Big Data And LiDAR: Developing A Framework To Process Country Wide LiDAR Point Cloud Data Sets Using Parallelization On A Computer Cluster

Duc Thang Nguyen
Supervisor: Nikolai Knapp
FIT 2015
20th April 2017

Outline

- Introduction
- Research Target
- Study Area
- Material
- Methodology
- Result
- Discussion

Introduction

Research Target

Study Area

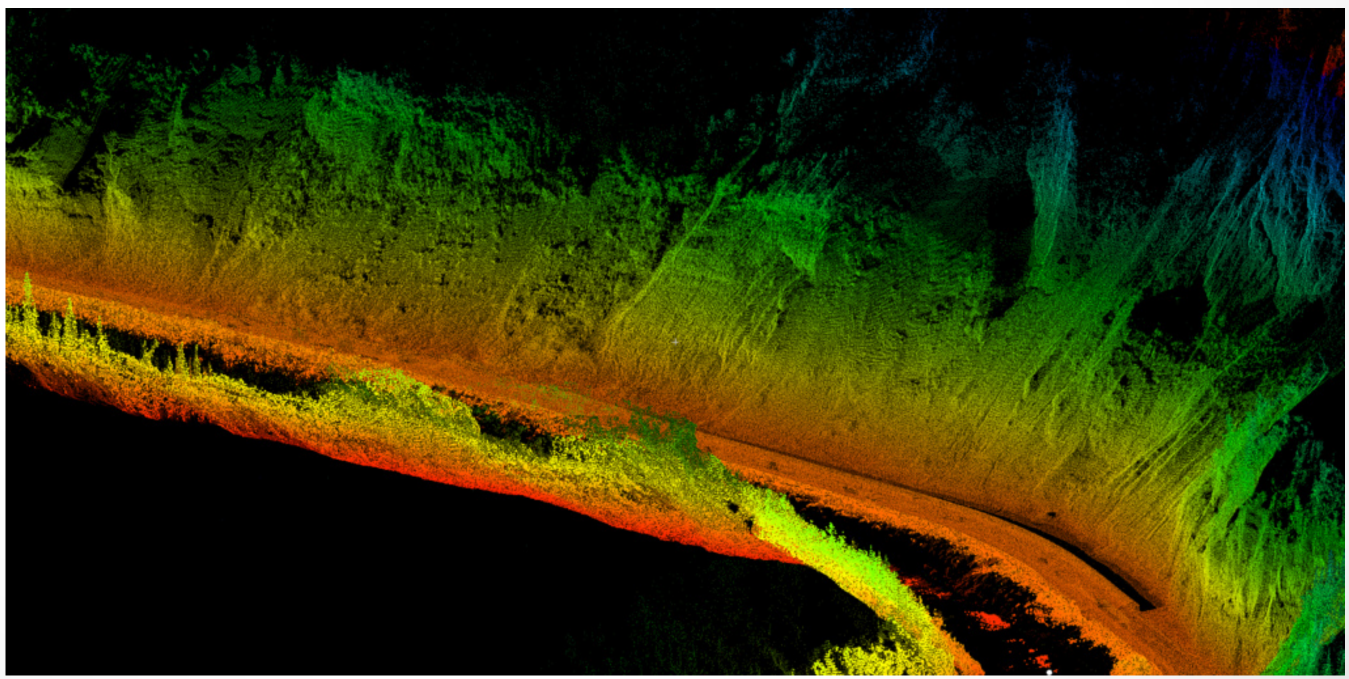
Material

Methodology

Result

Discussion

- Processing of multiple data for counties or statewide applications is challenging from modern geospatial software solutions (ArcGIS, ENVI, SAGA…)
- The data volume might reach tens or even bigger of Terabytes which stresses the limit of the storage of a single computer.
- Computation of Big data is another challenge for processing massive.



Source: Oregon State University
https://www.flickr.com/photos/oregonstateuniversity/8554891847

Introduction

Research Target

Study Area

Material

Methodology

Result

Discussion

1, How can we process LiDAR in a plenty of processes?

Developing a workflow processing LiDAR combines LAStools software

2, What are the final products of these processing?

Digital Surface Models (DSM),Digital Terrain Model (DTM) and normalized Digital Surface Models (nDSM).

3, How can we reduce the processing time of LiDAR data in the large scale?

Apply parallelization function in R software to efficiently handle big LiDAR datasets could be currently an optimum method.

FINAL GOAL:

To produce a generalized framework that can be applied for **any LiDAR** dataset format in the scale of state and country

