BIODIVERSITY RESEARCH

Diversity and Distributions   WILEY

Composite landscape predictors improve distribution models of ecosystem types

Trond Simensen[1,2] (iD) | Peter Horvath[1] (iD) | Julien Vollering[1,3] (iD) | Lars Erikstad[1,4] |
Rune Halvorsen[1] (iD) | Anders Bryn[1,5] (iD)

[1]Geo-Ecology Research Group, Department of Research and Collections, Natural History Museum, University of Oslo, Oslo, Norway

[2]Norwegian Environment Agency, Trondheim, Norway

[3]Department of Environmental Sciences, Western Norway University of Applied Sciences, Sogndal, Norway

[4]Norwegian Institute for Nature Research (NINA), Oslo, Norway

[5]Division of Survey and Statistics, Norwegian Institute of Bioeconomy Research, Ås, Norway

Correspondence
Trond Simensen, Department of Research and Collections, Natural History Museum, Univ. of Oslo, Oslo, Norway.
Email: trond.simensen@miljodir.no

Funding information
The Research Council of Norway; Norwegian Environment Agency; Norwegian Biodiversity Information Centre

Editor: Martin Jung

Abstract

Aim: Distribution modelling is a useful approach to obtain knowledge about the spatial distribution of biodiversity, required for, for example, red-list assessments. While distribution modelling methods have been applied mostly to single species, modelling of communities and ecosystems (EDM; ecosystem-level distribution modelling) produces results that are more directly relevant for management and decision-making. Although the choice of predictors is a pivotal part of the modelling process, few studies have compared the suitability of different sets of predictors for EDM. In this study, we compare the performance of 50 single environmental variables with that of 11 composite landscape gradients (CLGs) for prediction of ecosystem types. The CLGs represent gradients in landscape element composition derived from multivariate analyses, for example "inner-outer coast" and "land use intensity."

Location: Norway.

Methods: We used data from field-based ecosystem-type mapping of nine ecosystem types, and environmental variables with a resolution of 100 × 100 m. We built nine models for each ecosystem type with variables from different predictor sets. Logistic regression with forward selection of variables was used for EDM. Models were evaluated with independently collected data.

Results: Most ecosystem types could be predicted reliably, although model performance differed among ecosystem types. We identified significant differences in predictive power and model parsimony across models built from different predictor sets. Climatic variables alone performed poorly, indicating that the current climate alone is not sufficient to predict the current distribution of ecosystems. Used alone, the CLGs resulted in parsimonious models with relatively high predictive power. Used together with other variables, they consistently improved the models.

Main conclusions: Our study highlights the importance of variable selection in EDM. We argue that the use of composite variables as proxies for complex environmental gradients has the potential to improve predictions from EDMs and thus to inform conservation planning as well as improve the precision and credibility of red lists and global change assessments.

1 | INTRODUCTION

Human impact transforms nature all over the world (Ellis, Goldewijk, Siebert, Lightman, & Ramankutty, 2010), and the need for sustainable management of ecosystems is increasing (Díaz et al., 2019). To understand, monitor and manage Nature's diversity, that is, the variation in Earth's biotic and abiotic processes and features (see e.g. Zarnetske et al., 2019), we must know where this diversity is (Whittaker et al., 2005). Although management strategies such as red lists for ecosystems require systematic mapping of nature, that is, high-quality land cover maps of ecosystems (Keith et al., 2015), only a minor fraction of the Earth's surface has so far been mapped by field survey methods (Alexander & Millington, 2000). Remote sensing methods, although useful for a wide range of purposes, have not yet proven able to interpret community structure and species composition in ecosystems with the geographic and thematic accuracy required for many research and management purposes (Myers-Smith et al., 2020; Strand, 2013). Alternative, efficient pathways to information about the spatial distribution of ecosystems over large areas are therefore needed to enhance the precision and credibility of red lists and global change assessments. Recent guidelines and studies point to distribution modelling as a promising tool for this purpose (Bland, Keith, Miller, Murray, & Rodríguez, 2017; Horvath et al., 2019).

Distribution models are models that treat the geographic distribution of observable objects of a specific type (e.g. species) as a response to a set of supplied predictors (Halvorsen, 2012). Although single species are the most common target for distribution modelling (Henderson, Ohmann, Gregory, Roberts, & Zald, 2014), distribution modelling methods are, in principle, applicable to target objects of many kinds, for example, species assemblages or species groups (Pottier et al., 2013), patterns of species richness (Santos et al., 2020), plant communities (Franklin, 2013; Jiménez-Alfaro et al., 2018; Ovaskainen & Soininen, 2011), potential vegetation (Hemsing & Bryn, 2012), present and past "vegetation types" (Horvath et al., 2019; Janská et al., 2017; Longcore, Noujdina, & Dixon, 2019) and "ecosystem types" (Halvorsen, 2012). In this article, we use the term "ecosystem-level distribution modelling" (EDM) as an umbrella term for distribution modelling with units above the species level as modelling targets. The term applies to biotic communities, defined by species composition, as well as to ecosystems and their abiotic components. Compared with distribution modelling of single species, EDM has technical as well as practical advantages such as increased power to detect shared environmental patterns for multiple species, and, perhaps more importantly, enhanced potential to generate results relevant for management and decision-making (Ferrier & Guisan, 2006). Nevertheless, neither the encouragement of Ferrier and Guisan (2006) in their review of EDM studies nor the

increasing popularity of species distribution modelling (Araújo et al., 2019; Lobo, Jiménez-Valverde, & Hortal, 2010) has so far triggered extensive use of EDM. Several methodological questions in EDM therefore await closer examination.

Access to relevant predictor variables is pivotal for any application of distribution modelling (Araújo et al., 2019), but the theoretically optimal predictor set is difficult to identify as well as to obtain (Austin, 2002). Therefore, some important variables tend to be missing from most distribution models, reflecting: (a) lack of knowledge about which environmental factors cause the current distribution of a modelling target; and/or (b) lack of spatial data that represent processes and attributes known to be important (Austin, 2002; Barry & Elith, 2006). Few studies have explicitly addressed the relative suitability of different predictors for EDM (Halvorsen, 2012, but see Jiménez-Alfaro et al., 2018). Hence, a better understanding of which proximate variables cause the spatial distribution of different ecosystems is needed in EDM, along with better spatial proxies for these variables for applied modelling purposes.

Jiménez-Alfaro et al. (2018) concluded that any community type at any hierarchical level may be modelled at continental extent, provided it is consistently defined by species composition and constrained by environmental factors. Hierarchy theory has shown that the aggregation of similar components into fewer composite units (i.e. numerous species into fewer functional types or species into communities and ecosystems) may reduce the number of variables required to obtain models of a given quality, that is, with a certain predictive power (e.g. Allen & Starr, 2017). However, a shift of modelling target from single species to community or ecosystem may require a reformulation of the "ecological model," that is, the theoretical basis for the modelling process (Austin, 2002). With reference to ecosystems as modelling targets, we define "ecosystem types" as recurrent abstract "units of assessment that represent complexes of organisms and their associated physical environment within an area" (Keith et al., 2015, based on Tansley, 1935).

The concept of the environmental complex-gradient (Whittaker, 1956), that is a set of correlated environmental variables that act on the species in concert rather than one by one, is fundamental for describing and understanding variation in species' responses to the environment (Halvorsen, 2012). We hypothesize that the complex-gradient concept, commonly used to understand and describe species' relationships to the environment, can be extended to the landscape level as well as be implemented in studies of the distribution of ecosystem types. This extension implies that each level of ecological diversity contains subsystems at the level below (Turner & Gardner, 2015); landscapes contain ecosystems and other landscape elements, while ecosystems contain species and their environment (Halvorsen, Bryn, & Erikstad, 2016; Noss, 1990). Accordingly, we define a "complex landscape gradient" (CLG) as an "abstract continuous

variable that expresses more or less gradual, coordinated change in a set of more or less strongly correlated landscape variables." Thus, CLGs are composite variables expressing parallel, gradual or discontinuous variation in the presence and/or abundance of landscape elements. We define a "landscape element" as a "natural or human-induced object or characteristic, including spatial units assigned to types at an ecological diversity level lower than the landscape level, which can be identified and observed on a spatial scale relevant for the landscape level of ecological diversity" (Halvorsen et al., 2016). Composite landscape gradients can be obtained from multivariate analyses of landscape element compositional data undertaken to reduce the dimensionality of an *n*-dimensional landscape-level hyperspace (Erikstad, Uttakleiv, & Halvorsen, 2015). Furthermore, segments along two or more CLGs can be combined into "landscape types," defined as "more or less uniform areas characterized by their content of observable, natural and human-induced landscape elements."

Our definitions establish variation in landscape element composition along CLGs as a parallel to the spatiotemporal domain defined by Delcourt, Delcourt, and Webb (1982) as "meso-scale," capturing abiotic and biotic patterns that occur at spatial scales of approximately $10^6–10^{10}$ m^2 in response to processes operating at temporal scales of $10–10^4$ years (e.g. geomorphological processes, climatic fluctuations, human land use, fire regimes, etc.). Analyses of data from Norway indicate that response curves of landscape elements (including ecosystems) along CLGs bear resemblance to species response curves along local environmental complex gradients (see Figure 1; Erikstad, Halvorsen, & Simensen, 2019); most ecosystems appear to have distinct optima along CLGs, that is, intervals in which they reach maximum occurrence probability. If this is the case, such landscape gradients may potentially be useful as predictors of ecosystem types in EDMs.

