Environmental Impact of Agri-Environment Support for Organic Farming

REF: LM0495

Final Report - March 2021

Submitted to: Andrew Cole, Natural England

Prepared by: Paul Newell Price, Rachel Thorman, Andrew Crowe, Ian Adams, Sam McGreig, Naomi Jones

EXECUTIVE SUMMARY

Introduction

The Countryside Stewardship (CS) agri-environment scheme (AES) includes support for organic conversion and management. Organic management can have a positive impact on a range of environmental issues, and it is important to understand whether management under CS delivers expected environmental impacts. There is evidence that organic farming can contribute to improving soil quality. Organically managed landscapes are believed to be more complex than those managed conventionally, although evidence is based on analysis of a limited range of parameters and is equivocal.

This pilot study explored the feasibility of assessing the impact of organic options in CS on various soil characteristics and landscape character. The pilot focussed on OT3 – organic management of rotational land, which had greatest uptake in 2016/2017. The effects on soil quality were estimated through measurement of key soil quality indicators. Landscapes containing the organic agreements were compared to similar areas without organic options to assess any difference in landscape structure in terms of habitat parcel size, number and shape.

Methods

The sample comprised 30 fields from two geographic clusters. Fields under organic option were paired with 'counterfactual' fields of similar soil type in the local area, but under conventional management (i.e. 15 field pairs).

Soil properties

The 15 paired sites were selected from two clusters either sandy & light silty (8 pairs) or medium (7 pairs) soil types where desk assessment indicated some risk of soil erosion due to sloping ground. Counterfactual/conventional sites were matched with organic sites, as far as possible, by: region, land use, farm type, size and soil type. Sampling was undertaken between late January and March. Basic background information on farm management was collected over the telephone when arranging site visits.

Assessments were made in the field or samples were taken for lab analysis to measure:

- Soil bulk density
- Soil chemical analysis (soil pH; extractable P, K and Mg; total N; total organic C)
- Soil texture and organic matter (loss on ignition)
- Structural condition (Visual evaluation of soil structure) and dispersion ratio
- Earthworm populations
- Soil erosion risk (vegetation cover; management condition; water erosion risk & mitigation; and landscape connectivity).

Statistical analysis (analysis of variance) compared soil properties between organic and conventional management and models were fitted to determine linear relationships between dependent and independent variables.

In addition, metabarcoding, a DNA sequencing technology, was used to assess the bacterial, fungal, nematode, worm and other invertebrate populations of soil samples taken from the matched organic and conventionally farmed sites.

Landscape

Eight agreements containing organic options were used to assess whether landscapes which contain organic options have a different level of structural complexity than comparable landscapes under conventional management. A 5x5 grid of 1 km cells was placed with the central cell over the centroid of the organic options within the agreement. The UKCEH Land Cover Map was used to define spatial pattern of parcels in the central grid cell and in surrounding grid cells with similar habitat composition but without any organic options. The following metrics were calculated for each grid cell to describe landscape structure, with the metric scores for the central cell compared to the distribution of scores for the comparable neighbouring cells:

- Parcel count
- Average parcel size
- Mean area-perimeter ratio
- Edge density
- Habitat diversity.

Results

Management

Most organic sites had been under organic management for 10-15 years. The amount of change required to convert to an organic system varied, but crop rotation had changed at all organic sites. Organic manures were applied on all sites except two conventionally managed fields. Plough-based cultivation was more common under conventional management and insecticides and fungicides were applied respectively to half and all the conventional sites.

Soil properties

Although there were few significant differences in soil chemical properties between management systems, organic matter (loss on ignition) and extractable magnesium were significantly higher under OT3 compared to conventional management. An absence of a significant difference in clay content indicated successful selection of site pairings based on soil type.

There were no management level differences in soil bulk density or soil structural condition (visual assessments or soil dispersion ratio).

There were some differences between paired sites in the earthworm counts for the different ecotypes (endogeic and midden counts as a surrogate for anecic earthworms). However, overall there were no significant differences between OT3 and conventionally managed sites.

Vegetation/residue cover was higher under OT3 than conventional management and appeared to be related to differences in cropping.

Water erosion risk

The level of erosion risk was very similar for OT3 and conventional sites and the degree of erosion risk mitigation due to management was also similar. Most sites had low or very low landscape connectivity, thus the likelihood of eroded soil reaching a watercourse was low. However, scores for good management condition relating to watercourse protection were higher for sites managed under OT3 compared to conventional sites.

Soil biome (genetic analysis)

Significant differences between the organic and conventional groups were found in the alpha diversity of the bacterial populations but not the other organisms tested. A number of bacterial and fungal species were also found to be significantly associated with the organic farming practices. Genetic sequencing recorded no differences in the worm populations between the two farming practice groups.

Landscape

In some agreements the organic options were not concentrated at one location and in these cases each cluster of organic options was analysed separately, giving a total of 11 clusters across the eight agreements. For two of these clusters it was only possible to find a single comparable grid cell in the 5x5 neighbourhood, which meant that a statistical comparison could not be made in these cases.

Across the remaining nine clusters, only one set of landscape metrics for the focal cell showed any significant difference to the landscape metrics for the surrounding conventional cells. For that location the average parcel size was larger and the area to perimeter ratio was higher, which would be indicative of a lower landscape complexity. Ignoring significance testing, across the clusters there was no consistent pattern in the difference between the organic focal cell metrics and the metrics for the neighbourhoods. The nine clusters of organic options examined showed metric scores indicative of both higher and lower landscape complexity in focal cells than the mean metric scores for the comparable conventional cells in the neighbourhood.

Conclusions

There were few differences in soil quality between OT3 and counterfactual sites. Most of the OT3 and counterfactual sites had moderate to good levels of soil structural stability and quality. This was probably a reflection of the use of organic manures in recent years at most sites, whether they were organically or conventionally managed. However, vegetation/residue cover and soil organic matter content were higher on organic sites, probably due to the grass leys included in the organic rotation.

Species identified through genetic analysis as associated with organic farming practices have the potential to be organic farming biomarkers but this would need exploring in a larger dataset.

The absence of a consistent significant difference in landscape structure could be a result of the small sample size and/or the characteristics of the datasets used in this analysis. Equally, organic options may represent a weak driver of landscape structure compared to other historic and current factors driving decisions that affect management of features in the landscape. It is difficult to define an appropriate counterfactual to test differences in landscape structure as landscapes are the result of the interaction of natural and human processes over time, making each landscape unique. The results of this study have shown that even within the close neighbourhood of a location, the landscape composition and structure can vary considerably.

Recommendations

Many of the methods used in this pilot study would be suitable for a wider study, but some elements were constrained by resource and timeframe issues. Key recommendations for a national study:

- The range of soil assessments conducted in this study would be appropriate for use in a national study to identify any differences between farming systems. In this pilot study, the assessments failed to show differences between farming systems. This may be due to sample size or it may be that differences between management systems do not exist for some soil quality indicators.
- Timing of sampling:

    - Soil quality and risks to water quality are best carried out when soils are at field capacity (typically between mid-October and April)
    - Ideal timing for sampling of soil biota may be different, but will depend upon the species groups to be investigated
    - A narrow survey window maximises comparability but presents practical challenges and would likely require repeat visits to assess different factors.
- Collection of more detailed management information would allow better interpretation of results.
- Selection of paired sites would be improved by inclusion of more precise information on slope in addition to the datasets used in this study.
- There is potential for remote sensing to contribute to assessment of erosion, although issues such as image resolution and frequency of image capture require further investigation.
- Further investigation of potential biomarkers from genetic analysis.
- Use of hedge density in the landscape character assessments.
- Use of historic maps or undertaking pre-agreement surveys of the area to assess landscape complexity prior to the introduction of organic options.

CONTENTS

ACKNOWLEDGEMENTS

The project team would like to thank the agreement holders for allowing permission to access their land.

We are grateful to the following:

Field surveyors: Ryan Hickinbotham, Michael Morris and Dan Jakes (ADAS), Sam McDonough, Gilli Thorpe and Morgan Wodring (Fera).

Soil cores were processed for dry bulk density measurements, and earthworm biomass assessments were carried out by Ryan Hickinbotham, Michael Morris, Dan Jakes (ADAS), Morgan Wodring and Sam McDonough (Fera).

Dispersion ratio assessments were carried out by Helen Kingston (ADAS).

Sites were selected by Andy Frost (ADAS) and Lee Butler (Fera) contributed to the landscape analysis.

Invertebrate genetic analysis was carried out by: Benjamin Barrett, Chris Conyers, Jane Hall, Hollie Pufal, Katie Wetherall (Fera).

This project is supported by the Rural Development Programme for England, for which Defra is the Managing Authority, part financed by the European Agricultural Fund for Rural Development: Europe investing in rural areas.

1 INTRODUCTION

1.1 Background

The Countryside Stewardship (CS) agri-environment scheme (AES) includes support for organic conversion and management. Organic management can have a positive impact on a range of environmental issues, and it is important to understand whether management under CS delivers expected environmental impacts. A scoping study to identify appropriate methods for monitoring biodiversity, soil quality and landscape character under CS organic options (Carey et al., 2019) identified soil organic matter and erosion, water quality and landscape character as requiring further research.

There is an extensive evidence base that demonstrates how organic farming can contribute to improving soil quality (Lotter et al., 2003; Stolze et al., 2000; Seufert & Ramankutty, 2017). However, the specific contribution of AES Organic Management & Conversion options to this objective has yet to be systematically monitored and evaluated in England.

In addition, organically managed landscapes are believed to be more complex than those managed conventionally, although evidence is based on analysis of a limited range of parameters and is equivocal. Natural England project LM0458 identified that there was limited evidence of how support for organic farming in agri-environment schemes impacts soil quality and landscape character

This pilot study explored the feasibility of assessing the impact of organic options in CS on various soil characteristics and landscape character. The pilot focussed on OT3 – organic management of rotational land, which had greatest uptake in 2016/2017. The effects on soil quality were estimated through measurement of key soil quality indicators. Landscapes containing the organic agreements were compared to similar areas without organic options to assess any difference in landscape structure in terms of habitat parcel size, number and shape.

Natural England are considering funding a national monitoring programme to assess the effectiveness of the Countryside Stewardship agri-environment scheme support for organic conversion and management in improving soil quality and landscape character. To inform the feasibility and utility of such a monitoring programme, a pilot monitoring and evaluation project focusing on 30 sites was set up to assess some of the environmental impacts of organic farming, when supported by the Countryside Stewardship scheme.

This report presents the findings of this pilot monitoring and evaluation project. Section 2 focuses on the soil properties study and section 3 on the landscape characteristics study. The introduction, methods, results and conclusions from each study are reported separately in each section.

1.2 Objectives

The main objective of this pilot monitoring and evaluation project was to assess the feasibility of monitoring the impact of agri-environment support for organic farming on soil quality, water quality and landscape. The project focused on the following research questions:

1. Do soil properties differ between organic and conventional farms and if so, what are the implications of these differences for productivity and environmental outcomes?
2. Is there a difference in the quantity and frequency of soil erosion on organic and conventional farms?

3. How does soil biota compare on organic and conventional farms? What are the implications of any differences for productivity and environmental outcomes?
4. Do the risks and impacts on local surface water quality differ on organic and conventional farms?
5. Is there a difference in landscape character between organic and conventional farms?

The pilot focused on the organic land management option with the greatest uptake (accounting for around 5% of CS funding on management options in 2016/17: LM0460), namely: OT3 – rotational land. It targeted 2016 and 2017 agreements, which have had the most time for CS organic management to take effect. Outcomes from the pilot project will feed into design of future AES monitoring including monitoring of the new Environmental Land Management Scheme.

1.3 Methods

The impact of AES support for organic farming on soil and water quality was assessed through a field survey of 15 paired fields (30 fields in total) with one field in each pair in the OT3 option and one field with similar soil type not in the OT3 option (the counterfactual field). This was considered to be the minimum number of sites needed to detect any difference in soil properties between management systems. Counterfactual fields were located on neighbouring non-organic agreements, but the fields were not under any AES option. Sample fields were located in two clusters in two contrasting landscapes representative of lowland England in the East Midlands and East Yorkshire.

On each field, soil samples were taken to assess soil physical and biological properties, including use of DNA sequencing technology to assess soil biota (see section 2.2.1 for sampling period). Risk to local surface water quality was also assessed through assessment of soil erosion risk, landscape connectivity and observation of erosion features (see section 2.2.3.6 for methods).

Landscape characteristics were assessed through a desk exercise using Land Cover Map 2015 to identify land cover classes and to calculate various metrics describing landscape structure (see section 3.2 for methods).

2 SOIL PROPERTIES

2.1 Introduction

There is an extensive evidence base that demonstrates how organic farming can contribute to improving soil quality (Lotter *et al*., 2003; Stolze *et al*., 2000; Seufert & Ramankutty, 2017). However, the specific contribution of AES Organic Management & Conversion options to this objective has yet to be systematically monitored and evaluated in England. Before any national monitoring programme, it is important to investigate the feasibility of assessing this contribution at a regional scale.

The effect of CS organic land management options on soil properties relative to conventional farm management was assessed using methodologies developed as part of the CS Baseline project (LM0458) and Countryside Survey. Differences in soil properties between conventional and organic systems are likely to be driven by more frequent use of farmyard manures, legumes, cover crops and generally more diverse rotations, as well as the reduced use of agro-chemicals, particularly herbicides, pesticides, insecticides and fungicides in organic systems (Scullion *et al*., 1998; Hole *et al*., 2005; Marriott & Wander, 2006). Despite the greater use of tillage in some arable organic systems, compared with some conventional arable reduced tillage systems, one might therefore expect differences in soil organic matter content, soil structure, soil porosity, soil biology and soil aggregate stability between organic and some conventional systems. Soil chemical properties may also differ depending on the historical use of organic manures and manufactured fertilisers within the two systems. In this study the effects on soil quality were estimated at all sites through measurement of key soil quality indicators.

2.2 Methods

2.2.1 Sampling stratification

Soil physical and chemical properties were assessed at 30 sampling sites, based on 15 sites in the OT3 (rotational land) option and 15 counterfactual (CF) sites, with each counterfactual site paired to one of the OT3 option sites. The aim was for eight of the option sites and 8 of the counterfactual sites to be on *sandy or light silty* soils; and 7 option and 7 counterfactual sites to be on *medium* soils. Sampling was planned to be undertaken in January and February 2020, when soils were likely to be moist, but not frozen or too 'wet' (Table 1). In practice, sampling was extended into March due to very wet soil conditions which were unsuitable for taking samples.

A Geographical Information System (GIS) assessment was used to select sites that were dominated by '*sandy or light silty'* or '*medium'* soils, and predicted to have a moderate to high risk of erosion, i.e. sloping land (see section 2.2.3.6 - Defra, 2005).

Counterfactual sites were selected from land parcels associated with conventionally managed agreements and with linear boundary feature options, e.g. BE3 (management of hedgerows), SW1 (4-6 m buffer strip) and AB9 (winter bird food). Fields with parcel level options were excluded.

Table 1 Sampling stratification

Soil type	OT3 – rotational land	Counterfactual
Sandy and light silty (<18% clay)	8	8
Medium (>18% clay)	7	7
All	15	15

The selection of sites for both the organic (OT3 organic) and the counterfactuals (non-organic), were carried out in the same way using data analysis within GIS software. Site location options were selected from the Countryside Stewardship Scheme 2016 Management Options dataset provided by Natural England. The initial analysis selected sites in East Yorkshire and the East Midlands (Derbyshire and Nottinghamshire). These sites were further reduced in number by analysis using the © NSRI NATMAP Topsoil Texture dataset and selecting the sites that were located within sandy and light soils, and medium soils. Slope data was extracted from an Ordnance Survey 50 m digital terrain model to exclude sites that were situated on land that had a slope <2°. For practical purposes, groups of sites were selected that were within 5 km of each other.

Counterfactual sites were matched to their organic counterparts as far as possible in terms of the following factors:

- Geographical region - ideally National Character Area
- Land use type, i.e. rotational land
- Farm enterprise type
- Soil type and management practices that are key drivers of soil properties (e.g. the use of organic materials and cultivation techniques) so that the impact of organic certification on soil properties could be explored.

