EcoDes-DK15 Variable Descriptions

*Version 1.0 (28 September 2021)

Authors: Jakob J. Assmann, Urs A. Treier, Andràs Zlinsky, Jesper E. Moeslund and Singe Normand

Content

Dataset Description and Purpose

This dataset contains ecological and landscape descriptors extracted from the point clouds of Denmark's nationwide ALS/LiDAR dataset *'DHM/Punktsky'* collected in 2014/15. The raw point clouds can be accessed on the website of Kortforsyningen and documentation for the raw data is available here and here.

The purpose of this dataset is to provide a light-weight version of the nationwide ALS data condensed into easyly interpretable descriptors, which summarise the structure of the point cloud data for ecological and biological studies. As the data was collected in 2014/15 the data provides a snapshot of Denmark at that time, since then changes are likely to have occurred.

The extent of the dataset comprises the majority of the Danish land surface (including many of the small islands and Bornholm) split into 49 835 tiles. The data is provided as GeoTIFF rasters projected in ETR89 UTM 32 N based on the GRS80 spheroid (EPSG: 25832). NoData values are globally set to -9999, but please see the description on how to interpret the NoData cells for the individual variables. Masks for sea and small in-land water bodies are provided, but have to be applied manually where appropriate.

This document summarises the eighteen ecological and landscape descriptors extracted by us and how they were derived. We also highlight known issues relevant to the interpretation of these variables.

This document is mainly seen as an online complement to the manuscript accompanying the dataset: *Assmann et al.* ***in prep*** *- EcoDes-DK15: High-resolution ecological descriptors of vegetation and terrain derived from Denmark's national airborne laser scanning dataset* INSERT DOI!. In case of any discrepancies between this document and the manuscript, the description in the manuscript prevails unless otherwise stated.

Overview

The section provides a quick overview of all outputs and auxiliary files.

Point cloud derived variables

The data source for these variables are the raw point clouds provided by Kortforsyningen.

variable name	average file size
amplitude_mean and amplitude_sd	46 kb (2x)
canopy_height	42 kb
normalized_z_mean and normalized_z_sd	50 kb (2x)
point_counts	10 kb (28 x)
point_source_info	110 kb (4x)
proportions	19 kb (25x)

Terrain model derived variables

The data source for these variables are digital terrain model (DTM) rasters ultimately derived from the raw point clouds at a 0.4 m grain size. These 0.4 m DTM rasters are directly provided by Kortforsyningen and can be found here.

variable name	average file size
aspect	20 kb
dtm_10m	20 kb
heat_load_index	20 kb
openness_difference	20 kb
openness_mean	20 kb
slope	20 kb
solar_radiation	20 kb
twi	20 kb

Auxiliary Files

Files to support data access and handling.

file name	description
water masks	Sea and inland water masks for each tile
tile footprints	Tile footprints, allows for targeted subsetting of dataset
date_stamp	Date on which majority of points within a pixel were collected
vrt files	VRT files (virtual mosaic file) for each variable

Histogram Overview

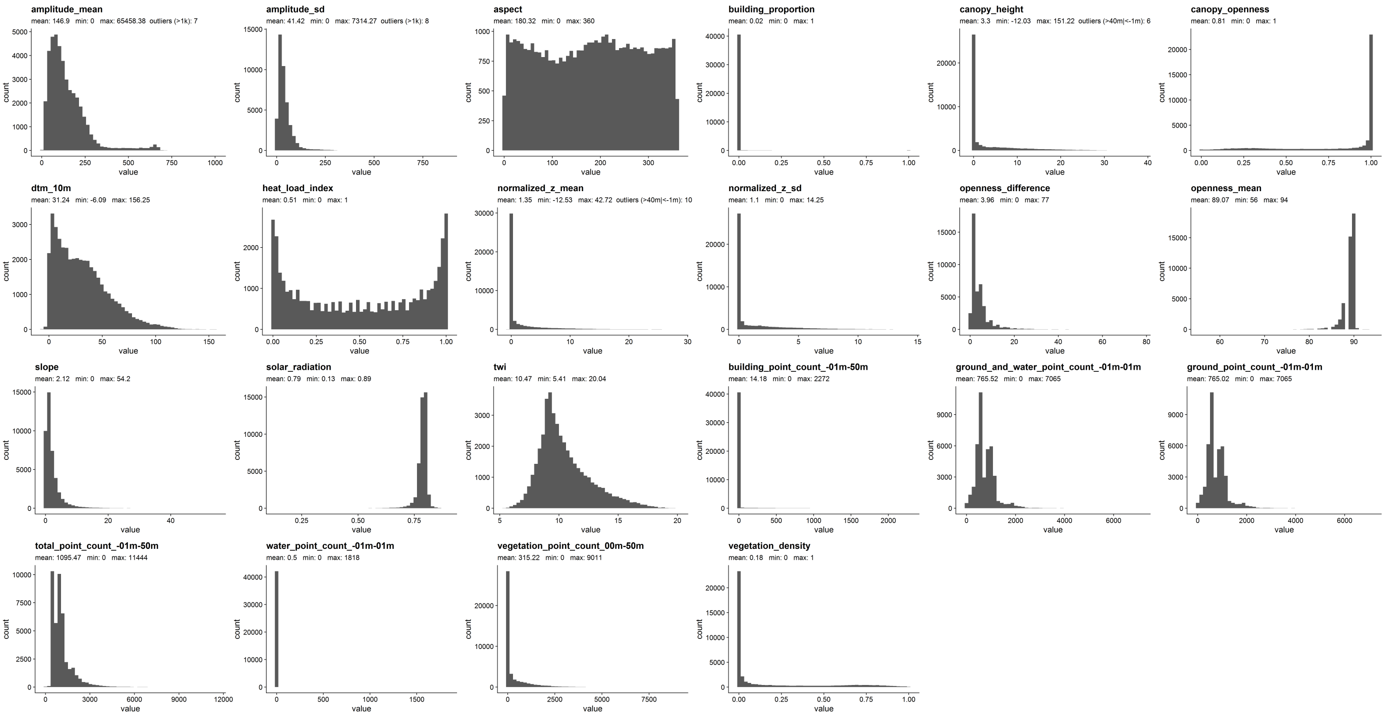


Figure 1: Overview of sample histograms (n = 50k) for all variables except the point source information variables and the vegetation point counts/proportions per height bin. Note that the histograms for the amplitude variable were curtailed above a value of 1000, and the canopy_height and normalized_z_mean were curtailed at 40 m. The number of outliers above these values is shown in the subtitle. The random geographic distribution of the point sample is shown in the figure below. Individual histograms are provided with each variable entry in the documentation. Samples that were covered by the inland and sea water masks were removed.

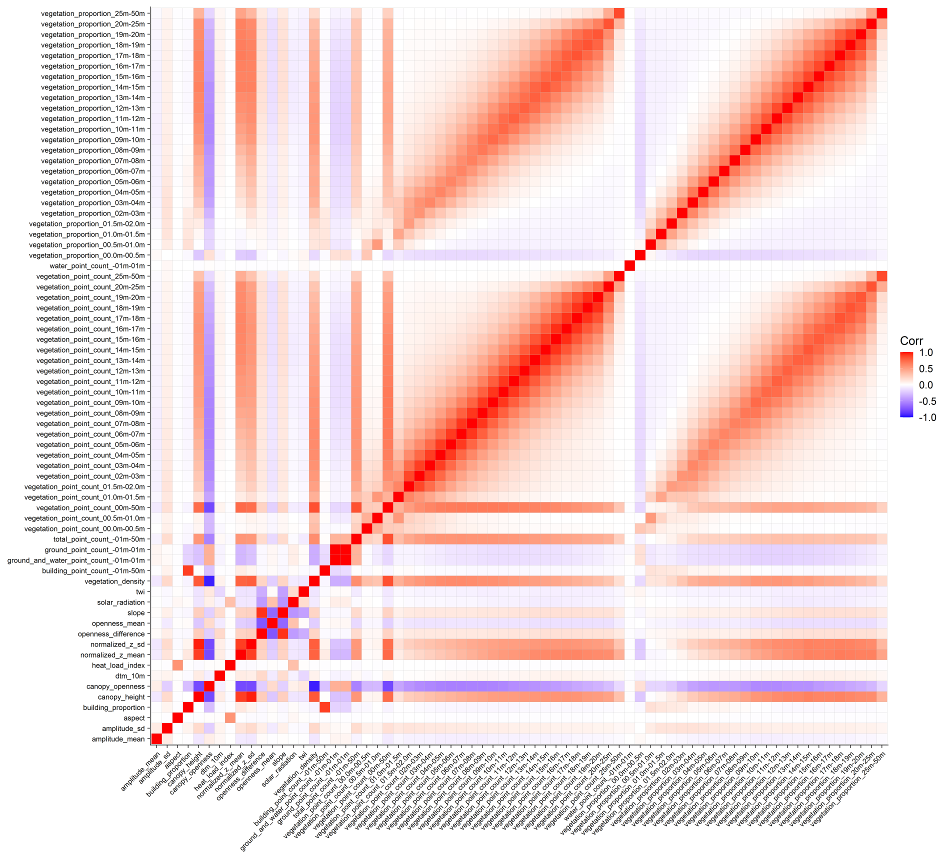


Figure 2: Correlation plot for all variables based on the same sample (n = 50k) as shown in the histogram plots / sample map. Samples that were covered by the inland and sea water masks were removed.

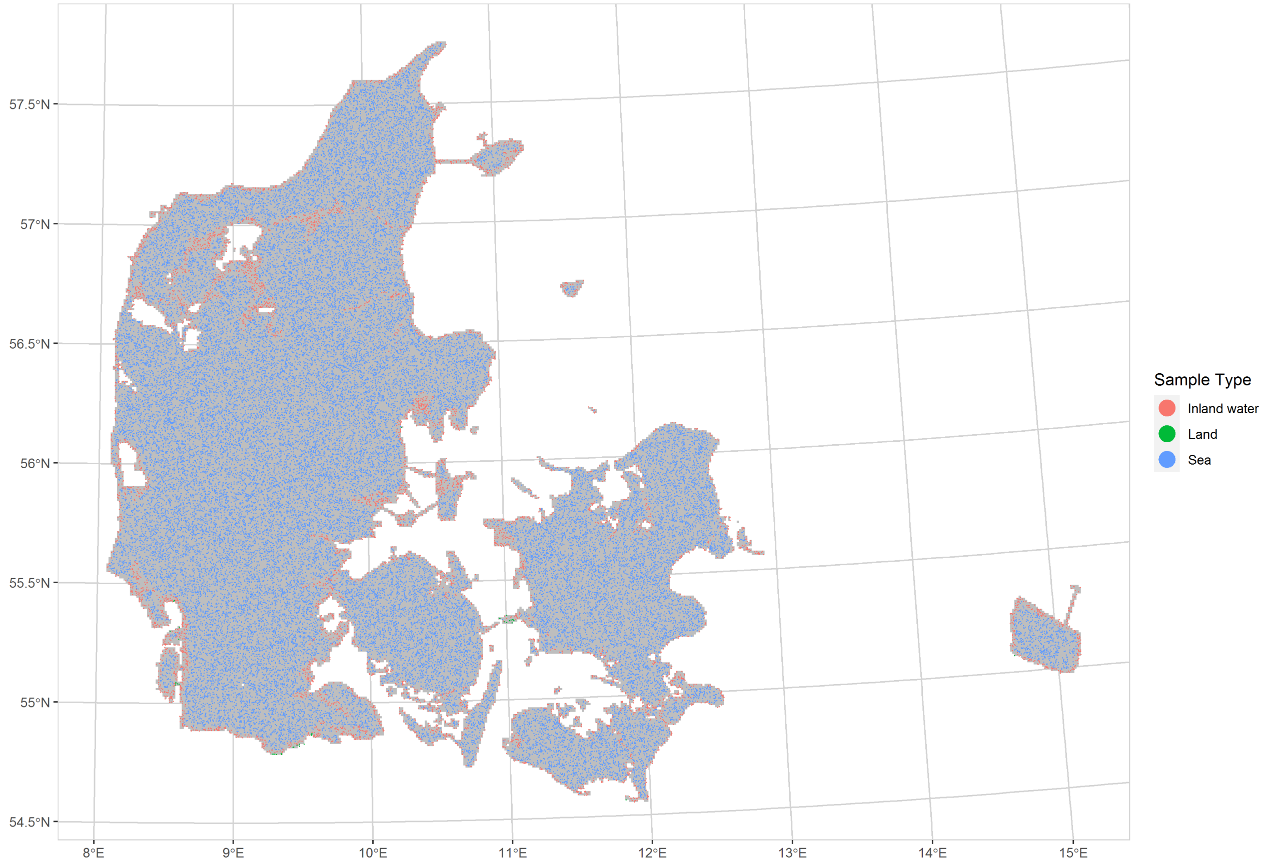


Figure: 3 Map of random sample locations (n = 50k) for data shown in the histogram and correlation plots. The colour scale indicates whether a sample location was covered by the sea or inland water masks.

Processing time

For all 49k tiles the processing took around 45 days for the first complete run through using 54 parallel threads on a dedicated machine with 2 x Intel Xenon Platinum 8180 @2.5GHz, 1.5 TB Ram (little used), a fast access RAID and Windows Server 2012 R2. An additional week and a half was needed for re-processing of some of the digital terrain model variables.

The total data volume of the outputs is around 80 GB (compressed 16.4 GB).

Completeness

The data set consists of 49832 tiles. Processing was almost completely successful with an average of only 34 tiles failing to be processed per variable. The majority of those tiles was located on the fringes of the data set. To ascertain that the data set has no gaps, we replaced any tiles that were missing due to processing failures with empty raster of the same extent and grain contain only nodata values (-9999). An overview of the number of tiles affected per variable can be found in /documentation/empty_tiles.csv on the GitHub repository. The tile_ids for the affected tiles for each variable can be found in the "empty_tiles_XXX.txt" file in the archive/folder of the variable.

Back to content.

Point cloud derived variables

amplitude_mean and amplitude_sd

Folder locations: `/outputs/amplitude/mean` and `/outputs/amplitude/sd`

File names: `amplitude_mean_xxxx_xxx.tif` and `amplitude_sd_xxxx_xxx.tif`

File type and units: `32-bit float, undefined`

Description:

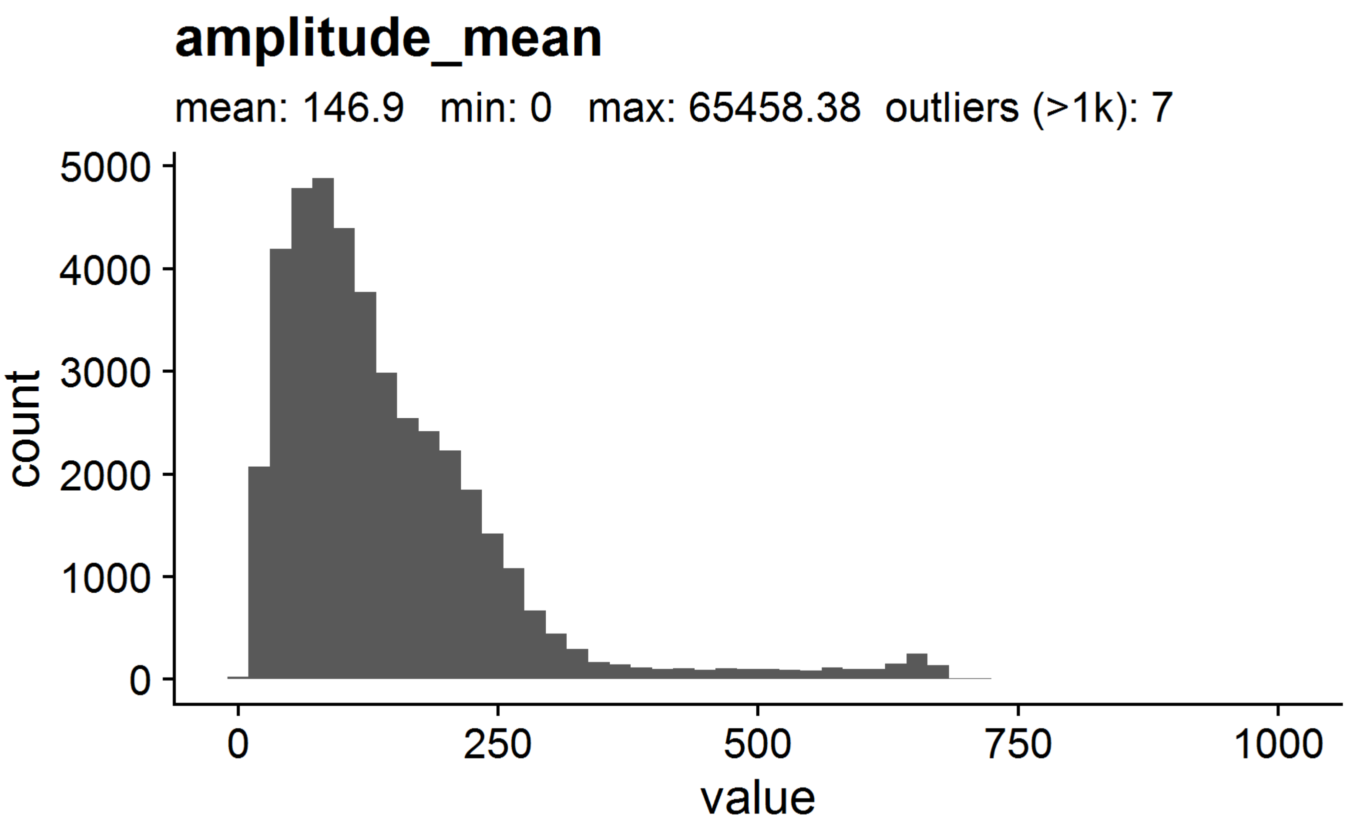
Arithmetic mean and standard deviation of the return amplitude for all points within a 10 m x 10 m grid cell.

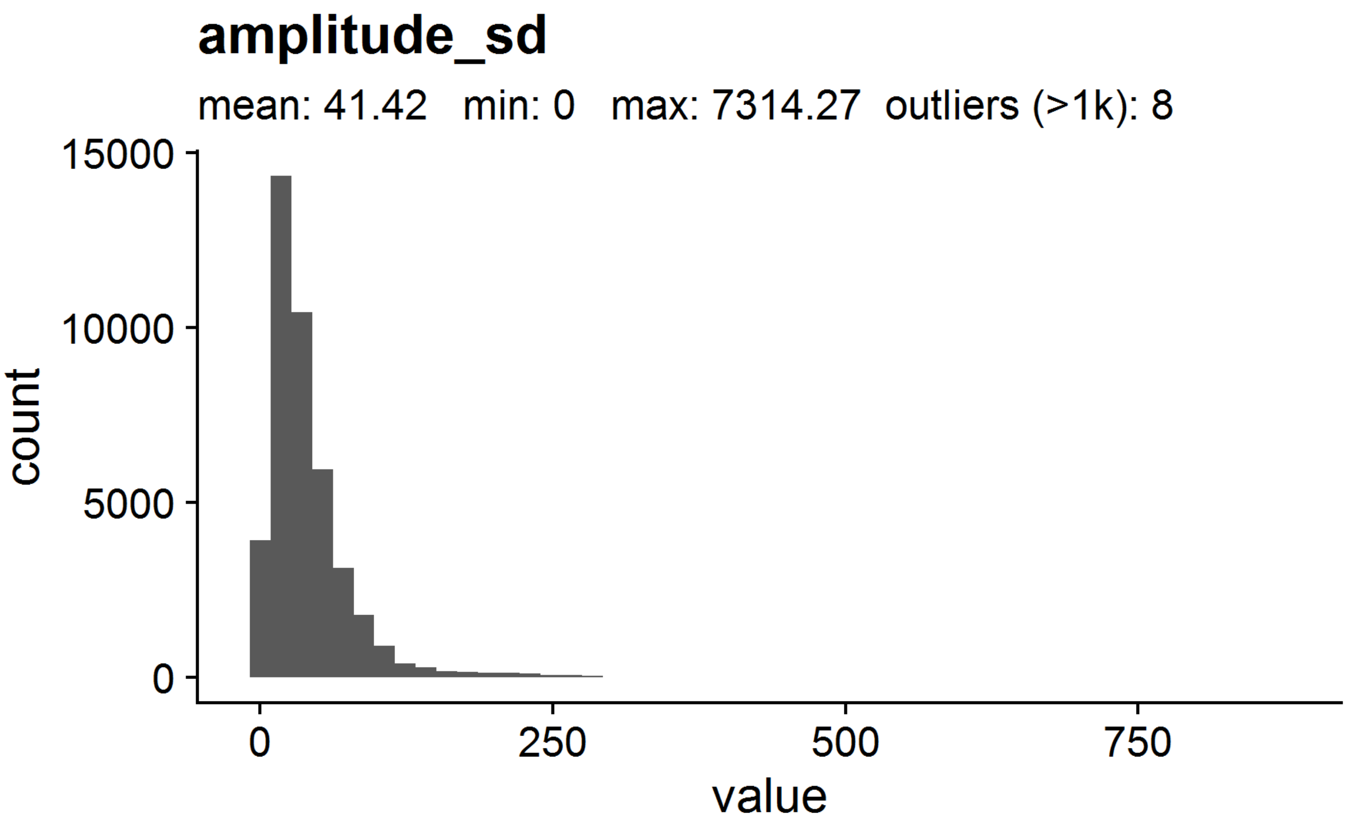
In the context of LiDAR, the amplitude represents the strength of the signal received by the sensor for each return. For this variable the arithmetic mean and standard deviation of the amplitude were calculated for all points within a 10 m x 10 m cell. Calculations were carried out for a single tile using the [OPALS Cell module](). Here, all points refers exactly to the following set of classes: ground, water, building, as well as low-, medium- and high-vegetation.