The aim of this study was threefold: (a) to explore how well distributions of ecosystem types can be predicted; (b) to compare the predictive power of different sets of predictors in EDM; and (c) to test if EDM can be improved by using of composite "landscape predictors" (CLGs and landscape types) as predictors.

2 | METHODS

2.1 | Study area

The study area comprised the entire mainland of Norway including coastal islands, but excluding the Svalbard archipelago, Jan Mayen and Bear Island, spanning latitudes from 57°57′N to 71°11′N and longitudes from 4°29′E to 31°10′E. Mainland Norway covers only 323,802 km^2, but comprises an exceptional range of natural variation, given its moderate size (Halvorsen et al., 2016), including both terrestrial, marine, limnic and snow and ice ecosystems. The study area is characterized by a wide range of climatic variation; all seven temperature-related vegetation zones commonly recognized in northern Europe (from boreo-nemoral to high alpine) occur in Norway (Bakkestuen, Erikstad, & Halvorsen, 2008). Norway has a high mineral and bedrock diversity (Ramberg, Bryhni, Nøttvedt, & Rangnes, 2008), and high diversity of landforms (Gjessing, 1978). In addition to natural variation, the diversity of ecosystems in Norway is enhanced by variation in human land use. Throughout history, most Norwegian ecosystems have been affected by land use activities such as domestic grazing, outfield fodder collection, heath burning, reindeer husbandry, forestry, and industrial, urban and recreational development (Almås, Gjerdåker, Lunden, Myhre, & Øye, 2004). The diversity of Norwegian ecosystems and landscapes is thoroughly described by the theoretical framework "Nature in Norway" (Halvorsen
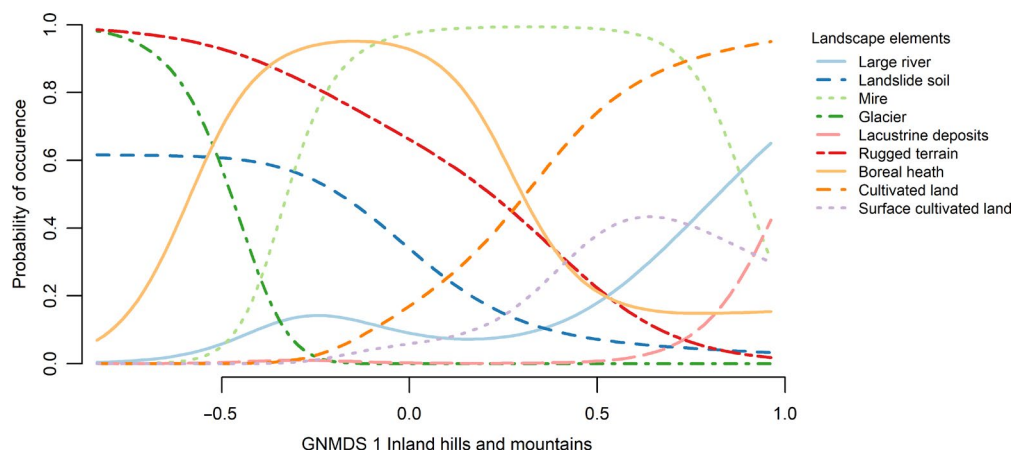

FIGURE 1 "Landscape element response plot," showing the distributions of eleven landscape elements along a "complex landscape gradient" within inland hills and mountains, as identified by ordination axis 1 obtained by use of global non-metric multidimensional scaling (GNMDS). The gradient reflects variation in abundance of landscape elements from steep rugged barren mountains (high alpine areas, left side of axis 1) towards areas with gentle slope in the lowland (right). The response curve is derived from ordination of 85 landscape elements recorded in 3,966 sampling units throughout Norway. Axes are scaled in half-change (H.C.) units: one unit corresponds to 50% turnover of landscape element composition

et al., 2016) from which terms, definitions and typologies applied in this study have been obtained.

2.2 | Response variables

The response variables in our study are (the occurrence of) nine terrestrial ecosystem types whose management (i.e. conservation planning, general land use planning and red-list assessments) would benefit specifically from better knowledge of their spatial distributions (NBIC, 2018). The nine ecosystem types in our study are difficult to map reliably by remote sensing methods (Erikstad et al., 2009; Strand, 2013), and none of them are currently included in full-coverage data sets for Norway.

Data from field-based vegetation and ecosystem-type mapping during the period 2004–2018 were used as training data for parameterization of EDM models for each type (Figure 2). We chose ecosystem types with equivalent definitions in two different systems of types, so that different sources of field data could be combined (Bryn, Strand, Angeloff, & Rekdal, 2018; Bryn & Ullerud, 2018; see Appendices S2 and S6). The raw data for the response variables were collected from three sources: (a) regional 9 × 9 km-grid surveys (AR 9 × 9, e.g. Bryn et al., 2015; Rekdal & Angeloff, 2013); (b) a subset of vegetation maps produced in the period 2004–2014 (NIBIO,

2018); and (c) data from ecosystem-type mapping conducted by the Norwegian Environment Agency (2019). From the raw response data, a post hoc processing was conducted to reduce bias related to spatial clustering of the training data (e.g. spatial autocorrelation, see Appendix S4). The training data used for each ecosystem type consisted of all presences remaining after post hoc processing and a random sample of ~10 000 true absences (Nad'o & Kaňuch, 2018; Table 1, Appendix S3).

2.3 | Predictors

We use the term "predictor" as a collective term for variables potentially accounting for variation in response variables, including continuous variables as well as categorical variables with several classes (e.g. land cover types). All our predictors were generated with a grain (pixel) size of 100 × 100 m, or adapted to this grain size by rasterization from vector formats or interpolation by kriging (see Table 2 and Appendix S7 for further details). Strongly correlated variables ($|\tau| > 0.7$) were omitted (Appendix S18). For model building, we used predictors from three qualitatively different predictor sets (basic climatic, geological, biological predictors, landscape predictors and neutral pseudo-predictors), and combinations of these (Table 2, Figure 2).
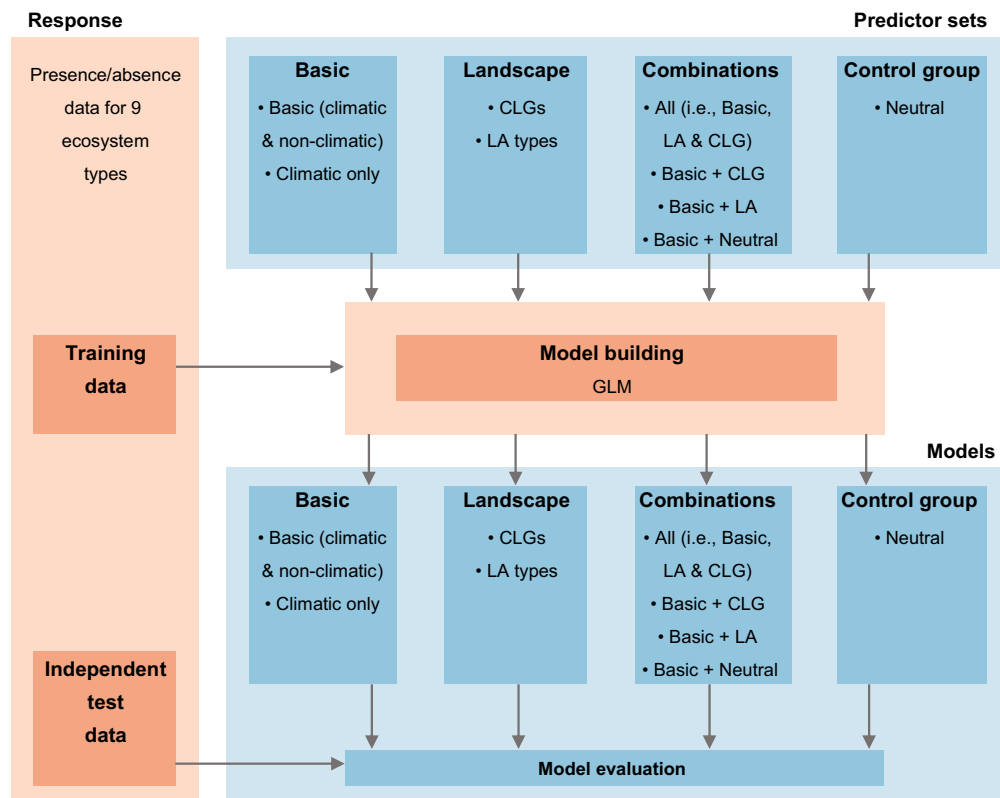

FIGURE 2 Methodological overview. For each of nine ecosystem types, we used one presence/absence response variable and predictors from qualitatively different predictor sets to build nine different models: two models from the 50 "basic" predictors (i.e. all single environmental variables including climatic variables and climatic variables only); two models from landscape variables (landscape types and complex landscape gradients, that is CLGs); four models from combinations of predictors from the other predictor sets; and one model from the control group of pseudo-predictors. The predictor sets represent different pools of predictors, available for model selection

The "basic predictors" set consisted of 50 single variables previously successfully used for distribution modelling of vegetation types in Norway (see Horvath et al., 2019, Appendices S7 and S9 for details). This set contains: (a) two geological variables (with a total of 37 categories of variation in bedrock and quaternary deposits); (b) 10 topographical variables; (c) five hydrological variables; (d) 32 climatic variables (mean and extreme values for temperature and precipitation, snow-covered area, snow water equivalent and 19 BIOCLIM variables); and (e) a land cover variable (9 classes of the "Land Resource Map" AR50).

