

Geoff Groom\*, Gregor Levin\*, Stig Svenningsen\*\*, Mads Linnet Perner\*

## **Dune Sand – Object based image analysis for vectorization of a dotted signature in Danish late 1800s maps**

*Keywords:* Denmark, topographic maps, dune sand, OBIA

*Summary:* The geography and history of the 300 km long coastal zone of western Jutland are marked by the role of dune sand. Tides, currents and waves deposit marine sand as beaches. During low water, the prevailing winds and frequent storms drive the beach sand inland. For much of this coast, a zone of high dunes forms a natural coastal defense, backed by several kilometers of landscapes of windblown sand, which are some of Denmark's main semi-natural wilderness areas. These have experienced widespread modern era changes, with establishment of plantation forests and holiday homes. Documentation of the changes and analysis of their effects re-quires digital geo-data of the former extents of the dune sand. The Danish Høje Målebordblade (HMB) set of historical maps surveyed from 1842-1899 record the earlier distribution of dune sand. In these maps, dune sand is signatored as fine black or dark grey dots with a relatively even spacing. The dune sand signature itself has no colouration, but dune sand dots locally overlay the colourations applied for the HMB heath and wetland legend items. A demonstration project of the possibilities for automated vectorization to digital geo-data of HMB map land categories has had dune sand as one of its target categories. Automated mapping of dune sand has used a sequence of object-based image analysis methods, including use of a raster of the distance to each dot. The resulting dune sand vector layer, for a 200 km<sup>2</sup> coastal test area in NW Jutland has, assessed against a 100 m spaced set of control points, a false-positive rate of under 5% and a false-negative rate of around 10%.

### **Introduction**

Production of digital geo-data of the HMB mapping of dune sand is significant since the geography and history of the 300 km long coastal zone of western Jutland are chronicles of the role of dune sand. Tides, currents and waves deposit marine sand as beaches. During low water, the prevailing winds and frequent storms drive the beach sand inland. For much of this coast, a zone of high dunes forms a natural coastal defense, backed by several kilometers of landscapes of windblown sand, which are some of Denmark's main semi-natural wilderness areas. These areas have experienced widespread changes, in particular during the past 100 years, with widespread establishment of plantation forests and holiday homes. Documentation of the changes and analysis of their effects requires machine-readable geo-data of the former extents of the dune sand.

Increasingly, the conversion of paper or image media historical maps to machine-readable digital geo-data layers is moving from time and resource demanding visual interpretation of the maps and manual digitisation to less time-consuming automated production methods. In a Danish context, analyses of historical maps have most often been elaborated only for a limited number of map sheets and relatively small study areas (e.g. Fritzboeger and Caspersen 2002; Jensen and Jensen 1979; Kristensen et al. 2009; Svenningsen et al. 2015) due to a reliance on visual, manual methods. There are, however, exceptions, such as Dam (2005), who, for all of Denmark, manually pro-

---

\* Aarhus University, Denmark. Corresponding author: Geoff Groom [gbg@bios.au.dk]

\*\* Royal Danish Library, Denmark

duced geographical information from late 18th century maps. In Denmark, there is a unique opportunity to develop and apply digital map processing methods for automated extraction of land categories from historical maps as all publicly produced topographical maps are freely available as high-resolution scans from the national mapping agency (SDFE). An automated extraction of machine-readable geo-data for land categories from historical topographic maps will provide the opportunity to generate countrywide map layers, representing the location and extent of different land categories at different points in time.

Digital map processing is a relatively young research area, which grew out of disciplines such as graphics recognition and document analysis (Chiang et al. 2016, Freeman and Pieroni 1980, Liu et al. 2019). Due to various graphical quality issues and the general complexity of the map contents, extraction of information from maps is generally more difficult compared to processing of other data sources, such as written documents or technical drawings (Lladós et al. 2002). In maps, layers of geographic features, such as roads, contour lines, text labels, and land categories, often overlap with each other, increasing the degree of content complexity and colour mixing. Furthermore, due to inconsistency of applied symbols and colours, processing of manually drawn maps is even more challenging (Leyk and Boesh 2009). It is not possible therefore to directly apply general techniques for graphics recognition and document analysis to map processing. In most historical maps, colours, symbols, linear features and text items, represent explicit, overlapping layers of geographic information. Colours often represent thematic information of land categories, such as forest or water. Symbols, and to some degree text items describe different types of geographic characteristics. Linear features can represent geographic features, such as roads or watercourses, boundaries of land categories, administrative boundaries or topographical characteristics, such as contour lines. A major objective of digital map processing is to isolate spatially these overlapping layers of geographic information from each other, and to filter out those map elements that obstruct the extraction of the geographical layer of interest.

