# Probing soil physical and biological resilience data from a broad sampling of arable farms in Scotland

 doi: 10.1111/sum.12214

# Probing soil physical and biological resilience data from a broad sampling of arable farms in Scotland

B. S. Griffiths[1], P. D. Hallett[2], T. J. Daniell[3], C. Hawes[3], G. S. Squire[3], S. M. Mitchell[3], S. Caul[3], T. A. Valentine[3], K. Binnie[3], A. J. Adeloye[4], R. Rustum[4] & I. Nevison[5]
[1]*Crop and Soil Systems Research Group, SRUC, West Mains Road, Edinburgh EH9 3JG, UK,* [2]*Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen AB24 3UU, UK,* [3]*Ecological Sciences, The James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK,* [4]*Institute for Infrastructure and Environment, Heriot-Watt University, Edinburgh EH14 4AS, UK, and* [5]*BioSS, King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, UK*

**Abstract**

Physical and biological soil stabilities (i.e. resistance and resilience) were measured on a range of arable farms across eastern Scotland under a range of management practices, with the objective of using a geographically restricted set of soils under similar land use to detect any underlying associations between soil stability, management factors and soil properties. Data were analysed using a combination of a stepwise fixed effects model selection within a linear mixed-model framework (LMM) and neural network analysis using a Kohonen self-organising map (KSOM). In general, physical and biological measures of stability were associated with both physical and biological soil properties, particularly bulk density, water retention characteristics, soil carbon and bacterial community structure. A strength of KSOM is its ability to fit more flexible models than the linear relationships of LMM. However, a weakness is that it does not have the ability of LMM to model the sampling design, which is likely to lead to overstating statistical significance. Consequently, KSOM identified more significant associations between soil properties and stability than LMM, while the latter identified significant associations at the between-farm level. The high-level land management decisions of farm type (conventional, organic, integrated), crop type or underlying soil type were not associated with stability at this regional scale, thus indicating that the effects of different management practices between farms were overridden by the soil properties on each farm. Management decisions on improving soil stability therefore need to be taken at the individual field scale.

**Keywords:** Bacterial community structure, land management, neural network analysis, field scale soil properties, resilience, resistance

## Introduction

The capacity of a soil to withstand and recover from external stresses underpins its capacity to perform a broad range of ecosystem services (Seybold *et al.*, 1999). Agricultural production causes additional stresses such as soil compaction, mechanical disturbance, pollutants, fertilisers and changes in plant communities. As soil is effectively a nonrenewable resource, increasingly under environmental pressure (Creamer *et al.*, 2010), considerable effort has been invested in understanding the stability of soil functions to disturbance (i.e. resistance, the immediate response to disturbance, and resilience, the recovery of function over time) (Griffiths & Philippot, 2013).

The main factors influencing biological stability are organic matter (OM) content and composition, aggregation and, to a lesser extent, clay content, pH and also the species-level characteristics of the soil microbial community (Griffiths & Philippot, 2013). Previous research has used a range of stresses imposed on controlled laboratory conditions to explore stability, including transient stresses such as heat, freezing or moisture fluxes (e.g. Tobor-Kaplon *et al.*, 2005; Kuan *et al.*, 2007) or persistent stresses such as heavy metal or pesticide additions, osmotic stress because of salt addition or rapid changes to pH (e.g. Degens *et al.*, 2001; Tobor-Kaplon *et al.*, 2006; Mertens *et al.*, 2007; Deng *et al.*, 2009). Factors affecting physical stability include

Correspondence: B. S. Griffiths. E-mail: bryan.griffiths@sruc.ac.uk
Received November 2014; accepted after revision July 2015

 

structural collapse because of weathering or compaction; therefore, physical stability assays examine the breakdown and recovery of pore structure because of weathering through cycles of wetting and drying or compression under stresses similar to agricultural implements (Gregory *et al.*, 2009). A survey of 26 different Scottish soils (Kuan *et al.*, 2007) ranging from sandy regosols to acidic peat soils with more than 20% carbon showed soils with a high OM content to be most resistant and resilient. Another long-term examination of field sites, with a narrower range of soil types but from diverse land uses, revealed a strong positive association between OM and stability, especially in grassland soils (Gregory *et al.*, 2009). Land management can alter soil stability through its effects on soil properties (Hueso *et al.*, 2011; Stockdale *et al.*, 2013; Zhang *et al.*, 2013),

A powerful approach for integrating disparate types of qualitative and quantitative data is the use of artificial neural networks (ANNs) (Schultz & Wieland, 1997; De la Rosa *et al.*, 2004). The Kohonen self-organising map (KSOM) is a subtype of ANN that is particularly useful for visualisation of multidimensional data. The input data for KSOM can have already been reduced in dimensionality by principal component analysis (PCA). Such an approach is ideally suited to soil survey data, which generally contains diverse physical, chemical and biological information, with some parameters best represented as the summarising metrics from multivariate analysis. KSOM was used to identify relationships between soil biological and chemical variables under managed or 'native' land use in Victoria, Australia (Mele & Crowley, 2008), a study not dissimilar to the regional assessment of factors related to soil stability described in this paper.

A previous data-mining approach applied to a range of soils in Scotland indicated that there were differences in the biological and physical stability of soils across regions of Scotland (Debeljak *et al.*, 2009). That study covered a wide range of land uses (including moorland, grassland, forest and arable) across different climatic zones. However, differences in land use and the inclusion of many soils very high in OM specifically restricted the data available to examine arable systems. Given the importance of arable soils for food production and the greater human-induced stresses they experience, a more focussed survey of arable soils across eastern Scotland was undertaken. This had the objective of using a geographically restricted set of soils under similar land use to detect any underlying relationships between soil stability, management factors and soil parameters. Data were analysed using a stepwise mixed-model approach and a complementary neural network analysis to explore the factors that underpin soil stability and resilience to disturbance. The study is highly relevant to the selection of indicators to characterise the capacity of soils to withstand and recover from disturbance as it provides an insight into major soil properties driving differences in stability.

## Materials and methods

### *Field sites*

Field sites were located on arable farms typical of enterprises across the east of Scotland as described by Hawes *et al.* (2010) and Valentine *et al.* (2012). These sites were selected to cover the three main geographical regions of arable production throughout the east coast of Scotland, which are as follows: northern (Inverness and Aberdeenshire), central (Tayside and Fife) and southern (Edinburgh and Borders). Within each of these regions, a range of farm types was selected to cover the spread of crop management intensity typical in each region. These included organic (certified 'organic' by the UK Soil Association (http://www.soilassociation.org/Whatisorganic/Organicstandards) and had been under organic cultivation for more than 5 yrs); conventional (whose primary management goal was to maximise yield of the main commodity crops and which did not follow formal integrated management or organic farming approaches); and integrated (an holistic pattern of land use which integrates natural regulation processes to achieve maximum replacement of inputs, Morris & Winter, 1999). Farms for this study were chosen from the pool of available sites to cover the same geographic regions (36% of the northern, 63% of the central and 53% of the southern sites) and farm types (61% of the organic, 38% of the integrated and 40% of the commodity farms). Analysis of soil texture from each field showed no significant difference in the average sand, silt and clay contents between the subset of sites selected here and the full complement of sites, confirming that they are representative of the wider pool.

At each farm, two paired fields were selected as representative of different stages in the crop rotation typical for that farm. Our sampling was targeted to collect a broad range of agricultural soil types, based on the National Soil Inventory of Scotland (1978–1988) (http://www.soils-scotland.gov.uk). Constraints of the statistical methods used meant that we could only use fields for which there was a complete data set, this limited the analysis to 16 of the farms within the survey where sufficient data were available (Table 1).

### *Soil sampling*

Soil samples were collected during September 2007 after the crop had been harvested but before the next crop was sown. Three samples were collected individually from random points within each field, and each sample comprised 1.5 L of well-mixed soil from within a 0.5 $m^2$ quadrat to a depth of 15 cm and sieved through a 10-mm mesh. An intact soil core 56 mm diameter × 40 mm height was also collected from the top 2–10 cm adjacent to each quadrat (Valentine *et al.*, 2012). Soil samples were stored at 5 °C until ready for processing.

