Briefing note

A new baseline of the area of semi-natural habitat in Wales for Indicator 43

Summary

The purpose of this note is to explain the method we have used to derive an estimate of the area of semi-natural habitat in Wales for publication by Welsh Government as National Well-being Indicator 43. Our latest estimate amounts to an area of 640,827 ha or 31% of the land area of Wales. The estimate is derived from a combined analysis of existing habitat maps and recent satellite imagery. Some elements of the analysis are quite novel and, as such, the estimate should be treated as experimental at this stage. The method will be refined further over the coming years.

Llinell sylfaen newydd o'r ardal o gynefin lled-naturiol yng Nghymru ar gyfer Dangosydd 43

Crynodeb

Diben y nodyn hwn yw esbonio'r dull a ddefnyddiwyd gennym i gael amcangyfrif o arwynebedd cynefin lled-naturiol Cymru i'w gyhoeddi gan Lywodraeth Cymru fel Dangosydd Lles Cenedlaethol 43. Mae ein hamcangyfrif diweddaraf yn cyfateb i arwynebedd o 640,827 hectar neu 31% o arwynebedd tir Cymru. Mae'r amcangyfrif yn deillio o ddadansoddiad cyfunol o fapiau cynefinoedd sy'n bodoli eisoes a delweddau lloeren diweddar. Mae rhai elfennau o'r dadansoddiad yn eithaf newydd ac, o’r herwydd, dylid trin yr amcangyfrif fel rhywbeth arbrofol ar hyn o bryd. Bydd y dull yn cael ei fireinio ymhellach dros y blynyddoedd nesaf.

Background

An initial estimate of the area of semi-natural habitat in Wales was reported as part of the *Well-being of Wales, 2016-17* report. This estimate was derived from the field-based Habitat Survey of Wales undertaken between 1979 and 1997 using (mainly) Phase 1 survey methods[1].

The sustained resource required to undertake a comprehensive field survey of this sort is unlikely to be available to repeat such an exercise in the foreseeable future, so we cannot reasonably expect to obtain updated figures using these same methods. In short, we have virtually no immediate (or even medium-term) prospect of being able to produce a directly comparable figure to this earlier published estimate.

In the work summarised here, we have developed an approach to deriving a new baseline estimate that can be updated in the coming years using consistent data sources and methodology. The approach makes use of evidence from satellite imagery to update our existing understanding of the distribution and extent of habitats across Wales.

While the work here employs a well-established metric, some aspects of the analysis are quite novel. We expect that further work will refine the approach and allow more detailed presentation of the results. In the meantime, we propose that the estimate be treated as an 'experimental statistic'.

Methodology

We start from the fortunate position of already knowing from previous mapping where the major stands of semi-natural habitat are in Wales. In addition to the above-mentioned Habitat Survey of Wales, we have a more recent update (the 'Phase 1 Update') of this mapping, produced through the incorporation of evidence from satellite imagery captured between 2003 and 2006. We also have evidence from more detailed field-based Phase 2 National Vegetation Classification (NVC) surveys of some habitats.

In simple outline, our approach involved the following steps.

Step 1: Assigning areas on the Phase 1 Update map to one of three broad classes:

- Semi-natural (SN) – includes for example blanket bog, heathland, unimproved grassland, broad-leaved woodland. (This also includes areas of water and estuaries.)
- Improved (I) – includes agriculturally improved land (grass and arable), conifer plantations, infrastructure, urban and bare rock.
- Possibly Improved (PI) – areas whose position either side of the semi-natural / Improved boundary is less clear, mainly grasslands labelled as 'poor improved' (B6), semi-improved acid (BI2) and neutral grassland (B22), and bracken stands

A full listing of the habitat types assigned to these classes is provided in Appendix I.

Because of their narrow width, hedges are not consistently observable in the imagery used in this analysis, and so are not included in our estimate of semi-natural habitat.

[1] Blackstock, T.H.; Howe, E.A.; Stevens, J.P.; Burrows, C.R.; and Jones, P.S. 2010 Habitats of Wales: A Comprehensive Field Survey, 1979-1997. University of Wales Press, Cardiff.

Step 2: Analysing satellite imagery from 2016 and 2017 to:

- check for losses from the SN class
- allocate a subset of areas in the PI class to a new 'candidate SN' class

(A fuller description of this remote sensing analysis is provided in Appendix II.)

