**Land sparing to make space for species dependent on natural habitats and High Nature Value farmland**

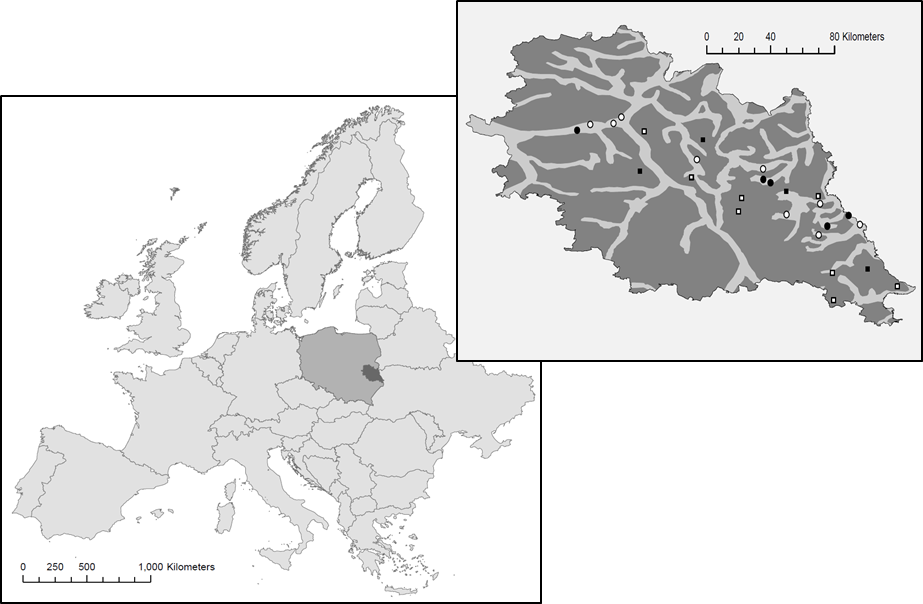
**Claire Feniuk, Andrew Balmford and Rhys E. Green**

**ELECTRONIC SUPPLEMENTARY MATERIAL**

1. **Study region and sites**

The study region covered an area of c.14,000 km2 of the western edge of the Polesian lowlands in the Lubelskie region of eastern Poland (figure S1). Soils associated with wetlands (histosols, fluvisols & gleysols – hereafter called ‘wetland soils’) cover 3,473 km2 (25% of the study region), mostly in low-lying areas and river valleys, while soils likely to be found in areas formerly with forest cover (podzols, cambisols, arenosols, luvisols and albeluvisols – hereafter called ‘forest soils’), cover 10,444 km2 (75%) [1]. Within the ‘cultivable area’ of the region (which excludes areas unsuitable for agriculture e.g. urban land), principal land uses include permanent arable land (54%), mosaics of mixed agriculture (16%), and grassland (13% including meadows, pastures and some natural floodplains). 16% of the region is classified as forest (natural and managed), and 2% as wetlands (marshes and peatbogs) [2].

Twenty-six 1 km x 1 km study sites were selected: 12 on forest soils and 14 on wetland soils (figure S1b). Sites were selected to cover a gradient of agricultural use. Zero-yielding natural habitat sites, all of which were within protected areas, comprised mixed/deciduous forests on forest soils (n=4), and floodplains and fen mires on wetland soils (n=5). Seventeen of the 26 study sites included farmed land. Farmed sites were chosen on both soil types to include a range from low-yielding through to high-yielding agricultural land (see tables S1, S2). Within soil types, sites were broadly matched in terms of altitude, topography and climate.



**Figure S1.** Map of the study area (dark grey), situated within the Lubelskie region of eastern Poland. Inset map shows the distribution of ‘forest soils’ (podzols, cambisols, arenosols, luvisols and albeluvisols – shown in dark grey) and ‘wetland soils’ (histosols, fluvisols and gleysols – light grey) within the study area. Study sites on forest and wetland soils are shown by squares and circles respectively. Filled symbols indicate zero-yielding natural habitat sites, and un-filled symbols indicate farmed sites covering a gradient of agricultural yields.

**Table S1.** Overview of main habitat types surveyed. Land cover statistics used for site selection were based on the CORINE land cover data [2].