Through 2019, resources were available to enable the authors to examine the possibilities of developing, undertaking and evaluating automated production of machine-readable geo-data layers from HMB historic map digital image sources, for five target land categories: open water bodies, dune sand, wetland (mainly meadow), heath and forest (Levin et al. 2020). The topographical maps 'Høje målebordsblade' (HMB) represent the first large-scale topographic mapping of Danish territory. The HMB map series consists of about 1200 sheets, each 31.4 x 37.7 cm (covering 47 km<sup>2</sup>) in a scale of 1:20,000, surveyed between the 1860s and 1899. The HMB mapping was undertaken by the Danish General Staff as a response to a growing need for detailed topographical information about the landscape for military purposes (Svenningsen 2016). Svenningsen et al. (2020) analyses how the HMB maps can be interpreted with respect to land categories.

The applied methodology for automated production of machine-readable geo-data was one of object-orientated image analysis (OBIA), largely reliant on expert knowledge derived rule-based image segmentation and object labelling operations, rather than statistical classification methods (Levin et al. 2020). OBIA methods of that form are widely used in production of geo-data layers from image data as they enable high levels of control over mapping results (Blaschke et al. 2014, Chen and Weng 2018, Chen et al. 2018). For the production of digital geo-data for water bodies and areas of forest, heath and wetland the HMB colourations represent the major basis for formation of image objects (Groom et al. 2020).

The HMB signature of areas interpreted as the target category dune sand, is particular with respect to the signature of the other four target categories. Rather than being a signature associated with a continuous colouration (with associated symbols), the HMB signature of dune sand comprises a

pattern of fine black or dark grey dots, with an approximately even spacing, most probably added free-hand (Figure 1). The HMB signature key includes the dot signature as representative for the categories sand-dune (Danish: "klit") and wind-blown sand (Danish: "flyvesand"). The dune sand dot signature occurs both with and without coincidence to the continuous colour signatures of HMB legend items such as heath and wetland (Figure 1, right). The same dot signature is also applied for beach sand areas (Figure 1, left). More widely spaced dots of a similar size are part of the signature of several other HMB legend areal items (Figure 1, right). Locally, the limit of an area of the dune sand towards "open land" map matrix parts is marked with a line of more closely spaced black dots (Figure 1, right). Thus, successful production of digital geo-data for the dune sand target category required quite different image and GIS analysis methods from those that were successfully applied for the continuous colour target categories (Groom et al. 2020). This paper presents the methods developed in a 2019 HMB digitisation pilot project for the automated production of digital geo-data for the dune sand category, and the associated accuracy of the product.

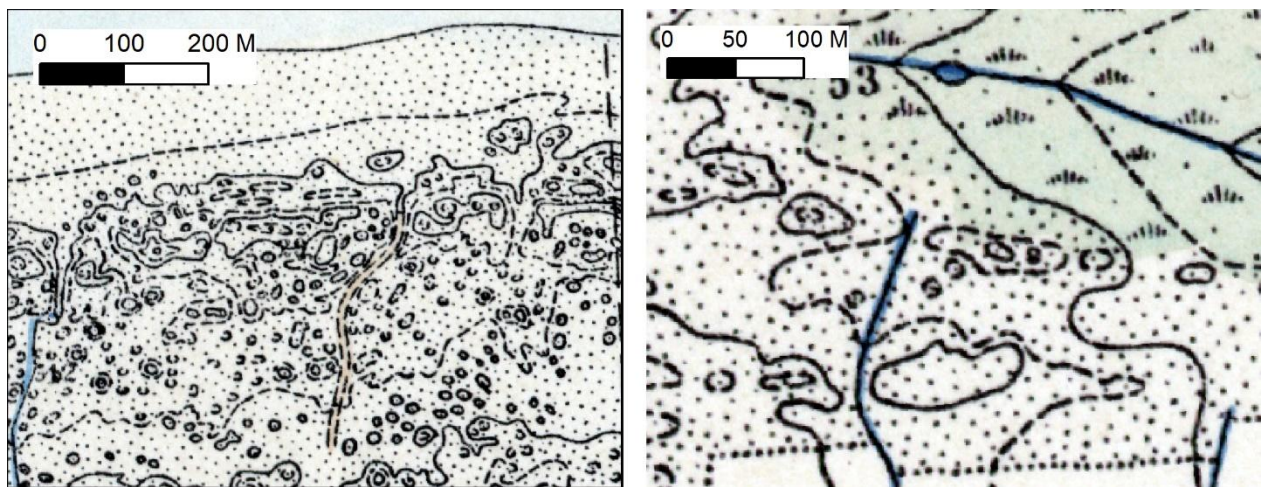


Figure 1. Examples from the HMB map series of the dot signature associated with the automated digitization target category "dune sand". Left: an example from sheet H157 with the signature applied for a beach area and an associated sand dune area. Right: an example from HMB sheet H324 of the signature applied with coincidence to the HMB continuous (light green) colouration signature for an area of "eng" (a component of the target land category, wetland). Both examples are of the KB data.

