Description and spatial inference of soil drainage using matrix soil colours in the Lower Hunter Valley, New South Wales, Australia.

Brendan P. Malone*

Sydney Institute of Agriculture, C81 Biomedical Building, The University of Sydney, New South Wales 2006, Australia.

brendan.malone@sydney.edu.au

Alex B. McBratney

Sydney Institute of Agriculture, C81 Biomedical Building, The University of Sydney, New South Wales 2006, Australia.

alex.mcbratney@sydney.edu.au

Budiman Minasny

Sydney Institute of Agriculture, C81 Biomedical Building, The University of Sydney, New South Wales 2006, Australia.

budiman.minasny@sydney.edu.au

***Corresponding author**

Abstract

Soil colour is often used as a general purpose indicator of internal soil drainage. In this study we developed a necessarily simple model of soil drainage which combines the tacit knowledge of the soil surveyor with observed matrix soil colour descriptions. From built up knowledge of the soils in our Lower Hunter Valley, New South Wales study area, the sequence of well-draining → imperfectly draining → poorly draining soils generally follows the colour sequence of red → brown → yellow → grey → black soil matrix colours. For each soil profile, soil drainage is estimated somewhere on a continuous index of between 5 (very well drained) and 1 (very poorly drained) based on the proximity or similarity to reference soil colours of the soil drainage colour sequence. The estimation of drainage index at each profile incorporates the whole-profile descriptions of soil colour where necessary, and is weighted such that observation of soil colour at depth and/or dominantly observed horizons are given more preference than observations near the soil surface. The soil drainage index, by definition disregards surficial soil horizons and consolidated and semi-consolidated parent materials.

With the view to understanding the spatial distribution of soil drainage we digitally mapped the index across our study area. Spatial inference of the drainage index was made using Cubist regression tree model combined with residual kriging. Environmental covariates for deterministic inference were principally terrain variables derived from a digital elevation model. Pearson's correlation coefficients indicated the variables most strongly correlated with soil drainage were topographic wetness index (-0.34), mid-slope position (-0.29), multi-resolution valley bottom flatness index (-0.29) and vertical distance to channel network (0.26). From the regression tree modelling, two linear models of soil drainage were derived. The partitioning of models was based upon threshold criteria of vertical distance to channel network. Validation of the regression kriging model using a withheld dataset resulted in a root mean square error of 0.90 soil drainage index units. Concordance between observations and predictions was 0.49. Given the scale of mapping, and inherent subjectivity of soil colour description, these results are acceptable. Furthermore, the spatial distribution of soil drainage predicted in our study area is attuned with our mental model developed over successive field surveys. Our approach, while exclusively calibrated for the conditions observed in our study area, can be generalised once the unique soil colour and soil drainage relationship is expertly defined for an area or region in question. With such rules established, the quantitative components of the method would remain unchanged.

Introduction

Soil colour is arguably one of the most obvious and easily observed soil morphological characteristics. Soil scientists use soil colour to differentiate genetic soil horizons as well as for the classification of soil types e.g. Isbell (1996). From a trained or untrained eye, some inference on soils may be made from observation of soil colour in relation to organic carbon content (Schulze et al. 1993, Aitkenhead et al. 2013, Pretorius et al. 2017), mineral composition (Schwertmann and Taylor 1977), soil water content and moisture regime (Bouma 1983, Blavet et al. 2000) ~~may be made from observation of soil colour~~. Our interest in this study is making inference of a soils' capacity to drain or soil drainage, based on observed characteristics of soil colour.

