Journal of Archaeological Science 40 (2013) 4688–4700

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: http://www.elsevier.com/locate/jas

Interpreting cultural remains in airborne laser scanning generated digital terrain models: effects of size and shape on detection success rates

Ole Risbøl [a,*], Ole Martin Bollandsås [b], Anneli Nesbakken [a], Hans Ole Ørka [b], Erik Næsset [b], Terje Gobakken [b]

[a] Norwegian Institute for Cultural Heritage Research, P.O. Box 736, Sentrum, N-0105 Oslo, Norway
[b] Norwegian University of Life Sciences, Department of Ecology and Natural Resource Management, P.O. Box 5003, N-1432 Ås, Norway

ARTICLE INFO

Article history:
Received 2 May 2013
Received in revised form
3 July 2013
Accepted 5 July 2013

Keywords:
Airborne laser scanning
Digital terrain models
Cultural remains
Forested areas
Detection success rates
Classification success rates
Physical properties of cultural remains

ABSTRACT

In this study, detection success rates were evaluated for cultural remains that were detected manually based on interpretation of digital terrain models (DTM) derived from airborne laser scanning data and with a resolution of 1, 5 and 10 points m^{-2}. The group of cultural remains included charcoal kilns, charcoal pits, hollow-roads, various pits, house foundations, tar kilns, grave mounds and pit-falls. The effects on the interpretation success of different types of cultural remains and their physical properties were studied: size, shape and elevation difference showing that the detection success rates varied considerably. The main tendency was that large cultural remains with clear geometrical shape (ovals and circles) and large elevation difference were much more successfully detected and classified compared to the smaller ones, especially those without a clear geometrical shape. The study also showed that it was the identification of the larger structures which profited most from an increased resolution of the DTM, and it was of no help to increase resolution in order to improve the identification of the irregularly shaped cultural remains.

1. Introduction

During the last decade, discrete pulse airborne laser scanning (ALS) has emerged as a useful technique for the archaeological community (Doneus and Briese, 2011) with regard to derivation of digital terrain models (DTMs) to be used for the detection of cultural remains and monuments. Technically, a DTM is obtained from ALS data by first classifying the echoes from each emitted pulse into ground echoes and off-ground echoes and subsequently terrain modelling from the ground echoes. Compared to other remote sensing techniques used for the detection of cultural remains, the main advantages of ALS are the possibilities to effectively identify and remove the vegetation response from the dataset and analyse the surface topography. Although removal of vegetation data from the dataset is an essential advantage, it is important to mention that dense vegetation can lead to reduced canopy penetration ability and consecutive lower point density on the ground resulting in a coarser DTM. In turn this will reduce the ability to identify cultural features. Still, the possibility of working with a 3D-model of the surface topography free from most vegetation is one of the main reasons for the widespread interest in ALS for archaeological purposes.

The point density of the ALS data will affect the quality of the DTM and thus also the detection success rates of cultural remains. Bollandsås et al. (2012) studied the relationship between detection success rates and point density (1, 5, and 10 points m^{-2}) for nine 500×500 m areas within their study area. On average 24% (ranging from 0 to 68%) of all cultural remains were detected with DTMs derived from ALS data with 1 point m^{-2}. Corresponding mean values for DTMs derived from ALS data with 5 and 10 points m^{-2} were 56% (11–90%) and 62% (11–87%), respectively. However, Bollandsås et al. (2012) did not analyze if there were differences in the detection success rates between specific types or categories (small/large, different shapes etc.) of cultural remains. Such analyses are highly interesting because correct classification and labelling of cultural remains is necessary for proper management. These analyses are also important for future studies on tuning automated algorithms for detection of cultural remains. To develop such algorithms, however, is out of the scope of the current study.

* Corresponding author. Tel.: +47 23355025.
E-mail address: ole.risbol@niku.no (O. Risbøl).

0305-4403/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jas.2013.07.002

The objective of the study at hand was twofold. The first objective was to assess if there were differences in detection success rates between different types of cultural remains that are manually detected using DTMs derived from ALS data. The data also enabled assessments of differences in detection success rates between different ALS point densities. The second objective was to assess differences in detection success rates between different categories based on the physical properties of the cultural remains. The categories were shape (circular, oval, square or irregular), size (maximum extension vertically as well as horizontally), and elevation (elevation difference from the top to the bottom of each remain).

2. Background

In many countries and regions, forested areas are those areas with the most sparse archaeological survey coverage today. Examples of the poor quality of surveys in forested areas can be found in former studies (Risbøl, 2005; Gustafsson et al., 2009). The quality of the field based surveys in forested areas with regard to proportion of known cultural remains is also lower compared to other land cover types. This makes ALS a highly relevant data source in order to improve the extent and quality of such surveys (Devereux et al., 2005; Doneus and Briese, 2006; Risbøl et al., 2006; Bofinger and Hesse, 2011; Georges-Leroy, 2011; Rutar and Črešnar 2011). The incomplete databases of cultural heritage represent a challenge to cultural heritage management, and this affects the way in which landscapes are understood by archaeologists as scenes of human activity (Risbøl, 2013). The poor quality of survey data in forested areas has a severe impact on the ability to exert an informed management in accordance with good preservation practices; the large number of new remains identified through ALS will have an important impact on the understanding and management of these areas. Some studies have indicated a substantial increase in the number of identified remains in forested areas when DTMs derived from ALS data are taken into use (Doneus and Briese, 2011; Georges-Leroy, 2011; Hesse, 2013).

It is evident from these studies that remains identified through ALS survey will rapidly come to constitute the majority of documented archaeology in forested areas. Further, given the difficulty of on the ground survey across large forested areas, it is suggested that 100% verification of the remains identified through the ALS surveys is neither practical nor desirable. It is therefore essential that the extent is established to which one can be confident in the interpretations of the ALS survey data. A number of studies have sought to establish confidence levels through comprehensive field checks. In a study carried out by Gallagher and Josephs (2008), 78% of 32 detected anomalies turned out to be cultural remains when verified in field. Furthermore, previous studies have indicated that detection success rates of cultural remains from DTMs differ for different types of cultural remains. In a Norwegian study (Risbøl, 2010), 74% of 62 charcoal pits in a forested area were detected but none of the six iron production sites (slag heaps) when ALS generated relief models were interpreted using a geographical information system (GIS). However, the results were improved when the DTM was interpreted with the use of the Quick Terrain Modeller (QTM) software (http://www.appliedimagery.com/) designed to easily handle and enhance ALS derived terrain models. Using this software in an attempt to detect cultural remains in an adjacent part of the aforementioned forested area, 83% of 20 charcoal pits and 1 out of 2 iron production sites were detected (Risbøl, 2010). These studies highlight two important points. First, that the reliability of the detection and identification of remains varies based on the character of the remains. Second, that the manner in which the data is visualized has a significant impact on the reliability of the interpretations. The present study makes a more formal and comprehensive analysis of these points.

Even though some studies have pointed out that small and less distinct remains are hard to detect (Doneus and Briese, 2006; Risbøl et al., 2006; Hesse, 2010; Shaw and Corns, 2011), no studies have been conducted so far that specifically address the relationship between detection success rates and physical properties with the cultural remains in forests, for example size and shape, although it is touched upon in some publications (Sittler and Schellberg, 2006; Gallagher and Josephs, 2008; Risbøl, 2010; Bofinger and Hesse, 2011; Georges-Leroy, 2011). In addition, some cultural remains with certain properties can be difficult to detect even by means of DTM interpretation and some of those cultural remains actually detected can be erroneously classified. For example, in some cases grave mounds have been misinterpreted as natural terrain elevations (Bofinger and Hesse, 2011) and similar confusions have occurred with slag heaps (Risbøl et al., 2006). However, Risbøl (2010) found that once an anomaly had been detected and defined as reminiscent of human activity, the classification accuracy was high.

As indicated earlier, interpretation of DTMs derived from ALS data in order to detect cultural remains is still a relatively new method in archaeology. Even though technically advanced approaches have been developed in recent years in order to enhance the visualising of DTMs, by, for instance, creating hill-shaded representations from multiple view directions (Devereux et al., 2008), or by the use of Local Relief Models (Hesse, 2010) or sky-view factor (Kokalj et al., 2011), most archaeologists using ALS generated DTMs in their work use hill-shaded images as their primary visual cue to ALS data (Štular et al., 2012). A comparative study of a series of different visualisation techniques have been published by Challis et al. (2011) and Bennett et al. (2012). Thus, more studies that assess the quality of such interpretations are needed, especially studies that assess the relationship between detection success rates and the different types of cultural remains as well as different categories according to size and shape. Furthermore, it is also important to increase knowledge about which types of cultural remains or monuments are most often confused. Misinterpretation can cause a poor understanding of the cultural history of the landscape. Accurate information on the presence and absence of the different cultural remain categories is of high relevance for improving the understanding of how landscapes were utilized by humans in the past and how single cultural remains enter into a larger context. Further, correct classification of cultural remains is also important for understanding the time-depth of a landscape and how human impact on the landscape changes with time. Additionally, the fact that some remains might be protected by law while others are not, emphasizes the importance of right classifications. Therefore, in many instances exact information is needed in order to comply with the legal regulations concerning the management of cultural heritage.