Introduction

Research Target

Study Area

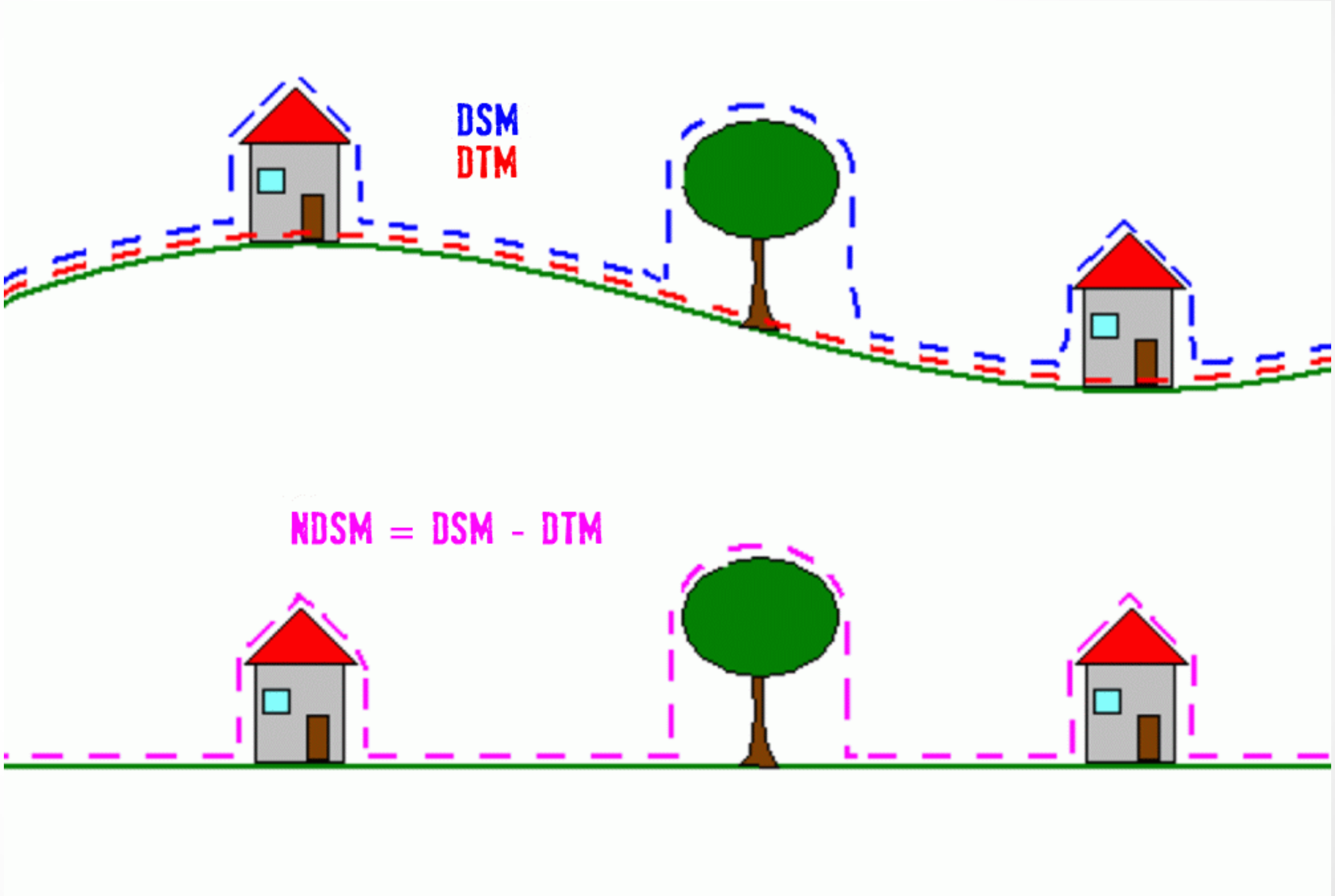
Material

Methodology

Result

Discussion

DSM, DTM and nDSM ?



Source: http://www.stadtentwicklung.berlin.de/umwelt/umweltatlas/ed610_03.htm

DSM, DTM and nDSM ?

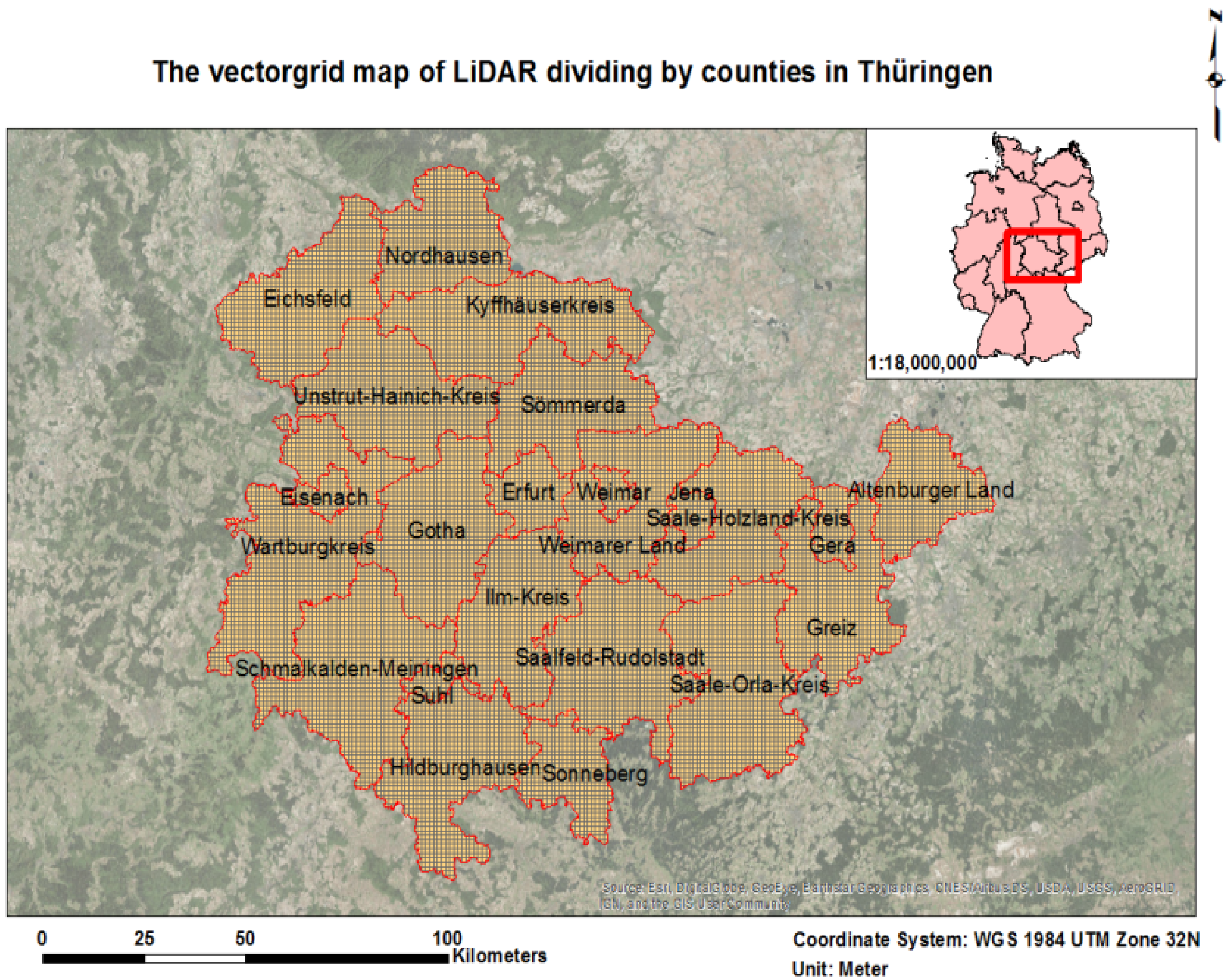
Digital Terrain Model (DTM) is a DEM that represents the elevation of the **bare earth** without taking into account any overground features (e.g. trees, buildings)

