DEVELOPING BIODIVERSITY PROXIES

Technical description

Technical Report from DCE – Danish Centre for Environment and Energy No. 123 2018

AARHUS UNIVERSITY
DCE – DANISH CENTRE FOR ENVIRONMENT AND ENERGY

[Blank page]

DEVELOPING BIODIVERSITY PROXIES

Technical description

Technical Report from DCE – Danish Centre for Environment and Energy No. 123 2018

Geoffrey Brian Groom
Jesper Bladt
Jesper Erenskjold Moeslund
Rasmus Ejrnæs

Aarhus University, Department of Bioscience

AARHUS UNIVERSITY
DCE – DANISH CENTRE FOR ENVIRONMENT AND ENERGY

Data sheet

Series title and no.: Technical Report from DCE – Danish Centre for Environment and Energy No. 123

Title: Developing biodiversity proxies
Subtitle: Technical description

Authors: Geoffrey Brian Groom, Jesper Bladt, Jesper Erenskjold Moeslund & Rasmus Ejrnæs
Institution: Aarhus University, Department of Bioscience

Publisher: Aarhus University, DCE – Danish Centre for Environment and Energy ©
URL: http://dce.au.dk/en

Year of publication: August 2018
Editing completed: June 2018

Referees: Flemming Skov
Quality assurance, DCE: Jesper R. Fredshavn

Financial support: The Danish Environmental Protection Agency

Please cite as: Groom, G.B., Bladt, J., Moeslund, J.E. & Ejrnæs, R. 2018. Developing biodiversity proxies. Technical description. Aarhus University, DCE – Danish Centre for Environment and Energy, 42 pp. Technical Report No. 123. http://dce2.au.dk/pub/TR123.pdf

Abstract: In 2014, Aarhus University and Copenhagen University in collaboration developed a biodiversity map for the Danish Nature Agency (Ejrnæs et al 2014). The map consists of two parts: 1) a national prioritisation of nature areas based on a complementarity analysis of species at 10×10 km scale and 2) a local prioritisation of nature areas based on a bioscore calculated as 10×10m pixels. Calculation of the bioscore is partly based on proxies for biodiversity. This report describes the methods developed and applied for the production of national datasets of certain new proxies for testing for potential use in the Biodiversity Map. The new proxies covered in this report take their basis in a mix of airborne laser survey data products, aerial orthophoto image data and certain national geoinformation data layers. The results of the testing of the new proxies for use in the Biodiversity Map are reported in the parallel report to this report, Ejrnæs et al. (2018).

Keywords: Biodiversity indicators, LIDAR, remote sensing

Layout: Graphic Group, AU Silkeborg
Front page photo: The biodiversity map and illustrations from the report.

ISBN: 978-87-7156-347-4
ISSN (electronic): 2245-019X

Number of pages: 42

Internet version: The report is available in electronic format (pdf) at
http://dce2.au.dk/pub/TR123.pdf

Contents

[Blank page]

1. Introduction

In 2014, Aarhus University and Copenhagen University in collaboration developed a biodiversity map for the Danish Nature Agency (Ejrnæs et al. 2014). The map consists of two parts: 1) a national prioritisation of nature areas based on a complementarity analysis of species at 10×10 km scale and 2) a local prioritisation of nature areas based on a bioscore calculated as 10×10m pixels. The aim of the bioscore is to make it possible to identify areas that are either known or potential habitats for redlisted species.

The bioscore consists of a species score, based on evaluations of known habitats for redlisted species, and a proxy score calculated as the sum of a number of proxies, or indicators, for biodiversity. If we had perfect knowledge of the distribution of biodiversity, the bioscore could be based on the species score alone. However, there exists no complete monitoring program for all species in Denmark, and for many species we have only limited sampling of their actual ranges. To counter this knowledge gap we include a number of proxies in the development of the bioscore focusing on improving our ability to evaluate nature areas for which we have limited knowledge of the species composition.

The proxies are included in the biodiversity map only if they have a known relation to biodiversity, if they can be mapped at national scale and if they contribute to improve our prediction of the occurrence of redlisted species.

The first version of the bioscore included 13 proxies selected from a list of 19 suggested candidates, such as proximity to the coast, steep slopes and forest continuity (Ejrnæs et al. 2014). The biodiversity map, specifically the local prioritisation was updated in 2015 (Bladt et al. 2016). This allowed us to update all species data to include new observations, but we also evaluated 4 new proxies to test if they could improve the statistical prediction of habitats for redlisted species. Two of the new proxies (‘Tree Height Variation’ and a proxy defined by areas with occurrences of habitats listed on Annex I in the Habitats Directive) passed the tests and were included in the bioscore at the expense of two other proxies.

In 2018 the Biodiversity Map is updated again. This involves harvesting and evaluation of available occurrence data on redlisted species, and updates of all proxies to reflect potential new data. The update also involves testing of a number of new potential proxy layers in the search of proxies that are even better at predicting occurrences of redlisted species than the previous set of proxies. This report describes the methods developed and applied for the production of national datasets of certain new proxies for testing for use in the biodiversity evaluation of bioscores in the Biodiversity Map. The new proxies covered in this report take their basis in a mix of airborne laser survey data products, aerial orthophoto image data and certain national geoinformation data layers. Previously, only one proxy derived from remote sensing data, ‘Tree Height Variation’, has been included in the Biodiversity Map based on the statistical tests. However, there appears to be a huge potential in utilizing remotely sensed data as these datasets are updated more and more frequently and in better and better spatial resolution with full national coverage.

Out of an initial bag of potential new proxies five main groups were selected for development: Urban Temporal Continuity (UTC), Sparse Vegetation Cover (SVC), Forest Canopy Roughness (FCR), Linear Microtopography (LNM), Non-Linear Microtopography (NLM). Different source data have been taken as the basis for each of these proxies: land parcel registration vector data (UTC), mosaic summer aerial orthophoto image raster data (SVC), canopy height model raster data (FCR), digital terrain model raster data (LNM & NLM). For each of these five groups of proxies, DK-wide analysis has been undertaken for production of vector irregular polygon (UTC, SVC, FCR), vector hexagonal polygon (FCR, LNM, NLM) and vector line (LNM, NLM) datasets. As well as representing a new proxy, the UTC work also represents a production of Ground Polygons for many of the parts of DK where these were absent from previous versions of the Biodiversity Map. All of these developments represent novel uses of source data and novel analysis and information delivery methods. The results of the testing of the new proxies for use in the Biodiversity Map are reported in the parallel report to this report, Ejrnæs et al. 2018.

2. Urban Temporal Continuity (UTC)

The previous versions of the Biodiversity Map (Ejrnæs et al. 2014, Bladt et al.2016) introduced the concept of Ground Polygons, comprising fields, protected areas (§ 3 areas) and subsets of forests with relatively homogeneous structure. Ground polygons have been used to define the extent of the Biodiversity Map (where the map is defined) but also in previous versions of the Biodiversity Map to define the spatial extent of point based species observations, where an observation was assigned to a ground polygon as a whole, rather than assigned to the map as a point. The mapping of Ground Polygons had many unmapped holes, the vast majority being all the parts (e.g. cities, towns, villages, and industrial or technical areas) with a high areal proportion of sealed surface (e.g. buildings, roads and similarly surfaced areas, building sites, derelict land). These areas can be important for biodiversity: Firstly, on account of the species that has the urban elements as their habitat, be it the gardens, parks, or buildings themselves. Secondly, as other habitats, with their biodiversity, exist in close proximity to the urban elements, either as adjacent areas or as elements within the urban matrix e.g. wooded areas (parks, cemeteries), wetlands, rivers, abandoned land. Therefore, it is vital to include the sealed surface parts of Denmark in the set of Ground Polygons to enable appropriate evaluation of the Bioscores.

The second justification for development in terms of Urban Temporal Continuity, is that the degree to which an area with a high areal proportion of sealed surfaces has been such, is itself seen as a key indicator of its value as a habitat for species. In its simplest terms, a “mature” urban area that comprises residential plots that were first created and build upon 50 years ago is more likely to be a better habitat for nature, overall, than an “immature” urban area that comprises residential plots first created and built upon just five years ago. This reasoning represents the essence of the Urban Temporal Continuity (UTC) proxy. Deriving urban spatial sub-units with relatively consistent degrees of UTC has been taken as the basis for mapping of Ground Polygons of the urban areas. It is in the nature of most patterns of urban development, particularly for residential land use, that a number of adjacent plots are created and built upon at a time. Thereby, spatial units based on merging of sets of adjacent individual property parcels build upon within a few years of each other can be formed as Ground Polygons with good representation of the UTC proxy.

A key DK dataset with information on the year when urban land plots were made available for urban land use are the so called “Jordstykke” vector polygon data. The Jordstykke data are one item of the MatrikelKort database; this work has used the version of those data as was available on 17th February 2014. The Jordstykke data (Figure 1) comprise vector polygons of all land parcels, i.e. 100% coverage of DK.

Figure 1. An example (Nakskov, Lolland) of the Matrikelkort Jordstykke parcel vector polygon data that was used for new Ground Polygons and for the Urban Temporal Continuity proxy.

The developed method also makes use of the following additional data layers:

GeoDanmark FOT data download of 20th March 2017:

- FOTbygning
- FOTvejmidte
- FOTvejmidteBrudt
- Biodiversity Map version 2015, Ground Polygons vector polygons
- BaseMap (http://www.dmu.dk/Pub/TR11.pdf) version 1, standard LULC legend, nearest neighbour resampled to 1 m raster cells.

Each layer was clipped to the required AOI extent. The method was developed and applied DK-wide as a workflow ("ruleset") in the Trimble eCognition v9.2 object based image analysis software. eCognition requires a raster data file to use as a basis for the cell size. The BaseMap dataset, clipped to the analysis extent and then resampled from 10 m to 1 m raster cell was used for this. Vector layers were clipped with Z and M values disabled in order to enable use in eCognition.

The UTC method was developed based on two 24x24 km areal units across Lolland, with 4 km E-W overlap between them in order to test for consistency in the results irrespective of the geographic extent of the input data (Figure 2).

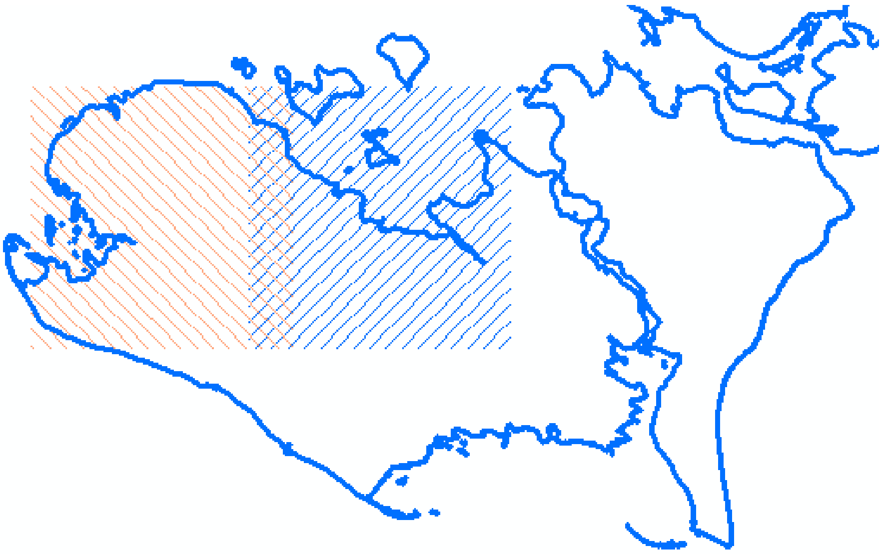


Figure 2. The two 24x24 km Lolland extents, with 4 km E-W overlay, that were the basis for UTC method development

The Jordstykke attribute data associated with every parcel includes data fields related to land use and four fields (RegistDato, GeomDato, SFE_Dato, PubliDato) that represent the year of events related to the parcel.

The key stages in the developed method are:

1. A set of objects are created from the Jordstykke polygon vector data.

2. Controls to isolate objects representing urban residential and commercial land parcels. This included exclusion of objects representing cemeteries, railway routes, farmland, class 1, 2 and 3 roads and land parcels without a building. These controls involve various uses of the three FOT layers and the BaseMap data.

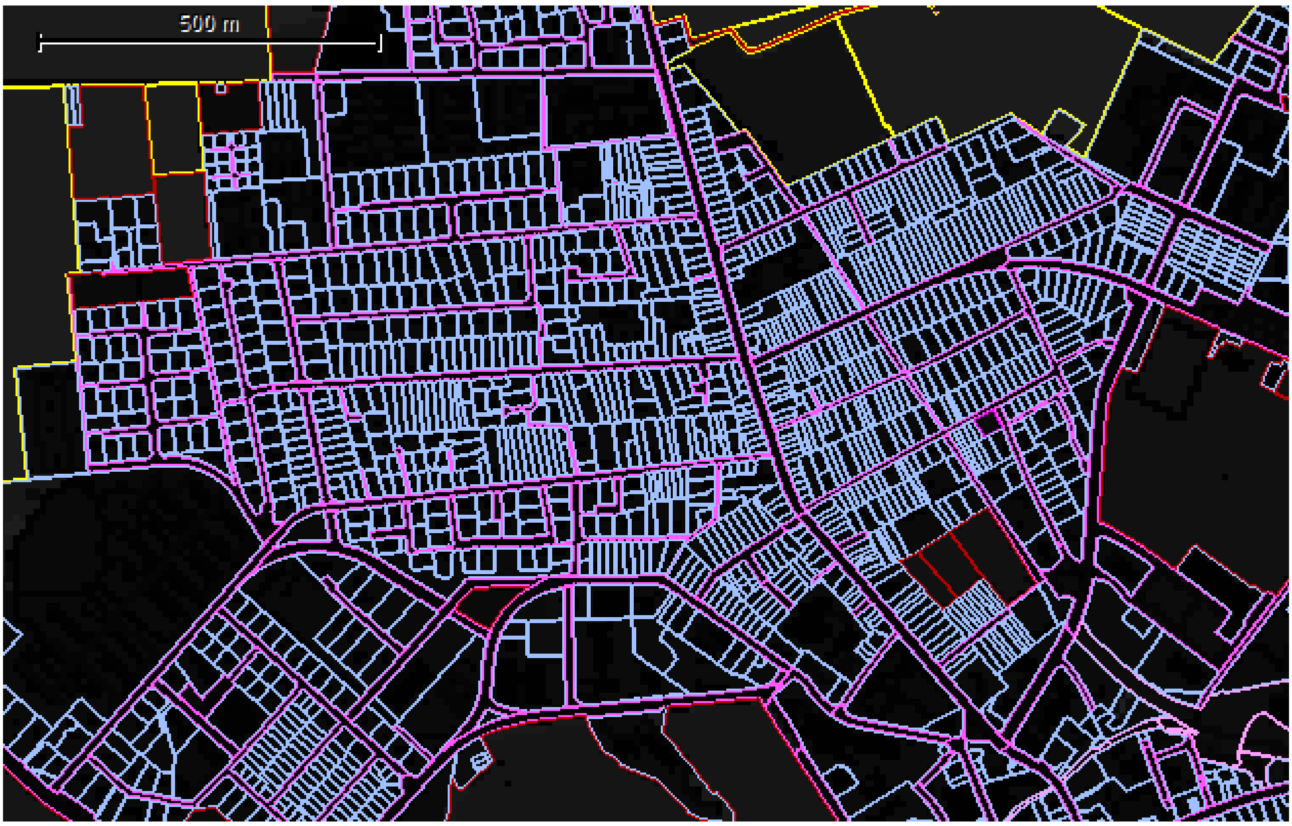


Figure 3. The Jordstykke land parcels (Nakskov, Lolland) with those that are to be merged based on date similarity in blue.

3. Merge remaining objects (Figure 3) based on the Jordstykke field RegistDato values: Taking a merge-object at random, a +/- 10 start and end years are defined and merge-objects with RegistDato values lying within the 20 year bracket are identified and where any are adjacent they are merged. This process continues until all merge-objects have been analysed in this way (Figure 4).

