

Article

# Mapping the Depth-to-Soil pH Constraint, and the Relationship with Cotton and Grain Yield at the Within-Field Scale

Patrick Filippi \*<sup>(D)</sup>, Edward J. Jones, Bradley J. Ginns, Brett M. Whelan, Guy W. Roth and Thomas F.A. Bishop

School of Life and Environmental Sciences, Sydney Institute of Agriculture, The University of Sydney, Sydney, NSW 2006, Australia; edward.jones@sydney.edu.au (E.J.J.); bgin8536@uni.sydney.edu.au (B.J.G.); brett.whelan@sydney.edu.au (B.M.W.); guy.roth@sydney.edu.au (G.W.R.); thomas.bishop@sydney.edu.au (T.F.A.B.) \* Correspondence: patrick.filippi@sydney.edu.au

Received: 20 March 2019; Accepted: 15 May 2019; Published: 21 May 2019



Abstract: Subsoil alkalinity is a common issue in the alluvial cotton-growing valleys of northern New South Wales (NSW), Australia. Soil alkalinity can cause nutrient deficiencies and toxic effects, and inhibit rooting depth, which can have a detrimental impact on crop production. The depth at which a soil constraint is reached is important information for land managers, but it is difficult to measure or predict spatially. This study predicted the depth in which a pH (H<sub>2</sub>O) constraint (>9) was reached to a 1-cm vertical resolution to a 100-cm depth, on a 1070-hectare dryland cropping farm. Equal-area quadratic smoothing splines were used to resample vertical soil profile data, and a random forest (RF) model was used to produce the depth-to-soil pH constraint map. The RF model was accurate, with a Lin's Concordance Correlation Coefficient (LCCC) of 0.63-0.66, and a Root Mean Square Error (RMSE) of 0.47–0.51 when testing with leave-one-site-out cross-validation. Approximately 77% of the farm was found to be constrained by a strongly alkaline pH greater than 9  $(H_2O)$  somewhere within the top 100 cm of the soil profile. The relationship between the predicted depth-to-soil pH constraint map and cotton and grain (wheat, canola, and chickpea) yield monitor data was analyzed for individual fields. Results showed that yield increased when a soil pH constraint was deeper in the profile, with a good relationship for wheat, canola, and chickpea, and a weaker relationship for cotton. The overall results from this study suggest that the modelling approach is valuable in identifying the depth-to-soil pH constraint, and could be adopted for other important subsoil constraints, such as sodicity. The outputs are also a promising opportunity to understand crop yield variability, which could lead to improvements in management practices.

Keywords: soil constraints; digital soil mapping; yield variability; precision agriculture

# 1. Introduction

Many soils possess a chemical or physical characteristic that constrains crop production, with the mechanism of growth inhibition varying depending on the nature of the constraint. The soils of the alluvial grain and 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 high levels of salinity, sodicity, alkalinity, and compaction [1], particularly in the subsoil, but still within the rooting zone of crops. Soil salinity [2,3], sodicity [4,5], and compaction [6,7], and their impact on crops in these regions have received much attention however, there has been little emphasis on soil alkalinity constraints.



While subsoil acidity is a widespread problem in other parts of Australia [8], subsoil alkalinity is a more localized issue, and has consequently received less consideration in the literature. Many of the soils in the cotton-growing valleys of eastern Australia are intrinsically alkaline due to the presence of calcium and sodium carbonates, and this generally increases with depth [9]. Soils with a high pH can limit the plant availability of certain nutrients, as well as cause toxicities [10]. At toxic levels, root growth is inhibited, decreasing the amount of water and nutrients that can be accessed, producing a detrimental impact on crop production [1]. The optimum pH (H<sub>2</sub>O) range for most crops is generally from 5.5 to 7, but this differs for individual crop types [11]. In soil with a pH (H<sub>2</sub>O) of nine or greater, it is likely that most crops would experience boron toxicity [12], as well as a considerable reduction in the availability of important macro and micronutrients, including phosphorus, nitrogen, calcium, magnesium, iron, copper, and zinc [13]. A pH (H<sub>2</sub>O) greater than nine also suggests the presence of sodium carbonates and high levels of soil sodicity, which indicates poor structural condition [12].

Soil pH, and the depth at which a soil pH constraint is reached, can significantly vary across fields [14], and this can result in spatially variable crop yields [15]. Digital soil mapping (DSM) has been used extensively for evaluating the spatial distribution of soil pH at a range of spatial extents, from the field [13], to the world [16]. Some studies solely produce pH maps for the topsoil (e.g., Kirk et al. [17]), which lacks important information about subsoil pH. Other studies create maps of several depth increments (e.g., Filippi et al. [18]), but this is often too much information from which to easily make management decisions. The typical broadacre farmer in Australia now has access to a plethora of data and information that relate to their farms [19], which is a great opportunity, but also a challenge to gain useful insights. 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 assist in identifying areas where crop rooting depth may be inhibited, and help farmers implement management practices to either rectify the issue or alter inputs according to the constrained production potential. It is important to know the depth at which a pH constraint is reached within a fine vertical resolution (e.g., 1 to 10 cm). However, most soil maps are produced at coarse resolutions, particularly deeper in the soil profile. For example, the Globalsoilmap specifications are at 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm [16], and this creates difficulty in identifying exactly at what depth pH is a limiting factor.

There has been limited focus in the literature on mapping the depth of soil constraints. The study reported here uses a novel modelling approach to predict and map the depth-to-soil pH constraint (high alkalinity: pH > 9 (H<sub>2</sub>O)) on a dryland cotton and grain farm in the Namoi Valley in northern New South Wales (NSW), Australia. A collection of spatial and temporal datasets were used for modelling and mapping the depth-to-soil pH constraint, and the relationship with grain and cotton yield mapping data was assessed.