## Materials

The work reported here relates to a study area, with approximately 150 km<sup>2</sup> of land, formed of the bounding-box of the parishes of Hirtshals, Horne and Asdal, in northern Jutland (Figure 2). (A second, larger, study area of the same project, in mid-Jutland, does not include HMB mappings of dune sand.) There are two forms of HMB source material that the work reported here, has used. One is from the digital archive of the Styrelsen for Dataforsyning og Effektivisering (SDFE, the Danish national mapping agency) of scannings made of one set of HMB paper maps. These source data are publically available via the SDFE web portal (<https://download.kortforsyningen.dk/>, last accessed in May2020), as georeferenced (ETRS\_1989\_UTM\_Zone\_32N, resampled to a pixel size of 2 m) image data that have been mosaicked and then cut into a set of 20 x 20 km tiles as .tif (8-bit unsigned integer, LZW lossless compressed) files. The other source is the archive of HMB paper sheets held by the Danish Royal Library (KB) and are the original paper sheets, including their margins. Use of items from this archive in this work required scanning and geo-referencing, which was applied to just the map

sheets needed in order to cover the project's study areas. Scanning of the KB sourced map sheets used a ScannTECH 600i-fb, with the standard settings of that scanner for colour artwork. Digitisation used a resolution of 600 dpi, which is appropriate for 1:20,000 scale source material (Tobler, 1988). The KB archive paper map sheets and the map sheets that were used for the SDFE scanning are not entirely the same raw material. Firstly, only approximately two-thirds of the sheets in the KB collection are full-colour with the remainder (mainly eastern Denmark) being just 3-colour (black, blue, brown), whereas the entire set that were scanned for the SDFE dataset are full colour. Secondly, where maps of both sources are full-colour, the actual colourations differ, with evidence that they relate to two sets of paper maps that had been coloured independently, rather than just the colours differences are due to the differential effects of time and storage conditions. The differences between the data from the two sources present interpretation challenges in a few places but, countering that, the situation is one of beneficial complementarity.

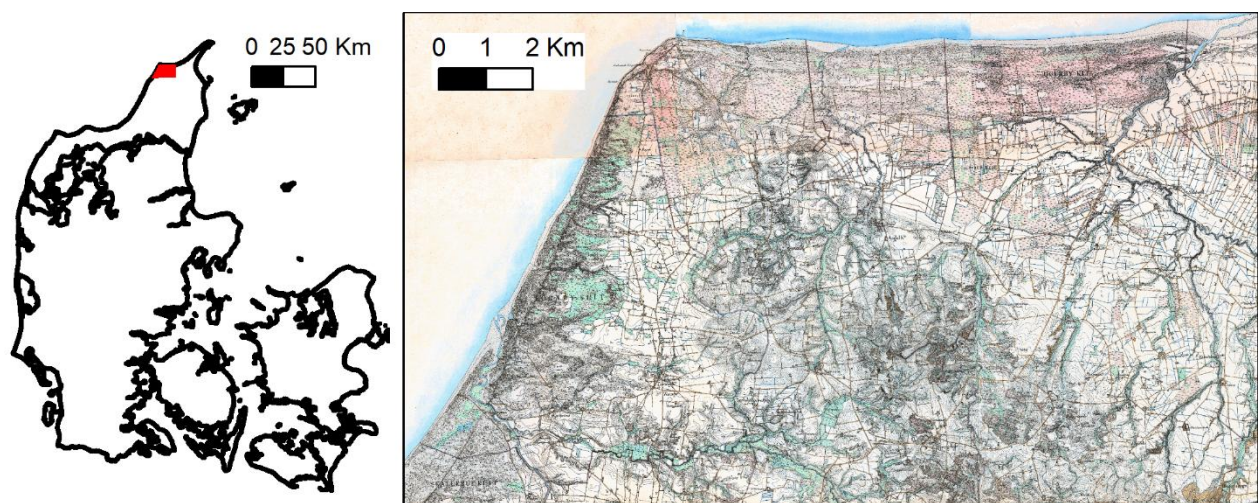


Figure 2. Location (left, red) and (right) HMB (SDFE source) overview of the north Jutland study area.

## Methods

The dune sand dot signature is carried relatively equally by the red, green and blue components of the scanned HMB map raster data. The start point of the workflow for digitisation of areas of dune sand was a derived image layer, being the Hue-Saturation-Intensity alternative colour space Intensity channel of the KB source data R-G-B image data. The Intensity channel data includes many items of the HMB map data that are, similarly to the dune sand dots, black or dark grey, e.g. text, linework, other symbols. Key components of the dune sand dot pattern that distinguish it from the other similar intensity items are (a) the size of the dots and (b) their relatively even spacing. Mapping of dune sand was made by a sequence of object-based image analysis (OBIA) processing steps that exploited those components.

The initial OBIA step isolated the dark parts, via grey-tone thresholding with 0.7 applied as a thresholding constant. An object size filter ( $< 21$  pixels) provided a set of initial dune sand dot candidate objects C1 with all other objects set as unclassified (Figure 3).



