Probability models, remote sensing and field observation: test for mapping some plant distributions in the Kongsfjord area, Svalbard

THIERRY BROSSARD and DANIEL JOLY

Brossard, T. & Joly, D. 1994: Probability models, remote sensing and field observation: test for mapping some plant distributions in the Kongsfjord area, Svalbard. *Polar Research 13*, 153–161.

The coupling of field observations of plants and data derived from satellite image classification makes it possible to build a statistical model which can be used to determine the probability of a vegetation character existing in a pixel of a known class. Probability maps indicating the presence and abundance of several vascular and non-vascular species were produced for lowland areas. These maps are used to establish an indirect relationship between vegetation characters and spectral signatures which give an indication of the habitat conditions through an overall classification of the image.

T. Brossard and D. Joly, Laboratorie de Géographie Physique, CNRS URA 908, 30 rue Mégevand, F-25030 Besancon, Cedex, France.

Introduction

Satellite imagery rarely provides direct information about plant cover composition and distribution in polar environments. Mineral components dominate in the landscape structure and for a large part determine spectral signatures (Franck & Isard 1986). Even when a tundra cover – which is easily detectable on the image – does develop, micro-local conditions induce scale differences smaller than a pixel (Rønning 1969; Brossard et al. 1984), so that plant groups, and to an even greater extent, plant species, escape satellite detection.

In an attempt to overcome this problem, we have used an indirect method, based on a model of empirical probabilities. The idea was to establish the statistical connection between field observation readings of plant species or groups, and the overall landscape types derived from satellite image classification (Tom & Miller 1984). The model thus built makes it possible, for each identified landscape type, to associate a given plant species or group with a probability value. On this basis, it is possible to draw up a map of the probable distribution of each of the plant components observed.

Material and study area

The image used for the experiment is from Landsat Thematic Mapper, 217–3, 20 July 1985, without geometrical correction. The pixel size is 30 × 30 metres.

The field observations were made during summer 1991 on both sides of Kongsfjorden. However, in order to show a greater number of results, the maps presented in this paper (Figs. 4–9) are restricted to Brøgger Peninsula (Fig. 1).

The peninsula is situated in the northwestern part of Spitsbergen. At Ny-Alesund 78.54°N, the mean temperature in July is 5.1°C and the mean annual precipitation is 385 mm. Geologically speaking, rocks belonging to the Cambrian-Ordovician periods bear clear indications of Caledonian orogeny and they are found in the mountain which forms the backbone of the peninsula. The major sedimentation periods were during the Permian-Carboniferous (sandstone, limestone) and then Tertiary (coal, sandstone) periods. These structures have been rejuvenated by recent Tertiary tectonic reactivation and Quaternary isostatic adjustment. Between the sea and the mountain, the strandflat is highly characteristic of these latitudes, with combinations of