Issues:

- The interpretation of the amplitude variable is not straightforward and its biological meaning can be complex. Nonetheless, the amplitude is sensitive to moisture and captures texture, it can therefore be highly useful for segmenting vegetated and non-vegetated surfaces.
- Amplitude is not directly comparable across point source ids due to potential differences in sensor unit etc. As some cells may contain returns from up to four different point source ids, we recommend using the amplitude variable in conjunction with information on the point source ids within each cell contained in the point_source_info variables.

Sample Histogram(s):





Note: The histograms are based on the sample of 50k cell cells shown in Figure 3. We removed all samples with a value larger than 1000 to illustrate the distributions independently of the outliers.

References:
No specific references available.

[Back to content.]()

---

canopy_height

Folder location: `/outputs/canopy_height`

File name: `canopy_height_xxxx_xxx.tif`

File type and units: `16 bit integer, metre x 100`

Description:

Canopy height calculated as the 95th-percentile of the normalised height above ground of all vegetation points within a 10 m x 10 m cell.

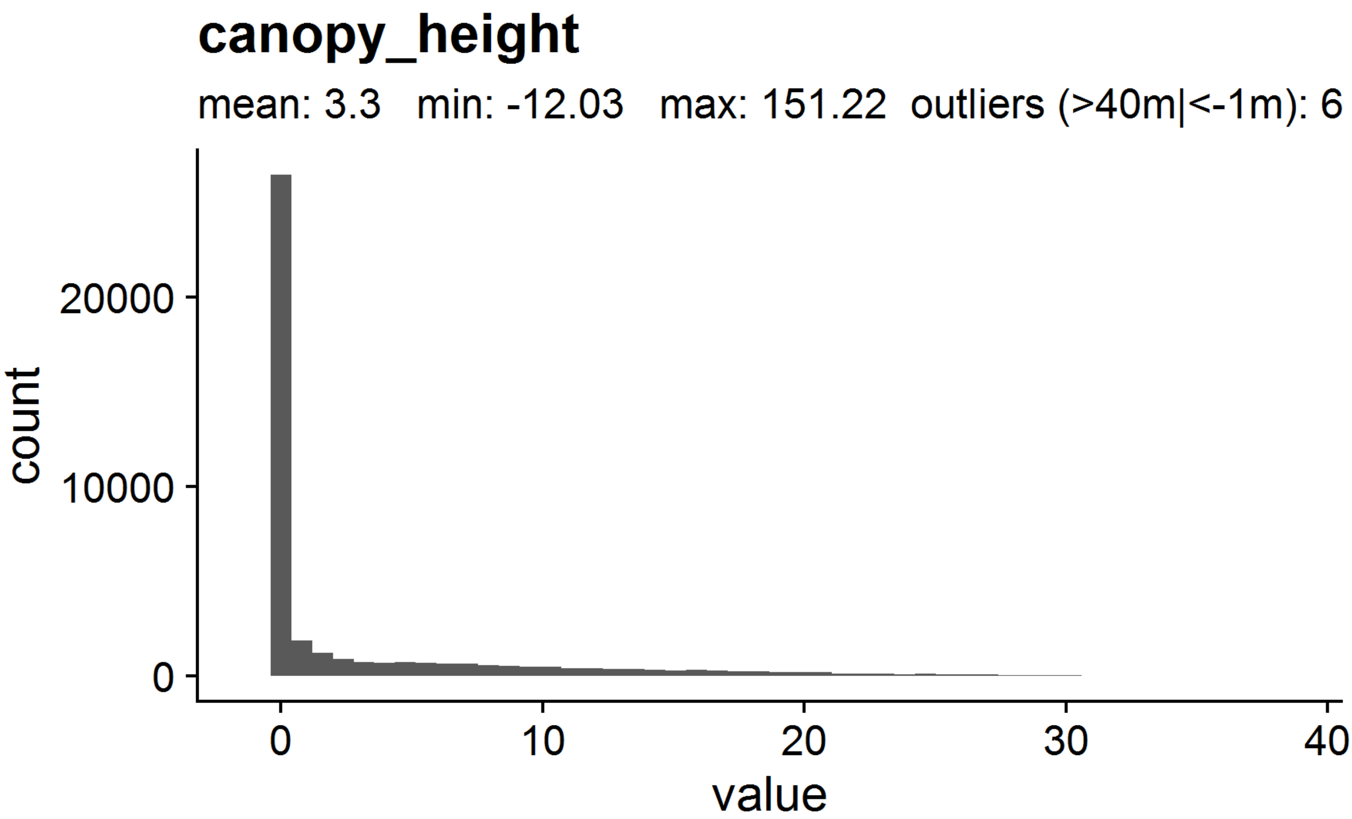
Calculated with OPALS Cell for each tile individually. Vegetation points consist of the following classes: low-, medium- and high- vegetation. Should there be no vegetation points in any given cell the value of the cell is set to zero.

Issues:

- If a cell contains no points the value is set to zero, not NA.
- In rare cases, this variable might lead to erroneous canopy-height readings if vegetation is found on

artificial structures. The canopy height is calculated even if there is only a small amount of vegetation returns in a cell. For example: A tall communications tower can be found just south of Aarhus. On top of this tower small patches of vegetation resulted in vegetation point returns. The canopy height for this cell is > 100 m. The building proportion variable may help in identifying such cases.

Histogram:



Note: The histogram is based on the sample of 50k cell cells shown in Figure 3. Outliers larger than 40 m were removed from the histogram.

References:
No specific references available.

[Back to content.]

---

normalized_z_mean and normalized_z_sd

Folder locations: `/outputs/normalized_z/mean` and `/outputs/normalized_z/sd`

File names: `normalized_z_mean_xxxx_xxx.tif` and `normalized_z_sd_xxxx_xxx.tif`

File type and units: `16-bit integer, metre x 100` and `16-bit integer, metre x 100`

Description:

Arithmetic mean and standard deviation of the mean height above ground (normalised z) for all points in a 10 m x 10 m grid cell.

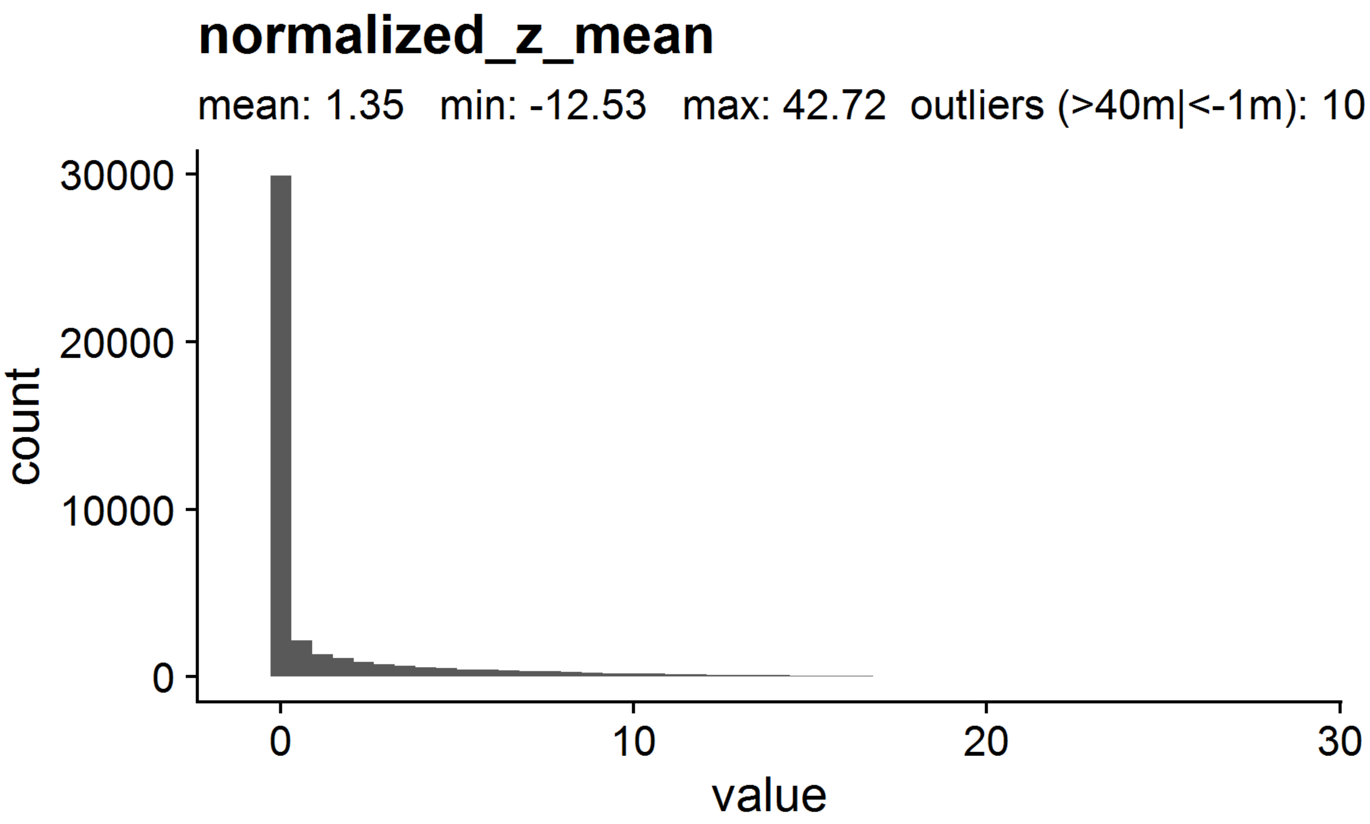
A normalised z attribute for each point were added to the point cloud of a single tile using [OpalsAddInfo](). To do so the absolute height above sea-level of a point was subtracted by the absolute height of the underlying cell of the 0.4 m digital terrain model. The arithmetic mean and standard deviation of the normalized_z were then calculated for all points within a 10 m x 10 m cell. These calculations were carried out with the [OPALS Cell module](). Here, all points refers to the following set of classes: ground, water, building, as well as low-, medium- and high-vegetation (classes 2,3,4,5,6,9).

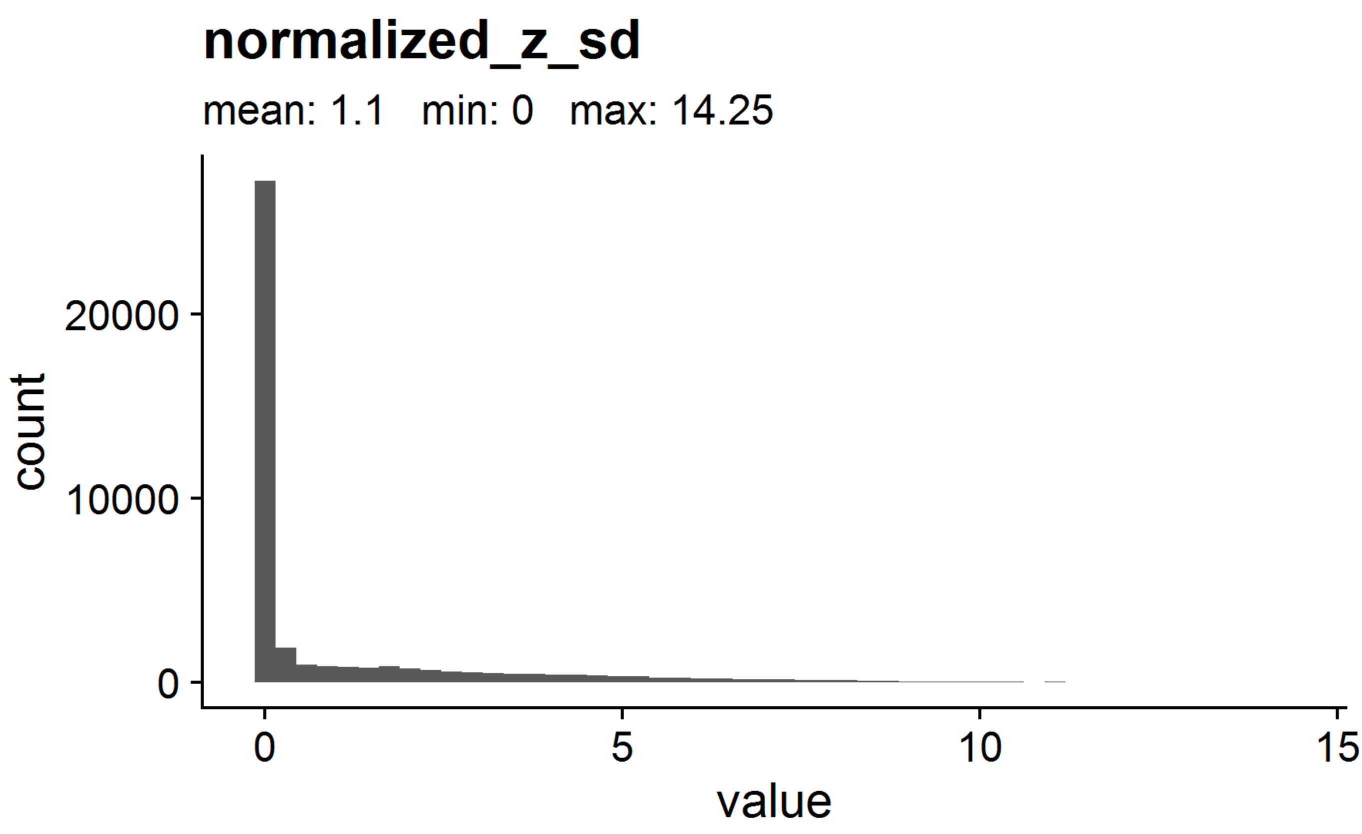
The American spelling of the variable name is kept for legacy reasons.

Issues:

- If a cell contains no points the value is set to zero, not NA.

Histogram(s):





Note: The histograms are based on the sample of 50k cell cells shown in Figure 3. Outliers larger than 40 m were removed from the normalised_z_mean histogram.

References:
No specific references available.

[Back to content.](#)

---

point_counts

Folder locations: `/outputs/point_count/point_count_name`

File names: `point_count_name_xxxx_xxx.tif`

File type and units: `16-bit integer, absolute count`

Description:

Absolute number of points within a 10 m x 10 m cell. Extracted for a combination of point classes and above height ranges specified below.

The "punktsky" point clouds were pre-classified by Geodatasyrelsen. The following point counts were extracted using the [OPALS Cell module](#) with filters applied for the respective height ranges and point classes.

General point counts:

name	height range	point classes
ground_point_count_-01m-01m	-1 m to 1 m	ground points (class 2)
water_point_count_-01m-01m	-1 m to 1 m	water points (class 9)
ground_and_water_point_count_-01m-01m	-1 m to 1 m	ground and water points (classes 2,9)
vegetation_point_count_00m-50m	0 m to 50 m	vegetation points (classes 3,4,5)
building_point_count_-01m-50m	-1 m to 50 m	building points (class 6)
total_point_count_-01m-50m	-1 m to 50 m	ground, water, vegetation and building points (classes 2,3,4,5,6,9)

Vegetation point counts for height bins:

name	height range	point classes
vegetation_point_count_00.0m-00.5m	0.0 m to 0.5 m	vegetation points (classes 3,4,5)
vegetation_point_count_00.5m-01.0m	0.5 m to 1.0 m	vegetation points (classes 3,4,5)
vegetation_point_count_01.0m-01.5m	1.0 m to 1.5 m	vegetation points (classes 3,4,5)
vegetation_point_count_01.5m-02.0m	1.5 m to 2.0 m	vegetation points (classes 3,4,5)
vegetation_point_count_02m-03m	2 m to 3 m	vegetation points (classes 3,4,5)
vegetation_point_count_03m-04m	3 m to 4 m	vegetation points (classes 3,4,5)
vegetation_point_count_04m-05m	4 m to 5 m	vegetation points (classes 3,4,5)
vegetation_point_count_05m-06m	5 m to 6 m	vegetation points (classes 3,4,5)
vegetation_point_count_06m-07m	6 m to 7 m	vegetation points (classes 3,4,5)
vegetation_point_count_07m-08m	7 m to 8 m	vegetation points (classes 3,4,5)
vegetation_point_count_08m-09m	8 m to 9 m	vegetation points (classes 3,4,5)
vegetation_point_count_09m-10m	9 m to 10 m	vegetation points (classes 3,4,5)
vegetation_point_count_10m-11m	10 m to 11 m	vegetation points (classes 3,4,5)
vegetation_point_count_11m-12m	11 m to 12 m	vegetation points (classes 3,4,5)
vegetation_point_count_12m-13m	12 m to 13 m	vegetation points (classes 3,4,5)
vegetation_point_count_13m-14m	13 m to 14 m	vegetation points (classes 3,4,5)
vegetation_point_count_14m-15m	14 m to 14 m	vegetation points (classes 3,4,5)
vegetation_point_count_15m-16m	15 m to 16 m	vegetation points (classes 3,4,5)
vegetation_point_count_16m-17m	16 m to 17 m	vegetation points (classes 3,4,5)
vegetation_point_count_17m-18m	17 m to 18 m	vegetation points (classes 3,4,5)
vegetation_point_count_18m-19m	18 m to 19 m	vegetation points (classes 3,4,5)
vegetation_point_count_19m-20m	19 m to 20 m	vegetation points (classes 3,4,5)
vegetation_point_count_20m-25m	20 m to 25 m	vegetation points (classes 3,4,5)
vegetation_point_count_25m-50m	25 m to 50 m	vegetation points (classes 3,4,5)

Additional information:

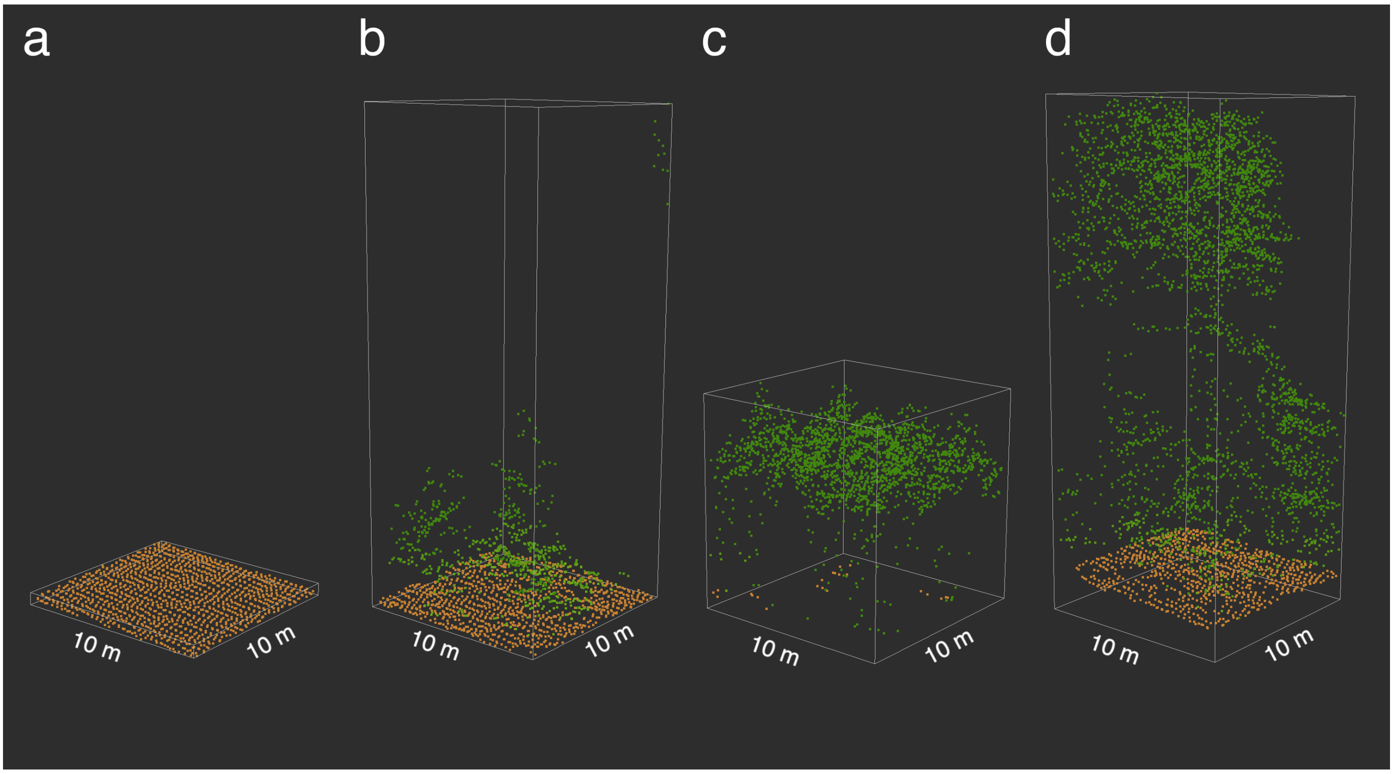


Figure 4: A set of canopy examples for visualisation of the point count variables. The approximate height of large bounding boxes is 25 m. a) agricultural field with no / very low vegetation, b) understory / shrub layer in mixed broadleaf woodland, c) dense young-ish coniferous forest (plantation?), d) old and tall mixed broadleaf woodland.

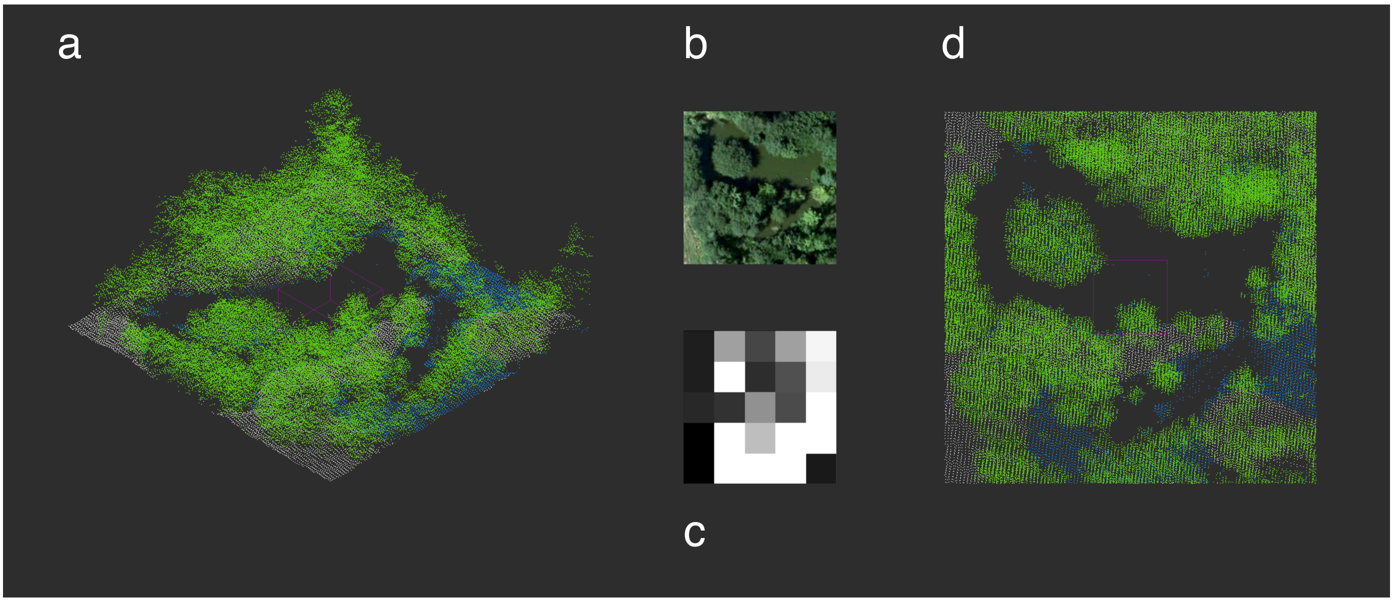
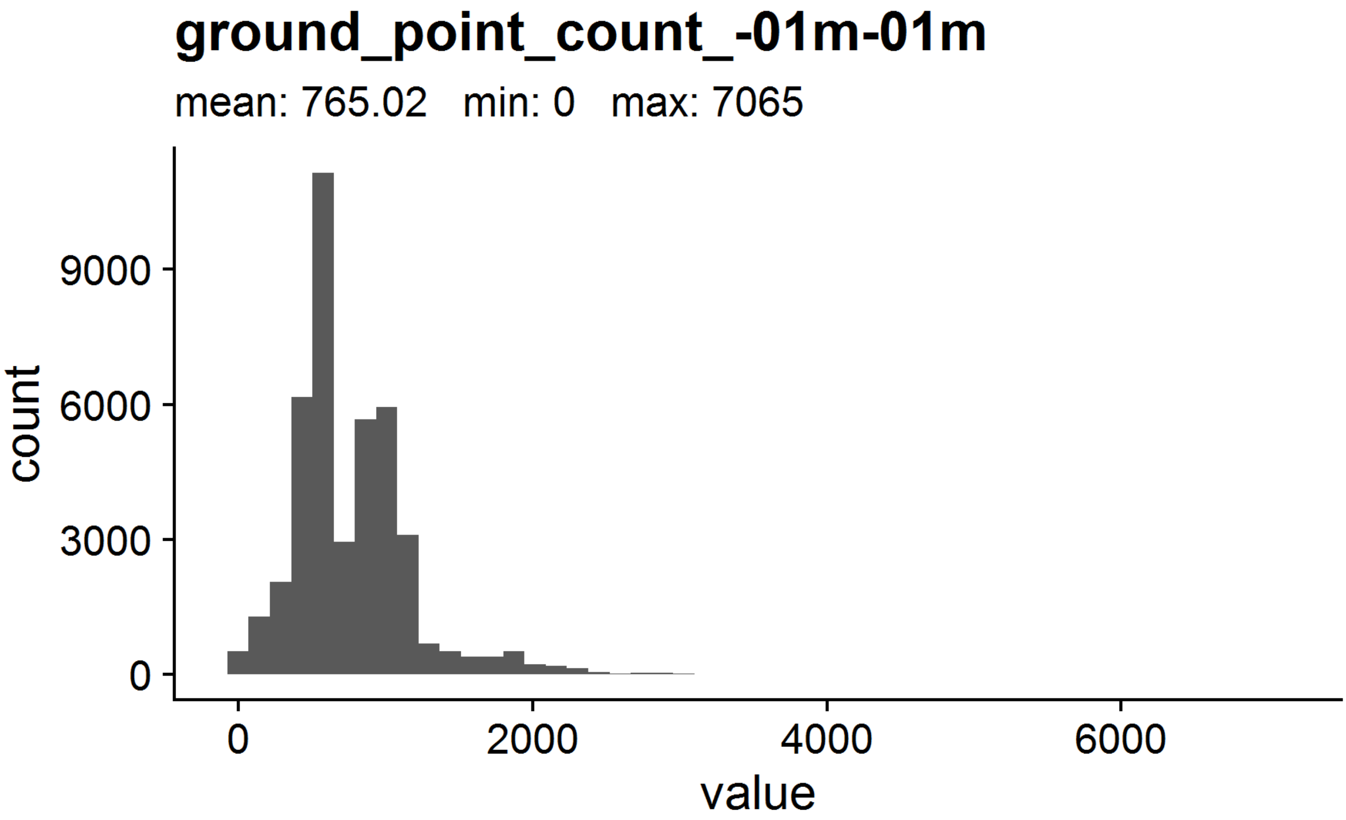


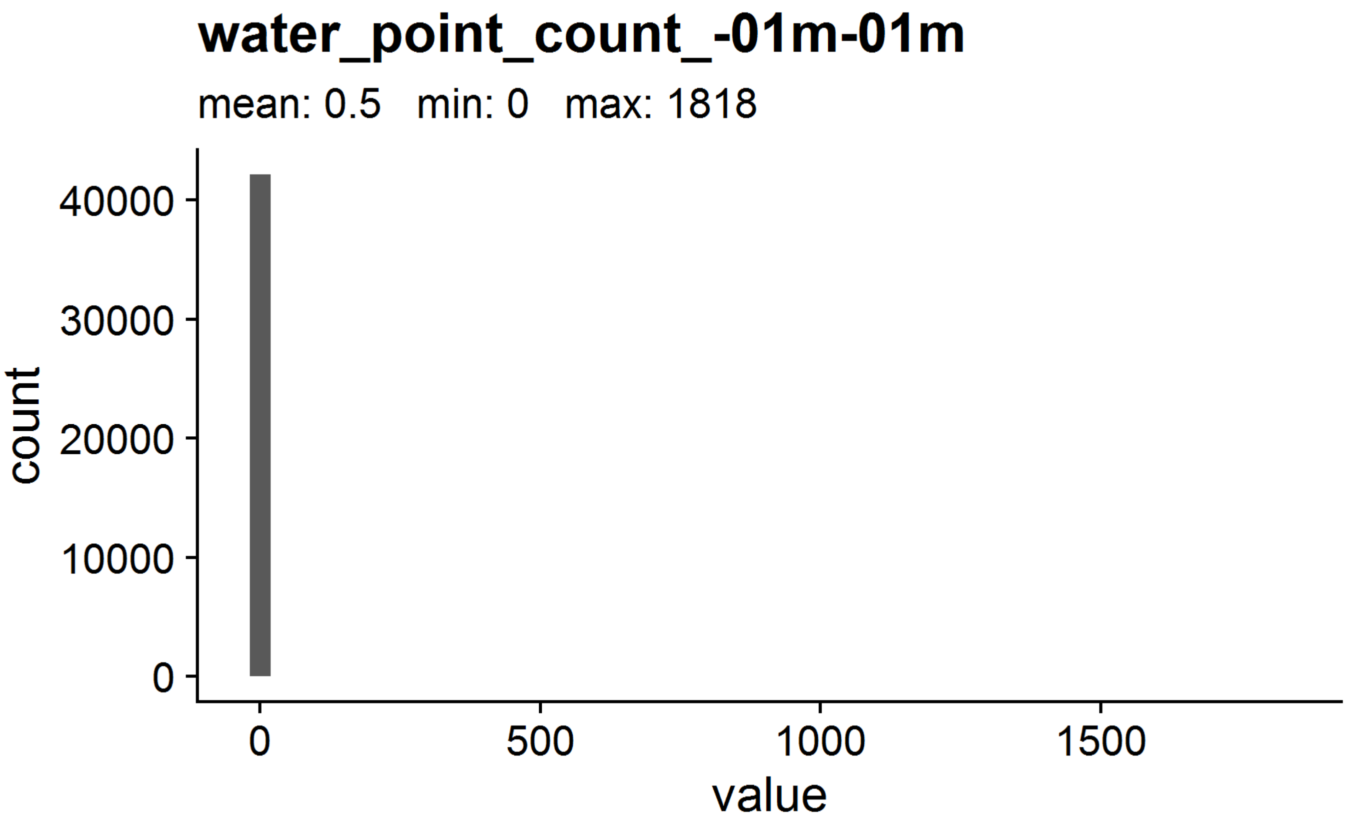
Figure 5: Example of the behaviour of returns from shallow waters in a forest pond / marsh area. a) perspective view of the forest pond (pink bounding box has a 10 m x 10 m footprint), b) orthophotograph at nadir view, c) point count intensity of the derived water point count variable (black = low count, white = high count), d) nadir view of point cloud. Please notice particularly the many missing returns from the regions in the pond with deep water.

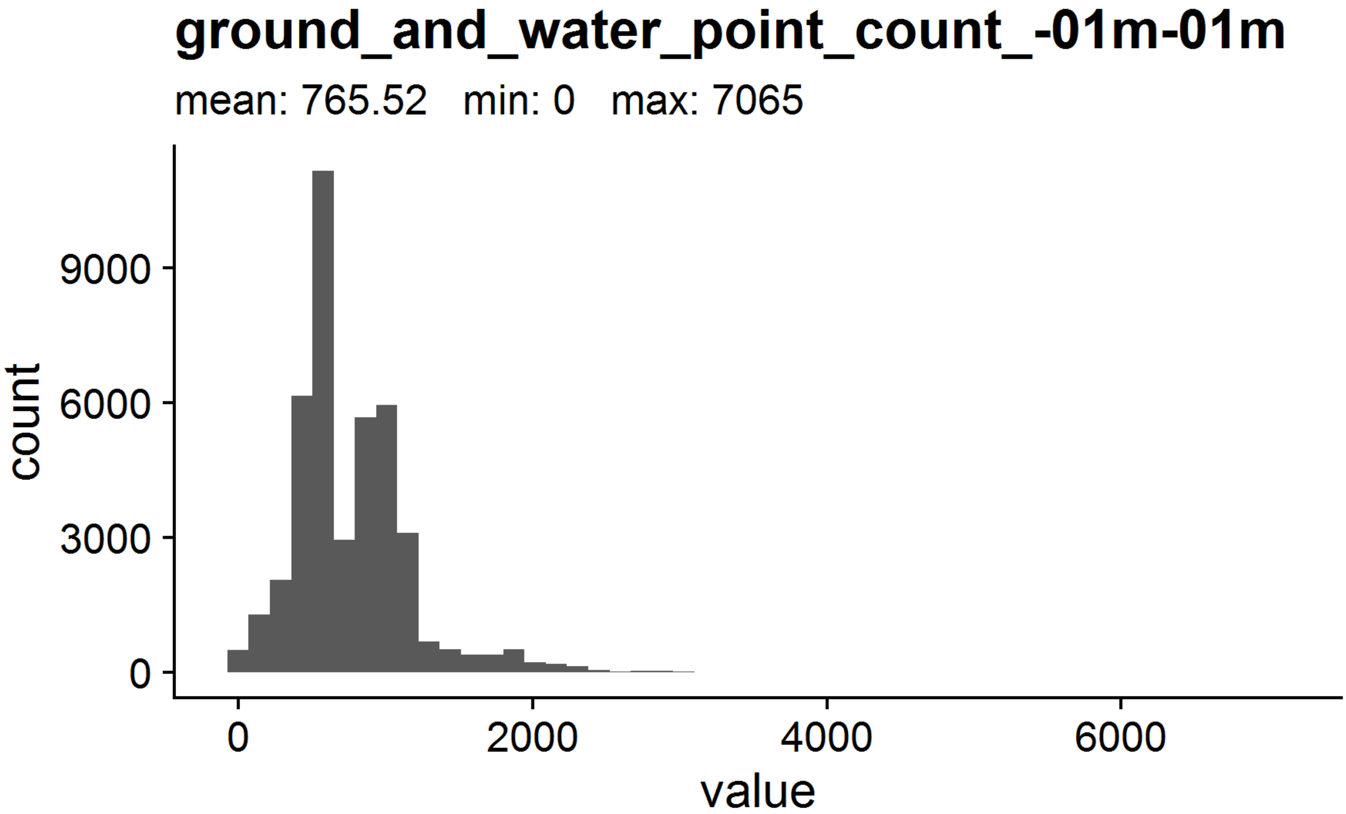
Issues:

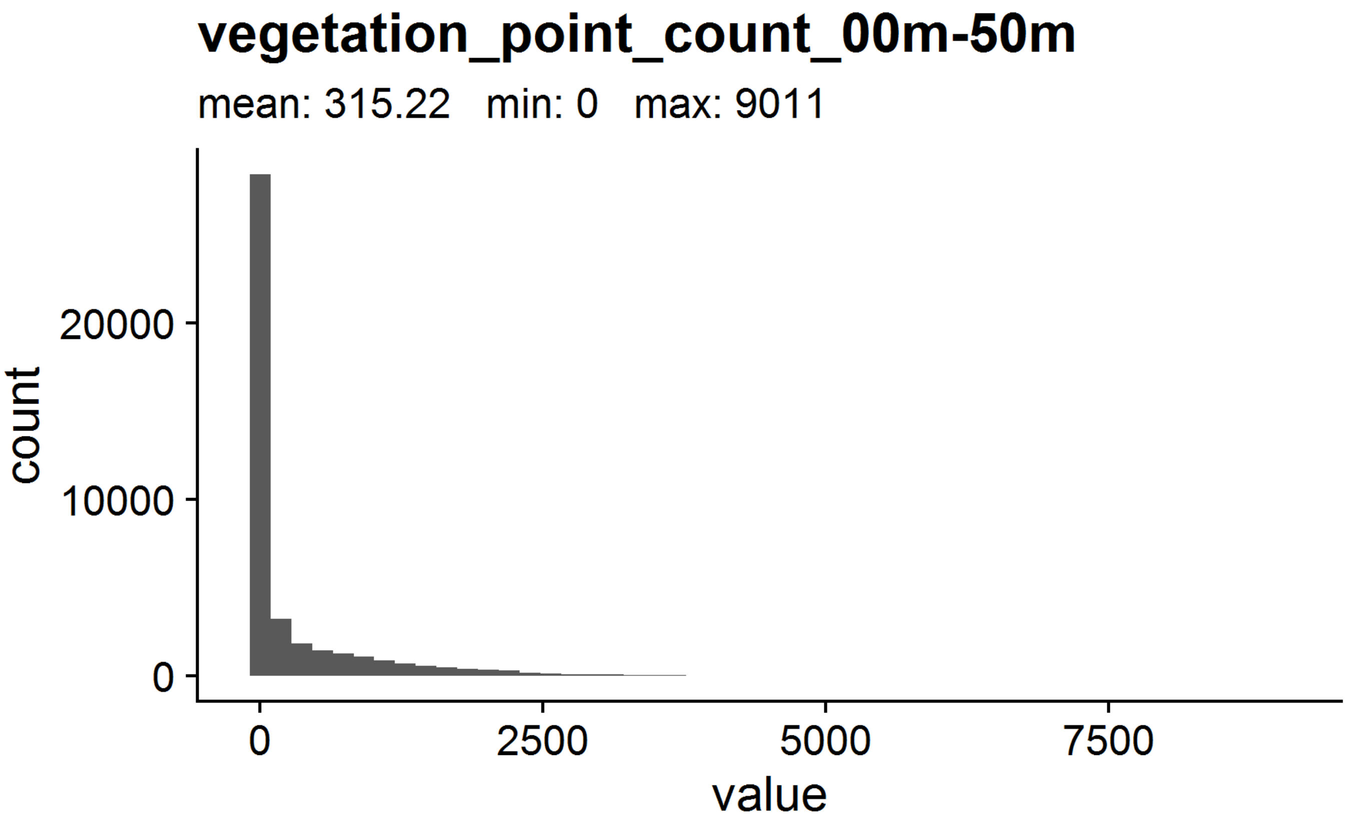
- Water returns only come from shallow water and even these may not be consistent.
- This might introduce inaccuracies and edge effects associated with water bodies.
- Any empty cell (e.g. over deep water) will return zero for all point counts and not NA.