For agricultural and environmental applications, soil drainage is an important property that affects plant growth, water flow and solute transport in soils (Kravchenko et al. 2002). It has long been established that soil colour patterns can be related to a soils' capacity to drain water (Evans and Franzmeier 1988; Pickering and Veneman 1984; Vepraskas and Wilding 1983). Naturally there are exceptions to this, but often~~,~~ soil colour can be interpreted as a reflection of oxidative and reductive soil processes. Reductive processes are caused by periodic or continuous water saturation. This could be due to position in the landscape (Chaplot et al. 2000) and/or the presence of a permanently or fluctuating water table near the soil surface. Described in Bouma (1983), reductive soil conditions occur when a soil is saturated. Microbial activity depletes the soil of any free oxygen (O_2) causing the soil to become anaerobic. Under anaerobic conditions, and in the presence of organic carbon, ferric iron (Fe^{3+}) is microbially converted to ferrous (Fe^{2+}). This process is referred to as iron reduction and causes the Fe pigmented coatings on soil particles (Fe^{3+} oxides) to dissolve off the particles and into the soil solution. This results in a washed out and ultimately, grey matrix soil colour, indicating the natural colour of the soil mineral grains. In addition, other redoximorphic features such as mottling and precipitation of manganese are symptomatic of soils which experience periodic or prolonged periods of soil saturation.

Explanations for the causes of soil to remain saturated for prolonged periods include the proximity of a watertable or a watercourse line. Related to these physical features is the topographical position of a particular site or landscape. For example soil saturation occurs in the landscape (from Chaplot et al. 2004), when the accumulated water flux, the product of the catchment area A_s and the area drainage flux q, passing across an element of contour length b, exceeds the product of local soil transmissivity T and the local surface gradient S

(O'Loughlin 1986). Thus terrain attributes such as slope gradient, elevation from, and distance to watercourse lines, and terrain wetness index are generally useful for understanding, but more importantly describing the spatial variation of saturated soils in a particular landscape. The other important variable, which determines soil drainage, is related to its permeability (transmissivity of water). Soil texture and pore size distribution are principal factors which determine the ability of soils to transmit water (Bouma 1983).

Soil drainage classes have been used widely in soil survey to characterise the wetness (drainage capacity) of soil and describe the fluctuations and proximity of the water table at site locations (Kidd et al. 2014). The distinctions between different drainage classes are based on tacit knowledge of the soil surveyor, or better, through physical measurements. These measurements may include observations of soil water tables via a well or core, and/or measurement of the soil water status. Measuring the soil water status for estimations of drainage class requires prolonged monitoring, and though possible, the procedure is complicated, costly, and time consuming (Bouma 1983). With these logistical issues, it is not surprising that soil colour and assessment of soil redoximorphic features are often used as an indicator for making assessments of soil drainage.

The implementation of quantitative indexes of soil drainage, inferred from soil colour and/or redoximorphic features is not a new concept. Some include that of Evans and Franzmeier (1988) which requires numeric indexing of Munsell notation, in addition to information regarding mottle characteristics and abundance. Blavet et al. (2002) also numericalised Munsell colour notations in addition to using of a soil redness index (Torrent et al. 1983) to derive a continuous index for describing the duration of water-logging. Chaplot et al. (2000) developed a continuous index (0-100) of soil hydromorphy based on the cumulative thickness of soil horizons with redoximorphic features, combined with information regarding the Munsell Hue and Value numbers. These studies exemplify the value of using low-cost soil morphological information for making inference of soil drainage characteristics.

Our interest in this study is to develop a different type of continuous index of soil drainage. It is necessarily simple, because the soil database we are using is limited in terms of direct measurements of soil drainage and is inconsistent, even unreliable in terms of descriptions of the abundance or even presence of redoximorphic features such as mottles. In the simplest terms, the drainage index we develop in this study combines some tacit knowledge with actual observations made in the field of the soil matrix colour (each genetic soil horizon), to derive a continuous whole-profile index of soil drainage.