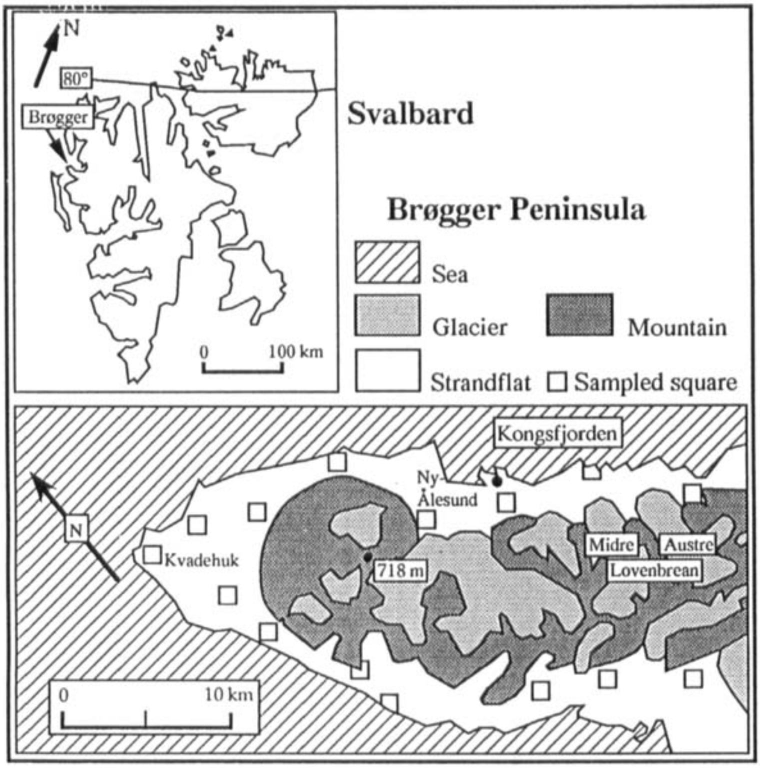

Fig. 1. The study area: Brøgger Peninsula, northwest Spitsbergen, Svalbard.

residual marine formations (raised beaches and sandbars), recent or active outwash plains and amphitheatre-shaped moraines deposited during the Little Ice Age (19th century). The landscape continues to be shaped by present-day processes including the congelifraction of rock shelves and a wide variety of ground deformation caused by frost action. Although generally sparse, vegetation is rarely absent even on the high peaks, and plants gain ground in juvenile moraines and on sandurs that are beginning to stabilize. The most advanced forms of vegetation are linked to *Dryadion* and moss tundra communities (Rønning 1969; Elvebakk 1985). The lowland vegetation of Brøgger Peninsula has been mapped by Brattbak (1981). Nesting, particularly by seabirds, provides the nitrogen which enables these areas to support a high biomass. This results in thick beds of waterlogged mosses.

Methods

Observation procedure

To ensure a valid statistical basis for the experiment, it was necessary to have a large number of field observation points providing a sample of all the area's biotopes. We were forced to exclude the mountainous area in the centre of the peninsula from the survey because of the amount of time needed to reach it. Our study deals only with the strandflat area (Fig. 1). Consequently, it must be emphasized that although the model has been applied to the whole peninsula, its validity varies according to the area in question. For this reason, the boundary of the area for which data were collected will be demarcated on the following maps.

Logistical constraints also affected the choice of a field sampling protocol (Hagget 1973): the distribution of control points according to a regular grid would have considerably lengthened the time needed to move from one point to another. This was not compatible with the time we had at our disposal. We therefore used the following spatial survey protocol which is an unaligned systematic sampling. An initial analysis of the image based on our knowledge of the terrain served to divide the area into main geographical units of comparable dimension. Fifteen representative

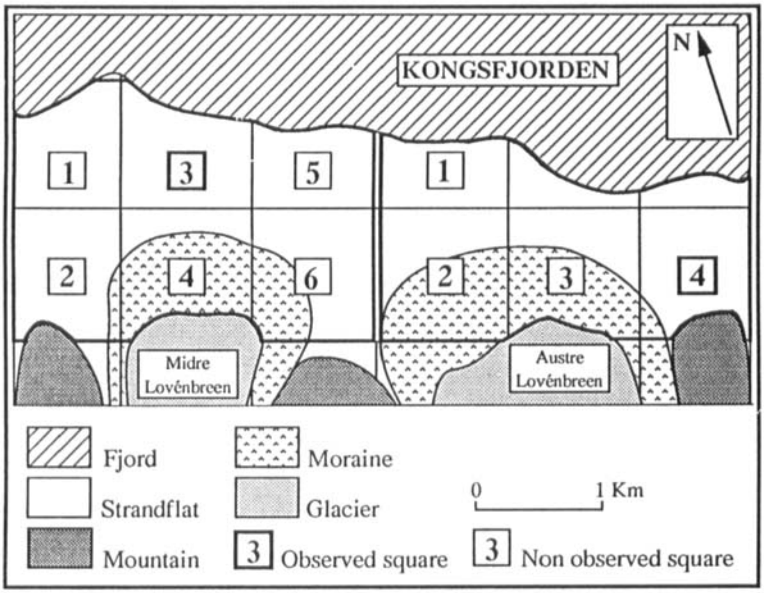

Fig. 2. Example of sampling procedure for two geographical units in front of Loven glaciers.

test areas were thus delimited on Brøgger Peninsula.

