Geoderma 376 (2020) 114545

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Eliciting experts' tacit models for the interpretation of soil information, an example from the evaluation of potential benefits from conservation agriculture

L.M. Chabala[a,1], J.G. Chimungu[b], R.M. Lark[c,*], F. Mtambanengwe[d], P.C. Nalivata[b], E. Phiri[a], G.M. Sakala[e]

[a] University of Zambia, School of Agricultural Sciences, Great East Road Campus, Lusaka, Zambia
[b] Lilongwe University of Agriculture and Natural Resources, Crop and Soil Sciences Department, Bunda Campus, Lilongwe, Malawi
[c] University of Nottingham, School of Biosciences, Sutton Bonington, Nottinghamshire LE12 5RD, UK
[d] University of Zimbabwe, Department of Soil Science and Agricultural Engineering, Mount Pleasant, Harare, Zimbabwe
[e] Zambia Agriculture Research Institute, Mt Makulu Central Research Station, Chilanga, Lusaka, Zambia

ARTICLE INFO

Keywords:
Elicitation
Probabilistic inversion
Conservation agriculture
Soil surveys

ABSTRACT

We examined a procedure to elicit the tacit models underlying expert opinions on environmental factors that affect the absolute yield benefits expected from the adoption of conservation agriculture (CA) practices in southern Africa. The procedure is based on expert evaluation of the expected improvement in crop yield on adoption of CA in a particular scenario or 'state', a state being a specified set of soil conditions captured by a standard soil profile description from a specified agroecological zone (AEZ) of Zambia. Mixed groups of scientists including soil scientists, agronomists, agricultural economists and other environmental scientists, facilitated by experienced senior researchers, were presented with multiple subsets each of three states, and asked to rank the states in each subset with respect to expected yield improvement under CA. The groups of scientists could be divided into two sets. Each set comprised two groups, and the agreement on ranking between groups within each set was larger than would be expected if the ranking were done at random. For both sets of groups the ranking could be modelled with respect to properties of the soil, and the contrast between AEZ. The models revealed two contrasting groups of conceptual assumptions. One group broadly expected larger absolute yield improvements from conservation agriculture in settings where water is most likely to be limiting and the carbon status of the soil is poor. By contrast, the other group expected larger improvements where water was less likely to be limiting. These contrasting views are relevant to current discussions as to whether conservation agriculture, which is promoted as a 'climate smart' strategy for cropping, is sufficiently attractive for smallholder producers in conditions where crop production is already challenging, and whether the potential benefits in areas where water availability is not of itself a common limitation should be considered. The elicited models could be translated directly into competing hypotheses to be tested, perhaps in on-farm trials of conservation agriculture practices over contrasting soils in the different AEZ. The method, based on modelling the ranking process, could be of more general interest for the elicitation of expert opinion about complex soil, crop and environmental systems.

1. Introduction

Conservation agriculture (CA) has been promoted as a strategy to improve food security in sub-Saharan Africa in the face of climate change (Thierfelder et al., 2017). Conservation agriculture entails zero or minimum tillage to reduce soil disturbance, the maintenance of soil cover by mulching, commonly with retained crop residues, and the diversification of cropping systems by intercropping or the use of rotations (Kassam et al., 2009). Uptake of CA in southern Africa has been variable in its success (Kassam et al., 2015), and it has been suggested (Giller et al., 2009) that CA has been over-promoted, as a panacea, at least for smallholder producers. In particular Giller et al. (2009) suggest

* Corresponding author.
E-mail address: murray.lark@nottingham.ac.uk (R.M. Lark).
[1] All authors contributed equally and are named in alphabetical order

https://doi.org/10.1016/j.geoderma.2020.114545
Received 18 December 2019; Received in revised form 1 May 2020; Accepted 15 June 2020

that greater attention should be paid to the physical and the socio-economic settings in which CA is most likely to deliver benefits for farmers who adopt it. They refer to the concept of the socio-ecological niche for particular farm practices (Ojiem et al., 2006): the set of conditions, physical, social and economic, which constrain the situations in which a practice is beneficial for farmers. Giller et al. (2009) suggest that, among other factors, soil conditions may limit the potential benefits from CA as an intervention, and so play a part in defining its particular niche. There have been attempts to identify conditions in which CA can be expected to be beneficial (e.g. Baudron et al., 2015). However, while there are several experimental trials incorporating CA practices in southern Africa (e.g. Ligowe et al., 2017), there is at present limited empirical evidence to support a robust definition of the soil and related factors which constrain the socio-ecological niche of CA. In this paper we explore the possibility of using expert opinion, extracted through formal elicitation, as a basis for a provisional identification of soil factors expected to influence the response to CA.

In soil science and land evaluation, we must often make an assessment of a site (e.g. its suitability for a particular crop) which, while in principle deducible from its biophysical and chemical properties, cannot be inferred in practice because of our partial understanding and the sparsity of available data. In such a situation an expert may base a judgement on their understanding and experience, effectively applying a tacit model of the system in which different factors are weighted intuitively. The value of such expert judgement for management of natural systems is widely recognized (e.g. McBride and Burgman, 2012; Stanton et al., 2018). In this paper we examine how expert opinion might be used to assess the expected benefits from the uptake of CA based on site information which might be inferred from a soil map. The approach which we examined entails the framing of elicitation questions in terms of ordinal measurement or ranking (e.g. Craig et al., 2009) which requires the subjects to rank small subsets of cases with respect to some property or possible outcome. These methods have their origin in psychology, specifically in Thurstone (1927) laws of comparative judgement which imply that one can infer implicit values of the psychophysical stimulus induced by different cases (e.g. the brightness of individual objects) from pairwise rankings of these objects. Methods to elicit expert opinion, or opinion of other special groups such as patients, from ranking of cases have been widely used in medicine and health economics (e.g. Craig et al., 2009; Ali and Ronaldson, 2012).

Neslo and Cooke (2011) present a promising approach to model individuals' tacit evaluations of states of a system from ranked subsets. This approach is based on the assumption that an individual making a judgement (Neslo and Cooke, 2011 call such an individual a 'stakeholder', we call them an 'expert') has a 'utility' for every state of the system, and that these utilities are reflected in the rankings of all subsets of states which the expert is asked to assess. Utility might therefore relate to a benefit or risk associated with a particular state of the system. Neslo and Cooke (2011) model the utilities of states as random quantities using methods called probabilistic inversion, either inferring different utilities for each state considered, or inferring the utility of any state as a function of a covariate set. In the latter case the estimated model could then be used to compute utilities for new cases where the values of the covariates are known. We summarize the method of Neslo and Cooke (2011) in the following theory section.

Neslo and Cooke (2011) presented an exemplar case study in which the utility of different health states for individuals were evaluated by a panel each of whom ranked subsets of states. The utility of particular states was modelled as a function of properties of the health state (e.g. individual mobility, or cognitive functioning). The approach was used by Flari et al. (2011) to model the risk associated with nanotechnology-enabled food products, by Halpern et al. (2013) to evaluate conservation management options for marine environments, to prioritize different exotic diseases which are potential threats to the Australian pig industry (Brookes et al., 2014a,b) and to identify features of emerging infectious diseases which make them threats to the safety of blood transfusions (Neslo et al., 2017).

In the study reported in this paper we undertook an elicitation exercise with teams of scientists engaged in research into CA in southern Africa. This was a desk-based study undertaken during a project workshop, and was based on evaluation of soil information presented in a standard format, rather than examination of soil profiles in the field. The states evaluated in this study were soils from Zambia, drawn from two of the country's three agroecological zones (AEZ) following Veldkamp et al. (1984). Each soil had a standardized profile description, based on legacy soil surveys undertaken in the country, and the particular AEZ was also identified. Participants were asked to rank subsets of these states with respect to the expected absolute improvement in crop yield expected after CA had been adopted, and allowed to establish. The approach of Neslo and Cooke (2011) was used to model the latent utility of the different cases with respect to soil properties and the agroecological zone. The utility is a variable, reflected in the ranking of states on the expected yield improvement expected for a particular soil in a specified AEZ on the introduction of CA practices. The objective was to identify the soil factors that experts regarded as important in making this evaluation, and to identify any conflicting assessments of the suitability of particular soils in agroecological zones for the promotion of CA. This would provide a provisional basis for identifying the most promising conditions for CA or, in the event of conflicting views, to identify the key competing hypotheses about the environmental limitations which constrain, along with other factors, the socio-ecological niche of CA implicit in expert opinion. This would help to inform the research agenda for future experimental work.