Thus, in the present study we wanted to evaluate the detection success rate for a range of cultural remains situated in a forested area.

3. Material and methods

3.1. Study area

The study area is situated in Eidsvoll and Nannestad municipalities, 40–50 km north of Oslo in the south-eastern part of Norway (60°15′N, 11°10′E, approximately 200 m above sea level (Fig. 1)). Within this area land-cover is mostly forest (70%) dominated by conifer species. The area was chosen because it was expected to hold a variety of cultural remains within a relatively

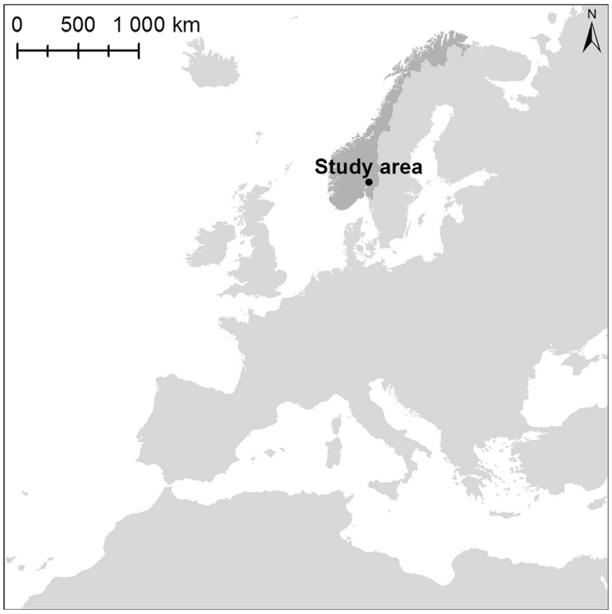

Fig. 1. The geographic location of the study area.

concentrated area. This presumption was based on existing knowledge from a few small scale surveys carried out in the area previously and large scale surveys and archaeological excavations in adjacent areas (Helliksen, 1997).

3.2. Field data

A systematic field survey was carried out in nine square blocks of land, each covering 500 × 500 m. This survey resulted in 334 detected cultural remains and monuments (Table 1), all situated in forest and belonging to nine different types and categories according to size and shape (Tables 2–4). The position of each cultural remain and monument were determined using a real time kinematic Global Positioning System (GPS) + Global Navigation Satellite System (GLONASS) receiver. Otherwise the field work was carried out in a conventional manner. For further details, see Bollandsås et al. (2012).

3.3. Previous human activity in the study area

The large amount of cultural remains within the study area indicates an intensive and varied human exploitation of forested areas throughout the Iron Age (500 B.C.–A.D. 1050), medieval (1050–1550) and historic times (after 1550). This is in fact a quite common phenomenon in areas covered by coniferous forest in mid-Scandinavia (Jacobsen and Follum, 1997; Skogsstyrelsen, 2003). Most of the remains found within the study area are related to utilization of the forest (the trees) as biomass fuel from the Iron Age and onwards. This is the case with the more than 200 kilns found in the area which were used for producing charcoal needed in iron production at the Eidsvoll ironworks established in 1624. This ironwork was operational more or less continuously for two centuries until it was closed down in 1822 (Holmsen, 1961a, 1961b). Large amounts of charcoal were needed in the blast furnaces in order to achieve sufficiently high temperatures when producing iron from ore.

There were seven house-foundations found in the study area, and all were quite small and situated near charcoal kilns. Some of them might have been in use by workers producing charcoal which was a task that required them to stay in the forest while the burning went on – a task that normally would last 2–3 weeks (Larsen, 1996).

Due to a larger number of excavated and dated charcoal pits in south-eastern Norway it is known that charcoal was produced in pits in the period A.D. 600–1500 (Larsen, 2009) and thus predates the production in kilns, proving a need for fuel prior to the time of the Norwegian ironworks. Charcoal produced in pits was either used in pre-industrial iron production or in late Iron Age/early medieval farm smithies (Larsen, 2009). In the study area no indications were found of pre-industrial iron production (i.e. bog-ore

Table 1
Summary of the different types of cultural remains registered by conventional field work. The column labelled "Shapes" displays all the different shapes that the different types had. The number of that specific shape appears in brackets. Total number of all the different types appears in the rightmost column.

Type	Shapes	Sum
Charcoal kilns	Circle (*197*), Oval (*1*), Square (*6*), Irregular (*1*)	205
Charcoal pits	Circle (*38*), Oval (*5*), Square (*5*)	48
Pitfalls	Circle (*3*), Oval (*21*), Square (*1*)	25
Grave mounds	Circle (*14*)	14
Tar kilns	Circle (*4*), Irregular (*10*)	14
Hollow roads	Line (*11*)	11
Various pits	Circle (*8*), Square (*1*), Irregular (*1*)	10
House foundation	Oval (*1*), Square (*6*)	7
Sum		334

Table 2
Number of field-observed cultural remains distributed on size- and elevation classes. Elevation is measured from the top to the bottom of the features.

		Elevation class			Sum
		0.0 to <0.5 m	0.5 to <1.0 m	≥1.0 m	
Size class	0 to <4 m	8	3	1	12
	4 to <8 m	14	37	6	57
	8 to <12 m	2	30	10	42
	12 to <16 m	8	19	6	33
	≥16 m	40	113	37	190
	Sum	72	202	60	334

Table 3
Number of field-observed cultural remains distributed on shape- and elevation classes.

		Elevation class			Sum
		0.0 to <0.5 m	0.5 to <1.0 m	≥1.0 m	
Shape	Line	7	4	0	11
	Circle	51	162	51	264
	Oval	0	23	5	28
	Square	11	7	1	19
	Irregular	3	6	3	12
	Sum	72	202	60	334

Table 4
Number of field-observed cultural remains distributed on size classes and shape.

		Shape					Sum
		Line	Circle	Oval	Square	Irregular	
Size class	0 to <4 m	0	7	1	3	1	12
	4 to <8 m	0	34	11	10	2	57
	8 to <12 m	0	20	16	3	3	42
	12 to <16 m	0	28	0	1	4	33
	≥16 m	11	175	0	2	2	190
	Sum	11	264	28	19	12	334

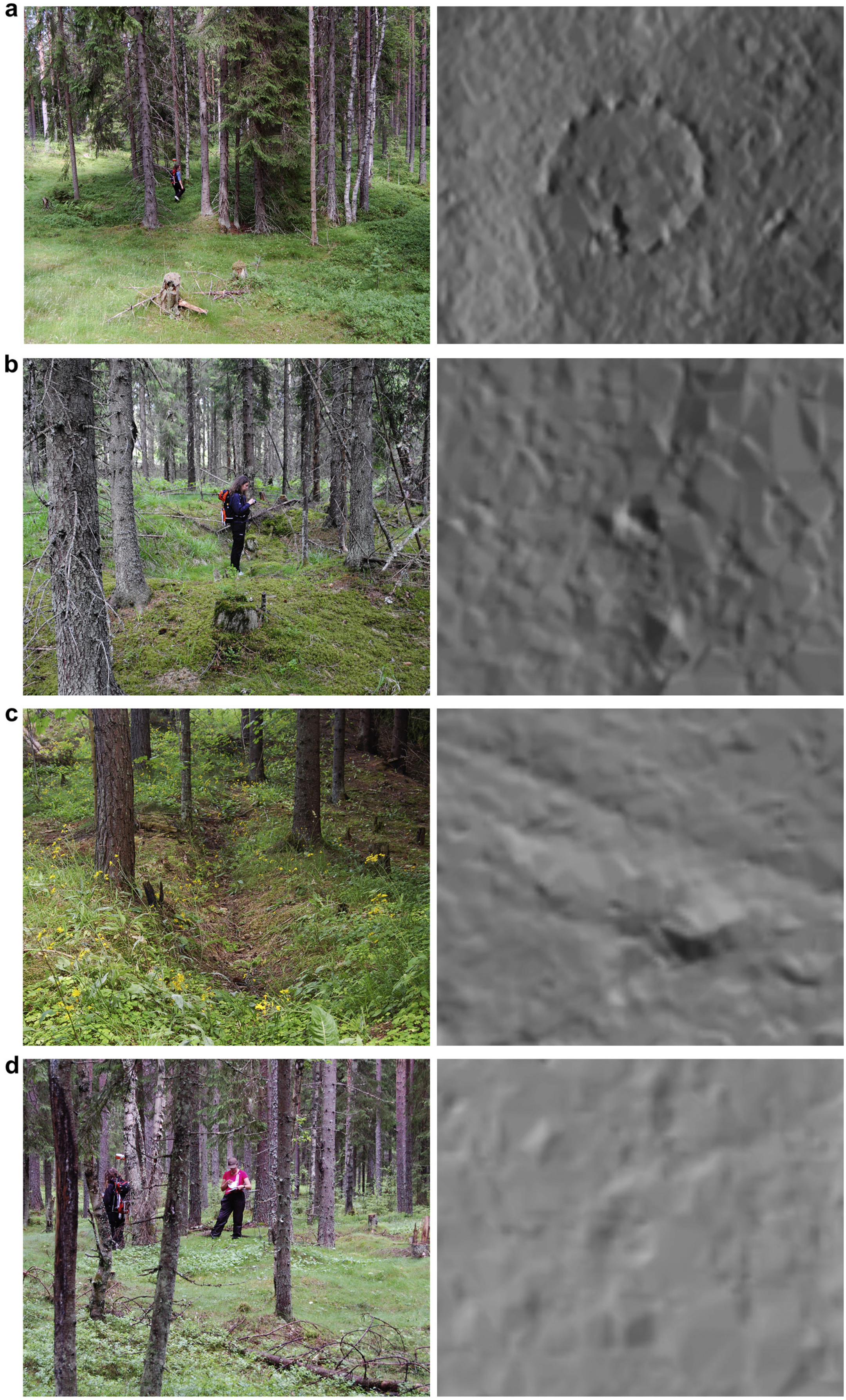

Fig. 2. a. A charcoal kiln measuring 20 m in diameter. b. A charcoal pit measuring 6.3 m in diameter and 0.7 m deep. c. A section of a hollow road. Length 7 m, width 1 m and depth 0.2 m. d. A house foundation measuring 5 × 8 m. e. A pitfall which is 0.9 m deep and measuring 7.7 × 10.2 m. f: A 1.5 m high grave mound measuring 11 m in diameter. g. A tar kiln with a diameter of 15 m. h. A 0.5 m deep pit measuring 1 m in diameter.

