Computers & Geosciences 35 (2009) 615–625

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Statistical consensus methods for improving predictive geomorphology maps

Mathieu Marmion[a,b,*], Jan Hjort[c], Wilfried Thuiller[d], Miska Luoto[a,b]

[a] *Thule Institute, University of Oulu, P.O. Box 7300, 90014 Oulu, Finland*
[b] *Department of Geography, University of Oulu, P.O. Box 3000, 90014 Oulu, Finland*
[c] *Department of Geography, University of Helsinki, P.O. Box 64, 00014 Helsinki, Finland*
[d] *Laboratoire d'Ecologie Alpine, UMR CNRS 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France*

ARTICLE INFO

Article history:
Received 25 September 2007
Received in revised form
28 January 2008
Accepted 9 February 2008

Keywords:
AUC
Uncertainty
Landforms
Predictive modelling

ABSTRACT

A variety of predictive models is currently used to map the spatial distribution of earth surface processes and landforms. In this study, we tested statistical consensus methods in order to improve the predictive accuracy of geomorphological models. The distributions of 12 geomorphological formations were recorded at a resolution of 25 ha in a sub-arctic landscape in northern Finland. Nine environmental variables were used to predict probabilities of occurrence of the formations using eight state-of-the-art modelling techniques. The probability values of the models were combined using four different consensus methods. The accuracy of the models was calculated using spatially independent test data by the area under the curve (AUC) of a receiver-operating characteristic (ROC) plot. The mean AUC values of the geomorphological models varied between 0.711 and 0.755 based on single-model techniques, whereas the corresponding values based on consensus methods ranged from 0.752 to 0.782. The weighted average consensus method had the highest predictive performance of all methods. It improved the accuracy of 11 predictions out of 12. The results of this study suggest that the consensus methods have clear advantages over single-model predictions. The simplicity of the consensus methods makes it straightforward to implement them in predictive modelling studies in geomorphology.

1. Introduction

Knowledge of the spatial distributions of earth surface processes, landforms and the underlying environmental factors affecting them has an important role in geomorphological research (Allen, 1997). However, data on the distribution of different processes and landforms are often scarce and can be difficult to acquire. One potential means to complement the insufficient information concerning the distribution of geomorphological phenomena and suitable physical environments for them is provided by predictive geomorphological modelling (Vitek et al., 1996; Luoto and Hjort, 2005).

Recently, spatial modelling has become one of the key issues in geomorphology, e.g. in assessing the stability of steep terrain (Dai and Lee, 2002; Guzzetti et al., 2006), mapping of the glaciated landscapes (Brown et al., 1998), mapping of soil and bedrock properties (Kheir et al., 2008) and periglacial processes (Mackay et al., 1992; Graff and Usery, 1993; Luoto and Seppälä, 2002; Hjort and Luoto, 2006). Previous studies have shown that modern spatial modelling techniques can provide useful forecasts of geomorphological phenomena in unsurveyed parts of landscapes (Luoto and Hjort, 2005), and can provide valuable contributions to theoretical (Walsh et al., 1998)

* Corresponding author at: Thule Institute, University of Oulu, P.O. Box 7300, 90014 Oulu, Finland. Tel.: +358 85531714; fax: +358 8 553 1693.
E-mail address: mathieu.marmion@oulu.fi (M. Marmion).

0098-3004/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2008.02.024

and applied research (Haeberli, 1992; Harris et al., 2001a b; Gude and Barsch, 2005).

Development of spatial modelling in geomorphology is based on three trends: growth in the availability of remotely sensed (RS) data and development of GIS techniques integrated with novel statistical methods (Walsh et al., 1998). In a methodological study, Luoto and Hjort (2005) compared different modelling techniques in predictive geomorphological mapping. Most importantly, although predictive models perform relatively accurately, they do not always provide robust spatial predictions. Such variability in modelling results is not surprising given that spatial models are correlative and therefore sensitive to the data and the mathematical functions utilized to describe the distributions of geomorphological phenomena in relation to environmental parameters. Process-based models using theoretical and experimental knowledge provide an alternative that is less dependent on empirical relationships. However, their implementation at the landscape level is difficult because of the complex processes and interactions that must be represented; and variability in forecasts is also common (Araújo and New, 2006). To overcome the problem of variability in predictions, the use of multiple models within a consensus modelling framework has been presented in various fields of research, e.g. in ecology (Huang and Lees, 2004; Thuiller, 2004; Araújo et al., 2005b; Huang and Lees, 2005), economy (Gregory et al., 2001), biomedicine (Nilsson et al., 2000), meteorology (Sanders, 1963), climatology (Benestad, 2004) and hydrology (Goswami and O'Connor, 2007).

In this study, eight state-of-the-art modelling techniques were utilized to predict the distribution of 12 geomorphological landform types in sub-arctic Finland. Next, the predictive performances of four consensus methods combining the model outputs (probability values) of eight modelling techniques were evaluated. We put special emphasis on model testing, and therefore we assessed the accuracy of the predictive models with spatially independent evaluation data (Fig. 1). The use of spatially independent data are of particular value since alternative approaches, including re-substitution and one-time data splitting, have been shown to lead to over-optimistic estimates of the model predictive capabilities in new areas and biased signals of the importance of different predictors (Fielding and Haworth, 1995; Peterson and Vieglais, 2001; Araújo et al., 2005a; Randin et al., 2006).

2. Study area

The study area is located in sub-arctic Finland (Fig. 1). The topography of the area is characterized by eroded fells with elevations ranging from ca. 200 to 640 m above sea level (a.s.l.). Geologically, the area belongs to