Within each unit, using a geometrical grid as a guide, we defined four to six areas of approximately one square kilometre each. We drew lots to designate which area would be observed in each unit; their positions are indicated in Fig. 1. An illustration of the sampling method is given by Fig. 2 where a zoom is made on two basic geographical units and their kilometric divisions. Within each test area, we drew transects across the landscape's structural elements, along which observation points were set up at regular intervals of 200 metres. These points were accurately detected on high-quality aerial photographs, ensuring accurate in-the-field detection. This protocol made it possible to reconcile the required sampling precision and the logistical constraints.

We then proceeded to record the plant species on a surface with an area equivalent to one pixel at each observation point. We only recorded the forty most common species in order to avoid problems due to such a large observation area. Information on topography, grain size and geomorphology (Satterwhite et al. 1984) is also taken into account. This information would be useful for the comments made later, but were not used for model building. A total number of 590 readings were taken. Fifteen species with a frequency less than 60/590 (±10%) were not kept for building the model.

Data preparation

Since all the observed sites were accurately detected, it was possible to locate their positions on the satellite image with a precision of a pixel and to plot them on the screen using a target. Subsequently, a data table was established in which the individuals (i.e., the observation points) are characterized by their spectral signatures in the different Thematic Mapper bands. Among them, thermal band No. 6 was not used for model building due to its different pixel scale.

Analysis of correspondences – a particular method of multivariate analysis based on chi square metrics (Benzecri 1980; Hill 1974) – was then applied to this table: as Principal component analysis, it provides a derived space which is defined by a different axis. One of the interesting properties of the resulting factorial space, called "distributional equivalence", is that it facilitates the simultaneous projection of characteristics (i.e., spectral signatures) and individuals (i.e., sampled pixels) on a single plane defined by one axis couple. The values giving the distance between individuals and characteristics were derived from the analysis of correspondences and were used to classify the elements of both categories in a hierarchical ascending order (Roux 1985). In this way, 8 types were defined. As each pixel is characterized by a given type, the projection of the typology on the different fac-

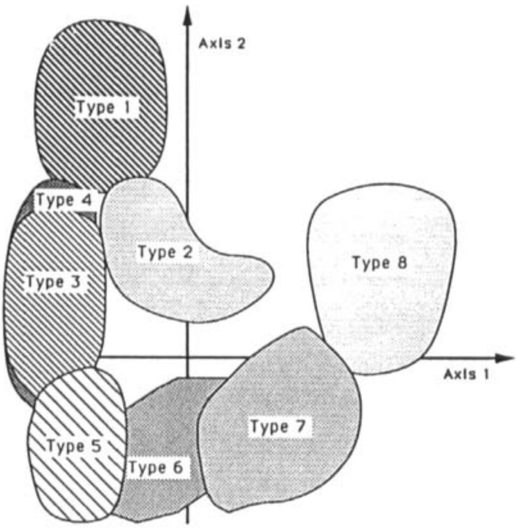

Fig. 3. Factorial structure according to axes 1–2: typology projection; types 3 and 4, which are superimposed, are separated by the third axis.

torial planes is made possible. Fig. 3 illustrates how the radiometric types thereby defined are distributed on the factorial structure of axes 1–2. It is to be noted that from type 1 to type 8, the analysis also classifies the signatures from the weakest to the strongest. Using this classification of the observed points, it is possible to classify all the pixels on the image by applying the criterion of the smallest factorial distance, hence Fig. 4 (Beaubien 1983; Howarth & Wickware 1983). Another point worth mentioning is that sea, shadow and ice, which had previously been defined by a first classification, were masked during processing; they correspond to additional types 9 (shadow, sea) and type 10 (ice) on Fig. 4.

As we also had field observation results of a varied nature, it was possible to interpret this typology and provide a description of its content. Following is a brief summary of the main types characteristics with reference to major plant habitats as they are classified by Elvebakk (1985). The type numbers refer to the boxes in the legend of Fig. 4.

Type 1. – Type 1 includes habitats with an abundant biomass and, most of the time, a high moisture level. These habitats correspond to mire, wetland and sometimes moss tundra. The nearby presence of bird sites such as cliffs and water stretches is also frequently observed. This accounts for the nitrophilous character of the vegetation (Sendstad 1978).