2. Theory

In this section we summarize the methodology presented by Neslo and Cooke (2011). We consider a number of possible 'states of affairs' or states. In our study a 'state' is defined by a soil profile description, standardized as described below, augmented where appropriate by observations on specific physical limitations, and set in a specific agroecological zone. There are N states in total in the analysis which we represent by $a_1, a_2, \ldots, a_N$. We adapt Neslo and Cooke (2011) terminology here, and assume that each expert has a tacit *latent ranking variable* (*lrv*) for each state, which they reflect in their rankings of subsets of the N states. The *lrv* of the jth state according to the kth expert out of L is treated as a random variable and denoted by $v_k(a_j)$. The *lrv* determines the predicted ranking of two states in a subset, thus if

$$v_k(a_j) < v_k(a_i)$$

then we expect state a_j to appear before state a_i in any ranking by the k^{th} expert of a subset which contains both states. In this study we have preferred to use an intuitive ranking in which the state expected to have the largest benefit from CA is ranked first. The value of the *lrv* for this state is therefore relatively small. For this reason we do not follow Neslo and Cooke (2011) in calling the *lrv* a utility, because in conventional usage a utility would be largest in the state with the largest expected benefit from CA. We therefore define the utility of a state $u_k(a_j)$ by

$$u_k(a_j) = -v_k(a_j), \tag{1}$$

we compute values of the *lrv*, and present them as utilities (e.g. in plots) by a change of sign. It follows, then, that the utility of a state in our model is a variable such that if $u_k(a_j) > u_k(a_i)$ it follows that the k^{th} expert expects a larger yield benefit from the adoption of CA in a situation represented by state a_j (with respect to climatic and soil conditions) than is expected in a situation represented by state a_i.

In the methodology of Neslo and Cooke (2011) we may estimate a value of the *lrv* for each state considered by an expert, such that an ordering of the states on the *lrv* values best approximates that expert's ordering of the states in each subset. However, one may also compute the *lrv* as a function of properties of the states. For simplicity consider a single property of the state, a covariate which takes value c_j for the jth state. This might be a particular soil property in our case. The *lrv* of the jth state according to the kth expert is modelled as a function of c_j:

$$v_k(a_j) = \alpha_k c_j, \tag{2}$$

where α_k is a coefficient, estimated empirically, for predicting the value of the *lrv* from values of the variable c. The value of the coefficient is found such that a ranking of the subsets of states based on the predicted *lrv* values most closely matches the actual ranking made by the experts. This approach based on covariates is of considerable practical use because it allows us to predict a value of the *lrv* for an unobserved state, h, from the value of the covariate for that state, c_h such that the predicted value of the *lrv* is $\alpha_k c_h$. If the potential benefit of CA practices at a site increase with increasing values of the covariate, c, we therefore expect α_k to have a negative value, indicating that the larger the value of c for a state the earlier we expect it to appear in rankings of subsets. We next consider in more detail how the estimation of expert *lrv* values, or appropriate coefficients to compute these from properties of the states, is done.

We consider P overlapping subsets of the N states, each subset comprising n of the states. For example, one might consider subsets each of three states: $S_1 = \{a_1, a_8, a_4\}$, $S_2 = \{a_2, a_3, a_4\}$, $S_3 = \{a_7, a_3, a_5\}$. Note that one state can appear in more than one subset. For every subset we consider R possible responses from the elicited expert. As the response is an ordering of the states in the subset, there are $R = n!$ possible responses – distinct orderings of the states.

We consider a $P \times R$ matrix $\mathbf{T}_k$ which tabulates the kth expert's responses for P ordered subsets. One element of the jth row of $\mathbf{T}_k$ is 1, the remaining elements in the row are zero. The column which contains the 1 corresponds to the recorded response for the jth subset. One may average the $\mathbf{T}$ matrices over a set of L responses to obtain the matrix $\mathbf{Q}_{\mathscr{R}}$:

$$\mathbf{Q}_{\mathscr{R}} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{T}_l. \tag{3}$$

The aim of the PI method is to find a set of values of the coefficients α_k, $k = 1, \ldots, L$ which match as closely as possible the mean responses recorded in $\mathbf{Q}_{\mathscr{R}}$. This is done by a stochastic model under which the coefficients are treated as random variables. One generates a (single) realization of a simple random model for these values, and then finds a set of projection weights which project from this sampled random variable onto the distribution of the α_k such that the difference between the $\mathbf{Q}$ matrix implied by the values of the utilities computed from the projected coefficients and the observed $\mathbf{Q}_{\mathscr{R}}$ is minimized. Two methods are considered to find the projection weights. These are the IPF (Iterative Proportional Fitting) algorithm (Fienberg, 1970) and the PARFUM (Parameter Fitting for Uncertain Models) algorithm (Du et al., 2006). The former always converges for 'feasible' $\mathbf{Q}_{\mathscr{R}}$, i.e. one where the results are entirely consistent with a particular set of real-valued *lrv* (i.e. orderings in overlapping subsets are consistent), and converges to a result with desirable properties (the mutual information of the random sample and the projected set of *lrv*s is minimized). Feasibility means, for example, that if state a_l is ranked above state a_m in one subset then it is ranked above that state in any other subset where both occur. However, experts might not always achieve this, and may rank two similar states inconsistently when they are compared with very different third states in two subsets. The PARFUM algorithm converges in all cases, and may be invoked when IPF fails to converge.

2.1. Algorithms

Both algorithms start with a set of m realizations of a random variable (in the case illustrated here where there is just one coefficient in the model for each expert), $\mathbf{b}_j$, $j = 1, \ldots, m$, which is proposed for the values of the coefficient α_k proposed for the k^{th} expert. For the lth realization one may compute an induced response matrix, $\mathbf{T}_l^{\mathrm{I}}$, and from all m of these a corresponding matrix $\mathbf{Q}^{\mathrm{I}}$ with the mean responses:

$$\mathbf{Q}^{\mathrm{I}} = \frac{1}{m}\sum_{l=1}^{m} \mathbf{T}_l^{\mathrm{I}}. \tag{4}$$

The objective of both algorithms is to obtain an $\mathbf{m} \times 1$ vector of projection weights, $\boldsymbol{\phi}$, which minimize the difference in some sense between the observed mean responses in $\mathbf{Q}_{\mathscr{R}}$ and the values

$$\mathbf{Q}^{\phi} = \sum_{l=1}^{m} \boldsymbol{\phi}[l]\mathbf{T}_l^{\mathrm{I}}. \tag{5}$$

The initial set of projection weights can be set to $\boldsymbol{\phi}^0 = \left[\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}\right]^{\mathrm{T}}$. These weights are then adjusted iteratively until convergence. Once the algorithm has converged one may then examine the projection weights and, for example, compute an estimate of the mean of the coefficient α_k as the weighted sum of the corresponding m random values in $\mathbf{b}_j[k]$, $j = 1, \ldots, m$ weighted by the final set of projection weights. The details of the IPF and PARFUM algorithms are described in the Appendix.

2.2. This study

As noted above, Neslo and Cooke (2011) use probabilistic inversion methods to fit models in which each individual expert's *lrv* for a state is predicted as a function of covariates. One simple linear form of the model, comparable to a simple linear regression, is given in Eq. (2) above. Such a model allows one to predict the utility for that expert of some new state, given values of the covariate. This model can be generalized to include more than one covariate, and to include categorical predictors. In this study we evaluated how far the rankings participants gave to states (soil within agroecological zones), with respect to absolute benefits obtained from the adoption of conservation agriculture, could be modelled as functions of properties of those states presented to the participants in the elicitation.

3. Methods

3.1. Definition of states

3.1.1. Agroecological zones

We used the classification of Zambia into broad natural regions due to Veldkamp et al. (1984). We follow Saasa (2003) in referring to these as agroecological zones (AEZ), of which there are three. AEZ II is sometimes subdivided into two subregions, but in this study we ignored this distinction.

AEZ I comprises the major valleys (e.g. Luangwa, Gwembe and Lusemfwa valleys) and southern parts of the Western and Southern provinces of Zambia. It is characterized by small amounts of rainfall (less than 800 mm a^{-1}) and a medium to high risk of drought. The growing season is typically 80–129 days in duration. In contrast, AEZ II, comprising the sandveld plateau of the Central, Eastern Lusaka and Southern provinces and the Kalahari sand plateau and Zambezi flood plains of Western Province has a medium to low drought risk with 800–1000 mm a^{-1} rainfall. The duration of the growing season is typically 100–140 days. AEZ III, in the north of the country, has much more rainfall (more than 1000 mm a^{-1}). Because the drought risk in AEZ III is very small, we did not consider any states situated in AEZ III.