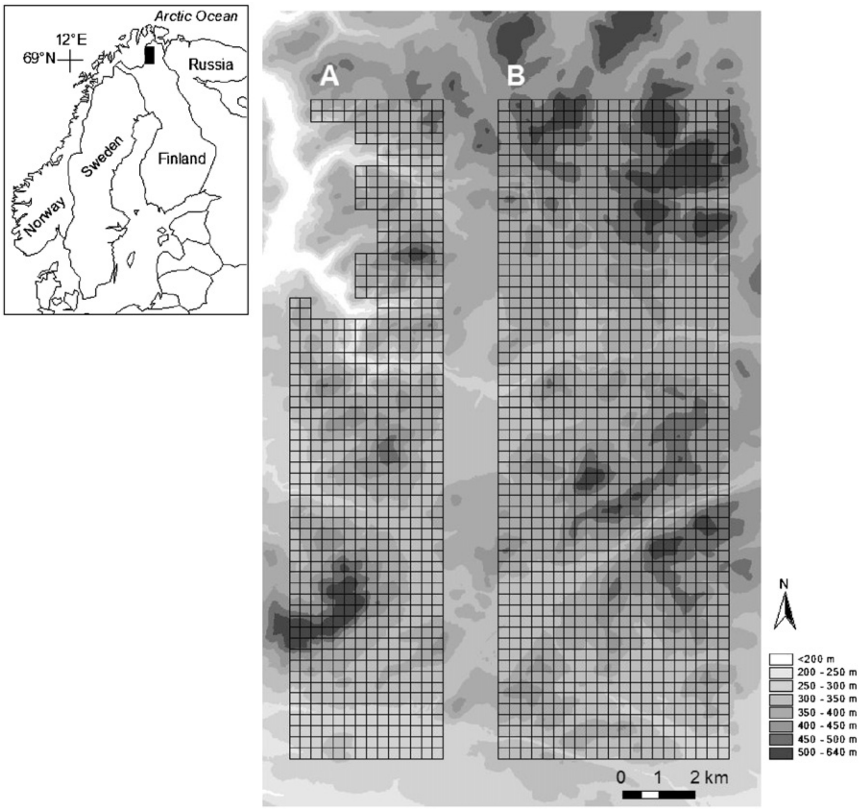

Fig. 1. Location of study area (black box) in northern sub-arctic Finland. Two spatially independent data sets "calibration" and "evaluation" are indicated by B and A. Total of 2032 grid squares are also shown on map.

a Pre-cambrian granulite complex about 1.9 billion years old (Meriläinen, 1976). Surface deposits consist of glacigenic till, peat, as well as sand and gravel deposits. The area lies within the zone of discontinuous permafrost (King and Seppälä, 1987). The mean annual air temperature was $-2.0\,^{\circ}$C and mean annual precipitation ca. 400 mm during the period 1962–1990 (Climatological Statistics in Finland 1961–1990, 1991). Botanically, the region lies to the north of the northern limit of the continuous Scots pine (*Pinus sylvestris* L.) forest in the Orohemiarctic Zone with mountain birch (*Betula pubescens* ssp. *czerepanovii*) as the prevailing tree species (Ahti et al., 1968). Mires belong to the palsa and subalpine mire types (Luoto and Seppälä, 2002). A more detailed description of the study region can be obtained from Hjort (2006).

3. Material and methods

3.1. Material

The 12 geomorphological landforms utilized in this study were peaty permafrost mounds (palsas), frost-formed fine-scale hummocks (convex non-sorted circles, earth hummocks and peat pounus), sorted patterned ground features (stone pits, sorted nets and sorted stripes), solifluction landforms (non-sorted solifluction terraces, sorted solifluction sheets and streams) as well as wind deflation sites (Hjort and Luoto, 2006). Despite the fact that landforms can be grouped, the features were treated as distinct types because different processes govern their formation (Washburn, 1979; French, 1996).

The landforms were mapped and converted to grid-based modelling data in a four-step process (Hjort, 2006):

(1) A detailed stereoscopic interpretation of black-and-white aerial photographs (1:31,000 scale) was performed to identify landforms from the study area before fieldwork.
(2) The features were mapped utilizing pre-mapping results in the field during the summers of 2002 and 2003. The positions of the landforms were located with a Global Positioning System (GPS) device (Garmin eTrex personal navigator; spatial accuracy ca. 10 m).
(3) The field-mapping results were digitized in a vector format on ortho-rectified aerial photographs utilizing GIS software.
(4) A binary variable (1 = present, 0 = absent), indicating the occurrence of the landform, was allocated to each modelling square using the geomorphological database.

The explanatory data utilized in the modelling were collected from three different information sources, namely a digital elevation model (DEM; Fig. 1), biotope database and digital soil map (Hjort and Luoto, 2006). Three topographical parameters, three soil-type variables and three vegetation variables were compiled using Arc/Info GRID at 500 m cell size resolution (25 ha; Table 1).

Table 1
Explanatory (environmental and spatial) variables utilized in statistical analyses and their description Moore et al. (1991)

Variable	Description
Mean altitude (m)	Temperature, snow distribution, potential energy, vegetation zone
Mean slope angle (deg)	Potential energy, water flow, snow distribution, radiation, soil thickness
Mean wetness index	Potential soil moisture, cold air distribution during inversion, silt content
Peat cover (100 m^2)	High water-holding capacity, nonconductor (dry peat)
Glacigenic deposit cover (100 m^2)	Frost susceptible, different size of soil particles
Sand and gravel cover (100 m^2)	Frost-resistant, dry, sparse ground layer vegetation
Cover of schrub (%) (a)	Snow thickness and distribution, soil moisture and temperature
Cover of canopy (%) (b)	Snow thickness and distribution, air temperature
Cover of alpine vegetation (%)	100−(a+b)

This rather coarse resolution was chosen based on the accuracy assessment of the used GIS data (Hjort and Luoto, 2006) and in an attempt to minimize the potential risks of spatial autocorrelation in statistical analyses (e.g. McCullagh and Nelder, 1989).

3.2. Evaluation of the models

The accuracies of the models and consensus methods (described in Sections 3.3 and 3.4) were calculated using spatially independent test data by the area under the curve (AUC) of a receiver-operating characteristic (ROC) plot (Fig. 1). The range of AUC values is from 0.0 to 1.0. A model providing excellent prediction has an AUC higher than 0.9, a fair model has an AUC between 0.7 and 0.9, and a model is considered as poor if it has an AUC lower than 0.7 (Swets, 1988). Based on AUC values, a "rank average" index indicates the average of the ranks of the modelling technique computed for each geomorphologic landform. In this study, eight modelling techniques and four consensus methods were tested. The rank values vary between 1 and 12, 12 indicating the highest model performance. A Wilcoxon signed ranks test was used to compare the statistical difference between the models.

3.3. Methods

All implemented modelling techniques were run in R environment[1] under the BIOMOD framework (Thuiller, 2003). These techniques can be assigned to three main categories: (1) regressive algorithms [generalized linear models (GLMs), generalized additive models (GAMs), multiple adaptive regression splines (MARS)], (2) classification techniques [classification tree analysis (CTA) and mixture discriminant analysis (MDA)] and (3) machine-learning

[1] R Development Core Team, 2004. R: a language and environment for statistical computing. Vienna, Austria. http://www.R-project.org.

methods [generalized boosting methods (GBMs), artificial neural networks (ANNs) and random forest (RF)]. It is important to stress that all models were used with a predictive rather than inductive goal in this study. In such circumstances, accuracy of model predictions is more important than the significance of particular environmental variables. We did not further investigate autocorrelation aspects or the relative importance of different variables (Legendre, 1993).

GLMs are mathematical extensions of linear models (McCullagh and Nelder, 1989). Recently, GLMs appear to be increasingly popular as the statistical model to be used. This is due to the ability of GLMs to handle nonlinear relationships and different types of statistical distributions characterizing spatial data. We used an automatic stepwise procedure based on the Akaike information criterion (AIC) in model calibration. Examples of the use of GLMs in geomorphological studies can be found in Atkinson et al. (1998), Rowbotham and Dudycha (1998), Dai and Lee (2002), Luoto and Seppälä (2002) and Luoto and Hjort (2004).