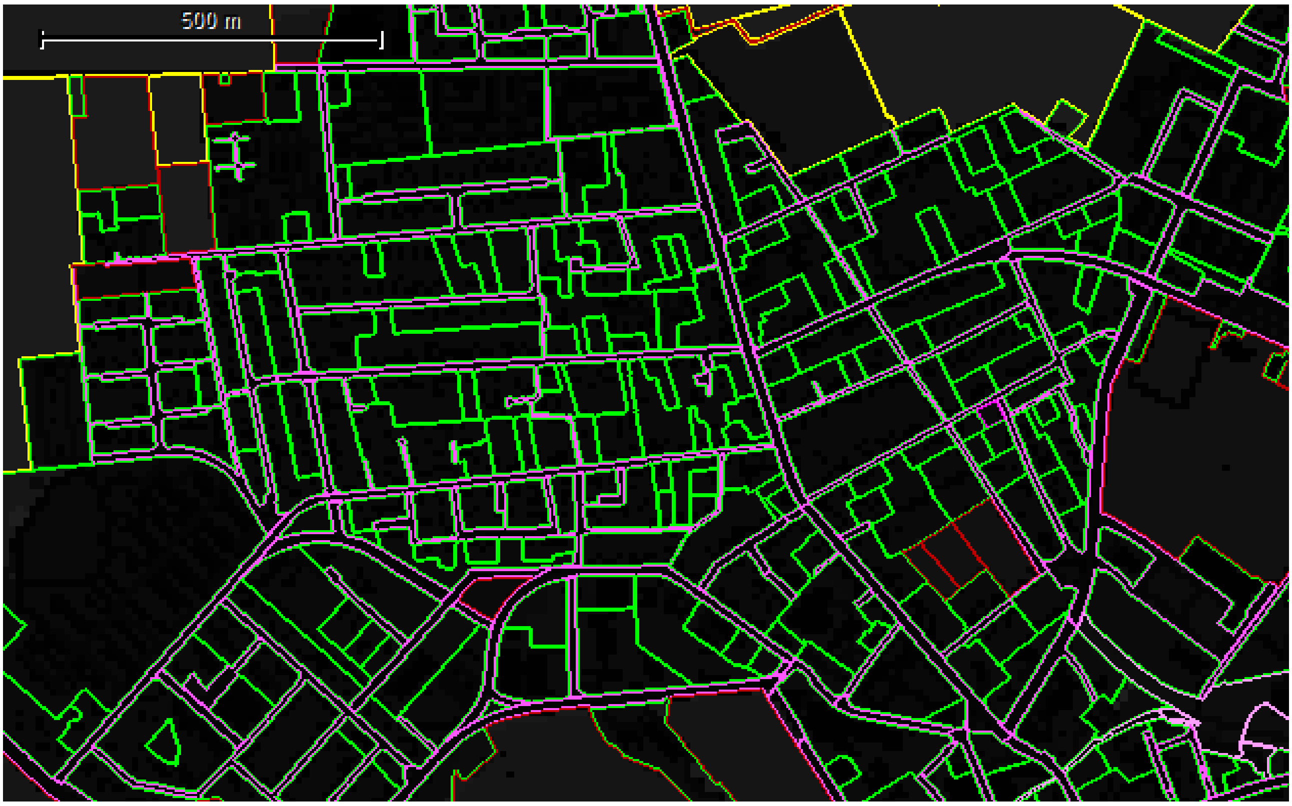


Figure 4. The result (green) of the date based merging of the land parcel objects (same extent as Figure 3).

4. Road objects are merged, and remaining land parcels that are not part of the previous versions of the Biodiversity Map set of Ground Polygons are labelled based on their composition of BaseMap classes (Figure 5).

In order to form a smaller number of larger new Ground Polygons (newGPs), some smaller land parcels have been merged with other adjacent larger land parcels that have a different history. Figure 6 shows that for most of the newGPs, the RegistDato-based age of the included land parcels is uniform. Thus, the newGP shown in red comprises eight land parcels with six having a RegistDato of 19310204 (4/2-1931), one having 19290802 and one having 18000101. And, the newGP shown in green comprises seven land parcels, with three having values of 19670823, and the other four values of 19680918, 19710601, 19720929 and 19760730.

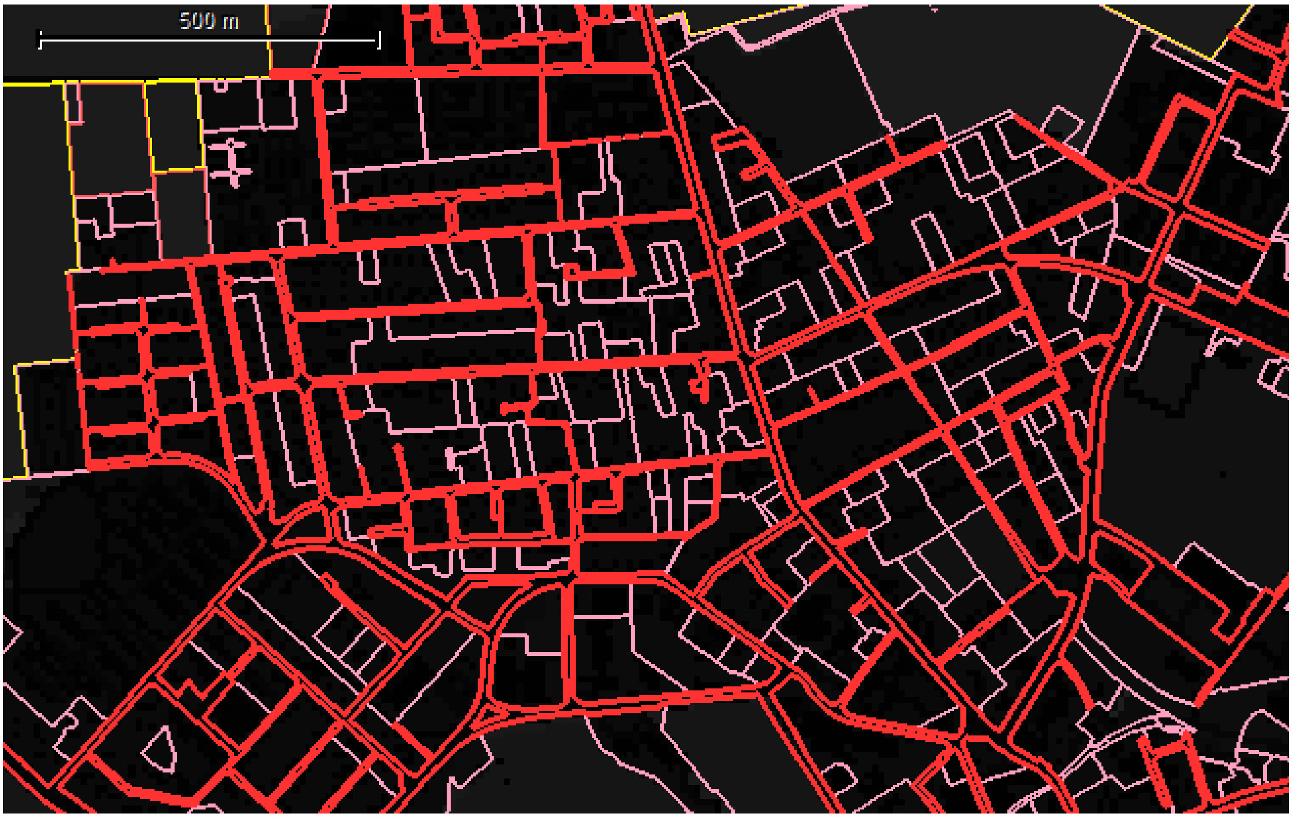


Figure 5. The new Ground Polygons in pink, previous GPs in yellow and road objects in red.

Figure 6. Standard deviation of the RegistDato value of the land parcels merged to each newGP. Grey-tone scaling applied here is between values of zero (black) and 250000 (white). However the sd values are distorted by the number scale of the RegistDato, whereby, for example, inclusion with a set of five 1950s parcels of one parcel from 1800 (sd = 55.9) will result in a greater influence on the sd than inclusion of two parcels from the 1920s (sd = 13.9). In hindsight, a better way would have been, for each newGP to allocate a common set of values to each of the different RegistDato present, e.g.
newGP-1 : comprising RegsistDato parts with values [18000101, 19000101, 1960101] → [1, 2, 3]
newGP-2 : comprising RegistDato parts with values [18400101, 19200101] → [1, 2].

As noted above, registDato is just one of four date fields in the Jordstykke attribute data. The other three are SFE_Dato, GeomDato and PubliDato. According to Kortforsyningen's 2011 data specification for the Matrikel data (https://kortforsyningen.dk/sites/default/files/old_gst/DOKUMENTATION/Data/matrikelkort_august_2011.pdf):

GEOMDATO : Dato for frigivelse i database
PUBLIDATO : Dato for frigivelse i database
REGISTDATO : Dato for frigivelse i database
SFE_DATO is only listed as one of the attributes of the Jordstykke dataset

For most of the land parcels the four date fields have the same value (Figure 7), particularly for the residential areas; roads, other technical areas and non-urban areas have date data value variation. For the five highlighted (in red) land parcels where the dates disagree, from left to right, the Geom. Publi, Regist and SFE dates are:

20090812	20090812	20090812	19770113
19850802	19850802	19850802	19760810
20030401	20030401	20030401	18000101
19870728	19870728	19870728	19651218
19880928	19880928	19880828	19770310

Hence we decided to use registDato for all land parcels rather than accounting for the few exceptions where the date fields deviated.

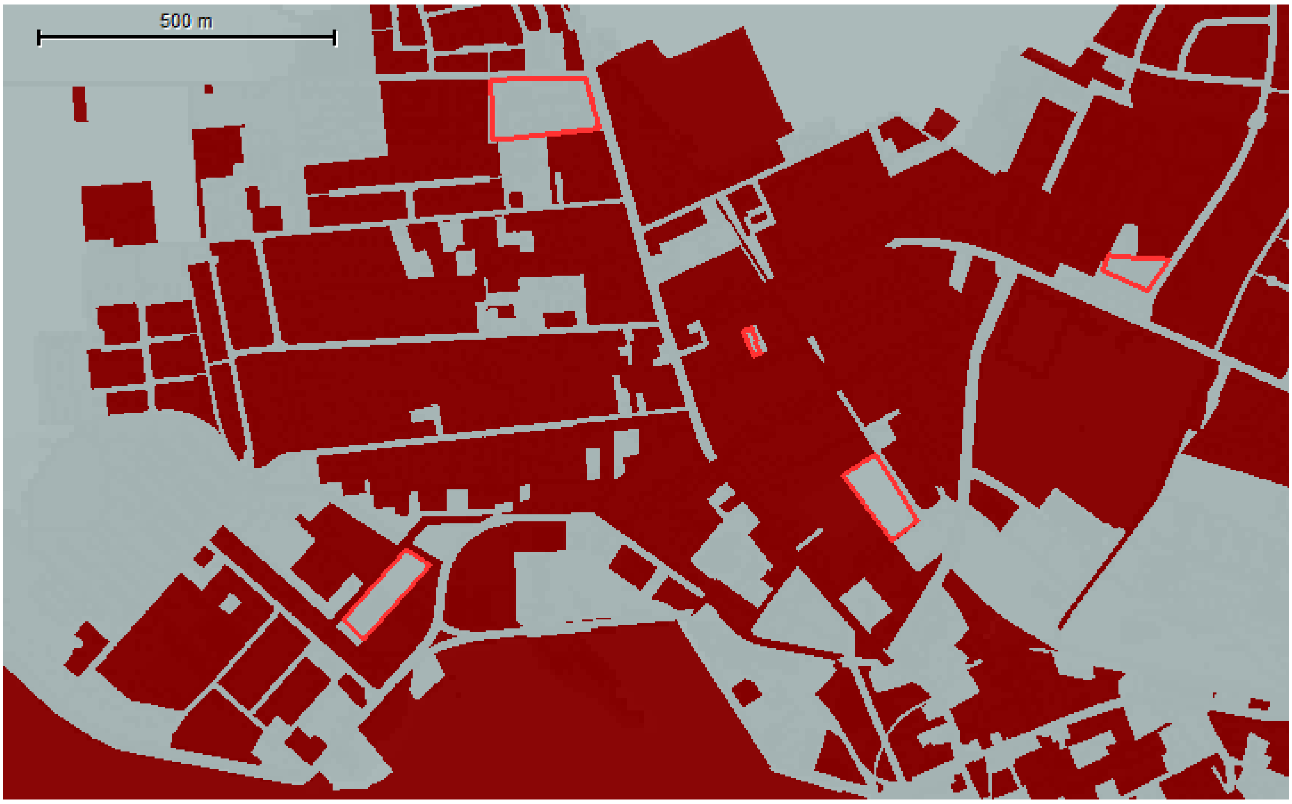


Figure 7. Agreement (red) versus disagreement (grey) at the level of the Jordstykke land parcels between the values in the four date related attribute table fields, with five parcels selected for date variation analysis as described in the main text.

The UTC analysis results for the overlap zone between the two 24x24 km Lolland test extents showed no differences.

The UTC analysis was applied DK-wide in terms of a set of approximately 8000 urban area cluster extents of varying sizes. These were formed based on the FOT 'Bypolygon' layer, each feature forming a cluster extend for the analysis. Outside these extends the resulting newGPs where supplemented with Jordstykke features overlapping the FOT layer 'Lav bebyggelse'.

The resulting newGPs with a Registdato for 1950 or older were used as proxy for old urban developed areas and tested for potential inclusion in the Biodiversity Map as the proxy 'Urban kontinuitet' (Ejrnæs et al. 2018). The proxy did not pass the statistical test though and was not included in the updated Biodiversity Map.

3. Sparse Vegetation Cover (SVC)

Areas with persistent low levels of vegetation cover may represent rare habitats that are important to specific plants and insects, as they represent environments with limitations such as in terms of nutrients, moisture, heat or disturbance factors. Examples of this situation in the Danish nature include grass heathlands, coastal grass and / or forb rich habitats, derelict land, low intensity technical areas, and banks alongside transportation routes. Sparse vegetation cover (SVC) represent a potential proxy for enhancement of the Biodiversity Map's Bioscores, but there has not existed any datasets that could serve as a ready basis for mapping an SVC proxy with the required spatial detail of < 100 m2 (10x10 m, 0.1 ha). Moreover, as SVC areas often include some larger woody plants and patches of bare ground, in addition to forbs and grasses, mapping of the components to an even fine spatial scale is needed, e.g. 0.0625 m2 (0.25x0.25 m). A mixture of cover types at a fine spatial scale is a pattern that distinguishes proxy significant SVC from proxy irrelevant SVC such as sown, intensively managed grasslands (e.g. municipal parks, lawns, sports facilities).

Satellite system remote sensing provides a basis for mapping vegetation levels via spectra band indices that can be interpreted as records of photosynthetic activity, such as the Normalised Difference Vegetation Index (NDVI). And, whilst several satellite systems now provide spectral image data with spatial resolutions between 0.5 and 2 meters (e.g. WorldView), it is not feasible, cost-wise, to acquire such data for multiple coverage of the entire extent of Denmark. Albeit that they now represent higher temporal frequency coverage (ca. 3 days at best) open access satellite image data for larger areas (Sentinel -2A, -2B), with spatial resolutions of ca. 10 m have been available for a few years, and the temporal fidelity of an entire Denmark coverage will be highly constrained by cloud cover conditions.

The mosaic dataset of summer aerial orthophotos imaged each year, supplied by SDFE to the public sector Consortium, which AU is a part of, for every 2nd year (even years) represents an alternative to satellite imagery. These image data have high spatial resolution (e.g. 0.16 m for 2012 and 2014, 0.2 m for 2016), high temporal fidelity (collected from early May - mid June for the entire country) and, since 2012, the spectral fidelity required to make the NDVI derivation. The radiometric qualities of these data are however very crude, on account of the sensing systems being digital frame cameras without any correction for incident light levels, the varying solar illumination condition under which they are collected (time of day, haze, etc. conditions), the colour balancing applied between image frames to improve them for visual analysis, and the delivery of the data as lossy compressed .ecw data. However, it has been possible to develop a method to use these data as a basis for mapping of SVC for all of Denmark. A key factor for that has been to ensure that the analysis workflow does not rely too heavily on accurate NDVI values.

In Figure 8 the SVC characteristic of two areas, one in a rural context, the other in an urban redevelopment context, are shown in terms of a NIR-VIS-VIS false-colour composite (FCC) representation of the 2012, 2014 and 2016 summer aerial orthophoto image data. In this FCC representation the intensity of redness can be crudely interpreted as the intensity of photosynthetic activity. In the former example (centre: utm 457800, 6174840), the area, approximately 50x50 m in extent, to the south of the trees has the appearance of relatively lower photosynthetic activity, with a patchy distribution of vegetation and bare ground on all three dates. The second area (centre : utm 576440, 6222520) is land reclaimed from the sea between 2000 and 2005 and then undergoing development as technical facilities, industrial units and offices. These image data indicate that between 2012 and 2014 much of the area was left to re-vegetate in a largely unmanaged way, leading to a SVC characteristic, but that since 2014 the vegetated extent has decreased as new roads and other hard surfaces have been developed.

