



Point of view: error estimation in field assignment of land-cover types

Eva L. Eriksen*, Heidrun A. Ullerud, Rune Halvorsen, Sigrun Aune, Harald Bratli, Peter Horvath, Inger K. Volden, Anders K. Wollan & Anders Bryn

Abstract

Questions: Substantial variation between observers has been found when comparing parallel land-cover maps, but how can we know which map is better? What magnitude of error and inter-observer variation is expected when assigning land-cover types and is this affected by the hierarchical level of the type system, observer characteristics, and ecosystem properties? **Study area:** Hvaler, south-east Norway. **Methods:** Eleven observers assigned mapping units to 120 stratified random points. At each observation point, the observers first assigned a mapping unit to the point independently. The group then decided on a 'true' reference mapping unit for that point. The reference was used to estimate total error. 'Ecological distance' to the reference was calculated to grade the errors. **Results:** Individual observers frequently assigned different mapping units to the same point. Deviating assignments were often ecologically close to the reference. Total error, as percentage of assignments that deviated from the reference, was 35.0% and 16.4% for low and high hierarchical levels of the land-cover-type system, respectively. The corresponding figures for inter-observer variation were 42.8% and 19.4%, respectively. Observer bias was found. Particularly high error rates were found for land-cover types characterised by human disturbance. **Conclusions:** Access to a 'true' mapping unit for each observation point enabled estimation of error in addition to the inter-observer variation typically estimated by the standard pairwise comparisons method for maps and observers. Three major sources of error in the assignment of land-cover types were observed: dependence on system complexity represented by the hierarchical level of the land-cover-type system, dependence on the experience and personal characteristics of the observers, and dependence on properties of the mapped ecosystem. The results support the necessity of focusing on quality in land-cover mapping, among commissioners, practitioners and other end users.

Keywords: Between-observer variation; classification; consistency; environmental gradient; error; land-cover mapping; observer effect; point sampling; reliability; repeatability; vegetation ecology; vegetation mapping.

Taxonomic reference: Lid & Lid (2005) for vascular plants.

Syntaxonomic reference: Halvorsen et al. (2015) for land-cover types.

Abbreviations: ED = Ecological distance; GLM = Generalised linear model; LCE = Local complex environmental variable; NiN = Nature in Norway; TPI = Topographic position index.

Submitted: 19 June 2018; first decision: 10 August 2018; accepted: 1 October 2018

Co-ordinating Editor: Erwin Bergmeier

Introduction

Society has a strong need for up-to-date, systematised knowledge about the geographical distribution of ecosystems and natural resources in rapidly changing landscapes (Fuchs et al. 2015; Mahmood et al. 2015). Field-

based land-cover maps are an important source of such knowledge, and substantial resources are allocated to land-cover mapping worldwide. Governmental agencies, planners, researchers, conservationists and private developers need land-cover maps for a broad spectrum of purposes, including resource and area management, to ob-

*Corresponding author's address: Natural History Museum, P.O. Box 1172, Blindern 0318 Oslo, Norway, e.l.eriksen@nhm.uio.no. Complete addresses of all authors can be found at the bottom of the paper.

tain statistics, to assess conservation value, and for monitoring changes over time (Alexander & Millington 2000; Pedrotti 2013; MNHN et al. 2014).

Land-cover mapping can be a complex task in a world where gradual natural variation is the norm and abrupt transitions are anthropogenic exceptions (Goodall 1963; Austin & Smith 1989; Collins et al. 1993; Foody 2002; Morgan et al. 2010). The common goal in land-cover mapping is to delineate the continuous variation in an area into polygons that each represents a generalised land-cover type. Systems of land-cover types can be built upon a wide spectrum of theoretical principles, tailored for different regions or user needs (e.g. NVC, Rodwell 1991–2000; Natura 2000, Sundseth 2008; EUNIS, EEA 2014). A consistent set of mapping units is required for land-cover mapping programmes. Nevertheless, observers may perceive the same natural variation differently and, accordingly, produce different land-cover maps for the same area (Cherrill & McClean 1999b). As a result, all land-cover maps are to varying degrees burdened with shortcomings in terms of errors and inter-observer variability. To reduce this, it may be useful to quantify and to systematise errors and inter-observer variability and analyse how these are linked to specific land-cover mapping programmes (Ullerud et al. 2018).

The terminology used to describe error and inter-observer variation in land-cover and vegetation maps varies among studies (e.g., Greco et al. 1994; Cherrill & McClean 1999a, 1999b; Hearn et al. 2011), and differs from standard statistical terminology. In this paper, the term ‘error’ is reserved for deviation from a ‘true’ reference (Foody 2002). Furthermore, we define ‘total error’ in the context of observers’ assignment of land-cover types, as a collective term for three components:

- (i) Context-dependent error, or irreducible error (Hastie et al. 2009), is the unavoidable variation in type assignment when more than one land-cover type has to be accepted as ‘true’. An important source of context-dependent error is shortcomings in definitions and descriptions of land-cover types. Lack of observable characteristics in the field that make confident assignment to land-cover type difficult can also contribute to this error component.
- (ii) Bias is the systematic deviation of land-cover-type assignments from an accepted ‘true’ reference type (Foody 2002). Bias may result from systematic overlooking of characteristics of land-cover types and systematic misidentification, or misjudgement, of the information that defines and characterises them. Underestimation bias has been found in visual estimation of vegetation cover and status (Gallegos Torell & Glimskär 2009; Couvreur et al. 2015).
- (iii) Residual error, the variation not explained by any available variable, is used for unsystematic deviations from the reference. Residual error arises be-

cause observers are different and comprises all kinds of unsystematic errors committed by them.

Of these three error components, context-dependent and residual errors are inherent properties of all land-cover maps, which can never be completely removed. Bias, on the other hand, represents systematic errors that may potentially be reduced or corrected (Archaux 2009).

While estimates of error indicate how trustworthy land-cover maps are, variation among observers indicates their consistency (sometimes referred to as ‘repeatability’). Quantitative assessments of total error in assignment of land-cover types require that a true reference type, or a proxy for such a truth, can be established (Foody 2002). When no ‘true’ reference is available, quantitative assessments of variation among observers is often used to assess map consistency instead. ‘Inter-observer variation’ is typically measured as the magnitude of disagreement in pairwise comparisons between different maps or type assignments of the same area (e.g. Greco et al. 1994; Cherrill & McClean 1999a, 1999b; Hearn et al. 2011).

Pairwise comparisons of maps made by different observers have revealed considerable variation in the magnitude of inter-observer variation in land-cover maps (Greco et al. 1994; Cherrill & McClean 1995, 1999a; Hearn et al. 2011). Comparative studies of land-cover maps have demonstrated higher inter-observer variation primarily when land-cover types on lower hierarchical levels of the systems of land-cover types are used. Systems with many land-cover types seemingly have a larger potential for incorrect identification than systems with fewer units (Ullerud et al. 2018).

Previous studies of inter-observer variation have been unable to distinguish clearly between variation in land-cover-type assignment, and variation in delineation of polygons (Morgan et al. 2010), i.e. the process of drawing a multitude of non-overlapping unique land-cover polygons covering the entire extent of the intended map. Cherrill and McClean (1995, 1999a) indicated that very different placement of boundaries in compared maps are likely due to differences in land-cover-type assignment rather than differences in the observers’ perception of transitions between types. They applied buffers around polygon boundaries and compared only the ‘core areas’ of polygons to separate the effect of type assignment. Buffering does not account for all subjectivity in polygon delineation, but may remove small discrepancies in delineation (Hearn et al. 2011) from a ‘true’ polygon boundary. Neither does buffering account for the effect that the mean value of a polygon or an area unit varies when the boundaries change (Openshaw 1984), i.e. the observer who made the map may not agree that the land-cover-type of the ‘core area’ of a given polygon is representative of the original, un-buffered polygon. This may affect to what degree polygons of different maps are comparable. Use of buffers is also problematic when the numbers of

polygons differ greatly between the compared maps. New methods are therefore required to separate the challenges related to polygon delineation from those of land-cover-type assignment.