The "landscape predictors" set consisted of 13 variables developed as a part of the new system for description and mapping of ecological diversity; Nature in Norway (Halvorsen et al., 2016; Appendix S11). Of these 13, the 11 CLGs represent variation in landscape element composition within three functional categories: (a) geo-ecological; (b) bio-ecological; and (c) land use-induced landscape variation. Specifically, the CLG variables were identified by multivariate analyses of data collected in a stratified sample of 100 test areas (25 × 25 km) that cover 56,400 km², or about one sixth, of mainland Norway. Within these test areas, a total of 3,966 observation units (landscape polygons 4–30 km²) were delineated based on geomorphological criteria. For each observation unit, 85 qualitative or quantitative landscape variables were recorded (see Appendix S10). These 85 landscape variables represent the occurrence or abundance of concrete, observable, landscape elements (e.g. mires, rivers, flat areas, etc.). Indirect drivers of, and ultimate causes for, the distribution of these elements (e.g. climatic variables) were avoided. We identified patterns of variation in landscape element composition by parallel use of detrended correspondence analysis (DCA; Hill & Gauch, 1980) and global non-metric multidimensional scaling (GNMDS; Minchin, 1987). To avoid circularity, the nine targets for EDM were deliberately selected not to be among the 85 variables used for identifying CLGs.

The Nature in Norway analyses (Erikstad et al., 2019) supported a division of the study area into six "major landscape types" identified by geomorphological criteria: inland hills and mountains; inland valleys; inland plains; coastal plains; coastal fjords and coastal hills and mountains. To further subdivide these major types into "minor landscape types," we first extracted groups of highly correlated variables as candidates for CLGs within functional variable categories (abiotic, biotic and land use-related variation). For each of these variable groups, we used constrained ordination (RDA, ter Braak, 1985) with forward selection of variables to obtain a parsimonious and orthogonal set of variables to represent each CLG. The CLGs obtained were subsequently divided into a number of discrete intervals, depending on the total length of each CLG as measured in units of compositional turnover. Each interval comprised a fixed amount (8%) of compositional turnover along a CLG. GIS-based proxies for the 11 CLGs were finally projected to the entire study area (Table 2, Appendices S12 and S20). Landscape types were obtained by combining segments along all CLGs identified as important for a given major landscape type; every unique combination of intervals along the set of relevant landscape gradients defined one landscape type

TABLE 1 Properties of modelled ecosystem types and model performance statistics (AUC and TSS). Ecosystem types (rows) are ordered by decreasing mean AUC for the n = 9 models for each ecosystem type

Ecosystem type*	Prevalence (% of study area)**	Red list status***	No of training points		No of evaluation points		Best AUC value	Best TSS value	Best predictor set, metric = AUC****
			Presence	Absence	Presence	Absence			
T22 Arctic-alpine dry grass heath (2b)	0.50	LC	490	9,904	133	9,960	0.927	0.796	Basic
T34 Coastal heath (10b)	0.50	EN	514	9,808	97	9,484	0.92	0.707	Basic + LA
T32 Semi-natural grassland (11b)	0.90	VU	1,754	10,053	294	10,132	0.898	0.682	Basic + CLG
T27 Boulder field (12b)	4.60	LC	1,008	10,085	719	10,152	0.863	0.606	CLG
T1 Bare rock (12c)	3.50	LC	268	10,113	572	10,184	0.862	0.576	Basic + CLG
V2 Mire & swamp forest (8b, 8c, 8d)	2.90	LC	3,529	9,879	1,201	9,961	0.824	0.521	Basic + CLG
T14 Exposed ridge (2c)	6.00	NT	4,030	9,901	1,661	9,951	0.809	0.507	All
V3 Bog (9a)	2.40	NT	3,460	9,868	929	9,946	0.796	0.456	All
V1 Open fen (9b, 9c)	6.10	LC	2,931	9,918	2,144	10,023	0.775	0.415	All

*Corresponding type in the Norwegian NIBIO type system in parentheses, **Based on estimates from Bryn et al. (2018), ***NBIC (2018).
Red-list criteria: LC, Least Concern; NT, Near Threatened; VU, Vulnerable; EN, Endangered. Several ecosystem types listed as LC include red-listed subtypes, ****See Table 2 for description of the predictor sets.

TABLE 2 Predictor sets, and properties of the predictors included in each set.

Predictor set	Functional variable category	Theme/data set	Description	Statistical variable category	Variables	Classes	Orig. Res.	Source
Basic	CLIM	Bioclim	Bioclimatic and climatic data	Cont.	32	–	1 km	SeNorge MET
	TOPO	Terrain variation	DEM and terrain variables derived from DEM	Cont.	10	–	100 m	NMA
	GEO	Geology	Bedrock and quaternary geology	Cat.	2	37	1:50 K - 1:250 K	NGU
	HYDRO	Hydrology	Proximity to hydrological surfaces	Cont.	5	–	1:50 K	NVE
	LC	Land cover	Land Resource Map (AR50)	Cat.	1	9	1:50 K	NIBIO
Landscape	LA	LA types	Major and minor landscape types	Cat.	2	290	1:50 K	NBIC
	Geo-eco.	CLG	Relief in hills and mountains	Cat	1	5	1:50 K	NBIC
	Geo-eco.	CLG	Relief in valleys and fjords	Cat	1	4	1:50 K	NBIC
	Geo-eco.	CLG	Relief in coastal plains	Cat	1	3	1:50 K	NBIC
	Geo-eco.	CLG	Distance to coast	Cont.	1	–	1:50 K	NBIC
	Geo-eco.	CLG	Inner-outer coast	Cat	1	5	1:50 K	NBIC
	Geo-eco.	CLG	Abundance of lakes	Cat	1	3	1:50 K	NBIC
	Geo-eco.	CLG	Abundance of wetlands	Cont.	1	–	1:50 K	NBIC
	Geo-eco.	CLG	Glacier presence	Cat	1	2	1:50 K	NBIC
	Bio-eco	CLG	Vegetation cover	Cat	1	4	1:50 K	NBIC
	Land use	CLG	Land use intensity	Cont.	1	–	1:50 K	NBIC
	Land use	CLG	Agricultural land use intensity	Cont.	1	–	1:50 K	NBIC
Neutral	NEU	NEU	Continuous neutral models	Cont.	7	–	100 m	-
	NEU	NEU	Discrete neutral models	Cat.	4	1,000	100 m	-

Note: All predictors were adjusted from their original resolutions into a grain size of 100 x 100 m (see also Appendices S7 and S9)

Abbreviations: cat., categorical; CLG, complex landscape gradient; CLIM, climatic; cont., continuous; GEO, geo-ecological; HYDRO, hydrological; IMR, Institute of Marine Research; LC, land cover; MET, Norwegian Meteorological Institute; NBIC, Norwegian Biodiversity Information Centre; NEU, Neutral landscape models; NGU, The Geological Survey of Norway; NIBIO, Norwegian Institute of Bioeconomy Research; NMA, Norwegian Mapping Authority; NVE, The Norwegian Water Resources and Energy Directorate; TOPO, topographic.

(Erikstad et al., 2019). Hence, the twelfth and thirteenth "landscape predictor" consisted of six major and 284 minor landscape types, respectively.

The "neutral predictors" set consisted of 11 variables derived from "neutral landscape models" (Gardner, Milne, Turner, & O'Neill, 1987; Appendix S13) as a control group for estimating the magnitude of differences in model performance that could arise by chance alone (see e.g. Fourcade, Besnard, & Secondi, 2018). These predictors, referred to as "pseudo-predictors," are completely artificial but show similar levels of spatial autocorrelation as the basic and landscape predictors.

2.4 | Model building

We fit generalized linear models (GLM; McCullagh & Nelder, 1989) with logit link function and binomial errors ("logistic regression") to the occurrence probability of each of the nine ecosystem types, as recommended for presence/absence data by Elith and Leathwick (2009). "Derived" predictors were obtained from original predictors by seven different transformation types: linear, monotonous, deviation, forward hinge, reverse hinge, threshold and binary (Vollering, Halvorsen, & Mazzoni, 2019). The effect of variable transformation is that the functional relationship between the occurrence of the modelled target and a predictor can be described more flexibly than if only the original predictors were allowed to enter the model (Vollering et al., 2019).

We selected variables by using an automated forward stepwise selection procedure based on F tests of nested models. First, a representative group of "derived predictors" was selected for each individual predictor based on their explanatory power. Single derived predictors were added to the model until no more derived predictors could be added that satisfied the pre-set threshold significance criterion (α = 0.001, see Vollering et al., 2019). Second, selection of predictors (each represented by a set of selected derived predictors) was performed by the same forward selection procedure (Vollering et al., 2019; see Appendices S8, S16 and S20 for details).

The full variable selection procedure was repeated for all predictor sets (1–9) described in Table 3. The resulting 81 models thus represented unique combinations of predictor variables and targeted ecosystem types. For models with two sets of predictors joined by a "+" sign, landscape or neutral variables (explaining significant amount of variation) were *added* to the best model derived from the basic predictors, to test the effect of adding the second set on model performance (see Bailey, Boyd, & Field, 2018). An example of models derived from different predictor sets is provided in Table 4, for the ecosystem-type T32 Semi-natural grasslands.

2.5 | Model evaluation

For evaluation of EDM models, we used a data set that was collected independently of the training data. This data set was obtained from the sample-based area frame field survey programme AR18×18 (Strand, 2013). The survey, which was conducted 2004–2014, included a systematic survey of 1,081 plots, each 0.9 km², distributed over the Norwegian mainland according to the 18 × 18 km LUCAS grid (Bryn et al., 2018; Strand, 2013). The survey encompassed variation along all major environmental gradients recognized in Norway, covering the full spatial extent of the study area (see Appendices S3, S4, S19 and S20).

