Received: 27 July 2021 | Revised: 26 April 2022 | Accepted: 13 May 2022

DOI: 10.1111/sum.12827

RESEARCH PAPER

Soil Use and Management | BRITISH SOCIETY OF SOIL SCIENCE | WILEY

Evidence of non-site-specific agricultural management effects on the score of visual soil quality indicators

Fernando Teixeira[1] | **Tatenda Lemann**[2] | **Carla Ferreira**[3,4,5] | **Matjaž Glavan**[6] | **Tóth Zoltán**[7] | **Tamás Hermann**[8] | **Jerzy Lipiec**[9] | **Magdalena Frąc**[9] | **Endla Reintam**[10] | **Minggang Xu**[11] | **Hongzhu Fan**[12] | **Coen Ritsema**[13]

[1]MED – Mediterranean Institute for Agriculture, Environment and Development, Institute for Advanced Studies and Research, Universidade de Évora, Évora, Portugal

[2]Centre for Development and Environment (CDE), University of Bern, Bern, Switzerland

[3]Department of Physical Geography and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden

[4]Navarino Environmental Observatory, Greece

[5]Research Centre for Natural Resources, Environment and Society (CERNAS), College of Agriculture, Polytechnic Institute of Coimbra, Coimbra, Portugal

[6]Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia

[7]Institute of Agronomy, Georgikon Campus, Keszthely, Hungarian University of Agriculture and Life Sciences, Gödöllő, Hungary

[8]Institute of Advanced Studies, Kőszeg, Hungary

[9]Institute of Agrophysics, Polish Academy of Sciences, Lublin, Poland

[10]Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, Tartu, Estonia

[11]Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences (IARRP, CAAS), Beijing, China

[12]Soil and Fertilizer Institute of the Sichuan Academy of Agricultural Sciences (SFI), Chengdu, China

[13]Department of Environmental Sciences, Soil Physics and Land Management, Wageningen University, Wageningen, The Netherlands

Correspondence
Fernando Teixeira, MED – Mediterranean Institute for Agriculture, Environment and Development, Institute for Advanced Studies and Research, Universidade de Évora, Pólo da Mitra, Ap. 94, 7006-554 Évora, Portugal.
Email: fteixeir@uevora.pt

Funding information
Swiss State Secretariat for Education, Research and Innovation; Chinese Academy of Sciences; Chinese Ministry of Science and Technology; Horizon 2020; European Union

Abstract

This study investigates 11 agricultural management practices (AMPs) and their effects on seven visual soil quality indicators and soil aggregate stability. The survey carried out across eight pedoclimatic zones in Europe and China was based on visual soil assessments (New Zealand VSA method) performed on soils subject to different soil management practices and nearby similar soils, under similar farming features, without the distinctive soil management practice (control). Fisher's exact test was used to test if the management treatment was independent of the score of each visual soil quality indicator and to test if the management treatment produced a higher frequency of the score 'good'. The results showed a statistically significant ($\alpha < .05$) higher frequency of the score 'good' for 'soil structure and consistency' and/or 'soil porosity' for six AMPs. For no-till AMP, the null hypothesis can also be rejected for 'susceptibility to erosion' and 'soil stability' and for 'mulching + permanent soil cover' AMP, for the 'presence of tillage pan' and 'soil colour'. The hypothesis that the management treatment was independent of the score of each indicator was rejected for 'soil structure and consistency' of three AMPs, for 'soil porosity' of three AMPs, for 'soil colour' of one AMP and for the ‘presence of tillage pan’ of one AMP. This study demonstrates that farming systems sharing a common influential soil management practice at different locations and with different soil types significantly affect the score of some visual soil quality indicators.

KEYWORDS
Fisher's exact test, New Zealand visual soil assessment method, pedoclimatic zones, soil management, soil structure, visual soil quality indicators

1 | INTRODUCTION

The assessment of visual soil quality indicators is often used to characterize soil's ability to function, namely, in an agricultural context, to assess tilth, that is, soil fitness for sowing, seed emergence and plant development. Visual soil quality indicators are, as the name implies, soil features that allow assessing their magnitude (score) based on standards and are correlated with different soil functions that collectively describe the ability of the soil to function. Many visual soil assessment schemes, including observation methods, standards and indices, have been proposed from the early works of Peerlkamp (1959) to the recent developments of this method proposed, for example, by Ball et al. (2007) and Guimarães et al. (2011), or the New Zealand Visual Soil Assessment method by Shepherd (2000). A commonly used visual indicator is observing soil friability, either by hand-manipulating the soil (Ball et al., 2007) or by a drop-shatter test (Shepherd, 2000). Other features that complement friability are soil macroporosity, the existence of compacted subsurface layers, changes in soil colour when compared to undisturbed soil, the presence and colour of mottles on the surface of peds, the existence of earthworm burrows or the number of earthworms in a given volume of soil, only to name a few. The visual soil quality indicators observed (should) reflect soil usage, for example, a pasture or arable soil, because the relative importance of each indicator will vary accordingly. Each method has its approach to summarize the score of each set of visual soil quality indicators observed and classifying them into a single category. One approach is by attributing a score to each indicator observed and applying a weight to the score that purportedly represents each indicator's relative importance with the purpose to calculate a single total score (Shepherd, 2000). Another approach focuses on a few select features that provide a baseline score (e.g. soil friability, soil aggregate features and porosity) and uses other features to refine the score up or down (Ball et al., 2007).

Many authors have reported different correlation coefficients between the ordered categories produced by different visual soil assessment schemes (Mueller et al., 2009; Murphy et al., 2013). One reason may lie in the number of visual soil quality indicators observed within each method and their relative weight on the total soil score. The weight of each indicator should be consistent not only with the ability of the soil to function but also with the set of indicators used to classify the soil health (the score or the total score). The choice of a set of weights is not a trivial problem. Different indicators may explain the same or partially the same phenomena. If we think about the ability of the soil to function as a mathematical equation, a clear parallel could be drawn with the multiple linear regression model, and the set of weights regarded as the coefficients of a set of explanatory variables that allow the best fit (not an arbitrary value or guess).