Inter-observer variation has been studied in the related context of botanical surveys, addressing precision in cover estimates and species identification (Hope-Simpson 1940; Klimeš et al. 2001; Scott & Hallam 2003; Archaux et al. 2007; Vittoz & Guisan 2007; Milberg et al. 2008). These studies address observers' ability to identify indicator species and estimate their abundance. Thus, they point to sources of error that may also appear in field-based land-cover mapping. Furthermore, comparative studies of botanical surveys indicate that inter-observer variation may depend on properties of the ecosystem, as higher rates of species misidentification have been recorded for ground vegetation than for tree species, and for grasslands than for forests (Archaux et al. 2006; Morrison 2016).

Errors in field-based and remotely sensed land-cover maps may lead to undesirable outcomes of management decisions based on such maps, including loss of biodiversity (Gorrod & Keith 2009), unsustainable forestry (Thompson et al. 2007), failure to achieve conservation goals (Tulloch et al. 2013), and increased cost in restoration projects (Gergel et al. 2007). Differences between repeated maps, aimed at detecting land-cover change, may not represent actual changes and must therefore be interpreted with great care (Kuroda et al. 2006). Knowledge about the magnitude and sources of error and inter-observer variation in maps is vital for efforts to improve map quality and decisions based upon these maps (Cherill 2016). Such knowledge is needed for each new system of land-cover types, because of differences in the spatial scales they address, the region they are intended for, and the method used to classify natural variation into land-cover types (Foody 2002). The only published attempt of quantitative assessment of error in field-based land-cover maps, disentangled from other sources of error, was made by Stevens et al. (2004). They used a single expert observer's opinion as 'truth'. There is an apparent lack of recognition that a 'true' reference is required, with which all single assignments can be compared. Furthermore, the effects on errors and inter-observer variation of hierarchical levels of land-cover types, observer characteristics, and properties of the mapped ecosystem, remain largely unknown.

The main aim of this study is to quantify total error in the assignment of land-cover types to points by comparing observations to a reference. Total error is further explored by measuring the 'Ecological Distance' (ED) between assigned and reference land-cover type. By designing the study around a set of spatially explicit observation points, our aim is to circumvent error and inter-observer variation arising from incongruent delineation of polygons. Unlike polygons or plots, a 'point' in this study has

no area, and the observers determine the land-cover type from the smallest possible spot around each point. Only one land-cover type can be assigned to a point. We use the new, official system for land-cover mapping in Norway, the NiN system, to address the following research questions:

- What magnitude of error and inter-observer variation can be expected in land-cover-type assignment using the Nature in Norway (NiN) system?
- Is error and inter-observer variation affected by the hierarchical level of the type system, observer characteristics, and ecosystem properties?

Study area

The study area was the Hvaler archipelago in Østfold County, situated near the mouth of the Oslo fiord in south-east Norway (Fig. 1) and belonging to the boreo-nemoral bioclimatic zone (Moen 1998). The climate is sub-oceanic with fairly warm summers (mean June–August temperature 15.6 °C, Aune 1993). Long periods of drought occur in occasional years. The annual mean temperature is 6.4 °C (Aune 1993) and the mean annual precipitation is 740 mm (1961–1990 averages, Førland 1993). The bedrock in the area consists mainly of granite (Berthelsen et al. 1996). Rounded hills (highest point 72 m a.s.l.), interrupted by rift valleys, dominate the terrain.

The natural vegetation of Hvaler spans the long-term successional gradient from bare rock to established forest, resulting from continuous land uplift since the last

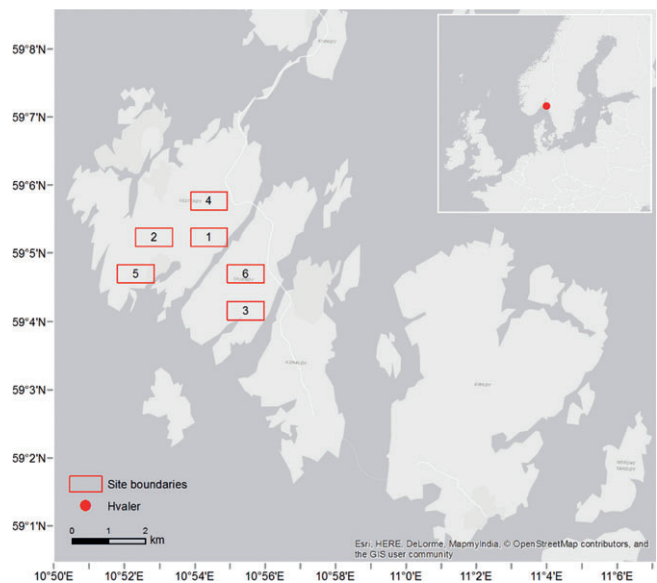


Fig. 1. Map of Hvaler archipelago. The six study sites are indicated by numbered rectangles. Inset: Northern Europe with the Hvaler archipelago location marked (WGS 1984, UTM zone 33).

glaciation 8000–9000 yrs B.P. The combination of weathering-resistant bedrock, exposure to strong winds and occasional sea-salt episodes retards succession to forest and explains the high cover of rock outcrops. Marine sediments enriched with remains of calcareous shell occur locally in depressions and give rise to soils rich in mineral nutrients. The forests are mostly coniferous, dominated by *Pinus sylvestris* and *Picea abies*. Some deciduous forest patches occur locally, with *Betula* spp., *Fraxinus excelsior*, *Populus tremula*, *Salix* spp. and *Sorbus aucuparia*. *Alnus glutinosa* has scattered occurrences in wet depressions.

The history of human settlement in the Hvaler archipelago dates back to the emergence of the islands from the sea. The prevailing occupation was small-scale agriculture, mostly combined with fisheries. During the 1850–1930 population peak, all suitable areas were used for grazing or haymaking. Even though agricultural use of the landscape is currently restricted to subsidiary farming and management for conservation, legacies of former use still characterise most of the Hvaler islands. Over the last 50 years, summer tourism has expanded substantially, and summer cabins appear scattered around the entire archipelago.

Methods

The NiN system

The research questions were addressed using the NiN system. NiN comprises two sets of land-cover types: a fundamental division of ecosystems into types based on ecological theory (Fig. 2), and an aggregation of the most detailed types into ‘mapping units’ for practical mapping at different predefined scales (Fig. 3; Halvorsen 2015a, 2015b; Halvorsen et al. 2015; Ullerud et al. 2018).

The NiN system is based on the gradient perspective on natural variation (Halvorsen 2012). Environmental factors that vary together, e.g. pH and the contents of calcium and mineral nutrients in the soil (Økland 1996), are combined to form ‘local complex environmental varia-

bles’ (LCEs) (Whittaker 1956). NiN compiles the most important LCEs that give rise to variation in species composition at fine scales, typically from one to a thousand meters. On the organisational level of ecosystems (Noss 1990), land-cover types on the two hierarchical levels ‘major types’ and ‘basic types’ are obtained. A major type comprises areas that can be characterised by the same set of LCEs and the same dominating structuring processes, such as environmental stress or human disturbance. Basic types are the most fundamental building blocks of NiN. They are defined, within each major type, by dividing each of the major type’s main LCEs into segments of standard size in terms of species compositional turnover. The number of segments is determined by the magnitude of compositional turnover along the major type’s LCEs. Basic types are thus combinations of segments along all the LCEs that are important for the major type in question (Halvorsen et al. 2015).

While most NiN land-cover types are defined by species composition, major types are also separated by other characteristics such as land use. Each major type belongs to one specific category of human disturbance intensity – natural, semi-natural, and strongly modified systems (Fig. 2). Natural systems present in the study area include forest, open fen and bare rock. Semi-natural grassland exemplifies a semi-natural major type. Among the strongly modified ecosystems, examples include ‘systems on hard substrates’, fields and permanent pastures.