|  |  |  |
| --- | --- | --- |
| Land-use type | No. of sites/  soil types | Description  (with CORINE land cover statistics where applicable) |
| Mixed/ deciduous forests  (natural habitat) | 4  Forest  Only | Mixed/deciduous forests are largely restricted to Nature Reserves. Site selection focussed on old-growth reserves larger than 1 km2 located within a larger forest complex (often lower quality plantation forest). Study sites were placed centrally within the reserve to maximise the distance to nearest adjacent land use, and sites where this distance was <100 m were rejected (typically this buffer was much larger). |
| Fen mires  (natural habitat) | 3  Wetland only | Site selection focussed on protected wetlands within National Parks, comprising of complexes of fens, transition mires and springs. Within these areas, sites were placed centrally within the areas of intact natural fen mire habitats in order to maximise buffer distances. These fen mires also contain natural hummocks of drier grassland and scrub, and small patches of birch/alder woodland. |
| Floodplains  (natural habitat) | 2  Wetland only | Sites were located within two of the region’s major river valleys: the middle course of the Bug River, and the lower course of the Wierpz River. Although a large proportion of both valleys are now farmed, study sites were placed within restricted areas of natural humid grassland and sedge meadows that are not managed for agriculture. Sites were placed centrally in the largest possible area of contiguous natural habitat in each valley. |
| Hay meadows  & pastures | 4  Wetland only | Sites comprised of large areas of grass meadows with clear indicators of agricultural management (annual mowing, maintained drainage channels). Such large contiguous areas of hay meadows were largely restricted to wetland soils.  Land use statistics of sites (CORINE):  Pastures >95% |
| Mixed farm  mosaics  (low/mid-yield) | 8  Forest: 6  Wetland: 2 | Sites comprised of mosaics of mixed farmland (arable, meadow, pasture) plus areas of non-crop habitat, including small patches of woodland/scrub and wet features (streams, ponds).  Major crop plants included oats, rye, triticale and potatoes.  Land use statistics of sites (CORINE):  Heterogeneous agriculture + pastures >50%  Total agricultural area >95% |
| Arable  Farmland | 5  Forest: 2  Wetland: 3 | Sites primarily comprised of arable land with no significant areas of non-crop habitat.  Major crops include wheat, barley, oilseed rape and maize.  Land use statistics of sites (CORINE):  Non-irrigated arable land >75%  Total agricultural area >95% |

**Table S2**. Summary of variation in yields of major crops (harvested tonnes ha-1 yr-1) in our mixed-farming and arable sites, compared with the wider study region, similar EU New Member States and the UK.

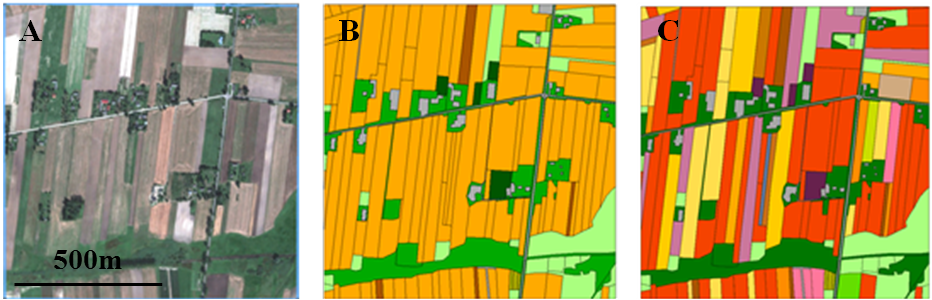
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mosaic/Arable Study Sites | | |  |  |  |
| Crop | Min. | Max. | Mean | Lubelskie region | EU Ref. | UK |
| Wheat | 1 | 5.9 | 3.7 | 3.7 | 4.5 | 7.4 |
| Barley | 2.5 | 4.8 | 3.8 | 3.2 | 3.7 | 5.8 |
| Oats | 1.7 | 4 | 2.9 | 2.7 | 2.5 | 5.5 |
| Rye | 1.8 | 3 | 2.6 | 2.5 | 3 | 5.6 |
| Triticale | 1.7 | 5 | 3.4 | 2.9 | 3.5 | 3.9 |
| Oilseed rape | 1.5 | 4 | 3.1 | 2.5 | 2.6 | 3 |

Yields foreight ‘mosaic’ study sites and five ‘arable’ study sites, compared with average yields for the Lublin region from ref. [3]. ‘EU Ref.’ refers to Poland, Hungary, Lithuania, Slovakia and Slovenia, and the numbers presented are the mean of national average yields in 2013 [4]. UK cereal yields for 2013 are also included for reference [5,6].

1. **Measuring agricultural yield**

**Data collection**

Of the 26 study sites, nine were zero-yielding natural habitat sites. The remaining 17 sites supported some level of agriculture. For each agricultural site a detailed field boundary map was produced in a GIS using a combination of satellite imagery and cadastral data. Crops were then mapped within each field by visiting the whole site (figure S2). In order to capture data on crop-rotation, crop mapping was completed in both 2013 and 2014. The extent and type of non-crop areas (including semi-natural habitat patches, abandoned land, roads and buildings) were also mapped.



**Figure S2.** An example showing the process for mapping crops within each study site. Based on satellite imagery (A) and aided by cadastral data, an approximate land cover map was drawn, including field boundaries (B). Detailed crop mapping was then carried through site visits in both 2013 and 2014 (C). From this, the total area under each crop in each site was calculated. Colours represent different land uses (B) and crop types (C).

Local yields were estimated from a combination of interviews with farmers, data submitted by agricultural advisors, and agricultural statistics published by the regional government. A total of 70 farmers were interviewed across all farmed sites. The complex pattern of land ownership meant that locating and interviewing farmers involved in the management of specific fields within some sites was challenging, so interviews with farmers based in the nearest village and farming locally (within 2 km) were also accepted. A minimum of two farmers were interviewed per site, with the exception of one site where only a single farmer was available. Information was gathered on crop area, harvest quantities.