The AEZ, either I or II, therefore entered into the definition of all states in this study.

3.1.2. Soil profiles

A state in this study consists of a description of a soil profile set in a specific AEZ (I or II). We obtained six profile descriptions for states set in AEZ I and eight for states set in AEZ II. Soil information provided was consistent across all states. We provided a profile description truncated at 1 m depth, with the depth of successive horizons provided and, for each horizon, a textural class following the system used in Zambia (Diestel, 1981). A Munsell colour code was given for each horizon and the available water content, pH (measured in water) and organic matter content (%). In addition we provided the Total Available Moisture (TAM: available water in mm over the top metre of soil), and, along with the TAM we named the TAM category as used by Bunyolo et al. (1982):— Very high (TAM > 155), High (115 <TAM⩽ 155), Moderate (75 <TAM⩽ 115), Low (35 <TAM⩽ 75), Very low (<35).

Muliokela (1995) lists general soil types characteristic of the AEZ of Zambia, but these are not named according to any particular soil classification. We used this list as a guide when assembling soil profile descriptions from legacy sources. Our aim was not to provide specific inferences about particular soils of interest, but rather to cover a range of soil conditions so that expert utilities could be modelled as functions of profile properties.

In addition to ensuring that we covered a range of soil conditions, it was necessary to provide the full list of soil properties listed in the first paragraph of this section. We were able to obtain descriptions for 11 soil profiles (out of a total of 14) from two sources, a compendium of information on soil physical properties in Zambia including soil profile descriptions and physical measurements (Maclean, 1970), and pedon descriptions from the excursion guide for the XIth International Forum on Soil Technology and Agrotechnology Transfer held in Zambia in 1985 (Woode, 1985). The extraction of information was fairly straightforward. Where soil analytical results were presented for fixed depth intervals alongside horizon-based descriptions, analytical data were attributed to the horizon from the depth interval that occupied most of the horizon. Soil organic matter (SOM) was recorded for each horizon, and where the original source provided soil organic carbon (SOC) SOM was obtained by the simple conversion SOM = SOC × 1.72. Although the proportions of organic carbon in SOM are known to vary, this factor has been widely used in the calculation of SOC from direct measurements of SOM, and so is appropriate for our purposes here (Landon, 1991).

In three states (Soils 1, 2 and 3) it was necessary to form compound profile descriptions by combining information from two sources where descriptions, incomplete with respect to the properties we required, corresponded to soils in the same AEZ, with comparable textures. These soils had to be included so as to give reasonable coverage of soil conditions expected in AEZ I. The sources were profile descriptions from a soil survey (Commissaris, 1973) and from the soil map of Zambia produced by Brammer (1976). Available water for the horizons was obtained from general values given for textural classes by Landon (1991), Table 6.12, inferring stone content from the profile descriptions. Details on the sources for each soil used in the elicitation are provided in the Supplementary Material to this paper (Table S1).

In addition to the information on each soil profile listed above, we provided a soil class name. This was based on the source of the description, and we made no attempt to identify a soil class in a single classification scheme. For most of the soil profiles described a classification was provided according to the Soil Map of Africa classification of D'Hoore (1964) — Lithosol, Vertisol, Ferruginous soil, Ferralitic soil, Paraferralitic soil — or Soil Taxonomy of Soil Survey Soil Survey Staff (1975) — Ultisol, Alfisol and Oxisol. Where one of these classifications was not used we provided a texture-based short description (e.g 'Loamy sand over gravelly sandy clay loam'). When the source material allowed some qualification or expansion of the simple class description this was included (e.g. 'Shallow Lithosol over saprolite').

In some states the original profile description included specific comments on physical limitations of the soil under cultivation (e.g. noting that it is prone to capping). Where these occurred they were added to the description of the state.

Each state was described on a 'soil card' which was headed with the AEZ and soil class, along with all the information by horizon listed in this section. In addition a schematic representation of the profile was printed on the card, showing the horizons by depth scale. The cards were produced using the graphics capabilities of theR platform (R Core Team, 2017), and the aqp package for R (Beaudette et al., 2013) was used to colour the horizons on the profile schematic according to the Munsell colour codes. The soil cards are presented in the Supplementary Information.

The soil descriptions used in the elicitation were circulated among all facilitators in advance of the elicitation itself (see below) and examined by them, including three senior and experienced Zambian soil scientists.

3.2. Preparation and execution

3.2.1. Materials, expert panels and facilitation

Having described a set of 14 states we then created 45 overlapping subsets, each of three states. This number was selected so that each state could appear, on average, in three subsets. Subsets were assembled manually, examining plots of soil properties in order to avoid 'dominated' subsets in which one soil clearly had better properties over all considered (e.g. larger TAM, pH and soil organic matter). At the same time we attempted to define subsets so as to span a range of soil conditions in each. Some of the subsets comprised states drawn from a single AEZ, and some subsets included states drawn from both the AEZs considered in this study. The list of subsets is provided in the Supplementary Information (Table S2).

The ranking of each subset was to be done, not by individuals but by six small groups each with 5 or 6 members and a facilitator. The facilitators were senior experienced scientists with expertise in soil science with agronomic applications. All had significant experience of research at post-doctoral level in sub-Saharan Africa, including interactions with national extension services and leading roles on farmer-oriented research and development programmes funded by agencies including the World Bank, the European Union, the Southern African Development Community, the United States Agency for International Development and the United Kingdom Department for International Development. Three of the facilitators were from Zambia, two from Malawi and one from Zimbabwe. All were Co-Investigators on the CEPHaS project (Strengthening Capacity in Environmental Physics, Hydrology and Statistics for Conservation Agriculture Research, see Acknowledgements) and are joint authors with the corresponding author of this paper. The groups consisted of CEPHaS project team members attending the project's Third Network Meeting in Lusaka (July 2019). All had expertise in soil science, agronomy, geophysics or hydrogeology, and the CEPHaS project is concerned with deploying this expertise to understand impacts of CA on the water cycle. The groups were set up to ensure that expertise at post-doctoral level in soil science and agronomy was evenly distributed between groups, and that all groups had similar numbers of Zambian participants with soil science and agricultural education at graduate level. Other team members were from the UK, Zimbabwe or Malawi. All groups included members with experience at the interface of research and extension or of on-farm participatory research in agronomy or soil science. Due to absence of some of the project team members, the number of groups was reduced from 6 to 5 on the day.

The general objectives of the elicitation exercise and the approach were discussed in person by the corresponding author and the Zambia-based facilitator team, and by email with members of the team based outside Zambia. The facilitator team were not familiar with the PI approach at this stage, and the specific modelling method was not discussed in the group so that the facilitators would not, consciously, approach the group task in terms of constructing state rankings from some function of soil properties.

Briefing notes were prepared in advance of the meeting, and shared and edited by the full facilitator team, before being shared with all meeting participants a week before the elicitation. The briefing notes (see Supplementary Material), set out the general objective of the elicitation, and the question that the team would be asked to consider when ranking states in each subset (see Section 3.2.2 below). The briefing notes also showed example soil cards, and an example of how a particular ranking would be recorded on the response sheets. The briefing notes also provided short summary notes on the soil classes used, the definition of the TAM classes according to Bunyolo et al. (1982) and the soil texture triangle used (Diestel, 1981).

3.2.2. Groupwork

Two and a half hours were allocated for the exercise. Prior to the groupwork, the briefing notes were presented as a reminder, and an opportunity was given for questions to be raised. The groups then went to separate locations at the meeting venue, where, under the guidance of the facilitator, they were reminded of the task. They considered each subset of states in turn, and considered the following instruction:

> *In each case you should assume that the soil is currently under conventional cultivation (ridges) for rainfed maize by a smallholder producer. The producer adopts conservation agriculture (zero till, retention of residues, and crop rotation). The question is: what absolute improvement in yield can he or she expect, averaged over varying seasons, from the third season after introduction onward?*

Participants were then reminded that the task was to rank the states in each subset from the one with the largest absolute increase in yield expected (or smallest decrease), to the one with the smallest absolute increase (or largest decrease). As the previous sentence makes clear, it was not assumed that CA would always result in an increased yield, a yield reduction is possible. By 'absolute' yield increase is meant a change in units of tonnes per hectare, as opposed to a proportional increase (proportion of yield prior to adoption of CA). The following comments in the task were included in the briefing notes:

Think about the following when making your rankings.

i. *We are interested in the yield improvement (from season three to allow for adjustment). In one particular case, where conditions are already very good for smallholder rainfed production, and drought risk is smaller, the absolute improvement may be smaller than for a case you are comparing it with where water availability is limiting. The best soil might therefore not automatically be highest-ranked, we are interested in which soil has the greatest potential improvement.*

ii. Information on cation exchange and nutrients such as P and K is not provided (as it was not available for most cases). You may regard clay content (texture group), pH and OM content as giving you, potentially, proxy information on soil fertility.

iii. Consider how the soil might respond to cessation of tillage (risk of capping) given its texture and the possibility of structural limitations given texture.

iv. Consider how soil pH might limit responses of the soil biota to improved conditions under CA, as well as crop response.