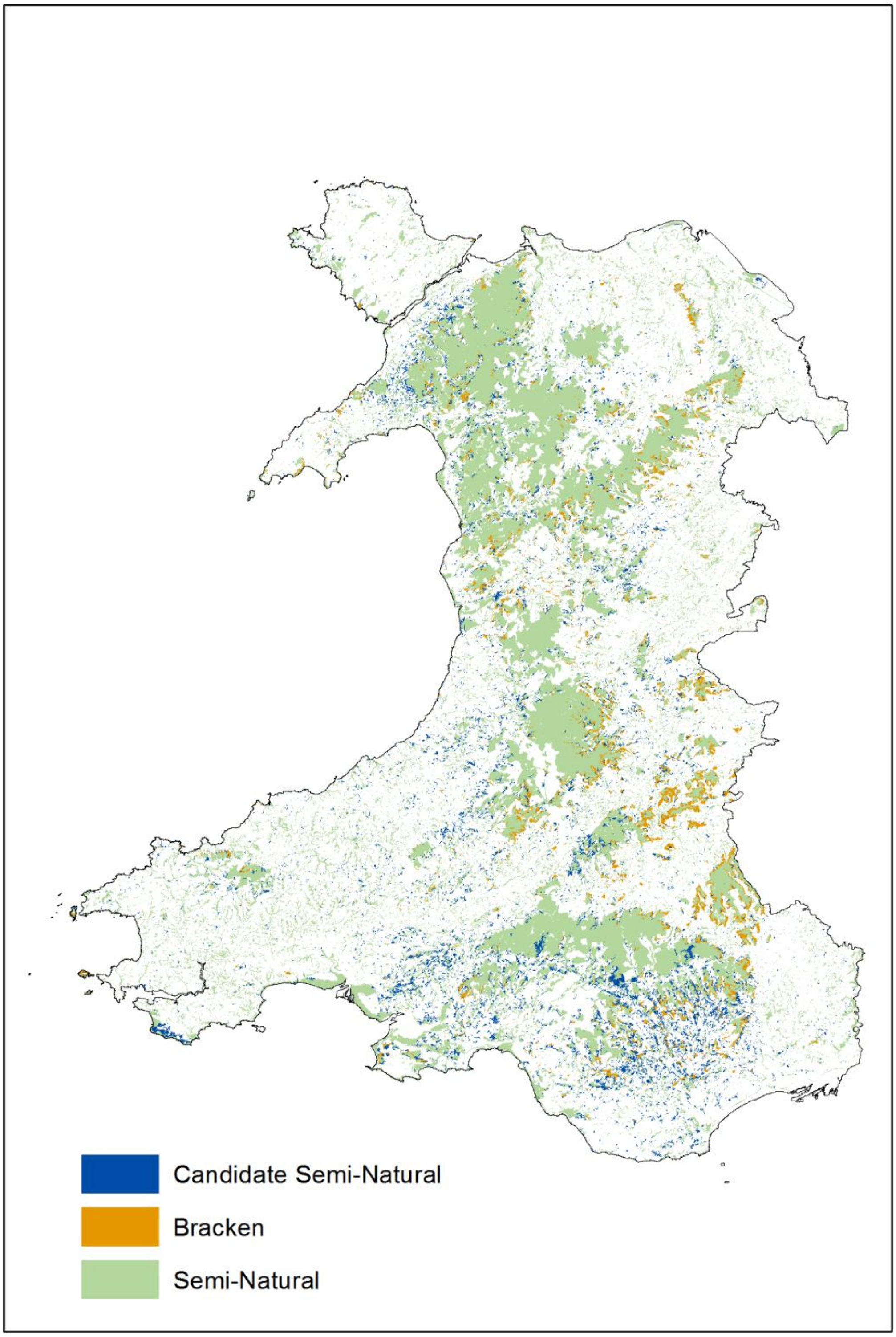
Step 3: Using evidence from the Integrated Administration and Control System (IACS) record – the crop type record from the system for managing subsidy payments to farmers - to check the accuracy with which agriculturally improved land was identified, and to refine the reallocation of PI to SN and I classes.

Results

The results of the analysis are summarised below. Several permutations are possible, but we have chosen to use the most expansive version (named as 'Total 1-3' in the table) as our estimate. This gives us a figure of 640,827ha or 31% of the land area of Wales.

	Class	Area (ha)	% land area (HWM)
1	Candidate semi natural	69,780	3.4
2	Bracken	47,569	2.3
3	Semi natural	523,478	25.2
	Total 1-3	640,827	30.8
	Total 2 and 3	571,047	27.5
	Total 1 and 3	593,257	28.6
	(Wales land area)	2,077,828	-

A corresponding map of the classes is shown on the next page.