In order to supplement and validate the information obtained through interviews, additional data were obtained from two further sources: estimates of average crop yields at a county level obtained from local agricultural advisors; and village-, county- and region-level agricultural statistics reported through the Central Statistical Office of Poland (Główny Urząd Statystyczny, GUS) via the Online Database of the Statistical Office of Lublin [3]. Where reliable local estimates were available from the farmer interview data they were used. Where reliable data were lacking, reported data from agricultural advisors and/or GUS were used. This accounted for less than 13% of the total surveyed area.

**Calculating food energy yield**

Using these data, annual energy production for each site was calculated as follows:

where *j* refers to the individual study site, *i* refers to each crop, *P* is the total energy production (in GJ yr-1), *A* is the cropped area (ha) obtained from crop mapping, *X* is the crop yield (t ha-1 yr-1) primarily from interview data, *E* is the proportion of each harvested crop that is edible (between 0 and 1, see table S3) and *G* is the energy content per unit weight of processed crop (GJ t-1, based on ref. [7]. To convert the total energy production (in GJ yr-1) into a measure of site-level yield (GJ ha-1 yr-1), total production of each site was divided by the area of the study site (100ha). This provides a landscape-scale measure of agricultural yield, reflecting the proportion of the site which is crop vs non-crop habitat, the types of crops grown within the site, and the in-field yield of those crops.

**Table S3.** Edible portions and energy content of edible mass for main crops.

|  |  |  |  |
| --- | --- | --- | --- |
| Crop | Energy content  (GJ t-1) | Edible Portion | Data Source |
| Wheat | 14.28 | 0.78 | [8] |
| Barley | 14.82 | 0.78 | [8] |
| Rye | 14.15 | 0.78 | [8] |
| Oats | 16.29 | 0.78 | [8] |
| Triticale | 14.07 | 0.78 | [8] |
| Maize | 15.28 | 0.79 | [8] |
| Buckwheat | 14.36 | 0.79 | [8]\*1 |
| Oilseed rape | 37.01 | 0.9 | [9] |
| Mustard | 37.01 | 0.9 | [9] |
| Potato | 3.22 | 0.9 | [8] |
| Sugar beet | 1.8 | 0.88 | [9] |
| Pumpkin | 1.09 | 0.7 | [10] |
| Cabbage | 1.05 | 0.8 | [10] |
| Onions | 1.67 | 0.9 | [8]\*2 |
| Carrots | 1.72 | 0.9 | [8]\*2 |
| Beetroot | 1.8 | 0.9 | [8]\*2 |
| Peas | 3.39 | 0.9 | [8]\*2 |
| Beans | 1.3 | 0.9 | [8]\*2 |
| Market garden | 1.72 | 0.9 | [8] |

Where data on edible portion of crops were not available (as indicated by asterisks), data for similar crops were used as surrogates as follows: (\*1) sorghum; (\*2) root vegetables. Energy content for land classified as a ‘market garden’ was calculated based on an average market garden, assuming equal proportions of the main crop types observed in this land use (potatoes, beans, onions, carrots, cabbages and strawberries). Energy content is based on ref. [7].

The inter-annual variation in production of each site due to crop rotation was relatively small, and the ranking of sites (from lowest to highest) based on total production did not change between years. The limited effect of crop rotation on total production is most likely to due to the large number of small fields within most 1 km2 sites – hence although crops grown in some fields changed, the total area of the whole site that was under each crop did not alter dramatically between years. For subsequent analysis, total energy yield data were therefore averaged over the two-year period.

**Special cases**

Beef cows were the primary recipients of fodder crops grown in our study sites. A feed conversion ratio (FCR) was used to estimate the weight of added body mass produced per unit weight of fodder for both hay (FCR = 11.2) and silage systems (FCR = 5.9, based on ref. [11] and assuming no additional dietary supplements). We then assumed that 72% of the carcass weight was edible (including meat and offal) [7,12]. Therefore, energy yield of fodder crops were expressed as the energy yield of edible beef per ha of fodder crop: 0.77GJ ha-1yr-1 for hay, and 1.45GJ ha-1yr-1 for silage.

Some farms, particularly those in lower-yielding mosaics, had small areas for mixed fruit and vegetable production close to the farmhouse (‘market gardens’). As the areas under each crop were often very small, occasionally with co-cropping, it was difficult to get an accurate estimate of individual crop areas. Therefore, the most common crops found in such land use were identified (potatoes, beans, onions, carrots, cabbages and strawberries) and production was estimated assuming equal areas of each of these crops.

1. **Measuring population densities of bird species**

Birds were surveyed at each site in the breeding seasons of 2013 and 2014. Two daytime visits per site were completed each year: one early visit (mid-April to late-May) and one late visit (early June to mid-July). Two parallel 1-km transects were walked, traversing the study site 500 m apart and 250 m from the nearest square edge, starting within 30 minutes of sunrise, and ending within 4 hours. Perpendicular distances from the transect line to the point of first detection of each bird were estimated and checked using a laser range finder. In addition to this, a single dusk survey was also completed at each site in order to capture data on crepuscular and nocturnal species, including Aquatic Warbler, *Acrocephalus paludicola*, and Great Snipe, *Gallinago media*. Evening surveys took place in the middle of the breeding season in 2013 (from mid-May to mid-June), with the aim of collecting data during the period of peak breeding activity. These surveys followed the same transects as the dawn surveys and began at one hour before sunset, continuing for ~2 hours after sunset – and so aimed to capture activity of both evening- and night-active birds.