For field-based mapping, the 443 terrestrial and wetland basic types are aggregated into broader mapping units. Criteria for aggregation and guidelines for NiN-based mapping are given in Bryn et al. (2018). The 277 mapping units adapted to mapping at scale 1:5000 (Bratli et al. 2016), nested within 58 major types, were used in the present study. Up to 199 mapping units within 43 major types were expected to potentially occur in the study area. A fact sheet is available for each mapping unit, with characteristics from LCEs, physiognomy, ecology, aerial photography, and a list of common and indicator species. The major type and mapping unit of each point observation were used to represent high and low hierarchical levels of the type system, respectively.

Example LCE	Agricultural management intensity										
	0	a	b	c	d	e	f	g	h	i	j
Segments of the LCE											
Category of human disturbance intensity	natural		semi-natural				strongly modified				
Example major type	T4 Forest		T32 Semi-natural meadow				T45 Ploughed meadow				

Fig. 2. Example of a local complex environmental variable (LCE) in the Nature in Norway system. LCEs are axes in a multidimensional ecological space, and can appear in multiple major types. The Agricultural management intensity LCE is segmented into eleven levels. The number of segments represents the amount of compositional turnover occurring due to different types of human agricultural disturbance over time.

	LCE segments	Basic types (1:500)	Mapping units (1:5000)
T12 Tidal meadow	c	A1 Tidal meadow in lower geolittoral zone	C1 Tidal meadow in lower and middle geolittoral zones
	d		
	e	A2 Tidal meadow in middle geolittoral zone	
	f		
	g	A3 Tidal meadow in upper geolittoral zone	C2 Tidal meadow in upper geolittoral and supralittoral zones
	h		
	i	A4 Tidal meadow in supralittoral zone	
	j		
	k		

Fig. 3. Example of a Nature in Norway (NiN) major type split into basic types which are aggregated to form mapping units. Basic types (A-series) can be mapped at 1:500 scale, while mapping units are aggregations that can be mapped at coarser scales (B-series 1:2500, C-series 1:5000, D-series 1:10 000, E-series 1:20 000). The main local complex environmental variable (LCE) in major type Tidal meadow (NiN code T12) is “Duration of exposure vs inundation” (NiN code TV). Note that the LCE segments start at “c” because the lower segments (“0”, “a” and “b”) are inundated > 50% of the time and thus lie outside of the Tidal meadow major type.

Study design

Six study sites, each $1 \times 0.5 \text{ km}^2$, were chosen to be representative for the natural variation in the study area. For each site, 20 observation points (Fig. 4) were distributed by a stratified random procedure based on a model of the Topographic Position Index (hereafter TPI, Jenness 2006). The TPI value of a pixel reflects the elevation of that pixel relative to the average elevation of neighbouring pixels. The TPI values were obtained for each $10 \times$

10 m^2 pixel by applying Topography Toolbox (Dilts 2015) in ArcMap (ESRI 2015, version 10.3.1) to a Digital Elevation Model (Kartverket.no, accessed 04.06.2015) covering the study area. Each pixel in the six sites was then classified into one of five site-specific TPI strata, representing local slope positions in each site. A balanced allocation of points to the different strata enabled over-sampling of depressions and under-sampling of ridges and hills relative to their areal cover. Thereby, the more diverse ecosystems of depressions were better repre-

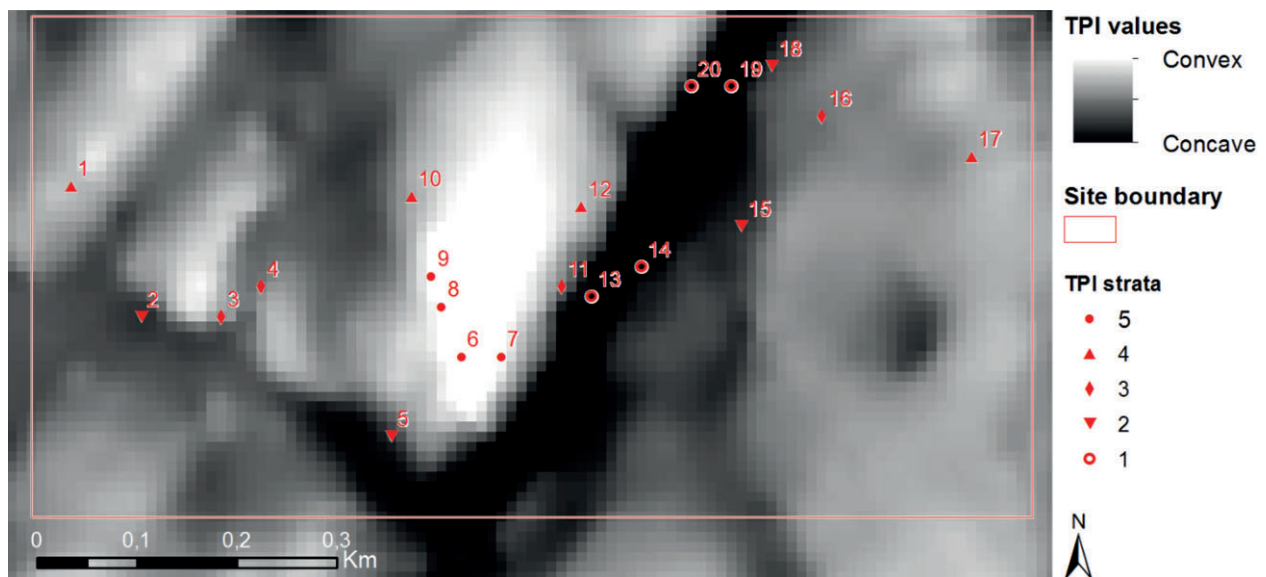


Fig. 4. Study site no. 4. Observation points (1–20) marked on top of the unstratified Topographic Position Index (TPI) layer. Darker colours represents lower TPI values (concave terrain features). Lighter colours represent higher TPI values (convex terrain features). Four random points ($10 \times 10 \text{ m}^2$ pixel centre coordinates) were chosen for each of five site-specific TPI strata, denoted by different point symbols.

sented in the sample while redundancy of bare rock sites, which cover large parts of the study area, was avoided.

ArcMap and Excel (Microsoft 2016) were used to place four points randomly in each stratum. The points were visually inspected on aerial photos (0.2 m resolution; norgebilder.no, accessed 25.05.2015) and relocated if they fell on water, private property, or in otherwise inaccessible areas. To avoid spatial autocorrelation, a minimum distance of 30 m was enforced between neighbouring points (mean Moran's I from TPI values = 0.374; Moran 1950). The points were located and marked by a red stick in the field by two of the authors before fieldwork took place.

Data collection

A group of 11 observers (referred to by capital letters A-K) participated in field work. The observers spanned variation in land-cover mapping experience, acquaintance with the NiN system and botanical field experience (Table 1). The observers' familiarity with NiN and the study area was assigned subjectively into three categories (low, medium, high), representing their level before field work.

Prior to fieldwork, observers A-J answered a 60-question online test of which 30 were related to assignment to NiN mapping units and 30 concerned the identification of indicator species relevant for land-cover-type assignment (supplement S4). Observer K constructed, and did not take, the test. For each observer, mapping-unit score

(0–30), species score (0–30) and total score (0–60) were recorded as the number of correct answers.

Fieldwork was conducted over five days in June 2016. The first two days were spent on training and calibration within the study area, but away from the observation points. Each observer was equipped with field instructions, descriptions of mapping units (Bratli et al. 2016) and diagrams showing the position of mapping units along LCEs within each major type (Bryn & Ullerud 2017). Fieldwork was carried out by all observers as a group. The group consequently visited each of the 20 points at each of the six study sites (Figs 1 and 4). At each point, observers individually and secretly considered the smallest possible spot around the point marker to determine the mapping unit, and independently wrote down their result. Only one mapping unit per point and observer was allowed. When all observers had recorded their assignment, the observers presented their decision to the group one by one. Thereafter, a group discussion took place to establish a reference mapping unit assignment to serve as proxy for the 'true' mapping unit at the point. In the few cases when no consensus was reached, even after lengthy discussions, a simple majority vote decided the reference mapping unit.