#### 2. Methods

#### 2.1. Study Site

This study was conducted on a mixed farming property—L'lara (30°15′18″ S, 149°51′39″ E), which is located near Narrabri in the Namoi Valley in northern NSW, Australia. Summers in Narrabri are very hot, while winters are cool. The long-term average precipitation for the study area is 658.5 mm, and is summer-dominant [20]. The farm consists of 780 hectares of uncropped dryland grazing on primarily native perennial pastures, and 1070 hectares of summer and winter dryland broadacre cropping (Figure 1). The cropped area was the focus of this study, where cotton (*Gossypium hirsutum* L.), and winter wheat (*Triticum aestivum* L.) are the dominant crops grown, with additional rotations of canola (*Brassica napus* L.), and chickpea (*Cicer arietinum* L.). The soils of the cropping fields at L'lara are classed as grey or brown Vertosols according to the Australia Soil Classification [21]. These soils are primarily derived from alluvial deposits of basaltic sediments from the western side of the Nandewar range.



**Figure 1.** The arrangement of the cropping fields at L'lara, and the location within northern New South Wales (NSW), Australia.

#### 2.2. Legacy Soil Data

Soil data from 80 locations sampled across the cropping fields in July 2017 were available. Two subsamples were extracted from the 0–0.1- and 0.4–0.5-m depth increments of each sampling location, resulting in a total of 160 subsamples. All soil subsamples were air dried, and ground to <2 mm fraction. Soil pH was analyzed in 1:5 soil:water extracts using a Mettler Toledo S220 SevenCompact<sup>™</sup> pH/Ion meter (Mettler Toledo, Colombus, OH, USA).

# 2.3. Additional Soil Data

Soil samples at an additional 30 sites were also collected. The locations of these soil sampling sites were determined with a stratified random sampling methodology utilizing k-means clustering for strata identification [22]. The spatial and temporal data used and the details of the clustering analysis and sampling/analysis are described in the following two subsections.

## 2.3.1. Data Used in Clustering Analysis

To guide where additional soil samples would be taken, spatial clusters were created using publicly available data. These data consisted of digital soil maps, a digital elevation model, air-borne gamma radiometric data, and remotely sensed satellite imagery. Digital maps of soil clay content at 90-m resolution for the 0-5-, 5-15-, 15-30-, 30-60-, and 60-100-cm depth increments were obtained from the Soil Landscape Grid of Australia (SLGA) [23]. Hydrologically corrected digital elevation model (DEM) data was used from the Shuttle Radar Topography Mission (SRTM) [24]. Air-borne gamma radiometrics data for the study area at 90-m spatial resolution were obtained from Geoscience Australia. Low-pass filtering was used on all radiometric products [25]. Google Earth Engine (GEE) [26] was used to access Landsat 7 Tier 1 surface reflectance satellite imagery. A cloud-masking filter was applied to remove all pixels that were affected by cloud cover, and the 10th, 50th and 90th percentile statistics were calculated from all images that fell within 1 January 2000, to 31 December 2017. These statistics were selected to not only represent the most common value, but the lower and upper distribution of the imagery. From this, two outputs were extracted; (1) the Normalized Difference Vegetation Index (NDVI), and (2) the red band. The NDVI was used to map changes in biomass throughout the production history and build a pattern of production potential across the farm. The red band was used separately as a representation of topsoil color. All spatial data were extracted onto a single 30-m grid using the nearest neighbor method.

#### 2.3.2. Clustering Analysis and Site Selection

To create clusters of the study area with the data sources described above, k-means clustering for strata identification was implemented using the statistical program JMP<sup>®</sup> Pro 11 (SAS Institute, Cary, NC, USA). The optimum number of clusters for the study area was found to be eight (Figure 2), and this was determined via the elbow method [27]. The elbow method looks at the relationship between the number of clusters and the proportion of variance explained and helps to identify the point where adding another cluster would provide marginal gain. The area of each cluster varied from 39 to 228 ha. The 30 soil sampling locations were selected randomly within each cluster, with the number of sites per cluster proportional to the area of each cluster (Figure 2).



**Figure 2.** The distribution of the eight clusters derived from a k-means clustering analysis, and the locations of the soil sampling sites across the study area.

#### 2.3.3. Sampling Details

Soil cores at the 30 sites were extracted to a 1.0-m depth in July 2018. The cores were then subdivided into five depth intervals (0–0.1, 0.1–0.3, 0.3–0.6, 0.6–0.8, and 0.8–1.0 m), resulting in a total of 150 subsamples. Equally to the legacy data, the subsamples were then air dried, and ground to <2 mm fraction, and soil pH analyzed in 1:5 soil:water extracts with a Mettler Toledo S220 SevenCompact<sup>™</sup> pH/Ion meter (Mettler Toledo, Colombus, OH, USA).

# 2.4. Vertical Depth Modelling/Resampling of Profiles

Equal-area quadratic smoothing splines were fitted to the soil pH data for each of the 110 individual soil sampling sites [28,29]. This was implemented using the 'ithir' package in R [30]. Rather than aggregating this for a few standard depth intervals as is common, data were stored at the 1-cm resolution in which it was fitted. This produced estimates of pH at 1-cm increments down each soil profile. Data were only predicted between the surface and deepest depth sampled. Consequently, the 30 soil profiles with 100 cm extracted resulted in 100 pH estimates per profile, and the 80 sites extracted to 50 cm resulted in 50 pH estimates per profile.