**Table 1** Mean soil and stability measurements per field used for analysis. Only fields for which there was a complete data set are included

| Farm properties | | | | | Measured soil properties | | | | | | | | | Bacterial community structure | | | | Physical and biological stability | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Farm | Field | Crop type | Farm type | MSSG | Water content at −5 kPa ($m^3/m^3$) | Bulk density (g/cm3) | Water retention, parameter a | Water retention, parameter b | Macroporosity (m3/m3) | Root elongation (mm) | Penetration resistance (MPa) | Carbon (%) | C:N | TRFLP PC1 blue dye | TRFLP PC2 blue dye | TRFLP PC1 green dye | TRFLP PC2 green dye | Shear modulus, G (Pa) | Thixotropic resistance | Thixotropic resilience | Resistance to CU (%Rcu) | Resistance to heat (%Rht) |
| 2 | a | 1 | Con | HIP | 0.25 | 1.50 | −2.50 | −6.79 | 0.16 | 16.33 | 2.23 | 1.87 | 12.57 | 0.18 | −0.01 | −0.11 | −0.01 | 71 959 | 0.29 | 0.95 | 52.57 | 103.18 |
| 3 | a | 1 | Org | HIP | 0.24 | 1.44 | −1.59 | −5.22 | 0.20 | 12.00 | 2.11 | 2.85 | 12.55 | 0.26 | −0.07 | −0.12 | 0.01 | 56 674 | 0.27 | 0.69 | 44.75 | 80.36 |
| 4 | a | 2 | Int | NCG | 0.32 | 1.05 | −2.86 | −8.27 | 0.20 | 21.50 | 1.13 | 3.21 | 13.40 | 0.11 | 0.02 | −0.06 | 0.02 | 10 2369 | 0.37 | 0.99 | 68.29 | 61.82 |
| 4 | b | 1 | Int | NCG | 0.35 | 1.17 | −3.11 | −10.56 | 0.13 | 10.33 | 1.58 | 3.96 | 11.56 | 0.18 | 0.07 | −0.04 | −0.05 | 90 069 | 0.41 | 0.95 | 76.24 | 75.52 |
| 5 | a | 4 | Int | BEG | 0.41 | 1.13 | −3.30 | −12.89 | 0.12 | 17.33 | 1.70 | 3.40 | 11.35 | 0.19 | 0.03 | −0.03 | −0.14 | 92 224 | 0.46 | 1.02 | 78.09 | 69.00 |
| 5 | b | 1 | Int | BEG | 0.46 | 1.14 | −3.66 | −15.88 | 0.09 | 21.00 | 1.73 | 3.87 | 12.74 | 0.21 | −0.01 | −0.08 | −0.17 | 52 870 | 0.13 | 0.52 | 103.97 | 88.82 |
| 6 | a | 1 | Org | NCG | 0.41 | 1.40 | −5.49 | −18.35 | 0.07 | 14.50 | 2.02 | 2.43 | 10.78 | 0.18 | −0.07 | −0.04 | −0.14 | 37 412 | 0.08 | 0.54 | 54.88 | 59.46 |
| 6 | b | 2 | Org | BEG | 0.32 | 1.34 | −3.10 | −9.48 | 0.16 | 14.33 | 1.76 | 2.33 | 11.94 | 0.04 | −0.09 | −0.05 | −0.10 | 17 703 | 0.06 | 0.33 | 65.90 | 63.56 |
| 8 | a | 5 | Int | BEG | 0.36 | 1.15 | −2.25 | −8.60 | 0.18 | 15.67 | 1.75 | 4.63 | 12.24 | −0.12 | 0.10 | 0.13 | −0.09 | 96 461 | 0.34 | 0.97 | 72.96 | 82.98 |
| 8 | b | 6 | Int | BEG | 0.33 | 1.05 | −4.00 | −11.54 | 0.16 | 16.00 | 1.32 | 3.03 | 12.82 | −0.07 | 0.15 | 0.14 | −0.11 | 75 825 | 0.21 | 0.86 | 75.10 | 83.21 |
| 9 | a | 1 | Int | NCG | 0.33 | 1.09 | −2.69 | −8.71 | 0.18 | 16.67 | 1.08 | 3.28 | 11.59 | 0.08 | −0.11 | −0.06 | −0.14 | 70 867 | 0.28 | 0.88 | 69.93 | 57.04 |
| 9 | b | 6 | Int | BEG | 0.33 | 1.06 | −2.10 | −7.64 | 0.21 | 33.00 | 1.01 | 3.59 | 12.10 | 0.28 | 0.02 | −0.01 | −0.01 | 34 881 | 0.12 | 0.59 | 68.46 | 68.15 |
| 12 | a | 5 | Con | BEG | 0.30 | 1.34 | −2.12 | −7.17 | 0.18 | 19.33 | 2.36 | 1.95 | 11.54 | 0.09 | 0.02 | 0.02 | 0.05 | 59 130 | 0.24 | 0.66 | 75.97 | 95.18 |
| 12 | b | 6 | Con | HIP | 0.32 | 1.42 | −2.44 | −8.18 | 0.15 | 17.00 | 1.53 | 2.75 | 12.01 | −0.18 | 0.21 | 0.05 | 0.08 | 53 936 | 0.35 | 0.90 | 71.62 | 85.29 |
| 13 | a | 5 | Con | BEG | 0.26 | 1.51 | −2.36 | −6.85 | 0.17 | 11.00 | 2.11 | 2.36 | 10.94 | 0.03 | −0.02 | 0.07 | 0.12 | 12 878 | 0.09 | 0.25 | 72.45 | 109.26 |
| 21 | a | 7 | Int | BE | 0.34 | 1.53 | −5.36 | −11.61 | 0.11 | 14.00 | 1.66 | 1.86 | 11.89 | −0.32 | 0.20 | −0.02 | −0.08 | 15 090 | 0.43 | 0.17 | 59.98 | 99.46 |
| 21 | b | 5 | Int | BE | 0.32 | 1.54 | −2.71 | −12.51 | 0.15 | 12.50 | 2.46 | 2.18 | 11.62 | −0.27 | 0.32 | 0.06 | −0.04 | 7002 | 0.08 | 0.24 | 65.17 | 81.16 |
| 23 | a | 4 | Org | BE | 0.33 | 1.34 | −4.25 | −11.51 | 0.10 | 9.00 | 3.12 | 3.18 | 12.49 | 0.06 | 0.13 | 0.04 | 0.05 | 36 299 | 0.28 | 0.87 | 66.46 | 77.30 |
| 23 | b | 2 | Org | BE | 0.38 | 1.19 | −3.78 | −12.90 | 0.10 | 9.33 | 1.41 | 3.70 | 11.48 | −0.03 | 0.15 | 0.09 | 0.04 | 96 401 | 0.49 | 1.04 | 73.24 | 90.62 |
| 26 | a | 1 | Org | BE | 0.28 | 1.52 | −4.48 | −10.89 | 0.11 | 13.33 | 3.05 | 1.86 | 12.05 | −0.48 | −0.54 | 0.00 | 0.11 | 15 325 | 0.05 | 0.36 | 65.48 | 76.02 |
| 31 | b | 7 | Org | HIP | 0.37 | 1.18 | −2.59 | −10.46 | 0.14 | 12.50 | 1.57 | 3.92 | 13.85 | 0.13 | 0.02 | −0.07 | −0.03 | 19 807 | 0.03 | 0.14 | 59.61 | 49.14 |
| 35 | a | 1 | Con | HIP | 0.38 | 1.08 | −2.67 | −10.10 | 0.16 | 17.33 | 1.30 | 4.21 | 12.01 | 0.05 | 0.05 | 0.11 | −0.02 | 52 607 | 0.26 | 0.69 | 79.15 | 60.94 |
| 35 | b | 6 | Con | NCG | 0.35 | 1.17 | −2.59 | −9.14 | 0.16 | 21.00 | 2.24 | 2.87 | 11.45 | −0.03 | 0.20 | 0.09 | −0.05 | 38 364 | 0.18 | 0.65 | 74.26 | 70.11 |
| 37 | a | 7 | Con | HIP | 0.18 | 1.48 | −1.15 | −4.00 | 0.25 | 7.00 | 1.12 | 1.97 | 13.55 | 0.03 | −0.17 | −0.14 | −0.04 | 58 154 | 0.05 | 0.27 | 32.12 | 47.26 |
| 37 | b | 1 | Con | HIP | 0.19 | 1.39 | −1.52 | −4.18 | 0.26 | 21.67 | 1.24 | 2.61 | 13.68 | 0.11 | −0.19 | −0.07 | 0.04 | 47 972 | 0.12 | 0.39 | 62.65 | 41.99 |
| 55 | a | 1 | Int | AS | 0.25 | 1.17 | −2.02 | −5.90 | 0.25 | 25.00 | 2.03 | 2.00 | 9.99 | 0.01 | −0.18 | −0.05 | 0.00 | 94 097 | 0.25 | 1.07 | 54.33 | 60.61 |
| 55 | b | 6 | Int | AS | 0.27 | 1.25 | −2.34 | −6.94 | 0.20 | 15.67 | 1.58 | 2.25 | 10.83 | 0.00 | −0.15 | −0.11 | −0.02 | 86 971 | 0.25 | 0.70 | 59.49 | 35.39 |

