



Soil health metrics reflect yields in long-term cropping system experiments

Catriona M. Willoughby¹ · Cairistiona F. E. Topp² · Paul D. Hallett³ · Elizabeth A. Stockdale⁴ · Robin L. Walker⁵ · Alex J. Hilton⁵ · Christine A. Watson^{5,6}

Accepted: 14 August 2023 / Published online: 4 September 2023
© The Author(s) 2023

Abstract

Soil health metrics with strong links to ecological function and agricultural productivity are needed to ensure that future management of agricultural systems meets sustainability goals. While ecological metrics and crop yields are often considered separately from one another, our work sought to assess the links between the two in an agricultural context where productivity is a key consideration. Here, we investigated the value of soil health tests in terms of their relevance to agricultural management practices and crop yields at contrasting long term cropping systems experiments. One site was on a sandy loam Leptic Podzol and the other on a sandy clay loam Endostagnic Luvisol. Furthermore, the experiments had different management systems. One contained legume-supported rotations with different grass-clover ley durations and organic amendment usage, while the other compared a range of nutrient input options through fertiliser and organic amendments on the same rotation without ley periods. Metrics included field tests (earthworm counts and visual evaluation of soil structure scores) with laboratory analysis of soil structure, chemistry and biology. This analysis included bulk density, macroporosity, pH, available phosphorus, exchangeable potassium, soil organic matter and potentially mineralizable nitrogen. Using a novel combination of long-term experiments, management systems and distinctive soil types, we demonstrated that as well as providing nutrients, agricultural management which resulted in better soil organic matter, pH, potassium and bulk density was correlated with higher crop yields. The importance of ley duration and potentially mineralizable nitrogen to yield in legume-supported systems showed the impact of agricultural management on soil biology. In systems with applications of synthetic fertiliser, earthworm counts and visual evaluation of soil structure scores were correlated with higher yields. We concluded that agricultural management altered yields not just through direct supply of nutrients to crops, but also through the changes in soil health measured by simple metrics.

Keywords Agriculture · Soil management · Soil structure · Soil health indicators · Yield

✉ Catriona M. Willoughby
c.willoughby@leeds.ac.uk

¹ School of Geography, University of Leeds, Leeds LS2 9JT, UK

² Department of Agriculture, Horticulture and Engineering Sciences, SRUC, West Mains Road, Edinburgh EH9 3JG, UK

³ School of Biological Sciences, University of Aberdeen, Aberdeen AB24 3UU, UK

⁴ NIAB, 93 Lawrence Weaver Road, Cambridge CB3 0LE, UK

⁵ Rural Land Use Department, Scotland's Rural College (SRUC), Craibstone Estate, Aberdeen AB21 9YA, UK

⁶ Department of Crop Production Ecology, Swedish University of Agricultural Sciences, 750 07 Uppsala, Sweden

1 Introduction

The declining health of agricultural soils has the potential to threaten both global food supplies and the functioning of the wider environment (Zwetsloot et al. 2021). Reductions in carbon stocks, declines in biodiversity and poor water and nutrient retention all have implications for primary productivity, climate change, off-site pollution and flooding (Gibbons et al. 2014; Muñoz-Rojas 2018). It is therefore vital that agricultural management ensures long-term maintenance of soil health to safeguard future sustainability. Soil health is defined here as the capacity of a soil to function as an ecosystem to sustain plants and animals (including humans) in the environment (Lehmann et al. 2020). Globally, there are growing policy pressures to measure soil health to support

improved land management and ensure compliance with relevant legislation regarding contaminants and pollutants, with results potentially linked to payments for ecosystem services (Department for Environment Food and Rural Affairs 2021; European Commission 2021). It is well recognised that easily usable soil health tests and interpretation frameworks that aid decision-making by land managers would benefit the health of agricultural soils (Stockdale et al. 2018).

Within agricultural systems a common goal of soil testing is to optimize yield, ensure efficient management of inputs and reduce nutrient losses. Use of an appropriate combination of soil health metrics can help support these goals, since soil properties that underpin better nutrient retention and root growth can result in more efficient nutrient use and abiotic stress resistance (Ding et al. 2020). However, determining universally relevant soil health indicators can be challenging due to soil and climate heterogeneity (Middleton et al. 2021; Rottler and Martyn 2021). Spatial and temporal heterogeneity of soil properties means that identification of suitable metrics for use in agriculture is challenging (Stockdale et al. 2019). It is also important to recognise that inherent soil characteristics, such as texture, are unresponsive to management change, but essential for interpreting results in context (Stockdale et al. 2019). Spatial variability can mean that a single field encompasses multiple contrasting soil types, which poses further challenges to interpretation.

Soil health assessments commonly include metrics that utilise a combination of physical, chemical and biological properties (Bünemann et al. 2018). Commercially available tests include broad spectrum nutrient analysis and indicators of soil carbon content, with analysis largely confined to laboratories (e.g. Crooks et al. 2019). Carbon is a critical soil health indicator and is often measured as soil organic matter (SOM) or soil organic carbon (SOC) through loss on ignition (LOI) (Bhogal et al. 2011). More detailed testing can include assays which fractionate carbon into different pools of decomposability (von Lützow et al. 2007). Detailed chemical analysis can include high-resolution imaging of structure and elemental composition (Farhate et al. 2020). To assess biological soil health, more detailed measures include DNA metabarcoding or soil community analysis; however, these have high associated costs (Panettieri et al. 2020). Potentially mineralizable nitrogen (PMN) is a more cost-effective indicator of soil biological health than sequencing approaches and is furthermore of agronomic relevance (Augarten et al. 2023). Earthworm counts have also been assessed in agricultural soils as a fast, cost-effective indicator of biological health, which have the additional benefit that they can be carried out by the practitioner. Due in part to more complex sampling requirements, physical assessments (bulk density, porosity) are not used as soil health metrics as widely as chemical analysis (Bünemann et al. 2018). As a result, inexpensive assessments of soil physical properties

specifically for agricultural land have been developed by multiple groups (Ball et al. 2007; National Soil Resources Institute 2001). An example of such a technique is the VESS (Visual Evaluation of Soil Structure) score, which provides an assessment of topsoil structure (Ball et al. 2007; Johannes et al. 2017).

Previous literature has sought to investigate links between management, soil health indicators and yields. Work carried out on long-term agronomic field trials previously found that results from three commercially available soil health tests did not differ according to contrasting management regimes on the field trials (Roper et al. 2017). The soil health tests used had included pH, broad spectrum nutrients and texture assessments alongside respiration, protein and organic carbon fractionation (Roper et al. 2017). While the authors were not able to assign any significant yield differences between plots with better soil health scores, they did show that there was a trend towards higher corn yields in plots with higher scores in the Comprehensive Assessment of Soil Health (CASH) in one of the three field trials assessed (Roper 2017). Building upon these findings, the selected indicators included in the CASH have been found to be responsive to management changes and yield in long-term cropping experiments with 3-year rotations (van Es and Karlen 2019). The comparison of crop yields between contrasting systems has been an issue highlighted by previous work (e.g. Stockdale et al. 2019; Roper et al. 2017). While researchers utilising field experiments may have the opportunity to compare between multiple replicated systems, this is not the case for farmers. Benchmarking yields with relevant local areas is therefore an alternative means of comparing performance between systems in contrasting geographical areas. Benchmarking soil health indicator results is recommended in industry (e.g. Soil Biology and Soil Health Partnership 2022), but few academic studies have utilised yield benchmarking in combination with soil health tests. Doing so gives us the opportunity to investigate the dynamics underlying agricultural management and soil health in a context applicable to commercial farms.

We aimed to link soil health metrics to agricultural productivity through quantification of a range of soil health indicators in contrasting cropping system experiments together with crop yield and management data (Fig. 1). Physical assessments were VESS score, bulk density and macroporosity. Chemical assessments were soil pH, available phosphorus (P), and exchangeable potassium (K), calcium (Ca), sodium (Na) and magnesium (Mg). Biological assessments were SOM, PMN and earthworm counts. Assessments were conducted on controlled long-term experiments with management practices including leys in the crop rotation and the application of organic fertilisers that may concurrently provide nutrients and boost soil health. Our aim was to investigate the value of soil health tests in terms

