Experimental Agriculture (2020), **56**, 506–527
doi:10.1017/S0014479720000125

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Longitudinal analysis of a long-term conservation agriculture experiment in Malawi and lessons for future experimental design

R. M. Lark[1,a*], I. S. Ligowe[2,a], C. Thierfelder[3,a], N. Magwero[4,b], W. Namaona[4,b], K. Njira[4,b], I. Sandram[4,b], J. G. Chimungu[4,c] and P. C. Nalivata[4,c]

[1]University of Nottingham, School of Biosciences, Sutton Bonington, Nottinghamshire LE12 5RD, UK, [2]Department for Agriculture Research Services, Chitedze Agricultural Research Station, P.O. Box 158, Lilongwe, Malawi, [3]CIMMYT, P.O. Box MP 163 Mount Pleasant, Harare, Zimbabwe and [4]Lilongwe University of Agriculture and Natural Resources, Bunda Campus, Lilongwe, Malawi
*Corresponding author. Email: murray.lark@nottingham.ac.uk

(Received 27 August 2019; revised 30 April 2020; accepted 27 May 2020; first published online 13 July 2020)

Abstract

Resilient cropping systems are required to achieve food security in the presence of climate change, and so several long-term conservation agriculture (CA) trials have been established in southern Africa – one of them at the Chitedze Agriculture Research Station in Malawi in 2007. The present study focused on a longitudinal analysis of 10 years of data from the trial to better understand the joint effects of variations between the seasons and particular contrasts among treatments on yield of maize. Of further interest was the variability of treatment responses in time and space and the implications for design of future trials with adequate statistical power. The analysis shows treatment differences of the mean effect which vary according to cropping season. There was a strong treatment effect between rotational treatments and other treatments and a weak effect between intercropping and monocropping. There was no evidence for an overall advantage of systems where residues are retained (in combination with direct seeding or planting basins) over conventional management with respect to maize yield. A season effect was evident although the strong benefit of rotation in El Niño season was also reduced, highlighting the strong interaction between treatment and climatic conditions. The power analysis shows that treatment effects of practically significant magnitude may be unlikely to be detected with just four replicates, as at Chitedze, under either a simple randomised control trial or a factorial experiment. Given logistical and financial constraints, it is important to design trials with fewer treatments but more replicates to gain enough statistical power and to pay attention to the selection of treatments to given an informative outcome.

Keywords: Climate-smart agriculture; Diversification; El Niño; Malawi; Sustainable intensification; Zero-tillage

Introduction

It is predicted that southern Africa will be substantially affected by increasing climate variability and change (Lobell *et al.*, 2008). Studies based on modelling of future climate effects suggest that the main climate hazards to crop production in the region are a delayed onset of the cropping

[a]Joint lead authors, listed alphabetically.
[b]Joint contributing authors, listed alphabetically.
[c]Joint senior authors, listed alphabetically.

© The Author(s), 2020. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

season, in-season droughts and dry spells, heat waves, erratic rainfall and early tailing-off of rainfall (Cairns *et al.*, 2012; Steward *et al.*, 2019), exacerbated by declining soil fertility (Kumwenda, 1998). The concept of climate-smart agriculture has been developed as an approach to cope with climate change (McCarthy *et al.*, 2011; Lipper *et al.*, 2014). To be regarded as climate-smart, a cropping system must adapt to negative effects of climate change, mitigate these effects (e.g. through reductions in greenhouse gas emissions or increased carbon sequestration) and lead to greater productivity and profitability (Thierfelder *et al.*, 2017).

In Malawi, a land-locked country in south-eastern Africa, the effect of climate change, population pressure and declining soil fertility has been felt for many years (Chinsinga and Chasukwa, 2018). Periodic droughts and floods have become more frequent in Malawi and there is great urgency to find viable solutions for farmers (Tesfaye *et al.*, 2015). The predominant land-use practice in Malawi is ridge tillage, where annual dug ridges of 20–50 cm height and spaced 75–90 cm are prepared (Bunderson *et al.*, 2017). Incorporation, burning or grazing of surface crop residues are the most common residue management strategies (Ngwira *et al.*, 2013). Maize (*Zea mays* L.) is the predominant food crop grown on approximately 80% of the smallholder land area with little diversification of the cereal with legumes (Smale, 1995).

Conservation agriculture (CA) is one of the climate-smart systems which has been heavily promoted in recent years in southern Africa (Thierfelder *et al.*, 2017). CA is a cropping system based on minimum soil disturbance, the retentions of crop residues and crop diversification (Kassam *et al.*, 2009). In addition, it requires complementary good agriculture practices for it to function (Thierfelder *et al.*, 2018). However, despite its transformational effect in the Americas and Australia in the past 40 years (Bolliger *et al.*, 2006; Kirkegaard *et al.*, 2011), mostly on large commercial farms, it was not adopted to a large extent in southern Africa (Kassam *et al.*, 2015) despite some successes in Zambia, Malawi and Zimbabwe (Aagaard, 2011; Bunderson *et al.*., 2017; Marongwe *et al.*, 2011). Currently CA-based systems are being practised on up to 10% of the farm land in southern Africa (Kassam *et al.*, 2015) with variable quality and duration of the practice.

Various benefits from CA have been reported (Thierfelder *et al.*, 2015), including studies that have shown yield benefits in Malawi (Ngwira *et al.*, 2012; Setimela *et al.*, 2018; Thierfelder *et al.*, 2016), which confirms that CA systems have the potential to increase productivity and profitability. Thierfelder and Wall (2009) suggest that this might result from improved infiltration and retention of water under CA. However, CA systems may also lead to slightly suppressed crop yields in the initial years (Giller *et al.*, 2009), to nitrogen lock-up through retention of cereal crop residues with a large C:N ratio (Mupangwa *et al.*, 2019) and to increase in certain pest and diseases carried over through crop residues, which can affect crop performance (Giller *et al.*, 2015). Waterlogging may also be a greater risk under CA (Thierfelder and Wall). The lack of an immediate yield benefit seems to be a major deterrent to widespread adoption alongside the culture and tradition of continued use of tillage practices (Thierfelder *et al.*, 2018).

It was against this background that a CA long-term trial was established in 2007 at the Chitedze Agriculture Research Centre in a randomised complete block design with seven treatments and four replications. The principal aim of the trial was to identify cropping systems under CA that would maintain or increase productivity over time and withstand soil moisture and fertility decline. We were further interested to study pest and disease dynamics, to explore the 'climate-smartness' of CA cropping systems, and to generate reliable soil, water and plant data in a controlled environment to calibrate crop simulation models (Ngwira *et al.*, 2014). Several studies of this trial have been published which address different agronomic questions about the systems in the experiment (Ligowe *et al.*, 2017; Ngwira *et al.*, 2014; Thierfelder *et al.*, 2013). However, there is a need for a synoptic analysis of the multi-season data to address targeted hypotheses about the treatments over the length of the trial, in particular, the joint effects of variations between the seasons and particular contrasts among treatments and groups of treatments. Furthermore, a long-term experiment such as this one offers the opportunity to address questions about the variability of treatment responses and the implications for design of future trials.

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

The objectives of the present study were therefore to: (a) assess the evidence provided by the maize yield data from the first 10 seasons of the Chitedze experiment for differences between treatments and seasons, examining specific hypotheses by means of orthogonal contrasts and (b) to learn lessons from the experiment in terms of power and design requirements for future long-term trials.

Materials and Methods

Study site description

The study was carried out at Chitedze Agriculture Research Station, located on the Lilongwe–Kasungu plains (−13.973 S, 33.654 E, 1145 m above sea level). The soils are ferruginous Latosols (WRB, 1998), (Alfisols – USDA Soil Taxonomy) which are deep and drain freely with a well-developed structure. The top soil is dark, reddish brown with pH ranging from 4.5 to 6.0. In the first year of establishment, baseline top soil analysis showed acid soil reaction (pH H_2O ranging from 5.1 to 5.2), medium SOM ranging from 3.2 to 3.5%, medium N ranging from 0.16 to 0.18%, low P ranging from 1.68 to 15.56 μg g^{-1} and medium K ranging from 0.33 to 0.56 cmolc kg^{-1} (Ligowe *et al.*, 2017).

The average long-term rainfall in the whole rainy season at Chitedze is approximately 900 mm, while the mean temperature ranges between 20 and 22°C. Rainfall measurements at the study site varied considerably from the long-term record with increasing frequency of droughts and low and erratic rainfalls in the last decade (Figure S1 in the supplementary material).

The cropping season usually started at the beginning of December and lasted until April in each year. The last years have been heavily affected by climate abnormalities such as El Niño anomalies in 2015/2016, a drought in 2014/2015 and high rainfalls in the La Niña year of 2016/2017.

The study site is dominated by maize-based farming systems cultivated on up to 80% of the land area of the Lilongwe–Kasungu plains under the dominant land-use practice of ridge tillage. Besides maize, a range of legumes (e.g. cowpea (*Vigna unguiculata* (L.) Walp), groundnuts (*Arachis hypogaea* L.) are grown as rotational crops. Pigeonpea (*Cajanus cajan* Millsp.), which features as a rotational crop in the Chitedze trial, is a common crop in many parts of Malawi but not systematically grown in the Lilongwe–Kasungu plains.

Trial history and experimental design

Prior to trial establishment, the site was under fallow for more than 5 years. The dominant natural fallow bush was *Tithonia diversifolia* (Deliya) and *Acacia polyacantha* (Mthethe). After land clearing, the trial establishment started with a uniform maize crop in 2006 and then trial layout and treatment implementation in 2007.

The trial was established in a completely randomised block design with four replications on-site. The plots were 24 m × 13.5 in dimensions and separated from all neighbouring plots by a passage of 1 m wide. The plots were laid out in four rows (the row extending in the direction of the long side of the plot) and in eight columns. Each block in the experimental design comprised two adjacent columns.

The treatments are described below. Note that by 'direct seeding', we mean that maize was planted directly into undisturbed soil, a standard zero-till practice.