We applied two discrimination metrics for evaluation of EDM models; the area under the receiver operator characteristic curve (AUC; Fielding & Bell, 1997) and the true skill statistic (TSS) maximized for model specificity and sensitivity (Liu, Berry, Dawson, & Pearson, 2005). Measurements by the AUC and TSS metrics were strongly correlated (Kendall's rank correlation coefficient, τ = 0.90, p < .001). Since the two metrics yielded similar results, models were classified based on AUC values according to the criterion of Araújo, Pearson, Thuiller, and Erhard (2005), as: excellent (AUC > 0.90); good (0.80 < AUC < 0.90); fair (0.70 < AUC < 0.80); poor (0.60 < AUC < 0.70); and fail (0.50 < AUC < 0.60). We assessed model performance across predictor sets by comparing key model properties—values of the AUC and TSS statistics, and the number of predictors included in the models. The Kruskal–Wallis test and multiple paired Wilcoxon rank-sum tests were used to test for significant differences in model performance across groups of models and between predictor sets. We assessed model parsimony by relating predictor number to model performance.

We used R version 3.5.2 for all statistical analyses (R Core Team, 2018; see tools and packages in Appendix S1) and followed the distribution modelling standards recommended by Araújo et al. (2019; Appendix S5).

3 | RESULTS

3.1 | Model performance

AUC values for the 81 models obtained for the nine ecosystem types ranged from 0.524 to 0.919 (Table 3). The best EDMs for seven out of nine ecosystem types were classified as "good" (AUC > 0.8), two of these as "excellent" (AUC > 0.9). The best models as judged by AUC were developed for T22 "Dry grass alpine heath," T34 "Coastal heathland" and T32 "Grassland" (Table 1). The ecosystem types discriminated with lowest success were V3 "Bog" and V1 "Open fen."

3.2 | Performance of the predictors

When all variables were available for model building (i.e. predictor set = "all"), the variables that were most often included in the models were as follows: "quaternary geology" (included in models for n = 8 ecosystem types); AR50 land cover (n = 6); "standard curvature" (a local morphometric terrain parameter, n = 6); CLG "distance to coast" (n = 5); and CLG "vegetation cover" (n = 5).

TABLE 3 Summary of model performance for each predictor set, based upon AUC.

Predictor set	Description	Model performance					Model parsimony	
		Mean AUC	Max AUC	Min AUC	Similar sets	Rank sum	Mean number of predictors	Mean number of derived predictors
Basic + CLG	Model derived from selection from climatic and non-climatic basic predictors; Complex landscape gradients (CLGs) explaining significant variation added subsequently	0.848	0.919	0.775	a	19	14.4	38.8
All (i.e., Basic, LA & CLG)	Model where all variables except pseudo-predictors were available for selection (i.e. climatic and non-climatic basic predictors, landscape types and CLGs)	0.846	0.911	0.775	a	25	10.9	40.6
Basic + LA	Model derived from selection from climatic and non-climatic basic predictors; Landscape types explaining significant variation added subsequently	0.837	0.92	0.757	ab	29	10.8	46.1
Basic	Model derived from a selection from climatic and non-climatic basic variables	0.835	0.927	0.756	ab	36	9.8	26.3
CLG	Model derived from complex landscape gradients only	0.821	0.897	0.713	b	42	5	13.6
Basic + Neutral	Model derived from selection from basic predictors; pseudo-predictors explaining significant variation added subsequently	0.807	0.918	0.72	bc	49	15.7	115
Clim	Model derived from climatic variables only	0.755	0.92	0.628	d	55	8.2	20.6
LA	Model derived from landscape types only	0.73	0.89	0.578	d	69	1.3	22.3
Neutral	Model derived from pseudo-predictors (neutral landscape models)	0.606	0.759	0.524	e	81	6.2	89.4

Note: Predictor sets represent non-overlapping "pools" of candidate predictors, from which any variable that explained a sufficient amount of variation could, in principle, be included in the models. Predictor sets containing a + sign represent cases for which selected variables from the last-mentioned predictor set were added to the final derived model for the ecosystem type, obtained solely by the first mentioned predictor set. Predictor sets (rows) are ordered by decreasing rank sum, based upon single-model AUC. Entries sharing a letter in the column "similar sets" (e.g. "ab") do not have significantly different median values at the $p < .05$ level (Wilcoxon paired rank-sum test). Each ecosystem type is modelled separately, and the number of ecosystem-type models is the same ($n = 9$) for all predictor sets

TABLE 4 Example showing nine models with different combinations of predictors for one out of the nine ecosystem types: T32 semi-natural grasslands.

Type example T32 Semi-natural grassland					
Semi-natural grassland includes meadows formed by forest or shrub clearance followed by livestock grazing and/or haymaking, subject to the additional condition of neither being subjected to ploughing nor reseeding nor heavy fertilization. The vegetation is dominated by graminoids and herbs, nitrophilous species are not prominent. Semi-natural grassland may be open (treeless) or, also when actively managed, have an open tree layer (wooded or coppice meadows). Land management intensity, lime richness and risk of drought are the most important LECs. Since the middle of the 20th century, traditional use of semi-natural grasslands has decreased and conversion into arable fields, agriculturally improved grassland or abandonment has taken place.					
Predictor set	**AUC-value**	**Max TSS-value**	**Number of predictors**	**Number of derived predictors**	**Predictors**
Basic + CLG	0.898	0.679	15	47	AR50 Land cover + valley depth + digital elevation model + proximity to coast + quaternary geology + precipitation seasonality + terrain ruggedness index + min. temp. September + aspect +6 CLGs (se row 3)
All (i.e., Basic, LA & CLG)	0.896	0.682	9	27	CLG Land use intensity + AR50 Land cover + visible sky + digital elevation model + CLG vegetation cover + CLG Agricultural land use intensity + quaternary geology + CLG distance to coast + distance to rivers
CLG	0.889	0.66	6	17	CLG Land use intensity + CLG Agricultural land use intensity + CLG vegetation cover + CLG freshwater lake properties + CLG terrain + CLG distance to coast
Basic	0.866	0.58	9	30	AR50 Land cover + valley depth + digital elevation model + proximity to coast + quaternary geology + precipitation seasonality + terrain ruggedness index + min. temp. September + aspect
Basic + LA	0.865	0.594	10	55	AR50 Land cover + valley depth + digital elevation model + proximity to coast + quaternary geology1 + precipitation seasonality + terrain ruggedness index + min. temp. September + aspect +landscape types (se row 7)
Basic + Neutral	0.827	0.545	15	113	AR50 Land cover + valley depth + digital elevation model + proximity to coast + quaternary geology1 + precipitation seasonality + terrain ruggedness index + min. temp. September + aspect +Neutral variables (see row 9)
LA	0.747	0.362	2	29	Minor landscape types + major landscape types
Clim	0.699	0.322	9	17	Growing season length + annual precipitation + temp. oscillations + Snow water equivalent in April + min. temp. September + mean temp. of driest quarter + min temperature of coldest month + maximum temperature in June + maximum temperature in October
Neutral	0.524	0.086	5	81	Random mosaic 1 + randomly shuffled landscape types + random mosaic 2 + randomly shuffled landscape types 2 + random gradient with midpoint displacement

Note: Predictor sets containing a + sign represent cases for which selected variables from the second predictor set were added to the final derived model for the ecosystem type, obtained solely by the first predictor set. Predictor sets (rows) are ordered by decreasing AUC. Corresponding tables for all ecosystem types are provided in Supporting Information.

Excluding the neutral predictor set, which produced low AUC values, model performance differed less among predictor sets than among ecosystem types (Figure 3). However, the Kruskal–Wallis rank-sum test revealed highly significant differences in AUC values also among predictor sets (χ^2 = 33.085, p < .001, df = 8). The best EDMs in terms of both AUC and TSS (Table 3; Appendix S14) were consistently obtained by adding landscape variables to the respective models derived from climatic and non-climatic basic predictors (Basic + CLG: mean AUC = 0.848, rank sum = 19). Adding selected