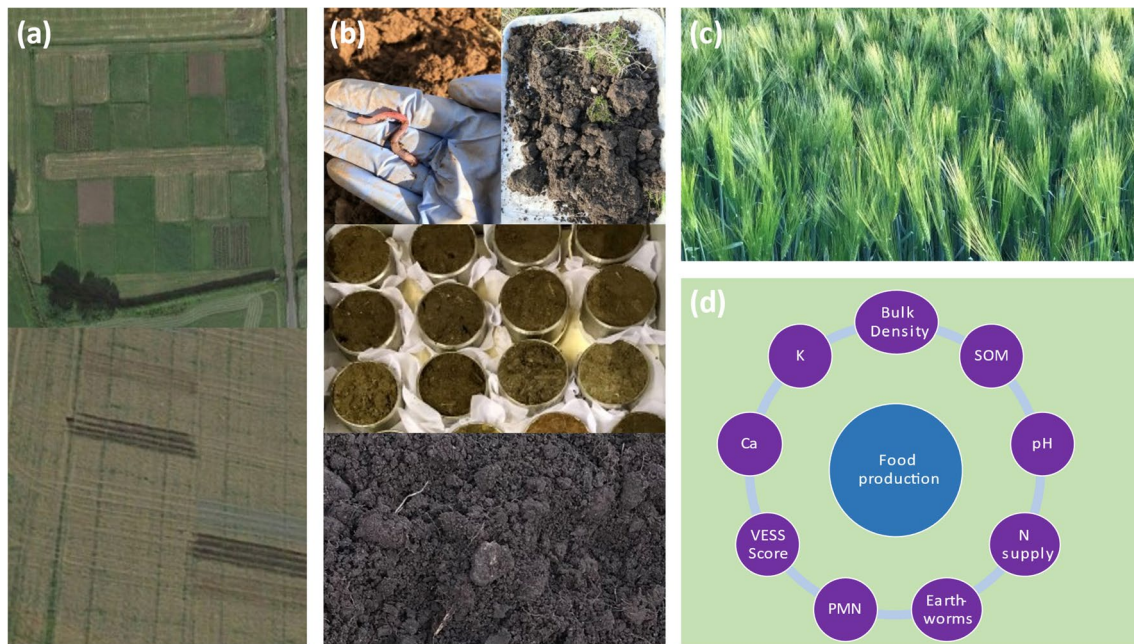


Fig. 1 A graphical summary highlighting **a** the two field sites utilised in the study, **b** the use of field tests, collection of intact cores, and bulk soil samples for analysis, **c** the use of yield data and benchmarks

to calculate relative yield percentage and **d** the soil health indicators which were predictive of relative yield percentage increases.

of their relevance to agricultural management practices and crop yields. This was to be met through achieving the following objectives: (i) establishing whether better soil health leads to higher yields; (ii) assessing how soil health indicator results change under contrasting agricultural management practices; (iii) analysing the relationship between agricultural management and yields; (iv) utilising benchmarking to compare yields between contrasting crops and sites. We hypothesised that plots with better soil health indicator results would have higher yields, and furthermore that both soil health test results and agricultural management factors would be correlated with yield. A benefit of this research is the identification of soil health indicators that are useful and relevant to environmental policy and farmers.

2 Materials and methods

The soil health metrics deployed combined in-field assessment together with the collection of topsoil (0–15 cm) bulk samples and intact cores for more detailed analysis. All fieldwork was carried out in March 2021, prior to cultivations to minimise impact of disturbance on the soil health metrics.

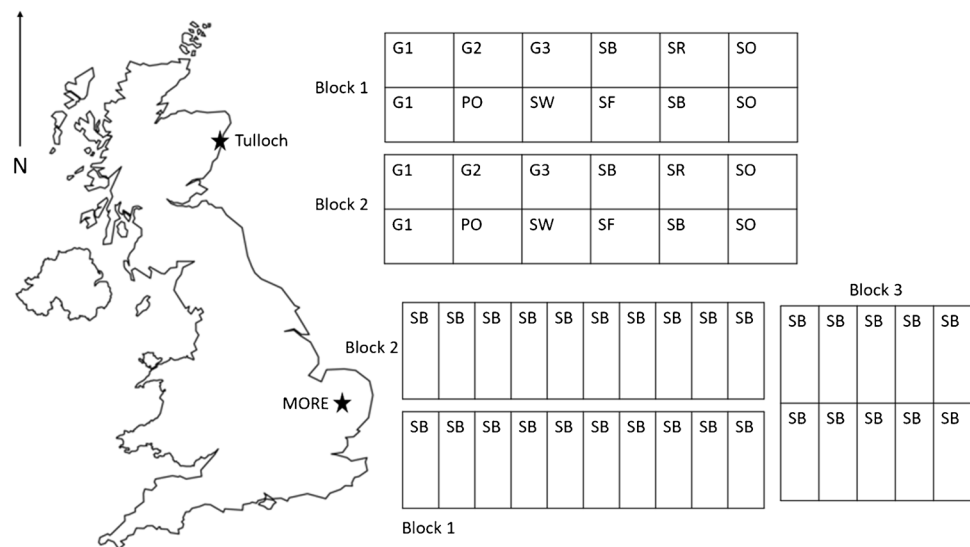
2.1 Sites

Controlled field experiments in the UK were used (Fig. 2). The first site was the Tulloch organic experiment

(subsequently referred to as Tulloch) in Aberdeenshire ($57^{\circ}10'33''\text{N}$, $002^{\circ}15'33''\text{W}$), managed by SRUC. This experiment had low inputs, organic amendments and consisted of a mixed and a stockless 6-course rotation (Fig. 2) and is described in detail by Willoughby et al. (2022). The site was arranged in two blocks, each block containing a replicate of the mixed and the stockless rotation. Within each rotation there were six plots, containing a single phase (crop), and all crops were present every year. All crops were spring sown. The grass and clover leys and undersown oats post-harvest were grazed by 4–6 sheep in the mixed rotations. Two cuts of silage were taken from the second-year ley, and one cut was taken from the third-year ley. To compensate for offtake in the mixed rotations, annual additions of organic cattle manure were applied. Cattle manure was added at 16 t ha^{-1} to the second-year ley, 10 t ha^{-1} to the third-year ley, and 12 t ha^{-1} to the swedes each year. In the stockless rotations, the grass and red clover ley was cut and mulched several times during the season prior to incorporation in the following spring. All crop residues in the stockless rotation, including straw, were chopped in the autumn, and mixed by ploughing to a depth of 20 cm in early spring.

The second site was the New Farming Systems Manure and Organic Replacement Experiment (subsequently referred to as MORE) in Norfolk ($52^{\circ}32'50''\text{N}$, $001^{\circ}02'18''\text{E}$), managed by NIAB. There were 30 plots, arranged into three blocks. All plots received typical synthetic fertiliser amendments for conventional farms, in addition to organic waste

Fig. 2 A map showing the locations, layouts and cropping sequences of the experimental sites, where G1 is the first year of grass and white clover (*Lolium perenne* Linnaeus/*Phleum pratense* L./*Trifolium repens* L.), G2 is the second year of grass and white clover, G3 is the third year of grass and white clover, SB is spring barley (*Hordeum vulgare* L.), SR is swede (*Brassica napus* L.), SO is spring oat (*Avena sativa* L.), PO is potato (*Solanum tuberosum* L.), SW is spring wheat (*Triticum aestivum* L.) and SF is spring fava beans (*Vicia faba* L.), derived from Willoughby et al. (2022).



amendments. All plots were under the same rotation, and the same crop was grown across the entire site each year (Fig. 2), but with different organic amendment treatments, namely green-waste compost (35 t ha⁻¹), turkey manure (8 t ha⁻¹), paper crumble (50 t ha⁻¹) and unamended (Table 1). Management was further sub-divided into treatments with augmented applications (every three years) and diminished applications (once) of amendments. For diminished treatments, amendments were applied in autumn in 2011. In the augmented treatments, further applications were made in 2014, 2017 and 2020. Each treatment had three replicates. All crop residues were incorporated in the autumn by ploughing the field to a depth of 20 cm. At both sites, samples were also collected from three areas selected at random in the field margins. These samples acted as controls, as they were not cropped and did not receive management inputs.

2.2 Field Sampling

Within each plot, the centre of the sampling area was recorded using a GPS mobile application. All samples were taken from within 5 m of this location. Composite samples for chemical and biological analysis were collected from each sampling area by combining and mixing five samples by hand mixed in a polythene bag. For each of the five samples, vegetation was cleared from the soil surface and

a Dutch auger with a 5-cm blade was used to obtain a bulk sample from 0 to 15 cm depth. Two intact cores were also collected from each plot to assess bulk density and macroporosity. Cores of 8 cm diameter × 5 cm height were taken from a 0–5 cm depth, using a driving dolly. All samples were stored in a cold room prior to analysis. The mean of the cores was used in the subsequent analysis.

2.3 Yield

Average yields obtained at Tulloch were 5.3 t ha⁻¹ for spring barley, 5.4 t ha⁻¹ dry matter (DM) for silage, 4.2 t ha⁻¹ for spring oats, 10.6 t ha⁻¹ (fresh weight) for potato, 2.1 t ha⁻¹ for field beans, 43.7 t ha⁻¹ for swedes (fresh weight) and 3.4 t ha⁻¹ for spring wheat. At MORE, the average spring barley yield was 5.75 t ha⁻¹. To ensure that results were comparable between sites and plots on which different crops were cultivated, relative yield percentage was used as an indicator of productivity. Relative yield percentage was calculated by dividing the yield achieved in the field with site-specific yield averages at each field trial location and multiplying the result by 100. Benchmarks used the same year as the achieved yield (2020) and were specific to each of the field trials utilised. At Tulloch, yields from another long-term field experiment incorporating the same crops in a neighbouring field were utilised to provide benchmarks. The

Table 1 The mean chemical composition of the organic amendments applied to the MORE trial, based on means from analysis in 2011, 2014 and 2017, expressed in kilogrammes per fresh tonne.

Amendment	DM	N	P ₂ O ₅	K ₂ O	MgO	SO ₃	Cu	Zn	Na ₂ O	Ca
Green waste compost	618	8.2	3.8	5.6	2.3	3.0	0.0	0.1	0.8	19.7
Turkey manure	455	25.1	16.0	13.7	3.7	6.2	29.0	127.8	1.9	40.9
Paper crumble	648	2.8	0.7	0.2	2.5	2.4	0.1	0.0	0.8	146.8

experiment used for benchmarking is described in detail in Lehtovirta et al. (2009) and Walker et al. (2015) with plots managed to a pH of 6.0 used for comparison, as these most closely matched the pH conditions at Tulloch (which has a pH of 6.0). Benchmark yields were therefore 7.8 t ha⁻¹ for spring barley, 6.6 t ha⁻¹ for spring oats, 7.6 t DM ha⁻¹ for hay, 65.2 t ha⁻¹ (fresh weight) for potato, 9.2 t ha⁻¹ for winter wheat and 7.7 t DM ha⁻¹ for swedes. All cereal yields were 85% DM. The MORE trial is situated within a working field. The average spring barley yield for this field was 5.75 t DM ha⁻¹, which was used as the benchmark for calculating relative yield percentages at MORE.