Histogram(s):


ground_point_count_-01m-01m
mean: 765.02 min: 0 max: 7065
9000
6000
3000
0
count
0
2000
4000
6000
value

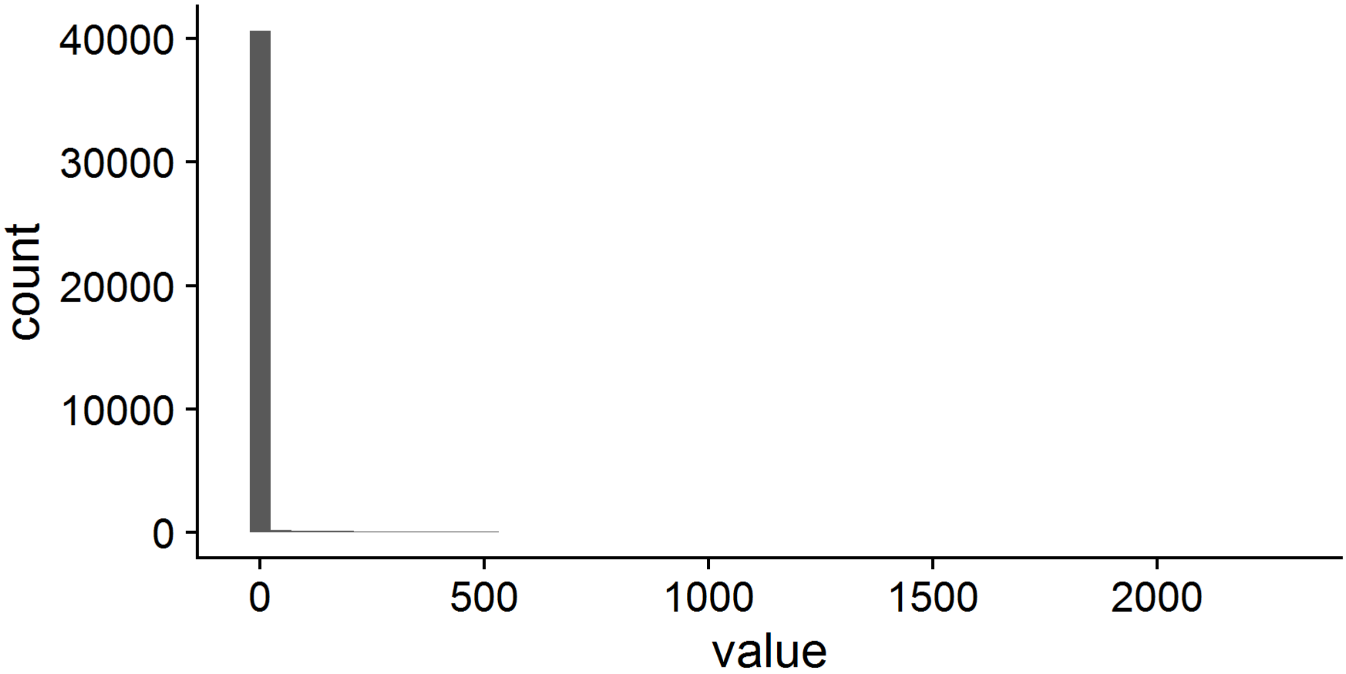

water_point_count_-01m-01m
mean: 0.5 min: 0 max: 1818
40000
30000
20000
10000
0
count
0
500
1000
1500
value


ground_and_water_point_count_-01m-01m
mean: 765.52 min: 0 max: 7065
count
9000
6000
3000
0
0
2000
4000
6000
value


vegetation_point_count_00m-50m
mean: 315.22 min: 0 max: 9011
count
20000
10000
0
0
2500
5000
7500
value

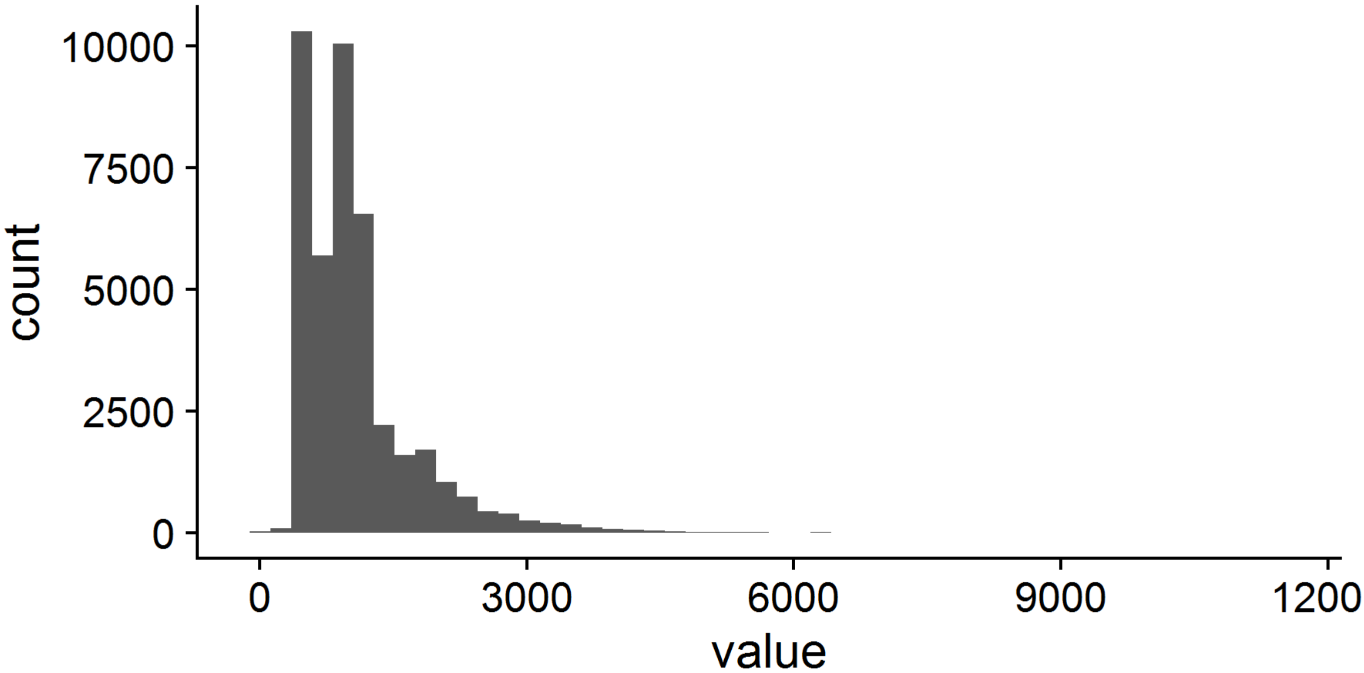
building_point_count_-01m-50m

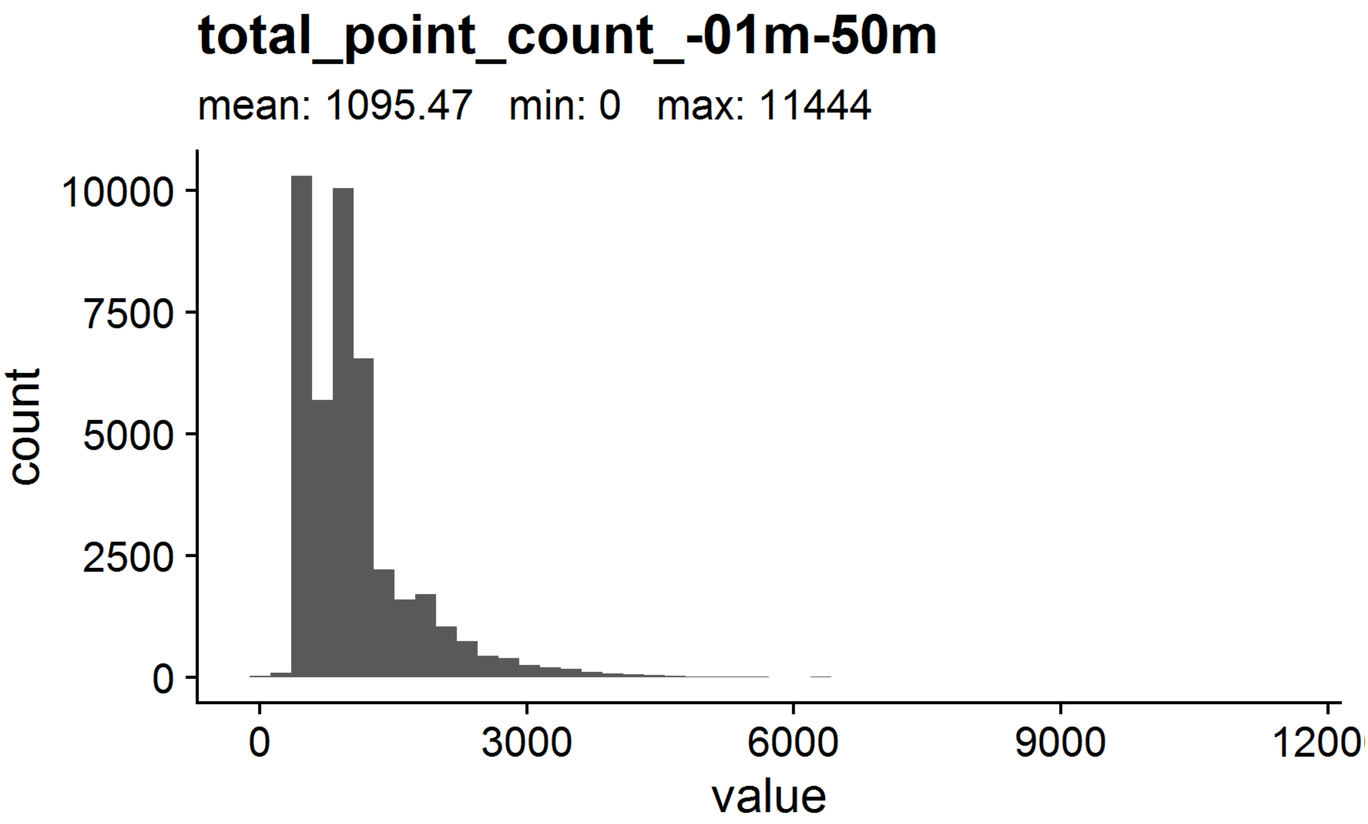
mean: 14.18   min: 0   max: 2272

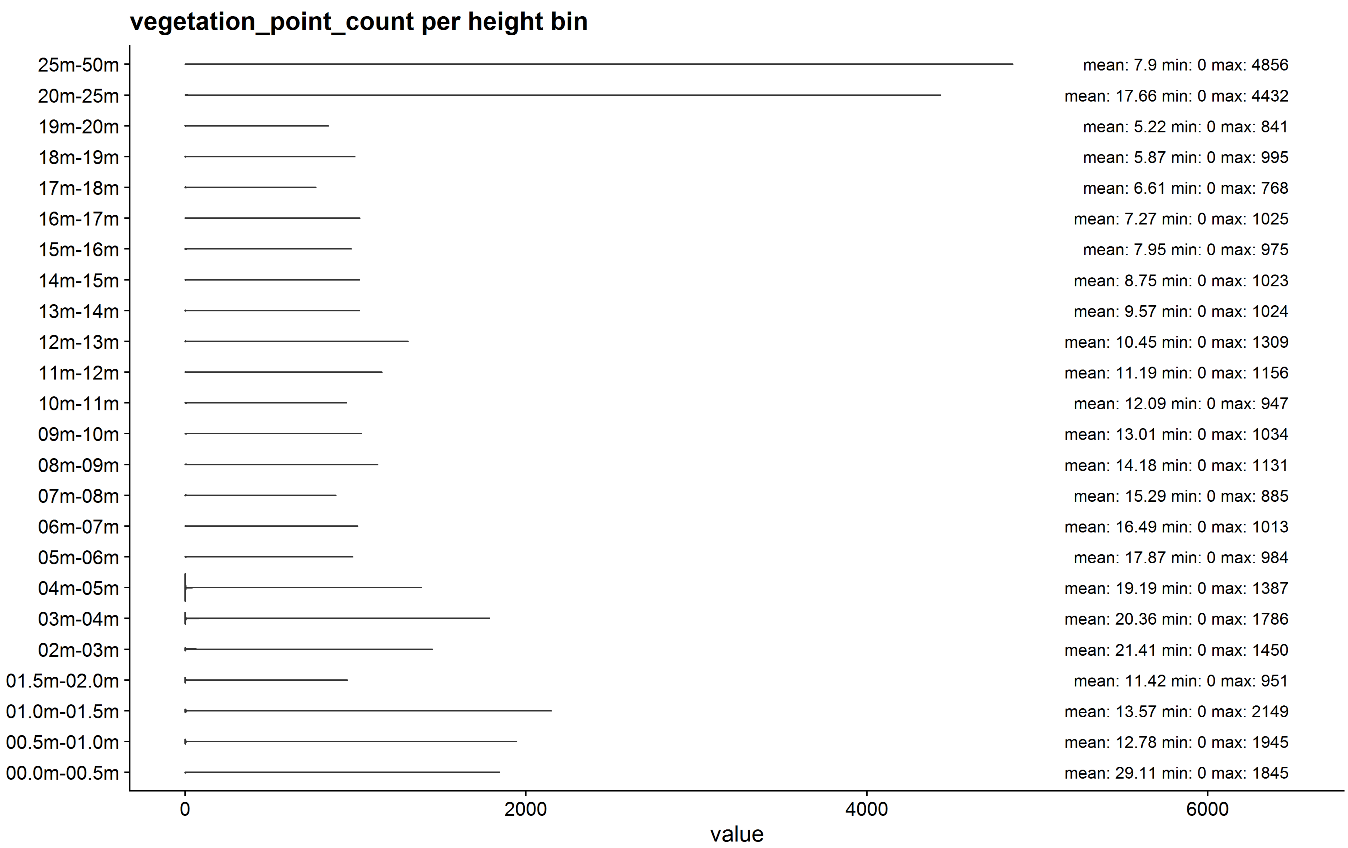


total_point_count_-01m-50m

mean: 1095.47   min: 0   max: 11444







Note: The histograms and violin plots are based on the sample of 50k cell cells shown in Figure 3.

References:

- Point classification outlined in: Geodatastyrelsen 2015. Danmarks Højdemodel, DHM/Punktsky Data version 2.0 - Januar 2015. Accessed online [7 March 2020]. https://kortforsyningen.dk/sites/default/file

s/old_gst/DOKUMENTATION/Data/dk_dhm_punktsky_v2_jan_2015.pdf

Back to content.

---

point_source_info

Folder locations:

- `/outputs/point_source_info/point_source_counts/`
- `/outputs/point_source_info/point_source_ids/`
- `/outputs/point_source_info/point_source_nids`
- `/outputs/point_source_info/point_source_proportion`

File name: `point_source_info_name_xxxx_xxx.tif`

File type and units:

- `point_source_counts - 16 bit integer, count`
- `point_source_ids - 16 bit integer, flight strip ID`
- `point_source_nids - 16 bit integer, count`
- `point_source_proportion - 16 bit integer, ratio x 10000`

Description:

Four descriptor variables for the points sources (flight strip ids) in each 10 m x 10 m cell.

This information may be helpful for interpreting any variable that might be affected by the flight strip id as a covariate. The flight strip id represents: a) differences between sensors / aircraft that may have been used during the nationwide LiDAR campaign, b) differences in acquisition time and date, c) differences in view point / acquisition angle of the cells.

- `point_source_counts` - Contains the number of points per 10 m x 10 m cell for each flight strip id in a tile. In this multi-layer raster each layer represents the point counts for one flight strip id in the tile. The order of layers matches those in the `point_source_ids` raster, which can be used for matching the point counts to the flight strip ids.
- `point_source_ids` - Multi-layer raster containing one layer for each flight strip found in a tile. The presence of a point of the relevant flight strip is indicated by the presence of a string containing the flight strip id in the cell. This layer can be used to match the layers of the `point_source_counts` and `point_source_proportions` layers to a flight strip id.
- `point_source_nids` - Single layer raster containing the number of different point source ids in each cell.
- `point_source_proportions` - Multi-layer raster containing the proportion of point counts for a given point source id per 10 m x 10 m cell. The order of layers corresponds to those in `point_source_ids`, which can be used to match the proportions to a given point source id.

Issues:
No known issues so far.

References:
No relevant references.

Back to content.

proportions

Folder locations: `/outputs/proportion/proportion_name`

File names: `proportion_name_xxxx_xxx.tif`

File type and units: `16-bit integer, ratio x 10000`

Description:

Simple ratios between selected point counts for each 10 m x 10 m cell.

