Airborne lidar data managament and homogenization

Žiga Kokalj[1,2] and Benjamin Štular[1]
with
Damien Vurpillot[3]

[1]ZRC SAZU
[2]SPACE-SI
[3]UFC

1. Data, Quality control and Metadata

presentation & practice

2. Rasterizing – making useful images

presentation & practice

3. Data homogenisation (multi-temporal data)

presentation & practice

4. Data Archiving

presentation & practice

5. Wrap up and Take-home-message

1. Data, Quality control and Metadata

presentation & practice

Discovery, management, analyses

- Why is it there?
- What is it?
- Is there anything beside it?
- Is there anything out there?

complexity

Basics about lidar data

file formats, file compression, spatial indexing, quality control

What types of data can you get?

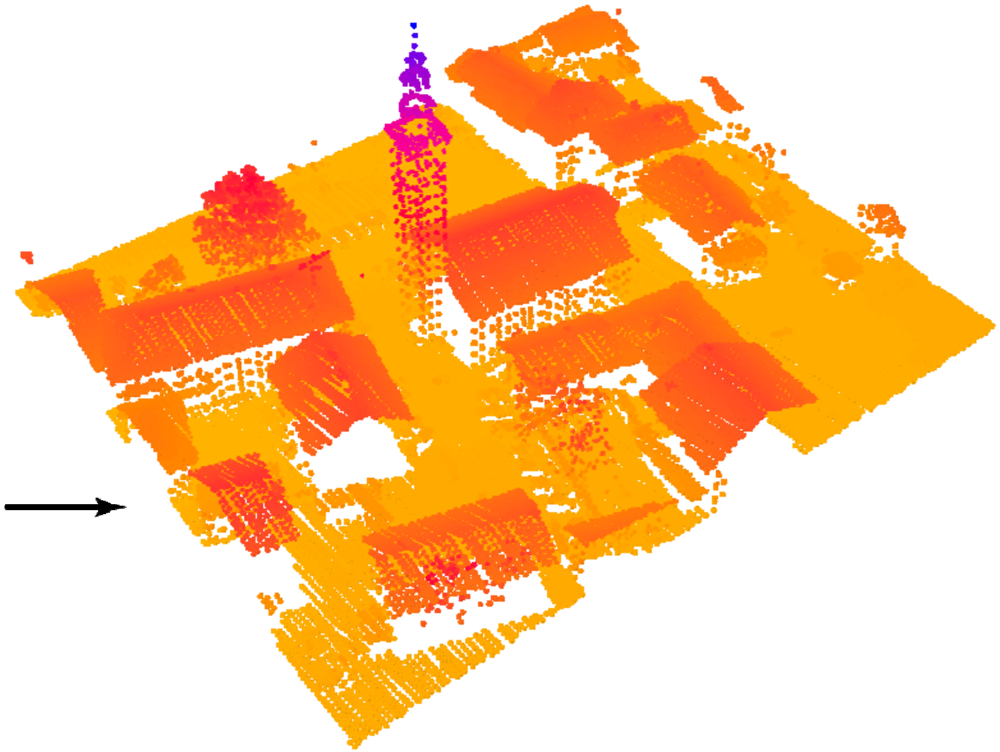
- research data (hi-res, small areas)
- nation wide ALS surveys (low-res, huge areas)
- „scraps“ (commonly data collected for engineering projects, low-res, small and awkward areas)
- other?

Critical questions

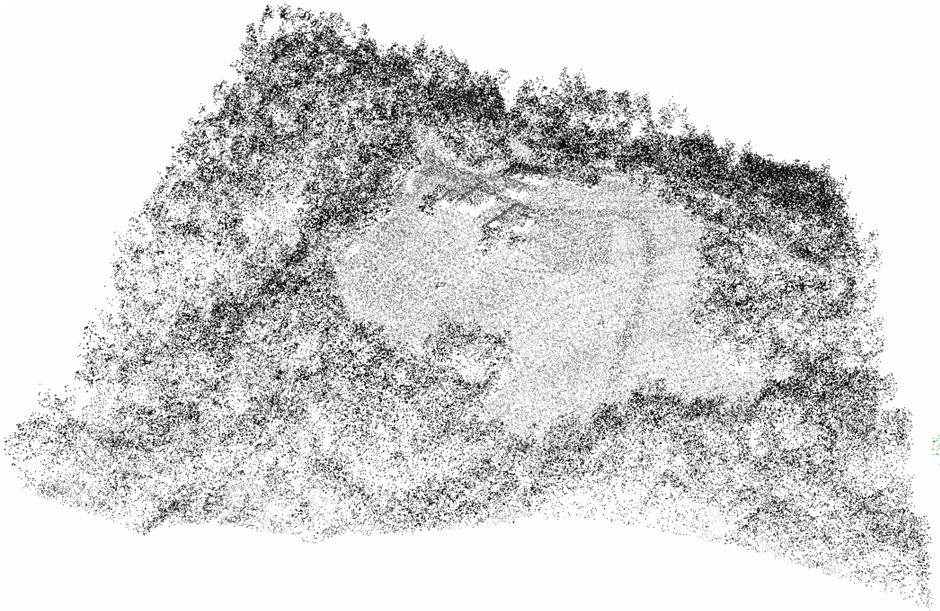
- What was the original application for which this data was collected?
- What is the quality of the original data?
- How much post-processing has been done?
- Is this data suited to my project?

Lidar data

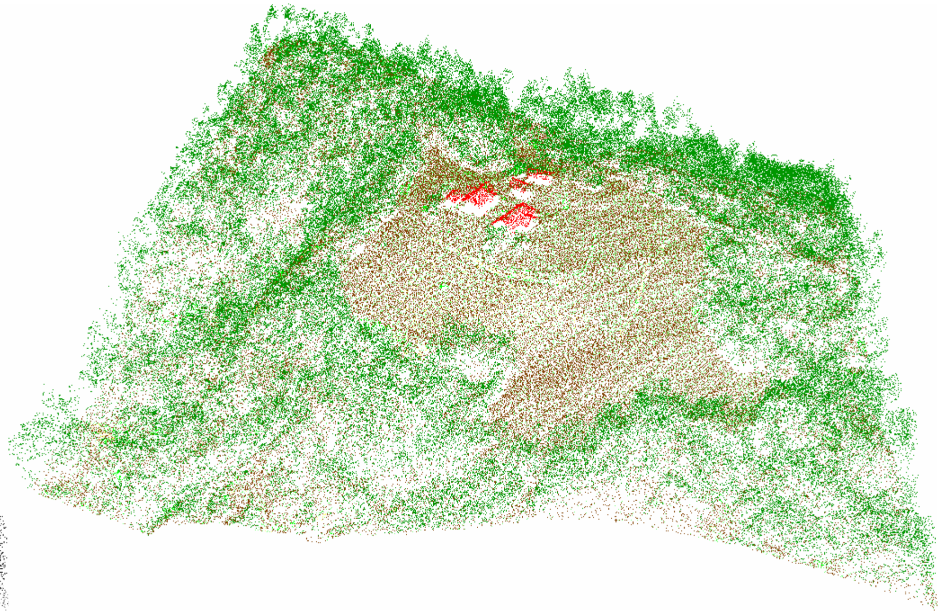
X	Y	Z	Intensity	echo no.
389838,44	126050,22	476,88	11	1
389838,45	126049,95	452,19	17	2
389838,49	126051,21	452,15	18	3
389838,32	126054,99	468,47	10	1
389838,31	126055,10	469,56	14	1
389838,34	126055,85	469,25	19	1
389838,44	126050,13	452,22	13	3
389838,48	126051,39	452,18	40	3
389838,30	126055,27	469,62	11	1
389838,35	126056,63	469,34	14	1
389837,72	126049,74	452,45	55	3
389837,74	126050,35	452,47	22	2
389837,77	126051,04	452,30	27	2
389837,79	126051,65	452,33	28	2
389837,81	126052,32	452,22	52	3
389837,84	126052,98	452,14	26	3
389837,86	126053,62	452,09	41	2
389837,88	126054,24	452,10	45	2
389837,91	126054,90	452,03	26	3
389837,93	126055,54	452,00	25	3
389837,96	126056,22	451,88	23	3
389837,98	126056,85	451,86	24	3
389838,04	126058,63	452,05	18	3
389838,06	126059,22	452,14	31	2
...	...	...	...	...



Point clouds

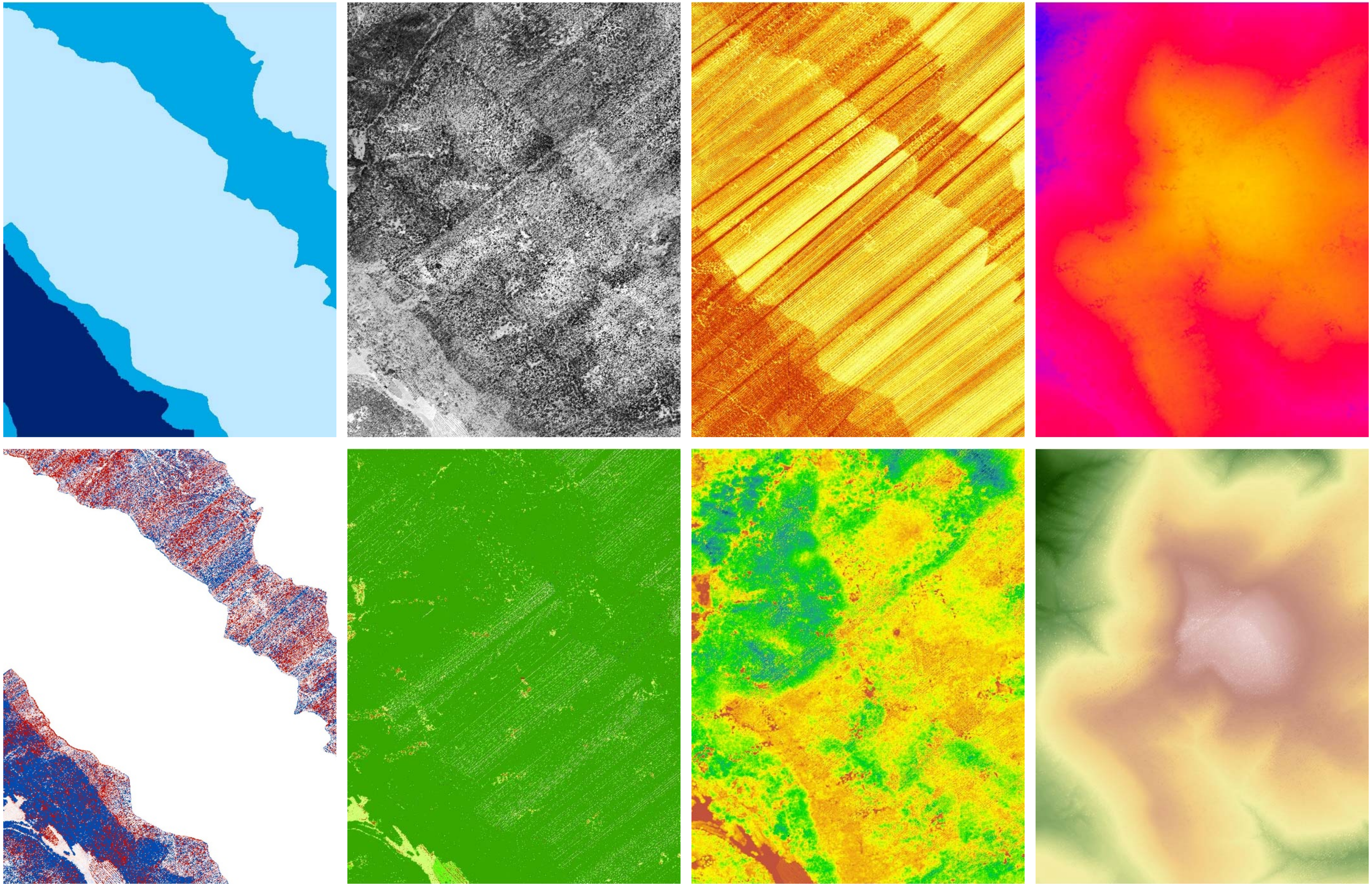


point cloud



classified point cloud

Different representations of data



ASCII

```
X,Y,Z
472302.5,117449.5,497.20
472303.5,117449.5,497.29
472304.5,117449.5,497.40
472305.5,117449.5,497.57
472306.5,117449.5,497.74
472307.5,117449.5,497.88
472308.5,117449.5,497.93
472309.5,117449.5,497.85
472310.5,117449.5,497.97
472311.5,117449.5,498.09
472312.5,117449.5,498.13
472313.5,117449.5,498.06
472314.5,117449.5,498.22
472315.5,117449.5,498.13
```

```
ncols 350
nrows 300
xllcorner 472300.000000
yllcorner 117150.000000
cellsize 1.000000
NODATA_value -9999.0
497.56 497.25 497.20 497.29 497.40 497.57 497.74 497.88
497.93 497.85 497.97 498.09 498.13 498.06 498.22 498.13
498.00 497.93 497.91 497.91 497.94 497.90 497.82 497.67
497.69 497.78 497.74 497.74 497.60 497.68 497.66 497.58
497.54 497.26 497.08 496.98 496.97 496.94 496.91 496.67
496.63 496.60 496.64 496.64 496.28 496.14 496.00 495.99
495.98 495.86 495.73 495.49 495.25 495.05 494.88 494.69
494.22 494.00 493.83 493.65 493.51 493.31 493.13 492.86
492.55 492.28
```

Data formats and sizes

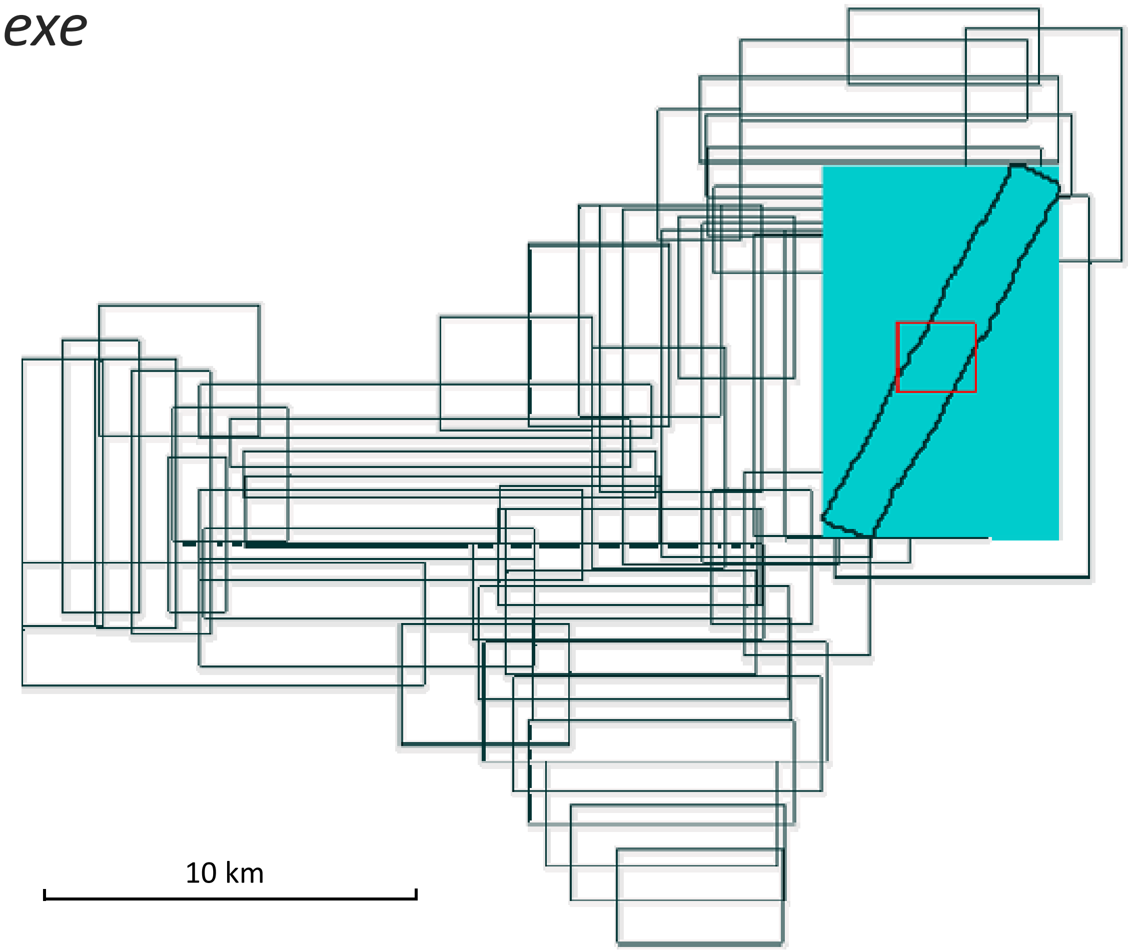
Group	File	kB
the same information	lidar_data.txt	74.400
	lidar_data.las	33.500
	lidar_data.zlas	8.100
	lidar_data.laz	7.400
the same information	lidar_data.xyz	2.700
	lidar_data.asc	700
	lidar_data.tif	400
	lidar_data_HS.tif	100

Compression of lidar data

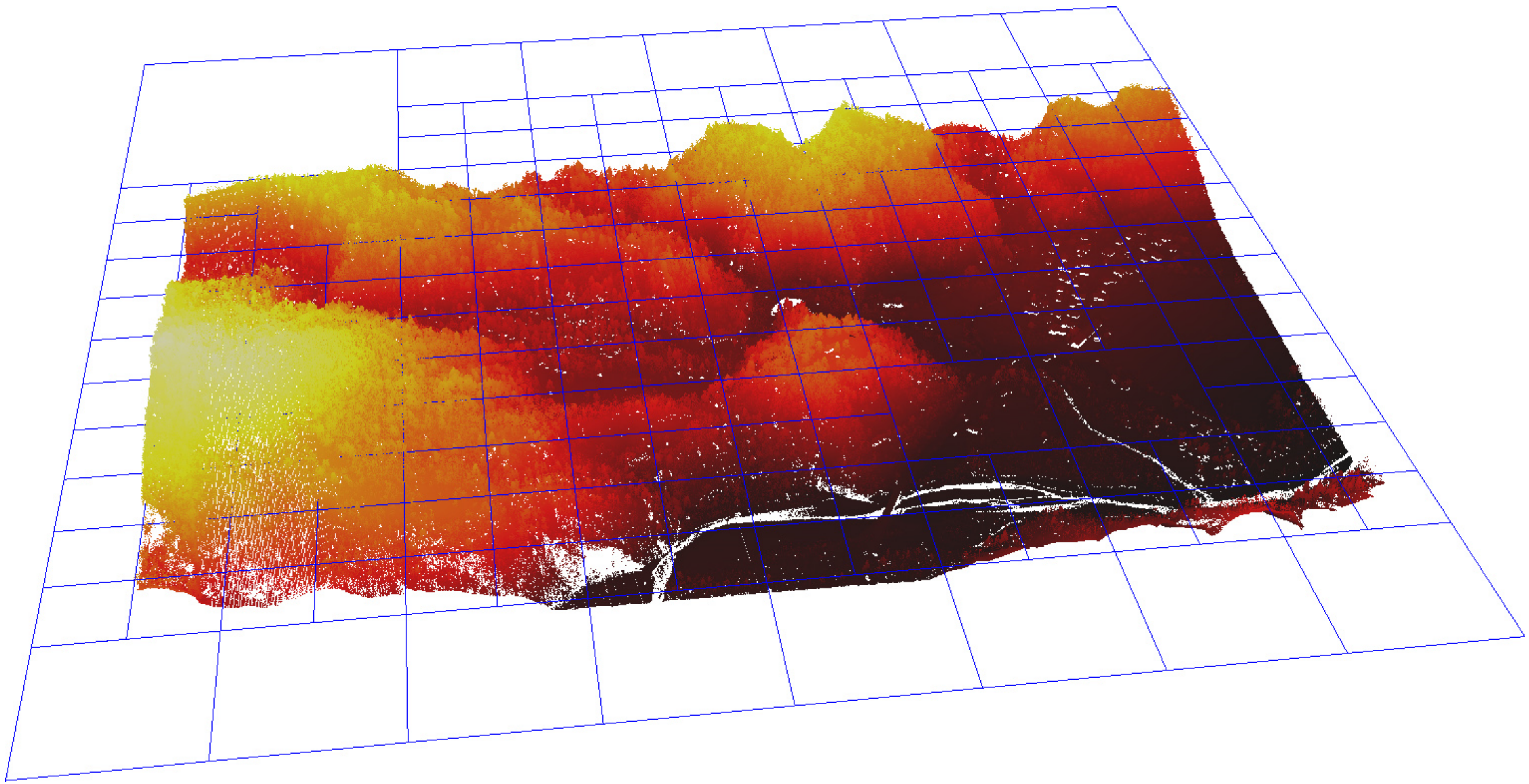
- *laszip.exe*
- LAZ is lossless compressed LAS
  - ≈ 9 % of original file size
- fastest encoding / decoding
- highest compression
- integrates with LAX spatial indexing
- open source --> free

Spatial indexing

- *lasindex.exe*



Spatial indexing



Density and accuracy of laser points

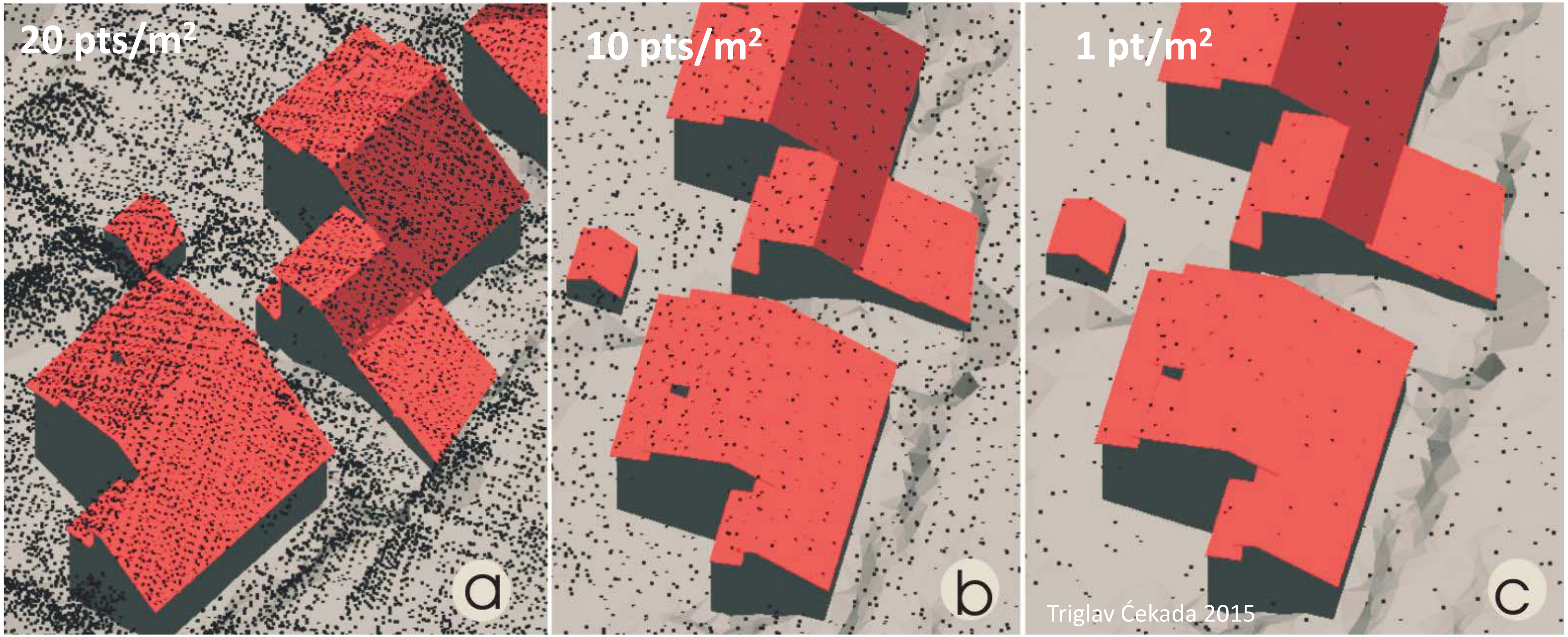
Accuracy

- the biggest influence:
    - accuracy of IMU (rotational angles of the platform)
    - accuracy of GNSS
    - accuracy of laser system
    - flying height
    - maximum scanning angle

- vertical accuracy 10-15 cm
- planimetric accuracy 30 cm

Point density

- 5-10 pts/m² intermediate density:
  - general DTM/DSM, e.g. for flood analyses
  - detailed topographic/geomorphological analyses
  - topographic mapping at scales 1:10,000 or 1:5,000



Lidar density

- density of what?
  - points/m^2
    - first returns
    - last returns
    - only returns
    - per flightline/combined
  - ground points/m^2
  - pulses/m^2

Metadata

Metadata

Metadata* is essential in order to maintain the scientific integrity of the entire process.

HOW, WHEN and WHY was the lidar data collected?
WHAT, HOW and WHY did we process?

*data that provides information about other data

Metadata – Data Collection

- lidar sensor, IMU and GNSS make and model,
- nominal scanning density,
- nominal swath overlap,
- flight height,
- date of data collection and
- purpose of data collection.