# 2.5. Modelling and Mapping of Soil pH

#### 2.5.1. Data used for Modelling/Mapping Soil pH

The data used for modelling/mapping soil pH across the study area consisted of data collected on-farm, and data from publicly available databases (Table 1; Figure 3). A proximal soil sensing survey was conducted to collect high-resolution apparent soil electrical conductivity (ECa) and gamma radiometrics data. Soil ECa was measured via electromagnetic induction using a DUALEM-21S instrument (Dualem Inc., Milton, ON, Canada). Gamma radiometric data were recorded using an RSX-1 gamma radiometric detector with a 4 L sodium iodine crystal (Radiation Solutions Inc., Mississauga, ON, Canada). The proximal soil sensing survey was conducted on 24-m swaths, and the position was recorded with differential GPS (DGPS) equipment. Continuous surface layers were obtained by kriging with local variograms [31] onto a standard 5-m grid through the software VESPER [32]. A 5-m resolution DEM was also obtained from Spatial Services, NSW Government [33]. The 90th percentile red band image from Landsat 7 described in the previous section was also used in this analysis. All spatial data were then extracted onto a single 5-m resolution grid using the nearest neighbor method.

		Data Type	Data Description	Resolution
	Ore former	ECa	Dual EM-21S (0-3.0 m)	5 m
	On-farm	Gamma radiometrics	Potassium (K)	5 m
	Dublic	Elevation	DEM	5 m
	Public	Landsat 7—red band	90th percentile (2000-2017)	30 m
(a)	E	Ca (mS/m) 237.0 88.0	(b) Gam	ma K (%) 7.75 0.15
(c)		DEM (m) 276.7 225.0 N	(d) Red	band 90 <sup>th</sup> 1813.0 1111.4

TT 1 1 1 D	6.1 1.	1 • .1	•	1 .
lable L. Description	of the data sources	used in the	manning ana	IVS1S
Tuble It Debeliption	of the data bources	abea m me	mapping and	1,010.

Figure 3. The data sources used for digital soil mapping. (a) Soil ECa (0-3.0 m) from a Dual EM-21S, (b) Gamma radiometrics K, (c) DEM, and (d) the 90th percentile (2000–2017) red band from Landsat 7.

# 2.5.2. Modelling/Mapping Procedure

At each of the 110 soil sampling locations, the corresponding on-farm and publicly available data described in Table 1 were collated. A model to predict the relationship between soil pH at 1-cm increments and these covariates was then created using a random forest model [34]. The 'ranger' package in the software R was used to create the model, which is essentially a fast implementation of random forests for high dimensional data [35]. Instead of one model for each depth increment, 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. More specifically, the central depth between the upper and lower depths was used (e.g., 0.5 cm for the 0–1-cm depth interval, and 99.5 for the 99–100-cm depth interval). This model was then used to predict onto the 5-m grid of the study area using the spatially distributed covariate dataset. 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 identified for each 5-m grid point. A pH of 9 ( $H_2O$ ) was selected, as this is an indicator of when significant constraints to growth for most crops is reached as described previously. This information was then mapped across the study area, showing the depth-to-high soil pH constraint.

The quality of the model was tested by using leave-one-site-out cross-validation (LOSOCV). This entailed iteratively 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 at all 110 sites, with 109 sites used to predict the remaining site each time. The results of the validation at every site were then combined, and the Lin's concordance correlation coefficient (LCCC) [36] and the RMSE (root-mean-square error) were used to assess the quality of the model predictions. This cross-validation procedure was performed in two ways: (1) at 1-cm vertical resolution (the splined observed pH data vs. the independently predicted pH data) and (2) at the original sampling depths of the observed data (observed pH data vs. predicted pH data at a 1-cm resolution aggregated to the original sampling depths). The rationale for this was that the splining procedure introduces some amount of uncertainty to the data [37] and validating by the second approach avoids this limitation as the predicted data are compared to the original, un-splined soil pH 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). This shows the amount by which the random forest model prediction accuracy would decrease if that particular variable is excluded. The larger the mean decrease in accuracy for a predictor variable, the more important that variable is deemed. All data analysis was performed in the open-source software R [38].

# 2.6. Crop Yield Data and the Relationship with Depth-to-Soil pH Constraint

# 2.6.1. Crop Yield Data

Yield data from an on-harvester monitoring system was used to analyze the relationship with the depth-to-soil pH constraint. This yield monitor data consisted of a variety of broadacre crops from two fields—Campey 4/5 (C4/5) which is 61 ha in size, and L'lara 2 (L2) which is 183 ha. For C4/5, canola was grown in 2016, and chickpea in 2017 (Figure 4a). For L2, wheat was grown in 2016, and a summer cotton crop in 2017/2018 (Figure 4b). The raw yield monitor data were processed by removing spurious and extreme values following the method of Taylor et al. [39]. VESPER software was then used to krige the yield data onto the same 5-m grid as the soil modelling/mapping covariate data. The final surfaces were corrected and standardized using field total yields (grain weight at silo, or sum of bales of cotton harvested). A 30-m internal buffered zone was applied around the paddock boundaries to remove the low-yielding edge effects, and a 20-m buffer was also placed on the contour banks in L2 (Figure 4a,b). The rationale for removing these yield values was because the cause of low yield in these areas would likely be related to other factors as well as potential high soil pH.



Figure 4. Cont.



**Figure 4.** (a), Canola yield for the 2016 season (left), and chickpea yield for the 2017 season (right) from a yield monitor for field Campey 4/5 (C4/5). (b), Wheat yield for the 2016 season (left), and cotton yield for the 2017/18 season (right) from a yield monitor for field L'lara 2 (L2).