Each group was provided with printed copies of the briefing notes. Each group also had a set of the soil cards, printed on stiff card and with each state on an individual card. This allowed the groups each to extract the cards for each subset in turn, and to place these on a table or the floor in the middle of the group to facilitate discussion. Each group recorded its agreed ranking for each subset on a results sheet which was returned at the end of the exercise.

3.3. Analysis

3.3.1. Exploratory analysis

Five groups, denoted A, B, C, D, E, participated in the elicitation. One group (A) did not complete the task, this was for external reasons (the facilitator had to leave for part of the assigned period) and its results are not considered further for purposes of modelling the ranking process. In exploratory analysis of the results from the remaining groups we computed a matrix to show pair-wise comparisons between groups with respect to the proportion of subsets for which they agreed on the ranking of all states. Under a null hypothesis in which the groups rank states at random in each subset, the expected proportion of subsets for which two groups would agree completely is $3!/3!^2 = 0.1\dot{6}$. We computed the 95% confidence interval for the proportion of agreed rankings for each pair of groups using the method of Blaker (2000) as implemented in the blakerci function in the PropCIs library (Scherer, 2018) for the R platform.

3.3.2. Modelling with PI algorithms

As reported below, there were larger levels of agreement on ranking of states between groups than would be expected under random ranking, but this was not true for every pair of groups. On the basis of the level of agreement, groups B and C, and groups D and E could be put together, the latter pair agreeing completely on the ranking of 19 of the 45 subsets, and the former two agreeing completely on the ranking of 18 of the subsets. For purposes of modelling with probabilistic inversion we combined these two pairs of groups and extracted the responses for the subsets on which they were in complete agreement on ranking.

The PI method was used to fit linear models for the *lrv* of each state. The model can be expressed in general form as

$$\mathbf{v} = \mathbf{X}\boldsymbol{\alpha}, \tag{6}$$

where $\mathbf{v}$ denotes a vector of fitted *lrv* for the N states, the matrix $\mathbf{X}$ is a $N \times P$ design matrix which includes, in its P columns the values of the covariates for the N states (e.g the values of soil pH in one column), and $\boldsymbol{\alpha}$ is a vector of length P which contains the model coefficients. The PI method finds, by means of the IPF or PARFUM algorithms, values of the coefficients in $\boldsymbol{\alpha}$ which give the closest matching between the observed rankings of states over all subsets and the rankings implied by the values of the *lrv* computed with Eq. (6).

The first step was to compute simple group *lrv* values for each state within each group, B-C or D-E. This is equivalent to the general model in Eq. (6) where the design matrix has N columns, one corresponding to each state, and where the terms on the main diagonal are 1 and all others are zero. In this case the elements of $\boldsymbol{\alpha}$ are direct estimates of the *lrv* for each state. The model does not allow us to compute the *lrv* for a new, unobserved, state. It was found, for both groups, that the IPF algorithm would not converge, so PARFUM was used. The next step was to form predictive functions of the properties associated with each state, with environmental variables appearing as predictors in the design matrix $\mathbf{X}$.

The first model we considered used a dummy variable, x_{AEZ}, as the covariate in $\mathbf{X}$, which took the value 0 for all states in AEZ I and 1 for all states in AEZ II. Under this model the corresponding coefficient in $\boldsymbol{\alpha}$ is the *lrv* for all states in AEZ II relative to the *lrv* of zero for all states in AEZ I. Note that, if states in AEZ I are generally expected to have a larger yield benefit from CA than states in AEZ II, then the *lrv* of AEZ I is the smaller (the utility is the larger) and so the coefficient in $\boldsymbol{\alpha}$ will have a positive value.

It is necessary to assess the evidence that a particular covariate is informative about the expert *lrv* for states in the examined subsets. To do this we used a permutation method by which the sets of of covariate values for the states were reallocated independently and at random to state labels, and a value of $\alpha_c$was re-estimated. This was repeated 1 000 times, and the interval bounded by the 2.5th and 97.5th percentiles of the 1 000 estimates of the model coefficients in $\boldsymbol{\alpha}$ were treated as a 95% confidence interval for the coefficient under a null hypothesis in which the variables in the design matrix $\mathbf{X}$ have no relationship to the expert *lrv* values for the states. Note that this permutation approach is effectively a random shuffling of the rows of the design matrix $\mathbf{X}$. The values of the covariates taken by a particular state are not separated from each other, and so correlations between the covariates are preserved.

If the permutation test described above gave a confidence interval for the coefficient which excluded the estimated value, then AEZ was retained in the model. The next step was to consider whether including TAM improved the model fit. We fitted a model with $P = 3$ columns in the design matrix, the first contained the AEZ dummy variable, the second contained values of TAM for all states in AEZ I, and the third contained values of TAM for all states in AEZ II. This is equivalent to a linear model with separate slopes for the regression line for two levels of a factor. The permutation method was applied to decide whether to retain TAM as a predictor for each AEZ (it could be dropped for one but retained for another). This process was then repeated for each of the following covariates in turn: OMm (mean organic matter content over the profile computed by weighting the value for each horizon by its thickness), pHm (mean soil pH over the profile, computed in the same way as OMm), and a dummy variable CMvF, the contrast between coarse and medium textures soils (top horizon) and fine ones — in this instance taking value 0 for Sand and Loamy Sand, and 1 for all other texture classes. Each of these covariates was tested in turn in a model with the dummy variable for AEZ included, and tested by the permutation method. After this was complete a model was fitted with all predictors selected in this way, and a permutation test was run to select those to be retained in the final model.

4. Results

All four groups B, C, D and E completed the exercise. The proportion of subsets on which each pair of expert groups agreed on all ranking positions is shown in Table 1a, along with the 95% confidence interval for this estimate. The closest agreement is between groups D and E, and the next closest between groups B and C. There is also evidence of agreement between groups B and D (the confidence interval for the proportion of subsets on which there is agreement excludes the value $0.1\dot{6}$), but the confidence intervals for the proportion of subsets with complete agreement for pairs {C,D} and {C,E} include $0.1\dot{6}$, the expected proportion under random ranking. For pair {B,E} the lower confidence bound for the proportion is not distinguished from $0.1\dot{6}$ at two places of decimals. Further modelling was undertaken using the 19 subsets on which groups D and E agreed, and their common rankings of these, and the 18 subsets on which groups B and C agreed.

Notes taken by observers showed that the time taken to complete a ranking decreased as the exercise progressed. Groups spent between 7 and 14 min ranking the first subset, but between 1 and 2 min to rank the 30th. This might be expected as the process becomes more familiar, and also, having discussed the ranking of one particular state in one subset, its ranking in others may be more quickly agreed. To test whether time pressure or fatigue might result in more erratic ranking later in the exercise we present in Table 1b a contingency table showing the numbers of subsets in which groups B and C were in complete agreement (row 1) and those on which they disagreed (row 2) from among (column 1) the first 22 subsets and (column 2) subsets 23–45. We present the corresponding contingency table for agreements between groups D and E in Table 1c. Note that there is no evidence of a reduced rate of agreement between the groups in the second group of subsets. A loglikelihood ratio test of the null hypothesis of random association between the rate of agreement and the group of subsets was conducted for each contingency table with the loglm command in the MASS library for the R platform (Venables and Ripley, 2002). There was no evidence to reject this null hypothesis in either case ($p = 0.9$ for groups B and C; $p = 0.44$ for groups D and E).

Note that Group A completed rankings for 21 subsets, these are not analysed further, except to note that the proportion of these on which their rankings matched those of groups B, C and D were, respectively, and with 95% confidence intervals shown in brackets, 0.57 [0.35,0.77], 0.38 [0.20,0.60], 0.29 [0.13,0.51] and 0.22 [0.15,0.55].