General proportions:

name	ratio
canopy_openness	ground and water points (-1 m to 1 m; classes 2,9) / all points (-1 m to 50 m; classes 2,3,4,5,6,9)
vegetation_density	vegetation points (0 m to 50 m; classes 3,4,5) / all points (-1 m to 50 m; classes 2,3,4,5,6,9)
building_proportion	building points (-1 m to 50 m; class 6) / all points (-1 m to 50 m; classes 2,3,4,5,6,9)

Vegetation proportions for height bins:

These proportions were calculated between the vegetation point count in the respective height bin and the total vegetation point count (0 m to 50 m) in a cell. Vegetation points refer to classes 3, 4 and 5.

name	height range
vegetation_proportion_00.0m-00.5m	0.0 m to 0.5 m
vegetation_proportion_00.5m-01.0m	0.5 m to 1.0 m
vegetation_proportion_01.0m-01.5m	1.0 m to 1.5 m
vegetation_proportion_01.5m-02.0m	1.5 m to 2.0 m
vegetation_proportion_02m-03m	2 m to 3 m
vegetation_proportion_03m-04m	3 m to 4 m
vegetation_proportion_04m-05m	4 m to 5 m
vegetation_proportion_05m-06m	5 m to 6 m
vegetation_proportion_06m-07m	6 m to 7 m
vegetation_proportion_07m-08m	7 m to 8 m
vegetation_proportion_08m-09m	8 m to 9 m
vegetation_proportion_09m-10m	9 m to 10 m
vegetation_proportion_10m-11m	10 m to 11 m
vegetation_proportion_11m-12m	11 m to 12 m
vegetation_proportion_12m-13m	12 m to 13 m
vegetation_proportion_13m-14m	13 m to 14 m
vegetation_proportion_14m-15m	14 m to 14 m
vegetation_proportion_15m-16m	15 m to 16 m
vegetation_proportion_16m-17m	16 m to 17 m
vegetation_proportion_17m-18m	17 m to 18 m
vegetation_proportion_18m-19m	18 m to 19 m
vegetation_proportion_19m-20m	19 m to 20 m
vegetation_proportion_20m-25m	20 m to 25 m
vegetation_proportion_25m-50m	25 m to 50 m

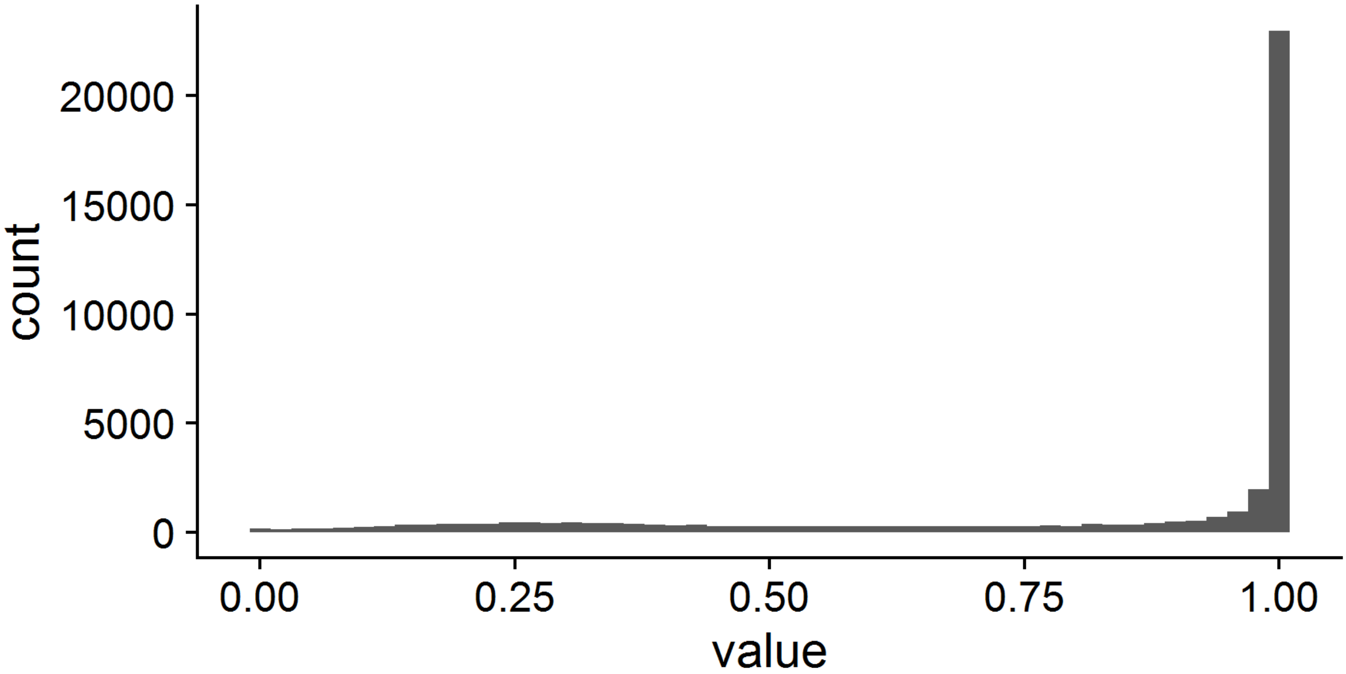
Issues:

- Mathematically invalid divisions may occur (i.e. division by zero) if a cell is empty for the point class in the denominator of the ratio. In this case a value of zero is assigned to the cell and not NA.

Histogram(s):

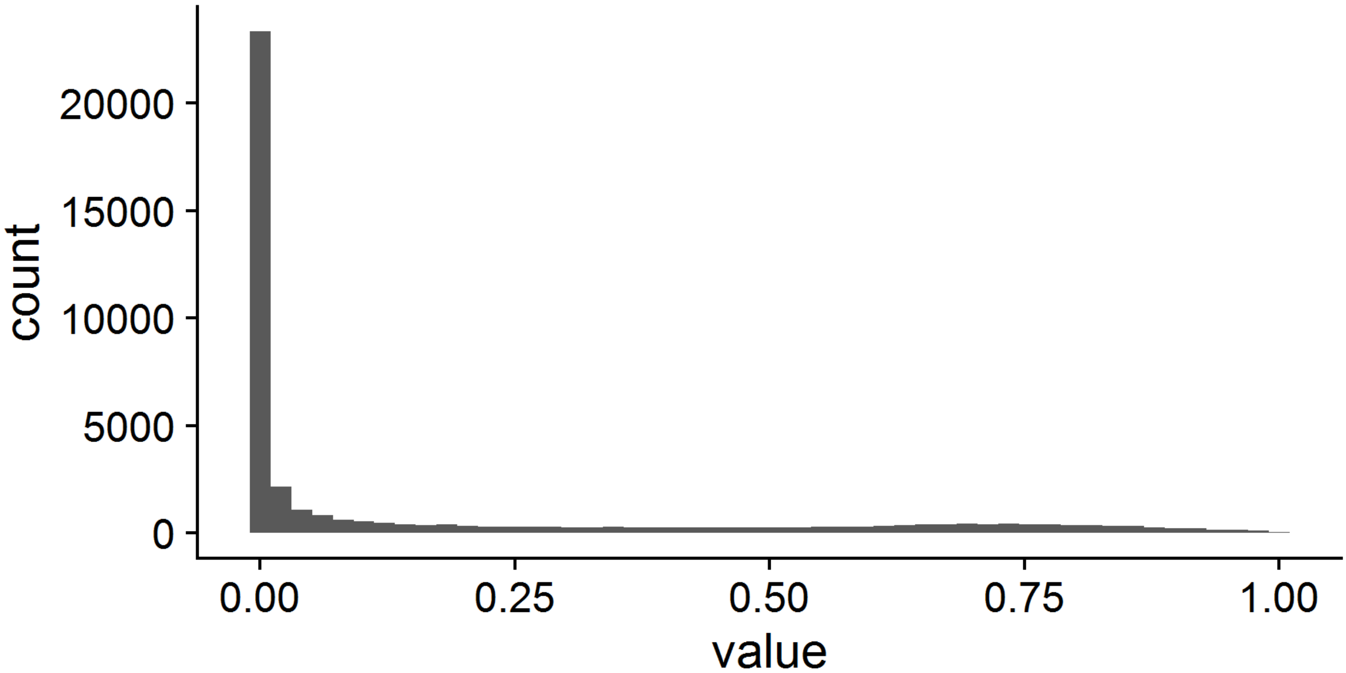
canopy_openness

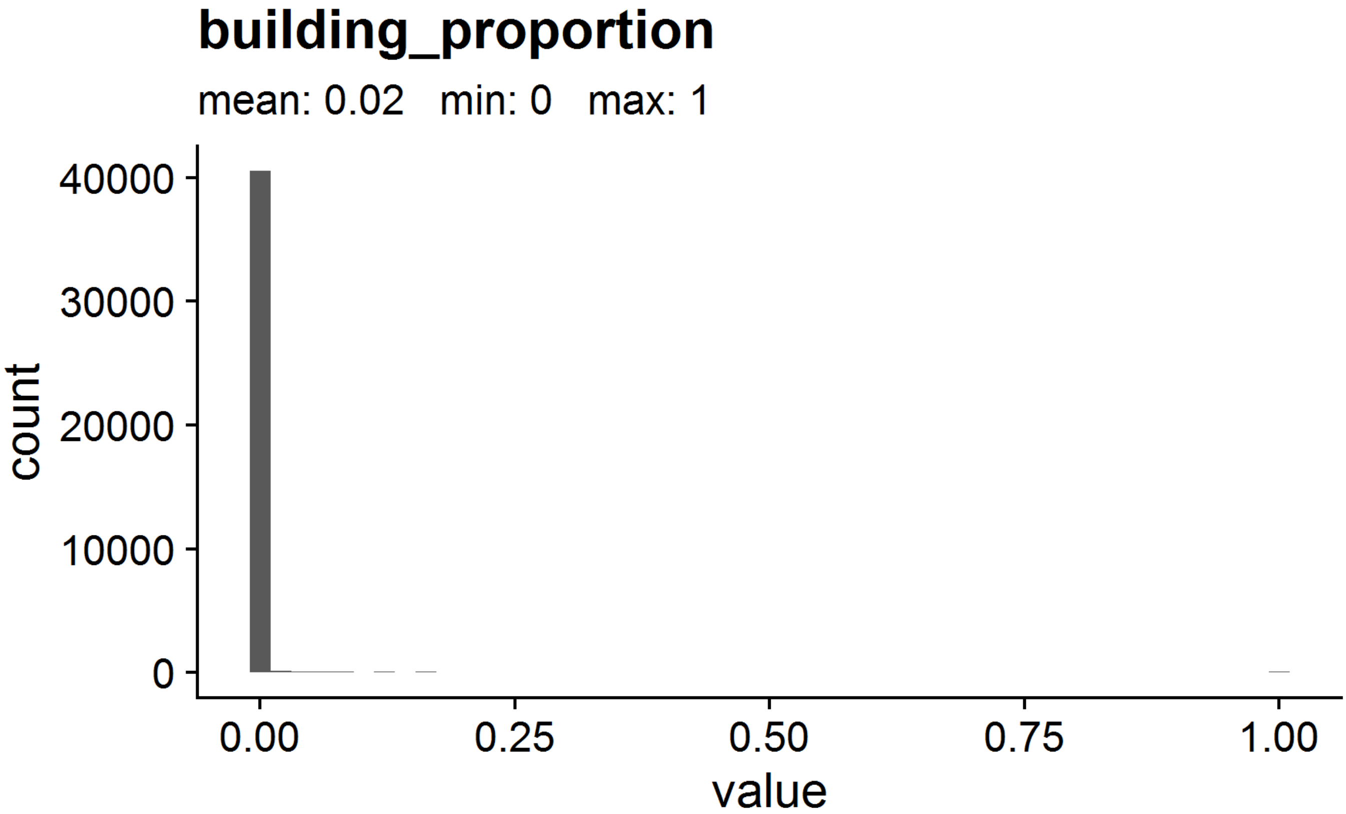
mean: 0.81   min: 0   max: 1

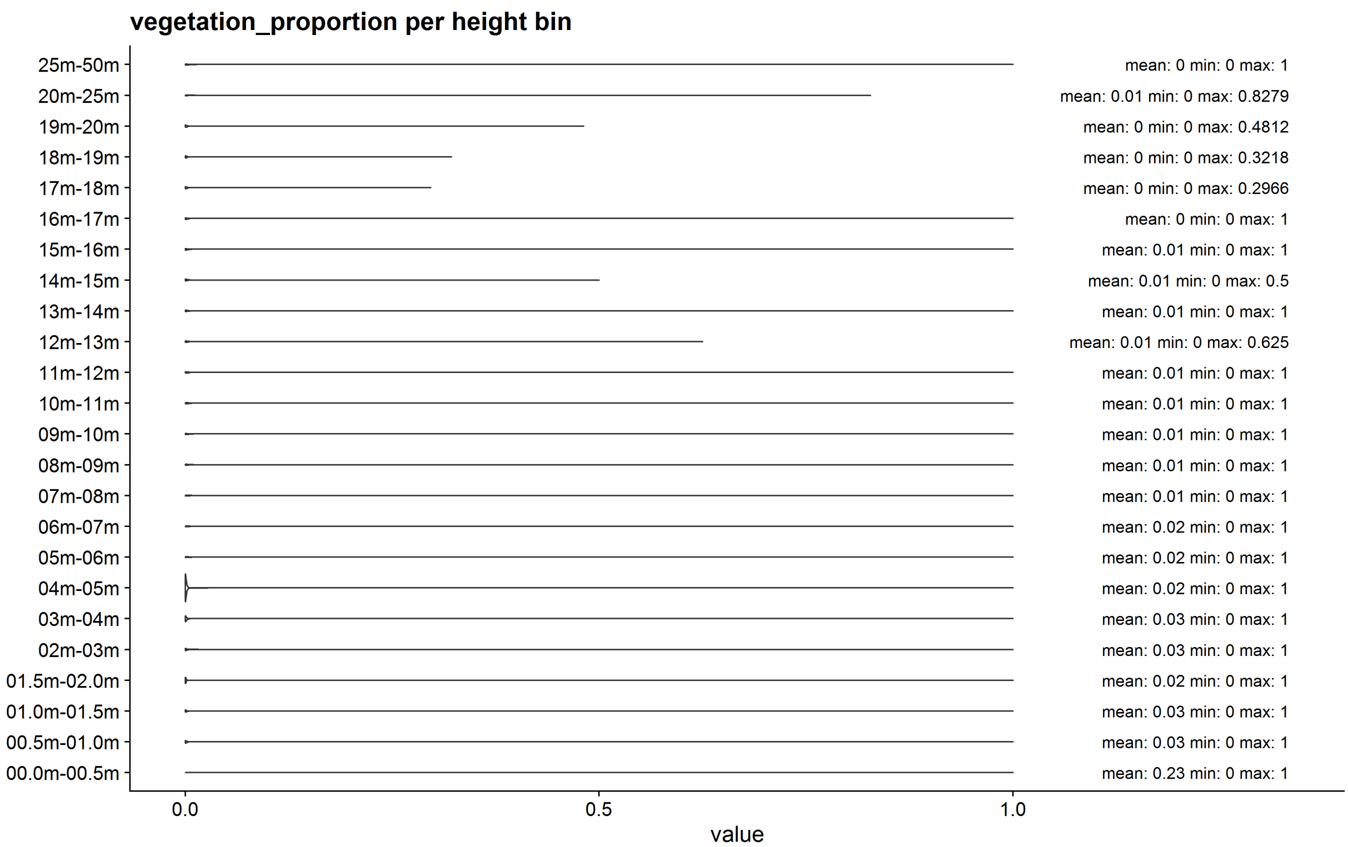


vegetation_density

mean: 0.18   min: 0   max: 1







Note: The histograms and violin plots are based on the sample of 50k cell cells shown in Figure 3.

References:

- van Leeuwen, M., Nieuwenhuis, M., 2010. Retrieval of forest structural parameters using LiDAR remote sensing. Eur J Forest Res 129, 749–770. https://doi.org/10.1007/s10342-010-0381-4
- Sasaki, T., Imanishi, J., Ioki, K., Morimoto, Y., Kitada, K., 2008. Estimation of leaf area index and canopy

openness in broad-leaved forest using an airborne laser scanner in comparison with high-resolution near-infrared digital photography. Landscape Ecol Eng 4, 47–55. https://doi.org/10.1007/s11355-008-0041-8

- Sasaki, T., Imanishi, J., Ioki, K., Song, Y., Morimoto, Y., 2016. Estimation of leaf area index and gap fraction in two broad-leaved forests by using small-footprint airborne LiDAR. Landscape Ecol Eng 12, 117–127. https://doi.org/10.1007/s11355-013-0222-y
- Melin, M., Hinsley, S.A., Broughton, R.K., Bellamy, P., Hill, R.A., 2018. Living on the edge: utilising lidar data to assess the importance of vegetation structure for avian diversity in fragmented woodlands and their edges. Landscape Ecol 33, 895–910. https://doi.org/10.1007/s10980-018-0639-7

Back to content.

Terrain model derived variables

aspect

Folder location: `/outputs/aspect`

File name: `aspect_xxxx_xxx.tif`

File type and units: `16-bit integer, degrees x 10`

Description:

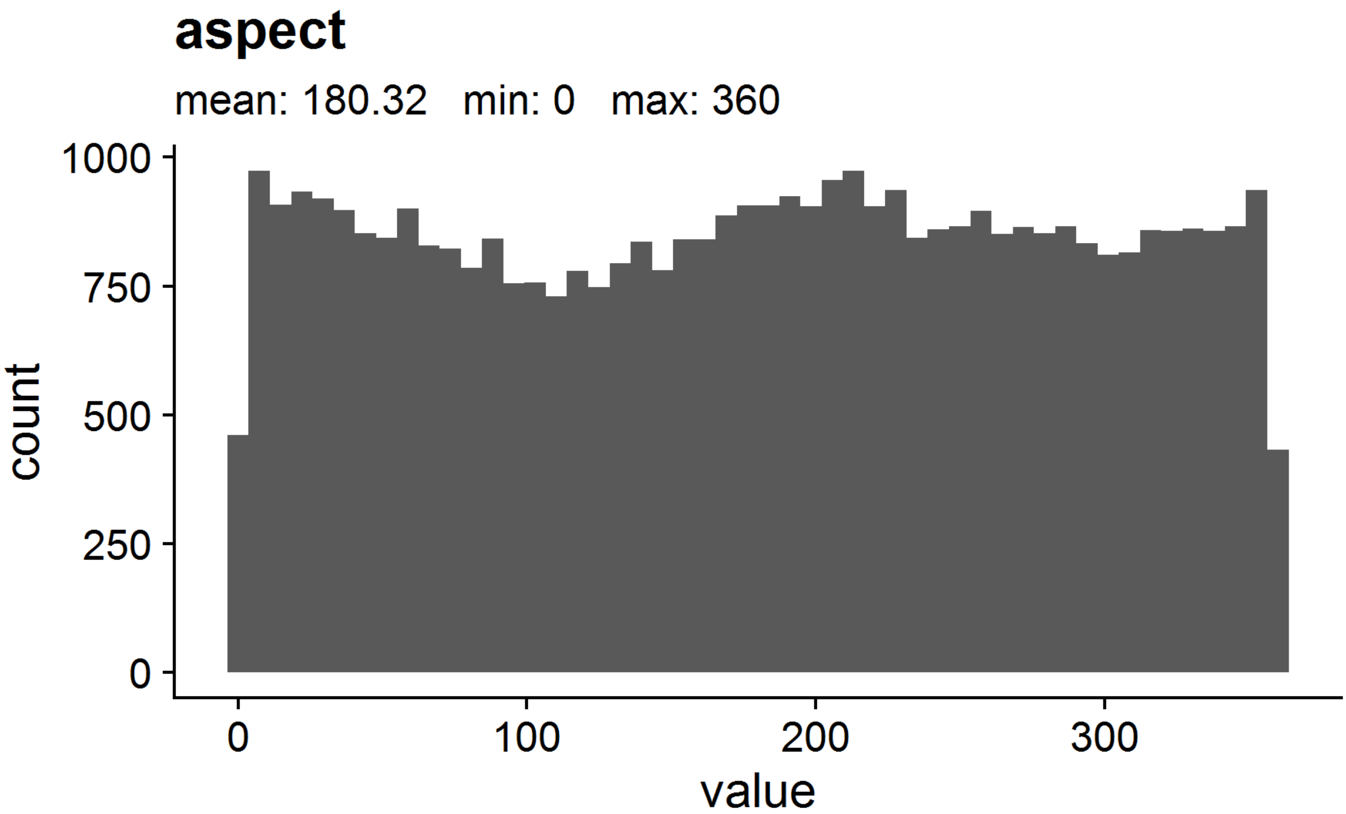
Aspect in degrees (stretched by a factor of 10) with 0° indicating North, 90° East, 180° South and 270° West. Flat areas were assigned an aspect of -1°. Values represent the aspect derived from a 10 m x 10 m aggregate of the DTM.

Calculated using `gdaldem aspect` (Horn's method) from the aggregated 10 m DTM rasters. To avoid edge effects, all calculations were done on a mosaic including the target tile and all available directly neighbouring tiles (maximum 8). The final value of each cell was converted from radian to degrees, stretched by a factor of 10 and rounded to the nearest integer. This results in a precision of the outputs of 0.1 degrees.

Issues:

- Should a neighbourhood mosaic be incomplete (i.e. less than eight neighbouring tiles), the aspect is not defined for the outermost cells of the focal tile along the incomplete edges. The nodata value (-9999) has been assigned to these cells.

Histogram:



Note: The histograms is based on the sample of 50k cell cells shown in Figure 3.

References:
No relevant references.

Back to content.