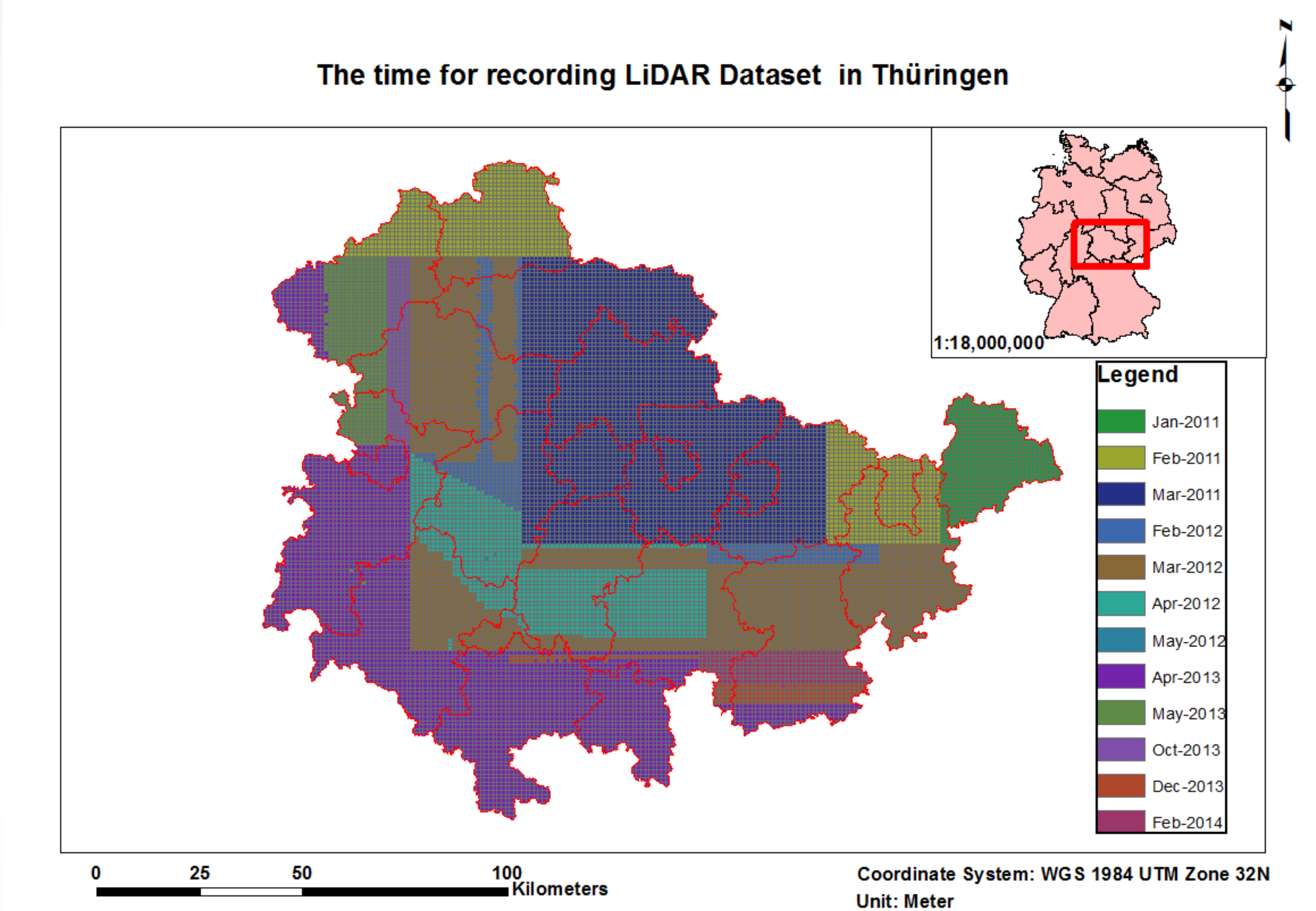
Digital Surface Model (DSM) contains elevations of natural terrain in addition to the natural (e.g. trees, shrubs) and man-made features (e.g. buildings)

Normalised Digital Surface Model (nDSM), result of the difference of the Surface and Terrain Models, highlights off-terrain objects.

- Introduction
- Research Target
- **Study Area**
- Material
- Methodology
- Result
- Discussion

These LiDAR datasets in Thüringen were published in the website of geo-spatial Thüringen state (http://www.geoportal-th.de/de-de/start.aspx).





- **16890 titles** of 1 km^2 in .LAZ format
- Storage about **410 Gb**

Introduction

Research Target

Study Area

Material

Methodology

Result

Discussion

- **Model Server Grid** (64 Cores) by UFZ

→ However, the optimum cores are 32

- A storage space of **10 Terabytes**
- **R software**: “doParallel” and “foreach” package
- **Lastools** software (Rapidlasso Company)

