A novel approach to map the depth to a soil pH constraint – a useful tool for understanding yield variability

Patrick Filippi[*1], Edward J. Jones[1], Bradley J. Ginns[1], Brett M. Whelan[1], Guy W. Roth[1], Thomas F.A. Bishop[1]

[1] *The University of Sydney, Sydney Institute of Agriculture, Sydney, New South Wales, Australia, patrick.filippi@sydney.edu.au*

Abstract

Subsoil alkalinity is a common issue in the alluvial cotton-growing valleys of northern NSW, Australia. This causes nutrient deficiencies, toxicity, and inhibits root growth, which can have a damaging impact on crops. The depth at which a soil constraint is reached is important information for farmers, however, this is hard to measure spatially. This study predicted the depth in which a pH constraint (pH > 9) was reached to a 1 cm vertical resolution for a 1 m soil profile on a dryland cropping farm in northern NSW, Australia. Equal-area quadratic smoothing splines were used to resample vertical soil profile data, and a random forest model was used to produce the depth-to-pH-constraint map. The model to spatially predict soil pH across the farm was accurate, with an LCCC of 0.63, and an RMSE of 0.47 when testing with leave-one-site-out-cross-validation. About 77% of the area was constrained by a pH greater than 9 within the top 1 m of soil. The relationship between the predicted depth-to-pH-constraint map and cotton and grain (wheat, canola, and chickpea) yield monitor data was analysed for individual fields. The deeper in the soil profile a pH constraint was reached, the greater the crop yield. A strong relationship was found for wheat, canola, and chickpea (Spearman's correlation (r_s) of 0.75, 0.66, and 0.58, respectively), and a moderate relationship for cotton (r_s = 0.37). The modelling approach presented could be used to identify the depth to other soil constraints, such as soil sodicity. The outputs are a promising opportunity to understand crop yield variability, which could lead to improvements in management practices.

Key Words

Soil constraints; digital soil mapping; yield variability; precision agriculture.

Introduction

The soils of the alluvial cotton-growing valleys of eastern Australia have many desirable physical and chemical characteristics that make them highly suitable for cropping, but soil constraints are still widespread and can significantly reduce crop yields. Some of the most common soil constraints in this region are sodicity, salinity, alkalinity, and compaction, and these are particularly prevalent in the subsoil (Dang *et al.* 2006). There is a need to quantify these constraints at a variety of spatial scales (within-field, to region), and to assess the impact that they have on crop yield. While subsoil acidity is a widespread problem in many parts of Australia, subsoil alkalinity is a more localised issue, and has accordingly received less attention. Alkalinity is a common issue in the soils of the cotton-growing valleys of eastern Australia, and is generally due to the presence of calcium and sodium carbonates, which increases with depth (Singh *et al.* 2003; Filippi *et al.* 2018a). Highly alkaline soils limit the accessibility of certain nutrients to plants, as well as cause toxicities. This can inhibit root growth, decreasing the amount of water and nutrients that can be utilised, which can negatively impact crop production (Dang *et al.* 2006). The optimum pH range for most crops is generally from 5.5 to 7 (pH in H_2O), and a soil with a pH of 9 would be a significant constraint for most agricultural crops (Hazelton and Murphy 2007).

Soil pH, and the depth at which a soil pH constraint (> 9) is reached can significantly vary laterally within fields (Adamchuk *et al.* 2007), which can result in spatially variable crop yields (Taylor *et al.* 2003). Digital soil mapping (DSM) has been used extensively for assessing the spatial distribution of soil pH at a variety of spatial extents, from the field, to the world. Some studies solely produce pH maps for the topsoil (e.g. Kirk *et al.* 2010), which lacks important information about subsoil pH. Other studies create maps of several depth increments (e.g. Filippi *et al.* 2018b), but this is often too much information to easily make management decisions. For a farmer that is experiencing a subsoil pH constraint in parts of their farm, a single map of the depth at which that pH constraint begins would be invaluable. This would help identify areas where crop rooting depth may be inhibited, and is a simple concept to help growers implement management practices to either rectify the issue, or alter inputs according to the constraint/production potential.