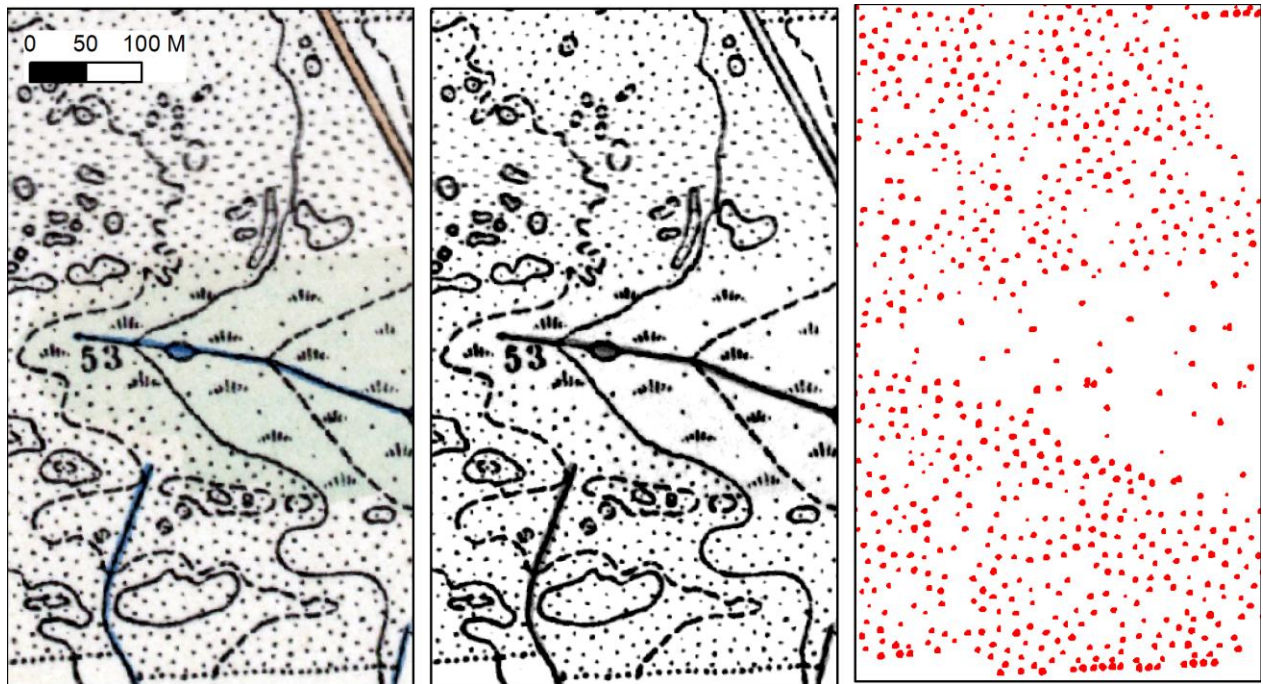


Figure 3. Left : an example sand dune area from HMB sheet H157. Centre : the KB R-G-B intensity derived image. Right : The initial set of dune sand C1 candidate objects (red).

A second filter removed any C1 object with a distance to the closest C1 object of over 20 m. A new object set (T1) was then formed as a 20-pixel coating around the C1 objects and for the extent of T1 a floating-point raster was formed of the distance of each pixel to a C1 object (Figure 4). (These two steps were the most computationally intensive of the entire workflow for the dune sand mapping, by a factor of ca. x100) The distance raster was subjected to a threshold-based segmentation (value < 11) with the resulting object set merged with the C1 objects, and all T1 objects were removed. The initial candidate object set of the dune sand area, C2, includes false-positives. Sets of clear-cut false-positive cases were filtered for via fixed threshold segmentation using other derived image layers (Figure 4).

In areas with marked sand dune development, the dune sand dots are obscured in the HMB maps by the HMB maps use of height isolines together with cross-lines between the isolines to represent steeply sloping ground. Sand dunes are therefore false-negative cases with respect to the C2 set of dune sand objects. In order to include sand dunes in the dune sand mapping, use was made of a mask image layer, also made with OBIA processing, of the HMB represented steeply sloping ground (Figure 5). Threshold segmentation was made of the mask image to give a set of objects, S, and a coating was made of the C2 objects into the S objects to give a set of C3 sand dune objects. Smaller unclassified objects that lay within C2 or C3 were merged to C2 or C3 respectively, and smaller C2 objects were merged with the unclassified objects. The objects both as a dune sand (C2) and a sand dune (C3) object set, and as a single combined object set (Figure 5) were exported as polygon data in ESRI shape format.