Type 2. – In spite of a high factorial proximity, the physiognomy of type 2 is very different from that of type 1. Water is still a relevant factor; it saturates the material – but there is a low vegetation cover. A part of the moraines and sandurs is classified here, as well as a few areas behind the alignments of former or current offshore bars.

Type 3. – Type 3 physiognomy closely resembles that of type 1, but with attenuated features notably regarding a less generalized saturation level. However, the biomass level remains high (moss tundra).

Type 4. – In type 4, vegetation cover becomes irregular or thin with a localization on gentle gelifluxion slopes, old sandur or marine terraces.

Type 5. – Most of the time, type 5 corresponds to accumulation forms with a small grain size (silty-sandy). The vegetation color is dark due to a dominance of a biotic black crust (sometimes in recent accumulation plains), combined with brown lichens.

Type 6. – In type 6, we encountered affirmed mineral characteristics including moraines and sandurs, but plant colonization is under way.

Type 7. – Type 7 shows even more marked mineral characteristics; either geomorphological processes are under way (frost action, outwash) or limiting factors have blocked the process of vegetal colonization (coarse material, late snow beds).

Type 8. – Type 8 is characterized by the highest reflectance values associated with a well-drained, often pale-coloured rock material (limestone). Found here are offshore bars with a homogeneous grain-size and units with bedrocks (intact or fragmented by frost). Most of type 8 corresponds to polar desert habitats.

Type 9. – Type 9 represents shadow and sea (masked areas).

Type 10. – Type 10 corresponds to ice or snow covered landscapes (masked areas).

Model building

The building of the model consisted of establishing a relationship between the classes (eight first landscape types described above) defined by signature processing and the distribution of the plant types observed within those classes. The frequencies found were used to calculate a value giving the empirical probability of a given plant component to exist in each pixel of a known class (Spiegel 1982). This model is presented as a matrix (Table 1) that crosslinks probability values for the classes and the considered objects (defined in term of vegetation characters). The accuracy of this approach to model building is measured by the fiducial threshold and range. One notices that these objects are recorded in Boolean form (i.e., in terms of presence/absence). When a given species occurred with a high frequency (>50%) and different cover rates, it was encoded into several abundance sub-categories giving distinct objects; among them, absence could be distinguished as an additional boolean character. In this way, absence probabilities are also likely to be obtained in order to determine a repulsion space for a given vegetation character.

Table 1 provides an example of the values obtained for several categories of vegetation characters. Thus, moss cover rate has been encoded into three different classes, regardless of the species criterion. A distinction has also been made with regard to the abundance of *Cetraria delisei* (and *C. islandica*), *Dryas octopetala*. The other species however, have been considered only on a presence-based criterion. Now, considering the column which corresponds to type 1, the table is to be read as to follow: if a given pixel is ranked in type 1, its probability values to have *C. delisei* with a cover of 0%, <10% and ≥10% are 33%, 35% and 32%, respectively.

Results: probability maps

Since the landscape type represented on each pixel is known for the whole image, this model can be used to extend the results obtained from individual observations to the whole space. This operation is therefore a form of interpolation leading to the edition of probability maps for each of the plant components observed. Consequently, the maps are a way of characterizing their potential distribution.

The selection of documents established from the model is not related to a given ecological problem, but rather aims at showing how probability maps illustrate different phyto-geographical aspects of plant distribution. Even if the chosen taxons are very common, their behaviours concerning space occupation and "strategy" are shown as very distinct.

Papaver dahlianum (Fig. 5). – There is little contrast in the distribution space of this species. It has a marked repulsion for wet habitats with a thick plant cover. But the probability of finding it on mineral environments is also low. The probability map thus provides a good illustration of the sporadic distribution of this pioneer plant.

Silene acaulis (Fig. 6). – There is a strong contrast in the distribution space of this species, which has

Table 1. One example of probabilities for some vegetation characters and the 8 types derived from classification.