2.2.2 Contacting farmers

Following site selection, land managers were contacted to arrange access and to obtain some background information on the site:

i. The length of time over which each organic farm had been managed organically.
ii. Whether organic manures were used in the farming system.
iii. The degree of system change when converting to an organic system:
    - Were cover crops used before conversion to organic?
    - Were organic manures used before conversion to organic?
    - Had the rotation changed as a result of conversion to organic?
iv. The cultivation system
    - Plough-based i.e. the majority of primary cultivations are carried out using a mouldboard plough)
    - Reduced tillage
    - No-till
    - Strip-till
    - Mixed (i.e. a more balanced use of cultivation systems)
    - Other
v. Previous crop. Note – only the crop from the previous year was required (i.e. 2019), but if available the previous 5 crops were recorded.
vi. Whether insecticides and/or fungicides were used in the previous crop (2019).
vii. An email address if the agreement holder would like to be sent a copy of the 2-page summary and a link to the final report.

2.2.3 Field survey

At all 30 sampling sites, the field contours were used to select a homogeneous area for sampling (i.e. similar soil type and condition) of no greater than one hectare. A Garmin© eTREK© "high sensitivity" GPS device (accurate to around 3 m) was used to mark the four corners of the sampling area.

2.2.3.1 Soil bulk density

Within the sampling area, a 'W' pattern was walked and a baseline bulk density (BD) sample (to a depth of 5 cm) was taken, and GPS located, at each of the five points on the 'W' (sample number = 5). Each sampling point was located well away from any tramlines or other atypical areas. Undisturbed soil cores were stored at around 4ºC before being processed.

Soil BD measurements were assessed relative to the topsoil BD 'trigger' levels (the level at which soil physical conditions may be an issue for production and further investigation is recommended) (Merrington, 2006).

To investigate the impacts of the OT3 option on topsoil BD, while controlling for the major effect of soil organic matter (SOM) content on BD, the following equation that predicts BD from SOM content (from Natural England monitoring data of semi-natural sites with topsoils within the SOM range 1.8% to 28%; Shepherd, 2017) was used to compare the residuals derived from predicted and measured values between option and counterfactual sites:

$BD = 1.1967e^{-0.052SOM}$ (n=130; R^2 = 0.76)

2.2.3.2 Soil chemical analysis and soil texture

To assess soil texture and chemical properties, 25 bulked soil cores were collected by walking the same 'W' pattern (as for soil BD) across each selected area to obtain a single composite sample for each area. Samples were taken to a depth of 15 cm in arable fields, short-term (<5 year) leys or grassland about to be ploughed and re-seeded, and to 7.5 cm depth in long-term grassland fields. Cores were kept cool (<4°C) before transport for analysis of:

- pH
- extractable phosphorus
- extractable potassium
- extractable magnesium
- total nitrogen (Dumas method)
- total carbon (Dumas method)
- total organic matter content (Loss on ignition)
- soil texture (percentage sand, silt and clay content; laser method).

2.2.3.3 Soil structural condition and aggregate stability

Visual Evaluation of Soil Structure (VESS)

Each site was assessed for soil structural condition and sampled for aggregate stability testing. Soil was sampled at three locations randomly selected along the first, second and fourth 'arms' of the soil sampling 'W'. Soil structural condition was assessed using the Visual Evaluation of Soil Structure (VESS) method (Guimaraes *et al*., 2011). The VESS scoring system (developed from Peerlkamp (1967)) provides an estimate of visual porosity and the uniformity of its distribution. The lowest score (Sq1) is given to the least compact and most porous condition, and the highest score (Sq5) to a very compact condition with very large and often platy aggregates with very low visible porosity. At each sampling point the following information was recorded:

- GPS location of VESS assessments (3 VESS assessments on each site/option)
- Individual and mean VESS score for the topsoil
- Individual and mean VESS score for the poorest layer in the topsoil (i.e. the layer, if a distinct layer is identified, with the highest VESS score/poorest structure and porosity; if no layering is identified the score for the whole topsoil block is used)
- Depth and thickness of the poorest layer in the topsoil at each assessment location

Dispersion ratio

Soil aggregate stability is a measure of the ability of soil aggregates to resist degradation when exposed to external forces such as water and wind erosion (e.g. Papadopoulos, 2009). Aggregate stability was assessed using the soil dispersion ratio test on soil collected from the three points where VESS assessments were made. At each point, a sample of 1.5 kg of soil was collected from the top 5-10 cm, placed in a plastic container and transported for analysis with the minimum of disturbance. The dispersion ratio test compares the proportion, by weight, of silt and clay suspended by mild slaking forces to the total amount present in the sample. The ratio has been found to be a valuable criterion for distinguishing between soils with different degrees of structural stability and has been widely used in Defra funded R&D and monitoring projects through using the method detailed in ADAS SOP SOILS/052 (Determination of Soil Stability by the Dispersion Ratio).

2.2.3.4 Earthworm sampling

Earthworms can be an indicator of good soil health. They create burrows while mixing, ingesting and excreting soil material, thereby modifying the physical structure and availability of soil resources, and fulfilling the role of 'ecosystem engineers' (e.g. Pulleman *et al*., 2012).

At the same three randomly selected locations as the VESS assessments, earthworm numbers and biomass were also measured on 'blocks' of soil using the AHDB GREATSOILS Factsheet, 'How to count earthworms' with earthworms identified to ecotype level (anecic, endogeic, epigeic). Earthworm assessments were made on soil blocks of 20 x 20 cm x 25 cm deep, with the sample taken well away from BD sampling sites.

The earthworms found during excavation were placed in a plastic box, containing moist paper towel or damp moss, and having respiration vents in the lid to prevent suffocation during transport for determination of earthworm numbers and biomass. The containers were stored in cool conditions prior to earthworm assessment within 24 hours of collection. Post assessments, earthworms were returned to agricultural land.

2.2.3.5 Midden counts

At the same time as the earthworm collection, the number of *Lumbricus terrestris* (LT) middens were counted and recorded from three 1 m^2 quadrat assessments carried out at randomly selected locations within 3 metres of each earthworm sampling site. LT form deep vertical burrows in the soil. Directly above their burrows is a midden, an accumulation of straw, tree leaves etc., with some worm casts.

2.2.3.6 Soil erosion risk

Soil erosion risk was assessed at all sites, based on field observations of vegetation/residue cover and erosion/surface runoff features using methodologies developed as part of the CS Baseline project (LM0458) to evaluate the risk of soil erosion and risk of delivery to water courses (see section 2.3.6 for detail of the methods and outputs).

Vegetation/residue cover percentage and score

At each site, the vegetation and crop residue cover was assessed by estimating the percentage cover in a 1 x 1 m quadrat at 10 points evenly spaced within the field. The mean (n=10) percentage cover was calculated for each site. The mean measured cover was then converted to an average cover score ranging from 0-3 (see section 2.3.6).

Management condition score

The management condition score ranging from 0 (poor) to 3 (very good) was assessed using a set of four scored factors:

- Evidence of soil erosion, including sediment fans, runnels, rills, or other eroded preferential paths on the land or soil leaving the field
- Signs of use for access by vehicles and presence of permanent ruts
- Grazing absence/presence and degree of poaching
- The vegetation/crop residue cover score

Water erosion risk score

Soil erosion risk by water was assessed using the categories in Table 2. Each erosion risk was scored ranging from 0 (lower risk) to 2 (high or very high risk). The soil type and slope angle of the site were measured using hand texturing (confirmed by laboratory analysis) and using a clinometer and ranging poles, respectively.

Table 2 Erosion risk categories (Defra, 2005)

Soil Erodibility Category	Steep slopes >7°	Moderate Slopes 3-7°	Gentle Slopes 2-3°	Level Ground <2°
Sandy and light silty soils	Very high	High	Moderate	Lower
Medium and calcareous soils	High	Moderate	Lower	Lower
Heavy/peaty soils	Lower	Lower	Lower	Lower

The topsoil textures that correspond to the soil erosion risk soil erodibility categories in Table 2 are presented in Table 3.

Table 3 Soil erodibility category soil textures

Soil Erodibility Category	Topsoil Textures
Sandy and light silty soils	Sand, Loamy Sand, Sandy Loam, Sandy Silt Loam, Silt Loam
Medium and calcareous soils	Sandy Clay Loam, Clay Loam, Silty Clay Loam
Heavy soils/peaty soils	Sandy Clay, Clay, Silty Clay, Peat / Peaty

Erosion mitigation score

The degree to which the water erosion risk was addressed by management was visually assessed and scored as 0 (not at all), 1 (partially), 2 (mostly) or 3 (completely).

Landscape connectivity score

Fields were also assessed in terms how likely it was that soil eroded from the field would reach a water course. Scores were assigned based on a number of factors:

- Evidence of soil loss from the field
- If there was a water course at the base of the slope
- Evidence of a surface link with a water course, road, drain or culvert
- Presence/absence of a buffer between the field and a water course, road, drain or culvert
- The proportion of the length of a water course covered by a buffer strip
- If there was a clear pathway for sediment between the field a gateway and a water course, road, drain or culvert
- Whether soils were compacted near a watercourse

The scores ranged from 1 (very unlikely) to 5 (highly likely in moderate events). The following descriptive guidance was provided to surveyors:

Very unlikely

- Gently to moderately sloping, no concentrated flow pathways, flat base of slope, no direct link with watercourse (riparian buffer > 4 m), good (≥ 30%) crop/vegetation cover in early autumn, winter and early spring.

Unlikely, but possible in high energy events

- Gently to moderately sloping, no concentrated flow pathways, flat base of slope with riparian buffer that could be breached to water course in high energy events, good crop/vegetation cover in early autumn, winter and early spring.

Unlikely, but possible in moderate events

- Gently to moderately sloping, concentrated flow pathways on midslope, flat base of slope with narrow (< 4 m) riparian buffer that could be breached to water course in moderate energy events.

Likely in moderate events

- Moderate to steeply sloping, sandy or light silty soil type, concentrated flow pathways to base of slope next to watercourse, low (<25%) crop/vegetation cover in early autumn, winter or early spring.

Highly likely in moderate events

- Moderate to steeply sloping, sandy or light silty soil type, concentrated flow pathways to base of slope with no or very narrow (<4 m) riparian buffer, low (<25%) crop/vegetation cover in early autumn, winter or early spring; evidence of soil loss to watercourse

2.2.3.7 Statistical analysis

Analysis of variance (ANOVA) was used to investigate any differences in soil physical or chemical properties between the organic OT3 (rotational land) option and counterfactual sites at the 95%

significance level. The analysis was carried out in Genstat (18th version, 2016 VSN International Limited), where the data was reviewed to check that the assumptions of ANOVA were satisfied, including checking for normal distribution and skewness. All the data fitted a normal distribution apart from the earthworm data, which was log transformed for analysis.

Regression analysis was also used to carry out a logistic regression on scores that had been converted to binomial data (i.e. either a 'good' or 'bad' score). The analysis tested if there was a difference in the proportion of the 'good' scores between the OT3 and counterfactual land management 'treatments'.

Summary statistics were used to describe site type soil characteristics including the soil type, organic matter content, soil pH, nutrient status, structural condition, aggregate stability, soil mineral nitrogen content and vegetation/residue cover.

Analysis of covariance models were fitted to determine linear relationships between dependent (e.g. soil organic matter content) and independent (e.g. clay content, geographical region, average annual rainfall, length of time in an organic system) variables.

2.2.4 Soil biota communities

The use of DNA based methods to monitor environmental sites has been growing for a number of years and recently reviewed by Porter & Hajibabaei (2018). The current project used metabarcoding, a DNA sequencing-based method to assess differences in soil microbiome and micro- /macro-fauna. The aim was to compare several matched organic and conventionally farmed sites and look for measurable differences and potential biomarkers. Based on previous experience gained during the CS-Baseline project, EU funded EMPHASIS project, Fera's Big soil project and other work for Defra plant health the following targets were chosen for metabarcoding of the soil samples:

1. 16S. This is a bacterial gene and is frequently used to assess bacterial populations. Fera has used this metabarcoding target in a range of different sample types including soil (Big soil community), bees (Budge *et al*., 2016) and faeces (Gaukroger *et al*., 2020)
2. ITS. This is a fungal gene used to assess fungal populations. Fera has previously used this target in soil (Big soil community) and spore traps (Ortega *et al*., 2020).
3. 18S. This gene targets eukaryotic organisms and Fera have used it to target nematodes (Ahmed *et al*., 2019).
4. Cytochrome oxidase 1 gene (COI). This gene is used to target invertebrates and Fera developed a MinION based protocol for its use in the EU-EMPHASIS and CS-Baseline projects. During the CS-Baseline project it was also observed that it allows the identification of earth worms.

The aim of this project was to extract DNA from all the biota in the soil samples and then sequence for all of the above genes allowing comparisons to be made on the biological diversity across the different sites.

2.2.4.1 Soil sampling

Soil samples were taken using a specified grid approach for assessment of soil biota using the metabarcoding technique used in the Fera "Big Soil Community" (BSC) project and developed as part of the AHDB funded soil health partnership. 25 sub-samples of soil were collected in a grid (Figure 1) over an area of approximately 1 hectare. From each sampling point, a sterile trowel was used to collect soil to a depth of 20-25 cm for arable sites and down to 7.5 cm for grassland. Trowels were washed and cleaned with a dilute (10%) bleach solution in the laboratory and transported to the field site in a clean, sealed plastic bag for sampling at a single site. The sub-samples were bulked, thoroughly mixed in a large, clean, sealed bag, and a portion of the sample (approx. 0.5 kg) transferred to a new collection bag. Soil samples were stored at c. 4°C before transport for analysis.

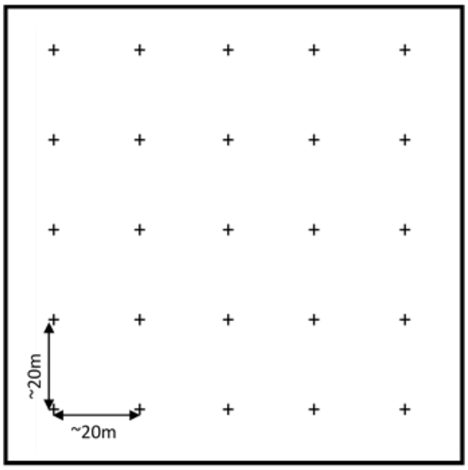

Figure 1 Soil biota sample design

2.2.4.2 Sequencing

DNA was extracted from 50 g of soil taken from the sites detailed in Table 1. The soil was disrupted using 2.5 cm ball bearing and a crude DNA extract was prepared using silica. The method is described in detail in (Woodhall *et al*., 2012). The crude DNA extract was then further purified using the power soil DNA extraction kit (Qiagen, UK).

The soil DNA was then amplified using primers for 16S (Caporaso *et al*., 2011), ITS (Toju *et al*., 2012) and 18S (Ahmed *et al*., 2019) and processed, indexed and sequenced on a MiSeq as described in (Illumina, 2013) yielding pairs of 300nt DNA sequences read pools per sample. For the COI sequencing the soil DNA was amplified using LCO 1490 and HCO 2198 primers (Folmer *et al*., 1994) and the amplicons processed, barcoded and sequenced using the SQK-LSK109 genomic sequencing kit (Oxford Nanopore) an a MinION R9.4 flowcell.

2.2.4.3 Bioinformatics

For 16S, ITS and 18S datasets, the Qiime2 (Bolyen *et al*., 2019) pipeline was used for primer trimming, quality control, denoising, chimera removal, sample filtering, taxonomic filtering (where relevant) and classification. Taxonomic assignments were made with a naïve bayes classifier, trained on the Silva 138 database (16S and 18S) and the Unite database (ITS). A confidence threshold of 0.7 was used for the 16S dataset, with a more conservative 0.9 being used for the ITS and 18S datasets.

For the COI dataset, reads were basecalled with Guppy (version 4.0.11) in high accuracy mode. Reads were then trimmed with Cutadapt (Martin, 2011) and then any reads which were too long or short for the expected amplicon size were removed. A custom COI database was built from sequences obtained from NCBI, and reads were subject to a BLASTn (Camacho *et al*., 2009) search against this database. The resulting reads were filtered so that only matches to sequences in the database that had a percentage identity of at least 80%, and an alignment length of at least 80% were included. Of these hits, only the hits with a bitscore within 20% of the highest scoring read were included in the final set of filtered reads. Finally, a lowest common ancestor approach was applied to the dataset, where if at least 75% of the assignments were in agreement, that taxonomic label was applied to the read. Otherwise, the next highest rank was considered, and the process was repeated until a label was assigned to the read. Along with invertebrates, worms were also detected in the COI data and this was used in the worm count analysis.