Data analyses

After fieldwork, the data were carefully proofread, and obvious mistyping errors corrected if possible. Binary variables, one for each observer, were constructed with one observation per point ($n = 120$), expressing agreement (1) or disagreement (0) between the mapping unit assigned by the observer, and the reference mapping unit. Additionally, 55 binary variables were constructed in the same way, expressing agreement (1) or disagreement (0) on each point ($n = 120$) between each possible pair of observers. These variables were used to calculate the fraction of assignments that corresponded with the reference (referred to as 'agreement with reference') and, for each pair of observers, the proportion of equal assignments (referred to as 'pairwise agreement'). The number of unique major types and mapping units in each set of 11 assignments was recorded for each of the 120 points.

All statistical analyses were done in R, version 3.4.2 (R Development Core Team 2017). Generalized linear models (GLMs; McCullagh & Nelder 1989) with log link function were used to find relationships between response and predictor variables and test them against null models of no relationship. A significance level of $\alpha = 0.05$, with Bonferroni correction (Legendre & Legendre 2012) in multiple testing situations, was used in forward model selection. Poisson-family GLM was used for count response variables, such as the number of unique mapping units per point. Binomial-family GLMs were used for binary variables like agreement with reference and pair-

Table 1. Observer (letters A-K) characteristics before field work: years of general botanical experience including mapping, age, and familiarity with the Nature in Norway (NiN) system and study area.

Observer	Years of experience	Age	Familiarity with NiN	Familiarity with study area
A	0	27	low	medium
B	3	28	medium	high
C	2	24	low	low
D	26	57	high	high
E	3	26	low	low
F	0	27	low	low
G	50	71	medium	high
H	10	53	medium	high
I	10	32	medium	medium
J	23	46	high	medium
K	41	59	high	high

wise agreement. For all GLM models, family = quasipoisson was used instead of family = Poisson (Ver Hoef & Boveng 2007) because of over-dispersion, defined as residuals deviating more than 20% from the residual degrees of freedom. The deviance D^2 was calculated for each model as a measure of the variation explained (Guisan & Zimmermann 2000). All mean values are arithmetic means.

Quantifying differences between mapping units

A method based on the NiN concept of ‘ecological distance’ (ED) was devised for quantifying deviation of individual mapping-unit assignments from the reference. The method used relationships between mapping units in the conceptual, multidimensional ecological ‘space’ of NiN’s LCE axes. The NiN mapping units by definition span the same standard interval of about one ED unit along each important LCE.

A deviation of one ED unit roughly corresponds to a 25% decrease in compositional similarity in noise-free data as assessed by the Bray-Curtis dissimilarity index (Halvorsen et al. 2015). For each observation point, the number of ED units between each observer’s mapping-unit assignment and the reference was obtained. ED was counted as the absolute value distance (Manhattan distance, or $L1$ norm, Goshtasby 2012) between mapping-unit midpoints along all LCEs used to define the units in question. The method is explained in full, with examples, in supplement S1.

Effects of observer and hierarchical level of the type system

Recording frequencies of major types and mapping units, per observer, were arranged in contingency tables. Observers’ binary agreement with reference, and ED to the reference, were used as indicators of total error. The chronological points sequence was used to test whether ED decreased during field work, indicating a learning effect. Pairwise binary agreement, and the number of unique major types and mapping units recorded at each single point, were used to quantify inter-observer variation. Performance in the knowledge test, observer identity, and years of experience (ln-transformed) were used as predictors of agreement with reference and pairwise agreement. In order to assess the importance of the hierarchical level of the type system, separate tests were performed for major types and mapping units.

Ecosystem effect

Average ED to the reference, calculated separately for each major type, was used to assess differences in total error among ecosystems. The reference major type was used as a predictor of the number of unique mapping units recorded per point, across all observers. The reference major type was also used as predictor of the number of observers who agreed with the reference mapping unit per point. Kendall’s rank correlation coefficient τ (Kendall 1938) was used to test if mean ED was related to the number of assignments to each major type.

Results

Observer characteristics

Considerable variation in performance was found among the ten observers who took the knowledge test. Species, major type and mapping unit mean scores were 16.8 (56.0% correct answers, range 23–83%), 18.5 (61.7% correct answers, range 20–90%) and 12.3 (41.0% correct answers, range 3–63%), respectively (Table 2). The test scores for species (Pearson’s $\rho = 0.757$, $t_8 = 3.272$, $p = 0.0113$), but not for mapping units ($\rho = 0.587$, $t_8 = 2.053$, $p = 0.0742$), were significantly related to years of experience.

Error and inter-observer variation

During fieldwork, 1320 point observations were made by the 11 observers, and 120 reference observations were obtained by the consensus procedure. A total of 23 unique major types and 59 unique mapping units were assigned to at least one point by one observer in the field. In contrast, the set of reference observations comprised 13 unique major types and 22 unique mapping units (Table 2). The three most commonly recorded mapping units were heather forest, heather-bilberry forest and open lime-poor shallow-soil heather-dominated ground.

Observers’ agreement with the reference was 83.6% (range 78.3–87.5%) for major types and 65.0% (range 56.6–72.5%) for mapping units. Mean pairwise agreement was lower, 80.6% (range 71.7–86.7%) for major types and 57.2% (range 47.5–70.0%) for mapping units, respectively (Table 2). All 11 observers agreed with each other on major type in 55.0% and on mapping unit in 22.5% of the points.

The overall mean ED between assigned and reference mapping units was 0.59 (range 0.44–0.77; Table 2). The frequency distribution of the 1320 single ED values was strongly right-skewed, with a mean value of 0.59 and standard deviation of 0.20 (Table 2, Fig. 5). An ED of one was found for 22.5% of the point observations, whereas

Table 2. Summary statistics for observer-specific assignments of major types and mapping units to points. Observers are indicated by capital letters A–K. Knowledge-test scores are given as number of correct answers out of 60 or 30. Point assignment results are given as numbers of unique major types and mapping units recorded and the numbers of points for which the assignment coincided with the reference (out of 120 points). Mean pairwise agreement refers to the mean fraction of points assigned to the same major type or mapping unit in a comparison between one specific observer and all other observers. Mean ED refers to the mean ecological distance between assigned, and reference, mapping unit, as explained in methods and supplement S1.

Observer	Knowledge test scores				Number of unique units		Agreement with reference		Mean pairwise agreement		Mean ED to reference
	Total	Species	Major type	Map-ping unit	Major type	Map-ping unit	Major type	Map-ping unit	Major type	Map-ping unit	
A	24	13	18	11	12	24	103	73	0.82	0.55	0.63
B	34	18	22	15	9	23	102	77	0.82	0.58	0.53
C	22	14	13	8	12	24	94	70	0.79	0.54	0.76
D	44	25	27	19	10	23	101	82	0.81	0.58	0.53
E	15	7	12	8	13	28	96	79	0.79	0.59	0.63
F	9	8	6	1	12	26	97	68	0.80	0.53	0.77
G	39	24	20	15	11	25	100	73	0.79	0.56	0.64
H	34	21	20	13	14	28	98	82	0.80	0.59	0.51
I	32	17	21	15	11	24	105	86	0.83	0.61	0.55
J	39	21	26	18	13	24	104	87	0.80	0.61	0.44
K	–	–	–	–	11	23	104	81	0.83	0.57	0.48
Mean	29.20	16.80	18.50	12.30	11.64	24.73	100.36	78.00	0.81	0.57	0.59
SD	11.32	6.25	6.50	5.44	1.42	1.93	6.66	13.55	0.02	0.26	0.20
Reference	60	30	30	30	13	22	120	120	1	1	0

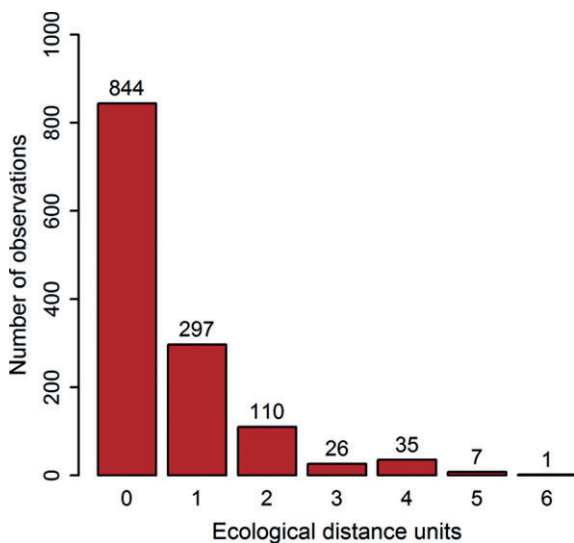


Fig. 5. Frequency distribution of ecological distance. Ecological distance is estimated between assigned and reference mapping units, based on all 1320 observations by 11 observers at 120 observation points. ED = 0 indicates that mapping unit is identical to the reference.

only 13.6% deviated more than this. There was a small, significant change in ED explained by point sequence (GLM: $F_{1,1318} = 6.45$, $p = 0.0112$, $D^2 = 0.00581$).