Candidate Semi-Natural
Bracken
Semi-Natural

Interpretation

The estimate provided by this method is inevitably different from the previously published one. It is not directly comparable to that estimate, given that it has been produced using different methods. Instead it represents a new baseline. Critically, the method presented here is one designed to be repeatable in the future to provide comparable figures with that presented in this note.

Of all habitats, grassland perhaps more than any other has presented us, in past exercises, with difficulties in drawing a consistent line between what is semi-natural and what is not. This is seen in the divergent estimates of the extent of grassland types made by different survey programmes (see, for example, Countryside Survey: Wales Results from 2007[2]). The method used in the analysis described in this note is free from the effects of observer variation that are thought to lie behind those differences, and with suitable calibration to account for annual variation, is considered to offer a credible approach for consistent demarcation between types.

In drawing the line using the Normalised Difference Vegetation Index (NDVI), we have an option to adopt a more generous interpretation of what we might consider to be semi-natural habitat, bringing in, for example, those semi-improved grasslands whose NDVI values are more typical of semi-natural vegetation than improved vegetation. Rather than offering a solely floristic perspective on vegetation quality, the approach here – using a measure of productivity – provides a means of including, in addition to those floristically richer stands, some areas of land that perhaps have the potential to function more like semi-natural habitat than habitats that have been subject to intensive agricultural improvement.

These areas of what we term 'candidate' semi-natural habitat might be viewed as presenting opportunities for exploring the enhancement of ecosystem resilience, given their frequent proximity to established areas of semi-natural habitat.

Further notes on the relationship between vegetation, management and NDVI are provided in Appendix III.

It is inevitable that there will be some amount of misclassification in our analysis. However, the distribution of candidate SN areas so often adjacent to extant SN habitat suggests the results are plausible in many instances. The method was developed, and the mapping produced, over a short period of time, allowing only a minimal amount of checking in the field. This serves to reinforce the earlier statement, that the work be considered as developmental and the estimate an 'experimental statistic'. If resources allow, we recommend field checking of areas classified as 'candidate' SN.

[2] Smart, S.M.; Allen, D.; Murphy, J.; Carey, P.D.; Emmett, B.A.; Reynolds, B.; Simpson, I.C.; Evans, R.A.; Skates, J.; Scott, W.A.; Maskell, L.C.; Norton, L.R.; Rossall, M.J.; and Wood, C. 2009 Countryside Survey: Wales Results from 2007. NERC/Centre for Ecology & Hydrology, Welsh Assembly Government, Countryside Council for Wales, 94pp. (CEH Project Number: C03259).

APPENDIX I

Classes	Phase 1 habitats	
Semi-natural (SN)	A1.1*	Broad-leaved woodland
	A1.2.1	Semi-natural conifer woodlands
	A1.3.1	Semi-natural mixed woodlands
	A2*	Scrub
	B1.1	Unimproved acid grassland
	B2.1	Unimproved neutral grassland
	B3*	All calcareous grasslands
	B5*	Marshy grassland
	C2	Upland species-rich ledges
	C3*	Tall herb and fern
	D*	Heathland
	E*	Mire
	F*	Swamp
	G*	Open water
	H*	Intertidal
	I1*	Natural rock exposure & waste
	J2*	Hedges (but not included on final map)
Improved (I)	A1.2.2	Conifer plantations
	A1.3.2	Mixed plantations
	A4*	Recently felled woodlands
	B4	Improved grassland
	I2*	Artificial rock exposure & waste
	J1*	Miscellaneous (includes arable)
	J3*	Built-up areas
	J4	Bare ground
Possibly Improved (PI)	A1.3	Mixed woodland
	A3*	Parklands
	B1.2	Semi-improved grasslands, acid
	B2.2	Semi-improved grasslands, neutral
	B6	Poor semi-improved grassland
	C1*	Bracken
	J5	Other

* = all sub-types

APPENDIX II: ANALYSIS METHOD

Satellite Imagery used in the analysis

The analysis makes use of the European Space Agency's Sentinel 2 (S2) satellite imagery and, in particular, the NDVI (normalised difference vegetation index) metric that can be derived from it.

Sentinel 2 satellite imagery is supplied as 100 x 100 km data 'granules', so for convenience Wales is divided into 8 corresponding project areas (Figure 1). These are given two-letter MGRS (military grid reference system) codes. As Figure 1 shows, most of Wales is covered by two granules, VD and VC.

These projects vary in size and overlap with neighbouring projects. Dividing up Wales leads to a more efficient analysis, reducing the amount of sea that needs to be included in the rectangular raster images and allowing comparisons between projects which have different dates of satellite imagery.