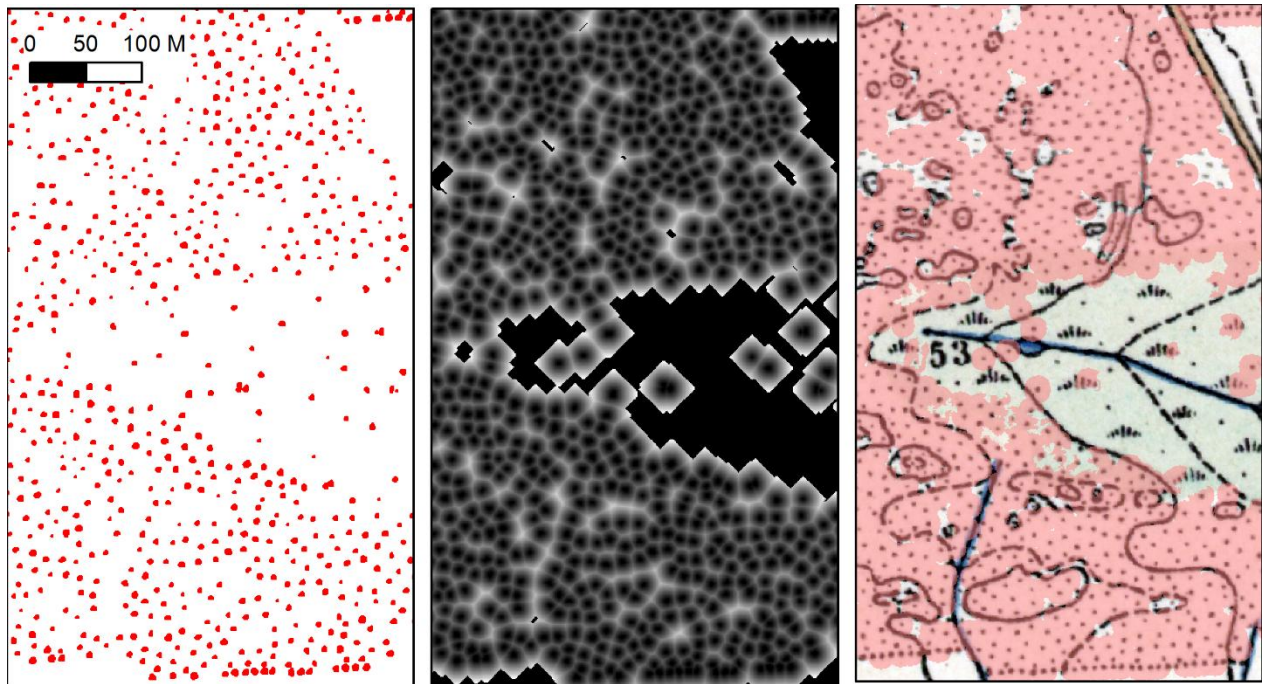


Figure 4. Left: C1 sand dune dot candidate objects. Centre: distance map to C1 objects. Right: The threshold segmented distance map objects, merged with C1 objects, and filtered for clear-cut false-positive parts to give the C2 candidate objects (pink) of the dune sand area, with the HMB KB map as the backdrop (N.B. This is not the finalised dune sand mapping.).