One common and promising use of visual soil assessment is to compare the effect of different agricultural management practices or different soil conditions on the ability of the soil to function. The current visual soil assessment schemes are useful for an on-farm approach. For which causes of high variability (e.g. soil type, topography and climate) can be deemed, *grosso modo*, to be controlled, that is, held constant, and the effects observed can be attributed to the distinct agricultural management practice or soil condition. However, no extrapolation to other locations can be done. Franco et al. (2019) performed a meta-analysis on 30 papers focusing on the use of Visual Evaluation of Soil Structure (VESS) (Guimarães et al., 2011), covering different climates (from temperate to tropical) and soil types and did not detect differences in the VESS score (Sq) in response to different soil management practices. The lack of differences in this study may be because of different reasons, individually or that concurred to a combined result. As discussed in the previous paragraph, the compound nature of the VESS score, which combines the observation of soil friability, aggregate shape and size, root system features, porosity and signs of anaerobic conditions, without really providing a weight to these attributes, seems to be one of the prime reasons.

Several questions must be answered before a ‘compound score’ is deemed to represent the ability of the soil to function. Basic questions, such as: ‘how does each visual soil quality indicator correlate with soil properties, climate and the different ecosystem services that soils provide?' remain to an extent unanswered. The motivation that a researcher, a technician or a farmer will have to use visual soil assessment schemes relies on the need to observe and compare. When promoting soil management options that purportedly avert a soil health issue, one should be clear on the general effects that they may induce on each visual soil quality indicator, even though these soil management practices may never have been tested locally. In this context, controlled experiments and/or soil surveys should report scores of individual visual soil quality indicators and not 'compound scores'. Two initial questions need answers to assess agricultural management practices that were found to be alternatives for soil health enhancement with visual soil assessment schemes: (i) How sensitive the visual soil quality indicators are to differences induced in the soil by the agricultural management practices? (ii) Are these differences site-specific, that is, are they the result of local interactions or are there detectable effects induced by the agricultural management practices that are not site-specific, although their strength may vary with location? The answers are especially important in regions with undeveloped rural extension services or for individual farmers.

To answer the above questions, we used the visual soil quality indicators of the New Zealand Visual Soil Assessment method (Shepherd, 2000) to assess the changes in the soil structure induced by 11 innovative agricultural management practices (AMPs) implemented across eight pedoclimatic zones. We observed the following visual soil quality indicators: soil friability, porosity, colour, the presence of tillage pan, earthworm counts in a given time, susceptibility to erosion and surface ponding. Additionally, aggregate stability field tests were performed. Each visual soil quality indicator was classified into one of three categories, 'poor', 'moderate' and 'good', according to the standards (Shepherd, 2000). Aggregate stability was also classified into one of three categories, 'poor', 'moderate' and 'good', according to standards (photographs, see Tables S2 and S3 in Appendix S1).

The main objective of this work was to test the null hypothesis that the score of seven visual soil quality indicators of the New Zealand Visual Soil Assessment method and the score of 'soil stability' were independent of 11 different agricultural management practices, used continuously for at least 5 years, irrespective of soil and climate. Additionally, the hypothesis that the distribution of the joint frequencies of the score 'good' for each visual soil quality indicator and agricultural management practice was not more extreme than one would expect if the distribution observed was because of mere chance was tested.

2 | MATERIALS AND METHODS

2.1 | Study sites and agricultural management practices

The data set used was recorded in a survey performed during the spring/summer of 2016 across Europe and China. Eleven AMPs of interest were selected for this study (Table 1). For further information on the AMPs, climate zones and soil types, see Barão et al. (2019) and, for a brief description of the AMPs, see Appendix S1 (Table S1). In addition, a control field or plot consisting of a similar soil sharing the same farming features except for the distinctive management practice was also recorded at each location.

2.2 | Soil properties and climate

The range of the soil properties and climate variables for each AMP is presented in Table 2. All locations were georeferenced. Local climate variables and indices for each location were estimated with the software 'Local Climate Estimator' New Loc_Clim (for more information, see FAO, 2005).

2.3 | Visual soil assessment

The visual soil quality indicators of the New Zealand VSA method (Shepherd, 2000) recorded were 'soil structure and consistency', 'soil porosity', 'the presence of tillage pan', 'soil colour', 'earthworm count', 'surface ponding' and 'susceptibility to wind and water erosion'. Additionally, the aggregate stability in water, given by the slake test, was recorded. At each location, the status of the visual soil quality indicators was assigned into categories 'good', 'moderate' or 'poor', according to the standards of the method (see Shepherd, 2000). The aggregate stability status was also assigned into three categories, 'good', 'moderate' or 'poor'; for a brief description, see Table S2 in Appendix S1.

2.4 | Statistical analysis

The categorical data used in this study have noise. The detection and removal of noise are beyond the scope of this study. The sources of noise are subjectivity and, although unlikely, confirmation bias. When assessing the magnitude (score) of the visual indicators, despite the use of standards, subjectivity may be of particular importance for those visual indicators that are multi-attribute

TABLE 1 Location of the case study sites (country and region, when applicable) and agricultural management practices (AMPs)

NT	MT	M	GM	MUL + PSC	CR	CC	LEG	IPM	IM	CLU
Portugal	Netherlands	Netherlands	Portugal	Portugal	France	Portugal	France	Netherlands	Portugal	Greece
Spain	France	Spain	Hungary	Spain	Portugal	Slovenia	Spain	Portugal	Spain	Slovenia
Greece	Portugal	Slovenia	China Qiyang	Hungary	Spain	Hungary	Slovenia	Spain	Slovenia	Romania
Poland	Spain	Hungary	China Suining	Estonia	Slovenia		Hungary	Slovenia	Romania	
Estonia	Slovenia	Poland		China Suining	Hungary		Romania	Poland	China Qiyang	
China Gong.[a]	Hungary	Estonia		China Gong.[a]	Romania		Poland	China Qiyang	China Gong.	
	Romania	China Qiyang			Poland					

Abbreviations: CLU, Change of land use; CR, Crop rotation; GM, Green manuring; IM, Irrigation management; IPM, Integrated pest management including organic agriculture; LEG, Leguminous crop in the rotation; M, Manuring; MT, minimum tillage; MUL, Mulching + Permanent soil cover; NT, no-till.