RIEGL LMS-Q780 laser scanner, IGI Aerocontrol Mark II.E 256Hz inertial navigation system (IMU) and Novatel OEMV-3 GNSS have been used.
Nominal scanning density 5 pulses/m^2, nominal swath overlap unknown (scanning angle ±30°).
Flight height 1200 do 1400 m above surface.
Date of data collection 14., 23. and 24. February 2014.
Data was collected as a part of the nation wide general purpose scanning.

Metadata – Raw data processing

- always implemented by the data collection provider,
- often black-boxed,
- if you can access the metadata, record it
- if not, then not ☹
- it would only become very important if raw data was to be reprocessed

Flight trajectory was calculated with GrafNav v8.50 using the dGPS method.
GNSS and IMU data were merged with AeroOffice v5.1f.
TerraMatch was used for line matching.
Raw data have been processed and exported to LAS 1.2 with RiPROCESS v1.5.9 (modules RiANALYZE v6.0.2 and RiWORLD v4.5.8) in D96/TM coordinate system with elipsoid heights.

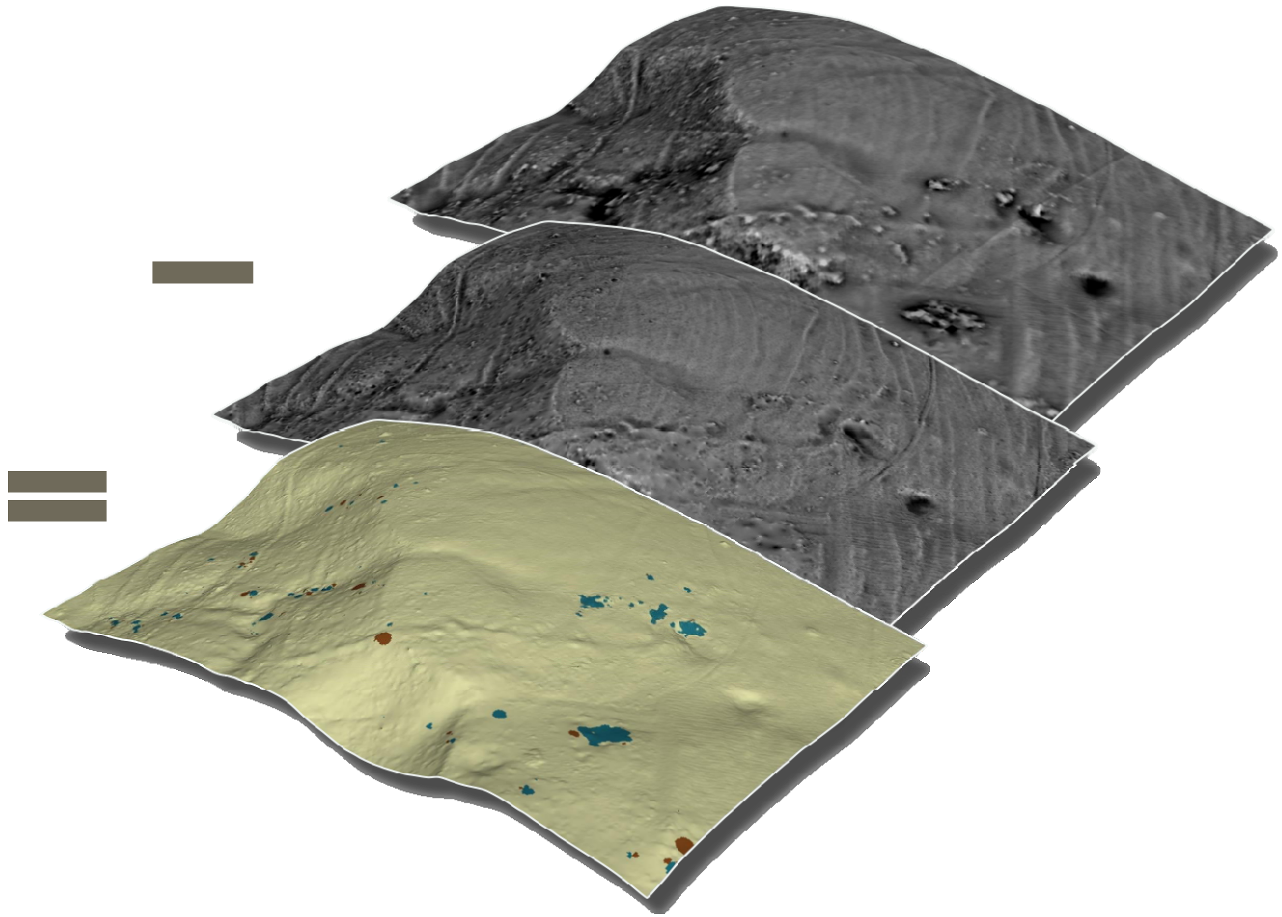
Metadata – Post-processing

- description of the processing goal (e.g. production of a terrain model, removing just the vegetation),
- method(s),
- software incl. parameter settings,
- DEM/DSM resolution.

Point cloud has been filtered for obtaining the ground surface suitable for archaeological interpretation (for method see Štular, Lozić 2016). Ground points have been detected using LAStools lasground module (settings: terrain type – forest or hills, granularity - ultra fine; ignore points with classification 7). 0,5 m DEM was interpolated with ordinary kriging in Surfer v.11 (settings: No. of sectors to search – 4, maximum No. of data from all sectors 64, maximum No. of data from each sector 16, maximum No. of data in all sectors – 8, blank node if more than 3 sectors are empty, radius 20 m).
(Data source: web service eVode at http://evode.arso.gov.si/indexd022.html?q=node/12, GKOT D48 files GK389_132, GK389_133, GK390_132 and GK390_133.

Metadata – Post-processing

- Manual interventions can be documented with difference map

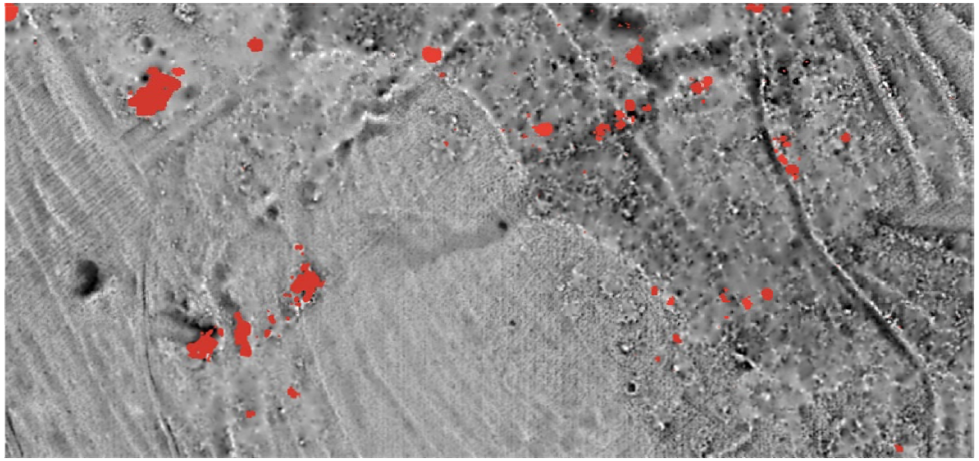


Metadata – DEM/DSM visualization

- description of the processing goal (e.g. multi-purpose visualization, on-screen archaeological interpretation, print publication),
- method(s),
- software incl. parameter settings

For printed publication complex multiple visualization merging RVT software derived visualizations sky-view factor and openes (settings for both: No. of search directions 16, radius 10 pixels, low level of noise removal) and WhiteboxGIS software derived deviation from mean elevation (settings: radius 7 cells) has been implemeted.

Metadata for scientific publication



RIEGL LMS-Q780 laser scanner, IGI Aerocontrol Mark II.E 256Hz inertial navigation system (INS) and Novatel OEMV-3 GNSS have been used. Nominal scanning density 5 pulses/m^2, nominal swath overlap unknown (scanning angle ±30°). Flight height 1200 do 1400 m above surface. Date of data collection 14., 23. and 24. February 2014. Data was collected as a part of the nation wide general purpose scanning.

Flight trajectory was calculated with GrafNav v8.50 using the dGPS method. GNSS and INS data were merged with AeroOffice v5.1f. TerraMatch was used for line matching. Raw data have been processed and exported to LAS 1.2 with RiPROCESS v1.5.9 (modules RiANALYZE v6.0.2 and RiWORLD v4.5.8) in D96/TM coordinate system with elipsoid heights.

Point cloud has been filtered for obtaining the ground surface suitable for archaeological interpretation (for method see Štular, Lozić 2016). Ground points have been detected using LAStools lasground module (settings: terrain type – forest or hills, granularity - ultra fine; ignore points with classification 7). 0,5 m DEM was interpolated with ordinary kriging in Surfer v.11 (settings: No. of sectors to search – 4, maximum No. of data from all sectors 64, maximum No. of data from each sector 16, maximum No. of data in all sectors – 8, blank node if more than 3 sectors are empty, radius 20 m).

(Data source: web service eVode at http://evode.arso.gov.si/indexd022.html?q=node/12, GKOT D48 files GK389_132, GK389_133, GK390_132 and GK390_133.

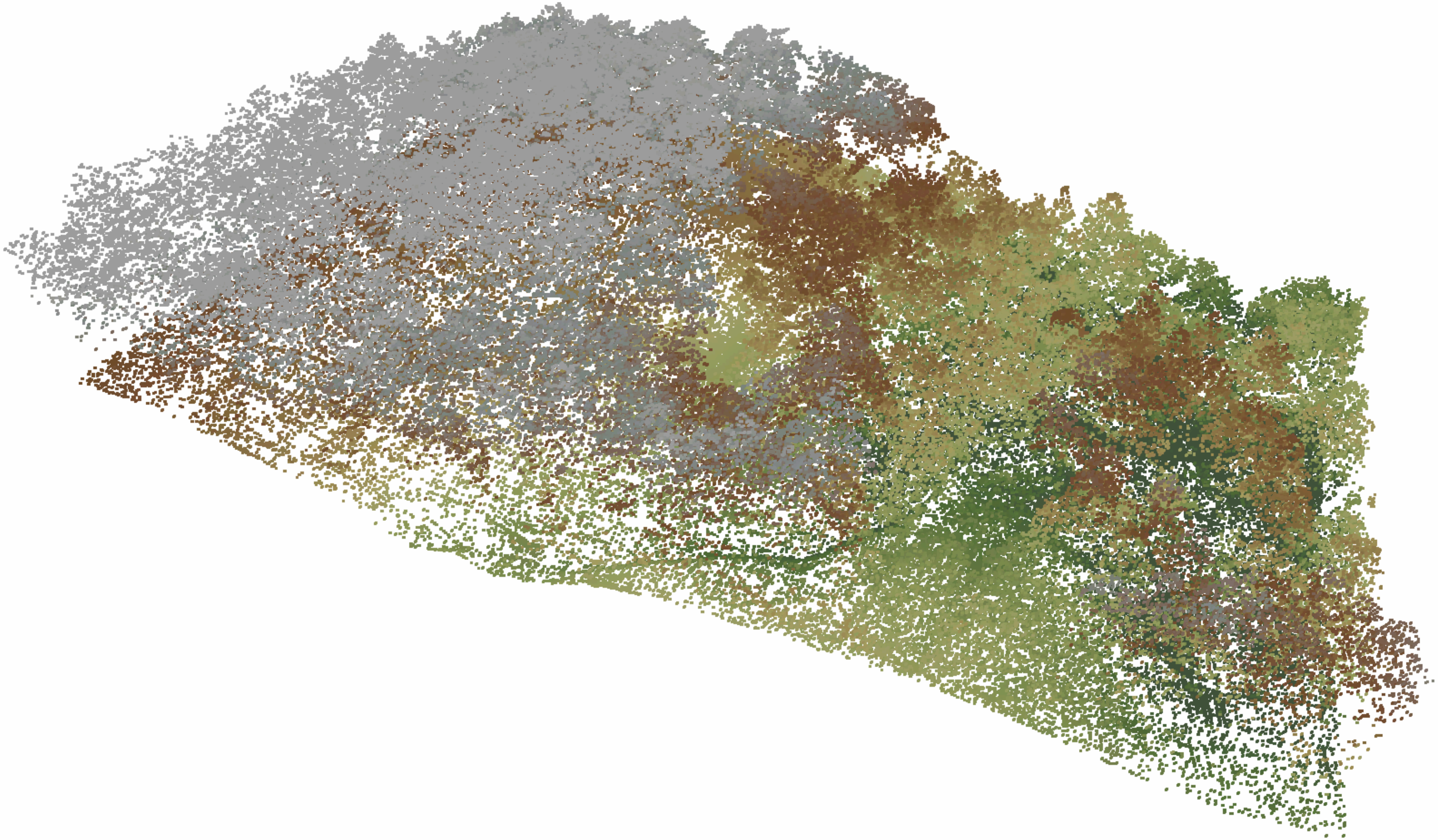
For printed publication complex multiple visualization merging RVT software derived visualizations sky-view factor and openes (settings for both: No. of search directions 16, radius 10 pixels, low level of noise removal) and WhiteboxGIS software derived deviation from mean elevation (settings: radius 7 cells) has been implemeted.

Metadata for scientific publication

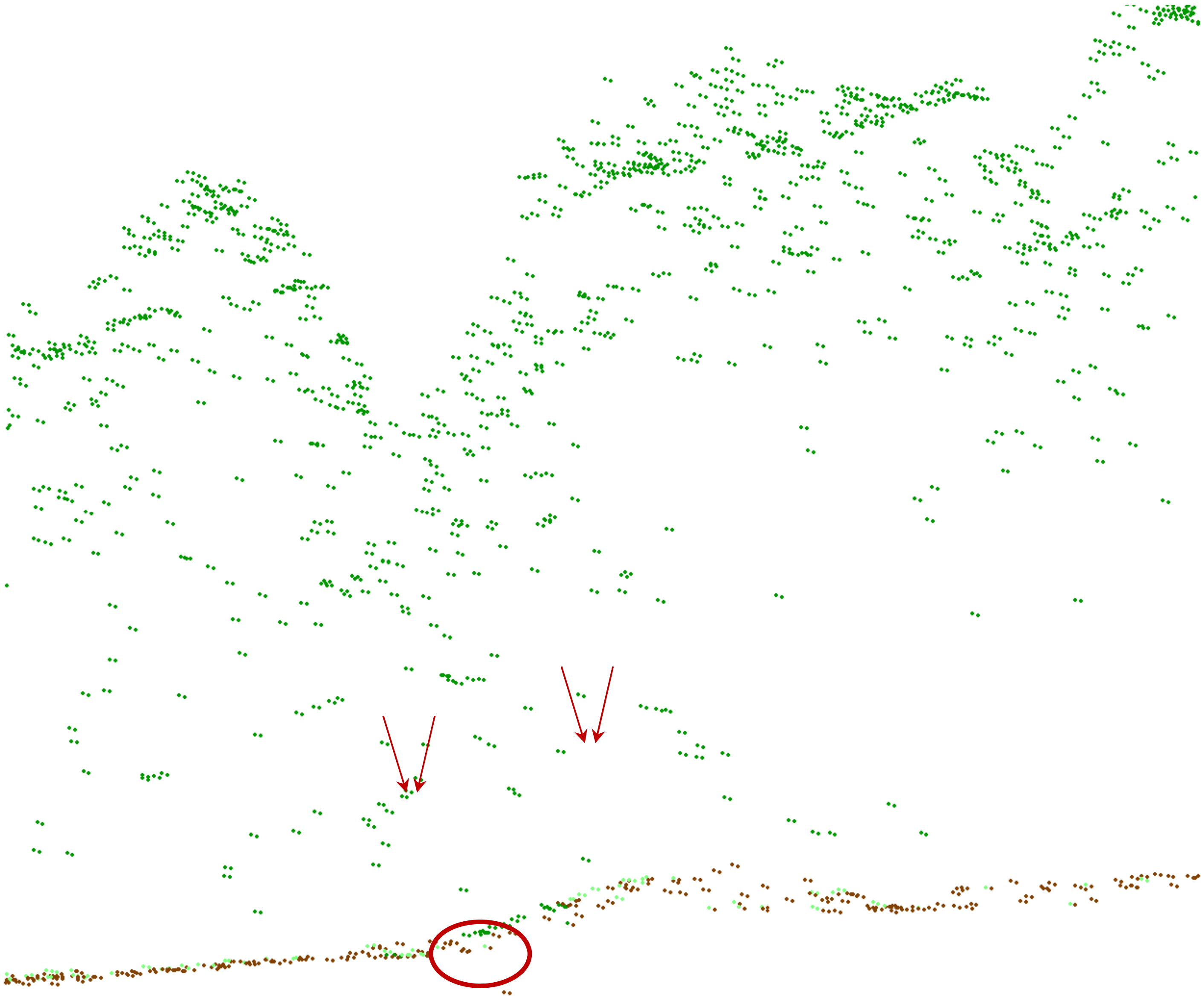
Lidar image.