Each project has a separate analysis, and some use a slightly different combination of Sentinel satellite images.

The analysis-ready version of S2 imagery used comes from the DEFRA 'Alpha' project and is atmospherically corrected and re-projected to fit the Ordnance Survey grid (EPSG: 27700). All data processing was carried out using this co-ordinate reference system.

Identification of objects

For each project, a set of objects was created by image segmentation using eCognition software and a spring S2 image, either from 20 April 2016 or 8 April 2017. These objects are the smallest units in the analysis. Each object is a group of pixels from the image (individual pixel size: 10m x 10m). In the analysis, the objects are edited to 'remove' marginal pixels to minimise confusing the properties of open fields with their boundaries (e.g. hedges). Pixels are not in fact deleted, just given an object ID of zero in data tables.

About NDVI

NDVI is the normalised difference vegetation index. This tells us, uniquely, about the photosynthetic activity of vegetation. It is one of the oldest (the earliest reference is 1974) and best known indexes used in remote sensing. While it has some problems with saturation at high values and a dependence on sun angle, these dependencies are well known and it has proved to be robust and reliable over 40 years.

NDVI simply contrasts the amount of reflected red and near infra-red light. This gives an index of the fraction of absorbed photosynthetically active radiation, which is related to plant productivity (if we monitor this over time) and plant health. Values close to 0 suggest low vegetation cover, little photosynthetic activity or poor health; values close to 1 suggest high photosynthetic activity, high cover and biomass and healthy vegetation. No other types of surface modify the light in this way, though all surfaces that reflect light can have an NDVI index which may be close to zero or, in the case of water, have negative values.