2.6.2. Relationship of Crop Yield Data with Depth-to-Soil pH Constraint Data

The predicted depth-to-high soil pH constraint data and the yield data were mapped on the same 5-m grid, allowing the relationship to be easily analyzed. Boxplots were used to assess this, where data were grouped in 10-cm intervals (depth-to-soil pH constraint), showing how the distribution of yield data changed as the depth-to-soil pH constraint changed for each paddock and crop/season. The Spearman rank-order correlation coefficient,  $r_s$ , was also used to assess this relationship. Spearman's correlation is a nonparametric measure of the direction and strength of a relationship that exists between two variables based on their ordered ranks. The median yield value was calculated for each 1-cm depth interval, and the  $r_s$  was then reported.

# 3. Results

#### 3.1. pH Variability in the Study Area

Soil pH values varied considerably across the study area in the top 100 cm, with a low of 6.1 at 0–10 cm, and a high of 10.2 at 30–60 cm (Figure 5). As depth increased, median soil pH values increased, and the variability of observations decreased. Soil pH was generally alkaline, with a mean soil pH of 8.2 in the shallowest layer (0–10 cm), rising to a mean pH of 9.3 in the deepest layer (80–100 cm).



**Figure 5.** Boxplots of the distribution of soil pH (1:5 H<sub>2</sub>O) with depth for 110 soil cores taken from cropping fields at L'lara. Solid black lines inside boxes represent median values, and the dashed, vertical grey line indicates the pH (H<sub>2</sub>O) threshold of nine, deemed to have potential constraining effects on crop growth.

#### 3.2. Depth-to-Soil pH Constraint

The depth-to-high soil pH constraint map showed considerable spatial variability across the study area (Figure 6). In total, 77% of the cropping fields had a strongly alkaline pH of nine somewhere within the top 100 cm (Table 2). 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 in constraint conditions, with areas that were constrained in the top 1–10 cm, and unconstrained within 100 cm, all within a short distance.



Figure 6. Digital soil map of the depth in which soil pH constraint (>9) is reached across the study area.

Depth (cm)	0–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80	81–90	91–100	Unconstrained within 100 cm
Area (ha)	11	16	183	339	128	51	23	32	37	4	245
Area (%)	1.0	1.5	17.1	31.7	12.0	4.8	2.2	3.0	3.5	0.3	23.0

# 3.3. Model Quality and Variable Importance

For the random forest model, the validated results showed an LCCC of 0.63, and an RMSE of 0.47 using LOSOCV when tested on the splined 1 cm data, and an LCCC of 0.66 and RMSE of 0.51 using LOSOCV when tested on the aggregated data at the original sampling depths (Table 3). These statistics suggest that the model could spatially predict soil pH accurately (within ~0.5 pH units). Visual examples of the validation of soil pH predictions at 1-cm increments are shown in Figure 7, where Figure 7a shows the mean observed and predicted soil pH of all sampling sites, and Figure 7b for a single randomly selected sampling site (Site 16). The soil pH predictions in these figures were created using an independent calibration. This demonstrates that the random forest model can predict the vertical distribution of soil pH.

**Table 3.** Validation prediction statistics for the soil pH model using leave-one-site-out cross-validation (LOSOCV) at a splined 1-cm resolution, and at the original sampling depth resolution.

Validation Resolution	Lin's Concordance Correlation Coefficient (LCCC)	Root Mean Square Error (RMSE)		
Splined 1 cm Original sampling depth	0.63	0.47		

0

2

40

80

8

Depth (cm)





80

80

and independently validated predicted data from the random forest model (blue circles) for the mean of all soil sites (**a**) and an example soil profile, Site 16 (**b**). The observed soil pH data at the sampled depth increments (solid black lines) for Site 16 is also shown in (**b**). The red dashed line represents the pH threshold of nine (H<sub>2</sub>O).

The plot of the predictor variable importance showed that proximally sensed gamma radiometric K was the most important predictor of soil pH (Figure 8). This was closely followed by depth, and then other horizontally variable predictors, such as DEM, and soil ECa. Landsat 7 data of the 90th percentile red band from the period 2000–2017 proved to be the least important predictor of soil pH, suggesting that this only represents a small component of soil spatial variation.



Figure 8. Mean decrease in accuracy showing the predictor variable importance in predicting soil pH.

# 3.4. The Relationship with Crop Yield and the Depth-to-Soil pH Constraint

Two fields, C4/5 and L2, were used to assess the relationship between yield monitor data and the predicted depth-to-soil pH constraint data. These fields were selected due to the availability of yield data, and the variation in the depth in which a pH of 9 was predicted across the field (Figure 6). Boxplots were used to describe this relationship, with the grouping of yield data at 10-cm depth-to-soil pH constraint intervals (Figure 9a,b). The 1–10 and 11–20 cm (to constraint) data are not used due to the very few locations reaching a pH greater than 9 in the upper 20 cm. It was clear that as a soil pH

constraint (>9) was deeper in the soil profile, the yield of most crops increased (Figure 9a,b). For all grain crops, the lowest median yield was observed where a soil pH constraint was reached in the shallowest, well-observed layer (21–30 cm), and the highest median yield was observed where soil was deemed unconstrained by pH in the top 100 cm. A clear spatial correlation can also be seen between the grain crop yield maps (Figure 4a,b) and the depth-to-soil pH constraint map (Figure 6). For cotton, the lowest median yield was observed where a soil pH constraint was reached in the shallowest layer (21–30 cm), and an increase in median yield was observed as the depth-to-soil pH constraint increased to 61–70 cm. However, this plateaued and slightly declined after this depth. The Spearman's correlation analysis revealed that the relationship between the predicted depth-to-soil pH constraint data and yield monitor data ranged from strong to weak (Table 4). 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 ( $r_s = 0.37$ ).