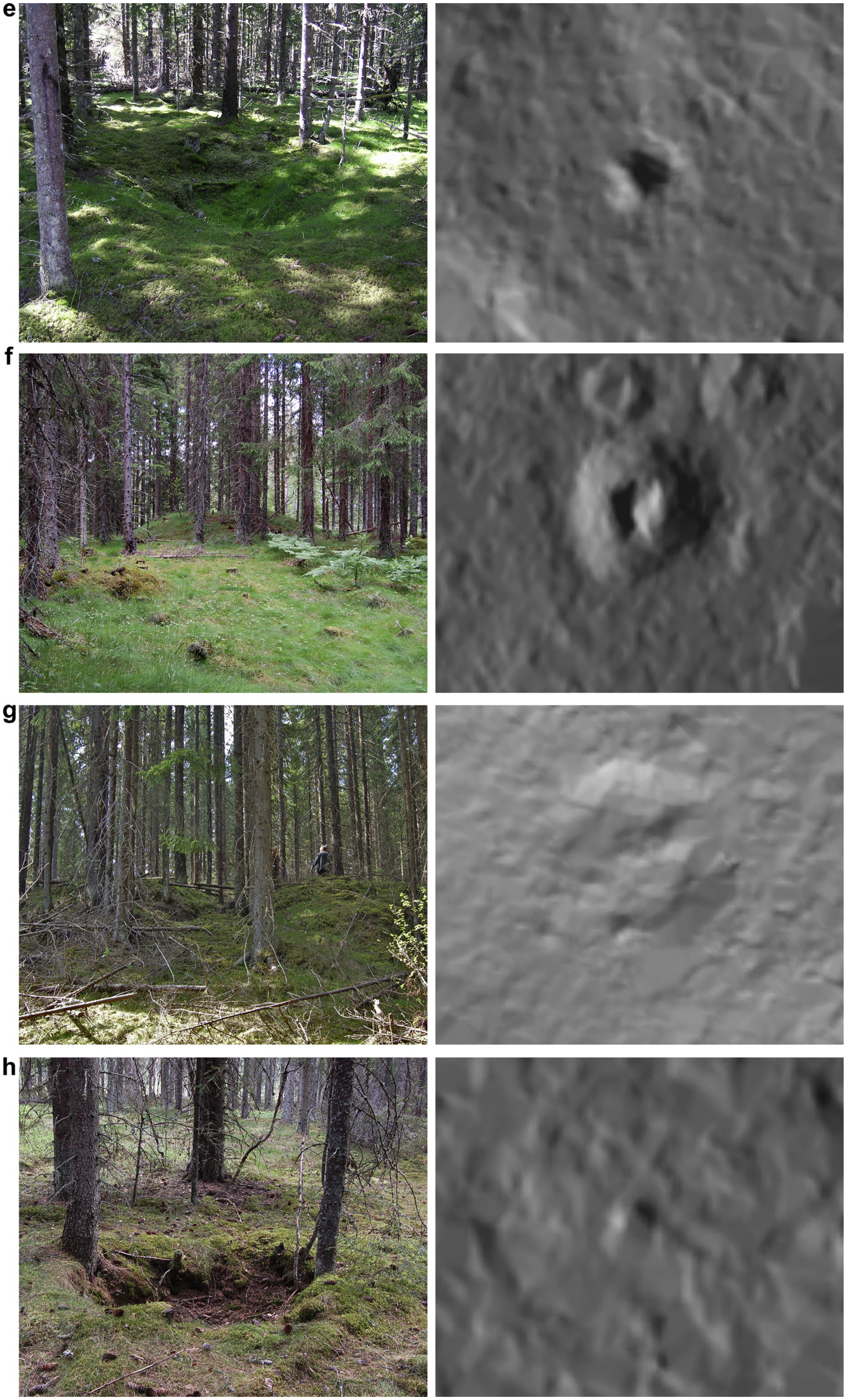

Fig. 2. (*continued*).

based iron production) and most likely the charcoal was primarily produced for forging in this area. Former studies have proven that the presence of charcoal pits in areas without traces of iron production can be explained by the production of charcoal for use in forges (Narmo, 1997). Thus, in this case it is uncertain from where the locals acquired the pig iron used in smithies. On the other hand pre-industrial iron production is registered not far from the study area. So far it is unclear what the charcoal produced in pits within the study area was actually used for.

Tar was also a highly valued commodity in Iron Age and medieval times, and the presence of a few tar kilns prove that tar was produced from resinous pinewood in this area. None of these kilns have been excavated and dated, but dating of similar tar kilns from other regions in south-eastern Norway indicate that they were in use in medieval times – approximately A.D. 1100–1500 (Rolfsen, 2002).

Due to deterioration the remains of pitfalls for elk today resemble charcoal pits, but are still somewhat wider and deeper. In the study area 25 pitfalls were found and mapped, some of them lying in a row as part of a catching system. Two pits situated in the study area have been dated indicating that in the early Iron Age from approximately 100 B.C. to A.D. 200, elk were caught by the use of such pits – or systems of pits in this area (Mølmen, 1989).

The presence of grave mounds prove that parts of what is today forested land used to be arable land with permanent settlement sometime during the Iron Age and perhaps later, but that settlement ceased at a certain point in history leaving the areas in question to reforestation. The co-localisation of farms, grave mounds and arable land is well known in Norwegian settlement archaeology (Jerpåsen, 1996; Stabbetorp et al., 2007).

Finally some hollow-roads were found in the area showing former infrastructure of unknown age but probably going back to the Iron Age. Hollow-roads are lanes appearing as terrain depressions and formed by a combination of traffic (mainly horses) and erosion (Hindle, 2008). Thus, the 334 remains and monuments originate from different kinds of human activity in this area throughout at least the last 2000 years and are important sources necessary in order to understand how humans utilized and marked the landscape.

3.4. Laser data

A Piper PA31 350 Navajo aircraft was used in the ALS data acquisition that was carried out in August and September 2007. The laser scanner, a Leica ALS50-II (which now must be considered as an older system), was set to operate with a pulse repetition frequency of 119 kHz and a half scan angle of 13°. The flying altitude was on average 790 m above ground level and the speed was 70 m/s. The target point density on the ground of the acquisition was 10 points m^{-2}.

3.5. DTM generation

The initial point density of the acquired ALS data was 10 points m^{-2}. The data were thinned to obtain two additional ALS datasets of 5 and 1 points m^{-2} using the method described by Magnussen et al. (2010) to enable analyses of the effect of point density on defection success rates. Creating a terrain dataset by thinning high density ALS data was chosen in order to isolate the effect of point density by eliminating all factors connected to the flying. If data with three different resolutions was collected from three different flights the resulting point clouds would be affected by changes in other sensor and acquisition settings such as flying altitude, pulse repetition frequency, and footprint size.

For each of the datasets with different point densities, three datasets with different levels of terrain smoothing were derived. Thus, DTMs were constructed from nine datasets covering the study area, all with different properties using the Terrascan software (Anon., 2010) that uses the progressive triangular irregular network densification algorithm of Axelsson (2000).

3.6. Experimental design and DTM interpretation

A randomized complete block design was used for the experiment. In the study nine experimental units (nine 500 × 500 m blocks) were used, from which nine different DTMs were prepared. As mentioned earlier, these DTMs had different combinations of point density (1, 5 and 10 points m^{-2}) and three levels of smoothing. With nine different DTMs for each of the nine experimental units, the total number of what were denoted as DTM-tiles added up to 81. Each of these DTM-tiles were interpreted by four archaeologists experienced in interpreting DTMs using the software QTM, who marked every cultural remain they could find digitally on screen. Ideally more than four interpreters would be desirable in a statistically based study like this in order to strengthen the assertions but due to economic and practical limitations the group was confined to four. The institutional backgrounds of the four archaeologists were quite similar and they all had general knowledge about the character of archaeology in forested areas. Examples of how the different categories of cultural remains typically appear in field and digitally on screen, is shown on Fig. 2a–h. The interpretations were carried out by use of QTM which is one of a range of available software tools developed with the purpose of handling, analysing and enhancing three-dimensional data, including ALS data. This allowed the interpreters to surf as well as shift light angle and a 360° change of light direction in real-time in manual analysis of the models. They also had the opportunity to exaggerate the elevation (the z-value) and make digital height profiles (cross-sections).