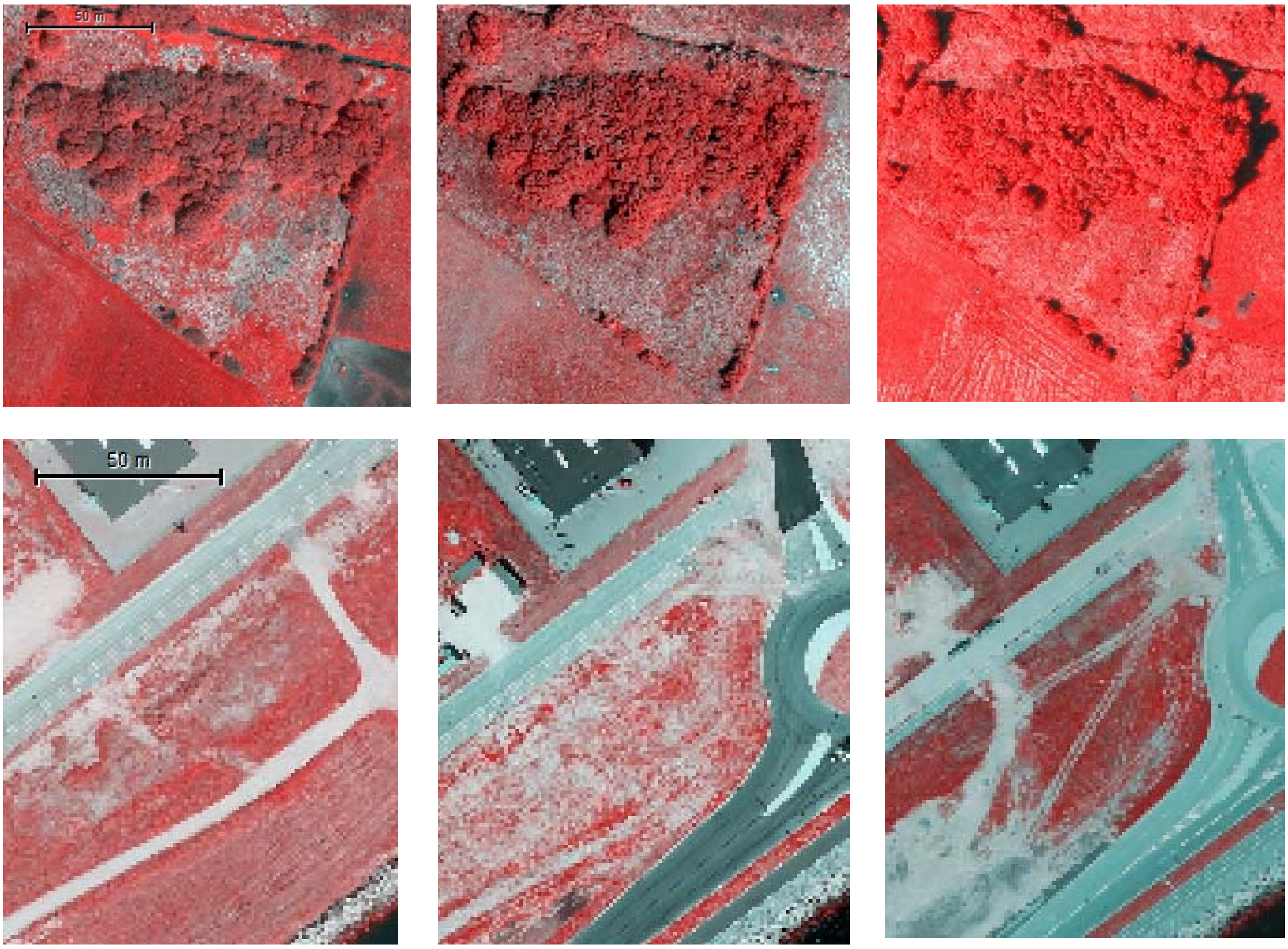


Figure 8. NIR-VIS-VIS false colour composite representations of the 2012 (left), 2014 (centre) and 2016 (right) summer orthophoto image data.

Figure 8 also illustrates the differences between these image data that are unrelated to the intrinsic scene target conditions, with the 2016 image data for the former example clearly displaying greater "redness" for all parts of the scene. It was therefore necessary to apply a standardization process to the pixel values. The standardization comprised two stages.

In the first stage, a 2-step z-normalisation was applied to the VIS and the NIR image pixel values:

1. (pixel value – mean) / sd → R_z
2. (R_z – min) / (max – min)

where, mean, sd, min, max are statistics calculated over all pixels. The "NDVI" is then derived for the standardized pixel values.

In the second stage, the z-normalised NDVI were balanced between the three years adjustment factors. Based on the 95 percentiles of the normalized NDVI values across 50x50 pixel spatial subunits, the adjustment factor were calculated as a pair of normalized difference calculations between the three years:

AF1 : (p95.NDVI.2012 – p95.NDVI.2014) / (p95.NDVI.2012 + p95.NDVI.2014)
AF2 : (p95.NDVI.2014 – p95.NDVI.2016) / (p95.NDVI.2014 + p95.NDVI.2016)

These standardizations are empirical in that they are affected by the compositions of the scenes in each year. However, they nonetheless increase the possibility to map SVC for each year with the same NDVI thresholds. Two pairs of an upper and a lower threshold were applied, one, being a looser pair of values, applied to map areas with SVC in all three years, and the other, being a tighter pair of values, applied to map SVC that was present in just two of the three years (Figure 9).

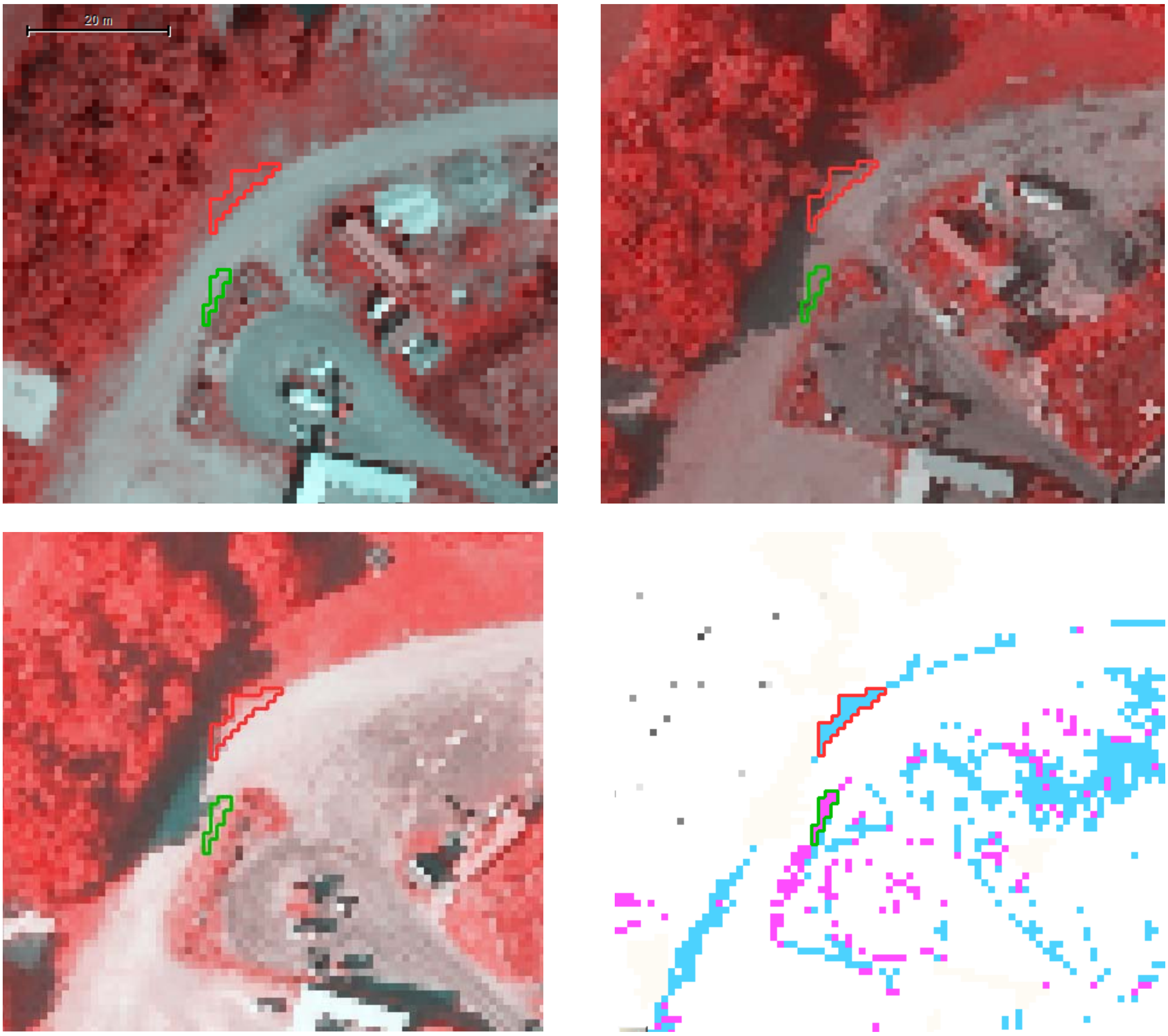


Figure 9. NDVI-based mapping of SVC, with (lower-right) cases of SVC for 2012, 2014 and 2016 in blue and cases of SVC in just two of the three years in purple. Also shown are NIR-VIS-VIS FCC of the SOF image data for 2012 (TL), 2014 (TR) and 2016 (BL), with one area of the 3-year SVC (red outline) and one area of 2-year SVC (green outline) highlighted.

It was necessary to screen the initial 2 / 3 year SVC areas for three forms of contra-case:

- The mappings of SVC from the image data can include cases that are not relevant for the Bioscores, such as agricultural land that was, by coincidence, in a SVC condition when the imaging was made in two or three of the years, or buildings associated with persistent SVC conditions, such as roof moss or algae. These parts were screened for using the HNV mapping of Cultivation Intensity (Brunbjerg et al. 2016) and the FOT bygning polygons respectively.
- Land in the process of development to a hard surface through the analysis period represents a special case, with the possibility of SVC conditions in the earlier years, but which should not be included in the final mapping of SVC as it is being changed to a permanently non-vegetated state. Thus, areas with SVC conditions for 2012 and 2014 but markedly low NDVI values for 2016 were also screened for.
- Areas maintained in a SVC condition due to intense management, such as lawns, municipal grasslands and sports facilities are also outside of the specification of SVC for application in the set of Bioscore proxies. Intense management, such as combinations of sowing with a relatively uniform seed mix, nutrient enrichment and frequent mechanical cutting of the vegetation increases the spatial homogeneity of the SVC. This condition was therefore screened for via an analysis of the local spatial variability in the NDVI pixel values of each year.

Conversely, the significance of SVC areas in terms of the Bioscore proxy is potentially increased by there being a spatial mosaic of SVC with patches of bare ground and / or taller vegetation. The former were analysed for in terms of the yearly NDVI values (i.e. as parts with low NDVI), and the latter in terms of the normalized Digital Surface Model data of the 2014-15 airborne laser survey national data set. Areas with bare ground or taller vegetation cover characteristic components were merged with the screened SVC areas. The presence of these additional cover components in mosaics with SVC was applied via the final screening, which removed any SVC+Bare+TallerVegetation (henceforth, “SVC_Final”) objects with limited extent and limited connectivity to other areas of such objects.

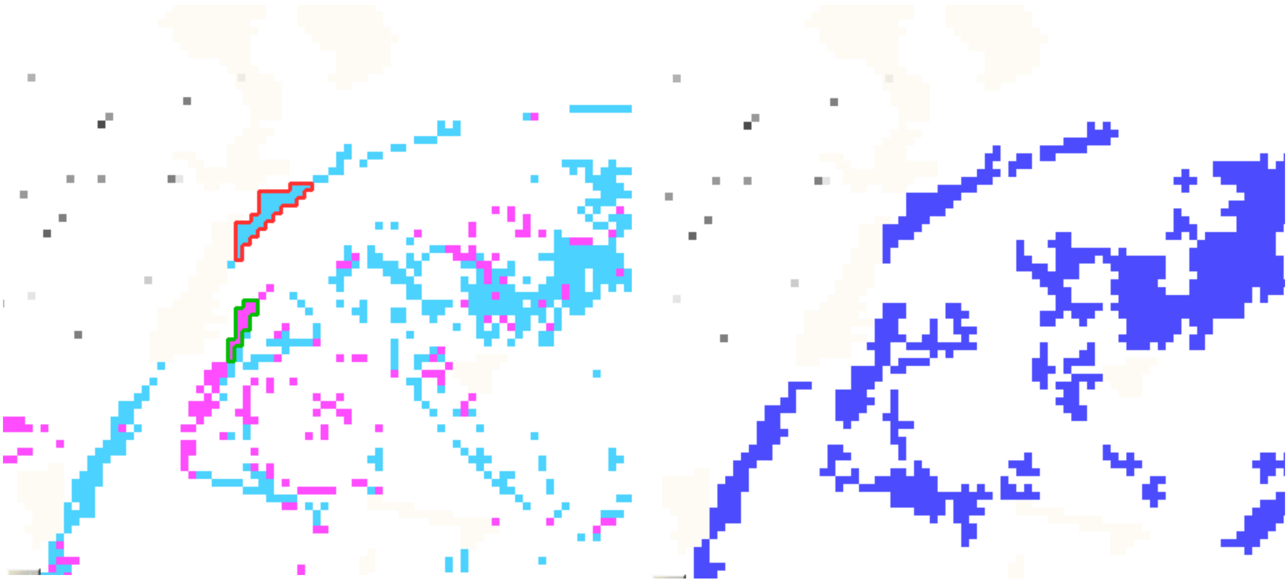


Figure 10. An example, same area as Figure 9, of, left, the initial 2 (purple) and 3-year (light blue) SVC mappings, and right, the SVC_final mapping after application of both negative and positive screening factors.

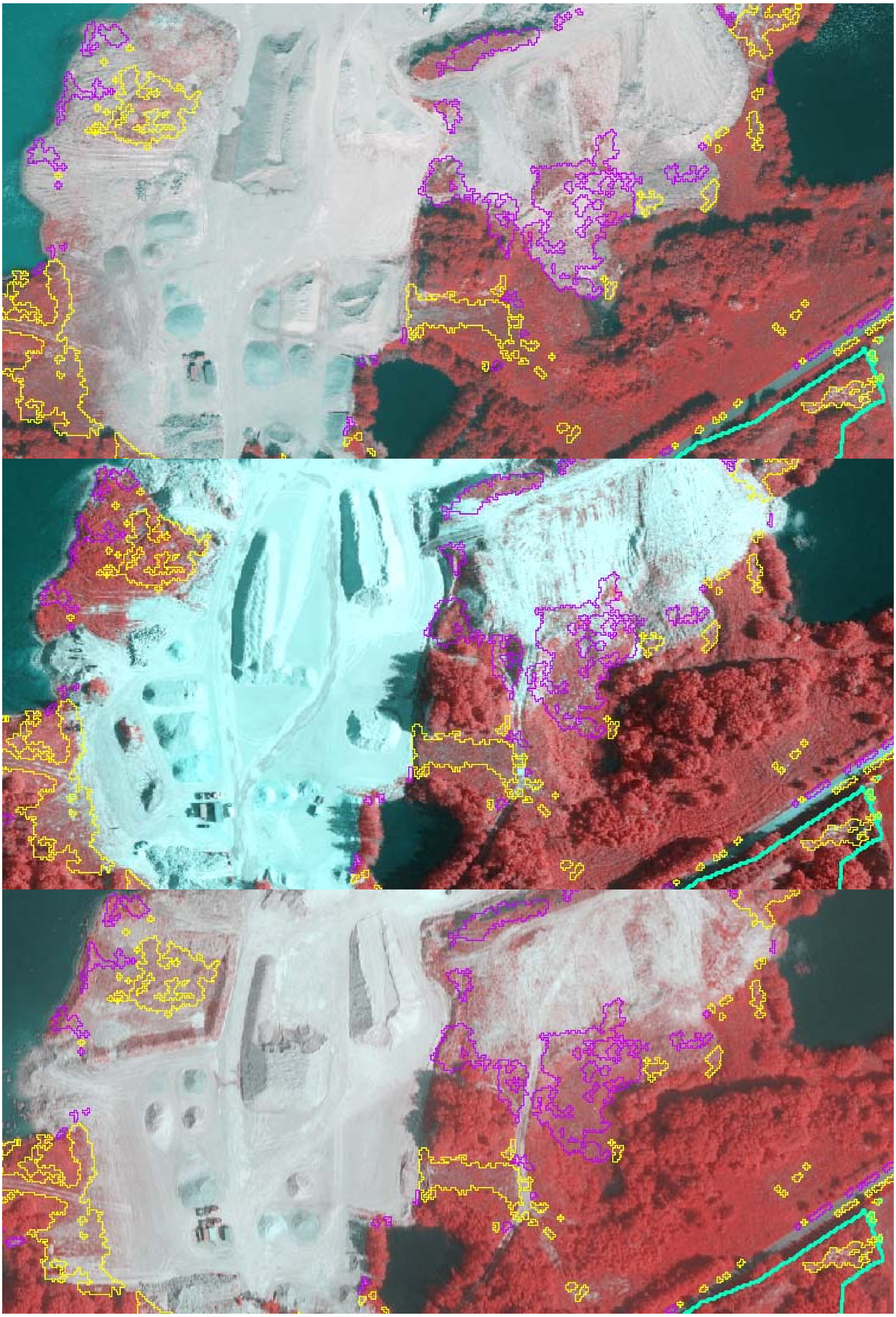


Figure 11. An area with considerable extent of SVC_Final, comprising both TwoYear (yellow) and AllYears (purple) cases, with just one case that coincided with an existing Ground Polygon (light green). The backdrop image data are, from top to bottom, NIR-VIS-VIS FCC renditions of the 2012, 2014, 2016 summer aerial orthophotos. Notes that roadsides, being lightly vegetated verges, are included in the SVC_Final set of areas.

The SVC method was applied DK-wide in terms of a set of 10645 mainly 2x2 km tiles (see Box 1).

Box 1

The SVC, FCR, LNM and NLM methods were applied DK-wide in terms of a set of 10645 non-overlapping mainly 2x2 km (4,000,000 m2) tiles. The tile size was adjusted to avoid the occurrence of tiles that are less than 2x2 km, such are along coastlines (Figure B1.1), by merges with adjacent tiles, resulting in 1251 tiles larger than 2x2 km (turquoise tiles in Figure B1.2), with the largest being approximately 10,000,000 m2, associated with the inclusion of small islands.