1. Conventional control practices (**CPM**); traditional farmers practice using the hand hoe (ridge and furrow system), maize as a continuous sole crop, no residue retention, stubble incorporated into the row for the following season.
2. Basin planting (**BAM**), planted in manually dug basins, maize as a continuous sole crop, residues retained *in situ* on the soil surface.

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

3. Direct seeding with a dibble stick (**DSM**), maize as a continuous sole crop, residues retained *in situ* on the soil surface
4. Direct seeded crop rotation (**DSRML**), planted with dibble stick, maize-legume rotation, residues retained *in situ* on the soil surface (both phases of the rotation were individual plots).
5. Direct seeded maize intercropping (**DSIMP**): direct seeding with dibble stick, maize with pigeonpea intercropping, residues retained *in situ* on the soil surface.
6. Direct seeded maize intercropping (**DSIMC**): direct seeding with dibble stick, maize with cowpea intercropping, residues retained *in situ* on the soil surface.
7. Direct seeded maize intercropping (**DSIMM**): direct seeding with dibble stick, maize with velvet bean intercropping (seeded 3 weeks after the maize), residues retained *in situ* on the soil surface.

Ridge tillage (75 cm ridge spacing and approximately 20–30 cm in high) and basin digging (of the size 15 cm × 15 cm × 15 cm) were dug each year during the dry winter season (usually in October) and then planted after the first effective rains. Note that basins were re-dug each year in the same positions to minimise disturbance. All dibble-stick planted direct seeding treatments (Treatment 3–7) were seeded on the same day as the CPM and BAM treatments.

Medium maturing commercial maize hybrids with yield potential of >10 t ha^{-1} were used as test crops in the trial (Table S1 in the supplementary material).

The maize was seeded at a target population of 53,000 plant ha^{-1} in all experimental maize plots including intercropping treatments. All maize crops were fertilised uniformly at the site following national recommendations for the agro-ecology. A recommended fertiliser rate of 69 kg ha^{-1} N, 21 kg ha^{-1} P_2O_5 and 4 kg ha^{-1} S ha^{-1} was applied to all the treatments. The application comprised 23 kg ha^{-1} N, 21 ha^{-1} kg P_2O_5 and 4 kg ha^{-1} S as a basal fertiliser dressing at planting and 46 kg ha^{-1} N in the form of urea 21 days after planting.

Weed control was carried out with an initial herbicide control of glyphosate (N-(phosphonomethyl) glycine) at seeding at a rate of 2.5 l ha^{-1} and manual weed control with hand hoes. Manual weeding was undertaken 3 to 4 weeks after planting, and again three to four weeks after the first weeding. A third manual weeding was undertaken after a further 3 to 4 weeks in wetter seasons when weeds achieved appreciable densities after the second weeding. Residues were applied in the first year of practices at a rate of 2.5–3 t ha^{-1} and maintained *in situ* after the first crop harvest.

It is important to note that, while seven treatments are listed above, there are eight plots in each block with two plots allocated to the rotation treatment DSRML. In any one season, one of these plots is in maize and the other under the legume. Legumes in the DSRML were initially only cowpea. From 2011 onwards, the rotational plot was split and both cowpeas and groundnuts were used in the rotation, but separate maize yields were not recorded for the two subplots. The rotational legume crops were seeded at 37.5 cm row spacing by 25 cm in-row spacing (106,666 plants ha^{-1}). Intercrops were planted 50 cm between planting stations (two seeds per station) except for the pigeonpeas which was seeded at 60 cm between planting stations in the same maize rows to avoid it being weeded out (two seeds per station and later thinned one). Both pigeonpea and cowpeas were planted at the same time as the maize. Velvet beans were planted 3 weeks after the maize to avoid competition between maize and the cover crop.

Harvesting procedures

Harvesting followed standard harvesting procedures taking 10 crop cuts from randomly selected quadrats (10 × 7.5 m^2) in each treatment. Maize was harvested after physiological maturity. Both cobs and biomass were weighed fresh and a cob and biomass subsample dried for moisture determination. The final above-ground biomass and grain were calculated in kg ha^{-1} based on a moisture content of 12.5%. Cowpea and groundnuts in rotation with maize as well as

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

intercropped cowpea, pigeonpea and velvet beans were also harvested at physiological maturity. Yields of velvet beans and pigeonpea were very low due to competition for light, moisture stress and pest attacks, respectively.

Statistical analysis

Exploratory analysis

The primary objective of exploratory analysis of data is to evaluate the plausibility of key assumptions in the planned analysis and to decide whether certain actions (e.g. data transformations) are required (Webster and Lark, 2019). To this end, a simple linear model was fitted to the data with blocks and a factorial combination of season and treatment, and the residuals (differences between fitted and observed values) were extracted. Summary statistics were computed for the residuals, and summary plots (histogram and box plots, and the plot of empirical quantiles against theoretical normal equivalents, the QQ plot). Among the summary statistics, we computed was the octile skewness (Brys, *et al.*, 2003). This is a measure of the apparent asymmetry of the distribution of data which is robust to a small number of outlying values. A plot of the residuals against fitted values was also examined to indicate whether there was evidence that the residuals were not homogeneous in their variance (Webster and Lark, 2019).

The residuals were examined for evidence of outlying observations. To do this, we used Tukey's (1977) 'outer fences' as thresholds. The lower outer fence is a threshold value set at three times the interquartile range below the first quartile of the data, and the upper outer fence is set at the same distance above the third quartile.

One issue in the analysis of any data set with repeated observations on the same units over time is the temporal correlation of the residuals. This is discussed further below, but the empirical variogram is a useful exploratory tool (Diggle, 1990). If $z_{i,t}$ is the residual in the ith plot at time t, then the marginal temporal variogram for time lag interval τ is estimated by:

$$\gamma_T(\tau) = \frac{1}{2N_\tau}\sum i = 1, m;\ t = 1,\ T - \tau\{Z_{i,t} - Z_{i,t+\tau}\}^2, \tag{1}$$

where N_τ is the total number of pair comparisons between observations within one plot over a time interval τ, and there are m plots and $T = 10$ seasons. This formula shows that the temporal variogram is computed only from comparisons between residuals for the same plot. A plot of the variogram against the lag interval indicates the magnitude of any temporal correlation in the residuals. If this is present, then the variogram is expected to increase with increasing lag time.

The longitudinal model and its fitting

We considered two linear mixed models. Both have the same fixed effects structure that reflects the experimental design. There are additive effects of blocking (four blocks) and then main effects of treatments, seasons and their interaction. Because repeated measurements are made on the same units (plots), the variance of repeated observations on the units, $\sigma^2{}_1$, must be estimated separately from the residual variance $\sigma^2{}_2$, and this is done in both models. The models differ in how the covariance of the repeated observations within the same plot is treated. In the first model, we assume that the variance of the difference between any two observations on the same plot is constant (sphericity assumption). In the second model, it is assumed that the variance of this difference depends on the interval in time between the two observations. The two models may therefore be written, respectively, as (assuming sphericity)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{y}$ is the length-n vector of observations, $\mathbf{X}$ is a design matrix which represents the fixed effects and $\boldsymbol{\beta}$ is a vector of fixed effects coefficients, representing block, treatment and season

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

effects and the interaction of the latter two. The matrix **Z** is a design matrix, with n rows and m columns where m is the number of plots. The elements of this design matrix are zeroes, with a 1 in element $\{i,j\}$ if and only if the ith observation is in the jth plot. The term is therefore a random variate of unit mean and covariance matrix $\sigma_1^2\mathbf{Z}\mathbf{Z}^{\mathrm{T}}$. The final term is an independently and identifically distributed random variable of variance $\sigma^2{}_2$.

The second model may be written as in Eq (2), and the difference is the definition of the covariance matrix for the term $\boldsymbol{\alpha}$ which now takes the form $\sigma_1^2\mathbf{R}$ where **R** is an $n \times n$ correlation matrix such that elements for the ith and jth observation are

$$\begin{aligned}\mathbf{R}\{i,j\} &= 1, \text{ if } i = j,\\ &= 0, \text{ if observations } i \text{ and } j \text{ are in different plots},\\ &= \rho(\tau) \text{ otherwise},\end{aligned} \tag{3}$$

where ρ (τ) is the absolute difference in time between the ith and jth observations and ρ is a correlation function. In this case, we used the Matérn correlation function (Stein, 1999) which has a smoothness parameter, κ, which determines the nature of variation over short-time intervals and a time parameter, φ, which controls the time interval over which the correlation decays. The models were fitted by minimising the negative log residual likelihood, following Diggle and Ribeiro (2007) using profiling over a few discrete values to find κ.

Note that the second model is more complex than the first, with two additional parameters (those of the correlation function). It will therefore always have a minimised negative log residual likelihood at least as small as that of the simpler model. To select between the models, it is necessary to account for this difference in complexity, and this can be done by comparing the values of Akaike's information criterion (AIC). A strategy of selecting the model for which AIC is smallest will minimise the expected Kullback–Leibler divergence between the estimated model and the process that generates the data over the set of models considered (Buckland *et al.*, 1997).

Inference

Inferences about the fixed effects in the model can be made based on the Wald statistic which may be compared with the F-distribution with specified numerator and denominator degrees of freedom. The latter degrees of freedom were found using the method of Kenward and Roger (1997, 2009) which accounts for the dependency within a linear mixed model. Models were fitted in this study using the lme4 package in R for estimation (Bates *et al.*, 2015), the afex package (Singmann *et al.*, 2018) and the pbkrtest package Halekoh and Højsgaard (2014) for inference and the emmeans package (Lenth, 2018) for estimation of effects.

While an overall effect of treatments can be examined by the methods described above, this is rarely of direct practical interest. Rather, particular treatments, or groups of treatments, may be compared to test particular hypotheses. These hypotheses should be determined in advance of the analysis, and, ideally, should be structured so that they map on to orthogonal contrasts among the treatments, that is to say into contrasts which constitute a partition of the overall treatment effect into independent components. The hypotheses identified for this analysis are now set out.