The Silva database was used for 16S and 18S analysis. The 18S database is annotated to a lesser extent than the 16S database, with annotations skipping the family and genus levels. As such, input to LEfSe was from the order level, which may mask some lower level differences. The Unite database was used for ITS analysis, and a custom COI database was built from the NCBI database.

2.3 Results

2.3.1 Field management

Most of the OT3 option sites had been organically managed for 10-16 years, although four sites had been organic since 1949. Of the 11 sites that had been organically manged for 10-16 years, 7 had required little change in the farming system to convert, although four had required a lot of change (Table 4), and before conversion all had used organic manures. Of the 11 sites more recently converted, only four had used cover crops prior to conversion. The crop rotation had changed following conversion to organic production at all the organic sites.

The site management details provided by agreement holders Indicated that fungicides had been used at all of the counterfactual sites in the previous year and insecticides had been applied at eight of the sites (Table 4).

All of the sites used organic manures in the farming system with the exception of two counterfactual sites. The majority of the conventional sites were using a plough-based cultivation system (i.e. the majority of primary cultivations were carried out with a mouldboard plough), although two sites used reduced tillage and two sites had a mixed system (i.e. systematic use of reduced tillage and ploughing across the rotation in approximately equal proportion). In contrast, half of the OT3 option sites used a plough-based system and half were mixed. At the counterfactual sites, winter cereals and oilseeds were the most common crop (nine sites) in the previous year (2019), whereas on the organic sites, grass was the most common previous crop (seven sites). Interestingly, two of the counterfactual sites were also in a grass ley rotation (Table 4).

Table 4 Details of site management provided by agreement holders

Pair CF/OT3	Current crop	Previous Crop (2019)	Length of time been organic (years)	Amount of system change required for organic conversion	Are organic manures used in the rotation?	Cultivation system	Insecticide applied in 2019	Fungicide applied in 2019
1 CF	W wheat	W barley	-	-	Yes	Plough	No	Yes
1 OT3	Grass	W wheat	16	Significant change	Yes	Mixed	No	No
2 CF	Stubble	W wheat	-	-	Yes	Reduced tillage	Yes	Yes
2 OT3	Cover crop	Forage rape	10-12	Little Change	Yes	Mixed	No	No
3 CF	Stubble	S barley	-	-	Occasionally	Plough	Yes	Yes
3 OT3	Kale & mustard	S barley	13	Little Change	Yes	Plough	No	No
4 CF	W wheat	No data	-	-	No	Plough	No	Yes
4 OT3	Grass	Grass	10-12	Little Change	Yes	Mixed	No	No
5 CF	W wheat	Peas	-	-	Yes	Reduced tillage	Yes	Yes
5 OT3	Grass	Grass	71	Unknown	Yes	Plough	No	No
6 CF	W wheat	Stubble turnip	-	-	Yes	Mixed	No	Yes
6 OT3	Grass	Grass	16	Significant change	Yes	Mixed	No	No
7 CF	W wheat	W barley	-	-	Yes	Plough	No	Yes
7 OT3	Cover crop	Whole pea crop	16	Significant change	Yes	Mixed	No	No
8 CF	W wheat	OSR	-	-	Yes	Plough	Yes	Yes
8 OT3	W wheat	Potatoes & S beans	71	Unknown	Yes	Plough	No	No
9 CF	W wheat	W wheat	-	-	Yes	Mixed	No	Yes
9 OT3	Grass	Grass	16	Significant change	Yes	Mixed	No	No
10 CF	W wheat	OSR	-	-	Yes	Reduced tillage	Yes	Yes
10 OT3	Stubble	W wheat	71	Unknown	Yes	Plough	No	No
11 CF	Oats	Barley	-	-	Yes	Plough	Yes	Yes
11 OT3	Grass	Grass	13	Little Change	Yes	Plough	No	No
12 CF	Stubble	W wheat	-	-	Occasionally	Plough	Yes	Yes
12 OT3	Grass	No data	13	Little Change	Yes	Plough	No	No
13 CF	W wheat	W wheat	-	-	Yes	Plough	Yes	Yes
13 OT3	Stubble	S barley undersown with grass ley	71	Unknown	Yes	Plough	No	No
14 CF	Stubble	No data	-	-	No	Plough	No	Yes
14 OT3	W wheat	Grass/red clover	10-12	Little Change	Yes	Mixed	No	No
15 CF	Cover crop	No data	-	-	No data	No data	No	Yes
15 OT3	Stubble	Arable intercrop	10-12	Little Change	Yes	Mixed	No	No

NB All ‘grass’ crops are temporary grass; W wheat = winter wheat; W barley = winter barley; S barley = spring barley; S beans = spring beans

2.3.2 Soil chemical analysis and soil texture

Table 5 summarises the soil chemical characteristics and soil textures at the 15 organic OT3 option sites and the 15 conventional counterfactual sites. The raw soil chemical analysis data are provided in the project database. There was a significant difference in extractable magnesium (Mg – mg/l) between sites in OT3 and the counterfactual sites (ANOVA: F = 9.80, d.f. = 1, 14, *P* = 0.007). ANOVA on the loss on ignition data also indicated that sites in the OT3 option had higher soil organic matter content than the counterfactual sites (ANOVA: F = 4.86, d.f. = 1, 14, *P* = 0.045). Soil total nitrogen (N - %) and soil organic carbon (SOC - %) values were also numerically (but not significantly) higher on OT3 sites than on counterfactual sites (ANOVA: F = 4.36, d.f. = 1, 14, *P* = 0.055, and ANOVA: F = 3.99, d.f. = 1, 14, *P* = 0.065 respectively). However, there were no significant differences (*P*>0.05) between OT3 and CF sites for soil pH, extractable P or extractable K concentrations (Table 5).

Table 5 Soil chemical characteristics and soil texture measurements: mean (standard error of the mean) and [range]

Site type	pH	P (mg/l)	K (mg/l)	Mg (mg/l)	Clay content (%)	Organic matter - Loss on ignition (% w/w dry basis)	Total nitrogen (% w/w basis)	Organic carbon - Dumas method (% w/w dry basis)
Counterfactual (n=15)	7.4 (0.3) [5.8-8.7]	19 (1.4) [10-31]	167 (20.4) [70-367]	77 (17.6) [23-261]	29 (2.3) [15-43]	4.8 (0.3) [3.3-6.4]	0.2 (0.0) [0.2-0.3]	2.0 (0.2) [1.2-3.0]
OT3 option (n=15)	7.2 (0.2) [6.3-8.5]	27 (4.1) [13-71]	198 (12.4) [104-260]	145 (37.4) [50-597]	24 (2.1) [7-36]	6.4 (0.7) [4.2-14.5]	0.3 (0.0) [0.2-0.7]	2.6 (0.3) [1.5-6.0]
***P* value**	0.523	0.103	0.175	0.007	0.187	0.045	0.055	0.065

2.3.3 Soil bulk density

There was no significant difference in dry bulk density (BD - g/cm^3) between areas in the OT3 option and the counterfactual areas (*P*>0.05) even when differences in soil organic matter content between sites were taken into account, using residual values derived from predicted and measured BD values as explained in section 2.2.3.1 (Table 6 and Table S1 in Appendix 1).

Table 6 Soil physical properties: mean (standard error of the mean) and [range]

Site type	Dry bulk density (g/cm^3)	Poorest layer VESS score	Dispersion ratio
Counterfactual	1.19 (0.03) [0.78-1.59] n=75	2.33 (0.12) [1.0-4.0] n=45	5.1 (0.2) [3.1-7.6] n=45
OT3 option	1.19 (0.03) [0.70-1.61] n=75	2.32 (0.12) [1.0-4.0] n=45	4.9 (0.4) [1.9-14.3] n=45
P value	0.978	0.948	0.648

2.3.4 Soil structural condition and aggregate stability

2.3.4.1 Visual Evaluation of Soil Structure (VESS)

Individual VESS scores for the poorest layer at each site ranged from 1.0 to 4.0 (Table 6), placing all but 6 out of the 90 soil blocks assessed in the ''friable' to 'firm' structural quality classes (Table 7).

Table 7 VESS soil structure quality scores (Guimaraes *et al*., 2011)

Class	VESS score	Interpretation
1	<2	Friable
2	2-3	Intact
3	3-4	Firm
4	4-5	Compact
5	5	Very compact

At 8 out of 15 paired sites, the OT3 option field had a higher VESS score than the paired counterfactual field. The counterfactual paired field was higher at 6 sites, and for one of the paired sites both fields had the same score, indicating overall that that there was no difference in VESS soil structural quality between the OT3 option fields and conventional counterfactual fields.

Notably, when the 3 'poorest layer' VESS scores for each site were averaged, none of the field sites had topsoil layers that were scored as 'compact' or 'very compact' (Table S2 in Appendix 1).

There was no difference between OT3 and counterfactual sites in the distribution of poorest layer VESS scores (3 measurements per site for all 30 sites – Figure 2), indicating that neither the OT3 option nor conventional management had a clear impact on soil structural quality as assessed by visual scoring.

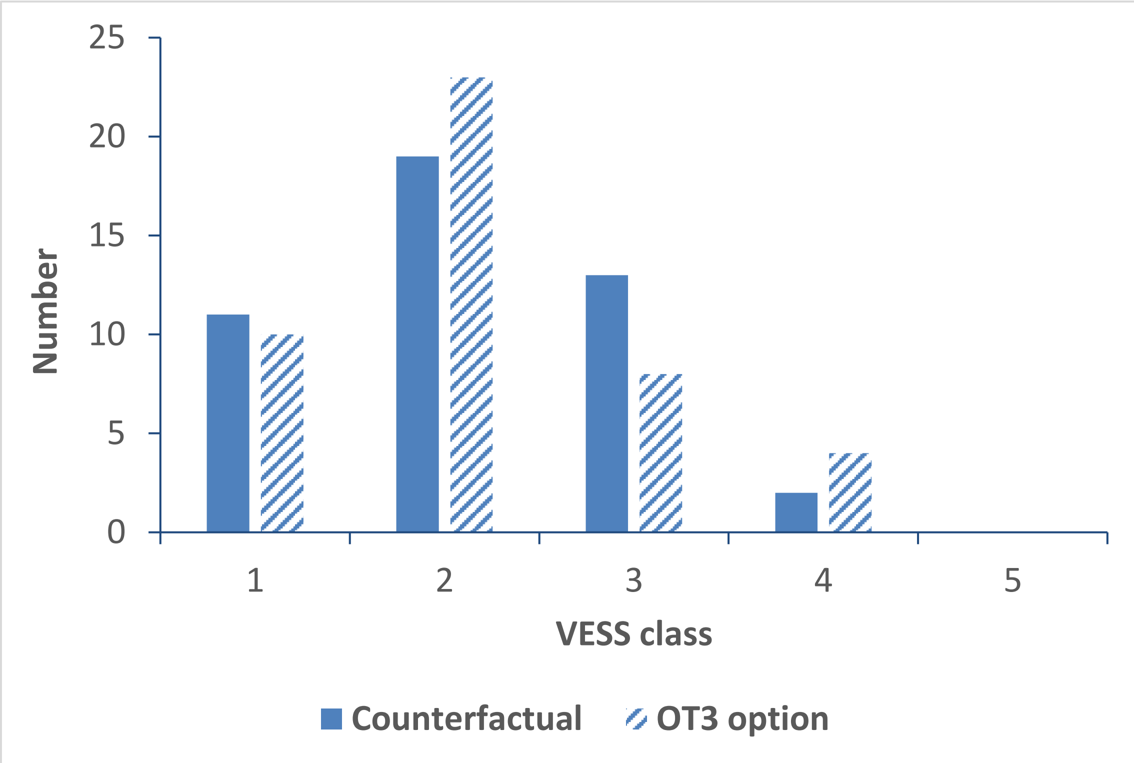

Figure 2 Distribution of poorest layer VESS scores: 3 measurements per site for all 30 sites (VESS class Table 7)

Analysis of variance (ANOVA) was used to investigate differences in VESS score between OT3 option fields and counterfactual fields; and between paired sites (Table 8).

The overall VESS mean for the poorest layer at both OT3 and counterfactual sites was 2.3. There was a significant difference in VESS score between the 15 paired sites (P=0.002, Table 8). In other words, soil structural quality varied between the 15 OT3-conventional pairs. However, there was no significant difference between land management 'treatments' (OT3 option vs counterfactual; P>0.05), indicating a spatial difference possibly relating to interactions between general agricultural management and abiotic factors rather than organic management.

Table 8 Results of ANOVA for the effect of paired site and land use on the VESS score for the poorest layer

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Pair	14	44.8722	3.2052	5.06	0.002
OT3 option/counterfactual	1	0.0028	0.0028	0.00	0.948
Residual	14	8.8722	0.6337		
Within site variation	60	5.3333	0.0889		
Total	89	59.0806			

There was also no significant difference in the overall soil block VESS score between land management 'treatments' (OT3 option vs counterfactual; $P>0.05$).

Additionally, the VESS scores for the poorest layer were converted into binomial data, where 1 was a VESS score of <3 and zero was a score of 3 or more. Regression analysis was then used to carry out a logistic regression on the binomial data and the proportion of the scores of <3 in each of the OT3 and counterfactual land management 'treatments'. In each analysis there was no significant difference between the two proportions ($P>0.05$), with 67% and 73% of the counterfactual and OT3 sites respectively scoring <3 (i.e. 'friable' to 'firm' structure).

2.3.4.2 Dispersion ratio

Mean dispersion ratio (DR – a measure of soil aggregate stability) values ranged from 1.8 to 11.6, placing them in the 'very stable' to 'fairly stable' stability classes (Table 9). Notably, there were no soil samples that were 'unstable' or 'very unstable'.

Table 9 Stability classes based on dispersion ratio determinations (ADAS, 1995)

Class	Dispersion ratio	Interpretation
1	<5	Very stable
2	5-10	Stable
3	10-15	Fairly stable
4	15-25	Slightly stable
5	25-30	Unstable
6	>30	Very unstable

Analysis of variance (ANOVA) was used to investigate differences in DR between the OT3 option fields and counterfactual fields; and between the paired sites (Table 8 and Table S3 in Appendix 1).

Even though in 12 out of the 15 paired sites, the OT3 option field had a lower DR than the counterfactual field (which would indicate that the OT3 option was resulting in more stable aggregates than conventional management), there was no significant difference in DR between the land management 'treatments' (OT3 option vs counterfactual; $P>0.05$; Table 8). The overall DR mean for the OT3 fields was 4.88 and for the counterfactual fields 5.13. There was also no significant difference between the 15 pairs ($P>0.05$; Table 8).

Table 8 Results of ANOVA for the effect of paired site and land use on DR

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Pair	14	139.3744	9.9553	1.52	0.223
OT3 option/counterfactual	1	1.4304	1.4304	0.22	0.648
Residual	14	91.9560	6.5683	7.52	
Within site variation	60	52.4020	0.8734		
Total	89	285.1628			

Although of low predictive value there was a significant negative relationship between DR and soil clay content (P=0.004; Figure 3) indicating that soil type may be an important factor influencing soil aggregate stability. This relationship was, however, highly influenced by one OT3 site, which had the highest DR (11.6) and the lowest clay content (7%). The regression analysis was re-run without this point and the percentage variance accounted for reduced to 7.9%.