Seven species of swifts, swallows, martins and terns were excluded from the analyses as it is difficult to obtain reliable estimates of population density of largely aerial birds using Distance sampling techniques. In addition, data on the Common Pheasant, *Phasianus colchicus*, were also excluded as this is an introduced alien species within the study region.

Distance analysis was completed using the Multiple-Covariate Distance Sampling (MCDS) engine in the Mark-Recapture Distance Sampling (MRDS) package in R [13]. Distance sampling was used in order to allow for differential detectability of species in different habitats [14–16]. Study site was set as a sample, and transects within each site as replicates. Species-specific truncation distances were selected, as some species could be detected reliably to considerable distances (>100m, particularly large, conspicuous species in open habitats), whilst others could not. Species with more than 30 observations were fitted with a species-specific detection function. Species with fewer than 30 observations were assigned to a detectability group based on assessments of their diet, habitat stratum, activity levels, behavioural traits and common mode of detection. Where appropriate, data for species that were analysed individually were also included in the multi-species analysis. In total, 66 individual species and 17 multi-species groups were analysed.

A key difference likely to affect detectability between study sites is the density of trees, which was estimated for each site based on tree survey data (obtained from a total of 576 25 m x 25 m tree survey plots over the 26 study sites). This was included as a continuous covariate (a scale factor) in the Distance analysis so that the probability of detection is dependent on the distance to the individual or cluster and on the density of trees, generating a site-specific estimate of effective survey area or ‘Effective Strip Width’ (ESW). This differs from the approach taken in previous similar studies (e.g. ref. [10]), in which broad habitat type was included as a categorical covariate for species with >20 observations in each habitat type, producing a habitat-based (but not site-specific) effective survey area. For species analysed in multi-species groups, a categorical variable to represent species identity was included as a factor, alongside tree density.

A broad range of models were considered for fitting detection functions (table S4), based on those recommended in ref. [14]. For species with more than 30 observations, models with series expansions were included in order to improve the model fit. For analysis of multi-species groups, only the key function was included due to constraints within the MRDS package. Some models failed to converge, whilst others were excluded based on assessment of Goodness of Fit tests (Kolmogorov-Smirnov probability test, Cramer-von-Mises probability test and *χ*2 test) and visual assessment of model fit and plausibility. Rather than selecting a single fitted model based on AICc, estimates of ESW were averaged across all plausible models offering a good fit, based on AICc weight. This model-weighted average ESW was then used to calculate site-specific population densities in subsequent analyses.

**Table S4.** Models fitted for Distance analysis of bird data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data | Model | | | |
| Covariate(s) | Key function | Series expansion | Adjustment Order |
| Species with ≥30 observations:  single-species detection functions | Tree density | Half Normal | Null | Null |
| Cosine | 2 |
| 3 |
| Hermite polynomial | 4 |
| 6 |
| Tree density | Hazard Rate | Null | Null |
| Cosine | 2 |
| 3 |
| Simple polynomial | 2 |
| 4 |
| Species with  <30 observations:  multi-species groups | Tree density + Species (as factor) | Half Normal | Null | Null |
| Hazard Rate | Null | Null |

Some models for some species/groups were excluded based on a failure to converge, poor goodness of fit, or implausibility of model fit. Model predictions of ESW were then averaged across all plausible models offering a good fit, based on AICc weight.

1. **Measuring population densities of sedge species**

Sedge surveys were conducted in May – June 2014, which coincides with the flowering period for a majority of species in the region. Twenty 100 m transects were established perpendicular to the two 1-km bird transects, spaced 100 m apart and on alternating sides of those transects. Five 1m2 quadrats were placed at 20 m intervals along each 100 m transect (100 quadrats per survey square). Within each quadrat, all sedges were identified to species. Where possible, this was done in the field but where necessary, voucher specimens were taken for subsequent identification. Identification was based on refs [17–24]. The percentage cover of each sedge species in the quadrat was estimated by eye using 10 cm x 10 cm markers to assist with coverage estimation [25]. The average proportion of each survey square covered by a given sedge species was estimated by taking the mean of cover values for the 100 quadrats in the square.

**(e) Fitting density-yield curves**

Methods for curve-fitting followed ref. [26], but adapted here because the data derive from study sites on two different categories of soil type.