**Table 1** (continued)

| Farm properties | | | | | Measured soil properties | | | | | | | | | Bacterial community structure | | | | Physical and biological stability | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Farm | Field | Crop type | Farm type | MSSG | Water content at –5 kPa (m3/m3) | Bulk density (g/cm3) | Water retention, parameter a | Water retention, parameter b | Macroporosity (m3/m3) | Root elongation (mm) | Penetration resistance (MPa) | Carbon (%) | C:N | TRFLP PC1 blue dye | TRFLP PC2 blue dye | TRFLP PC1 green dye | TRFLP PC2 green dye | Shear modulus, G (Pa) | Thixotropic resistance | Thixotropic resilience | Resistance to CU (%Rcu) | Resistance to heat (%Rht) |
| Overall statistics | | | | Mean | 0.32 | 1.29 | −2.88 | −9.27 | 0.16 | 15.76 | 1.82 | 2.89 | 11.95 | 0.02 | 0.00 | −0.01 | −0.02 | 56 432 | 0.24 | 0.68 | 66.99 | 73.56 |
| | | | | Std Dev | 0.06 | 0.18 | 1.13 | 3.53 | 0.05 | 6.24 | 0.81 | 0.81 | 0.87 | 0.18 | 0.18 | 0.09 | 0.09 | 33 564 | 0.16 | 0.31 | 13.66 | 19.86 |
| | | | | Max | 0.46 | 1.66 | −1.15 | −3.45 | 0.30 | 33.00 | 4.45 | 4.63 | 13.85 | 0.31 | 0.40 | 0.14 | 0.16 | 14 3996 | 0.81 | 1.26 | 103.97 | 111.71 |
| | | | | Min | 0.15 | 0.98 | −6.40 | −20.36 | 0.06 | 5.00 | 0.45 | 1.86 | 9.99 | −0.51 | −0.54 | −0.18 | −0.21 | 5117 | 0.02 | 0.03 | 18.31 | 33.20 |
| | | Upper quartile | | | 0.35 | 1.43 | −2.29 | −7.05 | 0.19 | 18.33 | 2.11 | 3.49 | 12.56 | 0.15 | 0.12 | 0.05 | 0.03 | 81 398 | 0.31 | 0.92 | 73.75 | 84.25 |
| | | Lower quartile | | | 0.28 | 1.15 | −3.48 | −11.53 | 0.13 | 12.50 | 1.37 | 2.22 | 11.51 | −0.03 | −0.08 | −0.06 | −0.09 | 35 590 | 0.10 | 0.38 | 59.80 | 60.77 |

Farm number refers to the farms described in Hawes *et al.* (2010). There are up to two fields (a, b) per farm. Crop types are as follows: 1, spring barley; 2, potatoes; 3, beans or peas; 4, barley undersown with peas; 5, winter barley; and 6, winter oilseed rape or calabrese; major soil sub-groups (MSSGs) are as follows: HIP, humus iron podzol; NCG, non-calcareous gley; BEG, brown earth with gleying; BE, brown earth; and AS, alluvial soil. Soil properties, bacterial community structure and soil stability measurements are as described in the text. Overall statistics were compiled using all replicate values.

*Background soil variables*

Major soil sub-group (MSSG) and texture were derived from the soil survey of Scotland, taken from map coordinates of each field. The individual soil samples were used for the determination of bacterial community structure by terminal restriction fragment length polymorphism (TRFLP), following collection of a subsample and DNA extraction (Deng *et al*., 2010) with both forward and reverse primers labelled with FAM (blue) and VIC (green), respectively. TRFLP data were Hellinger-transformed to reduce the effect of dominant peaks and yield-adjusted relative abundance data (Deng *et al*., 2010). The bacterial community structure data were then assessed by PCA to reduce dimensionality. This analysis was performed separately for each primer. PCA captured 17.4 and 12.3%, respectively, of the total variation in the first 2 of 34 dimensions of the analysis of the reverse primer labelled with VIC (green). Similarly, the first two of 50 dimensions of the FAM-labelled forward primer (blue) analysis captured 16.5 and 12.7% of the total variation. Root elongation of barley (*Hordeum vulgare*) was measured in individual intact cores of soil (Valentine *et al*., 2012). The assay measured the change in root length of a seedling over 48 h. Composite samples from each field were prepared by mixing the three individual samples and were used for the determination of percentage of sand and clay using a Mastersizer 2000 particle size analyzer (Malvern Instruments Limited, Malvern, UK); organic carbon and total nitrogen content using an Exeter Analytical CE440 Elemental Analyzer (EAI, Coventry, UK); and pH using a 1:2.5 solution of soil to water. The individual intact soil cores were saturated and then drained to 0, −0.5, −1 and −5 kPa water potential ($\psi$) on a ceramic tension table (ELE International, Leighton Buzzard, UK; Valentine *et al*., 2012). Water content, $\theta$, was assessed by weighing the soil at each water potential following oven-drying. The resulting water retention curve was used to calculate volumetric water content at field capacity (taken as −5 kPa) and macroporosity (>60 $\mu$m) all expressed on a per volume basis ($m^3/m^3$). As only a portion of the full water retention curve was measured, we used the simple relationship presented by Gardner *et al*. (1970) to describe its shape:

$$\psi = a(\theta)^{-b}, \tag{1}$$

where $\theta$ is the volumetric water content $m^3/m^3$, $\psi$ is the water potential (kPa), and $a$ and $b$ are fitting parameters (referred to hereafter as WR-$a$ and WR-$b$). Penetration resistance (PR; MPa) was measured on cores equilibrated to −20 kPa using a needle penetrometer (1 mm diameter, 30° cone angle, 4 mm/min penetration rate, readings; averaged at 1-mm intervals from 5- to 15-mm depth range) fitted to a mechanical test frame (Instron model 5544; Instron, MA, USA), with a 50-N load cell accurate to 2 mN at maximum load (Valentine *et al*., 2012). Bulk density ($g/cm^3$) was determined from the dry mass of soil in the core and the core volume.