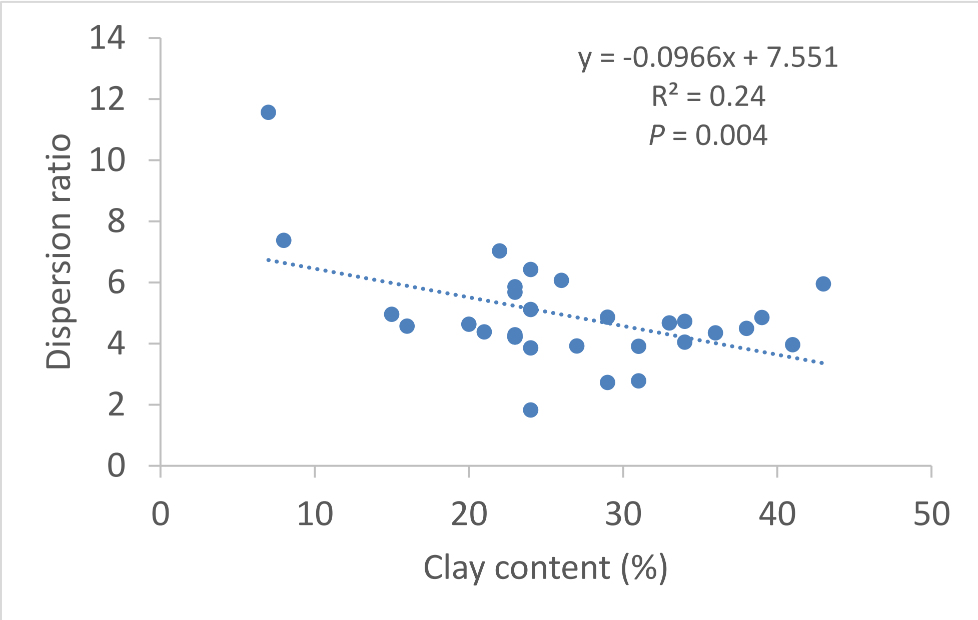

Figure 3 Relationship between dispersion ratio and soil clay content (%)

There was some indication, albeit with a very low predictive value (R^2 = 8.4%), of a significant negative relationship between DR and soil organic matter content (P=0.066) indicating that soil organic matter may have been a factor influencing soil aggregate stability. This relationship was greatly influenced by one OT3 site with a very high organic matter content (14.5%). Removal of this point, however, substantially increased the percentage variance accounted for (R^2 = 33%, Figure 4) and resulted in a highly significant (*P*<0.001) negative relationship (Figure 4). A further improvement in the percentage variance accounted for was observed with the additional removal of the highest DR (11.6) value (R^2 = 50.9%; *P*<0.001)). There was no relationship between DR and soil pH, soil extractable P, soil extractable K, soil extractable Mg, soil total-N content or soil organic carbon (*P*>0.05). Additionally, and somewhat surprisingly, there was no relationship between soil organic carbon content and soil clay content (*P*>0.05).

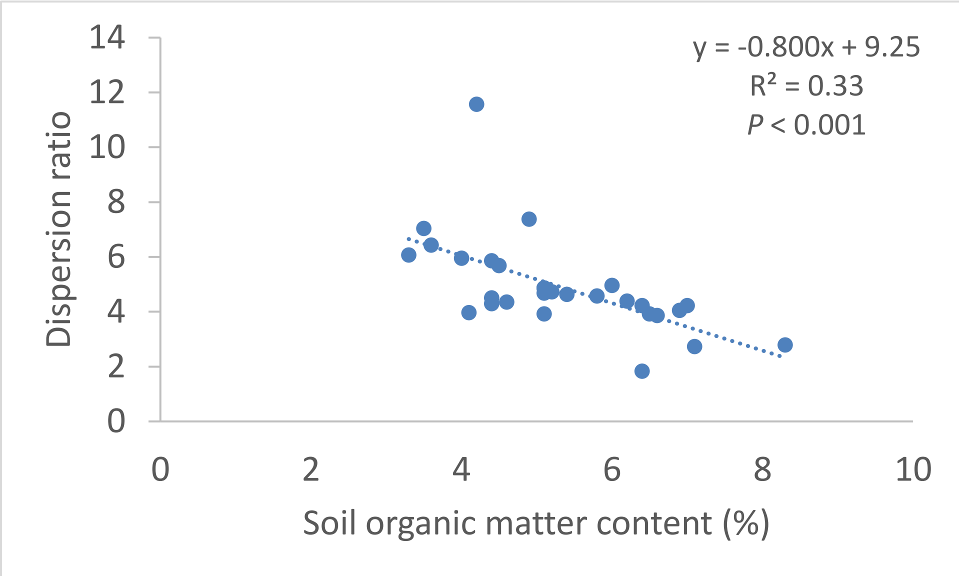

Figure 4 Relationship between dispersion ratio and soil organic matter content (%)

2.3.5 Earthworm sampling and midden counts

Analysis of variance (ANOVA) was used to investigate differences in total earthworm numbers, total earthworm biomass, number of juvenile worms, number of epigeic worms, number of endogeic worms, number of anecic worms and number of worm middens between the OT3 option fields and counterfactual fields; and between the paired sites. Prior to analysis, a log transformation was carried out and all mean values presented are back transformed.

There was a significant difference in the number of endogeic worms (P = 0.040) and the number of midden counts (P = 0.031) between the pairs (i.e. different locations) (mean = 77; P = 0.007). This significant difference in earthworm middens also is likely to indicate a difference in the population of anecic earthworms (Stroud *et al*., 2019). As the earthworm surveys took place in late February/early March, many anecic earthworms may still have been deep in the soil profile, so not counted in the survey as worms. The presence of middens, however, clearly indicates their presence and activity. There were also numerically more epigeic earthworms at the OT3 sites (mean = 1.27) when compared with the counterfactual sites (mean = 0.13), but the difference was not significant (P=0.062). Overall, there were no significant differences (P>0.05) in the earthworm variables measured between OT3 and counterfactual sites.

2.3.6 Soil erosion risk

2.3.6.1 Vegetation/residue percentage cover scores

At the counterfactual sites, the vegetation/residue cover did not exceed 30%, whereas the cover ranged from <20 to >90% cover on the OT3 option fields (Figure 5).

Analysis of variance (ANOVA) was used to investigate differences in vegetation/residue percentage cover between OT3 option fields and counterfactual fields; and between paired sites.

There was a significant difference in vegetation/residue cover (%) between OT3 (mean = 59%) and counterfactual sites (mean = 9%; P < 0.001), but no difference between the 15 pairs (P=0.770). At 14 out of 15 paired sites, the OT3 option field had a higher percentage vegetation/residue cover than the paired counterfactual field.

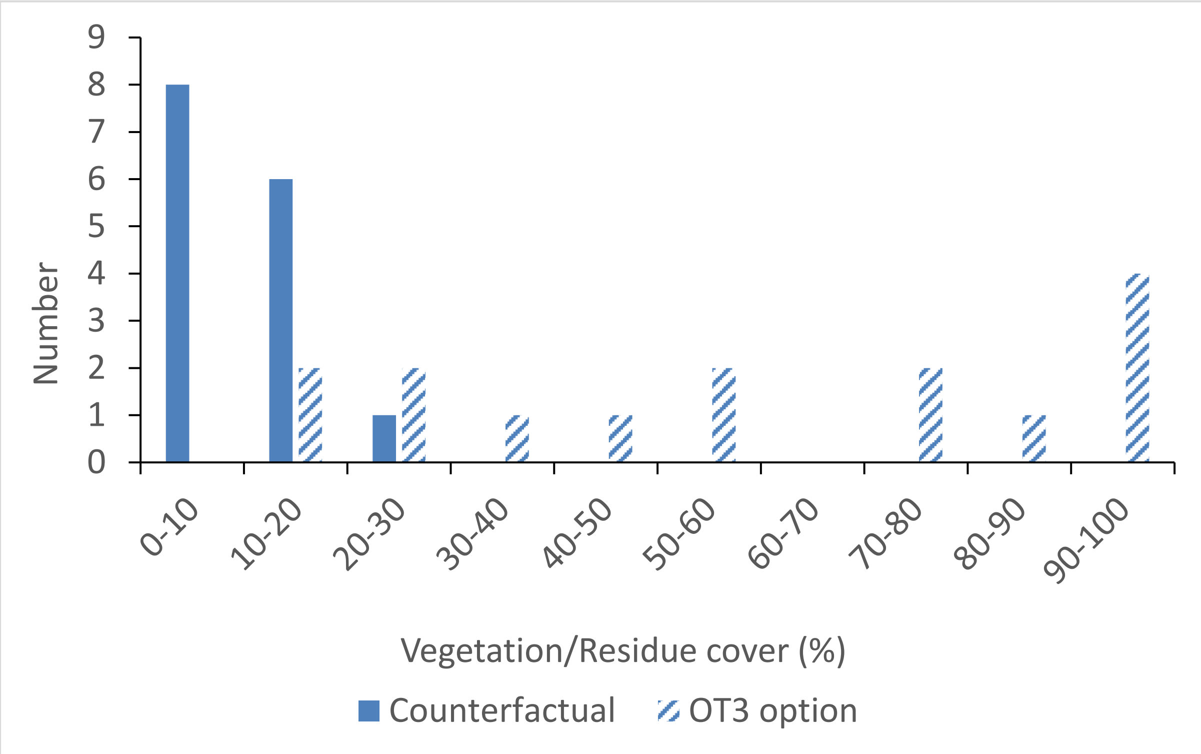

Figure 5 Distribution of percentage vegetation/residue cover for all 30 sites

Additionally, vegetation/residue percentage cover were assigned a score as used in the CS Baseline project (LM0458) and previous monitoring programmes (Table 9).

Table 9 Vegetation/residue cover scores (CS Baseline project - LM0458)

Score	Vegetation/residue cover (%)	Interpretation
3	95-100	Very High
2	80-94	High
1	65-79	Moderate
0	<65	Lower

Notably, all the counterfactual sites and over half of the OT3 sites scored 0 with a vegetation/residue cover of <65% (Figure 6). However, there was an imbalance in cropping between site types with 13 out of 15 OT3 sites in a cover crop, grass or stubbles and 10 out of 15 counterfactual sites in a cereal; 4 of the counterfactual fields were in stubbles and one had a cover crop.

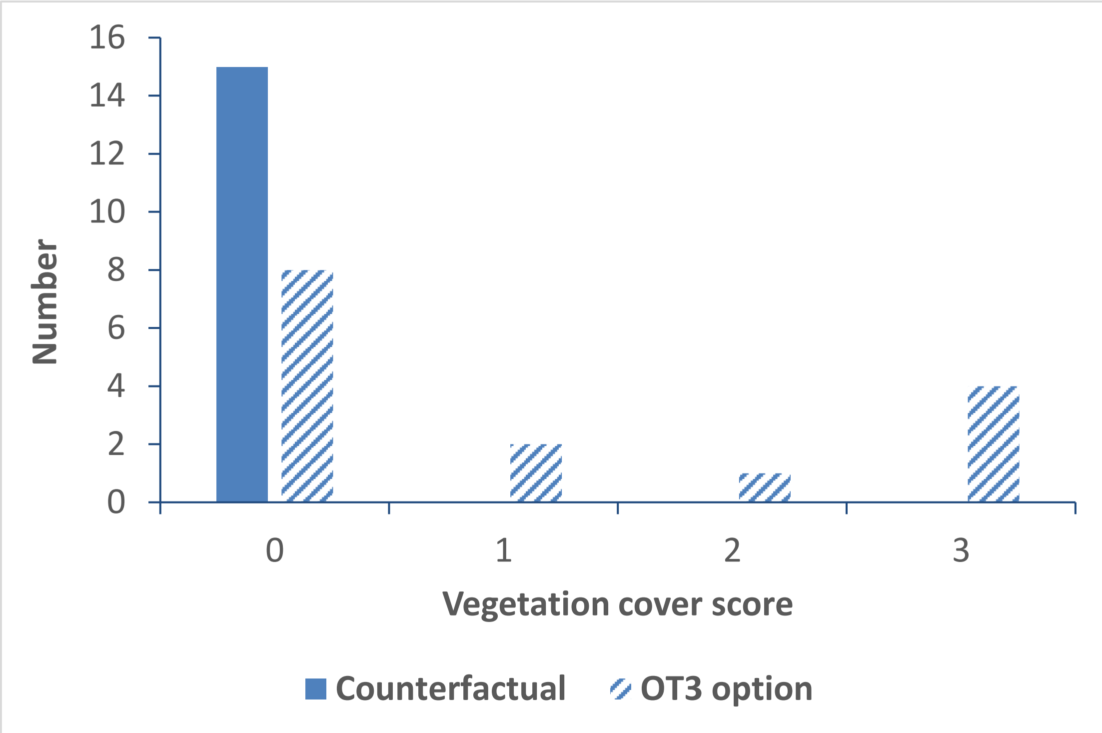

Figure 6 Distribution of vegetation/residue cover scores for all 30 sites

The vegetation/residue cover scores were converted into binomial data, where 1 was a vegetation/residue cover score of 1-3 and zero was a score of 0. Regression analysis was then used to carry out a logistic regression on the binomial data and the proportion of the scores of 1-3 in each of the OT3 and counterfactual land management 'treatments'. There was a significant difference between the two proportions (P<0.001), with 47% of the OT3 sites having a score of 1-3 (considered 'moderate' to 'very high'), whereas none of the counterfactual sites had a vegetation/residue cover score of 1-3.

2.3.6.2 Management condition scores

Based on the four scored factors described in section 2.5.6.2 (i.e. signs of erosion, signs of vehicle use, grazing pressure and vegetation/residue cover) the majority of counterfactual sites scored 0 or 1 i.e. the management was rated as 'poor' or 'adequate', and no fields were 'very good'. In contrast, most of the OT3 fields were scored as 2 or 3 i.e. 'good' or 'very good' (Table 10 and Table 11).

Table 10 Distribution of management condition scores

	Score			
	Poor	**Adequate**	**Good**	**Very good**
Site	**0**	**1**	**2**	**3**
Counterfactual	3	9	3	0
OT3 option	3	4	5	3

Table 11 Management condition scores (Jones *et al*., 2019; CS Baseline project - LM0458)

Score	Management condition	Interpretation
3	Very good	Management satisfies all the requirements/criteria to protect water
2	Good	Management satisfies all but one of the requirements/criteria to protect water
1	Adequate	Management results in some protection of water quality, but better management would have resulted in significantly improved protection
0	Poor	Poor management likely to result in water quality being compromised

The management condition scores were converted into binomial data, where 1 was a management condition score of 2 or 3 and zero was a score of 1 or 0. Regression analysis was then used to carry out a logistic regression on the binomial data and the proportion of the scores of 2 or 3 in each of the OT3 and counterfactual land management 'treatments'. There was a significant difference between OT3 and counterfactual sites (P=0.005), with 53% of OT3 sites scoring 'good' or 'very good', compared with 20% of counterfactual sites.

2.3.6.3 Water erosion risk

There was a very similar distribution of water erosion risk scores between OT3 option and counterfactual sites, with both treatments having relatively few fields at a 'lower' risk of water erosion (Table 12 and Table 13).

Table 12 Distribution of water erosion risk scores for all 30 sites

	Score		
	Lower	**Moderate**	**High or Very high**
Site	**0**	**1**	**2**
Counterfactual	3	5	7
OT3 option	3	4	8

Table 13 Water erosion risk scores (Defra, 2005)

Score	Water erosion risk
0	Lower risk
1	Moderate risk
2	High or very high risk

However, only 6 pairs had the same water erosion risk score, 5 pairs had a difference of 1 score, and 4 pairs had a difference of two scores (Table 14).

Table 14 The water erosion risk scores for the OT3 organic option and conventional counterfactual fields at 15 paired sites

Pair	OT3 option	Counterfactual
1	2	2
2	0	2
3	1	1
4	2	1
5	0	0
6	2	1
7	2	2
8	0	2
9	2	2
10	2	0
11	2	1
12	1	2
13	2	0
14	1	1
15	1	2

2.3.6.4 Erosion mitigation score results

Half of the sites had addressed the soil erosion risk ‘mostly’ or ‘completely’. However, there were 7 OT3 sites and 8 counterfactual sites that had addressed it only ‘partially’ or ‘not at all’ (Table 15 and Table 16).

Table 15 Distribution of erosion mitigation scores for all 30 sites

	Score			
	Not at all	Partially	Mostly	Completely
Site	**0**	**1**	**2**	**3**
Counterfactual	2	6	3	4
OT3 option	0	7	3	5

Table 16 Erosion mitigation scores (CS Baseline project - LM0458)

Score	Soil erosion mitigation
3	Completely
2	Mostly
1	Partially
0	Not at all

The erosion mitigation scores were converted into binomial data, where 1 was an erosion mitigation score of 2 or 3 and zero was a score of 1 or 0. Regression analysis was then used to carry out a logistic regression on the binomial data and the proportion of the scores of 2 or 3 in each of the OT3 and counterfactual land management 'treatments'. There was no significant difference between the two proportions ($P>0.05$), with 47% and 53% of the counterfactual and OT3 sites, respectively, having a score of 2 or 3 ('mostly' or completely' addressing the soil erosion risk).