2.4 Field tests

To obtain VESS scores, vegetation within 5 cm of the surface was cleared. A spade was used to loosen and excavate soil around three sides of a rectangle to minimise disturbance of the soil structure and surface upon excavation. To a spade depth, a 20 × 20 cm block was excavated and turned out onto a light-coloured plastic board for visual evaluation, using a VESS assessment recording sheet to identify and record horizontal layers, soil structure, macropores, roots and residues (Ball et al. 2007; Johannes et al. 2017). Scores were recorded for each layer and collated into a single block score accounting for the depth of each layer, with a higher score indicative of a poorer structure and a lower score indicating a better structure. Once the visual evaluation was completed, the block of excavated soil was sorted through by hand, and any earthworms found were collected in a plastic tub. Following the method used by Stroud (2019), a water bottle was used to gently wash earthworms when required to distinguish between adults and juveniles, and the number of adult worms in the block was recorded. Once the VESS score and earthworm count were completed, the worms and soil were carefully returned to the excavated pit. Two VESS blocks were recorded in each plot/control area.

2.5 Soil physical assessments

2.5.1 Water retention parameters and macroporosity

Intact cores were trimmed and cleaned prior to analysis. Fine mesh was placed on the underside of each core and secured with elastic bands. Prepared cores were placed mesh-side down on a suction plate and saturated in deionised water for 24 hours. Saturated samples were weighed and returned to the suction plate, which was adjusted to -50 hPa for 48 hours to ascertain the field capacity. To ensure soil water had equilibrated, two samples from each suction plate were selected at random, weighed and replaced for 2 h, then reweighed to check for changes. Once the weight had equilibrated, all cores were weighed. The difference in

volumetric water content at saturation and volumetric water content at -50 hPa was taken as the macroporosity (pores >50 µm) of each core. This procedure was repeated at -200 hPa and finally at -500 hPa. The volumetric water content of the cores at each water potential was calculated by dividing the weight of water by its density (1.0) and subsequently dividing this by the volume of the core.

2.5.2 Bulk density

After the water retention properties were assessed, cores were placed on an aluminium tray in a preheated oven at 105°C for 48 hours. Cores were allowed to cool completely in a desiccator before being weighed.

2.6 Soil chemical assessments

Bagged topsoil samples were used for the soil chemical tests. They were first mixed in the lab to improve homogeneity. Samples were then air dried, milled and sieved through a 2-mm mesh.

2.6.1 Available soil phosphorus (P), exchangeable calcium (Ca), sodium (Na), potassium (K) and magnesium (Mg)

To determine soil P, 10g of soil was extracted in Modified Morgan's solution, then analysed using a LaChat Quick-Chem Automated Ion analyser (Sinclair et al. 2015). The same solution was then analysed by ICP-OES to determine soil Ca, Na, K and Mg content.

2.6.2 Soil organic matter (SOM)

SOM was determined by loss on ignition (LOI). Soil was dried at 105°C for at least 24 hours to remove any residual moisture and subsequently placed in a PYRO microwave muffle system at 550°C for 2 hours; the loss on ignition was determined gravimetrically. A conversion factor of 2.0 was used to estimate soil carbon (C) content (Pribyl 2010).

2.6.3 Soil pH

pH was measured on a 1:2 suspension of soil in 0.01 M CaCl₂. To obtain the water equivalent pH value, 0.6 was added to the calcium chloride pH value. To ensure that pH measures were comparable between the two sites, we estimated the optimum pH based on the location; this was 6.2 at Tulloch (Crooks et al. 2019) and 7 at MORE based on the optimal range for arable rotations containing sugar beet (Agriculture and Horticulture Development Board 2019). The difference between the pH measured and the optimum pH was used as a soil health indicator, which we termed the

distance to optimal pH (e.g. if pH is 6.0 and optimal pH is 6.2, the distance to optimal pH is -0.2).

2.6.4 Potentially mineralizable nitrogen (PMN)

For PMN, 10 g of fresh bulk soil from each sample were placed into two 50 cm³ tubes: one incubated and one control. 25 cm³ of distilled deionised water was added to the incubated samples, and they were placed in an oven heated to 40°C for 7 days. The control samples were stored in a fridge at < 5°C for this 7-day period. Once the seven days elapsed, a 2 M KCl extraction was carried out on all the samples and a FiaStar 5000 was used to photometrically measure the ammonium N content of the extractant from each sample. The difference between the incubated and control samples was taken as the PMN (Waring and Bremner 1964).

2.7 Nutrient supply

The contribution of management decisions to the different rotation systems in the form of nutrients was quantified in terms of the kg ha⁻¹ of elemental N, P and K supplied to each plot. For each plot, rates of seed, fertiliser and organic amendments, deposition and fixation were used to calculate the total nutrient supply. The specific calculations utilised are described in detail in Willoughby et al (2022). The sheep did not receive any supplementary feed and resupplied grazed crop biomass to the plots in the form of manure; thus, we did not consider them to contribute to changes in the nutrient supply. Calculated N, P and K supply differed between the sites as MORE was managed with additions of synthetic fertilisers alongside organic amendment additions, while Tulloch was an organic field trial. The mean N supply at Tulloch was 254 kg ha⁻¹, while at MORE it was 759 kg ha⁻¹. The mean P supply at Tulloch was 91 kg ha⁻¹ and 250 kg ha⁻¹ at MORE. At Tulloch, the mean K supply was 103 kg ha⁻¹; this was 52 kg ha⁻¹ at MORE.

2.8 Statistical analysis

Statistical analysis was carried out in R 4.1.0 (R Core Team 2021). The packages “FactoMineR” (Lê et al. 2008), “factoextra” (Kassambara and Mundt 2020), “lavaan” (Rosseel 2012), “lme4” (Bates et al. 2015) and “lmerTest” (Kuznetsova et al. 2017) were used to carry out the analysis, while “tidyverse” (Wickham et al. 2019) and “ggpubr” (Kassambara 2020) were used to format the data and create graphs. Principal component analysis (PCA) was used to reduce the number of independent variables and identify issues of multicollinearity. Because indicator performance differed at each site, data was split, and each site-specific dataset was modelled separately. At Tulloch, where the rotations had different ley periods, ley duration was included as a variable in the

principal component analysis. The procedure used to carry out PCA scaled variables prior to analysis. Components with absolute eigenvalue >1 were extracted and analysed (Kaiser 1960), and the first 5 principal components were retained at each site (Table 2). The composition and correlations within the retained principal components were scrutinised through use of variable correlation plots and cos2 values. Retained principal component scores were used in subsequent modelling (Table 2). Control plots did not have recorded yields thus were not included in the PCA or modelling.

A backwards stepwise regression was used to identify predictors of yield from the retained principal components. In 2019, data from 1st cut silage yields, taken in mid-June, were not available at Tulloch; thus, records of weather data were used to establish previous years in which growing season length (defined as the longest period within a year that meets the following requirements: begins at the start of a period of five successive days where the daily-average temperature is greater than 5.0°C, ends on the day before of a period of 5 successive days when the daily average temperature is less than 5.0°C), wet (consecutive days with > 10 mm rainfall) and dry (consecutive days with 0 mm rainfall) periods and accumulated day degrees above 0°C from the 1st January to the 31st July (tsum) were similar to 2019. The averages of the 1st cut silage yields from these similar years were then used to replace the missing silage values in 2019 (Table 3). The weather data that informed this selection procedure is outlined in detail in Table 3.

Table 2 Eigenvalues, variances and cumulative variances determined by PCA analysis on Tulloch and MORE subsets, with PCs retained for subsequent modelling highlighted.

Principal component	PC1	PC2	PC3	PC4	PC5
Tulloch					
Eigenvalue	4.0	3.3	2.2	2.0	1.3
% Variance	23.4	19.4	12.7	11.6	7.8
Cumulative % variance	23.4	42.8	55.5	67.1	74.9
Starting model	Yield ~ 1 + PC1 + PC2 + PC3 + PC4 + PC5 + (1 block)				
BIC/AIC score	163.2/158.2				
Retained model	Yield ~ PC1 + PC4 + (1 block)				
BIC/AIC score	158.6/150.6				
MORE					
Eigenvalue	4.2	3.3	2.2	1.7	1.5
% Variance	26.0	20.5	13.6	10.9	9.3
Cumulative % variance	26.0	46.5	60.1	71.0	79.3
Starting model	Yield ~ 1 + PC1 + PC2 + PC3 + PC4 + PC5 + (1 block)				
BIC/AIC score	198.8/190.2				
Retained model	Yield ~ PC1 + PC5 + (1 block)				
BIC/AIC score	195.2/187.6				

Table 3 Selected years based on growing season, drought and wet periods and calculated tsum values from 1st January to 31st July with 1st cut silage yields from second (G2) and third (G3) year grass plots, from years at Tulloch selected due to their similarity in weather conditions to 2019 (t ha^{-1}). *Accumulated mean daily temperatures ($^{\circ}\text{C}$) above zero.