CLGs to the best model derived from "basic predictors" improved performance in 9 out of 10 cases, bringing about an average increase of AUC by 0.013 units (Wilcoxon paired samples test: V = 0, p = .0328). The fact that model performance *decreased* when selected pseudo-predictors were added to the models derived from basic predictors (mean AUC difference = −0.028; Wilcoxon paired samples test: V = 39, p = .0078) shows that the improvement of models by adding CLGs was not an effect of increasing number of predictors in the models (Appendix S15).
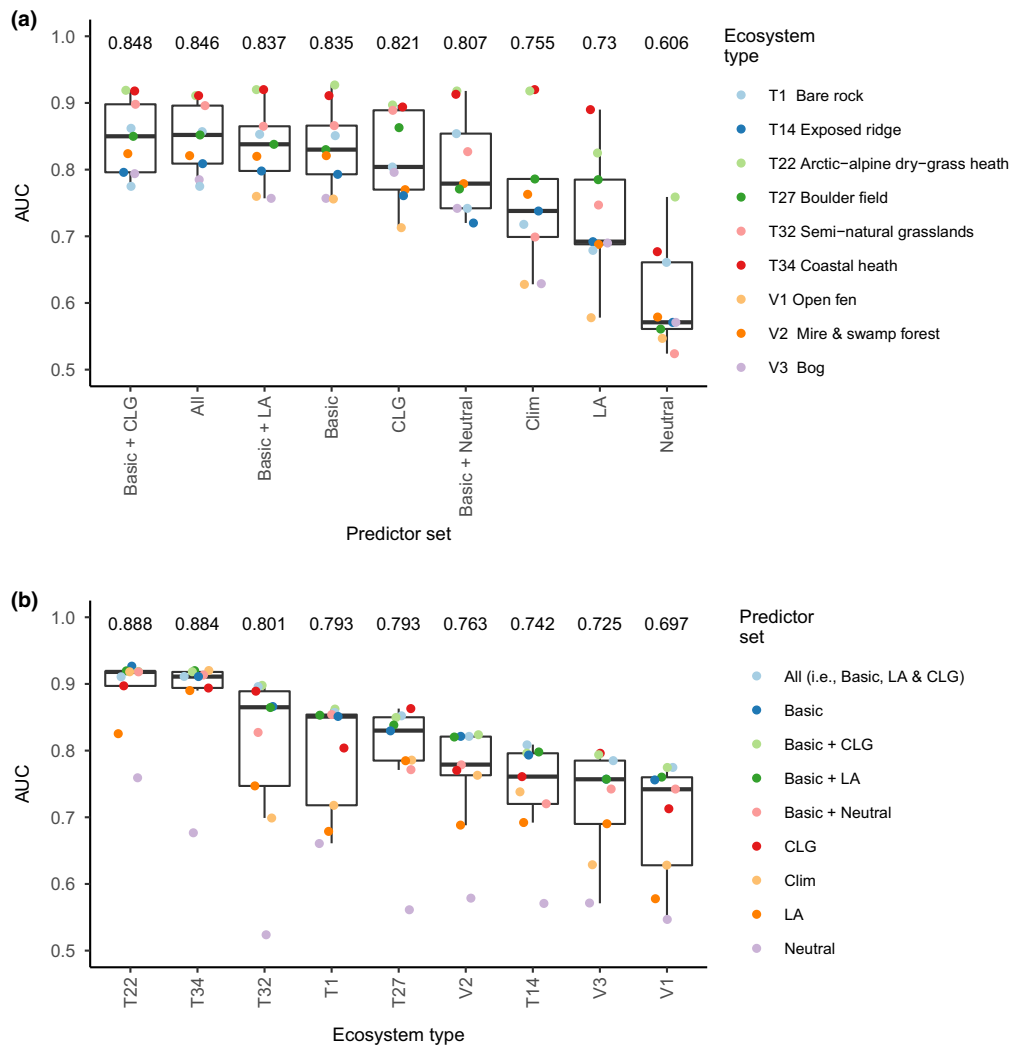
FIGURE 3 (a) Model performance measured as AUC (*y*-axis), for each of the nine predictor sets (*x*-axis). Each data point refers to one model, for one specific ecosystem type. Ecosystem types are indicated by different colours. Groups (predictor sets) are sorted by decreasing mean AUC for each group (values given above each box). (b) Model performance measured as AUC (*y*-axis), for each of the nine ecosystem types (*x*-axis). Predictor sets are indicated by different colours. Groups (ecosystem types) are sorted by decreasing mean AUC for each group (values given above each box)

The second-best predictor set ("All"; mean AUC = 0.846, rank sum = 25) consisted of models for which all except neutral predictors (both climatic and non-climatic basic predictors, CLGs and landscape types) were available for model building. Models derived solely from climatic and non-climatic "basic predictors" or solely from CLGs performed slightly less well (mean AUC = 0.835 and 0.821, rank sum = 36 and 42, respectively). Models built solely from climatic variables or from the coarser landscape-type variables (originally mapped as polygons 2–20 km^2) performed poorly (mean AUC = 0.730 and 0.755, rank sum = 55 and 69, respectively, see Table 3 and Figure 3).

The control group of pseudo-predictors gave rise to poor models for most ecosystem types (minimum, mean and maximum AUC values = 0.524, 0.606 and 0.759, respectively; rank sum = 81) despite the large number of derived predictors selected (mean = 89.4; Table 3).

3.3 | Model parsimony

The *simplest* models, obtained solely from landscape variables (Table 3, Figure 4 and Appendix S17: Figure S17), contained means of 1.3 predictors and 22.3 derived predictors when based upon landscape-type variables only and 5.0 predictors and 13.6 derived predictors when based upon CLGs. In contrast, the most complex models, obtained by adding pseudo-predictors to models derived from basic predictors, contained mean numbers of predictors and derived predictors of 15.7 and 115.0, respectively. Addition of landscape variables to the set of basic predictors also resulted in more complex models. Differences in model complexity (number of predictors) among models based upon different predictor sets were significant (Kruskal–Wallis rank-sum test, χ^2 = 89.657, *df* = 11, *p*-value <0.001). Bivariate plots of the relationship between model performance and model complexity

(Figure 4) indicated that the models obtained solely by CLGs resulted in the most parsimonious models when accounting for both model performance and number of predictors included in the models. Figure 4 shows that above a certain minimum number of predictors (≈10), the predictive power was negatively affected by including more predictors.

4 | DISCUSSION

4.1 | How well can ecosystem types be predicted?

Our results that EDMs for seven out of nine ecosystem types are "good" (AUC > 0.8) or "excellent" (AUC > 0.9) according to the criteria of Araújo et al. (2005), based upon independent model evaluation, show that discrimination of ecosystem types over large areas at a high spatial resolution (100 × 100 m) is possible. We thus consider our models to be of value for several practical management purposes, such as preparations for field-based mapping, as a supporting tool in red-list assessments, and as a knowledge base for land use and conservation planning at a regional scale in cases where detailed mapping is either not feasible or not necessary. Our results thus support the conclusion of Ferrier and Guisan (2006) that EDM deserves to be used more often, and more widely, as an alternative or a supplement to modelling of individual species.

The use of "pseudo-predictors" confirms that the model performance metrics reliably discriminate between models based upon randomly created gradients/patterns and models derived from real data, that is, environmental variables in the widest sense. Nevertheless, the substantial variation in predictive performance among pseudo-predictors calls for caution against over-interpretation of small differences in model performance also among sets of real predictors. However, the statistical tests demonstrate significant differences in the predictive power of different predictor sets in our study, even when the numerical differences in performance statistics are small. Moreover, due to inherent properties of the model performance metrics, relatively smaller differences represent more strongly significant model improvements in the upper range of AUC values, than at the lower values (see Fielding & Bell, 1997). We thus interpret the results of our analyses as sufficiently clear to represent general patterns in the overall ecological response of ecosystem types to different sets of explanatory variables.

4.2 | Finding the best predictors

Our results support the idea that identifying an ideal combination of predictors for each modelled target is critically important in EDM, like in species distribution modelling. The models with highest predictive power for each ecosystem type were characterized by a unique combination of predictors, not duplicated by models for any other ecosystem type in the study. This result is in

line with ecological theory, since the ecosystem types in our study are defined by differences in species composition as explained by variation along *different* sets of local environmental complex gradients and subject to the action of different structuring processes (Halvorsen et al., 2016; see also Appendix S2). Accordingly, variables to be used for EDM should be specifically selected for each ecosystem type.

The poor performance of climatic predictors alone indicates that current regional, that is climatic, variables alone are insufficient for explaining the distribution of ecosystem types at the resolution of our study. This is not surprising since the ecosystem types modelled in our study are defined by local complex variables and not by regional climatic conditions (see Appendix S2). Climatic predictors showed higher predictive power for ecosystem types well known to have a high probability of presence within a specific climatic region (e.g. T14 Exposed ridge and T22 Arctic-alpine dry grass heath). Furthermore, the current climate does not necessarily reflect the prevailing conditions in the period when the modelled target was established; there might for instance be time lag in species' responses to changes in environmental conditions (Bertrand et al., 2011; Guisan, Thuiller, & Zimmermann, 2017; Maiorano et al., 2013).

The fact that "basic predictors" (also including geological and land cover variables) performed significantly better than the climatic variables alone is in line with findings of, for example, Bailey et al. (2018) and Zarnetske et al. (2019), and confirm that both geodiversity and land use over time have strong effects on biodiversity patterns.

4.3 | Do "landscape predictors" improve distribution models of ecosystem types?

CLGs performed reasonably well as predictors of ecosystem types in our study: for a majority of ecosystem types, EDMs based upon CLG predictors efficiently accounted for considerable amounts of variation (i.e. high predictive power with few variables). The individual models (Table 4, Appendix S6) indicate that the CLGs enable flexible fitting of relationships between response and predictors. An example is the red-listed ecosystem-type "T32 Semi-natural grassland." The abundance of this ecosystem type increases gradually with the intensity of human land use to a certain point, before it decreases towards heavily utilized or highly urbanized areas (see frequency of observed presence-plot in Appendix S16). This unimodal relationship is well captured by the CLG "land use intensity," which includes the total abundance of buildings, infrastructure and man-induced land cover types. This information is lost in, for example, a traditional land cover map, in which the internal variability within, and the relationship between, the discrete, non-ordered classes are hidden.