Visual quality control

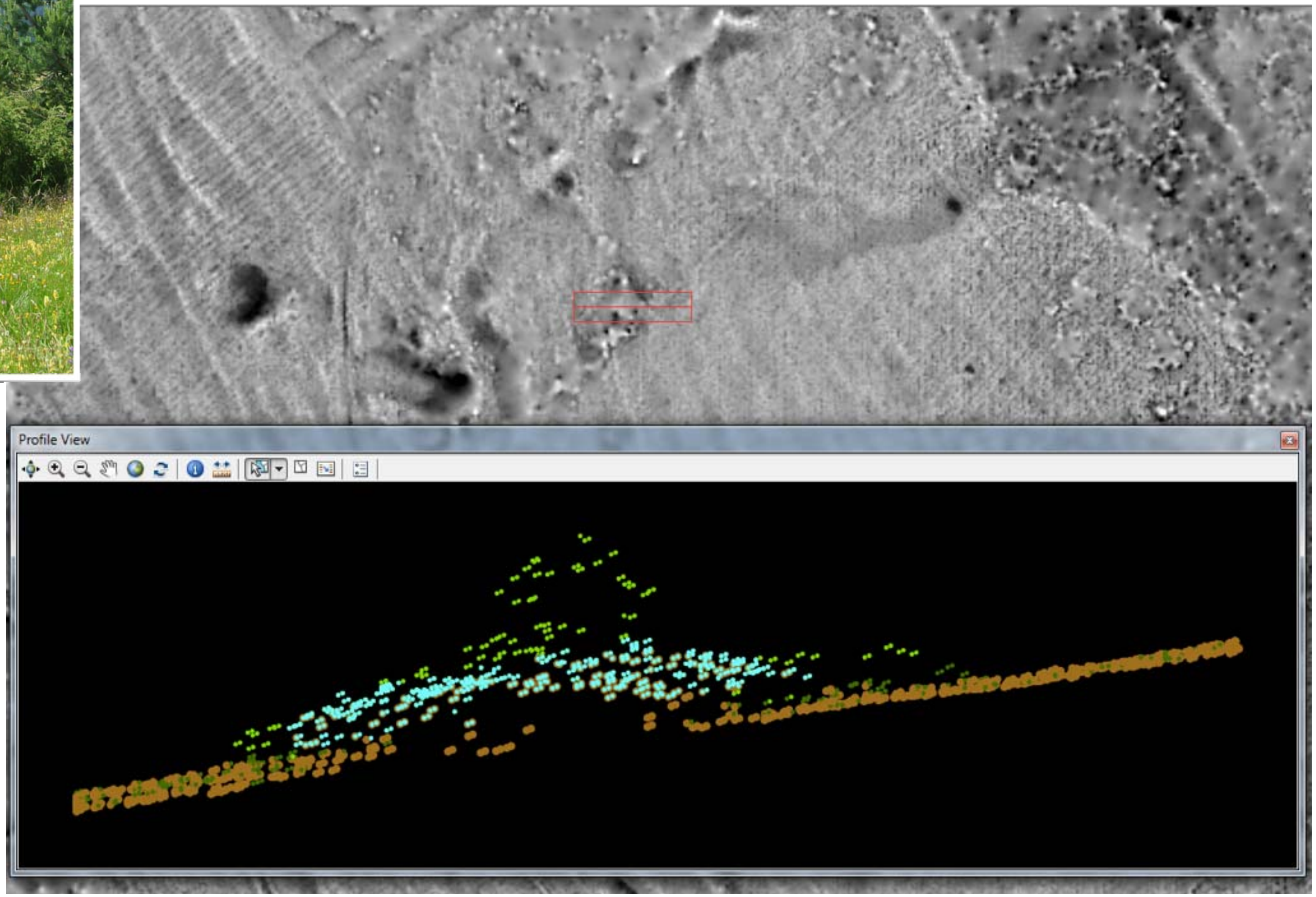
Visual quality control

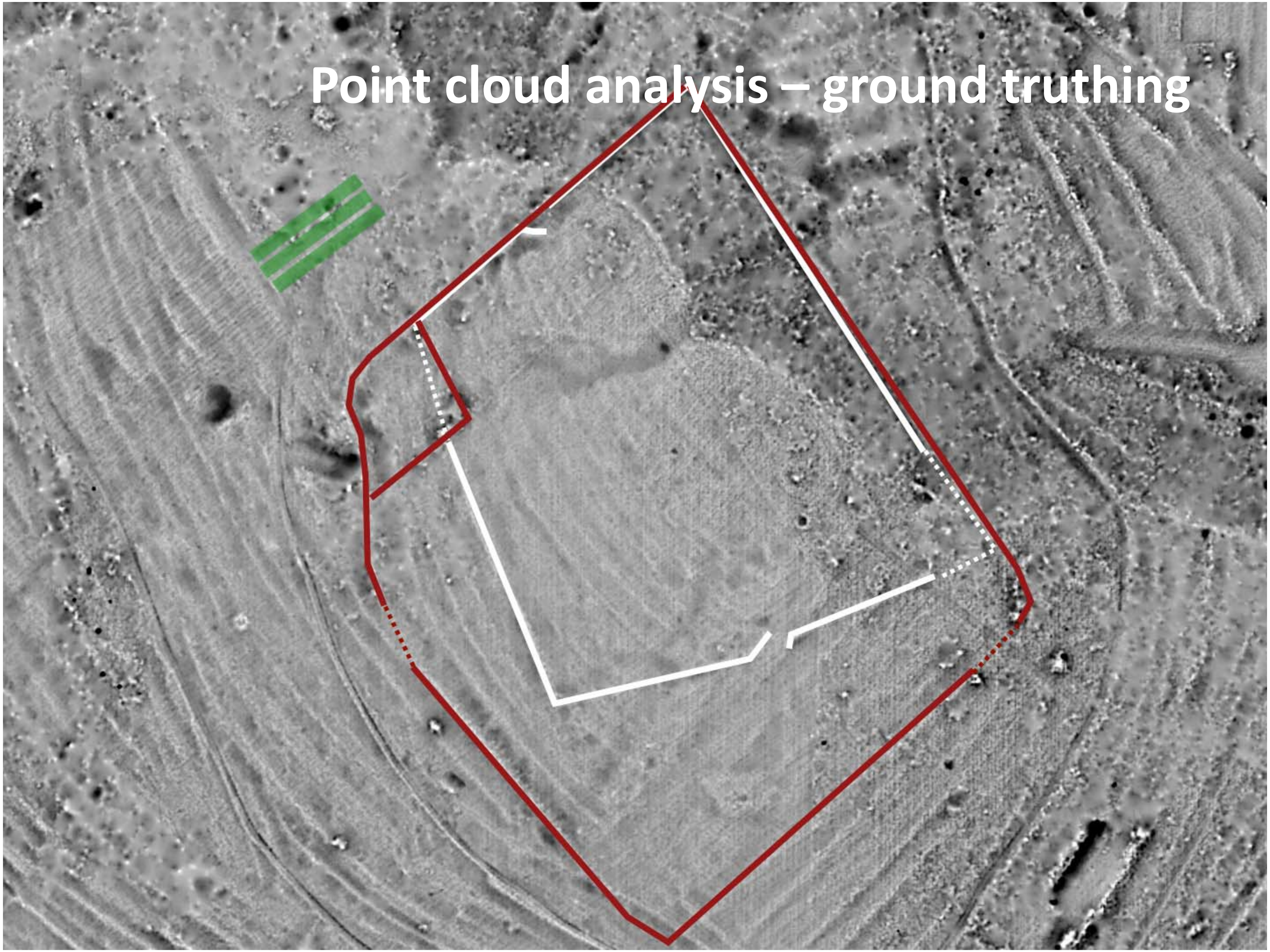


Point cloud analysis

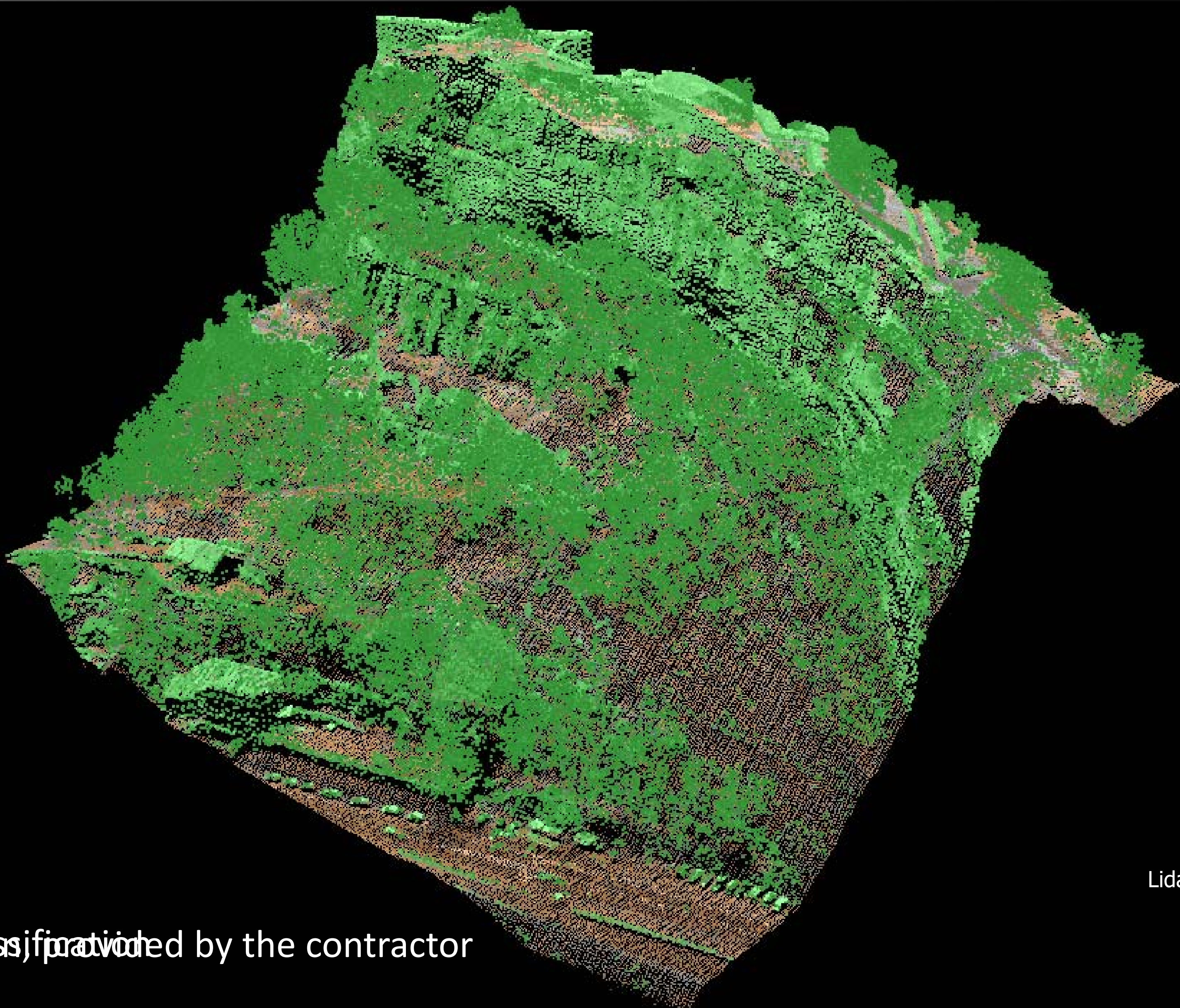


Point cloud analysis – ground truthing




Point cloud analysis – ground truthing

Observation of geological structures

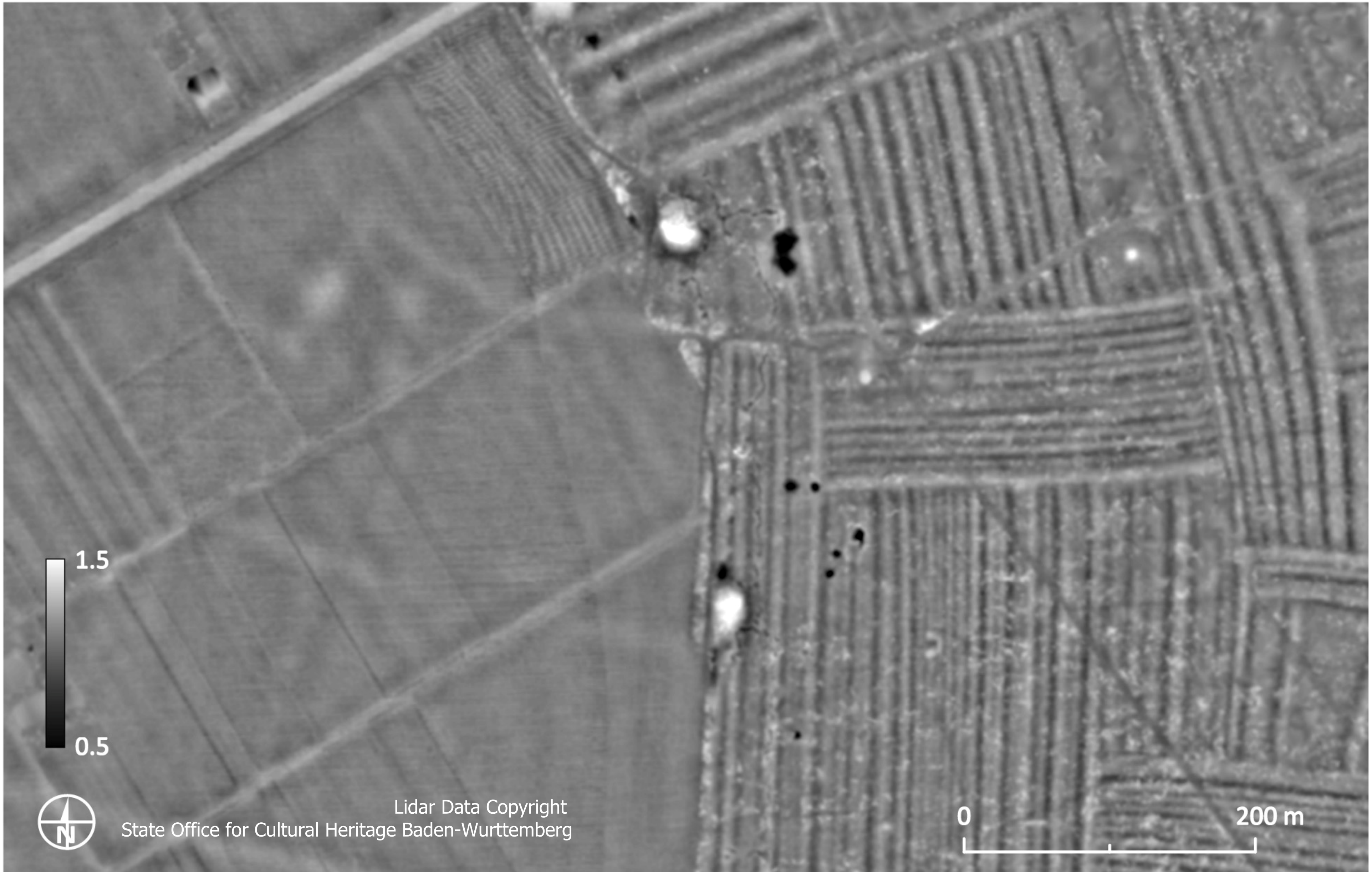


Lidar Data Copyright
UFC

classification, provided by the contractor

manual classification

Field system



Data processing and visualization tools

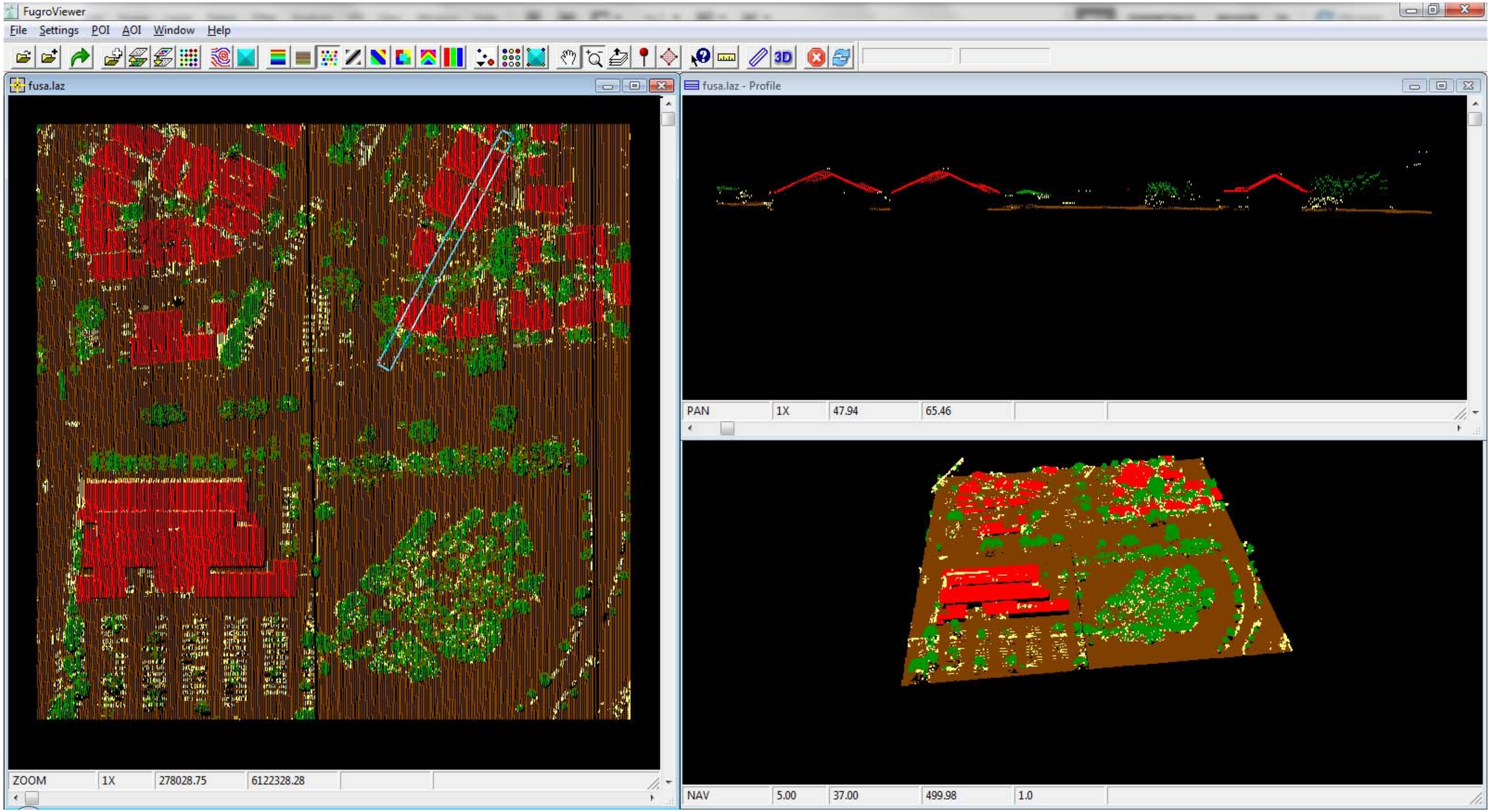
Data processing

- LASTools (Rapidlasso)
- OPALS (TU Vien)
- TerraScan in TerraModeller (Terrasolid)
- FUSION (USFS)
- PointVue LE (GeoCue)
- MARS FreeView (Merrick)
- gLidar, FERI Maribor
- GRASS
- ALDPAT
- MCC
- …

Point cloud viewing

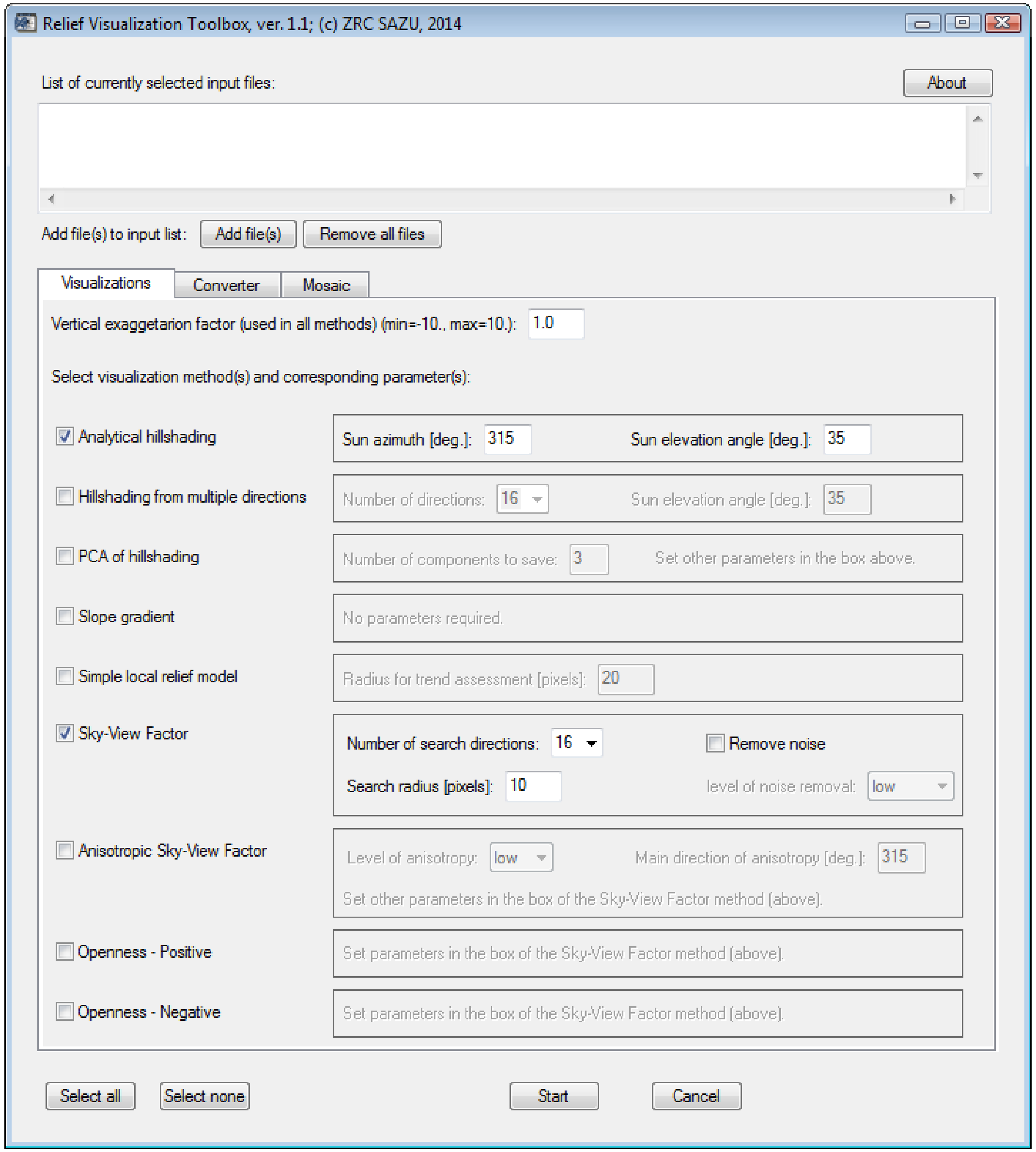
- FugroViewer, Fugro
- LasView, LAStools, Rapidlasso
- FUSION
- …

FugroViewer



http://www.fugroviewer.com/request/default.asp

Relief Visualization Toolbox (RVT)



iaps.zrc-sazu.si/en/rvt

- computation of elevation data visualizations
- free
- standalone
- very fast

Main characteristics

- narrow range of techniques with only the essential options
- very easy to use
- can process multiple files and all techniques in one go
- file format converter
- file merger
- results can be used in non-GIS software

LASTools

LASTools

- Marin Isenburg
- Rapidlasso GmbH

- streamline processing

- *http://lastools.org/download/lastools.zip*
- *http://groups.google.com/group/lastools*

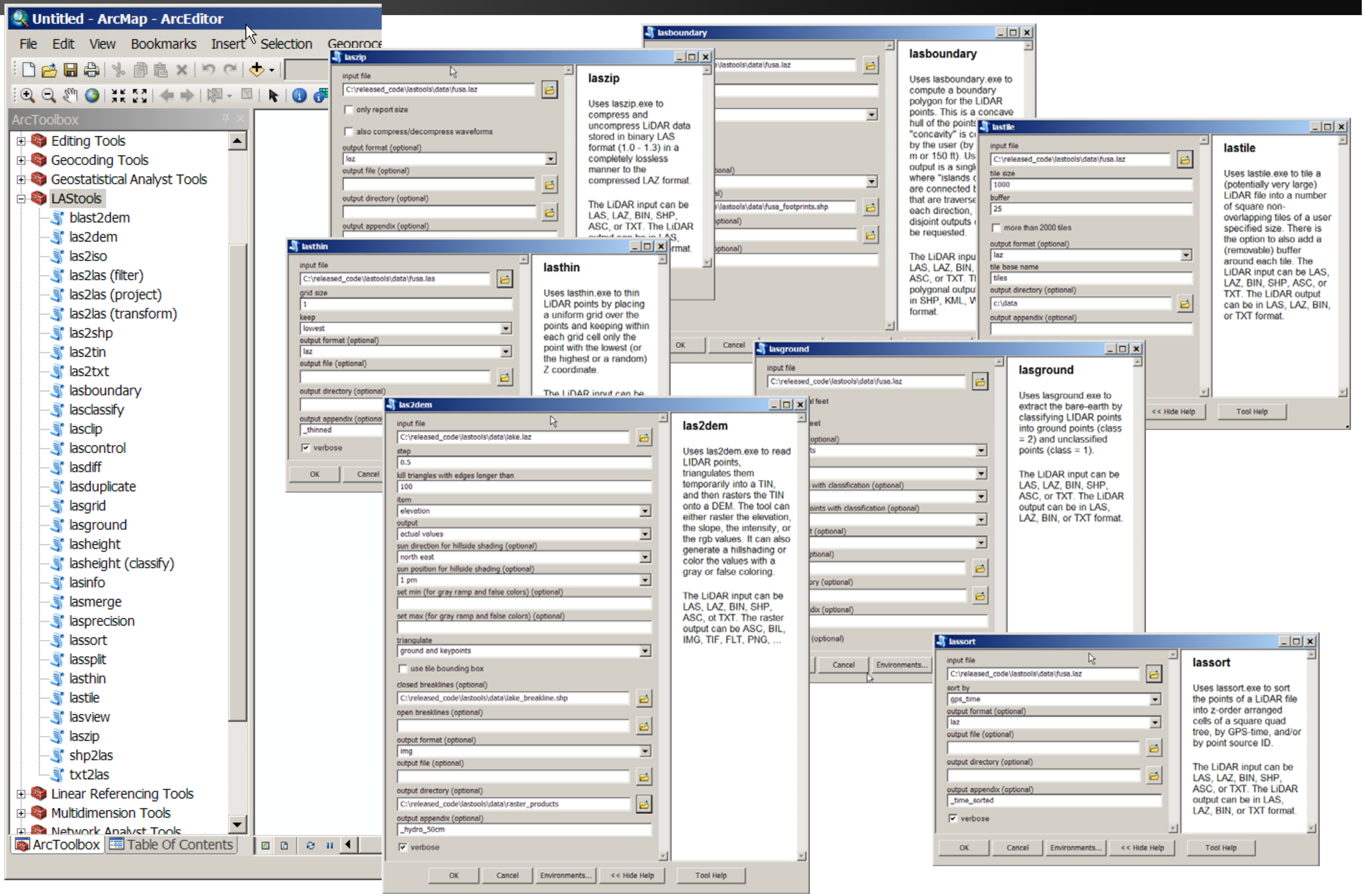
LASTools

- executables (.exe files)

- GUI (e.g. las2las *.las -gui)

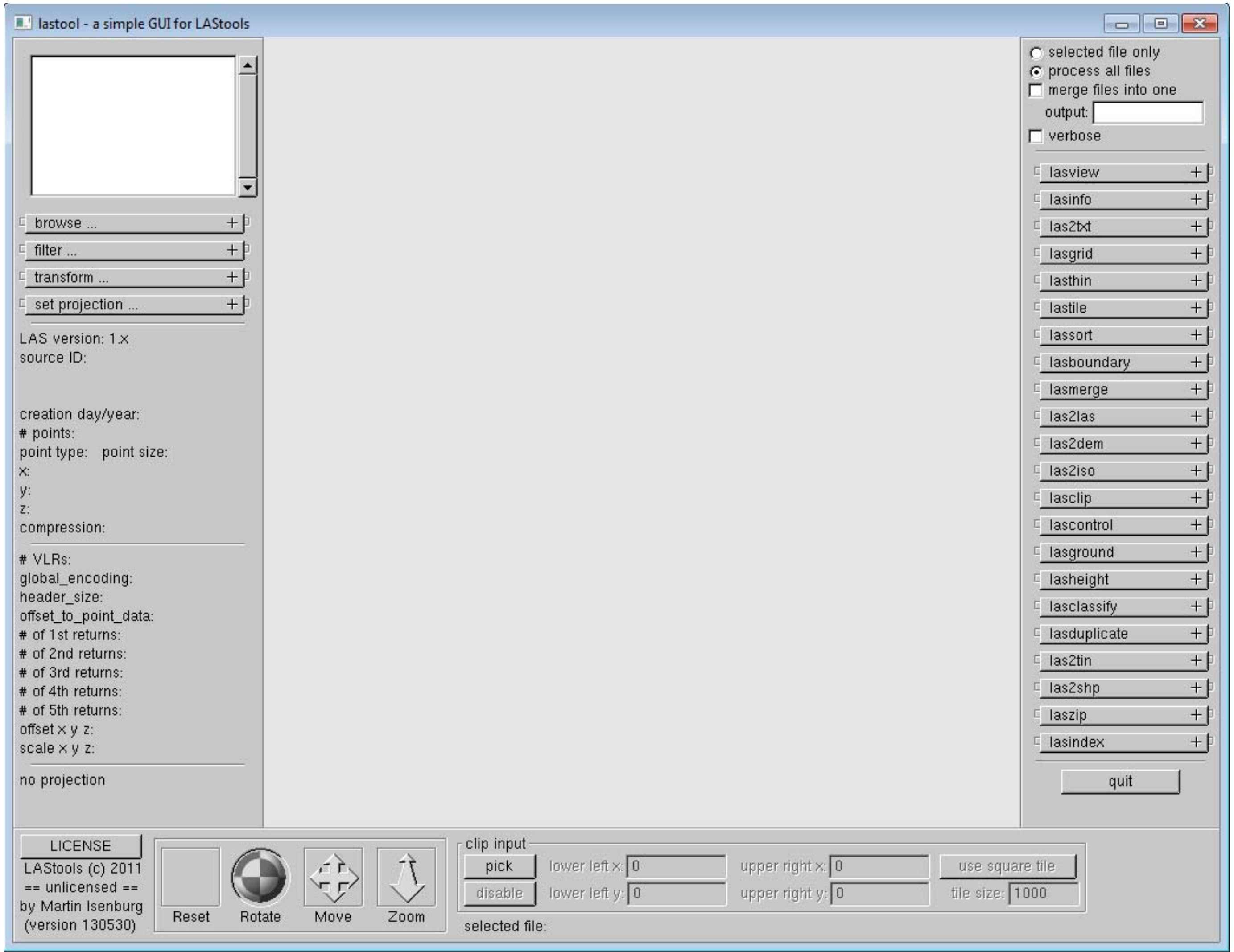
- Toolboxes
  - ArcGIS
  - QGIS
  - IMAGINE

ArcGIS toolbox



„One to rule them all“

- lastools.exe



- make Lastools available from any location

```
set PATH=%PATH%;D:\lastools\bin
```

- making the process more readable

```
lasgrid -i flightlines\*.laz -merged ^
        -density -step 10 -false -set_min_max 0 200 ^
        -o density.png
```

Pipelineing

Lister - [d:\LASTools\example_batch_scripts\typical_lidar_preparation_pipeline.bat]

Datoteka Uredi Možnosti Kodiranje Pomoč 55 %

```
::
:: a batch script for converting raw flight lines (not tiled) into
:: a number of products with a tile-based multi-core batch pipeline
::

::
:: specify parameters
::

:: allows you to run the script from other folders
set PATH=%PATH%;C:\Users\Public\lastools\bin;

:: here we specify the directory (e.g. the folder) in which the
:: original raw flight lines files are stored
set RAW_FLIGHT_LINES=E:\Raw_strips

:: here we specify in which format the original raw flight lines
:: files are stored in the RAW_FLIGHT_LINES folder
set RAW_FORMAT=las

:: here we specify the directory (e.g. the folder) in which the
:: temporary files are stored
set TEMP_FILES=C:\lastools_temp

:: here we specify the directory (e.g. the folder) in which the
:: resulting output files are stored
set OUTPUT_FILES=C:\lastools_output

:: here we specify where the GEOID transformation grid is stored
set GEOID=C:\Users\Public\Documents\WGS84_to_EGM2008_GRID_Adjustment\wgs84_2_egm0
8.asc

:: here we specify the number of cores we want to run on
set NUM_CORES=7

:: here we specify the target tile size of the tiling
set TILE_SIZE=1000

:: here we specify the target buffer size for each tile
set TILE_BUFFER=50

rmdir %TEMP_FILES% /s /q

::
:: start processing
::

:: transform the strips to GEOID EGM2008 orthometric heights

mkdir %TEMP_FILES%\orthometric_strips
lasheight -i %RAW_FLIGHT_LINES%\*.%RAW_FORMAT% ^
```