**Figure 9.** (a) Boxplots of the relationship of crop yield data with the predicted depth-to-high soil pH constraint data for field Campey 4/5. (b) Boxplots of the relationship of crop yield data with the predicted depth-to-high soil pH constraint data for field L'lara 2.

**Table 4.** The Spearman's correlation ( $r_s$ ) of the relationship between the depth-to-soil pH constraint map data and the median yield value for each 1-cm depth increment to 100 cm.

	Cam	pey 4/5	L′1	ara 2
	Canola '16	Chickpea '17	Wheat '16	Cotton '17/18
r <sub>s</sub>	0.66	0.58	0.75	0.37

#### 4. Discussion

# 4.1. Soil Alkalinity and Depth-to-Soil pH Constraint

Soil pH was generally alkaline and increased with depth. This is typical of Vertosols in the cotton-growing valleys, and other studies in similar areas concur [18,40]. The depth in which a soil pH greater than nine was reached was shown to be highly spatially variable across the study area. Over half of the total area reached a pH of 9 within the shallow subsoil (21–40 cm), and approximately 77% of the area possessed an alkalinity constraint somewhere in the top 100 cm of the soil profile. Soil alkalinity is often an inherent feature of the soils in these landscapes. However, it is possible that the distribution of soil pH is being impacted by cultivation practices, which could be bringing the more alkaline subsoils closer to the surface. Alkalinity in these soils are generally due to the presence of calcium carbonates, but the extremely high pH values (>9) also indicate that sodium carbonates are present. The presence of sodium carbonates also suggests that these soils possess high levels of soil sodicity [41]. While alkalinity can reduce nutrient accessibility, cause toxicities, and inhibit root growth for crops, soil sodicity is also known to have adverse impacts on crop productivity through waterlogging and soil structural decline.

# 4.2. Modelling/Mapping

#### 4.2.1. Modelling/Mapping Approach and Validation

The random forest model could predict the spatial distribution of soil pH well, and the approach proved successful in identifying the depth-to-soil pH constraint at a 1-cm vertical resolution to a 100-cm depth. The LOSOCV on the splined 1-cm data showed an LCCC of 0.63, and an RMSE of 0.47, and an LCCC of 0.66 and RMSE of 0.51 when tested on the data at the original sampling depths. The fact that these two cross-validation techniques showed very similar validation statistics is optimistic. The second cross-validation approach validates predictions with the original, un-splined data, which suggests that the splining procedure is creating a relatively small amount of uncertainty in the data, giving confidence in the developed mapping approach.

While this study was performed on a 1070-ha farm, there is promise for implementing this approach over larger areas. However, the fine-resolution splining of soil property data results in a much larger amount of data compared to typical DSM methods, and implementing this with very large datasets could become computationally restrictive. One opportunity to reduce the computational load of this approach is to fit the soil property at 5-cm increments with the splines as opposed to 1-cm increments. This would still be at a fine enough vertical resolution to inform management decisions for land managers but would significantly reduce the dataset size. An alternative would be to use the approach of Orton et al. [37] who presented a geostatistical approach to predict in 3D at any vertical resolution using observations from different vertical supports (e.g., soil horizons). This would remove the need for using splines to create the finely spaced vertical dataset. Adopting the approach by Orton et al. [37] also bypasses the uncertainty introduced by the splining process, although, as discussed, the results in the current study suggested that this was relatively small (Table 3). The uncertainty in imputed values is commonly ignored in DSM studies that use splined soil data.

Few studies have used DSM approaches to map the depth-to-soil constraints, but there has been considerable research in mapping the depth to bedrock [42–44]. The methods used in these studies are essentially traditional DSM approaches, and differ significantly from the current study. For example, data for the depth to which bedrock is reached is commonly available from mining exploration and bore hole drilling exercises, and Wilford et al. [44] used a database of these observations to create depth-to-bedrock maps of Australia using a Cubist model. In contrast, rather than use direct observations of the depth of the target variable/constraint, the current study uses splined soil pH data, and machine learning to predict the depth in which an alkalinity constraint is reached at each 1-cm vertical increment, which is then combined to create a depth-to-soil pH constraint map. It is typically

not easy to identify the depth at which a soil constraint occurs in the field, but this approach helps overcome this. Another advantage of the approach used in this study is that it could be applicable to any soil property, not just soil alkalinity.

#### 4.2.2. Predictor Variables

Only four spatial predictor variables were used for modelling and mapping (Table 1), whereas most other DSM studies use a much larger suite of predictor variables. The rationale for this was that the variables used in the current study were of a fine spatial resolution (5–30 m), with the on-farm collected data (ECa and gamma radiometrics) known to reflect soil spatial variability very well. The most important covariate for the soil pH random forest model was proximally sensed gamma radiometrics K (5 m). This covariate is often highly correlated with soil type and represents fine-scale soil spatial variability due to intensive measurements collected in this study. The second most important covariate for the soil pH random forest model was depth, which is logical, as pH in these soils increases as depth increases. The model could predict the vertical distribution of soil pH well, and this is due to the inclusion of depth as a predictor variable. This concept is demonstrated in Figure 7. The DEM data (5 m) represented broader trends across the study area and possessed lower short-range spatial variability (Figure 4), but this proved to be an important predictor variable. The ECa data (5 m) was the second least important predictor, as the high spatial variability (Figure 4) did not seem to represent the spatial variation of soil pH particularly well. The least important predictor was the 90th percentile red band (30 m), which could be due to the coarser spatial resolution compared to the other predictor variables, and that it did not sufficiently reflect the spatial distribution of soil pH.

