# Interpolation of Airborne LiDAR Data for Archaeology

Benjamin Štular, Edisa Lozić, Stefan Eichert

Title

# Interpolation of Airborne LiDAR Data for Archaeology

Authors

Benjamin Štular[a], Edisa Lozić[a,b,*], Stefan Eichert[c]

[a] Research Centre of the Slovenian Academy of Sciences and Arts, Novi trg 2, 1000 Ljubljana, Slovenia

[b] Institute of Classics, University of Graz, Universitätsplatz 3/II, 8010 Graz, Austria

[c] Natural History Museum Vienna, Burgring 7, 1010 Vienna; Austria

* Corresponding author. E-mail address: edisa.lozic@graz-uni.at (E. Lozić).

## Abstract:

The use of topographic airborne LiDAR data has become an essential part of archaeological prospection, and the need for an archaeology-specific data processing workflow is well established. However, interpolation, an important step in which a rasterized surface is derived from the classified point cloud, has received little attention from archaeologists. This processing step, also known as rasterization or gridding, has a direct impact on the accuracy and visual quality of the digital elevation model, but remains a challenge despite numerous studies. Numerous studies have compared the accuracy of different interpolators with conflicting results, but very few compare the visual precision. Also, there are no archaeology-specific studies. This paper addresses this problem by providing an archaeology-specific visual assessment of precision. Six of the most commonly used interpolators were tested at four test sites with an innovative use of triangular assessment method. Kriging was the best interpolator in undersampled areas and inverse distance weighting was a distant second. In other areas, triangulation with linear interpolation was marginally better than Kriging. However, when availability and computational costs are also taken into account, inverse distance weighting is the most suitable archaeology-specific interpolator. In addition, we propose a hybrid interpolator that combines the strengths of triangulation with linear interpolation and inverse distance weighting (QGIS plug-in).

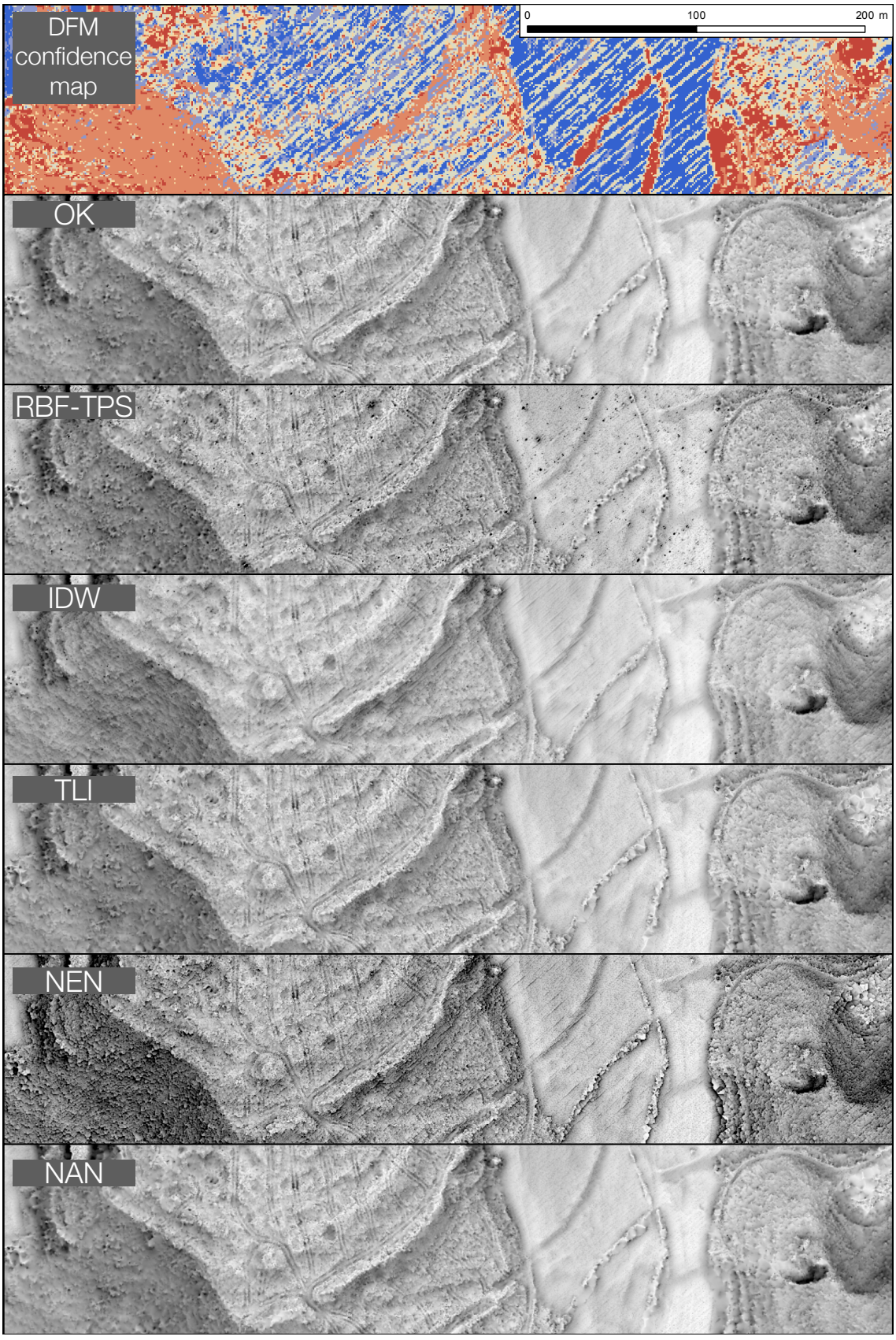
## Keywords:

Archaeology, airborne LiDAR, ALS, DEM, interpolation, rasterization, QGIS plug-in.

## Highlights

- Airborne LiDAR data is an essential part of archaeological prospection
- Archaeology-specific data processing workflow is essential
- Interpolation (also rasterization or gridding) is an important processing step with direct impact on the accuracy and visual quality of the digital elevation model
- Archaeology-specific visual assessment of precision for six interpolators has been performed
- Inverse distance weighting is the most suitable archaeology-specific interpolator
- Open source tool to calculate DFM confidence map in QGIS is included

## Graphical abstract



# 1. Introduction

As stated previously (Štular and Lozić, 2020), the use of topographic airborne LiDAR data (also known as airborne laser scanning or ALS) has become an essential part of archaeological prospection, e.g., (Chase et al., 2020; Doneus et al., 2020). Using a combination of perception and understanding, archaeologists interpret enhanced visualizations of high-resolution digital elevation models (DEMs) interpolated from classified point clouds (Challis et al., 2011; Crutchley, 2006; Lozić and Štular, 2021). The results have proven to be very efficient in detecting archaeological features and have already drastically changed our understanding of archaeological sites, monuments, and landscapes, especially in forested areas, e.g., recently, (Inomata et al., 2020; Laharnar et al., 2019; Menéndez Blanco et al., 2020; Stanton et al., 2020).

The need for an archaeology-specific data processing workflow is well established (Crutchley and Crow, 2010; De Boer et al., 2008; Doneus et al., 2020; Doneus and Briese, 2011; Fernandez-Diaz et al., 2014; Grammer et al., 2017), but we will briefly reiterate the reasons. First, visual inspection of enhanced raster visualization remains the main method of analysis in archaeology. Given that most machine learning approaches are based on enhanced visualization their importance is likely to remain central for the foreseeable future. This implies that data processing needs to focus on (visual) precision rather than (statistical) accuracy. Second, morphologically speaking archaeological features are anomalies, i.e., sporadically occurring atypical features of the terrain. Direct application of existing generic data processing methods is therefore not ideal. Third, the time, effort, equipment, and human resources invested in airborne LiDAR data processing are only a small part of a typical archaeological project. This means that some methods considered too time-consuming for general data processing, such as manual reclassification, are acceptable in archaeology. Finally, we are currently experiencing an unprecedented expansion of archaeological applications of general-purpose low- or medium-density airborne LiDAR. Therefore, any development for archaeology-specific data processing should be optimised for such data (Štular and Lozić, 2020).

In a recent review paper, the archaeology-specific data processing workflow was divided into 18 steps, ranging from raw data acquisition and processing, point cloud processing and product derivation, to archaeological interpretation, dissemination and archiving . Of these, only visualization (Lozić and Štular, 2021)(Challis et al., 2011; Hesse, 2010; Kokalj et al., 2011; Kokalj and Hesse, 2017; Kokalj and Somrak, 2019; Štular et al., 2012) and ground point filtering (Doneus et al., 2020, 2008; Opitz, 2013; Štular and Lozić, 2020) have received ample attention from archaeologists. However, between point cloud filtering and enhanced visualization lies an important step in which a rasterized surface is derived from the classified point cloud: interpolation. This step, also referred to as rasterization or gridding, has a direct impact on the accuracy and visual quality of a DEM, but remains challenging despite numerous studies. A recent review concluded that existing methods for generating DEM have difficulties when used in sharply changing terrain, in areas with dense non-ground features, and in complicated landscapes (Chen et al., 2017). Two main reasons are the large volume of airborne LiDAR datasets and inconsistent data point density. The latter imposes the need to create DEMs from data with a resolution locally smaller than the distance between data points (Chen et al., 2018). Given the inherent imperfections in any LiDAR-derived DEM, accuracy assessment is an important part of interpolation. A recent review paper addressing accuracy assessment has brought to light two facts of interest to archaeology: visual analysis is scarce and clear criteria for segmentation remain unresolved (Mesa-Mingorance and Ariza-López, 2020).

In contrast to the extensive literature on DEM interpolation, we are not aware of a single paper devoted exclusively to archaeology-specific DEM interpolation. First, archaeological DEM (Doneus et al., 2020), that we terme digital features model (DFM) (Štular and Lozić, 2020) is specific. In particular, it includes some off-terrain features that are very sensitive to the interpolation. Second, interpolation has only been briefly considered in a handful of studies. Boer and colleagues compared inverse

distance weighting, triangulation with linear interpolation (listed as TIN), and Kriging in a high point density scenario. They concluded that triangulation had the least smoothing effect and was therefore the most appropriate (De Boer et al., 2008). The same interpolators were evaluated in a study that focused on the unique conditions of Mesoamerica. Fernandez-Diaz and colleagues found that the three interpolators were comparable in a high point density scenario, but when point density decreased, Kriging performed best (Fernandez-Diaz et al., 2014). One of the spline interpolators was found to be superior to inverse distance weighing without validation by Riley and Tiffany (Riley and Tiffany, 2014). Rochelo and colleagues briefly mention that the natural neighbour was used as the most suitable interpolator (Rochelo et al., 2015) and we advocated Kriging (Štular and Lozić, 2016), especially compared to the nearest neighbour (Lozić and Štular, 2021). In an interesting study, Humme and colleagues proposed Kriging as a tool for enhanced visualization (Humme et al., 2006).

The lack of archaeological interest in interpolation partly reflects the notion that visualization and ground point filtering have a greater influence on the results of archaeology-specific data processing (Doneus et al., 2020; Kokalj and Somrak, 2019; Štular and Lozić, 2020). However, neglect is often a consequence of the blackboxing approach to airborne LiDAR data processing (Doneus et al., 2020; Doneus and Briese, 2011; Latour, 1999; Lozić and Štular, 2021)

This paper addresses the existing shortcomings by finding the most suitable interpolator (one or more) for archaeology-specific processing of airborne LiDAR data. Six of the most commonly used interpolators were described, their accuracy reviewed, and their archaeology-specific precision tested at four test sites. IDW interpolator was also evaluated for implementation in free and low-cost software.

