

USING PROXIMAL HYPERSPECTRAL SENSING TO MEASURE SOIL OLSEN P AND pH

Miles C.E. Grafton¹, Therese M.C. Kaul¹, Alan S. Palmer¹, Peter A. Bishop¹
and Michael White²

¹*School of Agriculture and Environment, Massey University
Private Bag 11 222, Palmerston North 4410, New Zealand*

²*Ravensdown Fertiliser Ltd. P.O. Box 1049, Christchurch 8042, New Zealand
Email: M.Grafton@massey.ac.nz*

Abstract

This paper reports on work undertaken to use a large data set of hyperspectral data measured on dry soil samples to obtain regression analysis which allows predictions of pH and Olsen P to be obtained from an independent data set. The large data set was obtained from 3,190 soil samples taken from the Ravensdown Primary Growth Partnership to a depth of 7.5cm. The spectra were measured using an Analytical Spectral Device which recorded 2,150 wavebands of 1nm resolution between 350nm and 2,500nm. Values for Olsen P and pH were provided from chemical analysis by Analytical Research Laboratories. The spectra were regressed using “R” statistical software which has the power to handle the data and report the wavebands with the most significance for the model. The data set for the prediction came from a stratified nested, grid soil sampling exercise which was used to find Olsen P stability at varying depths. This set had 400 samples from each of two data sets from different areas on Patitapu Station using a grid sample protocol. The 100 most significant wavebands from the PGP data set were used to regress the Patitapu data which were combined. These were regressed using “R” (Version 3.41, The R Foundation) and Statdata (Palisade, New York), which produced the same result. The partial least square regression of pH was very significant and was predicted well. Olsen P had a very significant correlation which was quite noisy, correlating the log10 of Olsen P was also undertaken and it would appear something is being measured that is associated with Olsen P. This work shows that it is possible to measure soil nutrient by proximal hyperspectral analysis which is transferable to an independent data set.

Keywords: soil testing, partial least squares regression, hyperspectral sensing, big data

Introduction

This work aims to examine the possibility of using proximal hyperspectral sensing to measure soil nutrient status of soil samples by seeing if a relationship could be established for either Olsen P or pH.

A large data set of chemically analysed soil samples was married to proximal hyperspectral data measured at the same laboratory. The work was undertaken by Analytical Research Laboratories Ltd. (ARL) an International Accreditation New Zealand (IANZ) accredited laboratory which is also (ISO 17025) certified. Spectra were measured between 350nm and 2,500nm at 1nm resolution using a Fieldspec 4 (ASD, Analytical Spectral Devices, Malvern

Panalytical, Malvern, U.K.). This type of device has been used for spectral analysis of soil cores in other work (Viscarra Rossel and Behrens, 2010; Nocita, *et al.*, 2014; Roudier, *et al.*, 2015). The data is a subset of soil samples taken at 7.5cm depth for the Ravensdown, Ministry of Primary Industries Primary Growth Project (PGP) and was supplied by Ravensdown. This data set had 3,189 samples.

A second independent data set of soil samples taken from a stratified, nested grid sample in an independent exercise (soil sampling depths; 3cm, 3-15cm, 15-30cm) at Patitapu Station near Alfredton, in the Central North Island (Kaul, *et al.*, 2017, Kaul and Grafton, 2017) was also analysed. This data set was analysed for Olsen P and pH at Massey University. The Olsen P was measured as per the method described in (Kaul and Grafton, 2017) and pH by a digital reading from a 10g dried sample in distilled water. These samples were also independently measured for spectra using an ASD probe at Massey University. This data set had 883 samples. Dry soil samples were analysed to avoid the effects of water on the spectra (Minasny, *et al.* 2011).

It was hypothesised that a partial least square regression (PLSR) could be undertaken on one data set and the equation derived would give a good prediction on an independent data set, to derive Olsen P and pH.