Vegetation characters	cover %	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8
Cetraria delisei	=0	33	76	30	15	51	63	61	62
(and *C. islandica*)	>10	35	18	23	51	38	32	34	38
	≥10	32	6	47	34	11	5	5	0
Dryas octopetala	=0	59	88	56	51	66	79	76	88
	<10	17	6	25	28	15	11	15	10
	≥10	24	6	19	21	19	9	10	1
Mosses	=0	10	18	9	10	8	15	23	39
	<20	32	35	19	31	30	17	20	10
	≥20	51	0	39	24	26	2	10	3
Papaver dahlianum	>0	2	12	4	12	2	16	10	16
Silene acaulis	>0	79	35	82	73	72	52	34	8

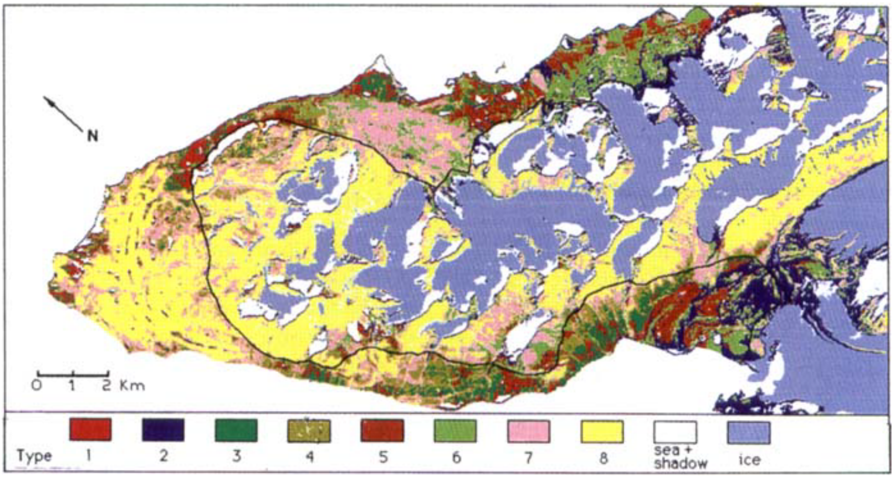

Fig. 4. Application of classification to the whole image.

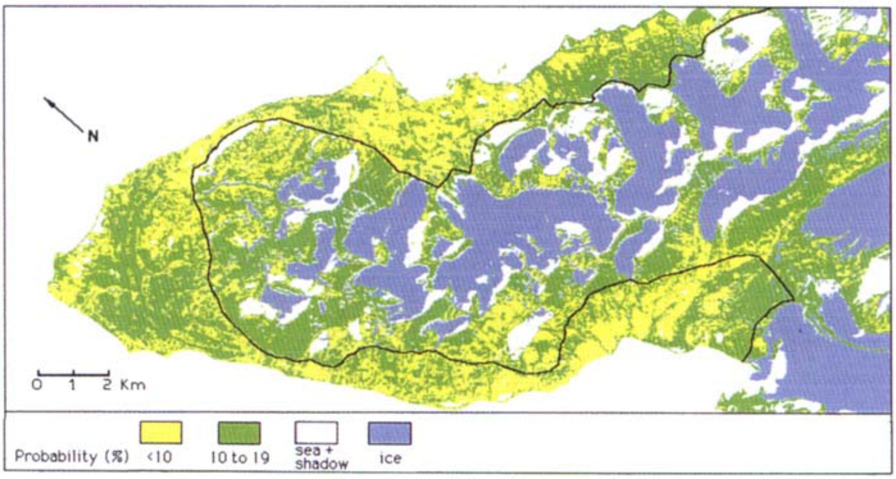

Fig. 5. Probability map: presence of *Papaver dahlianum*.