The interpretations were carried out in an individually randomized order in terms of the factors in the study (point density and smoothing). The randomization accounts for the potential effect that the interpretations became more accurate at the end of the experiment because the interpreters improved their skills. In addition to making a mark for each possible detected cultural remain, the interpreters also classified the type (charcoal kiln, grave mound etc.) of each detected object. The period of time the interpreters had at their disposal to go through all the tiles was rather limited compared to an ideal situation with as much time as needed to go through the same tiles more than once. The interpreters had three working days to interpret the 81 tiles; the time was limited due to economic and practical circumstances. The restricted use of time, when compared to an ideal situation with "unlimited" use of time, might have affected the achieved results in such a way that the detection and classifying figures should be understood as relative and not absolute figures. For instance one cannot disregard the potential danger that the smaller features were not given the same attention as the larger ones. It is important that these circumstances are noted if the study is compared to other similar studies.

3.7. Assessing differences in detection success rates between different types of cultural remains

The differences in detection success rates between different types of cultural remains were assessed using an error matrix (Table 5). In the error matrix the number of cultural remains of each specific type was compared to the number of correct interpretations of the respective types. In such a matrix, the number

Table 5

Error matrix. Shaded area: Number of detected and correctly classified cultural remains distributed on types (bold diagonal elements). Off-diagonal elements within the shaded area are number of detected, but misclassified cultural remains. White area: The last three columns (white area) display commission error (false positives), sum of all detections and user's accuracy (UA). The two columns under the heading Interpreted show the sum of all interpreted cultural heritage as well as user's accuracy when disregarding the commission errors (UACD). The two rows under the heading "Detected" show the sum of all detected cultural heritage and the producer's accuracy disregarding the omission error (PAD). The three rows under the heading "All" display the omission errors (false negatives), the sum of field observed remains and the producer's accuracy (PA). The four bottom lines show the PA distributed on the different interpreters.

		Type of field observed cultural remain									Interpreted		All		
		Pit-fall	Grave mound	Various pits	Hollow road	Charcoal pit	Charcoal kiln	Tar kiln	House foundation	Other	Sum	UACD (%)	Commission	Sum	UA (%)
Type of interpreted cultural remain	Pit-fall	**46**	0	0	0	17	2	0	0	0	66	71	13	78	59
	Grave mound	0	**261**	0	0	0	17	12	0	0	290	90	28	318	82
	Various pits	0	0	**0**	0	5	1	2	0	0	8	0	7	15	0
	Hollow road	0	0	0	**42**	0	0	0	0	0	42	100	51	93	45
	Charcoal pit	296	0	2	0	**267**	92	2	2	0	661	40	310	971	27
	Charcoal kiln	13	12	2	0	15	**4684**	7	0	0	4733	99	96	4829	97
	Tar kiln	3	20	0	0	0	38	**17**	0	0	78	22	5	83	20
	House ground	1	0	1	0	0	2	2	**0**	0	6	0	5	11	0
	Other	2	31	0	0	0	22	14	1	**0**	70	-	79	149	-
Detected	Sum	361	324	6	42	304	4858	56	3	0	5954	-	-	-	-
	PAD (%)	13	89	0	100	88	97	40	0	-	-	-	-	-	-
All	Omission	539	180	354	354	1424	2522	448	249	0	-	-	-	-	-
	Sum	900	504	360	396	1728	7380	504	252	0	-	-	-	-	-
	PA (%)	5	52	0	13	15	63	3	0	-	-	-	-	-	-
	PA1 (%)	0	71	0	9	22	70	0	0	-	-	-	122	-	-
	PA2 (%)	20	84	0	20	16	70	11	0	-	-	-	157	-	-
	PA3 (%)	0	19	0	0	16	62	0	0	-	-	-	132	-	-

of detected and correctly classified cultural remains will appear on the diagonal. The off-diagonal elements of the matrix are either detected cultural remains that were wrongly classified, cultural remains that were not detected (omission errors), or false detections (commission errors). The matrix also displays the producer's accuracy (PA) which is the number of correctly classified cultural remains in the percentage of the number of field-observed cultural remains of the specific type, i.e. the joint probability that a cultural remain of a certain type is detected and classified correctly. In the table, we also display the PAs for each of the interpreters (PA1 – PA4). To show also the probability that a cultural remain is correctly classified conditional on its detection - the matrix also displays the producer's accuracy of detected (PAD). Thus, as opposed to the PA, the PAD disregards the omission errors. In addition the so-called user's accuracy (UA) was calculated which is the number of correctly classified cultural remains in the percentage of the total interpretation of that specific type (includes commission errors), i.e., the probability that an anomaly found in a DTM and interpreted as a cultural remain is classified correctly. Similarly as for PA, the user's accuracy for correctly detected remains (UACD) was also calculated. The UACD disregards the commission errors.

3.8. Assessing differences in detection success rates according to different size, shape and elevation of cultural remains

In order to assess if there were differences in detection success rates between cultural remains of different categories of shape, size, and elevation, each cultural remain was assigned to each of these categories in field or based on field measurements. The shape categories were line, circle, oval, square and irregular (Table 1). Since none of the remains were rectangular this category was not included. The term irregular is used for those features with no clear geometrical shape. Five different categories of size based on measurements of the largest horizontal extent were used. Size classes 1–5 were 0 to $<$4 m, 4 to $<$8 m, 8 to $<$12 m, 12 to $<$16 m and $\geq$16 m, respectively (Table 2). The last category was elevation, based on measurements of elevation difference from the top to the bottom of the cultural feature. Elevation categories 1, 2 and 3 were 0 to $<$0.5 m, 0.5 to $<$1.0 m and $\geq$1.0 m, respectively (Tables 2 and 3).

To analyse if shape, size, and elevation had an effect on detection success rates, a Generalized Linear Model (GLM) was fitted (Nelder and Wedderburn, 1972) using the PROC GENMOD of the statistical software SAS (SAS Institute Inc., 2011). The GLM was chosen over an ordinary Analysis of Variance (ANOVA) because the data were quantitatively unbalanced due to the grouping into categories of shape, size and elevation. GLM is also very flexible in the sense that the response can have any distribution within the exponential family, in this case the binomial distribution (detected or not detected). A GLM used on this kind of data uses a link function to link the response (0 or 1) to the explanatory variables, and in this case the *logit* link function was used. The model was fitted to a dataset where each of the 334 cultural remains was registered detected (1) or not detected (0) for each of the nine ALS-data setups (three point densities and three smoothing levels) and for each of the interpreters. Thus, the dataset had 12,024 records. The explanatory variables were interpreter, point density, smoothing, block, shape, size, and elevation. Smoothing was not a factor of particular interest in this study, but it was still accounted for in the modelling because it was a part of the sampling design. The model was

$$\mathrm{TP} = \left(1 + e^{(\beta_0 + \beta_1 x_1 + \ldots + \beta_{25} x_{25})}\right)^{-1}$$

where TP is the response (true positive) which takes the value 1 if detection was successful and 0 if not. The Xs are the explanatory variables and the βs are parameters to be estimated. Since all variables are class-variables, they are represented by dummy variables. There is always one dummy variable less than the number of classes, for each explanatory variable. The parameters of the model were estimated using the method of maximum likelihood.

The estimated model was used to predict the mean probability of detecting cultural remains within each of the different categories of size and shape. These predictions are approximations of the mean values within each category as if the dataset was balanced, and they are referred to as least square means or LS-means (Goodnight and Harvey, 1978). The dataset was unbalanced in the sense that the number of cultural remains between categories, and levels within categories varies significantly (see Tables 1–4). For example, most of the circular cultural remains were large (Tables 3 and 4). An ordinary arithmetic mean TP for the circular cultural remains would therefore be biased if there was a positive relationship between TP and size. The predicted LS-means were obtained with each of the explanatory variables included in the model so that it would be possible to isolate the effect of each explanatory variable. Also, the difference between the LS-means of each level within each category was calculated. To test if the differences were significantly different from zero, t-tests were applied. Since multiple tests performed simultaneously increase the risk of falsely rejecting the null-hypothesis (Type I error), the t-tests were performed according to the Bonferroni approach, i.e., the significance level in each test was set to α divided by the number of tests.

4. Results

4.1. Assessing differences in detection success rates between different types of cultural remains

The error matrix (Table 5) shows results for all interpretations irrespective of smoothing, point density and all other explanatory variables. The most important information that can be derived from the matrix is therefore the relative "ranking" between different types of cultural remains. If Table 5 is considered first and with a focus on the observed cultural remains (vertically in the matrix), two types stand out with regard to PA, namely charcoal kilns (PA = 63) and grave mounds (PA = 52) which means that these types were those which had the highest joint probability to be detected and classified correctly. The high PAD values indicate that they were quite seldom misinterpreted. However, 34% (2522 out of 7380) and 36% (180 out of 504) of the observed charcoal kilns and grave mounds, respectively, were omitted. The matrix also shows that hollow roads were never misinterpreted (PAD = 100), but the omission rate was high. As many as 296 (33%) of the pitfalls were misinterpreted to be charcoal pits. Detected charcoal pits seem to be easier to classify correctly (PAD = 88) compared to pitfalls (PAD = 13). None of the house foundations or various pits were correctly classified and almost all were omitted.