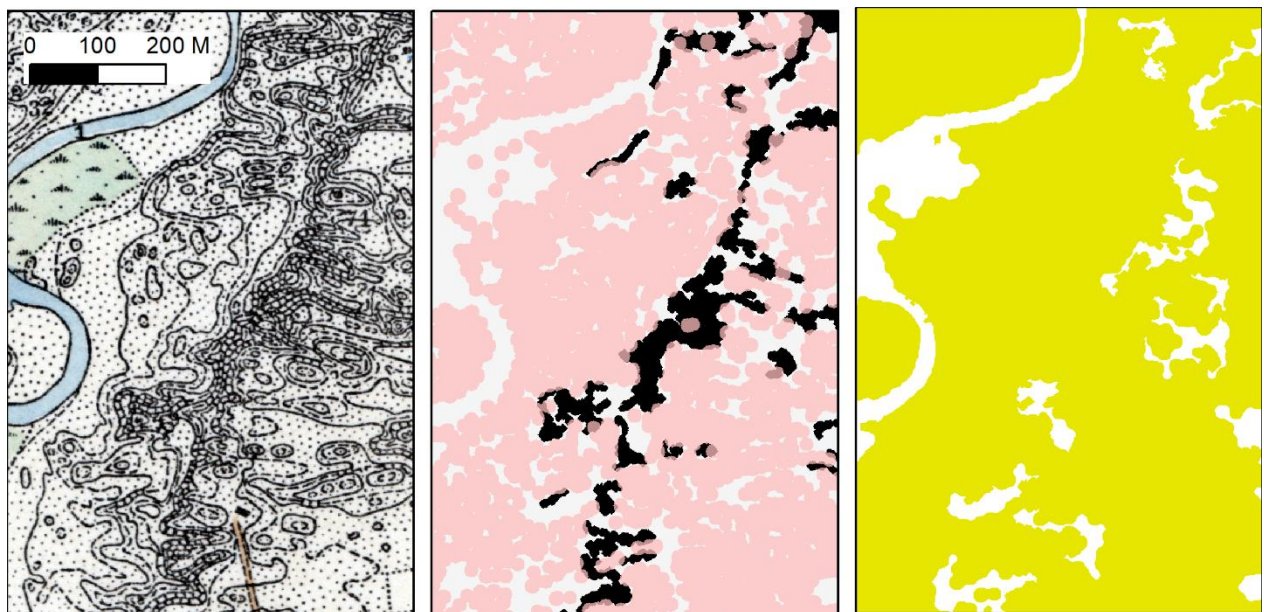


Figure 5. A part of HMB map sheet H157 (KB) with dune sand and probable sand dunes within (left). Centre: the C2 dune sand objects, with the mask image of steeply sloping ground (black). Right: the final dune sand object set (yellow), including sand dunes.

## Results

Processing, made as quarter-parts of the HMB map sheets, took ca. 15 minutes per part on a Core-i7, 32 GB RAM PC. The mapping of dune sand, including sand dunes, for the entire northern Jutland study area is shown in Figure 6. The accuracy of the geo-data layers produced for five target land categories (dune sand, water bodies, wetland, heath, forest) have been assessed with refer-



ence data derived by independent HMB map visual assessment of the target class membership for a 100 m spacing point over the central part of each of the project's two study areas (287 km<sup>2</sup>). Overall, for all five categories combined, for 17,201 reference points (points falling on black linework and text were discarded), the producer agreement level was 95.4 % with a false-positive level of 3.5 % (range: 1.0% – 6.2%) and a false negative level of 7.5% (range: 1.7% – 11.0%). For the category dune sand (including sand dune), with respect to the north Jutland study area reference data (91 km<sup>2</sup>, with 475 out of 4,999 retained points recorded as dune sand), the producer agreement level was 96.8% with a false-positive level of 3.2% and a false negative level of 11.3%. Compared to the overall accuracy, dune sand is associated with a slightly higher producer average, fewer false-positives but more false-negatives.

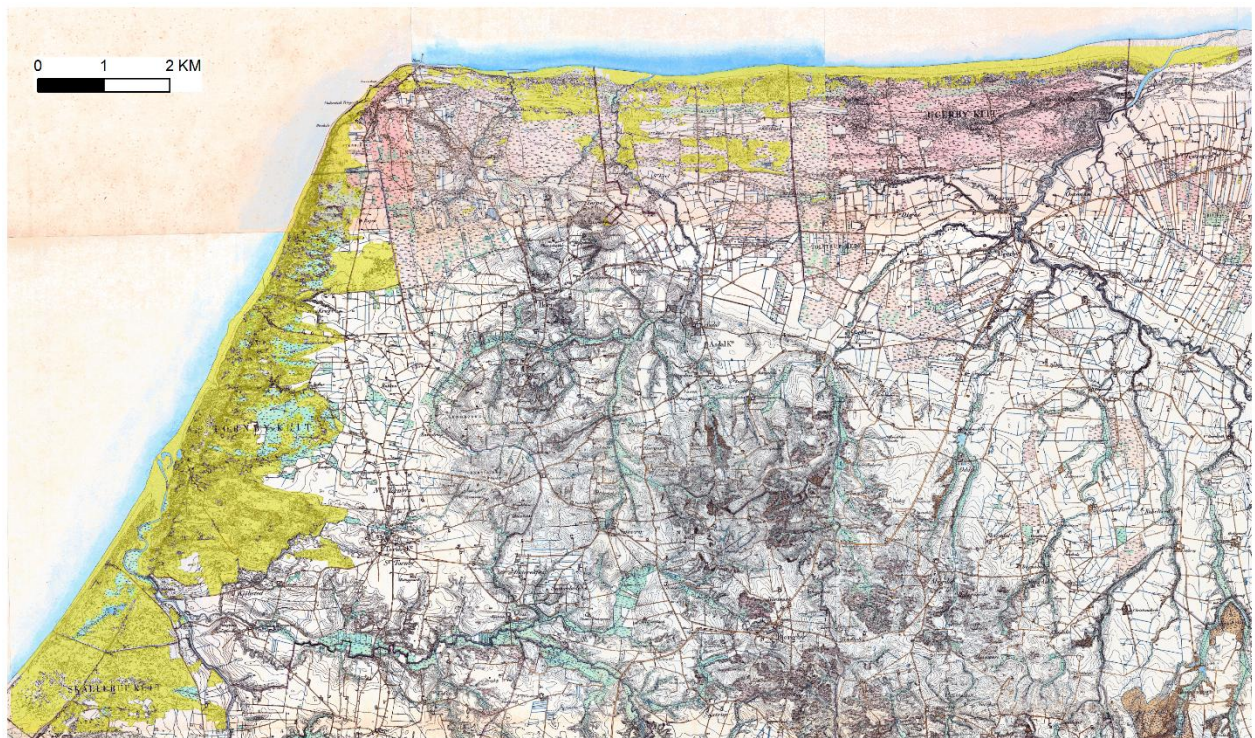


Figure 6: The HMB dune sand signature, including sand dunes, digital geo-data layer (yellow) for the entire northern Jutland study area, with the SDFE HMB map data as the backdrop.

