINE 2018 Land Cover:** A categorical map (Coordination of Information on the Environment) detailing land cover classes (urban, forest, water, etc.). This map was reprojected to UTM and resampled to 10 m (using nearest-neighbor to preserve class labels). It provides the land cover classification for each location in the study area, which will be used to stratify the analysis.

## Slide 3: Core Processing Parameters

* **Coherence Filter (γ ≥ 0.3):** Only InSAR points with coherence ≥ 0.3 were kept. Coherence is a measure of signal correlation between the radar image pair – values range 0–1. By setting γ\_min = 0.3, they discard very noisy or unstable points (where the interferometric phase was not reliable). This ensures the dataset quality by filtering out low-confidence elevations (the code used COHERENCE\_THRESHOLD = 0.3 and filtered points accordingly).
* **Grid Resolution – 10 m:** All data were gridded at 10-meter resolution. This matches the highest resolution reference DEM (TINITALY 10 m) and provides a common raster grid for comparison. Essentially, the study area is divided into 10 m cells; SAOCOM point data and references are aggregated or interpolated to this grid for analysis. A finer grid (10 m) captures terrain detail and allows direct cell-by-cell error calculations.
* **Resampling Methods:** When projecting and aligning datasets, different interpolation methods were applied:
  + Elevation data: **Cubic convolution** interpolation was used to resample DEMs (e.g., resampling Copernicus 30 m down to 10 m). Cubic convolution produces a smooth surface and avoids sharp artifacts, suitable for continuous terrain data.
  + Land cover data: **Nearest neighbor** resampling was used for the CORINE land cover raster. This method preserves the discrete class values (no averaging of classes). It ensures that land cover labels remain authentic (e.g., a forest stays labeled as forest after resampling, with no mixed classes).
* Why these choices: The cubic method smooths and minimizes resampling error in elevation surfaces, and nearest-neighbor keeps categorical maps intact. Together, these parameters (coherence cutoff, 10 m grid, proper resampling) set a consistent, quality-controlled framework for the validation.

## Slide 4: Data Preprocessing Workflow

**SAOCOM Point Data Processing:**

1. **Filter by Coherence ≥ 0.3:** All raw SAOCOM InSAR points were first filtered to include only those with a coherence value ≥ 0.3. This step, implemented in the code (df\_filtered = df[df['COHER'] >= 0.3]), removes low-quality points (e.g., over water or dense forest where the radar phase didn’t hold stable). It ensures subsequent analysis uses only meaningful elevation points.
2. **Remove Invalid Points:** Points with missing or zero coordinates or heights were discarded. For example, any entries where corrected latitude/longitude were 0 or non-numeric were dropped (dropna and zero checks in the code). This cleans up any junk data or processing artifacts.
3. **Spatial Outlier Removal (k-NN):** A k-nearest-neighbors algorithm was applied to eliminate isolated outliers. They computed the average distance to each point’s 5 nearest neighbors; if a point was far removed from any cluster (beyond a set distance threshold, e.g., 100 m), it was deemed an outlier and removed. This step (using the function remove\_isolated\_knn) catches spurious points that survived the coherence filter but are geographically isolated (likely noise or processing errors). After this, the remaining SAOCOM points form a clean, spatially consistent dataset. They then converted these points into a GeoDataFrame and projected it to UTM 32N for alignment with the DEM grid (the code adds x\_utm, y\_utm coordinates in meters).

**Reference DEMs Processing:**

1. **Reprojection to UTM 32N:** Both TINITALY and Copernicus DEM were originally in geographic coordinates. They were reprojected to the UTM Zone 32N projection (EPSG:32632) to match the SAOCOM data’s coordinate system and units. This ensures all elevation data is on the same plane coordinate grid (meters).
2. **Resample to 10 m Grid:** Using the defined target 10 m grid (covering the SAOCOM convex hull area), each DEM was resampled. The code used rasterio.warp.reproject with **cubic** resampling for the DEM values. TINITALY (native 10 m) was mostly just reprojected with minimal resampling, whereas Copernicus (native ~30 m) was interpolated down to 10 m. The result is that both reference DEMs are available as 10 m raster arrays aligned pixel-for-pixel for the region of interest.
3. **Clip to SAOCOM Data Boundary:** They defined the study area extent as the convex hull around all SAOCOM points (essentially the footprint of available InSAR data). Both DEM rasters were then clipped/masked to this hull. Any DEM cells outside where SAOCOM had coverage were set to NoData. This focuses the comparison only on areas where SAOCOM was present, and it ensures the error statistics aren’t skewed by large areas where no comparison can be made. In practice, they created a hull polygon and rasterized it as a mask (the code hull\_mask), then applied it to the resampled DEMs.

After these steps, the SAOCOM data and reference grids are prepared: SAOCOM points (filtered and cleaned) are in UTM coordinates, and reference DEMs are on the same 10 m UTM grid and masked to the SAOCOM extent. This unified dataset is ready for computing differences and statistics.

## Slide 7: Statistical Metrics Used

This slide defines the error metrics and explains the inclusion of robust statistics due to outliers in the data:

* **Bias (Mean Error, ME):** This is the average of SAOCOM minus reference elevation for all points – it measures systematic error. A positive bias means SAOCOM tends to overestimate elevation, whereas negative means underestimation. It’s sensitive to outliers; a few large errors can skew the mean.
* **RMSE (Root Mean Square Error):** It represents the overall magnitude of the error, combining both systematic and random components. It’s essentially the standard deviation of residuals when the mean error is zero. RMSE is heavily influenced by large errors (since errors are squared), so it gives a sense of worst-case performance as well. An RMSE is in the same units (meters) – e.g., RMSE = 5 m means on average, elevations deviate by ~5 m.
* **Median Error:** A robust measure of central tendency of errors. Instead of the mean, they look at the median of SAOCOM – reference differences. The median is much less affected by outliers, so it better represents the typical bias when outliers are present. If the data has some extreme bad points, the median will stay close to the majority of points.
* **NMAD (Normalized Median Absolute Deviation):** A robust measure of spread (dispersion) of the errors. They likely compute NMAD as 1.4826 × median(|error – median(error)|). This is analogous to a standard deviation but uses median distances from the median error. NMAD is insensitive to a few outliers, so it reflects the “core” error scatter. For normally distributed errors, NMAD ≈ standard deviation. Using NMAD helps because the data does contain outliers (some areas with very large errors due to decorrelation) – a traditional standard deviation would blow up, whereas NMAD gives a truer picture of typical error.
* **Why Robust Metrics:** The slide explicitly asks “Why robust?” because the SAOCOM dataset has outliers (e.g., a few points might have tens of meters error, especially in problematic land covers). The median and NMAD ignore extreme tails and therefore provide a more stable assessment of performance. They complement bias and RMSE: while RMSE tells you overall error including worst cases, NMAD tells you the error spread for the bulk of the data. In practice, they evaluated both classical and robust metrics in the analysis. For example, they might report something like “Mean error = +2.0 m but median error = +0.5 m” if a few outliers raised the mean, and similarly compare RMSE vs NMAD to illustrate outlier impact.

In summary, this slide sets the stage that the validation uses both traditional error stats and outlier-resistant stats, ensuring that the reported “accuracy” isn’t overly optimistic or pessimistic due to a few anomalies.

## Slide 8: Height Statistics Summary

* This slide presents fundamental distribution statistics for all the elevation datasets (SAOCOM and the references) prior to any calibration. Likely a table of min, max, mean, and median heights for each: e.g., raw SAOCOM, TINITALY DEM, Copernicus DEM.
* **Large Range of Raw SAOCOM Data:** The SAOCOM InSAR heights showed a much larger range (min to max) than the reference DEMs. This suggests some outliers or uncalibrated values. For instance, a few SAOCOM points might have wildly incorrect heights (perhaps due to phase unwrapping errors). It indicates the raw product is “noisy” and needs filtering or calibration.
* **Differences in Mean vs Median:** They observed that SAOCOM’s mean elevation vs median elevation differ notably, whereas for a well-calibrated dataset these would be closer. For example, if SAOCOM heights overall were offset (say consistently 10 m too high), the mean would be higher than the median of reference differences. The fact they underscore this difference means SAOCOM had a systematic bias relative to the reference. The median being different from the mean implies a skewed distribution (likely skewed by outliers or bias).
* **Need for Calibration:** Because of those discrepancies, they performed a calibration – essentially adjusting SAOCOM data to reduce systematic error. In practice, the notebook shows they calculated a constant offset: at high-coherence stable points, they took the median difference between SAOCOM and TINITALY, and subtracted that from all SAOCOM heights (“Method 1: Constant Offset”). This brings SAOCOM’s median in line with the reference and addresses the bias. For example, if raw SAOCOM was on the WGS84 ellipsoid and reference was on a geoid, there could be a tens-of-meters bias; the calibration offset accounts for that.
* The table likely demonstrates before-calibration numbers such as: SAOCOM min and max much wider (maybe some very low or high outlier values), SAOCOM mean vs median differing by a few meters (the bias). Meanwhile, the reference DEMs have their own distributions (which in a local area of Verona might be somewhat similar in min/max if covering the same region). Seeing this prompted the bias correction step on the next slides. In summary, Slide 8 communicates, “The raw SAOCOM data has peculiarities (wider spread, non-zero bias) that we must correct for to ensure a fair validation.”

## Slide 9: Reference DEM Comparison

* **Visual Differencing of TINITALY vs Copernicus DEM:** This slide addresses how the two reference elevation models differ from each other. It likely shows side-by-side maps or combined imagery (perhaps an 8-panel figure as indicated) highlighting differences in elevation between the 10 m TINITALY DEM and the 30 m Copernicus DEM over the area.
* **Spatial Patterns:** The maps probably show color-coded differences (TINITALY minus Copernicus). Spatial patterns of disagreement might appear in certain terrain: e.g., perhaps in steep mountainous areas, or where Copernicus’s 30 m resolution smooths features that TINITALY captures (causing differences along sharp ridges, river valleys, etc.). The slide allows one to see where one DEM is higher than the other. For instance, if Copernicus uses a geoid and TINITALY uses ellipsoid heights, there might be a roughly constant offset visible (though likely they are both put to the same datum in preprocessing). Additionally, differences can come from the source data and resolution (TINITALY might include finer man-made structures or levees that Copernicus blurs out).
* **Difference Histogram:** A histogram on the slide shows the statistical distribution of elevation differences (TINITALY – Copernicus) across the area. If both DEMs were perfect and on the same datum, this distribution would cluster around 0. Indeed, they found a strong correspondence: likely the mean difference is near zero or only a small bias. The spread (std or NMAD) might be on the order of a couple of meters, indicating overall they agree well.
* **Findings:** The text suggests these differences underscore a need for calibration. If TINITALY and Copernicus differ systematically (for example, imagine the histogram shows a consistent +5 m bias of one relative to the other), it reveals that using one or the other as “truth” can affect bias calculations. They noted what vertical datums each uses: Copernicus is on EGM2008 geoid, TINITALY presumably on WGS84 ellipsoid or a local geoid. This could contribute a constant offset. In any case, the key takeaway is that the two references are largely consistent (high correlation ~0.99, as indicated by “strong linear relationship”), with some spatially coherent differences (maybe due to different capture methods or datum).
* Having this comparison is important because it sets the baseline: any SAOCOM error must be interpreted in light of the small discrepancies the references themselves have. The slide effectively says, “We checked our two reference DEM sources against each other – they are very close (pattern and distribution shown here), so we can trust them. Now any big errors we see with SAOCOM are truly SAOCOM’s issues, not reference inconsistencies.”

(Note: Slide 9 content is duplicated in the provided text, likely a formatting issue. The explanation above covers the intended content.)