Quality checking

lasinfo

lasinfo -i *.laz -compute_density
summarize contents of all LAZ files

↓

lasview

lasview -i *.laz
visually inspect the LAZ files

↓

lasboundary

lasboundary -i *.laz -holes -disjoint -oshp
compute the boundary around (and holes in) the lidar

↓

lasgrid

lasgrid -i *.laz -density -step 3 -set_minmax 0 20 -false -otif
create a point density grid of the data and visualize it with false color

↓

lasduplicate

lasduplicate -i *.laz -unique_xyz -onil
determine if the data contains any points more than once

↓

lasoverlap

lasoverlap -i *.laz -step 3
check vertical and horizontal alignment of flight lines (where they overlap)

1. Data, Quality control and Metadata

Practice

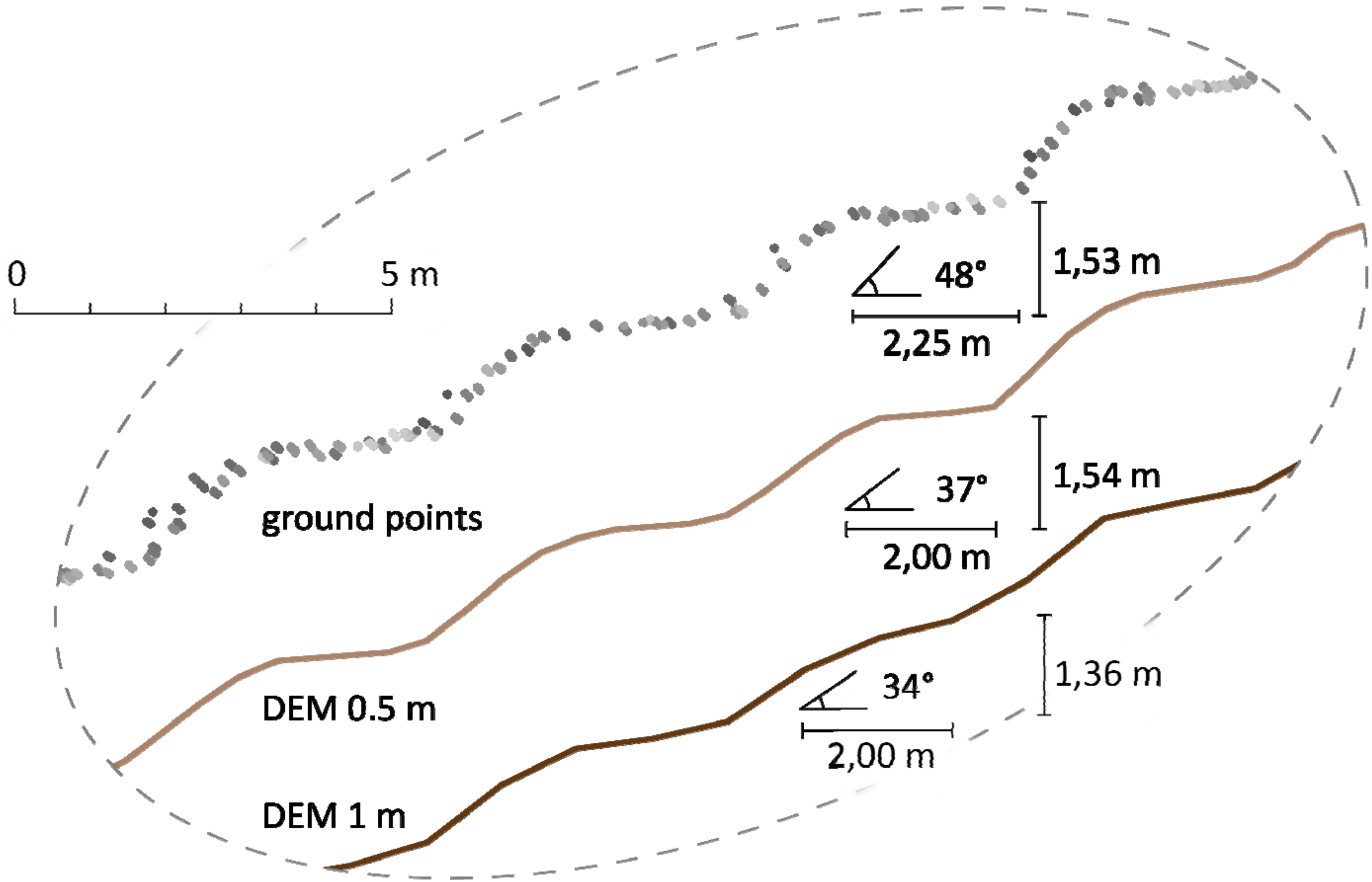
1. Data, Quality control and Metadata - practice

- Goals:
  - know your point cloud data
  - organize it
- Files:
  - PointCloud.laz (Sv.Helena)
- Tools:
  - lasinfo
  - lasoverlap
  - lasindex
  - lastile
  - lasboundary

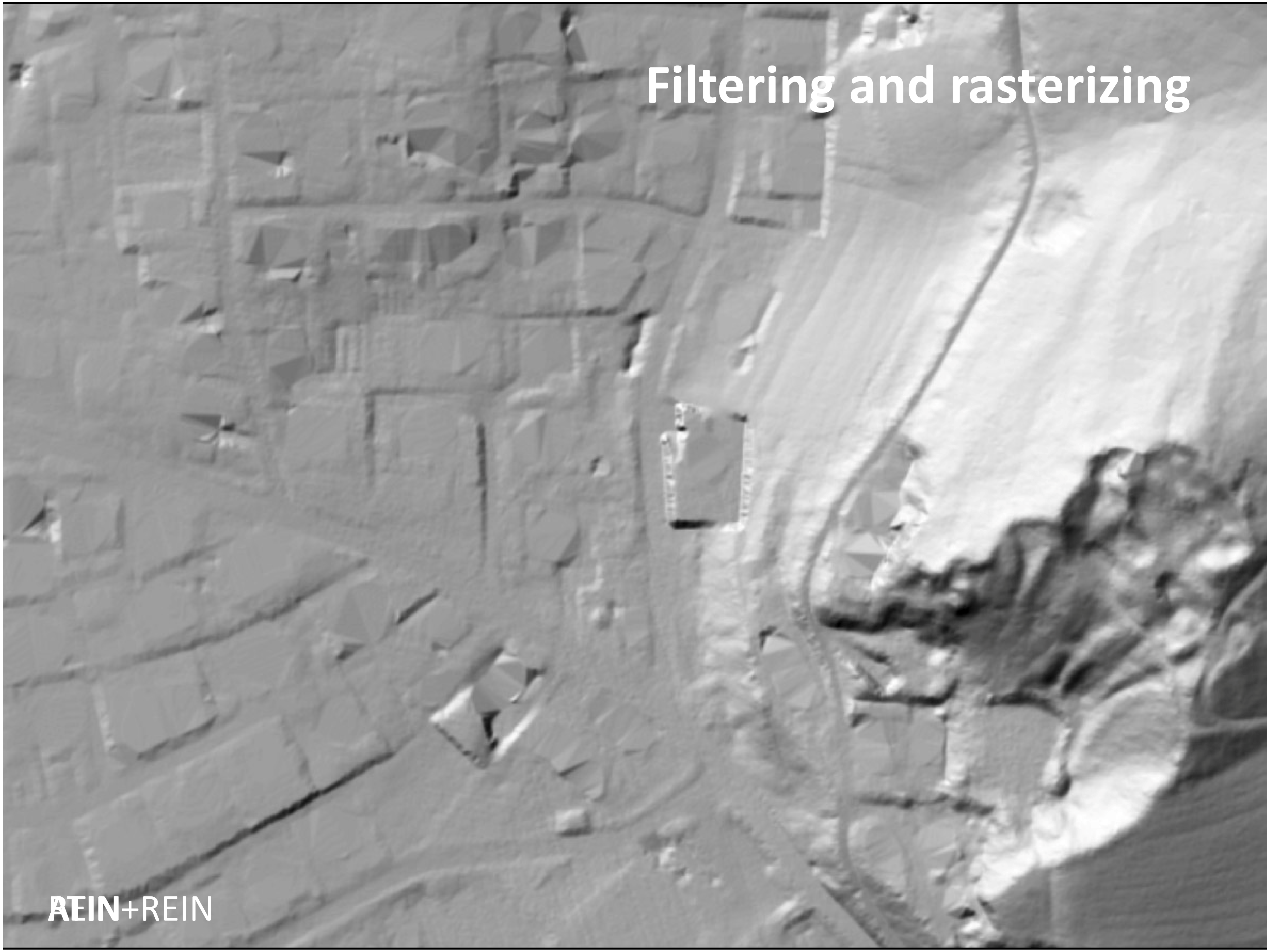
2. Rasterizing – making useful images

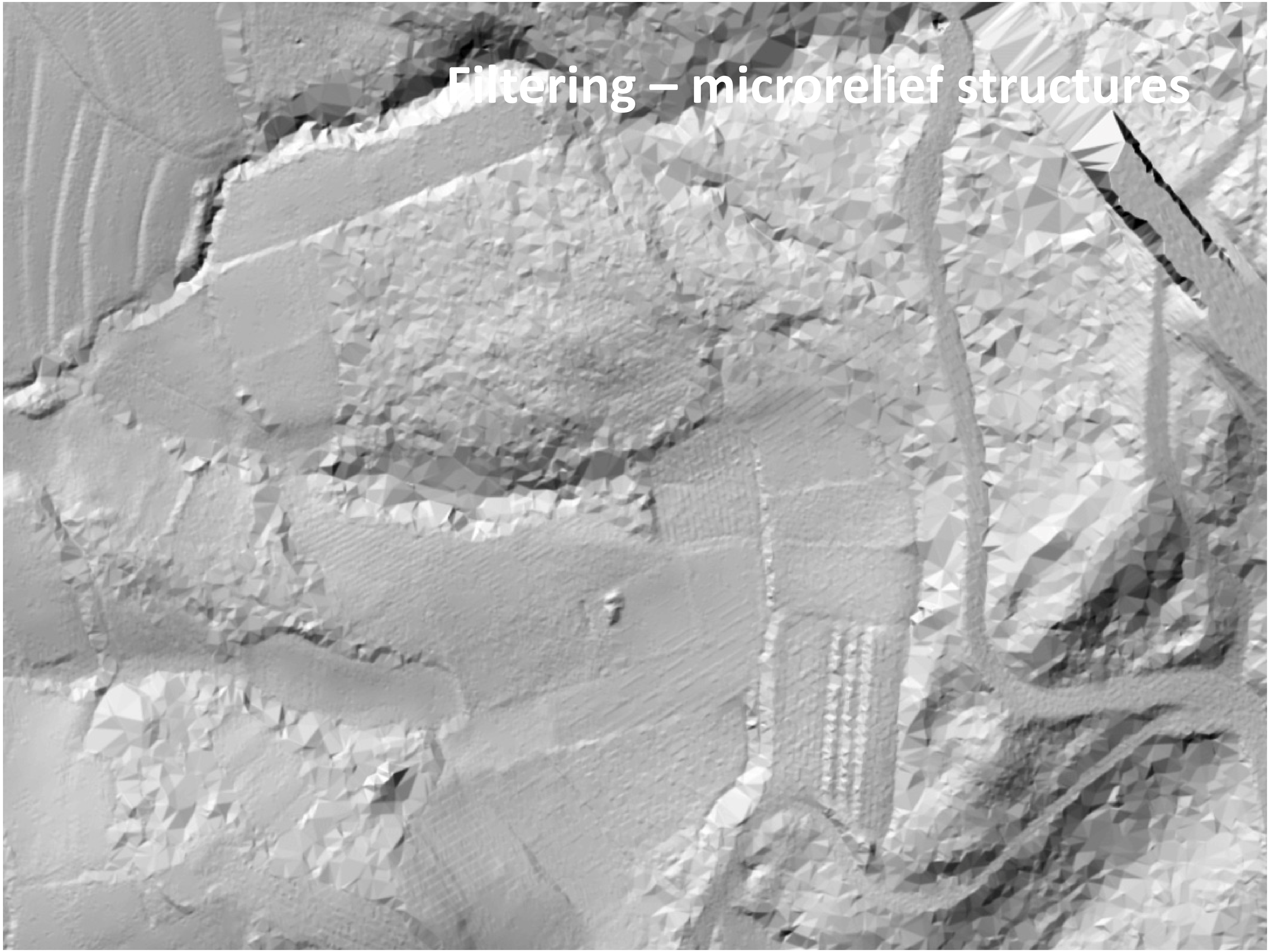
presentation & practice

Pixel size

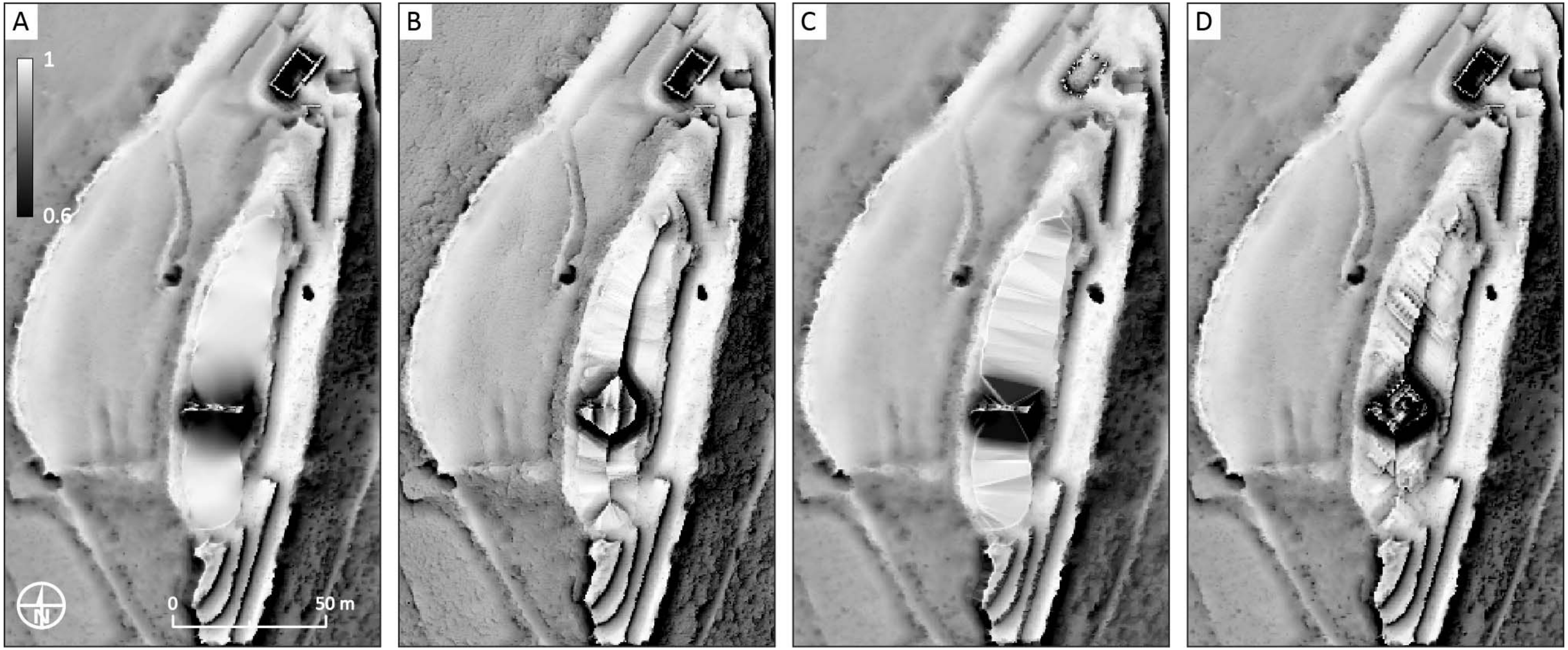

0
5 m
ground points
DEM 0.5 m
DEM 1 m
48°
2,25 m
1,53 m
37°
2,00 m
1,54 m
34°
2,00 m
1,36 m

Filtering and rasterizing




Filtering – microrelief structures

Interpolation methods



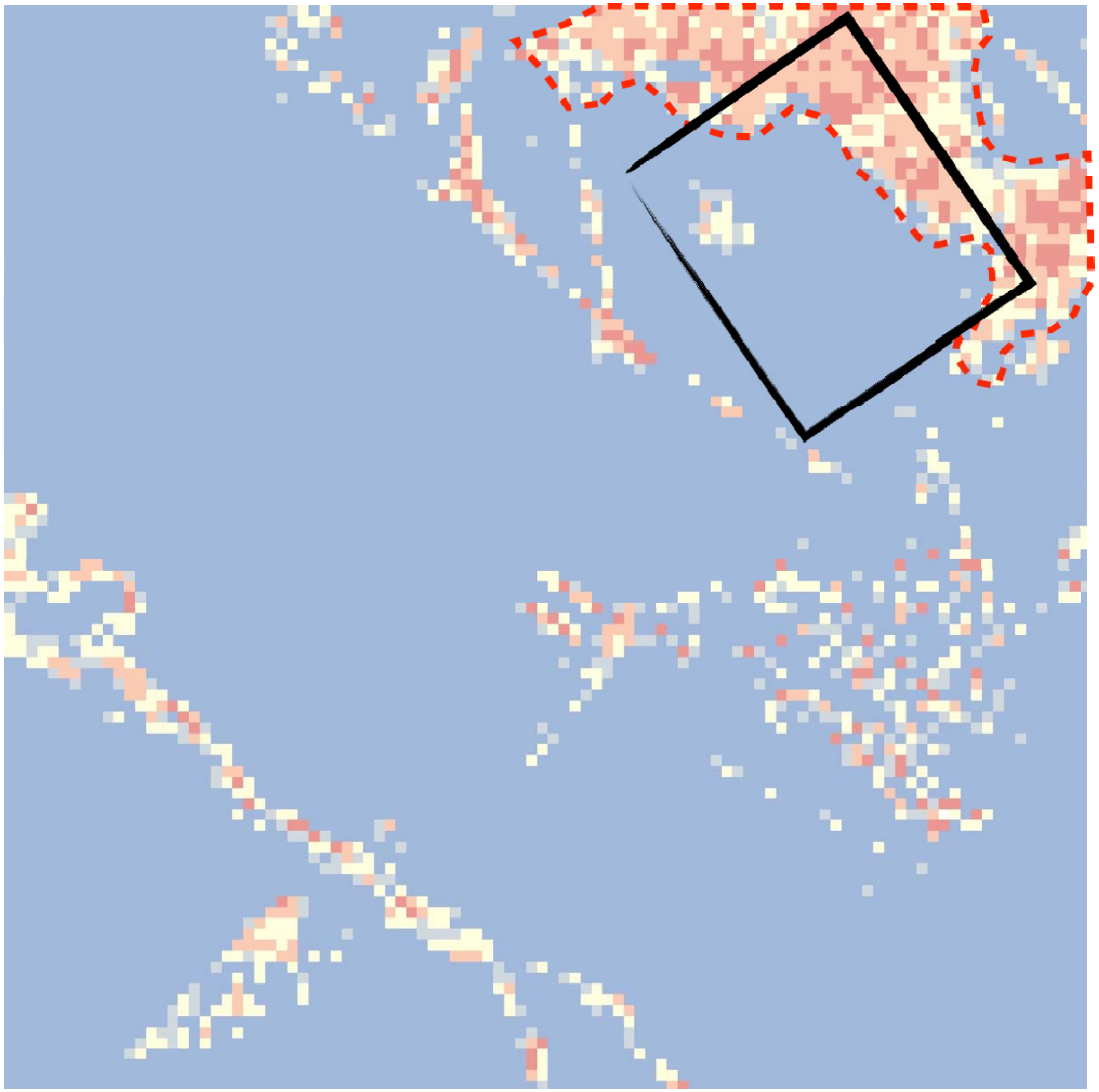
Natural Neighbours

Inverse Distance Weighted (IDW)

TIN with Repetitive Interpolation (REIN)