The current study uses a novel modelling approach to predict and map the depth to soil pH alkalinity constraint (pH > 9) on a dryland farm in the Namoi Valley in northern New South Wales (NSW), Australia. A collection of spatial and temporal datasets was used for modelling and mapping, and the relationship of the predicted depth-to-pH-constraint map with grain (wheat, canola, and chickpea) and cotton yield maps assessed.

Methods

Study site

This study was conducted on a 1070-ha dryland cropping farm "L'lara" near Narrabri in the Namoi Valley in northern NSW, Australia. Summer cotton and winter wheat are the dominant crops grown, with occasional rotations of winter canola, chickpea, faba bean, and field pea, and summer sorghum. The mean annual rainfall is 660 mm, and the soils are primarily grey and brown Vertosols.

Sampling and soil analysis

In total, 110 soil cores were extracted from the study area. At 80 of these sites, sub-samples were collected at 0-0.1, and 0.4-0.5 m. At the 30 remaining sites, sub-samples were extracted at (0-0.1, 0.1-0.3, 0.3-0.6, 0.6-0.8, and 0.8-1.0 m). All sub-samples were air dried, and ground to <2 mm fraction. Soil pH was analysed in 1:5 soil: water extracts using a Mettler Toledo S220 SevenCompact™ pH/Ion meter.

Vertical depth modelling

Equal-area quadratic smoothing splines were fitted using soil pH data from each individual soil core (Malone *et al.* 2009). This was implemented using the 'ithir' package in R (Malone 2015). The data was fitted and stored at 1 cm vertical resolution. For soil cores that had the full 100 cm profile extracted and analysed, this resulted in 100 pH values at 1 cm increments, and for sites with only 0-10 and 40-50 cm samples, this resulted in 50 pH values at 1 cm increments to 50 cm.

Soil modelling and mapping

At each of the 110 soil sampling locations, the corresponding on-farm and publicly-available data described in Table 1 was gathered. A model to predict the spatial distribution of soil pH was then created using this dataset and a random forest model (Breiman 2001). Instead of one model for each splined depth, all were modelled together. This was possible, as each depth (at a 1 cm increment) was stacked in the data frame, and depth was then included as a predictor variable. This model was then used to predict onto the 5 m covariate grid of the study area, and this was done at each 1 cm increment down to 100 cm, resulting in 100 maps. The depth in which the pH first reached a value of 9 or greater was then recorded for each 5 m grid point, as this is when significant constraints to growth for most crops is reached. This information was then combined into a single map of the study area, showing the depth to a soil pH constraint.

Table 1 – Data sources used in the modelling/mapping, and the spatial resolution

Data type	Data description	Resolution
Elevation	DEM	5 m
Gamma radiometrics (proximal)	Potassium (K)	5 m
Landsat 7 – red band	90th, percentile (2000-2017)	30 m
EC_a (apparent electrical conductivity)	EM-21S (0-3.0 m)	5 m

The quality of the model was tested by using leave-one-site-out cross-validation (LOSOCV). This entailed removing all soil data from each site, which ensured that data from different depth increments of the same soil core were not included in both the calibration and validation datasets. This LOSOCV was performed on all 110 sites, with 109 sites used to predict the remaining site each time. The results of the validation sites were then combined, and the Lin's concordance correlation coefficient (LCCC) (Lin 1989), and the RMSE (root-mean-square error) were used to assess the quality of the model predictions. Splined pH data at 1 cm increments was used to validate the predicted data. To test the importance of different predictor variables in the model, the mean decrease in accuracy was used, which is based on the mean square error (MSE).

Relationship with predicted depth-to-pH-constraint and crop yield

The relationship between the predicted depth-to-pH-constraint map, and crop yield monitor data (5 m) was assessed. The raw yield monitor data was processed by removing spurious and extreme values following the

method of Taylor *et al.* (2007). Data was used from two separate fields, for two seasons each. This consisted of canola in 2016, and chickpea in 2017 for field Campey 4/5 (C4/5), and wheat in 2016, and cotton in 2017/2018 for field L'lara 2 (L2). Boxplots were used to assess this relationship, where data was grouped in 10 cm intervals (depth-to-pH-constraint), showing how the distribution of yield data changed as the depth to a soil pH constraint changed for each paddock and crop/season. The Spearman rank-order correlation coefficient, r_s, was also used as a tool to assess the strength of this relationship.