## Slide 10: SAOCOM Spatial Coverage Analysis

* **Grid-Based Coverage Map:** Here they analyzed how much of the area SAOCOM actually covered with valid elevation points. After gridding to 10 m cells, each cell is marked as “SAOCOM data present” or “void (no data)”. The slide likely shows a map or statistic summarizing this.
* **Void Percentage = 87.0%:** This stark number means that 87% of the study area has no SAOCOM elevation data at all. In other words, nearly 9 out of every 10 pixels are voids. Only ~13% of cells had a usable SAOCOM height measurement. This result comes from counting grid cells within the convex hull of SAOCOM coverage: e.g., out of X total cells, only Y had a SAOCOM point, so Void% = ((X–Y)/X)\*100 ≈ 87%.
* **Implication:** The SAOCOM InSAR DEM is extremely sparse. The phrase “for every 10 pixels in the study area, nearly 9 have no SAOCOM data” emphasizes how incomplete the coverage is. This is a major limitation – large swaths of the map are blank. Any practical use of this DEM has to contend with these huge gaps.
* **Data Gaps Inquiry:** The slide sets up the question: what is special about those 87% void areas? Are they random, or do they correspond to specific terrain or land cover conditions? The last line (“What land cover types are in these voids?”) hints that the voids are not uniformly distributed – likely they cluster in certain land cover types (as we will see). This motivates the next analysis: overlaying the void mask with the land cover map to diagnose why SAOCOM failed there (e.g., is it all forest and water?).
* (Visually, the map on the slide might have colored areas where coverage exists and blank or dimmed areas where it’s void, illustrating that coverage is patchy and mostly absent except in certain spots – usually open or urban areas as one might suspect.)

## Slide 11: CORINE Land Cover

* **Land Cover Classification:** This slide describes how land cover was categorized in the study area using the 2018 CORINE dataset. They found **10 unique classes** present within the SAOCOM coverage region. These classes likely include: urban fabric (one or two types), industrial areas, various agricultural lands (e.g., arable land, vineyards, olive groves), forests (broadleaf, conifer, mixed), water bodies, etc. (The code listed a large CORINE classification dictionary, but only a subset occurs around Verona).
* **Processing of Land Cover Data:** The CORINE map (originally ~100 m resolution) was processed for the analysis: reprojected to match UTM coordinates and then resampled to the 10 m grid via nearest-neighbor. This means each 10 m cell got assigned the land cover code of whichever CORINE cell covered that location (no creation of mixed classes). Thus every point or pixel in the study area has an associated land cover label from the set of 10 classes.
* **Custom Color Palette:** They mention a custom color palette was applied to these classes for all maps and charts. In practice, they chose distinct, colorblind-friendly colors for each land cover category (the code defines specific RGB values for each CORINE code present). This ensures that in subsequent visualizations (maps of coverage, voids, etc.), the classes can be easily distinguished and consistently identified by color. For example, forests might be shades of green, urban areas purple/gray, water blue, etc., optimized so that even colorblind viewers can tell them apart.
* **Purpose:** Introducing the land cover data here signals that upcoming results will be stratified by these classes. Each class will be analyzed to see how SAOCOM performed (or didn’t) in those areas. The slide likely shows either a legend of the classes or a snippet of the land cover map itself, illustrating how the region is classified (perhaps showing Verona city in one color, Lake Garda (water) in another, the surrounding agricultural plains and forests in their respective colors). It sets the stage that land cover is a key explanatory variable in understanding SAOCOM’s data quality and availability.

## Slide 12: Linking Elevation to Land Cover

* **Point Tagging Process:** Every SAOCOM InSAR point was “tagged” with a land cover class. Concretely, for each filtered SAOCOM point (with known coordinates), they looked up what CORINE land cover code occupies that location. The code did this by finding the raster cell of the CORINE map in which the point falls (using the same 10 m grid). That land cover code was then attached as an attribute to the SAOCOM point.
* **Final Analysis Dataset:** As a result, they have a combined dataset where each SAOCOM elevation point knows its land cover type (urban, forest, etc.). This allows slicing and analyzing the DEM errors and coverage on a per-class basis. Slide 12 explains that this integration enables them to pose questions like:
  + “How does accuracy change between forests and urban areas?” – i.e. compute error metrics for the subset of points in forests vs the subset in urban, and compare.
  + “Which land cover types have the most data voids?” – i.e. overlay the void areas with land cover to see which classes are predominantly void of data.
* Essentially, this slide conveys that the study now has a way to correlate performance metrics with land cover. The elevation residuals (SAOCOM minus reference) and the coverage probability can be analyzed by category. Without this, one could report overall error, but now we can say, for example, “In class X the median error is Y and coverage is Z%.” This is crucial for understanding where the InSAR DEM works well and where it doesn’t.
* It turns the raw data into an “analysis dataset” with columns like: Height\_error, Coherence, LandCoverClass, etc., for each point. The following slides will leverage this dataset.

## Slide 13: Height Residuals by Land Cover

* **Error Statistics per Class:** This slide presents a table of error metrics for each land cover class (from the tagged points). “Height residual” means the difference between SAOCOM’s elevation and the reference DEM’s elevation at that point. For each class, they likely list metrics such as the median residual, NMAD, or RMSE.
* **Performance Varies Significantly:** The results show clear differences in accuracy from one land cover to another:
  + Urban areas: These show **lower error**. The slide specifically notes **NMAD < 3.5 m** for urban classes, meaning the spread of errors in urban/built-up zones is under 3.5 m. Urban points (high coherence targets like buildings, roads) gave very consistent, accurate results. Bias in urban areas may also be small after calibration. Essentially, SAOCOM DEM is quite trustworthy in towns and cities.
  + Forests: These show **much higher error**, with **NMAD > 5 m** (in some forest classes, NMAD might be substantially above 5 m, indicating very large variability). This is because in forests, the radar phase can come from the top of canopy one time, then deeper another time, or decorrelate entirely, leading to large height errors when it does produce a value. So there’s both bias and variability – e.g., SAOCOM might systematically overestimate ground height in forests (hitting the canopy) and do so inconsistently.
* By looking at the table one might see for example: Broad-leaved forest – median error +6 m, NMAD 6 m; versus Continuous urban – median error +1 m, NMAD 2.5 m. These numbers are hypothetical but illustrate the contrast the slide points out. Another class like “water” might not even have enough points for stats (since nearly all water is void, as we’ll see). Classes like “vineyards” or “pastures” might fall somewhere in between urban and forest in error magnitude.
* **Interpretation:** Land cover is clearly a controlling factor on InSAR elevation quality. The denser or more dynamic the cover (forests), the worse the performance. The table quantifies this: e.g., “Urban NMAD 3 m vs Forest NMAD 7 m” – that’s a big gap in reliability. This informs users that if they restrict to certain land covers, they can expect a certain error budget. It also shows where improvements are needed (forests have unacceptable errors for some applications).
* The phrase “performance varies significantly” is precisely what the table demonstrates – the DEM’s accuracy is not uniform but class-dependent.