# 4.3. Relationship between Depth-to-Soil pH Constraint and Crop Yield

The relationship between the depth-to-soil pH constraint map and cotton and grain yield was explored for fields C4/5 and L2. It was clear that the deeper in the soil profile a pH constraint greater than nine was reached, the yield of all crops generally increased. The relationship between yield and the depth-to-soil pH constraint data was stronger for the grain crops than for cotton. The Spearman's correlation was highest for wheat ( $r_s = 0.75$ ), followed by canola ( $r_s = 0.66$ ), chickpea ( $r_s = 0.58$ ), and then cotton ( $r_s = 0.37$ ). This suggests that cotton was relatively unaffected by an alkalinity constraint compared to the grain crops. Cotton is almost solely grown in the alkaline Vertosols of the alluvial valleys of eastern Australia and is therefore well-accustomed to these soils. This could be a possible reason for the weaker relationship. However, crop yield is a function of climate, soil, and management, and their interaction, and this could vary from season to season [45]. Future research should focus on using a longer time-series of yield data to assess the variation in the relationship between yield and the depth-to-soil pH constraint to account for this. The approach developed in this study should also be applied to a larger spatial extent in the future, as this information could be useful in identifying constraints to yield at a broader scale to guide policy decisions.

In field C4/5, similar spatial patterns were observed for canola in 2016 and chickpea in 2017. In contrast, the wheat yield maps in 2016 and the cotton yield maps in 2017/2018 showed some differing spatial patterns, with some of the highest yielding areas for cotton being the lowest yielding for wheat. These inconsistently yielding or flip-flop patterns have been commonly observed in other studies [45–47], and often point to a temporary constraint, rather than a permanent soil constraint such as high soil pH levels. Despite the important role that soil pH plays on crop productivity, few studies have looked at the impact of within-field soil pH variability on crop yield, let alone how the depth-to-soil pH constraint relates to this. Shatar and McBratney [48] assessed the relationship between soil pH (15–30 cm) and sorghum yield, and found that pH was a limiting factor in some areas of the field. A study by Schepers et al. [49] also found that management zones (MZ) with varying pH levels within a field displayed a marked difference in corn grain yield over several seasons. The MZ with the lowest mean pH (6.41) showed the highest yield, and the MZ with the highest mean pH

(7.43) displayed the lowest yields. In contrast, Adeoye and Agboola [50] saw a significant positive correlation between the soil pH and the relative yield of maize.

The results found in this study suggest that the depth-to-soil pH constraint can directly influence crop yield. However, as aforementioned, these very high levels of soil pH are a strong indicator that soil sodicity is also an issue. Soil sodicity is a common issue in the alluvial Vertosols of eastern Australia [4,51,52] and is deemed to be one of the biggest constraints to crop production in these regions [1]. High sodicity levels of soil have a greater capacity to inhibit root growth than high pH levels. This results in a smaller volume of soil that is accessible to crops, and consequently a smaller amount of water and nutrients available for crop development [53]. Further research should look at soil sodicity and the impact it has on crop yield variability in the study area. Future work should also use the approach presented to estimate the reduction in the soil water holding capacity and therefore yield potential, caused by soil constraints.

# 5. Conclusions

High levels of soil alkalinity were observed in the study area, particularly at depth. The random forest model to predict soil pH distribution showed high quality predictions when testing with LOSOCV, with an LCCC of 0.63–0.66, and an RMSE of 0.47–0.51. The overall approach to identify the depth at which a pH alkalinity constraint (>9) occurred at fine vertical resolution (1 cm) at a farm scale proved successful and showed promise for identifying other important subsoil constraints. The study revealed that the shallower in the soil profile a pH constraint was reached, the generally smaller the crop yield. A strong relationship was found for wheat, canola, and chickpea, with a weaker relationship for cotton. 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 crop management inputs to account for these constraints.

Author Contributions: Conceptualization, P.F. and T.F.A.B.; Formal analysis, P.F., E.J.J. and B.J.G.; Investigation, B.J.G.; Methodology, P.F., E.J.J., B.M.W. and T.F.A.B.; Project administration, P.F.; Resources, G.W.R.; Validation, P.F. and T.F.A.B.; Writing—original draft, P.F. and B.J.G.; Writing—review and editing, P.F., E.J.J., B.M.W., G.W.R. and T.F.A.B.

**Funding:** This research was partly funded by grants provided by the Cotton Research and Development Corporation (DAN1801, US1802), the Grains Research and Development Corporation (US00087), and the University of Sydney.

**Acknowledgments:** Thank you to the two anonymous reviewers for their helpful feedback, and to the Commercial Cropping Supervisor of L'lara—Kieran Shephard, for his assistance in gathering the data.

Conflicts of Interest: The authors declare no conflict of interest.