# 2. Materials and Methods

## 2.1 Test sites

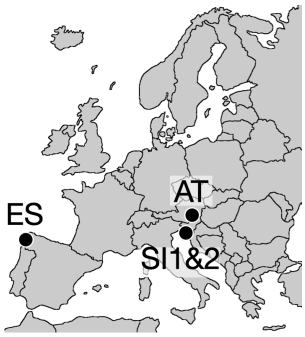
Four test sites were selected to simulate the most common archaeological use cases of airborne LiDAR data (Figure 1). Due to the availability of data and first-hand experience, all datasets are from Europe. For the comparison of the different steps of data processing, these are the same test sites that have already been used for ground point filtering analysis (Štular and Lozić, 2020). The description of the test sites is therefore only briefly repeated.

The test data from Austria (AT, ), Slovenia (SI1, SI2), and Spain (ES) were selected based on their archaeological and morphological similarity. Each site has a hilltop settlement (archaeology), buildings (modern), vegetation on steep slopes, and sharp discontinuities. An additional test site (SI2) was selected as an example of relentlessly dense low vegetation. Each test site is 1000 x 1000 m, but only the most relevant windows are shown in figures. All datasets are from nationwide data acquisitions and are in the public domain under various licences. The main difference between them is the average point density (Table 1). Therefore, the datasets represent respectively, high, medium, and low point density scenarios.

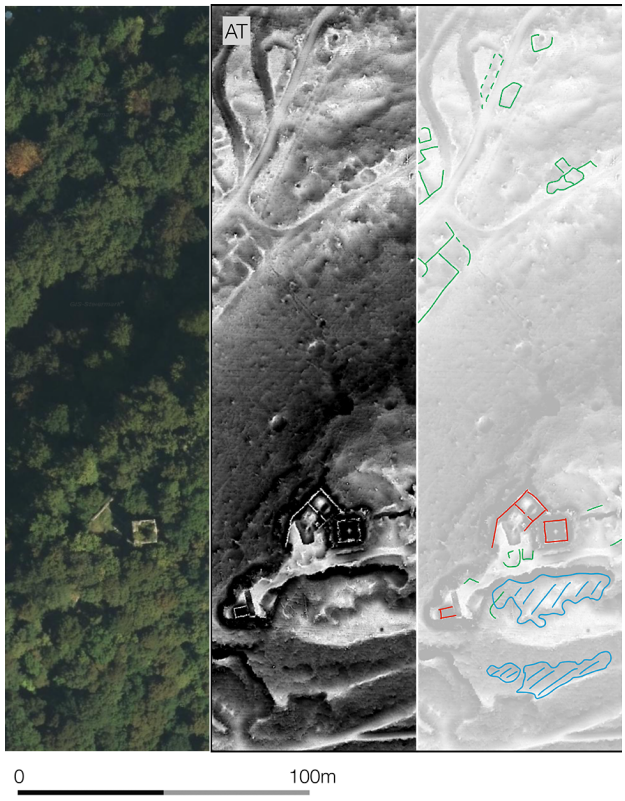
**Table 1.** Data point density for test sites: Pnts $10^6$– No. of all data points in millions; Pnts/m² – median point density per m² (average density is equal to Pnts $10^6$); Pnts $10^6$ class 2&6 – No. of points used for interpolation (ASPRS classes 2 and 6); Pnts/m² - median density of points used for interpolation per m² (ASPRS classes 2 and 6). Spacing – average spacing between the points used for interpolation (ASPRS classes 2 and 6).

| Filter | Pnts $10^6$ | Pnts/m² | Pnts $10^6$ (class 2&6) | Pnts/m² (class 2&6) | Spacing (m) (class 2&6) |
|---|---|---|---|---|---|
| AT | 12.11 | 15.54 | 7.52 | 8.96 | 0.33 |
| SI1[1] | 5.37 | 5.91 | 3.23 | 4.37 | 0.68 |
| SI2[1] | 5.39 | 5.81 | 4.65 | 5.23 | 0.62 |
| ES | 1.03 | 1.26 | 0.54 | 0.63 | 1.26 |

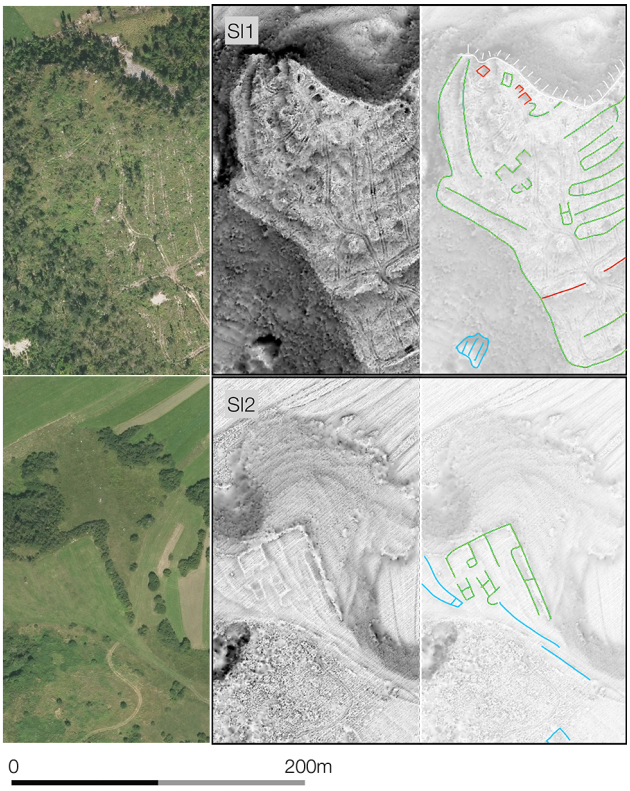
[1] Due to the specific raw data processing each point has a permutated 'shadow point'. This means that the average point density and point spacing are double of the actual data capacity. The values in the table reflect the actual data capacity, which is half of the values read from metadata.



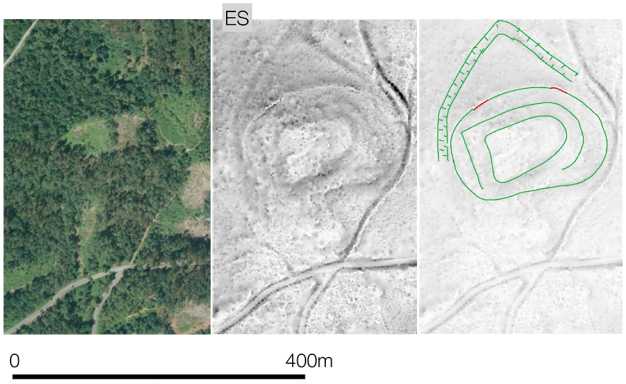

(a)




(b)




(c)




(d)

Figure 1 (after Štular and Lozić, 2020, Figure 1, CC-BY 4.0). Test data: **(a)** location of test sites: AT - N 46°53'05", E 15°30'48", SI1 - N 45°40'21", E 14°11'40", SI2 - N 45°40'56", E 14°12'25", ES - N 42°44'30", E 8°33'02"; **(b)** test site AT; **(c)** test sites SI1 and SI2; **(d)** test site ES. (b, c, d): shown at 100% crop size; left-digital orthophoto; middle-enhanced visualization of manually processed DFM; right-archaeological features (blue-embedded features, green-partially embedded features, red-standing features). DFMs are visualised with sky view factor (RVT 2.2, default settings).

## 2.2 DFM quality assessment

In this work, the DFM confidence map was used as a tool for visual assessment. DFM confidence map is based on classification and regression tree analysis (CART), which is commonly applied to remote sensing data (Bonham-Carter, 1994; Lawrence, 2001; Shao and Lunetta, 2012). We have modified the CART application to DEM uncertainty prediction map (Bater and Coops, 2009; Montealegre et al., 2015; Simpson et al., 2017) to the specific needs of archaeology-specific DFM (ŠtularEt2021). We have included an open source QGIS plug-in to calculate DFM confidence map (EichertEt2021).

## 2.3 Triangular assessment with ternary graph

The inconclusiveness of the previous studies shows that there is no best interpolator, but also that further statistical comparison of accuracy does not help the matter. Therefore, we departed from the established method of comparison. Our working assumption was that accuracy is roughly comparable across all interpolators and is not critical for archaeology-specific assessment. Therefore, we focused on archaeology-specific visual assessment of DFM precision. We assessed three key elements: precision, computational cost, and filter availability.

Precision was assessed using a knowledge-based descriptive approach based on visual inspection and comparison of datasets by two archaeologists with extensive experience in interpreting airborne LiDAR data. The assessment was based on segmentation according to the DFM confidence map and assessed. For example, areas with low confidence were assessed separately from the areas with high

confidence. This assessment is inherently subjective and it is expected to be biased towards archaeology-specific qualities. The weakness of this method is that there is no objective "truth" with which to compare the results. This was mitigated by first-hand experiential knowledge, i.e., field observation, for sites AT, SI1 and SI2 . The site ES was compared with vertical aerial photographs. The assessment was performed on 1 km$^2$ tiles, but only a smaller window is shown in the images.

The computational cost was estimated with time. Since only some interpolators are capable of parallel processing, times are not the best estimator of actual computational cost, but are a reliable indicator of computational practice.

The availability of filters is also a knowledge-based category, i.e., we queried suitable free and low-cost software packages known to us. Next, we compared them in practice to assess their performance. We focused on solutions capable of handling field-scale processing of airborne LiDAR data. The availability of each interpolator in the relevant software we are aware of is shown in the Table 2.

**Table 2.** Availability of interpolators in free or low cost software suitable for production processing of airborne LiDAR data.

| Software/Interpolator | OK | RBF-TPS | IDW | TLI | NEN | NAN |
|---|---|---|---|---|---|---|
| Golden Software Surfer v.19.2 | Y | Y | Y | Y | Y | Y |
| WhiteboxTools v.1.4.0[2] | | Y | Y | Y | Y | |
| lidr v.3.1.1[2] | Failed | | Y | | | |
| GRASS GIS v.7.8.3[1,2] | | | Y | | Y | |
| GDAL v.3.1.2[1,2] | | | Y | | Y | |
| SAGA GIS[1,2] | Failed | | Failed | Y | Y | Y |
| PDAL 2.1.0 | | | Y | | Y | |
| LAStools 1.4 | | | | Y | Y | |
| Global Mapper 21.0 | | | | Y | Y | |

[1] Does not accept LAS/LAZ as an input.

[2] See Table 2 for implementation details.

To compare precision, cost, and availability, we opted for the assessment triangle method. This method is broadly defined as a process of evidence-based reasoning from three underlying key elements (Pellegrino et al., 2001). We used precision, cost, and availability.

We modified the assessment triangle method by plotting the results on a ternary graph. The ternary graph is widely used in science. Archaeological examples include the age structure plot in zooarchaeology (Stiner, 1990) and the soil texture triangle (Saxton et al., 1986), which is commonly used in field archaeology. The ternary graph provides a two-dimensional representation of the ratios of three variables that sum to a constant. For example, each soil sample has different proportions of clay, silt, and sand, but the sum is always 100%. The latter seems inconsistent with the assessment triangle, since it implies that the result of each assessment is a constant. In our case, for example, plotting the results of the assessment triangle of OK and NEN always gives 100%, which would seem to make both equally good. However, our starting position is that each estimated interpolator is good enough to have been in use for decades, i.e., each is "equally" good. By plotting the results on a ternary graph, we estimate how balanced each interpolator is. And the best balanced one, we argue, is the most appropriate. Such an assessment triangle method with ternary graph can be applied to any three- or multi-dimensional problem, although the graphical representations become more complex as the number of dimensions grows.

