PREDICTING THE EFFECTS OF AGRICULTURAL LAND MANAGEMENT CHANGE ON SOIL QUALITY AND PRODUCTIVITY

Barry McDonald*, Neville Fowkes† and Penny Bilton

Abstract

MISG explored the relationships between eight important measures of soil quality and several predictor variables thought to influence soil quality. One measure, the penetration resistance of soil, was examined in a regression model for its relationship with soil order and soil moisture. The other quality measures were: bulk density in the top 15 cm of soil; carbon percentage in the top 15 cm; the biologically available carbon; phosphorus availability; and the size distribution of soil aggregates as represented by the mean weight, the percentage of very small soil aggregates (<0.85 mm) and the percentage of very large soil aggregates (>9.5mm). These last seven were examined for their inter-relationships and as response variables in a sequence of regression models. The models explored the effect of crops grown, and tillage methods used, at each farm site on soil quality. The impacts of crops and tillage were represented by indices, and at least one index was significant for six out of the seven quality measures studied. It was however found that background factors, including soil order and texture, land use and region, need to be taken into account when assessing the effect of the indices. As well as describing these regression models, this paper outlines some investigations made into: how the crop and tillage indices might be improved; how data on climatic and other variables might explain some regional differences; how interactions among background factors and indices might be assessed – in the absence of balanced data – and incorporated in the model; and how results might be presented.

*Institute of Information and Mathematical Sciences, Massey University Auckland, Albany, Private Bag 102 904, North Shore Mail Centre, New Zealand. Email: b.mcdonald@massey.ac.nz

†University of Western Australia, Western Australia, Australia.

1. Introduction

Much of New Zealand's income derives directly from agriculture, and for decades the New Zealand public, and agricultural industries, have invested in scientific research to enhance the productivity and improve the quality of agricultural production systems. One aspect of this is ongoing research into the health or quality of the soil and its importance to sustaining high levels of agricultural production.

Soil quality can be defined as the fitness of soil for a specific land use. The farm management practices employed under a given land use, including the type of crops grown, the tillage methods used, the fertiliser and irrigation applied, and the grazing practices, can affect the quality of the soil and the resulting economic outputs and environmental impacts. The focus of the MISG problem was to examine factors that affect eight important measures of soil quality relating to both short-term productivity and long-term sustainability of the soil. Of particular interest are the two management aspects most immediately under the control of the farmers: the tillage practices and the vegetation grown.

The MISG study formed one chapter in an ongoing study programme being undertaken by researchers from Crop & Food Research and the Sustainable Soil Management Promotion Group [1]. The eventual aim of the study programme is to produce a land management index that can be used by farmers to assess the likely effect of their farming practices on the soil and future productivity. An example would be the possible benefits or otherwise of replacing conventional ploughing with minimum-tillage techniques.

Soil scientists measure soil quality in several different ways, but for MISG purposes the study was mainly restricted to seven variables, which were grouped under four headings. Soil Compaction was measured by $BD15$, the bulk density of the top 15 cm of soil. Organic Matter was represented by two variables $C\%15$ (the carbon percentage in the top 15 cm) and HWC (hot water carbon - the biologically available carbon). Phosphorus Availability in the soil was measured by Olsen's P. The size of Soil Aggregates was represented by three variables: the stability of soil aggregates $AgStabMWD$, where MWD stands for mean weight distribution; and the extremes of the aggregate side distribution (ASD) as measured by the percentage of very small soil aggregates $ASD\% <$ 0.85 mm, reflecting a risk to erosion; and the percentage of very large soil aggregates $ASD\% >$ 9.5mm, reflecting soil compaction. More details of these variables are given in [1], Appendix III, p42-44. An eighth variable, penetration resistance, is also a measure of soil fitness for agricultural use. However it was the subject of study on its own, in section 4, and not treated alongside the other seven.

Predictors of the soil measures included: the soil order (i.e. soil classification) Allophanic, Brown, Gley, Granular, Melanic, Organic, Pallic or Recent soils; soil texture silt, sand or clay; the geographic region (Auckland / Waikato, Hawke's Bay, Canterbury, or Southland) and the land use or type of farming practiced (intensive cropping, mixed cropping, vegetable production, dairy farming, conventional sheep/beef farming, or high-tech intensive beef production). All these categorical variables were considered as background factors: of some interest in their own right but of less importance to the study than the effect of land management practices.

Land management was summarised by two indices: a tillage index calculated from 10 years of management data (tillage index high for many years of intensive cultivation, zero for undisturbed grass) and a Crop index reflecting the crop types grown. The actual formulas one should use to calculate the tillage and crop indices was a matter for discussion, and was considered during MISG (see section 3). However for most analysis considered here the index values were taken as given, based on formulas suggested by expert opinion and past experimental work [1].

Data for the MISG project was provided by the Sustainable Soil Management Promotion Group. Four members of the group were in attendance at MISG. The MISG contributed to the analysis of the soil data in several ways.

Exploratory data analysis was used to investigate inter-relationships between the soil measures and to conclude some variables should be analyzed on a transformed scale. Regression models were fitted to each soil measure in terms of the categorical variables land use (farm type), soil order, soil texture and geographic region. Results of these analyses are discussed in section 2.

Different aspects of the regression models were then examined in detail, to investigate potential improvements. First, the Crop and Tillage indices are based on empirical weightings of both the type of activity and how long ago the activity occurred. Some exploration was made of a method for better determining the weights. Second, significant regional differences were found in the soil measures even after adjusting for landuse, soil order and soil texture. The LENZ database (Land Environment of New Zealand) was interrogated to identify possible explanations for the regional differences. Third, there appear to be interactions between the effects of region, land use, soil order and soil texture on the soil measures, for example the effect of land use on the soil measures may

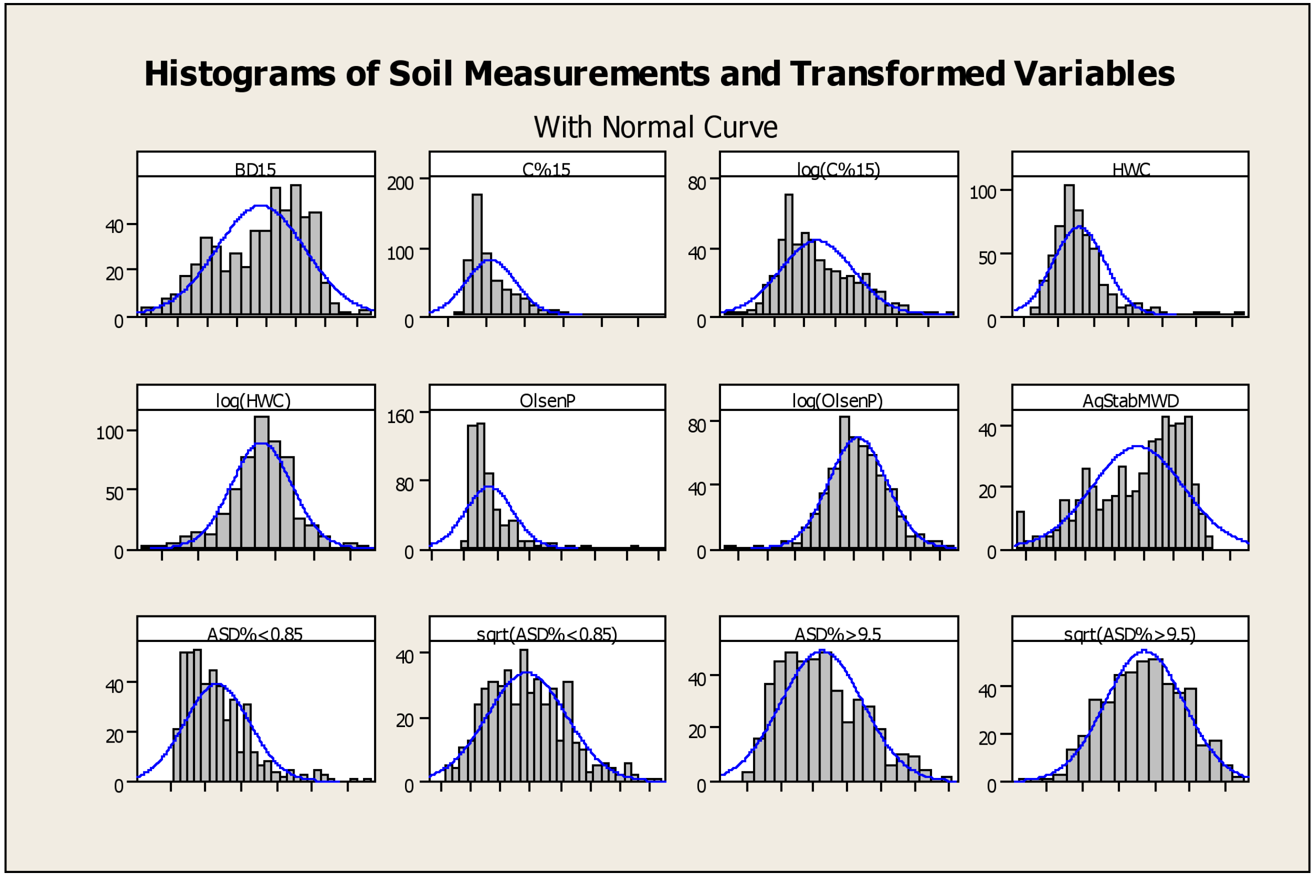

Figure 1. Histograms of soil measurements and transformed variables.

Table 1. Tally for Discrete Variables: SoilOrder, Texture, IfIndex, Landuse, Region

SoilOrder	Count	Percent
Allophanic	114	21.51
Brown	89	16.79
Gley	73	13.77
Granular	49	9.25
Melanic	5	0.94
Organic	14	2.64
Pallic	125	23.58
Recent	61	11.51
N=	530	

Texture	Count	Percent
Clay	89	16.79
Sandy	87	16.42
Silt	354	66.79
N=	530	

IfIndex	Count	Percent
No	177	33.40
Yes	353	66.60
N=	530	

Landuse	Count	Percent
Dairy	147	27.74
Intensive cropping	106	20.00
Mixed cropping	82	15.47
Sheep/beef	72	13.58
Techno Systems	18	3.40
Vegetable production	105	19.81
N=	530	

Region	Count	Percent
AucklandWaikato	172	32.45
Canterbury	164	30.94
Hawkes Bay	101	19.06
Southland	93	17.55
N=	530	

not be the same for all soil textures. These interactions were examined, using a somewhat unusual approach since not all cross-classifications of the categorical variables are available. Fourthly, suggestions were given as to how one could better express the data in a regression model for computation in Excel with new farmers' data, and how to present the results in a way easy to communicate to farmers. These four potential improvements to the analysis are discussed in section 3.

As mentioned above, an additional question concerned penetration resistance, a measure of soil compaction and therefore the difficulty roots have in entering the soil. The MISG group were asked to consider methods of correcting the penetration resistance data for differences in soil moisture. A simple model was proposed and some results are shown in section 4.

2. Regression Results

Histograms of the seven soil quality measures are shown in Figure 1, along with transformed versions. Although some of the transformed versions are roughly normal distributed, this was not the purpose of transformation. Indeed normality would be somewhat of a coincidence, since the sites examined in the study are not a random sample, and vary greatly in characteristics. Rather, the purpose of transformation was to reduce the effects of outliers on the regression modeling. The chosen transformations were to use log($C\%15$), log(HWC), log(OlsenP), sqrt($ASD\% < 0.85$) and sqrt($ASD\% > 9.5$). Alternatively one could use log(1+ '$ASD\% < 0.85$') and log(1+ '$ASD\% > 9.5$'), the '1+' being added to avoid problems at 0. However the log transformation was a problem for '$ASD\% > 9.5$' as it gave rise to extreme outliers on the left. A square root transformation is not uncommon for percentage data and avoids problems at zero [2]. It was decided that $BD15$ and $AgStabMWD$ would not be transformed as outliers were less of a problem.

The data consisted of observations from 530 sites. Figures for the various categorical variables are given in Table 1, along with an indicator ($IfIndex$) for which sites had crop index / tillage index information. The latter was available for two-thirds of the sites. Note in particular that there were very few Melanic and Organic soils, while Pallic and Allophanic soils were common. There were also few Techno Systems landuses, which were all in the Hawkes Bay. These limited numbers can affect significance in the analyses, especially for Melanic soils. Also no Techno Systems sites had crop / tillage index information so this category is left out of analyses using these indices.