Hierarchical level of the type system

Agreement with reference was 28.6% higher, and pairwise agreement 40.9% higher, for major types compared to mapping units. The mean number of unique major types recorded per point was 1.7, compared to 2.9 for mapping units. The knowledge test scores were 50.4% higher for major types than for mapping units (Table 2).

Observer effect

Seven observers agreed on the relative recording frequency of the three most commonly registered mapping units, which also corresponded to the reference ranking (see Table S2.1 in supplement S2). The most commonly

found mapping units also had the largest discrepancies in recording frequencies. For instance, open lime-poor shallow-soil lichen-dominated ground was recorded twice by observer G, but 25 times by observer K. A total of 24 mapping units were only recorded one or two times. On two occasions, the reference mapping unit matched only one other observers' recorded mapping unit.

Mean ED to the reference was significantly related to observers' total score in the knowledge test (GLM: $F_{1,8} = 10.36$, $p = 0.012$, $D^2 = 0.568$), but the total score could not predict mean agreement with reference (GLM: $F_{1,8} = 4.65$, $p = 0.063$, $D^2 = 0.368$). A distinction can be made between the observers with the least experience (A, C, E and F) and the more experienced ones (B, D, G, H, J and K). The former had a mean ED of 0.70 and test score range of 9–24, while the more experienced observers had a mean ED of 0.53 and test score range 32–44.

Ecosystem effect

The reference major type was a significant predictor of the number of unique mapping units recorded per point (GLM: $F_{12,107} = 6.07$, $p < 0.001$, $D^2 = 0.419$), and of the number of observers agreeing with the reference mapping unit (GLM: $F_{12,107} = 5.28$, $p < 0.001$, $D^2 = 0.331$). Broken down to single major types (Table 3), the highest mean ED between assigned and reference mapping units was found for strongly modified and semi-natural major types. Corresponding results for mapping units are found in Supplement S2, Table S2.2. Mean ED was not significantly correlated with the number of assignments to each major type (Kendall's $\tau = -0.444$, $p = 0.119$).

Table 3. Mean ecological distance (ED) between assigned and reference mapping unit, given separately for each of the nine major types recorded 11 or more times (i.e., once or more per observer, on average). Rows are sorted by mean ED.

NiN code	Major type	n	Mean ED
T39	Strongly modified hard substrate	13	2.62
T32	Semi-natural grassland	44	1.63
T35	Strongly modified unconsolidated substrate	23	1.50
V1	Open fen	60	1.13
T4	Forest	719	0.52
T2	Open shallow-soil ground	265	0.48
T44	Field	26	0.43
V2	Wetland forest	118	0.37
T1	Bare rock	131	0.04

Discussion

Error and inter-observer variation

The results confirm that substantial error and inter-observer variation can be expected in mapping unit assignment (Cherrill & McClean 1999a, 1999b; Stevens et al. 2004; Hearn et al. 2011), also when using the NiN system. Comparing observer's assignments with a 'true' reference enables estimation of error in addition to inter-observer variation (Foody 2002). Our measure of ecological distance adds important detail that assists interpretation of the erroneous assignments by quantifying how ecologically similar the deviating assignments are. The estimated magnitude of total error was 35.0% when comparing observers' assignments with the reference. Among the mapping unit assignments that deviated from the reference, 62.4% were assigned to an ecologically neighbouring mapping unit (1 ED, Fig. 5). Some of these very similar assignments almost certainly lie within a broader range of acceptable assignments. The total error must therefore be lower than 35.0%.

The mean agreement with reference was found to be higher than the mean pairwise agreement between observers. This accords with the basic property of stochastic variables that error propagates when stochastic variables are combined by addition or subtraction (Sokal & Rohlf 1995). Thus, the standard method for assessment of map quality by use of pairwise comparisons (Cherrill & McClean 1995, 1999a; Hearn et al. 2011) systematically underestimates map 'quality' as defined in accordance with standard statistical procedures as deviation from an ideal, 'true' value.

Estimates of inter-observer variation obtained in the present study (42.8% for mapping units) lie within the 17–82% range (mean 44.5%) of corresponding values in previous studies (Greco et al. 1994; Cherrill & McClean 1995, 1999a; Stevens et al. 2004; Hearn et al. 2011; Ullerud et al. 2018, see comparison of studies in Supplement S2, Table S2.3). Because the present study only considers error related to type assignment, not delineation, the results suggest that assignment error may constitute a larger problem than differing delineation of polygons in land-cover maps. A possible caveat to this may lie in what properties observers consider when assigning mapping units to points, in terms of scope, vegetation patches, and topography. This could mean that delineation was not completely separated from type assignment using the point observation method. Further research on issues of delineation is needed to quantify and compare error in these two aspects of land-cover mapping.

Hierarchical level of the type system

In accordance with the expectation that it is easier to hit a wider target, lower total error and better pairwise agreement was observed for assignments to major types compared to mapping units. The inter-observer variation estimates of 19.4 and 42.8% for the major type and mapping unit levels, respectively, show the same general pattern as previous estimates. Hearn et al. (2011) found 22.4, 65.8 and 81.5% inter-observer variation at the habitat, community and sub-community levels, and Cherrill & McClean (1999a) found 44.5 and 74.4% for Phase 1 Habitat Classification (JNCC 2010) maps using broad land-cover categories and land-cover types.

Using the broader units of higher hierarchical levels for land-cover mapping is a double-edged sword due to the trade-off between precision and error. The centre points of adjacent land-cover types are situated further apart in ecological space the fewer land-cover types there are, because more variation is included in each land-cover type (Ullerud et al. 2018). This entails that errors among adjacent land-cover types made on higher hierarchical levels of type systems involve larger deviations from the truth than errors on lower levels. Accordingly, the choice of hierarchical level should be guided by both the need for map consistency and need for detailed information in the resulting maps.

Observer effect

Observers' scores on the knowledge test partly explain their resulting ED from the field assignments, meaning that observers with better species knowledge make fewer errors in the field. The same relationship was not clear when using the binary 'agreement with reference' variable, indicating that our measure of ecological distance added insight to the analysis of error. Furthermore, the fact that the 11 observers rank the mapping units differently by frequency suggests that bias contributes to total error. Some of the most common land-cover types have the largest discrepancies between observers. This is contrary to the expectation that the common land-cover types that observers have more experience with, should be less prone to error (Scott & Hallam 2003; Archaux et al. 2009; Hearn et al. 2011). Bias may explain this result. Specifically, observers may have differing perceptions on what is considered the most conservative choice in borderline cases. Some observers consistently chose more drought-prone land-cover types, while others were biased towards less drought-prone land-cover types. Further research is needed to identify the causes of bias, and to quantify the relative contributions of bias and residual error to the total error.

Several observer attributes may contribute to total error and inter-observer variation, such as motivation, per-

sonal biases, experience, expectations, spatial scale of observation, fatigue, and facilitation or disturbance from co-workers (Morrison 2016). The knowledge test results (Table 2) indicate that differences in skills among observers are important in this respect, and suggest that observer skills can be assessed from their ability to recognise land-cover types and species from pictures. This supports the results of Hearn et al. (2011), who found that inter-observer variation decreased with experience of the habitat type, and Scott & Hallam (2003) who emphasise the importance of knowledge of the local flora.