Year	Crop	Mean yield	Length of growing season (days)	Max drought period (consecutive days with 0 mm rainfall)	Max wet period (consecutive days with > 10 mm rainfall)	Max tsum* (calc from 1st Jan)
2000	G2	6.1	305	18	3	2997
2000	G3	4.7	305	18	3	2997
2002	G2	5.9	310	13	3	3179
2002	G3	5.8	310	13	3	3179
2003	G2	4.8	313	19	1	3261
2003	G3	5.0	313	19	1	3261
2004	G2	4.5	314	10	2	3204
2004	G3	5.4	314	10	2	3204
2013	G2	5.1	319	19	2	3004
2013	G3	4.5	319	19	2	3004
2019	G2	5.3	312	11	4	3253
2019	G3	5.1	312	11	4	3253

At each step, variables were dropped based on BIC scores, and residual plots were checked. The model in which further exclusions did not result in an improved BIC score was the final model. The package “ggeffects” (Ludeke 2018) was used to generate predictions of yield based on retained principal component scores from the selected models.

3 Results and discussion

3.1 Soil health indicators linked to higher yields

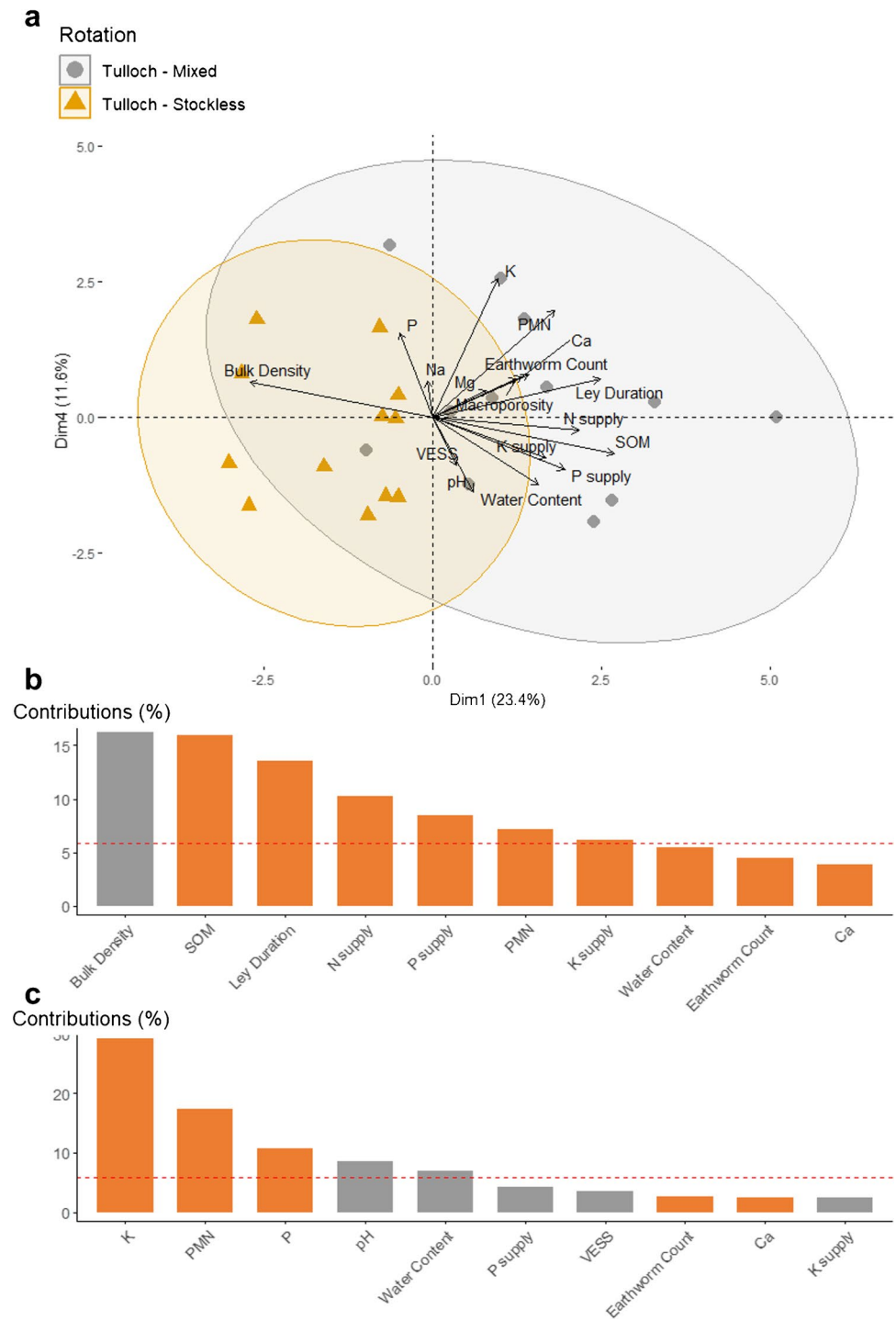
We met our first objective through using principal components to identify key soil indicators at both sites for yield prediction. The model selection procedure identified PC1 and PC4 at Tulloch and PC1 and PC5 at MORE as predictors of yield (Table 2). Scores of PC1 and PC4 at Tulloch were predicted by the retained model to lead to higher yields. Significant variable contributions (defined here as those higher than 6%, the average variable contribution of the retained PC scores used in model selection) and loadings of PC1 and PC4 at Tulloch corresponded to bulk density, SOM, ley duration, nutrient supply, K, PMN, P, distance from optimum pH and water content (Fig. 3b, c). Results from MORE showed that high PC1 scores were correlated with lower yields, but higher PC5 scores were correlated to yield increases. Because PC scores were used in the modelling, all variables are included in each of the retained components; thus, it was necessary to consider the contributions of each variable to each retained PC. Variable contributions and loadings of PC5 at MORE showed that this component was associated with earthworm counts, Na, VESS scores and Mg (Fig. 4b). PC1 at MORE was associated with soil

chemical indicators, with Mg, K, Ca, SOM, pH, N supply and Na being significant contributors (Fig. 4b).

Structural indicators (bulk density at Tulloch and VESS score at MORE) were contributors to the retained principal components and thus linked to yields. Results for specific indicators varied between the two sites (Fig. 3, Fig. 4). Results from PC1 at Tulloch showed that relative yield percentages were predicted to be higher in fields with lower bulk densities (Fig. 3b). The higher yields in plots with lower bulk densities showed that bulk density as an indicator was reflective of the soil’s suitability for development and growth of plant roots (Bünemann et al. 2018) and was thus highly agronomically relevant. In PC4 at Tulloch, the utility of VESS scoring as part of a soil health assessment was demonstrated through negative correlations with bulk density and macroporosity and a positive correlation with water content at field capacity (Fig. 3a) as found in previous work (Johannes et al. 2017). However, this was not consistent at MORE for PC5, for which correlations between VESS scores and other measures of soil physical properties varied (Fig. 4a). Using field testing adds weight to the results because few studies aimed at identifying minimum datasets for soil health testing have utilised field tests (de Paul Obade and Lal 2016; Rabot et al. 2018; Stone et al. 2016). The relevance of structural indicator results to yield and overall soil health means that they represent an opportunity to collect farm data on a larger scale and at a smaller financial cost than many of the other soil health metrics included in this analysis (Johannes et al. 2017; Stroud 2019).

Active earthworm populations (as indicated by high earthworm numbers) benefit many important soil environmental processes such as biological community activity and erosion resistance, so their positive correlation with yield