The first three hypotheses concern comparisons among broader groups of cropping systems. These are (H1): the mean yield under methods based on residue retention (either with direct seeding or with basins) exceeds that under conventional practices, (H2): mixed systems (intercropping or rotations) will have larger yields than monocropping systems (all with direct seeding) and (H3): the rotations will lead to larger maize yields than intercropping systems. The second three hypotheses concern specific comparisons within these groups which may provide a basis for specific advice about the choice of systems. These are (H4): basin planting will lead to larger yields than will direct seeding without this disturbance, (H5): yields will differ under intercropping with the late sown intercrop (velvet bean, sown 3 weeks after the maize) relative to cowpea and pigeon pea

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

sown with the maize and (H6): maize yields in intercropping systems will differ depending on whether the intercropped legume is cowpea or pigeon pea.

These hypotheses correspond to six orthogonal contrasts. These were as follows:

H1: Contrast C1. Control (CPM) plot versus the rest (comparing conventional treatment with various options which use zero-till and direct seeding with a dibble stick).

H2: Contrast C2. Within the direct seeding treatments: monocrop — BAM (basins) and DSM (dibble stick) versus mixed (rotations or intercrop)

H3: Contrast C3. DSRML (Rotation) versus DSIMC, DSIMP and DSIMM (intercrop)

H4: Contrast C4. BAM versus DSM

H5: Contrast C5. Late sown intercrop (DSIMM, Velvet bean) versus intercrop sown with maize – DSIMC (cowpea) or DSIMP (pigeon pea)

H6: Contrast C6. DSIMC versus DSIMP

Contrasts between the seasons were examined with prior orthogonal contrasts, but not a complete set (there are 9 degrees of freedom for season differences). Rather, we focused on hypothesised differences between seasons according to the NOAA Oceanic Niño Index (ONI). The value of El Niño Southern Oscillation as a predictor of crop yield in southern Africa has been reported previously (e.g. Cane *et al.*, 1994, Iizumi *et al.*, 2014). We identified as El Niño seasons those in which a strong positive ONI was recorded in at least one of the 3-month running averages from November to April prior to harvest. If there was a strong negative ONI in the same period, then the season was identified as La Niña (note that there was no conflict between these criteria in this period). Otherwise, a season was identified as neutral. This approach is comparable to that of Iizumi *et al.* (2014) who considered ONI in the 3 months prior to harvest as indicative of climatic conditions during the key reproductive phases of crop development. We hypothesised that yields would be smaller in El Niño seasons by this criterion, and that yields would be smaller in La Niña seasons than in neutral ones. These two hypotheses are encoded in the following two orthogonal contrasts:

Contrast S1: differences between El Niño seasons and La Niña or neutral seasons.

Contrast S2: differences between the La Niña and neutral seasons.

In order to better understand the season by treatment interaction, we considered the three 1-df components of the interaction: C1•S1, C2•S1 and C3•S2.

We treated the treatment contrasts, season contrasts and interaction terms as three separate families of tests for purposes of controlling family-wise error rate (FWER), which was done by the method of Holm (1979). We focused on the interactions of the first three treatment contrasts with the El Niño effect to limit the number of interaction terms for which the null hypothesis was tested and so to maintain the statistical power with which this particular family of tests was examined with FWER control.

Statistical power

One objective of this study is to use the Chitedze trial, a unique resource, to identify general lessons for the design of other experiments in the region. A key question is how much replication

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

such experiments require. The statistical power of a particular experiment and analysis is an important practical measure of its efficiency. Power is computed for a particular effect size (e.g. difference in means among treatments). The statistical power is the probability that a true effect of this size would be detected as significant (with a p-value below a specified threshold). Other things being equal, the power of an experiment can be increased by increasing replication or adopting blocking (Lawson, 2014) or by incorporating covariates into the data analysis that account for some of the residual variation (Rudolph *et al.*, 2016). The power for an experiment with a specified effect size and number of replicates can be computed from the non-centrality parameter of the variance ratio with the specified effect size and from the magnitude of the expected residual mean square (RMS). The daewr package for the R platform (Lawson, 2014) provides functions for doing this.

We computed single-season analyses of variance for the Chitedze trial and extracted the values of RMS. We then considered the problem of designing an experiment with just two treatments, a control and an intervention, with a specified effect size. This effect size was set at an increase of 25% on the mean yield of the experimental control plots over all seasons. This is equal to a yield increase of 1.16 t ha^{-1}. We computed the power to detect such an effect, for each season's RMS.

In fact, as at Chitedze, the power of an experiment must be evaluated over multiple seasons, and the effect of this on power will be one consideration when planning an experiment and in the interpretation of its outputs as it progresses. To this end, we considered an extension of the case above in which the experiment is run for 2, 3, 4 or 5 seasons. The power analysis was done by simulation, based on the random effects for the mixed model fitted for the Chitedze data. A data set was simulated with number of replicates set to 4, and number of seasons to 2 and the treatment means differing by 1.16 t ha^{-1}. The simulated data were analysed by a linear mixed model, and the p-value for the treatment effect extracted. This was repeated 2000 times and the proportion of cases in which $p < 0.05$ was computed as an estimate of power. This was repeated for all combinations of 2, 3, 4 or 5 seasons with 2, 4, 6, ..., 16 blocks.

Finally, we conducted a second power analysis by simulation for a factorial experiment in which two factors – diversification (with levels 'monocrop' or 'rotation') and cultivation (with levels 'conventional with residues removed' and 'zero till with residues retained') – were applied in full factorial combination. Such an experimental design includes one treatment which corresponds to a conventional cropping system ('monocrop' with 'conventional cultivation') and one which corresponds to full CA management ('rotation' with 'zero till'). The factorial structure allows the contributions of the two components of the CA system to be evaluated along with their interaction. In the simulation, we considered completely randomised blocks (four plots per block) and the same number of seasons and blocks used in the previous power analysis. Power analysis requires the specification of an effect size, and so this is somewhat speculative. For consistency with the previous power analyses, we assumed an overall effect size for the CA treatment over against conventional management of 1.16 t ha^{-1}. We assumed that each factor contributed an additive component of 0.5 t ha^{-1} to this effect, with the additional 0.16 t ha^{-1} the result of the interaction.

Note that all these power analyses are conditional on the variance components estimated from the Chitedze experiment, and so are conditional on the plot size used, as well as local conditions.

Results

Maize yields: treatment and season effects and interactions

The mean yields of maize in this experiment show variations between treatments and seasons (Figure 1). Plots of the residuals of the first exploratory model were examined (Figure 2), and their summary statistics are presented in Table S2 in the supplementary material (row 1). Note that one observation appears to be an outlier. The corresponding datum was removed,

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

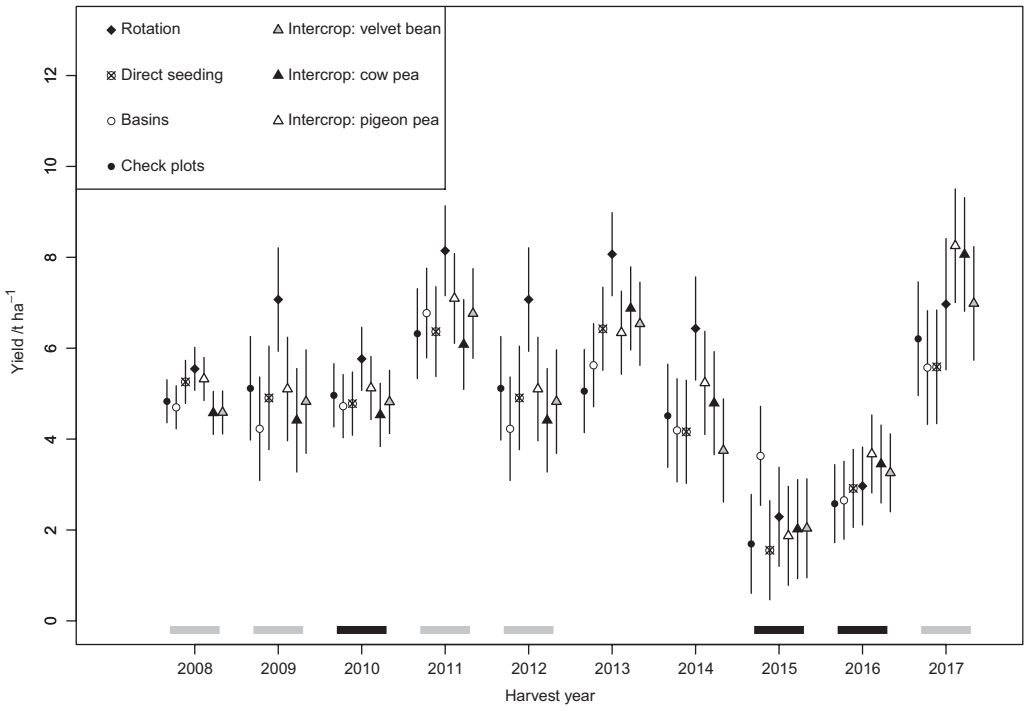

Figure 1. Treatment mean yields with 95% confidence interval for each harvest year (based on the pooled within blocks and treatments standard error in each season). Black horizontal lines at the base of the plot indicate those seasons in which a strong positive Oceanic Niño Index was recorded in at least one of the 3-month running averages from November to April prior to harvest. The grey horizontal lines indicate those seasons in which a strong negative index was recorded in the same period.

and the exploratory model was refitted. Summary plots for the second set of residuals were examined again (Figure S2 in the supplementary material) along with their summary statistics (Table S2 in the supplementary material, row 2). The absolute value of the skewness of the residuals after removal of the outlier is less than 0.5, smaller than the threshold value of 1 at which transformations are commonly considered (Webster and Oliver, 2013), and the absolute octile skewness is smaller than 0.2, an equivalent threshold (Rawlins *et al.*, 2005). These statistics, and the histogram, boxplot and QQ plot in Figure 2, suggest that the residuals may plausibly be regarded as normally distributed. The plot of the residuals against the fitted values in Figure S2 in the supplementary material suggests that the errors in the model can plausibly be treated as homogeneous in variance. On this basis, further analysis was done with the data in their original units (t ha^{-1}). The outlying observation was in plot 5 in block 1 (rotation) in the 2017 harvest year. The recorded yield was an order of magnitude smaller than for any other plots in the block, or plots in the same treatment, and Figure 2 shows how markedly different the corresponding residual for the fitted model is from those for the remaining observations. The experiment manager revisited the records for the season and concluded that the entry was likely to be an error of transcription, possibly through confusion with legume yield records for the associated plot in the rotation treatment. That single datum was therefore removed before all further analysis.