*Soil stability variables*

The individual, mixed soil samples were used for the determination of indicators of biological and physical stability. The physical assay was based on the soil's rheological behaviour when wet. Soils were first air-dried and then passed through a 400-$\mu$m mesh to remove larger particles that could cause interlocking during the test (Barré & Hallett, 2009). Dry soil was then placed in a 35 mm diameter × 2 mm height ring on a tension table, where it was saturated for 24 h and then dried to −0.05 kPa before testing using a parallel plate rheometer (Thermo Haake MARS, Karlsruhe, Germany). The soil-filled rings were loaded between parallel platens that had serrated faces to decrease slippage at the soil–platen interface. Platens were brought together until the normal force was 0.2 N, which corresponded ca. to a 2-mm gap spacing for all soils. A Peltier unit controlled platen temperature at 20 ± 0.1 °C during testing. After 1 min to allow for temperature equilibration (soils were close to 20 °C before testing), a 0.5-Hz oscillatory shear stress of 10 Pa was applied. This stress was found to be within the linear viscoelastic range and was used to determine the shear modulus, $G$, which describes the stiffness of the soil. From a practical standpoint, this parameter describes the susceptibility of the soil to deformation. An increased shear stress of 500 Pa was then applied for 30 s to cause liquefaction of soils, followed by a return to 10 Pa shear stress for 120 s to measure thixotropy. From these measurements, the response to increased stress measured stability as thixotropic resistance.

$$\text{thixotropic resistance} = \frac{G}{G_{500\,\text{Pa}}}. \tag{2}$$

Recovery following removal of the stress measured thixotropic resilience.

$$\text{thixotropic resilience} = \frac{G}{G_{\text{post}\,500\,\text{P}}}. \tag{3}$$

These measurements quantify the propensity of soil bonds (a product of mineral and OM associations) to resist elevated mechanical stresses and then reform once the stress has been removed (see Barré & Hallett, 2009).

Biological stability was assessed from the recovery of short-term decomposition following heat or copper disturbance (Griffiths *et al*., 2001). These assays integrate differences in microbial community structure, soil OM, pH

and clay content that are all implicated in stability (Griffiths *et al.*, 2008). Briefly, three replicates of each soil (i.e. one sample from each of the three replicate soil samples per field) were amended with 100 $\mu$L sterile distilled water and heat-stressed at 40 °C, three replicates were amended with 100 $\mu$L $CuSO_4{\cdot}5H_2O$ equivalent to 500 $\mu$g Cu/g soil, and three replicates were amended with 100 $\mu$L sterile distilled water as controls, all for 18 h. The soils were then mixed with 100 mg barley shoot powder (C/N 14; Seeking Vitality, Inkberrow, UK), and the production of $CO_2$ over the following 24 h at 16 °C was measured by infrared gas analyser. The resistance to heat (%Rht) and the resistance to copper (%Rcu) were calculated as follows:

$$\text{Resistance} = \left(\frac{\text{CO}_2 \text{ treatment}}{\text{CO}_2 \text{ control}}\right) \times 100. \quad (4)$$

### *Statistical analysis*

The complete data set for analysis (summarised in Table 1 and shown in full in Table S1) consisted of the following: crop at sampling; farm location (GIS-X, GIS-Y); crop type (agronomically similar crops such as spring barley and spring wheat, or carrots and potatoes, were assigned to the same type from Hawes *et al.* (2010)); farm type (i.e. integrated, conventional, organic); water content at −5 kPa,; bulk density; WR-*b*; WR-*a*; macroporosity; root elongation; PR; bacterial community structure (PC 1 and 2 of TRFLP, as these accounted a large proportion, 29%, of the variation, and preliminary ANOVA indicated significant effects of farming type and region associated with PC1 and PC2); %C; C:N; MSSG; and pedology texture.

### *Linear mixed modelling*

Stepwise fixed effects model selection with a linear mixed-model framework (LMM) was undertaken using the V-search procedure in the Biometris procedure library for Genstat (version 16) and was used to model physical stability as resistance and resilience to compression and rotation (*G*, ThixResist, ThixResil), and biological stability as resistance to copper or heat stress (%Rcu and %Rht). The stability variables were then subjected to analysis of covariance (ANCOVA) using a blocking structure of sample nested within field, which was nested within farm with no treatment structure and the soil properties as co-variates. This was to determine at what level any relationship was occurring (i.e. between farm, within farm or within fields on farm).

### *Neural network analysis*

We used a KSOM to analyse the multidimensional data array comprising measured soil properties [−500, bulk density, Visser-b, macroporosity, root length 1, PR (MPa)]: bacterial community structure as PC scores using 'blue' and 'green' dyes (TR1blue,TR2blue, TR1green, TR2green); basic soil properties but only a single measurement from the field (C, C:N, MSSG, pedology texture); physical stability as resistance and resilience to compression and rotation (*G*, ThixResist, ThixResil); and biological stability as resistance to copper or heat stresses (%Rcu and %Rht). The correlation coefficients were calculated between the observed data and the features, or the outputs of the KSOM. Every output parameter was compared with the observed parameter to evaluate the performance of the mapping process. This is performed using the Curve Fitting toolbox in MATLAB.

The network topology is described by the number of output neurons presented in the network and by the way in which they are interconnected. In this study, neurons in the output layer are arranged in a hexagonal grid where every neuron is connected to six neighbours, except for the ones at the edge of the grid. There is a need to point out that while the rows and the columns on the output layer are interpreted as co-ordinate axes to locate units and upon which the output of the KSOM can be interpreted, they do not have explicit meaning or relations to the variables of the input data set.

The number of neurons (Map size), $M$, may vary, where the number of neurons affects accuracy and generalisation capability of the KSOM. The self-organising map team at the Helsinki University of Technology offers guidance for determining the optimum number of neurons using the following heuristic formula (Vesanto *et al.*, 2000).

$$M = 5\sqrt{N}, \quad (5)$$

where $M$ is the number of map units or neurons, and $N$ is the total number of data samples or records. Once $M$ is known, the number of rows and columns in the KSOM can be determined. A guideline by the self-organising map team is in the following equation

$$\frac{l_1}{l_2} = \sqrt{\frac{e_1}{e_2}}, \quad (6)$$

where $l_1$ and $l_2$ are the number of rows and columns, respectively, $e_1$ is the biggest eigenvalue of the training data set, and $e_2$ is the second biggest eigenvalue.

Upon start of training, the initial values of the elements of the weight vectors in the grid are randomly assigned, usually numbers between zero and one. Then, the weight vectors are updated using two types of training algorithms: sequential training algorithms and batch training algorithms.

The multidimensional input data are first standardised by deducting the mean and then dividing the result by the standard deviation. This procedure ensures that every

variable has equal importance in training the KSOM, so that no components will have excessive influence or control of the training results by virtue of its higher absolute value. Then, a standardised input vector is chosen at random and presented to each of the individual neurons in the output layer or map for comparison with their code vectors to identify the code vector most similar to the presented input vector. The identification uses the Euclidian distance $D_i$, which is defined as follows:

$$D_i = \sqrt{\sum_{j=1}^{n}(x_{ij} - m_{ij})^2};\ i = 1, 2, \ldots, M, \tag{7}$$

where $D_i$ is the Euclidian distance between the input vector and the weight (or code) vector $i$; $x_{ij}$ is the $j$th element of the current input vector; $m_{ij}$ is the $j$th element of the weight vector $i$; $n$ is the dimensionality of both the input and the code vector; and $M$ is the number of neurons in the KSOM (or the size of the map). The neuron whose vector most closely matches the input data vector (i.e. for which the $D_i$ is minimum) is chosen as a winning node or the best matching unit (BMU). The weight vectors of this winning node and those of its adjacent neurons are then adjusted to match the input data using, thus bringing the weight vectors further into agreement with the input vector (Vesanto *et al*., 2000).

$$m_i(t+1) = m_i(t) + \alpha(t)\mathrm{h}_{ci}(t)[x(t) - m_i(t)], \tag{8}$$

where $t$ denotes time, $\alpha(t)$ is the learning rate at $t$, $h_{ci}(t)$ is the neighbourhood function centred in the winner unit $c$ at time $t$, and $h_{ci}(t)$ defines the region of the influence that the input sample has on the KSOM.

The nodes surrounding the winning node, its neighbourhood, are also updated so that they are made to look less like the input vector. The size of adjustment in the weight vector of the neighbouring neurons is dependent on the distance of those neurons from the winner in the output array. This adaptation procedure stretches the BMU and its topological neighbours towards the sample vector. In this manner, each node in the map internally develops the ability to recognise input vectors similar to itself.