GAMs are non-parametric extensions of GLMs. They provide a flexible data-driven class of models that permit both linear and complex additive response shapes, as well as the combination of the two within the same model (Hastie and Tibshirani, 1990). GAMs have been recently used in geomorphological studies (Hjort and Luoto, 2006; Brenning et al., 2007).

Multivariate adaptive regression splines (MARS) represent a relatively new technique that combines classical linear regression, mathematical construction of splines and binary recursive partitioning to produce a local model in which relationships between response and predictors are either linear or nonlinear (Friedman, 1991). An important feature of MARS is its sensitivity to outliers and to collinearity between the variables (Deichmann et al., 2002). Examples of the use of MARS in geomorphologic studies can be found in Luoto and Hjort (2005), in climatology in Corte-Real et al. (1995) and in geophysics in Deveaux et al. (1993).

CTA is an alternative to regression techniques, and uses a tree structure (Breiman et al., 1984). It is a rule-based method defined by binary decision splits about the values of predictors (Venables and Ripley, 2002). CTA is used rather frequently in geomorphological and environmental studies (Franklin, 2002; Luoto and Hjort, 2005).

Discriminant analysis is used in statistics to identify the linear combination of features which best separate two or more classes of object. MDA is an extension of the well-known linear discriminant analysis (LDA) (Venables and Ripley, 2002), in which classes are modelled as a mixtures of subclasses, with each subclass represented by a Gaussian distribution. An example of the use of MDA in geomorphology was presented by Merritt and Wohl (2003).

GBM is a sequential method based on binary trees (Ridgeway, 1999). GBM is considered as a machine-learning method using adaptive weighting of multiple outputs of numerous classification algorithms. The boosted classifier's prediction is based on an accuracy weighted vote across the estimated classifiers (Ridgeway, 1999). Boosting methods are novel statistical techniques, which have been used recently in ecological modelling (Elith et al., 2006). However, to the best of our knowledge GBM has not previously been used in geomorphological research.

ANN are powerful rule-based modelling techniques, which are frequently used in spatial modelling. ANN provide an alternative way to generalize linear regression functions (Venables and Ripley, 2002). Neural networks have received considerable attention as a means of building accurate models for prediction when the functional form of the underlying equations is unknown (Lek and Guegan, 1999). Luoto and Hjort (2005) evaluated the reliability of ANN for geomorphologic mapping, and Guzzetti et al. (1999) used this method to evaluate slope stability. ANN is also used more widely in geomorphologic topics such as the modelling of suspended sediment flux in rivers (Zhu et al., 2007).

RF is a classifier belonging to the machine-learning category based on the multiple trees method (Breiman, 2001). RF generates hundreds or thousands of trees forming a "forest". Each tree is grown by selecting randomly a training dataset, with replacement from the original dataset. In addition, the number of implemented explanatory variables in each tree varies randomly. To our knowledge, RF has never been used in geomorphology.

3.4. The consensus methods

As early as 1878, J. Willard Gibbs introduced into statistical mechanics the notion of ensemble. It is an idealization consisting of a large number of copies (i.e. predictions) of a system, considered all at once, each of which represents a possible state that the real system might be in at some specified time (Araújo and New, 2006). The emphasis in predictive modelling is on combining the individual predictions in order to obtain an enhanced one. Recently, Araújo et al. (2005b) and Thuiller (2004) proposed that consensus methods are more accurate than single-model predictions. This study presents four consensus methods: a non-selective median technique, two selective median techniques and a selective weighted average technique (Araújo et al., 2005b; Goswami and O'Connor, 2007).

Median(*all*) consensus method is the median value of the outputs of all eight modelling techniques. Median(all) has been used in an ecological context by Araújo et al. (2005b).

Median(*PCA*) is run after a selective processing for the eight single-model techniques based on a principal component analysis (PCA). The eight single-model techniques still project as previously, but an inner evaluation is performed to select them. The original data set of calibration (1316 grid squares; Fig. 1B) is divided into two inner data subsets using a 70%/30% random split. The two created subsets are called "inner calibration" (921 grid squares) and "inner evaluation" (395 grid squares). These both inner subsets are used to preselect four single-models among the eight disposal ones.

Median(PCA) consists of calculating for each geomorphologic formation the median value of only four modelling techniques, which were selected via a criterion based on a PCA. A PCA provides for each modelling technique a rate reflecting its ability to follow the general trend of projection of the eight modelling techniques (see Thuiller, 2004). An overlapped scatterplot of the eight individual predictions is prepared, referring to the environmental variables. The first principal component (PC1) reflects the general trend followed by the eight single-models, for each threatened plant species. The four models whose multi-variable scatterplot predictions along PC1 were the greatest were selected and the median value of these four single-models was computed. The models that were not following the general trend were then discriminated and not taken into account. This method was used in the past by Thuiller (2004), Araújo et al. (2005b) and Thuiller et al. (2005) in a biogeographical context.

Median(*AUC*) consensus method, like Median(PCA), is based on an inner validation process, however, the next step of selection criterion is rather different. Four modelling techniques are selected from among the eight ones on the basis of their inner validation AUC criterion (i.e. the single-models were calibrated and evaluated using the inner calibration and inner evaluation data subsets). The eight modelling techniques are ranked and the four best ones are selected. Then the median value of these four modelling techniques is calculated. An attempt of this nature was made by Hartley et al. (2006) in a biogeographic context.

The *weighted average* (WA) consensus method utilizes the predictive performance of the modelling techniques. Firstly, the four modelling techniques with highest accuracy are selected. Secondly, a WA is calculated based on inner evaluation AUC values of the selected modelling techniques, as described by following:

$$\mathrm{WA}_i = \frac{\mathrm{AUC}_{m1} \times m1_i + \mathrm{AUC}_{m2} \times m2_i + \mathrm{AUC}_{m3} \times m3_i + \mathrm{AUC}_{m4} \times m4_i}{\mathrm{AUC}_{m1} + \mathrm{AUC}_{m2} + \mathrm{AUC}_{m3} + \mathrm{AUC}_{m4}} \quad (1)$$

where $m1_i$, $m2_i$, $m3_i$ and $m4_i$ are the probability values of the *i*th geomorphological landform to be present in a given grid cell for the four selected single-models. This consensus method has been used in hydrology by Goswami and O'Connor (2007).

4. Results

4.1. Predictive accuracy of the eight modelling techniques

The predictive accuracies of the eight modelling techniques are presented in Tables 2 and 3. At maximum, two geomorphological formations out of 12 were excellently classified by the single-models. The mean AUC values of the modelling techniques varied from 0.711 for MDA and CTA, to 0.755 for GAM, with an average of 0.737. Palsas were projected with the highest accuracy, the AUC values varying from 0.813 (CTA) to 0.928 (GBM) with a mean value of 0.880. Figs. 2 and 3 represent the predictive distribution maps of earth hummocks and stone pits based on GAM (A) and ANN (B).