The empirical temporal variogram of the residuals in Figure S3 in the supplementary material shows no evidence for temporal dependence of the residuals within the experimental plots, with the variogram values at different time lags fluctuating around a value of approximately 0.5, and no evidence of an overall increase with time lag. The results from fitting the alternative linear mixed models to the data (Table 1) show that the negative residual log likelihood for the model with

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

Table 1. Fitted random effects parameters for alternative linear mixed models with negative log-likelihood (NLL) and AIC

Model	κ	φ	σ^2_1	σ^2_2	NLL	AIC
Exponential	2.0	5.34	0.376	0.688	−48.74	56.51
Sphericity	–	–	0.322	0.735	−47.77	54.47

σ^2_1 variance of repeated measurements on the same plot.
σ^2_2 residual variance.

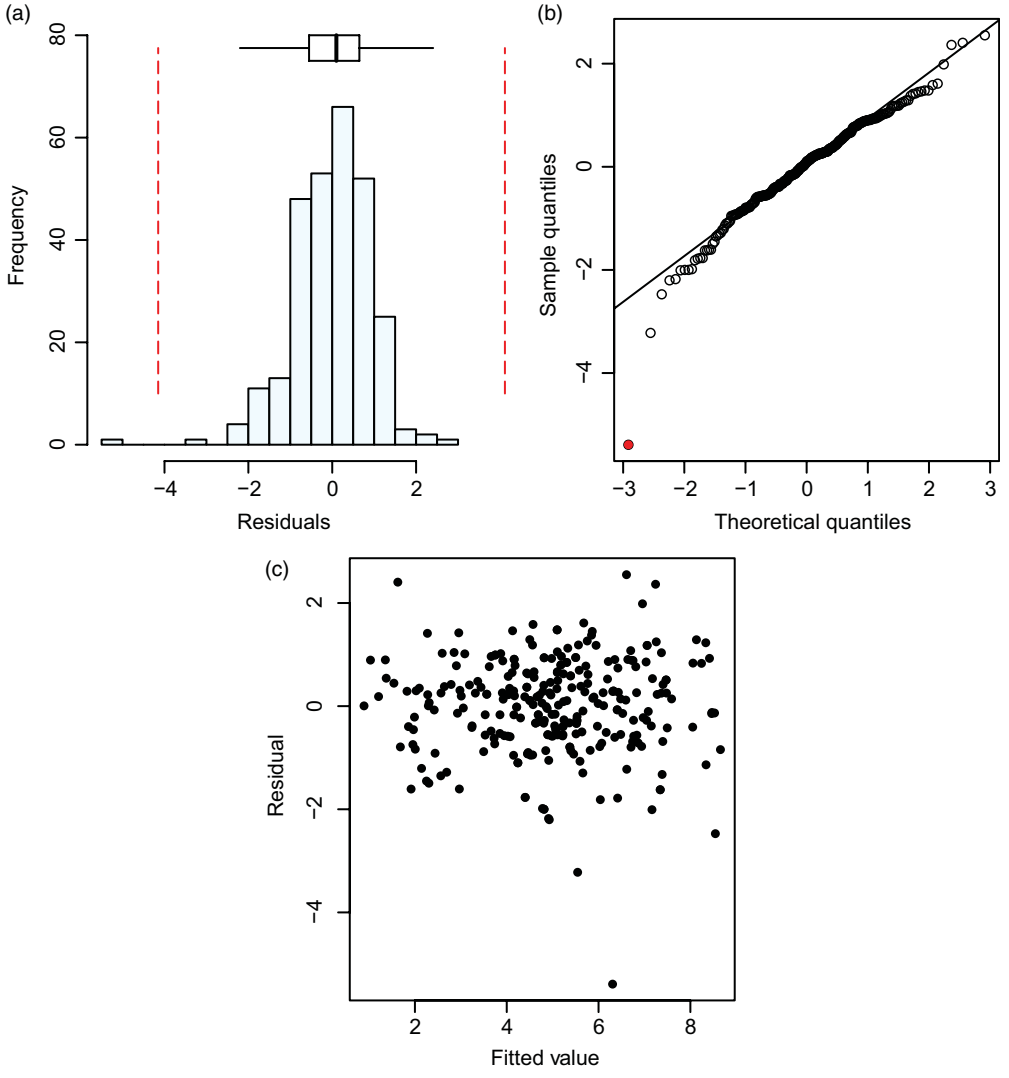

Figure 2. (A) Histogram for residuals from an exploratory model with a box-and-whisker plot. The vertical dashed line shows the threshold (lower) for probable outliers following Tukey (1977). (B) Plot of empirical and normal quantiles for residuals with normal line, and the solid disc shows a probable outlier according to the criteria of Tukey (1977). (C) Plot of residuals versus fitted values.

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

Table 2. Analysis of variance table (Kenward–Roger-adjusted Wald statistics with adjusted denominator degrees of freedom)

	Numerator Df	Denominator Df	Wald statistic	*p*
Block	3	20.83		
Season	9	191.86	87.85	$<2.2\times10^{-16}$
Treatment	6	20.18	4.72	0.0037
Season•treatment	54	178.11	2.01	0.0003

Table 3. Planned orthogonal contrasts among treatment means and three components of the treatment•season interaction

	Numerator Df	Denominator Df	Wald statistic	*p*	*p Holm adjustment*
C1 Check plots vs rest	1	19.44	1.88	0.1862	0.7447
C2 Monocrop vs mixed	1	19.69	6.18	0.0220	0.1100
C3 Rotation vs intercrop	1	24.25	15.75	0.0006	0.0034
C4 Basin vs DS	1	19.35	0.01	0.9063	1.0
C5 (CP and PP) vs VB	1	19.35	0.52	0.4783	1.0
C6 CP vs PP	1	19.35	0.77	0.3911	1.0
S1 El Niño vs (La Niña/neutral)	1	191.65	432.50	<0.0001	<0.0001
S2 La Niña vs neutral	1	186.01	3.03	0.0840	0.0836
C1•S1	1	184.46	1.20	0.2749	0.5498
C2•S1	1	184.46	0.52	0.4698	0.5498
C3•S1	1	202.63	13.01	0.0004	0.0012

exponential temporal dependence is only slightly smaller than that for the model assuming sphericity, and the AIC of the latter model is the smaller of the two. This is consistent with the empirical temporal variogram. On this basis, the model assuming sphericity was selected for subsequent analyses.

The Wald statistics for the main effects and their interaction (Table 2) provide strong evidence to reject the null hypothesis for both main effects (differences between seasons and between treatments) and their interaction, indicating that treatment effects differ between the seasons. Note that the denominator degrees of freedom computed by the method of Kenward and Roger (which is why they are not, in general, whole numbers).

The main effects and interactions are elucidated in more depth on the basis of the specific planned contrasts among levels of these factors (Table 3). The first orthogonal contrast, C1, between the conventionally managed check plots and the treatments with residue retention and direct seeding or basins considered together, is not significant. The mean effect size of the contrast is 0.46 t ha^{-1} (smaller yields on the conventional management), but the uncertainty of this estimate is large and the 95% confidence interval for the effect ranges from −1.17 to 0.23 t ha^{-1}. In summary, the experiment provides no evidence for an overall mean increase in maize yield with residue retention as a whole in comparison to conventional management.

Contrast C2 compares monocrop CA and mixed cropping systems. The estimated mean yield for the monocrop treatments was smaller than for the intercropping treatments by 0.67 t ha^{-1}, but, although the 95% confidence interval for this effect does not include zero, our evidence to reject the null hypothesis that the contrast (C2) is zero is weak when FWER is controlled at 0.05 within the group of contrasts among treatments (see the *p*-value with Holm adjustment in Table 3).

There is strong evidence for a difference in mean yield between the rotation plots and those with intercropping (contrast C3). The estimated effect sizes (Table 4) show that, averaged over seasons, the rotation treatment yielded 1.2 t ha^{-1} more than the intercropping treatments, and the null hypothesis of no effect can be rejected with $p = 0.003$ under FWER control.

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

Table 4. Estimated effect sizes for 1-df contrasts and their confidence intervals

Contrast[a]	Effects/t ha^{-1}	Df	SE/ t ha^{-1}	95% CI
C1 Check plots vs rest	−0.46	19.44	0.337	−1.17, 0.23
C2 Monocrop vs mixed	−0.67	19.69	0.268	−1.22, −0.11
C3 Rotation vs intercrop	1.21	24.25	0.304	0.58, 1.83
C4 Basin vs DS	−0.05	19.35	0.445	−0.98, 0.88
C5 (CP and PP) vs VB	−0.28	19.35	0.385	−1.08, 0.53
C6 CP vs PP	−0.39	19.35	0.445	−1.32, 0.54
S1 El Niño vs (La Niña/neutral)	−3.52	191.65	0.169	−3.85, −3.18
S2 La Niña vs neutral	−0.23	186.01	0.136	−0.50, 0.03

[a]NB first-named treatment has positive weight.