---

dtm_10m

Folder location: `/outputs/dtm_10m`

File name: `dtm_10m_xxxx_xxx.tif`

File type and units: `16-bit integer, metres x 100`

Description:

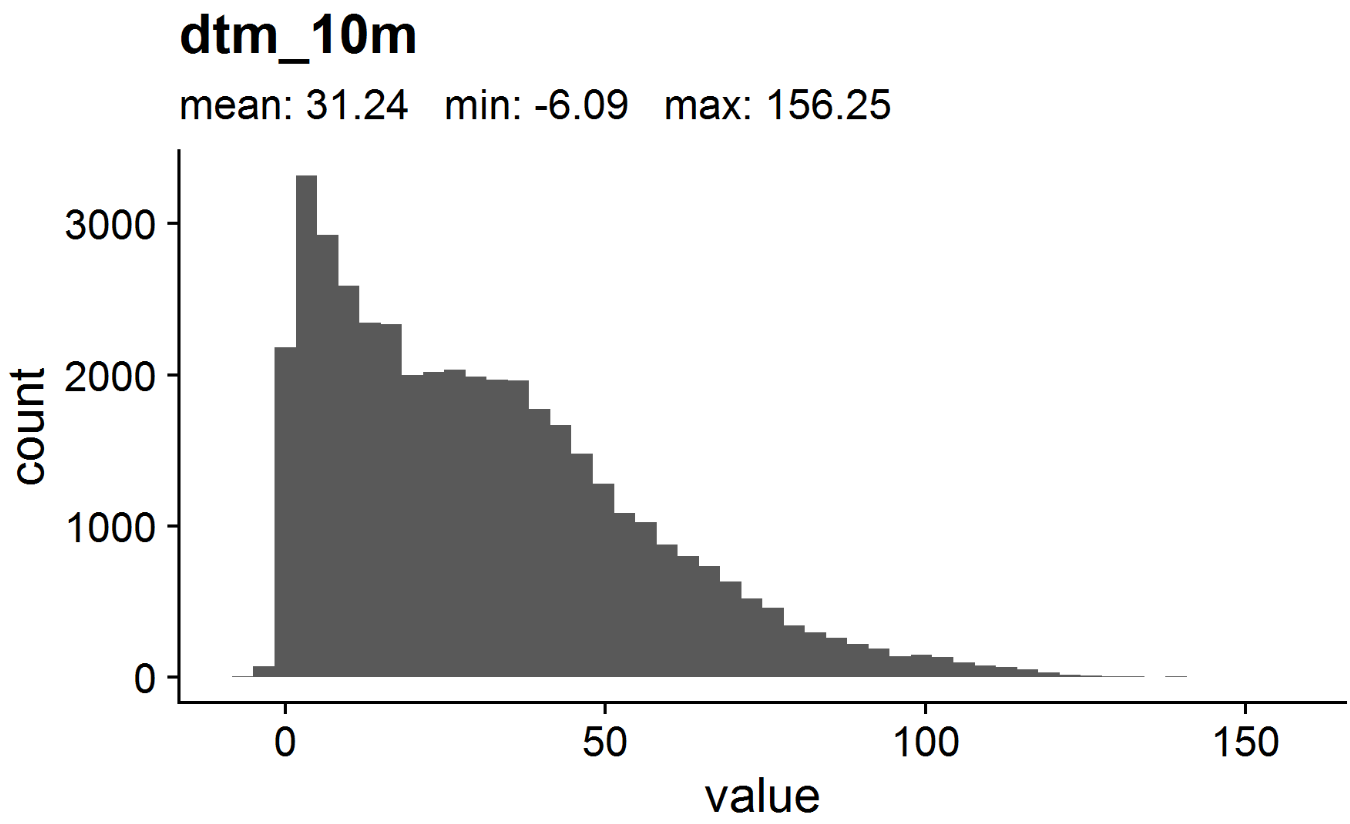
Digital Terrain Model (DTM) tiles of Denmark with a grain size of 10 m x 10 m describing the altitude above sea level for a given cell. The 10 m rasters are mean aggregates of the 0.4 m DTM provided by Kortforsyningen.

The outputs were stretched by a factor of 100 and stored as a 16-bit integer. The output values are therefore in cm - divide by 100 to convert back to metres.

Issues:

- No known issues.

Histogram:



Note: The histograms is based on the sample of 50k cell cells shown in Figure 3.

References:
No relevant references.

[Back to content.](#)

---

heat_load_index

Folder location: `/outputs/heat_load_index`

File name: `heat_load_index_xxxx_xxx.tif`

File type and units: `16-bit integer, unitless x 10000`

Description:

Heat load index calculated following McCune and Keon (2002). Index purely based on the aspect of a cell, ranging from zero (North slopes) to 1 (South slopes).

Note: The index *is not* the most meaningful measure of terrain-derived energy influx from the sun, please use the `solar_radiation` variable as an indicator for this (Figure below). Instead this variable seems to be a good indicator for soil moisture conditions (see Moeslund et al. 2019). The name of the variable is simply kept for legacy reasons.

Calculated from the 10 m aspect rasters following the equation specified in McCune and Keon (2002):

`heat_load_index = (1 - cos((radians(A)-45)))/2` where `A` is the aspect in degrees. The value was then stretched by a factor of 10000, rounded to the nearest integer and converted into a 16 bit integer. Please note that the input precision of the 10 m aspect raster is 0.1 degrees (see aspect).

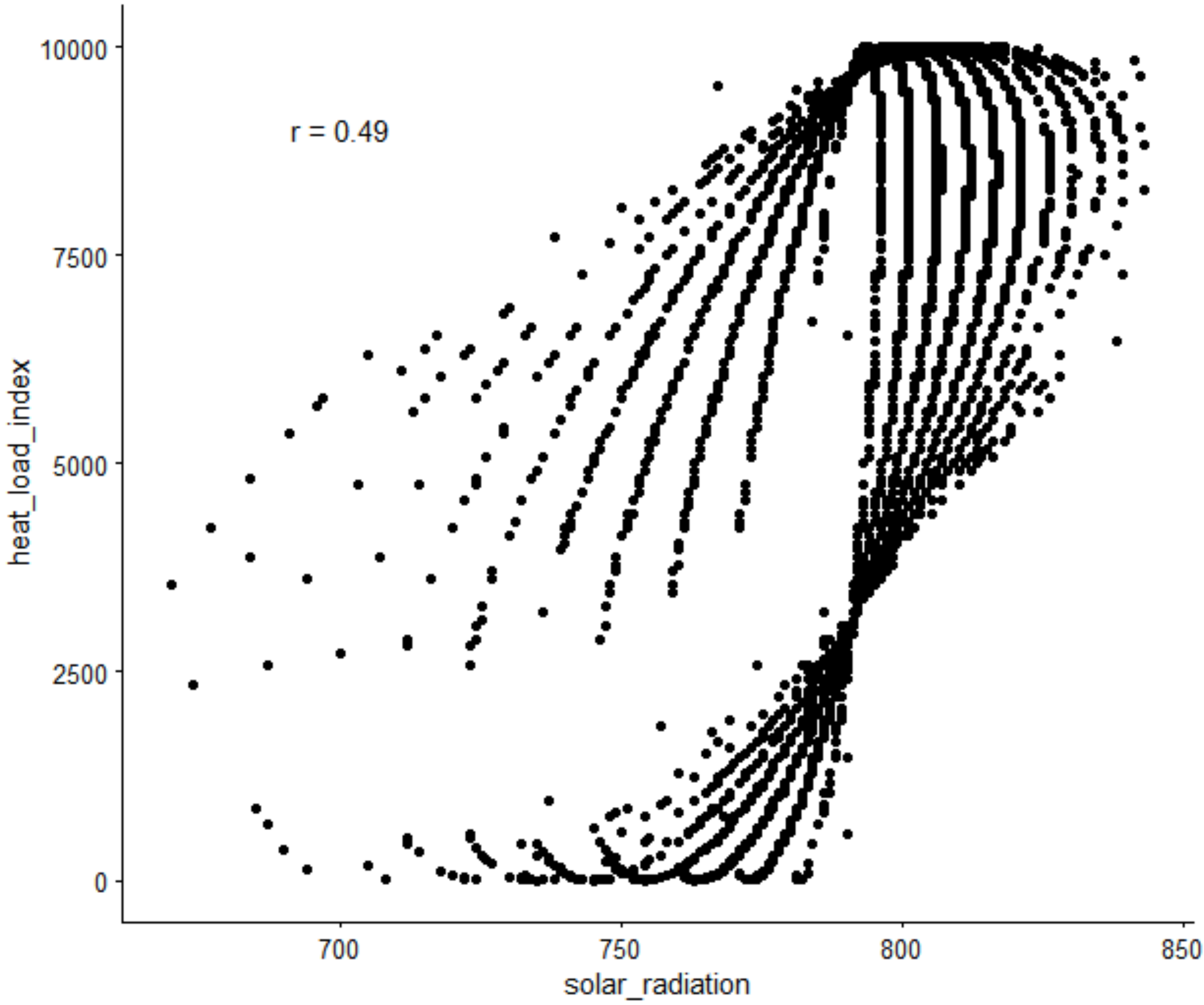
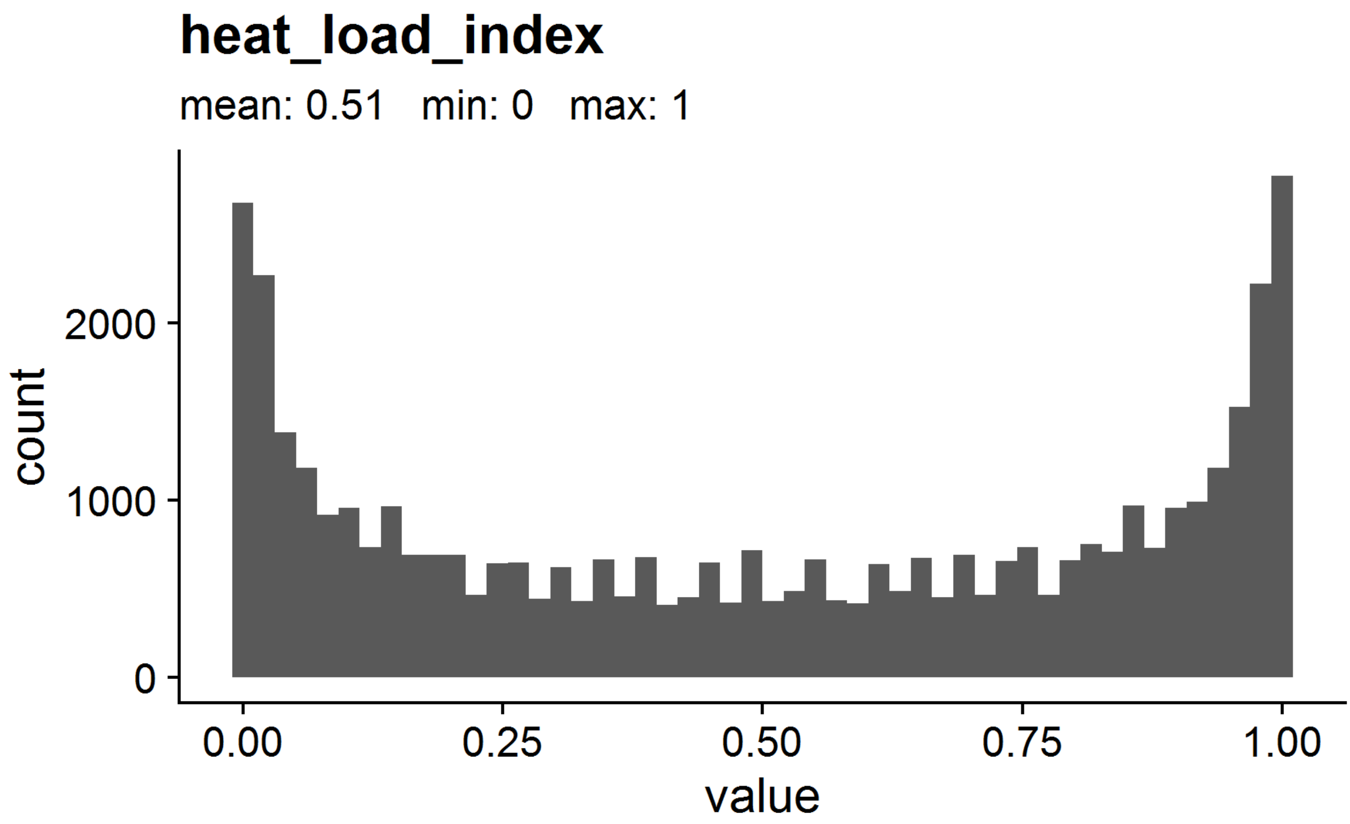


Figure 6: Illustrating the correlation between solar radiation and heat load index, both variables are moderately correlated (r = 0.49), but the solar radiation value seems to contain more information and is deemed better by the McCune and Keon (2002).

Issues:

- The index is not the best indicator for terrain-derived energy influx from the sun, please use the `solar_radiation` variable for this instead.
- The heat load index is not defined where the aspect is not defined or where the slope is zero (aspect = -1). The nodata value (-9999) has been assigned to all cells meeting either condition.

Histogram:



Note: The histograms is based on the sample of 50k cell cells shown in Figure 3.

References:

- McCune, B., Keon, D., 2002. Equations for potential annual direct incident radiation and heat load. Journal of Vegetation Science 13, 603–606. https://doi.org/10.1111/j.1654-1103.2002.tb02087.x
- Moeslund, J.E., Zlinszky, A., Ejrnæs, R., Brunbjerg, A.K., Bøcher, P.K., Svenning, J.-C., Normand, S., 2019. Light detection and ranging explains diversity of plants, fungi, lichens, and bryophytes across multiple habitats and large geographic extent. Ecological Applications 29, e01907. https://doi.org/10.1002/eap.1907

Back to content.

---

openness_mean

Folder location: `/outputs/openness_mean`

File name: `openness_mean_xxxx_xxx.tif`

File type and units: `16-bit integer, degrees`

Description:

Landscape openness calculated following Yokoyama et al. 2002 using the OPALS Openness module and a search radius of 150 m. Landscape openness is a landform descriptor that indicates whether a cell is located in a valley, depression or on a ridge.

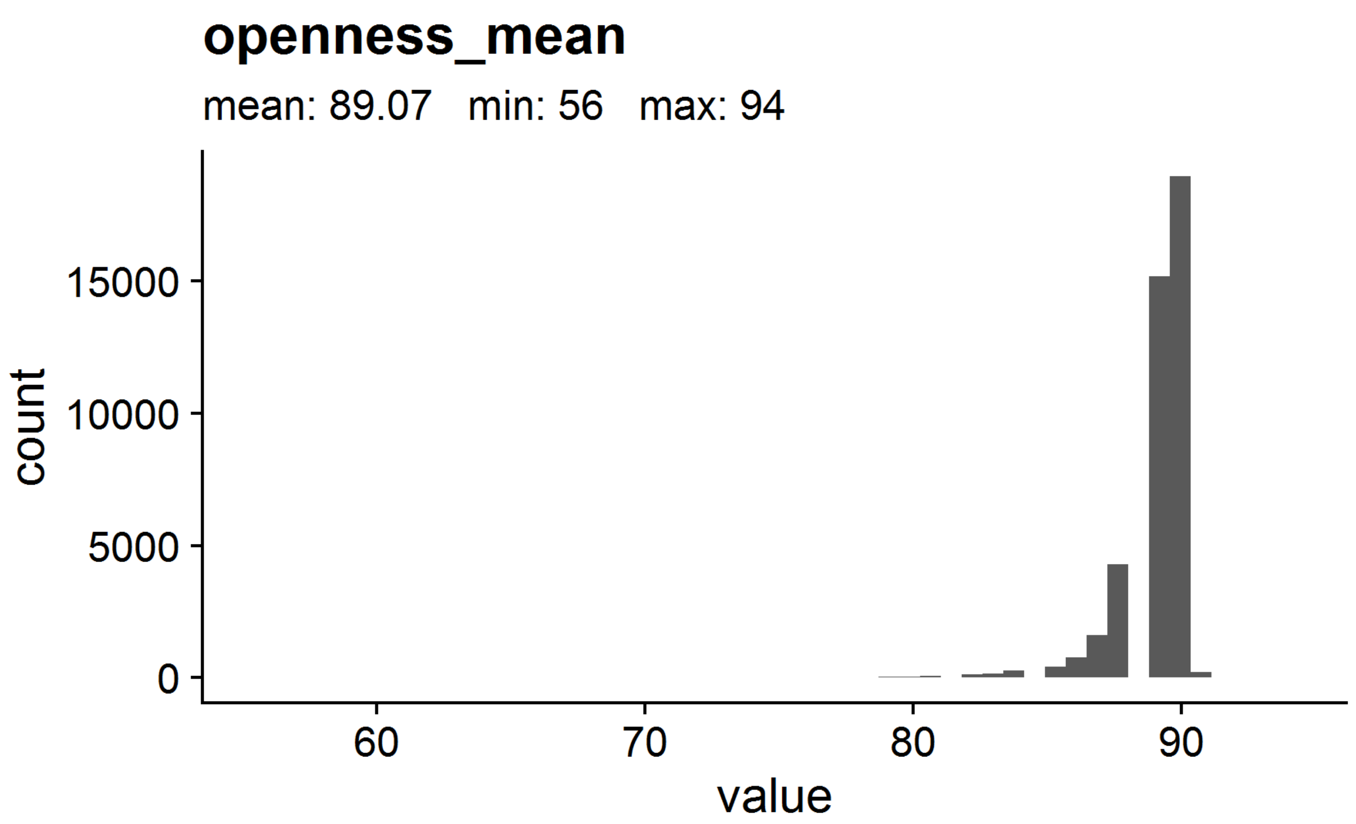
First, the 0.4 m DTM was aggregated to a grain size of 10 m. To reduce edge effects in subsequent calculations, this aggregation was carried out for a mosaic including the target tile and all available tiles in the direct neighbourhood (max. eight neighbouring tiles). The mean of the positive openness for all eight cardinal directions with search radius of 150 m was then calculated for all cells in the tile mosaic using the OPALS Openness module (feature = 'positive', kernelSize = 15 and selMode = 0). The output was cropped to the extent of the target tile.

If the neighbourhood mosaic was incomplete, i.e. contained less than eight neighbouring tiles, cells within the first 150 m of all edges where a neighbourhood tile was missing were masked out (set to NA). Finally, the mean openness per cell was converted from radians to degrees and rounded to the nearest full degree.

Issues:

- Cells with incomplete neighbourhoods will have NA values assigned for the first 15 cells (150 m) on the borders with missing neighbours.

Histogram:



Note: The histograms is based on the sample of 50k cell cells shown in Figure 3.

References:

- Yokoyama, R. / Shirasawa, M. / Pike, R.J. (2002): Visualizing topography by openness: A new application of image processing to digital elevation models. Photogrammetric Engineering and Remote Sensing, Vol.68, pp.251-266.