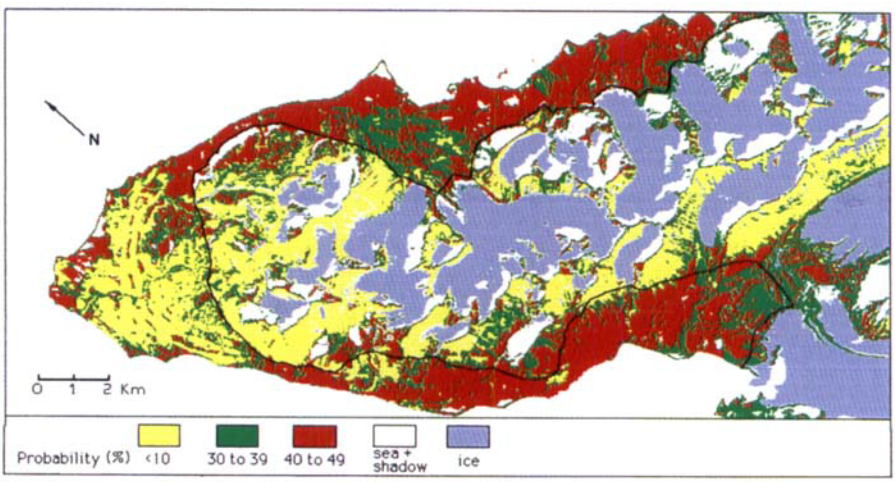

Fig. 6. Probability map: presence of *Silene acaulis*.

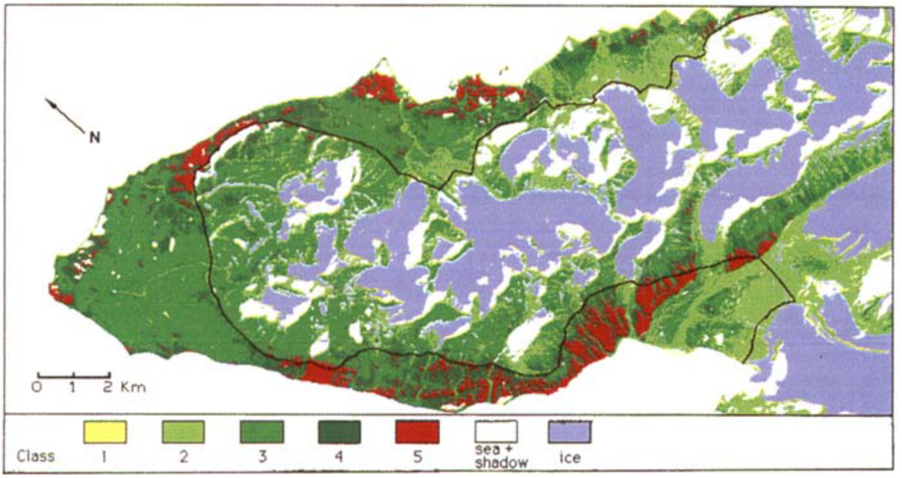

Fig. 7. Vegetation index (NDVI): (TM4 − TM3)/(TM4 + TM3). From class 1 (yellow) to class 5 (red), the index values increase.

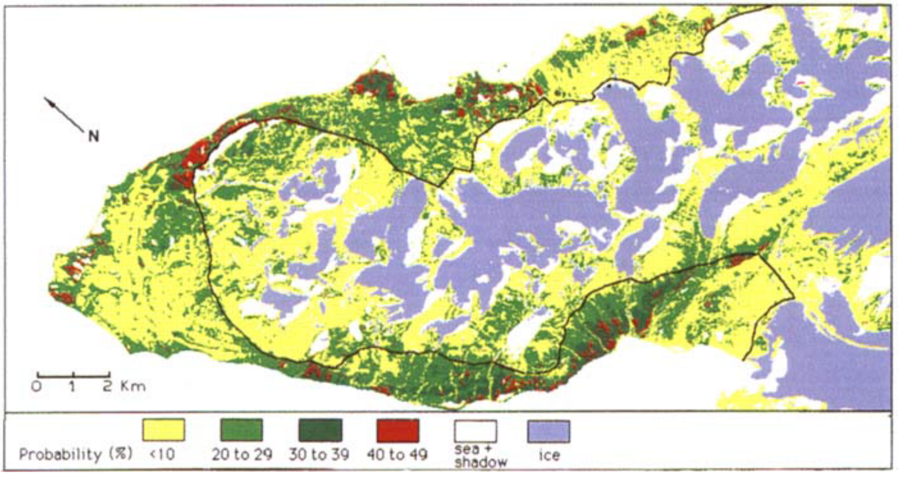

Fig. 8. Probability map: presence of moss with >20% cover rate.