## Discussion

The HMB map dune sand dot pattern was used in the HMB maps to denote marine beaches, seaward of the dune systems, and consequently the automated dune sand mapping includes beaches (Figure 7). That reduces the possibilities for directly comparing the former full areal extent of actual dune sand areas to modern extents. A reference for the seaward extent, such as the modern mean high-water line could be applied as a clipping mask to the HMB derived layer. However, soft coastlines, such as Jutland, are subject to relatively rapid rates of change, risking that use of modern data to edit historic data will introduce errors. There is included in the HMB maps a dashed line that lies along the shores, which represents the zero-height datum of the HMB survey (Figure 7, left). OBIA processing for expansion of the sea extent objects (already made as part of this project's preparatory processing) landward to the dash line could be applied as an HMB intrinsic reference for the dune sand extents. Greater use could also be made of the lines of closely



spaced dots that locate in the HMB maps the limit of dune sand areas towards other open land (Figure 1). The 49 points with a false-negative (FN) outcome include numerous reference points along the seaward edge of beaches (Figure 7, right), but also a number of FN point clusters. It is possible to postulate that the presence of less distinct HMB dune sand dots and broader areas with sand dunes both contribute to FN point clusters (Figure 7, right).

The methods described here have been developed and applied within a commercial software (Trimble eCognition ©, <https://geospatial.trimble.com/products-and-solutions/ecognition>). However, all of the steps could be applied via free and open-source programming languages and programming environments. The advantages of working within a commercial software are that the architecture of the component operations is programmed for fast processing, the different components are well integrated and the software's graphical user interface enables fast development of new ideas. The disadvantage, other than the cost of the software, is that the developer is limited to the set of and implementation of operations as they have been programmed within the software, some parts of which may be just barely documented.

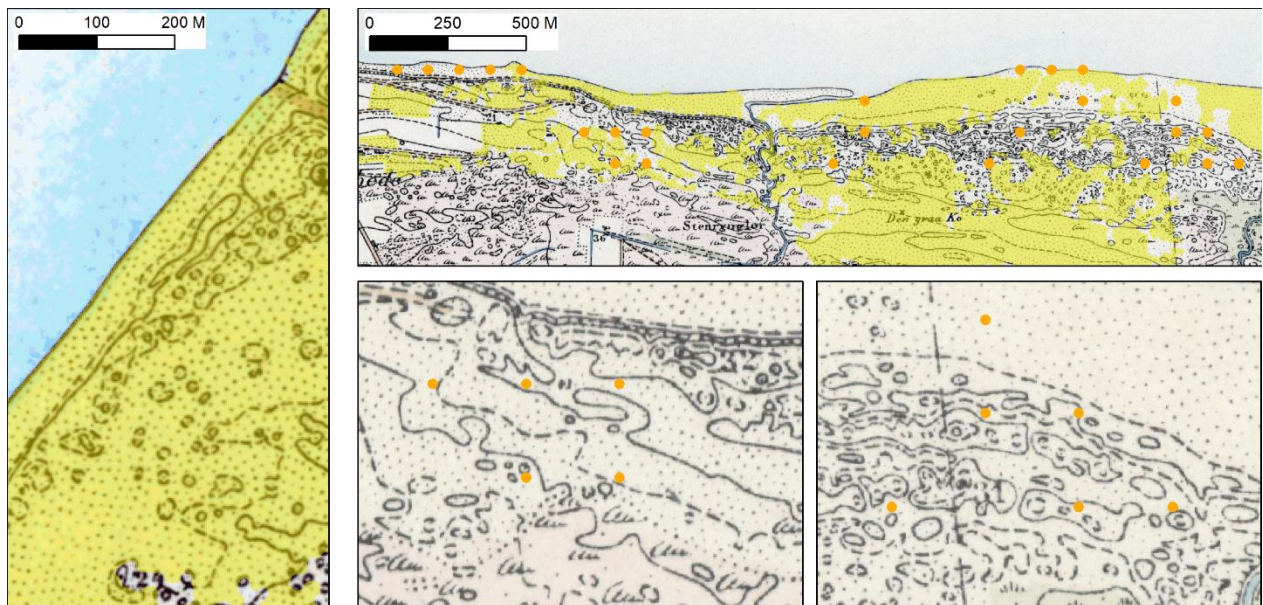


Figure 7. Left: Illustration (HMB map sheet H157, SDFE) of the inclusion of beaches in the automated production of the dune sand geo-data layer. Right: False-negative reference points (orange) for dune sand in the HMB map sheet H324, with (lower two parts) focus upon the west and east clusters of FN points; the backdrop is the KB HMB map image data.