The interpretations (horizontally in the matrix) were also studied. Similar as for PA, the user accuracy (UA) of charcoal kilns (UA = 97) and grave mounds (UA = 82) were higher than for the other types. This means that once the archaeologists interpreted a possible cultural remain to be either a charcoal kiln or a grave mound, these were most likely to be correctly classified. Also, pitfall classifications were quite often correct (UA = 58). Thus, while pitfalls were often misclassified as charcoal pits, those interpretations of pitfalls actually made were quite certain. Table 5 also shows that in 310 instances, natural features were erroneously classified as charcoal pits (commission error). The second most frequent type of misclassification of a natural feature was charcoal kilns (96 instances).

The PA for the interpreters also varies as shown in Table 5. There were no differences concerning house foundations and various pits and only small differences between the interpreters when it came to identifying charcoal kilns and charcoal pits. A little more variation was found as far as grave mounds and hollow roads were concerned and especially regarding pit-falls and tar kilns. The commission errors were relatively stable between the interpreters.

As aforementioned, the error matrix shows results irrespective of explanatory variable. However, it is also interesting to know how point density affects the PA, omission and commission errors. Thus, in Fig. 3 the PAs are shown, and the omission and the commission errors distributed on the three point densities used in the study. The commission errors for each type are shown relative to the total number of commission errors which were 45, 217 and 324 for the point densities 1, 5 and 10 points m^2, respectively. The main result was that increased point density improved detection success rates for all types, but the relative improvement for both PA and omission error varied between the different types. The greatest relative effect of increasing point density from 1 to 10 points m^{-2} with regard to PA was found for charcoal pits. Correspondingly, the greatest relative reduction of omission error was found for charcoal kilns which also had the greatest improvement in both PA and omission error in absolute terms. Thus, charcoal pits and kilns were the types that profited most from increasing point density from 5 to 10 points m^{-2}, both in terms of PA and omission error. The actual number of commission errors increased with increasing point density. Furthermore, the relative distribution of commission errors also changed as the point density was changed.

4.2. Assessing differences in detection success rates according to different size and shape of cultural remains

All effects of the fitted GLM model were statistically significant ($p < 0.0001$), except for smoothing ($p = 0.067$) (Table 6). Table 7 shows the least square mean detection probability of the different cultural remains after dividing them into five different categories according to their shape (see Table 1 for details). The table also shows the differences between the different LS-means of each category in the matrix to the right. All differences were significantly different ($p < 0.05$) unless marked with "ns". The results indicate that oval structures were most likely to be detected and classified correctly (LS-mean = 0.92). They were followed by circular and square structures which have a LS-mean of 0.32 and 0.30, respectively. Linear and irregular structures seemed to be most difficult to detect, with LS-means of 0.02 and 0.01, respectively. The difference in detection rate between lines and irregular shapes was not significant, neither was the difference in detection rate between circles and squares.

Table 8 shows LS-means and the difference in LS-means between the cultural remains after dividing them into five different size classes (see Table 2 or 4 for more information about the number of cultural heritage structures in each size class). The largest structures were most likely to be detected, with a LS-mean of 0.90. The LS-mean for cultural heritage of the second largest remains (12 to <16 m) was 0.41, while the LS-mean for structures measuring between 8 and <12 m was 0.09. These differences were all statistically significant. The two smallest size categories 4 to <8 m and 0 to <4 m had LS-means of 0.02 and 0.03, respectively. There was no significant difference in detection probability between these two categories. The overall tendency was that the probability for detection increased by size.

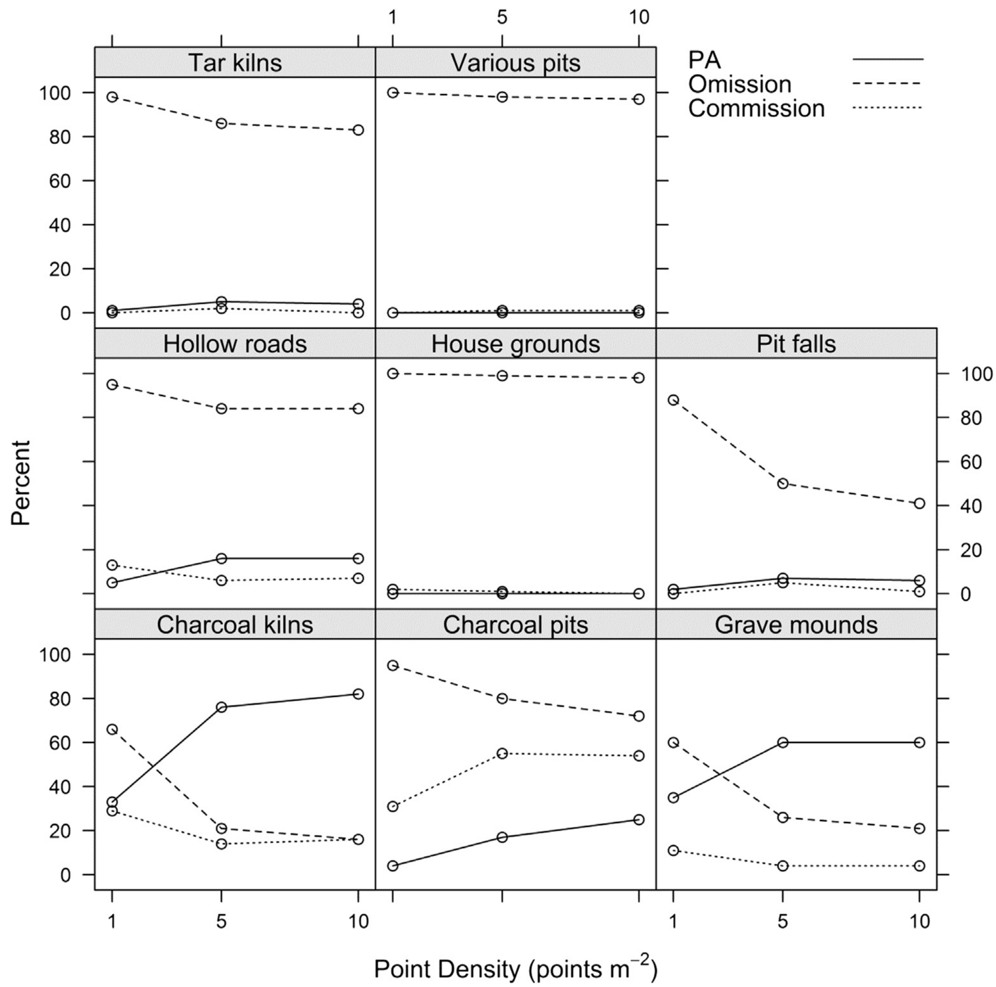

Fig. 3. Producer's accuracy (PA), omission- and commission errors for different point densities and different types of cultural remains. The commission errors are relative to the total number of commissions for each point density.

Table 9 shows the LS-means of the cultural remains after they were divided into three different elevation classes, and the difference between the LS-means of each class. Details on elevation classes in relation to size and shape can be seen in Table 2 or 3, respectively. Table 9 shows that the LS-mean increased with increasing elevation. For structures with elevation exceeding 1 m, the LS-mean was 0.22, for those between 0.5 m and 1.0 m, it was 0.11 whereas for those <0.5 m it was 0.06.

5. Discussion

Generally the results show, as one may expect, that large, regular and elevated cultural remains (charcoal kilns and grave mounds) are easier to detect and classify correctly than those which are small, low and irregular (charcoal pits, tar kilns, hollow roads, and especially various pits and house foundations). This is further underlined by the PA figures for each interpreter. However, PA for pitfalls turned out to be low compared to results from previous studies carried out in Norway (e.g. Risbøl, 2010); in addition this was a category which was only identified by one of the interpreters (see Table 5). Much of the reason for the low PA in the current study was that the pitfalls were often confused with charcoal pits which are quite similar in size and shape. They both consist of a pit surrounded by a low dyke, but usually pitfalls are somewhat deeper and larger compared to charcoal pits. Another complicating factor is that it was not uncommon to re-use pitfalls as charcoal pits (Amundsen, 2007). The opposite situation might also have

Table 6
Analysis of effects (Type 3) for the fitted GLM. The most important figures are marked with bold.

Effect	DF	Wald chi-square	pr > chi-square
Interpreter	3	362.8	<0.0001
Point density	2	1810.7	<0.0001
Smoothing	2	5.4	**0.067**
Block	8	151.3	<0.0001
Shape	4	545.4	<0.0001
Size	4	1708.8	<0.0001
Elevation	2	220.0	<0.0001

Table 7
Least square means (LS-means) and difference in least square means between cultural remains of different shapes. The most important figures are marked with bold.