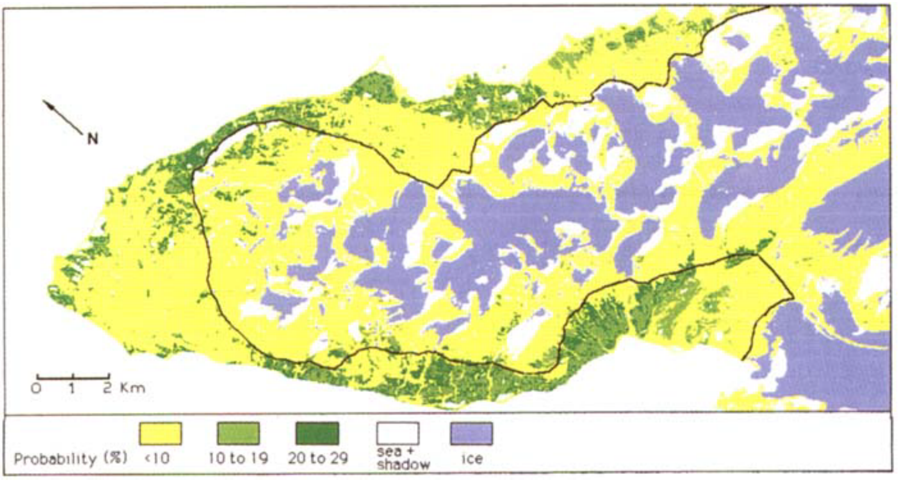

Fig. 9. Probability map: presence of *Dryas octopetala* with >10% cover rate.

a high presence probability on most landscape types. On the other hand, there is little likelihood of finding it in a drained mineral environment.

Vegetation index (Fig. 7) and mosses with a high cover rate (Fig. 8). – We present these two maps simultaneously to show how conventional image analysis processes such as the calculation of vegetation indices can be complemented with documents established from the proposed model. The index used here – (TM4 – TM3)/(TM4 + TM3) – is the widely-used normalized difference vegetation index (NDVI). A comparison of the two images shows a parallel variation of values. This is quite normal since we have only considered mosses with high cover rates; the importance of their contribution to the overall cover rate is thus highlighted. Nonetheless, some nuances can also be seen in this comparison. From class 1 to class 3 of Fig. 7, there is a lesser contribution of mosses to the biomass; other species appear which are more predominant, such as *Dryas octopetala*, *Cetraria delisei* and *Cassiope tetragona*. The model also shows that on the western end of the peninsula the presence probability of mosses with high cover rates is associated with wet depressions located behind the offshore bars; this escapes vegetation index calculations.

Model limitations

The indirect relationship established between spectral signatures and vegetation characters through an overall classification of the image sometimes highlights problems clearly illustrated by the *D. octopetala* map with a high cover rate (Fig. 9). For a large part, the indications on this map appear to be correct: *D. octopetala* is associated with types 1 and 3 whose abundant biomass mostly corresponds to mature plant groups. Yet, the results drawn from field observations show that this is not the case for certain geographical areas, notably the wet depressions surrounding lagoons on the western end of Brøgger. Not all the readings taken in this area include *D. octopetala*, even with low cover rates. The classification has determined as homogeneous wet environments some sets which are in fact a local combination of different elements. For instance, on the northern banks of Kongsfjorden, on some terraces and slope bottoms, large sorted stripes are found, that include a mosaic of extremely saturated environments and drier side folds where *D. octopetala* is abundant. Hence the mistakes made in the probability calculations when large-scale wet geographical areas do not include the above-described mosaic.

To overcome this difficulty, we should first conduct specific research on the relevant points and their signatures in order to find out whether two major subclasses can be distinguished. However, such case-by-case classification modifications may lessen the value of the model since this is a long procedure. We are investigating another solution which consists in complementing satellite images with other sources of digitalized information such as models of altitude, slope, orientation, solar energy, geology, etc., within the scope of a GIS-related approach. The typology of landscapes according to their ecological potential would thus be improved. Another alternative would be to introduce the criterion of geographical position in order to reinforce the value of nearby observations in the building of the model (Benali & Escoffier 1990; Lebart 1984). But in this case, we would have to modify our sampling protocol and go back to a homogeneous distribution of the observation points.