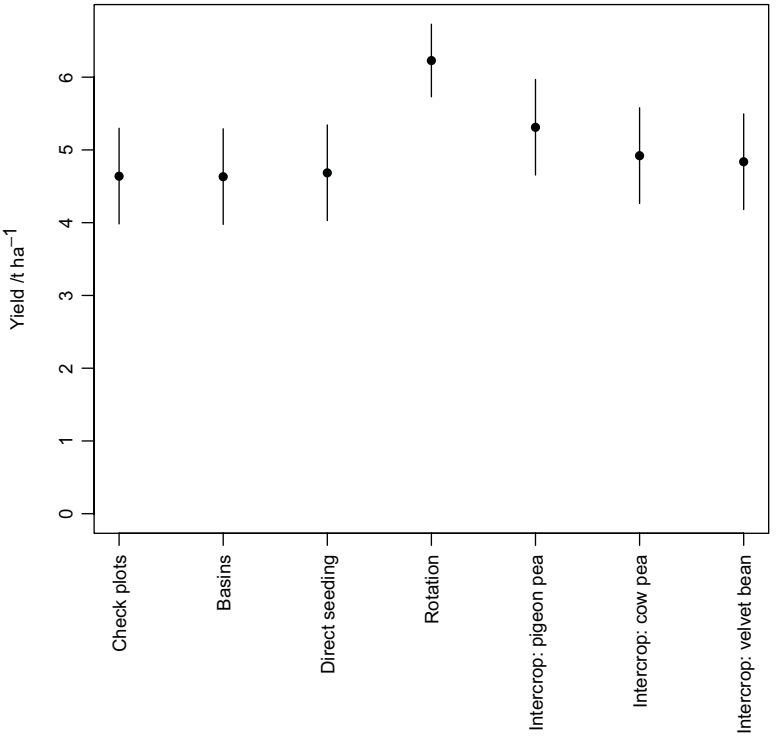

Figure 3. Estimated mean yields by treatment over all seasons, with 95% confidence interval.

The remaining contrasts comprising the treatment main effect all have small effect sizes with confidence intervals including zero, and the *p*-values are all large, indicating an absence of evidence to reject the null hypothesis. The treatment means, over all blocks and seasons, with their confidence intervals based on the random effects in the combined model are shown in Figure 3. Note that the mean yields for the check plots and monocrop CA plots are all very similar. The plots with CA and crop rotation for diversification clearly have the largest mean yield. There is strong evidence to reject the null hypothesis that overall mean yield does not differ between the El Niño seasons (harvest in 2010, 2015 and 2016) and the rest (Table 3 contrast S1), and Table 4 shows that, on average across treatments, the yield in the El Niño seasons is 3.52 t ha^{-1} less than the others. There is no evidence for an overall difference between La Niña and neutral seasons. Seasonal climate is expected to be an important factor in crop yield. Note also that there were infestations of wireworm (*Gonocephalum spp*) and white grub (*Heteronychus arator*) in the trial in the 2014/2015 and 2015/2016 seasons, which coincided with very dry years.

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

The overall treatment means, shown in Figure 3, obscure differences among seasons, which, as seen in Table 4, and the overall significant season by treatment interaction, can be substantial. The treatment main effects are of considerable practical relevance because they represent the mean yield that farmers can expect over a period of inter-season variation comparable to that for this experiment. That said, it is important to consider the interactions with season effects, because they may show, for example, that a particular treatment is most vulnerable to adverse seasonal conditions, or conversely, particularly robust in those conditions.

The interaction of C1 (conventional management vs all practices with residue retention) with the S1 season effect was not significant ($p = 0.55$, Table 3). The contrast C1 was not significant, and this component of the interaction shows that the experiment provides no evidence that, for example, the checkplots are at a particular disadvantage in the El Niño seasons.

Of the components of the interaction examined in more detail, only the difference between the rotational and intercropping treatments appears to differ between El Niño and other seasons. On examination of the raw treatment means in Figure 1, it would appear that the advantage of the rotational management is, in general, larger in La Niña and neutral seasons.

Statistical power

The values of RMS for successive one-season analyses of the trial data vary from year to year (Figure S4 in the supplementary material). Note that the smallest RMS was in the first season. This may reflect the effect of the varied treatments accumulating over time with the impact of direct seeding and residue incorporation, as well as rotations and intercropping on soil physical and chemical properties. The largest RMS is for the 2017 harvest. Note that this was a La Niña season in which there were also notable infestations of fall armyworm (*Spodoptera frugiperda*). Fall armyworm, a new invasive lepidopteran pest, reached Malawi in late 2016 and started affecting general crop yields. This had an unforeseen treatment effect on the longer-term maize yields.

The results for single-season power analysis (Figure S5 in the supplementary material) show that, with just four blocks, the power to detect the target effect never exceeds 0.8 (a common target power) and is more typically around 0.2. With 16 blocks, the 0.8 target is met or exceeded for all seasons apart from the 2017 harvest. Note the implication of our observation that the RMS was smallest in the first (2008) harvest season: the target power was exceeded for all cases with eight or more blocks. However, this is in the first season only, when treatment effects may be very limited.

Turning to the more realistic case, of multiple-season analyses, Figure 4 shows the power estimates obtained by simulation for the simple two-treatment experiment. Note that with 10 or more blocks and 2 or more seasons, the target power (0.8) is achieved. Six or fewer blocks are insufficient to achieve the target power with five or fewer seasons, but with eight blocks the target power is achieved if the experiment runs for four seasons or longer. The general point to note is that increasing replication by one block has a larger effect on power than extending the experiment by one season. When we ran this simulation to emulate the Chitedze experiment (10 seasons with four blocks), the power to detect the 1.16 t ha^{-1} target effect was 0.42.

Figure 5 shows the power estimates obtained by simulation for a simple factorial experiment with two factors each with two levels (diversification – monocrop or rotation; cultivation – conventional or zero till with retention of residues). In this analysis, we assumed that the overall treatment effect in the two-treatment power analysis was partitioned between two equal main effects and an interaction. Greater replication is necessary to detect these more complex effects. The main effect can be detected with target power with 16 blocks, or 12 blocks over 5 or more seasons. Detection of the interaction with target power, in these simulations, required 14 blocks and 4 or more seasons or 16 blocks and 3 or more seasons. When we ran this simulation to emulate the Chitedze experiment (10 seasons with 4 blocks), the power to detect the target main effect was 0.37, and the power to detect an interaction was the same.

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

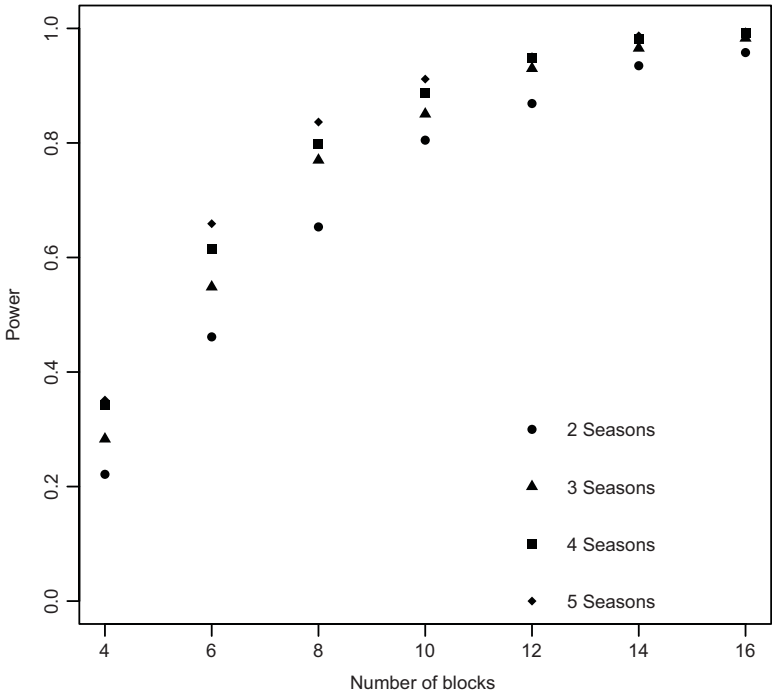

Figure 4. Power estimates obtained by simulation using random effect parameters from the Chitedze experiment for a two-treatment experiment over two or more seasons and with a target yield effect of 1.16 t ha^{-1}. Power is shown for differing numbers of blocks.

Discussion

Treatments and their effects

In this study, we focused on a long-term trial of CA systems to assess their effects on crop yield in 10 contrasting seasons. We set out to do this in a rigorous statistical framework to avoid the trap of focusing on interesting comparisons between particular treatments in particular seasons, which the human eye and brain are effective at detecting. Such effects may be real and may merit further investigation. For example, we note the apparent yield advantage of planting in basins in the 2015 El Niño harvest season, but the aim of any statistical analysis is to identify effects for which there is robust evidence against a background of natural variation. This requires the explicit control of the rate of false discovery in large and complex experiments, which is essential to avoid drawing optimistic but misleading conclusions.

There is a significant difference between yields from the treatments over the first 10 seasons (Table 2). The treatment means (Figure 3) represent the main effects of the treatments over all seasons. The interpretation of these requires some caution because of the interaction of treatment effects with seasonal differences (below), but nonetheless, the mean yield of a treatment over a decade is of considerable interest because it represents the long-term benefit to farmers of investing in a change of farming practice.

The specific contrasts between treatment means (Table 3) show some results of particular interest.