#### References

- Dang, Y.P.; Dalal, R.C.; Routley, R.; Schwenke, G.D.; Daniells, I. Subsoil constraints to grain production in the cropping soils of the north-eastern region of Australia: An overview. *Aust. J. Exp. Agric.* 2006, 46, 19–35. [CrossRef]
- Odeh, I.O.A.; Todd, A.J.; Triantafilis, J.; McBratney, A.B. Status and trends of soil salinity at different scales: The case for the irrigated cotton growing region of eastern Australia. *Nutr. Cycl. Agroecosyst.* 1998, 50, 99–107. [CrossRef]
- 3. Odeh, I.O.A.; Onus, A. Spatial analysis of soil salinity and soil structural stability in a semiarid region of New South Wales, Australia. *Environ. Manag.* **2008**, *42*, 265–278. [CrossRef] [PubMed]
- 4. McKenzie, D.C.; Abbott, T.S.; Chan, K.Y.; Slavich, P.G.; Hall, D.J.M. The nature, distribution and management of sodic soils in New-South-Wales. *Soil Res.* **1993**, *31*, 839–868. [CrossRef]
- Dodd, K.; Guppy, C.; Lockwood, P.; Rochester, I. The effect of sodicity on cotton: Plant response to solutions containing high sodium concentrations. *Plant Soil* 2010, 330, 239–249. [CrossRef]
- 6. McGarry, D. Soil compaction and cotton growth on a Vertisol. Soil Res. 1990, 28, 869–877. [CrossRef]

- 7. Antille, D.L.; Bennett, J.M.; Jensen, T.A. Soil compaction and controlled traffic considerations in Australian cotton-farming systems. *Crop Pasture Sci.* 2016, *67*, 1–28. [CrossRef]
- De Caritat, P.; Cooper, M.; Wilford, J. The pH of Australian soils: Field results from a national survey. *Soil Res.* 2011, 49, 173–182. [CrossRef]
- 9. Knowles, T.A.; Singh, B. Carbon storage in cotton soils of northern New South Wales. *Soil Res.* 2003, *41*, 889–903. [CrossRef]
- 10. Läuchli, A.; Grattan, S.R. Soil pH extremes. Plant Stress Physiol. 2012, 194-209.
- 11. Hazelton, P.; Murphy, M. Interpreting Soil Test Results: What Do All the Numbers Mean? CSIRO Publishing: Victoria, Australia, 2007.
- Slattery, W.J.; Conyers, M.K.; Aitken, R.L. Soil pH, aluminium, manganese and lime requirement. In *Soil Analysis: An Interpretation Manual*; Peverill, K.I., Sparrow, L.A., Reuter, D.J., Eds.; CSIRO: Collingwood, Australia, 1999; pp. 103–128.
- 13. McKenzie, N.; Jacquier, D.; Isbell, R.; Brown, K. *Australian Soils and Landscapes: An Illustrated Compendium*; CSIRO Publishing: Clayton, Australia, 2004.
- 14. Adamchuk, V.I.; Lund, E.D.; Reed, T.M.; Ferguson, R.B. Evaluation of an on-the-go technology for soil pH mapping. *Precis. Agric.* **2007**, *8*, 139–149. [CrossRef]
- 15. Taylor, J.C.; Wood, G.A.; Earl, R.; Godwin, R.J. Soil factors and their influence on within-field crop variability, Part II: Spatial analysis and determination of management zones. *Biosyst. Eng.* **2003**, *84*, 441–453. [CrossRef]
- Arrouays, D.; Grundy, M.G.; Hartemink, A.E.; Hempel, J.W.; Heuvelink, G.B.M.; Hong, S.Y.; Lagacherie, P.; Lelyk, G.; McBratney, A.B.; McKenzie, N.J.; et al. Globalsoilmap: Toward a fine-resolution global grid of soil properties. *Adv. Agron.* 2014, *125*, 93–134.
- 17. Kirk, G.J.D.; Bellamy, P.H.; Lark, R.M. Changes in soil pH across England and Wales in response to decreased acid deposition. *Glob. Chang. Biol.* **2010**, *16*, 3111–3119. [CrossRef]
- 18. Filippi, P.; Cattle, S.R.; Bishop, T.F.A.; Odeh, I.O.A.; Pringle, M.J. Digital soil monitoring of top- and sub-soil pH with bivariate linear mixed models. *Geoderma* **2018**, *322*, 149–162. [CrossRef]
- 19. Bramley, R.G.V.; Ouzman, J. Farmer attitudes to the use of sensors and automation in fertilizer decision-making: Nitrogen fertilization in the Australian grains sector. *Precis. Agric.* **2018**. [CrossRef]
- 20. Bureau of Meteorology. Monthly climate statistics—Narrabri West Post Office (053030). Available online: http://www.bom.gov.au/climate/averages/tables/cw\_053030.shtml (accessed on 18 December 2018).
- 21. Isbell, R.F. The Australian Soil Classification; CSIRO Publishing: Melbourne, Australia, 1996.
- Boydell, B.; McBratney, A.B. Identifying potential management zones from cotton yield estimates. *Precis. Agric.* 2002, 3, 9–23. [CrossRef]
- 23. Grundy, M.J.; Rossel, R.V.; Searle, R.D.; Wilson, P.L.; Chen, C.; Gregory, L.J. Soil and landscape grid of Australia. *Soil Res.* 2015, *53*, 835–844. [CrossRef]
- 24. Geoscience Australia. Geoscience Australia, 1 Second SRTM Digital Elevation Model (DEM). Bioregional Assessment Source Dataset. Available online: http://data.bioregionalassessments.gov.au/dataset/9a9284b6-eb45-4a13-97d0-91bf25f1187b (accessed on 4 December 2018).
- 25. Minty, B.R.S.; Franklin, R.; Milligan, P.R.; Richardson, L.M.; Wilford, J. The radiometric map of Australia. *Explor. Geophys.* **2009**, *40*, 325–333. [CrossRef]
- 26. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
- 27. Kodinariya, T.M.; Makwana, P.R. Review on determining number of Cluster in K-Means Clustering. *Int. J.* **2013**, *1*, 90–95.
- 28. Bishop, T.F.A.; McBratney, A.B.; Laslett, G.M. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* **1999**, *91*, 27–45. [CrossRef]
- 29. Malone, B.P.; McBratney, A.B.; Minasny, B.; Laslett, G.M. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* **2009**, *154*, 138–152. [CrossRef]
- Malone, B. Ithir: Soil Data and Some Useful Associated Functions, R Package Version 1.0. 2015. Available online: https://rdrr.io/rforge/ithir/ (accessed on 20 March 2019).
- 31. Haas, T.C. Kriging and automated variogram modelling within a moving window. *Atmos. Environ.* **1990**, *24*, 1759–1769. [CrossRef]