2.3.6.5 Landscape connectivity score results

Most sites were assessed as having 'low' or 'very low' landscape connectivity, with eroded soil either 'very unlikely' or 'unlikely' ('but possible in high energy events') to reach a water course. Only 2 fields were assessed as having high landscape connectivity with eroded soil 'likely' to reach a watercourse in 'moderate events'. Both of these fields were in the OT3 organic option. No fields had 'very high' connectivity (Table 17 and Table 18).

Table 17 Distribution of landscape connectivity scores for all sites

	Score				
	Very low	**Low**	**Moderate**	**High**	**Very high**
Site	**1**	**2**	**3**	**4**	**5**
Counterfactual	13	2	0	0	0
OT3 option	11	2	0	2	0

Table 18 Landscape connectivity scores (CS Baseline project - LM0458)

Score	**Likelihood**	**Landscape connectivity**
1	Very Unlikely	Very low
2	Unlikely, but possible in high energy events	Low
3	Unlikely, but possible in moderate events	Moderate
4	Likely in moderate events	High
5	Highly likely in moderate events	Very High

The majority of pairs (9) had the same landscape connectivity score; 4 pairs had a difference of one score, and 2 pairs had a difference of two scores (Table 19).

Table 19 The landscape connectivity scores for the OT3 organic option and conventional counterfactual fields at 15 paired sites

Pair	OT3 option	Counterfactual
1	2	1
2	2	1
3	1	1
4	1	1
5	1	1
6	1	1
7	4	1
8	1	2
9	1	2
10	1	1
11	1	1
12	1	1
13	1	1
14	4	1
15	1	1

2.3.7 Soil Biota Communities

Thirty samples were sequenced on an Illumina MiSeq for 16S, ITS and 18S metabarcoding, specifically looking for Bacteria, Fungi and Nematodes respectively. The same 30 samples were also sequenced on Oxford Nanopore's MinION, specifically for COI (insects/invertebrates) metabarcoding. The sequencing yielded over 11 million DNA sequences as shown in Table 20.

Table 20 The number of reads generated for each barcode gene. COI reads are not paired-end

Gene	Target Taxa	Reads
16S	Bacteria	3293397
ITS	Fungi	3480685
18S	Nematodes	2473319
COI	Insects/Invertebrates	2156258

The reads were assigned to taxa as described. The attached 'X_barplot.qza' files contain the results of this assignment for 16S, 18S and ITS and can be viewed with https://view.qiime2.org/. A relative abundance matrix can be extracted from each file by selecting the CSV option at the top left of the qiime2 view website. The relative abundance matrix for the COI data is attached as 'COI_abundance.tsv'.

Various analysis was then carried out on this abundance data looking for the ability to discriminate between the organic and conventional farming practice.

2.3.7.1 Diversity

Alpha diversity analysis was conducted with the Qiime2 package for 16S, ITS and 18S datasets. Alpha diversity can look at both richness (the number of taxa) and evenness (the distribution of taxa) within a treatment groups.

For alpha diversity, there was no significant difference between the organic and conventional treatment groups with regards to species richness in any of the datasets. For species evenness, however, a significant difference was observed between treatment groups in the 16S dataset, with organic samples generally having a more even spread of taxa (Figure 7).

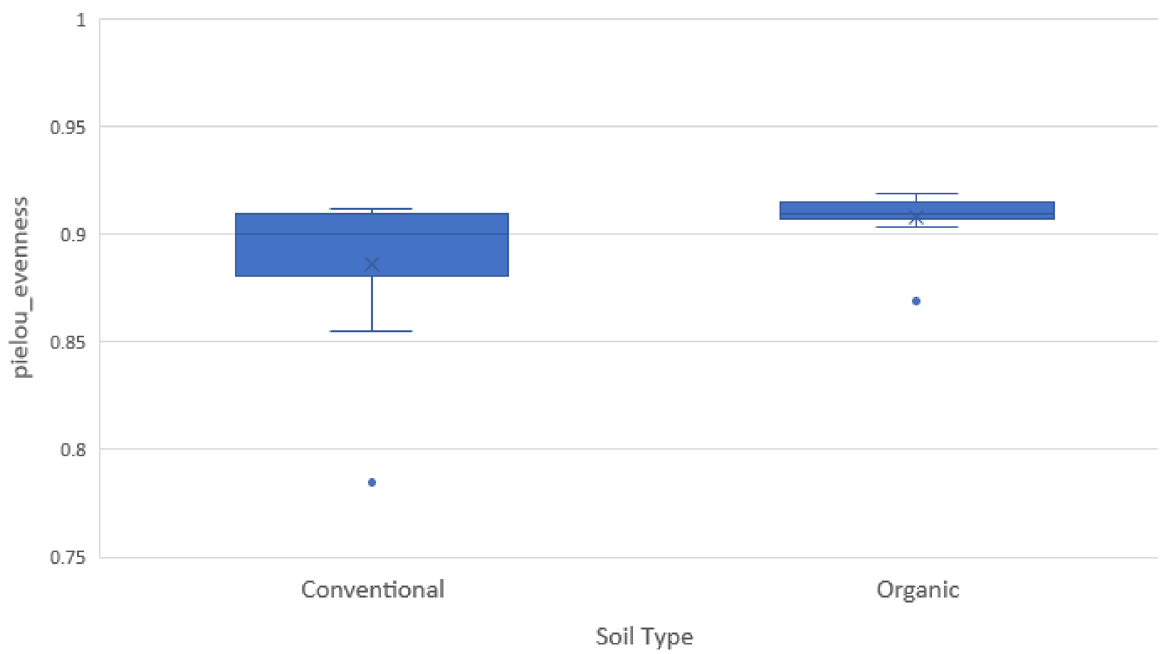

Figure 7 Boxplot displaying the 16S taxonomic evenness observed between the organic (n=13) and conventional (n=11) treatment groups, with taxa more evenly distributed in organic systems than in conventional systems (H=6.6469, d.f.=22, P=0.0099). X indicates the average evenness for each soil type.

2.3.7.2 Potential Biomarkers

Linear discriminant analysis Effect Size, or LEfSe (Segata *et al.*, 2011), is a method that aims to determine features, such as taxa, which are most likely to explain differences between treatment groups, such as organic/conventional farms. LEfSe analysis was carried out on all datasets with any ‘unassigned’ or ‘unresolved’ taxonomic labels removed as these are uninformative. As the sample size was small, stricter alpha values for the Kruskal-Wallis test (0.01), and the LDA threshold for discriminative features (4) were set. At these thresholds, no taxa were found in the COI and 18S datasets that could be confidently called ‘biomarkers’ for a particular treatment type. However, there are a number of potential biomarkers for the 16S and ITS datasets. A full list of these can be found in the accompanying ‘X-Biomarkers’ files.

There are a number of observations picked up by the LEfSe analysis for the 16S and ITS datasets. For the 16S data, most notable was the genus of bacteria, *Gaiella*, which appeared to be consistently more abundant in organic soils than in conventional soils. Hermans *et al*. (2017) reported that members of the *Gaiellaceae* family expressed a strong correlation with carbon-to-nitrogen levels. Similarly, members of the *Solibacteraceae* family and the genus *Pseudolabrys* were also identified as more relatively abundant in organic soils. The former plays an active role in protein and carbohydrate mineralisation (Wyszkowska *et al*., 2019), with the latter being hydrocarbon degraders and are incidentally often located in hydrocarbon rich soils (Miao *et al*., 2019). Interestingly, there were more bacterial biomarkers associated with organic soils than conventional soils, which may suggest a greater diversity of bacteria although this was not shown in the alpha diversity analysis.

For the ITS dataset, fewer biomarkers overall were identified. Where they were identified, the relative abundances were generally lower than the 16S dataset, thus any associations may be less reliable. However, the fungal genus *Hymenoscyphus* was positively associated with organic soils, whereas *Chaetomium* was positively associated with conventional soils

There were no biomarkers identified by LEfSe in the 18S or COI datasets for the parameters used in this analysis.

2.3.7.3 Worm Counts

COI based sequencing reads containing worms were split into *Annelida* and 'No *Annelida*' subsets (with unassigned/unresolved reads removed). These data were compared to worm counts conducted using traditional methods and described in section 2.2.3.4. Descriptive statistics can be seen in Table 21.

Table 21 Descriptive statistics for worm count and *Annelida* relative abundances

	Worm Count		*Annelida* Relative Abundance (%)	
	Organic	**Conventional**	**Organic**	**Conventional**
Mean	7.78	7.53	57.60	54.46
Median	7.33	7.67	59.88	58.68

Two generalised linear models (family = Poisson) were built to test the null hypotheses that A) Treatment groups have no significant impact on worm count (traditional data) and B) Treatment groups have no significant impact on the relative abundance of *Annelida* (sequencing data) in the samples. P-values of 0.838 and 0.328 were observed from hypotheses A and B respectively, both results suggesting that there is no significant difference between the number of worms observed in organic and conventional soils.

2.4 Discussion

It was notable that most of the soils were in generally moderate to good condition. Based on dispersion ratio tests and VESS assessments, all soils were in the 'very stable' to 'fairly stable' stability classes and 84 of the 90 soil blocks assessed had moderate to good structural quality (friable to firm); only six blocks had a compact layer; 4 from OT3 sites and 2 from counterfactual sites. This generally good level of structural stability and quality is probably a reflection of the use of organic manures in recent years at most sites, which can have a positive effect on soil physical quality (Bhogal *et al*. 2009, Johnston *et al*., 2009).

There were no differences between OT3 and counterfactual sites in soil dry bulk density, VESS scores, dispersion ratio (structural stability) scores, earthworm biomass/number or erosion risk mitigation. Indeed, dispersion ratio scores were more influenced by soil clay content and soil organic matter content. This lack of statistical differences may, in part, be due to the relatively small sample sizes in this initial study, whereas a larger sample size may detect differences. However, it is also possible that a larger sample size could potentially show fewer differences between systems.

The main differences between the site types were found in soil organic matter content, vegetation/residue covers and management condition, with OT3 sites having generally better scores than the counterfactual sites for these indicators. OT3 sites had higher soil organic matter content than the counterfactual sites despite the OT3 sites having lower clay content on average. This was probably a reflection of the contrasting rotations; all of the OT3 land was in an arable-grass ley rotation, whereas 14 out of 15 counterfactual sites were in an arable rotation. Arable-grass ley rotations are generally better than arable rotations at retaining or slowly increasing soil organic matter levels (Johnston *et al*., 2009).

The lower vegetation covers on the counterfactual sites were mainly due to differences in cropping. Thirteen out of the 15 OT3 sites were in a cover crop, grass or stubbles, while 10 out of 15 counterfactual sites were in a cereal: winter wheat and oats, which generally have a significant proportion of bare soil in late winter/early spring. Lower vegetation cover at the counterfactual sites was the main factor resulting in generally lower management condition scores, compared with the OT3 sites, although signs of erosion and vehicle use on the generally arable land were also contributing factors at some sites.

It is interesting, however, that the only two sites that were assessed as having high landscape connectivity with eroded soil 'likely' to reach a watercourse in 'moderate events' were OT3 sites. Landscape connectivity is a function of landscape configuration and how flow pathways are broken up by management interventions and features (e.g. hedgerows, in-field buffer strips and grass margins), and can be as important as soil erosion in determining impacts on water quality (Boardman *et al*., 2019), so it was interesting that OT3 management did not always afford high protection to local watercourses.

Genetic analysis was used to describe the microbiome and fauna of soil samples from conventional and organic systems. Both the LEfSe and alpha diversity analyses in this report had small sample sizes. LEfSe alpha values were adjusted to 0.01, which is in-line with the example for low-cardinality datasets (Segata *et al*., 2011). As such, only P values <0.01 were considered statistically significant. With a larger sample size, alpha values could have been relaxed to 0.05, where P values <0.05 would be considered statistically significant. Observations in significant alpha diversity differences could be similarly impacted by a small sample size and this should be considered when examining the results.

There were significant differences in alpha diversity with the organic system sites having significantly more evenness in their bacterial populations. This suggests that although there were similar numbers of bacterial taxa in both conditions the distribution of these species was more equal in the organic samples.

The LEfSe analysis identified a number of bacterial taxa, and to a lesser degree fungal taxa, associated with the organic samples. These may be potential biomarkers of organic farming and their significance should be explored in larger datasets, in order to examine their role and determine if they could be used to confirm organic status.

Although no significant difference in earthworm numbers was recorded between organic and counterfactual sites, the sample size was small: only 10 conventional samples and 12 organic samples had both sequencing and count data available for them. As such, any conclusions drawn from this analysis could differ greatly from a study with a much larger sample size. In addition, as noted above, due to the timing of sampling, many anecic earthworms were likely to be deep in the soil and therefore beyond the sampling depth.

Although organisms may be taxonomically very different, many may fall into similar ecological roles. As such, interpreting the functional role of identified taxa may lead to more insightful interpretation of the data. Tools such as FAPROTAX (Louca *et al*., 2016), FUNGuild (Nguyen *et al*., 2016) and NEMAGuild aim to provide this functional interpretation. These tools are currently still in the development stage. Future work could build on the data in this project and explore whether changes in microbial community structure leads to changes in community function.

3 LANDSCAPE CHARACTERISTICS

3.1 Introduction

There are numerous studies that have looked at the impact of organic land management and landscape complexity on biodiversity conservation (e.g. Aavik & Liira 2010, Carrié *et al*. 2017, Gabriel *et al*. 2010, Roschewitz *et al.* 2005, Rundlöf & Smith, 2006). These studies have shown that landscape structure does have impacts on biodiversity conservation, while the presence of organic management generally only enhances biodiversity conservation when located in more homogenous landscapes. The majority of these have considered management and landscape complexity as independent variables in the analysis rather than concentrating on whether organic landscapes have a different level of complexity to conventionally farmed landscapes. In addition, a number of these studies have used simple proxies such as the proportion of arable land (Fischer *et al*. 2011, Roschewitz *et al*. 2005, Winqvist *et al*. 2011) or the proportion of unimproved and improved grassland (Power, Kelly & Stout 2012, Purtauf *et al.* 2005) in place of true diversity metrics, and some broadly split landscapes into homogeneous and heterogeneous classifications based on these proxies. This means these studies consider landscape complexity in quite a simplistic way.

Organically managed landscapes are generally believed to be more complex than agricultural landscapes under conventional management (Krebs, 1999). However, there are relatively few studies that have tested the differences in the structure of land under organic and conventional management and the results are ambiguous. van Mansvelt, Stobbelaar & Hendriks (1998) found that land use and crop diversity was greater on organic farms than conventional farms, with larger number of structural features such as hedgerows and in-field trees. Gibson *et al*. (2007) found no significant difference in the landscape structure of organic and conventional farms. Norton *et al*. (2009) found a significant difference in the complexity of landscapes surrounding cereal cropping fields on organic farms compared to those on conventional farms, again both in terms of crop diversity and hedgerow density.

In this study we look at landscapes containing eight case study agreements to test whether areas with high densities of organic Countryside Stewardship options have significantly different landscape structure in terms of habitat parcel size, number and shape to similar areas where these options are not present.

3.2 Methods

3.2.1 Spatial analysis

The analysis uses the UKCEH Land Cover Map 2015 (LCM2015; Rowland *et al*. 2017) to define the spatial pattern of parcels in the landscape. LCM2015 classifies land into 21 habitat classes (Annex 1) and the spatial framework is derived from a combination of OS Mastermap and RPA Rural Land Registry parcels (for England). In this project we have used the vector version of LCM2015, which identifies the dominant land cover class for each parcel with a minimum size of 0.5 ha. Eight agreements (six in the Midlands and two in Yorkshire and the Humber; Figure 8) containing organic options were identified to act as case studies for the testing of the landscape structure analysis procedure. In most cases the organic options cluster around a single core location. However, in some cases the organic options are found in two or more clusters. The centroid defines the geographic centre of a distribution of points, and in this analysis is used to centre a 5x5 km grid (Figure 9) within which we compare the landscape structure of the central 1 km grid cell with similar neighbouring cells. However, when there are two or more distinct clusters of organic options within an agreement the centroid can be located in an area without organic options present meaning the comparison between the central cell and the neighbouring cells is not valid. In these cases, we sub-divided the organic options to produce spatially coherent groups and ran the analysis against each subset. This process means that the analyses have been performed on 11 clusters of organic options across the eight agreements.