		LS-mean	Shape				
			Line	Circle	Oval	Square	Irregular
Shape	Line	0.02		−0.30	−0.90	−0.28	0.01^{ns}
	Circle	**0.32**	0.30		−0.60	0.02^{ns}	0.31
	Oval	**0.92**	0.90	0.60		0.62	0.91
	Square	**0.30**	0.28	−0.02^{ns}	−0.62		0.29
	Irregular	0.01	−0.01^{ns}	−0.31	−0.91	−0.29	

^{ns}Not significant.

Table 8
Least square means (LS-mean) and difference in least square means between cultural remains of different size classes. The most important figures are marked with bold.

		LS-mean	Size class				
			0 to <4 m	4 to <8 m	8 to <12 m	12 to <16 m	≥16 m
Size class	0 to <4 m	0.03		0.01[ns]	−0.06	−0.38	−0.87
	4 to <8 m	0.02	−0.01[ns]		−0.06	−0.39	−0.87
	8 to <12 m	0.09	0.06	0.06		−0.33	−0.81
	12 to <16 m	**0.41**	0.38	0.39	0.33		−0.48
	≥16 m	**0.90**	0.87	0.87	0.81	0.48	

[ns]Not significant.

occurred. Thus it is quite difficult to distinguish between the two categories when interpreting DTMs and often this can only be solved by verification in field. One way of distinguishing pitfalls from charcoal pits without going out into the field, is to study the spatial context in which they appear, i.e., the mutual relations between the pits in terms of location. In most cases pitfalls are found in rows like pearls on a string, but with some distance between the single pits. Usually a precondition for studying such relationships is the ability to work with large landscape areas, since catching systems can consist of hundreds of pitfalls and stretch for several kilometres. The design of the present study, using spatially disjointed blocks measuring 500 × 500 m in size, might not be ideal in the sense that it could possibly have contributed to disguising the coherence between these pits as parts of a larger system. Such coherences are important premises needed in order to understand the purpose of the existence of such pits. In addition, the interpreters in this study were neither given information about the location from where the DTM tiles originated nor any background information about the historic use of the area. This was done in order to eliminate a potential effect of differences in local knowledge about the archaeology in the study area amongst the participants in the interpreter group. The lack of such *a priori* knowledge might also have affected the results. However, the UA for pitfalls was rather high (only exceeded by charcoal kilns and grave mounds) and indicates that once an object was identified and thought to be a pitfall, the right classification was applied quite successfully.

The results concerning hollow roads also deserve some attention. The study shows that such roads were quite difficult to detect, but when first detected they were correctly classified in all cases. There were no cases in which hollow roads were mixed up with other types of cultural remains but in some cases natural formations were interpreted as hollow roads, thus resulting in commission errors. This was in strong contrast to both charcoal pits and pitfalls which were mixed up with other cultural remains as well as natural formations.

Furthermore, the results show that tar kilns were often misinterpreted to be grave mounds or charcoal kilns. Occasionally this is not far from the truth owing to the fact that in some cases grave mounds were re-used as a foundation for the construction of tar kilns. However, in the study area only one of the 14 tar kilns was constructed on a grave mound.

Table 9
Least square means (LS-mean) and difference in least square means between cultural remains of different elevation classes. The most important figures are marked with bold.

		LS-mean	Elevation class		
			0.0 to <0.5 m	0.5 to <1.0 m	≥1.0 m
Elevation class	0.0 to <0.5 m	**0.06**		−0.05	−0.16
	0.5 to <1.0 m	**0.11**	0.05		−0.11
	≥1.0 m	**0.22**	0.16	0.11	

The effect of increased point density on PA and omission errors varies between the different types of cultural remains (Fig. 3). The general trend is that detection success rates improve substantially when point density is increased from 1 to 5 points m^{-2}, but the improvement is less pronounced or absent when the point density is increased to 10 points m^{-2} as also pointed out by Bollandsås et al. (2012). With regard to PA the charcoal kilns, charcoal pits, and grave mounds are the only types of cultural remains where increasing point density from 5 to 10 points m^{-2} to some extent improves the detection success rates. The omission error is also reduced for these types as a result of this increase in resolution. Further, pitfalls and to a lesser degree hollow roads, are also omitted less frequently. The effect of increased resolution is almost absent concerning the remains that seem most difficult to detect in the first place, especially house foundations and the group with various pits. Why the effects of increased point density more or less failed to occur for these types of remains can partly be explained by the circumstance that the target objects were either very low and diffuse and/or hidden beneath very dense undershrub and thus dampened the effect of increased point density (see also Crow et al., 2007). If dense ground vegetation explains the absence of the effects of increased point density it would be of no help to increase beyond 10 points m^{-2} owing to the fact that the laser pulses will be returned from the impenetrable vegetation nonetheless. On the other hand, if the explanation can be connected to the low elevation of the small remains, one cannot disregard the possibility that these would benefit from a denser set of ground points and consequently be detectable.

In many cases it is a challenging task to separate cultural remains from natural topographic elements, as highlighted by earlier studies (Risbøl et al., 2006; Bofinger and Hesse, 2011). The results show that increased point density not only results in increased detection success rates but also in an increased number of commission errors (see also Bollandsås et al., 2012). This can be explained by the fact that increased resolution will also make natural features that resemble cultural remains stand out more clearly. The greatest increase in commission errors when increasing point density from 1 to 5 points m^{-2} appears for the pits in general and for charcoal pits in particular. This can probably be attributed to the circumstances, in that natural pits resulting from wind-thrown trees or erosion are common in the study area. Also, glacial and periglacial processes have created geomorphological features such as drumlins and kettle holes which in some cases can be mistaken for man-made structures. Finally, the area is marked by modern forestry activities which also create various depressions in the terrain.

The oval shaped remains were most successfully detected, followed by the circular and square ones as judged by the LS-mean estimates (Table 7). The oval remains are mainly pitfalls. This may seem a contradictory result compared to that presented in Table 5 where pitfalls have a very low PA. However, the LS-mean is not affected by misinterpretation (detected, but classified wrongly) as is the case for PA. Thus, the high LS-mean value for ovals reflect that they were easily recognized as cultural remains, but not necessarily easy to classify correctly. One reason why oval remains are much easier to detect than other shape groups can be related to the fact that very few natural features appear oval. However, such a large difference between the oval and circular remains, as was the case in this study, requires attention. It is hard to give a full explanation for this finding, but as indicated above it might be that natural circular features are more common than natural oval features. The main message here, however, is that remains with a clear geometrical shape are more likely to be detected compared to irregular ones because they stand out compared to natural features, as also reported by Risbøl (2010). In the present study this applies almost solely to tar kilns which have an irregular shape and consequently a very low detection success rate. In contrast to most of the other remain categories, tar kilns consist of differently shaped and more randomly distributed elements such as pits, dykes and ditches that often constitute a totality with a rather irregular appearance.

That linearly shaped remains are at almost the same low level as the irregular ones when it comes to the detection success rates, also requires attention. This could have been explained by a function of lighting if fixed hill-shaded images were used but this was not the case in this study and the result must be explained otherwise. From earlier studies it is reported that linear features are amongst the easier ones to detect (Sittler and Schellberg, 2006; Schmidt et al, 2007; Ladefoged et al., 2011). According to a study conducted by Crow et al. (2007), linear structures are easier to detect than circular ones. In that particular study this could be related to size more than shape, since linear structures indicated rather substantial earthworks and the circular ones were charcoal platforms with only a little elevation. Still, when it is reported that cultural remains are often easier to observe on an ALS-generated DTM compared to standing on the actual spot in field (Schmidt et al, 2007; Millard et al., 2009; Bofinger and Hesse, 2011; Chase et al., 2011; Georges-Leroy, 2011), this statement is often applied to linear features (mainly paths, roads, lynchets plus ridges and furrows). The archaeological community uses the concept *linear feature* quite broadly but mainly it is used within a coarse-meshed topological understanding as opposed to points and polygons. The definition of topologies is not straightforward and different categories of the physical properties of cultural remains can belong to more than one topological group because they are compounded by different elements. To go further into this matter is a research topic that also involves, for example, perception studies, which are outside the scope of the current study.

In any case, this study does not support the findings reported by others that linear features are amongst the easier to detect. Alternatively this might be explained by the fact that the research design used in the present study fails to "isolate" the effect of linear shapes since the observed hollow roads are all shallow (<0.8 m) and with an average width of only 2.2 m. Further, they are also the only linear structures in the dataset. Operator biases might be another explanation, i.e., the potential risk that the persons who did the interpretations in this study developed an unconscious emphasis on visually detecting the dominant type of remains during the study of the DTM. The empirical data are to a large extent dominated by circular shaped remains (79%), while only 3% are linear. Gallagher and Josephs (2008) also experienced difficulties in detecting some linear features in their study and explained this by the presence of dense ground vegetation cover in combination with imperfectly adapted data acquisition and processing parameters. Based on the study at hand there is no reason to state that hollow-roads are more overgrown than other features. Most likely the fact that the hollow roads are quite shallow combined with the effect of operator biases could explain the rather poor detection success rate for this category.

Summarizing size, there is one clear conclusion: size matters – a conclusion supported by other studies (e.g. Gallagher and Josephs, 2008; Shaw and Corns, 2011). There is a very strong relationship between the size of a cultural remain and the detection success rates. The analysis estimates that as many as 90% of the cultural remains larger than 16 m in diameter can be expected to be successfully detected. This is followed by a middle-bracket with 41% (12–<16 m) and a heavy drop to less than 9% for those measuring less than 12 m in size.