Two parameters are used for training the KSOM: the learning rate [$\alpha(t)$] and the neighbourhood width parameter ($h_c$). The learning rate influences the size of the weight vector adjustments after each training step, whereas the neighbourhood width parameter determines to what extent the surrounded neurons are affected by the winner. Both the learning rate and the neighbourhood width are time dependent and are typically changed from large to small to provide the best performance with the smallest training time.

Linear learning function is used in this study

$$\alpha(t) = \alpha_0\left(1 - \frac{t}{T}\right), \tag{9}$$

where $T$ is the training length or the number of iterations, and $\alpha_0$ is the initial learning rate. In the KSOM toolbox of Matlab, $\alpha_0$ is specified as 0.5.

The neighbourhood function used was Gaussian centred in the winner unit $c$, such that:

$$h_{ci}(t) = \exp^{-(d_{ci}(t))^2/(2\sigma^2(t))} = \exp^{-(||r_c - r_i||))^2/(2\sigma^2(t))}. \tag{10}$$

In other words, all neurons $i$ located in a topological neighbourhood of the winning neurons $c$ will have their weights updated usually with a strength related to their distance $d_{ci}$ from the winning neuron, where $d_{ci}$ can be calculated using

$$d_{ci} = ||r_c - r_i||, \tag{11}$$

where $r_c$ and $r_i$ are the positions of nodes $c$ and $i$ on the KSOM grid known as the norm city-block distance.

$\sigma^2$ is the variance parameter specifying the spread of the Gaussian function.

Like the learning rate $\alpha(t)$, $\sigma(t)$ also decreases linearly as the number of iterations increases. In the early stages of training, the radius of the neighbourhood is large and most of the KSOM neurons belong to any node's neighbourhood. As the training progresses, the radius is reduced to allow good local ordering. In the KSOM toolbox in Matlab, the initial radius $\sigma_0$ is specified as max (1, $M$/4), where $M$ is the size of the map.

The quality of the trained KSOM is measured by the total average quantisation error and total topographic error. The quantisation error measures the quality of the map fitting to the data, that is the average distance between each data vector and its BMU at convergence. This error is calculated using:

$$q_\mathrm{e} = \frac{1}{N}\sum_{i=1}^{N}||x_i - m_c||, \tag{12}$$

where $q_\mathrm{e}$ is the quantisation error, $x_i$ is the $i$th data sample or vector $m_c$ is the prototype vector of the BMU for $x_i$, and $||\cdot||$ denotes Euclidian distance.

The topographic error, $t_\mathrm{e}$, is an indication of the degree of preservation of the topology of the data when fitting the map to the original data set. In other words, it measures the similarity between the neighbour on the model and on the input space. It is calculated as the proportion of sample vectors for which two best and the next BMUs for a given input vector are not adjacent, that is

$$t_e = \frac{1}{N}\sum_{i=1}^{N} u(x_i), \quad (13)$$

where $N$ is the number of samples, $x_i$ is the $i$th data sample, and $u(x_i)$ is a binary integer such that it is equal to 1 if the first and second BMUs of the map are not adjacent units; otherwise, it is zero. The results of this error measure are very easy to interpret and are also directly comparable between different models.

## Results

In the analysis, the final map size $M$, that is the number of neurons in the map, was 48, arranged into eight columns and six rows (8 × 6). Thus, it is apparent that the KSOM has significantly reduced the original large data array, 87 data points with 17 variables, to something much more manageable by eliminating the redundant elements. The final quantisation error was 2.56, and the total topographic error was 0.023. The variables used in the analyses are summarised in Table 1, and those identified as being significantly related to the measures of stability, by either LMM or KSOM, are shown in Table 2. The performance of KSOM in modelling measured values in terms of the correlation coefficient ($R$) is seen in Table 3 from which it is clear that the KSOM model could model the majority of the parameters as their $R$ values are over 70% (very strong for uncertain environmental data, Rustum & Adeloye, 2007; Rustum *et al*., 2008; Adeloye *et al*., 2012). The KSOM model failed to successfully model the measured values for thixotropic resistance. KSOM component plane analysis helps to visually illustrate the relationship between any variable and other variables (Figure 1) which can be viewed as a 'sliced' version of the KSOM and which indicated the relationships between the variables, but correlation coefficients were also calculated and compared with the LMM results as shown in Table 2.

### *Shear modulus (G)*

Linear mixed-model indicated a significant association of $G$ with bulk density (Table 2), and ANCOVA indicated that this association was significant ($P = 0.013$) at the between-farm level, as shown in Figure 2. KSOM also picked out bulk density, but additionally bacterial community structure, with the TRFLP principal components TR2G and TR1B, as being significantly associated with $G$ (Table 2).

### *Thixotropic resistance and resilience*

Although LMM showed significant correlation of thixotropic resistance and resilience with bacterial community structure, TR1G and TR2B (Table 2), ANCOVA indicated that the significant differences were within a field rather than between fields within a farm or even between farms. KSOM did not detect any relationship with bacterial community structure, but rather with water retention characteristics (WR-$b$ and water content at −5 kPa) for thixotropic resistance and with $G$ for thixotropic resilience (Table 2).

### *Resistance to heat*

No significant correlations with resistance to heat were identified by LMM, while KSOM identified correlations with both bacterial community structure (TR1B and TR1G) and soil physical parameters (bulk density and PR) (Table 2).

### *Resistance to copper*

Linear mixed-model revealed significant correlations with resistance to copper and water content at −5 kPa; bacterial

**Table 2** Soil parameters identified as significantly related to measures of soil stability from analyses using linear mixed model (LMM) and Kohonen self-organising map (KSOM)

| Stability | LMM (direction, $P$ value) | KSOM (correlation coefficient) |
|---|---|---|
| Shear modulus – $G$ | **Bulk density** (-ve. 0.05) | **Bulk density** (−0.69), TRFLP (PC2) green dye (−0.64), TRFLP (PC1) blue dye (+0.79) |
| Thixotropic – resistance | TRFLP (PC1) green dye (+ve, 0.003), TRFLP (PC2) blue dye (+ve, 0.033) | WR-$b$ (−0.82), water content −5 kPa (+0.78) |
| Thixotropic – resilience | TRFLP (PC1) green dye (+ve, 0.048) | Shear modulus – $G$ (+0.83) |
| Resistance to heat | – | TRFLP (PC1) blue dye (−0.72), bulk density (+0.62), penetration resistance (+0.87), TRFLP (PC1) green dye (+0.72) |
| Resistance to copper | **Water content −5 kPa** (+ve, 0.0025), **TRFLP (PC1) green dye** (+ve, 0.003), bulk density (-ve, 0.011), **%C** (+ve, 0.015) | WR-$b$ (−0.64), macroporosity (−0.61), **water content −5 kPa** (+0.78), **TRFLP (PC1) green dye** (+0.68), **%C** (+0.72) |

Parameters in bold are common to both analytical methods. A positive or negative direction of relationship in LMM is indicated by +ve or –ve, respectively. Soil properties, bacterial community structure and soil stability measurements are as described in the text.