## Slide 14: Void Analysis by Land Cover

* **Quantifying Voids per Class:** This slide contains a table assessing data gaps (voids) for each land cover class. Essentially, for each class it likely shows: the percentage of that class’s area that has no SAOCOM data, and that class’s contribution to the total void area.
* **Key Results:**
  + **Water bodies and forests are almost entirely voids:** In the table, water and forest classes will have a void percentage near 100%. For instance, “Water: 99% void” (meaning almost none of the water surface pixels yielded an InSAR measurement) and “Dense Forest: 90–100% void” (almost the entire forested area lacked data). These environments caused the radar coherence to drop so low that the processing failed to produce points. Practically, SAOCOM could not penetrate water (which acts like a moving mirror) or dense vegetation (random volume scattering), so those areas are blank.
  + **Vineyards and forests are the largest contributors to total void area:** This refers to absolute area of voids. Even if vineyards might have, say, 70% void rather than 100%, the Verona region has extensive vineyards. Similarly, forests cover a lot of terrain. Together these two classes cover a big portion of the void map. The table likely shows something like: forests contribute X% of all void pixels, and vineyards Y%, which are the top two contributors. For example, if forests cover large hills and are nearly all void, and vineyards cover large farmland and have moderate void rates, those two add up to the majority of void space.
* **Understanding the Void Landscape:** From this, one can conclude that if we want to improve coverage, tackling forests and vineyards would yield the biggest returns (since that’s where most voids lie). Water is also nearly 100% void but maybe occupies less area than forests in this region (e.g., mostly the lake or river surfaces – significant but not as widespread as vineyards/forest).
* The slide quantifies what Slide 20’s map will show visually. For example, it might say: Water – 99% void, contributes 5% of void area; Broadleaf Forest – 95% void, contributes 40% of void area; Vineyards – 80% void, contributes 25% of void area, etc. (illustrative numbers). So almost all water/forest pixels are void (worst relative coverage), and because forests (and to a lesser extent vineyards) cover broad regions, they dominate the void area in total.
* This analysis confirms the earlier suspicion: the data gaps are not random but are strongly tied to certain land covers – essentially those that cause low coherence (water, forest, even tall crops like vineyards during growing season). It tells us where the DEM is unreliable from a land cover standpoint.

## Slide 15: Error vs. Coherence

* **Relationship Plot:** This slide likely shows scatterplots or violin plots of height error as a function of coherence value. It captures how measurement quality improves with higher radar signal coherence.
* **Clear Trend:** As coherence (γ) increases, the elevation error decreases dramatically:
  + At low coherence (~0.3–0.4, which is near the minimum threshold used), the spread of errors is very large. The plot might show points widely dispersed (± tens of meters) and a bias possibly off zero. Essentially, low-coherence points are unreliable – they produce large random errors.
  + At high coherence (~0.8–1.0), points cluster tightly around zero error. The error distribution “tightens” – the violin plot narrows, meaning the majority of points have small deviations. Also, the median error approaches 0 as coherence approaches 1 (the red line or median in each bin is nearly at zero when γ is high). High-coherence points have very little bias and noise.
* **Interpretation:** Coherence is a strong predictor of accuracy. This validates the earlier filtering strategy (why they chose γ ≥ 0.3) and suggests even within the kept points, one could further trust only the higher-coherence subset for critical applications. The slide shows visually that a point with γ=0.9 will have a very small error bar, whereas one at γ=0.35 could be off by several meters or more.
* **Median to Zero:** The note “median error approaches zero” means any systematic bias present at lower coherences (perhaps due to phase unwrapping issues) disappears at high coherence – those points are essentially unbiased. Lower coherence points might have some residual bias (for example, if the algorithm tends to slightly overestimate heights when data is bad, you’d see median error maybe not at zero for the lowest bin). As coherence improves, that bias corrects itself (the processing finds the right phase cycle).
* In summary, this slide reinforces a key message: data quality (coherence) and accuracy go hand in hand. It provides empirical evidence that if one can increase coherence (say by multi-looking or stacking images), one would tighten the errors. It also suggests that any analysis of results should weight by coherence – for instance, an average error including many low-coherence points might look worse, whereas focusing on high-coherence points yields much better accuracy.

## Slide 16: Error by Land Cover

* **Distribution Plots per Class:** This slide visualizes the error distribution for each major land cover class, likely using violin plots or similar for each class side by side. This complements the numeric table (Slide 13) by showing the full error spread for each category.
* **Notable Shapes:**
  + Forests: The error distribution for forests is **wide and flattened**, indicating high uncertainty. The “violin” for forest classes will be very broad (large interquartile range and long tails). This means the residuals in forests vary from small to very large – a wide probability density. In practical terms, some forest points might have errors of only a few meters, but many have errors of 10 m or more, and it’s quite spread out. It visually confirms forests have both bias and variability issues.
  + Urban Fabric: The distribution for urban/built-up areas is **narrow and concentrated**. The violin for urban points is skinny, centered near zero. Most urban residuals cluster tightly around zero with a small spread (the error distribution might be a sharp peak). This indicates reliability – there are few outliers in urban area, and the typical error is low.
* **Comparison:** These contrasting shapes graphically demonstrate how much land cover matters. For example, an urban violin might mostly lie within ±3 m of zero (very narrow), whereas a forest violin might span ±10 m or more. The median line for forest could be offset (if there’s bias), and the overall shape can be asymmetric (if SAOCOM tends to overshoot ground in forests – e.g., mostly positive errors because it’s hitting tree tops). Urban’s median line will be essentially on zero with symmetry in the small spread around it.
* **Takeaway:** The narrow vs wide shapes echo the NMAD values given earlier. This visualization can also highlight any skew – e.g., if one class tends to consistently overshoot or undershoot. The note specifically says forest’s shape is “wide (high uncertainty)” vs urban’s “narrow (more reliable)”. So one can literally see uncertainty as the width of these violin plots. Any stakeholder looking at these would grasp that “if my area of interest is mostly forest, the SAOCOM DEM errors are all over the place; if it’s urban, the errors are tightly controlled.”
* This slide effectively communicates uncertainty by class in a way numbers alone might not. It reinforces the previous quantitative statements with a visual risk profile for each land cover.