The average percentage of species overlooked by one observer but not another (pseudo-turnover) in botanical surveys is estimated to 10–30% while misidentification occurs in about 5–10% of observations (Morrison 2016). Rare plants, or those belonging to certain taxonomic groups, are more likely to be missed (Burg et al. 2015). Because many terrestrial land-cover types are identified by their plant species composition, overlooking or misidentifying indicator species is likely to be an important source of error and inter-observer variation in assignment of such land-cover types. In addition, several modifying factors influence species' responses to specific environmental factors, and responses often differ between regions (Brown 1984; Austin et al. 1994; Dahl 1998). Thus, not only may different observers detect different species, but they may also interpret their observations differently. Familiarity with the study area is therefore likely to be important for correct assignment of land-cover types.

Ecosystem effect

Differing mean ED between assigned, and reference, mapping units in different major types, indicates that some ecosystems are more difficult to assess than others are. Our results point to strongly modified major types, which are conditioned on intensive human disturbance, as particularly difficult to assign correctly to mapping unit (Stevens et al. 2004). Problems with identifying the structuring processes that define these major types, as well as imprecise definitions, lack of consistent indicator species, or other factors relating to context-dependent error, are likely to contribute to the total error in these major types. Some strongly modified major types can resemble semi-natural major types in species composition, requiring historical information or substantial experience to identify the differences.

Differences in mean ED to the reference, estimated separately for each major type, suggest that some of the ecosystems targeted by Norwegian authorities as of particular conservation value (Lindgaard & Henriksen 2011), are also among the most difficult to confidently assign to land-cover type. This is particularly true for semi-natural grasslands, for which the second highest mean ED to the refer-

ence was found (1.63 ED units, Table 3). A likely contributor to this result stems from the fact that most of the semi-natural grasslands in the study area were in different stages of regrowth successions at the time of fieldwork. The agricultural landscape is undergoing large changes (Rønningen 1993; Fjellstad & Dramstad 1999; Sang et al. 2014) by the combined processes of abandonment and intensification (Robinson & Sutherland 2002; Jepsen et al. 2015; Plieninger et al. 2016). Successional processes cause species typical of intact semi-natural grasslands to be gradually replaced by species characterising the post-successional ecosystem, typically forest.

The ecosystem effect presented in this study implies that our estimates of error and inter-observer variation may only be representative of ecosystems that are present in the study area. Further research is needed on error and inter-observer variation in other ecosystems, e.g. alpine ecosystems.

The point observation method and use of a 'true' reference

The point observation method used in this study opens for estimation of total error and inter-observer variation in land-cover-type assignments. Relative to comparisons of complete wall-to-wall maps, use of spatially precise point observations circumvents variation related to boundary placement and heterogeneity within polygons. The two processes of mapping unit assignment and delineating a multitude of polygons may not be separated completely, but experience with the point observation method in this study suggests that it adds more, and other, information than the buffering method used in prior studies (Cherrill & McClean 1999a, 1999b; Hearn et al. 2011).

An advantage of using a group of observers to establish a 'truth' is that total error in addition to inter-observer variation can be estimated by comparing the individual land-cover-type assignments to the consensus-made 'true' reference. Stevens et al. (2004) used a similar approach, but we argue that several observers offer a much more robust 'true' reference than a single expert opinion. Using multiple observers allows more options and indicator species to be discovered, and extreme estimates to be adjusted (Sykes et al. 1983; Klimeš et al. 2001; Klimeš 2003).

There is potential to use the point observation method and establishment of a 'true' reference in evaluation of other kinds of land-cover maps. Only 23% of selected terrestrial habitat mapping projects in Europe implemented quality control (MNHN et al. 2014). Our method could be applied as part of evaluation programmes, e.g. for training and calibration of observers.

Our operationalisation of the point observation method, with the observers visiting the observation

points as a group and establishing the reference by a consensus process, may have some impact on the results. The small but significant decrease in ED over the course of the field work indicates an effect of learning. While it is positive to see that observers become more similar with training, this effect may lead to a slight underestimation of error and inter-observer variation in this study. It should be noted, however, that there was less variation in the mapping units present in the last of the six study sites, which makes the significance of the learning effect questionable. Another property of the method is that all observers' observations were made in the same order and may add a component of temporal autocorrelation.

Implications for NiN and land-cover mapping

Land-cover mapping using NiN types faces many of the same challenges as mapping by use of other land-cover-type systems. This is evident from the substantial inter-observer variation observed in our study, which is of similar magnitude as previous studies. It is possible that the NiN system incurs specific challenges by being primarily based on ecological theory rather than created for practical mapping purposes. On the other hand, where the land-cover types are correctly assigned they may give more information to end users because elements of land use are integrated into the land-cover type. For instance, correct assignment of a semi-natural land-cover type indicates the level of agricultural management intensity that is needed to maintain its character and species composition.

Failure to take potential error in land-cover maps into account may have unfortunate consequences for practical management, such as giving priority to low-value sites while sites of higher conservation value are lost. In a survey among British environmental professionals (Cherrill 2016), about one third of the respondents reported experience with net loss of biodiversity due to errors in assignment of land-cover types. In many more cases, mistakes incurred additional costs for the respondents' firms or clients. Furthermore, errors that propagate into biased areal cover estimates will reduce the quality of for instance red list assessments of ecosystems (Keith et al. 2015).

Conclusions

An encouraging finding is that, if the observers in this study are representative of ecologists in general, most deviations of land-cover-type assignments from the accepted truth are small. Our results point to several measures that may be taken to reduce error and inter-observer variation in land-cover mapping: (i) Aggregating 'difficult' land-cover types into broader categories, since the level of

agreement in land-cover-type assignment increases when broader and more general land-cover types are used (Hearn et al. 2011; Ullerud et al. 2018). Aggregation should be done with care, since use of broader land-cover types implies reduction of ecological precision. (ii) Allowing observers to assess their confidence in their land-cover-type assignments, akin to adding “plausible bounds” to cover estimates as suggested by Gorrod & Keith (2009). This may, however, lead to problems of overconfidence (Wintle et al. 2013) and underreporting of problems the observers are not aware of. Neither does it address bias. (iii) Integrating vegetation survey methods such as recording of species in sample plots into the mapping procedures, in order to improve awareness of the full species composition and the species’ value as indicators.

The results further suggest that extensive training and calibration is necessary to improve quality of land-cover maps. Moreover, strong evaluation programmes are mandatory in order to provide error estimates, and even more importantly, to provide insights into the reasons for the errors. Use of a ‘true’ reference and ecological distance to that reference can contribute new information about errors and be valuable tools in evaluation programmes. When the errors are known, the causes can be specifically addressed. Improving quality of land-cover maps is necessary to meet the societal demands of knowledge about the geographical distribution of ecosystems.

Author contributions

H.A.U., R.H., A.B. and E.L.E. contributed to the study design. H.A.U. and E.L.E. planned the fieldwork. E.L.E. and R.H. handled the statistical analysis, and E.L.E. led the writing. All authors participated in fieldwork and critically revised the manuscript.

Acknowledgements

Fieldwork was carried out by Johnny Hofsten and Espen S. Værland in addition to the authors.