In Table 2 are shown the coefficients for models fitted to the common responses of groups D and E. When AEZ only was considered as a predictor of rankings (Model 1) the coefficient fell outwith the 95% confidence interval for values under the null hypothesis obtained by the permutation method. Note that the coefficient is negative, which implies that the group expect larger absolute yield improvement under CA in AEZ II than AEZ I. In model 2 the coefficients for TAM both fall outside the 95% interval for the null hypothesis, and have opposite signs. The signs imply that expected yield benefits are larger on soils

Table 1a
Proportion of 45 subsets in which groups agreed completely on ranking of states, with the 95% confidence interval for these estimates shown in square brackets.

	B	C	D	E
B	1.00	0.40 [0.26,0.55]	0.38 [0.24,0.53]	0.29 [0.17,0.44]
C		1.00	0.16 [0.07,0.29]	0.11 [0.04,0.24]
D			1.00	0.42 [0.29,0.57]

Table 1b
Numbers of subsets in which group B and C are in complete agreement on ranking among subsets 1–22 and 23–45.

	Subsets 1–22	Subsets 23–45
Complete agreement	9	9
Disagreement	13	14

Table 1c
Numbers of subsets in which group D and E are in complete agreement on ranking among subsets 1–22 and 23–45.

	Subsets 1–22	Subsets 23–45
Complete agreement	8	11
Disagreement	14	12

Table 2
Models fitted to responses from groups D and E for the 19 subsets for which they were in full agreement*.

Model	Predictor	Estimated coefficient*	95% bounds†
1	AEZ	−0.187	−0.185,0.187
2	AEZ	−0.031	−0.100,0.084
	TAM(AEZ I)	0.113	−0.098,0.087
	TAM(AEZ II)	−0.161	−0.097,0.085
3	AEZ	−0.391	−0.100,0.078
	OMm (AEZ I)	0.004	−0.102,0.080
	OMm (AEZ II)	−0.085	−0.103,0.081
4	AEZ	−0.211	−0.103,0.081
	pHm (AEZ I)	0.078	−0.104,0.083
	pHm (AEZ II)	−0.108	−0.100,0.079
5	AEZ	−0.363	−0.116,0.108
	CMvF (AEZ I)	0.010	−0.119,0.103
	CMvF (AEZ II)	0.024	−0.117,0.106
6	AEZ	0.043	−0.082,0.070
	TAM(AEZ I)	0.260	−0.081,0.071
	TAM(AEZ II)	−0.159	−0.084,0.069
	pHm (AEZ II)	0.347	−0.082,0.069

*Note that these model coefficients produce values of the *lrv* (negative utilities) as they predict a ranking variable in which the state with the largest expected yield benefit is ranked 1.

† From 1000 permutations.

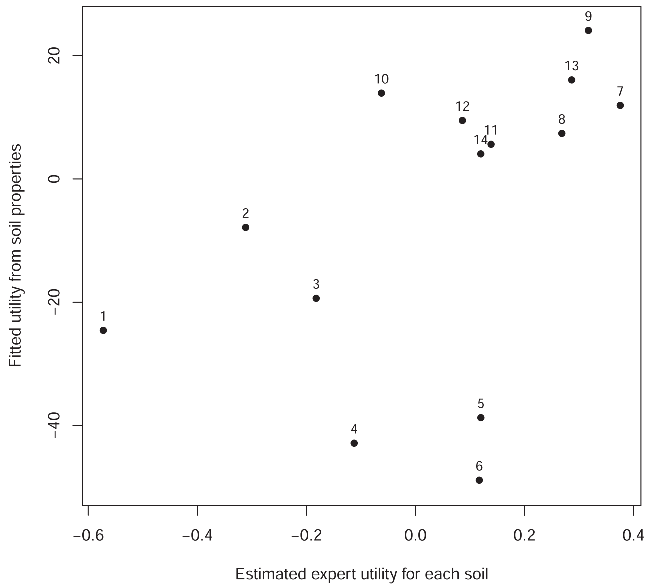

Fig. 1. Plot of (abscissa) estimated expert utilities for each state (negative *lrv* value) against the predicted utilities (negative fitted *lrv* values, ordinate) based on the selected function of soil properties, group D and E using subsets for which they were in full agreement.

with a smaller available water content in AEZ I, but conversely, will be larger on soils with a larger available water content in AEZ II. The results for models 3 and 5 show that the coefficients for mean profile organic matter content and for the dummy variable distinguishing soils with coarse or medium-textured topsoil from fine-textured topsoil fall within the 95% interval obtained by permutation, so there is no evidence that they account for the group's rankings. However, there is evidence (model 4) that the group expect larger absolute yield improvement on soils of smaller pH in AEZ II, but no pH effect in AEZ I.

When all the covariates which appeared informative in models 1 to 5 were combined in a single model there was evidence that all contribute to the group's ranking, although there is a change of sign in the coefficient for pH, which may result from correlations between the soil properties of the states (model 6). Within AEZ II pH and TAM are positively correlated. It therefore appears that, with TAM included in the model, there is evidence that group D-E expects larger benefits from CA on more acid soils.

A plot (Fig. 1) of the estimated group utilities for each state against those obtained from model 6 in Table 2, (computed in each case from the fitted *lrv* values following Eq. (1)), shows that these are generally linearly related with the exception of states 4, 5 and 6. It is notable that these are three states out of the four in total (the other is with state 14), for which notes on the soil card recorded that the soil is at risk of structural instability (puddling or capping). Note that states 14, 5 and 6 have very similar estimated expert utilities. This implies that, while the soil properties used for modelling the expert rankings imply a small benefit from CA for these soils, in AEZ I, the panel appears to interpret these soils as being likely to benefit rather more from CA interventions than others in AEZ I (states 1,2 and 3), possibly because CA specifically protects soils against direct impact of rainfall that might cause puddling, and the cultivations that might cause capping.

Table 3 presents model coefficients for the common responses of groups B and C. Again, AEZ considered as a single covariate appears to be predictive of the group rankings. However, the coefficient is positive here, implying that group B-C expect a larger benefit from CA in AEZ I than in AEZ II. This is contrary to the result for groups D-E (Table 2). When TAM is considered as a predictor the coeffient for states in AEZ I falls within the range obtained with random permutation, but there is evidence that TAM is predictive of rankings for sites in AEZ II, the

Table 3
Models fitted to responses from groups B and C for the 18 subsets for which they were in full agreement*.

Model	Predictor	Estimated coefficient*	95% bounds[†]
1	AEZ	0.136	−0.111,0.112
2	AEZ	−0.023	−0.091,0.081
	TAM(AEZ I)	−0.038	−0.090,0.079
	TAM(AEZ II)	0.119	−0.088,0.080
3	AEZ	0.100	−0.086,0.077
	OMm (AEZ I)	0.237	−0.086,0.080
	OMm (AEZ II)	0.276	−0.085,0.078
4	AEZ	0.017	−0.078,0.084
	pHm (AEZ I)	0.280	−0.079,0.086
	pHm (AEZ II)	0.372	−0.078,0.087
5	AEZ	0.172	−0.073,0.063
	CMvF (AEZ I)	0.011	−0.069,0.067
	CMvF (AEZ II)	0.166	−0.070,0.065
6	AEZ	0.348	−0.052,0.051
	TAM(AEZ II)	0.001	−0.053,0.052
	pHm (AEZ I)	0.319	−0.054,0.052
	pHm (AEZ II)	0.526	−0.052,0.052
	OMm (AEZ I)	0.614	−0.054,0.052
	OMm (AEZ II)	0.230	−0.053,0.053
	CMvF (AEZ II)	−0.038	−0.054,0.053
7	AEZ	0.147	−0.062,0.068
	pHm (AEZ I)	0.359	−0.063,0.068
	pHm (AEZ II)	0.557	−0.063,0.064
	OMm (AEZ I)	0.590	−0.063,0.069
	OMm (AEZ II)	0.184	−0.065,0.065

*Note that these model coefficients produce values of the *lrv* (negative utilities) as they predict a ranking variable in which the state with the largest expected yield benefit is ranked 1.