Figure 2 shows scatterplots with lowess smoothers of the seven soil response measures, and their relationship to the Crop and sqrt(Tillage) indices. 'Lowess' stands for stands for LOcally-WEighted Scatterplot Smoother, and provides a convenient visualization of underlying linearity or curvature in the bivariate relationship. The corresponding Correlation and p-values are given in Table 2. Note that Tillage, and more strongly sqrt(Tillage), is significantly negatively related to Crop Index. $AgStabMWD$, log(HWC) and log($C\%15$) strongly increased with increasing Crop Index and with decreasing sqrt(Tillage) (again the correlation was stronger than using Tillage Index). The prevalence of very small aggregates, represented by sqrt($ASD\% < 0.85$), and the log(OlsenP) decrease significantly but weakly with higher Crop Index and lower Tillage. There is no overall relationship between the Crop and Tillage indices and $BD15$ and sqrt($ASD\% > 9.5$).

The results of the regression analyses are summarised in Table 3. The actual implications for each response variable are discussed one at a time in what follows, using more detailed regression output.

BD15 (Bulk Density at 15 cm) The histogram for $BD15$ (Figure 1) showed the distribution is notably bimodal with the main peak around 1.2 and the lesser peak (around 0.75). The latter peak corresponds mainly to the Allophanic soils of Auckland/Waikato. The model summaries in Table 3(a) show $BD15$ varied significantly with soil order, texture and region, and with landuse (but not when attention was restricted to crops only). The goodness of prediction (as measured by the coefficient of determination, R^2) was higher when the data was restricted

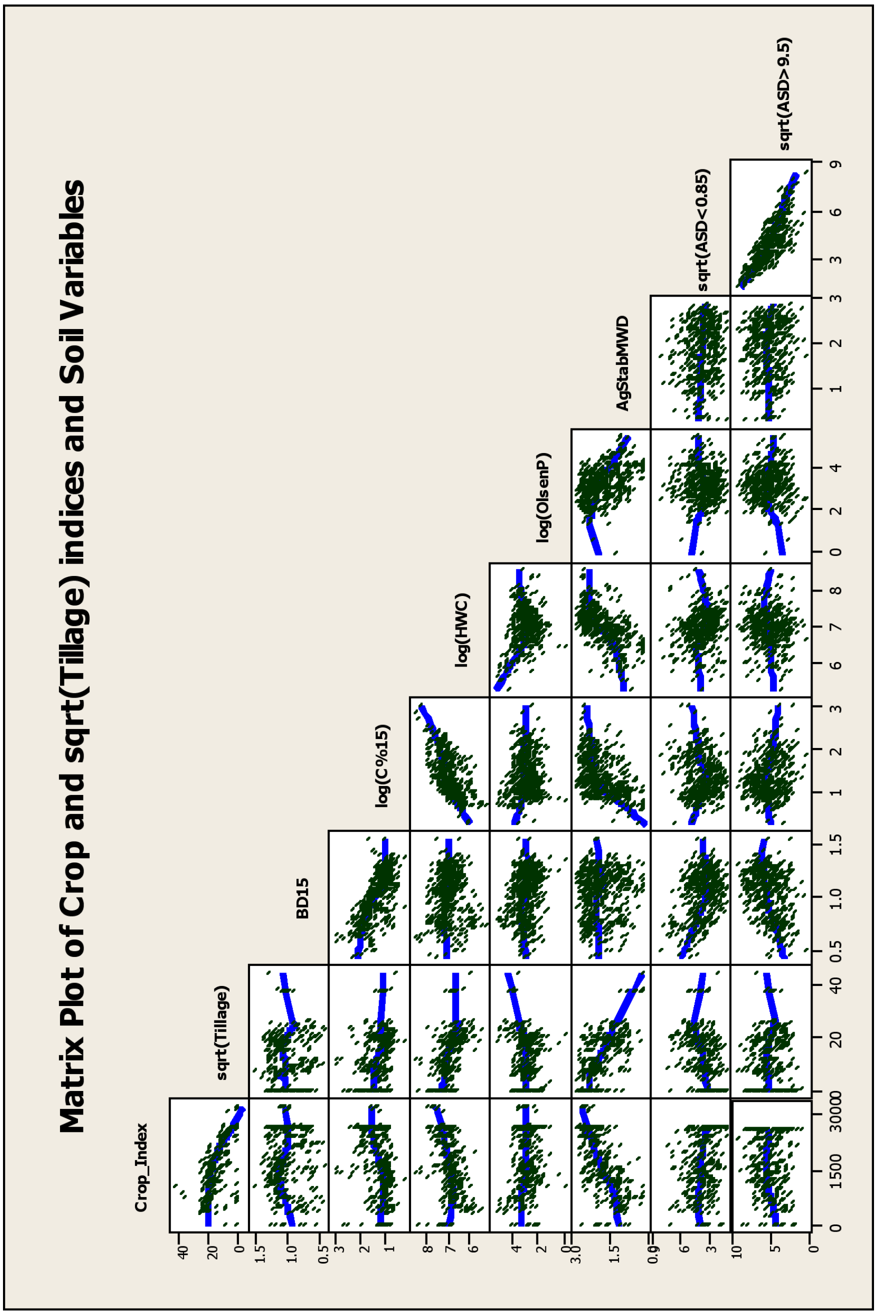

Figure 2. Matrix plots with lowess smoother

Table 2 Correlations: Crop_Index, sqrt(Tillage, BD15, log(C%15), log(HWC), ...

	Crop_Index	sqrt(Tillage)	BD15	log(C%15)	log(HWC)	log(OlsenP)	AgStabMWD	sqrt(ASD%<0.85)
sqrt(Tillage)	-0.626							
	0.000							
BD15	-0.068	0.001						
	0.203	0.979						
log(C%15)	0.335	-0.322	-0.697					
	0.000	0.000	0.000					
log(HWC)	0.449	-0.525	-0.097	0.637				
	0.000	0.000	0.026	0.000				
log(OlsenP)	-0.192	0.252	-0.243	-0.048	-0.184			
	0.000	0.000	0.000	0.272	0.000			
AgStabMWD	0.692	-0.667	-0.094	0.478	0.572	-0.388		
	0.000	0.000	0.032	0.000	0.000	0.000		
sqrt(ASD%<0.85)	-0.243	0.287	-0.357	0.127	-0.209	0.051	-0.126	
	0.000	0.000	0.000	0.007	0.000	0.281	0.007	
sqrt(ASD%>9.5)	0.072	-0.050	0.324	-0.195	0.096	-0.024	-0.025	-0.787
	0.211	0.384	0.000	0.000	0.041	0.607	0.588	0.000

Cell Contents: Pearson correlation
P-Value

Table 3 Model Summaries for Soil Quality Measures

Table 3 (a) Model Summaries for BD15

All Data Variable	(n=528) P	All Data Variable	(n=351) P	Crops Only Variable	(n=293) P	Crops Only Variable	(n=200) P
SoilOrder	0.000	SoilOrder	0.000	SoilOrder	0.000	SoilOrder	0.000
Texture	0.000	Texture	0.001	Texture	0.007	Texture	0.006
Landuse	0.000	Landuse	0.086	Landuse	0.269	Landuse	0.733
Region	0.000	Region	0.000	Region	0.000	Region	0.000
		Crop_Index	0.201			Crop_Index	0.781
		sqrt(Tillage)	0.086			sqrt(Tillage)	0.041
R^2	71.6 %	R^2	72.7 %	R^2	77.8 %	R^2	78.9 %

Table 3 (b) Model Summaries for log(C%15)

All Data Variable	(n=528) P	All Data Variable	(n=351) P	Crops Only Variable	(n=292) P	Crops Only Variable	(n=199) P
SoilOrder	0.000	SoilOrder	0.000	SoilOrder	0.000	SoilOrder	0.000
Texture	0.000	Texture	0.000	Texture	0.007	Texture	0.003
Landuse	0.000	Landuse	0.028	Landuse	0.269	Landuse	0.020
Region	0.000	Region	0.000	Region	0.000	Region	0.000
		Crop_Index	0.000			Crop_Index	0.003
		sqrt(Tillage)	0.714			sqrt(Tillage)	0.671
R^2	67.5 %	R^2	68.5 %	R^2	69.7 %	R^2	71.4 %

Table 3 (c) Model Summaries for log(HWC)

All Data Variable	(n=529) P	All Data Variable	(n=352) P	Crops Only Variable	(n=292) P	Crops Only Variable	(n=199) P
SoilOrder	0.000	SoilOrder	0.000	SoilOrder	0.000	SoilOrder	0.000
Texture	0.000	Texture	0.000	Texture	0.000	Texture	0.021
Landuse	0.000	Landuse	0.000	Landuse	0.001	Landuse	0.172
Region	0.012	Region	0.000	Region	0.006	Region	0.001
		Crop_Index	0.000			Crop_Index	0.018
		sqrt(Tillage)	0.000			sqrt(Tillage)	0.004
R^2	57.9 %	R^2	60.4 %	R^2	56.0 %	R^2	52.0 %

Table 3 (d) Model Summaries for log(OlsenP)

All Data Variable	(n=528) P	All Data Variable	(n=352) P	Crops Only Variable	(n=292) P	Crops Only Variable	(n=199) P
SoilOrder	0.000	SoilOrder	0.002	SoilOrder	0.001	SoilOrder	0.107
Texture	0.000	Texture	0.023	Texture	0.001	Texture	0.028
Landuse	0.000	Landuse	0.000	Landuse	0.000	Landuse	0.000
Region	0.000	Region	0.000	Region	0.002	Region	0.008
		Crop_Index	0.009			Crop_Index	0.137
		sqrt(Tillage)	0.064			sqrt(Tillage)	0.835
R^2	46.7 %	R^2	42.5 %	R^2	59.1 %	R^2	53.5 %

Table 3 (e) Model Summaries for AgStabMWD

All Data	(n=527)	All Data	(n=351)	Crops Only	(n=293)	Crops Only	(n=199)
Variable	P	Variable	P	Variable	P	Variable	P
SoilOrder	0.000	SoilOrder	0.061	SoilOrder	0.000	SoilOrder	0.027
Texture	0.000	Texture	0.001	Texture	0.000	Texture	0.009
Landuse	0.000	Landuse	0.026	Landuse	0.000	Landuse	0.009
Region	0.000	Region	0.003	Region	0.000	Region	0.057
		Crop_Index	0.000			Crop_Index	0.000
		sqrt(Tillage)	0.000			sqrt(Tillage)	0.047
R^2	60.9 %	R^2	63.9 %	R^2	50.7 %	R^2	55.1 %

Table 3 (f) Model Summaries for sqrt(ASD%<0.85)

All Data	(n=457)	All Data	(n=304)	Crops Only	(n=293)	Crops Only	(n=200)
Variable	P	Variable	P	Variable	P	Variable	P
SoilOrder	0.000	SoilOrder	0.000	SoilOrder	0.000	SoilOrder	0.012
Texture	0.005	Texture	0.013	Texture	0.650	Texture	0.286
Landuse	0.000	Landuse	0.190	Landuse	0.482	Landuse	0.496
Region	0.000	Region	0.000	Region	0.000	Region	0.000
		Crop_Index	0.051			Crop_Index	0.632
		sqrt(Tillage)	0.003			sqrt(Tillage)	0.011
R^2	35.8 %	R^2	41.8 %	R^2	31.7 %	R^2	35.1 %

Table 3 (g) Model Summaries for sqrt(ASD%>9.5)

All Data	(n=457)	All Data	(n=304)	Crops Only	(n=293)	Crops Only	(n=200)
Variable	P	Variable	P	Variable	P	Variable	P
SoilOrder	0.000	SoilOrder	0.000	SoilOrder	0.000	SoilOrder	0.003
Texture	0.080	Texture	0.172	Texture	0.883	Texture	0.311
Landuse	0.011	Landuse	0.147	Landuse	0.541	Landuse	0.329
Region	0.000	Region	0.000	Region	0.000	Region	0.000
		Crop_Index	0.227			Crop_Index	0.435
		sqrt(Tillage)	0.175			sqrt(Tillage)	0.036
R^2	20.4 %	R^2	27.6 %	R^2	23.1 %	R^2	32.3 %

to crop sites (Intensive Cropping, Vegetable Production or Mixed Cropping). If the rise in R^2 had been small then there would be some question as to whether it was simply a random change due to the much smaller sample size for crop-only sites. However in the case of $BD15$ the rise in R^2 is about 6%, which does suggest that the model for $BD15$ is indeed fitting better for crop sites. For pastoral landuse sites the R^2 is around 64.5%.