Notably, the CLGs are indirect (distal) gradients that reproduce local landscape gradients at meso-scale, but that do not allow direct mechanistic modelling of the processes which give rise to the observed patterns. However, since
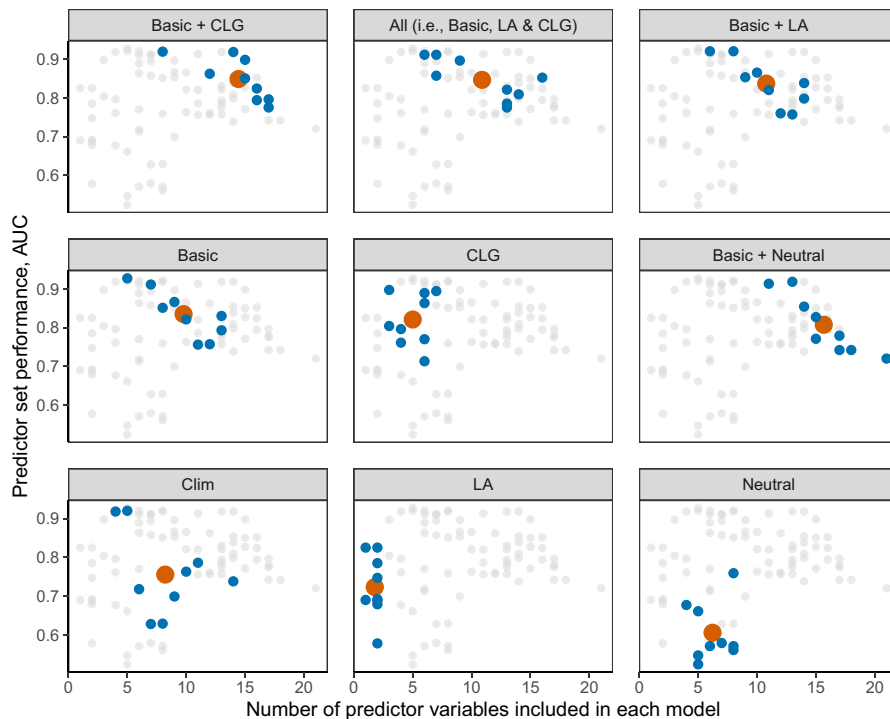
FIGURE 4 Model parsimony for the nine different predictor sets. Blue dots show all models derived from the same predictor set, depicted in separate panels with all 81 models (grey dots) shown in the background as a reference. Model complexity, measured as the number of predictors included in each model, on the *x*-axes. Model performance measured as AUC on the *y*-axes. The centroid (mean values on both axes) for each predictor set is indicated by orange dot. A model placed in the upper left part of a figure combines the desirable properties of small set of predictors and high predictive power (high parsimony), while a model placed in the lower right corner combines many variables with low predictive power (low parsimony). A corresponding figure showing model performance plotted against the number of derived predictors, see Appendix S17

direct (proximal) predictors that represent multiple drivers, operating over long time spans, are difficult or impossible to represent by adequate proxies, the aggregated patterns that result from such processes (CLGs) may in some cases serve as better surrogates for these processes than, for example, the current climate so often used for species distribution modelling. Our study demonstrates that CLGs extracted by ordination of landscape data may be of predictive significance for single ecosystem types that were not subject to prior ordination. The good performance of the CLGs demonstrated in this study provides an interesting parallel to distribution modelling studies at the species level for which it has been demonstrated that gradients obtained by multivariate methods (e.g. ordination axes) are better predictors of species diversity than single environmental variables (e.g. Ejrnæs, 2000; Margules, Nicholls, & Austin, 1987; Santos et al., 2020).

Gradient analysis at the landscape level is to a large extent still unexplored (but see Luck & Wu, 2002) and the "universality" of the CLGs and their potential for model extrapolation in time and space is therefore largely unknown. The composite and "data-driven" nature of the CLGs may theoretically limit their potential use for model projections and extrapolations in time and space. However, many of the landscape gradients recognized in our study are probably widely distributed and likely to remain stable over long timespans, even in a changing climate (e.g. terrain gradients, coast-inland gradients, gradients in human land use intensity, gradual changes in vegetation cover, abundance of lakes, etc.). We suggest that the distributions of ecosystems along complex gradients in the landscape should be explored across larger areas, as such studies may yield new insights of

both theoretical and practical importance (e.g. for extrapolatability of model predictions).

4.4 | Missing predictors

In line with, for example, Ullerud, Bryn, and Klanderud (2016) and Horvath et al. (2019), we find that some ecosystem types are more easily discriminated by means of available predictors than others. This is expected, since different ecosystem types will typically have different response curves to underlying environmental gradients. Our results show that there is inherently more difficult to discriminate presence from absence for ecosystem types with a broad and flat-topped (platykurtic) response curves along identified gradients, rather than types with sharper-peaked (leptokurtic) response curves. However, differences in model performance may also arise due to "missing predictors," that is, lack of spatial data representing attributes known to be important (see Barry & Elith, 2006). Even when the conditioning mechanisms are well understood, as in the case of the bog and fen ecosystem types in our study (V1 and V3, see Appendix S6: Table S6), quantitative data that can be used for modelling may currently be unavailable.

4.5 | Model complexity

After initial model selection based on internal evaluation (see Vollering et al., 2019), simple models performed consistently better in external evaluation data than more complex models (Figure 4). Our results indicate that the use of CLGs as predictors in EDM may
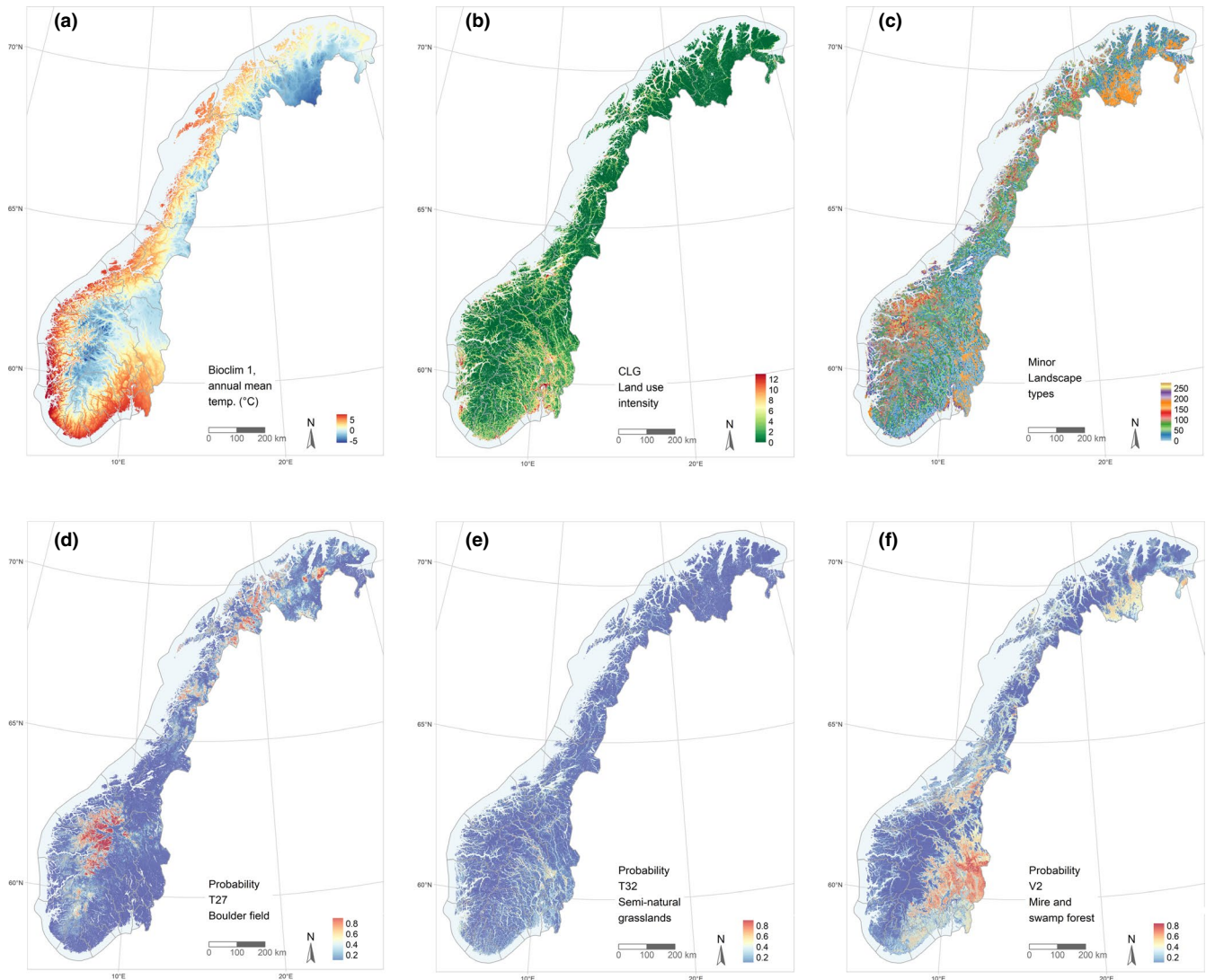
FIGURE 5 Examples of predictors and spatial prediction maps at a resolution of 100 × 100 m for the entire study area, the mainland of Norway, spanning 13 latitudinal and 27 longitudinal degrees. (a) "Bioclim 1, annual mean temperature," one of the 50 "Basic predictors," (b) "CLG land use intensity," one of the 11 complex landscape gradients (CLGs). (c) "Minor landscape types." (d–f) Spatial prediction of the ecosystem types: "T27 Boulder field"; "T32 Semi-natural grasslands"; and "V2 Mire and swamp forest"

constitute a route to statistically simpler models by reducing the number of model parameters, without introducing unsupported assumptions about simple cause–effect relationships. Statistically, the identification of complex gradients can be considered as a dimension-reduction method that coerces several correlated variables into a parsimonious set of composite, orthogonal variables. Although the principle of parsimony should not be used to underpin an unsupported belief that nature is simple, parsimony may have a true epistemic value in the evaluation of correlative and predictive models, as simpler models are less prone to overfitting (Coelho, Diniz-Filho, & Rangel, 2019; Halvorsen, 2012). One might argue that complex gradients (e.g. CLGs) derived from multiple correlated variables are by definition not simple. However, it is very difficult to provide GIS coverage for directly causal variables in distribution modelling at the scale of our study. Any variable used as predictor in distribution modelling will therefore to some extent serve as

an indirect proxy for inherently complex cause–effect relationships (see Austin, 2002).