First, there is no evidence from this experiment for an average yield difference between conventional management and the plots with residue retention and direct seeding or basins, considered as a whole. The estimated mean yield advantage of the plots with residue retention is 0.46 t ha^{-1} but the confidence interval for this is wide including zero, and the evidence provided

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

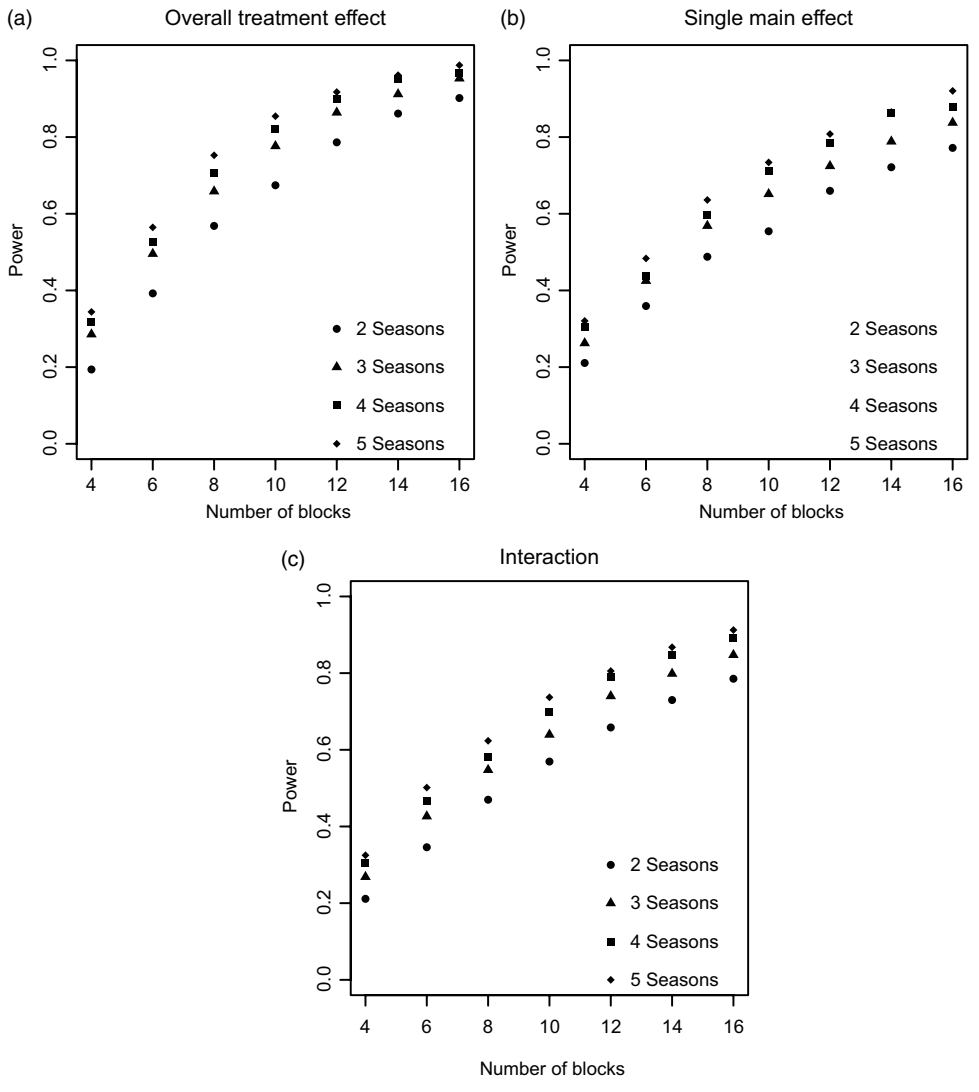

Figure 5. Power estimates obtained by simulation using random effect parameters from the Chitedze experiment for a 2 × 2 factorial experiment over two or more seasons. The additive effects of the two factors (rotation as opposed to monocropping, and zero till with retention of residues as opposed to conventional cultivation) are each 0.5 t ha^{-1} with an additional 0.16 t ha^{-1} for the full CA treatment (interaction effect) Power is shown for differing numbers of blocks and (A) for the overall treatment effect (3-df), (B) for a single main effect (power is the same for both factor) and (C) for the interaction.

by the p-value for contrast C1 does not allow us to reject the null hypothesis of no effect ($p = 0.74$ on adjustment to control FWER).

Second, there is strong evidence for a difference in maize yield between the rotational treatment and intercropping (contrast C3), and there is weak evidence (particularly when the FWER is controlled) for an overall difference between the monocropping and mixed treatments (contrast C2). The mean yield under rotations is the largest of all treatments (Figure 3), consistent with the inference about contrast C3. This result is of interest as many smallholder farmer fields in Malawi practice intercropping. Other findings have suggested that yield benefits due to reduction of pests and diseases and the accumulation of residual nitrogen are larger under rotation than

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

intercropping (Thierfelder *et al.*, 2012). It should be noted, however, that the yield advantage of rotation is not so large that a producer could produce the same amount of maize from a given area of land with half of it in the non-maize phase of the rotation as under intercropping

It is notable that over these 10 seasons, the mean yields from the control treatments and from the monocropping treatments with direct seeding (basin sowing and direct seeding) are very similar (Figure 3). The difference between treatment means for the basins and direct seeding monocrop (DSM) is very small (−0.05 t ha^{-1}), and the corresponding contrast in the analysis of variance (C4) shows that there is no evidence for any difference between these treatments ($p = 0.91$). While soil disturbance through yearly ridge and basin making might be expected to influence water retention and also nutrient dynamics in the soil, the Chitedze trial provides no evidence for effects on maize yield when basins were used, other factors (monocropping and direct seeding) held constant. This is consistent with the study by Thierfelder *et al.* (2018), who argue that zero tillage with retention of residues without crop diversification will not give benefits in southern African conditions.

Not surprisingly, there is a marked effect of season (Table 2) on maize yield. This main effect may incorporate various factors, such as differences among the maize varieties (Table S1), although all varieties used were in the same maturity group and such a strong effect is not expected from a change from one commercial hybrid to the another. However, there is very strong evidence (Table 3) for a contrast between the El Niño seasons (defined in the Methods section with respect to 3-month running mean anomalies) and the 'normal' cropping seasons. This contrast is particularly marked in the low yields for the 2015 and 2016 harvest years, which corresponds with the build-up and the full El Niño years. As noted in the results section, the 2015 and 2016 harvest seasons also showed wire worm and white grub infestations. Interestingly, there is no evidence for a difference in yield between the La Niña seasons and neutral ones.

There is evidence for an interaction of season and treatment effects (Table 2), indicating that differences among the treatments depend on seasonal conditions. Some such effect is expected, because there will be long-term changes in soil conditions on the plots after conversion from fallow, and carry-over effects of treatments are expected between seasons. Note, for example, the marked increase in the difference between the rotational and intercropping yields from season 1 (when the rotational plots did not have a preceding legume crop) and season 2 (when they did). The family of interactions among the treatment and season contrasts (Table 3) show that the difference between the rotational and intercropping treatments itself differs between El Niño seasons and the rest (the C3•S1 component of the interaction with $p = 0.0012$). There is no evidence that the other treatment contrasts considered (control plots vs the rest, or monocropping vs mixed treatments) differ between El Niño and non-El Niño seasons.

Previous analyses (Steward *et al.*, 2018) have shown that there may be larger benefits from CA systems relative to alternatives under conditions of drought and heat than under more evenly distributed rainfall. However, the Chitedze trial provides no evidence that the C1 contrast (CPM vs systems with residue retention) interacts with the El Nino component of the season effect (S1), $p = 0.27$ (Table 3). Furthermore, season and treatment interactions are likely to be complex, and here the key component of the interaction is (C3•S1). Inspection of Figure 1 shows that in 8 of the 10 seasons, the mean yield for rotational plots exceeded the mean yields of the intercropping treatments. However, this difference was much reduced in the El Niño predecessor year 2014/2015, and in 2015/2016 El Niño season. In both years, the rotational plots had smaller mean yields than all intercropping treatments, an effect which persisted in 2016/2017.

The Chitedze experiment fails to show a general advantage of systems with residue retention as a whole over conventional management. This may, in part, reflect local conditions. For example, Thierfelder and Wall (2012) have previously shown that decline of fertility under conventional management may be buffered at sites which are initially very fertile, and that could apply here particularly as the experimental plots were under fallow immediately prior to the experiment.

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

The use of recommended rates of fertiliser on all plots may also reduce differences between conventional management and CA practices.

The findings of this analysis may also point to additional lessons which should be learned about experimental design for cropping systems in the region. Note that the *post hoc* power analysis for a proposed effect size of 1.16 t ha^{-1} in a simple two-treatment trial with the same number of blocks and seasons was small (0.42). The amount of replication in the experiment is probably insufficient. We therefore consider next the findings of this study that are relevant to design of future experiments

Variability of the experimental material in space and time

A key focus of this study is on the statistical model used for the longitudinal analysis of the whole 10-harvest data set, and its implications for experimental design in CA research. One key feature of the statistical model is the lack of evidence for pronounced temporal correlation of the within-plot random effect, suggested by the empirical temporal variogram in Figure S3, and corroborated by the selection of the model assuming sphericity (Table 1). This implies that the interaction of seasonal variation with spatial variations across the site is sufficiently complex to mask any carry-over of effects that lead, for example, for a large positive within-plot random effect in one season into subsequent seasons. Note also that the variance for repeated measures on the same plot is somewhat less than half the residual variance.

The RMS for single-season analysis (Figure S4) is smallest in the first season and somewhat erratic in subsequent seasons with the largest residual variance in the 2017 harvest (which might reflect the new impact of Fall armyworm on crop yields which started in cropping season 2016/2017). A topic for further work is whether there is generally an increase in residual variance after the introduction of contrasting treatments on a site, previously under fallow (as at Chitedze) or uniform cultivation. This could arise from non-linearities in the effects of treatments and their interaction with within-site variation. If this is so, then it implies that experiments on interventions where such interactions with inherent variation of the site are expected might require more replication than other experiments under conventional management (e.g. fertiliser or variety trials), and that data from the latter should be treated with caution in power analysis for CA trials. An important task for further research is therefore to identify available data specifically from CA trials and to analyse these appropriately to develop robust guidelines for experimental design.

The power analyses in Figure S5 show that, for single-season trials, the probability of detecting an increase of 25% over the mean conventional yield (an increase of 1.16 t ha^{-1}), based on the RMS for each season in turn, is small with fewer than 16 replicates (with the exception of season one). The relationship between power and the RMS for some effect size is not linear, and the difference between the first season and subsequent ones with respect to power is quite pronounced, underlining the point above that CA trials, once established, might not be very sensitive at conventional levels of replication.