Introduction

Research Target

Study Area

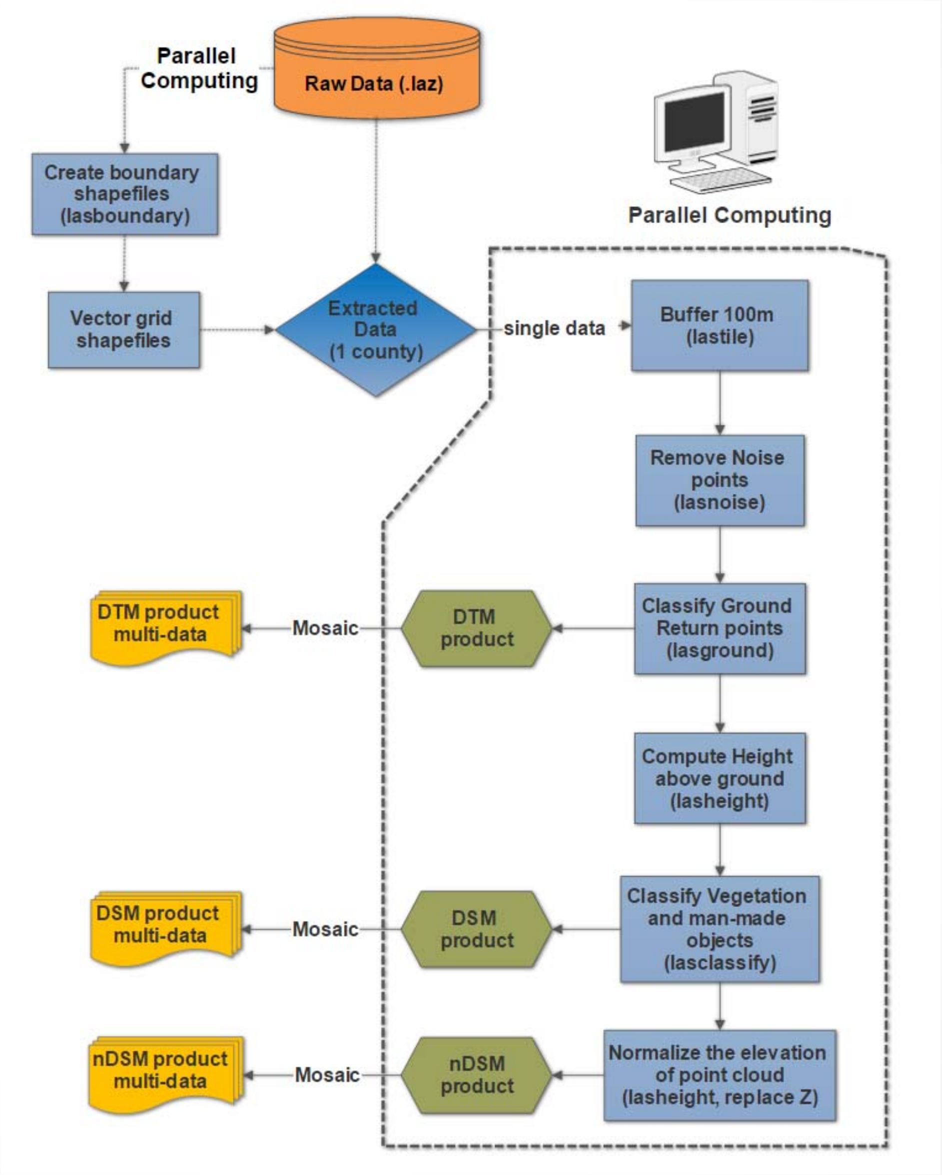
Material

Methodology

Result

Discussion

The workflow of the Big LiDAR processing



A part of R codes for processing LiDAR data

```
#Remove the noise object in LiDAR
infile <- paste0(fn, ".laz")
outfile <- paste0(fn, "_noise.laz")
system2(command = paste0(laspath, "lasnoise.exe"), args = c("-i", wd, infile, "-remove_noise", "-olaz", "-o", outfile))

#Remove the synthetic points
infile <- paste0(fn, "_noise.laz")
outfile <- paste0(fn, "_synpoint.laz")
system2(command = paste0(laspath, "las2las.exe"), args = c("-i", wd, infile, "-keep_class 0 1 2 13 20", "-olaz", "-o",

#Classify ground return
infile <- paste0(fn, "_synpoint.laz")
outfile <- paste0(fn, "_grclass.laz")
system2(command = paste0(laspath, "lasground.exe"), args=c("-i", wd, infile,"-metro", "-olaz", "-o", outfile))

# Calculate height of each point above ground
infile <- paste0(fn, "_grclass.laz")
outfile <- paste0(fn, "_height.laz")
system2(command=paste0(laspath, "lasheight.exe"), args=c("-i", wd, infile, "-olaz", "-o", outfile))

# Classify vegetation and man-made objects
infile <- paste0(fn, "_height.laz")
outfile <- paste0(fn, "_class.laz")
system2(command=paste0(laspath, "lasclassify.exe"), args=c("-i", wd, infile, "-olaz", "-o", outfile))

# Create a point cloud with only ground returns
infile <- paste0(fn, "_class.laz")
outfile <- paste0(fn, "_ground.laz")
system2(command=paste0(laspath, "lasclassify.exe"), args=c("-i", wd, infile, "-olaz", "-keep_class 2", "-o", outfile))

# Normalize for terrain
infile <- paste0(fn, "_class.laz")
outfile <- paste0(fn, "_norm.laz")
system2(command=paste0(laspath, "lasheight.exe"), args=c("-i", wd, infile, "-olaz", "-replace_z", "-o", outfile))

# Derive a digital terrain model (DTM) from the ground return point cloud
infile <- paste0(fn, "_ground.laz")
outfile <- paste0(fn, "_DTM.tif")
system2(command=paste0(laspath, "las2dem.exe"), args=c("-i", wd, infile, "-kill 200", "-elevation", "-o", outfile))

# Derive a normalized digital surface model (nDSM) from the normalized point cloud
infile <- paste0(fn, "_norm.laz")
outfile <- paste0(fn, "_nDSM.tif")
system2(command=paste0(laspath, "lasgrid.exe"), args=c("-i", wd, infile, "-elevation", "-highest", "-o", outfile))
```

Introduction

Research Target

Study Area

Material

Methodology

Result

Discussion

County	Number of LiDAR tiles with 100m buffer titles (unit)	Time spent for creating 100m buffer tiles (hours)	Time spent for LAStools processing (hours)	Total recording time (hours)	The time spent per 1 Core (hours)
Alterbuger Land	653	0.91	4.10	5.01	0.16
Gotha	1049	1.60	5.21	6.81	0.21
Ilmkreis-Suhl	1085	1.28	4.02	5.30	0.17
Eichsfeld	1059	1.40	3.77	5.17	0.16
Erfurt-Sommerda	1207	2.60	4.48	7.08	0.22
Gera-Greiz	1125	4.29	4.05	8.34	0.26
Hildburghausen	1081	2.20	3.14	5.34	0.17
Jena-SaaleHolzlandKreis	1038	1.90	2.62	4.52	0.14
Kyffhaeuserkreis	1178	1.79	4.94	6.73	0.21
Nordhausen	817	1.08	2.80	3.88	0.12
Saale-Orla-Kreis	1292	1.53	3.03	4.56	0.14
Saalfeld-Rudolstadt	1204	1.73	3.61	5.34	0.17
Schmalkalden-Meiningen	1362	2.05	4.53	6.58	0.21
Sonneberg	528	0.85	1.55	2.40	0.07
Unstrut-hainich-kreis	1100	2.30	4.57	6.87	0.21
Wartburgkreis-Eisenach	1606	2.50	5.91	8.41	0.26
Weimar-WeimarerLand	999	1.21	1.94	3.15	0.10