We evaluated six of the most common algorithms (OK, RBF-TPS, IDW, TLI, NEN, and NAN) using field scale data from four test sites.

# 3. Theory

Interpolation, also called rasterization or gridding, is a process of fitting a surface to point elevation data by effectively draping a grid of defined cell size over the point data. In mathematics it is an approximation procedure and in statistics it is an estimation issue. This means that the points are derived from the original data and do not consist of the actual data points. The value of each cell is predicted by a specific variable based on measurements at points within the area of interest. The variables are known as interpolators. Interpolators can be exact or inexact, depending on whether or not the interpolated area passes through the points, but no method of interpolation is 100% accurate except at the points through which the function was fitted. Two implicit assumptions are that the terrain surface is continuous and smooth and that there is a high correlation between adjacent data points (Dong and Chen, 2018; Moore et al., 1991; Shepard, 1968).

There are many different interpolation methods used in geosciences to interpolate grid surfaces from airborne LiDAR point clouds. They can be divided into deterministic and geostatistical methods. The former assumes that each input point has a local influence that decreases with distance, for example, while the latter considers the statistical relationship between sample points (Xiaoye, 2008). The most commonly used geostatistical interpolator for terrain gridding is Kriging (Bonham-Carter, 1994; Longley et al., 2015; Oliver and Webster, 1990). Among deterministic the most common are radial basis functions with thin plate splines (RBF-TPS) (Chen and Li, 2019; De Smith et al., 2018; Desmet, 1997), inverse distance weighting (IDW) (Longley et al., 2015; Shepard, 1968; Wood and Fisher, 1993), triangulated irregular network with linear interpolation (TLI) (Bonham-Carter, 1994; Chew, 1989; Lee and Schachter, 1980), nearest neighbor (NEN) (Bonham-Carter, 1994; Chen and Li, 2019), and natural neighbor (NAN) (Abramov and McEwen, 2004; Habib, 2021; Sambridge et al., 1995). These methods are very well known so they are only presented summarily (Table 3).

**Table 3.** Key characteristics of tested interpolators.

| Interpolator | Type | Search | Strength | Weakness | Best for |
|---|---|---|---|---|---|
| Kriging (OK, etc.) | Inexact | Variogram | prediction | comp. cost | irregular undersampled data |
| Thin plate splines (RBF-TPS) | Exact | Distance | prediction | comp. cost | irregular undersampled data |
| Inverse distance to a power (IDW) | Exact | Distance | comp. cost | artefacts | regular or oversampled data |
| Tri. w/ Linear Interpolation (TLI) | Exact | None | comp. cost | artefacts | regular data |
| Nearest Neighbor (NEN) | Assigned | None | comp. cost | lacks smoothness | regular data |
| Natural Neighbor (NAN) | Exact | Area | prediction | comp. cost | irregular undersampled data |

The very fact that so many methods have been successfully working together for decades is a clear signal that there is no single method that is best for rasterizing airborne LiDAR point clouds, let alone that there is a universally best interpolator. Therefore, many studies have compared the efficiency of these interpolators, but even studies using comparable methods and materials often came up with conflicting results. Despite the similarities, it is important to mention that most of the comparative studies were conducted with a specific objective in mind, so they cannot be compared one-to-one.

We looked at 18 comparative studies (Table 4). The first indication of the popularity of each method is the frequency with which they were included in the comparative studies. Although Kriging and

spline are each a generic term for several methods, their dominance along with IDW is evident (14, 13, and 14 times, respectively). NAN seems to have gained traction since 2009. Most of the studies presented use only quantitative statistical comparisons of accuracy. Kriging and spline are also most often found to be the most accurate (six and five times, respectively), followed by IDW (three times).

**Table 4.** Comparison of interpolation methods in selected literature presented in chronological order: B(est in the test), T(ested), "."(not tested). Studies cited are: 1 - (Hutchinson and Gessler, 1994); 2 - (Desmet, 1997); 3 - (Doucette and Beard, 2000); 4 - (Lloyd and Atkinson, 2002); 5 - (Abramov and McEwen, 2004); 6 - (Aguilar et al., 2005); 7 - (Su and Bork, 2006); 8 - (Xiaoye, 2008); 9 - (Erdogan, 2009); 10 - (Heritage et al., 2009); 11 - (Bater and Coops, 2009); 12 - (Guo et al., 2010); 13 - (Razak et al., 2013); 14 - (Chu et al., 2014); 15 - (Chen et al., 2015); 16 - (Chaplot et al., 2006); 17 - (Chen and Li, 2019); (Habib, 2021).

| **Interpolator** | **1** | **2** | **3** | **4**[2] | **5** | **6** | **7** | **8** | **9** |
|---|---|---|---|---|---|---|---|---|---|
| Kriging (OK, etc.) | B | T | T | B | . | . | T | T | T |
| Spline (TPS, etc.) | B | B | B | . | T | T | T | T | B |
| Inverse distance to a power (IDW) | . | . | T | T | . | T | B | B | T |
| Tri. w/ Linear Interpolation (TLI) | . | . | . | . | T | . | . | . | . |
| Nearest Neighbor (NEN) | . | T | . | . | B | . | . | . | . |
| Natural Neighbor (NAN) | . | . | . | . | T | . | . | . | . |
| Other | . | T | T | . | . | B | . | . | . |
| L(iDAR) / O(ther) data[4] | OF | OF | OF | L | OP | OP | L | L | OF |
| Field scale data (Y/N) | N | N | N | N | Y | N | Y | 0[3] | Y |
| Visual assesment[1] (Y/N) | N | Y | N | N | Y | N | N | 0[3] | N |

**Table 4 continued.**

| **Interpolator** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** |
|---|---|---|---|---|---|---|---|---|---|
| Kriging | B | . | B | . | B | T | B | T | T |
| Spline | T | T | T | . | . | . | T | B | . |
| Inverse distance to a power (IDW) | T | T | T | B | T | . | T | T | T |
| Tri. w/ Linear Interpolation (TLI) | B | . | T | . | T | . | . | T[4] | . |
| Nearest Neighbor (NEN) | . | . | . | . | . | . | . | . | . |
| Natural Neighbor (NAN) | . | B | T | T | . | T | . | T | B |
| Other | T | T | . | T | . | B | B | T | . |
| L(iDAR) / O(ther) data[4] | L | L | L | L | L | OP | OM | L | OS |
| Field scale data (Y/N) | N | N | N | Y | Y | N | N | Y | N |
| Visual assesment[1] (Y/N) | N | N | N | Y | Y | Y | N | Y | Y |

[1] Visual assessment is always accompanied by the Quantitative Evaluation of the Accuracy, but not vice versa.
[2] DSM rather than DEM is assessed.
[3] No experiment, results are based on literature analysis.
[4] Other data: field measurements (OF), photogrammetry (OP), derived from scanned maps (OM), synthetic dataset (OS).

Only two studies include a comprehensive visual assessment of precision (Chen and Li, 2019; Razak et al., 2013), but the result is again a split between Kriging and spline. The general lack of qualitative visual analysis of precision was noted in a recent review of nearly two hundred papers (Mesa-Mingorance and Ariza-López, 2020). Why is this important? DEMs are used for analysis, visualization, or both (Ali, 2004). Accuracy is the most important factor for analysis and precision for visualization. Since the primary application of DFM in archaeology is visual analysis, precision is more important. Therefore, not only is there no interpolator that is consistently the most accurate, but there is also a lack of studies that would focus on evaluating the visual precision of interpolators. Therefore, an archaeology-specific evaluation of interpolators is needed.

# 4. Results

## 4.1 Visual assessment of precision

In the first step, we compared the algorithms directly in Golden Software Surfer®, a popular low-cost software of choice for geoscientists (Table 5).

**Table 5.** Interpolation algorithms tested in Golden Software Surfer v.19.2.

| Filter | Variables[1] | Time (sec)[2] AT/SI1/SI2/ES | Multicore |
|---|---|---|---|
| Ordinary Kriging (OK) | Type: point; Drift type: none | 573/135/113/26 | Y |
| Radial Basis Function – Thin Plate Spline (RBF-TPS) | Shape factor ($R^2$): 0.086<br>Radius: 20 m | xy/3756/3668/31 | Y |
| Inverse distance to a power (IDW) | Power: 2; Smooth: 0 | 56/15/11/3 | N |
| Tri. w/ Linear Interpolation (TLI) | N/A | 54/37/51/4 | N |
| Nearest Neighbor (NEN) | N/A | 23/9/8/1 | N |
| Natural Neighbor (NAN) | N/A | 266/87/65/12 | N |

[1] The following (default) settings for search neighborhood have been used for all algorithm: no. of sectors to search 4; max. data to use from all sectors: 64; max. data to use from each sector: 16; min. data in all sectors: 8; assign no data if 3 sectors are empty. The search radius was set to (default) global.
[2] Processed on Windows 10.1803, 2.6 GHz 6-Core Intel Core i7-9750H, 32MB RAM.

The results were as expected and are consistent with the results of other studies discussed above (Figures 5; 6; 7; 8). The largest differences between the different interpolators were visible in "red" areas with DFM confidence 1, defined by a point density of less than ¼ of the grid density or by low point density and very steep slopes. In "red" areas, the terrain could not be accurately predicted by any algorithm and what was observed were various algorithm artefacts (Albani et al., 2004). Most noticeable was the triangular pattern of TLI and NAN (this implementation of NAN uses Delaunay triangulation in the first step) and most distracting was the "scaling" of NEN. IDW introduced a relatively dense pattern of negative interpolation artefacts, i.e., errors in the form of exaggerated negative values. As expected, OK and RBF-TPS handled such situations best and are similar, but in our implementation the latter introduced more pronounced negative artefacts. In contiguous "red" areas, archaeological features smaller than about four times the size of the grid cell were not detectable.

"Orange" areas with DFM confidence two were defined by steep slopes (between 22.5°and 42.5°) and a point density lower than the grid density. Such areas were observed on the slopes of prehistoric hillforts in the ES and SI1 test sites. The scaling of NEN was so bad that interpretative mapping was not possible. The second worst, somewhat surprisingly, was RBF-TPS, which produced a dense pattern of positive artefacts. These were present at OK and NAN in much more attenuated form. NAN's and TLI's triangular artefacts were still present at times. IDW was most pleasing to the eye as there were no positive artefacts, but there were other issues. First, there were very small negative artefacts. Second, there was a slight scaling effect on slopes. This effect was not too visually disturbing, but it can lead an inexperienced interpreter (or an experienced interpreter without sufficient metadata) to erroneously conclude that the slope was terraced. This is further exacerbated when IDW settings were optimised for archaeology (Figure 6). However, both of these problems were barely visible on the high-density data from the AT test site (Figure 5: A). This was the only difference that depended on the size of the grid cells and/or the density of the data points we recorded. In "orange" areas, therefore, all interpolators safe NEN can be used for interpretative mapping. For example, a quarry is identifiable in sufficient detail in all interpolators except NEN to determine type and relative age (Figure 3: A).