- 32. Whelan, B.M.; McBratney, A.B.; Minasny, B. Vesper 1.5–Spatial Prediction Software for Precision Agriculture. In Proceedings of the 6th International Conference on Precision Agriculture, ASA/CSSA/SSSA, Madison, WI, USA, 14–17 July 2002; ASA/CSSA/SSSA: Madison, WI, USA, 2002; Volume 179.
- 33. Department of Finance, Services and Innovation. NSW Foundation Spatial Data Framework-Elevation and Depth-Digital Elevation Model. Available online: https://data.nsw.gov.au/data/dataset/8f73f5ca-4f7f-4707-bfe2-0efbb9027107 (accessed on 4 December 2018).
- 34. Breiman, L. Random forests. Mach. Learn. 2001, 45, 5-32. [CrossRef]
- 35. Wright, M.N.; Ziegler, A. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J. Stat. Softw. 2017, 77, 1–17. [CrossRef]
- 36. Lin, L.I.K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **1989**, *45*, 255–268. [CrossRef]
- 37. Orton, T.G.; Pringle, M.J.; Bishop, T.F.A. A one-step approach for modelling and mapping soil properties based on profile data sampled over varying depth intervals. *Geoderma* **2016**, *262*, 174–186. [CrossRef]
- 38. R Core Team. *R: A Language and Environment for Statistical Computing;* R Foundation for Statistical Computing: Vienna, Austria, 2017.
- 39. Taylor, J.A.; McBratney, A.B.; Whelan, B.M. Establishing Management Classes for Broadacre Agricultural Production. *Agron. J.* **2007**, *99*, 1366–1376. [CrossRef]
- 40. Singh, B.; Odeh, I.O.A.; McBratney, A.B. Acid buffering capacity and potential acidification of cotton soils in northern New South Wales. *Soil Res.* **2003**, *41*, 875–888. [CrossRef]
- 41. Day, A.D.; Ludeke, K.L. Soil Alkalinity. In *Plant Nutrients in Desert Environments*; Adaptations of Desert Organisms; Springer: Berlin/Heidelberg, Germany, 1993.
- 42. Shafique, M.; Meijde, M.; Rossiter, D.G. Geophysical and remote sensing-based approach to model regolith thickness in a data-sparse environment. *Catena* **2011**, *87*, 11–19. [CrossRef]
- 43. Wilford, J.R.; Thomas, M. Predicting regolith thickness in the complex weathering setting of the central Mt Lofty Ranges, South Australia. *Geoderma* **2013**, *206*, 1–13. [CrossRef]
- 44. Wilford, J.R.; Searle, R.; Thomas, M.; Pagendam, D.; Grundy, M.J. A regolith depth map of the Australian continent. *Geoderma* **2016**, *266*, 1–13. [CrossRef]
- 45. Blackmore, S.; Godwin, R.J.; Fountas, S. The Analysis of Spatial and Temporal Trends in Yield Map Data Over Six Years. *Biosyst. Eng.* **2003**, *84*, 455–466. [CrossRef]
- 46. Blackmore, S. The interpretation of trends from multiple yield maps. *Comput. Electron. Agric.* **2000**, *26*, 37–51. [CrossRef]
- 47. Guo, W. Spatial and temporal trends of irrigated cotton yield in the Southern High Plains. *Agronomy* **2018**, *8*, 298. [CrossRef]
- 48. Shatar, T.M.; McBratney, A.B. Empirical modeling of relationships between sorghum yield and soil properties. *Precis. Agric.* **1999**, *1*, 249–276. [CrossRef]
- Schepers, A.R.; Shanahan, J.F.; Liebig, M.A.; Schepers, J.S.; Johnson, S.H.; Luchiari, A., Jr. Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. *Agron. J.* 2004, *96*, 195–203. [CrossRef]
- 50. Adeoye, G.O.; Agboola, A.A. Critical levels for soil pH, available P, K, Zn and Mn and maize ear-leaf content of P, Cu and Mn in sedimentary soils of South-Western Nigeria. *Fertil. Res.* **1985**, *6*, 65–71. [CrossRef]
- 51. Northcote, K.H.; Skene, J.K.M. *Australian Soils with Saline and Sodic Properties*; CSIRO Soil Publication No. 27; CSIRO Division of Soils: Melbourne, Australia, 1972.
- 52. Filippi, P.; Cattle, S.R.; Bishop, T.F.A.; Pringle, M.J.; Jones, E.J. Monitoring changes in soil salinity and sodicity to depth, at a decadal scale, in a semiarid irrigated region of Australia. *Soil Res.* **2018**, *56*, 696–711. [CrossRef]
- 53. Soil Conservation Service. Soil Conservation Service. Soil Survey Division Staff Soil survey manual. In *U.S. Department of Agriculture Handbook 18*; US Government Printing Office: Washington, DC, USA, 1993.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).