Introduction

Research Target

Study Area

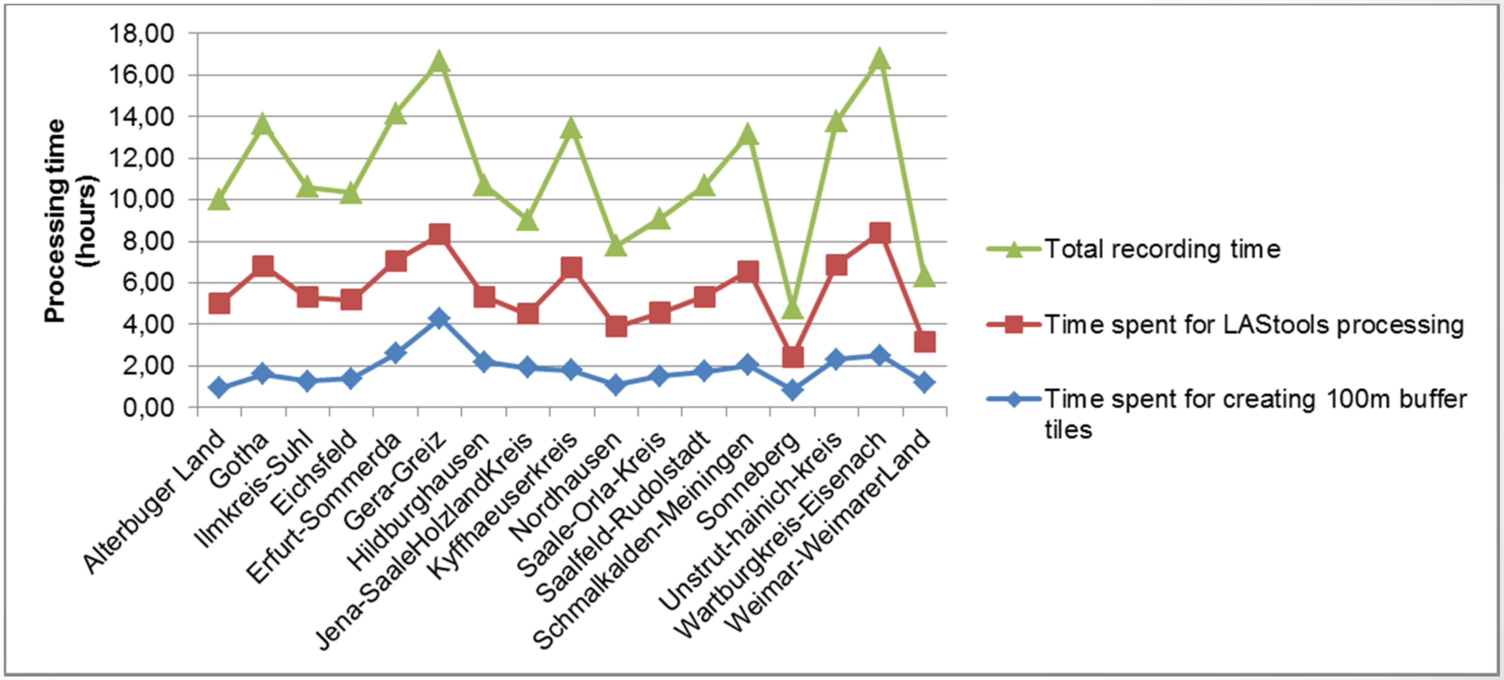
Material

Methodology

Result

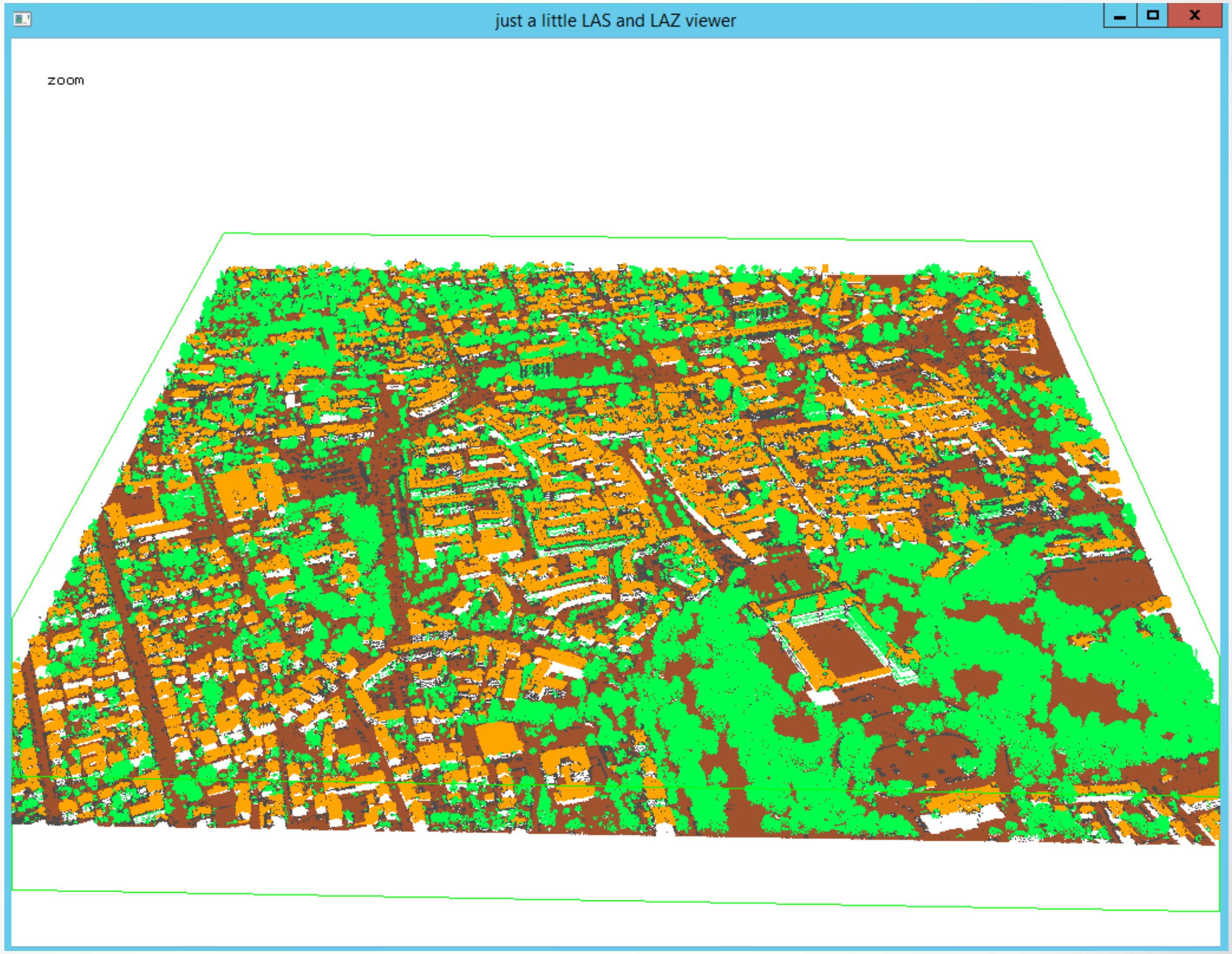
Discussion

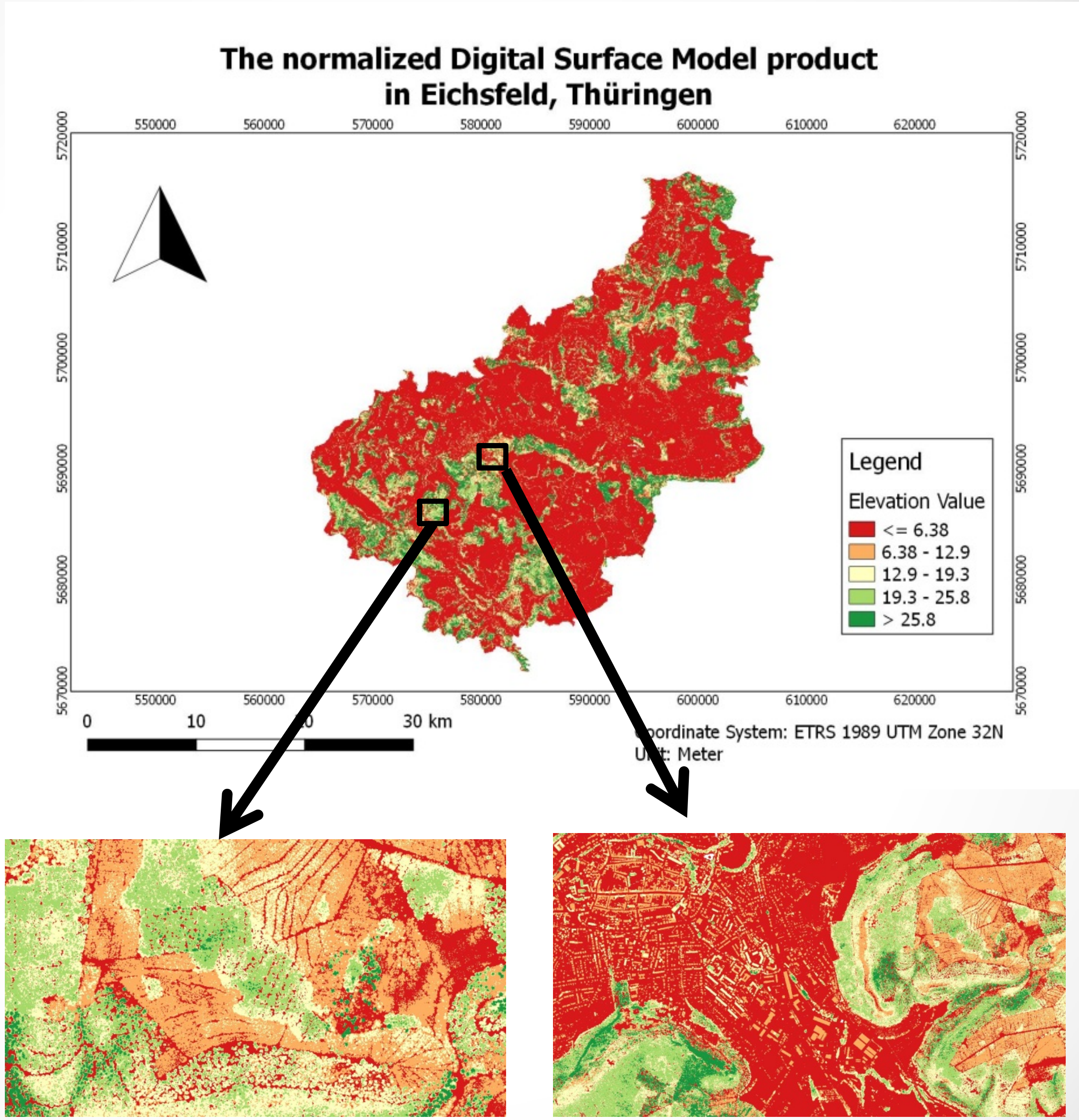
The diagram of spending time comparison for processing data in 17 counties



-> 6.22 Terabytes of the storage space were consumed.