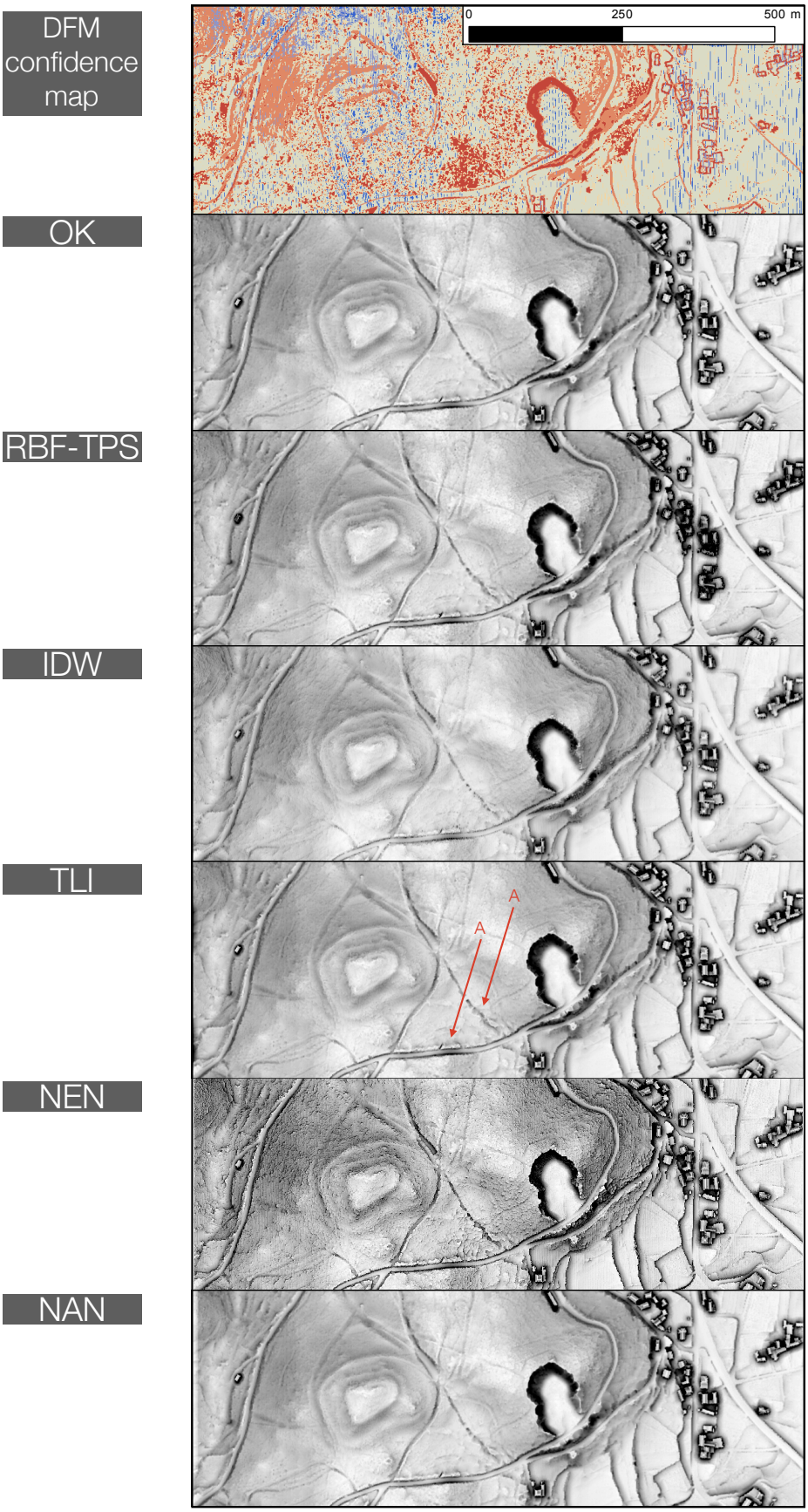



Figure 2. Tested interpolators for the site ES, DFM resolution 1 m (the processing pipeline except for interpolation is the same for all instances: point cloud processing according to ŠtuLoz2020; Sky view factor visualization with default settings in RVT v.2.2).

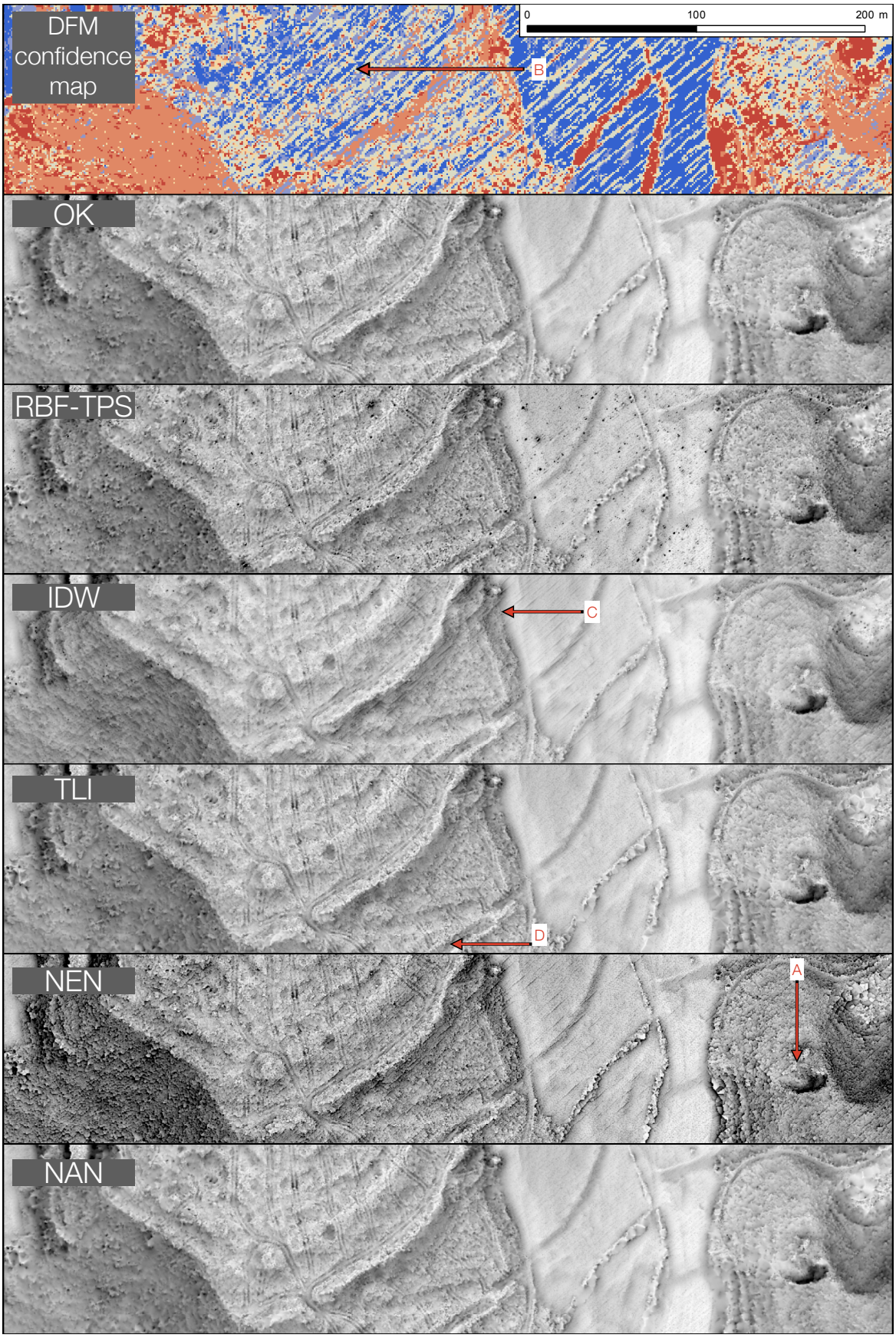



Figure 3. Tested interpolators for site SI1, DFM resolution 0.5 m (processing pipeline except interpolation is the same for all instances: point cloud processing according to ŠtuLoz2020; Sky view factor visualization with default settings in RVT v.2.2).

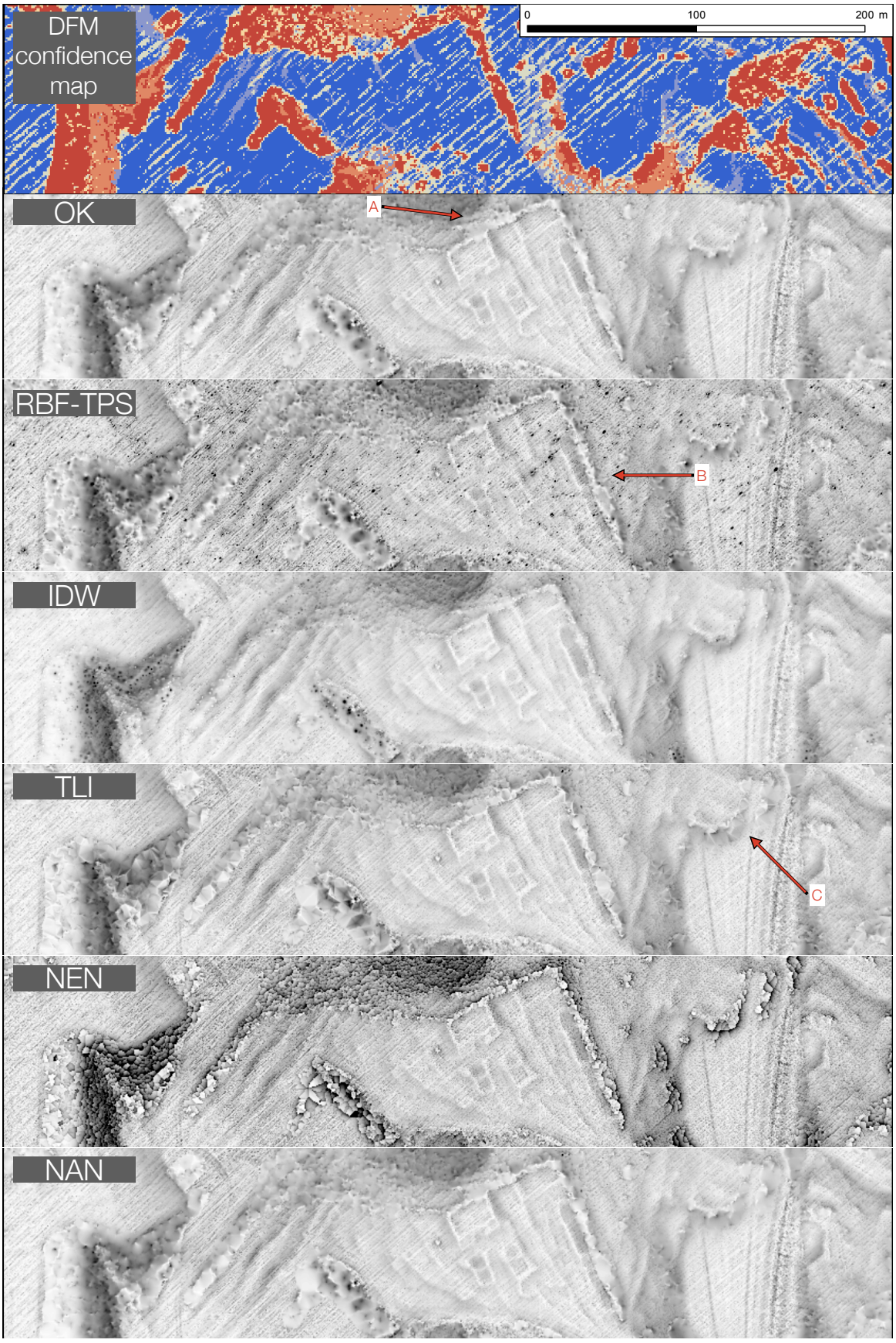



Figure 4. Tested interpolators for site SI2, DFM resolution 0.5 m (the processing pipeline except interpolation is the same for all instances: point cloud processing according to ŠtuLoz2020; Sky view factor visualization with default settings in RVT v.2.2).