[a]China Gong.: Gongzhuling.

and for which there are no unambiguous rules to compound a score, for example, Shepherd's 'susceptibility to wind and water erosion' (Shepherd, 2000). Other instances where subjectivity arises are when the features do not allow a clear classification into one or another category—borderline cases—or in soils for which there are no representative standards. Regarding confirmation bias, a systematic error favouring a particular soil management practice and score of a visual indicator, that is, increasing a particular joint frequency, is implausible because the researchers come from different countries and professional backgrounds, that is, people with different cognitive biases. However, the occurrence of confirmation bias cannot be completely ruled out. Increasing the objectivity of the soil visual indicators, that is, increasing the independence of the score of the visual indicators from individual subjectivity, would allow diminishing the errors present in the data set.

Fisher's exact test was used to test the null hypotheses. The criteria to use the test were the management practice (AMP and control) and the score of each visual soil quality indicator (good, moderate and poor). The level of significance of the analysis was set to $\alpha \leq .05$. Fisher's exact tests for 2×2 contingency tables, a one-tailed test after combining two score categories (moderate + poor), were calculated to test the hypothesis that the management practice produced the frequencies of the scores that are better than one would expect by mere chance. Fisher's exact tests for 2×3 contingency tables (two soil management practices × three scores), two-tailed tests, were also calculated to test if the frequencies are different from chance when both directions are considered (both tails). The decision to present both tests resides in the following: If the analysis seeks solely to test if a treatment (AMP or control) produces better (or worse) scores than expected because of chance, a one-tailed test is adequate, but it will not give any information on what may happen in the other tail of the sampling distribution, in the other region of rejection. If what is happening in the other tail will not affect the adoption of the soil management practice, either because the marginal frequency of that score is too small or deemed irrelevant, a one-tailed test is appropriate. From a scientific point of view, especially when there are three groups (scores), two-tailed tests are important because they allow unveiling differences that would pass unnoticed.

Both Fisher's exact tests were calculated using the function FISHERTEST of the Excel add-in Real Statistics (Real Statistics software, n.d.). All calculations were performed using Excel (Microsoft Office 2016). The data presented in Table S4 of the Appendix S1 allow the reader to check the results.

TABLE 2 Soil properties and climate variables range (minimum, maximum and median values)

		Sand		Silt		Clay		PR		pH		SOM							
		AMP	C	AMP	C	AMP	C	AMP	C	AMP	C	AMP	C	T year	P year	PET year	AI	NPP lim	GCI
NT	Max.	55.6	54.9	72.5	70.6	47.5	52.1	4.1	3.1	8.3	8.3	7.2	6.9	18.7	1014	1507	1.12	1470	80
	Median	33.5	32.0	39.4	36.8	23.6	18.5	2.9	2.2	7.3	7.2	4.8	2.4	16.7	563	1144	0.62	936	21
	Min.	6.7	7.5	24.9	23.7	4.8	5.3	1.1	1.3	5.2	5.1	0.9	0.8	4.9	356	553	0.29	631	10
MT	Max.	91.7	67.7	77.8	69.8	35.0	43.7	5.2	5.5	8.7	8.7	4.0	4.3	17.8	1393	1228	2.24	1470	45
	Median	50.9	43.8	31.4	28.7	16.1	16.7	1.7	1.8	7.1	7.1	2.6	2.2	10.5	665	728	0.96	1071	27
	Min.	9.2	10.2	5.8	19.8	0.0	6.4	0.7	0.5	5.2	5.6	0.6	1.0	4.7	250	510	0.20	459	2
M	Max.	91.5	90.3	79.6	71.2	41.5	38.4	4.4	3.7	8.7	8.7	4.0	5.2	18.2	1393	1228	2.24	1794	80
	Median	52.2	52.7	32.6	29.5	10.0	11.6	2.3	2.2	6.9	6.5	1.9	1.8	9.4	601	725	0.91	981	29
	Min.	9.2	17.9	5.8	9.4	0.0	0.0	0.6	0.5	4.8	5.0	0.6	0.5	4.8	250	504	0.20	459	12
GM	Max.	47.4	62.4	68.7	71.2	39.2	30.9	2.1	1.4	8.4	8.3	2.7	1.3	18.2	1372	1068	1.36	1794	67
	Median	28.2	32.4	33.5	37.4	33.4	23.9	1.5	1.3	7.0	7.1	2.2	1.3	16.5	1014	1060	1.12	1470	39
	Min.	20.7	19.2	23.2	19.2	10.6	9.7	1.0	1.1	5.3	4.7	1.2	1.1	9.2	665	725	0.91	1071	10
MUL + PSC	Max.	80.2	78.0	73.8	71.2	29.4	30.4	3.5	3.1	8.7	8.4	42.8	14.4	17.9	1014	1228	1.4	1470	80
	Median	51.9	54.8	27.4	24.7	13.5	13.1	1.4	1.4	6.7	6.6	2.4	2.0	9.2	665	728	1.0	1071	29
	Min.	15.8	16.7	10.3	8.9	4.4	3.4	0.7	0.6	5.2	5.0	1.2	0.8	4.7	250	504	0.2	459	10
CR	Max.	91.6	80.2	77.9	59.9	37.8	43.1	5.5	5.9	8.4	8.7	4.4	4.3	18.2	1393	1228	2.24	1794	67
	Median	50.9	52.9	32.2	28.7	12.8	14.8	2.2	2.2	6.6	6.8	2.1	1.8	9.5	665	703	0.91	1036	28
	Min.	9.2	16.7	8.3	17.8	0.0	0.0	0.9	0.5	4.7	5.4	0.5	0.3	4.7	250	510	0.20	459	10
CC	Max.	47.4	62.4	60.1	51.4	29.4	18.4	3.5	4.6	7.8	7.4			15.7	1393	1060	2.24	1470	29
	Median	41.7	55.0	46.5	45.0	11.7	15.6	3.0	2.6	7.5	7.0			9.7	1014	728	0.96	1383	29
	Min.	39.9	32.9	23.2	19.2	0.0	0.0	2.1	1.1	6.4	7.0			9.2	665	622	0.91	1071	10
LEG	Max.	80.2	78.0	52.6	57.9	28.4	37.1	5.5	5.6	8.7	8.5	5.4	2.7	17.8	1393	1228	2.24	1383	38
	Median	50.9	50.5	29.8	28.8	16.4	14.8	1.9	2.2	6.7	6.6	2.8	1.8	9.2	665	728	0.91	1071	29
	Min.	39.6	30.0	10.5	8.9	0.0	0.0	0.9	0.6	5.7	5.6	0.9	0.9	7.1	356	622	0.29	631	10
IPM	Max.	89.8	92.7	71.9	72.7	41.5	39.3	4.4	3.5	8.1	8.6	5.1	5.2	18.2	1393	1228	2.24	1794	67
	Median	26.0	31.1	38.7	38.8	9.5	12.3	2.1	2.2	6.3	6.3	2.5	1.8	15.7	914	1001	1.11	1365	27
	Min.	18.6	16.7	9.4	7.3	0.8	0.0	1.1	1.1	5.2	4.2	0.8	0.5	7.3	304	567	0.29	548	2
IM	Max.	79.8	65.3	50.4	64.6	42.5	51.3	5.8	5.4	8.4	8.4	3.7	2.0	18.2	1393	1228	2.24	1794	76
	Median	48.3	39.7	27.9	32.3	21.5	21.4	2.5	2.2	7.4	7.5	2.6	1.8	10.5	477	896	0.53	814	38
	Min.	15.0	4.1	10.8	22.0	0.0	0.0	1.0	1.3	5.4	4.7	0.7	1.1	5.3	356	622	0.29	631	2