4.2. Predictive accuracy of the consensus methods

On the basis of the results presented in Table 2, the Median(All) consensus method did not perform better than the single-model techniques. The AUC range varied from 0.590 to 0.905 with a mean of 0.752. The "rank average" index of Table 3 reflects this behavior. For a given projection technique (single-model techniques and consensus methods) this number is the average of the projection modelling technique's rank for each geomorphologic formation. The rank average of Median(All) is 7.6, which is lower than that of GAM (7.7). GAM provides equal or even better projection than Median(All). The mean AUC value of GAM (0.755) and Median(All) (0.752) for the 12 geomorphological formations underlines the

Table 2
AUC values of predictions based on evaluation data of eight modelling techniques and four consensus methods (Median(All), Median(PCA), Median(AUC) and WA)

	ANN	CTA	GAM	GBM	GLM	MARS	MDA	RF	Median (All)	Median (PCA)	Median (AUC)	WA
Palsa	0.924	0.813	*0.866*	***0.928***	*0.871*	0.871	0.846	*0.924*	0.905	*0.923*	0.892	0.916
CNSC	*0.660*	0.506	*0.607*	0.592	*0.607*	*0.562*	0.605	0.637	0.608	*0.641*	0.624	**0.713**
SEC	*0.635*	0.625	*0.659*	*0.635*	*0.656*	0.627	0.607	0.612	0.644	*0.652*	0.655	**0.669**
Earth hummock	0.878	*0.868*	0.886	*0.888*	0.883	*0.877*	0.826	*0.880*	0.897	*0.886*	0.896	**0.906**
Peat pounu	0.834	*0.852*	*0.879*	*0.878*	*0.864*	0.867	0.816	0.852	0.878	*0.880*	0.887	**0.891**
Stone pit	0.693	0.718	0.739	*0.740*	0.738	*0.696*	*0.711*	*0.707*	0.761	*0.742*	0.758	**0.774**
Sorted net	0.629	*0.519*	*0.626*	*0.573*	*0.614*	0.581	0.568	0.553	0.605	*0.606*	0.615	**0.641**
Sorted stripe	*0.832*	0.846	*0.849*	0.851	*0.849*	*0.832*	0.842	0.839	0.845	*0.843*	0.842	**0.876**
NSS terrace	0.600	**0.618**	0.599	0.601	*0.570*	*0.602*	*0.568*	*0.577*	0.605	*0.586*	0.592	0.607
SS sheet	*0.903*	0.789	*0.859*	*0.890*	*0.860*	0.859	0.814	0.886	0.859	*0.893*	0.867	**0.909**
SS stream	0.859	0.807	*0.881*	*0.830*	*0.888*	*0.827*	0.823	0.829	0.832	*0.886*	0.886	**0.893**
Deflation site	0.562	0.567	***0.614***	*0.581*	*0.592*	*0.578*	0.505	0.567	0.590	*0.590*	0.590	0.594
Mean	0.751	0.711	0.755	0.749	0.749	0.732	0.711	0.739	0.752	0.761	0.759	**0.782**

Italic values reflect modelling techniques used for Median(PCA). Underlined values reflect those selected for Median(AUC) and WA. Bold values indicate best modelling technique/consensus method for a given geomorphologic case. "Mean" is mean value of each modelling/consensus technique over all 12 geomorphologic formations (CNSC = convex non-sorted circle; SEC = stony earth circle; NSS terrace = non-sorted solifluction terrace; SS sheet = sorted solifluction sheet; SS stream = sorted solifluction stream).

Table 3
Number of models in different accuracy classes of AUC based on eight modelling techniques and four consensus methods

	ANN	CTA	GAM	GBM	GLM	MARS	MDA	RF	Median(All)	Median(PCA)	Median(AUC)	WA
Excellent	2	0	0	1	0	0	0	1	1	1	0	3
Fair	4	7	7	6	7	6	7	6	6	6	7	5
Poor	6	5	5	5	5	6	5	5	5	5	5	4
Rank average	6.1	3.5	7.7	7.3	7.0	4.1	2.2	4.7	7.6	8.2	8.1	11.5

"Rank average" indicates average of ranks of modelling techniques computed for each geomorphologic case (12 indicates best model). Excellent = AUC > 0.9, fair = 0.7 < AUC < 0.9 and poor = AUC < 0.7.

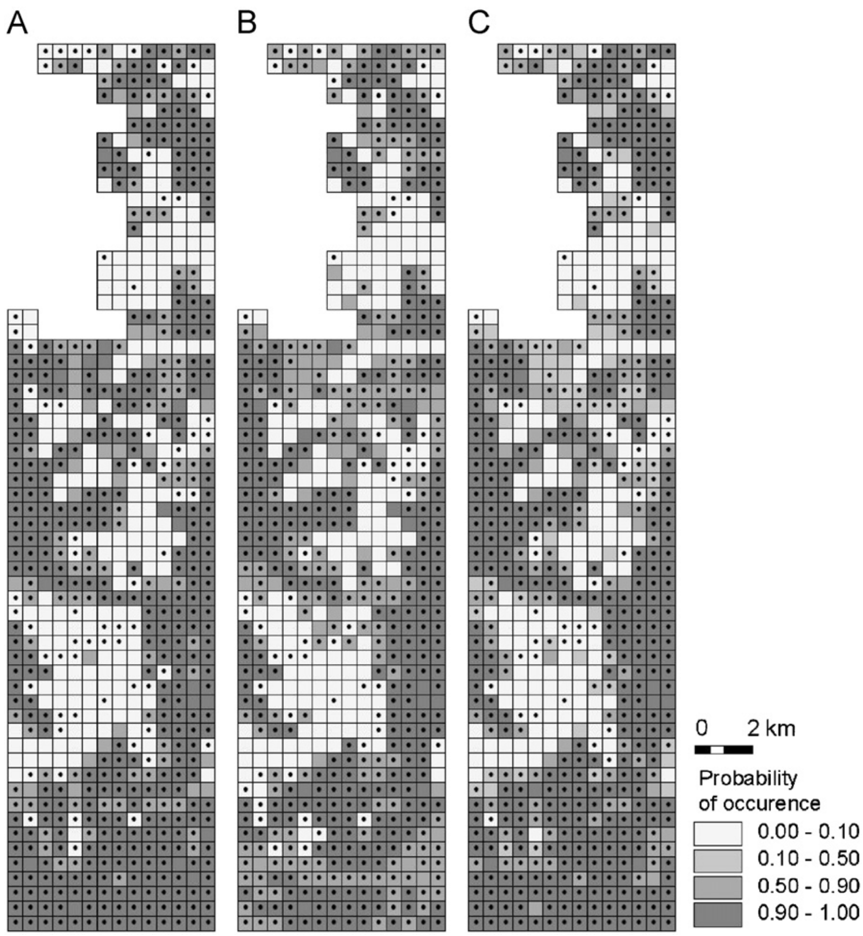

Fig. 2. Estimated spatial distribution of earth hummocks in sub-arctic Finland by GAM (A; AUC = 0.886), ANN (B; AUC = 0.878) and WA (C; AUC = 0.906). Gray levels represent different levels of probability of presence. Recorded presences (true) are shown with black dots.

equivalence of the results. A statistical analysis based on the Wilcoxon signed ranks test confirms these similar performances (p-value of 0.689).