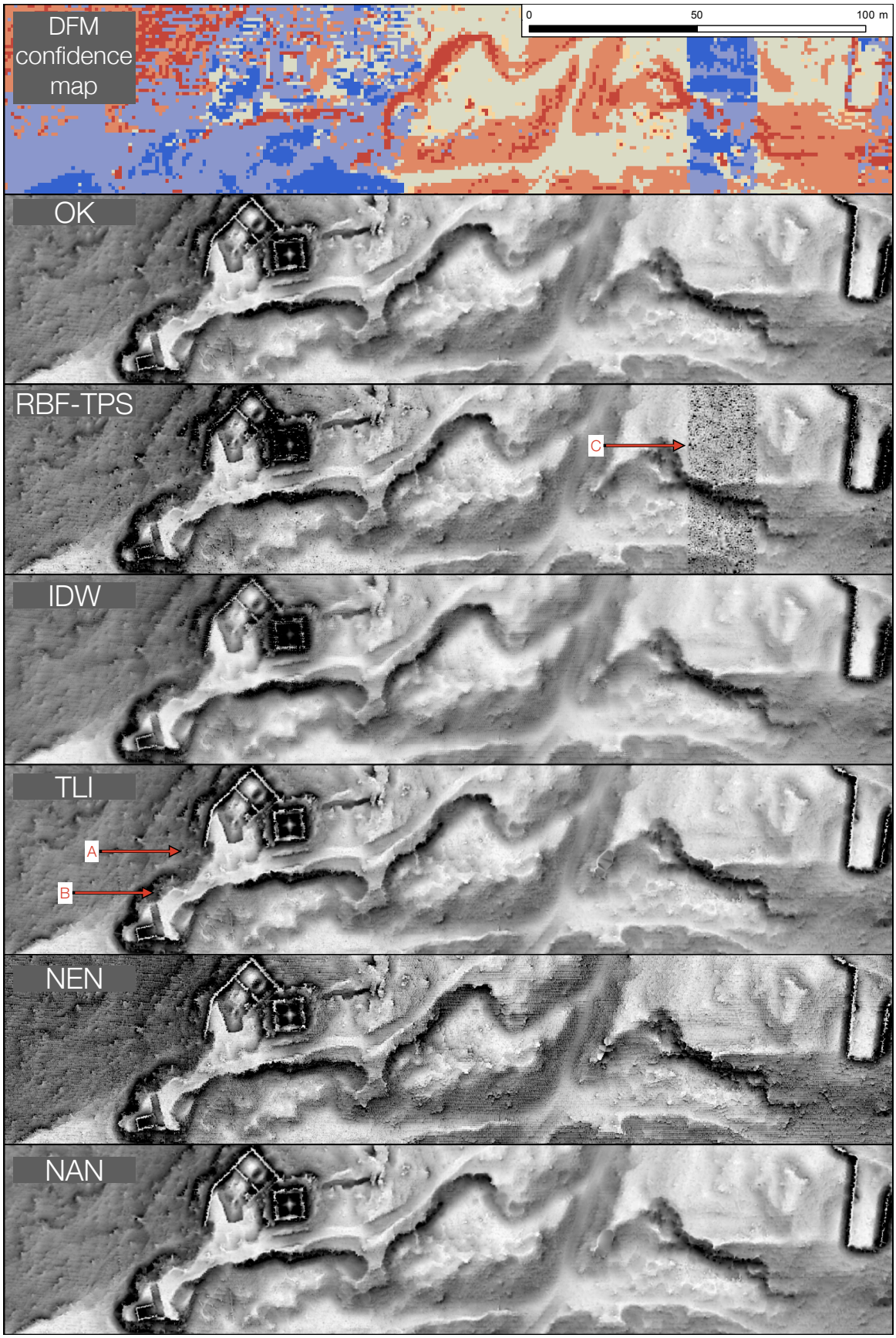



Figure 5. Tested interpolators for the site AT, DFM resolution 0.25 m (the processing pipeline except interpolation is the same for all instances: point cloud processing according to ŠtuLoz2020; Sky view factor visualization with default settings in RVT v.2.2).

"Transitional" areas of DFM confidence three or four were defined by low point density (between ¼ and one times the grid density) or high density of low vegetation. The interior of the hillfort in test site SI1 was one such area (Figure 3: B). Again, NEN performed the worst, although it was useful for interpretative mapping. RBF-TPS introduced relatively many small negative artefacts that were visually distracting. IDW tended to show data noise resulting from the pattern of measurement points (Figure 3: C). This is not distracting to experienced interpreters, but could be mistaken for geological features or similar. OK , NEN and TLI were very similar, with TLI being slightly more accurate. We turn, for example, to the tracks left by heavy machinery clearing the forest after the ice rain disaster. These tracks were about as wide as the size of the grid cell, but more pronounced than typical archaeological features. They were clearly detected with OK, NEN and TLI, with TLI being slightly more accurate (Figure 3: D). Similar observations can be made for the test sites ES (Figure 2: A) and AT (Figure 5: B).

As expected, all interpolators produced the best results in "blue" areas. These were areas with DFM confidence five or six, defined by a point density at least as high as the grid density, and not located on densely vegetated slopes. NEN was significantly noisier than the rest, but not to an extent that would prohibit interpretative mapping. RBF-TPS was still plagued by visually distracting small, low artefacts, and it was unable to process an area of overlapping flight strips on high-density data (Figure 5: C). IDW exhibited a slight degree of oversmoothing. Depending on the circumstances, this can be either beneficial or detrimental for interpretative mapping, but in principle this behaviour is not optimal for archaeology-specific interpolation. OK , NAN and TLI were the best and nearly indistinguishable, although TLI was only slightly more accurate at representing linear features as wide as cell size in "blue" areas.

Also of great interest were "blue" areas interspersed with patches of "red" cells, which we termed "mixed" areas. A typical example was the ruins of a Roman villa in test site SI2. The outer walls of the villa were used as a densely vegetated hedge, which means that the average point density was high but severely limited locally. On the north slope a possible diagonal wall (Figure 4: A) was detectable on IDW but only partially on OK, TIN and NAN; it was not visible on RBF-TPS and NEN in sufficient detail for interpretative mapping. The eastern outer wall (Figure 4: B) could be detected by all interpolations, but accuracy varied: IDW was best, followed by OK, NAN and TLI; NEN and RBF-TPS were again of insufficient quality. For example, the area of a possible additional two-room building was detectable on OK, IDW and NAN, but only partially on TLI (Figure 4: C). Thus, in "mixed" areas, RBF-TPS and NEN in particular are again substandard.

To sum up the visual analysis, OK was the best interpolator in "red" areas and IDW was a distant second. In "orange" areas the results for all interpolators improved markedly and interpretative mapping of all but the smallest archaeological features was possible. OK and IDW performed best, while the artefacts in NAN and TLI were still distracting. As expected, all methods achieved their respective best results in "blue" and "transition" areas. The differences between OK, TIN and NAN were negligible, while IDW was slightly oversmoothed. From the point of view of the individual interpolators, TLI was the best archaeology-specific interpolator for "blue" and, somewhat surprisingly, "transition" areas. OK was best in "red" and "orange" areas. IDW"s strength was its overall robustness. In our opinion, OK was the best on average, followed by IDW. The optimal interpolator would be a hybrid between Kriging for "red" and "orange" areas and TLI for "transition" and "blue" areas. Whether "blue" or "red", RBF-TPS was plagued with low artefacts and NEN with noise, but both can be used for interpretative mapping in "transition" and "blue" areas. This is particularly important for NEN, which is used by some providers of large general-purpose datasets (for example, in the Slovenian nationwide dataset: (Triglav Čekada and Bric, 2015).

However, significant differences between the execution of the same algorithms on different software have been observed both in GIS in general (Mattivi et al., 2019) and in archaeology-specific airborne LiDAR data processing (Štular and Lozić, 2020). Therefore, we tested all available implementations of the IDW interpolator. We are aware of six suitable implementations, but PDAL's "writers.gdal" is an instance of GDAL and was excluded. For the five filters (Table 6), the same values

for the "power" and "smoothing" variables were used (two and zero, respectively). Nevertheless, there were notable differences in the results (Figure 6).

**Table 6.** IDW processing of test site SI2 in selected free or low cost software packages: filters, processing, and variables. The tile size is 1 km$^2$ with 9.3 x 10$^6$ processed points (classes 2 and 6); 0.5 m grid has been interpolated.

| Software | Filter | Time (sec)[1] | las/laz | Multicore |
|---|---|---|---|---|
| Golden Software Surfer v.19.2 | Inverse Distance to a Power | 11 | Y | Y |
| WhiteboxTools v.1.4.0[2] | LidarIdwInterpolation | 13 | Y | Y |
| GRASS GIS v.7.8.3[2] | v.surf.idw | 43 | N | N |
| GDAL v.3.1.2[2] | Grid (IDW w/ NN) | 289 | N | Y |
| lidr v.3.1.1[3] | knnidw | 37 | Y | N |

**Table 5 continued.** Variables.

| Software | No. Points | Power | Radius (m) | Smoothing |
|---|---|---|---|---|
| Golden Software Surfer v.19.2 | 64[4] | 2 | 707 | 0 |
| WhiteboxTools v.1.4.0[2] | N/A | 2 | 2.5 | 0 |
| WhiteboxTools - optimized | N/A | 3 | 5 | 0 |
| GRASS GIS v.7.8.3[2] | 12 | 2 | N/A | N/A |
| GDAL v.3.1.2[2] | 12 | 2 | 5 | 0 |
| lidr v.3.1.1[3] | 10 | 2 | 10 | 0 |

[1] Processed on MacOS 10.15.7 (for Surfer see Table 5), 2.6 GHz 6-Core Intel Core i7, 32MB RAM.
[2] Accessed trough QGIS v.3.16.1 Processing Toolbox.
[3] Accessed trough Rstudio v.1.4.1103.
[4] No. of sectors to search 4; max. data to use from all sectors: 64; max. data to use from each sector: 16; min. data in all sectors: 8; assign no data if 3 sectors are empty.

On the slopes in the "red" and "orange" areas, GRASS and GDAL show a clear "scaling" comparable to that of NEN (Figure 6: A, B). The two are visually identical in all areas, leading us to conclude that they are likely instances of the same filter. The "scaling" artefacts were much weaker, but still present in Whitebox. There were significant differences in one "mixed" area, for example, how the small terrace to the east of the potential two-room building is shown (Figure 6: C). Surprisingly, the main archaeology-specific differences are in the "blue" areas. First, Whitebox and GDAL/GRASS have very small negative artefacts that are perceived as noise. Second, the archaeological features appear progressively less pronounced in GDAL/GRASS, lidr and Surfer, and finally in Whitebox (Figure 6: D). This difference can be attributed to different values of the variables "maximum number of points" and "search radius". The higher the number of points used and/or the larger the radius, the smoother the result. Since these two variables are not implemented in the same way in all filters and are not always accessible to the user, they were left at default settings for the test (Table 5). Therefore, the above results can be adjusted by making appropriate settings in each of the tested filters. As an example, the result of Whitebox Tools' "LidarIdwInterpolation" with the settings optimised for archaeology (weight 3, radius 5 m) is included (Figure 6).

In summary, all tested IDW implementations can be used for DFM rasterization, but from the perspective of visual quality, Surfer and lidr are superior and Whitebox is second best.