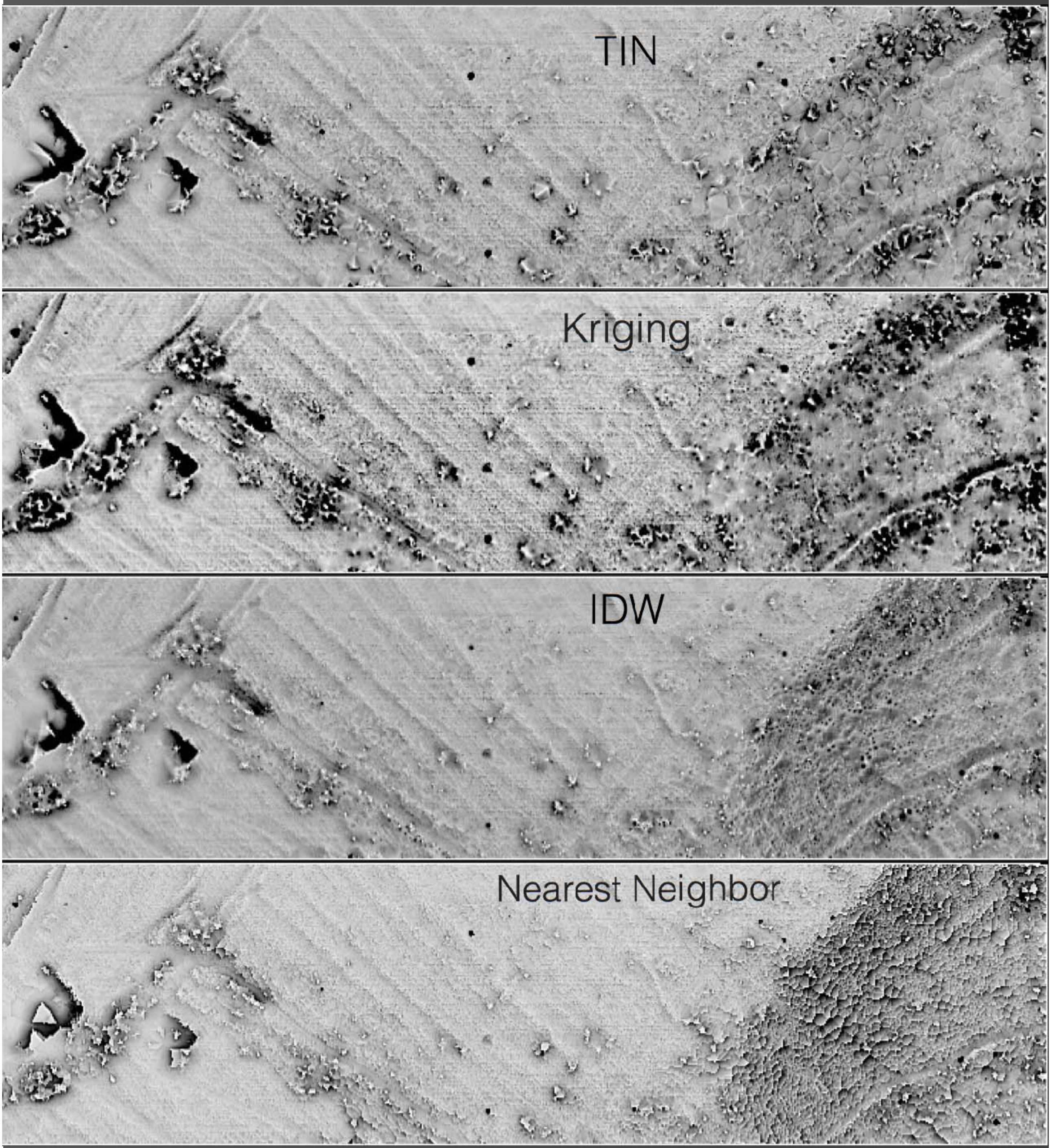
Splines

Interpolation methods

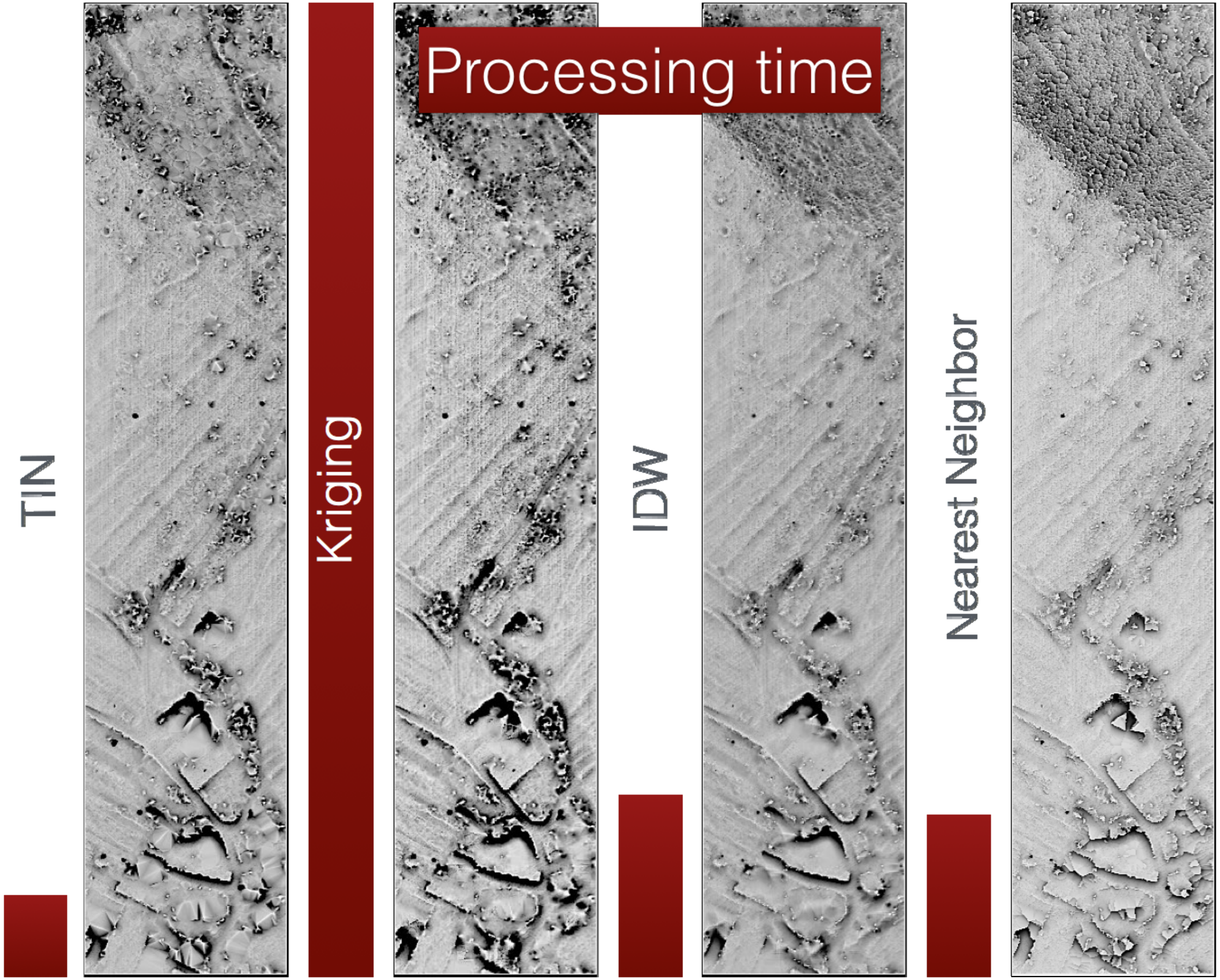


Point density

- 0,3 m
- 0,5 m
- ≦ 1 m

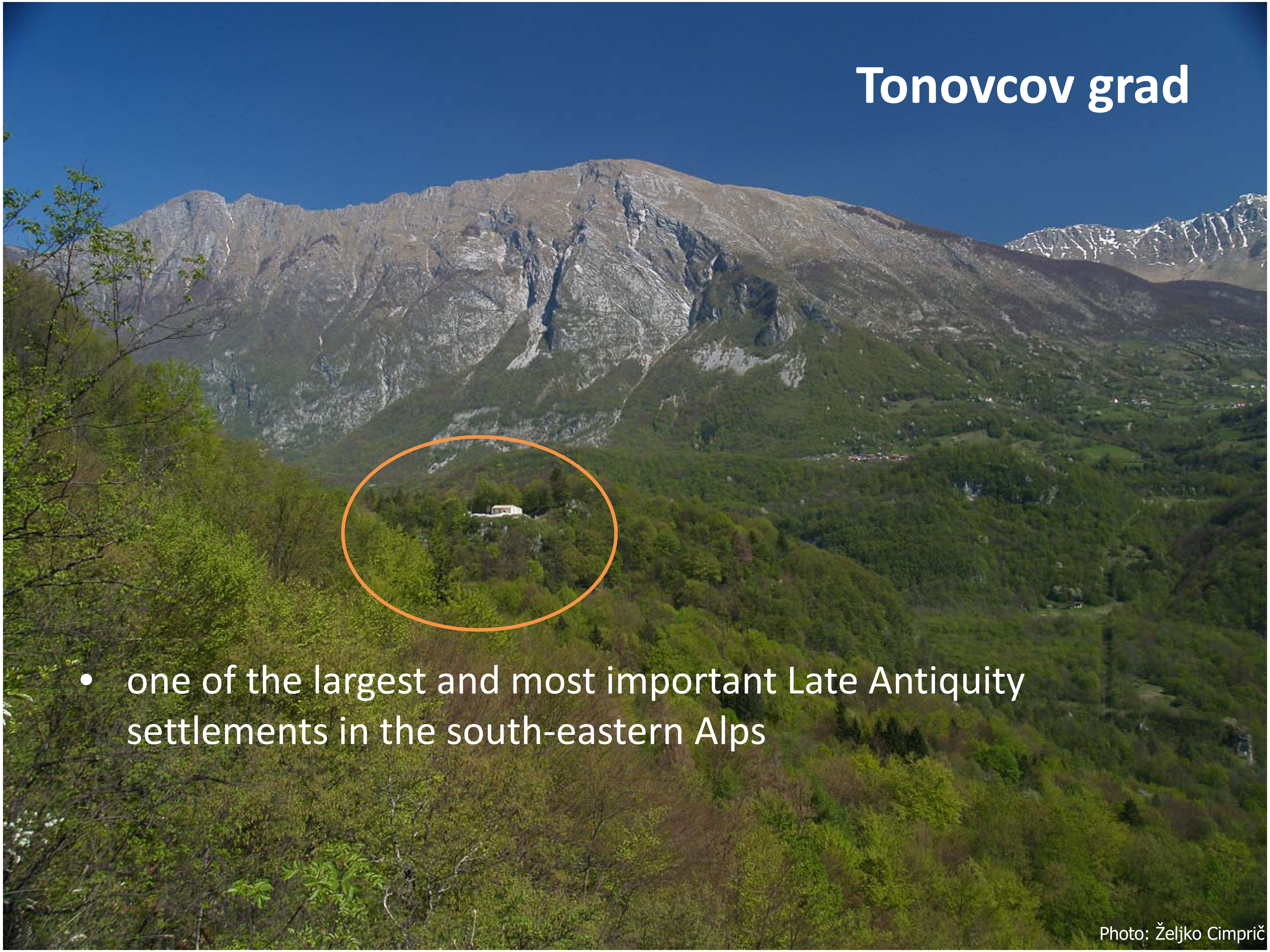

TIN
Kriging
IDW
Nearest Neighbor

Interpolation methods



Raster data visualizations

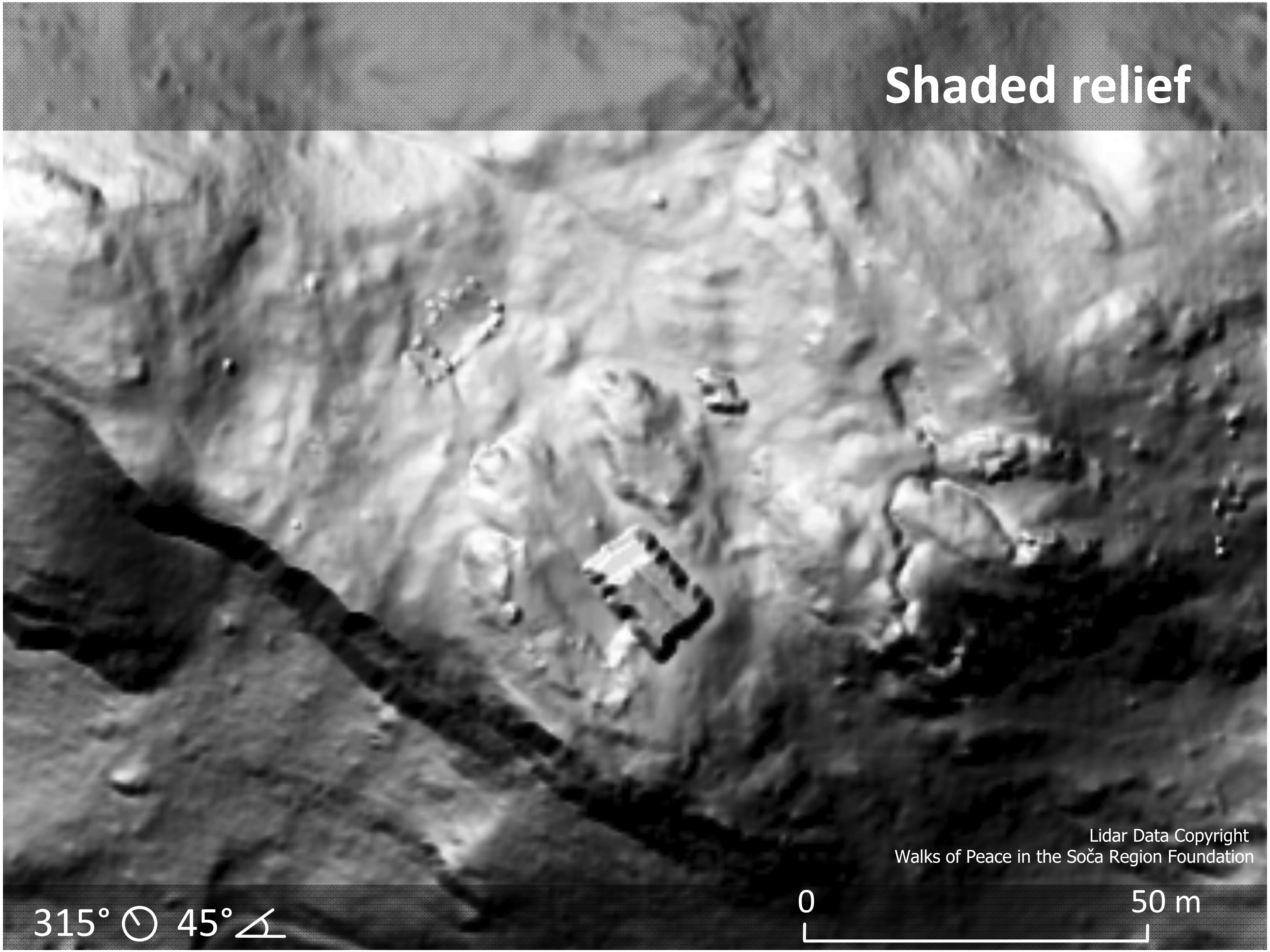
Tonovcov grad



- one of the largest and most important Late Antiquity settlements in the south-eastern Alps

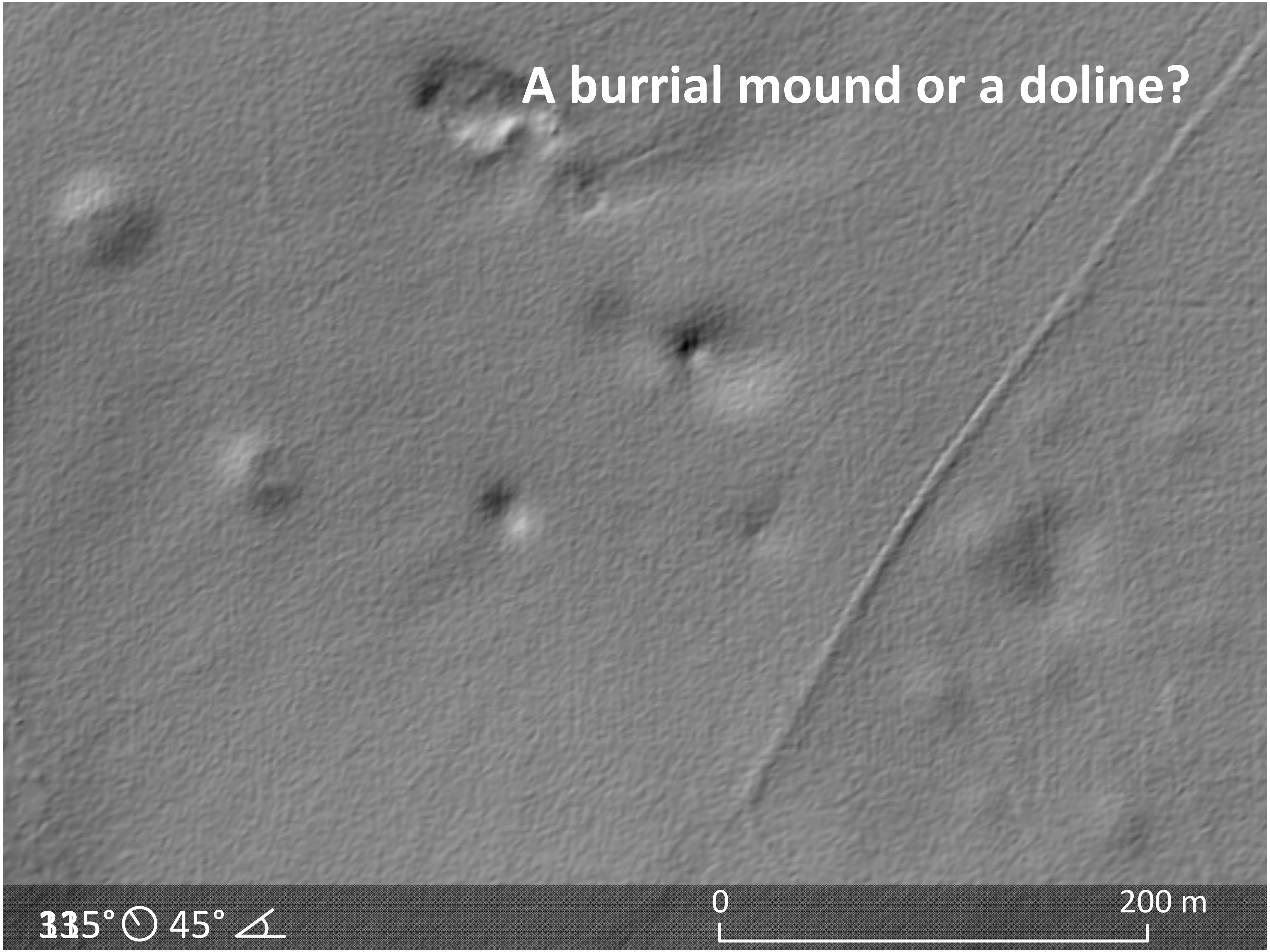
Photo: Željko Cimprič


Orthophoto
0
50 m

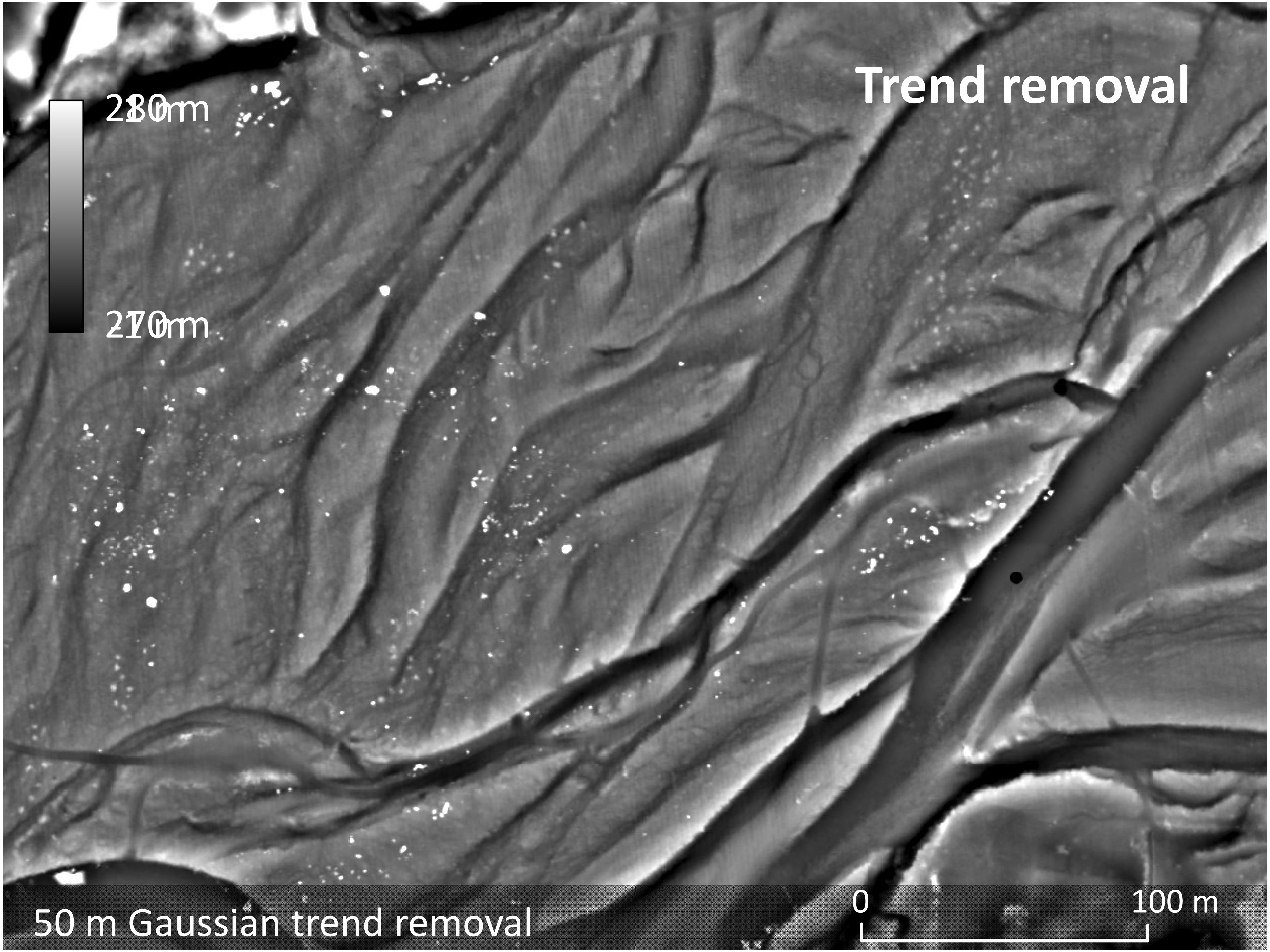


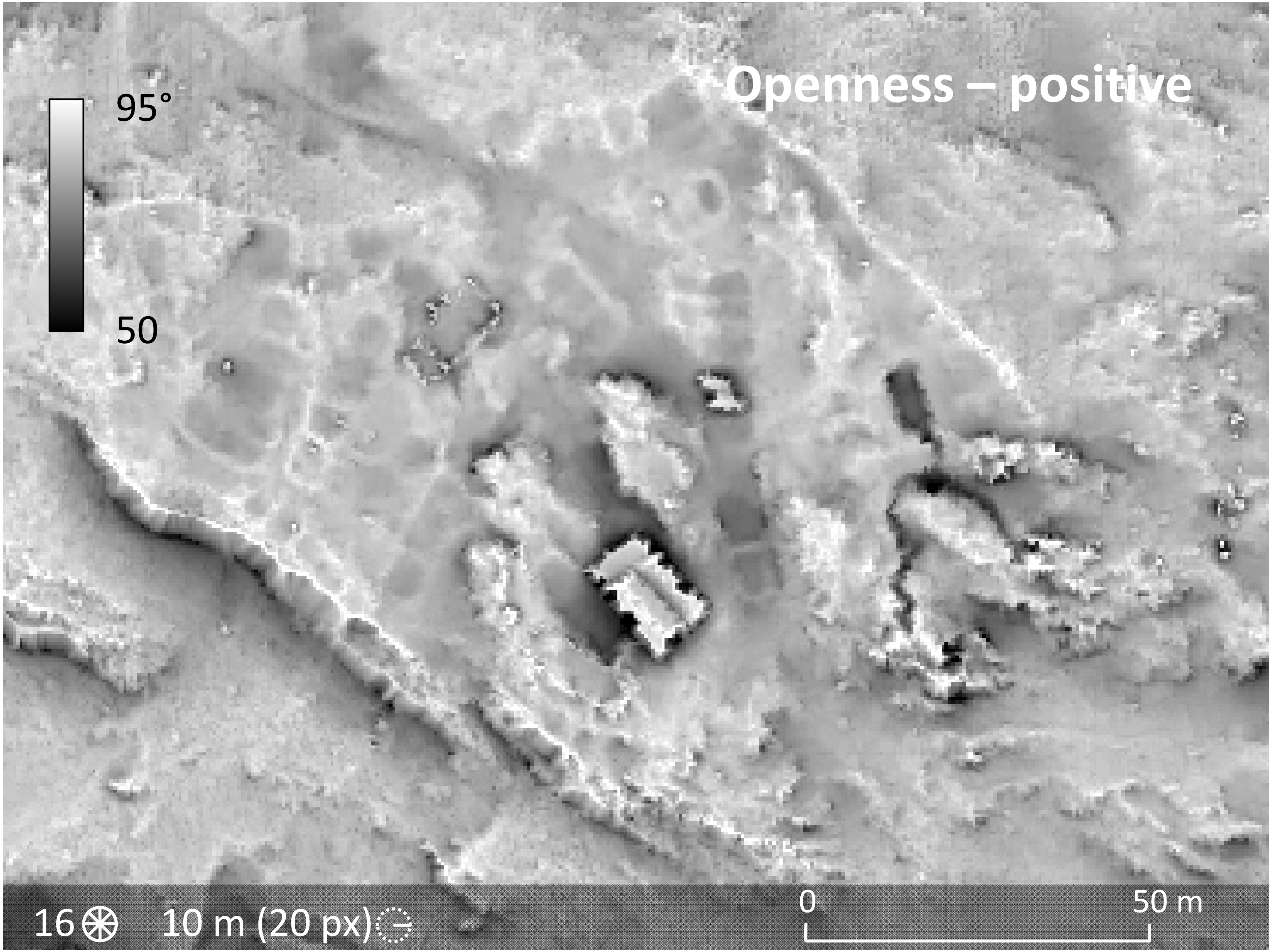
Shaded relief
Lidar Data Copyright
Walks of Peace in the Soča Region Foundation
315° 45°
0
50 m