**Table 3** The performance of a Kohonen self-organising map in modelling measured values in terms of correlation coefficient

| Measured value ($X$) | Correlation | Linear model Poly1: KSOMX = p1 × $X$ + p2 Coefficients (with 95% confidence bounds) |
|---|---|---|
| Water content −5 kPa | 0.92 | p1 = 0.79 (0.71, 0.86)<br>p2 = 0.067 (0.042, 0.092) |
| Bulk density | 0.89 | p1 = 0.64 (0.57, 0.71)<br>p2 = 0.44 (0.35, 0.53) |
| Fitting parameter (WR-*b*) | 0.87 | p1 = 0.65 (0.57, 0.73)<br>p2 = −3.23 (−4.05, −2.42) |
| Macroporosity | 0.89 | p1 = 0.73 (0.65, 0.81)<br>p2 = 0.048 (0.035, 0.061) |
| Root elongation | 0.61 | p1 = 0.26 (0.18, 0.33)<br>p2 = 11.29 (10.07, 12.5) |
| Penetration resistance | 0.64 | p1 = 0.32 (0.24, 0.41)<br>p2 = 1.16 (1.01, 1.32) |
| TRFLP PC1 blue dye | 0.80 | p1 = 0.42 (0.35, 0.49)<br>p2 = 0.012 (0.0004, 0.025) |
| TRFLP PC2 blue dye | 0.87 | p1 = 0.56 (0.49, 0.63)<br>p2 = 0.0026 (−0.0086, 0.014) |
| TRFLP PC1 green dye | 0.74 | p1 = 0.42 (0.34, 0.51)<br>p2 = −0.0073 (−0.014, 9.2e−005) |
| TRFLP PC2 green dye | 0.78 | p1 = 0.47 (0.38, 0.55)<br>p2 = −0.013 (−0.020, −0.0062) |
| Carbon (%) | 0.89 | p1 = 0.69 (0.61, 0.76)<br>p2 = 0.91 (0.68, 1.1) |
| C:N | 0.69 | p1 = 0.27 (0.20, 0.33)<br>p2 = 8.83 (8.09, 9.57) |
| Shear modulus ($G$) | 0.76 | p1 = 0.43 (0.35, 0.52)<br>p2 = 3.06e+004 (2.55e+004, 3.56e+004) |
| Thixotropic resistance | 0.43 | p1 = 0.17 (0.094, 0.25)<br>p2 = 0.0096 (−0.011, 0.031) |
| Thixotropic resilience | 0.78 | p1 = 0.57 (0.47, 0.66)<br>p2 = 0.28 (0.21, 0.35) |
| Resistance to copper (%Rcu) | 0.71 | p1 = 0.41 (0.32, 0.49)<br>p2 = 39.75 (33.78, 45.72) |
| Resistance to heat (%Rht) | 0.81 | p1 = 0.50 (0.42, 0.58)<br>p2 = 35.6 (29.75, 41.45) |

The correlation coefficients were calculated between the observed data and the features, or the outputs of the KSOM. Every output parameter was compared with the observed parameter to evaluate the performance of the mapping process. This is performed using the Curve Fitting toolbox in MATLAB. Soil properties, bacterial community structure and soil stability measurements are as described in the text.

community structure (TR1G); bulk density; and carbon content (%C) (Table 2), while ANCOVA indicated significant differences between farms ($P = 0.003$) as shown in Figure 3. There was no clear link between resistance to copper and farm type; crop type (i.e. barley, potatoes etc.); or soil type, but the east–west location of the farm was significantly correlated (Figure 4). The correlation was influenced by the western most two farms although no single factor from their analysis was responsible (such as %C, C:N, soil type). KSOM also identified water content at −5 kPa and bacterial community structure (TR1G) as being significantly correlated with resistance to copper, as well as macroporosity and WR-*b* (Table 2).

## Discussion

We used a KSOM, the principal goal of which is to transform an incoming signal pattern of arbitrary dimension into a two-dimensional discrete map. It involves clustering the input patterns in such a way that similar patterns are represented by the same neuron or one of its close

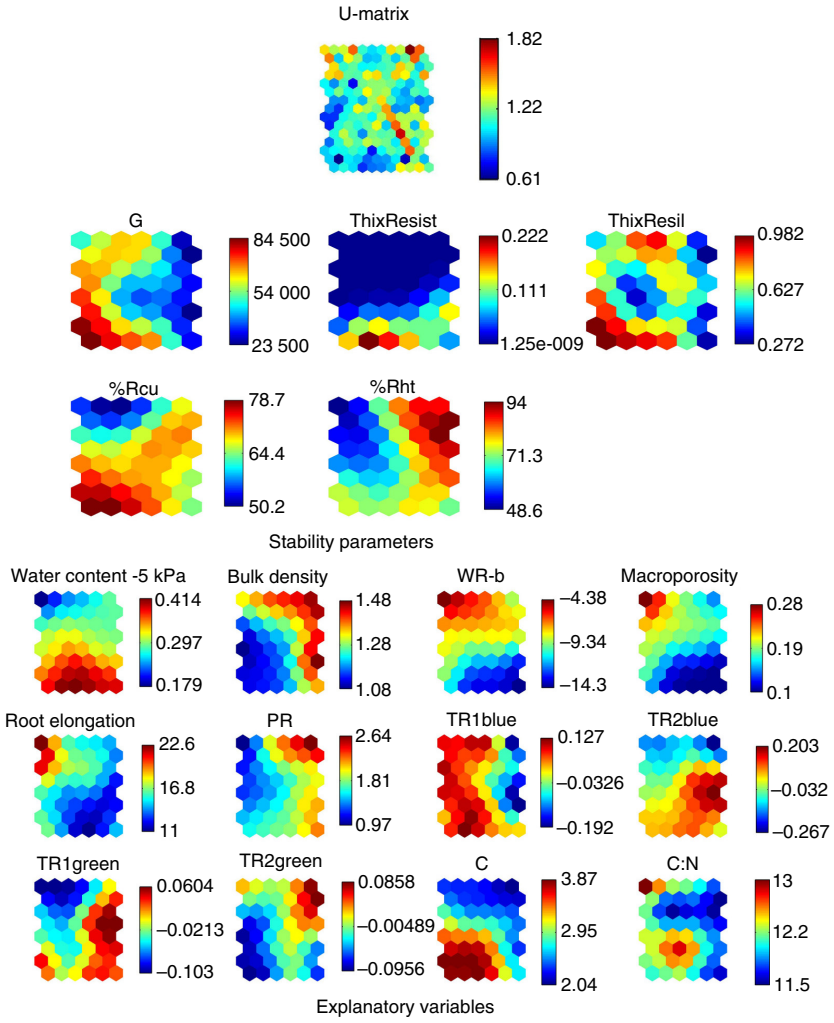


**Figure 1** Component planes of the Kohonen self-organising maps of the stability parameters [as described in the text : shear modulus (*G*), thixotropic (Thix) resistance (resist) and resilience (resil), % resistance to copper (%RCu) and heat (%Rht)] and explanatory soil variables[(as described in the text includes water retention parameter *b* (WR-*b*), penetration resistance (PR), PC 1 and 2 from TRFLP of bacterial community structure using blue and green dyes (TR)] from arable farms across east and north-east Scotland. The component planes help to visually illustrate the relationship between any variable and other variables considered in this analysis. For example, the plane for water content −5 kPa is highly negatively correlated with WR-*b* and highly positively correlated with ThixResist. The U-matrix shows the cluster structure of the KSOM, representing the Euclidean distance between neighbouring units.

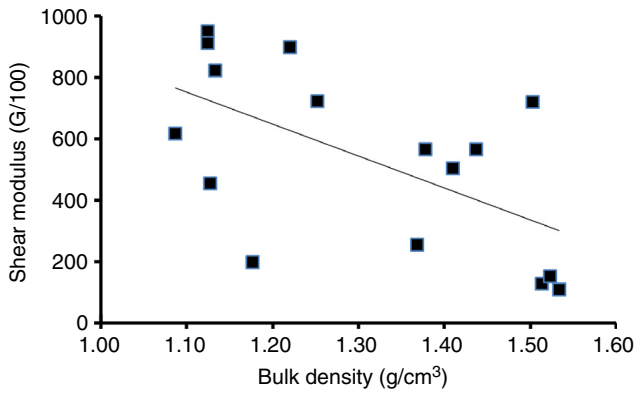


**Figure 2** Relationship at the farm level between soil bulk density and soil shear modulus (*G*), an indicator of physical stability. Means of 16 farms. $R^2 = 0.3497$.