There did not appear to be any relationship with Crop Index, but a weak relationship with Tillage Index. The effect of Tillage was slightly stronger when expressed on a square root scale, in most models but especially for the modelling of HWC. The table therefore shows the square root of the Tillage Index for all models. A possible interpretation of such a transformation is that there is a law of diminishing returns in the effect of tillage index on the soils. Be that as it may, the main benefit of using sqrt(Tillage) is to reduce the influence of certain very high tillage sites on the regression. Using a shifted logarithmic transformation log(C + Tillage Index) might achieve the same purpose but then some arbitrary $C > 0$ would have to be chosen for those sites with no tillage, in order to avoid turning those points into high-influence points instead.

In the regression equations for $BD15$ (Figure 3) the control level (intercept) represents Canterbury Pallic Silt soils with Sheep/beef grazing. This same control level is used in all regressions, and each significant regression coefficient implies a significant difference to the corresponding control category. Thus Allophanic and Organic soils had significantly lower mean $BD15$ than the control soil order (Pallic) while the other soil orders did not differ significantly from Pallic in $BD15$. Clay soils had significantly higher mean $BD15$ than the control texture (Silt) but Sandy soils did not differ significantly from Silt for $BD15$. Soils used for vegetable production had significantly lower $BD15$ than sites where Sheep/beef landuse occurred, but no other landuse gave a mean $BD15$ significantly different to that of Sheef/beef farming. Sites from the other three regions all had significantly lower mean $BD15$ than sites from Canterbury. When Crop and Tillage indices were taken into account, Vegetable production became non-significant. This may be because the tillage index "stole its thunder" (high tillage also being related to lower $BD15$) or because the loss of 170 data points impacted disproportionately on this category of landuse. Note that Crop and Tillage Indices were not recorded for sites with Techno System landuse, so that category drops out of any analysis involving the indices. When attention was restricted to Crop-only sites the type of landuse (i.e. type of cropping) became unimportant.

Regression Details for BD15

```
The regression equation is
BD15 = 1.21 - 0.157 Allophanic + 0.0068 Brown - 0.0206 Gley
       - 0.0510  Granular - 0.0100 Melanic - 0.381 Organic + 0.0327 Recent
       + 0.0562 Clay - 0.0286 Sandy + 0.0265 Dairy - 0.0089 Intensive
       cropping  - 0.0079 Mixed cropping + 0.0415 Techno Systems - 0.0689
       Vegetable production  - 0.293 AucklandWaikato  - 0.0884 Hawkes Bay
       - 0.104 Southland
```

Predictor	Coef	SE Coef	T	P
Constant	1.20792	0.01841	65.61	0.000
Allophanic	-0.15732	0.02597	-6.06	0.000
Brown	0.00684	0.01713	0.40	0.690
Gley	-0.02060	0.01998	-1.03	0.303
Granular	-0.05098	0.03197	-1.59	0.111
Melanic	-0.00995	0.05891	-0.17	0.866
Organic	-0.38108	0.04092	-9.31	0.000
Recent	0.03271	0.02516	1.30	0.194
Clay	0.05617	0.01733	3.24	0.001
Sandy	-0.02858	0.01763	-1.62	0.106
Dairy	0.02650	0.01801	1.47	0.142
Intensive cropping	-0.00885	0.01910	-0.46	0.643
Mixed cropping	-0.00790	0.02010	-0.39	0.694
Techno Systems	0.04152	0.03515	1.18	0.238
Vegetable production	-0.06895	0.01964	-3.51	0.000
AucklandWaikato	-0.29334	0.02522	-11.63	0.000
Hawkes Bay	-0.08841	0.02437	-3.63	0.000
Southland	-0.10427	0.01666	-6.26	0.000

```
S = 0.120514   R-Sq = 71.6%   R-Sq(adj) = 70.6%
```

Regression Analysis for BD15 including Indices

```
The regression equation is
BD15 = 1.25 - 0.102 Allophanic + 0.0058 Brown - 0.0007 Gley - 0.0392
       Granular + 0.0004 Melanic - 0.456 Organic + 0.0144 Recent +
       0.0732 Clay - 0.0331 Sandy + 0.0378 Dairy + 0.0081 Intensive cropping
        + 0.0078 Mixed cropping - 0.0442 Vegetable production
        - 0.337 AucklandWaikato - 0.0765 Hawkes Bay - 0.113 Southland
        - 0.000018 Crop_Index - 0.00207 sqrt(Tillage)
```

Predictor	Coef	SE Coef	T	P
Constant	1.25103	0.04031	31.04	0.000
Allophanic	-0.10197	0.03585	-2.84	0.005
Brown	0.00578	0.01996	0.29	0.772
Gley	-0.00071	0.02267	-0.03	0.975
Granular	-0.03925	0.04259	-0.92	0.357
Melanic	0.00044	0.06121	0.01	0.994
Organic	-0.45639	0.06507	-7.01	0.000
Recent	0.01444	0.04159	0.35	0.729
Clay	0.07315	0.02310	3.17	0.002
Sandy	-0.03310	0.02299	-1.44	0.151
Dairy	0.03783	0.02242	1.69	0.092
Intensive cropping	0.00814	0.03238	0.25	0.802
Mixed cropping	0.00783	0.02920	0.27	0.789
Vegetable production	-0.04421	0.03858	-1.15	0.253
AucklandWaikato	-0.33724	0.03241	-10.41	0.000
Hawkes Bay	-0.07646	0.03499	-2.19	0.030
Southland	-0.11330	0.01879	-6.03	0.000
Crop_Index	-0.00001786	0.00001394	-1.28	0.201
sqrt(Tillage)	-0.002068	0.001203	-1.72	0.086

```
S = 0.120111   R-Sq = 72.8%   R-Sq(adj) = 71.3%
```

Figure 3. Regression analysis of $BD15$

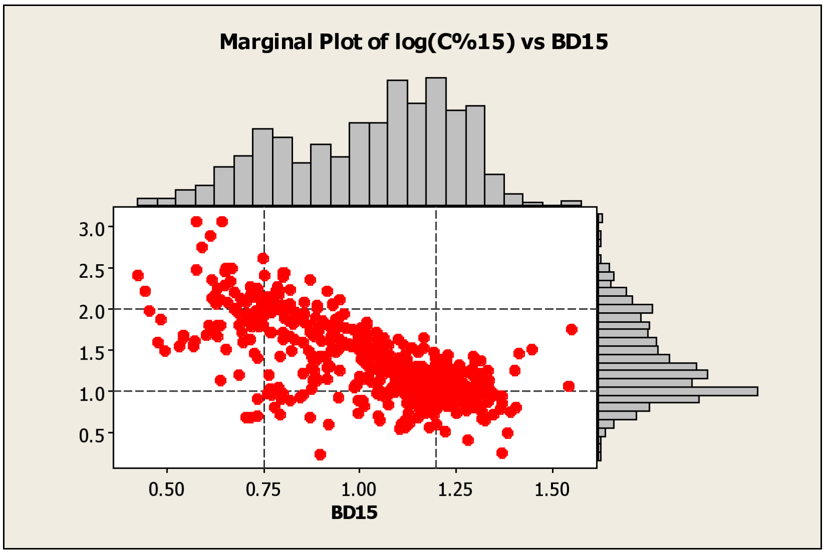

Figure 4. Marginal plot of log(C%15) vs BD15

C%15 (Carbon percentage at 15 cm) The histogram in Figure 1 showed the C%15 data are very right-skewed, but a log transformation reduced the skewness and mitigated most outliers. Again we note a slight bimodality in the log(C%15) distribution, and the marginal plot in Figure 4 shows the peaks are associated with the peaks in the BD15 distribution. The grid reference lines are at $BD15 = 0.75$ and 1.2, and $\log(C\%15) = 1.0$ and 2.0.

As for BD15, the log(C%15) is very significantly related to soil order, texture, landuse and region. The regression details (Figure 5) show Organic, Allophanic, and to a lesser extent Gley and Granular soils had significantly higher log(C%15) levels than Pallic soils, while Brown, Melanic and Recent soils did not differ significantly from Pallic. Clay soils had significantly lower log(C%15), and Sandy soils significantly higher log(C%15) than Silt. Cropping (whether intensive, vegetable or to a lesser extent mixed cropping) gave rise to significantly lower levels of log(C%15) than Sheep/beef farming while the other animal-based landuses gave similar readings to Sheep/beef. The other regions all had significantly higher log(C%15) than Canterbury. When the Crop index (and non-significant Tillage index) were taken into account the main change was that the distinctions between land uses become non-significant. Figure 6 suggests this is because Sheep/Beef and Dairy sites tended to have higher levels of Crop Index than crop/vegetable sites, and log(C%15) generally increases with Crop Index. Another change with

Regression Details for log(C%15)

```
The regression equation is
log(C%15) = 1.12 + 0.509 Allophanic + 0.0320 Brown + 0.131 Gley + 0.190
            Granular - 0.041 Melanic + 1.32 Organic - 0.0557 Recent
            - 0.150 Clay + 0.117 Sandy + 0.0451 Dairy - 0.170 Intensive
             cropping - 0.0912 Mixed cropping - 0.0102 Techno Systems
            - 0.333 Vegetable production + 0.273 AucklandWaikato
            + 0.171 Hawkes Bay + 0.259 Southland
```

Predictor	Coef	SE Coef	T	P
Constant	1.12134	0.04225	26.54	0.000
Allophanic	0.50902	0.05981	8.51	0.000
Brown	0.03204	0.03919	0.82	0.414
Gley	0.13090	0.04585	2.85	0.004
Granular	0.19044	0.07333	2.60	0.010
Melanic	-0.0414	0.1348	-0.31	0.759
Organic	1.31718	0.09380	14.04	0.000
Recent	-0.05566	0.05766	-0.97	0.335
Clay	-0.15046	0.03944	-3.81	0.000
Sandy	0.11683	0.04012	2.91	0.004
Dairy	0.04512	0.04129	1.09	0.275
Intensive cropping	-0.17032	0.04391	-3.88	0.000
Mixed cropping	-0.09118	0.04609	-1.98	0.048
Techno Systems	-0.01023	0.08059	-0.13	0.899
Vegetable production	-0.33327	0.04502	-7.40	0.000
AucklandWaikato	0.27254	0.05827	4.68	0.000
Hawkes Bay	0.17123	0.05594	3.06	0.002
Southland	0.25897	0.03807	6.80	0.000

Regression Analysis for log(C%15) including Indices

```
The regression equation is
log(C%15) = 0.755 + 0.479 Allophanic + 0.0298 Brown + 0.141 Gley
            + 0.291 Granular - 0.071 Melanic + 1.52 Organic + 0.216 Recent
            - 0.129 Clay + 0.165 Sandy + 0.0201 Dairy
            + 0.0429 Intensive cropping + 0.0543 Mixed cropping
            - 0.130 Vegetable production + 0.246 AucklandWaikato
            + 0.180 Hawkes Bay + 0.285 Southland + 0.000143 Crop_Index
            - 0.00095 sqrt(Tillage)
```

Predictor	Coef	SE Coef	T	P
Constant	0.75467	0.08645	8.73	0.000
Allophanic	0.47908	0.07750	6.18	0.000
Brown	0.02982	0.04275	0.70	0.486
Gley	0.14057	0.04867	2.89	0.004
Granular	0.29094	0.09131	3.19	0.002
Melanic	-0.0709	0.1308	-0.54	0.588
Organic	1.5235	0.1393	10.94	0.000
Recent	0.21616	0.08914	2.42	0.016
Clay	-0.12911	0.04891	-2.64	0.009
Sandy	0.16528	0.04871	3.39	0.001
Dairy	0.02012	0.04807	0.42	0.676
Intensive cropping	0.04288	0.06932	0.62	0.537
Mixed cropping	0.05425	0.06251	0.87	0.386
Vegetable production	-0.13026	0.08290	-1.57	0.117
AucklandWaikato	0.24591	0.07062	3.48	0.001
Hawkes Bay	0.17979	0.07509	2.39	0.017
Southland	0.28492	0.04014	7.10	0.000
Crop_Index	0.00014283	0.00002980	4.79	0.000
sqrt(Tillage)	-0.000953	0.002596	-0.37	0.714