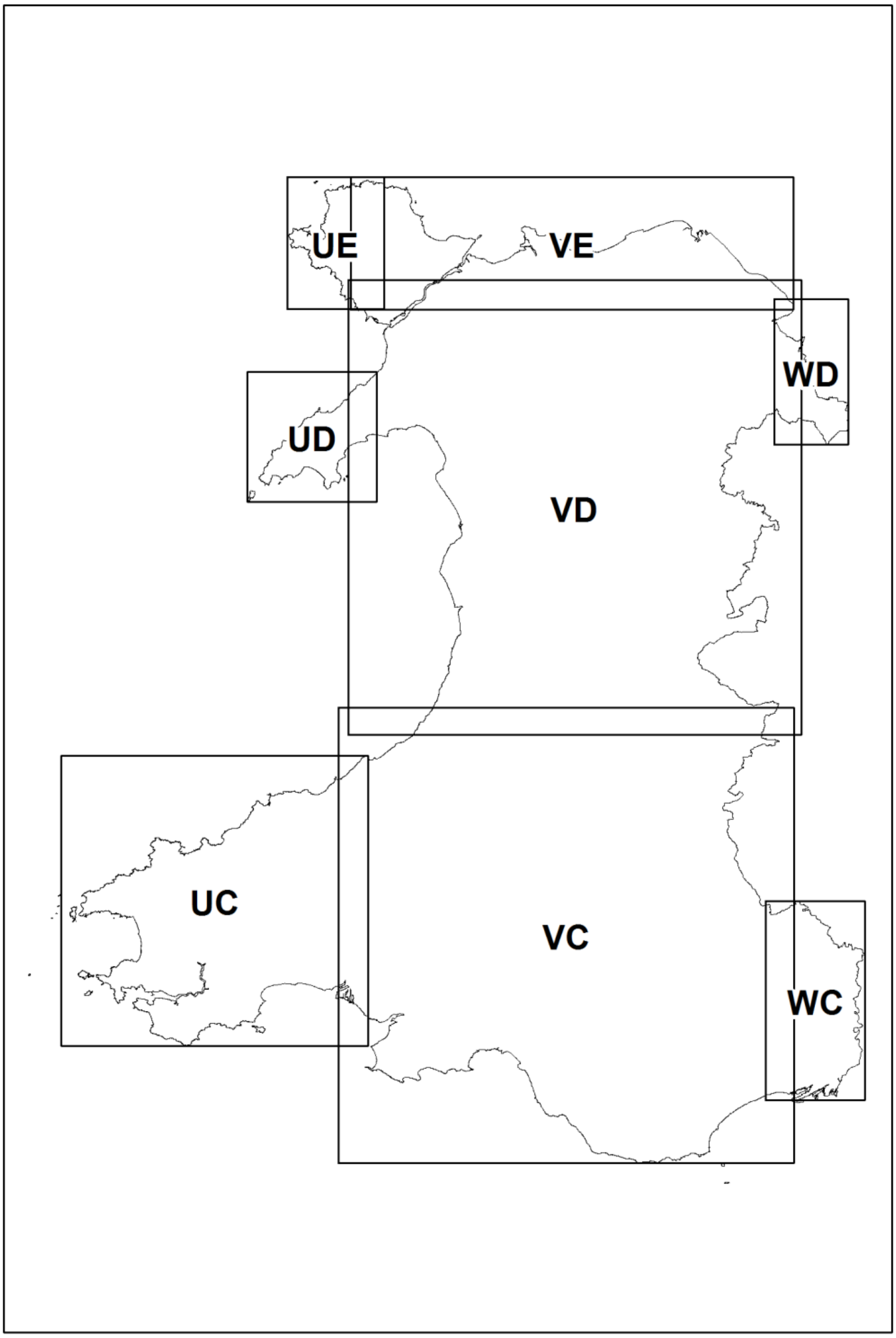


Figure 1. The project areas, corresponding to S2 data granules, that cover Wales.

Data compilation for the objects

Layers of data from the Phase 1 map, a digital elevation model (height, slope, aspect) and satellite images are ‘stacked’ on top of a raster version of the objects, using raster functions in the R software environment. These functions calculate statistics for every pixel in the stack which corresponds with each object. For example, if an object was converted to 45 pixels once marginal pixels have been removed, the height of each pixel in the elevation model is looked up and the 45 heights are averaged for this object.

The result is a data table with every object forming a row, every attribute a column. These tables include not only at-surface reflectance of green, red and near infra-red light, but the amount of Semi-Natural (SN), possibly improved (PI) and the bracken component of PI, taken from the Phase 1 map. They also have NDVI values which show the photosynthetic activity of the vegetation cover in spring (March and April), a time at which semi-natural vegetation contrasts with the early 'greening-up' of agricultural grasses in improved grasslands. A combination of images from 2016 and 2017 has been used, though not all images are available for all projects. In vegetated areas, NDVI varies from 0 to 1, with values closer to 1 showing more absorption of red light in photosynthesis.

These attribute tables are then joined to the spatial data in the vector version of the set of objects to give a table that can be viewed in GIS.

Analytical approach

A key analytical challenge in this work was to use evidence from satellite imagery (principally NDVI), taken from objects of known class (SN, I), to allocate objects of PI class to one of these two other classes. Apart from mixed woodlands, these objects of PI class are all grasslands or bracken.

To differentiate between classes, we used a Bayesian approach to explore relationships between NDVI values and class membership and to identify thresholds that can be used to assigned objects in the PI class to either the SN or I class.

To apply this method, we needed to define two components: a ‘prior’ probability which represents our starting belief about the most likely class, and a ‘likelihood function’ which tells us how to modify this belief, using evidence from the histogram frequencies.

As we have no existing information to start with, following Laplace, our prior belief must be 0.5 for each of two classes: SN and I. For a probability function, this is the way we say ‘we don’t know’ or are indifferent.

We used the density distribution of NDVI values (Figure 2) to derive a likelihood function. This is a normalised version of counts brought together on the same frequency scale by making the total number of objects add up to 100. For project VC, the density of SN and Improved objects is roughly the same around NDVI = 0.68. Below this value, the likelihood function favours SN.