(Slide 16 content was repeated in the text, likely indicating two figures for error distributions – possibly one figure per reference or splitting classes across two slides. The description above captures the general idea to avoid duplication.)

## Slide 17: Individual Land Cover Overlay

* **Spatial Distribution of a Specific Class:** This slide shows a map where one particular land cover class is highlighted on top of a satellite (or base map) image. The example given is the “Vineyards” class – a key agricultural land cover in the region.
* **What the Map Shows:** All areas classified as Vineyards are probably colored or outlined (using the custom color/hatch for vineyards) on a high-resolution satellite image of the region. This gives context about where that class is located geographically (e.g., perhaps on the gently rolling hills north of Verona). One can see how extensive it is and in what terrain.
* **Extensive Vineyards:** The note mentions the vineyards class is extensive, meaning viticulture covers a significant part of the study area. This is important because earlier analyses showed vineyards had a high proportion of voids and contributed a lot to missing data. Seeing them on the map, one might notice they often coincide with certain topography (e.g., hill slopes) which might affect radar line-of-sight or coherence (maybe the vines create periodic row structures that could cause some decorrelation).
* **Purpose of Overlay Maps:** These overlays help connect the abstract analysis back to the real world. For instance, now when we say “vineyards had many voids,” the audience can picture where those voids occur – in those green-colored vineyard regions on the actual landscape. It also allows visual inspection: perhaps some vineyard areas did have data (maybe younger vineyards or different geometry) visible as not void on later maps.
* Each class overlay (not just vineyards) was prepared similarly, but vineyards are shown as a representative example. The map is likely annotated or color-coded with a legend for that class. If the satellite image is visible beneath, one can correlate land cover classification accuracy as well (e.g., the colored overlay aligns with obvious vineyard patterns on the photo).
* In summary, Slide 17 is illustrative: it highlights one important land cover and its spatial reach, serving as a prelude to examining how SAOCOM performed in that particular class and others. It emphasizes that vineyards are a “key agricultural feature” – meaning results for this class have practical importance (e.g., if someone wanted to use the DEM for vineyard management or hydrology, they should pay attention to how it fared).

## Slide 18: Gridded SAOCOM Residuals

* **Continuous Residual Maps:** This slide describes a process where the pointwise errors (SAOCOM minus reference at each point location) were interpolated into a continuous surface. By doing this, they create an error map covering the whole area, not just at the discrete point locations.
* **Visualization of Spatial Error Patterns:** Converting scattered residuals into a grid (via interpolation) allows one to see broad patterns: where SAOCOM’s DEM tends to overshoot (positive residuals) or undershoot (negative residuals) geographically. For example, after interpolation one might see entire ridges showing a red hue (meaning SAOCOM consistently high there) or valleys in blue (SAOCOM low).
* **Methodology:** The slide indicates they did this for multiple panels (perhaps different reference comparisons or different slices). The placeholder “6-Panel” suggests they might show:
  1. Residual map (SAOCOM – TINITALY) for the entire area.
  2. Residual map (SAOCOM – Copernicus) similarly.
  3. Maybe separate panels highlighting where differences are significant vs negligible (like a mask of areas beyond certain error thresholds).
  4. Or possibly maps broken down by region (if the area is large, maybe NW, NE, etc.).  
     Each panel likely uses a color scale (e.g., red = SAOCOM higher than ref, blue = SAOCOM lower than ref, white = close to zero difference).
* **What it Reveals:** These residual heatmaps can show if errors cluster by terrain type. For instance, one might observe most red areas coincide with forests (meaning SAOCOM’s phase measured something like canopy top – making SAOCOM heights too high relative to ground). Blue spots might coincide with specific conditions, or perhaps negative errors where phase unwrapping went to the wrong cycle on some hill causing an underestimate. Patterns could also indicate if there’s any spatial bias – e.g., an entire swath of the image has positive bias (could hint at a residual uncorrected atmospheric effect or geoid difference if any).
* **Allows Visual QA:** By examining these maps, the researchers ensure there are no large-scale distortions (like a tilt or bowl-shaped error pattern). If most errors look randomly distributed with respect to only land cover and not, say, gradually increasing eastward, that confirms the geolocation and calibration are fine.
* Since the text says “visualize where SAOCOM DEM is higher or lower than the reference”, the main idea is these interpolated maps show exactly that in an intuitive way. Likely the commentators would say, “The red patches here correspond to forested hills – SAOCOM reported higher elevations (hitting canopy) than the reference ground DEM – while the flatter open areas show minimal color, meaning good agreement.” It’s a diagnostic and summary visualization of spatial error distribution.