Figure 8 Case study locations

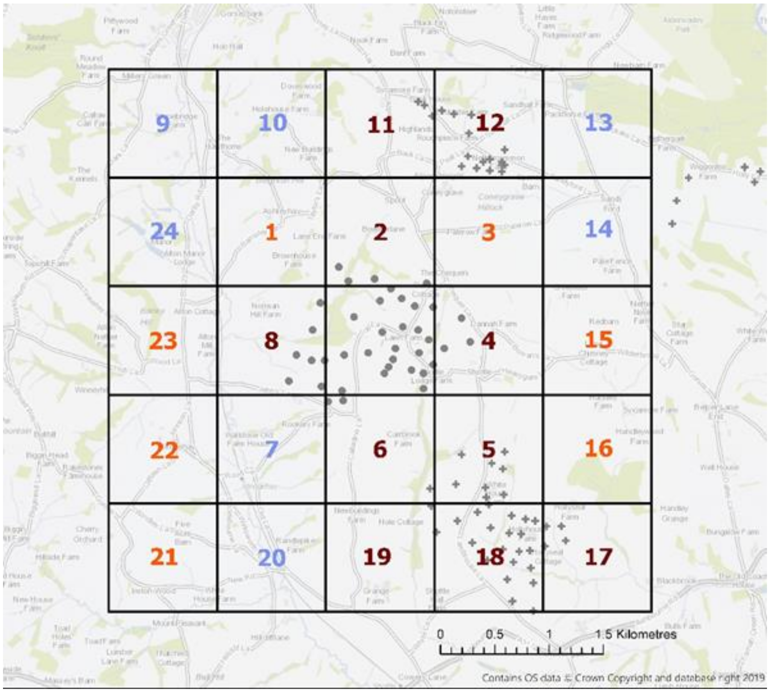

Figure 9 Example of 5x5 grid (Case Study 7, South cluster) used to compare landscape structure between central organic grid cell and conventional landscapes in the surrounding cells. • focal organic CS options, + organic CS options not associated with the focal group. Orange grid cells: used in the comparison. Brown grid cells: discounted cells due to containing organic options. Blue grid cells: discounted due to differences in habitat composition.

For each agreement (or option cluster), the centroid of the organic options was determined and the 5x5 grid of 1 km^2 cells centred on the centroid was created (Figure 9). The grid IDs for cells are then joined to each parcel in the land cover map 2015 (LCM2015) which they intersect. This creates duplicates of the LCM2015 parcels where more than one grid cell intersects the parcel. The spatially joined LCM2015 parcels are then dissolved based on grid ID and habitat and the proportion of each habitat associated with the grid cell is calculated. These proportions are used to calculate the multidimensional distance between the land cover in the central focal cell and the land cover in each of the 24 surrounding cells. Cells with a land cover distance greater than the threshold of 0.25 were dropped from the analysis to ensure the comparisons of landscape structure were made between cells with similar habitat compositions.

Figure 10 provides a flow diagram of the process for selection of comparison cells. Prior to the eight case studies being identified, the landscape methodology was developed on an agreement containing organic options as a development case. This development case sat near the edge of moorland and represents a landscape with a variety of habitat composition within the neighbouring cells. The similarity threshold of 0.25 was chosen based on an examination of the multidimensional distances for the development case and should not be taken as the definitive threshold for future studies until validated against a wider range of landscape types. The remaining cells that contain organic options were identified and dropped from the analysis to produce the final set of grid cells to compare to the landscape structure of the focal cell.

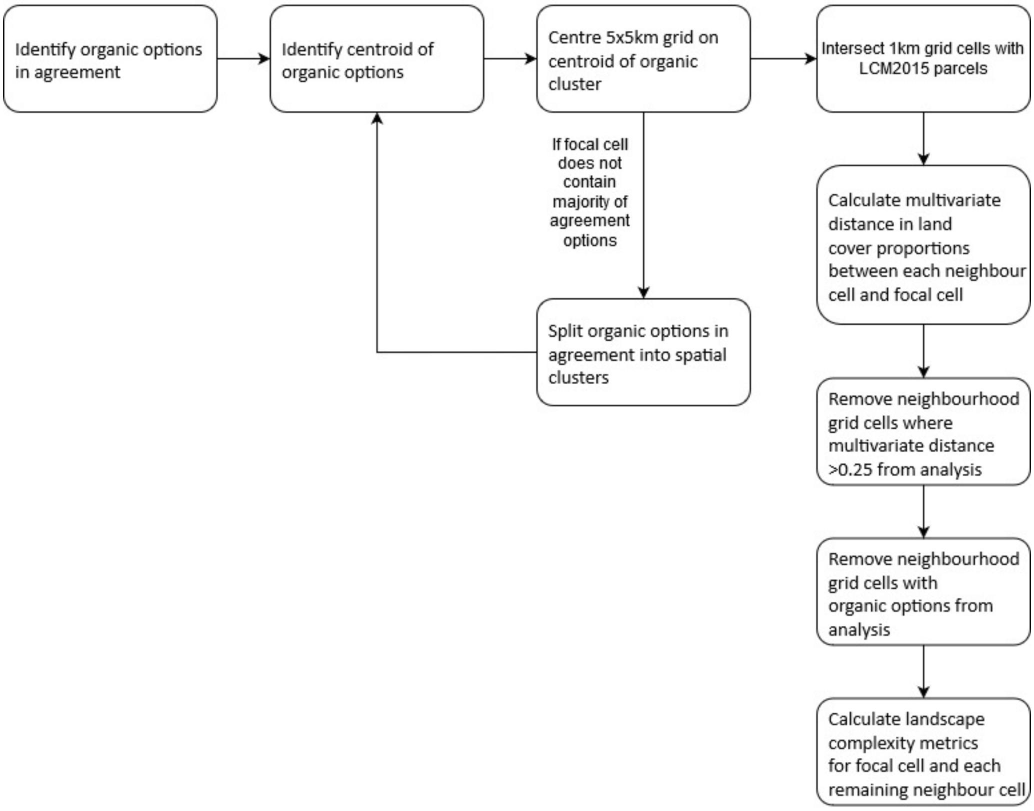

Figure 10 Flow diagram of process to perform the comparison in landscape complexity between the focal cell containing organic options and similar non-organic grid cells in the neighbourhood

3.2.2 Landscape Metrics

Metrics describing the structure of the landscape were calculated for each grid cell as a whole and, in the case of parcel count, mean parcel size and mean area-perimeter ratio, also for each habitat present within the grid cell. The whole grid metrics for the comparable cells were then summarised to give the mean, median and upper and lower 95% (t_{n-1}*standard deviation; where t_{n-1} is the 95% t-value for sample size-1 degrees of freedom) sample intervals about the mean. Significance is tested by checking if the metric of the focal cell falls outside of the 95% sample interval (SI) for the metric in the neighbourhood.

1. Parcel Count (Whole grid and habitat)

The parcel count represents the total number of parcels intersected by a grid cell. Parcels can subdivide larger habitat patches so the count can include multiple contiguous parcels of the same habitat.

2. Average Parcel Size (Whole grid and habitat)

The area of parcels intersected by a grid cell are used to calculate the mean parcel size for that grid cell. A parcel will contribute its full area to the calculation of the parcel mean area for each grid cell, leading the total parcel area associated with each grid cell to be in excess of 1 km^2. This is done to prevent artificial parcel boundaries delimited by the positioning of the edge of the 1km grid cell from biasing the parcel size metric.

3. Mean Area-Perimeter Ratio (Whole grid and habitat)

The area-perimeter ratio for each parcel is calculated and the mean value taken from all the parcels that intersect a grid cell. As with the Average Parcel Size, the full parcel is used in the calculations rather than clipping the parcels to the 1 km^2 grid cell. As with the parcel area using the whole parcel area and perimeter is done to prevent introducing artificial parcel boundaries delimited by the positioning of the edge of the 1km grid cell from biasing the parcel area-perimeter ratio metric.

4. Edge Density (Whole grid)

Edge density (m/km^2) is calculated by first clipping the parcels to the 1 km^2 grid cell and then calculating the total boundary length for the clipped parcels. A value of 4000 m is subtracted from the perimeter total to account for the artificial external boundary length created by clipping the parcels to the grid cell. The remainder is halved to account for each internal boundary being double counted (caused by each boundary being represented in the perimeter measures of two parcels). Unlike the previous two metrics, the edge density metric is not biased by the intersection of the grid cell boundary and the parcel boundaries and so can be calculated appropriately using only the edges within the grid cell.

5. Habitat Diversity (Whole grid)

Habitat diversity is calculated from the total area of each habitat associated with the parcels intersecting the grid cell and converting this to the proportion of the total parcel area (p_i) for each habitat. The proportions are used to calculate the Shannon diversity index according to Equation 1.

$$-\sum p_i \cdot \ln(p_i) \qquad \text{Equation 1}$$

3.3 Results

Each of the case studies listed below represents one of the agreements identified in Figure 8. The area of interest for each case study is one (or occasionally two) 5x5 km squares centred on the organic options present in the agreement. Each of these squares is split into 25 1km grid cells as described in section 3.2.1. The whole grid metrics are summarised in the tables below and in the figures presented in Annex 2, with the more detailed breakdown by grid cell and land cover class presented in Annexes 3 and 4.

3.3.1 Case Study 1

As the locations of the organic options for this agreement cluster into two distinct spatial groups, the options were split into north-west (NW; 15 options) and south-east (SE; 25 options) clusters. The central grid cell of the NW cluster contains two land cover classes (arable and horticulture, and improved grassland) while the larger area of interest contains five additional land cover classes (broadleaf woodland, freshwater, neutral grassland, suburban, and urban). For this cluster, six of the neighbouring cells (1,9,13,14,19,23) were suitable for comparison as a non-organic landscape of a similar habitat composition as the central grid cell. The central grid cell of the SE cluster contains four land cover classes (arable and horticulture, broadleaf woodland, improved grassland, and suburban) while the larger area of interest contained three additional land cover classes (freshwater, neutral grassland, and urban). For the SE cluster seven of the neighbouring cells (11,12,13,15,16,17,23) were suitable for comparison as a non-organic landscape of a similar habitat composition as the central grid cell. None of the landscape metrics for the focal cell in either the NW cluster (Table 22) or SE cluster (Table 23) were significantly different from the expectation based on the distribution of the metrics from the neighbouring cells.

Table 22 Landscape metrics of the central focal cell and the comparable non-organic neighbouring cells (n=6) for the NW cluster of organic options for Case Study 1

		Parcel Count	Average Parcel Size (ha)	Average AP Ratio	Edge Density (m/km^2)	Shannon Diversity
Focal cell		37	4.59	47.1	9,794	0.63
Neighbourhood	Mean	49.3	3.63	39.2	12,066	0.65
	Median	53.0	3.39	37.3	12,194	0.66
	Lower SI	23.5	1.84	28.5	7,256	0.47
	Upper SI	75.1	5.42	49.9	16,875	0.82

Table 23 Landscape metrics of the central focal cell and the comparable non-organic neighbouring cells (n=7) for the SE cluster of organic options for Case Study 1

		Parcel Count	Average Parcel Size (ha)	Average AP Ratio	Edge Density (m/km^2)	Shannon Diversity
Focal cell		50	3.45	37.2	12,900	0.77
Neighbourhood	Mean	44.9	4.22	40.9	11,339	0.68
	Median	44	4.36	42.25	11,058	0.71
	Lower SI	34.6	2.76	33.1	9,173	0.42
	Upper SI	55.1	5.68	48.8	13,505	0.94

3.3.2 Case Study 2

The central grid cell contains four land cover classes (arable and horticulture, broadleaf woodland, improved grassland, and suburban) with an additional four land cover classes (calcareous grassland, coniferous woodland, inland rock, and urban) present in the larger area of interest. Seven of the neighbouring cells (5,7,17,18,19,20,23) were suitable for comparison as a non-organic landscapes of a similar habitat composition as the central grid cell. None of the landscape metrics for the focal cell in this Case Study were significantly different from the expectation based on the distribution of the metrics from the neighbouring cells (Table 24).

Table 24 Landscape metrics of the central focal cell and the comparable non-organic neighbouring cells (n=7) for Case Study 2

		Parcel Count	Average Parcel Size (ha)	Average AP Ratio	Edge Density (m/km^2)	Shannon Diversity
Focal cell		47	3.53	40.1	11,576	0.78
Neighbourhood	Mean	62.6	2.61	32.5	13,926	0.90
	Median	64.0	2.32	31.0	14,398	0.91
	Lower SI	34.8	1.07	23.3	9,461	0.61
	Upper SI	90.4	4.14	41.8	18,390	1.19

3.3.3 Case Study 3

The central grid cell contains four land cover classes (arable and horticulture, broadleaf woodland, improved grassland, and suburban) with an additional three land cover classes (freshwater, inland rock, and urban) present in the larger area of interest. Ten of the neighbouring cells (3,9,11,13,15,17,20,21,23,24) were suitable for comparison as a non-organic landscapes of a similar habitat composition as the central grid cell. None of the landscape metrics for the focal cell in this Case Study were significantly different from the expectation based on the distribution of the metrics from the neighbouring cells (Table 25).

Table 25 Landscape metrics of the central focal cell and the comparable non-organic neighbouring cells (n=10) for Case Study 3

		Parcel Count	Average Parcel Size (ha)	Average AP Ratio	Edge Density (m/km^2)	Shannon Diversity
Focal cell		50	3.69	37.9	12,655	1.08
Neighbourhood	Mean	45.2	4.93	39.4	11,493	0.90
	Median	44.5	4.23	38.4	11,108	0.87
	Lower SI	15.5	0.16	26.2	6,703	0.51
	Upper SI	74.9	9.71	52.5	16,283	1.30

3.3.4 Case Study 4

The central grid cell contains three land cover classes (arable and horticulture, broadleaf woodland, and improved grassland) with an additional two land cover classes (inland rock, and suburban) present in the larger area of interest. Seventeen of the neighbouring cells (1,3,4,5,8,9,10,11,12,13,15,16,18,19, 22,23,24) were suitable for comparison as a non-organic landscape of similar habitat composition as the central grid cell. None of the landscape metrics for the focal cell in this Case Study were significantly different from the expectation based on the distribution of the metrics from the neighbouring cells (Table 26).

Table 26 Landscape metrics of the central focal cell and the comparable non-organic neighbouring cells (n=17) for Case Study 4

		Parcel Count	Average Parcel Size (ha)	Average AP Ratio	Edge Density (m/km^2)	Shannon Diversity
Focal cell		27	7.83	55.9	8,204	0.30
Neighbourhood	Mean	23.5	10.47	60.6	7,113	0.31
	Median	21.0	11.24	64.4	6,229	0.27
	Lower SI	7.0	3.54	37.7	3,184	0
	Upper SI	39.9	17.40	83.6	11,042	0.77

3.3.5 Case Study 5

The central grid cell contains four land cover classes (arable and horticulture, broadleaf woodland, improved grassland, and suburban) with an additional two land cover classes (freshwater, and urban) present in the larger area of interest. One of the neighbouring cells (16) was suitable for comparison as a non-organic landscape of similar composition as the central grid cell (Table 27).

Table 27 Landscape metrics of the central focal cell and the comparable non-organic neighbouring cells for Case Study 5

		Parcel Count	Average Parcel Size (ha)	Average AP Ratio	Edge Density (m/km^2)	Shannon Diversity
Focal cell		42	3.75	38.9	11,198	0.81
Neighbourhood	1 cell	58	2.83	33.4	13,244	1.14

3.3.6 Case Study 6

The central grid cell contains four land cover classes (arable and horticulture, broadleaf woodland, improved grassland, and suburban) with an additional three land cover classes (coniferous woodland, freshwater, and urban) present in the larger area of interest. Five of the neighbouring cells (7,21,22,23,24) were suitable for comparison as a non-organic landscape of similar composition as the central grid cell. None of the landscape metrics for the focal cell in this Case Study were significantly different from the expectation based on the distribution of the metrics from the neighbouring cells (Table 28).