Back to content.

openness_difference

Folder location: `/outputs/openness_difference`

File name: `openness_difference_xxxx_xxx.tif`

File type and units: `16-bit integer, degrees`

Description:

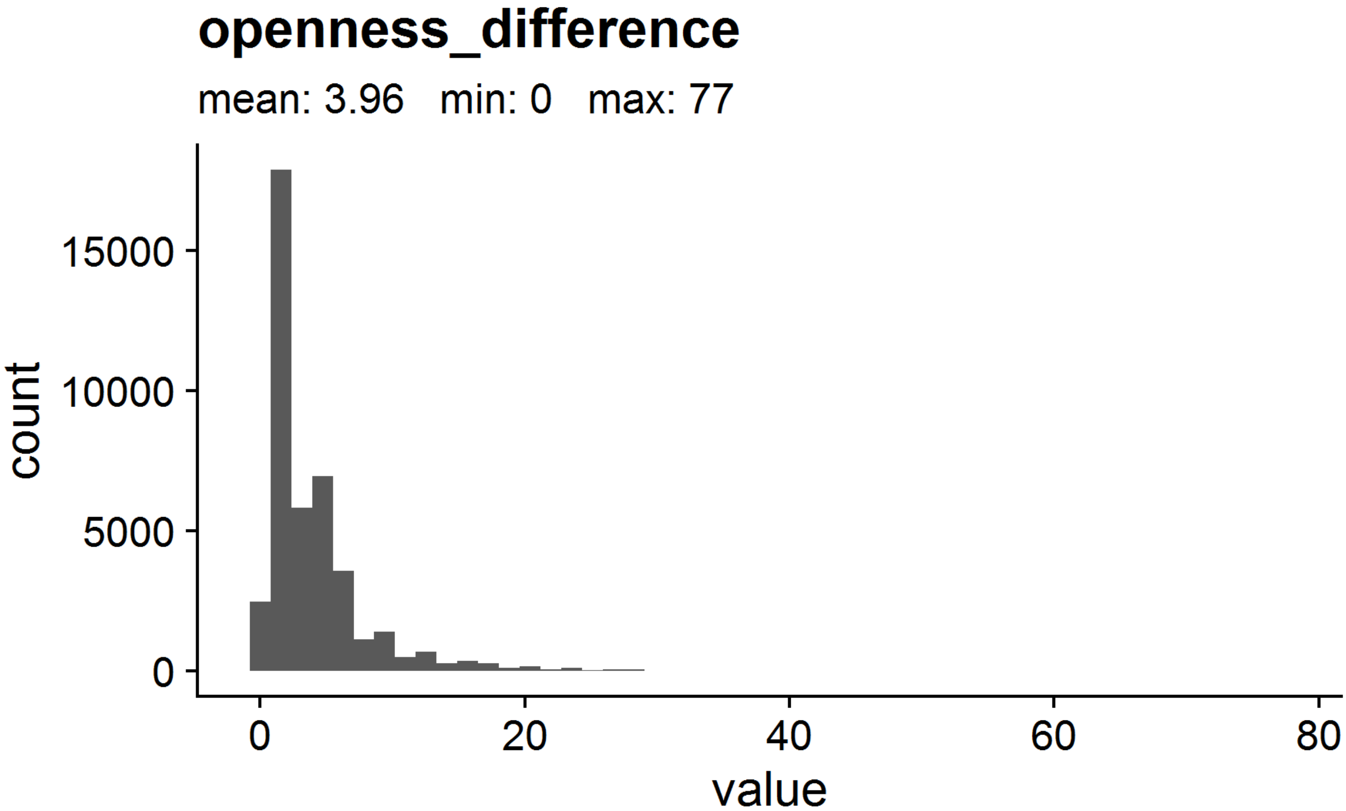
Min/max difference in landscape openness based on Yokoyama et al. 2002 calculated using the OPALS Openness module and a search radius of 50 m. Indicates presence / absence of linear landscape features.

First, the 0.4 m DTM was aggregated to a grain size of 10 m. To reduce edge effects in subsequent calculations, this aggregation was carried out for a mosaic including the target tile and all available tiles in the direct neighbourhood (max. eight neighbouring tiles). The min and max of the positive openness for all eight cardinal directions within a search radius of 50 m were then calculated for all cells in the tile mosaic using the [OPALS Openness module](#) (feature = 'positive', kernelSize = 5 and selMode = 1/2). Next, the min & max values were converted from radians to degrees, the difference was calculated and the result rounded to the nearest full degree. The output was cropped to the extent of the target tile. If the neighbourhood mosaic was incomplete, i.e. contained less than eight neighbouring tiles, cells within the first 50 m of all edges where a neighbourhood tile was missing were masked out (set to NA).

Issues:

- Cells with incomplete neighbourhoods will have NA values assigned for the first 5 cells (50 m) on the borders with missing neighbours.

Histogram:



Note: The histograms is based on the sample of 50k cell cells shown in Figure 3.

References:

- Yokoyama, R. / Shirasawa, M. / Pike, R.J. (2002): Visualizing topography by openness: A new application of image processing to digital elevation models. Photogrammetric Engineering and Remote Sensing, Vol.68, pp.251-266.

[Back to content.]

---

slope

Folder location: `/outputs/slope`

File name: `slope_xxxx_xxx.tif`

File type and units: `16-bit integer, degrees x 10`

Description:

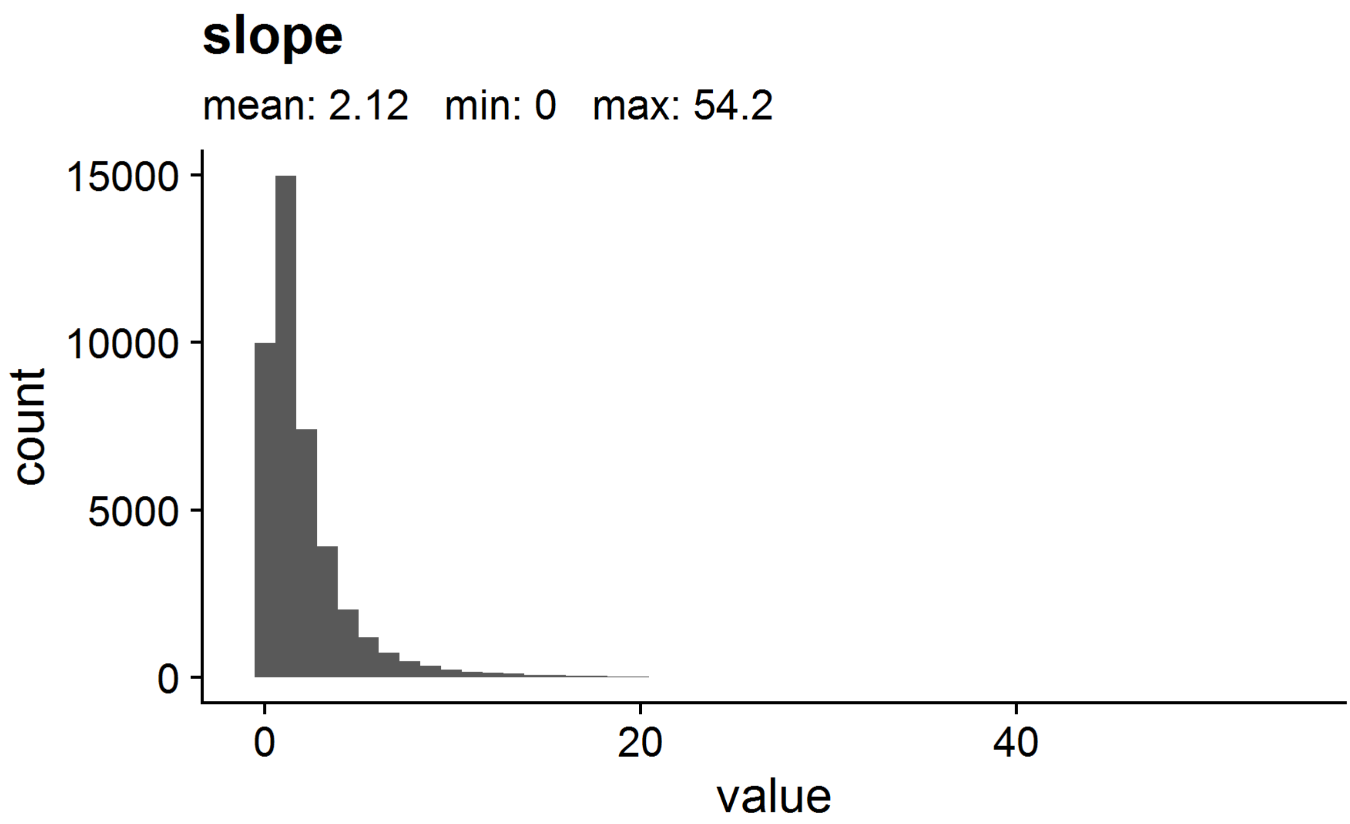
Slope in degrees at 10 m grain size derived from the 10 m DTM.

Calculated using `gdaldem slope` (Horn's method) on an aggregated 10 m grain size DTM raster. To avoid edge effects all calculations were done on a mosaic including the target tile and all available directly neighbouring tiles (maximum eight). The outputs values were stretched by a factor of 10 and then rounded to the nearest integer, giving the slope values a precision of 0.1 degrees.

Issues:

- Should a neighbourhood mosaic be incomplete (i.e. less than eight neighbouring tiles), the slope for the outermost cells of the focal tile along the respective edges is not defined. For these cells the slope value is set to nodata (-9999).

Histogram:



Note: The histograms is based on the sample of 50k cell cells shown in Figure 3.

References:
No relevant references.

Back to content.

---

solar_radiation

Folder location: `/outputs/solar_radiation`

File name: `solar_radiation_xxxx_xxx.tif`

File type and units: `16-bit integer, ln(MJ x cm^-2 x yr^-1) x 1000`

Description:

Incident solar radiation estimated following McCune and Keon (2002).

Calculated from the 10 m aspect and slope rasters using equation 3 specified by McCune and Keon (2002):

```
solar_radiation = 0.339 + 0.808 x cos(radians(L)) x cos(radians((S/10))) - 0.196 x sin(radians(L)) x sin(radians(S/10)) - 0.482 * cos(radians(180 - absolute(180 - (A/10)))) x sin(radians((S/10)))))
```

where `L` is the centre latitude of the cell in degrees, `S` is the slope of the cell in degrees and `A` is the aspect of the cell in degrees. The division of aspect and slope by a factor of 10 is required due to the stretched nature of the variables in the aspect and slope rasters. Please note that aspect and slope have an accuracy of 0.1 degrees. The final value was stretched by a factor of 1000, rounded to the nearest integer and converted into a 16 bit integer.

Additional Information:

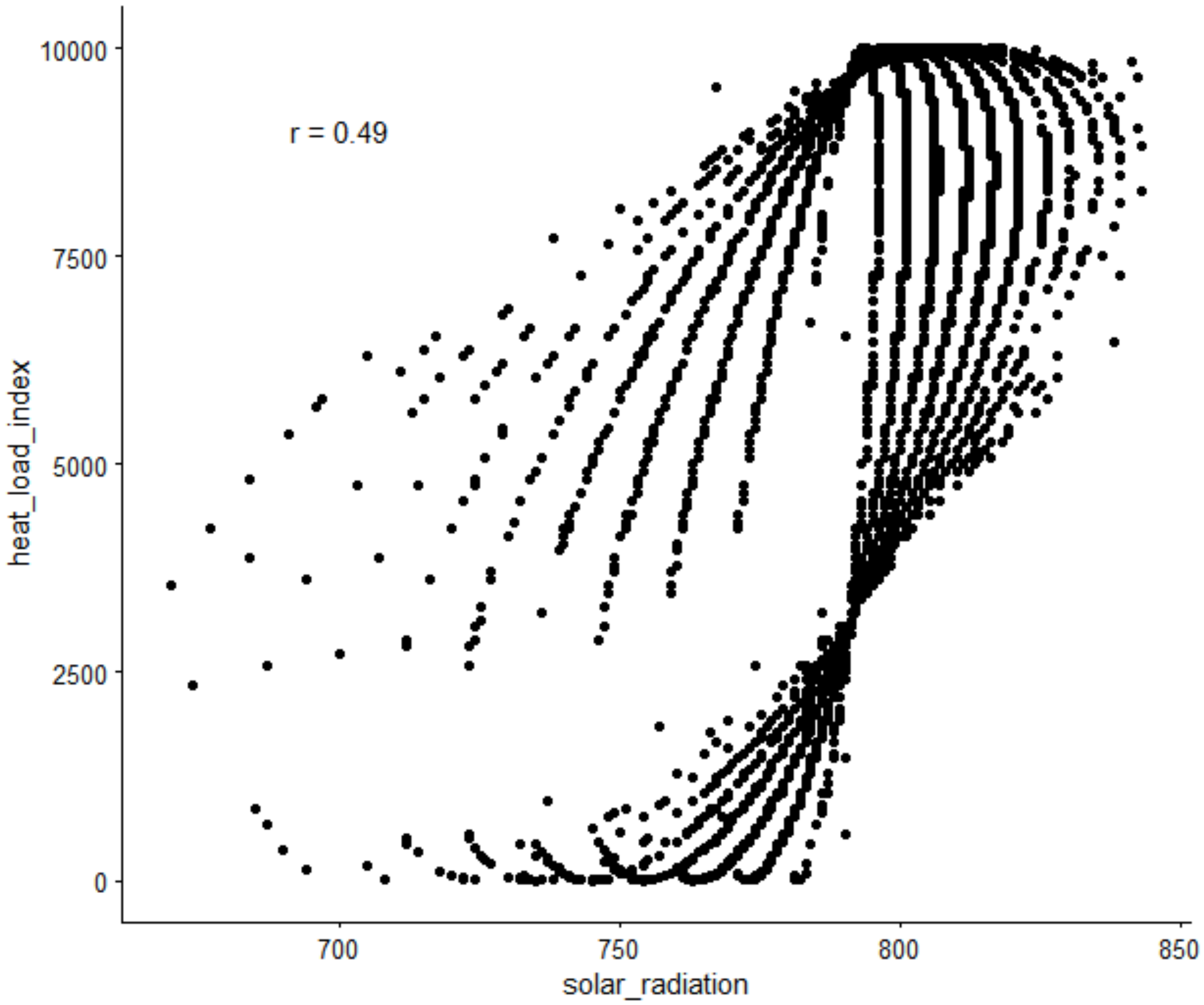
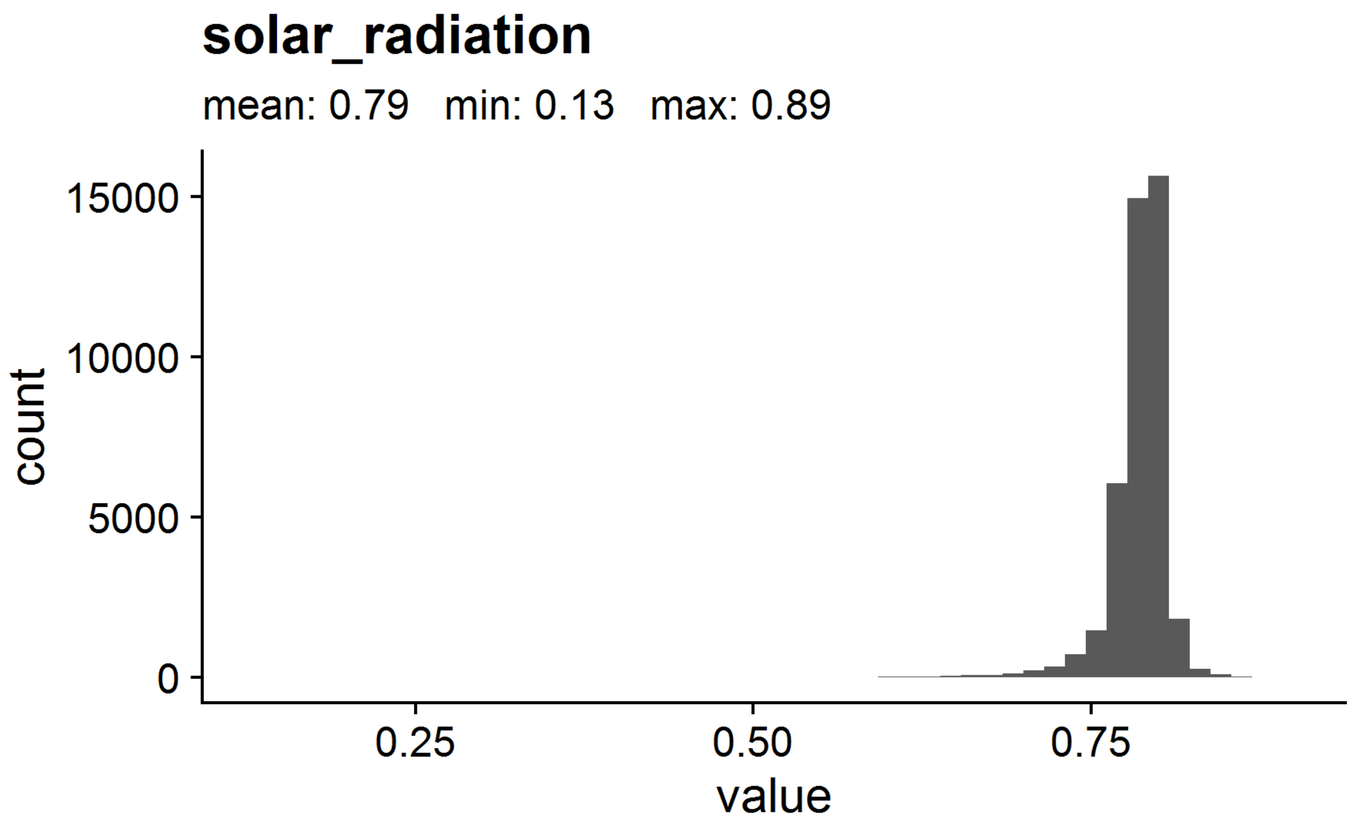


Figure 6: Illustrating the correlation between solar radiation and heat load index, both variables are moderately correlated (r = 0.49), but the solar radiation value seems to contain more information and is deemed better by the authors.

Issues:

- No data values are propagated from the aspect and slope rasters.

Histogram:



Note: The histograms is based on the sample of 50k cell cells shown in Figure 3.

References:

- McCune, B., Keon, D., 2002. Equations for potential annual direct incident radiation and heat load. Journal of Vegetation Science 13, 603–606. https://doi.org/10.1111/j.1654-1103.2002.tb02087.x

[Back to content.]()

---

twi

Folder location: `/outputs/twi/`

File name: `twi_xxxx_xxx.tif`

File type and units: `16 bit integer, unitless index x 1000`

Description:

Topographic wetness index (TWI) calculated following Kopecký et al. 2020. Calculations were done on the aggregated 10 m DTM neighbourhood mosaic (max. 8 neighbours). As such the index value calculated here only considers a catchment the size of one tile and all its neighbours (for non-edge tiles this is a 3 km x 3 km catchment, for edge tiles it is smaller). The resulting output was then cropped to the target tile, stretched by a factor of 1000 and rounded to the next full integer. Calculation were done using SAGA GIS v. 7.8.2. A workflow procedures is provided below. See the respective pages in the [SAGA GIS v. 7.8.2 documentation]() for a detailed description of the modules used.