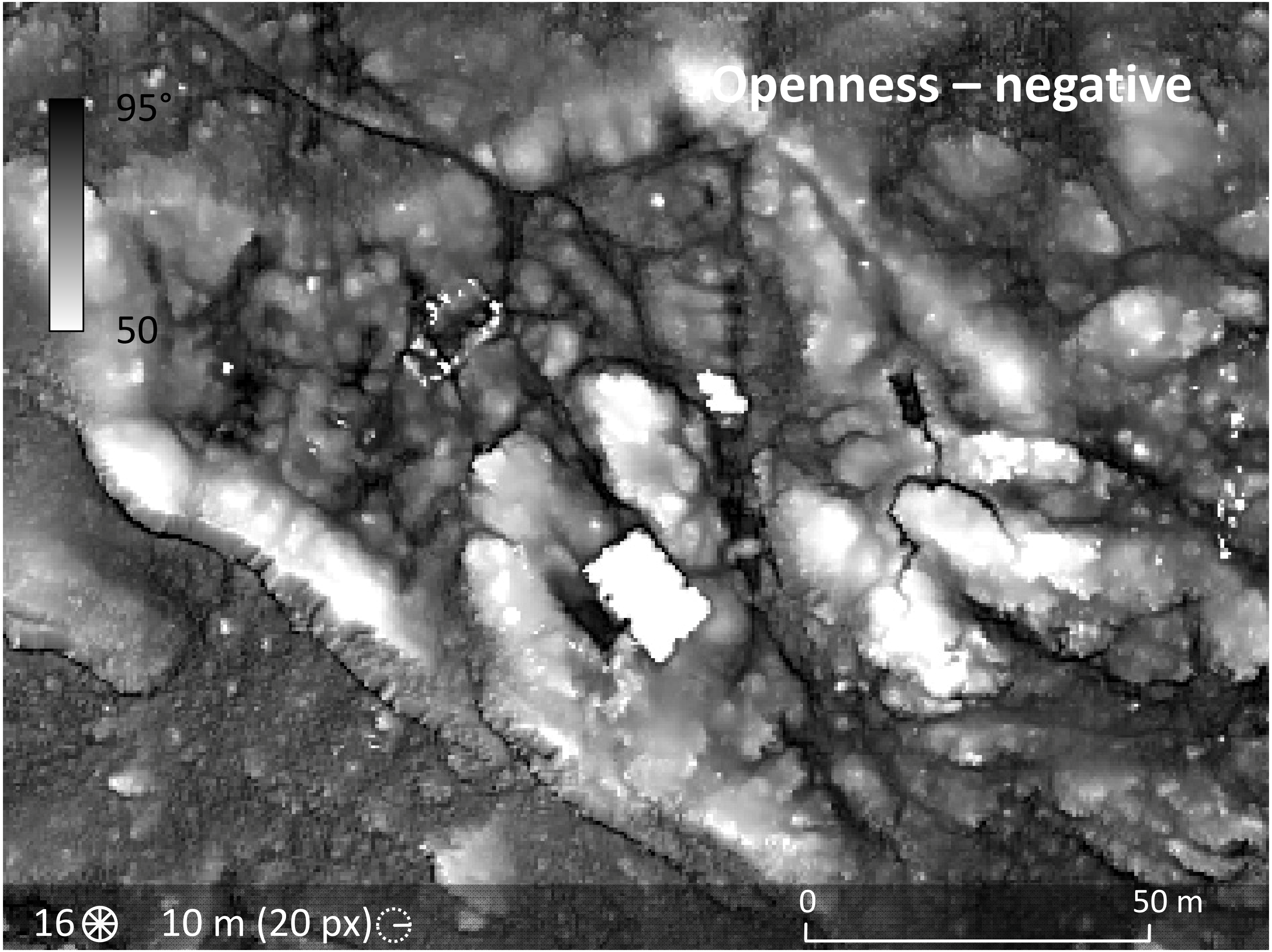
Typical ridge and furrow case study

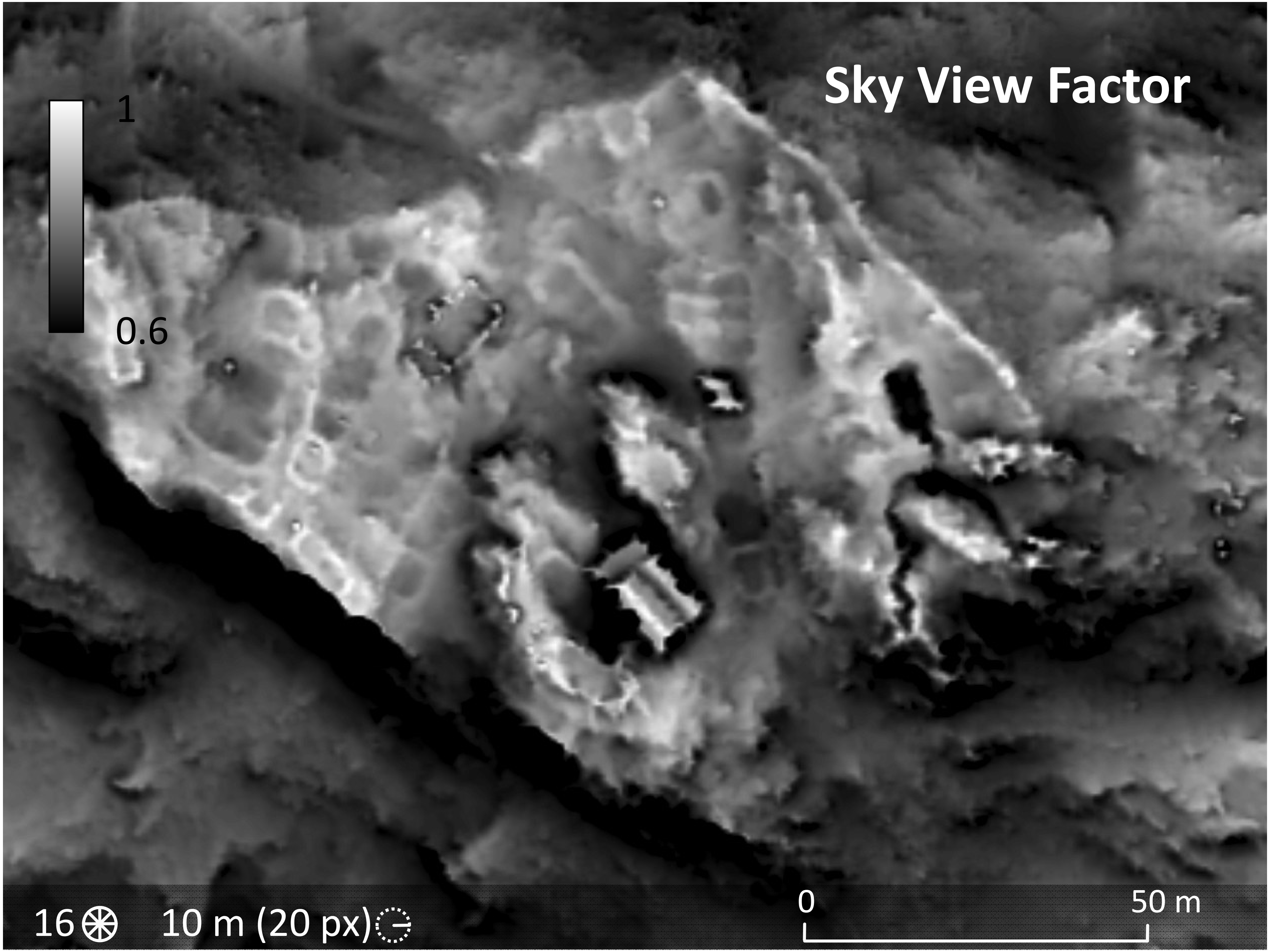


A burrial mound or a doline?
335° 45°
0
200 m


Trend removal
50 m Gaussian trend removal
0
100 m

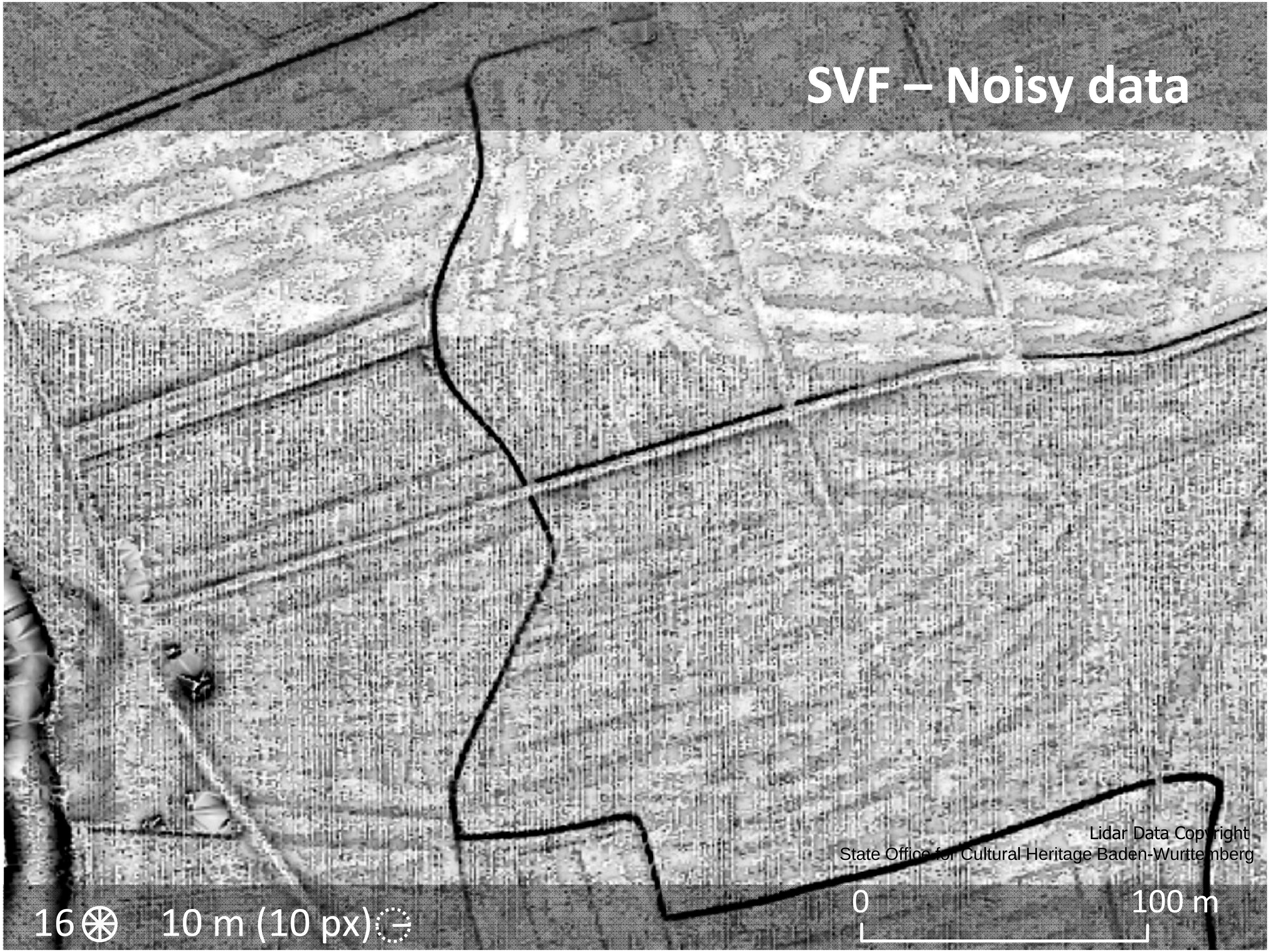

Openness – positive
95°
50
16
10 m (20 px)
0
50 m


Openness – negative
95°
50
16
10 m (20 px)
0
50 m



Sky View Factor
1
0.6
16
10 m (20 px)
0
50 m

SVF – Noisy data



Not only for lidar DEMs

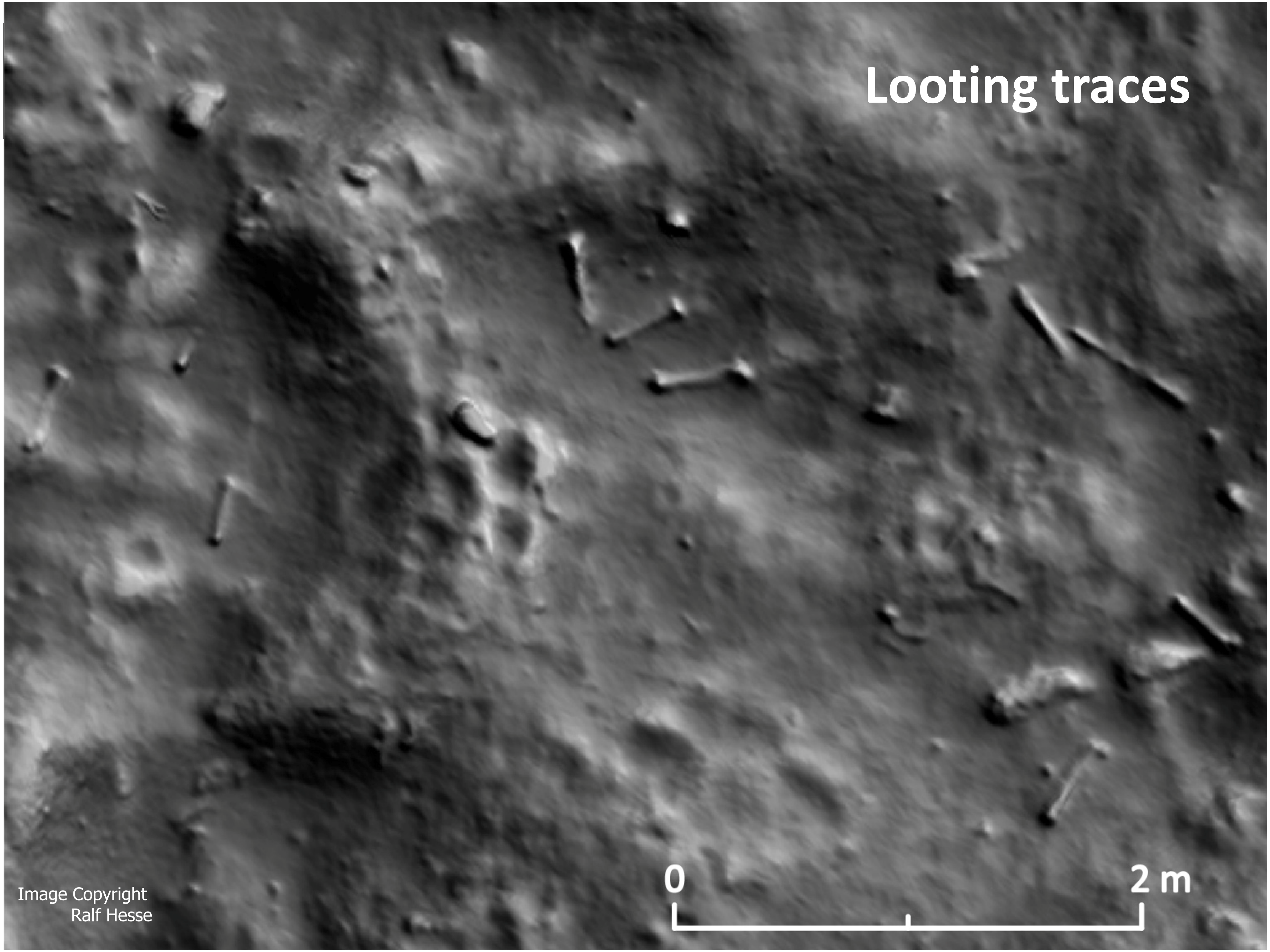
Applicable to all gridded elevation data

- data sources
  - airborne lidar
  - TLS
  - SfM
  - SLS
- scale
  - continental
  - regional
  - local
- observation
  - landscape
  - site
  - object