For data on birds derived from counts, we fitted Maximum-likelihood (M-L) univariate parametric regression models for each species with the count of clusters (groups of individuals) detected in a survey plot *n*, as the dependent variable and annual food energy production per unit area *x* as the independent variable. We included an offset term in all models of bird data to account for variation among plots in the effective survey area per visit *a* (derived from the Distance analysis for birds- see above), the number of survey visits *e* and the mean cluster size (individuals per cluster) for that species *s*. We did this because we wished to obtain a model of population density *y* (mean number of individual birds per unit area per survey visit) in relation to yield. The mean density of individuals in a plot is given by

*y*= (*n*\**s*) / (*a*\**e*),

and therefore

*y* = *n* / *v*,

where *v* = (*a* \* *e*) / *s).* By including loge(*v*) as an offset term in the Poisson regression model with *n* as the dependent variable and a logarithmic link function, our fitted models were made equivalent to models of population density in relation to yield. The formulae in table S5 represent the relationships we assumed to hold between population density y and yield *x* which represents annual food energy yield (in GJ ha-1 yr-1). The parameter estimates *b*0, *b*1, *b*2, *b*W and *α* were estimated from the data by the Poisson regression method described above and are models of population density *y*, rather than the count of clusters in a plot *n*, through the inclusion in the regression of the offset term. For example, for Model A in table S5, the expression shown for population density is

*y = exp(b*0 + *b*1 (*α*)).

It follows that

loge (*y*) = *b*0 + *b*1 (*α*).

Given that *y* = *n* / *v* (see above) and loge (*y*) = loge (*n*) - loge (*v*), substitution and rearranging gives

loge (*n*) = loge (*v*) + *b*0 + *b*1 (*α*),

which is the formulation of the Poisson model with a log link, cluster count per plot as the dependent variable and an offset.

We fitted four alternative formulations of the density-yield model (A-D), as detailed in table S5. This was done in order to account for any differences in the density-yield relationship on different soil types, models A and B were fitted to the forest (n=12) and wetland (n=14) datasets separately. Models C and D were fitted to the full dataset (forest and wetland data combined, n=26), but soil type was included in the model through the parameter *bW* and the binary variable *WET* (where *WET*=0 for sites on forest soils, *WET*=1 for sites on wetland soils). A special case of model A occurred when a species was only present in zero-yielding natural habitat plots and not in the farmed plots on one of the soil types. In such cases, the mean population density was assumed to be that observed in the natural habitat and the parameter *b0* of model Awas taken to be the natural logarithm of that mean density and parameters *b*1 and *α* (see Table S5) could not be estimated. The modelled density of such species in farmed habitats was zero.

**Table S5.** Univariate parametric models of population density *y* in relation to agricultural yield *x*.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | | Data | Number of fitted parameters |
| A | *y* = exp(*b*0 + *b*1 (*α*)) | Curves fitted for Forest and  Wetland data separately.  Sum of ln(*L*) or *RSS* (forest + wetland) used to calculate model AICc | 6  (3 Forest +  3 Wetland) |
| B | *y* = exp(*b*0 + *b*1 ( *α*) +*b*2 (2*α*)) | 8  (4 Forest +  4 Wetland) |
| C | *y* = exp(*b*0 + (*b*W\**WET*) + *b*1(*α*)) | Curves fitted to Forest and  Wetland data combined.  WET = 1 (Wetland data)  WET = 0 (Forest data) | 4 |
| D | *y* = exp(*b*0 + (*b*W\**WET*) + *b*1(*α*) +*b*2 (2*α*)) | 5 |

Maximum-likelihood Poisson (for count data) or Ordinary Least Squares (for extent data) univariate parametric regression models were fitted for each species with cluster count (birds) or cover (sedges) as the dependent variable and annual food energy production per unit area as the independent variable. The variable *x*represents yield (in GJ ha-1 yr-1) and *b*0, *b*1, *b*2, *bW*and *α* areparameters estimated from the data.

Maximum-Likelihood (M-L) estimates for the parameters were calculated assuming a Poisson error structure, using the Nelder-Mead numerical optimisation method from the R package ‘alabama’ [27]. These model formulations were chosen because they give a wide range of curve-shapes. Model B and D curves were often hump-shaped, with an asymmetrical shape. This asymmetry is well described by the inclusion of the shape parameter *α*. In all models, the value of *α* was constrained to be positive and not to exceed 4.6. This maximum value was used because for species with high *α*, the likelihood of the data was usually approximately constant with increasing *α* beyond this value, making a precise M-L model impossible to identify. The shape of the functions determined by models with high *α* varied little as *α* was changed. For such hump-shaped functions, the best-fitting model sometimes predicted a peak population density value that fell in a gap between sites in the distribution of the yield variable. Occasionally, the value of this prediction was much greater than the maximum observed density at any site. We considered such models to be unrealistic, and therefore constrained the model parameter values to predict peak densities no greater than 1.5 times the maximum observed density, following ref. [26].