Table 28 Landscape metrics of the central focal cell and the comparable non-organic neighbouring cells (n=5) for Case Study 6

		Parcel Count	Average Parcel Size (ha)	Average AP Ratio	Edge Density (m/km^2)	Shannon Diversity
Focal cell		60	2.62	35.6	13,460	0.56
Neighbourhood	Mean	64.4	2.46	32.3	14,517	0.32
	Median	65.0	2.36	31.4	14,913	0.33
	Lower SI	45.0	1.37	23.0	10,197	0
	Upper SI	83.8	3.56	41.6	18,837	1.00

3.3.7 Case Study 7

The options were split into a North cluster (11 options) and a South cluster (38 options). The central grid cell for the North cluster contains six land cover classes (arable and horticulture, broadleaf woodland, calcareous grassland, coniferous woodland, improved grassland, and suburban) with an additional six land cover classes (acid grassland, freshwater, heather, inland rock, neutral grassland, and urban) present in the larger area of interest. For this cluster, one of the neighbouring cells (4) was suitable for comparison as a non-organic landscape of similar composition as the central grid cell (Table 29).

The central grid cell for the South cluster contained three land cover classes (arable and horticulture, broadleaf woodland, and improved grassland) with an additional seven land cover classes (acid grassland, calcareous grassland, coniferous woodland, freshwater, inland rock, suburban, and urban) present in the larger area of interest. For this cluster, 12 of the neighbouring cells (1,3,7,9,10,13,14,15,20,21,22,23) were suitable for comparison as a non-organic landscape of similar composition as the central grid cell.

The focal cell for the South cluster has a significantly larger parcel size with a corresponding higher area-perimeter ratio when compared to the parcel characteristics of the neighbourhood grid cells (Table 30).

Table 29 Landscape metrics of the central focal cell and the comparable non-organic neighbouring cells for North cluster of Case Study 7

		Parcel Count	Average Parcel Size (ha)	Average AP Ratio	Edge Density (m/km^2)	Shannon Diversity
Focal cell		78	1.90	29.0	15,812	1.25
Neighbourhood	1 cell	66	2.34	32.7	14,536	0.95

Table 30 Landscape metrics of the central focal cell and the comparable non-organic neighbouring cells (n=12) for South cluster of Case Study 7

		Parcel Count	Average Parcel Size (ha)	Average AP Ratio	Edge Density (m/km^2)	Shannon Diversity
Focal cell		55	**3.09**	**34.9**	13,273	0.46
Neighbourhood	Mean	77.8	1.92	29.0	16,180	0.44
	Median	75.5	1.95	28.9	15,801	0.50
	Lower SI	49.1	1.17	24.1	12,452	0
	Upper SI	106.5	2.68	34.0	19,908	1.00

3.3.8 Case Study 8

The options were split into an East cluster (18 options) and a West cluster (26 options). The central grid cell for the East cluster contains four land cover classes (arable and horticulture, improved grassland, suburban, and urban) with an additional five land cover classes (broadleaf woodland, coniferous woodland, freshwater, heather grassland, and urban) present in the larger area of interest. For this cluster, seven of the neighbouring cells (1,3,15,16,21,22,23) were suitable for comparison as a non-organic landscape of similar composition as the central grid cell.

The central grid cell for the West cluster contains four land cover classes (arable and horticulture, broadleaf woodland, improved grassland, and suburban) with an additional four land cover classes (freshwater, inland rock, neutral grassland, and urban) present in the larger area of interest. For this cluster, seven of the neighbouring cells (2,15,16,17,18,19,21) were suitable for comparison as a non-organic landscape of similar composition as the central grid cell. None of the landscape metrics for the focal cell in in either the East cluster (Table 31) or West cluster (Table 32) were significantly different from the expectation based on the distribution of the metrics from the neighbouring cells.

Table 31 Landscape metrics of the central focal cell and the comparable non-organic neighbouring cells (n=7) for East cluster of Case Study 8

		Parcel Count	Average Parcel Size (ha)	Average AP Ratio	Edge Density (m/km^2)	Shannon Diversity
Focal cell		50	4.46	39.8	12,388	0.82
Neighbourhood	Mean	36	5.48	44.9	10,385	0.80
	Median	32	5.28	44.1	9,663	0.76
	Lower SI	8	0.97	22.0	4,309	0.47
	Upper SI	63	10.00	67.8	16,460	1.14

Table 32 Landscape metrics of the central focal cell and the comparable non-organic neighbouring cells (n=7) for West cluster of Case Study 8

		Parcel Count	Average Parcel Size (ha)	Average AP Ratio	Edge Density (m/km^2)	Shannon Diversity
Focal cell		44	4.04	36.3	12,551	0.77
Neighbourhood	Mean	46	4.06	37.4	12,436	0.81
	Median	47	4.03	36.9	11,902	0.80
	Lower SI	29	2.10	22.8	8,348	0.54
	Upper SI	64	6.01	52.0	16,523	1.08

3.4 Discussion

The selection process for cells is undertaken to minimise the variation between the focal organic cell and cells being used as counterfactuals. However, it should be noted that this analysis does not explicitly account for the amount of conventional land in the focal cell. We also did not have a record of all organic land present in each case study area, so there may also be organically managed land outside of Countryside Stewardship that could confound the analysis. If the study had been designed with experimental control we would select multiple pairs of farms where the main difference is whether the farm is conventionally or organically managed (Weibull, Östman & Granqvist 2003) and other factors such as farm size, type and situation in the landscape are controlled for. However, a control study would require a much larger number of paired farms than it was possible to include in this project. Instead we follow a similar logic to Norton *et al*. (2009), where spatial proximity between the focal (organic) study area and the counterfactual (conventional) areas substitutes for the variety of environmental and socioeconomic variables that would be considered in pairing up sites in a control study. Even so there can be significant changes in habitat composition over relative short distances, for example grassland systems transitioning to moorland at an altitudinal limit or pastoral systems transitioning to agro-forestry. Therefore we have a step to exclude cells where the habitat composition is significantly different to the focal cell, as the drivers leading to and associated with a different habitat composition are likely to have a larger influence on landscape structure than whether the land is under conventional management.

For two of the clusters of options in this study (Case Study 5 and North cluster of Case Study 7), there was only one comparable conventional cell in the neighbourhood of the focal cell. This highlights the difficulty of finding appropriate counterfactuals at the landscape scale. The size of the neighbourhood could be increased but as the distance from the focal cell increases, so it is expected that the differences in historic and contemporary drivers for the landscape will become more significant. For the whole analysis there is only one option cluster where metrics for the focal cell were significantly different (at α=0.05) from the mean for the comparable non-organic cells. This was the South cluster of Case Study 7 which has a significantly larger mean parcel size and mean area-perimeter ratio for the central focal cell compared to the comparable non-organic cells in the neighbourhood. However, given the number of individual statistical tests (9 sites * 5 metrics = 45 tests for significance) being undertaken and the covariance between different landscape metrics, these two metrics being significantly different from the neighbourhood means may well be down to random chance.

None of the focal cells have a parcel count that is statistically significantly different from the comparable non-organic grid cells in the neighbourhood. There were six instances where the parcel count was less than the mean of the neighbourhood and four instances where the parcel counter was greater than the mean for the neighbouring cells, with one case where the parcel count was equal to the mean. Only the South cluster of Case Study 7 had a mean parcel size that was significantly different from the comparable neighbouring cells, with a total of five instances where the mean parcel size for the focal cell was larger than the neighbourhood mean, and six instances where it was smaller than the neighbourhood mean. Similarly, for the area-perimeter ratio only the focal cell in the South cluster of Case Study 7 was significantly different to the neighbourhood mean, with six instances where the area-perimeter ratio of the focal cell is greater than the mean of the neighbourhood, and five instances where it is less than the mean for the neighbourhood. Both the edge density and habitat diversity measures have the same split of six instances where the edge density and habitat diversity are greater than the neighbourhood mean and five instances where they are less than the neighbourhood mean. Overall there is no indication from the grouping of results of a consistent difference in landscape structure where organic options are a major component of the landscape.

Future studies may find hedge density a more useful metric than edge density derived from parcel boundaries from LCM2015. The boundaries of LCM2015 parcels represent a variety of boundary types including culturally and biologically important features (such as hedges, stone walls and hedge banks), simple fence lines (which contribute little to the functioning of the landscape) and purely notional boundaries between habitats with no physical feature present where the boundary line has been placed. The study had originally planned to look at hedge density using the RPA mapping of hedgerows recorded for the Basic Payment Scheme and Countryside Stewardship. However, it was not possible to include the dataset within the timeframe of this project.

While Norton *et al*. (2009) found a significant difference in landscape complexity between organic and conventional farms, their approach answers a slightly different set of questions to that addressed here. They compared national averages for the two types of management rather than examining how the complexity of the local landscape differs under organic versus conventional management. Additionally, they had performed ground surveys of the habitats and features associated with the fields within the study, and so had finer scale data with which to test differences.

4 CONCLUSIONS & RECOMMENDATIONS

4.1 Conclusions

Most of the OT3 and counterfactual sites had moderate to good levels of soil structural stability and quality. This was probably a reflection of the use of organic manures in recent years at most sites, whether they were organically or conventionally managed. OT3 sites had higher soil organic matter content than the counterfactual sites despite the OT3 sites having lower clay content on average. This was probably a reflection of the contrasting rotations; all of the OT3 land was in an arable-grass ley rotation, whereas most of the counterfactual sites were in an arable rotation.

OT3 sites also had higher vegetation/residue percentage covers than the counterfactual sites, which was also reflected in higher management condition scores. This was also most probably due to differences in cropping between the sets of OT3 and counterfactual sites. Most of the OT3 sites were under a grass or cover crop, whereas most of the counterfactual sites were in a winter cereal. Despite these differences, it was clear from field observations that OT3 management did not always afford adequate protection to local watercourses due to high landscape connectivity (a combination of inherent risk and management) at two sites.

There were very few differences in soil physical and chemical properties between fields in the organic OT3 (rotational land) option and conventional counterfactual fields.

Differences were detected in the biome of soils collected from organic and conventional farming and potential biomarker species identified but the small scale of the sampling means that these can only be preliminary findings at this stage.

This landscape study did not detect a consistent significant difference between the structure of landscapes where organic options are present and those landscapes of comparable habitat composition without organic options. This could be due to a small sample size and reliance on the LCM2015 which is mainly derived from existing parcel mapping from OS and the RPA and which may not be best suited to identifying all significant landscape features. However, it is more likely that organic options are a relatively weak driver of landscape structure compared to the landscape history and the common technological and business constraints that drive contemporary management decisions. Norton *et al* (2009) noted that the small number of farms where the organic management had been in place for decades strongly influenced the statistics for the organic farms as a whole, with those farms that had recently converted to organic having very variable field sizes compared to smaller fields observed on long-term organic farms. This effect of time since conversion to organic management may explain the variation we see in the case studies, where both higher and lower levels of complexity are observed in the landscape structure for the organic focal cell when compared to neighbouring cells.

A larger study that combines sub-parcel mapping of habitats and features (such as hedges, stone walls, in-field trees, etc.) with information on landscape history (for example how long the land has been under organic management and identifying other organic land not under option) along similar lines to that undertaken by Norton *et al.* (2009) may be able to detect differences in landscape structure between organic and conventional landscapes. The costs involved with collecting sub-parcel habitat information, either through ground survey or very high-resolution aerial and satellite imagery, would be high. This cost would be compounded by the additional resources required to collate information on the landscape history and the scaling up required to produce a statistically valid sample. However, there has been little indication from this study to suggest that a more intensive study with a larger sample size would produce more definitive results than presented here.

Overall, the range of soil assessments conducted in this study would be appropriate for use in a national study to identify any differences between farming systems. The pilot study indicated that there may be improvements in soil quality (as indicated by soil organic matter content) and land management condition that result from support for organic farming. A larger study would provide a more reliable indication of whether these differences or any other differences are consistent across the population of organic agreements or whether differences are mainly driven by other biotic, abiotic or management factors.

4.2 Recommendations

4.2.1 Timing of sampling

Sampling and survey for assessing soil quality and risks to water quality are best carried out when soils are at field capacity and field drains are running, typically between mid-October and mid-April. The survey period for this pilot project was constrained to the months of January to March, which was probably the best option for assessing erosion features and landscape connectivity as it came towards the end of the autumn to winter period when vegetation covers are generally lower and erosion and runoff events are most likely to occur. Any erosion features resulting from rainfall events earlier in the season are 'preserved' in the landscape and can still be observed at this time. Nevertheless, from practical (limited opportunities to carry out field work due to varying weather conditions) and scientific (most of the soil quality variables do not change significantly while the soil is at field capacity) perspectives, a larger scale survey would need to be carried out from October to April to cover the main period when soil quality assessments can be most effectively carried out, erosion events are more likely to occur and to allow for interruptions to the field work programme due to inclement weather. Consideration could be given to focusing earthworm sampling in October-November and March-April to maximise opportunities for recording earthworms (Singh *et al.*, 2021; Stroud, 2019). Timing of sampling will also be important for interpretation of genetic analysis of the soil microbiome.

4.2.2 Collection of management data

For this pilot study it was only possible to collect limited management data. However, this information was extremely useful in explaining some of the differences found between OT3 and counterfactual sites, putting the measurements and observations into a more general, but equally relevant context. A more detailed survey questionnaire with more specific information about field cropping and management history would provide additional data to help explain the variation in soil quality, management condition and risks to water quality.

4.2.3 Range of assessments

The sampling structure and range of assessments carried out were sufficient to determine differences in soil quality or soil erosion risk mitigation between the 15 site pairs and between OT3 and counterfactual sites. A larger sample size may, however detect differences not measured in this pilot project, and would allow more investigation of the relationship between abiotic factors and the soil quality and resource protection variables being investigated.

4.2.4 Selecting paired sites

A GIS assessment was used to select paired sites with similar soil type and erosion risk. However, in the resulting sample there were differences in the baseline water erosion risk score in 9 out of 15 pairs, and a difference in landscape connectivity score in 6 out of 15 pairs. This could potentially be improved through using digital elevation modelling data. However, it is important to note that slope is only one of many factors that influence erosion rate and the risk of soil and sediment losses to water. The key point is that sites should present some risk to local water quality, due to a degree of slope and proximity to a watercourse. Sites with more erodible soil types (e.g. soils with sandy or light silty topsoil texture) should also be favoured for investigation. Obtaining paired sites with identical levels of erosion risk is probably not possible whichever database or remote sensing technique is used to select them, as a degree of 'ground truthing' will always be needed.

4.2.5 Use of remote sensing in site selection and assessment

A fieldwork approach has been shown to be effective in assessing soil quality and risks to water quality on land under contrasting agricultural management systems. However, field work requires considerable resource and could be complemented by using GIS and remote sensing methods. ADAS and other organisations have been exploring the potential of using Sentinel-2 and PlanetScope satellite imagery to investigate and map soil erosion and landscape features at different temporal and spatial scales. Satellite and drone imagery can be used to observe erosion features over time at field and catchment scale. However, part of the challenge is that erosion is very sporadic both spatially and temporally (Evans *et al*., 2016), so to obtain a reliable picture of soil erosion risk in the landscape, imagery will be required over a number of years. It is also important to determine the minimum level of image resolution that is needed to spot erosion features and provide some assessment of landscape connectivity.

Satellite imagery may be able to complement field survey by aiding sample selection and providing additional data for extrapolation, but is unlikely to be able to replace it entirely. For example, Archer *et al*. (2014) concluded that "earth observation cannot replace field sampling but can complement it by: i) helping to improve sampling schemes, and ii) providing an exhaustive covariate (available across large parts of the landscape) that if used in a statistical model can substantially reduce the uncertainty in predictions". Earth observations could potentially be used to help detect change over time, but this has not been demonstrated. For example, Archer *et al*. (2014) stated that "there have been no published studies to date which have demonstrated that a change in SOC or total N can be measured [at national scale], using only remotely sensed methods".

4.2.6 Assessing landscape connectivity

Boardman *et al*. (2019) have asserted that landscape connectivity is more important than erosion rates in determining off-site impacts of erosion and runoff. Landscape connectivity is therefore an important consideration in assessing the impact of land management on water quality. This pilot project used a methodology developed by ADAS that was first employed at national scale in the CS Baseline project (LM0458). It is a survey method relying on field observations and surveyor experience and expertise that could be developed further or complemented by imagery. For example, Google Earth and other images can be very helpful for spotting and delineating field runoff and erosion features, and flow between fields. However, the timing of the image is critical to picking up such features.