Materials and methods

A PLSR was undertaken using “R” version 3.41 (R, 2016; Mevik and Wehrens, 2007) on the larger data set. Powerful statistical software was required as the data set had 2,151 columns and 3,190 rows. From this analysis the 100 columns which represented the wavebands that were most significant for Olsen P and pH were found. These were used to undertake a PLSR using both “R” and StatTools (Pallisade, NY, U.S.A.) which is a statistical software plug in for Excel. StatTools has a limit of 100 independent variables for a PLSR. The regression was undertaken in both packages to ensure that the same result was obtained. For practical purposes and to avoid autocorrelation and co-linearity 100 independent variables was considered a maximum. This becomes the case if the number of rows in the data is less than the number of columns used in the regression. A PLSR in StatTools was undertaken on both data sets for pH and Olsen P, using 100 independent variables for each.

The PLSR equation derived from the smaller data set was used to predict values of pH and Olsen P on the larger data set. These predictions from the equation were graphed as a second series overlay on the PLSR scatter plot of the larger data set.

The pH data from both sets had a normal distribution that meets the requirement of a regression analysis. Whereas the Olsen P data have Chi^2 distributions, which reduces the accuracy of regression analysis. This is also the reason the standard deviation for Olsen P is high as the data has a skewed distribution. To improve the normality requirement a \log_{10} distribution of Ravensdown Olsen P data was regressed.

Results

The mean and standard deviations of the data are displayed in Table 1.

Table 1. Mean and standard deviations of the two data sets.

Data set	Mean	Standard Deviation
Massey Olsen P	18.1	13.9
Massey pH	5.55	0.33
Ravensdown Olsen P	23.1	15.4
Ravensdown pH	5.72	0.32

The statistical summary of the PLSR provided a clear correlation for all regressions; although the results for Olsen P had much more noise, see Table 2. All analyses were statistically significant, the probability of results being random was less than 0.0001.

Table 2. Summary of regression analyses.

Data set	R ²	Adj. R ²	Std. error	F	P value
Ravensdown Olsen P ^a	0.8273	0.4706	11.21	2.319	<2.2 × 10 ⁻¹⁶
Ravensdown pH ^a	0.9064	0.7131	0.17	4.688	<2.2 × 10 ⁻¹⁶
Massey pH ^b	0.4828	0.4167	0.25	7.3	<2.2 × 10 ⁻¹⁶
Massey pH ^c	0.4828	0.4167	0.25	7.3	<.0001
Massey Olsen P ^c	0.6051	0.5546	9.03	11.99	<.0001
Ravensdown pH ^d	0.4427	0.4247	0.24	24.53	<.0001
Ravensdown Olsen P ^d	0.2887	0.2656	13.21	12.53	<.0001
Ravensdown Log OP ^d	0.3607	0.34	9.03	17.42	<.0001

^a R regression 2,148 and 1,040 degrees of freedom; ^b R regression 100 and 782 degrees of freedom;

^c StatTools regression 100 and 782 degrees of freedom; ^d StatTools regression 100 and 3088 degrees of freedom.

Figures (1 -4) displays the PLSR from the StatTools scatter plots.