Figure B1.2

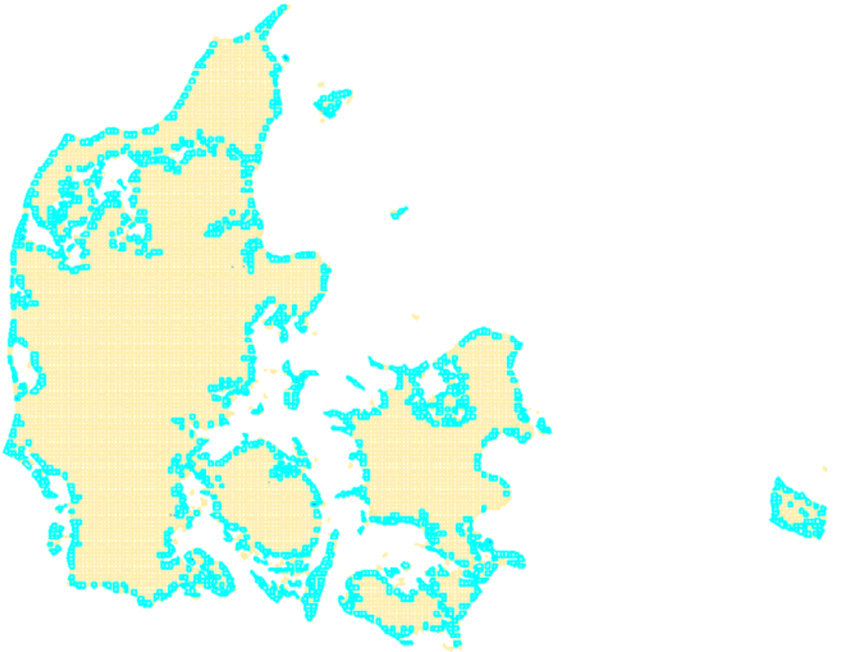


Figure B1.2

The resulting mapped SVC_final extents were provided for analysis as a proxy layer (Figure 11). After manual evaluation we included all polygons with SVC of at least two years in the proxy (see Ejrnæs et al. 2018). The proxy did not pass the statistical test though and was not included in the updated Biodiversity Map.

4. Forest Canopy Roughness (FCR)

The three dimensional form of an area of taller woody vegetation can represent a relevant variable of the Biodiversity Map Bioscores through a number of process relationships. A more complex 3D canopy form can be indicative of an area where more natural processes of taller woody vegetation development occur, with natural tree falls and/or branch losses give rise to a "more structured" forest with the possibility for a larger number of ecological niches. Conversely, a simply, more uniform 3D form, may be taken as indicative of a more managed area, with a maintenance of even aged trees of a single species (e.g. Fagus sylvatica). Alternatively, a more structured canopy but where the structure has a marked spatial pattern could be indicative of plantation forest, with trees planted in rows. Analysis of the 3D form of a vegetation canopy requires data that records the heights of canopy components. Two remote sensing methods are at present a basis for that: stereographic processing of overlapping aerial orthophoto image data and airborne lidar survey. Both provide, as initial data sets, point cloud data which can be either used directly or applied after rendering as raster data, i.e. raster digital elevation model data. While, the former can provide very dense (i.e. many points per areal unit) point clouds and the possibility for good integration of simultaneously acquired elevation and spectral information, airborne laser survey (ALS) represents a more direct recording of heights and greater possibilities for analysis in terms of multi elevation components. Furthermore, DK-wide ALS data have been acquired in 2014-2015 and processed to raster expressions of the terrain height (digital terrain model, DTM) and over-terrain surfaces (taller vegetation, buildings, etc, digital surface model, DSM) with a cell size of 0.4 m (Rosenkranz and Lund 2015). The latter have been applied to derive taller woody vegetation canopy structure (FCR "forest canopy roughness") proxy, via a normalised DSM (nDSM), i.e the surface heights corrected for variations in the ground heights, as DSM-DTM.

Marked variations in canopy structure are clearly displayed in the nDSM data (Figure 12).

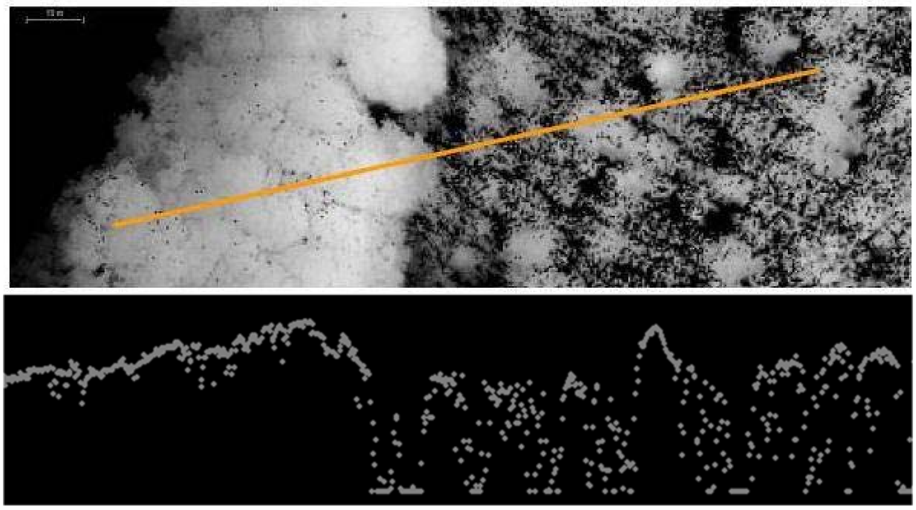


Figure 12. Upper: Example of the 2014-15 ALS nDSM data. Lower: nDSM data value profile along the orange line. The image and profile display the presence of canopy parts with markedly different forms, one (SW) of higher trees with larger and more coalesced crowns, the other (NE) of mostly lower, trees with smaller crowns and marked separation between the crowns.

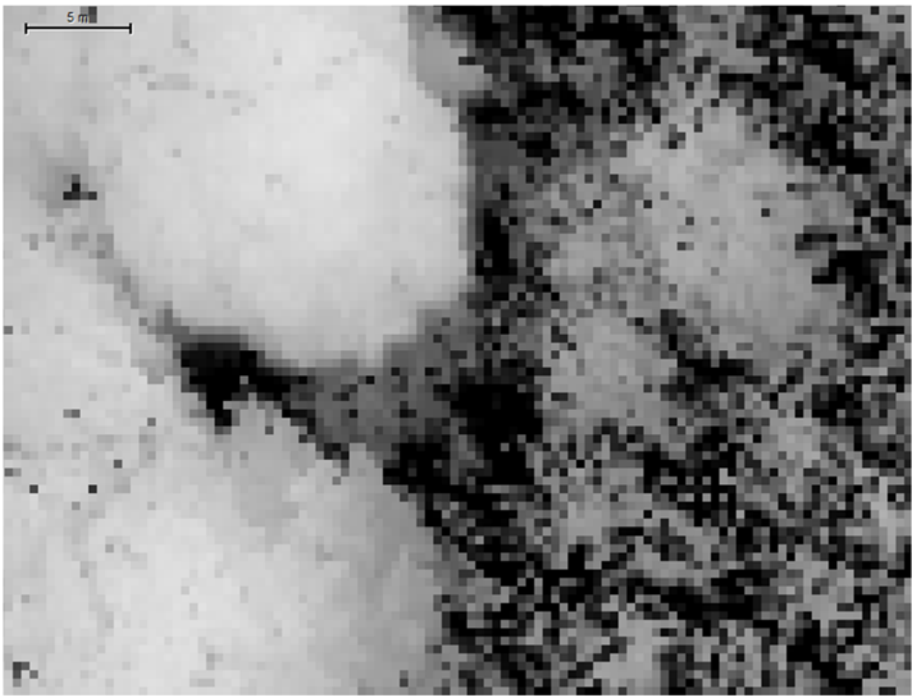
However, at broader scales, the diversity of canopy 3D forms presents a number of challenges for analysis as a single expression of roughness:

- height variations : should local height variations of 3 m within a canopy of 25 m high trees be evaluated the same as 3 m height differences in a canopy comprised mainly of 10 m high trees?
- canopy gaps: if all gaps between trees are seen as aspects of the local canopy roughness, what degree of gap size and 2D spatial form represents the transition from a rough canopy to a more "tree savanna" / parkland situation?
- understorey layers : at what point does a difference in height of the canopy components represent the presence of two separate canopy layers with open space between them, as when a younger closed canopy grows up in an area with standard trees with crowns that start 10 m higher than the young trees
- micro scale nDSM variations : The nDSM is a derivative of the ALS point cloud data, but the method used and its design criteria are not known. It may be that the extent to which it can be applied to analysis more localised height variations in vegetation is less than the extent to which it can be applied for broader scale vegetation height variations or for the 3D form of more solid, more regular structures such as buildings (Figure 13).

Many standard evaluations methods for canopy roughness, such as in terms of nDSM statistical variability (e.g. coefficient of variation, entropy), or in terms of derivatives such as kernel based edge filters (e.g. lee sigma, canny) are markedly influenced by the "extreme" values represented by the gaps between the crowns. Therefore, it is necessary to ask, should the gaps be ignored from expressions of the "canopy" roughness? Should gaps be filled to reduce the effect of the extreme nDSM values therein? What height should they be filled as?

The novel method developed here has been to describe the canopy structural complexity in terms of linear zones where the general height levels within the canopy height alter. As well as their lateral extent, the height change zones have the magnitude of the height change as an attribute. From the mapping of these zones, the canopy structure complexity is evaluated in terms of the density and variety of the height change zones per areal unit (hexagons). The spread of zone occurrence across hexagon segments enables evaluation, for sets of adjacent hexagons, of the degree to which observed canopy structures are part of a repeated broader scale pattern.

Figure 13. An example from the nDSM data. In the higher vegetation to the left the general canopy form is apparent. In the TR a more solitary higher tree with a large crown is apparent, but simple interpretation of the data in the LR part as distinct vegetation structures is problematic.



Working from a 3x3-pixel kernel median value smoothing of the nDSM to reduce the effect of single pixel noise, with parts corresponding to FOTbygning vector polygons excluded, the workflow begins with simplification of the canopy complexity via a set of objects of relatively uniform height (Figure 14). The inner-workflow for the formation of these canopy "platform" objects is a step-wise thresholding, starting with the tile nDSM maximum and ending when the step lower threshold is 2 (i.e. 2 meters, thereby, not forming platform objects that include low vegetation or unvegetated gaps).

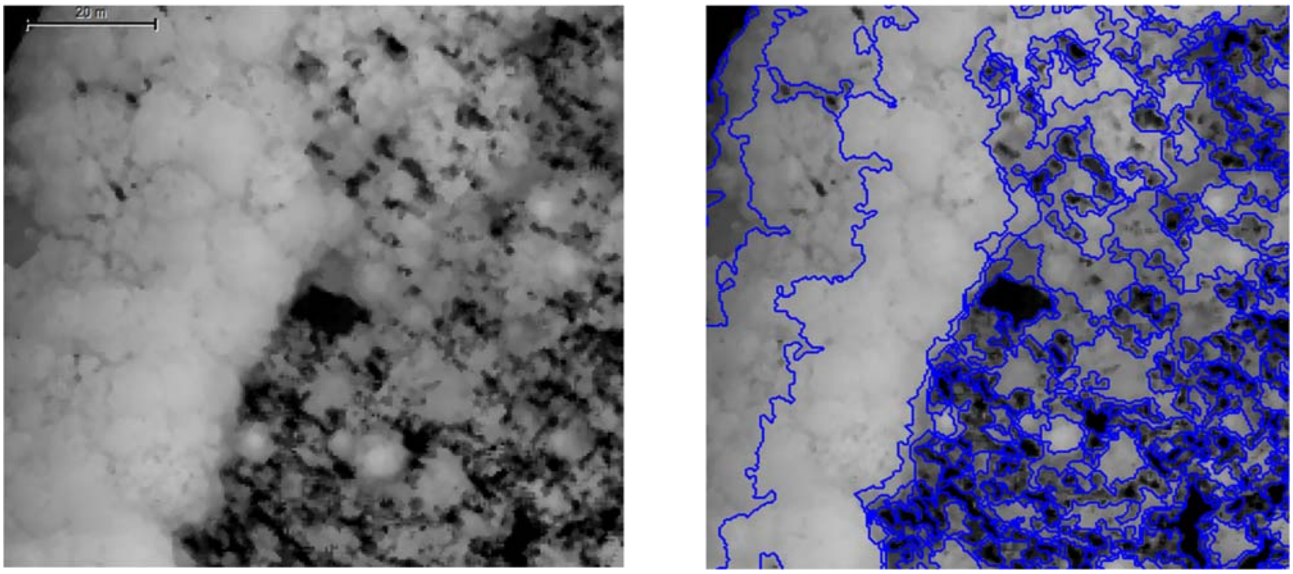


Figure 14. Objects ("platforms") comprising relatively uniform nDSM heights.

At each step (Si), there is also (a) iterative pixel-wise object growth for adjacent nDSM image pixels with values a little less than the initial lower threshold but contained by not including parts with a marked canopy slope (based on a simple canopy slope image) and (b) inclusion of non-platform object with an extent of less than 40 m2 and greater than 0.75 relative common border to an Si platform object.

The interface zones of platform pairs that have height nominal differences of 3-6 m (2-step), 6-9 m (3-step) and 9-12 m (4-step) are then delimited and labelled (Figure 15). Information is retained as a zone object variable of which specific height step the zone rates to, e.g. such that a 2-step zone between a 24-27 m platform and a 18-21 m platform can be distinguished from a 2-step zone between a 21-24 m platform and 15-18 m platform. The platform interface zones have the same width throughout; where two platform interface zones of the same step-dimension meet, the zones are merged, locally giving rise to zones with double width.

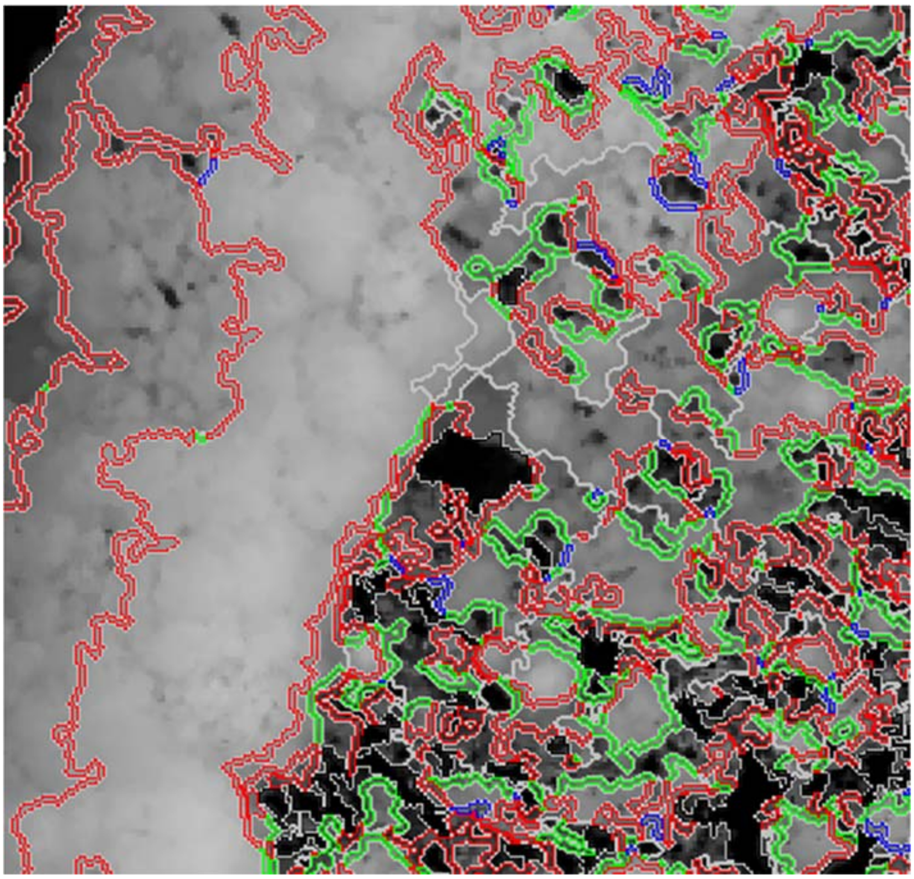


Figure 15. Platform interface zones : red: 2-step, green: 3-step, blue: 4-step The marked contrast between the high and even canopy to the left and the more complex canopy structure to the right is apparent in the platform interface zone density and variety patterns.