Of course, single-season power analysis is not a direct guide to practice as most agronomic trials run over several seasons concurrently. The power results in Figure 4, for the same effect size of 1.16 t ha^{-1}, are based on variance components from the model in Table 1. Other things being equal, the lack of temporal correlation of the repeated measures on plots should mean that adding seasons to a designed trial has a larger effect on power than if the correlation was large. However, because the variance of repeated measures is smaller than the residual, it is notable in Figure 4 that the effect of adding one season to the trial is notably smaller than the effect of adding one additional block. A target power of 0.8 can be achieved with eight replicates if the trial runs for at least four seasons. This information could be used when planning trials. Costs entailed in added replicates (more land requirement, more inputs and labour) must be balanced with respect to the costs of increasing the length of the trial (land rental, opportunity costs). It is also necessary to consider the time taken for effects of any treatments which are manifested through changes in soil

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

conditions, which may take 2 to 5 years (Thierfelder *et al.*, 2015). Nonetheless, the benefits of a longer trial (to capture a range of climatic conditions) will be limited if there is not sufficient replication of the experiment in space.

Similar conclusions were reached from the power analysis for a factorial experiment. While such an experiment might be more effective than a randomised control trial for gaining insight into interventions like CA with several components, the power to detect main effects and interactions in our example was somewhat less than to detect a difference between two treatments with the same amount of replication. It is important to identify an informative set of treatments for an experiment, but they must also be adequately replicated.

Implications

Given the importance of understanding the impacts of CA practices and their potential role for food security under climate change, these analyses suggest that experimenters should pay careful attention to questions of experimental design, including power. It is better to have adequately replicated experiments with a small number of carefully focused treatments than to spread resources too thinly. We propose further in-depth meta-analysis of CA trials in southern Africa. Rather than focusing on understanding CA systems *per se*, they should be focused specifically on the spatial and temporal variability of responses and their implication for the statistical power of experiments as a function of length (number of seasons) and replication.

It is also important to pay attention to the treatments selected for experimentation. While ostensibly an experiment with a control and a range of proposed interventions allows these to be tested, it should be recalled that a set of comparisons between treatments and a single control do not constitute orthogonal contrasts (Dunnett, 1955). Special treatment is needed for such tests, with a cost in statistical power. Furthermore, such simple contrasts may not be very informative about how components of a CA system contribute to its outcomes. For example, the Chitedze experiment does not allow us to separate out the effects of direct seeding (with residue retention) and of diversification. This is because these two components are not studied in full factorial combination (we have no plots which use rotations, for example, under conventional cultivation). Fisher (1926) observed that factorial experiments, in which treatments consist of combinations of levels of distinct factors, may be very informative. That is the motivation for the power analyses undertaken in this study on a notional 2×2 factorial design which would allow the disentangling of cultivation and diversification effects. In our proposed example, however, rather more replication again would be needed to achieve a target power of 0.8.

We therefore suggest that further thought is given to the treatment structure in CA experiments. As discussed above, a factorial experiment may be more informative about the components of a CA response than a set of comparisons between treatments and a control. We have proposed one factorial experiment, based on the observation in this study that yields under rotation are larger than any of the other treatments. As is often noted, CA is neither simply zero till with mulching nor it is simply the diversification of cropping systems. Factorial experiments would allow the main effects and their interactions to be teased out in more detail. First attempts in this regards have been reported from Thierfelder *et al.* (2013) and Ngwira *et al.* (2014).

A new generation of CA experiments, with adequate replication and careful selection of treatments based on past and recent work, would be very informative. Other issues raised by the Chitedze experiment include the question of prior landuse on experimental sites (at Chitedze it was fallow, which is not generally representative) and the management of the plots (at Chitedze, they are kept free of weeds, and fertiliser is applied in accordance with standard recommendations). Consideration should be given as to whether such good agronomic practice is sufficiently representative for smallholder farmers' conditions which are more cash and input constrained to reap the full benefits of CA in their own field fields. However, as this was not the

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

objective of this study but to understand processes and statistical designs, these considerations may be secondary for this particular case.

Conclusions

The analyses presented in this paper provide a synoptic account of the differences among agronomic systems in one of the few long-term experiments on CA systems in Southern Africa. The experiment does not provide evidence that there is an overall advantage with respect to maize yield for strategies with residue retention as a whole by comparison to conventional management. The clearest effect is the benefit of using rotation rather than intercropping as a diversification strategy with beneficial effects on reductions in pests and diseases expected. The mean maize yield over the whole trial under direct seeding with rotation was notably larger than under direct seeding with intercropping. There is a need to investigate further the potential for crop rotation as a strategy for diversification within intercropping systems. Maize yield alone is not the only consideration, given that land is taken out of production for the staple crop. Yields of alternative legume break crops require further investigation, and the place that different crops might have in providing livelihoods and food security for producers.

As would be expected, the effects of drought seasons on crop yield are substantial, and the specific contrasts among seasons confirm previous work on El Niño impacts. However, there was no evidence of difference between La Niña and neutral seasons, drought seasons have a more pronounced effect than wet seasons, highlighting the importance of adaptation to drought and heat limitations for long-term food security.

The effects of season and agronomic system are not simply additive, but interact. In the Chitedze experiment, the most pronounced component of the interaction among those we examined was the effect of El Niño conditions in reducing the benefits of rotations. This underlines the complexity of the joint effects of components of CA systems. If drought effects are very strong, then the benefit of rotation is reduced. The interaction points to the ongoing, continued risk from extreme weather conditions.

The longitudinal analysis of this experiment provides information on the variation of treatment responses in space and time, and a trial of this length under a consistent design over a decade is a nearly unique resource to be exploited for information and statistical analysis. This might be of less interest to smallholder farmers than the treatment effects, but is of considerable importance to the experimenter, considering how to design cost-effective but sensitive trials for further research. The yields of maize in this experiment show complex variation in time and space, the variance of repeated measures within plots is substantial, but the modelling suggests that the effect is not temporally correlated, and carry-over effects at plot-within-block scale are limited. Analysis of single seasons and the longitudinal analysis both show that the amount of replication to achieve adequate power, as evidenced by this long-term trial, is substantial, and larger than in the original experiment. The results presented in Figures 4 and 5 should, we suggest, be carefully considered by workers planning comparable long-term trials elsewhere in the region. The response to adding a replicate to the experiment with respect to power is larger than the effect of extending the trial by a season (given an established experiment).

The experiment at Chitedze considered a large number of treatments in addition to the conventional control. We suggest that further experiments, as well as increasing replication, also require careful attention to the treatment structure. A factorial design would allow greater insight into how components of a CA system contribute to its overall behaviour. Power analysis, based on the random effects model from our analysis, suggested that replication requirement for factorial designs would be somewhat larger than for simple comparisons between a treatment and a control.

In a situation where available land and resources to fund trials are limited, one should therefore give careful thought to how many experimental treatments are examined and how they are

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

structured. It is better to have a focused experiment with sufficient replication of a few carefully chosen treatments than to examine many treatments in a wide-ranging but inadequately replicated trial.

Supplementary Materials. To view supplementary material for this article, please visit https://doi.org/10.1017/S0014479720000125.

Funding. This study was funded by the Natural Environment Research Council (UK Research and Innovation) as part of the Global Challenges Research Fund [grant number NE/P02095X/1].

Acknowledgements. The analysis undertaken in this study, and the contributions of RML, PN, JC, IS, KN, WN and NM were part of the GCRF-CEPHaS project supported by the Natural Environment Research Council (UK Research and Innovation) as part of the Global Challenges Research Fund [grant number NE/P02095X/1]. GCRF-CEPHaS is also providing funding for ongoing site security and instrumentation of the Chitedze experiment. We thank the Department for Agricultural Research Services (Malawi Government) which manages the Chitedze Experiment, and for supporting the contribution by IL. We thank the donors of the CGIAR Research Program on MAIZE (www.maize.org) and the Feed the Future Project Africa RISING for providing time for CT to participate in this study. As CIMMYT was the originator of this trial, funding for establishing the trial by the International Fund for Agriculture Development (IFAD) as well as funds for continuation of the trials from CIMMYT and the University of Leeds under GCRF-AFRICAP (UKRI-GCRF) is also acknowledged.

Conflict of Interest. The authors declare that they have no conflict of interest.

References

Aagaard P. (2011). The practice of conventional and conservation agriculture in east and southern Africa. Conserv. Farm. Unit Zambia, Lusaka, Zambia.

Bates D., Maechler M., Bolker B. and Walker S. (2015). Fitting linear mixed effects models using lme4. *Journal of Statistical Software* **67**, 1–48.

Bolliger A., Magid J., Amado T.J.C., Scora Neto F., Dos Santos Ribeiro M.D.F., Calegari A., Ralisch R. and De Neergaard A., (2006). Taking stock of the Brazilian 'zero-till revolution': a review of landmark research an farmers' practice. *Advances in Agronomy* **91**, 47–110.

Brys G., Hubert M. and Struyf A. (2003). A comparison of some new measures of skewness. In Dutter R., Filzmoser P., Gather U. and Rousseeuw P.J. (eds), *Developments in Robust Statistics*. Heidelberg: Physica-Verlag, pp. 98–113.

Buckland S.T., Burnham K.P. and Augustin N.H. (1997). Model selection: an integral part of inference. *Biometrics* **53**, 603–618.

Bunderson W.T., Jere Z.D., Thierfelder C., Gama M., Mwale B.M., Ng'oma S.W.D., Museka R., Paul J.M., Mbale B., Mkandawire O. and Tembo P. (2017). Implementing the principles of conservation agriculture in Malawi: crop yields and factors affecting adoption. In Kassam A., Mkomwa S. and Friedrich T. (eds), *Conservation Agriculture for Africa: Building Resilient Farming Systems in a Changing Climate*. Wallingford, UK: CABI Publishing.

Cane M.A., Eshel G. and Buckland R.W. (1994). Forecasting Zimbabwean maize yield using eastern equatorial Pacific seas surface temperature. *Nature* **370**, 204–205.