## Slide 19: Height Correlation

* **Scatter Plots of Elevations:** This slide shows direct comparisons of elevation values: SAOCOM vs reference on an X–Y plot. Each point on the scatter is one location’s reference DEM height (x-axis, presumably) and the SAOCOM height (y-axis). If SAOCOM were perfect, all points would lie exactly on the 1:1 diagonal line.
* **Strong Linear Relationship:** The points do cluster along a line, indicating SAOCOM’s DEM is strongly correlated with the actual topography. The correlation coefficient “r” is very high (close to 1). Essentially, SAOCOM captures the shape of the terrain – higher areas in the reference are high in SAOCOM, lower areas are low, etc., with not much systematic distortion in horizontal positioning or overall relief.
* **Spread Around 1:1 = Error:** However, the points are not perfectly on the line; there’s a cloud of dispersion around it. The vertical deviation of points from the diagonal represents the error at those points (SAOCOM minus reference). The slide notes that the “spread… visually represents the error.” In practice, you might see a fuzzy band around the line – the thickness of that band is roughly the typical error magnitude. If the band is thin (~a few meters), that’s good; a thick band means big deviations.
* For example, if one zooms in on the scatter, a reference height of 200 m might correspond to SAOCOM points mostly between, say, 195 m and 205 m (which would be a ±5 m error band). In forest areas it might be wider. The correlation being high means this band is centered on the line – just spread. If there was a bias, the whole cloud would be offset above or below the line (but calibration removed major bias, so it’s centered).
* **Interpretation:** This plot is another way to evaluate accuracy: one can visually confirm no major non-linear biases (like if low elevations had different bias than high elevations – that would show up as curve or slope change). A strong linear scatter with constant scatter indicates errors are more or less random (with some dependence on cover as we know, but not e.g. worse at higher altitudes specifically).
* In summary, Slide 19 assures that overall, SAOCOM DEM correlates well with ground truth (the “shape” of terrain is right). The remaining issue is the random/scatter error around that line – which we quantified earlier with RMSE, NMAD, etc. The visual check via scatter plot likely accompanies a statement like “Correlation coefficient ~0.99. The trend line is unity (slope ~1), but there is visible scatter corresponding to up to a few meters of error.” It’s a sanity check on the linear fidelity of the DEM.

## Slide 20: Land Cover Inside the Voids

* **Masking Voids by Land Cover:** This slide presents a map that highlights only the areas where SAOCOM failed to produce data (the voids) and then color-codes those areas by their land cover type. Essentially, they take the void mask (87% of area) and within that mask, display the CORINE land cover classification. The rest (non-void areas) might be shown in grey or omitted, focusing attention on void zones only.
* **Clear Domination by Forest, Vegetation, Water:** The map vividly shows that the void areas are almost entirely comprised of certain classes: primarily forests (various shades of green, if multiple forest types), vegetated natural areas, and water bodies (blue). In contrast, land covers like urban or open fields are scarcely visible in the void map (because those had coverage). For instance, the white areas in previous maps (voids) now appear, say, green (forest) or blue (water) on this map. This confirms quantitatively what Slide 14’s numbers indicated: the reason those places have no data is because they are forests, water, etc.
* Example: If one looks at a mountain range area on this map, it will be colored as “Broad-leaved forest” or “Mixed forest” – demonstrating that those mountains are forest-covered and that’s why they’re void. The Lake or river appears as a blue patch of void – water class. Areas that were vineyards likely show up in whatever color was assigned to vineyards, indicating many voids are in those zones too.
* **Decorrelation as the Cause:** The caption explicitly states these classes dominate the gaps, “pointing to decorrelation as the cause.” This is important – it ties the phenomenon back to the physics: In SAR interferometry, heavy vegetation causes temporal decorrelation (the radar signals change between passes due to vegetation growth/movement), and water causes very low return coherence (a water surface changes constantly and reflects radar away). Thus, it wasn’t an issue of, say, satellite coverage or processing area – it was these land cover types inherently made the InSAR measurement impossible or invalid.
* **Takeaway:** If someone sees this map, the takeaway is straightforward: “Wherever we don’t have data, it’s basically because it’s forest or water or similar.” Conversely, “If it’s urban or simple farmland, we mostly did get data.” It visually correlates land cover with data voids one-to-one. This kind of map could also serve a user planning data fusion – it tells which areas one would have to fill in using other sources (almost exactly the forested and water regions).
* In sum, Slide 20 is powerful evidence connecting cause (land cover) and effect (voids). It validates the earlier analytical approach by direct observation: the blank areas on the SAOCOM DEM align with specific challenging land cover categories.

## Slide 21: Quantifying Void Contributors

* **Bar Charts of Void Impact:** This slide presents two charts to pinpoint which land cover classes are most problematic regarding voids: one by relative impact and one by absolute impact.
  + **Worst Relative Coverage (Left Chart):** This bar chart ranks classes by the percentage of their area that is void (as calculated earlier). For example, Water likely tops this list at ~100% void (virtually no coverage). You might see Forests (broad-leaved, coniferous) also near the top with ~90%+ void. Perhaps Wetlands or certain vegetated mixes also high. This chart basically mirrors the “Pct\_LC\_is\_Void” column of the table: it tells us which land cover types SAOCOM consistently failed in. The higher the bar, the worse SAOCOM’s coverage in that class. So a bar at ~100% (water) means SAOCOM did not work over that class at all.
  + **Largest Contributors to Total Void Area (Right Chart):** This chart ranks classes by how much of the total void **area** they account for. Here, Forests likely come out on top. If, say, 50% of all void pixels are in forests, the forest bar will be highest. Vineyards might be second if they cover a big region that is largely void. Water might have a high void percent but might not contribute as much total area (if the water bodies in the study area are limited in extent). This chart highlights, for resource allocation, which voids matter most in aggregate. For instance, if forests constitute, say, 40% of all voids and vineyards 20%, together those two classes make up the majority (~60%) of missing data – indicating these should be priority targets for improvement or alternate data sources.
* **Interpretation:**
  + The left chart’s tallest bars confirm the classes with inherently poor InSAR performance (water, dense forests, possibly marshes or snow if present). These are classes where you expect trouble due to physics (e.g., water’s coherence ~0, heavy forest’s coherence very low).
  + The right chart’s tallest bars tell us where most of the void problem lies region-wise. Likely Forests are number 1; maybe Agricultural mosaic or vineyards next. For example, suppose Broad-leaved forest contributed 30%, Coniferous forest 15%, Vineyards 15% – those three would be ~60% of all voids. That means if one could somehow retrieve data in forests and vineyards, the void percentage would significantly drop.
* **Water vs. Forest Example:** Water might have a bar near 100% on the left (worst coverage) but possibly a smaller bar on the right (if water bodies cover, say, only ~5% of the area, then they contribute 5% to voids). Conversely, broadleaf forest might have ~95% void on the left and also a huge bar on the right (because they cover, e.g., 20% of area, contributing a big chunk of voids). This two-pronged view is helpful: the left says “are we ever getting data here?” the right says “how much does it matter overall?”
* By presenting both, the slide communicates: Water is essentially always void but maybe limited in area; forests are almost always void and very extensive, making them the biggest concern. And indeed, that’s likely why “water” leads in relative void, while “forests” lead in total void area, exactly as noted.
* This quantification supports decision-making for future work (e.g., focus on forest improvements). It also neatly summarizes the void issue in a class-by-class manner for the audience.