[†] From 1000 permutations.

positive sign implying that states where the soil has a large TAM will have smaller benefits from CA. There are positive coefficients all outwith the 95% intervals obtained by permutation for mean profile organic matter and for mean profile pH. The values of these coefficients do not differ markedly between AEZ I and AEZ II, and imply that larger benefits from CA are expected by this group in states with more acid soils and smaller organic carbon content. The contrast between states with coarse or medium and fine-textured topsoil appears to be predictive of ranking in AEZ II, with smaller benefits expected for the finer-textured soils. All covariates which appeared to be predictive in the single covariate models (in the separate AEZ), were included in a single model. In this case some of the coefficients fell within the 95% interval obtained by random permutation, so were dropped, leaving model 7. Under this model there are differences between AEZ and effects of profile mean pH and OM. The signs are unchanged from the earlier models.

A plot of the estimated group utilities and the fitted utilities from model 7 in Table 3, (again obtained by a change of sign from *lrv* values) shows (Fig. 2) a linear relationship, with states 2 and 8 as outliers, both given smaller utilities by the group than were predicted from the fitted model. It is notable that state 2 is a shallow soil over saprolite, something not expressed in the covariates used for modelling, but which might have affected the assessment made by group B-C of the site's potential.

5. Discussion

The model results presented in Tables 2 and 3 indicate that, for the subsets in which they concurred, soil properties in the different states, along with AEZ, were predictive of group rankings as cofficients for some covariates fell outwith the 95% range obtained by permutation. It was notable that, while the fitted utilities and those estimated for the

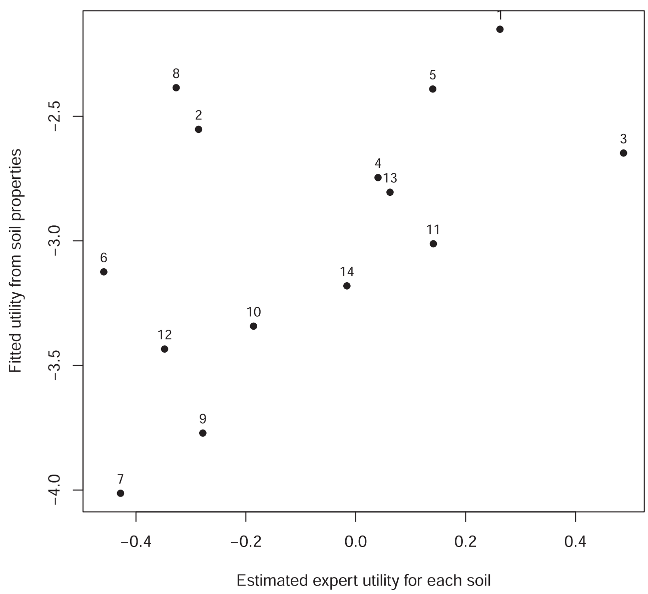

Fig. 2. Plot of (abscissa) estimated expert utilities for each state (negative *lrv* value) against the predicted utilities (negative fitted *lrv* values, ordinate) based on the selected function of soil properties, group B and C using subsets for which they were in full agreement.

states directly were mostly consistent, differences could be explained by particular features of the soil included in descriptive notes, in particular the states where the soils have potential structural problems in the case of group D-E.

A marked feature of these results is that, in so far as groups B and C concurred in their assessment of subsets, they ranked them differently to groups D and E in their consistent rankings. Broadly the models estimated by PI indicate that group D-E expected larger yield benefits from CA states in which conditions for rain-fed production would be most favourable (AEZ II with a smaller drought risk, and larger improvements for states with larger available water contents in AEZ II). In contrast, group B-C expected larger benefits from CA in the more drought-prone AEZ I, and on soils within AEZ II with smaller available water content (TAM, model 2 in Table 3). Similarly, they expect larger benefits from CA on soils with smaller organic matter content. The two groups concur, in the final joint models (Table 2, model 6; Table 3, model 7) in expecting larger benefits from CA on more acid soils. It is important to remember that the task given the groups was to rank on expected absolute improvement in yield under CA, while group D-E expected larger yield benefits from CA in less droughty conditions and soils with larger available water content but this does not mean they do not expect CA to be advantageous in conditions more challenging for rain-fed production.

Whilst participants from Zambia were distributed across all groups, it is possibly significant that the facilitators for groups B and C (in general expecting larger benefits from adoption of CA in AEZ I and on soils with less available water and organic matter), were both from Zambian institutions. The facilitators of groups D and E were from Zimbabwean and Malawian institutions. Furthermore, the facilitator of group A was also from a Zambian institution and it is notable that the rankings for this group, in the subsets that they completed, were more frequently in agreement with groups B and C (with a confidence interval for the estimated proportion exceeding $0.1\dot{6}$) than with groups D and E (with a confidence interval for the estimated proportion including $0.1\dot{6}$). It is possible that this reflects common outlooks on CA and its potential in a Zambian setting, within Zambian institutions.

The fundamental difference between groups B-C and D-E is whether states in which factors such as water availability are expected to be most challenging for crop production would have the largest absolute benefit from CA adoption, or whether (without excluding the expectation that CA will be beneficial in these circumstances), larger benefits are to be expected where conditions are more favourable for production in general. It is interesting to note that the Zambian Agriculture Research Institute (ZARI) is currently examining whether there are benefits from CA adoption in AEZ III (Siulemba, pers. comm.), where the drought risk is regarded as 'almost nil' (Saasa, 2003).

It is important to remember, of course, that we have here been modelling expert opinion rather than experimental or observational data. That there should be contrasting opinions is no surprise, particularly given the considerations in the previous paragraph. One value of this approach is that it allows us to probe differing expert expectations in ways that should allow the establishment of hypotheses for experimental testing. For example, both models of expert opinion in this study expect there to be differences between AEZ with respect to benefits from CA. Group D–E, in their model 6 (Table 2) appear to hypothesize that there would be an interaction between AEZ and available water in the profile, with larger benefits from CA over soils with less available water in AEZ I, and larger benefits over soils with more available water in AEZ II. Both groups appear to concur in expecting larger benefits from CA over acid soils, and group B-C appear to expect larger benefits from CA adoption over soils with smaller organic matter content.

On this basis one might design a network of on-farm trials, to test some of these implicit hypotheses. For example, the hypotheses may be stated as follows (where:

H1. The effects of CA on absolute yield improvement differ between AEZ I and AEZ II, with (following groups B and C) a larger benefit on AEZ I

H2. There is an effect of soil TAM on yield improvement on adoption of CA.

H3. There is an interaction between soil TAM and AEZ (I or II), with larger yield improvements on soils with smaller TAM in AEZ I, and on soils with larger TAM in AEZ II (following groups D and E).

To test these hypotheses one requires a factorial design in which soils with varying TAM are used in the contrasting AEZ. One might start with a map of soil TAM inferred from the SoilGrids data (Hengl et al., 2017) by means of pedotransfer functions applied to the basic soil property data. Following Bunyolo et al. (1982) one would then identify all locations in AEZ I and in AEZ II where expected TAM is low or very low, and those where it is high or very high. One could then select equal numbers of sites from the low-very low and high-very high TAM sites in each of AEZ I ad AEZ II, aiming for sample balance with respect to other soil properties such as pH by means of the the `cube` algorithm of Deville and Tillé (2004) as implemented in the `BalancedSampling` library for the R platform (R Core Team, 2017; Grafström and Lisic, 2016). It would then be necessary to visit the sites on the ground, both to ensure that the predicted values of the key soil properties are reasonably reliable and to obtain informed consent from a farmer at each site to participate in the trial. At each site a field under conventional cultivation would be selected, and a set of adjacent plots would be established, one to be maintained under conventional production and the other for conversion to one or more CA strategies, perhaps following the proposal of Rusinamhodzi et al. (2011) to employ a factorial design to examine the joint effects of tillage practice, mulching and rotation. Because of the factorial design at site-level it would then be possible to analyse these data by an analysis of variance (reflecting the nested structure of any sub-plots within each site) to test hypotheses H1, H2 and H3 above by examining, respectively, the main effect of AEZ, the main effect of soil TAM (high or very high versus low or very low) and the interaction of these two factors. Interactions with any factorial design implemented at within-site level (e.g. factorial effects of tillage, mulching and rotation), could also be examined.