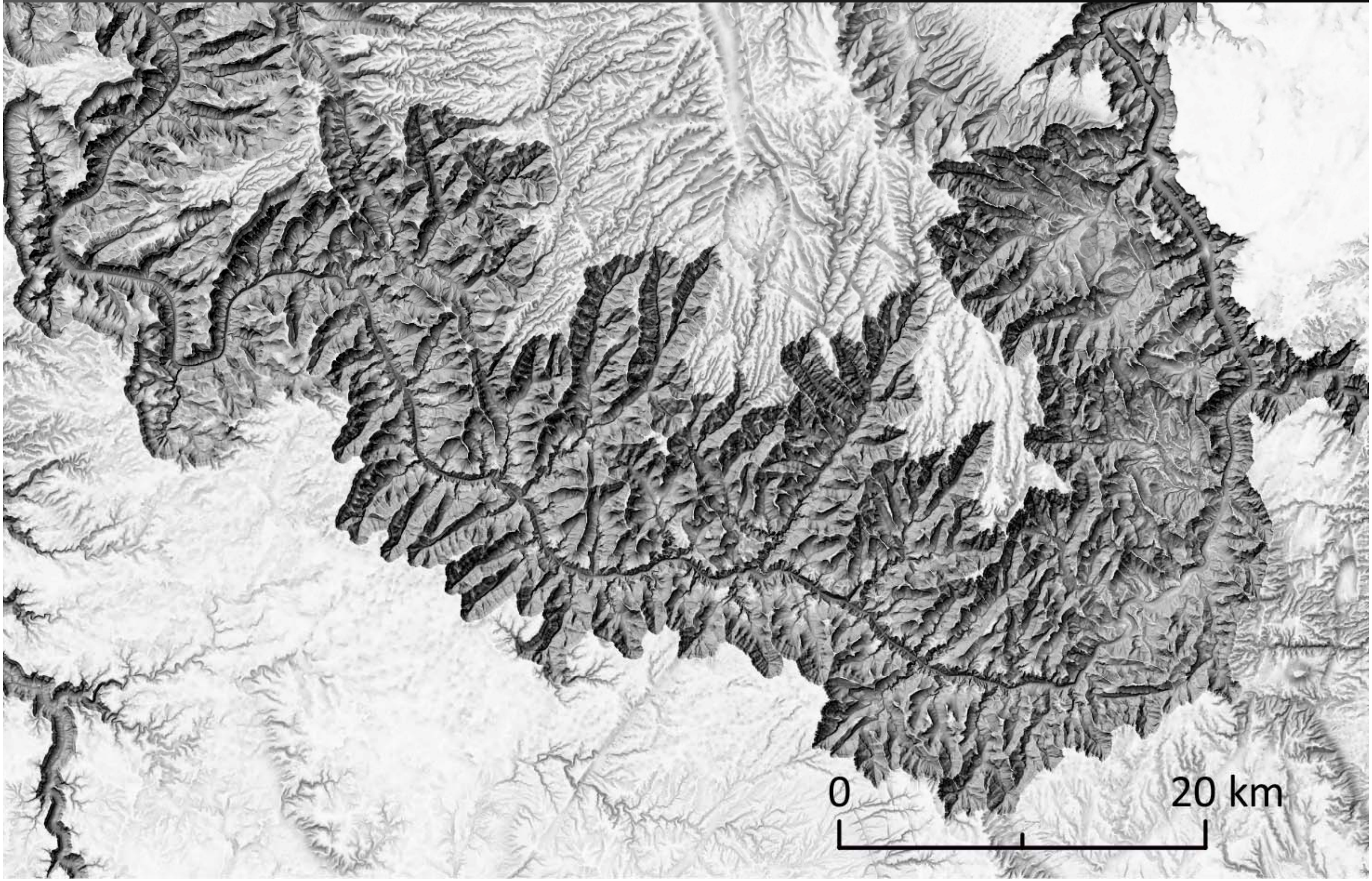
Decorations

- SfM, Newgrange kerbstone

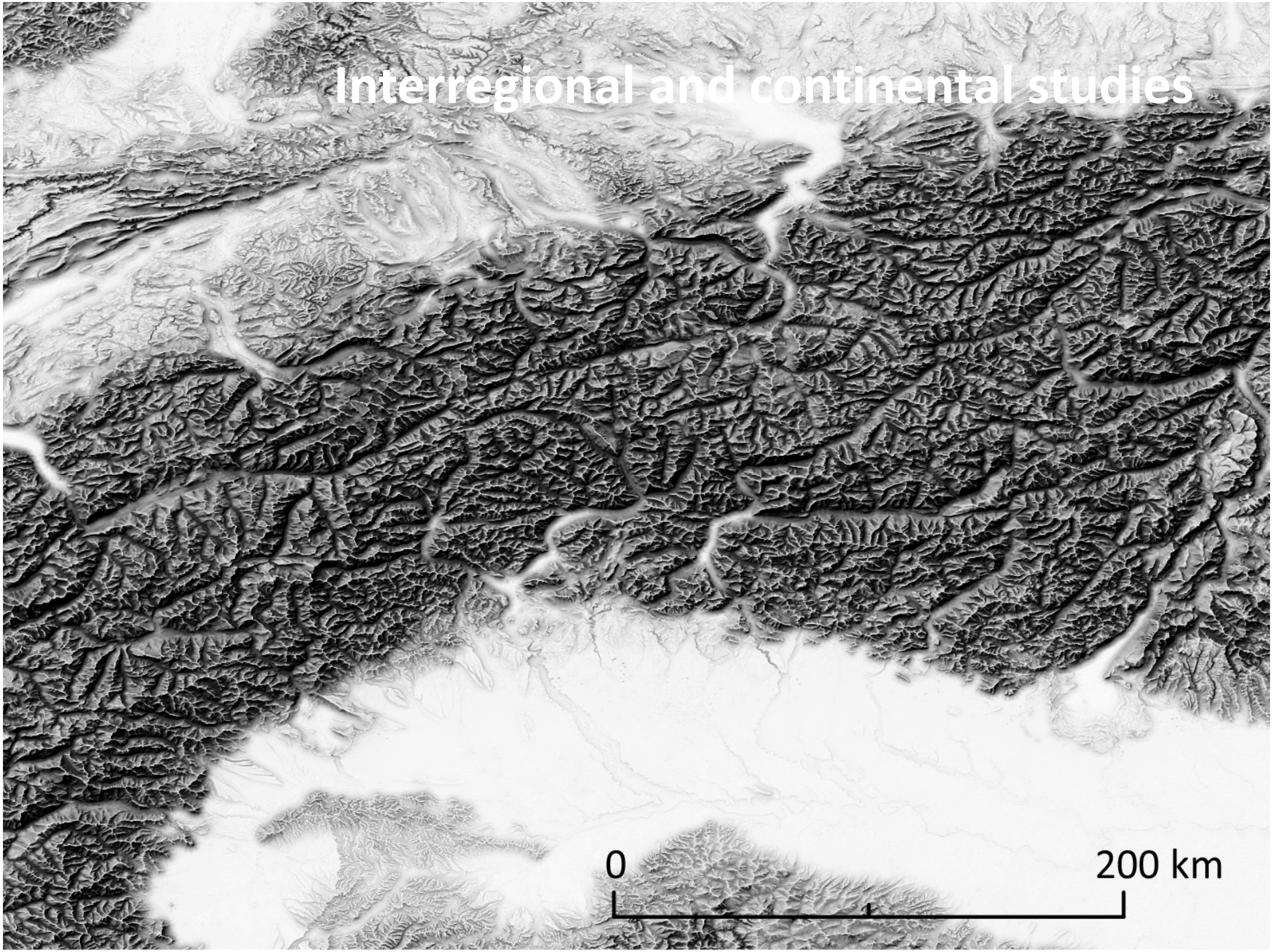
Sky View Factor


Looting traces
0
2 m
Image Copyright
Ralf Hesse

Regional studies



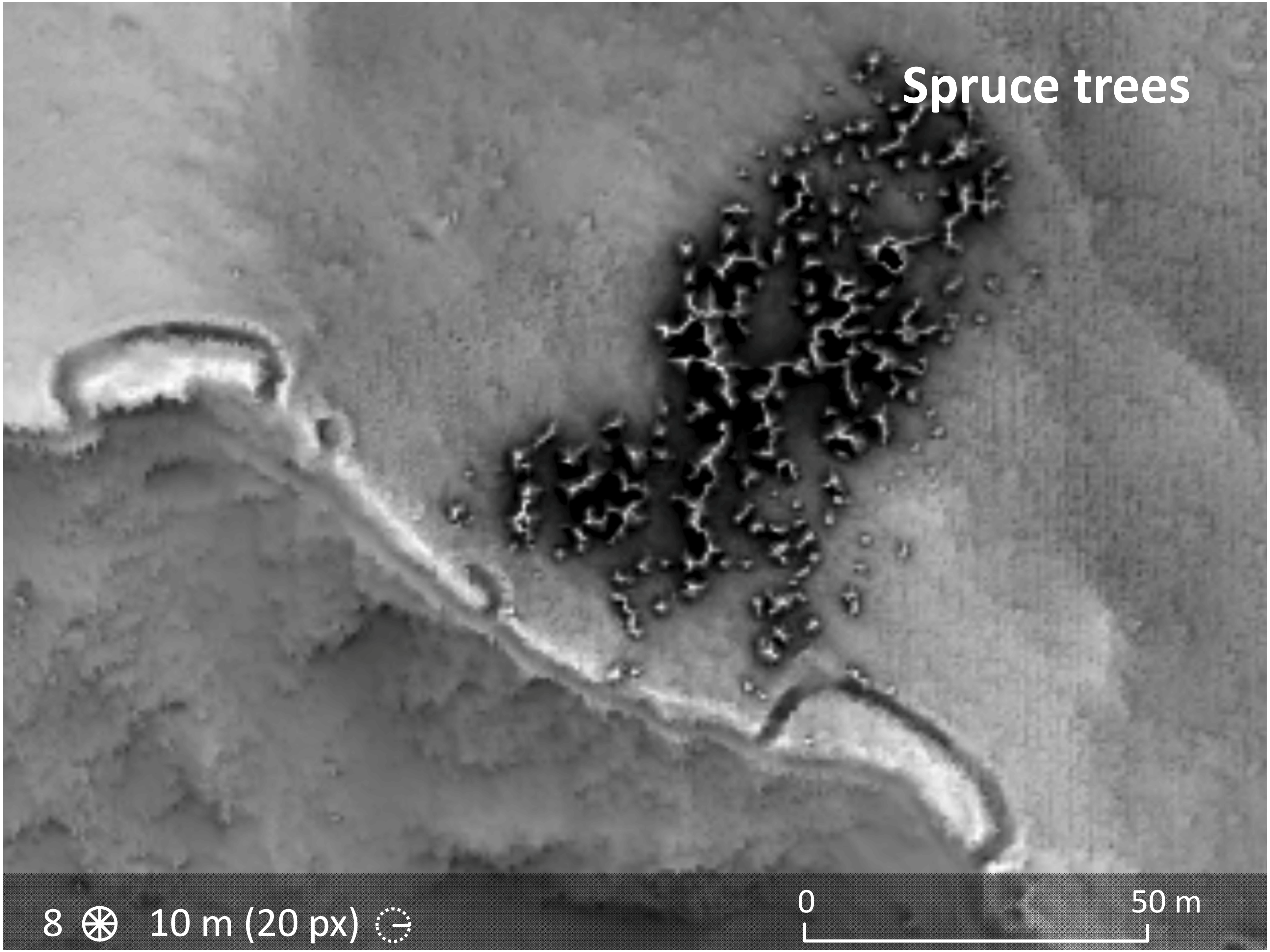
Interregional and continental studies

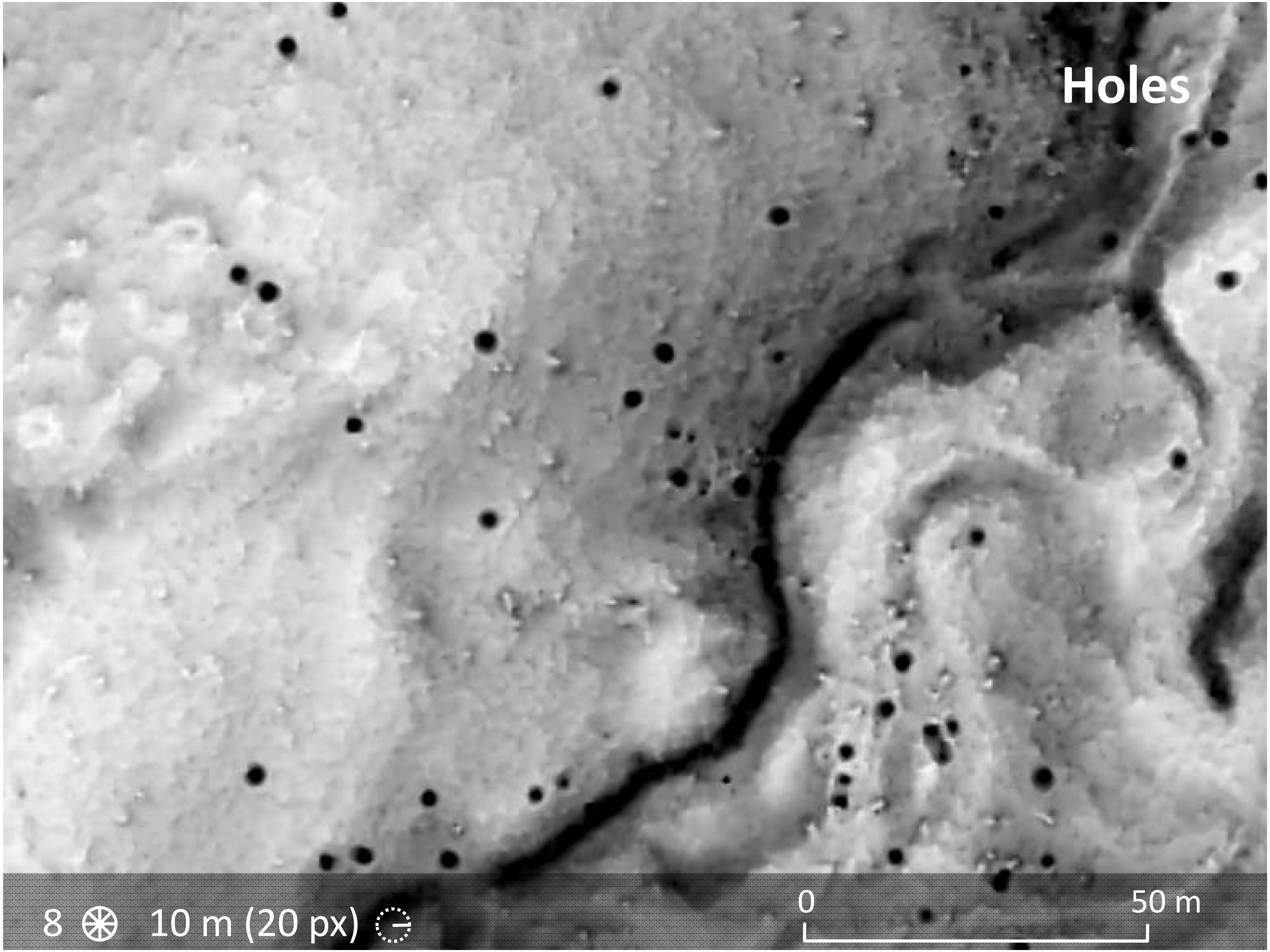


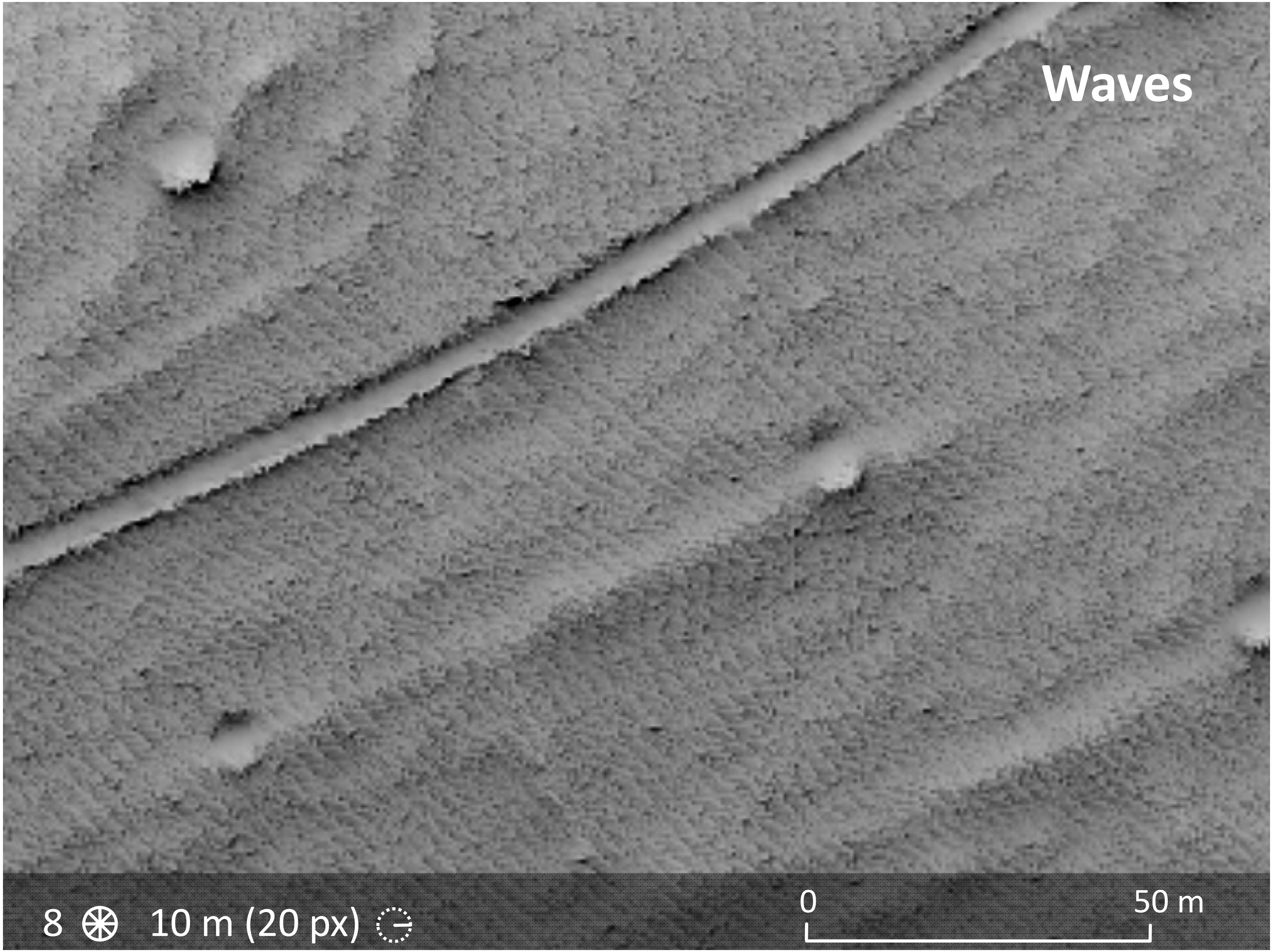
Artefacts

Artefacts in data

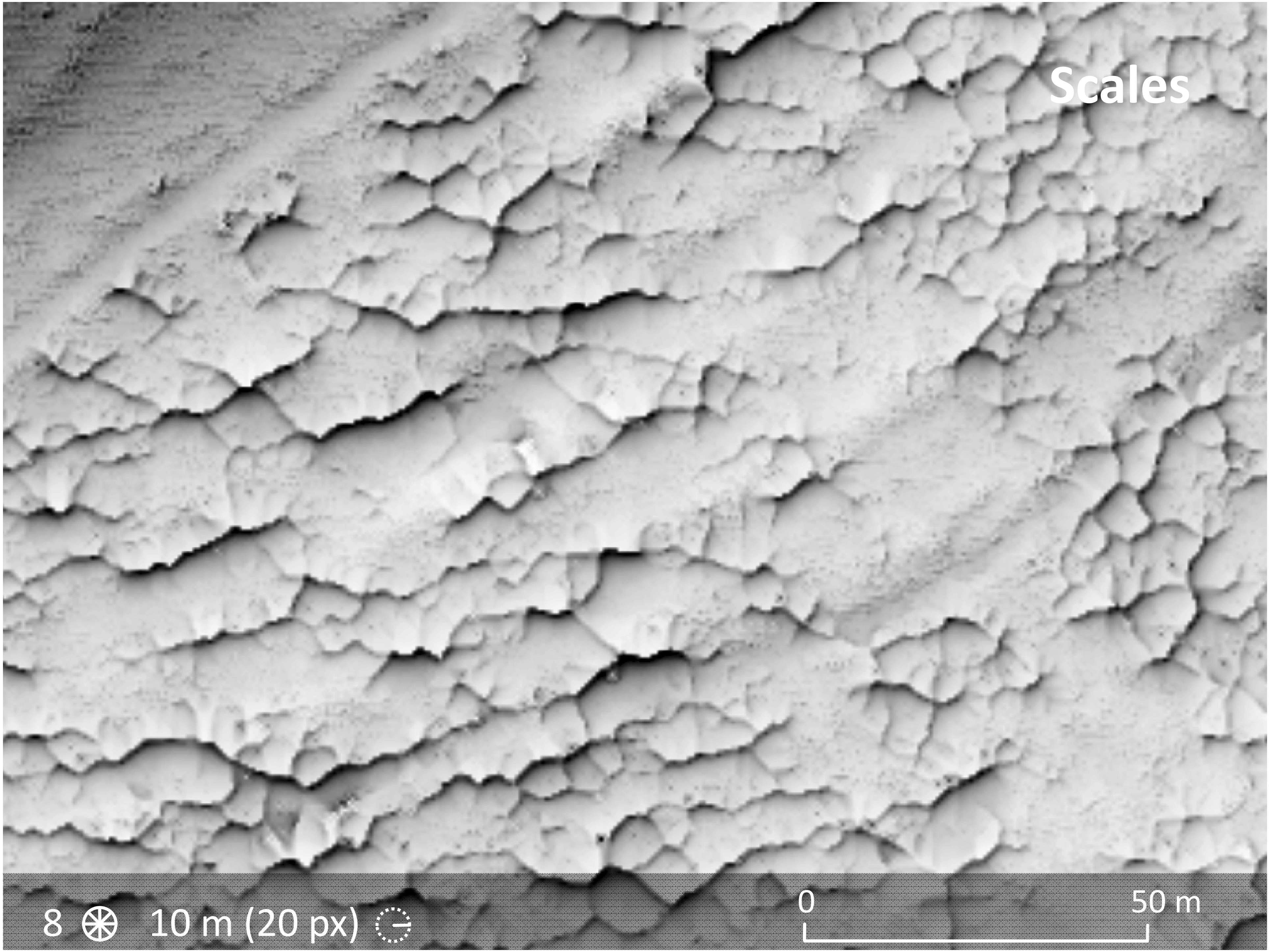
- acquisition
- processing


Spruce trees
8
10 m (20 px)
0
50 m


Holes
8
10 m (20 px)
0
50 m



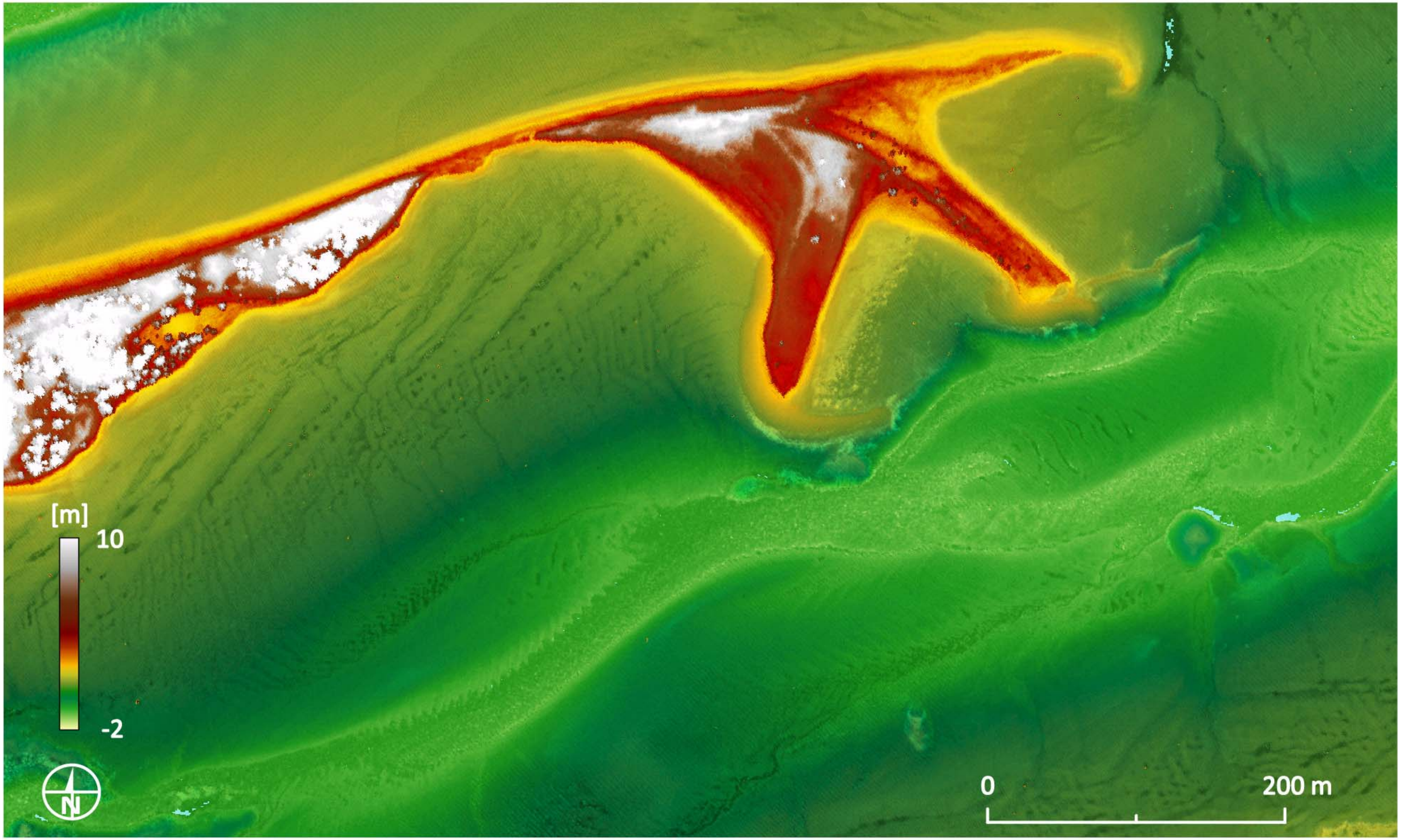
Waves
8
10 m (20 px)
0
50 m


Scales
8
10 m (20 px)
0
50 m


Stars
8
10 m (20 px)
0
50 m

Culbin sands

Culbin sands



2. Rasterizing – making useful images

Practice

2. Rasterizing – making useful images

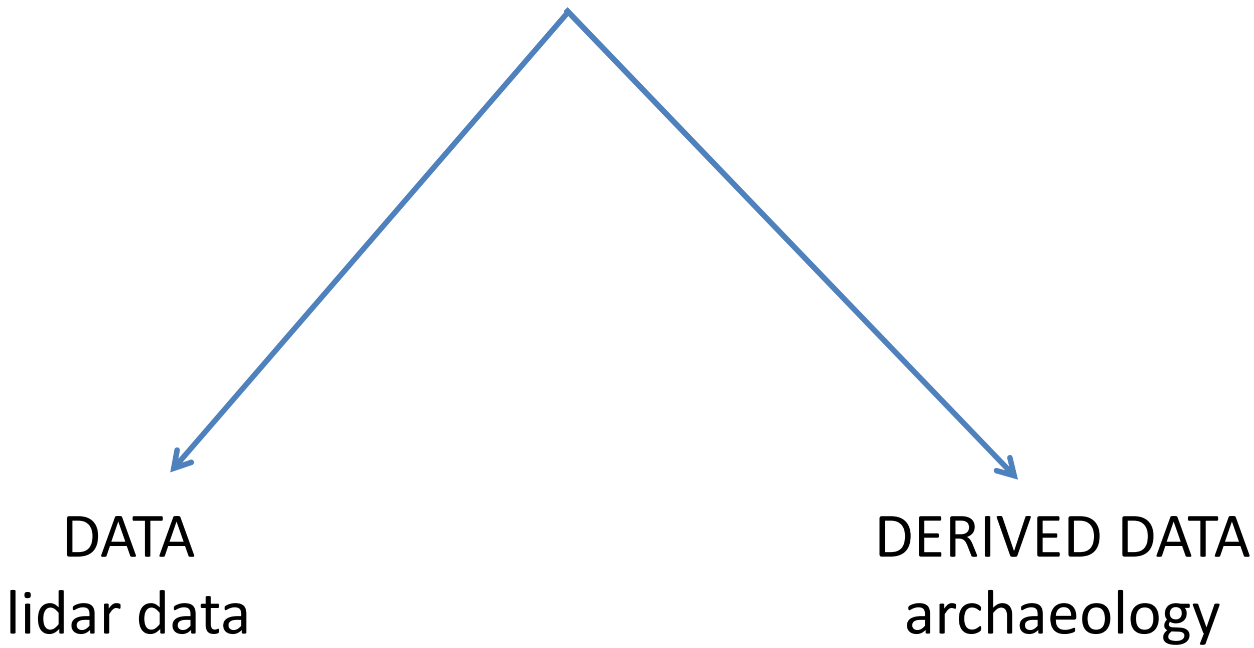
- Goals:
  - rasterizing
  - what a difference a visualization makes
- Files:
  - PointCloud
  - DEM (Sv.Helena)
- Tools:
  - RVT

3. Data Homogenization

presentation & practice

Data homogenization

- *standardization* – make use of standard formats etc.
- *homogenization* – make uniform, integrate, fuse
- *harmonization* – coordinate, fit together, i.e. LOD



Lidar data homogenization

- *standardization* – LAZ, GeoTIFF, TXT
- *homogenization* – fuse disparate data in GIS
- *harmonization* – transparent data archive

Homogenization – multi-temporal data

DTM change detection - why?

- monitoring of archaeological sites and landscapes
    - central heritage management task: “to monitor the condition of cultural monuments”
- record, understand and quantify the impact of natural and anthropogenic processes
- detection of threats

Homogenization – multi-temporal data

Multi-temporal datasets have been almost always:

- acquired with different scanners
- raw data processed with different software (versions)
- acquired with different nominal density etc.

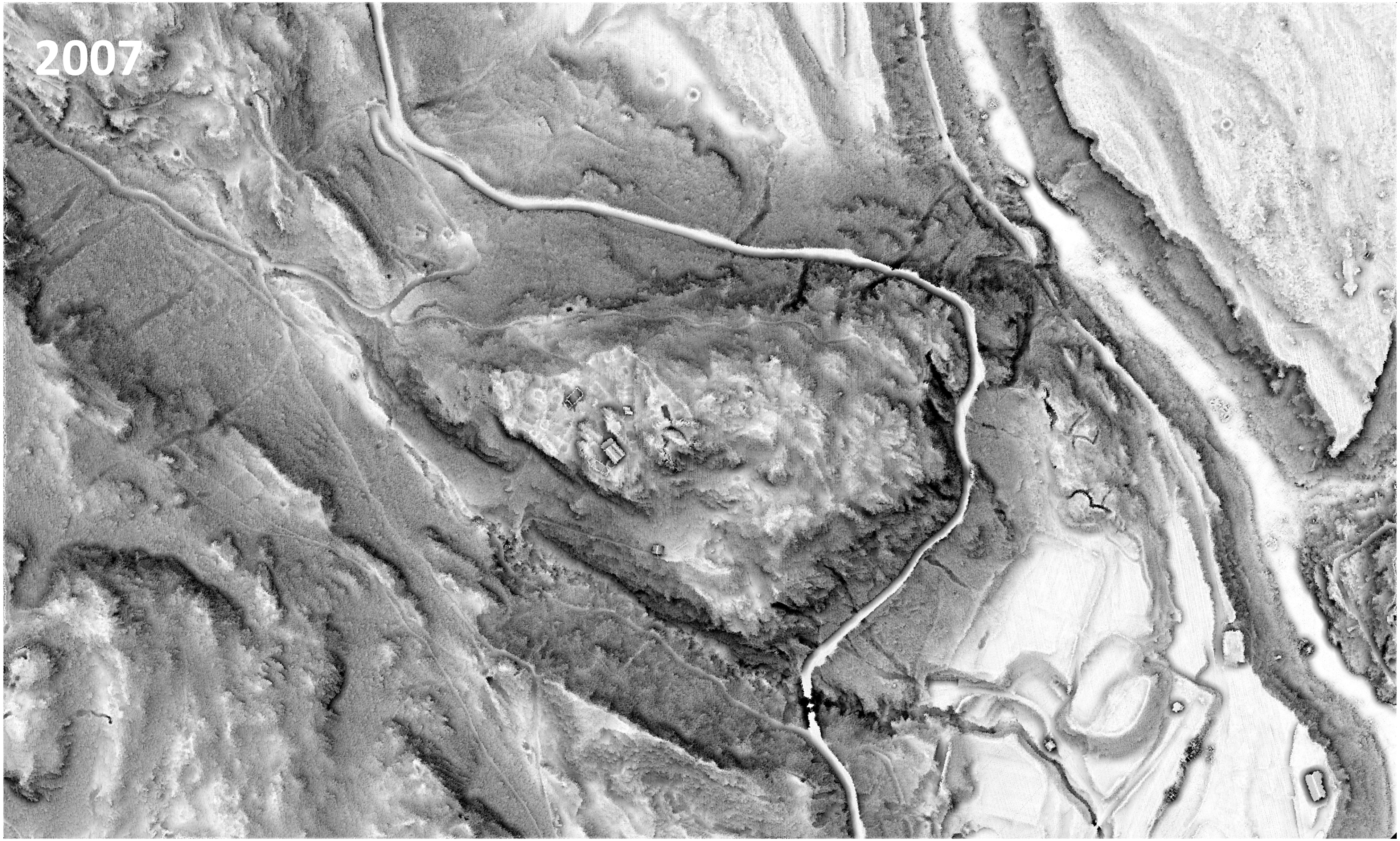
Homogenization – multi-temporal data

- direct point cloud comparison rarely makes sense
- re-process as much as you can (usually from the point cloud classification onwards) using exactly the same parameters for both (all) datasets
- compare DEM's/DSM's
  - simple in principle, demanding in practice

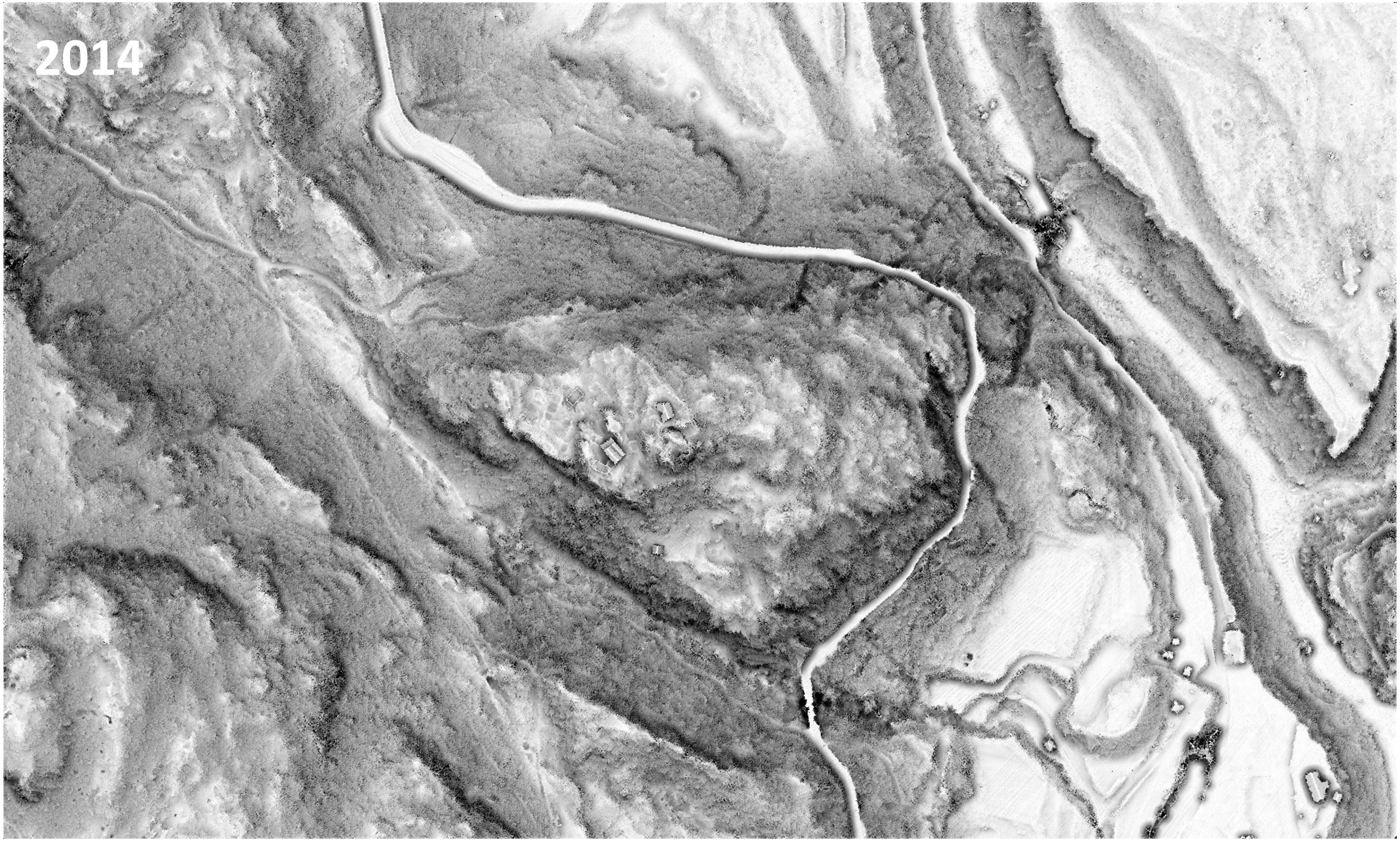
Homogenization – multi-temporal data

- subtract younger from older, obviously
- for multi-temporal - 3 or more time slices - many additional techniques are available (c.f. methods for multi-spectral satellite imagery)