## Slide 22: Coverage and Voids Across the Landscape

* **Composite Land Cover & Void Map:** This slide shows the entire study area with all land cover types drawn (each class in its designated color), and it indicates SAOCOM voids explicitly. They mention white areas indicating voids. Likely, the map is fully colored by land cover, but wherever SAOCOM had no data, they left it blank or white. So one can see a patchwork of colors with many white holes.
* **Powerful Summary Image:** This single image encapsulates the spatial distribution of all classes and highlights exactly where data is missing. The fact that 87% is void is visually evident – the map would appear mostly white, with only fragments of color. It dramatizes how fragmented the actual coverage is. For example, you may see colored polygons (urban, fields) as islands surrounded by white (forested hills or water), giving an intuitive “swiss cheese” appearance.
* **Fragmentation:** The term “fragmented nature of the coverage” is apt – coverage is not in one contiguous block but in scattered patches. You might notice, for instance, thin colored corridors where roads or towns are (where SAOCOM gleaned some data), cutting through large white regions (forests) where it didn’t. Or open farmland (some color) dotted within mostly void surroundings.
* **87% Void Statistic Visualized:** The map provides a spatial feeling for that 87% number. Nearly nine-tenths of the map is white, reinforcing how incomplete the dataset is if one looks across the landscape. It’s one thing to say “87% void”; it’s another to see an almost ghosted map with just a few colored bits of data – it really drives home the limitation.
* **Interpretation:** After seeing this, one understands that using SAOCOM DEM “as-is” for the whole area is untenable – the data coverage is too sparse. It also highlights exactly where one has data (so if your area of interest coincides with those colored patches, you’re in luck; if not, there’s a gap).
* This image is likely what the presenters call a “powerful summary.” It succinctly shows both the distribution of land covers and how voids carve them out. For instance, one could point to a large white chunk and note “that’s a forest – no data there,” then to a small adjacent colored area “that’s a town – data exists there.” The viewer can correlate with known geography (maybe see Verona city area covered, Lake Garda area white, etc.).
* In summary, Slide 22 communicates the grand overview: **the SAOCOM DEM coverage is highly fragmented and largely absent (white) in the study area, with coverage only in certain land cover pockets.** This sets the context for the concluding points about the dataset’s utility and limits.

## Slide 23: Coverage vs. Voids – A Closer Look

* **Zoomed-In Class Example:** This slide (and possibly a continuation on another Slide 23 if duplicated) zooms into one particular class or region to examine within that class how SAOCOM coverage versus voids are distributed. They mention it’s essential for assessing data utility for specific applications like forestry. So likely, they zoom into a forested region.
* **Partitioning One Land Cover into Covered vs Void Areas:** For example, consider a large forested area: this map would highlight that forest area and differentiate parts of it where SAOCOM did manage to get some coverage vs parts that remained void. Perhaps they use two colors or overlay patterns: one shade for “forest areas with data” and another (or white) for “forest areas without data.” The result is a patchy pattern within the forest zone.
* **Why This Matters (Use-case perspective):** If one’s interest is forestry, you’d want to know “did the DEM capture any of my forest, or is it mostly missing?” This zoomed view shows that even in a single class, coverage can be sporadic. For instance, maybe along the edges of the forest or in younger plantations the radar got a return (those appear as covered), but in the dense core of the forest it’s void. Such detail is crucial if, say, you planned to use the DEM for forest management – you’d instantly see which parts of your forest have data and which you’d have to survey by other means.
* **Likely Visual:** The map might have the forest area outlined, with, say, green shading where SAOCOM provided a height and transparent or hatched fill where it did not. The rest of the classes might be greyed out or not shown to focus on that class. Possibly, they chose one of the major void classes (like broadleaf forest) for this demonstration.
* Since this slide specifically references forestry as an example, it suggests they zoomed in on a known forest region. They demonstrate how within that forest, only a few points (if any) have coverage – reinforcing that for forestry applications the L-band InSAR dataset alone is insufficient without those gaps filled.
* **Conclusion from Closer Look:** The audience can appreciate that even within one land cover patch, the InSAR coverage is uneven. Some micro-areas of that forest might have slightly better conditions (maybe a clearing or a seasonally dry period between images) that yielded data, while most did not. It underscores the previous findings on a local scale.
* If Slide 23 was duplicated, possibly two examples were given – e.g., one zoom on a forest area, another on an agricultural area like vineyards. Each would show that within those classes, coverage is partial. The notes for both would be analogous: demonstrating the patchiness of data within a homogeneous land cover.
* Overall, this closer look connects the broad statistics to real-world scenarios: for a given land cover (like a specific forest or vineyard block), how does the presence/absence of SAOCOM data look, and what does that mean for using the DEM in that context.

(The content for Slide 23 appears twice in the provided text, suggesting perhaps two separate zoom-in slides for two different classes or areas. The explanation above covers the general approach of these “closer look” slides.)