The final analysis is made in terms of a nested coarse set and fine set of hexagons, and their associated triangular segments (Figure 16). Hexagons are used as analysis spatial units rather than quadrats as they can be expected to edge-wise associate better with the basic units of forest vegetation, namely roughly circular trees crowns, and also due to the geometric limitations of quadrats (e.g. distance inconsistencies between sideways and cornerwise neighbours). The hexagons are generated as ESRI vector polygons as parts of a single nominal DK-wide hex-net; thus, hexagons of both sizes mesh along the edges of analysis units such as the sets of approximately 2x2 km tiles (see above). The two hexagon sizes used were side-lengths of 25 m and 50 m. The triangular segments were also made as ESRI vector polygons for each hexagon. The hexagon and segment production was made by a bespoke Python+arcpy script. The possibility to form hexagon objects is provided as a locked customised eCognition algorithm and the possibility to form hexago-

nal ESRI vector polygons is provided by the ESRI ArcMap addon tool Create_Hexagon_Tessalation
(https://www.arcgis.com/home/item.html?id=03388990d3274160afe240ac54763e57). However the script developed here for hexagon and segment creation is seen, from experience, as a marked improvement over both of those options.

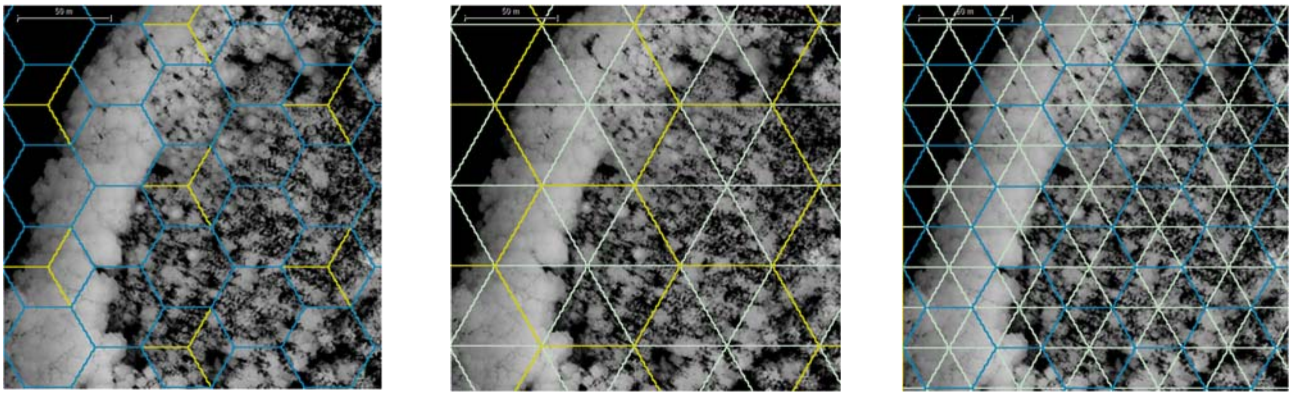


Figure 16. Illustrative examples of the hexagons (left, 50 m in yellow, 25 m in blue), 50 m hexagon segments (middle) and 25 m hexagon segments) used as for canopy roughness proxy spatial analysis.

For the final part of the canopy structure complexity workflow, three analysis were made for both hexagon size object sets:

- Relative Area of the edge zones (all step magnitudes) per hexagon, RAE (Figure 17 and 18).
- The Variety of edge types per hexagon, VAR, in terms of the number of different step + platform-pair edge units are present, for those edge cases with at least a nominal relative presence in the hexagon.
- The degree of Spatial Spread of the edges within the hexagon: this was evaluated in terms of the hexagon segments (triangles), such that any segment with a nominal minimum relative extent of nDSM values > 0.5 + a nominal minimal relative extent of edge objects were counted and expressed as a proportion of the number of segments that met the first criteria. So, with the possible maximum of 6, a hexagon with just 5 segments that comprised vegetation for > the minimal relative extent, and with just 3 of those five for which the minimal relative extent of edge objects was exceeded, the Spatial Spread value would be 3 / 5, i.e. 0.6.

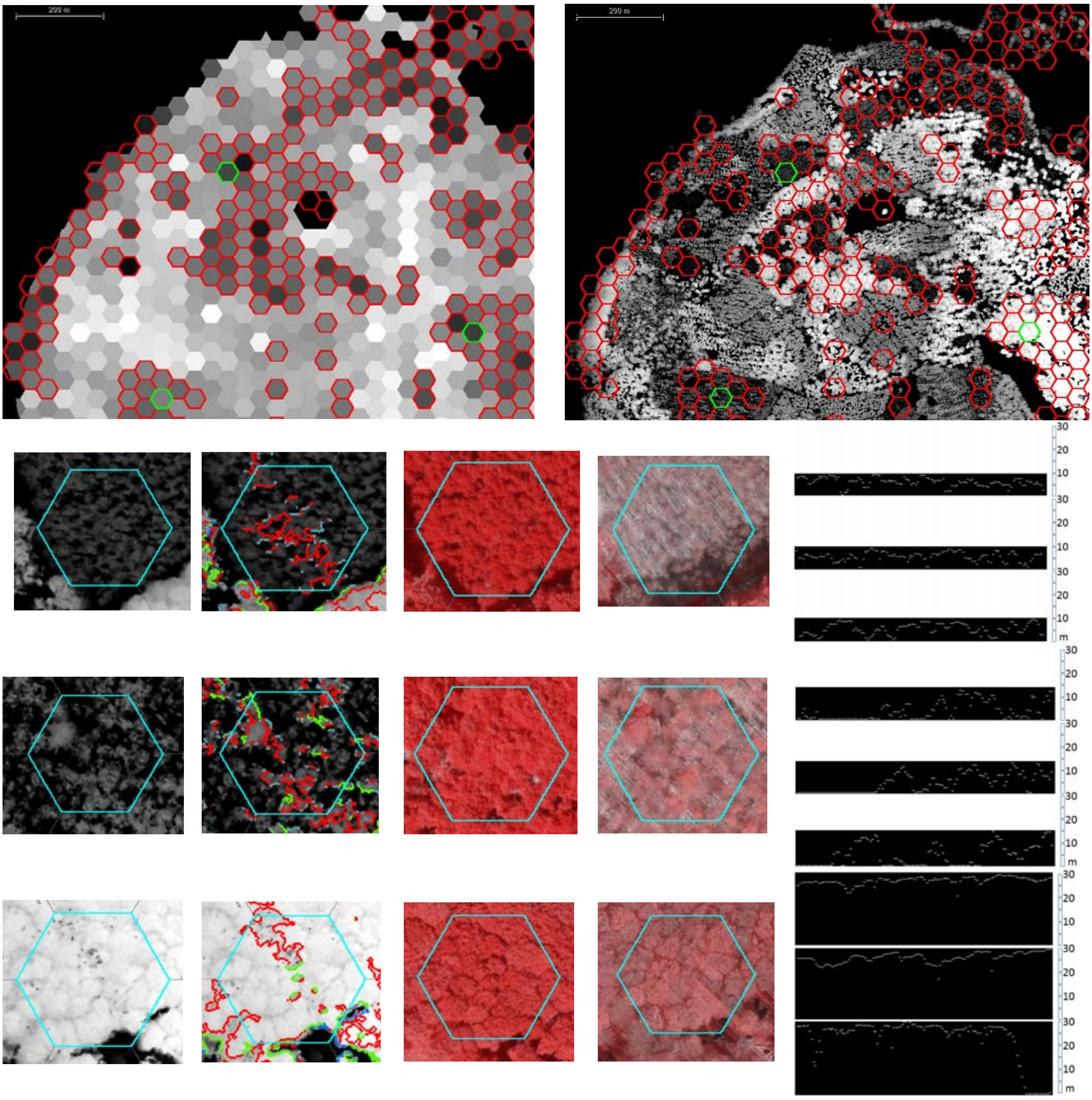


Figure 17. Relative Area of Edges: three 25 m hexagon cases with RAE < 2.0. Top row left : grey-scale representation of the per-hexagon RAE values (low→high : dark→light); Top row right : grey scale representation of the nDSM data (low→high : dark→light) : vector overlay (red or green) of hexagons with RAE < 2.0. Rows 2 to 4 : illustration of the three single hexagon shown in green in the top row: left to right: nDSM, platform interface zones, 2014 summer aerial orthophoto image data, 2016 spring aerial orthophoto image data, nDSM profiles for the three principle axes of the hexagon.

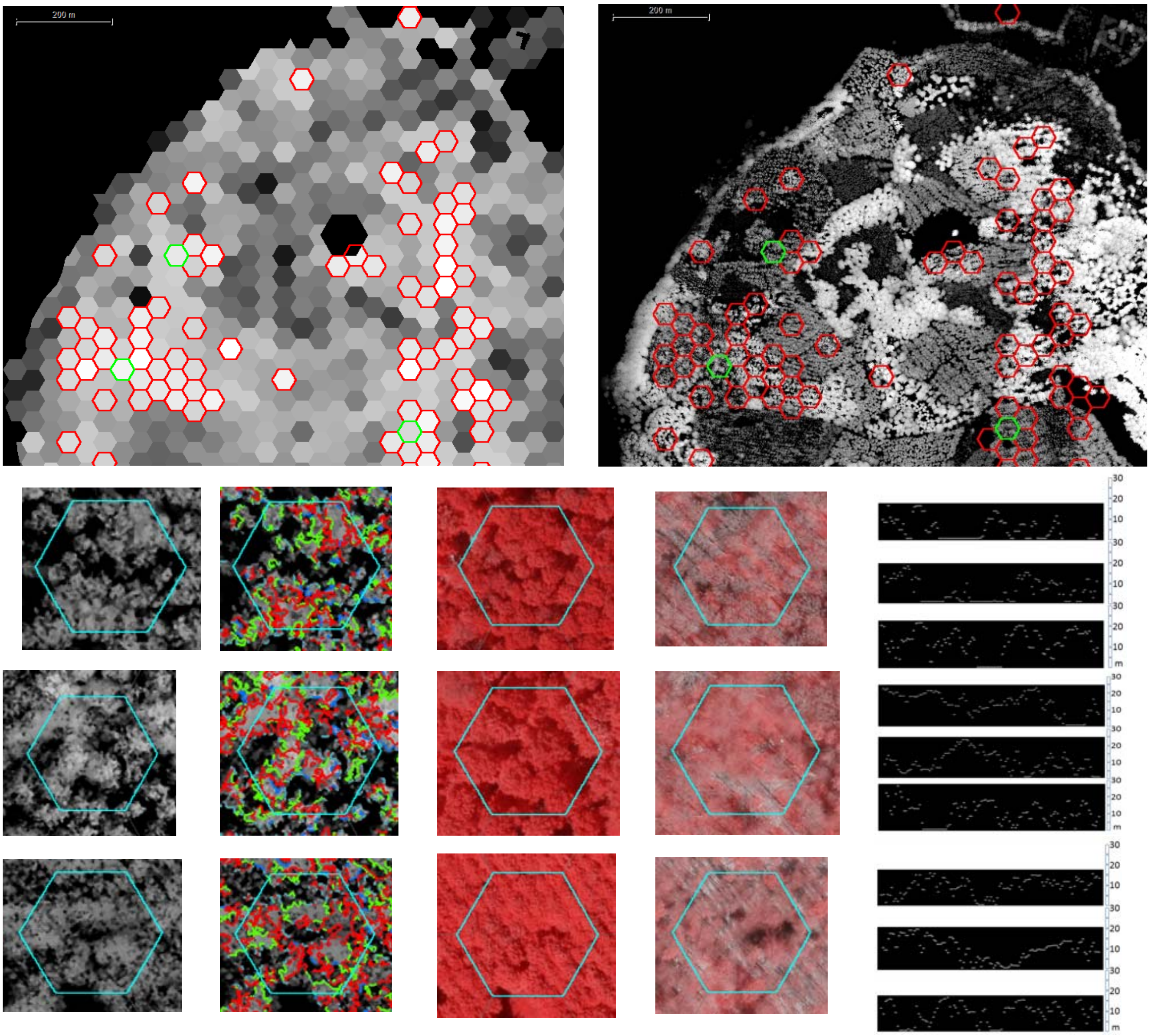


Figure 18. Relative Area of Edges: three 25 m hexagon cases with RAE > 3.0. (Same part arrangement as for Figure 17, except : vector overlay (red or green) of hexagons with RAE > 3.0).

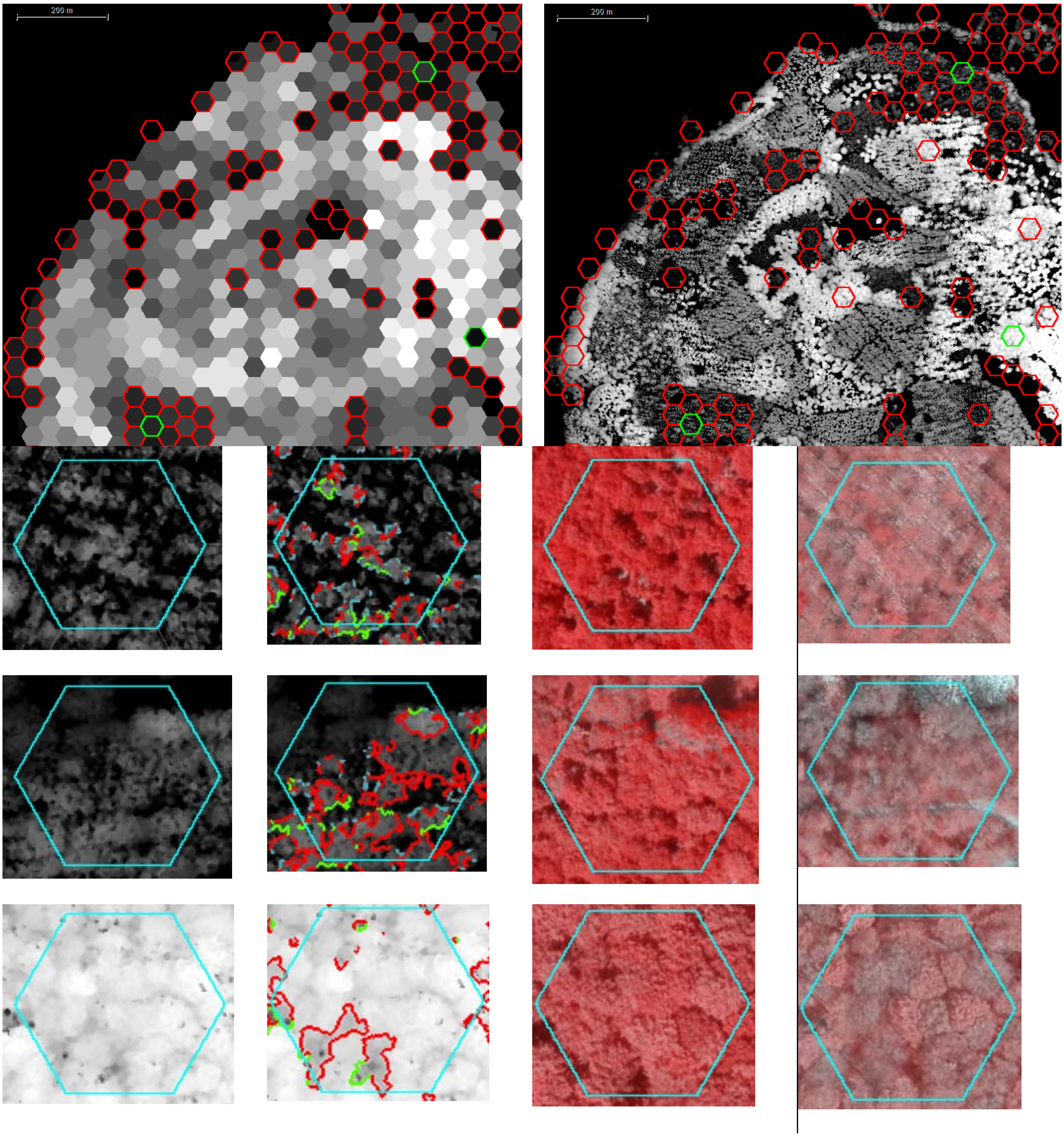


Figure 19. Variety of Edges: three 25 m hexagon cases with VAR < 5. Top row left: grey-scale representation of the per-hexagon VAR values (low→high: dark→light); Top row right: grey scale representation of the nDSM data (low→high: dark→light): vector overlay (red or green) of hexagons with VAR < 5. Rows 2 to 4: illustration of the three single hexagon shown in green in the top row: left to right: nDSM, platform interface zones, 2014 summer aerial orthophoto image data, 2016 spring aerial orthophoto image data.