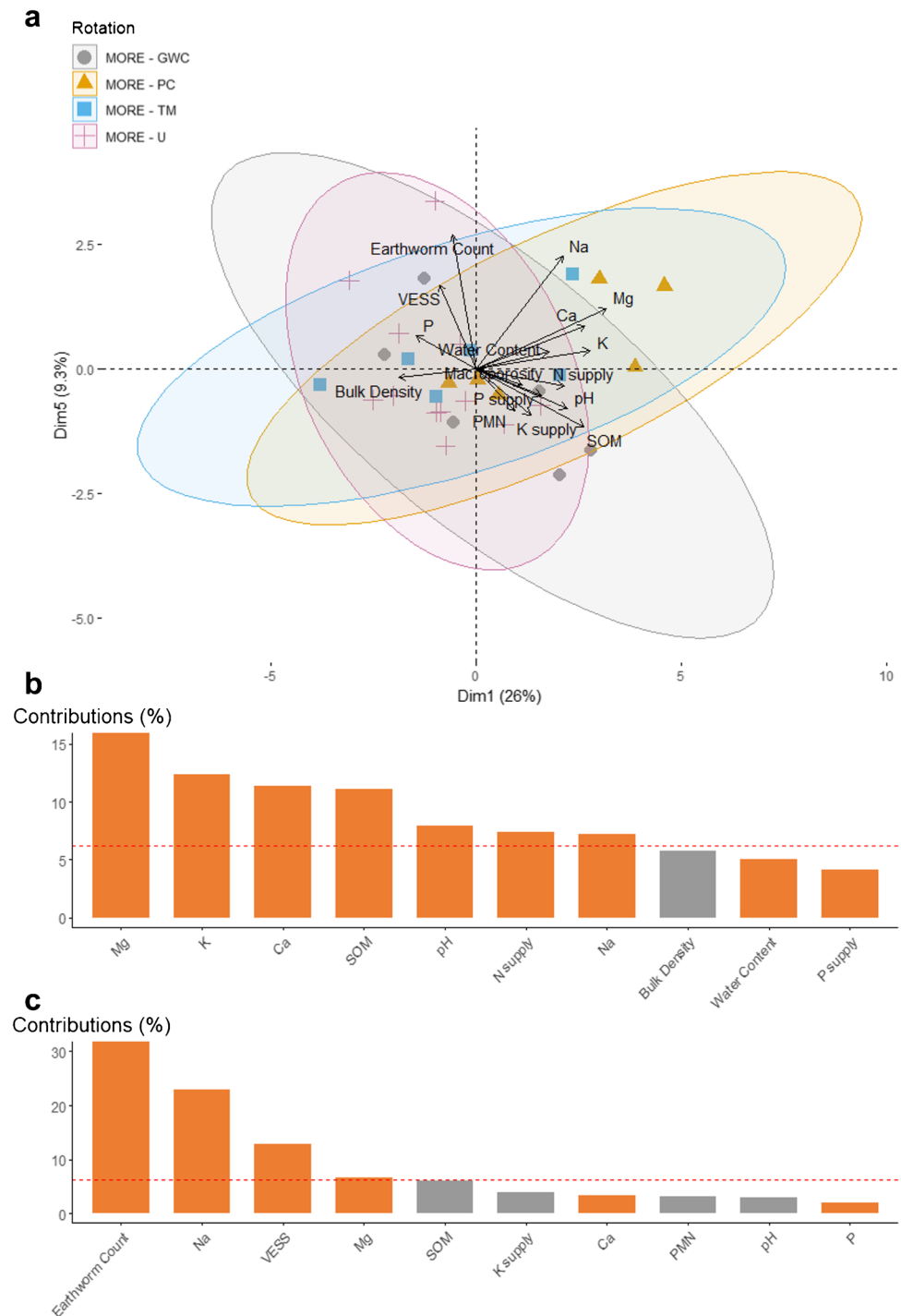
Fig. 3 Biplot of loadings and scores from retained principal components from Tulloch (**a**), contribution of variables to the retained principal components at Tulloch (**b** is principal component 1, **c** is principal component 4), coloured according to loadings (orange is a positive loading and grey is negative). The red line represents the average variable contribution to each principal component.



at MORE highlights their potential to boost crop productivity (van Groenigen et al. 2014). Additions of earthworms to contrasting agricultural systems have been shown to increase PMN and overall soil microbial biomass in perennial cropping systems (Subler et al. 1997); however, this trend was not reflected in rotational systems where changes to PMN and microbial biomass were highly spatially variable (Subler et al. 1997). Increases in earthworm populations can increase

soil mineral N concentrations (Zhang et al. 2010), which may be due to the chemical composition of their casts (Kawaguchi et al. 2011). The retained principal components at MORE, earthworm counts, were generally associated with poorer VESS scores and lower soil macroporosity (Fig. 4a). This is inconsistent with results at Tulloch, where earthworms were negatively correlated with bulk density (Fig. 3a), and with previous literature which has found that earthworm increases

Fig. 4 Biplot of loadings and scores from retained principal components from MORE (a), contribution of variables to the retained principal components at MORE (b is principal component 1, c is principal component 5), coloured according to loadings (orange is a positive loading and grey is negative). The red line represents the average variable contribution to each principal component.



are associated with improved soil physical structure (Bai et al. 2018; Berdeni et al. 2021). The results at MORE are due to the lower earthworm numbers overall at this site (Fig. 5b) meaning that any beneficial impacts on soil structure were not enacted (Zhang et al. 2010).

The metrics that did not make a significant contribution to yield prediction in our analysis may still be useful predictors of yield in further studies. In this study, responses

at the sites differed in that both SOM and PMN were influential in the retained principal components from the Tull-och dataset (Fig. 3b, c), while only SOM was influential in the retained PCs from the MORE dataset (Fig. 4b, c). This can be attributed to the fertiliser applications on the MORE plots, meaning that their nitrogen requirements were met through inputs and there was little variability in yield due to soil nitrogen supply to the crop.

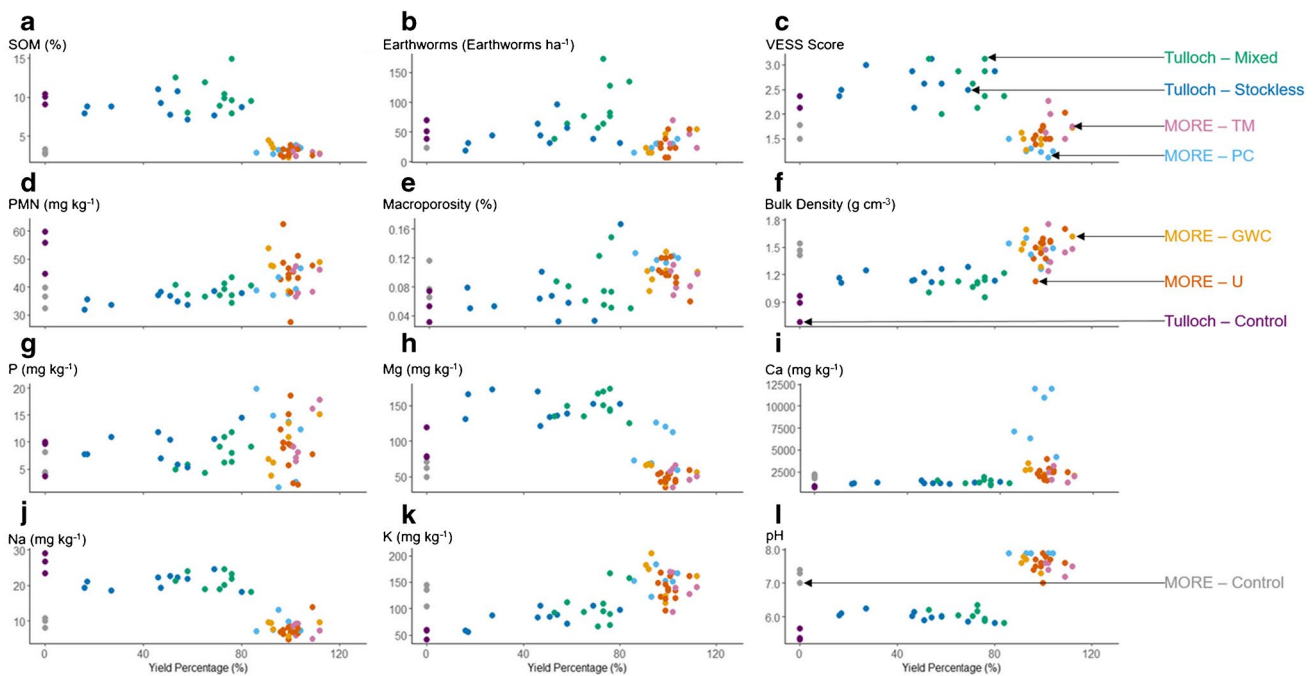


Fig. 5 Plots of soil health metrics and relative yield percentage, coloured according to rotational management with green waste compost (GWC), paper crumble (PC), turkey manure (TM), unamended (U) and control plots at MORE and mixed, stockless and control plots at Tulloch.

3.2 Soil health responses to agricultural management

The importance of agricultural management, particularly ley duration, to soil health improvement is highlighted in the results at Tulloch. Previous literature has found that ley inclusion can lead to reduced soil disturbance, enhanced root biomass and penetration and increased earthworm numbers compared with arable rotations without ley (Berdeni et al. 2021; Rollett et al. 2006; Zani et al. 2022). There were more earthworms in rotations with multi-year ley periods and cattle manure at Tulloch (Fig. 5b). Our results further support the findings of previous studies by showing that yields were higher and response of yields to soil health improvements was more pronounced in plots which had longer leys than those which had single-year leys (Martin et al. 2020). In our study, soil biological performance was found to be a higher contributor to relative yield percentage than nutrient supply (Fig. 3b, c, Fig. 4b, c). Specifically, when farming systems were intensified with no leys and higher inputs, soil chemistry and nutrient status became key to maintenance of crop yield and soil health (Fig. 4b, c). We therefore demonstrated that agricultural management improved crop yields both through cropping sequences, supply of chemical inputs and the alteration of soil health properties.

The effect of ley inclusion and duration was not limited to those indicators which were predictive of yield at Tulloch. Principal component analysis showed that bulk density was

higher, and SOM was lower in rotations with 1-year leys compared to those with 3-year leys (Fig. 3a). This was due to decreased disturbance from cultivations in the rotations with longer leys, in addition to the development and permeation of roots from the grass and clover mixture. The increases in yield (as %) and soil health with longer duration leys are in line with previous works which have shown the benefits to subsequent yields with ley durations longer than a single year (Zani et al. 2022). However, although MORE did not include any ley periods, relative yield percentage performance was higher at MORE than at Tulloch (Fig. 5). This is reflective of the differences in nutrient management between the systems, as plots at MORE received a higher input of N, P and K than those at Tulloch, and additions of synthetic fertilisers meant that nutrients were more highly available.