Cairns J.E., Sonder K., Zaidi P.H., Verhulst N., Mahuku G., Babu R., Nair S.K., Das B., Govaerts B., Vinayan M.T., Rashid Z., Noor J.J., Devi P., San Vicente F. and Prasanna B.M. (2012). Maize production in a changing climate: impacts, adaptation, and mitigation strategies. In Sparks D. (ed), *Advances in Agronomy*. Burlington: Academic Press, pp. 1–58.

Chinsinga B. and Chasukwa M. (2018). Narratives, climate change and agricultural policy processes in Malawi. *Africa Review* **10**, 140–156.

Diggle P.J. (1990). *Time Series Analysis: A Biostatistical Introduction*. Oxford: Oxford University Press.

Diggle P.J. and Ribeiro P.J. (2007). *Model-Based Geostatistics*. New York: Springer.

Dunnett C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* **50**, 1096–1121.

Fisher R.A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain* **33**, 503–513.

Giller K.E., Andersson J.A., Corbeels M., Kirkegaard J., Mortensen D., Erenstein O. and Vanlauwe B. (2015). Beyond conservation agriculture. *Frontiers in Plant Science* **6**, 870.

Giller K.E., Witter E., Corbeels M. and Tittonell P. (2009). Conservation agriculture and smallholder farming in Africa: the heretic's view. *Field Crops Research* **114**, 23–34.

Halekoh U. and Højsgaard S. (2014). A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models – the R package pbkrtest. *Journal of Statistical Software* **59**(9), 1–30.

Holm S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65–70.

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

Iizumi T., Luo J.-J., Challinor A.J., Sakurai G., Yokozawa M., Sakuma H., Brown M.E. and Yamagata T. (2014). Impacts of El Niño Sourthern Oscillation on the global yields of major crops. *Nature Communications* **5**, 3712. https://doi.org/10.1038/ncomms4712

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

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

Kenward M. and Roger J. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**, 983–997.

Kenward M.G. and Roger J.H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis* **53**, 2583–2595.

Kirkegaard J., Conyers M., Hunt J., Kirby C., Watt M. and Rebetzke G. (2011). Sense and nonsense in conservation agriculture: principles, pragmatism and productivity in Australian mixed farming systems. *Resilient Food Systems for a Changing World*. ACIAR, GRDC, Brisbane Convention and Exhibition Centre.

Kumwenda J.D.T., Waddington S.R., Snapp S.S., Jones R.B. and Blackie M.J. (1998). Soil fertility management in Southern Africa. In Byerlee D. and Eicher C.K. (eds), *Africa's Emerging Maize Revolution*. Colorado, USA: Lynne Rienner Publishers, p. 305.

Lawson J. (2014). *Design and Analysis of Experiments with R*. Boca Raton, FL: Chapman and Hall.

Lenth R. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. *R package version 1.1*. Available at https://CRAN.R-project.org/package=emmeans (accessed 25 June 2020).

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

Lipper L., Thornton P., Campbell B.M., Baedeker T., Braimoh A., Bwalya M., Caron P., Cattaneo A., Garrity D. and Henry K. (2014). Climate-smart agriculture for food security. *Nature Climate Change* **4**, 1068–1072.

Lobell D.B., Burke M.B., Tebaldi C., Mastrandrea M.D., Falcon W.P. and Naylor R.L. (2008). Prioritizing climate change adaptation needs for food security in 2030. *Science* **319**, 607–610.

Marongwe L.S., Kwazira K., Jenrich M., Thierfelder C., Kassam A., Friedrich T. (2011). An African success: the case of conservation agriculture in Zimbabwe. *International Journal of Agricultural Sustainability* **9**, 153–161.

McCarthy N., Lipper L. and Branca G. (2011). Climate-smart agriculture: smallholder adoption and implications for climate change adaptation and mitigation. *Mitigation of Climate Change in Agriculture Working Paper* **3**, 1–37.

Mupangwa W., Thierfelder C., Cheesman S., Nyagumbo I., Muoni T., Mhlanga B., Mwila M., Sida T. and Ngwira A. (2019). Effects of maize residue and mineral nitrogen applications on maize yield in conservation-agriculture-based cropping systems of Southern Africa. *Renewable Agriculture and Food Systems* **35**, 1–14.

Ngwira A.R., Aune J.B. and Mkwinda S. (2012). On-farm evaluation of yield and economic benefit of short term maize legume intercropping systems under conservation agriculture in Malawi. *Field Crops Research* **132**, 149–157.

Ngwira A.R., Aune J.B. and Thierfelder C. (2014). DSSAT modelling of conservation agriculture maize response to climate change in Malawi. *Soil and Tillage Research* **143**, 85–94.

Ngwira A.R., Thierfelder C. and Lambert D.M. (2013). Conservation agriculture systems for Malawian smallholder farmers: long-term effects on crop productivity, profitability and soil quality. *Renewable Agriculture and Food Systems* **28**, 350–363.

Rawlins B.G., Lark R.M., O'Donnell K.E., Tye A. and Lister T.R. (2005). The assessment of point and diffuse soil pollution from an urban geochemical survey of Sheffield, England. *Soil Use and Management* **21**, 353–362.

Rudolph S., Wonglee C., Lark R.M., Marchant B.P., Garré S, van der Kruk J., Kongkaew T., Vereecken H. and Weihermüller L. (2016). Soil apparent conductivity measurements for planning and analysis of agricultural experiments: a case study from Western-Thailand. *Geoderma* **267**, 220–229.

Setimela P., Gasura E., Thierfelder C., Zaman-Allah M., Cairns J.E. and Boddupalli P.M. (2018). When the going gets tough: performance of stress tolerant maize during the 2015/16 (El Niño) and 2016/17 (La Niña) season in southern Africa. *Agriculture, Ecosystems & Environment* **268**, 79–89.

Singman H., Bolker B., Westfall J. and Aust F. (2018). afex: analysis of factorial experiments. *R package version 0.19-1*. Available at https://CRAN.R-project.org/package=afex (accessed 25 June 2020).

Smale M. (1995). 'Maize is life': Malawi's delayed green revolution. *World Development* **23**, 819–831.

Steward P.R., Dougill A.J., Thierfelder C., Pittelkow C.M., Stringer L.C., Kudzala M. and Shackelford G.E. (2018). The adaptive capacity of maize-based conservation agriculture systems to climate stress in tropical and subtropical environments: a meta-regression of yields. *Agriculture, Ecosystems & Environment* **251**, 194–202.

Steward P.R., Thierfelder C., Dougill A.J. and Ligowe I., (2019). Conservation agriculture enhances resistance of maize to climate stress in a Malawian medium-term trial. *Agriculture, Ecosystems & Environment* **277**, 95–104.

Tesfaye K., Gbegbelegbe S., Cairns J.E., Shiferaw B., Prasanna B.M., Sonder K., Boote K., Makumbi D. and Robertson R. (2015). Maize systems under climate change in sub-Saharan Africa: potential impacts on production and food security. *International Journal of Climate Change Strategies and Management* **7**, 247–271.

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125

Thierfelder C. and Wall P.C. (2009). Effects of conservation agriculture techniques on infiltration and soil water content in Zambia and Zimbabwe. *Soil and Tillage Research* **105**, 217–227.

Thierfelder C., Baudron F., Setimela P., Nyagumbo I., Mupangwa W., Mhlanga B., Lee N. and Gérard B. (2018). Complementary practices supporting conservation agriculture in southern Africa. A review. *Agronomy for Sustainable Development* **38**, 16.

Thierfelder C., Cheesman S. and Rusinamhodzi L. (2012). A comparative analysis of conservation agriculture systems: benefits and challenges of rotations and intercropping in Zimbabwe. *Field Crops Research* **137**, 237–250.

Thierfelder C., Cheesman S. and Rusinamhodzi L. (2013). Benefits and challenges of crop rotations in maize-based conservation agriculture (CA) cropping systems of southern Africa. *International Journal of Agricultural Sustainability* **11**, 108–124.

Thierfelder C., Chivenge P., Mupangwa W., Rosenstock T.S., Lamanna C. and 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.

Thierfelder C., Matemba-Mutasa R., Bunderson W.T., Mutenje M., Nyagumbo I. and Mupangwa W. (2016). Evaluating manual conservation agriculture systems in southern Africa. *Agriculture, Ecosystems & Environment* **222**, 112–124.

Thierfelder C., Rusinamhodzi L., Ngwira A.R., Mupangwa W., Nyagumbo I., Kassie G.T. and Cairns J.E. (2015). Conservation agriculture in Southern Africa: advances in knowledge. *Renewable Agriculture and Food Systems* **30**, 328–348.

Thierfelder C. and Wall P.C. (2009). Effects of conservation agriculture techniques on infiltration and soil water content in Zambia and Zimbabwe. *Soil and Tillage Research* **105**, 217–227.

Thierfelder C. and 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.

Tukey R.W. (1977). *Exploratory Data Analysis*. Reading, Mass: Addison Wesley.

Webster R. and Lark R.M. (2019). Analysis of variance in soil research: examining the assumptions. *European Journal of Soil Science* **70**, 990–1000. https://doi.org/10.1111/ejss.12804

Webster R. and Oliver M.A. (2007). *Geostatistics for Environmental Scientists*. Chichester: John Wiley & Sons.

WRB (1998). *World Reference Base on Soils*. Rome, Italy: FAO-ISRIC.

Cite this article: Lark RM, Ligowe IS, Thierfelder C, Magwero N, Namaona W, Njira K, Sandram I, Chimungu JG, and Nalivata PC (2020). Longitudinal analysis of a long-term conservation agriculture experiment in Malawi and lessons for future experimental design. *Experimental Agriculture* **56**, 506–527. https://doi.org/10.1017/S0014479720000125

Downloaded from https://www.cambridge.org/core. IP address: 201.141.164.60, on 31 Aug 2020 at 15:00:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0014479720000125