The Median(PCA) and Median(AUC) methods with mean AUC values of 0.761 and 0.759 had similar predictive abilities. The results of the inner validation processes are summarized in Table 4. The selected single-model techniques were different for these two consensus methods. On average, two or three selected modelling techniques were common. Median(PCA) had an AUC range from 0.590 to 0.923 with a mean of 0.761. For Median(AUC), the corresponding three values were 0.590, 0.896, and 0.759. In Table 3, the rank average of these two consensus methods reflects the similarity of the consensus methods [8.1 for Median(AUC) and 8.2 for Median(PCA)]. However, considering all eight single-model techniques, none of these consensus methods provided the best projection for all geomorphological

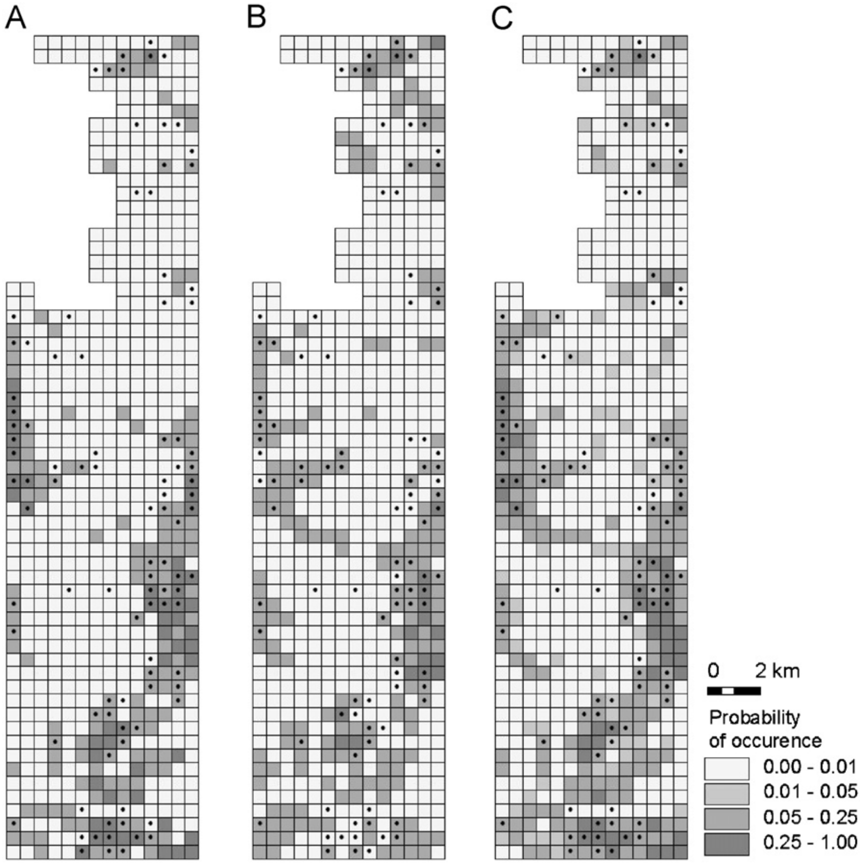

Fig. 3. Estimated spatial distribution of stone pits in sub-arctic Finland by GAM (A; AUC = 0.739), ANN (B; AUC = 0.693) and WA (C; AUC = 0.774). Gray levels represent different levels of probability of presence. Recorded presences (true) are shown with black dots.

Table 4
AUC values of eight modelling techniques computed during inner validation process

	ANN	CTA	GAM	GBM	GLM	MARS	MDA	RF
Palsa	0.861	0.851	***0.984***	*0.978*	*0.981*	0.851	0.978	*0.972*
CNSC	*0.869*	0.793	*0.866*	0.877	*0.874*	*0.869*	0.825	**0.884**
SEC	*0.636*	0.606	***0.706***	*0.673*	*0.702*	0.682	0.643	0.673
Earth hummock	0.903	*0.868*	0.908	*0.913*	0.909	*0.911*	0.879	***0.915***
Peat pounu	0.866	*0.839*	*0.928*	***0.931***	*0.924*	0.927	0.918	0.928
Stone pit	0.750	0.536	0.799	***0.817***	0.795	*0.728*	*0.775*	*0.811*
Sorted net	0.764	*0.721*	*0.834*	*0.829*	***0.851***	0.710	0.742	0.799
Sorted stripe	*0.909*	0.842	***0.951***	0.924	*0.948*	*0.888*	0.887	0.900
NSS terrace	0.628	0.635	0.677	**0.704**	*0.660*	*0.654*	*0.659*	*0.677*
SS sheet	*0.911*	0.874	***0.925***	*0.913*	*0.924*	0.912	0.903	0.908
SS stream	0.839	0.808	***0.924***	*0.909*	*0.923*	*0.910*	0.880	0.903
Deflation site	0.788	0.740	*0.806*	***0.828***	*0.813*	*0.820*	0.791	0.805

Italic values are those selected by PCA and underlined ones are four best ones considering AUC criteria. Bold values indicate best models for a given geomorphologic formation (CNSC = convex non-sorted circle; SEC = stony earth circle; NSS terrace = non-sorted solifluction terrace; SS sheet = sorted solifluction sheet; SS stream = sorted solifluction stream).

formations. Both methods were statistically compared with GAM and significant differences were not observed [p-values of 0.790 and 0.346 for Median(PCA) and Median(AUC), respectively].

The WA consensus method was the only consensus method, which performed statistically better (p-value 0.011) than the most accurate single-model technique (GAM). Fig. 4 illustrates the differences between the predictive distribution maps based on GAM and WA for four geomorphological landforms. For 11 geomorphological formations out of 12 WA provided higher accuracy than GAM (Tables 2 and 3). AUC values of WA varied from 0.594