For each site, we calculated the population density (individuals per unit area per survey) expected from the M-L model described above and multiplied it by the value of the area surveyed (*a*\**e*) for that square, giving the expected total number of individuals recorded at that site. We then used the observed count of individuals in each plot and this expected count of individuals to calculate, for each plot, the natural logarithm of the Poisson probability of the observed total number of individuals, given the expected number from the model. We thene summed this across all sites to give the log-likelihood (ln(*L*)) of the data. AIC adjusted for small sample size (AICc) was then calculated for each model. For models A and B, this was based on the sum of log-likelihoods for the fit to each separate dataset (forest and wetland), and the sum of the number of fitted parameters (see table S5). This allows direct comparison with models C and D, as the same data were being used overall, based on all 26 study sites. The best-fitting model was then selected based on AICc. Our assumption of a Poisson error distribution for the analysis of bird counts, when combined with the use of AICc for selecting among models, may have resulted in us being more likely to select models with more fitted parameters (e.g. Model B over Model A or Model D over Model C). That is because we did not attempt to account for overdispersion, which was technically difficult because many species in our dataset had sparse data. Overdispersion in numbers of individuals was present to some extent because clusters of more than one individual occurred for some species, but this did not have a large effect because mean cluster size was 1 or close to 1 for most species. The main reason for overdispersion was residual differences among plots in the density of clusters that were probably due to unmeasured environmental variables, in addition to yield. Despite this, we considered it preferable to base the bird models on counts, rather than to model densities using least-squares methods. This was because the use of Poisson models renders it straightforward to allow for the sample size of individuals counted in each plot and because the use of the logarithmic link function precludes the prediction of negative densities.

To derive weighted mean density-yield curves representative of the whole study region (which comprises 75% forest soils and 25% wetland soils), for each species the estimates of its population density on forest and wetland soils across the yield gradient were multiplied by 0.75 (forest) and 0.25 (wetland) respectively and then summed. This effectively produced as composite density-yield curve for each species within the study region, reflecting regional proportions of the different soil types.

For extent data (i.e. proportional cover for sedges), we followed a very similar approach to that described for count data, with the following exceptions: univariate parametric regression models were fitted with extent (coverage per unit area) as the dependent variable and annual food energy production per unit area as the independent variable. We fitted the same four alternative formulations of the model (as described above), but such that:

where *y* is coverage per unit area, *c* is the coverage and *a* is the survey area.

Ordinary Least Squares estimates for the parameters were then calculated assuming a normal error structure, again using the Nelder-Mead numerical optimisation method in R package ‘alabama’ [27]. AICc was then calculated for each model from the sum of the residual sum of squares (*RSS*) for the fit to each dataset, as described above, and the best-fitting model selected based on AICc. Again composite density-yield functions were produced by multiplying the model predictions by 0.75 and 0.25 for forest and wetland data respectively.

Models with the lowest AICc value for each bird and sedge species, and the parameter values for those fitted models are given in supplementary files. These files give, for each species and soil type, the model selected (denoted A – D, see table S5, or \* for species-soil combinations in which the species only occurred in natural habitat, or X for species-soil combinations for which no individuals were recorded). The tables also give estimates for the model parameters given in table S5. In all cases, the column of the files for the parameter b0 gives the natural logarithm of the modelled value of the mean density in natural habitat, which is the estimate of b0 in the case of models A and B and the estimate of b0 + bW in the case of models C and D.

**(f) Production scenarios for 2050**

We explored the effects of alternative approaches to food production across a wide range of values of Total Annual Agricultural Output, *AAOTOTAL*, 1 GJ ha-1 yr-1 across the entire region to 99 GJ ha-1 yr-1. This highest production level would require all cultivable land to be farmed at our assumed maximum attainable yield, which we took to be 1.25 times the highest yield we recorded in our study squares. However, for illustrative purposes, we focussed on two production scenarios for 2050:

(1) Business as Usual

For this scenario, we assumed that total regional food energy production continues to increase to 2050 in line with trends observed in the period 2005–2014. Regional production data (for the whole Lubelskie region) were obtained from the Central Statistical Office of Poland [3] for eight major crops (wheat, barley, rye, oats, triticale, potato, sugar beet, and oilseed rape) for the period 2005–2014. Annual production (by mass) was converted into annual food energy production for each crop in each year based (as above) on energy content per unit weight of processed crop (from ref. [7]). The mean annual change in regional food energy production (averaged across all eight crops) for 2005–2014 was calculated as +0.8 GJ ha-1 yr-1, which we then projected forward to 2050. Overall, the Continued Growth scenario for 2050 represents a 72.5% increase from 2014 levels.

From 2004–2011, only a marginal increase in per capita consumption of +0.4% per annum has been observed [3]. The current trend in agricultural output is driven by reasonably steady growth in wheat and triticale production (largely for export) alongside a significant increase in demand for biofuel feedstocks (particularly oilseed rape and sugar beet). Between 2014 and 2050, regional human population is predicted to decrease by more than 20% from 2014 levels [3]. Increases in production to 2050 under this scenario are therefore assumed to contribute primarily to national and international exports, and the burgeoning market for biofuels.

(2) Lower Bound

Under this scenario, we assumed that combined demand from the agricultural sector (comprising consumption, exports and biofuel production) is capped at 2014 levels. In addition, we assumed significant efforts are made to tackle edible food waste. Although estimates vary between crops and regions, roughly 30–40% of food is currently wasted at some stage in the food chain [8,28,29]. We adopted a mid-point value of 35% as a representative estimate of the proportion of edible food that is wasted in 2014. We assume an ambitious target of a 50% reduction in food waste by 2050, representing a likely upper-limit which is in line with other studies (e.g. ref. [30]). Overall, this ‘zero-growth plus waste reduction’ scenario assumes a 17.5% decrease in *AAOTOTAL* from 2014 levels by 2050.