The general finding concerning elevation is that the effect of increased elevation difference is evident in terms of increased probability for detecting and correctly classifying the cultural remains. Compared to structures with an elevation difference of 1 m or more, those of 0.5–<1 m are half as likely to be seen, while those with the smallest elevation difference are approximately half as likely to be detected as those measuring between 0.5 and 1 m. All these results are statistically significant. Other studies concur, for example the finding that cultural remains with a certain depth, such as various pits, are easier to detect than shallow ones (Gallagher and Josephs, 2008). A study of several deep pitfalls for reindeer in northern Norway (Risbøl, 2010) also supports this conclusion. The same study also points to the problem that in many cases none of the laser pulses will reach all the way to the very bottom of deep pits with quite narrow openings due to the angle of the beam which results in a more shallow appearance in the DTM than in reality. The point coverage of the more or less vertical slopes inside a pit will be dependent on where the pit is situated in relation to the airplane and the scan swath when the ALS is conducted. If the pit is situated right below the plane, the chances for the pulses to reach all the way to the very bottom is high compared to a position at the outskirt of the swath where the laser beam will hit the ground at an angle. ALS has been successfully used for the detection of cave openings and for documenting their morphology, including the depth of the caves (Weishampel et al., 2011). The average depth was 21 m, and with an average opening diameter of 15 m it was possible to document the depths of the caves using a 1 m resolution DEM.

The results in the current study regarding the effect of elevation shows that a higher elevation difference will often give a more pronounced outline of the cultural remains and their constituents – at least as long as the maximum height and maximum depth are not too far apart horizontally. The sharper the edges are, the more likely they are to be detected on the DTM, while even gradually descending slopes will often result in a more unclear shape of the cultural remains, making them less visible.

6. Conclusion

This study has shown that the detection success rates when using ALS-derived DTMs for detecting and classifying cultural remains, as measured by PA, PAD, UA, UACD, omission error and commission error, differ between different types of cultural remains.

For all cultural remains the positive effect of increasing the point density from 1 to 5 points m^{-2} was relatively large but almost absent when increasing from 5 to 10 points m^{-2}. The larger the remains are, the more they benefit from a denser set of ground points in terms of detecting and classifying them. However, increasing point density beyond 10 points m^{-2} could potentially contribute to better identification of the smaller features. As dense a set of ground points as possible is advantageous, but usually there are practical limitations, such as those connected to economy, restricting the range of action and a recommendation of approximately 5 points m^{-2} seems to be reasonable in areas with similar landscape and with similar remains as dealt with in this study.

When it comes to identifying which aspect of the remains affects the detection and classifying rate most, this study shows a very strong relationship between size and detection success rates. Larger structures are more easily detected than smaller ones. That remains with an irregular shape often tend to be confused with natural features was another find. The results concerning elevation show that cultural remains with a certain elevation will often have more pronounced edges and thus are easier to detect and classify correctly than shallow ones. In short, large features with a clear geometrical shape, characterized by a clear elevation difference are most easily detected and given the right classification.

ALS is increasingly being used by archaeologists as a method for detecting and interpreting cultural remains in forests, and will eventually contribute to an improved archaeological record in wooded areas. This will affect the cultural historic understanding of how humans relate to landscape in the past. In addition it has implications for heritage management policy which, to an increasing degree, will have to act in accordance with improved knowledge about larger parts of the landscape.

The study presented here indicates that a record based on the interpretation of ALS data will be somewhat biased due to the differences in detection success according to the size and shape of the cultural remains. By and large ALS is very usable for the purpose of identifying and listing cultural remains in forests, but the quality of the record improves if the ALS data interpretations are combined with ground truthing.

Acknowledgements

The authors would like to thank the Norwegian Directorate for Cultural Heritage for funding the main part of the project. We also greatly appreciate the financial contribution from The Norwegian Forest and Landscape Institute. Archaeologist Kristin Os from NIKU deserves special thanks for her participation in the field work in Eidsvoll and Nannestad in the summer of 2010. In addition we want to give our thanks to Dr. Liviu T. Ene from the Norwegian University of Life Sciences for his contribution with processing the data used in the study. Finally, we are indeed grateful to the four archaeologists who did the interpretations: Kjetil Skare, Hedmark County Municipality; Christer Tonning, Vestfold County Municipality; Rut Langebrekke Nilsen and Hans Marius Johansen, Sør-Trøndelag County Municipality. Without their admirable efforts we would not have been able to conduct this study. Finally we want to thank the six referees whose comments have contributed to a substantial improvement of the paper.

References

Amundsen, T., 2007. Fangstgroper for elg (Pitfalls for elk). In: Amundsen, T. (Ed.), Elgfangst og bosetning i Gråfjellområdet (Elk catching and settlment in the Gråfjell area), Gråfjellprosjektet, bind II. Varia 64. Kulturhistorisk Museum, Universitetet i Oslo, pp. 63–136 (in Norwegian).

Anon., 2010. TerraScan User's Guide. Terrasolid Ltd., Jyvaskylä, Finland. http://www.terrasolid.fi/system/files/tscan_1.pdf (last date accessed 02.10.11.).

Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing 33 (B4), 111–118.

Bennett, R., Welham, K., Hill, R.A., Ford, A., 2012. A comparison of visualization techniques for models created from airborne laser scanned data. Archaeological Prospection 19 (1), 41–48.

Bofinger, J., Hesse, R., 2011. As far as the laser can reach…: laminar analysis of LiDAR detected structures as a powerful instrument for archaeological heritage management in Baden-Württemberg, Germany. In: Cowley, D.C. (Ed.), Remote Sensing for Archaeological Heritage Management. Europae Archaeologiae Consilium (EAC), Brussel, pp. 161–171.

Bollandsås, O.M., Risbøl, O., Ene, L.T., Nesbakken, A., Gobakken, T., Næsset, E., 2012. Using airborne small-footprint laser scanner data for detection of cultural remains in forests: an experimental study of the effects of pulse density and DTM smoothing. Journal of Archaeological Science 39, 2733–2743.

Challis, K., Forlin, P., Kincey, M., 2011. A generic toolkit for the visualization of archaeological features on airborne LiDAR elevation data. Archaeological Prospection 18 (4), 279–289.

Chase, A.F., Chase, D.Z., Weishampel, J.F., Drake, J.B., Shrestha, R.L., Slatton, K.C., Awe, J.J., Carter, W.E., 2011. Airborne LiDAR, archaeology and the ancient Maya landscape at Caracol, Belize. Journal of Archaeological Science 38, 387–398.

Crow, P., Benham, S., Devereux, B.J., Amable, G.S., 2007. Woodland vegetation and its implications for archaeological survey using LiDAR. Forestry 80 (3), 241–252.

Devereux, B.J., Amable, G.S., Crow, P., Cliff, A.D., 2005. The potential of airborne lidar for detection of archaeological features under woodland canopies. Antiquity 79, 648–660.

Devereux, B.J., Amable, G.S., Crow, P., 2008. Visualisation of LiDAR terrain models for archaeological feature detection. Antiquity 82, 470–479.

Doneus, M., Briese, C., 2006. Full-waveform airborne laser scanning as a tool for archaeological reconnaissance. In: Campana, S., Forte, M. (Eds.), From Space to Place: 2nd International Conference on Remote Sensing in Archaeology (Proceedings of the 2nd International Workshop, CNR, Rome, Italy, December 2–4, 2006), BAR International Series, vol. 1568, pp. 99–107.

Doneus, M., Briese, C., 2011. Airborne laser scanning in forested areas – potential and limitations of an archaeological prospection technique. In: Cowley, D.C. (Ed.), Remote Sensing for Archaeological Heritage Management. Europae Archaeologiae Consilium (EAC), Brussel, pp. 59–76.

Gallagher, J.M., Josephs, R.L., 2008. Using LiDAR to detect cultural resources in a forested environment: an example from Isle Royale National Park, Michigan, USA. Archaeological Prospection 15, 187–206.

Georges-Leroy, M., 2011. Airborne laser scanning for the management of archaeological sites in Lorraine (France). In: Cowley, D.C. (Ed.), Remote Sensing for Archaeological Heritage Management, Europae Archaeologiae Consilium (EAC). EAC Occasional Paper No. 5, Brussel, pp. 229–234.

Goodnight, J.H., Harvey, W.R., 1978. Least Square Means in the Fixed-effect General Linear Model. Technical Report R-103. SAS Institute, Cary NC, pp. 1–12.

Gustafsson, L., Dahlberg, A., Green, M., Henningsson, S., Hägerhäll, C., Lars-son, A., Lindelöw, Å., Lindhagen, A., Lundh, G., Ode, Å., Strengbom, J., Ranius, T., Sandström, J., Svensson, R., Widenfalk, O., 2009. Konsekvenser för kulturarv, friluftsliv, landskapsbild och biologisk mångfald. Faktaunderlag till MINT-utredningen (Consequences for cultural heritage, outdoor life, landscape and biodiversity. Basic facts to the MINT–review), pp. 1–208. SLU, Rapport. (in Swedish).