References

- Alexander, R. & Millington, A.C. 2000. *Vegetation mapping*. Wiley, Chichester, EN.
- Archaux, F. 2009. Could we obtain better estimates of plot species richness from multiple-observer plant censuses? *Journal of Vegetation Science* 20: 603–611.
- Archaux, F., Berges, L. & Chevalier, R. 2007. Are plant censuses carried out on small quadrats more reliable than on larger ones? *Plant Ecology* 188: 179–190.
- Archaux, F., Camaret, S., Dupouey, J.-L., Ulrich, E., Corcket, E., Bourjot, L., (...) & Touffe, J. 2009. Can we reliably estimate species richness with large plots? An assessment through calibration training. *Plant Ecology* 203: 303–315.
- Archaux, F., Gosselin, F., Berges, L. & Chevalier, R. 2006. Effects of sampling time, species richness and observer on the exhaustiveness of plant censuses. *Journal of Vegetation Science* 17: 299–306.
- Aune, B. 1993. Klima – *Temperaturnormaler normalperiode 1961–1990* [Normal temperature normal period 1961–1990]. Norwegian Meteorological Institute [Report no. 2-93], Oslo, NO.
- Austin, M.P., Nicholls, A.O., Doherty, M.D. & Meyers, J.A. 1994. Determining species response functions to an environmental gradient by means of a beta-function. *Journal of Vegetation Science* 5: 215–228.
- Austin, M.P. & Smith, T.M. 1989. A new model for the continuum concept. *Vegetatio* 83: 35–47.
- Berthelsen, A., Olerud, S. & Sigmond, E.M.O. 1996. *Geologisk kart over Norge, berggrunnskart OSLO 1:250 000* [Geological map of Norway, bedrock map OSLO 1:250,000]. Norges geologiske undersøkelse, Trondheim, NO.
- Bratli, H., Halvorsen, R., Bryn, A., Bendiksen, E., Jordal, J.B., Svalheim, E.J., (...) & Øien, D.-I. 2016. *Dokumentasjon av NiN tilrettelagt for praktisk naturkartlegging i målestokk 1:5000. Artikkelen 8 (versjon 2.0.2)* [Documentation of NiN adapted for mapping to scale 1:5000. Article 8 (version 2.0.2)]. Norwegian Biodiversity Information Centre, Trondheim, NO.
- Brown, J.H. 1984. On the relationship between abundance and distribution of species. *The American Naturalist* 124: 255–279.
- Bryn, A., Halvorsen, R. & Ullerud, H.A. 2018. *Hovedveileder for kartlegging av terrestrisk naturvariasjon etter NiN (2.2.0)* [Guidelines for mapping terrestrial ecosystems based on NiN (2.2.0)]. 1st edition. Natural History Museum, University of Oslo, Oslo, NO.
- Bryn, A. & Ullerud, H.A. 2017. *Veileder for arealdekkende kartlegging av terrestrisk naturvariasjon etter NiN i målestokk 1:5 000 og 1:20 000. Kortveileder (versjon 2.1.0b)* [Field guide for land-cover mapping of terrestrial ecosystems based on NiN to scales 1:5000 and 1:20,000 (version 2.1.0b)]. Natural History Museum, Oslo, NO.
- Burg, S., Rixen, C., Stöckli, V. & Wipf, S. 2015. Observation bias and its causes in botanical surveys on high-alpine summits. *Journal of Vegetation Science* 26: 191–200.
- Cherrill, A. 2016. Inter-observer variation in habitat survey data: investigating the consequences for professional practice. *Journal of Environmental Planning and Management* 59: 1813–1832.
- Cherrill, A. & McClean, C. 1995. An investigation of uncertainty in field habitat mapping and the implications for detecting land cover change. *Landscape Ecology* 10: 5–21.
- Cherrill, A. & McClean, C. 1999a. Between-observer variation in the application of a standard method of habitat mapping by environmental consultants in the UK. *Journal of Applied Ecology* 36: 989–1008.
- Cherrill, A. & McClean, C. 1999b. The reliability of ‘Phase 1’ habitat mapping in the UK: the extent and types of observer bias. *Landscape and Urban Planning* 45: 131–143.
- Collins, S.L., Glenn, S.M. & Roberts, D.W. 1993. The hierarchical continuum concept. *Journal of Vegetation Science* 4: 149–156.
- Couvreur, J.M., Fiévet, V., Smits, Q. & Dufrêne, M. 2015. Evaluation of the «observer effect» in botanical surveys of grasslands. *Biotechnologie Agronomie Societe Et Environnement* 19: 132–142.
- Dahl, E. 1998. *The phytogeography of northern Europe: British Isles, Fennoscandia and adjacent areas*. Cambridge University Press, Cambridge, EN.

- EEA 2014. *EUNIS habitat classification*. European Environment Agency, Copenhagen, DK.
- Fjellstad, W.J. & Dramstad, W.E. 1999. Patterns of change in two contrasting Norwegian agricultural landscapes. *Landscape and Urban Planning* 45: 177–191.
- Foody, G.M. 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment* 80: 185–201.
- Fuchs, R., Herold, M., Verburg, P.H., Clevers, J.G.P.W. & Eberle, J. 2015. Gross changes in reconstructions of historic land cover/use for Europe between 1900 and 2010. *Global Change Biology* 21: 299–313.
- Førland, E.J. 1993. Klima – Nedbørnormaler normalperiode 1961–1990 [Precipitation normals, Normal period 1961–1990]. Norwegian Meteorological Institute [Report no. 39/93], Oslo, NO.
- Gallegos Torell, Å. & Glimskär, A. 2009. Computer-aided calibration for visual estimation of vegetation cover. *Journal of Vegetation Science* 20: 973–983.
- Gergel, S.E., Stange, Y., Coops, N.C., Johansen, K. & Kirby, K.R. 2007. What is the value of a good map? An example using high spatial resolution imagery to aid riparian restoration. *Ecosystems* 10: 688–702.
- Goodall, D.W. 1963. The continuum and the individualistic association. *Vegetatio* 11: 297–316.
- Gorrod, E.J. & Keith, D.A. 2009. Observer variation in field assessments of vegetation condition: Implications for biodiversity conservation. *Ecological Management & Restoration* 10: 31–40.
- Goshtasby, A.A. 2012. *Image registration: principles, tools and methods*. Springer, New York, US.
- Greco, S., Petriccione, B. & Pignatti, F. 1994. Vegetation mapping: a numerical comparative study of six maps of Białowieża forest. *Phytocoenosis* 6: 105–113.
- Guisan, A. & Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147–186.
- Halvorsen, R. 2012. A gradient analytic perspective on distribution modelling. *Sommerfeltia* 35: 1–165.
- Halvorsen, R. 2015a. *Grunnlag for typeinndeling av natursystemnivået i NiN – Analyser av generaliserte artslistedatasett. Artikkel 2 (versjon 2.0.2) [Basis for typification at the ecosystem level in NiN – analyses of generalised species-list datasets. Article 2 (version 2.0.2)]*. Norwegian Biodiversity Information Centre, Trondheim, NO.
- Halvorsen, R. 2015b. *NiN – Typeinndeling og beskrivelsessystem for natursystemnivået. Artikkel 3 (versjon 2.0.2) [NiN – typification and attribute system for the ecosystem level. Article 3 (version 2.0.2)]*. Norwegian Biodiversity Information Centre, Trondheim, NO.
- Halvorsen, R., Bryn, A. & Erikstad, L. 2015. *NiNs systemkjerne – Teori, prinsipper og inndelingskriterier. Artikkel 1 (versjon 2.0.2) [NiN core – theory, principles and criteria for division into types. Article 1 (version 2.0.2)]*. Norwegian Biodiversity Information Centre, Trondheim, NO.
- Hastie, T., Tibshirani, R. & Friedman, J. 2009. *The elements of statistical learning*. 2nd edition. Springer, New York, US.
- Hearn, S.M., Healey, J.R., McDonald, M.A., Turner, A.J., Wong, J.L.G. & Stewart, G.B. 2011. The repeatability of vegetation classification and mapping. *Journal of Environmental Management* 92: 1174–1184.
- Hope-Simpson, J.F. 1940. On the Errors in the Ordinary use of Subjective Frequency Estimations in Grassland. *Journal of Ecology* 28: 193–209.
- Jepsen, M.R., Kuemmerle, T., Müller, D., Erbe, K., Verburg, P.H., Haberl, H., (...) & Reenberg, A. 2015. Transitions in European land-management regimes between 1800 and 2010. *Land Use Policy* 49: 53–64.
- JNCC 2010. *Handbook for Phase 1 habitat survey – a technique for environmental audit*. Joint Nature Conservation Committee, Peterborough, EN.
- Keith, D.A., Rodríguez, J.P., Brooks, T.M., Burgman, M.A., Barrow, E.G., Bland, L., (...) & Spalding, M.D. 2015. The IUCN Red List of Ecosystems: Motivations, Challenges, and Applications. *Conservation Letters* 8: 214–226.
- Kendall, M.G. 1938. A new measure of rank correlation. *Biometrika* 30: 81–93.
- Klimeš, L. 2003. Scale-Dependent Variation in Visual Estimates of Grassland Plant Cover. *Journal of Vegetation Science* 14: 815–821.
- Klimeš, L., Dančák, M., Hájek, M., Jongepierová, I. & Kučera, T. 2001. Scale-Dependent Biases in Species Counts in a Grassland. *Journal of Vegetation Science* 12: 699–704.
- Kuroda, A., Ikeda, S., Mukai, S. & Toyohara, G. 2006. Successive mapping of secondary pine forests affected by pine wilt disease and subsequent forest management in Miyajima Island, SW Japan. *Phytocoenologia* 36: 191–212.
- Legendre, P. & Legendre, L. 2012. *Numerical ecology*. 3rd edition. Elsevier, Amsterdam, NL.
- Lid, J. & Lid, D.T. 2005. *Norsk flora [Norwegian flora]*. 7th edition. Samlaget, Oslo, NO.
- Lindgaard, A. & Henriksen, S. (eds.) 2011. *The 2011 Norwegian Red List for Ecosystems and Habitat Types*. Norwegian Biodiversity Information Centre, Trondheim, NO.
- Mahmood, R., Pielke, R.A.S. & McAlpine, C.A. 2015. Climate-Relevant Land Use and Land Cover Change Policies. *Bulletin of the American Meteorological Society* 97: 195–202.
- McCullagh, P. & Nelder, J.A. 1989. *Generalized linear models*. 2nd edition. Chapman & Hall, London, EN.
- Milberg, P., Bergstedt, J., Fridman, J., Odell, G. & Westerberg, L. 2008. Observer bias and random variation in vegetation monitoring data. *Journal of Vegetation Science* 19: 633–644.
- MNHN, EEA, Ichter, J., Evans, D. & Richard, D. 2014. *Terrestrial habitat mapping in Europe: an overview*. European Environment Agency Museum national d'Histoire naturelle, FR [Report no. 1/2014], Luxembourg.
- Moen, A. 1998. *Nasjonalatlas for Norge: Vegetasjon [National atlas for Norway: vegetation]*. Statens kartverk, Hønefoss, NO.
- Moran, P.A.P. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37: 17–23.
- Morgan, J.L., Gergel, S.E. & Coops, N.C. 2010. Aerial photography: a rapidly evolving tool for ecological management. *BioScience* 60: 47–59.
- Morrison, L.W. 2016. Observer error in vegetation surveys: A review. *Journal of Plant Ecology* 9: 367–379.
- Noss, R.F. 1990. Indicators for Monitoring Biodiversity: A Hierarchical Approach. *Conservation Biology* 4: 355–364.
- Openshaw, S. 1984. Ecological fallacies and the analysis of areal census data. *Environment and Planning A* 16: 17–31.
- Pedrotti, F. 2013. *Plant and vegetation mapping*. Springer, Berlin Heidelberg, DE.
- Plieninger, T., Draux, H., Fagerholm, N., Bieling, C., Bürgi, M., Kizos, T., (...) & Verburg, P.H. 2016. The driving forces of landscape change in Europe: A systematic review of the evidence. *Land Use Policy* 57: 204–214.
- Robinson, R.A. & Sutherland, W.J. 2002. Post-war changes in arable farming and biodiversity in Great Britain. *Journal of Applied Ecology* 37: 157–176.
- Rodwell, J.S. 1991–2000. *British Plant Communities*. Cambridge University Press, Cambridge, EN.