An example of displaying 3D LiDAR dataset with Classification in Gotha City (lastools)




The normalized Digital Surface Model product
in Eichsfeld, Thüringen
550000
560000
570000
580000
590000
600000
610000
620000
5720000
5710000
5700000
5690000
5680000
5670000
Legend
Elevation Value
<= 6.38
6.38 - 12.9
12.9 - 19.3
19.3 - 25.8
> 25.8
0
10
30 km
Coordinate System: ETRS 1989 UTM Zone 32N
Unit: Meter

Introduction

Research Target

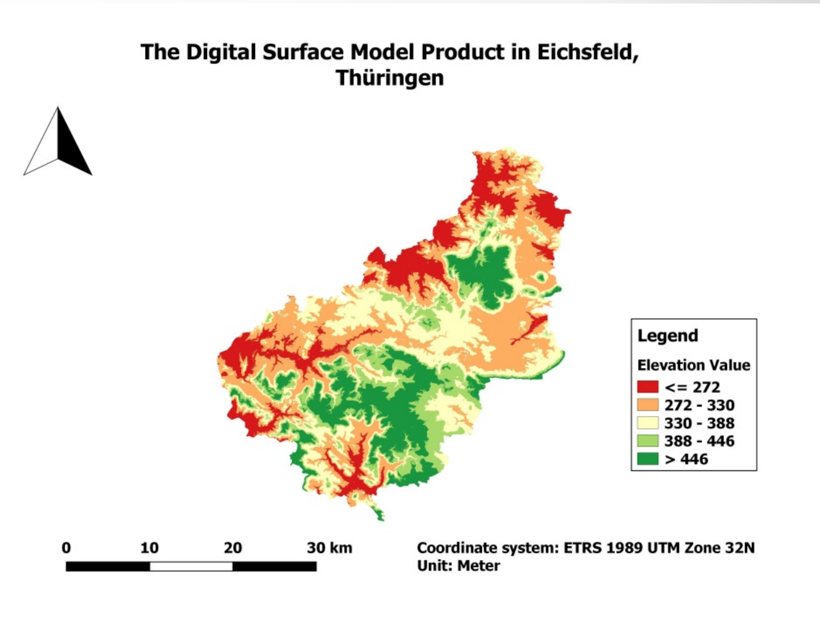
Study Area

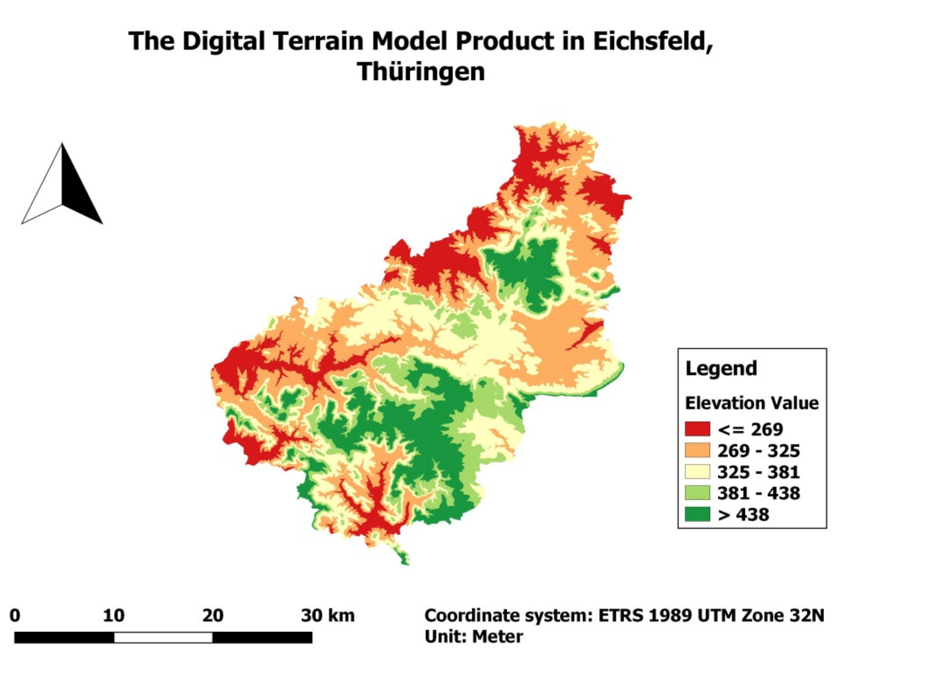
Material

Methodology

Result

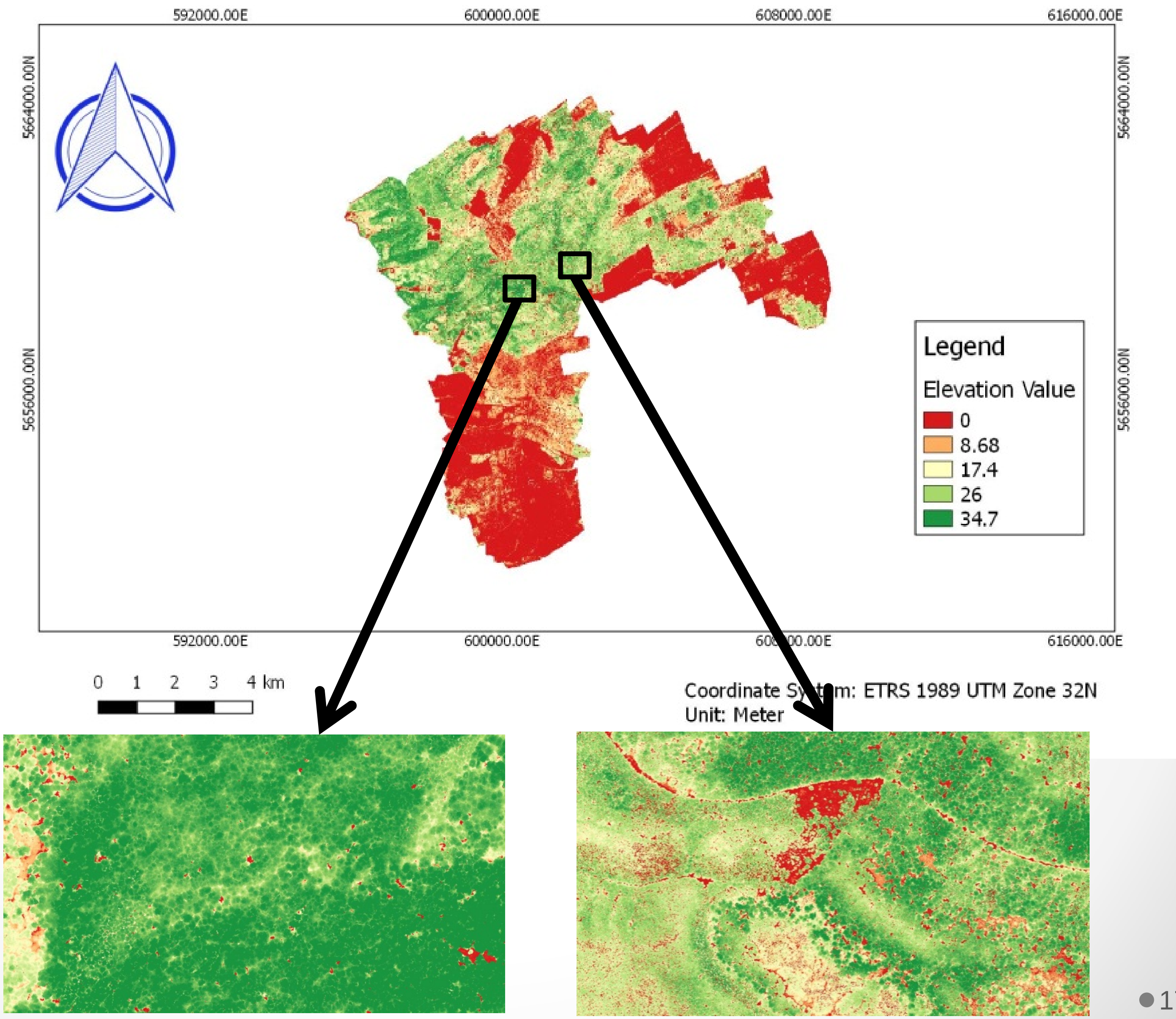
Discussion





- Introduction
- Research Target
- Study Area
- Material
- Methodology
- **Result**
- Discussion