(Continues)

TABLE 2 (Continued)

		Sand		Silt		Clay		PR		pH		SOM							
		AMP	C	AMP	C	AMP	C	AMP	C	AMP	C	AMP	C	Tyear	P year	PET year	AI	NPP lim	GCI
CLU	Max.	68.8	63.4	46.7	45.3	56.4	30.1	3.2	4.8	7.9	8.2	8.6	7.5	18.7	1393	1507	2.24	1618	38
	Median	46.1	49.6	31.6	31.3	17.2	18.4	2.6	4.2	7.4	7.6	5.5	3.9	18.7	636	1507	0.52	1025	21
	Min.	13.9	35.4	21.2	18.6	0.5	1.5	2.2	2.4	6.3	6.5	2.4	2.0	9.7	477	622	0.33	814	21

Note: Sand, silt and clay (%); AI = Aridity index = P annual mean/PET annual mean (dimensionless); P year = mean annual precipitation (mm); PET mean = mean annual potential evapotranspiration (mm); pH = inverse logarithm of the proton concentration; PR = penetration resistance (MPa); NPP = net primary production potential, NPP Lim = limiting value, NPP temperature or NPP precipitation (g [DM] m^{-2} $year^{-1}$); SOM = soil organic matter (%); Tyear = mean annual temperature (°C).

Abbreviations: AMP, agricultural management practice; C, control; CLU, Change of land use; CR, Crop rotation; GCI, Gorczynski continentality index; GM, green manuring; IM: irrigation management; IPM: Integrated pest management including organic agriculture; LEG: Leguminous crop in the rotation; M, manuring; MT, minimum tillage; MUL, Mulching + Permanent soil cover; NT: no-till.

3 | RESULTS AND DISCUSSION

3.1 | Results of the Fisher's exact test for 2 × 2 contingency tables

The results in Table 3 showed that the null hypothesis-that the higher frequency of the score 'good' was because of chance- was rejected for 'soil structure and consistency' and 'soil porosity' for five and four AMPs, respectively. For three AMPs, the null hypothesis was rejected simultaneously for both indicators. Furthermore, in the case of no-till (NT), the null hypothesis was also rejected for 'soil stability' and 'susceptibility to wind and water erosion', and, for mulching (MUL + PSC), it was rejected for the 'presence of tillage pan' and 'soil colour'.

3.2 | Results of the Fisher's exact test for 2 × 3 contingency tables

When both directions were considered (two-tailed test, Table 4), a generalized increase of the *p*-value led to the acceptance of the null hypothesis for some visual soil quality indicators and AMPs. The rejection of the null hypothesis for 'soil structure and consistency' and for 'soil porosity' was observed for three AMPs, with the simultaneous rejection of the null hypothesis for both indicators observed only with manuring (M). For no-till (NT), the null hypothesis was accepted for 'soil stability' and 'susceptibility to wind and water erosion', while for mulching (MUL + PSC), the null hypothesis was accepted for 'soil colour'. However, the opposite was also observed, a decrease in the *p*-value leading to the rejection of the null hypothesis, in the case of 'soil colour' and manuring (M). To understand this result, we must analyse the joint frequencies (Table S4 in Appendix S1). The joint frequencies of the score 'good' were 15 and 13 for manuring and control groups, respectively, totalling 28, and in the other direction, the joint frequencies of the score 'poor' were 0 and 6, respectively, totalling 6. The latter is an extreme combination of the score 'poor', and if one collapses the categories 'good' and 'moderate' in a single category and redo the test, the *p*-value is .01 for a one-tailed test, and thus, one would reject the null hypothesis that the frequency of the score 'poor' is independent of the management practice.