S = 0.257611   R-Sq = 68.5%   R-Sq(adj) = 66.8%

Figure 5. Regression analysis for log(C%15)

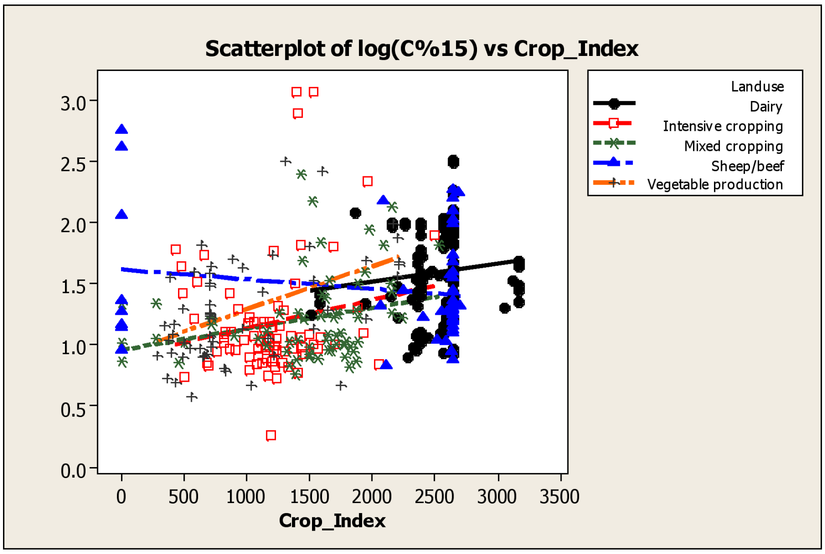

Figure 6. Scatterplot of log(C%15) vs Crop_Index

including the indexes is that Recent soils now show significantly higher levels of log(C%15) than Pallic soils. This is possibly just an adjustment effect: recent soils were generally quite low in Crop index, and so were predicted to be low in log(C%15) so a significant upward adjustment was needed to return these soils to their correct mean. Restricting attention to Crop-only data did not change the R^2s much, so there may be little to gain by using a separate model for these data.

HWC (Hot Water Carbon) The histograms in Figure 1 showed the HWC data are very right-skewed, but taking logarithms transformed the data to something like a normal distribution with a few extra outliers. The log(HWC) is very significantly related to soil order, texture, land use and region. In the regression details (Figure 7) without the indices, Granular and Melanic soils had significantly lower levels of log(HWC) than Pallic soils, while Organic soils had much higher levels. Allophanic, Brown, Gley and Recent soils did not differ significantly from Pallic soils. Clay-textured soils had significantly lower log(HWC) than Silt, and Sandy soils marginally more. Dairy landuse was associated with significantly higher log(HWC) than Sheep/beef production but there was no difference between the latter and Techno Systems landuse. All types of cropping gave rise to significantly lower log(HWC) levels. Auckland/Waikato and Hawkes Bay sites were not significantly

Regression Details for log(HWC)

```
The regression equation is
log(HWC) = 7.21 - 0.0500 Allophanic - 0.0648 Brown - 0.0217 Gley
           - 0.296 Granular - 0.469 Melanic + 1.16 Organic - 0.0379 Recent
           - 0.262 Clay + 0.0828 Sandy + 0.179 Dairy - 0.435 Intensive
           cropping - 0.251 Mixed cropping + 0.0175 Techno Systems
           - 0.415 Vegetable production - 0.0054 AucklandWaikato
           - 0.0464 Hawkes Bay + 0.123 Southland
```

Predictor	Coef	SE Coef	T	P
Constant	7.21150	0.04791	150.52	0.000
Allophanic	-0.05005	0.06791	-0.74	0.461
Brown	-0.06479	0.04446	-1.46	0.146
Gley	-0.02167	0.05204	-0.42	0.677
Granular	-0.29602	0.08330	-3.55	0.000
Melanic	-0.4695	0.1532	-3.06	0.002
Organic	1.1620	0.1066	10.90	0.000
Recent	-0.03794	0.06550	-0.58	0.563
Clay	-0.26241	0.04481	-5.86	0.000
Sandy	0.08284	0.04559	1.82	0.070
Dairy	0.17912	0.04684	3.82	0.000
Intensive cropping	-0.43497	0.04989	-8.72	0.000
Mixed cropping	-0.25098	0.05237	-4.79	0.000
Techno Systems	0.01748	0.09157	0.19	0.849
Vegetable production	-0.41451	0.05115	-8.10	0.000
AucklandWaikato	-0.00541	0.06620	-0.08	0.935
Hawkes Bay	-0.04643	0.06353	-0.73	0.465
Southland	0.12326	0.04322	2.85	0.005

S = 0.313968   R-Sq = 57.9%   R-Sq(adj) = 56.5%

Regression Analysis for log(HWC) including Indices

```
The regression equation is
log(HWC) = 6.88 + 0.0373 Allophanic - 0.0680 Brown + 0.0167 Gley
           - 0.103 Granular - 0.497 Melanic + 1.17 Organic + 0.164 Recent
           - 0.196 Clay + 0.0872 Sandy + 0.244 Dairy - 0.0372 Intensive
           cropping + 0.0379 Mixed cropping - 0.0116 Vegetable production
           - 0.124 AucklandWaikato + 0.130 Hawkes Bay + 0.155 Southland
           + 0.000124 Crop_Index - 0.0122 sqrt(Tillage)
```

Predictor	Coef	SE Coef	T	P
Constant	6.87847	0.09703	70.89	0.000
Allophanic	0.03730	0.08696	0.43	0.668
Brown	-0.06800	0.04789	-1.42	0.157
Gley	0.01674	0.05455	0.31	0.759
Granular	-0.1029	0.1025	-1.00	0.316
Melanic	-0.4967	0.1469	-3.38	0.001
Organic	1.1684	0.1565	7.47	0.000
Recent	0.1636	0.1001	1.63	0.103
Clay	-0.19628	0.05494	-3.57	0.000
Sandy	0.08725	0.05471	1.59	0.112
Dairy	0.24433	0.05380	4.54	0.000
Intensive cropping	-0.03716	0.07786	-0.48	0.634
Mixed cropping	0.03787	0.07022	0.54	0.590
Vegetable production	-0.01156	0.09301	-0.12	0.901
AucklandWaikato	-0.12384	0.07932	-1.56	0.119
Hawkes Bay	0.13004	0.08429	1.54	0.124
Southland	0.15512	0.04502	3.45	0.001
Crop_Index	0.00012355	0.00003347	3.69	0.000
sqrt(Tillage)	-0.012182	0.002912	-4.18	0.000

S = 0.289392   R-Sq = 60.4%   R-Sq(adj) = 58.2%

Figure 7. Regression analysis for $\log(HWC)$

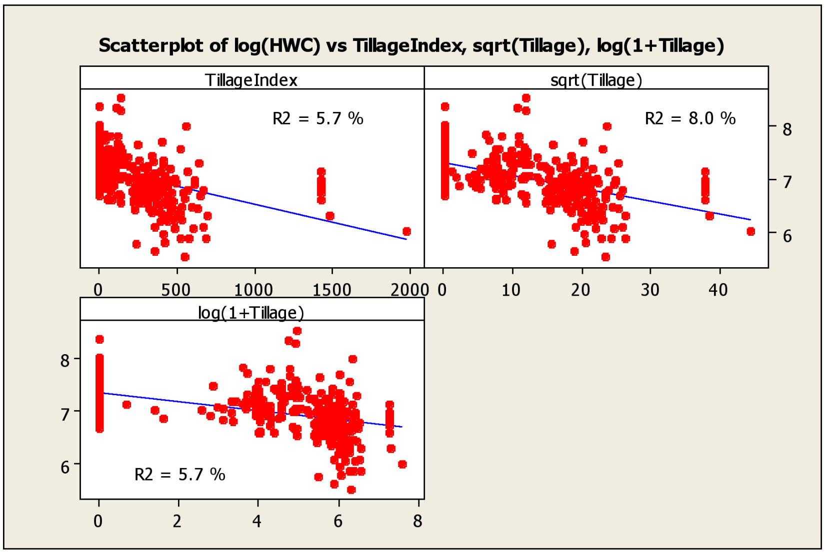

Figure 8. Scatterplot of log(HWC) vs TillageIndex, sqrt(Tillage), log(1+Tillage)

different to Canterbury sites but Southland sites had significantly higher log(HWC).

When the Crop index and sqrt(Tillage) index were taken into account these were both highly significant: log(HWC) rose with increasing Crop index and fell with increasing sqrt(Tillage) index. The main changes to the model were that Granular soils ceased to be significantly different to Pallic, and the cropping landuses ceased to be significantly different to Sheep/Beef.

Similar relationships are found if one uses the untransformed Tillage index or log(1+ Tillage index), but both these representations of the Tillage effect are subject to outliers at one end or other of the Tillage range, as shown in Figure 8. Using sqrt(Tillage) reduces the influence of outliers and tends to give slightly more significant models with slightly bigger R^2.

Restricting attention to Crop-only data gave a reduction in R^2. It is not clear why.

Olsen P The histograms in Figure 1 showed the OlsenP data must also be transformed by taking logarithms, which produces something very like a normal distribution. The log(OlsenP) is very significantly related to soil order, texture, landuse and region, but the model fit (R^2) is rather poorer than for the previous soil quality measures. Model fit

Regression Details for log(OlsenP)

```
The regression equation is
log(OlsenP) = 2.55 - 0.416 Allophanic + 0.0511 Brown - 0.150 Gley
              - 0.099 Granular - 0.469 Melanic + 0.699 Organic + 0.144 Recent
              + 0.159 Clay - 0.214 Sandy + 0.418 Dairy
              + 0.219 Intensive cropping + 0.0975 Mixed cropping
              + 0.048 Techno Systems + 1.10 Vegetable production
              + 0.690 AucklandWaikato + 0.053 Hawkes Bay + 0.206 Southland
```

Predictor	Coef	SE Coef	T	P
Constant	2.54739	0.08603	29.61	0.000
Allophanic	-0.4165	0.1219	-3.42	0.001
Brown	0.05108	0.07983	0.64	0.523
Gley	-0.15010	0.09344	-1.61	0.109
Granular	-0.0987	0.1496	-0.66	0.510
Melanic	-0.4692	0.2751	-1.71	0.089
Organic	0.6993	0.1914	3.65	0.000
Recent	0.1439	0.1176	1.22	0.222
Clay	0.15895	0.08046	1.98	0.049
Sandy	-0.21383	0.08186	-2.61	0.009
Dairy	0.41846	0.08411	4.98	0.000
Intensive cropping	0.21909	0.08959	2.45	0.015
Mixed cropping	0.09746	0.09404	1.04	0.300
Techno Systems	0.0479	0.1644	0.29	0.771
Vegetable production	1.09987	0.09186	11.97	0.000
AucklandWaikato	0.6897	0.1189	5.80	0.000
Hawkes Bay	0.0531	0.1141	0.47	0.642
Southland	0.20614	0.07761	2.66	0.008

S = 0.563793   R-Sq = 46.7%   R-Sq(adj) = 44.9%

Regression Analysis for log(OlsenP) including Indices

```
The regression equation is
log(OlsenP) = 2.97 - 0.520 Allophanic + 0.0133 Brown - 0.226 Gley
              - 0.208 Granular - 0.512 Melanic + 0.648 Organic - 0.071Recent
              + 0.071 Clay - 0.256 Sandy + 0.425 Dairy
              - 0.204 Intensive cropping - 0.172 Mixed cropping
              + 0.637 Vegetable production + 0.845 AucklandWaikato
              + 0.037 Hawkes Bay + 0.249 Southland - 0.000174 Crop_Index
              + 0.0108 sqrt(Tillage)
```