4.6 | Conservation and management implications

In the procedure and criteria for red-list assessments of ecosystem types (IUCN, 2016), spatial distribution plays a key role. For many ecosystem types, the detailed distribution is unknown, and in practice, the assessment is therefore accomplished by expert judgements (NBIC, 2018). A good EDM can provide information about the distribution that may support and improve these expert judgments considerably. Together with risk modelling, cause–effect modelling and modelling of changes in distribution over time, EDM has been suggested as a preferred method for such assessments (Bland et al., 2017). Development of high-quality prediction maps is a cost-effective tool

for mapping the distribution of species and ecosystems, allowing for cost-efficient field efforts. The prediction maps developed in our study (Figure 5, Appendix S21) are suitable for use as a supporting tool in the upcoming red-list assessments for Norway (NBIC, 2018) as well as for planning of field-based mapping of red-listed ecosystem types (NBIC, 2018; Norwegian Environment Agency, 2019).

5 | CONCLUSIONS

The current spatial distribution of ecosystem is the result of climatic, geological, biological and human land use-related processes that have been acting in concert over thousands of years. Identifying variables that can predict the outcome of such aggregated processes is inherently difficult, but critically important in EDM. Our results indicate that ecosystems, just like species, may have distinct optima (maximum probability of occurrence) within specific segments of broad-scale gradients in the landscape. We suggest that improvements in EDM may be achieved by combining the development of better proxies for missing predictors with more knowledge about the distributions of ecosystems along complex gradients in the landscape, at several scales and levels. EDMs may complement field-based mapping and remote sensing in improving our knowledge about the spatial distribution of ecosystems. Such knowledge is essential for planning and management purposes, since the biodiversity of our planet cannot be managed species by species.

DATA AVAILABILITY STATEMENT
R-scripts, spatial predictions and raster data for the composite landscape predictors and the neutral predictors supporting the results in the paper are available for download from the DRYAD database at https://doi.org/10.5061/dryad.cjsxksn33. Other spatial data produced in this work are available on request from the authors. Due to the restrictions with ownership of the original area frame survey data (AR18X18), these data are not openly available from the authors.

ORCID
Trond Simensen https://orcid.org/0000-0002-7022-1349
Peter Horvath https://orcid.org/0000-0002-6017-5385
Julien Vollering https://orcid.org/0000-0002-7409-2898
Rune Halvorsen https://orcid.org/0000-0002-6859-7726
Anders Bryn https://orcid.org/0000-0003-4712-8266

REFERENCES
Alexander, R., & Millington, A. C. (Eds.) (2000). *Vegetation mapping*. Chichester, UK: John Wiley & Sons.

Allen, T. F., & Starr, T. B. (2017). *Hierarchy: Perspectives for ecological complexity*. Chicago, IL: University of Chicago Press.

Almås, R., Gjerdåker, B., Lunden, K., Myhre, B., & Øye, I. (2004). *Norwegian agricultural history*. Trondheim, Norway: Tapir.

Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., ... Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, *5*(1), eaat4858. https://doi.org/10.1126/sciadv.aat4858

Araújo, M. B., Pearson, R. G., Thuiller, W., & Erhard, M. (2005). Validation of species–climate impact models under climate change. *Global Change Biology*, *11*(9), 1504–1513. https://doi.org/10.1111/j.1365-2486.2005.01000.x

Austin, M. P. (2002). Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. *Ecological Modelling*, *157*(2–3), 101–118. https://doi.org/10.1016/s0304-3800(02)00205-3

Bailey, J. J., Boyd, D. S., & Field, R. (2018). Models of upland species' distributions are improved by accounting for geodiversity. *Landscape Ecology*, *33*(12), 2071–2087. https://doi.org/10.1007/s10980-018-0723-z

Bakkestuen, V., Erikstad, L., & Halvorsen, R. (2008). Step-less models for regional environmental variation in Norway. *Journal of Biogeography*, *35*(10), 1906–1922. https://doi.org/10.1111/j.1365-2699.2008.01941.x

Barry, S., & Elith, J. (2006). Error and uncertainty in habitat models. *Journal of Applied Ecology*, *43*(3), 413–423. https://doi.org/10.1111/j.1365-2664.2006.01136.x

Bertrand, R., Lenoir, J., Piedallu, C., Riofrío-Dillon, G., de Ruffray, P., Vidal, C., ... Gégout, J.-C. (2011). Changes in plant community composition lag behind climate warming in lowland forests. *Nature*, *479*(7374), 517–520. https://doi.org/10.1038/nature10548

Bland, L., Keith, D., Miller, R., Murray, N., & Rodríguez, J. (2017). *Guidelines for the application of IUCN Red List of Ecosystems Categories and Criteria, version 1.1*. Gland, Switzerland: International Union for the Conservation of Nature. Retrieved from https://portals.iucn.org/library/sites/library/files/documents/2016-010-v1.1.pdf

Bryn, A., Kristoffersen, H.-P., Angeloff, M., Nystuen, I., Aune-Lundberg, L., Endresen, D., ... Rekdal, Y. (2015). Location of plant species in Norway gathered as a part of a survey vegetation mapping programme. *Data in Brief*, *5*, 589–594. https://doi.org/10.1016/j.dib.2015.10.014

Bryn, A., & Ullerud, H. A. (2018). *Veileder for kartlegging av terrestrisk naturvariasjon etter NiN (2.2.0) i målestokk 1: 5.000 og 1: 20.000. Kort versjon til bruk i felt*. Trondheim, Norway: Artsdatabanken.

Bryn, A., Strand, G.-H., Angeloff, M., & Rekdal, Y. (2018). Land cover in Norway based on an area frame survey of vegetation types. *Norsk Geografisk Tidsskrift – Norwegian Journal of Geography*, *72*(3), 131–145. https://doi.org/10.1080/00291951.2018.1468356

Coelho, M. T. P., Diniz-Filho, J. A., & Rangel, T. F. (2019). A parsimonious view of the parsimony principle in ecology and evolution. *Ecography*, *42*(5), 968–976. https://doi.org/10.1111/ecog.04228

Delcourt, H. R., Delcourt, P. A., & Webb, T. (1982). Dynamic plant ecology: The spectrum of vegetational change in space and time. *Quaternary Science Reviews*, *1*(3), 153–175. https://doi.org/10.1016/0277-3791(82)90008-7

Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Agard, J., Arneth, A., ... Zayas, C. N. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science*, *366*(6471), eaax3100. https://doi.org/10.1126/science.aax3100

Ejrnæs, R. (2000). Can we trust gradients extracted by detrended correspondence analysis? *Journal of Vegetation Science*, *11*(4), 565–572. https://doi.org/10.2307/3246586

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Ellis, E. C., Goldewijk, K. K., Siebert, S., Lightman, D., & Ramankutty, N. (2010). Anthropogenic transformation of the biomes, 1700 to 2000. *Global Ecology and Biogeography*, 19(5), 589–606. https://doi.org/10.1111/j.1466-8238.2010.00540.x

Erikstad, L., Bakkestuen, V., Hanssen, F., Evju, M., Statbbetorp, O. E., & Aarrestad, P. A. (2009). *Evaluation of nationwide satellite-based vegetation maps* (77pp.). NINA report 448. Retrieved from https://www.nina.no/archive/nina/PppBasePdf/rapport/2009/448.pdf

Erikstad, L., Uttakleiv, L. A., & Halvorsen, R. (2015). Characterisation and mapping of landscape types, a case study from Norway. *Belgian Journal of Geography*, 3, 1–15. https://doi.org/10.4000/belgeo.17412

Erikstad, L., Halvorsen, R., & Simensen, T. (2019). *Natur i Norge (NiN) versjon 2.2.* Inndelingen i landskapstyper. Trondheim, Norway: Artsdatabanken. Retrieved from https://www.artsdatabanken.no/nin/landskap

Ferrier, S., & Guisan, A. (2006). Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, 43(3), 393–404. https://doi.org/10.1111/j.1365-2664.2006.01149.x

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24, 38–49. https://doi.org/10.1017/S0376892997000088

Fourcade, Y., Besnard, A. G., & Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, 27(2), 245–256. https://doi.org/10.1111/geb.12684

Franklin, J. (2013). Mapping vegetation from landscape to regional scales. In E. van der Maarel, & J. Franklin (Eds.), *Vegetation ecology* (pp. 486–508). Oxford, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118452592.ch16

Gardner, R. H., Milne, B. T., Turner, M. G., & O'Neill, R. V. (1987). Neutral models for the analysis of broad-scale landscape pattern. *Landscape Ecology*, 1(1), 19–28. https://doi.org/10.1007/bf02275262

Gjessing, J. (1978). *Norges landformer.* Oslo, Norway: Universitetsforlaget.