3.3 | Detected effects induced in soils by the AMPs and possible causes

A higher frequency of the score 'good' for 'soil stability' was found in no-till AMP than in control. This finding

TABLE 3 One-tailed Fisher's exact test for 2 × 2 contingency tables (categories moderate and poor were collapsed)

	NT	MT	M	GM	MUL + PSC	CR	CC	LEG	IPM	IM	CLU
STR	**.00**	.50	**.00**	.73	**.03**	.20	.78	**.04**	.15	.05	**.01**
POR	**.02**	.40	**.00**	.73	**.00**	.14	.50	.09	**.03**	.10	.08
STA	**.04**	.27	.05	.73	.38	.09	.26	.16	.50	.72	.50
PAN	.10	.50	.07	.50	**.02**	.40	.26	.67	.50	.13	.68
COL	.34	.20	.40	.09	**.03**	.39	.26	.67	.35	.28	.50
EAR	.10	.10	.24	.23	.35	.50	.26	.76	.50	1.00	.50
ERO	**.04**	.28	.38	.23	.62	.61	.22	.50	.50	.72	.15
PON	.24	.27	.37	1.00	.50	.50	1.00	.50	.71	.50	.76

Note: The *p*-values written in bold are statistically significant ($\alpha \leq .05$).

Abbreviations: Col, soil colour; CLU, change of land use; CR, crop rotation; LEG, leguminous crop in the rotation; IM, irrigation management; IPM, integrated pest management including organic agriculture; M, manuring; MT, minimum tillage; GM, green manuring; MUL + PSC, Mulching + Permanent soil cover; NT, no-till. Ear, earthworm count; Ero, susceptibility to wind and water erosion; Pan, presence of a tillage pan; Pon, surface ponding; Por, soil porosity; Sta, soil stability (Slake Test); Str, soil structure.

TABLE 4 Two-tailed Fisher's exact test for 2 × 3 contingency tables

	NT	MT	M	GM	MUL + PSC	CR	CC	LEG	IPM	IM	CLU
	***p*-value**										
STR	**.00**	.83	**.00**	1.00	.11	.48	1.00	.08	.16	.10	**.00**
POR	.05	.16	**.00**	1.00	**.00**	.07	1.00	.17	**.04**	.19	.09
STA	.06	.57	.16	1.00	.22	.21	.52	.08	.86	.40	1.00
PAN	.23	1.00	.15	1.00	**.04**	.24	.52	1.00	.70	.27	1.00
COL	.43	.08	**.04**	.18	.07	.37	.52	.19	.65	.56	1.00
EAR	.27	.31	.32	.29	.60	.78	.52	1.00	1.00	1.00	.71
ERO	.08	.77	.67	.45	1.00	1.00	.44	1.00	1.00	1.00	.30
PON	.48	.55	.54	1.00	.67	1.00	1.00	.59	1.00	1.00	1.00

Note: The *p*-values written in bold are statistically significant ($\alpha \leq .05$).

Abbreviations: CLU, change of land use; CR, crop rotation; IM, irrigation management; IPM, integrated pest management including organic agriculture; LEG, leguminous crop in the rotation; GM, green manuring; MT, minimum tillage; M, manuring; MUL + PSC, mulching + permanent soil cover; NT, no-till. Col, soil colour; Ear, earthworm count; Ero, susceptibility to wind and water erosion; Pan, presence of a tillage pan; Pon, surface ponding; Por, soil porosity; Sta, soil stability (Slake Test); Str, soil structure.

may be explained by Tisdall and Oades (1979) results, which established a relationship between aggregate stability and the size of the population of vesicular-arbuscular mycorrhizal hyphae (hyphal length), that are obligate symbionts with host plants. Thus, the survival of the mycorrhizal propagules (hyphae) when the host plants are not present and/or the ease of infection of host plants can be severely impaired by tillage (see Kabir, 2005 for an extensive review on the subject). No-till also had a higher frequency of the score 'good' for 'susceptibility to wind and water erosion' than control. This result may be explained by the protection provided by the plant residues to the direct impact of rain or irrigation water drops on superficial soil aggregates, protecting the soil from developing a surface crust (allowing higher infiltration rates), soil detachment (diminishing laminar soil erosion) and the preservation of the continuity of macropores from the surface to higher depths (non-biological and biological continuous macropores: earthworms burrows, channels created by decaying roots, cracks etc.) enhancing infiltration (Kairis et al., 2021; Ranaivoson et al., 2017; Zachmann et al., 1987).

Concerning 'soil colour' (Table 4), the frequency of the score 'poor' for manuring (0) was much lower than for control (6), and the *p*-value is associated with this region of rejection (see Discussion at the end of Section 3.2). This result may suggest that the 'soil colour' score was related to increased soil organic matter (SOM) driven by the organic matter amendment. However, no correlation was found between the scores of 'soil colour' and SOM content (Teixeira et al., 2021), which allows speculating that other variable(s) may play a role.

Mulching (MUL + PSC) presented higher frequencies of the score 'good' for both the 'presence of tillage pan' and 'soil colour'. Although correlations have been reported, for example, between mulching and the temperature, and between mulching and the water content of the soil, the thickness of the mulch, the mulch material, local rain and temperature regimes, the development of the crop, all contribute differently to measured soil temperature and water content (Cook et al., 2006; Mulumba & Lal, 2008). The causes of the effects detected in this study on these visual soil quality indicators are not clear.