**(g) Estimating total regional populations at 2050 (two-compartment model)**

For all values of *AAOTOTAL*within the feasible range (described above), we calculated the total regional population of each species by multiplying the area under each land-use by the predicted population densities at the relevant level of agricultural yield, derived from the fitted density-yield models such that:

*POPTOTAL* = (A – (*AAOTOTAL*/*x*)) *f(0*) + (*AAOTOTAL*/*x*) *f(x)*

Where *POPTOTAL* is the total regional population of a species, *A* is the total cultivable area of the region*, AAOTOTAL* is the Total Annual Agricultural Output of the region, and *x* is the yield of farmed land. *f(0)* and *f(x)* are expected values of population density at zero yield and the yield *x* (derived from fitted density-yield curves). The value of *x* will vary between different land-use strategies (Sharing, Sparing or Intermediate): under Sharing, *x* is the minimum yield necessary to achieve *AAOTOTAL* through farming the entire cultivable area; under Sparing, *x* is the maximum attainable yield (*xMAX* - in this study, 99 GJ ha-1 yr-1, based on agricultural survey data); and under Intermediate strategies it can be any value between the two.

We calculated species’ population density on zero-yielding natural habitats (*f(0)*) as the sum of their densities in forest (*f(forest)*) – and wetland (*D(wetland)*) habitats multiplied by the proportion of the total area that could support these habitats, based on soil type (0.75 and 0.25 respectively):

*f(0*) = 0.75 (*f(forest)) +* 0.25 *(f(wetland))*

**(h) Estimating total regional populations at 2050 (three-compartment model)**

We proposed an alternative formulation of the land sparing model, such that the land area required to achieve a given value of *AAOTOTAL* is minimised by farming at the highest feasible yield (x*MAX*), and that unfarmed land (Au) is divided between unfarmed natural habitats (A(0)) and very extensively-managed farmland, often characterised as High Nature Value farmland (*AHNVf*).

In the three-compartment model, the total province-wide population size of a given species was evaluated as:

*POPTOTAL* = *A(0) f(0)* + *AHNVf f(xHNVf)* + (*AAOTOTAL*/*x*) *f(xMAX)*

Where *f(0)*, *f(xHNV)* and f(x*MAX*) are expected values of population density at zero yield and the yields xHNVf and x*MAX* assumed to apply to the HNVf and high-yielding farmed parts of the region respectively.

For the purposes of estimating population densities for *AHNVf*using species’ density-yield curves, we assume management of this ‘compartment’ reflects low-yielding farming with a theoretical yield of *xHNVf*. This could be any value outside the feasible range of yields for a given value of *AAOTOTAL*. For this study, we assumed *xHNVf* to be the mean yield of all the study sites where average yields fell below the minimum feasible yield for 2050 - i.e. below the minimum yield for our Lower Bound scenario (8GJ ha-1 yr-1, n=9). These sites are characterised by farmland with a high proportion of semi-natural habitat; mosaics of low-intensity agriculture and natural structural elements; and/or farmland supporting rare species or species of conservation concern – meeting the criteria for the three types of HNV farming described by the European Environment Agency [31].

Although we assign HNVf a theoretical yield value for the purposes of estimating population densities, we conservatively assume that HNVf makes zero contribution to *AAOTOTAL*, as this effectively de-couples the yield of this compartment from *AAOTOTAL* and theoretically allows for a range of management options, including management solely for conservation purposes (e.g. through targeted conservation projects or through conservation grazing). We therefore note that the combined extent of natural habitat *A(0)* and HNVf (*AHNVf*) is set by the area spared by high-yielding farming, so that

*A(0)* + *AHNVf* =*A*-(*AAOTOTAL*/*xMAX*)

Subject to this constraint, any values of *A(0)* and *AHNV* could be assumed, so the proportion of spared land (*Au*) that was HNVf (*AHNV* / *A(0)* + *AHNVf*) could take any value between zero and one.

In this model, we set *AHNV* as that which maximised the geometric mean of the ratio of the predicted total population to the total population in estimated in 2014, averaged across all species of birds and sedges. The 2014 reference population size for each species was calculated by multiplying the area currently under each land-use type (based on the CORINE land cover dataset [2]) by the population density at the mean food energy yield for this land cover type. This optimisation was done for birds and sedges combined at each assumed production level. We then counted the number of species for which the highest population occurred with farming at the lowest permissible yield in the two-compartment model (land sharing), the highest attainable yield linked to land sparing in the two-compartment model and the highest attainable yield linked to land sparing in the three-compartment model.

**Supplementary references**

[1] Panagos P. 2006 The European soil database. *GEO: connexion* **5,** 32–33.

[2] European Environment Agency. 2013 Corine Land Cover 2006 raster data.

https://www.eea.europa.eu/data-and-maps/data/clc-2006-raster-3

[3] Urząd Statystyczny w Lublinie. 2014 http://stat.gov.pl/bdlen/app/strona.html

[4] FAO. 2015 FAOSTAT Statistical databases. Rome: Food and Agriculture Organization of the United Nations. http://faostat3.fao.org/home/E

[5] DEFRA. 2014 Agriculture in the United Kingdom 2013. DEFRA.