Conclusion

Model building using the probability approach combined with satellite data processing can be proposed for vegetation cartography. With this objective in view, the hypothesis proposed is that satellite data provides significant information on habitat characteristics which determine plant cover composition and distribution. Thus, the relationship between vegetation and spectral signatures is formalized in indirect terms. Image classification is used to recognize different landscape-habitat types to which observed objects (i.e., different plant taxons) are correlated by statistics, in order to build up empirical probability models. The practical results of this process of image classification are maps showing the probable distributions of observed vegetation characters according to the typology. When compared with the ground truth, the accuracy of the model fails when the typology keeps in the same class habitats for which the ecological characteristics are not homogeneous for a given plant character. This can be explained by the fact that only one information source was taken into account for recognizing the habitat types. In such a case, the satellite image would need to be completed by other data layers available in a geographical information system. In order to obtain a more precise characterization of habitat types, elevation models could be helpful as they allow us to derive additional data, for example on altitude, gradient, exposure, landform and solar radiation.

Moreover, maps established by this method and integrated with geographical information systems constitute levels of information likely to provide an aid to arctic environment inventory and management.

It is also interesting to state that all the processes and documents we are presenting have been established on standard micro-computer hardware. They can therefore be decentralized and used as near as possible to the observation field. The maps, thus edited, provide research workers interested in environmental issues with a tool for basing their approach of space on empirical methods.

References

Beaubien, J. 1983: Une méthode de réhaussement d'image Landsat pour la classification du couvert végétal. Pp. 559–566 in *Proceedings of the 8th Canadian Symposium on Remote Sensing*. Metrolitho Inc., Sherbrooke, Canada.

Benali, H. & Escoffier, B. 1990: Analyse factorielle lissée et analyse factorielle des différences locales. *Revue Statistique Appliquée 38–2*, 55–76.

Benzecri, J. P. 1980: L'analyse des données. Dunod, Paris. 617 pp.

Brattbakk, I. 1981: *Vegetasjonkart 1:10,000*. Norsk Polarinstitutt, Oslo.

Brossard, T., Deruelle, S. & Nimis, P. L. 1984: An interdisciplinary approach to vegetation mapping in High Arctic Environment, Ny-Ålesund, Svalbard. *Phytocoenolgia 12–4*, 433–453.

Elvebakk, A. 1985: Higher phytosociological syntaxa on Svalbard and their use in subdivision of the Arctic. *Nordic Journal of Botany 5*, 273–284.

Franck, T. D. & Isard, S. A. 1986: Alpine vegetation classification using high resolution aerial imagery and topoclimatic index values. *Photogrammetric Engineering and Remote Sensing 52*, 381–388.

Hagget, P. 1973: L'analyse spatiale en géographie humaine. Armand Colin, Paris. 430 pp.

Hill, M. O. 1974: Correspondence analysis: a neglected multivariate method. *Appl. Statist. 23–3*, 340–354.

Howarth, P. J. & Wickware, G. M. 1983: Multivariate analysis techniques for land cover classification using multispectral scanner data. Pp. 463–473 in *Proceedings of the 8th Canadian Symposium on Remote Sensing*. Metrolitho Inc., Sherbrooke, Canada.

Lebart, L. 1984: Correspondence Analysis of graph structure. *Bulletin technique du Centre de Statistique et Informatique appliquées, 2*. Paris p. 5–19.

Roux, J. 1985: *Algorithmes de classification*. Masson, Paris.

Rønning, O. 1969: Features of the ecology of some arctic Svalbard (Spitsbergen) plant communities. *Arctic and Alpine Research 1–1*, 29–44.

Satterwhite, M., Rice, W. & Shipman, J. 1984: Using landform and vegetative factors to improve the interpretation of Landsat imagery. *Photogrammetric Engineering and Remote Sensing. 50* 83–91.

Sendstad, E. 1978: Notes on the biology of an Arctic bird rock. *Årbok 1977*. Norsk Polarinstitutt, Oslo.

Spiegel, M. R. 1982: *Theory and problems of statistics*. McGraw-Hill, New-York.

Tom, C. H. & Miller, L. E. 1984: An automated land-use mapping comparison of the Bayesian maximum likelihood and linear discriminant analysis algorithms. *Photogrammetric Engineering and Remote Sensing 50*, 193–207.