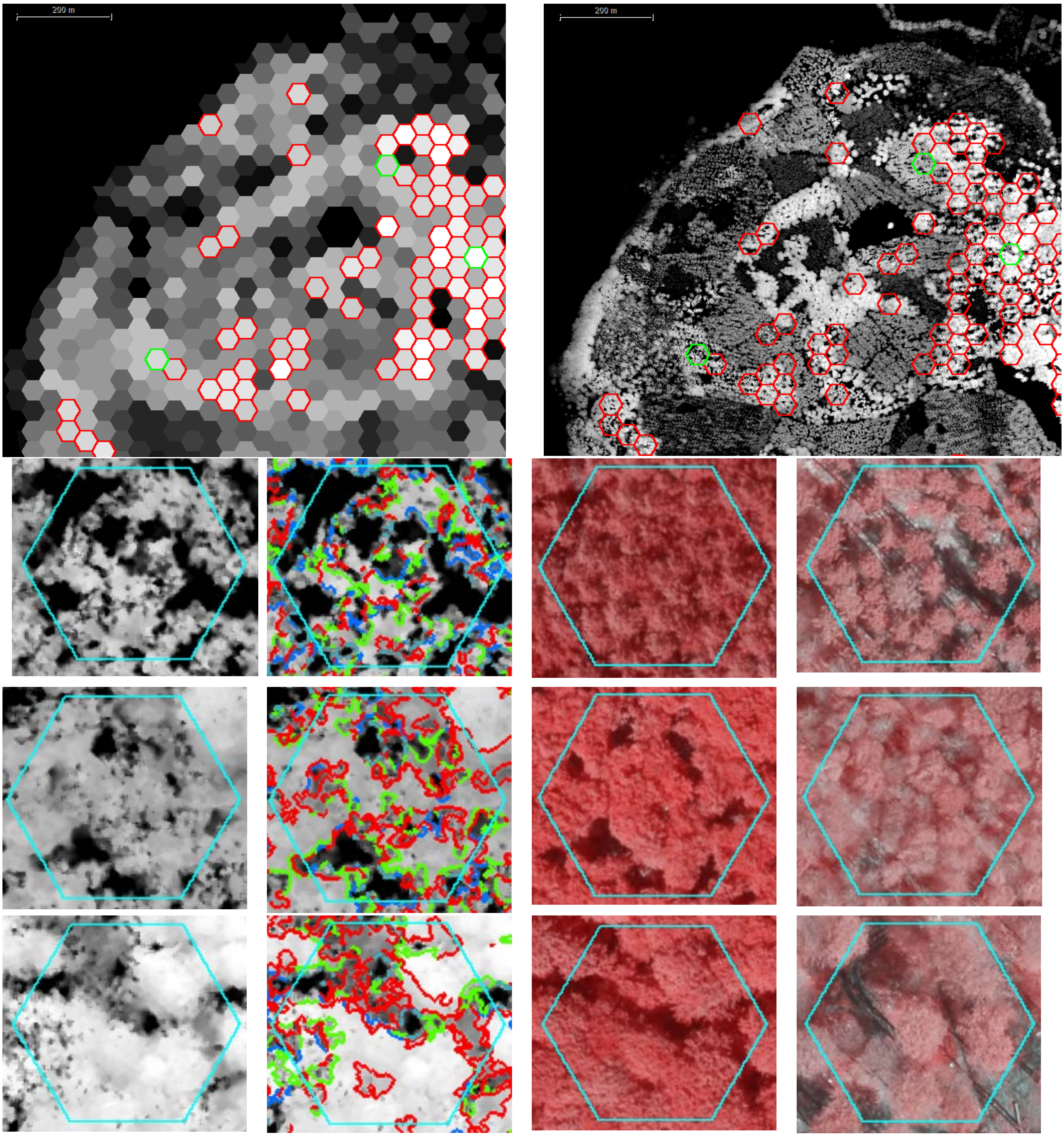


Figure 20. Variety of Edges: three 25 m hexagon cases with VAR > 15. (Same part arrangement as for Figure 19, except: vector overlay (red or green) of hexagons with VAR > 15).

The spatial spread of edges (SSE) within hexagons index was developed to be a means of identifying forest parts with marked canopy complexity but where that was mainly related to artificial management practices, such as linear tree planting in plantations. Such cases might be expected to have particular spatial arrangements of edges, which could be detected in terms of the spread of edges between the hexagon segments. The ability of the SSE analysis developed here to record that patterning was weak, and SSE was dropped as a means of making that forest form interpretation.

As an alternative to the edge based method developed here, the canopy complexity of the hexagonal units was also evaluated in terms of a form of Foliage Height Diversity index (FHD) based on the nDSM values. The FHD is a Shannon-Weiner index that was developed by the mid-20th century North American ecologist Robert MacArthur, with the hypothesis that bird species diversity is related to the vertical structure of the vegetation (https://web.stanford.edu/group/stanfordbirds/text/essays/Birds_in_the_Bush.html). MacArthur's FHD index involved use of a white board, mounted at different heights on a pole, with the proportion of it obscured by leaves at each height recorded. The proportions are combined into a single number that is high if roughly the same amount of vegetation is found at each height (grass, shrubs, and trees intermixed), and low if the foliage is concentrated at a single height -- as in a grassland or in a forest with no undergrowth. The standard formula for FHD is:

FDH = -SUM[(pi) * ln(pi)]

where pi is the proportion of the total foliage which lies in the ith of the chosen horizontal layers.

Here, the horizontal layers were taken as 3 m high. On implementational grounds the FHD(nDSM) calculation was made top→ bottom from 45 m, being a height higher than the tallest trees in Denmark.

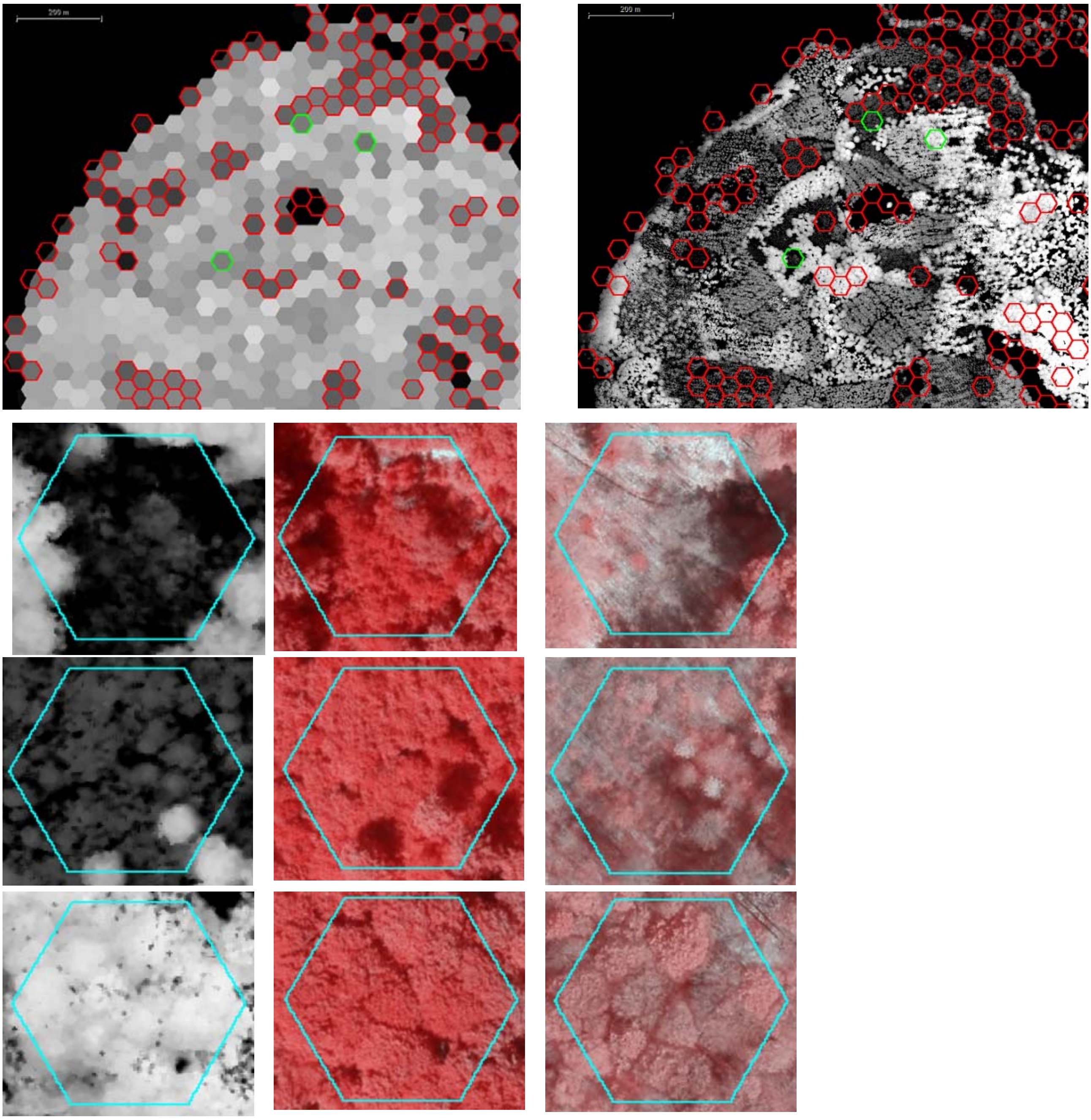


Figure 21. Forest Height Diversity (nDSM) : three 25 m hexagon cases with FHD(nDSM) < 1.5. Top row left: grey-scale representation of the per-hexagon FHD(nDSM) values (low→high : dark→light); Top row right : grey scale representation of the nDSM data (low→high : dark→light) : vector overlay (red or green) of hexagons with FHD(nDSM) < 1.5. Rows 2 to 4: illustration of the three single hexagon shown in green in the top row: left to right: nDSM, 2014 summer aerial orthophoto image data, 2016 spring aerial orthophoto image data.

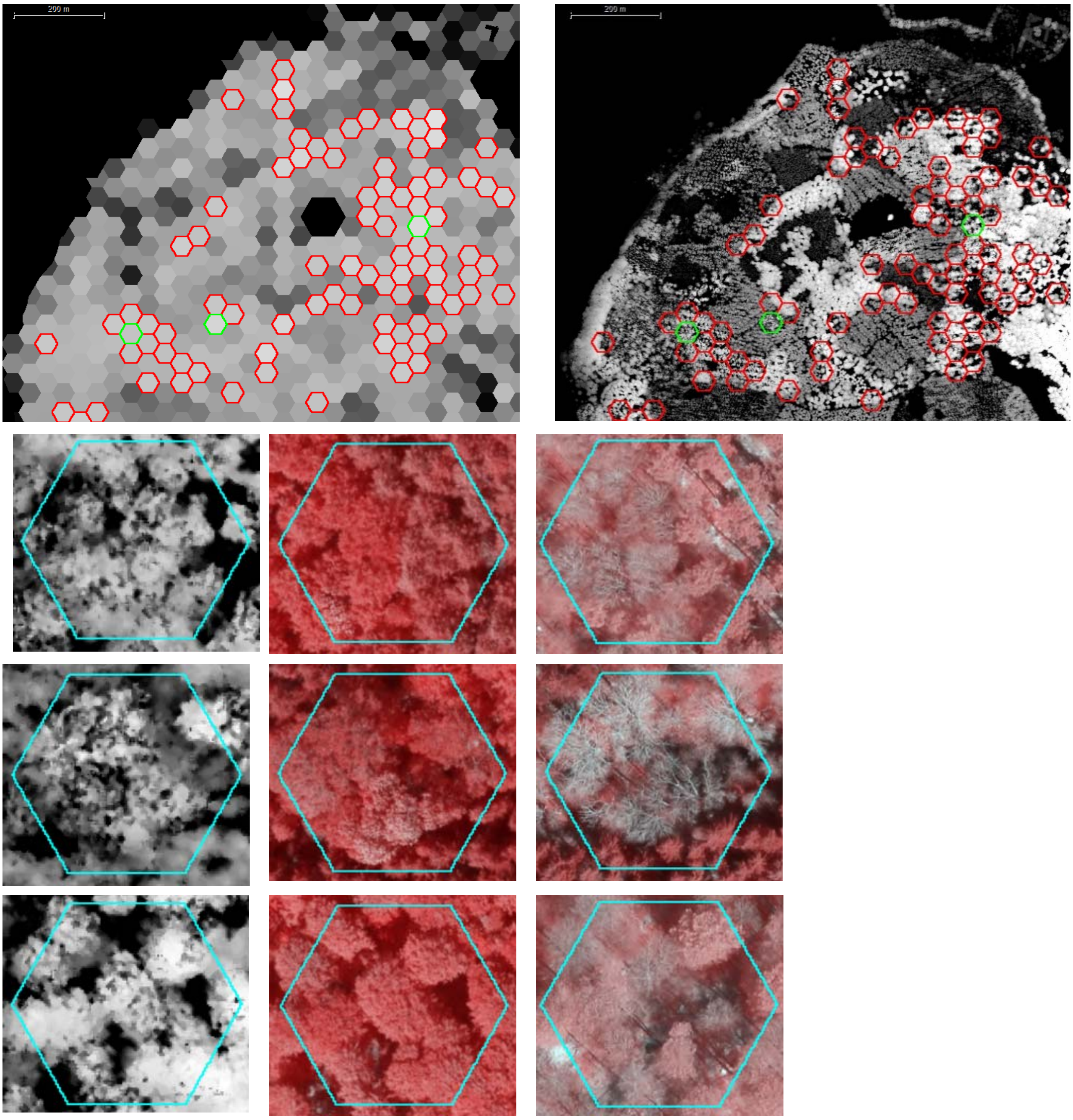


Figure 22. Forest Height Diversity (nDSM): three 25 m hexagon cases with FHD(nDSM) > 2. (Same part arrangement as for Figure 21, except: vector overlay (red or green) of hexagons with FHD(nDSM) > 2).

Initially 12 different subsets of 25m and 50m hexagons were initially tested as proxies in the Biodiversity Map based on the top 5% values of VAR, RAE and FDH, with and without a limit of at least 50% of the hexagon being covered by canopy. In the initial analyses VAR of the 50m hexagons were the best performing proxy. Subsequently we tested top 10%, 15%, 20% and 25% of VAR at the 50m hexagons for use as a proxy. The best of these (top 25% VAR values in 50m hexagons) is described as ‘Strukturskov’ in Ejrnæs et al (2018). Any overlap between the hexagons and fields with non-tree crop types were removed (i.e. field not cultivated with willow, populus etc.). The FCR proxy passed the statistical test and was included in the updated Biodiversity Map.

5. Microtopography

Whilst larger topographic features, such as banks, ditches, roads and paths, represent one layer of information that can relate to biodiversity in terms of the relative presence or absence of cultural influences (farming, forestry, transportation, etc.), other signs of such influences are related to more micro-scale topographic patterns, such as ploughing lines and plantation tree planting lines. In the current Biodiversity Map proxy developments, the larger topographic feature have been investigated via topographic map data (the proxy ‘Linjetæthed’), and the microtopography patterns via analysis of the ALS digital terrain model data (DTM). The methods developed for the latter are described in this section. One analysis of the DTM data has been undertaken to detect microtopographic features with marked linearity, and a second analysis has been undertaken to detect microtopographic features that are markedly non-linear. A high local occurrence rate of linear microtopographic features is interpreted negatively for biodiversity, and a high local occurrence of non-linear features with few linear features is interpreted positively.

5.1 Linear Microtopographic Features (LNM)

The Trimble eCognition Line Extraction algorithm was applied with 15 angular degree steps, i.e. for 8 angular line orientation ranges, to the raw DTM image data and a version of the DTM with enhancement of edges. The edge enhancement was made as the difference between the raw DTM and a 5x5 pixel kernel median filter image of the raw DTM (Figure 23). The outputs of the Line Extraction operations are a pair of new image layers, with lines represented by higher pixel values (Figure 24). Figure 24 illustrates that a higher proportion of the minor non-landscape-edge-related line features are extracted from the edge enhanced image than the raw DTM.

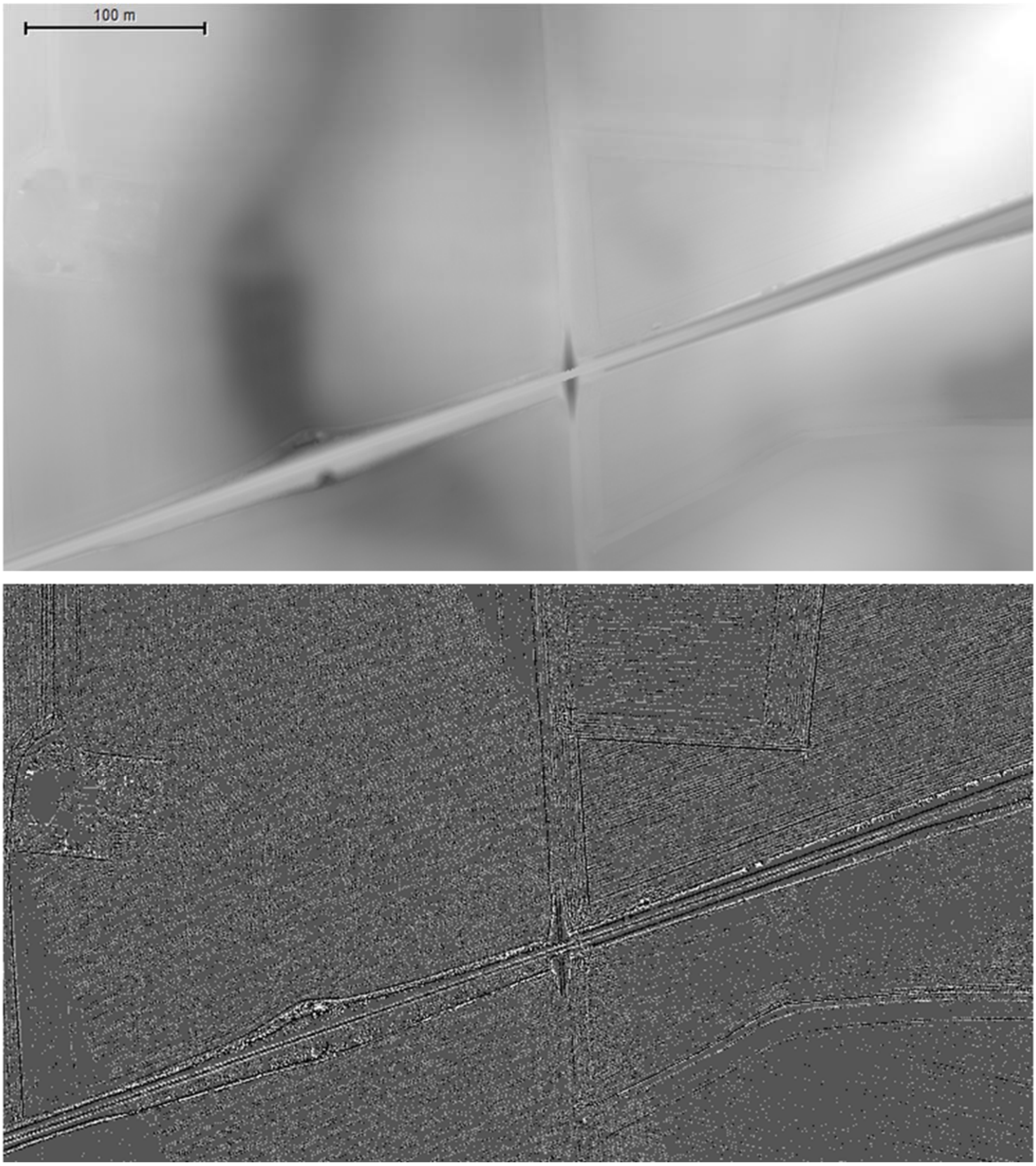


Figure 23. Example of the two image data layers analysed with the Line Extraction algorithm. Top: the raw DTMdata. Bottom: the edge enhancement of the DTM.