Predictor	Coef	SE Coef	T	P
Constant	2.9716	0.1929	15.40	0.000
Allophanic	-0.5204	0.1729	-3.01	0.003
Brown	0.01333	0.09523	0.14	0.889
Gley	-0.2258	0.1085	-2.08	0.038
Granular	-0.2083	0.2038	-1.02	0.307
Melanic	-0.5124	0.2921	-1.75	0.080
Organic	0.6481	0.3111	2.08	0.038
Recent	-0.0714	0.1991	-0.36	0.720
Clay	0.0708	0.1092	0.65	0.517
Sandy	-0.2563	0.1088	-2.36	0.019
Dairy	0.4247	0.1070	3.97	0.000
Intensive cropping	-0.2035	0.1548	-1.31	0.190
Mixed cropping	-0.1719	0.1396	-1.23	0.219
Vegetable production	0.6374	0.1849	3.45	0.001
AucklandWaikato	0.8448	0.1577	5.36	0.000
Hawkes Bay	0.0371	0.1676	0.22	0.825
Southland	0.24888	0.08952	2.78	0.006
Crop_Index	-0.00017398	0.00006656	-2.61	0.009
sqrt(Tillage)	0.010778	0.005791	1.86	0.064

S = 0.575409   R-Sq = 42.5%   R-Sq(adj) = 39.4%

Figure 9. Regression analysis for log(OlsenP)

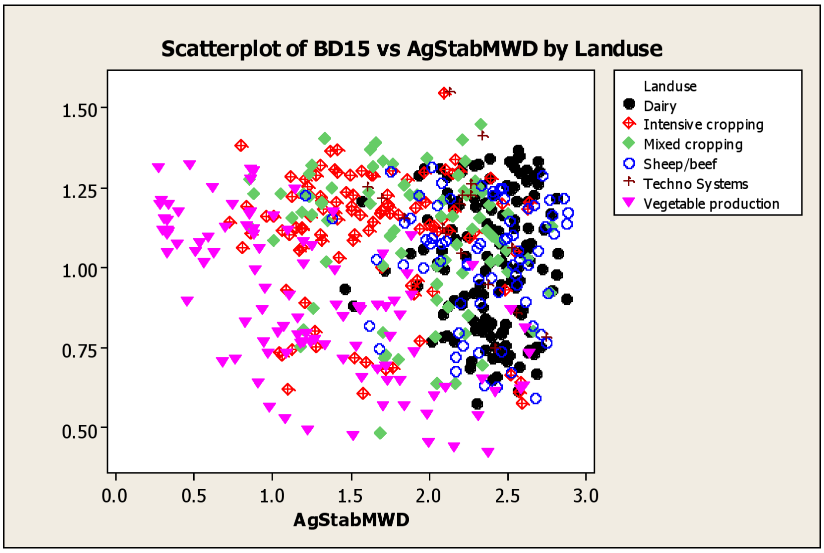

Figure 10. Scatterplot of $BD15$ vs $AgStabMWD$ by land use

(R^2) was better when the model was restricted to Crop sites. Curiously the fit was worse when restricted to sites which had crop and tillage indices, suggesting an element of sampling bias. To explore this possible bias further, consider the crop-only sites. The median OlsenP was 36.02 for the 93 crop-only sites without a Crop/Tillage index, whereas the median OlsenP was 20.83 for the 199 sites *with* the index, which is a significant difference (Mann-Whitney test, $p < 0.001$). It is left to the soil scientists to suggest an explanation for this sampling bias.

In the regression without indices (Figure 9), Organic soils had significantly higher log(OlsenP) and Allophanic soils significantly lower levels, and the other soils are much the same as Pallic soils. With more data it is possible Melanic soils would exhibit significantly lower log(OlsenP). Clay texture soils had more log(OlsenP) than Silt-textured soil, while sandy-textured soils had significantly less. Dairy, Intensive cropping and Vegetable landuses were associated with higher log(OlsenP) levels than Sheep/beef landuse. Auckland/Waikato and Southland sites had significantly higher log(OlsenP) than Canterbury, while Hawkes Bay sites were similar to Canterbury sites.

The log(OlsenP) fell significantly as the Crop Index rose, and there was an apparent (non-significant) trend for it to rise with sqrt(Tillage). With these indices in the model, the Gley soils appeared significantly lower, while Clay, and Intensive Cropping landuse became non-significant.

Regression Details for AgStabMWD

```
The regression equation is
AgStabMWD = 2.44 + 0.144 Allophanic - 0.0834 Brown - 0.0473 Gley
           + 0.201 Granular - 0.086 Melanic + 0.363 Organic - 0.231 Recent
           - 0.128 Clay + 0.164 Sandy - 0.0611 Dairy - 0.768 Intensive
           cropping  - 0.517 Mixed cropping + 0.174 Techno Systems
           - 1.04 Vegetable production - 0.129 AucklandWaikato
           - 0.399 Hawkes Bay + 0.145 Southland
```

Predictor	Coef	SE Coef	T	P
Constant	2.43805	0.06108	39.91	0.000
Allophanic	0.14379	0.08694	1.65	0.099
Brown	-0.08340	0.05688	-1.47	0.143
Gley	-0.04726	0.06645	-0.71	0.477
Granular	0.2009	0.1063	1.89	0.059
Melanic	-0.0864	0.1952	-0.44	0.658
Organic	0.3628	0.1359	2.67	0.008
Recent	-0.23147	0.08375	-2.76	0.006
Clay	-0.12835	0.05717	-2.25	0.025
Sandy	0.16389	0.05818	2.82	0.005
Dairy	-0.06105	0.05975	-1.02	0.307
Intensive cropping	-0.76816	0.06358	-12.08	0.000
Mixed cropping	-0.51666	0.06673	-7.74	0.000
Techno Systems	0.1739	0.1189	1.46	0.144
Vegetable production	-1.03631	0.06519	-15.90	0.000
AucklandWaikato	-0.12906	0.08461	-1.53	0.128
Hawkes Bay	-0.39915	0.08111	-4.92	0.000
Southland	0.14546	0.05525	2.63	0.009

```
S = 0.400059   R-Sq = 60.9%   R-Sq(adj) = 59.6%
```

Regression Analysis for AgStabMWD including Indices

```
The regression equation is
AgStabMWD = 1.84 + 0.119 Allophanic - 0.0680 Brown - 0.0115 Gley
           + 0.176 Granular - 0.078 Melanic + 0.455 Organic - 0.115 Recent
           - 0.0302 Clay + 0.242 Sandy - 0.109 Dairy - 0.219 Intensive
           cropping - 0.111 Mixed cropping - 0.362 Vegetable production
           - 0.215 AucklandWaikato - 0.249 Hawkes Bay + 0.104 Southland
           + 0.000267 Crop_Index - 0.0155 sqrt(Tillage)
```

Predictor	Coef	SE Coef	T	P
Constant	1.8414	0.1254	14.69	0.000
Allophanic	0.1192	0.1124	1.06	0.289
Brown	-0.06805	0.06209	-1.10	0.274
Gley	-0.01146	0.07048	-0.16	0.871
Granular	0.1759	0.1324	1.33	0.185
Melanic	-0.0777	0.1898	-0.41	0.682
Organic	0.4551	0.2022	2.25	0.025
Recent	-0.1150	0.1294	-0.89	0.374
Clay	-0.03022	0.07099	-0.43	0.671
Sandy	0.24225	0.07072	3.43	0.001
Dairy	-0.10896	0.06962	-1.56	0.119
Intensive cropping	-0.2194	0.1006	-2.18	0.030
Mixed cropping	-0.11062	0.09074	-1.22	0.224
Vegetable production	-0.3616	0.1202	-3.01	0.003
AucklandWaikato	-0.2153	0.1025	-2.10	0.036
Hawkes Bay	-0.2490	0.1089	-2.29	0.023
Southland	0.10411	0.05834	1.78	0.075
Crop_Index	0.00026718	0.00004324	6.18	0.000
sqrt(Tillage)	-0.015504	0.003765	-4.12	0.000

```
S = 0.373846   R-Sq = 63.9%   R-Sq(adj) = 61.9%
```

Figure 11. Regression details for $AgStabMWD$

AgStabMWD The histogram of $AgStabMWD$ (Figure 1) shows the distribution is left-skewed and apparently bimodal though this may simply be a sampling artifact. Figure 10 indicates that there is no clear association between $BD15$ and $AgStabMWD$, but most pastoral landuses are associated with very high $AgStabMWD$ and most cropping landuses with low $AgStabMWD$. Soils used for Vegetable Production is likely to have both low $BD15$ and low $AgStabMWD$.

The model summaries in Table 3(e) show that $AgStabMWD$ varied significantly with Soil order, texture and region, and land use. Also, $AgStabMWD$ rises significantly with Crop Index and falls significantly with sqrt(Tillage). In the regression equations (Figure 11) the Organic soils had significantly higher $AgStabMWD$ than Pallic soils. Recent soils had significantly lower $AgStabMWD$ than Pallic soils: although some Recent and Pallic soils had high $AgStabMWD$, out of the 22 sites with lowest $AgStabMWD$, 20 were Recent and two Pallic. Clay textured soils tended to have lower $AgStabMWD$ than silt, and sandy soils higher $AgStabMWD$. All forms of cropping were associated with lower $AgStabMWD$ than pastoral sites, and Hawkes Bay sites had significantly lower $AgStabMWD$ than Canterbury sites, and Southland sites higher $AgStabMWD$. When Crop and Tillage indices were taken into account, Recent and Clay soils ceased to be significant, as did Mixed Cropping landuse and Southland location. However, Auckland/Waikato sites showed up as significantly lower in $AgStabMWD$ than Canterbury sites. Again it is not clear whether these changes are due to the indices or simply due to the loss of 170 data points from the analysis.

sqrt(ASD%<0.85) The sqrt($ASD\% < 0.85$) is very significantly related to SoilOrder, Texture, Landuse and Region but these background variables are relatively poor predictors of the response ($R^2 = 35.8\%$).

The sqrt($ASD\% < 0.85$) is significantly negatively related to Crop Index and positively related to sqrt(Tillage), meaning that high tillage is associated with more very small aggregates and hence more propensity to erosion. However the R^2 only reaches 41.8%. The model does not fit any better when attention is focused on crop-only sites. From the regression details (Figure 12) Allophanic, Organic and (possibly) Melanic soils have higher mean sqrt($ASD\% < 0.85$) than Pallic or other soils do; as do sandy textured soils, and all sites used for cropping. So these soils are more likely to blow away or be washed away. Dairy sites have significantly lower mean sqrt($ASD\% < 0.85$). Soils in Hawkes Bay and Southland had significantly lower mean sqrt($ASD\% < 0.85$) than Canterbury but this was not true of Auckland/Waikato sites. After allowing for Crop Index and Tillage, the above differences became nonsignificant apart from the regional ones and the fact that Organic soils have considerably higher sqrt($ASD\% < 0.85$) than other soils.

sqrt(ASD%>9.5) The ASD%> 9.5 was the worst-described variable among the soil quality measures. The sqrt($ASD\% > 9.5$) was significantly related to soil order, land use and region but not to texture, and these background categories only described 20.4% of the variation in sqrt($ASD\% > 9.5$). It did not appear to be related to either the crop index or tillage index among all sites, though there was a significant negative relationship of tillage to sqrt($ASD\% > 9.5$) in the case of crop-only sites. In the regression without indices (Figure 13), Allophanic, Brown and Organic soils are less likely to have $ASD\% > 9.5$ than Pallic or other soils, while Dairy sites, Hawkes Bay sites and especially Southland sites have higher mean sqrt($ASD\% > 9.5$). After allowing for the Crop and Tillage indices, several variables changed their significance and the R^2 rises from 20.4% to 27.6%, which is a little odd since the Crop and Tillage indices were not themselves significant. Specifically, Dairy, (and additionally) Mixed Cropping and Intensive Cropping had significantly higher sqrt($ASD\% > 9.5$) than Sheep/beef sites; while Brown, Organic and (additionally) Recent soils (but not Allophanic) had lower sqrt($ASD\% > 9.5$). Perhaps some of these changes are due to the loss of 153 sites from the sample. Again, there is some suggestion that sampling bias (in terms of selecting sites with or without crop/tillage indices) may be related to the response. This time there is little difference in cropping sites, but for the 164 pastoral sites, the median $ASD\% > 9.5$ was 41.5 for the 60 sites without a crop/tillage index, as compared to 27.8 for the 104 sites with a crop/tillage index, which is almost significant (Mann-Whitney $p = 0.053$). The effects represented in the model are therefore somewhat uncertain and need to be confirmed.