The normalized Digital Surface Model product in Hainich National Park, Thüringen



Introduction

Research Target

Study Area

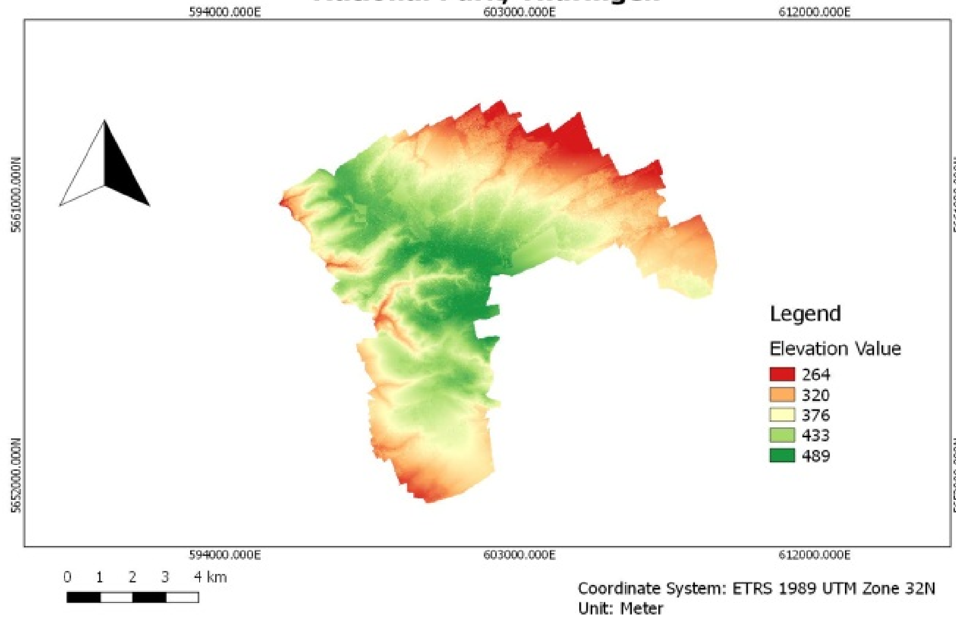
Material

Methodology

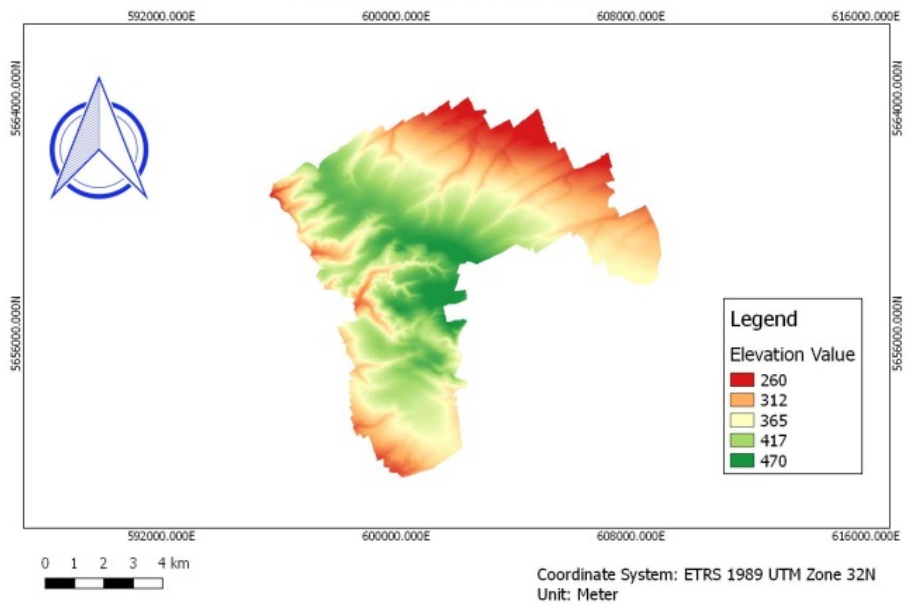
Result

Discussion

The Digital Surface Model product in Hainich National Park, Thüringen

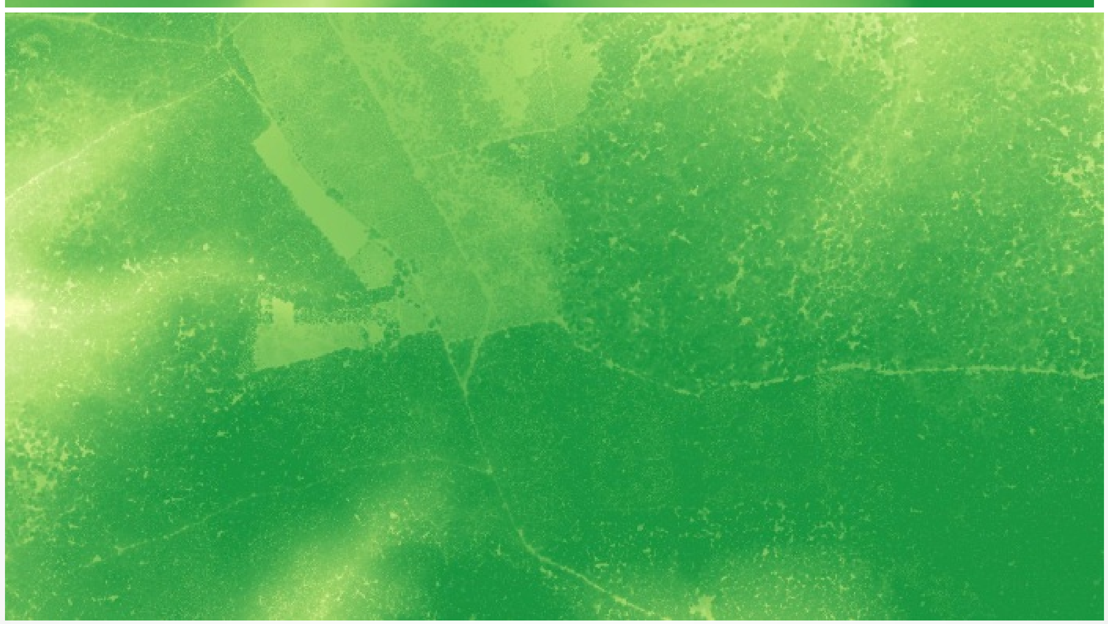


The Digital Terrain Model product in Hainich National Park, Thüringen



Comparison on the 2 products

Snapshot of zoom in DTM product



Snapshot of zoom in DSM product

Introduction

Research Target

Study Area

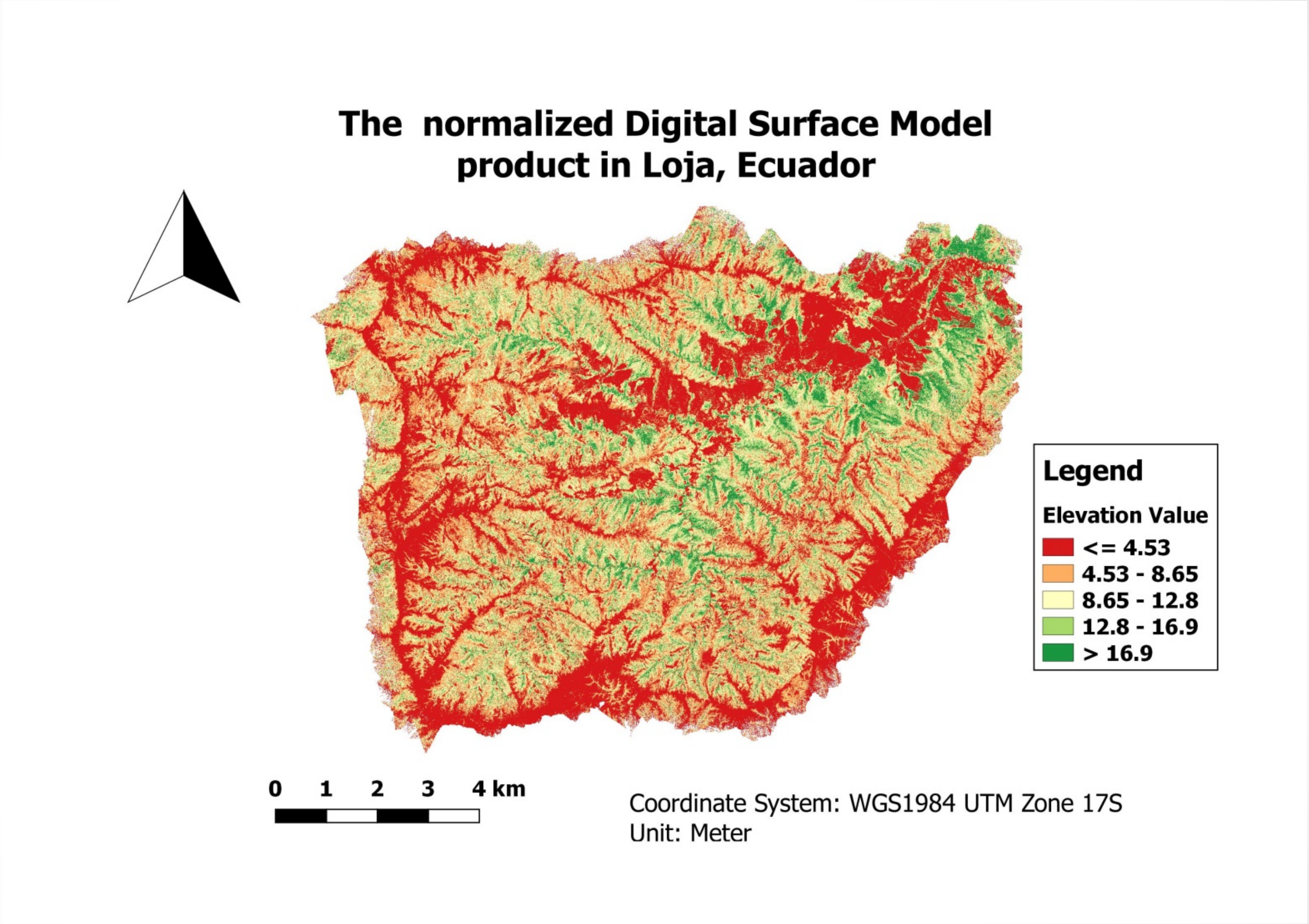
Material

Methodology

Result

Discussion

Applying framework to another LiDAR data in Ecuador



→ Proving: The framework was succesful for processing any LiDAR data composition.

The framework was proven to efficiently process large-scale LiDAR dataset in parallel in a highly scalable distributed environment by using LiDAR processing tools.

LAStools were included to perform particular LiDAR processing steps due to their efficiency, reliability and popularity.

Limitation:

- Some problems of defining some man-made objects (windmills, power lines)
- The final products provided some small holes, because there are no point cloud from water body, swamp and boundary of building → apply interpolation

Future research:

Developing the framework which contains more applications such as calculating LiDAR metrics: Leaf Area Index (LAI), above ground biomass (AGB)…

REFERENCE

- Baltsavias, E.P. 1999. Airborne laser scanning: Existing systems. firms. and other resources. ISPRS Journal of Photogrammetry and Remote Sensing. 54(2-3):164-198.
- Guan. H., Li. J., Zhong. L., Yongtao. Y., & Chapman. M. (2013). Process virtualization of large-scale lidar data in a cloud computing environment.Computers & Geosciences. 60. 109-116.
- Isenburg., M. (2012). LAStools-Efficient tools for LiDAR processing. Available at: http://www.cs.unc.edu/~ isenburg/lastools.
- Leach. C. (2014). Introduction to parallel computing in R. Available at: http://michaeljkoontz.weebly.com/uploads/1/9/9/4/19940979/parallel.pdf
- Lefsky M.A., W.B. Cohen. G.G. Parker and D.J. Harding. 2002. Lidar remote sensing for ecosystem studies. Bioscience 52(1) 19-30
- Li Z., Hodgson M., Li W. (2017), A General-purpose Framework for Parallel Processing of Large-scale LiDAR Data, International Journal of Digital Earth, DOI:10.1080/17538947.2016.1269842
- Weston. S., Calaway. R. (2015). Getting Started with doParallel and foreach. Available at: https://cran.r-project.org/web/packages/doParallel/vignettes/gettingstartedParallel.pdf

Thanks for your attention!

Further question or information,
please contact: duc-thang.nguyen@hnee.de