129 The motivation for deriving a soil drainage index is that we are particularly interested in
130 understanding its spatial distribution across the landscape, as this is probably more useful
131 from a land management and assessment perspective. Studies such as Schaetzl et al. (2009)
132 demonstrate this. Furthermore, after a number of years surveying the area described in this
133 study, we have developed a mental concept of how soil drainage varies across the landscape.
134 It is a useful exercise to validate such mental models with empirical information. Given the
135 relationship between topography and soil saturation, there is considerable benefit in applying
136 digital soil mapping methods (McBratney et al. 2003) for inferring the spatial distribution of
137 soil drainage. A number of studies have constructed soil spatial inference models of soil
138 drainage class using topographical variables (e.g. Kravchenko et al. 2002; Campling et al.
139 2002). Bell et al. (1992) used multivariate discriminant analysis using topography and
140 geological information to spatially predict drainage classes. Chaplot et al. (2004) were
141 interested in mapping the soil hydromorphic index using topographic indices derived from
142 the land surface and saprolite upper boundary. Less invasive techniques of mapping soil
143 drainage classes through the use of remote sensing platforms have also been demonstrated
144 (Peng et al. 2003; Cialella et al. 1997)
145 The aims of this study are threefold: 1) To develop an index of soil drainage combining
146 tacit knowledge and empirical information of soil matrix colour. 2) To determine whether an
147 empirical relationship exists between the estimated drainage index and landscape features. 3)
148 To develop a soil spatial prediction function for estimating the spatial distribution of soil
149 drainage across our study area in the Lower Hunter Valley, NSW (New South Wales,
150 Australia).

151 **Materials and methods**

152 *Study area*
153 The area of this study is the Hunter Wine Country Private Irrigation District (HWCPID),
154 situated in the Lower Hunter Valley, NSW. The HWCPID covers approximately 220 km^2 and
155 encompasses the localities of Pokolbin and Rothbury, NSW (32.83°S 151.35°E), which are
156 approximately 140 km north of Sydney, NSW (Fig. 1). Topographically, this area consists
157 mostly of undulating hills that ascend to low mountains to the south-west. The underlying
158 geology of the HWCPID is predominantly Early Permian, with some Middle and Late
159 Permian formations. Described in the Newcastle Coalfield Regional 1:100000 Geology Map
160 (Hawley et al. 1995), the most extensive formation is the Rutherford Formation (Early
161 Permian) which consists of siltstones, marl, and some minor sandstone. Much of the southern

and eastern part of the HWCPID is underlain by the Rutherford Formation. Other extensive formations include: the Mulbring Siltstone (Late Permian siltstones), the Branxton Formation (Middle Permian conglomerates, sandstones and siltstones) and the Farley Formation (Early Permian silty sandstones). These formations occupy the north-western extents of the HWCPID. In terms of landuse, dryland agricultural grazing systems are predominant, followed by an expansive viticultural industry. While most of the land has been dedicated for these uses, tracts of remnant natural vegetation (dry forest) are apparent, particularly towards the south-western area—which is bordered by Broken Back Range, Werakata National Park situated to the east, and some areas situated in the northern extents.

[Insert figure 1 here]

Figure 1: The Hunter Wine Country Private Irrigation District (shaded green) and surrounding localities. Sampling locations of the three survey campaigns: Malone et al. (2011) 34 soil profiles, Odgers et al. (2011) 251 soil profiles and 1261 soil profiles from annual surveys. Black lines indicate roads. Blue lines are major watercourses.

Our knowledge of the soils across the HWCPID was first informed from legacy soil survey which is described in detail within the Soil Landscapes of the Singleton 1:250,000 Sheet Map and Report (Kovac and Lawrie 1990). This knowledge has since evolved through annual soil surveying campaigns by students and members of our research group, which began in 2001 and continue to the present time. These annual surveys, while concentrated to the south of the study area, form a densely populated database of soil information and descriptions. This information and soil knowledge has been supplemented with two area-wide soil surveys of the HWCPID which have been described in Malone et al. (2011) and Odgers et al. (2011). Based on these various soil surveying campaigns we have found—based on the sub-order level of the Australian Soil Classification system (Isbell 1996)—that the most dominant soils across the HWCPID are both Brown and Red Dermosols and Chromosols. Generally, Dermosols and Chromosols are the most prolific; there are few Kurosols by comparison. Hydrosols and Rudosols are few, but generally concentrated near watercourse lines. Calcarosols are also few, yet exist in areas where the Rutherford Formation exists, particularly where the occurrence of the calcareous marl parent material is present. Corresponding WRB (FAO 1998) soil classes to the ASC soil orders are: ~~Calicisols~~ Calcisols (Calcarosols), Luvisols (Chromosols and some Dermosols), Acrisols (Kurosols and some Dermosols), Fluvisols (Hydrosols), and Regosols (Rudosols).