It is likely that the diverging views that our elicitation uncovered among experienced scientists engaged with CA research in Southern Africa reflects, in part, how the scientific literature on CA in the region increasingly highlights the complexity of CA systems, and the need for nuance in understanding and promoting them (Giller et al., 2009). It is known that CA can have marked effects on crop yields where rainfall is unreliable and soil fertility is poor (e.g. Thierfelder et al., 2015), and studies have shown improved water infiltration under CA (e.g. Thierfelder and Wall, 2012). However, Rusinamhodzi et al. (2011), in a meta-analysis of CA studies in Southern Africa, found that while yields improved under CA on well-drained soils, mulching could reduce yields (due to waterlogging) under heavy rainfall. Benefits of CA were found to be strongly dependent on the use of appropriate crop rotations, good weed control and adequate plant nutrition. If nutrient supplies are non-limiting then rainfall remains the main yield determinant, and no treatments can offset the effects of extremes of drought or flooding. Rusinamhodzi et al. (2011) recommended that the promotion of CA required better targeting to those conditions where it had the greatest potential. They also suggested the need for a network of long-term field experiments to improve understanding of how these environmental factors interact with aspects of the CA system. We suggest that the approach outlined in this paper could be used, as exemplified in the previous paragraph, to focus and prioritize the environmental factors that such a network of experiments incorporates into its design. The meta-analysis of Rusinamhodzi et al. (2011) also emphasizes the importance of factors such as weed control and access to fertilisers. We suggest that this elicitation process could be extended to consider such management factors, and broader socio-economic drivers of the success or failure of CA, and other agronomic interventions proposed to improve cropping resilience under climate change.

This approach to dealing with expert opinion contrasts with approaches used in formal elicitation practice. The approach of Cooke (1991) starts with a process of 'calibrating' experts on test problems, the results being used to weight the divergent opinions from a panel when forming an elicited output. In 'behavioural elicitation' (Reagan-Cirincione, 1994; O'Hagan et al., 2006) the objective, in so far as possible, is to arrive at a consensus view after individual opinions have been independently elicited and then shared, through a facilitated process of discussion. While both these approaches are pragmatic when the objective is an expert view to be used for a practical task, the case study reported here is a reminder that experts, with considerable experience, can have very divergent views. The PI modelling approach brings these into focus by starting with relatively simple ranking tasks, and then looking for underlying explanatory factors of the ranking. Where the objective is setting scientific priorities and developing hypotheses this may be more fruitful than the pursuit of consensus, or a weighting approach that conceals, at least to some extent, divergent views.

This is, to the best of our knowledge, the first example of the application of this ranking and modelling approach to problems in soil science and agronomy. We suggest that it could be of wider interest, for example to illuminate the factors that experts bring to bear in judgements about soil quality, which could be relevant to the long-standing challenge of defining appropriate indicators for soil monitoring (Ritz et al., 2009).

6. Conclusions

The process of ranking subsets of states, defined on soil and environmental properties, followed by modelling the latent utilities hypothesized to underly the rankings, is a promising method by which to extract the tacit model that experts deploy when making assessments. In this study we have found contrasting views on the situations in which the absolute yield improvements from the adoption of conservation agriculture will be largest. This is consistent with, and may well reflect, a growing awareness in the literature of the complexity of the factors that influence yield responses to the adoption of CA. The different modelled latent utilities of CA can be used to set up competing hypotheses, which could be tested experimentally in experiments, perhaps on-farm, at locations selected to span the range of conditions with respect to soil properties of interest (e.g. available water) in contrasting agroecological zones.

The modelling procedure allowed us to make use of categorical and continuous soil information from legacy sources. It was also possible, by direct estimation of expert utilities, and comparison with utilities predicted from covariates, to highlight specific features of particular soils (here the resilience of soil structure, particularly under cultivation) which experts treated as significant.

While the use of expert opinion is essential in tackling many complex problems related to soil management and crop production, this study is a reminder that experts can have conflicting views, and we suggest that this approach to modelling how those opinions are used in the ranking task may be more informative than elicitation methods which aim to produce a single outcome as a consensus or weighted judgement.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by UK Research and Innovation (UKRI) [Grant No. NE/P02095X/1] through the GROW programme of the Global Challenges Research Fund. The project is entitled *CEPHaS - Strengthening Capacity in Environmental Physics, Hydrology and Statistics for Conservation Agriculture Research*. We acknowledge the contribution of all members of the CEPHaS project team who participated in this study at the project's Third Network Meeting in July 2019, Lusaka. We are grateful to two anonymous reviewers, and to the associate editor who handled this paper for their careful reading and helpful suggestions.

Appendix A. Appendix A. The IPF and PARFUM algorithms

A.1 The IPF *algorithm*. At the start of the hth iteration of the algorithm the current projection matrix, $\boldsymbol{\phi}^h$, is set equal to $\boldsymbol{\phi}^{h-1}$. Within the hth iteration we cycle through the $p = 1, 2, \ldots, P$ subsets in turn. At the beginning of each of these P cycles we update the mean response matrix

$$\mathbf{Q}^{\phi}_{h,p} = \sum_{l=1}^{m} \boldsymbol{\phi}^h[l]\mathbf{T}^{\mathrm{I}}_l, \tag{7}$$

where $\boldsymbol{\phi}^h$is the current version of the projection matrix.

Within each of the P subsets we cycle through the $l = 1, 2, \ldots, m$ realizations of the randomly-generated latent variables. If the response for the pth subset in the jth realization is the rth out of the R responses for that subset then we compute

$$\phi_j = \boldsymbol{\phi}^h[j]\frac{\mathbf{Q}_{\mathscr{R}}[p, r]}{\mathbf{Q}^{\boldsymbol{\phi}}_{l,p}[p, r]}, \tag{8}$$

and then substitute this value for the jth element of $\boldsymbol{\phi}^h$.

When the projection matrix $\boldsymbol{\phi}^h$ has been updated over all realizations within the current subset we move to the $(p + 1)$ th subset. Once this has been done over all subsets the final mean response matrix is computed as

$$\mathbf{Q}^{\boldsymbol{\phi}}_h = \sum_{l=1}^{m} \boldsymbol{\phi}^h[l]\mathbf{T}^{\mathrm{I}}_l. \tag{9}$$

The algorithm is deemed to have converged if $\mathbf{Q}^{\boldsymbol{\phi}}_l = \mathbf{Q}_{\mathscr{R}}$ within some degree of tolerance.

A. 2 The PARFUM *algorithm*. Within the hth iteration we cycle through the $p = 1, 2, ..., P$ subsets in turn. In the PARFUM algorithm the projection matrix $\boldsymbol{\phi}^h$ is not updated continuously but once in every iteration. A separate projection matrix is computed for each realization $\mathbf{a}^L_j$, and the resulting matrices are averaged over all realizations. We retain the separate projection vectors for the realizations within the hth iteration in the $m \times P$ matrix $\boldsymbol{\Phi}_h$. Within each of the P subsets we cycle through the $l = 1, 2, ..., m$ realizations of the randomly-generated latent variables. If the response for the pth subset in the jth realization is the rth out of the R responses for that subset then we compute a value for the projection matrix specific to this subset and realization:

$$\boldsymbol{\Phi}_h[j, p] = \boldsymbol{\phi}^{h-1}[j]\frac{\mathbf{Q}_{\mathscr{R}}[p, r]}{\mathbf{Q}^{\boldsymbol{\phi}}_{h-1}[p, r]}. \tag{10}$$

When the projection matrix has been computed for all realizations within the current subset we move to the $p + 1$th subset. Once this has been done over all subsets the updated projection vector is computed as the average of projection vectors computed for each subset:

$$\boldsymbol{\phi}^h = \frac{1}{P}\boldsymbol{\Phi}_h\mathbf{1}_m, \tag{11}$$

where $\mathbf{1}_m$ is a vector length m of ones, and then the final mean response matrix is computed as

$$\mathbf{Q}^{\boldsymbol{\phi}}_h = \sum_{l=1}^{m} \boldsymbol{\phi}^h[l]\mathbf{T}^{\mathrm{I}}_l. \tag{12}$$

The algorithm is deemed to have converged if $\mathbf{Q}^{\boldsymbol{\phi}}_h \approx \mathbf{Q}_{\mathscr{R}}$ within some degree of tolerance.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.geoderma.2020.114545.

References

Ali, S., Ronaldson, S., 2012. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. British Medical Bulletin 23, 21–44.

Baudron, F., Thierfelder, C., Nyagumbo, I., Gérard, B., 2015. Where to target conservation agriculture for African smallholders? How to Overcome Challenges Associated with its Implementation? Experience from Eastern and Southern Africa. Environments 2, 338–357.

Beaudette, D.E., Roudier, P., O'Geen, A.T., 2013. Algorithms for Quantitative Pedology: A Toolkit for Soil Scientists. Computers & Geosciences. 52, 258–268.

Blaker, H., 2000. Confidence curves and improved exact confidence intervals for discrete distributions. Canadian Journal of Statistics 28, 783–798.