## Conclusion

A method has been described and demonstrated to capture as geo-data polygons a historic map legend category represented by a hand produced dot pattern. In the light of the relevance of this land category for nature in Denmark, the developed methodology can provide a useful and relatively accurate means for a countrywide vectorisation of the past extent and distribution of dune sand. The same symbol is applied also for representation of dune sand in the subsequent Danish 1:20,000 topographic map series, the Lave Målebordsbladet maps from the first decades of the 1900s, opening-up the possibility for production of vector geo-data to follow the trajectories of national and local distribution changes of this land category.



## References\

- Blaschke, T., G.J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Queiroz Feitose, F. van der Meer, H. van der Werff, F. van Coillie and D. Tiede (2014). Geographic Object-Based Image Analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing* 87(0): 180-191.
- Fritzbøger, B. and O.H. Caspersen (2002) Long Term Landscape Dynamics - a 300-Years, Case Study from Denmark. *Danish Journal of Geography* Special Issue 3: 13–27
- Chen, G. and Q. Weng (2018). Special issue: Remote sensing of our changing landscapes with Geographic Object-based Image Analysis (GEOBIA). *GIScience & Remote Sensing* 55(2): 155-158.
- Chen, G., Q. Weng, G.J. Hay, and Y. He (2018). Geographic object-based image analysis (GEOBIA): emerging trends and future opportunities. *GIScience & Remote Sensing* 55(2): 159-182.
- Chiang, Y.-Y., S. Leyk, N.H. Nazari, S. Moghaddam and T.X. Tan (2016) Assessing the impact of graphical quality on automatic text recognition in digital maps. *Computers & Geosciences* 93, 21–35.
- Groom G, G. Levin, S.R. Svenningsen and M.L. Perner (2020). Machine learning helps us over the map vectorisation crux. In: K. Irás. (ed) *Automatic vectorization of historical maps: International workshop organized by the ICA Commission on Cartographic Heritage into the Digital*, Budapest, 13<sup>th</sup> March 2020. Budapest: ELTE Eötvös Loránd University: 91-100.
- Freeman H, and G.G. Pieroni, Eds. (1980) *Map Data Processing*. London: Academic Press. pp. 384
- Jensen, R.H. and K.M. Jensen (1979). Kulturlandskabet i Borris og Sdr. Felding — En kortbladsanalyse af et Vestjysk landbrugssamfund og en dokumentation for dets udvikling. *Danish Journal of Geography* 78-79: 61–99.
- Kristensen, S.B.P., A. Reenberg and J.J.D. Peña (2009) Exploring Local Rural Landscape Changes in Denmark: A Human-Environmental Timeline Perspective. *Danish Journal of Geography* 109 (1): 47–67.
- Levin, G., G. Groom, S.R. Svenningsen and M. Perner (In prep) *Automated production of spatial datasets for land categories from historical maps - Method development and results for a pilot study of Danish late-1800s topographical maps*. Aarhus (DK): Aarhus University, DCE - Danish Centre for Environment and Energy.
- Leyk, S. and R. Boesch (2009) Extracting Composite Cartographic Area Features in Low-Quality Maps, *Cartography and Geographic Information Science*, 36: 71-79.
- Liu, T., P. Xu and S. Zhang (2019) A review of recent advances in scanned topographic map processing. *Neurocomputing* 328: 5–87.
- Lladós, J, E. Valveny, G. Sanchez and E. Marti (2002) Symbol recognition: Current advances and perspectives. In: D. Blostein and Y.-B. Kwon (eds.) *Graphics Recognition Algorithms and Applications* (Lecture Notes in Computer Science, Vol. 2390). Berlin: Springer, 104–128.
- Svenningsen, S.R., G. Levin and M.R. Jepsen (2015) Decrease in Danish semi-natural grassland – a social construct or a real-world change? *Danish Journal of Geography*, 115 (2): 157-166.
- Svenningsen, S.R. (2016). Mapping the Nation for War: The Landscape in Danish Military Cartography 1800–2000. *Imago Mundi*, 68(2), 196–201.

Svenningsen, S.R., M.L. Perner, G. Levin and G. Groom (2020). “Investigating land area categories in large-scale historical topographic maps in relation to analyzing land use and land cover changes”. In A. Tsorlini, C. Boutoura (eds) *Proceedings of 15th ICA Commission Conference on Digital Approaches to Cartographic Heritage and 22nd Conference of the Map & Geoinformation Curators Group (MAGIC)*. Online, 6-7 May, pg. 164-178, ISSN 2459-3893.

Tobler, W. (1988). Resolution, Resampling, and All That. In: H. Mounsey and R. Tomlinson (eds.), *Building Data Bases for Global Science*, London: Taylor and Francis, 129-137.