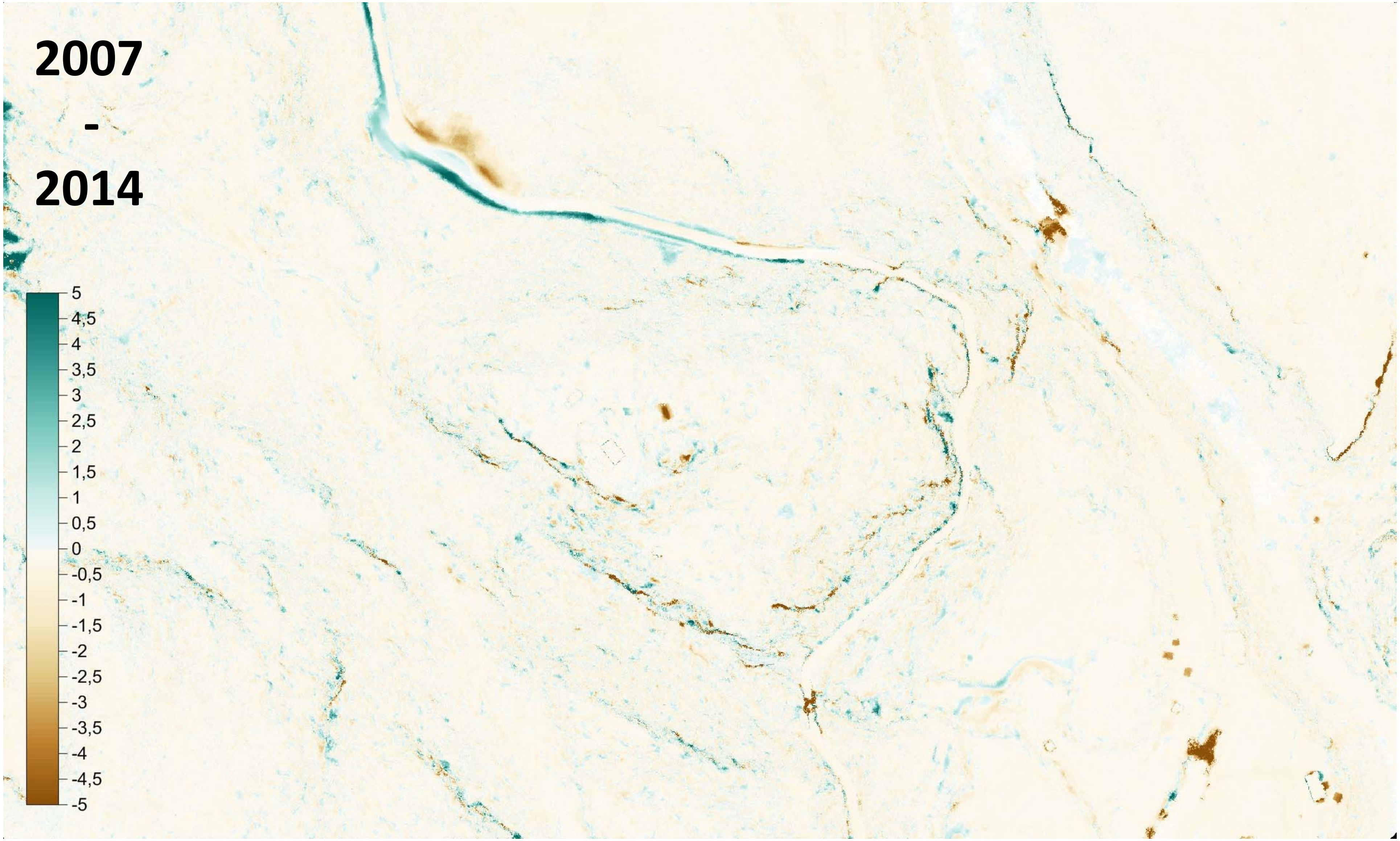
Homogenization – multi-temporal data



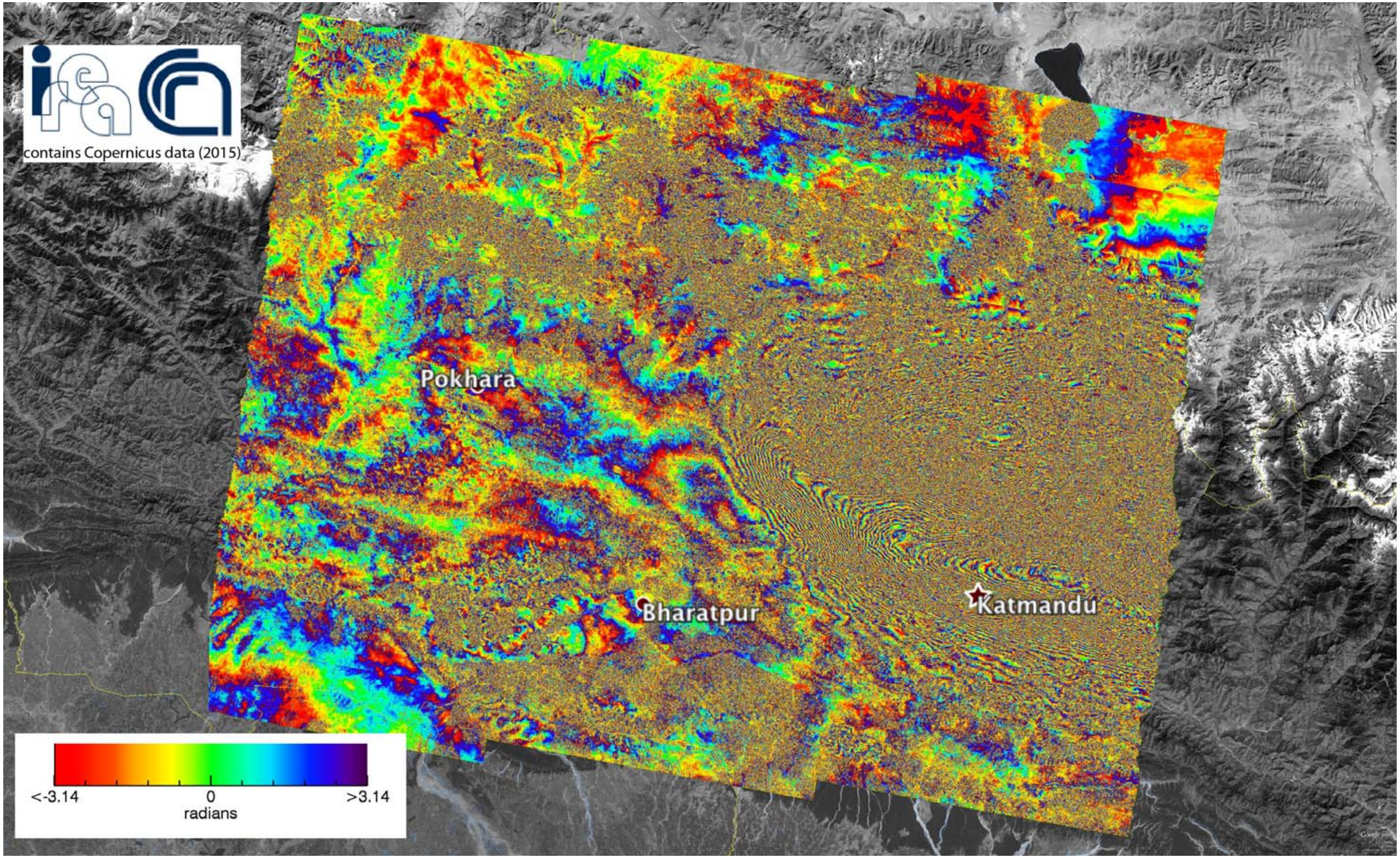
Homogenization – multi-temporal data



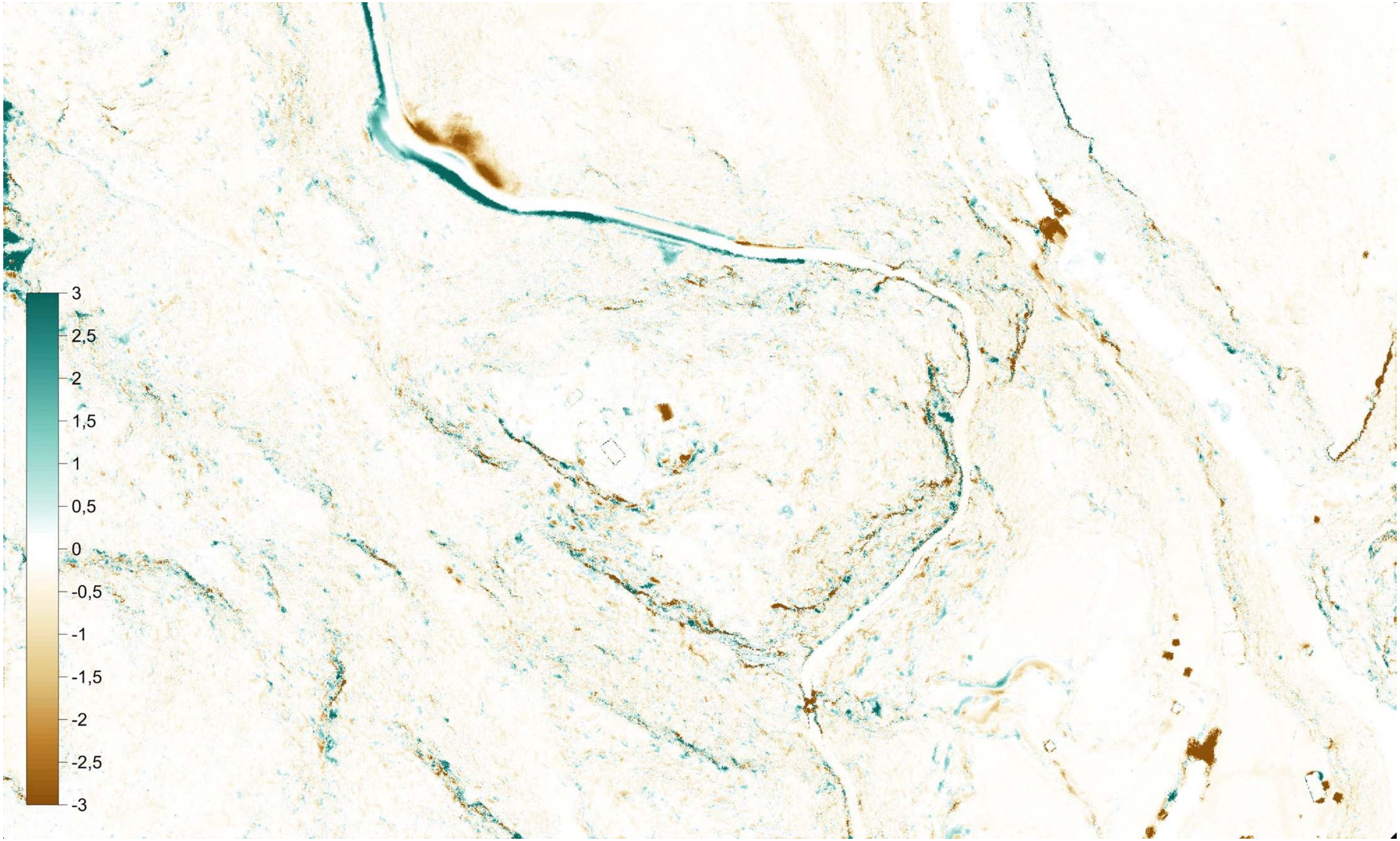
Homogenization – multi-temporal data



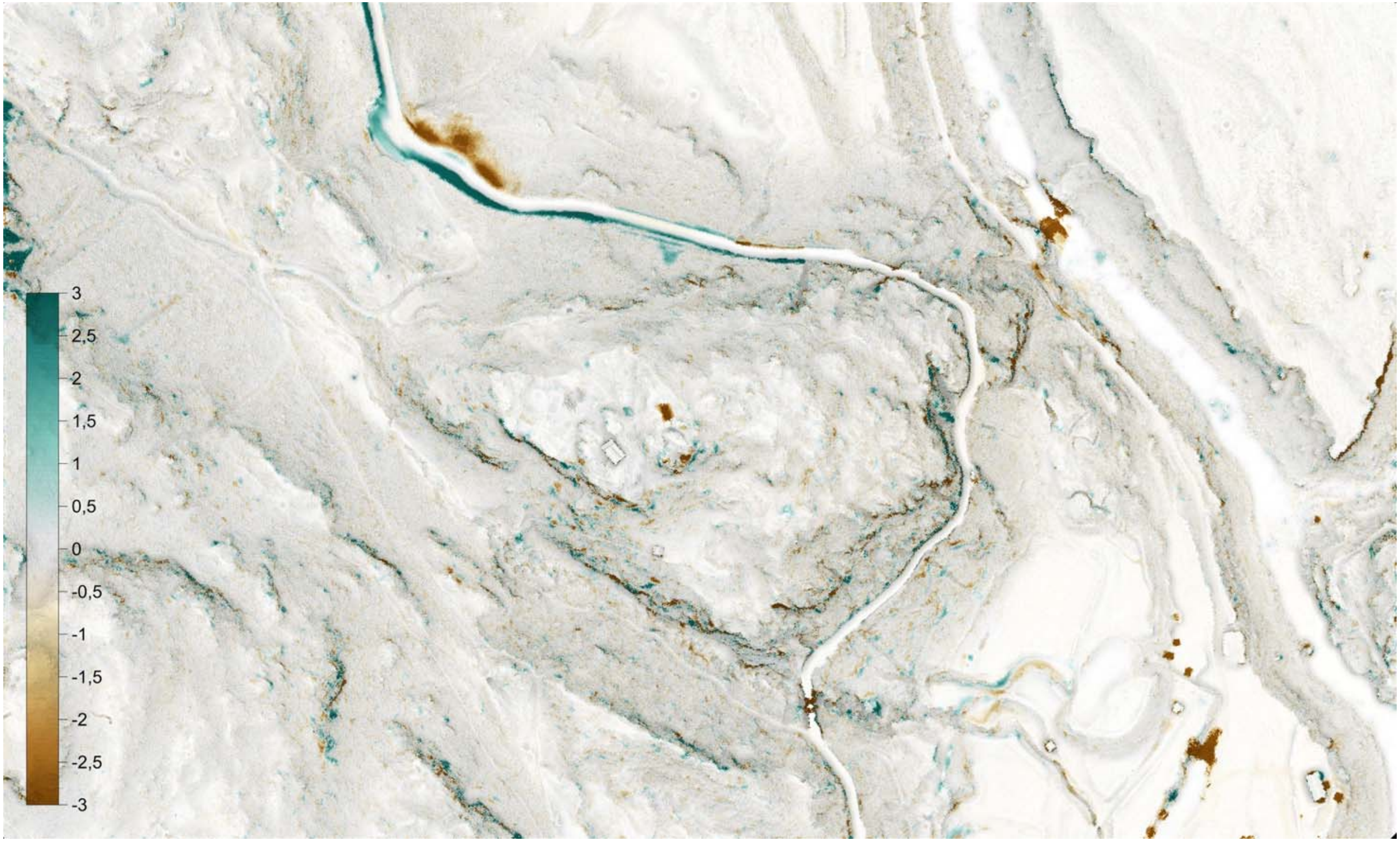
Homogenization – Multi-temporal data



Homogenization – Multi-temporal data



Homogenization – Multi-temporal data



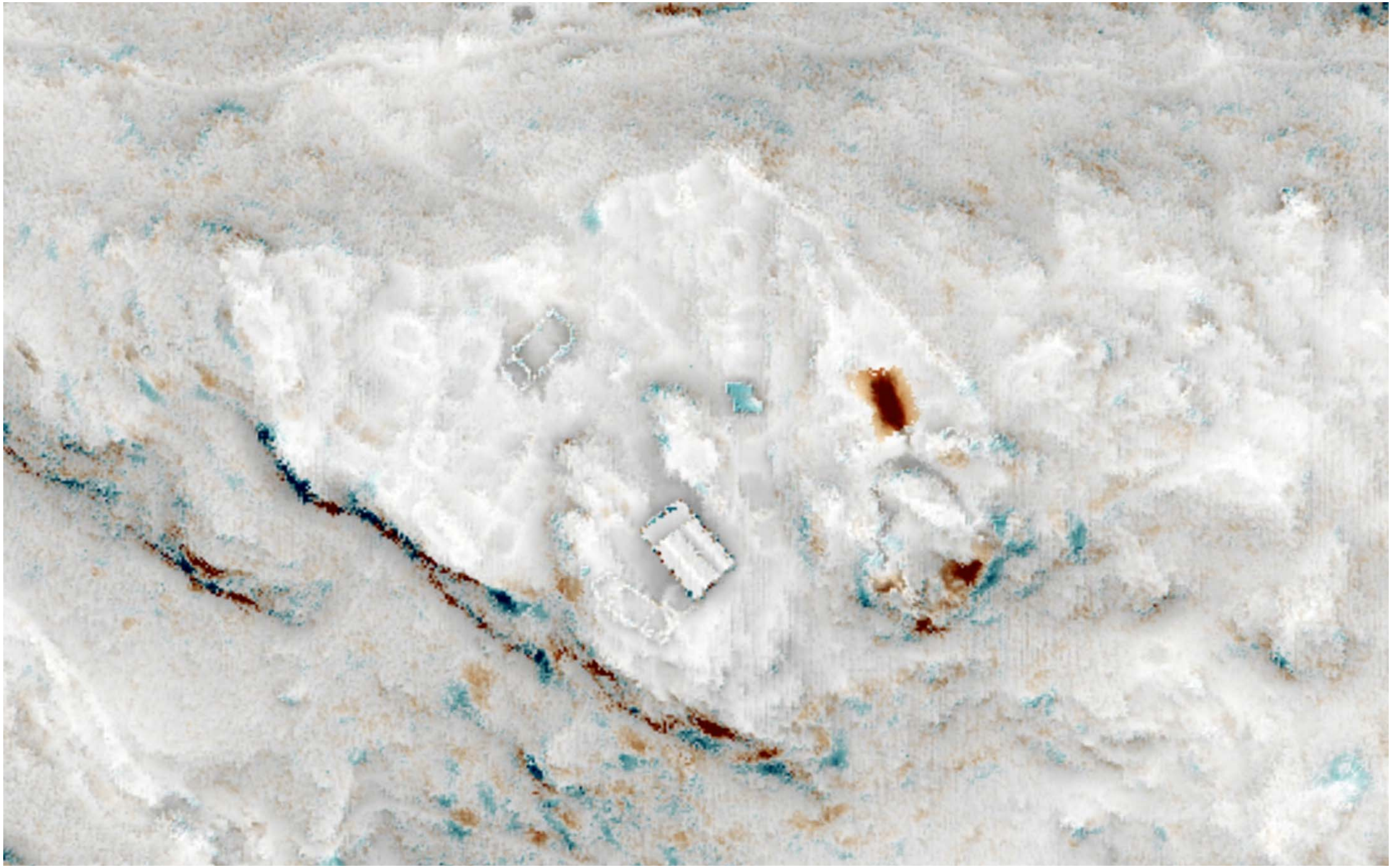
3. Data Homogenization

practice

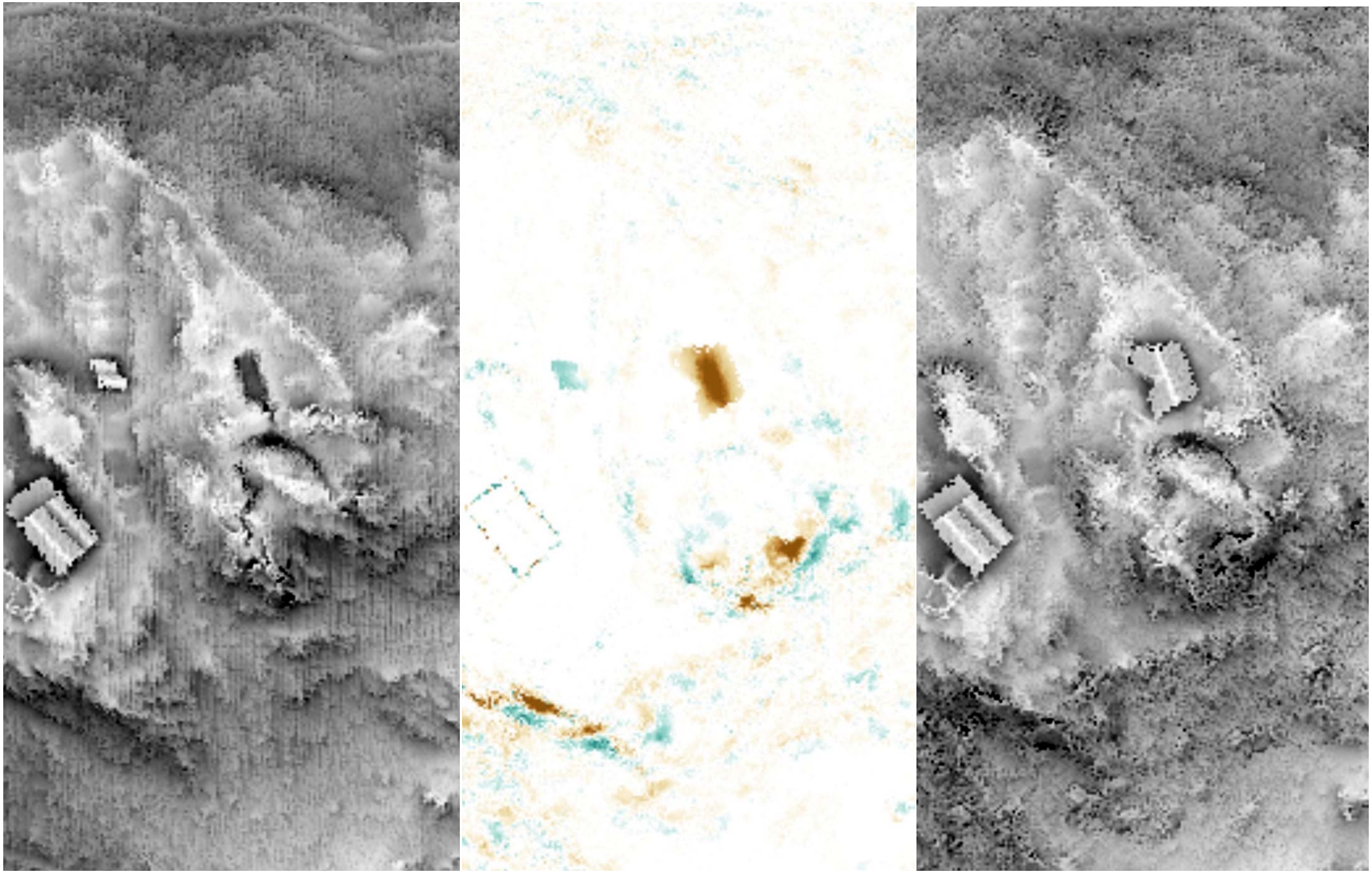
Practice – multi-temporal data

- Goal:
  - compare disparate data
- Files:
  - dem2007.tiff
  - dem2014.dem
  - svf2007.tiff
  - svf2014.tiff
  - Diff_Map.jpg (use in case of emergency)
- Task:
  - create and visualize the difference map
  - find information meaningful for archaeology

Homogenization – Multi-temporal data



Homogenization – multi-temporal data



Homogenization – multi-temporal data

Homogenization – Multi-temporal data

Homogenization – multi-temporal data

Homogenization – Multi-temporal data

- More often than not there is no archaeologically meaningful information in the comparison of the multi-temporal ALS derived DEM
- No. 1 reason in practice : lack of precision
- Solution: harmonization (raw data processing), (much) better data
- Take home message: comparing 2 lidar (ALS) datasets will rarely yield archaeologically meaningful results unless acquired specifically for this task or in cases of big changes that are observable even without the comparison

4. Data Archiving

presentation & practice

Archiving Data – What to keep?

- Original data (the data you received, e.g. classified point cloud) transcribed in most sensible format (e.g. LAZ)
- Final version(s) of point cloud data in - LAZ
- Final version(s) of raster data - GeoTIFF
- Visualizations that you find most useful (usually 2 or 3) + HS - GeoTIFF
- Optionally (but highly recommendable): ALL of the final maps – TIFF, PDF, …
- METADATA - TXT

Archiving Data – Practical Advice

- Be mindful of the back-up's and whatnot
- The best storage is still very cheap compared to data acquisition and processing – MORE is MORE
- Follow the usual good practices for digital archiving (e.g. http://guides.archaeologydataservice.ac.uk)
- SHARE, SHARE, SHARE! If you didn't start working on that article in 2 years, chances are you never will. But someone else might! Think of it this way: share the data and you might get the co-author credit; hoard the data and you get diddly-squat!

5. Final words

Final words

- Know your data!
- Record and publish metadata.
- Archive your data. Data dump IS NOT data archive.
- Share your data!

Final words

- Know your data!
- Record and publish metadata.
- Archive your data. Data dump IS NOT data archive.
- Share your data!
- If you liked this workshop buy us a beer the next time we meet☺