Brammer, H., 1976. Soils of Zambia. Memoir and map at 1: 2,500,000. Department of Agriculture (Land Use Branch), Ministry of Rural Development, Lusaka.

Brookes, V.J., Hernández-Jover, M., Cowled, B., Holyoake, P.K., Ward, M.P., 2014a. Building a picture: prioritization of exotic diseases for the pig industry in Australia using multi-criteria decision analysis. Preventive Veterinary Medicine 113, 103–117.

Brookes, V.J., Hernández-Jover, M., Neslo, R., Cowled, B., Holyoake, P.K., Ward, M.P., 2014b. Identifying and measuring stakeholder preferences for disease prioritization: A case study of the pig industry in Australia. Preventive Veterinary Medicine 113, 118–131.

Bunyolo, A.M., Commissaris, A., Hennemann, G.R., 1982. Rating of land qualities. Soil moisture and climatic conditions and their conversion into land suitability in Zambia. Soil Survey Unit, Land Use Branch Soil Bulletin No. 5. Department of Agriculture, Mt Makulu, Chilanga.

Commissaris, A., 1973. Semi-detailed soil survey Sasare Block, Proposed refugee settlement area. Land Use Services Division Soil Survey Report No 12. Ministry of Rural Development, Mt Makulu, Chisanga.

Cooke, R.M., 1991. Experts in uncertainty: opinion and subjective probability in science. Oxford University Press, Oxford.

Core Team, R., 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria http://www.R-project.org/.

Craig, B.M., Busschbach, J.J.V., Salomon, J.A., 2009. Keep it simple: Ranking health states yields values similar to cardinal measurement approaches. Journal of Clinical Epidemiology 62, 296–305.

Deville, J.C., Tillé, Y., 2004. Efficient balanced sampling: the cube method. Biometrika 91, 893–912.

D'Hoore, J.L., 1964. Soil Map of Africa Scale 1 to 5000000. Commission for Technical Cooperation in Africa Publication No 93. CCTA, Lagos.

Diestel, H., 1981. Water retention values of Zambian soils synthesized from other soil physical data. Soil Survey Unit Bulletin No. 7. Department of Agriculture, Mt. Makulu.

Du, C., Kurowicka, D., Cooke, R.M., 2006. Techniques for generic probabilistic inversion. Computational Statistics and Data Analysis 50, 1164–1187.

Fienberg, S.E., 1970. An iterative procedure for estimation in contingency tables. Annals of Mathematical Statistics 41, 907–917.

Flari, V., Chaudhry, Q., Neslo, R., Cooke, R., 2011. Expert judgment based multi-criteria decision model to address uncertainties in risk assessment of nanotechnology-enabled food products. Journal of Nanoparticle Research 13, 1813–1831.

Giller, K.E., Witter, E., Corbeels, M., Tittonell, P., 2009. Conservation agriculture and smallholder farming in Africa: The heretics view. Field Crops Research 114, 23–34.

Grafström, A., Lisic, J. 2016. BalancedSampling: Balanced and Spatially Balanced Sampling. R package version 1.5.2.

Halpern, B.S., Longo, C., McLeod, K.L., Cooke, R., Fischoff, B., Samhouri, J.F., Scarborough, C., 2013. Elicited preferences for components of ocean health in the California Current. Marine Policy 42, 68–73.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLOS ONE 12, 1–40.

Kassam, A., Friedrich, T., Shaxson, F., Pretty, J., 2009. The spread of conservation agriculture: justification, sustainability and uptake. International Journal of Agricultural Sustainability 7, 292–320.

Kassam, A., Friedrich, T., Derpsch, R., Kienzle, J., 2015. Overview of the worldwide spread of conservation agriculture. Field actions science reports. The Journal of Field Actions 8, 1–10.

Landon, J.R., 1991. Booker Tropical Soil Manual: A Handbook for Soil Survey and Agricultural Land Evaluation in the Tropics and Subtropics. Routledge, London.

Ligowe, I.S., Nalivata, P.C., Njoloma, J., Makumba, W., Thierfelder, C., 2017. Medium-term effects of conservation agriculture on soil quality. African Journal of Agricultural Research 12, 2412–2420.

Maclean, A.H., 1970. Moisture characteristics of some Zambian soils. Soil Science Research Section Soil Survey Report No. 6. Department of Agriculture, Mt Makulu.

McBride, M.F., Burgman, M.A., 2012. What is expert knowledge, how is such knowledge gathered and how do we use it to address questions in landscape ecology? In: Perera, A.H., Drew, C.A., Johnson, C.J. (Eds.), Expert Knowledge and Its Applications in Landscape Ecology. Springer, New York.

Muliokela, S.W., 1995. Zambia Seed Technology Handbook. Ministry of Agriculture, Food and Fisheries, Zambia.

Neslo, R.E.J., Cooke, R.M., 2011. Modeling and validating stakeholder preferences with probabilistic inversion. Applied Stochastic Models in Business and Industry 27, 115–130.

Neslo, R.E.J., Oei, W., Janssen, M.P., 2017. Insight into 'calculated risk': an application to the prioritization of emerging infectious diseases for blood transfusion safety. Risk Analysis 37, 1783–1795.

O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., Rakow, T., 2006. Uncertain Judgements: Eliciting Expert Probabilities. John Wiley, Chichester.

Ojiem, J., Ridder, N., Vanlauwe, B., Giller, K.E., 2006. Socio-ecological niche: a conceptual framework for integration of legumes in smallholder farming systems. International Journal of Agricultural Sustainability 4, 79–93.

Reagan-Cirincione, P., 1994. Improving the accuracy of group judgement: a process intervention combining group facilitation, social judgement analysis and information technology. Organizational Behavior and Human Decision Processes 58, 246–270.

Ritz, K., Black, H.I.J., Campbell, C.D., Harris, J.A., Wood, C., 2009. Selecting biological indicators for monitoring soils: A framework for balancing scientific and technical opinion to assist policy development. Ecological Indicators 9, 1212–1221.

Rusinamhodzi, L., Corbeels, M., van Wijk, M.T., Rufino, M.C., Nyamangara, J., Giller, K.E., 2011. A meta-analysis of long-term effects of conservation agriculture on maize grain yield under rain-fed conditions. Agronomy for Sustainable Development 31, 657–673.

Saasa, O.S., 2003. Agricultural intensification in Zambia, the role of policies and policy processes. Lund University and the Institute of Economic and Social Research, University of Zambia https://www.keg.lu.se/en/sites/keg.lu.se.en/files/a9.pdf.

Scherer, R. 2018. PropCIs: Various Confidence Interval Methods for Proportions. R package version 0.3-0. https://CRAN.R-project.org/package=PropCIs.

Soil Survey Staff, 1975. Soil Taxonomy, a basic system of soil classification for making and interpreting soil surveys. Soil Conservation Service, U.S. Department of Agriculture, Agriculture Handbook No. 436.

Stanton, C.Y., Mach, K.J., Turner, P.A., Lalonde, S.J., Sanchez, D.L., Field, C.B., 2018. Managing cropland and rangeland for climate mitigation: an expert elicitation on soil carbon in California. Climatic Change 147, 633–646.

Thierfelder, C., Wall, P.C., 2012. Effects of conservation agriculture on soil quality and productivity in contrasting agro-ecological environments of Zimbabwe. Soil Use and Management 28, 209–220.

Thierfelder, C., Mutenje, M., Mujeyi, A., Mupangwa, W., 2015. Where is the limit? Lessons learned from long-term conservation agriculture research in Zimuto Communal Area, Zimbabwe. Food Security 7, 15–31.

Thierfelder, C., Chivenge, P., Mupangwa, W., Rosenstock, T.S., Lamanna, C., Eyre, J.X., 2017. How climate-smart is conservation agriculture (CA)? — its potential to deliver on adaptation, mitigation and productivity on smallholder farms in southern Africa. Food Security 9, 537–560.

Thurstone, L.L., 1927. A law of comparative judgement. Psychological Review 34, 273–286.

Veldkamp, W.J., Muchinda, M., Delmotte, A.P., 1984. Agroclimatic Zones in Zambia. Soil Bulletin No. 9, Soil Survey Unit, Land Use Branch, Department of Agriculture, Republic of Zambia.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, Fourth ed. Springer, New York.

Woode, P.R. (Ed.) 1985. Proceedings of the XIth International Forum on Soil Technology and Agrotechnology Transfer. Part II Excursions and Appendices.