Conclusion In summary the regression analysis shows that the soil quality measures (with the exception of $ASD\% > 9.5$) are significantly related to the crop and/or tillage indices. Furthermore the relationships are in the anticipated direction, with soil quality increasing with higher crop index and lower tillage. It is further shown that the soil quality measures are significantly related to soil order, texture, region and landuse, in addition to the indices. The discussion has revealed which particular levels of the categorical variables cause them to be significant. The analysis has suggested appropriate transformations to use to represent the inter-relationships among the variables. The analysis has also uncovered some limitations in the data, in that not all sites have crop and tillage data and this may potentially bias the regression results. It is hoped

Regression Details for sqrt(ASD%<0.85)

```
The regression equation is
sqrt(ASD%<0.85) = 3.46 + 1.19 Allophanic + 0.270 Brown - 0.144 Gley
                  - 0.268 Granular + 0.942 Melanic + 0.826 Organic
                  + 0.012 Recent - 0.185 Clay + 0.431 Sandy - 0.439 Dairy
                  + 0.759 Intensive cropping + 0.591 Mixed cropping
                  + 0.071 Techno Systems + 0.968 Vegetable production
                  - 0.247 AucklandWaikato -0.503 Hawkes Bay -0.987 Southland
```

Predictor	Coef	SE Coef	T	P
Constant	3.4615	0.1944	17.80	0.000
Allophanic	1.1898	0.2531	4.70	0.000
Brown	0.2704	0.1675	1.61	0.107
Gley	-0.1437	0.1968	-0.73	0.466
Granular	-0.2677	0.3067	-0.87	0.383
Melanic	0.9416	0.8086	1.16	0.245
Organic	0.8262	0.3811	2.17	0.031
Recent	0.0120	0.2500	0.05	0.962
Clay	-0.1848	0.1702	-1.09	0.278
Sandy	0.4305	0.1703	2.53	0.012
Dairy	-0.4392	0.1948	-2.26	0.025
Intensive cropping	0.7592	0.1905	3.99	0.000
Mixed cropping	0.5915	0.1997	2.96	0.003
Techno Systems	0.0714	0.3256	0.22	0.827
Vegetable production	0.9679	0.1926	5.03	0.000
AucklandWaikato	-0.2467	0.2454	-1.01	0.315
Hawkes Bay	-0.5032	0.2382	-2.11	0.035
Southland	-0.9865	0.1672	-5.90	0.000

```
S = 1.09050   R-Sq = 35.8%   R-Sq(adj) = 33.3%
```

Regression Analysis for sqrt(ASD%<0.85) including Indices

```
The regression equation is
sqrt(ASD%<0.85) = 3.93 + 0.538 Allophanic + 0.214 Brown - 0.248 Gley
                  - 0.313 Granular + 0.864 Melanic + 2.64 Organic + 0.305
Recent
                  - 0.352 Clay + 0.373 Sandy - 0.453 Dairy
                  - 0.052 Intensive cropping + 0.065 Mixed cropping
                  + 0.014 Vegetable production + 0.390 AucklandWaikato
                  - 1.11 Hawkes Bay - 0.871 Southland - 0.000240 Crop_Index
                  + 0.0325 sqrt(Tillage)
```

Predictor	Coef	SE Coef	T	P
Constant	3.9281	0.3629	10.82	0.000
Allophanic	0.5384	0.3268	1.65	0.101
Brown	0.2136	0.1841	1.16	0.247
Gley	-0.2482	0.2097	-1.18	0.238
Granular	-0.3130	0.3841	-0.81	0.416
Melanic	0.8644	0.7673	1.13	0.261
Organic	2.6398	0.5562	4.75	0.000
Recent	0.3048	0.3565	0.85	0.393
Clay	-0.3518	0.2103	-1.67	0.096
Sandy	0.3732	0.2078	1.80	0.074
Dairy	-0.4534	0.2344	-1.93	0.054
Intensive cropping	-0.0515	0.2956	-0.17	0.862
Mixed cropping	0.0647	0.2663	0.24	0.808
Vegetable production	0.0142	0.3534	0.04	0.968
AucklandWaikato	0.3897	0.2914	1.34	0.182
Hawkes Bay	-1.1128	0.3018	-3.69	0.000
Southland	-0.8715	0.1759	-4.95	0.000
Crop_Index	-0.0002405	0.0001227	-1.96	0.051
sqrt(Tillage)	0.03250	0.01076	3.02	0.003

```
S = 1.01101   R-Sq = 41.8%   R-Sq(adj) = 38.1%
```

Figure 12. Regression analysis for sqrt($ASD\% < 0.85$)

Regression Details for sqrt(ASD%>9.5)

```
The regression equation is
sqrt(ASD%>9.5) = 5.31 - 1.38 Allophanic - 0.790 Brown - 0.005 Gley
                -0.506 Granular + 0.36 Melanic - 1.17 Organic - 0.293 Recent
                + 0.200 Clay - 0.392 Sandy + 0.783 Dairy
                + 0.347 Intensive cropping + 0.425 Mixed cropping
                + 0.836 Techno Systems - 0.028 Vegetable production
                + 0.514 AucklandWaikato + 0.685 Hawkes Bay + 1.19 Southland
```

Predictor	Coef	SE Coef	T	P
Constant	5.3105	0.2688	19.76	0.000
Allophanic	-1.3817	0.3499	-3.95	0.000
Brown	-0.7905	0.2316	-3.41	0.001
Gley	-0.0051	0.2721	-0.02	0.985
Granular	-0.5058	0.4240	-1.19	0.234
Melanic	0.358	1.118	0.32	0.749
Organic	-1.1659	0.5268	-2.21	0.027
Recent	-0.2930	0.3456	-0.85	0.397
Clay	0.2005	0.2354	0.85	0.395
Sandy	-0.3920	0.2354	-1.67	0.097
Dairy	0.7832	0.2692	2.91	0.004
Intensive cropping	0.3469	0.2633	1.32	0.188
Mixed cropping	0.4251	0.2761	1.54	0.124
Techno Systems	0.8364	0.4502	1.86	0.064
Vegetable production	-0.0282	0.2662	-0.11	0.916
AucklandWaikato	0.5138	0.3393	1.51	0.131
Hawkes Bay	0.6848	0.3293	2.08	0.038
Southland	1.1873	0.2312	5.14	0.000

S = 1.50760   R-Sq = 20.4%   R-Sq(adj) = 17.3%

Regression Analysis for sqrt(ASD%>9.5) including Indices

```
The regression equation is
sqrt(ASD%>9.5) = 4.86 - 0.615 Allophanic - 0.664 Brown + 0.173 Gley
                 - 0.415 Granular + 0.58 Melanic - 2.94 Organic - 1.49
      Recent
                 + 0.260 Clay - 0.371 Sandy + 0.699 Dairy
                 + 0.877 Intensive cropping + 0.779 Mixed cropping
                 + 0.703 Vegetable production - 0.243 AucklandWaikato
                 + 1.75 Hawkes Bay + 1.08 Southland + 0.000206 Crop_Index
                 - 0.0203 sqrt(Tillage)
```

Predictor	Coef	SE Coef	T	P
Constant	4.8582	0.5020	9.68	0.000
Allophanic	-0.6154	0.4520	-1.36	0.174
Brown	-0.6640	0.2546	-2.61	0.010
Gley	0.1731	0.2900	0.60	0.551
Granular	-0.4152	0.5313	-0.78	0.435
Melanic	0.579	1.061	0.55	0.586
Organic	-2.9392	0.7692	-3.82	0.000
Recent	-1.4924	0.4930	-3.03	0.003
Clay	0.2602	0.2909	0.89	0.372
Sandy	-0.3713	0.2874	-1.29	0.197
Dairy	0.6985	0.3242	2.15	0.032
Intensive cropping	0.8772	0.4088	2.15	0.033
Mixed cropping	0.7789	0.3684	2.11	0.035
Vegetable production	0.7030	0.4888	1.44	0.152
AucklandWaikato	-0.2426	0.4031	-0.60	0.548
Hawkes Bay	1.7472	0.4174	4.19	0.000
Southland	1.0841	0.2433	4.46	0.000
Crop_Index	0.0002056	0.0001697	1.21	0.227
sqrt(Tillage)	-0.02025	0.01489	-1.36	0.175

S = 1.39834   R-Sq = 27.6%   R-Sq(adj) = 23.0%

Figure 13. Regression analysis for sqrt($ASD\% > 9.5$)

that this incompleteness in the index data may in time be remedied, to ensure the validity of inferences drawn.

In conclusion, the results to this point hold promise towards development of a useful land management index, but much work remains to be done. The next section suggests some directions for improvement of the model.

3. Possible modifications to the model

3.1. Tweaking the indices

In the above, the crop and tillage indices were taken as a given. However the calculations that went into these indices used empirical weightings to measure the combined effect of crop or tillage events over the last ten years. There is some experimental data behind these weights, considering separately the impact of particular events and then the damping of those impacts over time. An alternative approach, suggested at MISG, is to use Structural Equation Modelling [4] to determine the weights. A conceptual example of the model is given in Figure 14. The boxes $Till1, Till2, \ldots, Till10$ refer to tillage effects in each of the last 10 years. These effects will be correlated, as represented by the double-headed arrows on the left. These effects combine together in a latent "Tillage Impact" variable. The effect is assumed to be causal but with error ($u1$), with causality represented by single-headed arrows. The unobserved "Tillage Impact" then causally affects the soil quality measures ($SQ1, SQ2, SQ3$). Again there is some error involved ($e1, e2, e3$). Linear regression is used to model the causality. Conceptually it is feasible to make the structural equation model much bigger, to separate out the effect of individual tillage techniques at each year, and to include crop effects and background variables, all in the same model. The limitations are the size of the dataset relative to the model, and the accuracy of the data.

The SEM approach was tried at MISG, using the computer program AMOS. Two simple models (as in Figure 14) were able to be estimated, but the fit was inadequate. Probably more background variables need to be allowed in the model, provided this can be done in a parsimonious way. In principle the approach is promising.

3.2. Use of LENZ data

Significant regional differences were found in the soil measures even after adjusting for land use, soil order, soil texture, and farm management indices. The strength of regional differences after these adjustments was

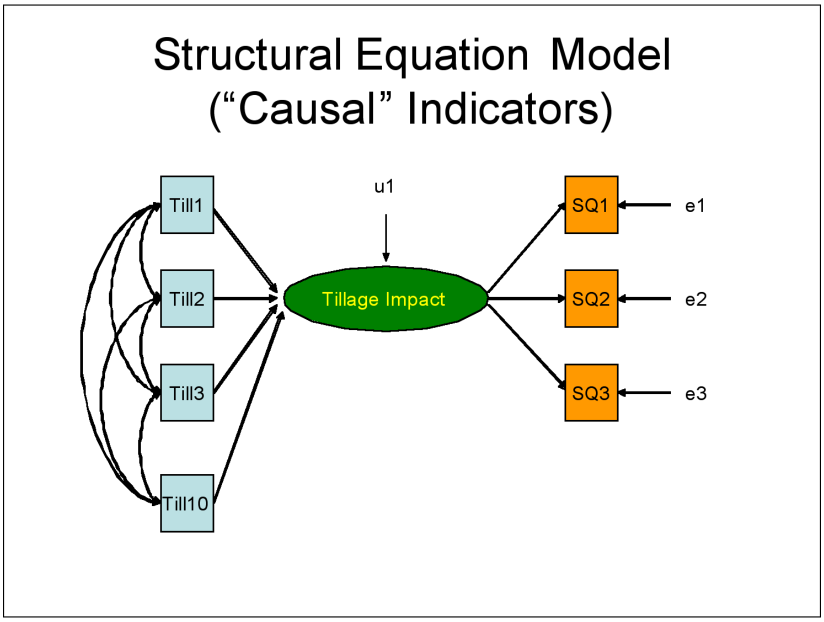

Figure 14. Structural equation model

surprising. The LENZ database (Land Environment of New Zealand) was interrogated to identify possible explanations for the regional differences. LENZ categorises sites according to a variety of measures, and data are available to describe most agricultural locations. Some measures are depicted in Figure 15, showing considerable regional variation. Climatic variations (annual rainfall, mean temperature and solar radiation) offered a partial explanation, significantly related to $C\%15$ and HWC, but having less influence on $BD15$, $AgStab$ (Mean weight), $ASD < 0.85$ and $ASD > 9.5$. It may be that other components of LENZ may help explain soil order differences. It was concluded that LENZ data are likely to provide a useful component of a Land Management Index.