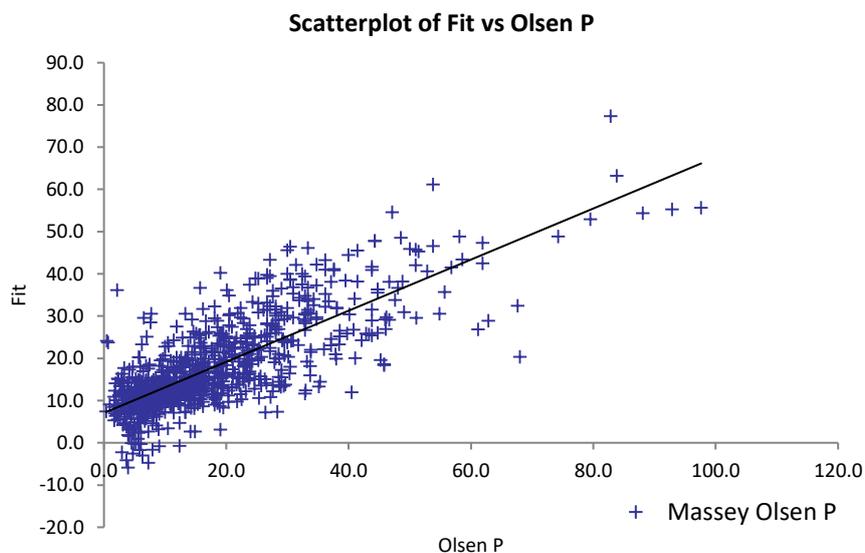


Figure 1: Linear regression of Olsen P using Massey data

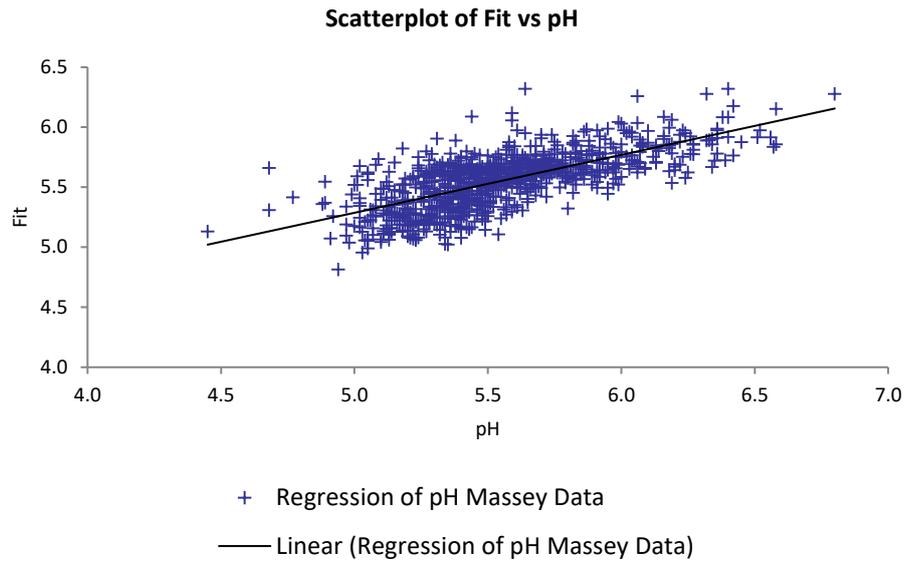


Figure 2: Linear regression of pH using Massey data

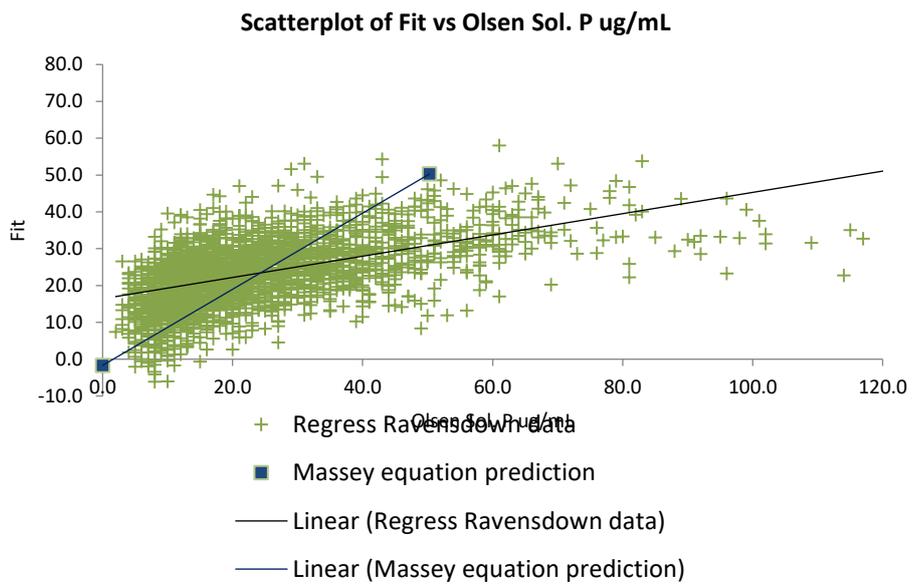


Figure 3: Ravensdown data PLSR of Olsen P and with predictions using the equation from the Massey PLSR