Conceptual model of the spatial distribution of soil drainage

In the HWCPID, we have often observed a common sequence of soils down hillslopes, which indicate varying degrees of soil drainage. Morphologically, this sequence can be observed as changes in the matrix soil colour. For example, red coloured soils are observed a lot on hilltops and crests. Brown and yellow soils can be found further down the hillslope, and often grey and black soils are found on the foot slopes near watercourse lines. This sequence of soil colour ~~are~~ is not uncommon down a hillslope in other parts of the world (e.g. Simonson and Boersma 1972; Bouma 1983; Kravchenko 2002).

In Table 1, site and soil morphological data ~~is~~ are provided for three soils, developed from the same parent material (siltstone), at different positions of a hillslope in the HWCPID. These data are not in isolation; rather they represent a common occurrence, not of soil type, but soil colour change and presumably soil drainage. On the crest is a Red Dermosol, which then grades into a Brown Dermosol at the mid-slope position, followed by a Grey Dermosol on the flat, near a watercourse line. Terrain variables: Topographic Wetness Index (TWI), Multi-resolution Valley Bottom Flatness Index (MRVBF), and Vertical Distance to Channel Network (VDCN) highlight some topographical information which may provide further explanatory evidence for describing this sequence of soils and associated soil drainage. For example TWI and MRVBF, both indices for describing the movement and concentration of water in the landscape, increase down the hillslope i.e. soils in the mid-to-low parts of the hillslope accumulate and concentrate more water than soils on or near the hillcrests.

[Table 1 here]

Table 1. Site and soil morphological information for three soil profiles down a catena in the HWCPID. Topographic Wetness Index (TWI), Multi-resolution Valley Bottom Flatness Index (MRVBF), and Vertical Distance to Channel Network (VDCN).

Similar to a soil drainage index, the soil hydromorphic index by Chaplot et al. (2000) requires information regarding redoximorphic features i.e. mottling for its derivation. Because we cannot rely on this data in our database (as such features have not been consistently nor accurately recorded), we need to derive another index, based exclusively on the soil matrix colour. Further in the discussion we propose an approach how to incorporate such features within our simple index. Our drainage index ranges continuously between and including the values of 5 and 1. The conceptual model of soil water drainage in the HWCPID and exemplified with the data in Table 1, is that "red" soils have the highest drainage index value of 5, "brown" soils (4), "yellow" soils (3), "grey" soils (2), and "black" soils (1). This index implies that "red" soils drain better than "brown", which drain better than "yellow" soils and so on. "Black" soils are the poorest in terms of soil drainage because it is these soils

that appear to be saturated permanently and as a consequence have accumulated carbon. The soil drainage index has been designed for where descriptions have been made for each genetic soil horizon of a soil profile (but it may also be applied where soil is observed as regular or at specific depth intervals). Derivation of the soil drainage index is now described in the following methodological sections.

Derivation of the drainage index

The data

In this study we use soil data collected from three major soil surveying campaigns conducted in the HWCPID. In total, these campaigns have amounted to 1546 individual soil profile observations and descriptions (Figure 1). The breakup of these profiles is: 34 come from the work by Malone et al. (2011); 251 from the work of Odgers et al. (2011); and 1261 from annual soil survey work for the years between 2001 and 2011. For each of these soil profiles, data was recorded for each genetic horizon. Our primary interest is in the matrix soil colour of each horizon, particularly the moist colour, which was recorded on the basis of matching the observed soil colour with a colour chip on a Munsell HVC (Hue, Value, Chroma) colour chart. We disregarded horizon descriptions where the lower boundary did not exceed 40cm from the top of the soil profile. We also disregarded horizon descriptions of semi- and unconsolidated parent materials which in Australian soil nomenclature are described as B/C and C horizons respectively (The National Committee on Soil and Terrain 2009). For example, if the first horizon of a particular soil profile was 0-55cm, then it would be included in the drainage index model. If a soil profile had a sequence of horizons measuring: 0-30cm, 30-75cm, 75-120cm, and >120 (which was found to be bedrock), then the drainage index would only consider the observed data from 30-120cm. After this filtering process, we ended up with 3731 soil horizon data with moist soil colour descriptions to work with.