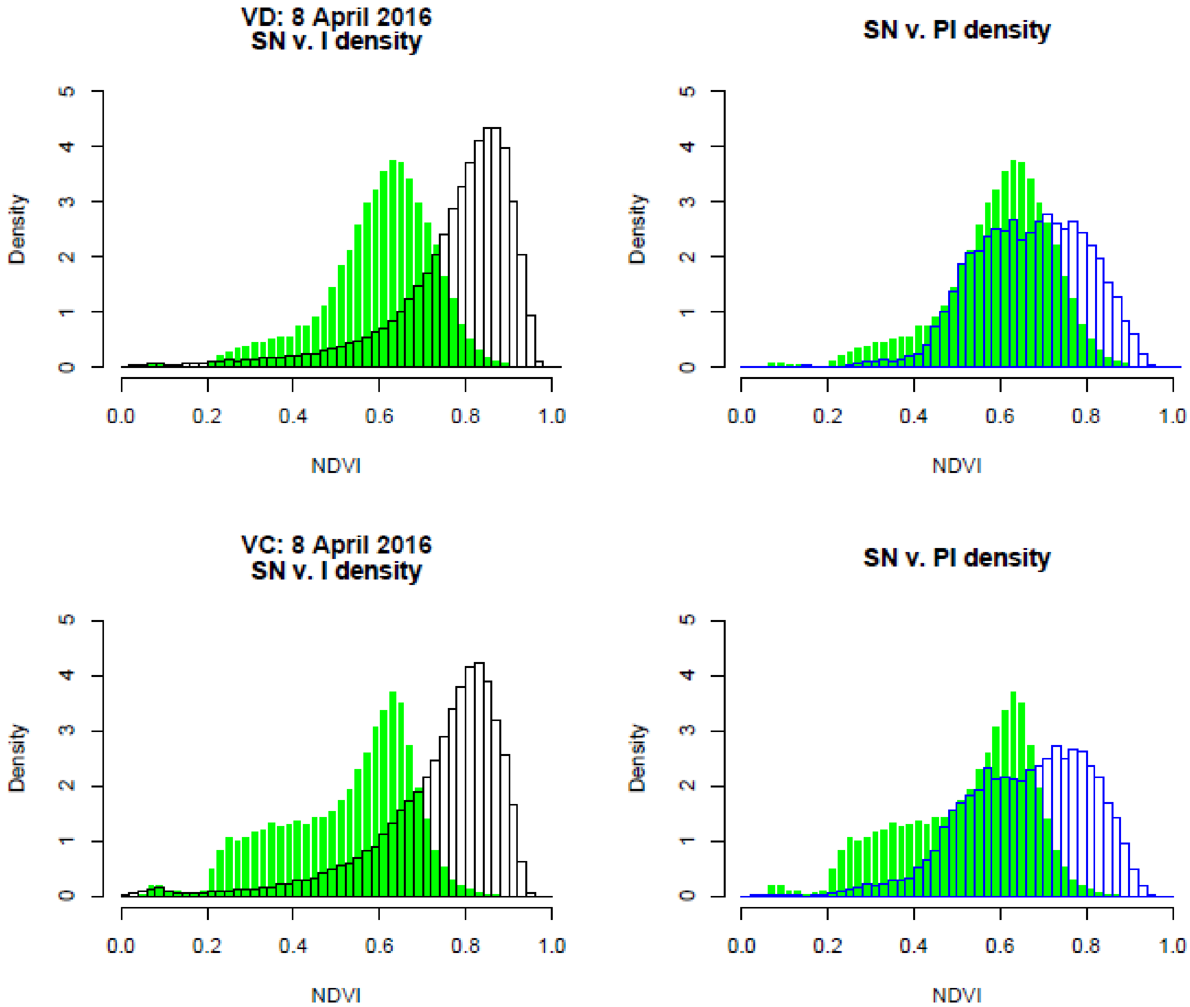


Figure 2. Density distribution for, in the left-hand graphs, SN (green) and I (black) classes and, in the right-hand graphs, SN and PI (blue) classes.

Figure 3 illustrates the results for PI objects of applying the SN:I likelihood function and prior probability to the NDVI values calculated from imagery taken on 8 April 2017. The likelihood function classifies just over 53% of PI objects as candidate semi-natural objects. The top row shows the distribution of NDVI values for objects labelled as SN and PI. The bottom row shows NDVI for objects labelled as Improved. The NDVI values have been converted to percentages. The middle row shows the probabilities for each NDVI value, with a horizontal line at 0.5. Probabilities above this line classify objects as SN; below this line as Improved. To emphasise this, the truth value, {0,1} is shown as red circles. A vertical red line shows the threshold below which objects with a lower NDVI value are classified as SN. The same thresholds are superimposed on the distributions in the top and bottom rows. The percentage 'assigned to class SN' is the area under the curve in the top row, between the two thresholds.

As this diagram illustrates, the statistical classification gives an upper threshold value, below which objects labelled as PI can be regarded as candidate semi-natural habitats. These thresholds were used in GIS queries to produce a map of candidate semi-natural habitats.