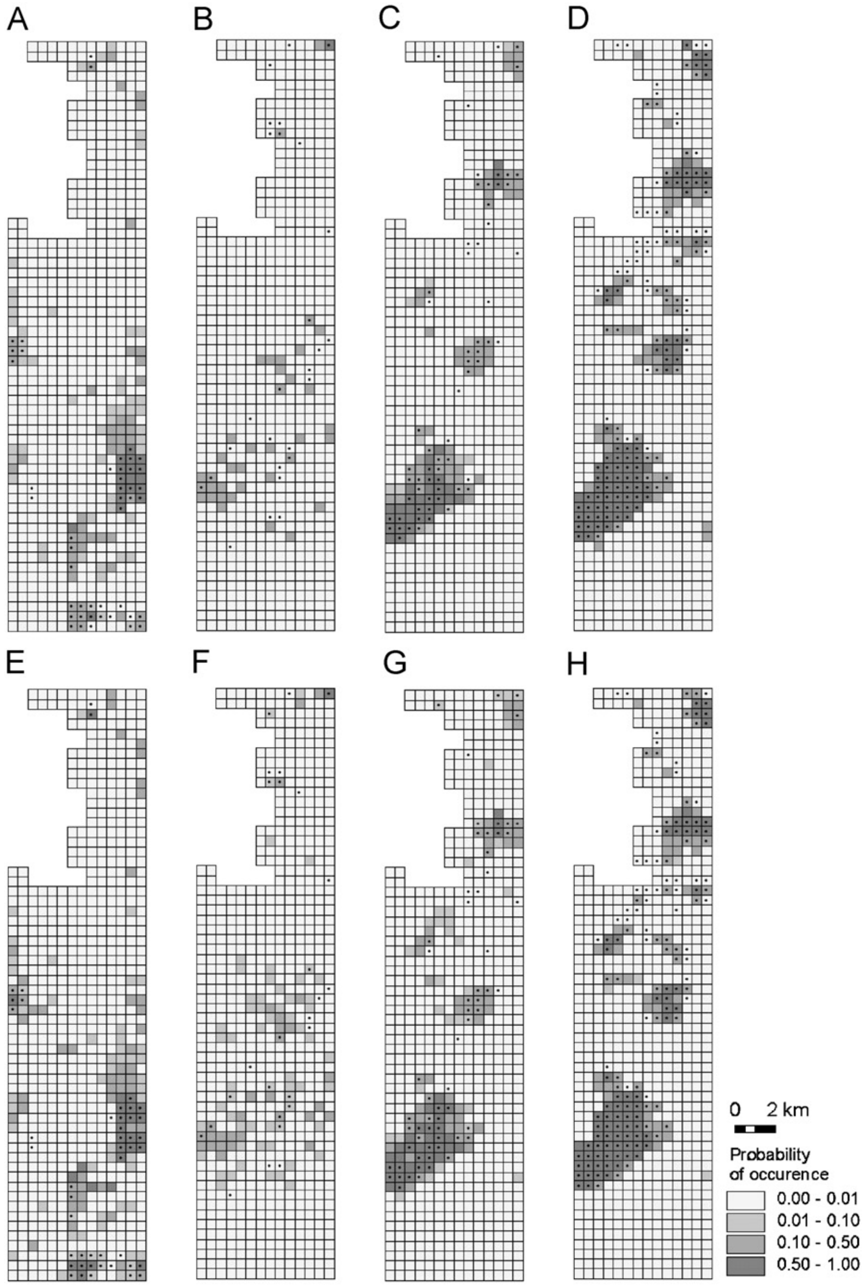

Fig. 4. Estimated spatial distribution of palsa mires (A and E), convex non-sorted circles (B and F), sorted stripes (C and G) and sorted solifluction sheets (D and H) in sub-arctic Finland by GAM (top row; AUC values of 0.866, 0.607, 0.849 and 0.859, respectively) and WA (bottom row; AUC values of 0.916, 0.713, 0.876 and 0.909, respectively). Gray levels represent different levels of probability of presence. Recorded presences (true) are shown with black dots.

to 0.916 with a mean value of 0.782, which is the highest among the four consensus methods [Median(All): 0.752; Median(PCA): 0.761; Median(AUC): 0.759]. The rank average of 11.5 highlights the performance of WA. Furthermore, the WA consensus method provides the highest number of excellent (3 of 12) and the least poor (4 of 12) predictions. The convex non-sorted circles illustrate this performance. The AUC of WA (0.713) was clearly higher than the value of Median(PCA) (0.641). The maps presented in Figs. 2 and 3 represent the spatial distribution of earth hummocks and stone pits estimated by WA (C).

5. Discussion

Predictive modelling has been increasingly utilized in several research topics (Walsh et al., 1998; Araújo et al., 2005a; Heikkinen et al., 2006; Thuiller et al., 2006; Pearson et al., 2007). In geomorphology, predictive modelling has the advantage of providing relevant and useful information on earth surface processes and landforms over extensive areas, such data being unavailable through more conventional survey methods (Atkinson et al., 1998; Rowbotham and Dudycha, 1998; Etzelmüller et al., 2001; Nilsson et al., 2002; Gurney and Bartsch, 2005). Spatial modelling can be used to develop predictions of the locations of the most suitable sites for a given geomorphological feature in unsurveyed parts of landscapes and to improve the targeting of efforts and resources to such sites (Walsh et al., 1998; Luoto and Hjort, 2005). Spatial modelling studies in geomorphology have usually been conducted by employing only one modelling approach (Atkinson et al., 1998; Luoto and Hjort, 2004; Walsh and MacNally, 2003). Some variation is expected from using different techniques, because different models use different of assumptions, algorithms and parameterizations. Thus, when studies use a single-modelling technique there is no information about whether the selected method provides the best predictive accuracy for the particular data set used (Araújo and New, 2006). In a global change study, Araújo et al. (2005a) emphasized the current need to reduce the uncertainties of model predictions. As a response to the variability in model performance between different methods, two recent developments to reduce the uncertainty in spatial modelling have been defined (Araújo and New, 2006). Rather than using a single-modelling technique, investigators can use different consensus methods: (i) choose a framework including different methods and models for each response variable and select the most accurate technique using different evaluation methods, or (ii) take a majority vote criterion approach among multiple models thus deriving a single projection that represents the central tendency across all models considered (Huang and Lees, 2004; Heikkinen et al., 2006).

In this study, we tested four statistical consensus methods to improve the modelling accuracy of geomorphological models. The predictive performances of the different modelling techniques were somewhat variable because of the variety of the mathematical algorithms on which the methods are based. In addition to the technical differences, distribution, abundance and environmental specificity of the modelled geomorphological phenomena varied considerably in the study area, which affected the ability of the models to predict the occurrences of different periglacial landforms (Hjort and Luoto, 2006; Hjort et al., 2007).

The median consensus method did not improve the performances of the eight individual modelling techniques. This was due to the median mathematical function itself. The Median(All) method does not have any pre-selective process of the models' predictions. Only the techniques with the fourth and the fifth largest probabilities of the eight were directly taken into account to calculate the new presence probability value, for each grid square. Consequently, the resulting probability of occurrence can be over- or underestimated. This partly explains the limits of the method. Nevertheless, considering the rank average index, its average performances were at the same level as GAM, which was on average the best-performing single-model technique.

Median(PCA) has been used to increase the reliability of forecasts of species extinction under climate change scenarios. As in the study carried by Araújo et al. (2005b), Median(PCA) provided in general more reliable predictions than single-model techniques. The same observation also held for Median(AUC). Both methods provide the same predictions of accuracy and the pre-selective processing of the predictions' models avoids the effects of the low performing single-model techniques on the predictions. The median function combined the second and the third largest probability values among the selected techniques to obtain a new projection. Our results confirm the good performance of median consensus methods obtained by Araújo et al. (2005b).