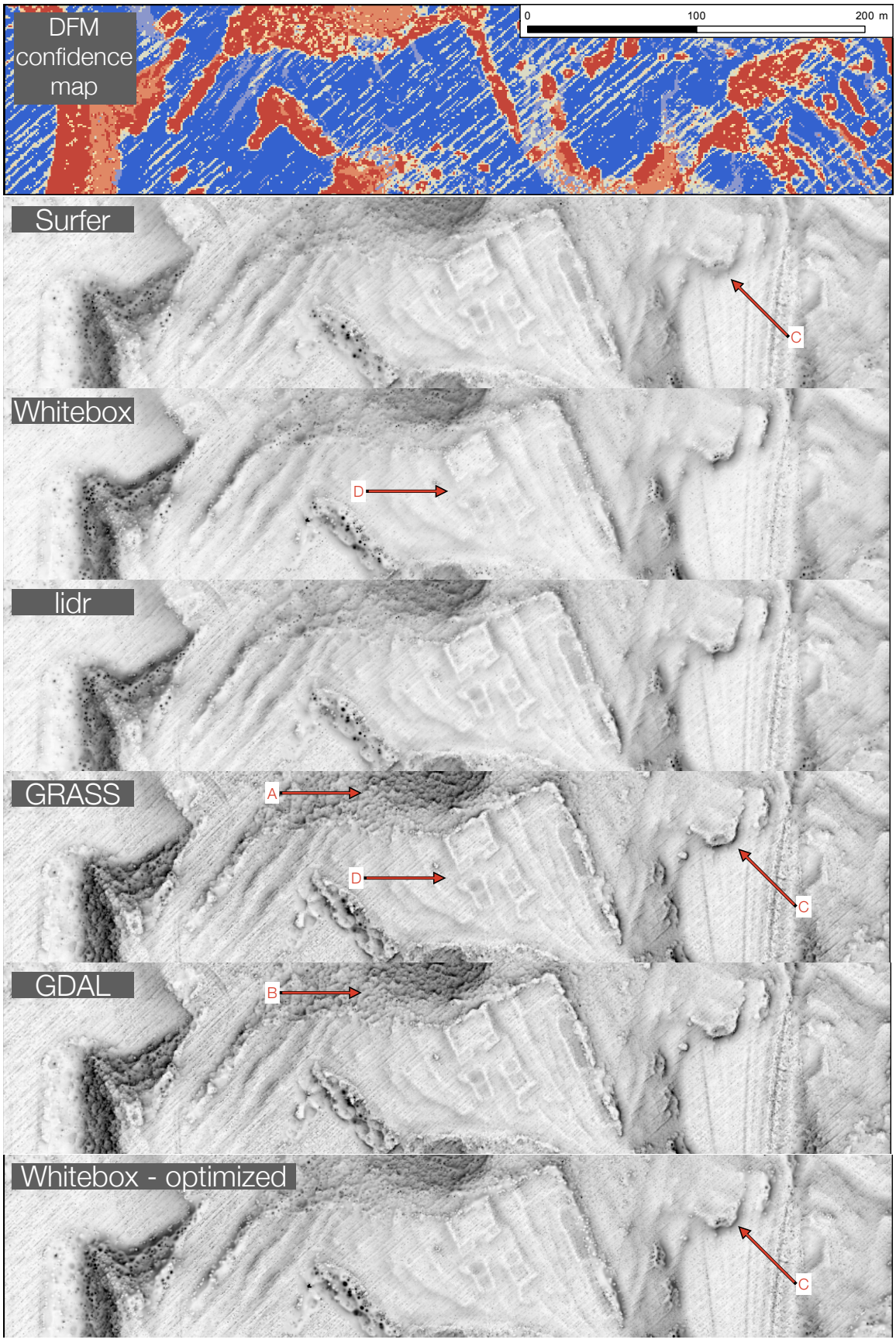



Figure 6. IDW interpolator calculated with different software (Table 5) for SI2 site, DFM resolution 0.5 m (processing pipeline except interpolation is the same for all instances: point cloud processing according to ŠtuLoz2020; Sky view factor with default settings in RVT v.2.2).

## 4.2 Cost and availability

In our test, we estimated the computational cost by timing the execution of all interpolators with the same equipment (Table 5). The only surprise is RBF-TPS. It is known for its high computational cost, but 30x longer times than OK were not expected. OK and NAN were significantly more expensive than IDW and NEN and TLI was in the middle. The difference between OK and IDW was 10x. Before the implementation of parallel processing in this particular software package, this difference was much higher (by a factor of 12x on our hardware). From the perspective of archaeology-specific pipeline (Štular and Lozić, 2020), with parallel processing enabled, the computational cost for all interpolators except RBF-TPS is acceptable.

Regarding the availability, we focused on free and low-cost solutions suitable for processing field scale airborne LiDAR data. The availability of each interpolator in relevant software known to us is shown in the table (Table 2). NAN was only available in one software package, while TLI and NEN were the interpolators of choice for most LiDAR-oriented software packages. The results for OK were disappointing. We knew of three seemingly suitable filters, but two failed to finish in a reasonable time (defined here as 100x the time achieved by Surfer). IDW was the most widely available interpolator. "LidarIdwInterpolation" by Whitebox Tools is the most time efficient and arguably the most widely available.

## 4.3 Triangular assessment

As mentioned earlier, the goal of the triangular assessment was to find the interpolator that was best suited for the archaeology-specific DFM gridding. This was achieved by plotting the results of the assessment triangle on a triangular graph (Figure 7). The most appropriate method was the one that was most balanced and plotted closest to the centre, which is IDW. OK was the most precise one, but lacked availability more than anything else. This may change in an instance.

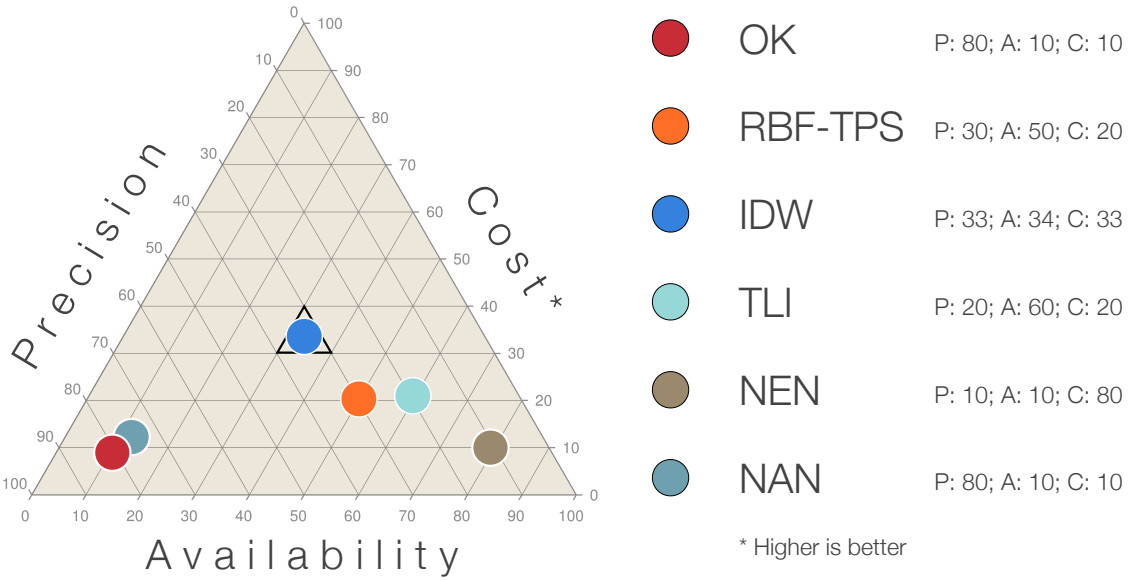



Figure 7. Most suitable archaeology-specific interpolator according to the results of the assessment triangle plotted on the triangular graph. Results closest to the centre of the triangle are best.

# 5. Discussion & Conclusion

The multitude of accuracy assessment analyses shows that choosing the appropriate interpolator is difficult, as each method has its own advantages and disadvantages. Traditional wisdom states that Kriging-based methods are capable of producing more accurate DEMs, but they are computationally intensive and time consuming. IDW, NEN, and TLI are simple and fast methods to generate relatively accurate DEMs, but their performances decrease and become sensitive to topographic variations as the spatial resolution of DEM increases. Spline seems to offer a trade-off between computation time and accuracy at high resolutions, but becomes less reliable at low resolutions (Guo et al., 2010).

The stated main objective of this paper was to find the most suitable interpolator (one or more) for rasterizing archaeology-specific DFM from airborne LiDAR data. The result of the triangular assessment was clear: IDW is currently the most suitable archaeology-specific interpolator. However, this may change if either OK or a TLI-OK/IDW hybrid becomes available. In particular a hybrid between the two exact interpolators TLI and IDW based on DFM confidence map segmentation is feasible and relatively easy to implement (Figure 8; Eichert et al 2021).

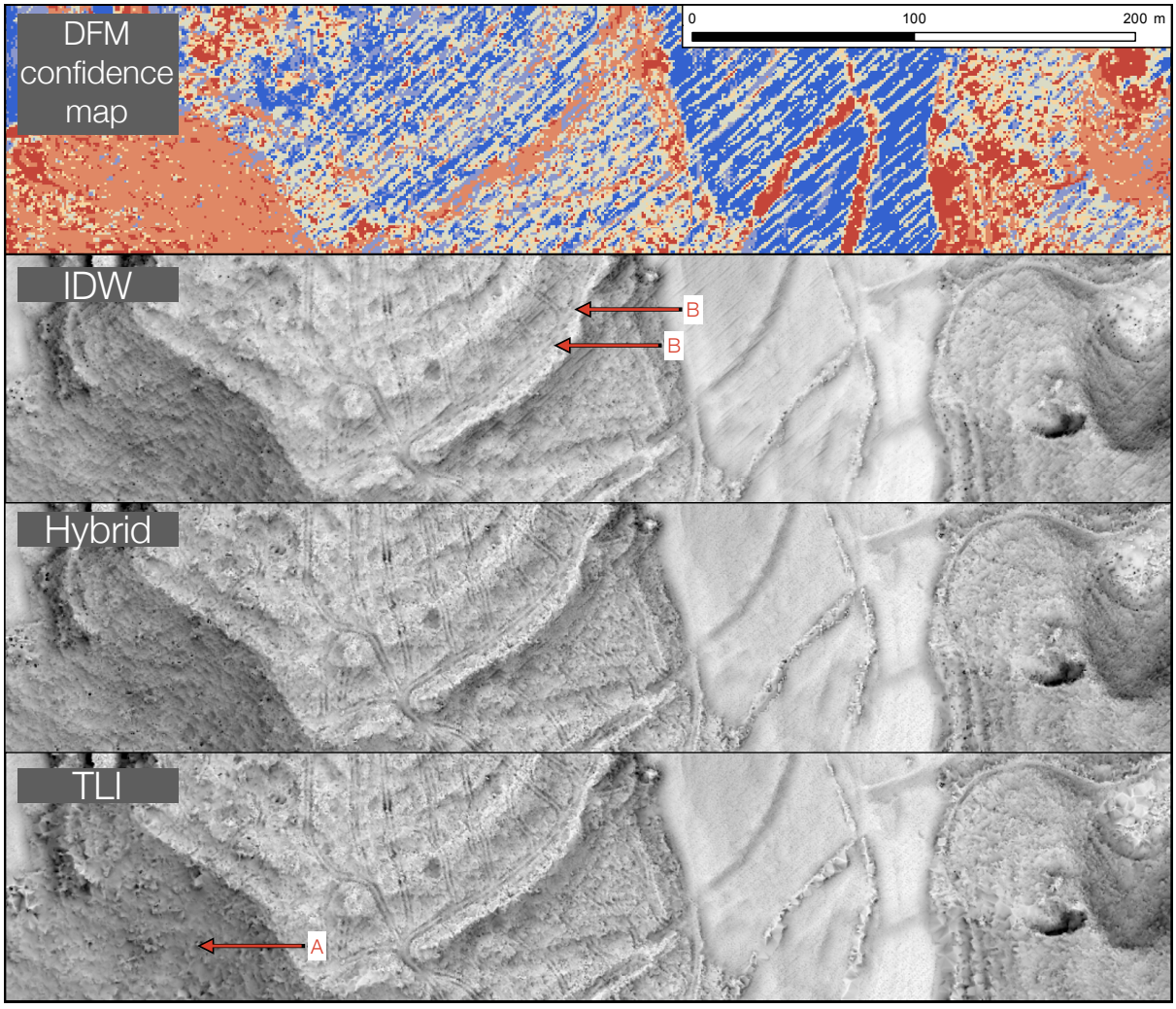



Figure 8. Hybrid interpolator shown with DFM confidence map, IDW, and TLI. Hybrid interpolation removes TLI artefacts on "red" zones (A), but brings maximum precision in "blue" zones (B). DFM resolution 0.5 m (processing pipeline except interpolation is the same for all instances: interpolation calculated in Golden Software Surfer; point cloud processing according to ŠtuLoz2020; Sky view factor with default settings in RVT v.2.2).