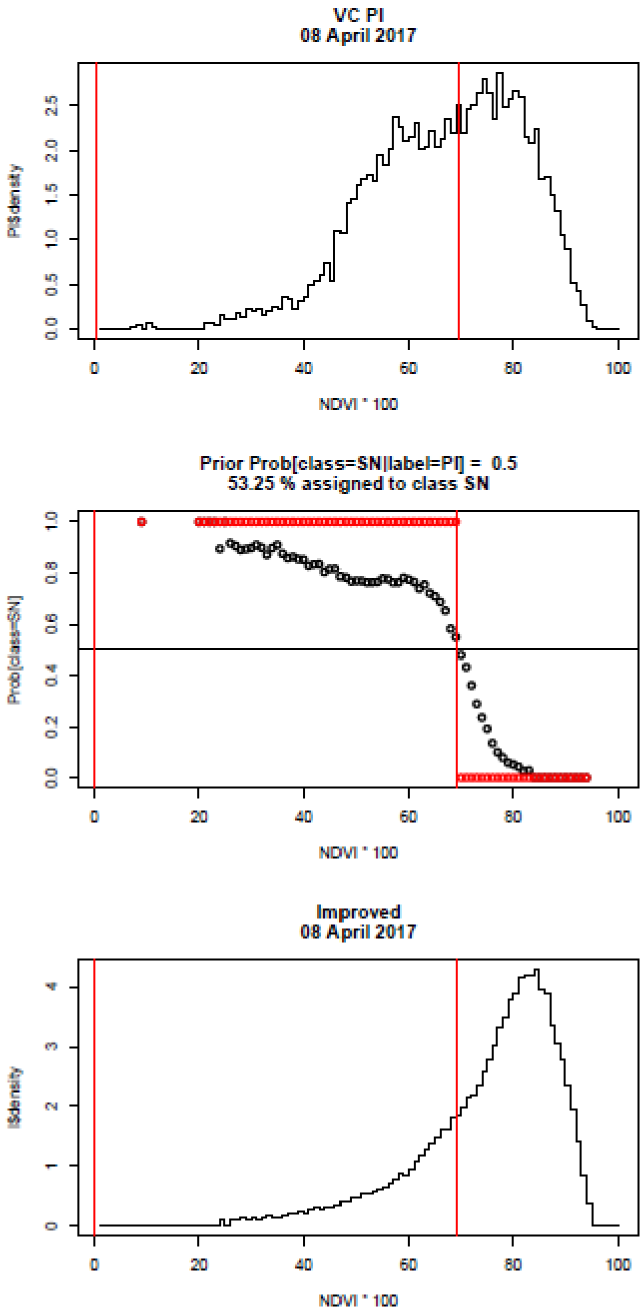


Figure 3. The classification of PI objects, using the SN:I likelihood function.

The rule for classifying PI as candidate SN or Improved depends on a Bayesian analysis, giving a probability, followed by a truth function assigning the object to the most likely of the two classes. Ideally we want to combine inferences from more than one image because this gives us more confidence in the result. Some images may have incomplete evidence because of partial cloud cover, while images from different times of year can help confirm the classification. In particular, we need to combine key spring images to test whether the NDVI of PI objects falls below the upper threshold for SN objects. Since some low-NDVI objects are urban or infrastructure, we also need to confirm that later in the summer the NDVI shows productive vegetation, in which case we need to check against a lower threshold value. A further check was made against legacy IACS data to rule out any objects which had been recorded as crops or re-seeded between 1993 and 2016

This combined inference is achieved through the use of Boolean logical functions. These automatically allow for errors and lack of evidence (defaults) from clouds and haze, where some tests can only be failed, not passed.

The default search rule for candidate SN, applied to data tables, is:

SN < 5 AND	Not already counted as SN
(SN + PI) > 5 AND	Overlap with SN and PI classes more than 50%
NDVI0420 < 0.64 AND	Low spring photosynthetic activity
NDVI0408 < 0.69 AND	
NDVI0314 < 0.59 AND	
NDVI0719 > 0.6 AND	High summer photosynthetic activity
IACScrop = 0	No reported agricultural land use in IACS (except as permanent grassland)

Appendix III

Habitat management and agricultural improvement

The aim of agricultural improvement is to exploit the vegetation and harvest the products of photosynthesis, either as a crop or through stock grazing. This is most beneficial on the better lowland soils, where climax forests can be replaced by earlier successional grasslands and arable crops. Most of the very large areas of semi-natural habitat in Wales are left in the uplands and along the coast. In the lowlands there are very many small stands of woodlands, relatively unimproved grasslands and patches of mire vegetation on wet soils. We know that very little habitat in Wales is completely natural, perhaps only some stands of coastal heaths and vegetation on inaccessible cliff ledges. Everywhere else is modified by agriculture or woodland management: for example, extensive sheep-grazing on the uplands suppresses native tree growth. However, here we are looking for semi-natural rather than wholly natural habitats. This can be defined as habitats in which the vegetation over is dominated by native species and which have organically functioning soils.

We can draw a contrast between the extensive agricultural management of semi-natural habitats and intensive agricultural improvement. Sheep can be allowed to graze on a mountain without any other attempt to modify the soil conditions. By contrast, the farmer might drain, lime and fertilise the lower slopes of the mountain to convert grasslands and heaths dominated by native species into agriculturally improved grasslands. These have a high cover of non-native grasses (such as *Lolium perenne* and *Cynosurus cristatus*) and cultivars which take advantage of the higher nutrient status and tend to out-compete the native species. With sustained effort, the soils can be permanently modified and only very slowly revert to their previous nutrient status if the artificial management is given up.