Results

Depth to soil pH constraint

The depth-to-pH-constraint map showed considerable spatial variability across the study area (Fig. 1). In total, 77% of the cropping fields had a strongly alkaline pH of 9 somewhere within the top 100 cm. The eastern section of the study area was largely unconstrained, with much of the middle section becoming constrained at 31-40 cm. The south-western section of L'lara had high spatial variability, with areas that were constrained in the top 1-10 cm, and unconstrained to 100 cm, all within a short distance.

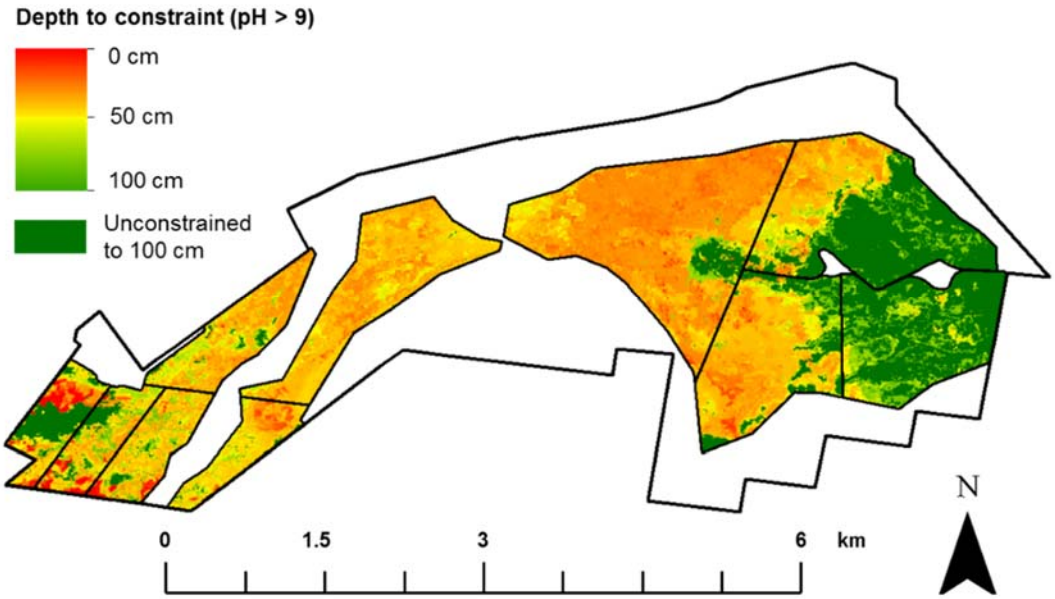

Fig. 1 – Digital soil map of the depth in which soil pH constraint (> 9) is reached across the study area

Model quality and variable importance

For the random forest model, the validated results showed values of an LCCC of 0.63, and an RMSE of 0.47 using LOSOCV (Table 2), suggesting that the model could spatially predict soil pH accurately. Proximally-sensed gamma radiometric K was the most important predictor of soil pH. This was closely followed by mid-depth, and then other horizontally variable predictors, such as DEM, and soil ECa. Landsat 7 data of the 90th percentile red band from 2000-2017 proved to be the least important predictor of soil pH.

Table 2 – Prediction statistics of modelled soil pH against measured soil pH at all depths using leave-one-site-out cross-validation (LOSOCV)

Validation technique	LCCC	RMSE
Leave-one-site-out cross-validation	0.63	0.47

The relationship with crop yield and the depth to soil pH constraint

Results showed that as a soil pH constraint (> 9) was deeper in the soil profile, the yield of most crops increased. The Spearman's correlation analysis revealed that the relationship between the predicted depth-to-pH-constraint data and yield monitor data ranged from strong to weak (Table 3). The strongest relationship was found with wheat (r_s = 0.75), followed by canola (r_s = 0.66), chickpea (r_s = 0.58), and then cotton (r_s = 0.37). Although not shown here, the lowest median yield was observed where a soil pH constraint was reached in the shallowest layer, and the highest median yield was observed where soil was deemed unconstrained by pH in the top 100 cm for all grain crops.