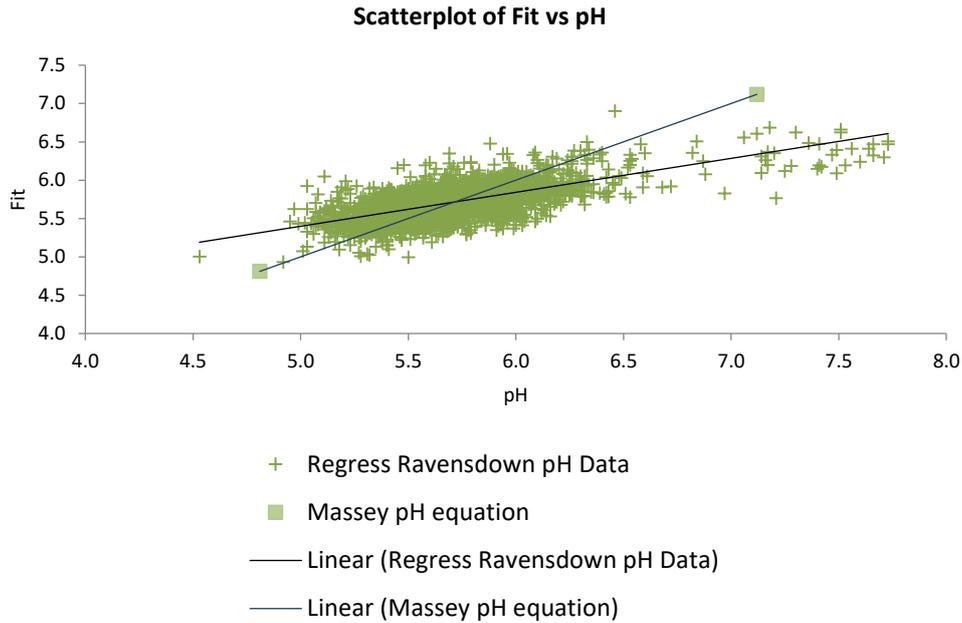


Figure 4: Regression of Ravensdown pH and prediction derived from Massey pH PLSR

The linear nature of the regression is more obvious for pH than for Olsen P, as per the summary in Table 1. The equations developed from the PLSR of the Massey data although at different gradients to the regression trend lines however do bisect the data points in the scatter plots evenly. A PLSR of the logs of Olsen P was undertaken as it appeared the relationship may be non-linear. The summary is in Table 1 and the scatter plot of the PLSR appears in Figure 5.