For the field survey, additional written guidance could be provided to surveyors to look out for features indicating flow into water courses or between fields. It is particularly important to look carefully at field boundaries, especially hedges, for signs of flow between fields and to look for drains/culverts between fields and under roads and tracks.

Digital elevation models (DEMs) could also be combined with soil maps to provide an assessment of the degree of landscape connectivity. DEMs of sufficient resolution can give good information on flow routing connectivity between a field, any sedimentation areas and a watercourse. However, a DEM cannot determine critical depths for over-topping, whether flow can leave a field via a drain or culvert or how landscape elements -- including field boundaries, roads, tracks and ditches -- act as 'valves' to control the flow of runoff and sediment.

4.2.7 Assessment of Landscape Characteristics

The main issue which makes the assessment of the impact of management on landscape character difficult is that landscapes are the outcome of natural and human processes over time, making the landscape at each location the unique product of the processes at that location. In addition, it also means that any management changes require enough time *in situ* to produce a detectable change. The usual way to deal with this type of observational situation is to use a repeated measures or panel data approach where measurements for each location are recorded at multiple points through time. In this way it is possible to consider each location in reference to the initial measurements made at that location under the assumption that the influence of unmeasured variables remains constant over the period of observation and is independent of the treatment variables. This type of study design is difficult to implement retrospectively following the historic introduction of organic management, although sometimes enough historic data are available to construct the panel. Future studies may want to identify holdings where historic maps and farm plans are available which would allow an assessment of the historic landscape character to be undertaken for comparison with the characteristics of the current landscape. Additionally, a panel of farms that have recently entered organic management, or are planning to enter organic management in the near future, could be identified. The landscape character metrics could then be measured prior to, and over the course of, the organic agreement.

Hedge density (data unavailable for this project) may be a more useful metric than edge density derived from parcel boundaries from LCM2015. The boundaries of LCM2015 parcels represent a variety of boundary types including culturally and biologically important features (such as hedges, stone walls and hedge banks), simple fence lines (which contribute little to the functioning of the landscape) and purely notional boundaries between habitats with no physical feature present where the boundary line has been placed. Hedges represent a structurally complex physical boundary feature which require ongoing management. They are recognised as important both culturally and ecologically for their influence on landscape character, but are also under pressures from agricultural intensification both directly from a drive to increase field size for more efficient use of mechanisation and indirectly from the use of agro-chemicals which can impact on their ecological functioning. Identifying if hedgerow density varies between organic and conventional landscapes would be useful for informing policy on the retention and restoration of this habitat.

5 REFERENCES

Aavik T & Liira J. (2010). Quantifying the effect of organic farming, field boundary type and landscape structure on the vegetation of field boundaries. *Agriculture, Ecosystems & Environment*, 135:178-186.

ADAS. (1995). Structural Stability Assessment Using the Dispersion Ratio Method. ADAS Experimental Protocol.

Ahmed M, Back MA, Prior T, Karssen G, Lawson R, Adams I & Sapp M. (2019). Metabarcoding of soil nematodes: the importance of taxonomic coverage and availability of reference sequences in choosing suitable marker(s). *Metabarcoding and Metagenomics,* 3.

Archer A, Rawlins B, Grebby S, Marchant B & Emmett B. (2014). Identify the opportunities provided by developments in earth observation and remote sensing for national scale monitoring of soil quality. British Geological Survey Internal Report, SP1316_C. 32pp.

Bhogal A, Nicholson F A & Chambers B J. (2009). Organic Carbon Additions: Effects on Soil Bio-Physical and Physico-Chemical Properties. *European Journal of Soil Science*, 60:276-286.

Boardman J, Vandaele K, Evans R & Foster DL. (2019). Off-site impacts of soil erosion and runoff: Why connectivity is more important than erosion rates. *Soil Use and Management,* 35 (2):245-256: https://doi.org/10.1111/sum.12496

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H ….. & Walters W, (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology,* 37:852-857.

Budge GE, Adams I, Thwaites R, Pietravalle S, Drew GC, Hurst GDD, Tomkies V, Boonham N & Brown M. (2016). Identifying bacterial predictors of honey bee health. *Journal of Invertebrate Pathology,* 141:41-44.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, and Madden TL. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10**,** 421.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N & Knight R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences,* 108:4516-4522.

Carey P, Dimambro M, Rayns F. (2019). Countryside Stewardship organic management and conversion options: A scoping study to establish a monitoring protocol: Methods, power analysis and costing. Report to Natural England.

Carrié R, Andrieu E, Ouin A & Steffan-Dewenter I. (2017). Interactive effects of landscape-wide intensity of farming practices and landscape complexity on wild bee diversity. *Landscape Ecology*, 32**:**1631-1642.

Defra (2005). Defra Controlling Soil Erosion: A Manual for the Assessment and Management of Agricultural Land at Risk of Water Erosion in Lowland England (revised 2005, PB4093). Available to download from: http://www.defra.gov.uk/environment/land/soil/publications.htm.

Evans R, Collins AI, Foster IDL, Rickson RJ, Anthony SG, Brewer T, Deeks L, Newell Price JP, Truckell IG and Zhang Y. (2016). Extent, frequency and rate of water erosion of arable land in Britain –

benefits and challenges for modelling. *Soil Use and Management* 32(1), 149-161. doi: 10.1111/sum.12225.

Fischer C, Flohre A, Clement LW, Batáry P, Weisser WW, Tscharntke T & Thies C. (2011). Mixed effects of landscape structure and farming practice on bird diversity. *Agriculture, Ecosystems & Environment*, 141:119-125.

Folmer O, Black M, Hoeh W, Lutz R, and Vrijenhoek R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol,* 3:294-299.

Gabriel D, Sait SM, Hodgson JA, Schmutz U, Kunin WE & Benton TG. (2010). Scale matters: the impact of organic farming on biodiversity at different spatial scales. *Ecology Letters*, 13**:**858-869.

Gaukroger CH, Stewart CJ, Edwards SA, Walshaw J, Adams IP & Kyriazakis I. (2020). Changes in Faecal Microbiota Profiles Associated With Performance and Birthweight of Piglets. *Frontiers in Microbiology* 11.

Gibson RH, Pearce S, Morris RJ, Symondson WOC & Memmott J. (2007). Plant diversity and land use under organic and conventional agriculture: a whole-farm approach. *Journal of Applied Ecology*, 44**:**792-803.

Guimaraes RML, Bal, BC & Tormena CA. (2011). Improvements in the visual evaluation of soil structure. *Soil Use and Management,* 27:395–403.

Hermans SM, Buckley HL, Case BS, Curran-Cournane F, Taylor M & Lear G. (2017). Bacteria as Emerging Indicators of Soil Condition. *Applied and Environmental Microbiology* 83**,** e02826-02816.

Hole D G, Perkins A J, Wilson JD, Alexander IH, Grice PV & Evans AD. (2005). Does organic farming benefit biodiversity? *Biological Conservation*, 122:113-130.

Illumina (2013). *16S Metagenomic Sequencing Library Preparation Guide* [Online]. Available: http://emea.support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf [Accessed].

Johnston AE, Poulton PR, Coleman K. (2009). Soil Organic Matter: Its importance in Sustainable Agriculture and Carbon Dioxide Fluxes. In D.L. Sparks (ed.) Advances in Agronomy 101. Burlington: Academic Press, 2009:1-57.

Jones N, Conyers S, Crowe A, Elliott J, Cao Y, Newell Price P, Gooday R, O'Seanechain D, Haigh D, Forster Brown C & Adams I. (2019). CS Baseline project - LM0458 Final Report. The Environmental Effectiveness of the Countryside Stewardship scheme; Establishing a baseline agreement monitoring sample. REF: ECM47452/22965 (ITT 1630).

Krebs JR, Wilson JD, Bradbury RB & Siriwardena GM. (1999). The second silent spring? *Nature*, 400:611-612.

Lotter DW. (2003). Organic agriculture. *Journal of Sustainable Agriculture*, 21:59-128.

Louca S, Parfrey LW & Doebeli M. (2016). Decoupling function and taxonomy in the global ocean microbiome. *Science,* 353:1272-1277.

Marriott EE & Wander MM. 2006. Total and labile soil organic matter in organic and conventional farming systems. *Soil Science Society of America Journal*, 70:950-959.

Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* 17**,** 3.

Merrington G, Fishwick S, Barraclough D, Morris J, Preedy N, Boucard T, Reeve M, Smith P & Fang C. (2006). The development and use of soil quality indicators for assessing the role of soil in environmental interactions. Environment Agency Science Report SC030265, Bristol, UK.

Miao Y, Johnson NW, Gedalanga PB, Adamson D, Newell C & Mahendra S. (2019). Response and recovery of microbial communities subjected to oxidative and biological treatments of 1,4-dioxane and co-contaminants. *Water Research,* 149:74-85.

Nguyen NH, Song Z, Bates ST, Branco S, Tedersoo L, Menke J, Schilling JS & Kennedy PG. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology,* 20:241-248.

Norton L, Johnson P, Joys A, Stuart R, Chamberlain D, Feber R, Firbank L, Manley W, Wolfe M, Hart B, Mathews F, Macdonald D & Fuller RJ. (2009). Consequences of organic and non-organic farming practeces for field, farm and landscape complexity. *Agriculture, Ecosystems & Environment*, 129:221-227.

Ortega SF, Ferrocino I, Adams I, Selvestri S, Spadaro D, Lodovica Gullino M & Boonham N. (2020). Monitoring and surveillance of aerial mycobiota of rice paddy through DNA metabarcoding and qPCR. Journal of Fungi, 6:372 https://doi.org/10.3390/jof6040372.

Papadopoulos, A., Bird, N. R. A., Whitmore, A. P. and Mooney, S. J. (2009). Investigating the effects of organic and conventional management on soil aggregate stability using X-ray computed tomography. *European Journal of Soil Science* 60 (3), 360–368. doi:10.1111/j.1365-2389.2009.01126.x. ISSN 1351-0754.

Peerlkamp PK. (1967). Visual estimation of soil structure. In: de Boodt, M., de Leenherr, D.E., Frese, H., Low, A.J., Peerlkamp, P.K. (Eds.), West European Methods for Soil Structure Determination, vol. 2, no. 11, State Faculty Agric. Sci., Ghent, Belgium, pp. 216–223.

Porter TM & Hajibabaei M. (2018). Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. *Molecular Ecology,* 27: 313-338.

Power EF, Kelly DL & Stout JC. (2012). Organic farming and landscape structure: effects on insect-pollinated plant diversity in intensively managed grasslands. PloS one, 7(5).

Pulleman, M., Creamer, R., Hamer, U., Helder, J., Pelosi, C., Pérès, G. & Rutgers, M. (2012). Soil biodiversity, biological indicators and soil ecosystem services - an overview of European approaches. *Curr. Opin. Environ. Sustain.* 4, 529–538. doi: http://dx.doi.org/10.1016/j.cosust.2012.10.009.

Purtauf T, Roschewitz I, Dauber J, Thies C, Tscharntke T & Wolters V. (2005). Landscape context of organic and conventional farms: influences on carabid beetle diversity. *Agriculture, Ecosystems & Environment*, 108:165-174.

Roschewitz I, Gabriel D, Tscharntke T & Thies C. (2005). The effects of landscape complexity on arable weed species diversity in organic and conventional farming. *Journal of Applied Ecology*, 42:873-882.

Rowland CS, Morton RD, Carrasco L, McShane G, O'Neil AW & Wood CM. (2017). Land Cover Map 2015 (vector, GB). NERC Environmental Information Data Centre.https://doi.org/10.5285/6c6c9203-7333-4d96-88ab-78925e7a4e73.

Rundlöf M & Smith HG. (2006). The effect of organic farming on butterfly diversity depends on landscape context. *Journal of Applied Ecology*, 43:1121-1127.

Scullion J, Eason W & Scott E. 1998. The effectivity of arbuscular mycorrhizal fungi from high input conventional and organic grassland and grass-arable rotations. *Plant and Soil*, 204:243-254.

Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS & Huttenhower C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol,* 12:R60.

Seufert V & Ramankutty N. (2017). Many shades of gray—The context-dependent performance of organic agriculture. *Science advances,* **3,** e1602638.

Shepherd M. (2017). Taking the long view - an introduction to Natural England's long term monitoring network 2009 – 2016: Long Term Monitoring Network soils data and protocols. Natural England: ISBN 978-1-78354-448-6.

Singh J, Cameron E, Reitz T, Schädler M, Eisenhauer N. (2021). Grassland management effects on earthworm communities under ambient and future climatic conditions. *Eur J Soil Sci.* 2021;72:343–355. https://doi.org/10.1111/ejss.12942.

Stolze M, Piorr A, Häring AM & Dabbert S. (2000). *Environmental impacts of organic farming in Europe,* Universität Hohenheim, Stuttgart-Hohenheim.

Stroud JL. (2019). Soil health pilot study in England: Outcomes from an on-farm earthworm survey. PLoS ONE 14(2).

Toju H, Tanabe AS, Yamamoto S & Sato H. (2012). High-Coverage ITS Primers for the DNA-Based Identification of Ascomycetes and Basidiomycetes in Environmental Samples. *PLoS ONE* 7**,** e40863.

van Mansvelt JD, Stobbelaar DJ & Hendriks K. (1998). Comparison of landscape features in organic and conventional farming systems. *Landscape and Urban Planning*, 41:209-227.

Weibull AC, Östman Ö & Granqvist Å. (2003). Species richness in agroecosystems: the effect of landscape, habitat and farm management. *Biodiversity & Conservation*, 12**:**1335-1355.

Winqvist C, Bengtsson J, Aavik T, Berendse F, Clement LW, Eggers S, ... & Pärt T. (2011). Mixed effects of organic farming and landscape complexity on farmland biodiversity and biological control potential across Europe. *Journal of Applied Ecology*, 48:570-579.

Woodhall JW, Webb KM, Giltrap PM, Adams IP, Peters JC, Budge GE & Boonham N. (2012). A new large scale soil DNA extraction procedure and real-time PCR assay for the detection of Sclerotium cepivorum in soil. *European Journal of Plant Pathology,* 134:467-473.

Wyszkowska J, Borowik A, Olszewski J & Kucharski J. (2019). Soil bacterial community and soil enzyme activity depending on the cultivation of *Triticum aestivum*, *Brassica napus*, and *Pisum sativum* ssp. *arvense*. *Diversity,* 11:246.

6 APPENDIX 1 – SUPPLEMENTARY TABLES

Table S1 Dry bulk density (BD - g/cm^3) table of means (of three samples) for the OT3 organic option and conventional counterfactual fields at 15 paired sites

Pair	OT3 option	Counterfactual
1	1.21	1.14
2	1.28	1.27
3	1.34	1.37
4	1.21	1.36
5	1.46	0.97
6	1.30	1.24
7	1.09	1.17
8	0.95	0.99
9	1.25	1.03
10	0.99	1.10
11	0.76	1.27
12	1.42	1.27
13	0.88	0.90
14	1.30	1.33
15	1.35	1.41

Table S2 'Poorest layer' VESS table of means (of three samples) for the OT3 organic option and conventional counterfactual fields at 15 paired sites

Pair	OT3 option	Counterfactual
1	2.0	1.8
2	3.5	3.3
3	2.0	3.0
4	2.0	2.7
5	2.0	1.5
6	2.0	1.6
7	2.6	2.1
8	2.3	1.8
9	1.8	0.9
10	1.0	2.0
11	1.8	2.0
12	1.9	1.7
13	1.5	2.0
14	2.8	3.0
15	3.3	3.2

VESS score: Friable 1; Intact 2; Firm 3; Compact 4; Very compact 5

Table S3 Soil dispersion ratio table of means (of three samples) for OT3 organic option and counterfactual conventional fields at 15 paired sites

Site	OT3 option	Counterfactual
1	2.7	5.0
2	4.4	4.7
3	11.6	6.1
4	4.6	7.0
5	5.9	4.5
6	3.9	4.2
7	2.8	4.6
8	3.9	4.0
9	1.8	4.4
10	4.0	6.0
11	5.1	6.4
12	7.4	4.7
13	4.2	4.9
14	4.3	5.7
15	3.9	4.9