While supply of adequate nutrients is indisputably vital to ensuring crop yield performance and resilience in the long term (Cuvaradic et al. 2004; Loide 2019), our results showed that the influence of nutrient supply depended on the management and site context. Inclusion of these variables allowed comparison of different cropping system types. At both sites, the high influence of nutrient supply in the retained PCs showed the contribution of agricultural management to yields (Fig. 2 and Fig. 3). Increasing PC1 scores at MORE were predicted to lead to lower yields. PC1 at MORE was associated with Mg, K, Ca, SOM, pH, N supply and Na (Fig. 4b). The application of organic amendments with varying nutrient compositions is likely to be a

significant contributor to yield and soil chemistry results from MORE: for example, paper crumble amended plots had higher contents of Ca and Mg than the other amendments at MORE (Fig. 5h, i). Mg content was higher in control plots at Tulloch than at MORE, indicating higher baseline Mg at Tulloch, and at Tulloch all plots which were managed had higher Ca than controls (Fig. 5i). In PC1 from MORE, soil Na was positively correlated with distance from optimal pH, and in PC5 soil Ca is negatively correlated with distance from optimal pH (Fig. 4b, c). The differing composition of the organic amendments applied is also a potential driver for both soil chemistry and pH results particularly regarding yield, as these differed across MORE (Table 1).

PMN was an important contributor to higher yields at Tulloch (Fig. 3c). This close connection between yield and biological health has previously been noted in organic systems (Stockdale and Watson 2009). Results showed a consistently negative association between SOM and bulk density, indicating the influence of cropping sequence, residue management and external additions on SOM overall (Riley et al. 2022). This is expected given the smaller particle density of SOM than soil minerals and SOM impacts on soil aggregation (Dexter et al. 2008). At Tulloch, earthworm counts were higher in plots which received additions of animal-derived amendments, which also had longer ley duration (Fig. 3a, Fig. 5b), (Steiner et al. 2007; Bhogal et al. 2011).

3.3 Inter-site variance in soil health metrics

Our selected soil health indicators encompassed a range of analytical methods and have previously been shown to relate to chemical, physical and biological soil functions (Svoray et al. 2015; Tully and McAskill 2020). The variation in soil health indicator correlations indicated the potential for error that can occur if relationships between variables do not account for site and soil heterogeneity. Control plot results showed that the soil health metrics which varied between sites were largely dominated by chemical indicators, specifically Na, K and pH (Fig. 5j, k, l). We also found that control plots had lower bulk densities at Tulloch than at MORE (Fig. 5f). Variations in the relationships between soil health indicators and yields at Tulloch and MORE reflect the inter-site, soil type and management variations found across the literature (Payne 2006; Turmel et al. 2015). Identifying indicators that were sensitive to management change and predictors of yield at both sites was an important step towards understanding and recommending soil health indicators pertinent to both agricultural management and environmental sustainability. SOM and PMN were well represented in the principal components, and their importance to soil health overall has previously been endorsed in the literature (Brock et al. 2013; de Paul Obade and Lal

2016; Karlen et al. 2013); thus, their inclusion is also warranted to ensure a well-rounded assessment of overall soil health. Bulk density and SOM were highly influential in PC1 at both sites and PC4 at Tulloch (Fig. 3, Fig. 4). The consistent importance of chemical indicators regardless of site and soil type differences was a strong indication of their function as soil health metrics (Karlen et al. 2013); thus, their inclusion is also warranted to ensure a well-rounded assessment of overall soil health. At both sites, bulk density was strongly negatively correlated with SOM; both of which were highly influential in PC1 at both sites and PC4 at Tulloch (Fig. 3, Fig. 4). Control plots showed that the SOM was higher at Tulloch than at MORE (Fig. 5a). The consistent importance of chemical indicators regardless of site and soil type differences was a strong indication of their function as soil health metrics.

The analysis incorporated two sites, so we expected that soil type and underlying geology would lead to varying results in addition to management differences (Rawlins et al. 2012; Paterson et al. 2011). For example, the difference in PMN between the cultivated and control plots was much larger at MORE than at Tulloch (Fig. 5d), while rotational management did not lead to site-specific changes in pH (Fig. 5l). Results showed that soil P was more variable at MORE plots than Tulloch, while soil K was typically greater at MORE than at Tulloch, although some mixed plots had similar levels to those at MORE (Fig. 5g, k). Although between site heterogeneity is an important consideration, the indicators which behaved similarly across both sites were consistent across the contrasting local climates, soil types, land use histories and management systems.

3.4 Benchmarking yield and soil properties for cross-site comparisons

Comparison of yields between contrasting sites and management types can be a confounding issue in the broader applicability of much agricultural research. The need for context-specificity of yield comparisons between contrasting systems is recognised, particularly between organic and conventionally managed systems such as the field experiments considered here (Seufert et al. 2012). Calculating yield ratios between contrasting crops and systems is one possible solution (Rondanini et al. 2012); however, this does not negate the need for context specificity if comparing different sites where growing season, management expectations and crop varieties may be expected to vary (Knapp and van der Heijden 2018). We accounted for this through site-specific benchmarking and subsequent calculation of relative yield percentages so that the rotations within each trial are compared against an appropriate indication of expected performance. The higher relative yield percentages obtained from the management systems in the MORE trial compared with

the Tulloch trial are therefore a reflection of site-specific management performance.

Other studies have accounted for inter-site variation either by replicating management and analysis across contrasting geographic regions (Krupek et al. 2022) or characterising underlying features through mapping (Kooch et al. 2022). An increasingly popular approach is the compilation of indicator scores through weighted totals which account for “better” and “poorer” results depending upon site specific context (Cao et al. 2023). Our analysis has accounted for site specificity through benchmarking yields and recording pH results against the optimum pH at each field site, thus allowing for differences in inherent pH due to location, geology and soil type, in addition to benchmarking of yields. This approach may also be used for a range of soil health metrics where there is reason to believe that site differences may be influencing results. There are many advantages to utilising approaches which set results into such a context, as they aid interpretability to both researchers and the land managers for whom the results are most relevant (Soil Biology and Soil Health Partnership 2022). Benchmarking also allows for ease of comparison between widely different sites and management, potentially creating opportunities to collate and interpret soil data internationally, thus enhancing the scope of future research (Maharjan et al. 2020). We consider that in achieving our objective to utilise benchmarking for cross-site comparisons, we are building the evidence base for future research to carry out such approaches in future farming system research.

4 Conclusions

Analysing a suite of soil health indicators at contrasting sites enabled us to identify relationships between the indicators themselves, as well as their relevance to yield. Alongside management practices, we identified several soil health indicators that were predictive of yields. Across both sites, those identified were earthworm counts, PMN, SOM, K, Ca, pH, bulk density and VESS scores with agricultural management characteristics. The value of field tests as yield-predictive indicators was an encouraging finding, and results confirm the need to encompass physical and biological assessments alongside the more widely utilised chemical tests. Additionally, we showed that in sites where leys were included, increased ley duration from 1 year to 3 years was associated with both higher yields and better soil health. Increases in yield due to nutrient supply were more pronounced in systems which had lower inputs overall. Inclusion of nutrient supply as a means of comparing management types allowed the interaction between soil health and N, P and K management to be accounted for, thereby improving yield predictions. We conclude that both nutrient supply and soil health testing were pertinent to yields, although caution should be

taken when choosing relevant and responsive indicators, which will vary depending upon site and management.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13593-023-00919-3>.

Acknowledgements We would like to acknowledge the technical staff and field teams at SRUC and NIAB who have been integral to the maintenance and continued function of the long-term field experiments.

Authors' contributions Conceptualization: C.A.W., E.A.S., C.F.E.T., and C.M.W. Methodology: C.A.W., E.A.S., C.F.E.T., P.D.H., and C.M.W. Investigation: C.M.W. Formal analysis: C.M.W. and C.F.E.T. Writing—original draft: C.M.W. Writing—review and editing: C.A.W., E.A.S., C.F.E.T., P.D.H., R.L.W., and A.J.H. Funding acquisition: C.A.W. and E.A.S. Resources: C.A.W., E.A.S. and P.D.H.

Funding Catriona Willoughby was supported by a PhD studentship co-funded by NIAB and SRUC. The long-term experiments at SRUC were financially supported by the Scottish Government RESAS Strategic Research programme (JHI-D3-1 Healthy Soils for a Green Recovery). The long-term experiments at NIAB are supported by a charitable grant from The Morley Agriculture Foundation.