- Rønningen, K. 1993. Agricultural policies and landscape management. Some examples from Norway, Great Britain and Germany. *Norsk Geografisk Tidsskrift – Norwegian Journal of Geography* 47: 93–104.
- Sang, N., Dramstad, W.E. & Bryn, A. 2014. Regionality in Norwegian farmland abandonment: Inferences from production data. *Applied Geography* 55: 238–247.
- Scott, W.A. & Hallam, C.J. 2003. Assessing species misidentification rates through quality assurance of vegetation monitoring. *Plant Ecology* 165: 101–115.
- Sokal, R.R. & Rohlf, F.J. 1995. *Biometry: the principles and practice of statistics in biological research*. 3rd edition. Freeman, New York, US.
- Stevens, J.P., Blackstock, T.H., Howe, E.A. & Stevens, D.P. 2004. Repeatability of Phase 1 habitat survey. *Journal of Environmental Management* 73: 53–59.
- Sundseth, K. 2008. *NATURA 2000: protecting Europe's biodiversity*. European Commission, Oxford, UK.
- Sykes, J.M., Horrill, A.D. & Mountford, M.D. 1983. Use of Visual Cover Assessments as Quantitative Estimators of Some British Woodland Taxa. *Journal of Ecology* 71: 437–450.
- Thompson, I.D., Maher, S.C., Rouillard, D.P., Fryxell, J.M. & Baker, J.A. 2007. Accuracy of forest inventory mapping: Some implications for boreal forest management. *Forest Ecology and Management* 252: 208–221.
- Tulloch, V.J., Possingham, H.P., Jupiter, S.D., Roelfsema, C., Tulloch, A.I.T. & Klein, C.J. 2013. Incorporating uncertainty associated with habitat data in marine reserve design. *Biological Conservation* 162: 41–51.
- Ullerud, H.A., Bryn, A., Halvorsen, R. & Hemsing, L.Ø. 2018. Consistency in land cover mapping: influence of fieldworkers, spatial scale and classification system. *Applied Vegetation Science* 21: 278–288.
- Ver Hoef, J.M. & Boveng, P.L. 2007. Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology* 88: 2766–2772.
- Vittoz, P. & Guisan, A. 2007. How reliable is the monitoring of permanent vegetation plots? A test with multiple observers. *Journal of Vegetation Science* 18: 413–422.
- Whittaker, R.H. 1956. Vegetation of the Great Smoky Mountains. *Ecological Monographs* 26: 1–80.
- Wintle, B.C., Fidler, F., Vesik, P.A., L. Moore, J. & Armsworth, P. 2013. Improving visual estimation through active feedback. *Methods in Ecology and Evolution* 4: 53–62.
- Økland, T. 1996. Vegetation-environment relationships of boreal spruce forest in ten monitoring reference areas in Norway. *Sommerfeltia* 22: 1–349.

Author addresses

Eriksen, E.L. (Corresponding author, e.l.eriksen@nhm.uio.no)¹, **Ullerud, H.A.** (h.a.ullerud@nhm.uio.no)¹, **Halvorsen, R.** (rune.halvorsen@nhm.uio.no)¹, **Aune, S.** (sigrun.aune@nibio.no)^{2,1}, **Bratli, H.** (harald.bratli@nhm.uio.no)¹, **Horvath, P.** (peter.horvath@geo.uio.no)¹, **Volden, I.K.** (ingerkvolden@gmail.com)¹, **Wollan, A.K.** (a.k.wollan@nhm.uio.no)¹, **Bryn, A.** (anders.bryn@nhm.uio.no)¹

¹ Natural History Museum, P.O. Box 1172, Blindern 0318 Oslo, Norway

² Norwegian Institute of Bioeconomy Research, Klæbuveien 153, 7031 Trondheim, Norway

Electronic supplements

Supplementary material associated with this article is embedded in the article's pdf. The online version of Phytocoenologia is hosted at the journal's website www.schweizerbart.com/journals/phyto. The publisher does not bear any liability for the lack of usability or correctness of supplementary material.

Supplement S1: Method for estimating Ecological Distance

Supplement S2: Additional Tables

Supplement S3: Translations of NiN terms

Supplement S4: Knowledge test questionnaire

Please save the electronic supplement contained in this pdf-file by clicking the blue frame above. After saving rename the file extension to .zip (for security reasons Adobe does not allow to embed .exe, .zip, .rar etc. files).