As a side note, the biggest differences between the interpolators were observed in the undersampled areas, while it was the opposite when comparing different implementations of IDW. This comparison was also a reminder of how different filters of a relatively straightforward algorithm give noticeably different results. Software implementation must be taken into account, especially when making statements such as IDW power variable two is best for DSM and power three for DEM (Montealegre et al., 2015).

Third, we have addressed a wider field of DEM accuracy studies. Our results suggest that the inconclusiveness of DEM accuracy assessments is at least partly due to the approach: searching for the single best or most accurate interpolator. In comparison, a similar search for the best archaeology-specific DEM visualization initially showed that different visualizations are optimal under different circumstances (Challis et al., 2011; Štular et al., 2012). Recently, it seems that a single hybrid visualization is able to satisfy the vast majority of use cases (Kokalj and Somrak, 2019). To some extent, this can also be applied to interpolators. Different interpolators are best suited for different levels of DFM (or DEM) confidence. A hybrid solution that successfully combines an exact TLI interpolator with a smooth interpolator OK or IDW seems to be the best choice at present.

To this end, the results of our analysis of visual precision coupled with DFM confidence map provide clear criteria for segmentation, which is one of the stated unresolved issues of DEM interpolation (Mesa-Mingorance and Ariza-López, 2020).

## Author Contributions:

Conceptualization, E.L.; methodology, B.Š.; formal analysis, E.L.; writing and original draft preparation, B.Š.; writing—review and editing, B.Š., E.L., S.E.; data processing, E.L., S.E.; programming, S.E.; project administration, B.Š.; funding acquisition, B.Š. and E.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by ARRS SLOVENIAN RESEARCH AGENCY, grant numbers N6-0132 and P6-0064, and FWF AUSTRIAN SCIENCE FOUND grant number 13992.

Institutional Review Board Statement: Not applicable. Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable (all data used are open access).

Acknowledgments: The authors want to thank to the reviewers for the contribution to the final version of the article.

Conflicts of Interest: The authors declare no conflict of interest.

## References

Abramov, O., McEwen, A., 2004. An evaluation of interpolation methods for Mars Orbiter Laser Altimeter (MOLA) data. International Journal of Remote Sensing 25, 669--676.

Aguilar, F.J., Agüera, F., Aguilar, M.A., Carvajal, F., 2005. Effects of Terrain Morphology, Sampling Density, and Interpolation Methods on Grid DEM Accuracy. Photogrammetric Engineering & Remote Sensing 71, 805–816.

Albani, M., Klinkenberg, B., Andison, D.W., Kimmins, J.P., 2004. The choice of window size in approximating topographic surfaces from Digital Elevation Models. International Journal of Geographical Information Science 18, 577–593.

Ali, T.A., 2004. Proceedings of the American Congress on Surveying and Mapping (ACSM) Conference 2004, Nashville, TN, USA, 2004. ACSM.

Bater, C.W., Coops, N.C., 2009. Evaluating error associated with lidar-derived DEM interpolation. Computers & Geosciences 35, 289–300.

Bonham-Carter, G.F., 1994. Geographic Information Systems for Geoscientists: Modelling with GIS, Geographic Information Systems for Geoscientists: Modelling with GIS. Pergamon.

Challis, K., Forlin, P., Kincey, M., 2011. A Generic Toolkit for the Visualization of Archaeological Features on Airborne LiDAR Elevation Data. Archaeol. Prospect. 18, 279–289. https://doi.org/10.1002/arp.421

Chaplot, V., Darboux, F., Bourennane, H., Leguédois, S., Silvera, N., Phachomphon, K., 2006. Accuracy of interpolation techniques for the derivation of digital elevation models in relation to landform types and data density. Geomorphology 77, 126–141.

Chase, A.S.Z., Chase, D., Chase, A., 2020. Ethics, New Colonialism, and Lidar Data: A Decade of Lidar in Maya Archaeology. Journal of Computer Applications in Archaeology 3, 51–62. https://doi.org/10.5334/jcaa.43

Chen, C., Li, Y., 2019. A Fast Global Interpolation Method for Digital Terrain Model Generation from Large LiDAR-Derived Data. Remote Sensing 11, 1324.

Chen, C., Li, Y., Yan, C., Dai, H., Liu, G., 2015. A Robust Algorithm of Multiquadric Method Based on an Improved Huber Loss Function for Interpolating Remote-Sensing-Derived Elevation Data Sets. Remote Sensing 7, 3347--3371.

Chen, C., Li, Y., Zhao, N., Guo, B., Mou, N., 2018. Least Squares Compactly Supported Radial Basis Function for Digital Terrain Model Interpolation from Airborne Lidar Point Clouds. Remote Sensing 10, 587.

Chen, Z., Gao, B., Devereux, B., 2017. State-of-the-Art: DTM Generation Using Airborne LIDAR Data. Sensors 17, 150.

Chew, P.L., 1989. Constrained Delaunay Triangulations. Algorithmica 4, 97–108.

Chu, H.-J., Wang, C.-K., Huang, M.-L., Lee, C.-C., Liu, C.-Y., Lin, C.-C., 2014. Effect of point density and interpolation of LiDAR-derived high-resolution DEMs on landscape scarp identification. GIScience & Remote Sensing 51, 731--747.

Crutchley, S., 2006. Light detection and ranging (lidar) in the Witham Valley, Lincolnshire: an assessment of new remote sensing techniques. Archaeol. Prospect. 13, 251–257. https://doi.org/10.1002/arp.294

Crutchley, S., Crow, P., 2010. The Light Fantastic: Using Airborne Lidar in Archaeological Survey. English Heritage, Swindon.

De Boer, A.G., Laan, W.N., Waldus, W., Van Zijverden, W.K., 2008. LiDAR-based surface height measurements: applications in archaeology, in: Frischer, B., Dakouri-Hild, A. (Eds.), Beyond Illustration: 2D and 3D Digital Technologies as Tools for Discovery in Archaeology, BAR International Series. Archaeopress, New York, pp. 76–84, 154–156.

De Smith, M.J., Goodchild, M.F., Longley, P., 2018. Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools. Troubador Publishing Ltd.

Desmet, P.J.J., 1997. Effects of Interpolation Errors on the Analysis of DEMs. Earth Surface Processes and Landforms 22, 563–580.

Doneus, M., Briese, C., 2011. Airborne Laser Scanning in forested areas – potential and limitations of an archaeological prospection technique, in: Cowley, D.C. (Ed.), Remote Sensing for Archaeological Heritage Management. Europae Archaeologia Consilium (EAC), Brussels, pp. 59–76.

Doneus, M., Briese, C., Fera, M., Janner, M., 2008. Archaeological prospection of forested areas using full-waveform airborne laser scanning. Journal of Archaeological Science 35, 882–893. https://doi.org/10.1016/j.jas.2007.06.013

Doneus, M., Mandlburger, G., Doneus, N., 2020. Archaeological Ground Point Filtering of Airborne Laser Scan Derived Point-Clouds in a Difficult Mediterranean Environment. Journal of Computer Applications in Archaeology 3, 92–108. https://doi.org/10.5334/jcaa.44

Dong, P., Chen, Q., 2018. LiDAR Remote Sensing and Applications. CRC Press, Taylor & Francis Group.

Doucette, P., Beard, K., 2000. Exploringthe Capability of Some GIS Surface Interpolators for DEM Gap Fill. Photogrammetric Engineering & Remote Sensing 66, 881–888.

Erdogan, S., 2009. A comparision of interpolation methods for producing digital elevation models at the field scale. Earth Surface Processes and Landforms 34, 366–376.

Fernandez-Diaz, J., Carter, W., Shrestha, R., Glennie, C., 2014. Now You See It… Now You Don't: Understanding Airborne Mapping LiDAR Collection and Data Product Generation for Archaeological Research in Mesoamerica. Remote Sensing 6, 9951–10001. https://doi.org/10.3390/rs6109951

Grammer, B., Draganits, E., Gretscher, M., Muss, U., 2017. LiDAR-guided Archaeological Survey of a Mediterranean Landscape: Lessons from the Ancient Greek Polis of Kolophon (Ionia, Western Anatolia). Archaeological Prospection 24, 311–333. https://doi.org/10.1002/arp.1572

Guo, Q., Li, W., Yu, H., Alvarez, O., 2010. Effects of Topographic Variability and Lidar Sampling Density on Several DEM Interpolation Methods. Photogrammetric Engineering & Remote Sensing 76, 701–712.

Habib, M., 2021. Evaluation of DEM interpolation techniques for characterizing terrain roughness. CATENA 198, 105072.

Heritage, G.L., Milan, D.J., Large, A.R.G., Fuller, I.C., 2009. Influence of survey strategy and interpolation model on DEM quality. Geomorphology 112, 334–344.

Hesse, R., 2010. LiDAR-derived Local Relief Models–a new tool for archaeological prospection. Archaeological Prospection 17, 67–72. https://doi.org/10.1002/arp.374

Humme, A., Lindenbergh, R., Sueur, C., 2006. Revealing Celtic fields from lidar data using kriging based filtering, in: Proceedings of the ISPRS Commission V Symposium. Citeseer.

Hutchinson, M.F., Gessler, P.E., 1994. Splines - more than just a smooth interpolator. Geoderma 62, 45–67.

Inomata, T., Triadan, D., López, V.A.V., Fernandez-Diaz, J.C., Omori, T., Bauer, M.B.M., Hernández, M.G., Beach, T., Cagnato, C., Aoyama, K., Nasu, H., 2020. Monumental architecture at Aguada Fénix and the rise of Maya civilization. Nature 582, 530–533.

Kokalj, Ž., Hesse, R., 2017. Airborne laser scanning raster data visualization: A Guide to Good Practice. Založba ZRC, Ljubljana.

Kokalj, Ž., Somrak, M., 2019. Why Not a Single Image? Combining Visualizations to Facilitate Fieldwork and On-Screen Mapping. Remote Sensing 11, 747. https://doi.org/10.3390/rs11070747

Kokalj, Ž., Zakšek, K., Oštir, K., 2011. Application of sky-view factor for the visualisation of historic landscape features in lidar-derived relief models. Antiquity 85, 263–273.

Laharnar, B., Lozić, E., Štular, B., 2019. Rural Settlement: Relating buildings, landscape, and people in the European Iron Age, in: Cowley, D.C., Fernández-Götz, M., Romankiewicz, T., Wendling, H. (Eds.), . Sidestone Press, pp. 263–271.

Latour, B., 1999. Pandora's hope: essays on the reality of science studies. Harvard University Press.

Lawrence, R.L., Wright A., 2001. Rule-based classification systems using classification and regression tree (CART) analysis. Photogrammetric Engineering and Remote Sensing 67, 1137--1142.