Concerning 'soil structure and consistency', the statistically significant *p*-value for three AMPs showed that the score is not independent of the management practice (Table 4) and, if one focuses solely on the probability of the AMP performing better than the control (Table 3), the number of AMPs that affect the score rises to 5. The higher scores of 'soil structure and consistency' with some AMPs may be due to different reasons. As discussed above, no-till had a higher frequency of the score 'good' for 'soil stability' than control. One may speculate that a better 'soil stability' will affect the score of 'soil structure and consistency', a drop-shatter test that can be defined as a test to assess the magnitude of soil's friability. Regarding minimum tillage, the *p*-values for the contingency tables of 'soil structure and consistency' were .83 and .50, for the 3×2 and 2×2 contingency tables, respectively, and for 'soil stability', they were .57 and .27, which suggest that soil disruption by minimum tillage creates a soil profile that is not very dissimilar to be observed within the control group (topsoil inversion). Minimum tillage leads to soil aggregates breakdown, higher topsoil aeration and compaction of deeper layers by the multiple passes of tillage implements and the pressure load of heavy machines. Still, lower soil-bearing capacity compared to no-till. Controlled experiments, at different locations, showed that the effect of minimum tillage on aggregate stability, when compared to conventional tillage, is dynamic, varying through the year (Daraghmeh et al., 2009), but others show clear higher aggregate stability under minimum tillage (Kasper et al., 2009). Although no effect was detected on the scores of 'soil structure and consistency', this does not mean that at certain locations, one will not find a minimum tillage effect on aggregate stability (Kasper et al., 2009) and on 'soil structure and consistency'. Interactions of minimum tillage with, for example, climate, soil properties and crops may play an essential role in this indicator's status.

The broad definition of change of land use (CLU), Table S1 in Appendix S1, causes this AMP to encompass practices that have as a single common feature lowering the pressure load affecting the soils (e.g. avoiding livestock trampling or managed fallow). Amidst all the noise that a broad definition brings to the analysis, the detected effect of the AMP on the scores of 'soil structure and consistency' suggests that the change of soil structure because of lower surface soil compaction is the probable cause (Pietola et al., 2005).

Detected effects of manuring on the scores of 'soil structure and consistency' (Tables 3 and 4) may be related to better aggregate stability, although the *p*-values for 'soil stability' were not statistically significant ($p = .05$ and $p = .16$, Tables 3 and 4, respectively). The material used as manure and the amounts applied to soil seems to have different effects on aggregates stability, from no-effect to positive effect (e.g. Roldán et al., 1996). This manuring effect on aggregates' stability may be because of the availability of carbon-rich organic materials causing fast microbial development and/or the development of fungal hyphae (e.g. Abiven et al., 2007; Bertagnoli et al., 2020; Roldán et al., 1996). Regarding green manuring (GM), no effect was detected on the scores of 'soil structure and consistency' and 'soil stability'. Green manuring effects on aggregate stability have been detected in controlled experiments, and they seem to be related to the plant species (or mix) being used (Breland, 1995), but, in some instances, no effect was detected (MacRae & Mehuys, 1987). Kamran et al. (2021) found a higher percentage of aggregates in the range size from 0.25 to 2 mm for treatments with massive amounts of green manuring (22.5–37.5 Mg ha^{-1}) of Chinese milkvetch (*Astragalus sinicus* L.) and mineral fertilization—green manuring alone (15 Mg ha^{-1}) was statistically significantly less effective than mineral fertilization alone for this range of aggregates size. Interactions between green manuring (plant species), total amount incorporated into the soil, climate and soil properties may explain the lack of effects on the visual soil quality indicators in the present study.

The effect of mulching on the score of 'soil structure and consistency' may be because of higher aggregate stability, although the effect on the frequency of the score 'good' of 'soil stability' was not statistically significant, *p*-value = .38 (Table 3). Mulching has been found to positively correlate with aggregate stability and aggregate size (Mulumba & Lal, 2008) and, as previously discussed in the case of no-till, mulch protects surface aggregates from the direct impact of raindrops. Mulching has also been found to attenuate the wetting–drying cycles of the soil (Cook et al., 2006), which in turn diminishes aggregate breakdown, that is, avoids slacking (Le Bissonnais et al., 1989).

Regarding the incorporation of leguminous crops (LEG) in the crop rotation, although the null hypothesis that the scores of 'soil stability' are independent of the management treatments cannot be rejected, it has been found, in controlled experiments, that some leguminous crops not only increase aggregate stability but also fungal hyphal length within the aggregates when compared to non-legumes (Haynes & Beare, 1997).

Regarding 'soil porosity', the null hypothesis was rejected for manuring (M), mulching (MUL+PSC) and integrated pest management (IPM) (Tables 3 and 4) and also, in the case of no-till (NT), for the one-tailed test (Table 3). Observable soil macropores, the voids between aggregates that allow water and air movement and the growth of the plant roots, are a key feature of soil structure. The score of 'soil porosity' might correlate, among others, with aggregates stability, soil tillage intensity, the pressure loads soils withstand and plant species. If one focuses on Table 3, no-till (NT), manuring (M) and mulching (MUL+PSC), the null hypothesis that the frequencies of the score 'good' are because of chance is rejected for 'soil porosity' and simultaneously for 'soil structure and consistency', and thus, one may suggest that the explanatory mechanisms of the effect of the AMPs on 'soil structure and consistency' (see discussion above) may also explain, at least partially, 'soil porosity' status. Differently, for integrated pest management (IPM), the null hypothesis cannot be rejected for 'soil structure and consistency', given the *p*-values of .15 and .16 for Tables 3 and 4, respectively. In contrast, the null hypothesis is rejected for 'soil porosity' with *p*-values of .03 and .04. One may speculate that the detected effect on 'soil porosity' is because of the higher biodiversity and high abundance of organisms that affect the macroporosity of the soils with the AMP treatment (Ferris & Tuomisto, 2015).

3.4 | Future work

The fact that no statistically significant effect of the AMPs on the scores of most visual soil quality indicators was detected does not mean that, at specific locations, because of interactions of the AMPs with, for example, soil properties, climate, crops etc., an effect cannot be detected and even its strength measured. Thus, the hypothesis that soil management, measurable soil properties, climate and possible interactions will affect the scores of the visual soil quality indicators must be tested with a different approach, for example, by studying the relationships (correlations) between the scores and these variables (Teixeira et al., 2021) and assessing the variability of the scores (do not confound with the variability in quantitative variables) that these variables explain. These studies may allow building statistical models to predict the score of different visual soil quality indicators, accounting for the effect of soil management on those scores.