Table 3 – The Spearman's correlation (r_s) of the relationship between the depth-to-pH-constraint map data and the median yield value for each 1 cm depth increment to 100 cm

	Campey 4/5		L'lara 2	
	Canola '16	**Chickpea '17**	**Wheat '16**	**Cotton '17/18**
r_s	0.66	0.58	0.75	0.37

Conclusion

High levels of soil alkalinity were observed in the study area, particularly at depth. The random forest model to predict the spatial distribution of soil pH was accurate, with a validated LCCC value of 0.63, and an RMSE of 0.47 with LOSOCV. The equal-area quadratic smoothing splines proved useful for fitting the soil pH data to vertical 1 cm increments of the 100 cm soil profile. It is typically not easy to identify the depth at which a soil constraint occurs in the field, but this approach helps overcome this. An advantage of this approach is that it could be applicable to any soil property, not just soil alkalinity. The study revealed that the deeper in the soil profile a pH constraint was reached, the greater the crop yield. A strong relationship was found for wheat, canola, and chickpea (Spearman's correlation (r_s) of 0.75, 0.66, and 0.58, respectively), and a moderate relationship for cotton (r_s = 0.37). The output of a single map showing the depth at which a soil alkalinity constraint occurs is a valuable, concise piece of information for farmers and land managers, and is a promising avenue to guide the remediation of soil constraints, or the tailoring of management inputs to account for these constraints. Future research should use this developed approach for other important soil constraints, such as sodicity.

References

Adamchuk, V.I., Lund, E.D., Reed, T.M., Ferguson, R.B. (2007) Evaluation of an on-the-go technology for soil pH mapping. *Precision Agriculture* **8**, 139-149.

Breiman, L. (2001) Random forests. *Machine Learning* **45**, 5-32.

Dang, Y.P., Dalal, R.C., Routley, R., Schwenke, G.D., Daniells, I. (2006) Subsoil constraints to grain production in the cropping soils of the north-eastern region of Australia: an overview. *Australian Journal of Experimental Agriculture* **46**, 19-35, https://doi.org/10.1071/EA04079

Filippi, P., Cattle, S.R., Bishop T.F.A., Pringle, M.J., Jones, E.J. (2018a) Monitoring changes in soil salinity and sodicity to depth, at a decadal scale, in a semiarid irrigated region of Australia. *Soil Research* **56**, 696-711.

Filippi, P., Cattle, S.R., Bishop, T.F.A., Odeh, I.O.A., Pringle, M.J. (2018b) Digital soil monitoring of top- and sub-soil pH with bivariate linear mixed models. *Geoderma* **322**, 149-162.

Hazelton, P., Murphy, M. (2007) Interpreting Soil Test Results: What Do All the Numbers Mean? *CSIRO Publishing*, Collingwood, Victoria

Kirk, G.J.D., Bellamy, P.H., Lark, R.M. (2010) Changes in soil pH across England and Wales in response to decreased acid deposition. *Global Change Biology* **16**, 3111-3119.

Lin, L.I.K. (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 255-268.

Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M. (2009) Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* **154**, 138-152.

Malone, B. (2015) ithir: Soil data and some useful associated functions. R package version 1.0.

Singh, B., Odeh, I.O.A., McBratney, A.B. (2003) Acid buffering capacity and potential acidification of cotton soils in northern New South Wales. *Soil Research* **41**, 875-888.

Taylor, J. C., Wood, G. A., Earl, R., Godwin, R. J. (2003) Soil factors and their influence on within-field crop variability, Part II: Spatial analysis and determination of management zones. *Biosystems Engineering* **84**, 441-453.

Taylor, J.A., McBratney, A.B., Whelan, B.M. (2007) Establishing Management Classes for Broadacre Agricultural Production. *Agronomy Journal* **99**, 1366-1376.