Munsell HVC soil colour descriptions are not conducive for quantitative studies. Therefore, we performed a conversion from the Munsell HVC colour space to the CIELAB colour space (Robertson 1977; CIE 1978). The CIELAB colour space can describe any uniform colour space by the three variables: L^*, a^*, and b^*. Each variable represents the lightness of the colour ($L^* = 0$ yields black and $L^* = 100$ indicates diffuse white), its position between red/magenta and green (a^*, negative values indicate green while positive values indicate magenta) and its position between yellow and blue (b^*, negative values indicate blue and positive values indicate yellow). The non-linear equations for converting from Munsell HVC to CIELAB are described in Viscarra Rossel et al. (2006). First Munsell HVC are

263 converted to the CIE XYZ colour space based on a fitted neural network model of known
264 XYZ values and corresponding Munsell soil colour chips, which are derived from the
265 Munsell Conversion program Version 6.41 (http:// www.gretagmacbeth.com). Standard CIE
266 (1978) equations are then used to transform from CIE XYZ to CIELAB. Because a model
267 based approach (neural networks) — rather than a physical relationship or direct
268 correspondence — are used to transform from Munsell HVC to CIE XYZ, the prediction will
269 inevitably be uncertain to some degree. The extent of this uncertainty is not known. Viscarra
270 Rossel et al. (2006) do state however that the conversion was adequate.
271 One of the problems with descriptions of soil colour is that they are subjective and can be
272 ambiguous estimates. Each individual's perception of colour is different, which will result,
273 for the same soil, often quite different predictions of soil colour. In order to work with this
274 type of data, our drainage index is rooted in fuzzy set theory (Zadeh 1965), meaning that
275 some of the ambiguity and uncertainty in soil colour prediction can be dealt with by
276 allocating each observation, membership to multiple defined classes. Therefore the first step
277 in defining a drainage index entails designating centroids or archetypal soils for each soil
278 colour/drainage class. Using the unconverted data (i.e. the Munsell HVC colours), we
279 designated each observation to a particular colour class based on the colour groupings of
280 Northcote (1979). From summary statistics, we came up with 3 centroids (the 3 most
281 frequently observed) for each colour class to represent the reference or archetypal soil colours
282 (Table 2). Three reference colours (15 in total) for each colour class was a pragmatic decision
283 based on the fact that we wanted to derive an appropriate configuration of centroids within
284 the L^*, a^*, and b^* feature space.

285 **[Insert table 2 here]**

286 Table 2. Reference soil colours for each soil colour class in both Munsell HVC and CIELAB
287 colour space notation. Note that the reference colours are grouped following the colour
288 groupings that were created by Northcote (1979).

289

290 With the reference colours established, we then estimated the Mahalanobis distance of
291 each observation to each reference colour:

$$d_i = \sqrt{(\boldsymbol{x}_i - \boldsymbol{c}_j)^T \boldsymbol{S}^{-1} (\boldsymbol{x}_i - \boldsymbol{c}_j)}$$

292 $i = 1, \ldots, N;\ j = 1, \ldots, C$

Equation 1

where d is the Mahalanobis distance between the multivariate vector $\boldsymbol{x}$ (here an observed L^*, a^*, and b^* vector) and reference colour vector $\boldsymbol{c}$ (L^*, a^*, and b^*). $\boldsymbol{S}$ is the variance-covariance matrix of N observed $\boldsymbol{x}$. The result here is an $N \times C$ matrix $(\boldsymbol{D})$ where each element d_{ij} represents the Mahalanobis distance of each observed horizon colour i to each reference colour j.