3.3. Exploration of interactions

There appear to be interactions between the effects of region, land use, soil order and soil texture on the soil measures, for example the effect of land use on the soil measures may not be the same for all soil textures. Unfortunately the categorical data was unbalanced: each categorical variable had at least one category that was not present at all levels of some other categorical variable. For example each region had at least one missing soil order, land use or texture. This made the ex-

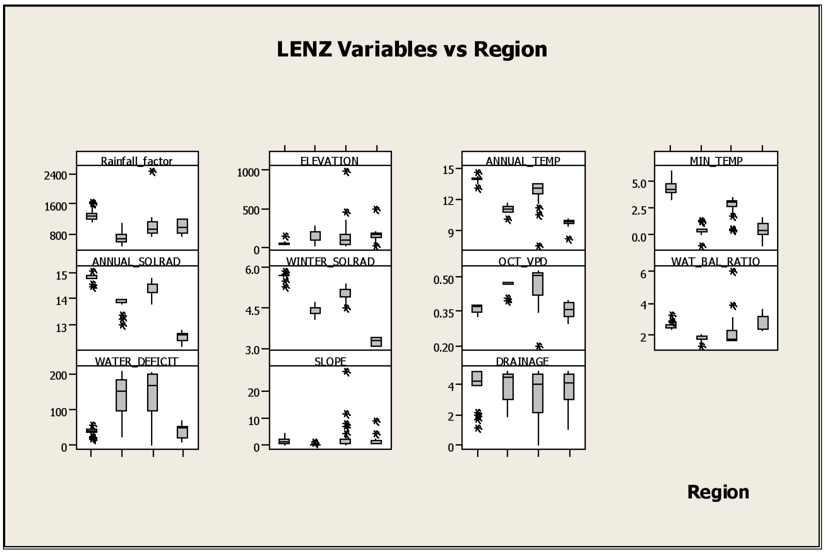

Figure 15. LENZ variables vs region

ploration of interactions difficult, as formally one could not fit a general model with interactions except using REML. A graphical method was therefore proposed and used to identify those interactions that seemed to be the most important. Specifically, the residuals from the main effects model were plotted against each background factor, as in Figure 16. Confidence intervals deviating considerably from the zero line indicate likely interactions. Interaction variables could then be created specifically to deal with the aberrant category, and the alleged interaction tested in a regression model. The intention would be to only include interactions when absolutely necessary, as for prediction purposes one needs to avoid overfitting to the sample data. Also some alleged interactions may just be the result of data errors, and the technique aided considerably in identifying these unusual cases. Of course it may be that some of these interactions will go away when more variables are included in the model e.g. LENZ data.

3.4. Suggestion for presentation

Suggestions were given as to how one could best express the data in a regression model for computation in Excel with new farmers data, and further how to present the results in a way easy to communicate to farmers. The latter is illustrated by the mock-up in Figure 17. In this mock-up, one would conclude that, all other things being equal, if

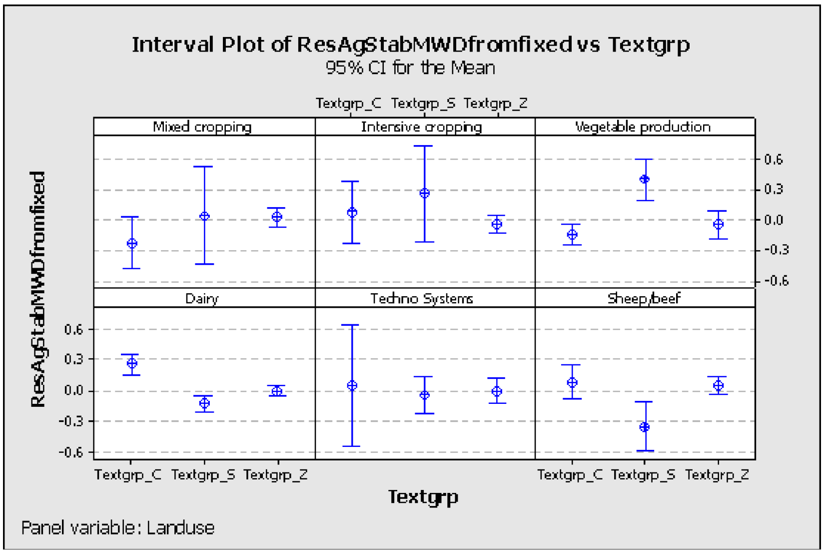

Figure 16. Interval plot of $ResAgStabMWDfromfixed$ vs $Textgrp$

a farmer can increase his/her crop score the stability will increase (with average 95% confidence intervals shown for each combination), and if he/she can reduce the tillage score, the stability should increase.

4. Analysis of penetrometer data

Roots cannot penetrate soils that are too hard (>2.5 MPa). The root system is inhibited and production from crops is therefore reduced. Penetrometers attempt to measure the resistance of soil to penetration and thus provide a further useful soil quality measure [1], in addition to the seven measures discussed in section 2.

A heuristic measure of penetration resistance (PR) is called the thumb test. If the thumb pressed into the soil penetrates no further than one thumbnail, this corresponds to a PR of about 1 MPa. Pentration to the first knuckle corresponds to a PR of 0.5 MPa. Penetration to the second knuckle corresponds to about 0.25 MPa.

Modern Penetrometers (as illustrated in Figure 18) measure the force (MPa/cm^2) required for penetration to a prescribed depth (15 cm). There are practical issues which affect the accuracy of the instruments, including the actual depth, the speed of insertion of the probe, and so on. However this study focused on the specific issue of water content. It is believed that penetrability is strongly dependent on the water con-

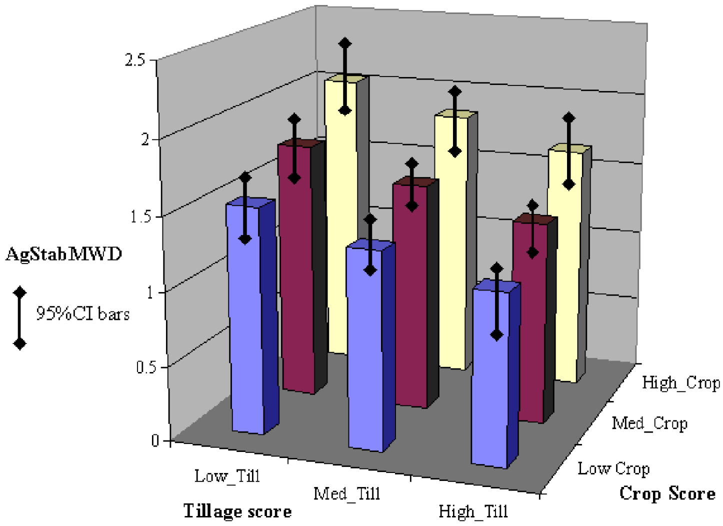

Figure 17. Proposed method of data presentation

tent of the soil, with PR changing by a factor of 2-4. The questions considered were:

- How to model the relationship between PR and water content?
- Is the relationship the same for all soil types?
- How then could one adjust field measurements of PR for water content?

Figure 19 illustrates the nature of the relationship that we are trying to model. As moisture content increases, the PR decreases.

Two models were proposed and examined at MISG, and one model subsequently. Let x denote the water content. The models are

(a) Exponential model $Pr_{field} = ae^{-bx}$

(b) Inverse model $Pr_{field} = \frac{1}{a+bx}$

(c) Power model $Pr_{field} = ax^b$

Model (a) was proposed on theoretical (dimensionality) grounds. It was anticipated that for a particular soil type s

$$Pr_{field} = a(\frac{\rho}{\rho_x})^n e^{-bx} \Rightarrow \log(Pr_{field}) = A + C_s - bx$$

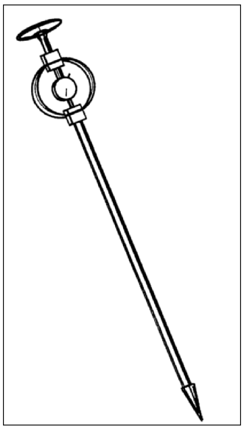

Figure 18. A modern penetrometer [3]

where ρ is the soil density and a, b, n are soil-dependent constants. Note that after taking logarithms the dependence on x is 'separated out'. It was of great interest to know whether b is soil-independent. If so, this would be a boon to soil researchers, as it would lead to a single water adjustment factor that could be used for soil types, regions, orders, and so on.

The data provided for the study were comprised of 529 pairs of readings for both '$PR15UnAdj$' (unadjusted penetration resistance to 15 mm) and '$SM15\%w\_w$' (soil moisture to 15 mm, percentage weight for weight). The soil order for each pair was also recorded, hence giving the scatterplot in Figure 19. Figure 20 shows comparative boxplots for variables by soil order. The Allophanic and Organic soils had highest median moisture content.

During MISG the exponential and inverse models were explored. The best-fitting model discovered at that time represented $\log(PR15UnAdj)$ as linear in $SM15\%w\_w$ with intercept varying according to soil order: model (a). However the model fitted poorly, with an R^2 of only 20%. It was left till after MISG to consider whether the nature of the relationship between PR and soil moisture depended soil order i.e. is there an interaction? Investigation showed there was indeed a significant interaction ($p = 0.014$). Penetration resistance falls faster than average (steeper slopes in $SM15\%w\_w$) for Melanic, Brown, and Pallic soils, and slower for Recent, Gley, Allophanic, Organic and lastly Granular soils.

Another question raised at MISG was whether some of the poor fit was due to mismeasurement at low PR levels? The scatterplots of $\log(PR15UnAdj)$ and $1/PR15UnAdj$ (Figures 21 and 22) seem to sug-