[6] DEFRA. 2015 Agriculture in the United Kingdom 2014. DEFRA.

[7] US Department of Agriculture. 2015 Composition of foods raw, processed and prepared. USDA National Nutrient Database for Standard Reference, Release 27 (revised). Version Current: May 2015. USDA Agricultural Research Service, Nutrient Data Laboratory. http://www.ars.usda.gov/ba/bhnrc/ndl

[8] FAO. 2011 Global food losses and food waste - extent, causes and preventions. Rome: Food and Agriculture Organization of the United Nations.

[9] Wirsenius S. 2003 Efficiencies and biomass appropriation of food commodities on global and regional levels. *Agric. Syst* **77,** 219–255. (doi:10.1016/S0308-521X(02)00188-9)

[10] Hulme MF et al. 2013 Conserving the birds of Uganda’s banana-coffee arc: land sparing and land sharing compared. *PLoS ONE* **8,** e54597. (doi:10.1371/journal.pone.0054597)

[11] Petit HV, Flipot PM. 1992 Source and feeding level of nitrogen on growth and carcass characteristics of beef steers fed grass as hay or silage. *J. Anim. Sci.* **70**, 867–875.

[12] Wirsenius S, Azar C, Berndes G. 2010 How much land is needed for global food production under scenarios of dietary changes and livestock productivity increases in 2030? *Agric. Syst.* **103,** 621–638. (doi:10.1016/j.agsy.2010.07.005)

[13] Laake J, Borchers D, Thomas L, Miller D, Bishop J. 2015 mrds: Mark-Recapture Distance Sampling. R package version 2.1.12. http://CRAN.R-project.org/package=mrds

[14] Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L. 2001 *Introduction to distance sampling*. Oxford: Oxford University Press.

[15] Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L. 2004 *Advanced distance sampling*. Oxford: Oxford University Press.

[16] Buckland S, Russell R, Dickson B, Saab V, Gorman D, Block W. 2009 Analyzing designed experiments in distance sampling*. J. Agric. Biol. Environ. Stat.* **14,** 432–442. (doi: 10.1198/jabes.2009.08030)

[17] Fitter R, Fitter A, Farrer, A. 1984 *Grasses, sedges, rushes and ferns of Britain and northern Europe – Collins pocket guide*. London: Collins.

[18] Fitter R, Fitter A, Blamey M. 1974 *The wild flowers of Britain and northern Europe*. Glasgow: William Collins Sons & Co Ltd.

[19] Jager EJ, Muller F, Ritz CM, Welk E, Wesche K. 2013 *Rothmaler exkursionsflora von Deutschland*. Berlin: Springer Spektrum.

[20] Jermy AC, Simpson DA, Foley MJY, Porter MS. 2008 Sedges of the British Isles. 3rd ed. London: Botanical Society of the British Isles.

[21] Poland J, Clement E. 2009 *The vegetative key to the British flora*. Southampton: John Poland and The Botanical Society of the British Isles.

[22] Rose F. 1981 *The wild flower key of the British Isles and N.W. Europe*. London: Frederick Warne Publishers Ltd.

[23] Rose F. 1989 *Colour identification guide to the grasses, sedges, rushes and ferns of the British Isles and north-western Europe*. London: Penguin Books Ltd.

[24] Rose F. 2006 *The wild flower key*. Revised and expanded edition. London: Frederick Warne Publishers Ltd.

[25] Coulson SJ, Bullock JM, Stevenson MJ, Pywell RF. 2001 Colonization of grassland by sown species: dispersal versus microsite limitation in responses to management. *J. Appl. Ecol.* **38,** 204–216. (doi: 10.1046/j.1365-2664.2001.00585.x)

[26] Phalan B, Onial M, Balmford A, Green RE. 2011 Reconciling food production and biodiversity conservation: land sharing and land sparing compared. *Science* **333,** 1289–1291. (doi: 10.1126/science.1208742)

[27] Varadhan R. 2015 alabama: Constrained Nonlinear Optimization. R package version 2015.3-1. http://CRAN.R-project.org/package=alabama

[28] Foley JA. 2011 Can we feed the world & sustain the planet? *Sci. Am.* **305,** 60–65.

[29] Parfitt J, Barthel M, Macnaughton S. 2010 Food waste within food supply chains: quantification and potential for change to 2050. *Phil. Trans. R. Soc. B.* **365,** 3065–3081. (doi:10.1098/rstb.2010.0126)

[30] Lamb A et al. 2016 The potential for land sparing to offset greenhouse gas emissions from agriculture. *Nat. Clim. Change* **6,** 488–492. (doi: 10.1038/NCLIMATE2910)

[31] Paracchini ML, Petersen J-E, Hoogeveen Y, Bamps C, Burfield I, van Swaay C. 2008 High Nature Value farmland in Europe. An estimate of the distribution patterns on the basis of land cover and biodiversity data. Luxemburg: Office for Official Publications of the European Communities. (doi: 10.2788/8891)