The measure of similarity (or membership) of each horizon observation to each reference colour is estimated as:

$$u_{i,j} = \frac{1}{1 + \sum_{l=1, l \neq j}^{C} \left(\frac{d_{ij}}{d_{il}}\right)^{\frac{1}{m-1}}}$$

$i = 1, \ldots, N;\ j = 1, \ldots, C;\ l = 1, \ldots, C\text{-}1$

Equation 2

where $u_{i,j}$ is the similarity of horizon observation i to reference colour j, and where d_{il} is the Mahalanobis distance of i to the other reference colours d_{il}. Thus observations close to (as determined by the Mahalanobis distance) a reference colour will have a higher similarity than those observations more distant. The fuzzy exponent m determines the level of similarity fuzziness of i to each reference colour. A value of 1 for m will result in all $u_{i,j}$ converging to either 0 or 1, which implies a crisp partitioning of the observations to the reference colours. Conversely, m values approaching infinity will create similarities with complete overlap such that an observation will have equal similarity to all reference colours. In this study, we pragmatically set m to 1.5 on the basis that we did not desire to crisply partition the observations, yet still allow for some overlap to the reference colours.

Estimation of drainage index for each horizon and subsequently each soil profile

The drainage index ranges continuously between and including the values of 5 and 1. As per the conceptual model of soil water drainage in the HWCPID, *red* soils have the highest drainage index value of 5, *brown* soils (4), *yellow* soils (3), *grey* soils (2), and *black* soils (1). For each horizon the drainage index is calculated as a weighted average based on the degree of similarity to each reference colour. Such that:

$$DI_i = \sum_{i=1}^{N} u_{ij} RC_j$$

Equation 3

where *DI* is the drainage index and *RC* refers to the reference soil colour *j*. While not considered in this study, the presence of mottles could potentially be included in this index with the following equation:

$$DI_{i\,(rx)} = DI_i \times Rp$$

Equation 4

Here $DI_{i\,(rx)}$ is the drainage index incorporating information about the proportion of mottles within the soil matrix (expressed as a percentage), and Rp is simply 1- the observed proportion. Of course this equation would need to be tested against real data.

Continuing on from equation 3, because we need to derive a whole profile drainage index value we need to aggregate each *DI* calculated at each horizon for each soil profile *P*. However, we want to preferentially weight the observed *DI* values such that observations at depth are given more weight to those higher up the soil profile. Firstly, for each horizon in soil profile *P*, *a* vector based on the observed upper and lower horizon boundaries is created and then summed. For example, in *P*, a particular horizon is observed to occur from 45-75cm. The summed vector of this sequence (i.e. 45, …, 75) is 1860. We may denote this summed vector as *SV*. Therefore the whole-soil profile drainage index value can be calculated as:

$$DI_P = \sum \frac{SV_h}{\sum_{h=1}^{Z} SV_h} \cdot DI_h$$

$h = 1, \ldots, Z$

where DI_P and DI_h are the drainage index value/s for the whole-profile and genetic soil horizons respectively of a soil profile *P*.

Correlation of the drainage index with environmental variable

The ultimate aim of this paper is to derive a drainage index map for the HWCPID. As a preliminary step we wanted to investigate the relationship (using Pearson's coefficient of correlation) of the derived drainage index with a suite of environmental covariate information. In this study, this covariate information is exclusively derived from a digital elevation model (25m ground resolution) sourced from the NSW Government. Informed from similar work of mapping soil drainage and hydromorphy (such as Kravchenko et al. 2002; Campling et al. 2002; and Chaplot et al. 2004), from the digital elevation model we derived a

number of potentially useful primary and secondary terrain variables: Elevation (E), slope gradient (S), slope length (SL), slope height (SH), mid-slope position (MSP), terrain wetness index (TWI), vertical distance to channel network (VDCN), multi-resolution valley bottom flatness index (MRVBF), analytical hillshading (AH). These indices were derived using the terrain analysis modules of SAGA GIS (http://www.saga-gis.org/), and described in more detail in Table 3.

[Insert table 3 here]

Table 3. Description of topographical variables used in this study.