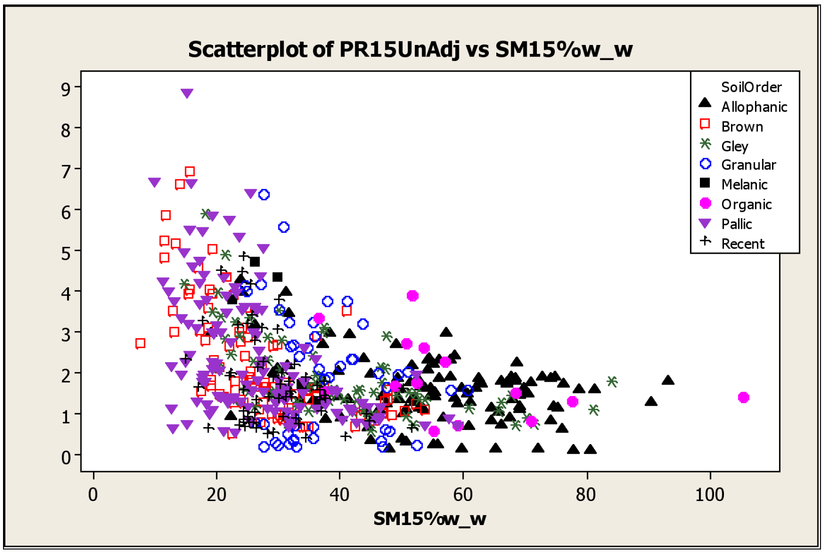

Figure 19. Scatterplot of $PR15UnAdj$ vs $SM15\%w\_w$

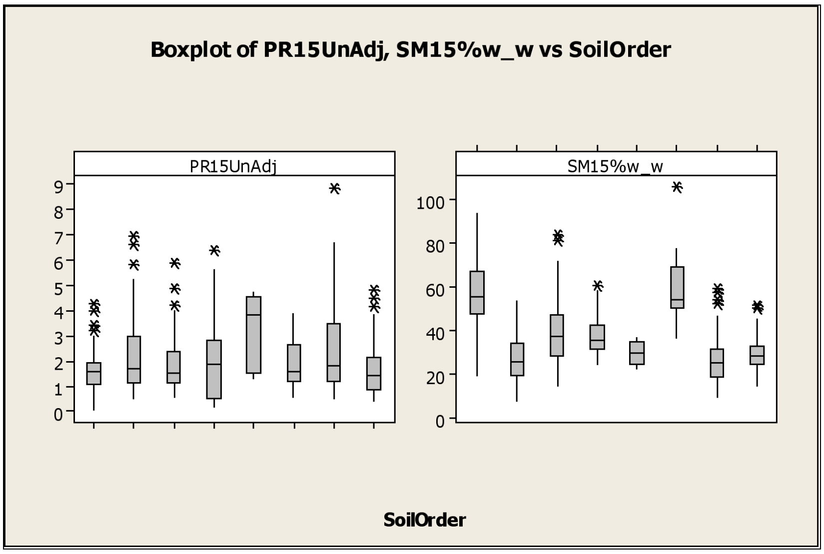

Figure 20. Boxplot of $PR15UnAdj, SM15\%w\_w$ vs $SoilOrder$

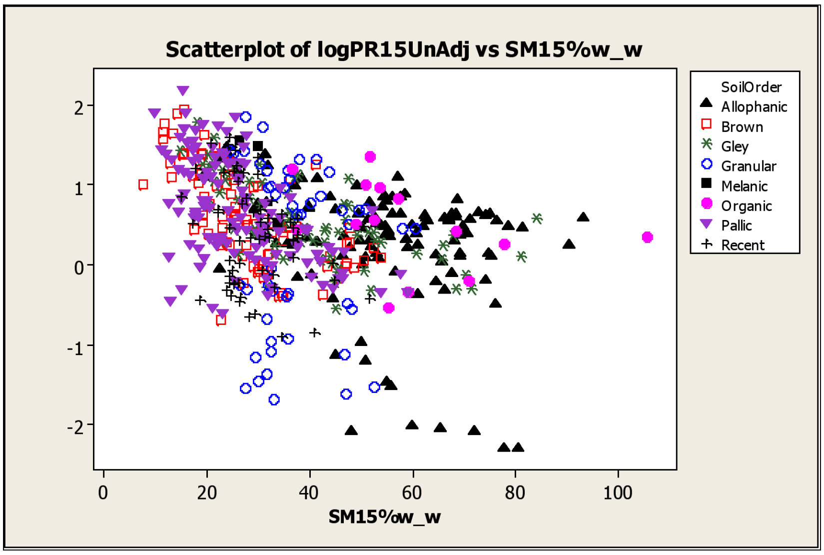

Figure 21. Scatterplot of $logPR15UnAdj$ vs $SM15\%w\_w$

gest this may be so. Discussions with soil scientists at MISG indicated measurement may be problematic at low PR. At the very least it is clear that any modelling is likely to be heavily influenced by very low PR values, especially for Allophanic and Granular soils.

Two modifications to the MISG model were subsequently considered: Firstly the data with $\log(PR) < -1$ were excluded from modelling. Secondly, a smoother of $\log(PR)$ versus Soil Moisture suggests curvature (Figure 23). Therefore the alternative model

$$\log(PR) = \alpha_s + \beta \log(\text{Soil moisture})$$

was considered. This simplifies to the power model. This had the twin benefits of reducing the curvature and at the same time led to a model with no significant interaction (p-value = 0.138), i.e. no necessity for separate βs slopes for each soil order.

Results of linear regression with this log-log model are shown in Figure 24. The regression model is still a poor fit ($R^2 = 27.2\%$) but at least permits a simple predictive equation without interaction (p-value for no interaction = 0.128). The model is an improvement on the log-linear one with the same points removed ($R^2 = 21.7\%$) . The results can therefore be summarized as $PR = A_s x^{-0.82}$ where $A_s = \exp(3.2961 + coef + s^2/2)$, the $s^2/2$ term being a bias correction and x being the moisture content, $x = SM15\%w\_w$. Hence $PR = 46.5x^{-0.82}$ for Allophanic soils, $PR = 30.05x^{-0.82}$ for Brown soils, $PR = 35.83x^{-0.82}$

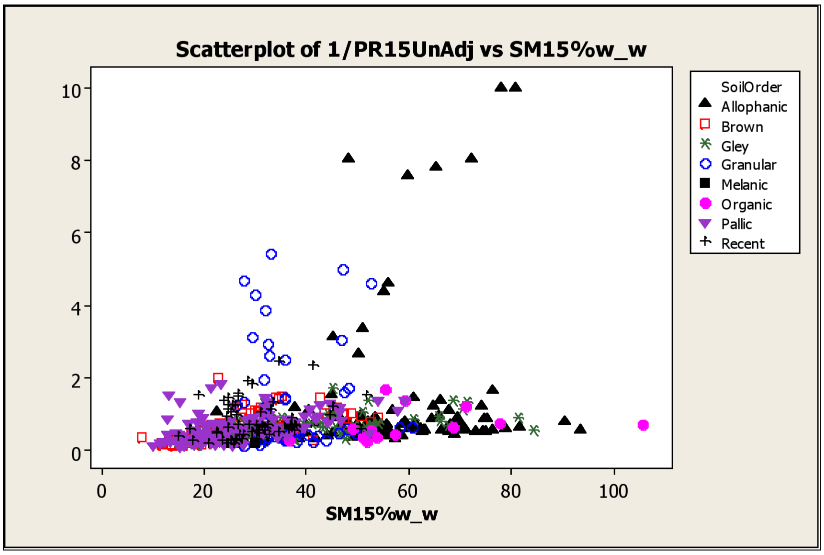

Figure 22. Scatterplot of $1/PR15UnAdj$ vs $SM15\%w\_w$

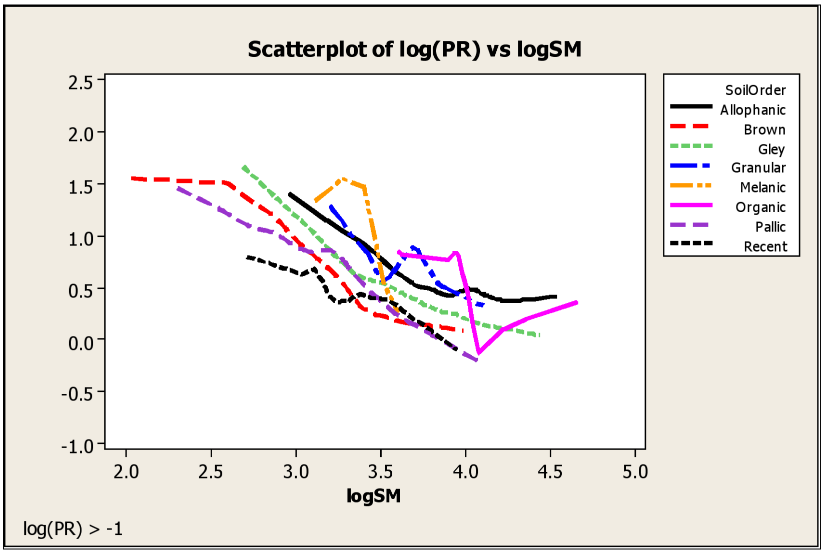

Figure 23. Scatterplot of $\log(PR)$ vs $logSM$

```
The regression equation is
log(PR) = 3.30 - 0.820 logSM + 0.421 Allophanic - 0.0151 Brown + 0.161 Gley
          + 0.292 Granular + 0.502 Melanic + 0.499 Organic - 0.151 Recent

Predictor        Coef   SE Coef       T      P
Constant       3.2961    0.2116   15.58  0.000
logSM        -0.81964   0.06451  -12.70  0.000
Allophanic    0.42136   0.08198    5.14  0.000
Brown        -0.01510   0.06877   -0.22  0.826
Gley          0.16078   0.07711    2.09  0.038
Granular      0.29174   0.09485    3.08  0.002
Melanic        0.5025    0.2255    2.23  0.026
Organic        0.4990    0.1492    3.35  0.001
Recent       -0.15114   0.07852   -1.92  0.055

S = 0.493823   R-Sq = 27.2%   R-Sq(adj) = 26.0%
```

Figure 24. Regression analysis for log(PR)

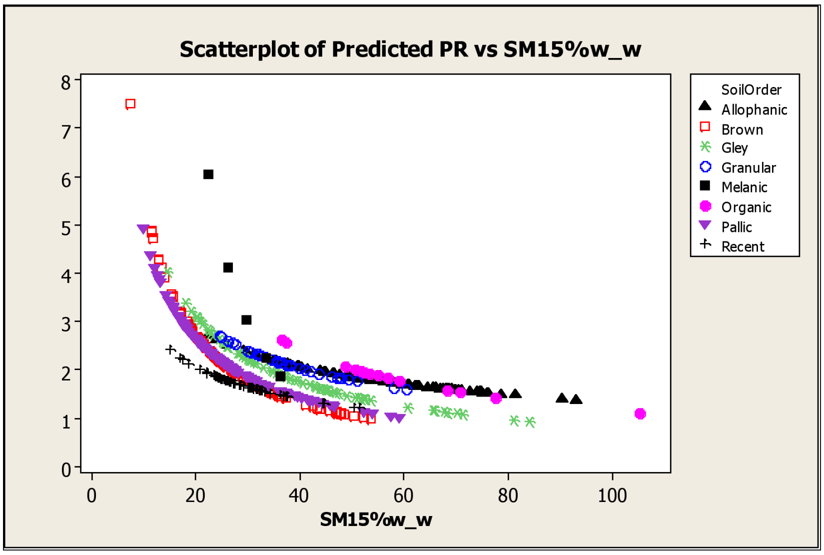

Figure 25. Scatterplot of predicted PR vs $SM15\%w\_w$

for Gley, $PR = 50.43x^{-0.82}$ for Granular, $PR = 50.25x^{-0.82}$ for Melanic, $PR = 26.22x^{-0.82}$ for Recent, and $PR = 30.51x^{-0.82}$ for Pallic soils.

Of course these formulae make no adjustment for Texture, Region, Landuse or Cropping and Tillage, which can lift the explained variation in $\log(PR)$ further. However this seems an adequate starting point for future modelling. The fitted curves are displayed in Figure 25.

5. Conclusion

The MISG tackled a very extensive problem, with eight important measures of soil quality as response variables, and several significant predictors including indices representing the impact of crops grown and tillage methods used at each farm site. It was found that background factors, including soil order and texture, land use and region, need to be taken into account when assessing the effect of the indices. Some suggestions were made as to how the crop and tillage index construction might be improved, how data on climatic and other variables might explain some regional differences, how interactions between factors and indices might be assessed and incorporated in the model, and how results might be presented. The relationship between soil moisture and penetration resistance was clarified. The extensive nature of the problem meant that development of a full land management index was never going to be possible in such a short time as MISG. The contribution of MISG was to bring several creative and expert minds together to explore many different aspects of the problem. As a result, basic quantitative knowledge was added by the finding of significant relationships between variables, by the identification of possible model limitations, and by the exploration of ways to resolve them, towards development of a fully-fledged model.

Acknowledgements

Thank you to the MISG participants Kaye Marion (RMIT), Mini Ghosh (Massey University), Walt Davis, Simon Leong and Victoria Wei (Statistics NZ). Special thanks to Kaye for the representation in 3.4 and to Walt for the Structural Equation Modelling.

Thank you also to the industry representatives: Dr Mike Beare (email: bearem @crop.cri.nz), Erin Lawrence (email: lawrencee@crop.cri.nz), Jeromy Cuff (email: Jeromy.cuff@ecan.govt.nz) and Esther Meenken (email: meekene@crop.cri.nz).

References

[1] Beare M. H., Lawrence E. J., Tregurtha C. S., Harrison-Kirk T., Pearson A. and Meenken E. D., Progress towards the development of the Land Management Index 2004-05 project report *Crop & Food Research Confidential Report No. 1408*, New Zealand Institute for Crop & Food Research Limited, Private Bag 4704, Christchurch, New Zealand (2005), *http://www.maf.govt.nz/sff/about-projects/search/02-125/02125-finalreport.pdf*

[2] Chatterjee S., Hadi A., and Price B. *Regression Analysis by Example*, 3rd Ed., John Wiley & Sons, (2000)

[3] Penetrometer gif is from *http://tinyurl.com/ykx5ff*.

[4] Byrne B. M. *Structural Equation Modelling with AMOS: Basic Concepts, Applications and Programming*, Lawrence Erlbaum Associates Inc New Jersey, (2001)