Data availability The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agriculture and Horticulture Development Board (2019) RB209 section 1 principles of nutrient management and fertiliser use. The AHDB Nutrient Management Guide (RB209). https://projectblue.blob.core.windows.net/media/Default/Imported%20Publication%20Thumbs/RB209/2023/NutManGuideRB209S1_230526_WEB.pdf. Accessed 20 Oct 2022
- Augarten AJ, Malone LC, Richardson GS, Jackson RD, Wattiaux MA, Conley SP, Radatz AM, Cooley ET, Ruark MD (2023) Cropping

- systems with perennial vegetation and livestock integration promote soil health. *Agric Environ Lett* 8:e20100. <https://doi.org/10.1002/ae12.20100>
- Bai Z, Caspari T, Gonzalez MR, Batjes NH, Mäder P, Bünemann EK, de Goede R, Brussaard L, Xu M, Ferreira CSS, Reintam E, Fan H, Mihelič R, Glavan M, Tóth Z (2018) Effects of agricultural management practices on soil quality: a review of long-term experiments for Europe and China. *Agric Ecosyst Environ* 265:1–7. <https://doi.org/10.1016/j.agee.2018.05.028>
- Ball BC, Batey T, Munkholm LJ (2007) Field assessment of soil structural quality - a development of the Peerlkamp test. *Soil Use Manag* 23:329–337. <https://doi.org/10.1111/j.1475-2743.2007.00102.x>
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67. <https://doi.org/10.18637/jss.v067.i01>
- Berdeni D, Turner A, Grayson RP, Llanos J, Holden J, Firbank LG, Lappage MG, Hunt SPF, Chapman PJ, Hodson ME, Helgason T, Watt PJ, Leake JR (2021) Soil quality regeneration by grass-clover leys in arable rotations compared to permanent grassland: effects on wheat yield and resilience to drought and flooding. *Soil Tillage Res* 212:105–137. <https://doi.org/10.1016/j.still.2021.105037>
- Bhogal A, Nicholson FA, Young I, Sturrock C, Whitmore AP, Chambers BJ (2011) Effects of recent and accumulated livestock manure carbon additions on soil fertility and quality. *Eur J Soil Sci* 62:174–181. <https://doi.org/10.1111/j.1365-2389.2010.01319.x>
- Brock C, Franko U, Oberholzer HR, Kuka K, Leithold G, Kolbe H, Reinhold J (2013) Humus balancing in Central Europe-concepts, state of the art, and further challenges. *J Plant Nutr Soil Sci* 176:3–11. <https://doi.org/10.1002/jpln.201200137>
- Bünemann EK, Bongiorno G, Bai Z, Creamer RE, de Deyn G, de Goede R, Fleskens L, Geissen V, Kuyper TW, Mäder P, Pulleman M, Sukkel W, van Groenigen JW, Brussaard L (2018) Soil quality – a critical review. *Soil Biol Biochem* 120:105–125. <https://doi.org/10.1016/j.soilbio.2018.01.030>
- Cao Y, Li X, Qian X, Gu H, Li J, Chen X, Shen G (2023) Soil health assessment in the Yangtze River Delta of China: method development and application in orchards. *Agric Ecosyst Environ* 341:108190. <https://doi.org/10.1016/j.agee.2022.108190>
- Crooks B, Sinclair A, Edwards T (2019) Technical note TN726 fertiliser recommendations for grassland. <https://www.fas.scot/downloads/tm726-fertiliser-recommendations-for-grassland-scotland/>. Accessed 16 Feb 2023
- Cuvaric M, Tveitnes S, Krogstad T, Lombnæs P (2004) Long-term effects of crop rotation and different fertilization systems on soil fertility and productivity. *Acta Agric Scand B Soil Plant Sci* 54:193–201. <https://doi.org/10.1080/09064700410030258>
- de Paul OV, Lal R (2016) Towards a standard technique for soil quality assessment. *Geoderma* 265:96–102. <https://doi.org/10.1016/j.geoderma.2015.11.023>
- Department for Environment Food and Rural Affairs (2021) Sustainable farming incentive: how the scheme will work in 2022. <https://www.gov.uk/government/publications/sustainable-farming-incentive-how-the-scheme-will-work-in-2022/sustainable-farming-incentive-how-the-scheme-will-work-in-2022>. Accessed 26 Jan 2023
- Dexter AR, Richard G, Arrouays D, Czyz EA, Jolivet C, Duval O (2008) Complexed organic matter controls soil physical properties. *Geoderma* 144:620–627. <https://doi.org/10.1016/j.geoderma.2008.01.022>
- Ding Z, Kheir AMS, Ali MGM, Ali OAM, Abdelaal AIN, Lin X, Zhou Z, Wang B, Liu B, He Z (2020) The integrated effect of salinity, organic amendments, phosphorus fertilizers, and deficit irrigation on soil properties, phosphorus fractionation and wheat productivity. *Sci Rep* 10:1–13. <https://doi.org/10.1038/s41598-020-59650-8>
- European Commission (2021) Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. URL <https://op.europa.eu/en/publication-detail/-/publication/6a2cfdb8-2101-11ec-bd8e-01aa75ed71a1/language-en>. Accessed 2 Dec 2022
- Farhate CVV, de Souza ZM, Cherubin MR, Lovera LH, de Oliveira IN, Carneiro MP, la Scala N (2020) Abiotic soil health indicators that respond to sustainable management practices in sugarcane cultivation. *Sustainability (Switzerland)* 12:1–19. <https://doi.org/10.3390/su12229407>
- Gibbons JM, Williamson JC, Williams AP, Withers PJA, Hockley N, Harris IM, Hughes JW, Taylor RL, Jones DL, Healey JR (2014) Sustainable nutrient management at field, farm and regional level: soil testing, nutrient budgets and the trade-off between lime application and greenhouse gas emissions. *Agric Ecosyst Environ* 188:48–56. <https://doi.org/10.1016/j.agee.2014.02.016>
- Johannes A, Weisskopf P, Schulin R, Boivin P (2017) To what extent do physical measurements match with visual evaluation of soil structure? *Soil Tillage Res* 173:24–32. <https://doi.org/10.1016/j.still.2016.06.001>
- Kaiser HF (1960) The application of electronic computers to factor analysis. *Educ Psychol Meas* 20:141–151. <https://doi.org/10.1177/001316446002000116>
- Karlen DL, Cambardella CA, Kovar JL, Colvin TS (2013) Soil quality response to long-term tillage and crop rotation practices. *Soil Tillage Res* 133:54–64. <https://doi.org/10.1016/j.still.2013.05.013>
- Kassambara A, Mundt F (2020) factoextra: extract and visualize the results of multivariate data analyses. R package version 1.0.7. <https://rpkgs.datanovia.com/factoextra/index.html>
- Kassambara A, 2020. ggpubr: “ggplot2” based publication ready plots. R package version 0.4.0. <https://rpkgs.datanovia.com/ggpubr/>
- Kawaguchi T, Kyoshima T, Kaneko N (2011) Mineral nitrogen dynamics in the casts of epigeic earthworms (Metaphire hilgendorfi: Megascolecidae). *Soil Sci Plant Nutr* 57:387–395. <https://doi.org/10.1080/00380768.2011.579879>
- Knapp S, van der Heijden MGA (2018) A global meta-analysis of yield stability in organic and conservation agriculture. *Nat Commun* 9:3632. <https://doi.org/10.1038/s41467-018-05956-1>
- Kooch Y, Ghorbanzadeh N, Haghverdi K, Francaviglia R (2022) Soil quality cannot be improved after thirty years of land use change from forest to rangeland. *Sci Total Environ* 159–132. <https://doi.org/10.1016/j.scitotenv.2022.159132>
- Krupek FS, Redfearn D, Eskridge KM, Basche A (2022) Ecological intensification with soil health practices demonstrates positive impacts on multiple soil properties: a large-scale farmer-led experiment. *Geoderma* 409:115594. <https://doi.org/10.1016/j.geoderma.2021.115594>
- Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest package: tests in linear mixed effects models. *J Stat Softw* 82(13):1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lê S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. *J Stat Softw* 25(1):1–18. <https://doi.org/10.18637/jss.v025.i01>
- Lehmann J, Bossio DA, Kogel-Knabner I, Rillig MC (2020) The concept and future prospects of soil health. *Nat Rev Earth Environ* 1:544–553. <https://doi.org/10.1038/s43017-020-0080-8>
- Lehtovirta LE, Prosser JI, Nicol GW (2009) Soil pH regulates the abundance and diversity of Group 1.1c Crenarchaeota. *FEMS Microbiol Ecol* 70:367–376. <https://doi.org/10.1111/j.1574-6941.2009.00748.x>
- Loide V (2019) The results of an NPK-fertilisation trial of long-term crop rotation on carbonate-rich soil in Estonia. *Acta Agric Scand B Soil Plant Sci* 69:596–605. <https://doi.org/10.1080/09064710.2019.1628292>
- Ludeke D (2018) ggeffects: tidy data frames of marginal effects from regression models. *J Open Source Softw* 3:772. <https://doi.org/10.21105/joss.00772>