4 | CONCLUSION

The results of this study show that, with few exceptions, 'soil structure and consistency' and 'soil porosity' are the most sensitive visual soil quality indicators among those included in the New Zealand VSA investigated, allowing the detection of the effects induced in soils by the agricultural management practices surveyed irrespective of soil properties and climate.

Soil management effects on the scores of visual soil quality indicators are intertwined with the effects of soil properties, climate and their interactions. The hardship of separating those effects has led many researchers to consider the soils' ability to function, for practical purposes, as a local emergent phenomenon. Thus, site specificity of the effects of soil management has been a hurdle when promoting innovative soil management practices. For the first time, this work allowed us to identify soil management effects of an important number of soil management practices on visual soil quality indicators, independent of soil properties and climate. It also allowed identifying which visual soil quality indicators are likely to be sensitive to those soil management practices. However, the results of this work do not allow us to predict an effect of a given soil management practice on the score of the visual soil quality indicators at a given location.

Another important conclusion is that a soil management effect must be considered when modelling the score of the visual soil quality indicators of the New Zealand VSA method, at least for those indicators and management practices for which the independence hypothesis was rejected. This does not mean that, for the other indicators, soil management effects are inexistent. Taken as the main effects, they might be low, but it does not rule out the existence of important soil management effects in the frame of interactions with soil properties and/or climate variables.

Presently, the compound nature of the visual soil quality indexes does not convey any guidance for adopting soil management practices. Although with a limited scope, this work allows farmers and technicians at different pedoclimatic zones to identify soil management options that may improve the score of those indicators, based on the scores of the visual soil quality indicators of their soils. Further questions need to be answered before the accurate prediction of the score of each visual soil quality indicator at a given location becomes a reality, namely, 'how the score of visual soil quality indicators correlate with different climate variables and soil properties (both inherent and manageable properties)?' and 'which interactions affect the score of the indicators?'. On another note, by answering these and other questions, visual soil assessment methods may become a reliable way to assess soils' ability to function, allowing for assessing or controlling the impact of different policies on soil quality.

ACKNOWLEDGEMENTS

The database used in this study resulted from the iSQAPER research project (Interactive Soil Quality Assessment in Europe and China for Agricultural Productivity and Environmental Resilience), which ended in April 2020. The project was funded by the European Union's Horizon 2020 Program for Research & Innovation, under grant agreement n° 635750, the Chinese Ministry of Science and Technology (grant n° 2016YFE011270), the Chinese Academy of Sciences (grant n° 16146KYSB20150001) and the Swiss State Secretariat for Education, Research and Innovation (contract: 15.0170-1). Special thanks to the case study site teams who performed the field assessment and laboratory measurements. The authors thank the two anonymous referees and editors for the constructive comments and criticism of the previous versions of the manuscript.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the Supplemental Materials of this article.

ORCID

Fernando Teixeira https://orcid.org/0000-0003-1920-3871
Minggang Xu https://orcid.org/0000-0002-7308-0633

REFERENCES

Abiven, S., Menasseri, S., Angers, D. A., & Leterme, P. (2007). Dynamics of aggregate stability and biological binding agents during decomposition of organic materials. *European Journal of Soil Science*, *58*(1), 239–247. https://doi.org/10.1111/j.1365-2389.2006.00833.x

Barão, L., Alaoui, A., Ferreira, C., Basch, G., Schwilch, G., Geissen, V., Sukkel, W., Lesmesle, J., Garcia-Orenes, F., Morugán-Coronado, A., Mataix-Solera, J., Kosmas, C., Glavan, M., Pintar, M., Tóth, B., Hermann, T., Vizitiu, O., Lipiec, J., Reintam, E., … Wang, F. (2019). Assessment of promising agricultural management practices. *Science of the Total Environment*, *649*, 610–619. https://doi.org/10.1016/j.scitotenv.2018.08.257

Ball, B., Batey, T., & Munkholm, L. (2007). Field assessment of soil structural quality – A development of the Peerlkamp test. *Soil Use and Management*, *23*(4), 329–337. https://doi.org/10.1111/j.1475-2743.2007.00102.x

Bertagnoli, B. G. P., Oliveira, J. F., Barbosa, G. M. C., & Colozzi Filho, A. (2020). Poultry litter and liquid swine slurry applications stimulate glomalin, extraradicular mycelium production, and aggregation in soils. *Soil and Tillage Research*, *202*, 104657. https://doi.org/10.1016/j.still.2020.104657

Breland, T. A. (1995). Green manuring with clover and ryegrass catch crops undersown in spring wheat: Effects on soil structure. *Soil Use and Management*, *11*(4), 163–167. https://doi.org/10.1111/j.1475-2743.1995.tb00950.x

Cook, H. F., Valdes, G. S. B., & Lee, H. C. (2006). Mulch effects on rainfall interception, soil physical characteristics and temperature under Zea mays L. *Soil and Tillage Research*, *91*(1–2), 227–235. https://doi.org/10.1016/j.still.2005.12.007

Daraghmeh, O. A., Jensen, J. R., & Petersen, C. T. (2009). Soil structure stability under conventional and reduced tillage in a sandy loam. *Geoderma*, *150*(1–2), 64–71. https://doi.org/10.1016/j.geoderma.2009.01.007

FAO [Food and Agriculture Organization of the United Nations] (2005). Locate climate estimator (New_LocClim). http://www.fao.org/land-water/land/land-governance/land-resources-planning-toolbox/category/details/en/c/1032167/

Ferris, H., & Tuomisto, H. (2015). Unearthing the role of biological diversity in soil health. *Soil Biology and Biochemistry*, *85*, 101–109. https://doi.org/10.1016/j.soilbio.2015.02.037

Franco, H. H. S., Guimarães, R. M. L., Tormena, C. A., Cherubin, M. R., & Favilla, H. S. (2019). Global applications of the visual evaluation of soil structure method: A systematic review and meta-analysis. *Soil and Tillage Research*, *190*, 61–69. https://doi.org/10.1016/j.still.2019.01.002