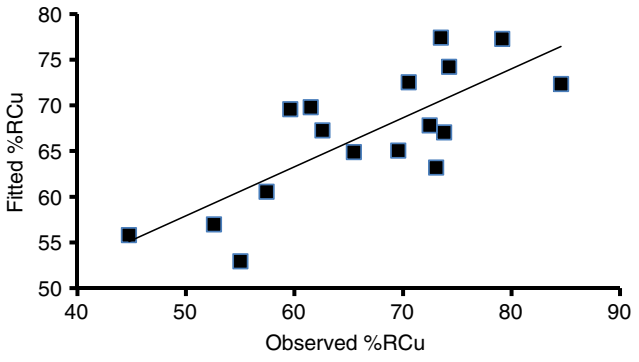


**Figure 3** Plot of observed versus fitted resistance to copper (%RCu) at the farm level from the linear mixed model (LMM) based on fixed effects only [46.51 + (66.3 × water content −5 kPa) + (54.56 × TRFLP PC1 green dye]. Means of 16 farms. Adjusted $R^2 = 0.5418$.

neighbours on the map (Back *et al.*, 1998), thus converting a complex, nonlinear statistical relationship between large dimensional data into a simple relationship on a small dimensional display (Kohonen *et al.*, 1996). All the agricultural fields in this study fell within a specific geographical area and on a limited range of soil types. For this reason, pH and clay–sand content were not used in this

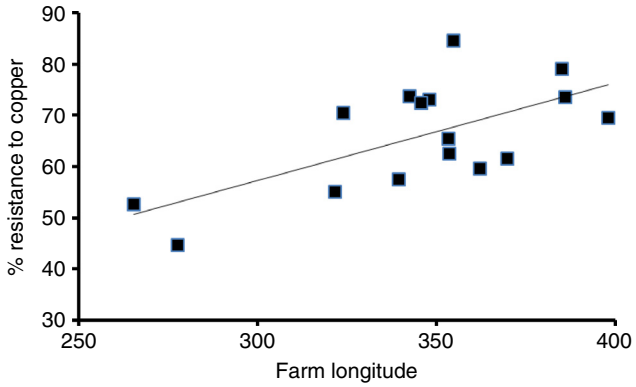


**Figure 4** Average farm resistance to copper plotted against farm longitude (GIS_X/1000). $R^2 = 0.4238$.

analysis. pH varied from 5.2 to 6.1, much less than from natural soils (i.e. the wider survey of resilience in Scottish soils had a range from 3.5 to 7.2, Kuan *et al*., 2007), while clay–sand content was related to MSSG and pedology texture which were included. The KSOM analysis identified a greater number of significant associations between soil parameters and the measures of stability than the LMM. That in itself is not surprising because of the more complex (nonlinear) nature of the analysis, which is a strength of KSOM. A weakness is that, unlike LMM, it does not have the ability of LMM to model the hierarchy of sample within field within farm in the sampling design and hence treats each observation as independent. This is likely to lead to overstating statistical significance. However, the fact that many of the same parameters were identified as being significantly correlated with stability indicates that they are important driving variables. The other important message obvious from the two analyses is that both biological and physical soil parameters are contributing to both biological and physical soil stability. This result from field samples supports the experimental work from Griffiths *et al*. (2008) and Zhang *et al*. (2010) which concluded that biophysical interactions determine soil stability.

From the analyses in this study, bulk density, water content at −5 kPa and %C were identified as significant variables by both modelling schemes (Table 2) and could therefore be taken as the most appropriate for explaining the causal relationship between soil properties and soil stability. In a wider survey, Emmett *et al*. (2010) observed a negative correlation between bulk density and soil carbon in UK arable soils and that arable soils had the greatest soil bulk density (compared with other land uses). This implies that arable soil will be the land use most at risk from reducing soil stability. Soil structure must be critical as the distribution of pores will affect water retention properties and bulk density in the field. %C will indirectly contribute to soil structure but should also be directly related to soil mechanical behaviour [as measured by *G* in this paper or compression characteristics in Debeljak *et al*. (2009) and Gregory *et al*. (2009)]. The KSOM did indicate a positive contribution of %C to *G* (correlation coefficient +0.54) but that was less than the other measures shown in Table 2. The measurements of %C and water content at −5 kPa were positively correlated (0.79) and %C and bulk density negatively correlated (−0.86). %C will also be directly involved in stability to copper because of its sequestration of heavy metals and reduction in the concentration of available copper (Kuan *et al*., 2007). Biological parameters, here represented as changes in bacterial community structure, were also correlated with physical stability (*G*). There is no indication whether the different bacterial community structures influence physical stability through the production of extra-cellular compounds (i.e. mucilage) binding soil particles together (Czarnes *et al*., 2000) and alterations to soil structure (Helliwell *et al*., 2014) or simply that a different soil structure supports different bacterial community structures (Griffiths *et al*., 2008). The identification of a role for bacterial community structure in soil stability is a first step in understanding the interactions involved and opens the way to measure the correlation between soil stability and specific microbial groups with identified functions.

In previous research exploring soil stabilisation by a range of biological compounds, the rheological approach found impacts of soil moisture and biological exudates on *G* and thixotropic resistance, but recovery measured as thixotropic resilience was negligible, presumably because of irreversible damage to interparticle bonds (Barré & Hallett, 2009). Using the much broader range of soils in the current study, the same tests found much better resilience, suggesting the rapid development of interparticle bonds from mineral, organo-mineral and capillarity forces following the removal of stresses from some soils. The LMM analysis indicated that while there were significant differences observed in thixotropic resistance and resilience, they were within a field rather than between fields within a farm or even between farms. This is likely to mean the data are dominated by a few very positive correlations, but that there is no overall consistent pattern. KSOM also indicated that thixotropic resilience was related to the initial shear modulus (*G*), a measure of mechanical stability. Resistance and resilience to heat and copper have been used previously as measures of biological stability (e.g. Griffiths *et al*., 2001), although here we used resistance alone: this has been shown to be strongly related to resilience and so is an accurate measure of stability (Gregory *et al*., 2009). The lack of correlation between resistance to heat and any of the soil parameters by LMM mirrors previous results (Kuan *et al*., 2007), and analysis using multiobjective regression trees also failed to detect any relationships between soil physico-chemical parameters and stability to heat (Debeljak *et al*., 2009). KSOM, however, did show correlations for stability to heat

with indicators of soil structure and microbial community structure. Resistance to copper showed highly significant correlations at the farm level, and further analysis indicated that there was also a correlation with farm location (east–west axis). This would need to be confirmed in a subsequent survey, but our analysis did not detect any consistent trends in the soil variables of these particular farms, although there is a consistent change in climate over the region (Birse & Dry, 1970). The results do support the observations that biological and physical variables jointly contribute to soil stability.

## Conclusions

The high-level land management decisions, that is farm types (conventional, organic, integrated) or crop types, were not picked out at this regional scale study, nor was the underlying soil type. The broad nature of the current survey, even though it was focussed on fields under arable production from a similar geographic region, was only designed to highlight the basic underlying factors. The fact that for stability to both compression ($G$) and copper, the LMM data could be grouped by farm which might indicate that the different management practices between farms were overridden by the underlying soil parameters on each farm. The experimental evidence strongly indicates that within a farm, stability can be improved by soil amendments and by incorporation of a grass rotation (Gregory *et al*., 2009; Stockdale *et al*., 2013), but that at this regional scale, other factors determine the overall pattern of stability. Management decisions on improving soil stability therefore need to be taken at the individual field scale.

## Acknowledgements

The study was partly funded by the Rural and Environment Science and Analytical Services Division of the Scottish Government. The authors thank the anonymous reviewers for their helpful insights and comments.

## References

Soil Association organic soil standards. Available at: http://www.soilassociation.org/Whatisorganic/Organicstandards, accessed 28/07/2015.

Adeloye, A.J., Rustum, R. & Kariyama, I.D. 2012. Neural computing modeling of the reference crop evapotranspiration. *Environmental Modelling & Software*, **29**, 61–73.

Back, B., Sere, K. & Hanna, V. 1998. Managing complexity in large database using self organising map. *Accounting, Management and Information Technologies*, **8**, 191–210.

Barré, P. & Hallett, P.D. 2009. Rheological stabilization of wet soils by model root and fungal exudates depends on clay mineralogy. *European Journal of Soil Science*, **60**, 525–538.

Birse, E.L. & Dry, F.T. 1970. *Assessment of climatic conditions in Scotland 1: based on accumulated temperature and potential water deficit*. The Macaulay Institute for Soil Research, Aberdeen.