- Maharjan B, Das S, Acharya BS (2020) Soil health gap: a concept to establish a benchmark for soil health management. *Glob Ecol Conserv* 23:e01126. <https://doi.org/10.1016/j.gecco.2020.e01126>
- Martin G, Durand JL, Duru M (2020) Role of ley pastures in tomorrow's cropping systems. A review. *Agron Sustain Dev* 40:17. <https://doi.org/10.1007/s13593-020-00620-9>
- Middleton TE, McCombs AL, Gailans SR, Carlson S, Karlen DL, Moore KJ, Liebman MZ, Kaspar TC, Al-Kaisi MM, Laird DA, Wiedenhoft MH, Delate K, Cambardella CA, Thompson ML, Heaton EA, McDaniel MD (2021) Assessing biological soil health through decomposition of inexpensive household items. *Appl Soil Ecol* 168. <https://doi.org/10.1016/j.apsoil.2021.104099>
- Muñoz-Rojas M (2018) Soil quality indicators: critical tools in ecosystem restoration. *Curr Opin Environ Sci Health*. 5:47–52. <https://doi.org/10.1016/j.coesh.2018.04.007>
- National Soil Resources Institute (2001) A guide to better soil structure. Silsoe, UK. URL https://www.landis.org.uk/downloads/downloads/structure_brochure.pdf. Accessed 2 Dec 2022
- Panettieri M, Guigue J, Chemidlin Prevost-Bouré N, Thévenot M, Lévêque J, Le Guillou C, Maron PA, Antoni AL, Ranjard L, Mounier S, Menasseri S, Viaud V, Mathieu O (2020) Grassland-crop-land rotation cycles in crop-livestock farming systems regulate priming effect potential in soils through modulation of microbial communities, composition of soil organic matter and abiotic soil properties. *Agric Ecosyst Environ* 299:106973. <https://doi.org/10.1016/j.agee.2020.106973>
- Paterson E, Campbell CD, Coull MC, Shand CA (2011) Geochemical atlas for Scottish topsoils. The Macaulay Land Use Research Institute. ISBN 978-0-902701-00-7
- Payne RW (2006) New and traditional methods for the analysis of unreplicated experiments. *Crop Sci* 46:2476–2481. <https://doi.org/10.2135/cropsci2006.04.0273>
- Pribyl DW (2010) A critical review of the conventional SOC to SOM conversion factor. *Geoderma* 156:75–83. <https://doi.org/10.1016/j.geoderma.2010.02.003>
- R Core Team (2021) R: a language and environment for statistical computing. Vienna, Austria. URL <https://www.R-project.org/>
- Rabot E, Wiesmeier M, Schlüter S, Vogel HJ (2018) Soil structure as an indicator of soil functions: a review. *Geoderma* 314:122–137. <https://doi.org/10.1016/j.geoderma.2017.11.009>
- Rawlins BG, McGrath SP, Scheib AJ, Breward N, Cave M, Lister TR, Ingham M, Gowing C and Carter S (2012) The advanced soil geochemical atlas of England and Wales. 70–74. ISBN 978-0-85272-709-6
- Riley H, Henriksen TM, Torp T, Korsaaeth A (2022) Soil carbon under arable and mixed dairy cropping in a long-term trial in SE Norway. *Acta Agric Scand B Soil Plant Sci* 72:648–659. <https://doi.org/10.1080/09064710.2022.2047770>
- Rollett A, Sparkes DL, Wilson P (2006) The legacy of stockless organic conversion strategies. *Ann Appl Biol* 150:107–113. <https://doi.org/10.1111/j.1744-7348.2006.00109.x>
- Rondanini DP, Gomez NV, Agosti MB, Miralles DJ (2012) Global trends of rapeseed grain yield stability and rapeseed-to-wheat yield ratio in the last four decades. *Eur J Agron* 37:56–65. <https://doi.org/10.1016/j.eja.2011.10.005>
- Roper WR, Osmond DL, Heitman JL, Waggoner MG, Reberg-Horton SC (2017) Soil health indicators do not differentiate among agronomic management systems in North Carolina soils. *Soil Sci Soc Am J* 81:828–843. <https://doi.org/10.2136/sssaj2016.12.0400>
- Rosseel Y (2012) lavaan: an R package for structural equation modeling. *J Stat Softw* 48(2):1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rottler CM, Martyn TE (2021) Determining an ideal sampling density for microbial community analyses of farm fields. *Agrosyst Geosci Environ* 4. <https://doi.org/10.1002/agg2.20156>
- Seufert V, Ramankutty N, Foley JA (2012) Comparing the yields of organic and conventional agriculture. *Nature* 485:229–232. <https://doi.org/10.1038/nature11069>
- Sinclair A, Crooks B, Edwards T, Coull M (2015) Technical note TN668. URL <https://www.sruc.ac.uk/media/ootl1mnc/tn668.pdf>. Accessed 2 Dec 2022
- Soil Biology and Soil Health Partnership (2022) Soil Health scorecard approach Sampling protocol and benchmarking tables England and Wales Version 1.0. URL [https://projectblue.blob.core.windows.net/media/Default/Research%20Papers/AHDB/2022/AHDB%20Soil%20Health%20Scorecard%20Protocol%20and%20Benchmarking%20-%20England%20and%20Wales%20\(v1.0\).pdf](https://projectblue.blob.core.windows.net/media/Default/Research%20Papers/AHDB/2022/AHDB%20Soil%20Health%20Scorecard%20Protocol%20and%20Benchmarking%20-%20England%20and%20Wales%20(v1.0).pdf). Accessed 13 Feb 2023
- Steiner C, Teixeira WG, Lehmann J, Nehls T, de Macêdo JLV, Blum WEH, Zech W (2007) Long term effects of manure, charcoal and mineral fertilization on crop production and fertility on a highly weathered Central Amazonian upland soil. *Plant Soil* 291:275–290. <https://doi.org/10.1007/s11104-007-9193-9>
- Stockdale E, Griffiths B, Hargreaves P, Elphinstone J, Bhogal A (2018) Developing a practical and relevant soil health toolkit for UK agriculture. *International Fertiliser Society Proceedings* 1–28. ISBN 978-0-85310-461-2
- Stockdale EA, Griffiths BS, Hargreaves PR, Bhogal A, Crotty FV, Watson CA (2019) Conceptual framework underpinning management of soil health—supporting site-specific delivery of sustainable agroecosystems. *Food Energy Secur* 8. <https://doi.org/10.1002/fes3.158>
- Stockdale EA, Watson CA (2009) Biological indicators of soil quality in organic farming systems. *Renewable Agric Food Syst* 24(4):308–318. <https://doi.org/10.1017/S1742170509990172>
- Stone D, Ritz K, Griffiths BG, Orgiazzi A, Creamer RE (2016) Selection of biological indicators appropriate for European soil monitoring. *Appl Soil Ecol* 97:12–22. <https://doi.org/10.1016/j.apsoil.2015.08.005>
- Stroud JL (2019) Soil health pilot study in England: outcomes from an on-farm earthworm survey. *PLoS ONE* 14:1–16. <https://doi.org/10.1371/journal.pone.0203909>
- Subler S, Baranski CM, Edwards CA (1997) Earthworm additions increased short-term nitrogen availability and leaching in two grain-crop agroecosystems. *Soil Biol Biochem* 29:413–421. [https://doi.org/10.1016/S0038-0717\(96\)00173-3](https://doi.org/10.1016/S0038-0717(96)00173-3)
- Svoray T, Hassid I, Atkinson PM, Moebius-Clune BN, van Es HM (2015) Mapping soil health over large agriculturally important areas. *Soil Sci Soc Am J* 79:1420–1434. <https://doi.org/10.2136/sssaj2014.09.0371>
- Tully KL, McAskil C (2020) Promoting soil health in organically managed systems: a review. *Org Agric* 10:339–358. <https://doi.org/10.1007/s13165-019-00275-1>
- Turmel MS, Speratti A, Baudron F, Verhulst N, Govaerts B (2015) Crop residue management and soil health: a systems analysis. *Agric Syst* 134:6–16. <https://doi.org/10.1016/j.agry.2014.05.009>
- van Es HM, Karlen DL (2019) Reanalysis validates soil health indicator sensitivity and correlation with long-term crop yields. *Soil Sci Soc Am J* 83:721–732. <https://doi.org/10.2136/sssaj2018.09.0338>
- van Groenigen JW, Lubbers IM, Vos HMJ, Brown GG, de Deyn GB, van Groenigen KJ (2014) Earthworms increase plant production: a meta-analysis. *Sci Rep* 4:6364. <https://doi.org/10.1038/srep06365>
- von Lütow M, Kögel-Knabner I, Ekschmitt K, Flessa H, Guggenberger G, Matzner E, Marschner B (2007) SOM fractionation methods: relevance to functional pools and to stabilization mechanisms. *Soil Biol Biochem* 39:2183–2207. <https://doi.org/10.1016/j.soilbio.2007.03.007>
- Walker R, Watson C, Edwards A (2015) Aspects of pH gradient effect on crop productivity and crop quality parameters in an eight course ley-arable rotation started in 1961. *Asp Appl Biol* 128:63–68

- Waring SA, Bremner JM (1964) Ammonium production in soil under water-logged conditions as an index of nitrogen availability. *Nature* 201:951–952. <https://doi.org/10.1038/201951a0>
- Wickham H, Averick M, Bryan J, Chang W, D'Agostino McGowan L, Francois R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Lin Pederson T, Miller E, Milton Backe S, Muller K, Ooms J, Robinson D, Paige Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to the tidyverse. *J Open Source Softw* 4(43):1686. <https://doi.org/10.21105/joss.01686>
- Willoughby C, Topp CFE, Hallett PD, Stockdale EA, Stoddard FL, Walker RL, Hilton AJ, Watson CA (2022) New approach combining food value with nutrient budgeting provides insights into the value of alternative farming systems. *Food Energy Secur*. <https://doi.org/10.1002/fes3.427>
- Zani CF, Lopez-Capel E, Abbott GD, Taylor JA, Cooper JM (2022) Effects of integrating grass-clover leys with livestock into arable crop rotations on soil carbon stocks and particulate and mineral-associated soil organic matter fractions in conventional and organic systems. *Soil Use Manag* 38:448–465. <https://doi.org/10.1111/sum.12754>
- Zhang S, Chao Y, Zhang C, Cheng J, Li J, Ma N (2010) Earthworms enhanced winter oilseed rape (*Brassica napus* L.) growth and nitrogen uptake. *Agric Ecosyst Environ* 139:463–468. <https://doi.org/10.1016/j.agee.2010.08.007>
- Zwetsloot MJ, van Leeuwen J, Hemerik L, Martens H, Simó Josa I, van de Broek M, Debeljak M, Rutgers M, Sandén T, Wall DP, Jones A, Creamer RE (2021) Soil multifunctionality: synergies and trade-offs across European climatic zones and land uses. *Eur J Soil Sci* 72:1640–1654. <https://doi.org/10.1111/ejss.13051>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.