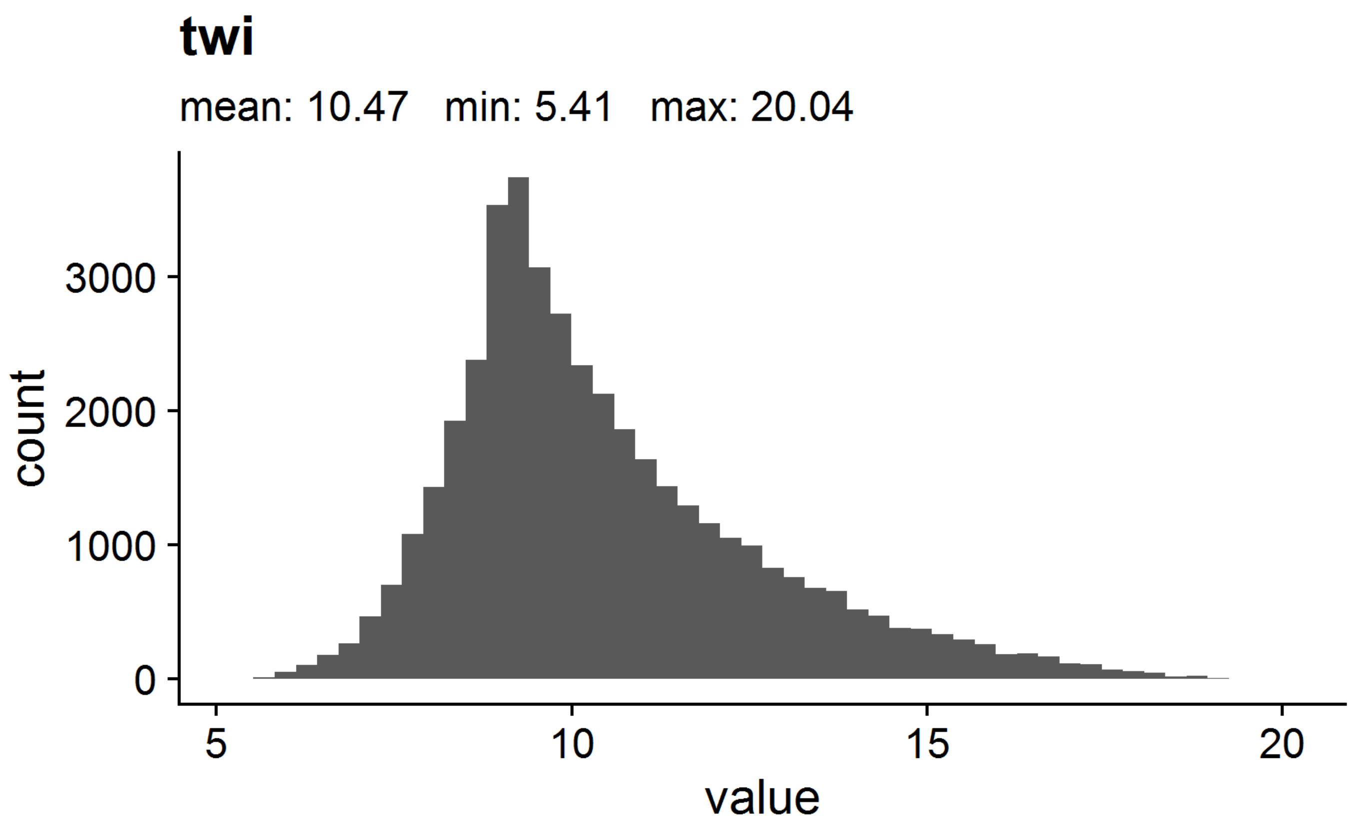
Calculation procedure:

1. Sink filling of 10 m DTM neighbourhood mosaic using the `ta_preprocessor 5` module and option `-MINSLOPE 0.01`. See Wang and Liu 2006.
2. Calculation of *flow accumulation* based on the *sink filled DTM mosaic* (produced in step 1) using the `ta_hydrology 0` module with options `-METHOD 4` and `-CONVERGENCE 1.0`. See Freeman 1991 and Quinn et al 1991.
3. Calculation of *flow width* and *specific catchment area* using the module `ta_hydrology 19` from the *sink filled DTM mosaic* and the *flow accumulation* (produced in steps 1 and 2). See Gruber and Peckham 2008 and Quinn et al. 1991.
4. *Slope* calculation based on the *sink filled DTM mosaic* (produced in step 1) using the `ta_morphometry 0` module with option `-METHOD 7`. See Harlick 1983.
5. *TWI* calculation using module `ta_hydrology 20`, based on *specific catchment area* (step 3) and *slope* (step 4). See Beaven and Krikby 1979, Boehner and Selige 2006 and Moore et al. 1991.

Issues:

- Tiles with incomplete neighbourhoods (i.e. less than 8 direct neighbours are available) will suffer from edge effects in the direct vicinity of the relevant border.
- *Flow accumulation* is only calculated for the tile neighbourhood. Even in the ideal case of the neighbourhood being complete, for most cells *flow accumulation* is therefore calculated only within a 30 km x 30 km catchment.
- General concerns regarding the TWI as a proxy for plant relevant soil moisture apply. See e.g. Kopecký et al. 2020 for more detail.

Histogram:



Note: The histograms is based on the sample of 50k cell cells shown in Figure 3.

References:

- Kopecký, Martin, Martin Macek, and Jan Wild. 2020. 'Topographic Wetness Index Calculation Guidelines Based on Measured Soil Moisture and Plant Species Composition'. *Science of The Total Environment*. 143785. https://doi.org/10.1016/j.scitotenv.2020.143785.
- Wang, L. & H. Liu. 2006. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. International Journal of Geographical Information Science, Vol. 20, No. 2: 193-213.
- Freeman, G.T. 1991. Calculating catchment area with divergent flow based on a regular grid. Computers and Geosciences, 17:413-22. ScienceDirect.
- Quinn, P.F., Beven, K.J., Chevallier, P. & Planchon, O. 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. Hydrological Processes, 5:59-79. ResearchGate.
- Gruber, S., Peckham, S. 2008. Land-Surface Parameters and Objects in Hydrology. In: Hengl, T. and Reuter, H.I. [Eds.]: Geomorphometry: Concepts, Software, Applications. Developments in Soil Science, Elsevier, 33:293-308. https://www.elsevier.com/books/geomorphometry/hengl/978-0-12-374345-9.
- Haralick, R.M. 1983. Ridge and valley detection on digital images. Computer Vision, Graphics and Image Processing, Vol.22, No.1, p.28-38.
- Beven, K.J., Kirkby, M.J. 1979. A physically-based variable contributing area model of basin hydrology. Hydrology Science Bulletin 24(1), p.43-69.
- Boehner, J., Selige, T. 2006. Spatial Prediction of Soil Attributes Using Terrain Analysis and Climate Regionalisation. In: Boehner, J., McCloy, K.R., Strobl, J.: 'SAGA - Analysis and Modelling Applications', Goettinger Geographische Abhandlungen, Vol.115, p.13-27.
- Moore, I.D., Grayson, R.B., Ladson, A.R. 1991. Digital terrain modelling: a review of hydrogical, geomorphological, and biological applications. Hydrological Processes, Vol.5, No.1.

Back to content.

Auxiliary files

---

water masks

Folder locations: `/outputs/masks/inland_water_mask` and `/outputs/masks/sea_mask`

File names: `inland_water_mask_xxxx_xxx.tif` and `sea_mask_xxxx_xxx.tif`

File type and units: `16-bit integer, binary (1 = land; no data = water)`

Description:

Sea and inland water masks for each tile.

Generated from polygon shapefiles assembled by Jesper Moeslund (AU Department of Bioscience - Biodiversity and Conservation). Fore each tile the polygon geometries were burned into the 10 m x 10 m grid using `gdal_rasterize`.

Issues:

- Shape, outline, presence and absence of small water bodies and coastlines may fluctuate over time.

The masks were chosen to present a snapshot of the water bodies as close to the time point of the LiDAR data acquisition as possible (winter 2014/2015), but inaccuracies may still arise.
- Location of inland water bodies and coastlines might have change since then.
- The inland water masks was produced to be as comprehensive as possible, but some small ponds and water bodies might have been missed.

References:
No relevant references.

Back to content.

---

tile footprints

Folder location: `/outpus/tile_footprints`

File names: `tile_footprints.shp`, `tile_footprints.dbf`, `tile_footprints.prj` and `tile_footprints.shx`

File type: `ESRI Shapefile`

Description:

Tile footprint geometries (polygons) for all processed tiles. This shapefile is particularly useful for identifying which tiles overlap with an area of interest should only a subset of the dataset be required for an analysis.

The file was generated based on the finished products for the `dtm_10m` variable using the `extract_tile_footprints.R` script.

Issues:

Currently no know issues.

References:

No relevant references.

Back to content.

---

date_stamp

Folder locations: `/outputs/date_stamp`

File names: `date_stamp_xxxx_xxx.tif`

File type: `32-bit integer, date in fromat YYYYMMDD`

Description:

The date_stamp variable provides an indication when the majority of points within a pixel were collected. This information is of use for ecological studies where accounting for within-season differences in timing of the ALS point aquistion is required.

Using the OPALS Cell module and the "majority" option, we exported the GPS time stamp with the highest frequency (mode) of all time stamps within the pixel. The GPS time stamp was then converted into Central European Time (CET) and reduced into a eight-digit integer with the format YYYYMMMDD, where YYYY is the year, MM the month and DD the date of collection.

Issues:

- Applying the "majority" estimator of the OPALS cell module on the GPS time stamp identifies the mode of points with the same microsecond time stamp. This is precision is potentially counter-beneficial, as in some cases it might not correctly identify the mode of points with the same date. Extracting the mode of the date directly would be more ideal, but also more challenging to implement with OPALS (GPS time to date conversion required before export). While this is worth considering for future version of the dataset, the current version of the date_stamp will likely capture the mode of the date sufficiently in most cases.
- If no points are present in a cell, then a date in September 2011 is assigned (for some reason between 14-20 September 2011). For future versions of this function we should look into why this is the case.

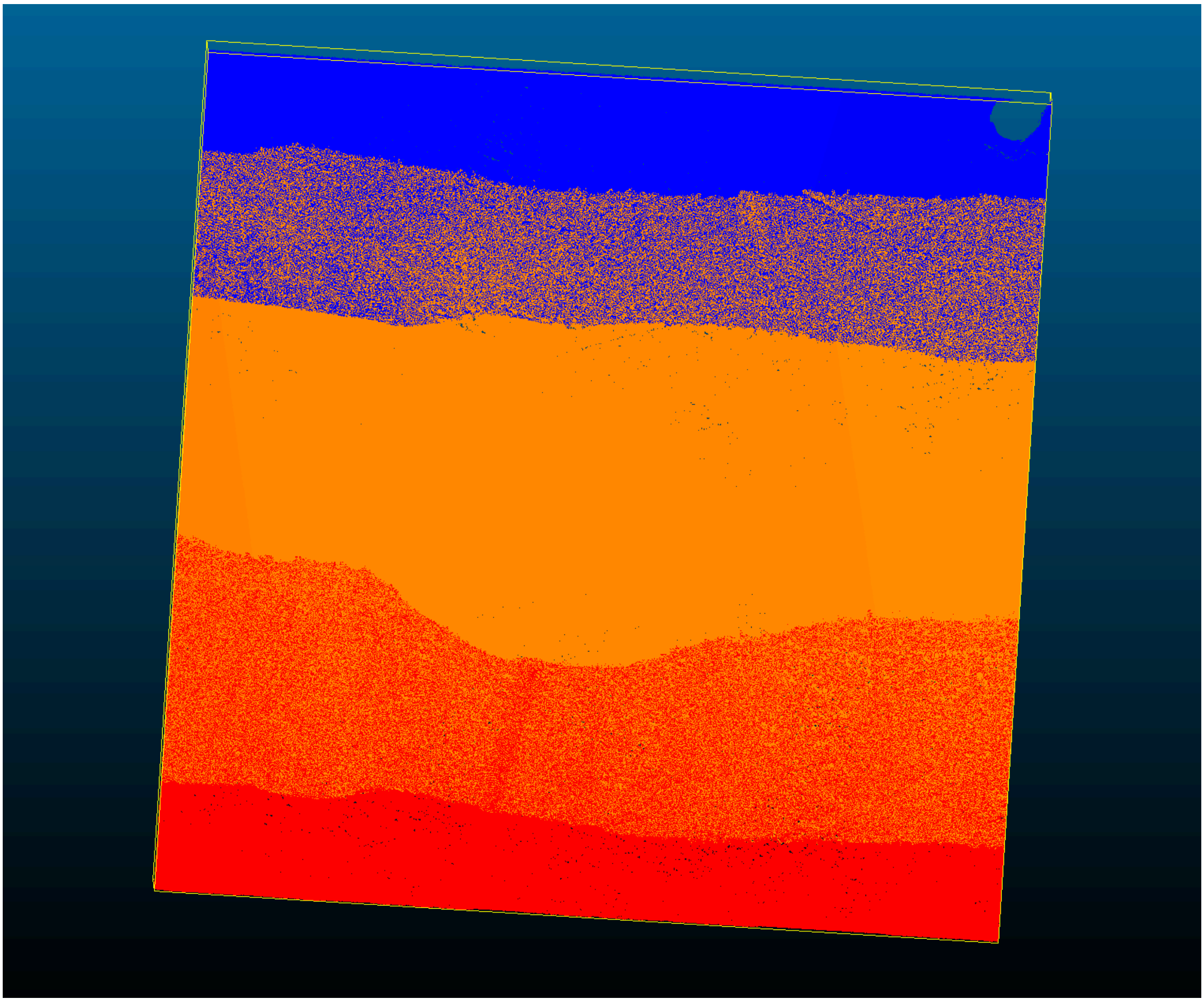


Figure: GPS time stamps in milliseconds per point for tile 6239_447.

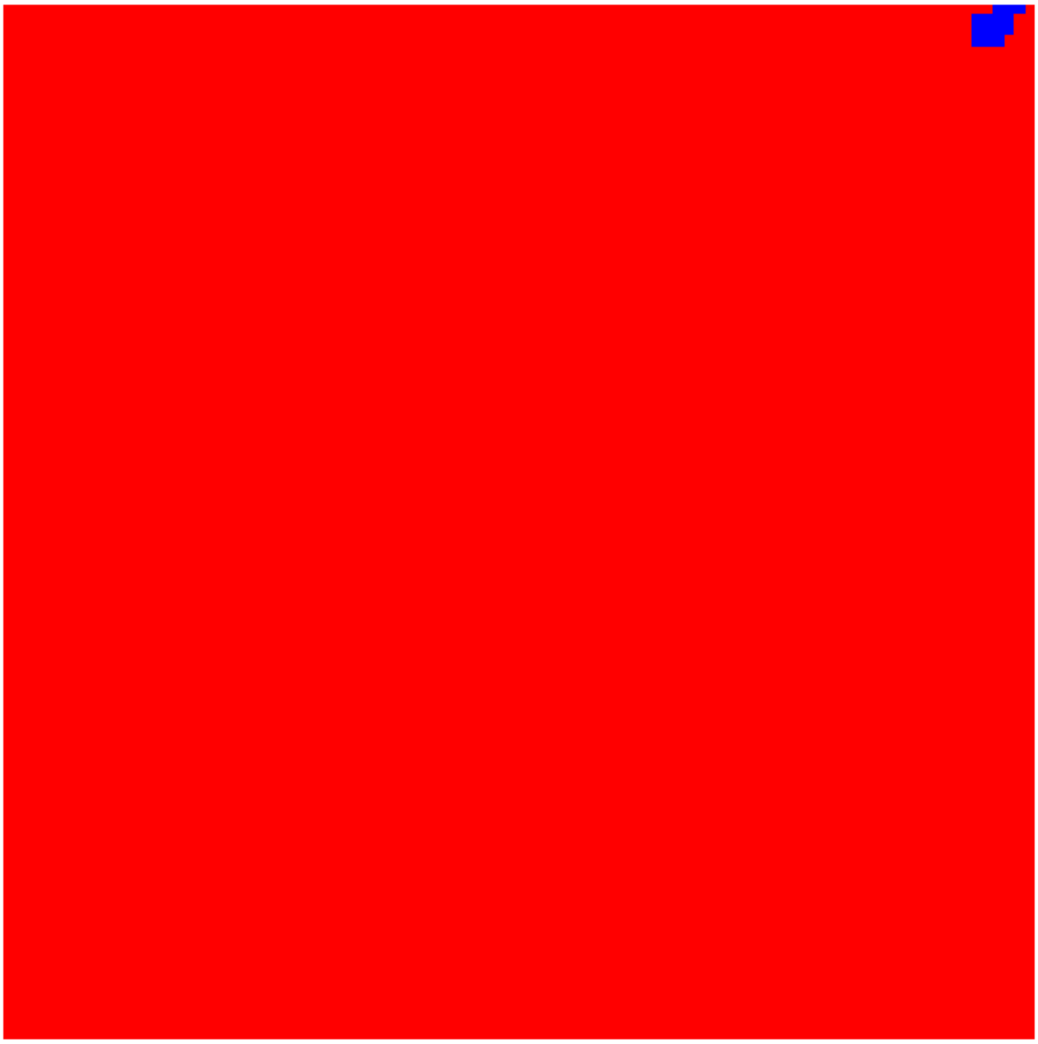
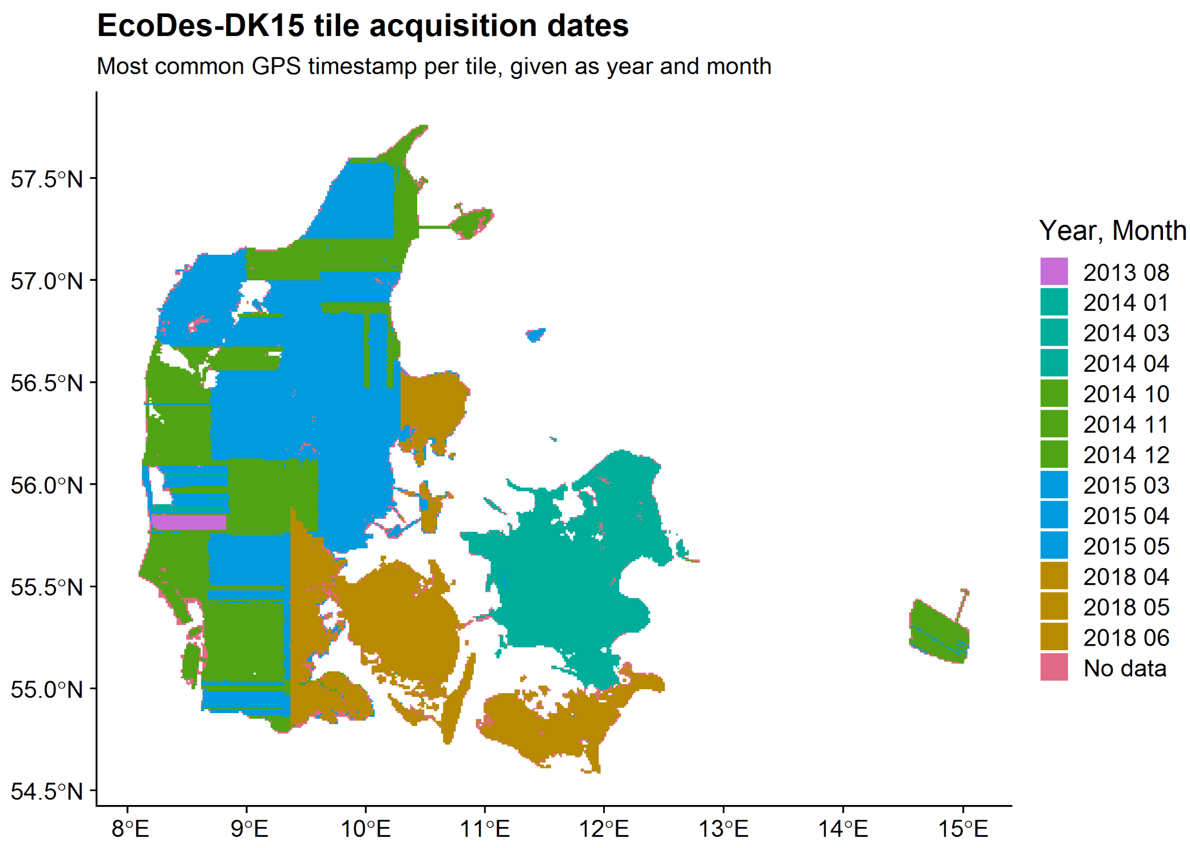


Figure: Date stamps per pixel extracted for tile 6239_447 (note September dates assigned to empty cells in top right corner).

- The dataset seems to contain not only data from 2014/15, but also from 2013 and 2018:



References:

No relevant references.

[Back to content](#).

---

vrt files

Folder locations: `/outputs/*`

File names: `*.vrt`

File type: `VRT file`

Description:

Each variable folder/archive contains a *.vrt file, where * is the variable name (same as the variable folder/archive). These [VRT files](#) allows the user to access all tiles for any one variable in a single virtual mosaic without the need of carrying out actual mosaicing of the rasters.

The files were generated using the `make_vrt_subfolders.bat` script.

Issues:

- On some platforms older RStudio R session are unable to open these files, likely due to memory / file number limitations placed on the R session by RStudio. Should you encounter this problem we recommend updating R studio or starting an R session in a console or using an alternative IDE.
- The files can be slow to handle in interactive GIS applications due to the size. We recommend generating pyramids in ArcMap or QGIS upon opening the VRTs in these applications.

References:

No relevant references.

Back to content.