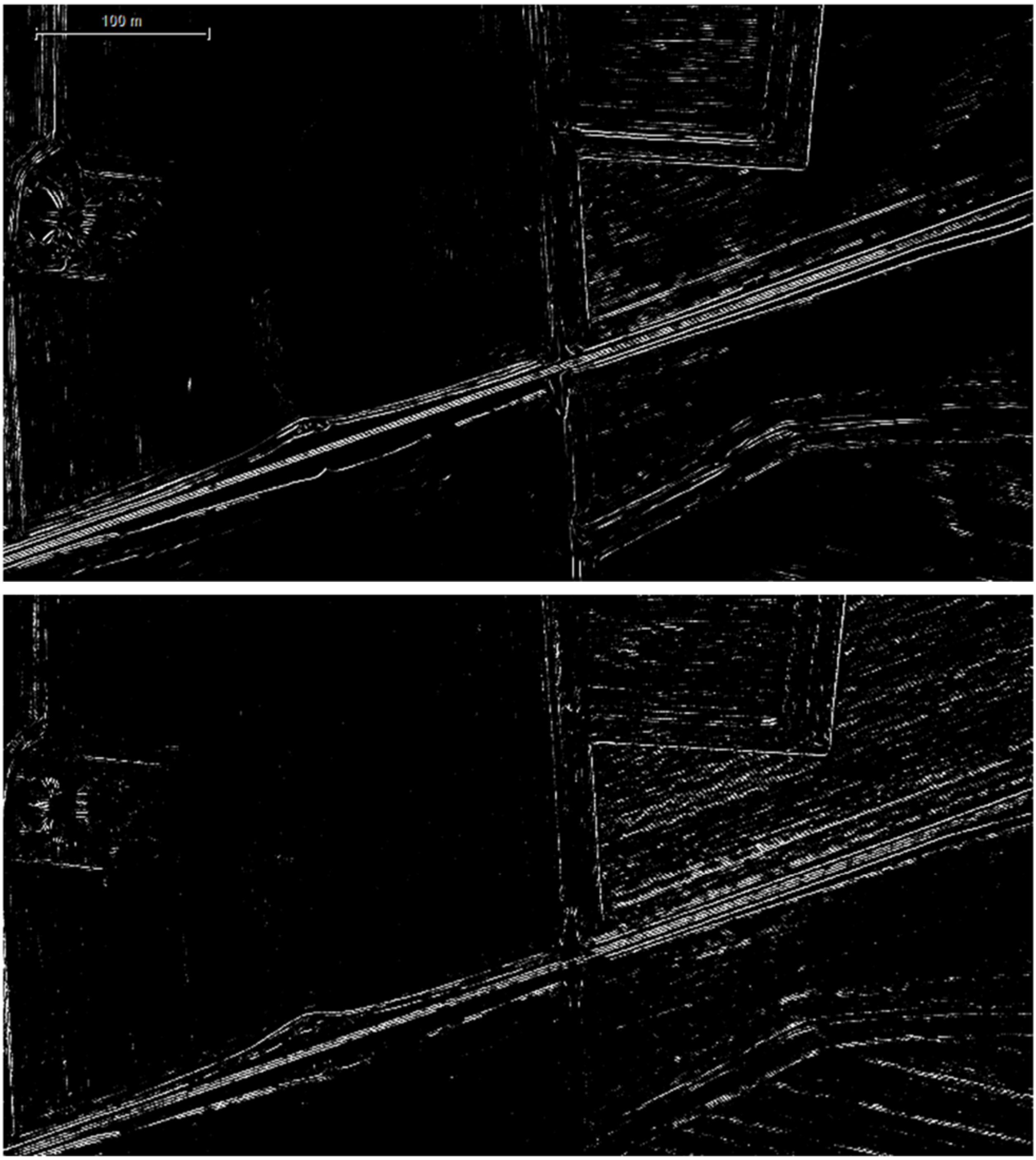


Figure 24. The linear elements found by the Line Extraction algorithm applied to the raw DTM(top) and the edge enhanced (bottom) image layers.

The two Line Extraction output images were summed and parts in the summed image with pixel values > 5 were converted to a set of objects. Objects with lengths < 5 m were then removed (Figure 25) to exclude more marked linear landscape elements such as roadsides and ditches.

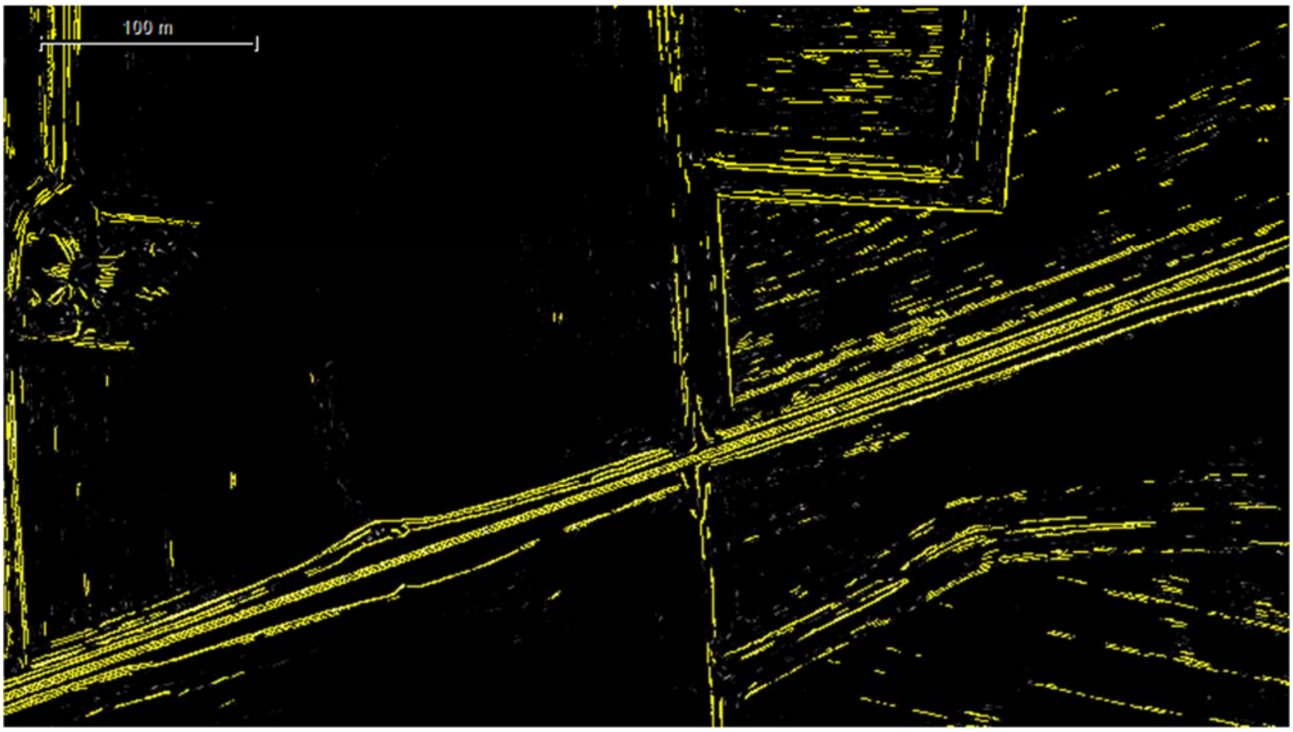


Figure 25. The extracted lines as objects, with objects of length < 5 m removed.

Filters were then applied to the set of line objects to remove several subsets that represented remaining spurious landscape features, such as broad lines, and lines with marked curvature (Figure 26). The spatial occurrence of line objects was then expressed in terms of the 25 m side length hexagons (Figure 27), for:

a) the relative area per hexagon of all line objects
b) the proportion of all lines per hexagon with main directions in each of 12 15-degree angular orientation categories (with low angular variety being interpreted as evidence of cultural management practices, such as ploughing).

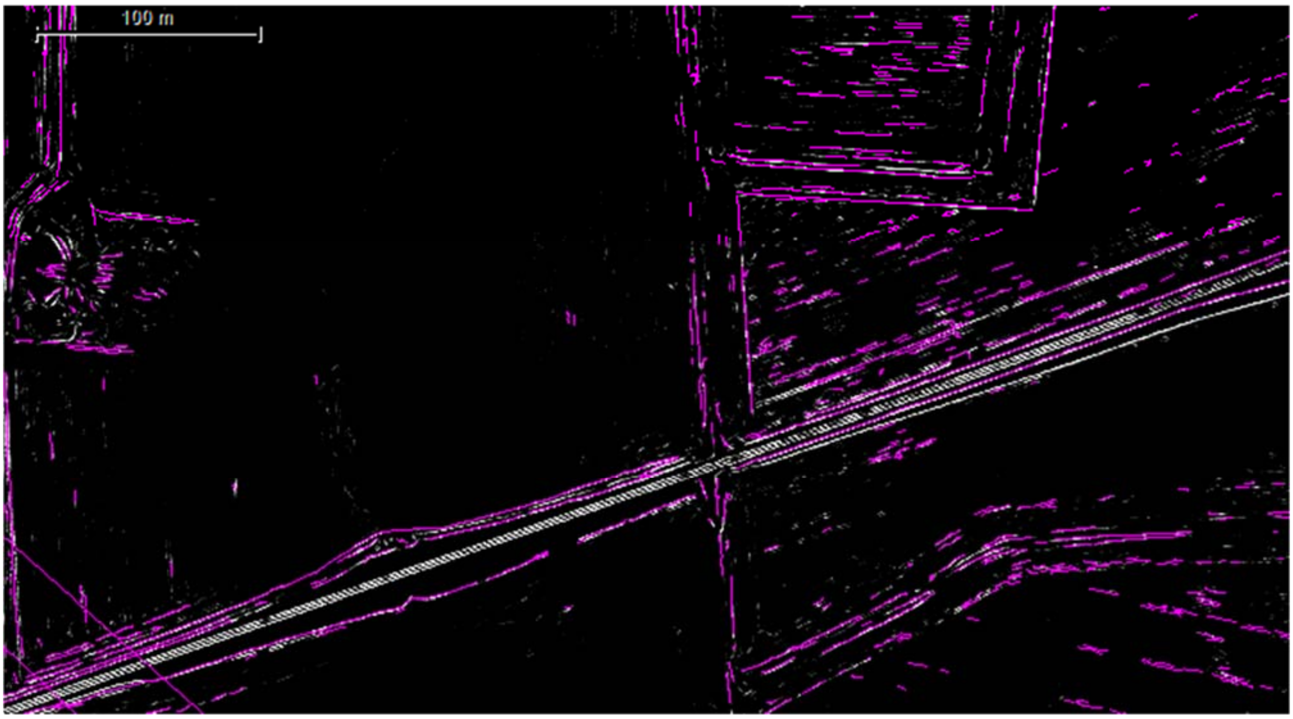


Figure 26. In purple, the linear DTM data element remaining after application of filters.

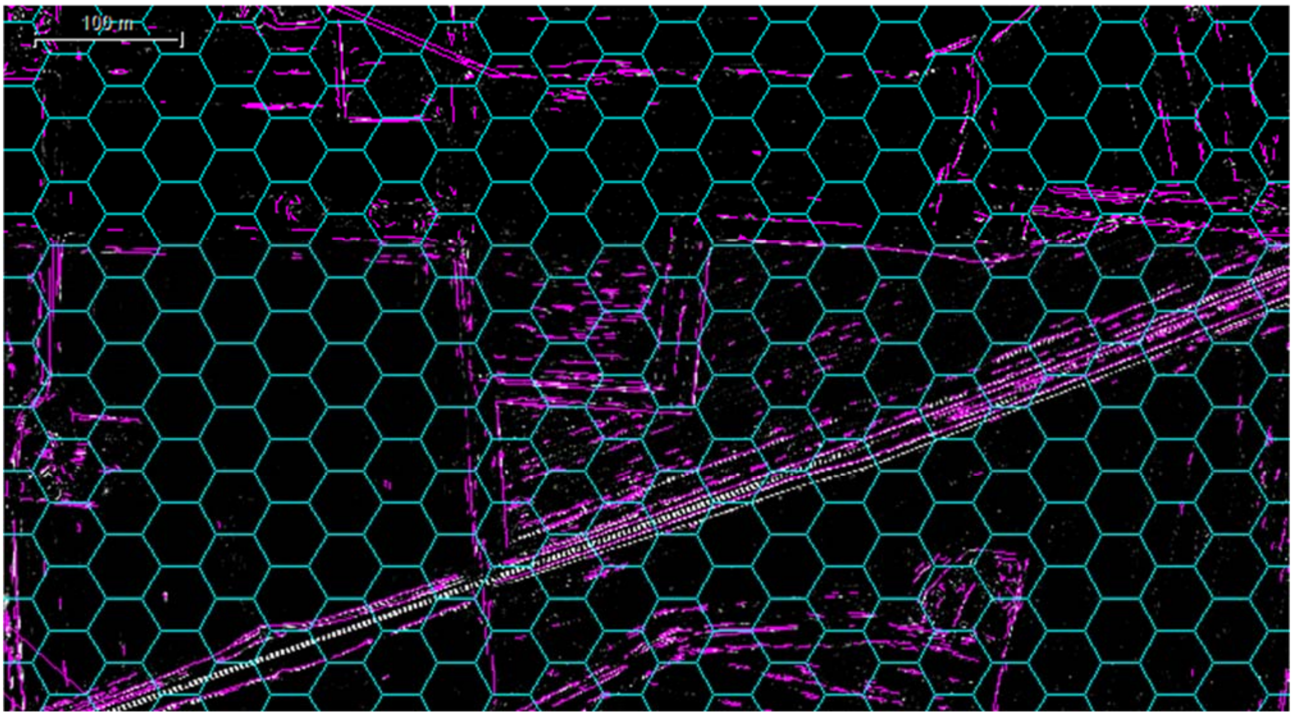


Figure 27. The LNM objects (purple) overlain with the 25 m side length hexagon analysis units.

Four subsets of hexagons were tested as proxies in the Biodiversity map based on values in the relative area of all line segments per hexagon and based on proportion of lines in two neighbouring angle classes, thereby aiming at identifying hexagons with few linear features and / or hexagons where the lines are not mainly parallel. The best performing LNM proxy actually performed fairly well in the statistical tests (Ejrnæs et al. 2018). However, it appeared that the layer included some effect of 'flightlines' (see section 5.2 for a more detailed description). Therefore, it was decided not to include the proxy in the updated Biodiversity Map until this issue has been investigated further.

5.2 Non-Linear Microtopographic Features (NLM)

To capture microtopographic features without marked linearity, a simple thresholding was applied to the edge enhanced version of the DTM data (Section 5.1), creating a set of objects of image data with values either < -2 or > +2 (Figure 28). These objects were sub-setted first in terms of their area, with the objects of =< 1 m2 then sub-setted to also exclude those smaller cases lying adjacent to or close to (=< 0.4 m) at least one of the objects larger than 1 m2 (Figure 28). In most locations, a large number of non-linear microtopographic feature objects remained after these filterings, typically > 500,000 per km2 (Figure 28). Hexagon-wise analysis was then made, with hexagons having a relative area of the remaining NLM features of > 0.01 being classified as hexagons with the NLM feature present (Figure 28). Denmark-wide analysis was made in terms of the nominally 2x2 km2 tiles, as have been described above.