Guimarães, R., Ball, B., & Tormena, C. (2011). Improvements in the visual evaluation of soil structure. *Soil Use Management*, *27*, 395–403. https://doi.org/10.1111/j.1475-2743.2011.00354.x

Haynes, R. J., & Beare, M. H. (1997). Influence of six crop species on aggregate stability and some labile organic matter fractions. *Soil Biology and Biochemistry*, *29*(11–12), 1647–1653. https://doi.org/10.1016/s0038-0717(97)00078-3

Kabir, Z. (2005). Tillage or no-tillage: Impact on mycorrhizae. *Canadian Journal of Plant Science*, *85*(1), 23–29. https://doi.org/10.4141/p03-160

Kairis, O., Aratzioglou, C., Filis, A., van Mol, M., & Kosmas, C. (2021). The effect of land management practices on soil quality indicators in Crete. *Sustainability*, *13*, 8619. https://doi.org/10.3390/su13158619

Kamran, M., Huang, L., Nieb, J., Geng, M., Lub, Y., Liaob, Y., Zhou, F., & Xu, Y. (2021). Effect of reduced mineral fertilization (NPK) combined with green manure on aggregate stability and soil organic carbon fractions in a fluvo-aquic paddy soil. *Soil and Tillage Research*, *211*, 105005. https://doi.org/10.1016/j.still.2021.105005

Kasper, M., Buchan, G. D., Mentler, A., & Blum, W. E. H. (2009). Influence of soil tillage systems on aggregate stability and the distribution of C and N in different aggregate fractions. *Soil and Tillage Research*, *105*(2), 192–199. https://doi.org/10.1016/j.still.2009.08.002

Le Bissonnais, Y., Bruand, A., & Jamagne, M. (1989). Laboratory experimental study of soil crusting: Relation between aggregate breakdown mechanisms and crust structure. *Catena*, *16*(4–5), 377–392. https://doi.org/10.1016/0341-8162(89)90022-2

MacRae, R. J., & Mehuys, G. R. (1987). Effects of green manuring in rotation with corn on the physical properties of two Quebec soils. *Biological Agriculture & Horticulture*, *4*(4), 257–270. https://doi.org/10.1080/01448765.1987.9755112

Mueller, L., Kay, B. D., Hu, C., Li, Y., Schindler, U., Behrendt, A., Shepherd, G., & Ball, B. (2009). Visual assessment of soil structure: Evaluation of methodologies on sites in Canada, China and Germany. *Soil and Tillage Research*, *103*(1), 178–187. https://doi.org/10.1016/j.still.2008.12.015

Mulumba, L. N., & Lal, R. (2008). Mulching effects on selected soil physical properties. *Soil and Tillage Research*, *98*(1), 106–111. https://doi.org/10.1016/j.still.2007.10.011

Murphy, B. W., Crawford, M. H., Duncan, D. A., McKenzie, D. C., & Koen, T. B. (2013). The use of visual soil assessment schemes to evaluate surface structure in a soil monitoring program. *Soil and Tillage Research*, *127*, 3–12. https://doi.org/10.1016/j.still.2012.06.00

Peerlkamp, P.K. (1959). A visual method of soil structure evaluation. Meded. V.d. Landbouwhogeschool en Opzoekingsstations van deStaat te gent. XXIV No. 24, pp. 216–221

Pietola, L., Horn, R., & Yli-Halla, M. (2005). Effects of trampling by cattle on the hydraulic and mechanical properties of soil. *Soil and Tillage Research*, *82*(1), 99–108. https://doi.org/10.1016/j.still.2004.08.004

Real Statistics software. https://www.real-statistics.com/free-download/

Ranaivoson, L., Naudin, K., Ripoche, A., Affholder, F., Rabeharisoa, L., & Corbeels, M. (2017). Agro-ecological functions of crop residues under conservation agriculture. A review. *Agronomy for Sustainable Development*, *37*(4), 17. https://doi.org/10.1007/s13593-017-0432-z

Roldán, A., Albaladejo, J., & Thornes, J. B. (1996). Aggregate stability changes in a semiarid soil after treatment with different organic amendments. *Arid Soil Research and Rehabilitation*, *10*(2), 139–148. https://doi.org/10.1080/15324989609381428

Shepherd, T. (2000). *Visual soil assessment. Volume 1. Field guide for cropping and pastoral grazing on flat to rolling country*. Horizons.mw & Landcare Research.

Teixeira, F., Basch, G., Alaoui, A., Lemann, T., Wesselink, M., Sukkel, W., Lemesle, J., Ferreira, C., Veiga, A., Garcia-Orenes, F., Morugán-Coronado, A., Mataix-Solera, J., Kosmas, C., Glavan, M., Zoltán, T., Hermann, T., Vizitiu, O., Lipiec, J., Frac, M., … Fleskens, L. (2021). Manuring effects on visual soil quality indicators and soil organic matter content in different pedoclimatic zones in Europe and China. *Soil &Tillage Research*, *212*, 105033. https://doi.org/10.1016/j.still.2021.105033

Tisdall, J. M., & Oades, J. M. (1979). Stabilization of soil aggregates by the root systems of ryegrass. *Australian Journal of Soil Research*, *17*(3), 429. https://doi.org/10.1071/sr9790429

Zachmann, J. E., Linden, D. R., & Clapp, C. E. (1987). Macroporous infiltration and redistribution as affected by earthworms, tillage, and residue. *Soil Science Society of America Journal*, *51*(6), 1580–1586. https://doi.org/10.2136/sssaj1987.03615995005100060032

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Teixeira, F., Lemann, T., Ferreira, C., Glavan, M., Zoltán, T., Hermann, T., Lipiec, J., Frąc, M., Reintam, E., Xu, M., Fan, H., & Ritsema, C. (2022). Evidence of non-site-specific agricultural management effects on the score of visual soil quality indicators. *Soil Use and Management*, *00*, 1–11. https://doi.org/10.1111/sum.12827