As in the hydrological study presented by Goswami and O'Connor (2007), the WA consensus method also provided the most robust predictions in the present study. Goswami and O'Connor (2007) compared this method to other consensus methods and the obtained predictions were promising. In the present study, in contrast to the other evaluated consensus methods, WA was a combination of the four best single-model techniques based on calibration data, which made it more robust compared with the other consensus methods.

In this study, it appeared clearly that the consensus methods were as accurate or better than the best single-model (see Table 5). It confirms the relevance of the use of consensus algorithms. Huang and Lees (2004, 2005) studied the ability of consensus methods based on majority and weighted votes for forest mapping. In their study, methods based on weighted votes improved the accuracy of single-models, which agrees with our results. However, the improvement of accuracy is not similar for all landforms. The predictive accuracy of palsa mire was not improved by any consensus methods, whereas the

Table 5
Advantages and drawbacks of tested consensus methods

	Pre-selection	Rank	Statistical improvement	Complexity
Median(All)	None	1 Least accurate method	None. *P*-value: 0.689	+
Median(PCA)	Principal component analysis (four models)	3	None. *P*-value: 0.790	+++
Median(AUC)	AUC criterion	2	None. *P*-value: 0.346	++
WA	AUC criterion	4 Best method	Yes. *P*-value: 0.011	++

Column "rank" indicates predictive accuracy of methods. Column "statistical improvement" reflects ability of methods to improve statistically accuracy of ensemble of predictions based on Wilcoxon signed ranks tests. Column "complexity" indicates computational elaborateness of methods.

prediction of stone pit was more accurate using consensus methods. Huang and Lees (2004) defined a confidence index of predictions based on the similarity of the predictions. However, in our study, the ability of consensus methods to improve the prediction of single-models was not correlated with the confidence index.

6. Conclusion

There is a strong need for robust predictions in geomorphologic research. Consensus methods based on the predictions of several single-models appear to be a powerful approach to improve the reliability of geomorphologic predictions. The weighted average consensus method provided superior predictions when compared with the other consensus methods and single-model techniques. The weighted average consensus method, which has not previously been used in a geomorphologic context, improved the accuracy of 11 predictions out of 12. Spatial modelling is used in various topics of physical geography. Although improved accuracy can be delivered through the traditional tasks of trying to build better models with improved data and statistical techniques, we propose that consensus methods should be utilized more often in theoretical and applied research projects in geomorphology.

Acknowledgments

MM was funded by the Academy of Finland (Project Grant 116544). WT was partly funded by the EU FP6 MACIS specific targeted project (Minimization of and adaptation to climate change: impacts on biodiversity No. 044399) and EU FP6 ECOCHANGE integrated project (Challenges in assessing and forecasting biodiversity and ecosystem changes in Europe). M. Bailey helped with correction of the English text. Kevo Subarctic Research Institute provided good facilities during the field surveys. The comments by J.J. Royer and a second unknown referee helped greatly in improving this paper.

References

Ahti, T., Hämet-Ahti, L., Jalas, J., 1968. Vegetation zones and their sections in northwestern Europe. Annales Botanici Fennici 5, 169–211.

Allen, P.A., 1997. Earth Surface Processes. Blackwell Science, Cambridge, 404pp.

Araújo, M.B., New, M., 2006. Ensemble forecasting of species distributions. Trends in Ecology and Evolution 22, 42–47.

Araújo, M.B., Pearson, R.G., Thuiller, W., Erhard, M., 2005a. Validation of species-climate impact models under climate change. Global Change Biology 11, 1504–1513.

Araújo, M.B., Whittaker, R.J., Ladle, R.J., Erhard, M., 2005b. Reducing uncertainty in projections of extinction risk from climate change. Global Ecology and Biogeography 14, 529–538.

Atkinson, J., Jiskoot, H., Massari, R., Murray, T., 1998. Generalized linear modelling in geomorphology. Earth Surface Processes and Landforms 23, 1185–1195.

Benestad, R.E., 2004. Tentative probabilistic temperature scenarios for northern Europe. Tellus 56A, 89–101.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Breiman, L., Friedman, F., Olshen, F., Stone, C., 1984. Classification and Regression Trees. Wadsworth Statistics/Probability Series. Wadsworth, Belmont, CA, USA, 358pp.

Brenning, A., Grasser, M., Friend, D.A., 2007. Statistical estimation and generalized additive modeling of rock glacier distribution in the San Juan Mountains, Colorado, United States. Journal of Geophysical Research 112, F02S15.

Brown, D., Lusch, D., Duda, K., 1998. Supervised classification of types of glaciated landscapes using digital elevation data. Geomorphology 21, 233–250.

Corte-Real, J., Zhang, X., Wang, X., 1995. Downscaling GCM information to regional scales: a non-parametric multivariate regression approach. Climate Dynamics 11, 413–424.

Dai, F.C., Lee, C.F., 2002. Landslide characteristics and slope instability modeling using GIS, Lanatau Island, Hong Kong. Geomorphology 42, 213–338.

Deichmann, J., Eshghi, A., Haughton, D., Sayek, S., Teebagy, N., 2002. Application of multiple adaptive regression splines (MARS) in direct response modeling. Journal of Interactive Marketing 16, 15–27.

Deveaux, R.D., Gordon, A.L., Comiso, J.C., Bacherer, N.E., 1993. Modeling of topographic effects on Antarctic sea-ice using multivariate adaptive regression splines. Journal of Geophysical Research-Oceans 98, 20307–20319.

Elith, J., Graham, C., Anderson, R., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehman, A., Li, J., Lohman, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. Overton, J., Peterson, T.A., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.

Etzelmüller, B., Ødegård, R.S., Berthling, I., Sollid, J.L., 2001. Terrain parameters and remote sensing data in the analysis of permafrost distribution and periglacial processes: principles and examples from southern Norway. Permafrost and Periglacial Processes 12, 79–92.

Fielding, A., Haworth, P., 1995. Testing the generality of bird-habitat models. Conservation Biology 9, 1466–1481.

Franklin, J., 2002. Enhancing a regional vegetation map with predictive models of dominant plant species in chaparral. Applied Vegetation Science 5, 135–146.

French, H.M., 1996. The Periglacial Environment, second ed. Longman, London, 341pp.

Friedman, J., 1991. Multivariate adaptive regression splines. Annals of Statistics 19, 1–141.

Goswami, M., O'Connor, K.M., 2007. Real-time flow forecasting in the absence of quantitative precipitation forecasts: a multi-model approach. Journal of Hydrology 334, 125–140.

Graff, L.H., Usery, E.L., 1993. Automated classification of generic terrain features from digital elevation models. Photogrammetric Engineering and Remote Sensing 59, 1409–1417.

Gregory, A.W., Smith, G.W., Yetman, J., 2001. Testing for forecast consensus. Journal of Business and Economic Statistics 19, 34–43.

Gude, M., Barsch, D., 2005. Assessment of geomorphologic hazards in connection with permafrost occurrence in the Zugspitze area (Bavarian Alps, Germany). Geomorphology 66, 85–93.