Helliksen, W., 1997. Gård og utmark på Romerike 1100 f.Kr. – 1400 e. Kr. Gardermoprosjektet (Farm and outfield at Romerike 1100 B.C. – 1400 A.D. The Gardermoen project). Varia 45. Universitetets Oldsaksamling, Oslo, pp. 1–175 (in Norwegian).

Hesse, R., 2010. LiDAR-derived Local Relief Models – a new tool for archaeological prospection. Archaeological Prospection 17 (2), 67–72.

Hesse, R., 2013. The changing picture of archaeological landscapes: lidar prospection over very large areas as part of a cultural heritage strategy. In: Opitz, R.S., Cowley, D.C. (Eds.), Interpreting Archaeological Topography. 3D Data, Visualisation and Observation, Occasional Publication of the Aerial Archaeological Research Group, No. 5. Oxbow Books, Oxford, pp. 171–183.

Hindle, P., 2008. Medieval Roads and Tracks. Shire Publications, Buckinghamshire, pp. 1–64.

Holmsen, A., 1961a. Eidsvoll bygds historie. Bygdehistorien til omkring 1700 (The History of the Rural District Eidsvoll. The History of a Parish Until About 1700), bind I, 1. del. Eidsvoll Bygdebokkomité, Oslo, pp. 1–388 (in Norwegian).

Holmsen, A., 1961b. Eidsvoll bygds historie. Bygdehistorien 1700–1914 (The History of the Rural District Eidsvoll. The History of a Parish 1700–1914) bind I, 2. del. Eidsvoll Bygdebokkomité, Oslo, pp. 1–719 (in Norwegian).

Jacobsen, H., Follum, J.-R., 1997. Kulturminner og skogbruk (Cultural Monuments and Forestry). Skogbrukets Kursinstitutt, Biri, pp. 1–248 (in Norwegian).

Jerpåsen, G., 1996. Gunnerød – an analysis of an archeological landscape. In: Ihse, M. (Ed.), Landscape Analysis in the Nordic Countries. Integrated Research in a Holistic Perspective, Proceedings from the Second Seminar of Nordic Landscape Research, Lund 13–14 May 1994. Forskningsrådsnämnden, Sockholm, pp. 149–159.

Kokalj, Ž., Zakšek, K., Oštir, K., 2011. Application of the sky-view factor for the visualization of historic landscape features in lidar derived relief models. Antiquity 85 (327), 263–273.

Ladefoged, T.N., McCoy, M.D., Asner, G.P., Kirch, P.V., Puleston, C.O., Chadwick, O.A., Vitousek, P.M., 2011. Agricultural potential and actualized development in Hawai'i: an airborne LiDAR survey of the leeward Kohala field system (Hawai'I Island). Journal of Archaeological Science 38, 3605–3619.

Larsen, J.M., 1996. Det var kølatrafikken som heldt liv ti dom (The charcoal transportation kept them alive). Norsk Skogbruksmuseums årbok nr. 14, 1993–1996, Elverum, pp. 54–126 (in Norwegian).

Larsen, J.H., 2009. Jernvinneundersøkelser (Investigation of Iron Production). In: Faglig program, bind 2. Varia 78. Kulturhistorisk Museum, Universitetet i Oslo, pp. 1–250 (in Norwegian).

Magnussen, S., Næsset, E., Gobakken, T., 2010. Reliability of LiDAR derived predictors of forest inventory attributes: a case study with Norway spruce. Remote Sensing of Environment 114, 700–712.

Millard, K., Bruke, C., Stiff, D., Redden, A., 2009. Detection of a low-relief 18th-century British siege trench using LiDAR vegetation penetration capabilities at Fort Beauséjour-Fort Cumberland national historic site, Canada. Geoarchaeology: An International Journal 24 (5), 576–587.

Mølmen, Ø., 1989. Den gamle elgfangsten i Mathiesen-Eidsvold Værks skoger (The Old Elk Catching in the Forests of Mathiesen-Eidsvold Værk), pp. 1–118, Eidsvold Værk. (in Norwegian).

Narmo, L.E., 1997. Jernvinne, smie og kullproduksjon i Østerdalen. Arkeologiske undersøkelser på Rødsmoen i Åmot 1994–1996 (Iron Production, Smithy and Charcoal Production in Østerdalen. Archaeological Investigation at Rødsmoen in Åmot 1995–1996). Varia 43. Universitetets Oldsaksamling, Oslo, pp. 1–201 (in Norwegian).
Nelder, J., Wedderburn, R., 1972. Generalized linear models. Journal of Royal Statistical Society, Series A 135 (3), 370–384.
Risbøl, O., 2005. Kulturminner i Gråfjell – kulturhistoriske resultater fra registreringsprosjektet (Cultural remains in Gråfjell – cultural historic results from the survey project). In: Stene, K., Amundsen, T., Risbøl, O., Skare, K. (Eds.), "Utmarkens grøde" – mellom registrering og utgraving i Gråfjellområdet, Østerdalen (Cultural remains in Gråfjell – cultural-historical results from the survey project). Varia 59. Kulturhistorisk Museum, Universitetet i Oslo, pp. 5–26 (in Norwegian).
Risbøl, O., 2010. Towards an improved archaeological record through the use of airborne laser scanning. In: Forte, M., Campana, S., Liuzza, C. (Eds.), Space, Time and Place, 3rd International Conference on Remote Sensing in Archaeology British Archaeological Review, International Series 2118, pp. 105–112.
Risbøl, O., 2013. Cultivating the "Wilderness" – how Lidar can improve archaeological landscape understanding. In: Opitz, R.S., Cowley, D.C. (Eds.), Interpreting Archaeological Topography, 3D Data, Visualisation and Observation Occasional Publication of the Aerial Archaeological Research Group, No. 5. Oxbow Books, Oxford, pp. 53–64.
Risbøl, O., Gjertsen, A.K., Skare, K., 2006. Airborne laser scanning of cultural remains in forests: some preliminary results from a Norwegian project. In: Campana, S., Forte, M. (Eds.), From Space to Place: 2nd International Conference on Remote Sensing in Archaeology (Proceedings of the 2nd International Workshop, CNR, Rome, Italy, December 2–4, 2006), BAR International Series, vol. 1568, pp. 107–112.
Rolfsen, P., 2002. Tjæremiler i Norge – med utgangspunkt i en tjæregrop på Hovden i Bykle (Tar kilns in Norway – with a starting point with a tar pit at Hovden in Bykle). In: Høigård Hofseth, E. (Ed.), UKM – En mangfoldig forskningsinstitusjon. Universitetets kulturhistoriske museer, Skrifter Nr. 1, Oslo, pp. 255–265 (In Norwegian).
Rutar, G., Črešnar, M., 2011. Reserved optimism: preventive archaeology and management of cultural heritage in Slovenia. In: Cowley, D.C. (Ed.), Remote Sensing for Archaeological Heritage Management, EAC Occasional Paper No. 5. Europae Archaeologiae Consilium (EAC), Brussel, pp. 259–264.
SAS Institute Inc., 2011. SAS OnlineDoc® 9.3. SAS Institute Inc., Cary, NC.
Schmidt, S., Bofinger, J., Keller, R., Kurz, S., 2007. LIDAR – high resolution raster data as a survey tool. In: Figueiredo, A., Velho, G. (Eds.), The World is in Your Eyes. CAA 2005. Computer Applications and Quantitative Methods in Archaeology, Proceedings of the 33rd Conference, Tomar, March 2005. Tomar, pp. 255–260.
Shaw, R., Corns, A., 2011. High Resolution LiDAR specifically for archaeology: are we fully exploiting this valuable resource? In: Cowley, D.C. (Ed.), Remote Sensing for Archaeological Heritage Management, EAC Occasional Paper No. 5. Europae Archaeologiae Consilium (EAC), Brussel, pp. 77–86.
Sittler, B., Schellberg, S., 2006. The potential of LIDAR in assessing elements of cultural heritage hidden under forest canopies or overgrown by vegetation: possibilities and limits in detecting microrelief structures for archaeological surveys. In: Campana, S., Forte, M. (Eds.), From Space to Place: 2nd International Conference on Remote Sensing in Archaeology (Proceedings of the 2nd International Workshop, CNR, Rome, Italy, December 2–4, 2006). Archaeopress, Oxford, pp. 117–122.
Skogsstyrelsen, 2003. Kulturmiljövård i skogen. Att känna och bevara våra kulturminnen (Cultural heritage in the forest. To recognize and preserve cultural monuments). Riksantikvarieämbetet, Stockholm & Jönköping, pp. 1–259 (in Swedish).
Stabbetorp, O.E., Bøe Sollund, M.-L., Brendalsmo, J., Norderhaug, A., 2007. Layers of the past: a theory and method for historical landscape analysis. Landscape Research 32 (4), 463–479.
Štular, B., Kokalj, Ž., Krištof, O., Nuninger, L., 2012. Visualization of lidar-derived relief models for detection of archaeological features. Journal of Archaeological Science 39, 3354–3360.
Weishampel, J.F., Hightower, J.N., Cjase, A.F., Chase, D.Z., Patrick, R.A., 2011. Detection and morphologic analysis of potential below-canopy cave openings in the karst landscape around the Maya polity of Caracol using airborne LiDAR. Journal of Cave and Karst Studies 73 (3), 187–196.