## Slide 24: Conclusions

* **Accuracy Depends on Coherence & Land Cover:** The first conclusion is that SAOCOM can produce accurate elevation data **under favorable conditions**, but its accuracy is **highly variable** depending on signal coherence and what type of land cover is present. In practice, this means: in high-coherence areas (flat, stable surfaces with good reflection – e.g., urban, bare soil), the vertical accuracy was quite good (errors of only a few meters or less). However, in low-coherence areas (thick forests, water, very rough terrain with changing scatterers), the performance degrades severely – if data exists at all, it has large errors. This dependency was demonstrated by the error vs. coherence trends and the class-by-class error stats. So SAOCOM’s DEM is **not uniformly accurate**: it excels in certain terrains and fails in others.
* **Coverage is the Primary Limitation (87% void):** Perhaps the biggest takeaway: the SAOCOM InSAR DEM had data for only ~13% of the area, leaving **87% as voids**. These voids were not randomly scattered; they were concentrated in specific environments – notably **forested areas, complex vegetation, and bodies of water** (as the analysis showed). This means the dataset’s usefulness is fundamentally limited by its lack of coverage in the majority of the landscape. Even if the 13% covered parts are accurate, the fact that nearly nine-tenths of the area are blank makes it problematic as a standalone elevation source. Essentially, the technology struggled wherever coherence was low (which corresponded to those land covers). This is identified as the “primary limitation” because no matter how accurate the good points are, one cannot ignore that most of the map has no points.
* **Best Suited vs Less Reliable Use Cases:** Given the above, the SAOCOM DEM is **best suited for areas like urban regions and open agricultural land** – places with high signal stability. In such areas, we saw good accuracy and at least some coverage. These are scenarios where the L-band can bounce off stable structures or ground and maintain coherence (for example, city infrastructure or plowed fields in dry conditions). In contrast, the DEM is **much less reliable in vegetated or complex terrain** – essentially any area with tall, dense vegetation (forest canopy) or dynamic surfaces (water, wetlands) cannot be well measured by a single-pass L-band InSAR like this. In those areas, either no data was obtained, or the error margins are so large that the heights can’t be trusted for fine applications. For instance, using this DEM for detailed topography in a forested mountain or for bathymetry of water bodies would be futile. But using it for an urban floodplain’s elevation might be fine.
* **Summary of Findings:** In one sentence – **SAOCOM’s InSAR DEM works where the radar signal remains coherent (open, man-made surfaces) and fails where it decorrelates (forest, water), leaving most voids there**. Thus, one can rely on it in certain land cover contexts and must find alternatives in others. The conclusions emphasis likely aligns with expectations for L-band: it penetrates more vegetation than shorter wavelengths, but clearly still had major issues with dense forest cover and long temporal baselines that caused decorrelation.
* These conclusions wrap up the validation: they highlight where the product met requirements (accuracy in ideal spots) and where it didn’t (coverage gaps, vegetated terrain). This informs any potential user or stakeholder about the inherent limitations and appropriate use of the data.

## Slide 25: Future Work

* **Data Fusion to Fill Voids:** The first item proposes combining SAOCOM DEM data with other sources to address the extensive gaps. “Data fusion” could involve integrating a secondary DEM or height source in void areas. For example, they might merge SAOCOM with optical stereoscopic DEMs, LiDAR data, or the Copernicus DEM for those empty regions. The idea is to **“intelligently” fill the 87% voids** by leveraging information from elsewhere – perhaps weighting by reliability. This might entail, say, using Copernicus 30 m heights in forest voids (with some adjustment) or using multi-temporal SAOCOM acquisitions to get some coverage where a single pass failed. Essentially, since single-pass left huge holes, draw on other sensors or multiple passes to plug those holes so you end up with a more complete DEM.
* **Test Different InSAR Processing Parameters:** This suggests that the processing choices (coherence threshold, filtering techniques, unwrapping methods, timeframe between image pairs, etc.) could be varied to see if coverage or accuracy improves. For instance, what if they lowered the coherence threshold a bit to retrieve a few more points (albeit noisy ones)? Or used a multi-look approach to raise coherence at the expense of resolution? Or tried shorter temporal baseline data (images taken closer together in time to reduce decorrelation in vegetation)? They could also experiment with processing in summer vs winter (leaf-off) or using alternate interferometric combinations. The goal would be to **increase coverage or accuracy by tuning the algorithm** – maybe capturing more marginal points or reducing outliers. They might explore algorithms like filtering the phase differently or using external data in phase unwrapping to avoid errors. This line of work acknowledges that the current parameters were a first attempt and that by tweaking them, one might recover some of the 87% void or reduce error spread.
* **Analyze Influence of Seasonality/Temporal Changes on Coherence:** This is particularly aimed at understanding (and potentially mitigating) the decorrelation that caused voids. Different seasons can drastically change coherence for L-band SAR: e.g., in winter deciduous forests lose leaves, possibly allowing higher ground coherence; in dry season fields have less biomass or water content, possibly improving coherence. Also, shorter time intervals between radar acquisitions yield higher coherence (less change on ground). Future work would involve studying coherence values and DEM results obtained at different times of year or with varying temporal gaps. For instance, one could attempt an interferometric pair during a stable weather period vs one spanning heavy rain and growth. By **quantifying how coherence — and thus DEM completeness — varies with season or time**, they can strategize data collection for better results (e.g., task SAOCOM to image forests in late autumn after leaf fall). It also helps predict where and when L-band InSAR is viable.
* **Overall Aim:** These future work items are targeted at overcoming the identified weaknesses. Filling voids is a direct response to the 87% gap problem. Tuning processing and timing is a response to the coherence/land-cover issues. The underlying message is optimistic: We can likely improve upon these results. For example, using multi-temporal integration (like persistent scatterer or multi-pass averaging) could raise coherence in vegetation and yield some data where now there is none. If a second SAOCOM pass a few days apart was used, maybe some currently void pixels would have become coherent and provided elevation. Or combining ascending/descending passes might resolve some layover issues.
* In summary, Slide 25 lays out a roadmap: **1)** Combine SAOCOM with other data to get a more complete and robust DEM, **2)** Optimize InSAR processing (and perhaps acquisition strategy) to improve coverage/quality, and **3)** Study how natural changes affect L-band performance so future data acquisitions can be timed or processed to minimize decorrelation. These steps aim to make the next iteration of the DEM validation much more successful in challenging areas.