Gurney, S.D., Bartsch, A., 2005. Mapping the spatial distribution of geomorphological processes in the Okstindan area of northern Norway, using Geomorphic Process Units as derived from remote sensing and ground survey. Fennia 183, 1–14.

Guzzetti, F., Carrara, A., Cardinali, M., Reichenbach, P., 1999. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology 31, 181–216.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., Galli, M., 2006. Estimating the quality of landslide susceptibility models. Geomorphology 81, 166–184.

Haeberli, W., 1992. Construction, environmental problems and natural hazards in periglacial mountain belts. Permafrost and Periglacial Processes 3, 111–124.

Harris, C., Davies, M.C.R., Etzelmüller, B., 2001a. The assessment of potential geotechnical hazards associated with mountain permafrost in a warming global climate. Permafrost and Periglacial Processes 12, 145–156.

Harris, C., Haeberli, W., Mühll, D.V., King, L., 2001b. Permafrost monitoring in the high mountains of Europe: the PACE Project in its global context. Permafrost and Periglacial Processes 12, 3–11.

Hartley, S., Harris, R., Lester, P.J., 2006. Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant. Ecology Letters 9, 1068–1079.

Hastie, T., Tibshirani, R., 1990. Generalized Additive Models. Chapman & Hall, London, 352pp.

Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W., Sykes, M.T., 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. Progress in Physical Geography 30, 751–777.

Hjort, J., 2006. Environmental Factors Affecting the Occurrence of Periglacial Landforms in Finnish Lapland: A Numerical Approach. Shaker Verlag, Aachen, 162pp.

Hjort, J., Luoto, M., 2006. Modelling patterned ground distribution in Finnish Lapland: an integration of topographical, ground and remote sensing information. Geografiska Annaler 88A, 19–29.

Hjort, J., Luoto, M., Seppälä, M., 2007. Landscape scale determinants of periglacials features in subarctic Finland: a grid-based modelling approach. Permafrost and Periglacial Processes 18, 115–127.

Huang, Z., Lees, B.G., 2004. Combining non-parametric models for multisource predictive forest mapping. Photogrammetric Engineering and Remote Sensing 70, 415–425.

Huang, Z., Lees, B.G., 2005. Representing and reducing error in natural-resource classification using model combination. International Journal of Geographical Information Science 19, 603–621.

Kheir, R.B., Chorowicz, J., Abdallah, C., Dhont, D., 2008. Soil and bedrock distribution estimated from gully form and frequency: a GIS-based decision-tree model for Lebanon. Geomorphology 93, 482–492.

King, L., Seppälä, M., 1987. Permafrost thickness and distribution in Finnish Lapland—results of geoelectrical soundings. Polarforshung 57, 127–147.

Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? Ecology 74, 1659–1673.

Lek, S., Guegan, J., 1999. Artificial neural networks as a tool in ecological modelling, an introduction. Ecological Modelling 120, 65–73.

Luoto, M., Hjort, J., 2004. Generalized linear models in periglacial studies: terrain parameters and patterned ground. Permafrost and Periglacial Processes 15, 327–338.

Luoto, M., Hjort, J., 2005. Evaluation of current statistical approaches for predictive geomorphological mapping. Geomorphology 67, 299–315.

Luoto, M., Seppälä, M., 2002. Modelling the distribution of palsas in Finnish Lapland with logistic regression and GIS. Permafrost and Periglacial Processes 13, 17–28.

Mackay, D.S., Robinson, V.B., Band, L.E., 1992. Classification of higher order topographic objects on digital terrain data. Computers, Environment and Urban Systems 16, 473–496.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, second ed. Chapman & Hall, New York, 511pp.

Meriläinen, K., 1976. The granulite complex and adjacent rock in Lapland, northern Finland. Bulletin of Geological Survey of Finland 281, 1–129.

Merritt, D.M., Wohl, E.E., 2003. Downstream hydraulic geometry and channel adjustment during a flood along an ephemeral, arid-region drainage. Geomorphology 52, 165–180.

Moore, I.D., Grayson, R.D., Ladson, A.R., 1991. Digital terrain modelling: a review of hydrological, geomorphological and biological applications. Hydrological Applications 5, 3–30.

Nilsson, J., Persson, B., von Heijne, G., 2000. Consensus prediction of membrane protein topology. FEBS (Federation of European Biochemical Societies) Letters 486, 267–269.

Nilsson, J., Persson, B., Von Heijne, G., 2002. Prediction of partial membrane protein topologies using a consensus approach. Protein Science 11, 2974–2980.

Pearson, R.G., Raxworthy, C.J., Nakamura, M., Townsend Peterson, A., 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of Biogeography 34, 102–117.

Peterson, A.T., Vieglais, D.A., 2001. Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. Bioscience 51, 363–371.

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmerman, N.E., Zappa, M., Guisan, A., 2006. Are species distribution models transferable in space? Journal of Biogeography 33, 1689–1703.

Ridgeway, G., 1999. The state of boosting. Computing Sciences and Statistics 31, 172–181.

Rowbotham, D.N., Dudycha, D., 1998. GIS modelling of slope stability in Phewa Tal watershed, Nepal. Geomorphology 26, 151–170.

Sanders, F., 1963. On subjective probability forecasting. Journal of Applied Meteorology 2, 191–201.

Swets, K., 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285–1293.

Thuiller, W., 2003. BIOMOD—optimizing predictions of species distributions and projecting potential future shifts under global change. Global Change Biology 9, 1353–1362.

Thuiller, W., 2004. Patterns and uncertainties of species' range shifts under climate change. Global Change Biology 10, 2020–2027.

Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T., Prentice, I.C., 2005. Climate change threats to plant diversity in Europe. In: Proceedings of the National Academy of Sciences, 102, pp. 8245–8250.

Thuiller, W., Broennimann, O., Hughes, G.O., Alkemade, J.M.R., Midgley, G.F., Corsi, F., 2006. Vulnerability of African mammals to anthropogenc climate change under conservative land transformation assumptions. Global Change Biology 12, 424–440.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics With S, fourth ed. Springer, Berlin, 495pp.

Vitek, J.D., Giardino, T.R., Fitzgerald, J.W., 1996. Mapping geomorphology: a journey from paper maps, through computer mapping to GIS and virtual reality. Geomorphology 16, 233–249.

Walsh, C., MacNally, R., 2003. Hierarchical partitioning. R project for statistical computing 〈http://cran.r-project.org/〉.

Walsh, S.J., Butler, D.R., Malanson, G.P., 1998. An overview of scale, pattern, process relationships in geomorphology: a remote sensing and GIS perspective. Geomorphology 21, 183–205.

Washburn, A.L., 1979. Geocryology—A Survey of Periglacial Processes and Environments. Arnold, London, 406pp.

Zhu, Y.M., Lu, X.X., Zhou, Y., 2007. Suspended sediment flux modeling with artificial neural network: an example of the Longchuanjiang River in the upper Yangtze catchment, China. Geomorphology 84, 111–125.