Creamer, R.E., Brennan, F., Fenton, O., Healy, M.G., Lalor, S.T.J., Lanigan, G.J., Regan, J.T. & Griffiths, B.S. 2010. Implications of the proposed Soil Framework Directive on agricultural systems in Atlantic Europe – a review. *Soil Use and Management*, **26**, 198–211.

Czarnes, S., Hallett, P.D., Bengough, A.G. & Young, I.M. 2000. Root- and microbial-derived mucilages affect soil structure and water transport. *European Journal of Soil Science*, **51**, 435–443.

De la Rosa, D., Mayol, F., Diaz-Pereira, E. & Fernandez, M. 2004. A land evaluation decision support system (MicroLEIS DSS) for agricultural soil protection. *Environmental Modelling and Software*, **19**, 929–942.

Debeljak, M., Kocev, D., Towers, W., Jones, M., Griffiths, B.S. & Hallett, P.D. 2009. Potential of multi-objective models for risk-based mapping of the resilience characteristics of soils: demonstration at a national level. *Soil Use and Management*, **25**, 66–77.

Degens, B.P., Schipper, L.A., Sparling, G.P. & Duncan, L.C. 2001. Is the microbial community in a soil with reduced catabolic diversity less resistance to stress or disturbance? *Soil Biology & Biochemistry*, **33**, 1143–1153.

Deng, H., Li, X.-F., Cheng, W.-D. & Zhu, Y.-G. 2009. Resistance and resilience of Cu-polluted soil after Cu perturbation, tested by a wide range of soil microbial parameters. *FEMS Microbiology Ecology*, **70**, 137–148.

Deng, H., Zhang, B., Yin, R., Wang, H., Mitchell, S., Griffiths, B.S. & Daniell, T.J. 2010. Long-term effect of re-vegetation on the microbial community of a severely eroded soil in sub-tropical China. *Plant and Soil*, **328**, 447–458.

Emmett, B.A., Reynolds, B., Chamberlain, P.M., Rowe, E., Spurgeon, D., Brittain, S.A., Frogbrook, Z., Hughes, S., Lawlor, A.J., Poskitt, J., Potter, E., Robinson, D.A., Scott, A., Wood, C. & Woods, C. 2010. Countryside Survey: Soils Report from 2007. Technical Report No. 9/07 NERC/Centre for Ecology & Hydrology. 192 pp. (CEH Project Number: C03259).

Gardner, W.R., Hillel, D. & Benyamini, Y. 1970. Post irrigation movement of soil water: I. Redistribution. *Water Resources Research*, **6**, 851–861.

Gregory, A.S., Watts, C.W., Griffiths, B.S., Hallett, P.D., Kuan, H.L. & Whitmore, A.P. 2009. The effect of long-term soil management on the physical and biological resilience of a range of arable and grassland soils in England. *Geoderma*, **153**, 172–185.

Griffiths, B.S. & Philippot, L. 2013. Insights into the resistance and resilience of the soil microbial Community. *FEMS Microbiology Reviews*, **37**, 112–129.

Griffiths, B.S., Ritz, K., Wheatley, R., Kuan, H.L., Boag, B., Christensen, S., Ekelund, F., Sørensen, S.J., Muller, S. & Bloem, J. 2001. An examination of the biodiversity-ecosystem function relationship in arable soil microbial communities. *Soil Biology & Biochemistry*, **33**, 1713–1722.

Griffiths, B.S., Hallett, P.D., Kuan, H.L., Gregory, A.S., Watts, C.W. & Whitmore, A.P. 2008. Functional resilience of soil microbial communities depends on both soil structure and microbial community composition. *Biology and Fertility of Soils*, **44**, 745–754.

Hawes, C., Squire, G.R., Hallett, P.D., Watson, C.A. & Young, M. 2010. Arable plant communities as indicators of farming practice. *Agriculture, Ecosystems and Environment*, **138**, 17–26.

Helliwell, J.R., Miller, A.J., Whalley, W.R., Mooney, S.J. & Sturrock, C.J. 2014. Quantifying the impact of microbes on soil structural development and behaviour in wet soils. *Soil Biology & Biochemistry*, **74**, 138–147.

Hueso, S., Hernández, T. & García, C. 2011. Resistance and resilience of the soil microbial biomass to severe drought in semiarid soils: the importance of organic amendments. *Applied Soil Ecology*, **50**, 27–36.

Kohonen, T., Hynninen, J., Kangas, J. & Laaksonen, J. 1996. SOM_PAK: The self-organizing map program package. Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo.

Kuan, H.L., Hallett, P.D., Griffiths, B.S., Gregory, A.S., Watts, C.W. & Whitmore, A.P. 2007. The biological and physical stability and resilience of a selection of Scottish soils to stresses. *European Journal of Soil Science*, **58**, 811–821.

Mele, P.A. & Crowley, D.E. 2008. Application of self-organizing maps for assessing soil biological quality. *Agriculture, Ecosystems and Environment*, **126**, 139–152.

Mertens, J., Ruyters, S., Springael, D. & Smolders, E. 2007. Resistance and resilience of zinc tolerant nitrifying communities is unaffected in long-term zinc contaminated soils. *Soil Biology & Biochemistry*, **39**, 1828–1831.

Morris, C. & Winter, M. 1999. Integrated farming systems: the third way for European agriculture? *Land Use Policy*, **16**, 193–205.

National soil inventory for Scotland. Available at: http://www.soils-scotland.gov.uk, accessed 28/07/2015.

Rustum, R. & Adeloye, A.J. 2007. Replacing outliers and missing values from activated sludge data using Kohonen self-organizing map. *Journal of Environmental Engineering*, **133**, 909–916.

Rustum, R., Adeloye, A.J. & Scholz, M. 2008. Applying Kohonen self-organizing map as a software sensor to predict biochemical oxygen demand. *Water Environment Research*, **80**, 32–40.

Schultz, A. & Wieland, R. 1997. The use of neural networks in agroecological modelling. *Computers and Electronics in Agriculture*, **18**, 73–90.

Seybold, C.A., Herrick, J.E. & Brejda, J.J. 1999. Soil resilience: a fundamental component of soil quality. *Soil Science*, **164**, 224–234.

Stockdale, E.A., Banning, N.C. & Murphy, D.V. 2013. Rhizosphere effects on functional stability of microbial communities in conventional and organic soils following elevated temperature treatment. *Soil Biology & Biochemistry*, **57**, 56–59.

Tobor-Kaplon, M.A., Bloem, J., Romkens, P.F.A.M. & de Ruiter, P.C. 2005. Functional stability of microbial communities in contaminated soils. *Oikos*, **111**, 119–129.

Tobor-Kaplon, M.A., Bloem, J. & de Ruiter, P.C. 2006. Functional stability of microbial communities from long-term stressed soils to additional disturbance. *Environmental Toxicology and Chemistry*, **25**, 1993–1999.

Valentine, T.A., Hallett, P.D., Binnie, K., Young, M.W., Squire, G.R., Hawes, C. & Bengough, A.G. 2012. Soil strength and macropore volume limit root elongation rates in many UK agricultural soils. *Annals of Botany*, **110**, 259–270.

Vesanto, J., Himberg, J., Alhoniemi, E. & Parhankagas, J. 2000. SOM Toolbox for Matlab 5, Report A57. Available at: http://www.cis.hut.fi/projects/somtoolbox/. Accessed 28/08/2015.

Zhang, B., Deng, H., Wang, H., Yin, R., Hallett, P.D., Griffiths, B.S. & Daniell, T.J. 2010. Does microbial habitat or community structure drive the functional stability of microbes to stresses following re-vegetation of a severely degraded soil? *Soil Biology & Biochemistry*, **42**, 850–859.

Zhang, B., Wang, H., Yao, S. & Bi, L. 2013. Litter quantity confers soil functional resilience through mediating soil biophysical habitat and microbial community structure on an eroded bare land restored with mono *Pinus massoniana*. *Soil Biology & Biochemistry*, **57**, 556–567.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Table S1.** Complete data set from farm survey.