In a similar way, areas of semi-natural woodland can be converted to improved grasslands. Other losses of semi-natural heaths, bogs and grasslands can take place, notably to conifer plantations with the loss of native species cover. These conversions are dramatic, involving a high investment of management activity and may take a long time to reverse. Other interventions are less damaging. Larch, which is deciduous, may have less disastrous effects on the ground flora, though for the purposes of this mapping still count as silvicultural improvement. Heathland cover can also be lost as a result of small-patch burning or accidental large fires, but this is temporary with no replacement by non-native species and will not be counted as loss of semi-natural habitat.

NDVI and the loss of semi-natural habitat

Ecological theory tells us that plants compete for light, and that — within the limits of available soil nutrients, water and light — the most competitive plants will come to dominate the habitat. While there are other competitive strategies, as a general principle this suggests that the undisturbed natural vegetation ought to maximise its use of photosynthetically available radiation. In Wales, natural succession would mostly result in deciduous woodlands where an increasing proportion of captured sunlight is converted and stored in building plant structures. Of course, what is achievable locally on nutrient-poor acid soils and in cooler upland climates will be less than what is possible on nutrient-rich soils in the lowlands.

Contrasts in the NDVI vegetation index convey signals of agricultural (or silvicultural) intensification and improvement. In grasslands, we expect an increase in productivity with a correspondingly higher NDVI. If there is no better conversion of sunlight into energy, the improvement has been unsuccessful. In theory, we would measure this using a 'dense time series' of NDVI measurements, giving a measure of the productivity over the growing season. While future analyses may help track vegetation growth through radar imagery, the optical instruments which are blocked by cloud typically do not give us a complete time series even with Sentinel 2 satellites passing over some parts of Wales every 3 days. There is also a problem with the saturation of the NDVI index in the best part of the summer growing season, during which most grasslands and woodlands have values close to 1.

If a time series is unavailable, we can focus (as we have in this analysis) on the spring season. In order to get increased productivity, agriculturally improved grasslands will 'green up' much earlier in the year than those dominated by native species. The higher yield is partly the result of these improved grasslands growing almost all year round. This suggests that the conversion of semi-natural, unimproved grasslands to agriculturally improved grasslands will be shown by a contrast in the NDVI during the spring growing season. Not only do we expect a difference, but a recently converted grassland ought to show a stronger contrast with before, more than other grasslands that may have been converted some while ago and now need maintaining. Similar strong contrasts should be seen in the conversion of deciduous woodland to improved grasslands.

We can also say something about where we expect conversion to take place. For example, agricultural improvement of a small, isolated area of upland acid grassland distant from the enclosed fields is much less plausible than the improvement of a large, regularly shaped patch of land next to enclosed fields or other recently improved upland takes. These areas will typically (but not always) have straight boundaries on the image corresponding with post and wire fence lines on the ground. Some sort of access track and gateways may be visible on recent air photos.

The conversion of upland grasslands and heathlands to plantations will eventually be seen as obvious growth in tree cover. Mature conifer forests have a high NDVI even in winter, contrasting very clearly with the low NDVI of deciduous woodlands before the tree canopy comes into leaf. The early stages of plantation will depend on the groundworks, but would typically show a decrease in NDVI as vegetation ground cover is lost. It could be argued that if there is no difference in ground cover, the habitat is still semi-natural until the tree crop starts to become dominant.

Generally speaking, any agricultural improvement will lead to an increase in NDVI, greater than the semi-natural habitat being replaced. Often this contrast will be a step change, for example clearly absent in the spring of one year, clearly showing in the next year.

Conversion of semi-natural habitats to an arable crop rotation, such as a cereal crop or *Lolium* ley, will have a different seasonal pattern of NDVI depending on the crop. Spring cereals will have a very low winter and spring NDVI as the crop cover is very low, rising as the crop greens up and puts on biomass, then dropping again as the crop ripens and transfers energy from green photosynthetic tissues into grain. The NDVI of grass leys drops rapidly as they are cut for silage, while intensively grazed improved fields may have a constant but low NDVI.

Bracken canopies can act as a woodland surrogate, and on the upland fringes typically occupy unimproved soils. If bracken stands are converted to improved fields, the contrast in NDVI is most obvious in the spring, when the bracken litter has a very low NDVI, rather than the summer when the bracken canopy has a high NDVI similar to a productive improved grassland.