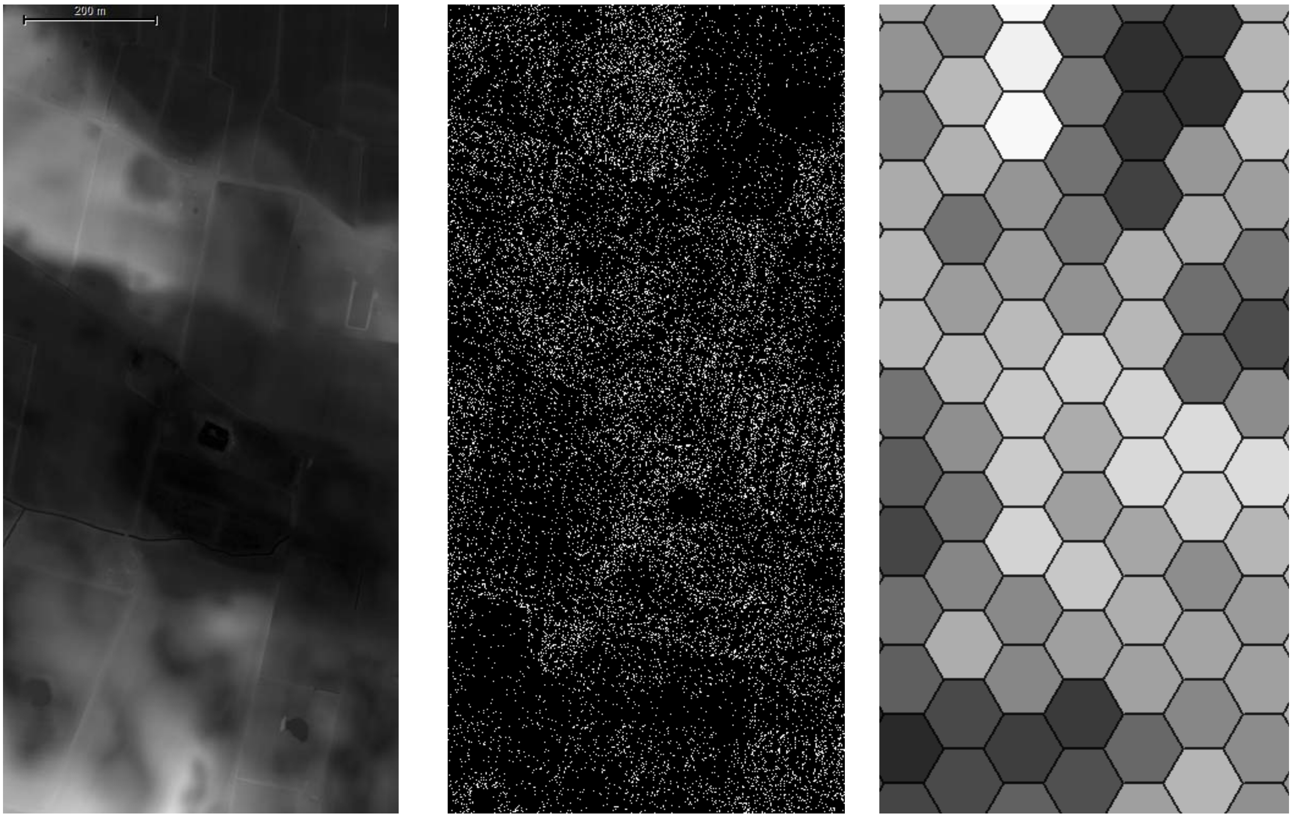


Figure 28. An example of the DTM data (left) with the set of post-filtering NLM objects (centre) and the 50 m hexagon relative area of NLM objects valuation (right) for the corresponding area, with light grey hexagons having a high relative areal of NLM objects.

When sets of neighbouring 2x2 km tiles were viewed, it became apparent that the distribution of hexagons with the NLM feature had marked spatial patterning, with this being particularly marked at a more regional scale (Figure 29). This patterning was inherent from the filtered NLM object set and also the edge-enhanced DTM data (Figure 30).

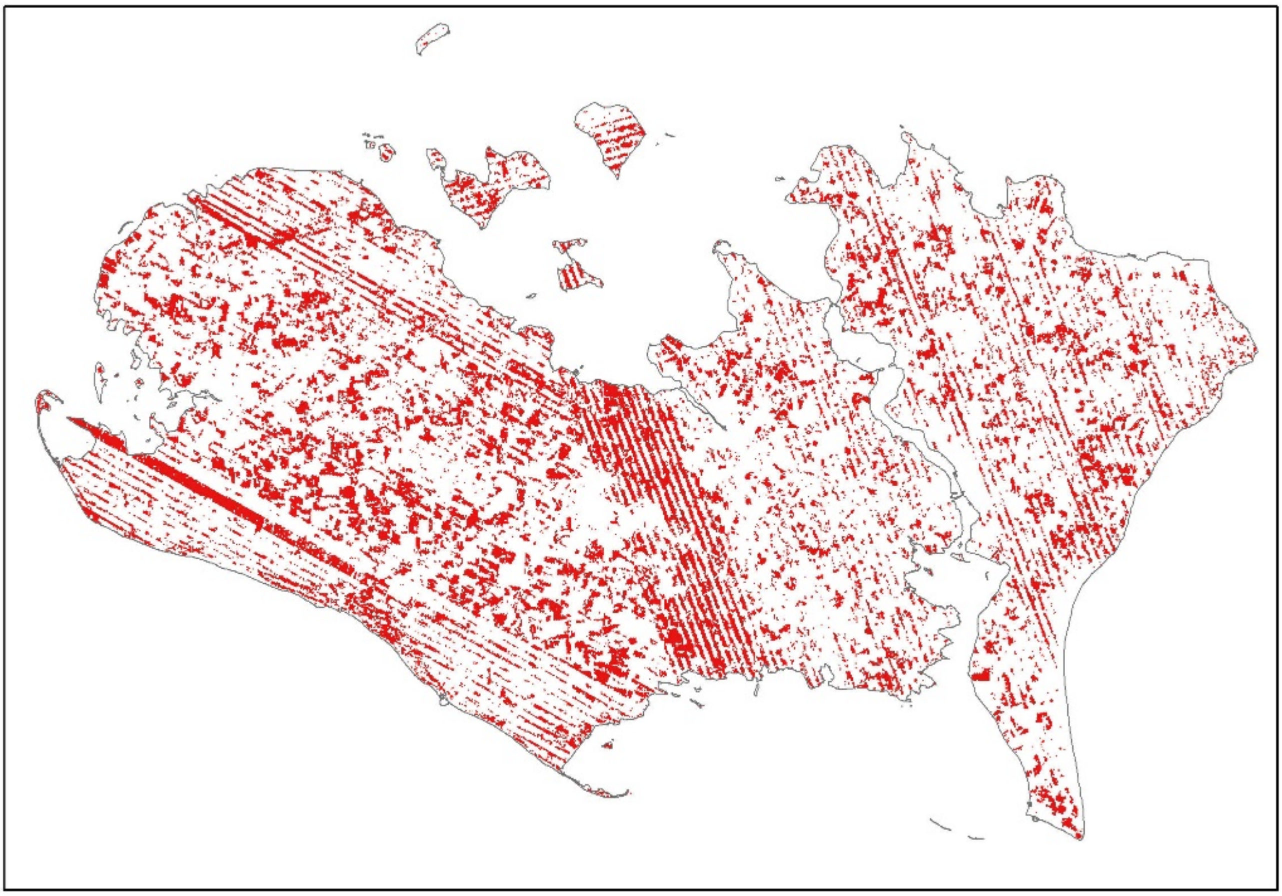


Figure 29. Spatial arrangement of 50 m hexagons with large numbers of non-linear microtopographic features for Lolland and Falster.

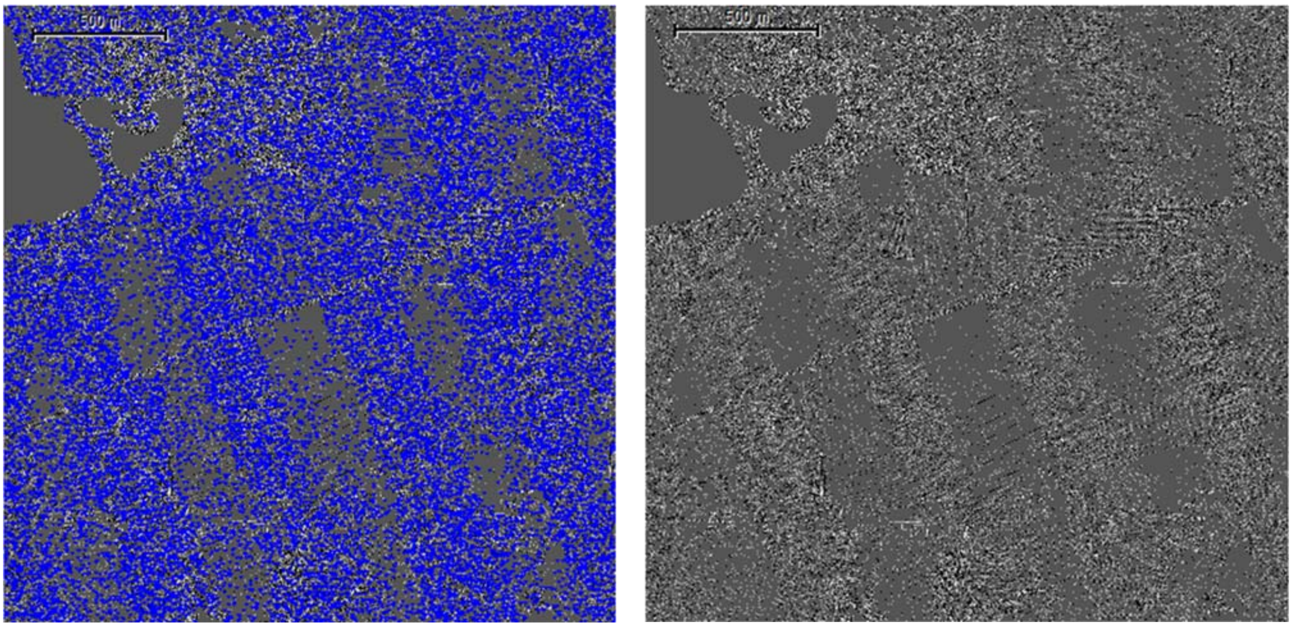
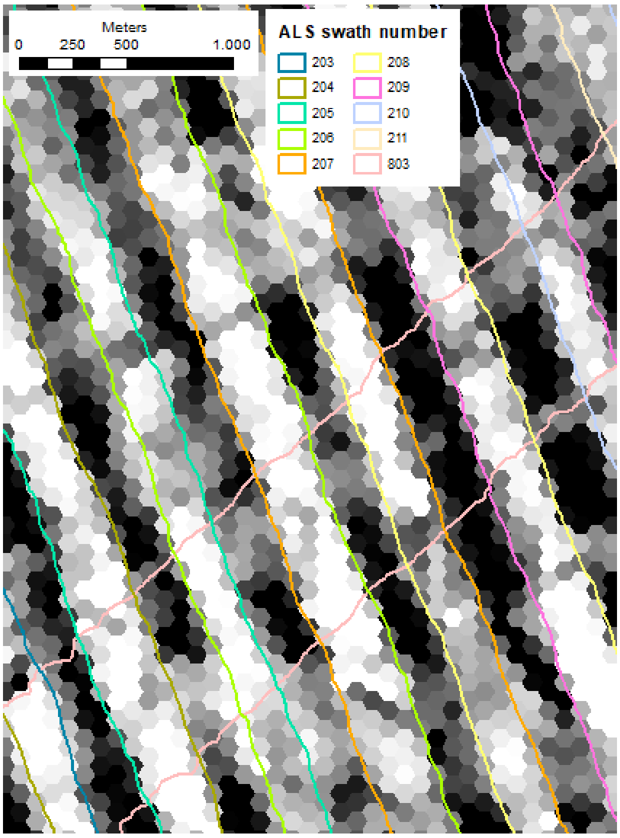


Figure 30. The NLM post-filtering objects (left) and the edge enhanced DTM data (right) for one 2x2 km tile.

The patterning appears to relate to the ALS. The ALS was undertaken to give partial (nominally 60%) sideways overlaps between swaths scanned by adjacent flightlines, resulting in along-swath zones without swath overlap and zones with swath overlap (Figure 31); cross-swaths such as swath #803 (in Figure 31), made for ALS calibration purposes, give additional overlaps, with some areas covered by three or more swaths. It seems likely that higher total ALS point cloud point densities that are associated with areas with swath overlap, result in greater local spatial detail in the DTM data. This is a known characteristic that is seen in many ALS point clouds (pers.comm A.Zlinszky, April 2018). The NLM analysis made here is sensitive to these point density related variations in the DSM spatial detail. The effect on the NLM result is so marked that it was decided that it was not meaningful to proceed with testing of the non-linear microtopogrphic feature proxy. Whilst the LNM statistical analysis gave a weak positive result (see Ejrnæs et al.2018), swath related striping was also visually apparent in the LNM data, albeit not as marked as it was in the NLM data. On this basis it was decided to not include the LNM data as a proxy in the Biodiversity Map. Analyses of the FCR were also based on DSM/DTM data but focused on much larger topographic differences. We detected no effect of the swath overlap in the FCR analyses and hence the effect appears to be relevant only when focusing on topographic differences of a few centimeters. The swath overlap related variations in ALS point density should be taken into consideration in other analysis focused on small local DTM differences.



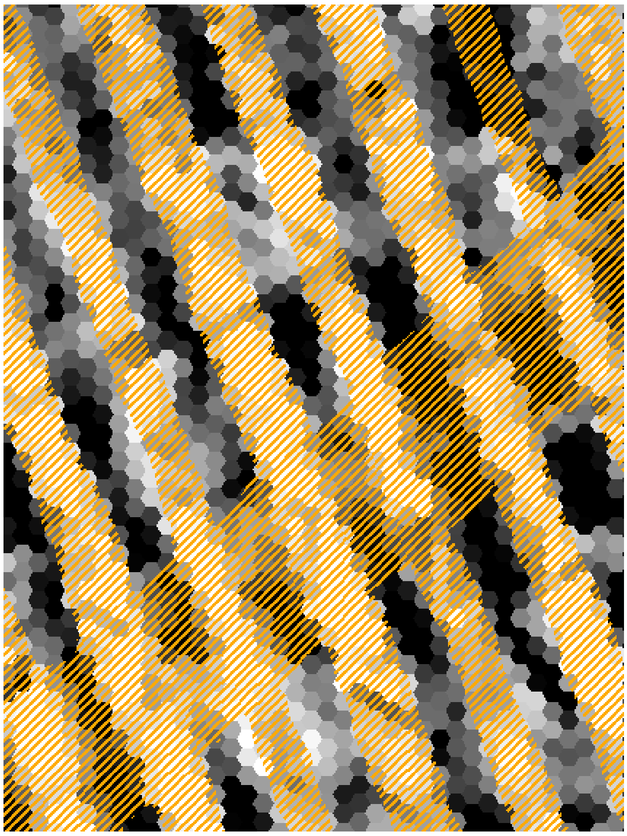


Figure 31. For a central part of the area shown in Figure 29, the 50 m hexagon relative area of NLN result overlain with the individual ALS swath (left) and the zones (with hatching) where there is swath overlap (right).

6. References

Bladt, J., Brunbjerg, A.K., Moeslund, J.E., Petersen, A.H. & Ejrnæs, R. 2016. Opdatering af lokal bioscore for biodiversitetskortet for Danmark 2015. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, 20 s. - Teknisk rapport fra DCE - Nationalt Center for Miljø og Energi nr. 74.

Brunbjerg. A.K., Bladt. J., Brink, M., Fredshavn, J., Mikkelsen, P., Moeslund, J.E., Nygaard, B., Skov, F. & Ejrnæs, R., 2016, Development and implementation of a High Nature Value (HNV) farming indicator for Denmark. Ecological Indicators. Vol 61, Part 2, p.274-281.

Ejrnæs, R, Moeslund, J. E., Groom, G. B., Brunbjerg, A. K. & Bladt, J. . 2018. Videreudvikling af lokal bioscore for biodiversitetskortet for Danmark. Aarhus Universitet DCE. Teknisk rapport fra DCE

Ejrnæs, R., Petersen, A.H., Bladt, J., Bruun, H.H., Moeslund, J.E., Wiberg-Larsen, P. &

Rahbek, C. 2014. Biodiversitetskort for Danmark. Videnskabelig rapport fra DCE - Nationalt Center for Miljø og Energi nr. 112

Rosenkranz, B. C. & J. Lund. 2015. Danmarks Højdemodel - én model med et utal af anvendelser. Geoforum Perspektiv 26

DEVELOPING BIODIVERSITY PROXIES

Technical description

In 2014, Aarhus University and Copenhagen University in collaboration developed a biodiversity map for the Danish Nature Agency (Ejrnæs et al 2014). The map consists of two parts: 1) a national prioritisation of nature areas based on a complementarity analysis of species at 10×10 km scale and 2) a local prioritisation of nature areas based on a bioscore calculated as 10×10m pixels. Calculation of the bioscore is partly based on proxies for biodiversity. This report describes the methods developed and applied for the production of national datasets of certain new proxies for testing for potential use in the Biodiversity Map. The new proxies covered in this report take their basis in a mix of airborne laser survey data products, aerial orthophoto image data and certain national geoinformation data layers. The results of the testing of the new proxies for use in the Biodiversity Map are reported in the parallel report to this report, Ejrnæs et al. (2018).

ISBN: 978-87-7156-347-4
ISSN: 2245-019X