Lee, D.T., Schachter, B.J., 1980. Two Algorithms for Constructing a Delaunay Triangulation. International Journal of Computerand Information Science 9, 219–242.

Lloyd, C.D., Atkinson, P.M., 2002. Deriving DSMs from LiDAR data with kriging. International Journal of Remote Sensing 23, 2519–2524.

Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W., 2015. Geographic Information Science and Systems, Geographic Information Science and Systems. Wiley.

Lozić, E., Štular, B., 2021. Documentation of Archaeology-Specific Workflow for Airborne LiDAR Data Processing. Geosciences 11, 26.

Mattivi, P., Franci, F., Lambertini, A., Bitelli, G., 2019. TWI computation: a comparison of different open source GISs. Open Geospatial Data, Software and Standards 4.

Menéndez Blanco, A., García Sánchez, J., Costa-García, J.M., Fonte, J., González-Álvarez, D., Vicente García, V., 2020. Following the Roman Army between the Southern Foothills of the Cantabrian Mountains and the Northern Plains of Castile and León (North of Spain): Archaeological Applications of Remote Sensing and Geospatial Tools. Geosciences 10, 485.

Mesa-Mingorance, J.L., Ariza-López, F.J., 2020. Accuracy Assessment of Digital Elevation Models (DEMs): A Critical Review of Practices of the Past Three Decades. Remote Sensing 12, 2630.

Montealegre, A., Lamelas, M., Riva, J., 2015. Interpolation Routines Assessment in ALS-Derived Digital Elevation Models for Forestry Applications. Remote Sensing 7, 8631–8654.

Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. Hydrological Processes 5, 3–30.

Oliver, M.A., Webster, R., 1990. Kriging: a method of interpolation for geographical information systems. International Journal of Geographical Information Systems 4, 313–332.

Opitz, R.S., 2013. An overview of airborne and terrestrial laser scanning in archaeology, in: Opitz, R.S., Cowley, D.C. (Eds.), Interpreting Archaeological Topography: Airborne Laser Scanning, 3D Data and Ground Observation, Occasional Publication of the Aerial Archaeology Research Group. Oxbow Books, Oxford, pp. 13–31.

Pellegrino, J.W., Chudowsky, N., Glaser, R., 2001. Knowing What Students Know: The Science and Design of Educational Assessment, Knowing What Students Know: The Science and Design of Educational Assessment. National Academy Press.

Razak, K.A., Santangelo, M., van Westen, C.J., Straatsma, M.W., de Jong, S.M., 2013. Generating an optimal DTM from airborne laser scanning data for landslide mapping in a tropical forest environment. Geomorphology 190, 112–125.

Riley, M.A., Tiffany, J.A., 2014. Using LiDAR data to locate a Middle Woodland enclosure and associated mounds, Louisa County, Iowa. Journal of Archaeological Science 52, 143–151. https://doi.org/10.1016/j.jas.2014.07.018

Rochelo, M.J., Davenport, C., Selch, D., 2015. Revealing pre-historic Native American Belle Glade earthworks in the Northern Everglades utilizing airborne LiDAR. Journal of Archaeological Science: Reports 2, 624–643. https://doi.org/10.1016/j.jasrep.2014.11.009

Sambridge, M., Braun, J., McQueen, H., 1995. Geophysical parametrization and interpolation of irregular data using natural neighbours. Geophysical Journal International 122, 837–857.

Saxton, K.E., Rawls, W.J., Romberger, J.S., Papendick, R.I., 1986. Estimating Generalized Soil-water Characteristics from Texture. Soil Science Society of America Journal 50, 1031–1036.

Shao, Y., Lunetta, R.S., 2012. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. ISPRS Journal of Photogrammetry and Remote Sensing 70, 78–87.

Shepard, D., 1968. A two-dimensional interpolation function for irregularly-spaced data. Presented at the Proceedings of the 1968 23rd ACM national conference on -, ACM Press, pp. 517–524.

Simpson, J., Smith, T., Wooster, M., 2017. Assessment of Errors Caused by Forest Vegetation Structure in Airborne LiDAR-Derived DTMs. Remote Sensing 9, 1101.

Stanton, T.W., Ardren, T., Barth, N.C., Fernandez-Diaz, J.C., Rohrer, P., Meyer, D., Miller, S.J., Magnoni, A., Pérez, M., 2020. 'Structure' density, area, and volume as complementary tools to understand Maya Settlement: An analysis of lidar data along the great road between Coba and Yaxuna. Journal of Archaeological Science: Reports 29, 102178. https://doi.org/10.1016/j.jasrep.2019.102178

Stiner, M.C., 1990. The Use of Mortality Patterns in Archaeological Studies of Hominid Predatory Adaptations. Journal of Anthropological Archaeology 9, 305–351.

Štular, B., Kokalj, Ž., Oštir, K., Nuninger, L., 2012. Visualization of lidar-derived relief models for detection of archaeological features. Journal of Archaeological Science 39, 3354–3360.

Štular, B., Lozić, E., 2020. Comparison of Filters for Archaeology-Specific Ground Extraction from Airborne LiDAR Point Clouds. Remote Sensing 12, 3025.

Štular, B., Lozić, E., 2016. Digitalni podatki, in: Ciglič, R., Geršič, M., Perko, D., Zorn, M. (Eds.), . Geografski inštitut antona Melika ZRC SAZU, pp. 157–166.

Su, J., Bork, E., 2006. Influence of Vegetation, Slope, and Lidar Sampling Angle on DEM Accuracy. Photogrammetric Engineering & Remote Sensing 72, 1265--1274.

Triglav Čekada, M., Bric, V., 2015. Končan je projekt laserskega skeniranj Slovenije. Geodetski vestnik 59, 586–592.

Wood, J.D., Fisher, P.F., 1993. Assessing interpolation accuracy in elevation models. IEEE Computer Graphics and Applications 13, 48--56.

Xiaoye, L., 2008. Airborne LiDAR for DEM generation: some critical issues. Progress in Physical Geography 32, 31–49.

# Appendix A: Description of tested interpolators

Interpolation, also called rasterization or gridding, is a process of fitting a surface to point elevation data by effectively draping a grid of defined cell size over the point data. In mathematics it is an approximation procedure and in statistics it is an estimation issue. This means that the points are derived from the original data and do not consist of the actual data points. The value of each cell is predicted by a specific variable based on measurements at points within the area of interest. The variables are known as interpolators. Interpolators can be exact or inexact, depending on whether or not the interpolated area passes through the points, but no method of interpolation is 100% accurate except at the points through which the function was fitted. Two implicit assumptions are that the terrain surface is continuous and smooth and that there is a high correlation between adjacent data points (Shepard 1968; MooreEt1991; Xiaoye2008; Crutchley; Dong 2018, 49-53).

There are many different interpolation methods used in geosciences to interpolate grid surfaces from airborne LiDAR point clouds. They can be divided into deterministic and geostatistical methods. The former assumes that each input point has a local influence that decreases with distance, for example, while the latter considers the statistical relationship between sample points (Xiaoye2008). The most commonly used interpolators are Kriging, radial basis functions (RBF), inverse distance weighting (IDW), triangulated irregular network with linear interpolation (TLI), natural neighbor (NAN), and nearest neighbor (NEN) (ChenEt2019). These methods are briefly described below, including strengths and shortcomings relevant to our purpose.

Kriging is a geostatistical approach to interpolation (MitasEt1999). The method was originally developed by Krige and Matheron to interpolate a surface from scattered sample data for the mining industry, but is now widely used in spatial data analysis. It relies on spatial autocorrelation between sample data to derive the unknown values of data points. In other words, it takes into account both the distance and the degree of variation between sample data (OliverEt1990; BonhamCarter1994, 154-160; LongleyEt2015, 337). The extent to which the spatial autocorrelation assumption holds is examined in a computed variogram (Chaplot et al., 2006). Kriging can also be understood as a weighted average procedure. However, its weights depend not only on the distances between sample points and estimation locations, but also, on the mutual distances between sample points (AndersonEt2005). It is generally accepted that Kriging-based approaches are more accurate in estimating DEMs than widely used tools such as IDW (LloydEt2002a; Lloyd2002b). In particular, robust Kriging approaches such as ordinary Kriging (OK) are advantageous with increasing sample distance. However, the high computational cost of OK was often cited as prohibitive when processing LiDAR data (AndresonEt2005).

RBF interpolation is a large family of exact interpolators that use a variational approach (MitasEt1999). It makes no assumptions about the data points other than that they are non-co-linear. It estimates the values of various mathematical functions to minimize the overall curvature of the surface, resulting in a smooth surface that passes exactly through the data points (GuoEt2010). It has been described that it is like bending a rubber sheet to pass through all the data points (Xiaoye2008). For terrain modelling, the so-called multiquadric function was shown to be particularly effective, as have thin plate splines (TPS) (DeSmith2018, 393-394). We have tested RBF-TPS, which produces a function that passes as close to the data points as possible while still maintaining some degree of smoothness (Desmet1997). Its advantages are ease of use, flexibility, and high interpolation accuracy, which was found to be similar to Kriging in some tests (HutchinsonEt1994; ChenEt2019). However, it is often not suitable when there are large changes in surface values within a short horizontal distance (RobinsonEt2006) and its application to LiDAR data is hindered by the huge computational cost, which is an area of active development (e.g., ChenEt2019).

The following methods use a local neighbourhood approach to interpolation based on the assumption that each point affects the resulting surface only up to a certain finite distance (MitasEt1999).

IDW is a deterministic interpolator that assumes that the closer a data point is to the cell, the more influence it has on the predicted value (Andersonet2005; Xiaoye2008). IDW estimates that a point value at any point P in the plane is a weighted average of the values at the data points $D_i$ (Shepard1968). IDW has been shown to work well for dense and uniformly distributed data points, but not so much for sparse or unevenly distributed data points (WoodFisher1993). Another disadvantage of IDW is that it cannot make estimates that fall outside the range of minimum and maximum data point values, such as for ridges and valleys (LongleyEt2015, 335). However, this is less of an issue with very dense data.

Triangulated irregular network (TIN), as mentioned, is an alternative approach to terrain representation. It partitions a surface into a set of contiguous, non-overlapping triangles, usually using Delaunay triangulation. It requires less data storage than raster structures and is particularly useful for surfaces with discontinuities (Chew1989; BonhamCarter1994, 80). However, gridded DEMs have much more use in modern GIS and TINs are often interpolated into a grid by recording the elevation for each triangle node, while the elevations between nodes are estimated by linear interpolation. This method is called triangulation with linear interpolation (TLI) and is an exact interpolator. It is best suited for uniformly distributed data points, as datasets containing sparse areas result in distinct triangular facets on the map (LeeEt1980; GuibasEt1985; Goldensoftware2013).

In the nearest neighbor (NEN) method, each cell takes the value of the nearest data point. The method is primarily used for re-gridding discrete attributes, such as rock type or binary viewshed. The NEN method can also be used with ratio- and interval-based measurements, such as DEM, but the surfaces produced lack smoothness (BonhamCarter1994, 109). It is known for its very low computational cost and has therefore gained popularity in LiDAR data processing (ChenEt2019), especially when combined with TIN. In some experiments, it produced DEMs with the highest accuracy (SterenczakEt2016).

The natural neighbor method (NAN) is very similar to IDW, but applies weighting to individual data points using a weighted average method based on Thiessen polygons. In some implementations, Delaunay triangulation can be used in the first step. NAN is an exact local interpolator that is particularly suitable for continuous surfaces. It has the ability to deal with very irregular distributions of data points. These properties make the method ideal for rasterizing data with irregular spacing, such as DEMs from LiDAR (SambridgeEt1995). Under certain circumstances, NAN yielded higher accuracy than OK or IDW (AbramovEt2004; Habib2020).