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models: With applications in R.* London, UK: Cambridge University Press. https://doi.org/10.1017/9781139028271

Halvorsen, R. (2012). A gradient analytic perspective on distribution modelling. *Sommerfeltia*, 35(1), 1–165. https://doi.org/10.2478/v10208-011-0015-3

Halvorsen, R., Bryn, A., & Erikstad, L. 2019 (2016). *NiN systemkjerne – teori, prinsipper og inndelingskriterier.* Versjon 2.2. – Natur i Norge (NiN) Systemdokumentasjon, 1 (Version 2.2.0), 1–291. Trondheim, Norway: Artsdatabanken. Retrieved from https://www.artsdatabanken.no/Files/29717/Artikkel_1___NiNs_systemkjerne___teori,_prinsipper_og_inndelingskriterier.pdf

Hemsing, L. Ø., & Bryn, A. (2012). Three methods for modelling potential natural vegetation (PNV) compared: A methodological case study from south-central Norway. *Norsk Geografisk Tidsskrift – Norwegian Journal of Geography*, 66(1), 11–29. https://doi.org/10.1080/00291951.2011.644321

Henderson, E. B., Ohmann, J. L., Gregory, M. J., Roberts, H. M., & Zald, H. (2014). Species distribution modelling for plant communities: Stacked single species or multivariate modelling approaches? *Applied Vegetation Science*, 17(3), 516–527. https://doi.org/10.1111/avsc.12085

Hill, M. O., & Gauch, H. G. (1980). Detrended correspondence analysis: An improved ordination technique. *Vegetatio*, 42(1), 47–58. https://doi.org/10.1007/BF00048870

Horvath, P., Halvorsen, R., Stordal, F., Tallaksen, L. M., Tang, H., & Bryn, A. (2019). Distribution modelling of vegetation types based on area frame survey data. *Applied Vegetation Science*, 22(4), 547–560. https://doi.org/10.1111/avsc.12451

IUCN. (2016). *An introduction to the IUCN red list of ecosystems: The categories and criteria for assessing risks to ecosystems* (14 pp.). Gland, Switzerland: IUCN. Retrieved from https://doi.org/10.2305/IUCN.CH.2016.RLE.2.en

Janská, V., Jiménez-Alfaro, B., Chytrý, M., Divíšek, J., Anenkhonov, O., Korolyuk, A., ... Culek, M. (2017). Palaeodistribution modelling of European vegetation types at the Last Glacial Maximum using modern analogues from Siberia: Prospects and limitations. *Quaternary Science Reviews*, 159, 103–115. https://doi.org/10.1016/j.quascirev.2017.01.011

Jiménez-Alfaro, B., Suárez-Seoane, S., Chytrý, M., Hennekens, S. M., Willner, W., Hájek, M., ... Tsiripidis, I. (2018). Modelling the distribution and compositional variation of plant communities at the continental scale. *Diversity and Distributions*, 24(7), 978–990. https://doi.org/10.1111/ddi.12736

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(3), 214–226. https://doi.org/10.1111/conl.12167

Liu, C., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28(3), 385–393. https://doi.org/10.1111/j.0906-7590.2005.03957.x

Lobo, J. M., Jiménez-Valverde, A., & Hortal, J. (2010). The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, 33(1), 103–114. https://doi.org/10.1111/j.1600-0587.2009.06039.x

Longcore, T., Noujdina, N., & Dixon, P. J. (2019). Landscape modeling of the potential natural vegetation of Santa Catalina Island, California. *Western North American Naturalist*, 78, 617–632. https://doi.org/10.3398/064.078.0406

Luck, M., & Wu, J. (2002). A gradient analysis of urban landscape pattern: A case study from the Phoenix metropolitan region, Arizona, USA. *Landscape Ecology*, 17(4), 327–339. https://doi.org/10.1023/a:1020512723753

Maiorano, L., Cheddadi, R., Zimmermann, N. E., Pellissier, L., Petitpierre, B., Pottier, J., ... Guisan, A. (2013). Building the niche through time: Using 13,000 years of data to predict the effects of climate change on three tree species in Europe. *Global Ecology and Biogeography*, 22(3), 302–317. https://doi.org/10.1111/j.1466-8238.2012.00767.x

Margules, C. R., Nicholls, A. O., & Austin, M. P. (1987). Diversity of Eucalyptus species predicted by a multi-variable environmental gradient. *Oecologia*, 71(2), 229–232. https://doi.org/10.1007/BF00377288

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed., Vol. 37). London, UK: Chapman and Hall.

Minchin, P. R. (1987). An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio*, 69(1), 89–107. https://doi.org/10.1007/BF00038690

Myers-Smith, I. H., Kerby, J. T., Phoenix, G. K., Bjerke, J. W., Epstein, H. E., Assmann, J. J., ... Wipf, S. (2020). Complexity revealed in the greening of the Arctic. *Nature Climate Change*, 10(2), 106–117. https://doi.org/10.1038/s41558-019-0688-1

Nad'o, L., & Kaňuch, P. (2018). Why sampling ratio matters: Logistic regression and studies of habitat use. *PLoS ONE*, 13(7), e0200742. https://doi.org/10.1371/journal.pone.0200742

NIBIO. (2018). *Vegetasjonskart*. Retrieved from https://nibio.no/tema/landskap/vegetasjonskart

NBIC. (2018). *Red list for ecosystems and habitat types*. Norwegian Biodiversity Information Centre. Retrieved from https://www.artsdatabanken.no/rodlistefornaturtyper

Norwegian Environment Agency. (2019). *Naturbase*. Retrieved from https://www.miljodirektoratet.no/verktoy/naturbase/

Noss, R. F. (1990). Indicators for monitoring biodiversity: A hierarchical approach. *Conservation Biology*, 4(4), 355–364. https://doi.org/10.1111/j.1523-1739.1990.tb00309.x

Ovaskainen, O., & Soininen, J. (2011). Making more out of sparse data: Hierarchical modeling of species communities. *Ecology*, 92(2), 289–295. https://doi.org/10.1890/10-1251.1

Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C. F., ... Guisan, A. (2013). The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecology and Biogeography*, *22*(1), 52–63. https://doi.org/10.1111/j.1466-8238.2012.00790.x

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramberg, I. B., I. Bryhni, A. Nøttvedt, & K. Rangnes (Eds.) (2008). *The making of a land – Geology of Norway*. Trondheim: Geological Society of Norway.

Rekdal, Y., & Angeloff, M. (2013). *Utmarksbeitet i fjellområdet mellom Valdres og Gausdal*. Report 7. Ås, Norway: Norwegian forest and landscape Institute.

Santos, J. P., Sobral-Souza, T., Brown, K. S., Vancine, M. H., Ribeiro, M. C., & Freitas, A. V. L. (2020). Effects of landscape modification on species richness patterns of fruit-feeding butterflies in Brazilian Atlantic Forest. *Diversity and Distributions*, *26*(2), 196–208. https://doi.org/10.1111/ddi.13007

Strand, G.-H. (2013). The Norwegian area frame survey of land cover and outfield land resources. *Norsk Geografisk Tidsskrift – Norwegian Journal of Geography*, *67*(1), 24–35. https://doi.org/10.1080/00291951.2012.760001

Tansley, A. G. (1935). The use and abuse of vegetational concepts and terms. *Ecology*, *16*(3), 284–307. https://doi.org/10.2307/1930070

ter Braak, C. J. (1985). Correspondence analysis of incidence and abundance data: Properties in terms of a unimodal response model. *Biometrics*, *41*(4), 859–873. https://doi.org/10.2307/2530959

Turner, M. G., & Gardner, R. H. (2015). Introduction to landscape ecology and scale. In *Landscape ecology in theory and practice: Pattern and process* (pp. 1–32). New York, NY: Springer. https://doi.org/10.1007/978-1-4939-2794-4

Ullerud, H. A., Bryn, A., & Klanderud, K. (2016). Distribution modelling of vegetation types in the boreal–alpine ecotone. *Applied Vegetation Science*, *19*(3), 528–540. https://doi.org/10.1111/avsc.12236

Vollering, J., Halvorsen, R., & Mazzoni, S. (2019). The MIAmaxent R package: Variable transformation and model selection for species distribution models. *Ecology and Evolution*, *9*(21), 12051–12068. https://doi.org/10.1002/ece3.5654

Whittaker, R. H. (1956). Vegetation of the Great Smoky Mountains. *Ecological Monographs*, *26*(1), 1–80. https://doi.org/10.2307/1943577

Whittaker, R. J., Araújo, M. B., Jepson, P., Ladle, R. J., Watson, J. E. M., & Willis, K. J. (2005). Conservation biogeography: Assessment and prospect. *Diversity and Distributions*, *11*(1), 3–23. https://doi.org/10.1111/j.1366-9516.2005.00143.x

Zarnetske, P. L., Read, Q. D., Record, S., Gaddis, K. D., Pau, S., Hobi, M. L., ... Finley, A. O. (2019). Towards connecting biodiversity and geodiversity across scales with satellite remote sensing. *Global Ecology and Biogeography*, *28*(5), 548–556. https://doi.org/10.1111/geb.12887

BIOSKETCHES

Trond Simensen's research is focused on landscape ecology, distribution modelling and the study of patterns of diversity in Earth's biotic and abiotic processes and features.

The members of the **Geo-Ecology research group** at the **Natural History Museum, University of Oslo** are interested in describing patterns of variation in nature, understanding the processes that generate them and facilitating practical use of ecological knowledge to the wider benefit of society.

Author contributions: RH, LE and TS conceived the research idea; TS and RH with contributions from AB, JV and PH developed the methodology; TS, JV and PH curated the data and performed the modelling and the statistical analyses; AB participated in all parts of the AR18×18 data set, including field work; TS, AB and RH wrote the draft manuscript which was critically reviewed, revised and finally approved by all authors.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.