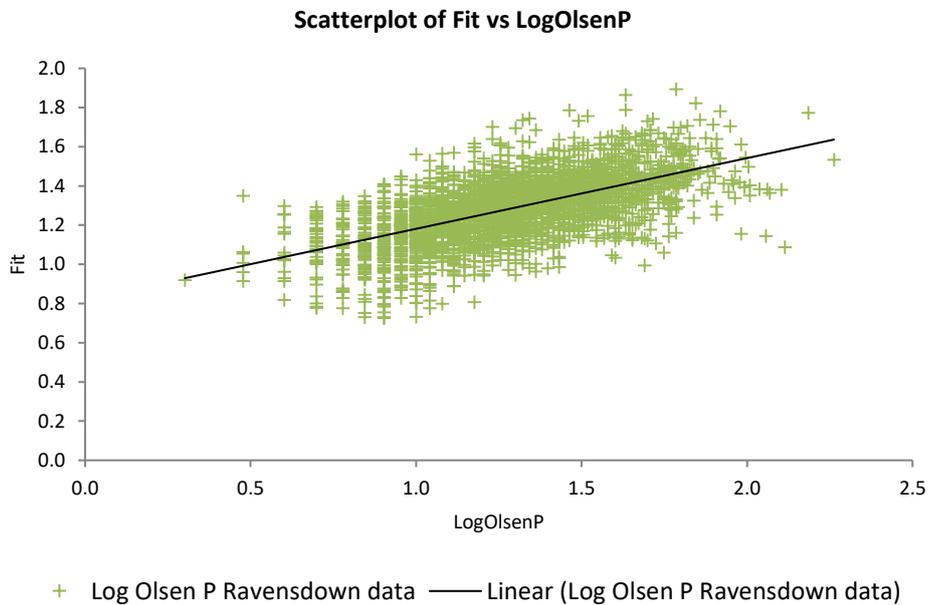


Figure 5: Scatter plot of Log of Olsen P of Ravensdown data

Discussion

This work demonstrates that there is potential to use proximal hyperspectral sensing to measure nutrients in soils. The PLSR of pH was linear and less variable than Olsen P; this reflects that

the pH data met the normality test. The wavebands used for the analyses were gleaned from the Ravensdown data set; which were the 100 most significant found by a regression on 2,150 wavebands on 3,189 soil samples.

The selected wavebands for Olsen P and pH produced statistically very significant relationships in the Massey data; and the equations derived from the PLSR, proved to be nearly as well correlated as the PLSR on the Ravensdown data.

The shape and variance in the Olsen P data results from the spectral data having a Chi^2 distribution rather than a normal distribution, which results in a less, linear relationship. This appears to be the case as the linear relationship improves when the logs of Olsen P were fitted. However, the data is still variable so the PLSR may be finding a relationship with soil components that have a relationship with phosphorus.

Conclusion

This work with large data sets shows there is promise in using proximal sensing to measure soil nutrients. This may lead to more intensive soil testing at a cheaper cost to the farmer and provide an opportunity to deliver fertiliser more efficiently based on better information.

References

- Kaul, T.M.C., Grafton, M.C.E., Hedley, M.J., and Yule, I.J. (2017). Understanding soil phosphorus variability with depth for the improvement of current soil sampling methods. In: *Science and policy: nutrient management challenges for the next generation*. <http://flrc.massey.ac.nz/publications.html>. Occasional Report No. 30. Fertilizer and Lime Research Centre, Massey University, Palmerston North, New Zealand: ISSN 0112-9902, 11 pages.
- Kaul, T.M.C., Grafton, M.C.E., (2017). Geostatistical Determination of Soil Noise and Soil Phosphorus Spatial Variability. *Agriculture*, [https:// DOI: 10.3390/agriculture7100083](https://doi.org/10.3390/agriculture7100083) 10 Pages.
- Minasny, B., McBratney, A.B., Bellon-Maurel, V., Roger, J.-M., Gobrecht, A., Ferrand, L. *et al.* (2011). Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. *Geoderma*. Vol 67, 118–124.
- Mevik, B.H. & Wehrens, R. (2007) The pls package: principal component and partial least squares regression in R. *Journal of Statistical Software.*, Vol 18, 1 – 24.
- Nocita, M., Stevens, A., Toth, G., Panagos, P., van Wesemael, B., Montanarella, L., (2014). Prediction of soil organic carbon content by diffuse reflectance spectroscopy a local partial least square regression approach. *Soil Biology and Chemistry*. Vol 68, pp 337 – 347.
- Viscarra Rossel, R.A., Behrens, T., (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, Vol 158, pp 46 – 54.
- R Core Team. *R: A Language and Environment for Statistical Computing*. (2016), R Foundation for Statistical Computing, Vienna.
- Roudier, P., Hedley, C.B., Ross, C.W., (2015). Prediction of Volumetric Soil Organic Carbon from Field Moist Intact Soil Cores. *European Journal of Soil Science*, Vol 66:4 pp 651 – 660.