Mapping the drainage index

We use a digital soil mapping (McBratney et al. 2003) framework for the spatial interpolation of the drainage index across the HWCPID to the same resolution as the topographic variables (25m). The dataset of 1546 profiles was randomly split into calibration (70%) and validation (30%) datasets. For calibration, the soil spatial prediction function employed here was a regression kriging model. Using the covariates described above, we used cubist models to identify any deterministic relationship with the drainage index at each of the observed soil profiles. Cubist is a prediction-oriented regression model that is based mostly on work by Quinlan (1992). Although it initially creates a tree structure, it collapses each path through the tree into a rule. A regression model is fitted for each rule, based on the data subset defined by the rules. The set of rules are pruned or possibly combined, and the candidate variables for the linear regression models are the predictors that were used in the parts of the rule that were pruned away. The residuals from the cubist model were investigated for spatial autocorrelation as a means to detect any additional (random) spatial trend of the drainage index not detected from the covariates. We used geostatistics and locally fitted variograms (based on the exponential model) for spatial interpolation (kriging) of the residuals across the entire HWCPID. The sum of the outputs from the deterministic modelling and residual kriging resulted in a final drainage index map.

The validation dataset was withheld from the calibration procedure. Using measures such as the root mean square error (RMSE) and concordance coefficient we compared the regression kriging predictions at each of the validation profiles with their 'observed' value.

The RMSE measures the differences between predicted and observed values and is estimated by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(z_{pi}(s) - z_i(s)\right)^2}{n}}$$

where $z_{pi}(s)$ and $z_i(s)$ are the predicted and observed values of validation point i and n is the number of validation points. The concordance coefficient measures the fidelity of the observations and the predictions to a 1:1 line (Lin 1989).

The implementation of methods in this study (where previously not already stated) were carried out using the R statistical software (R Core team 2015) for general statistical analyses and mapping. The R package "Cubist" (Kuhn et al. 2016) was used for fitting the cubist model. VESPER geostatistical software (Minasny et al. 2005) was used for the local fitting of variograms and kriging.

Results

Pearson's coefficient of correlation between the derived soil profile drainage index and each of the covariate data sources from highest to lowest were: TWI (-0.34), MSP (-0.29), MRVBF (-0.29), VDCN (0.26), SH (0.22), SL (-0.18), S (0.11), E (0.09), and AH (-0.03). These correlation coefficients indicate some general features of soil drainage in the HWCPID, for example there is a positive correlation of the drainage index with vertical proximity to watercourse lines. Indices such as TWI and MRVBF, which inform us about the hydrological characteristics of the area, are negatively correlated with the drainage index. Thus based on landscape position, where the soil is more prevalent to concentration of water, the drainage index is also lower. The correlations of the slope indices S, SH, and MSP with the drainage index indicate a relationship whereby gentle slopes (relatively low S and SH and high MSP) soils more likely to have a lower drainage index. Similarly, longer slopes (SL) result in a negative correlation with the drainage index.

Fitting of the Cubist model to the calibration data resulted in the partitioning of two simple rules for the spatial distribution of the drainage index. Each rule defining a different regression model:

Rule 1

$if\ VDCN \leq 7.8$m

$then$

$$DI = 6.58 + 0.06(VDCN) - 0.20(TWI) - 0.02(E) - 0.80(AH)$$

Rule 2

$if\ VDCN\ > 7.8\text{m}$

$then$

$$DI = 6.58 - 0.11(TWI) - 0.01(E) - 0.09(MSP) + 0.001(VDCN)$$

These simple linear regressions are pre-empted by a recursive split of all data based on a threshold value of 7.8 for the VDCN. Essentially this means that vertical proximity to a watercourse line is a defining characteristic of soil drainage. Common parameters to each linear model were VDCN, TWI, and E, while AH was only included in the first rule and MSP was only included in the second rule.

Examining where each Cubist rule was applied shows clearly the relationship of the rules with proximity to watercourse lines (Figure 2a). For approximately one-third of the area, rule 1 was applied.

[Figure 2 here]

Figure 2. Spatial map of the application of Cubist rules across the HWCPID (a). Map of the soil drainage index across the HWCPID (b). Black lines indicate roads. Blue lines indicate watercourse lines

The associated drainage index map which resulted from the regression kriging model is shown in Figure 2b. With the blue lines indicating the watercourse lines, it is clear from the map that proximity to them has a considerable effect on the soil drainage. From a basic statistical analysis, where rule 1 was applied, the mean drainage index was 2.70 with 95% of the area between 1.85 and 3.60. Where rule 2 was applied, the mean drainage index was 3.40 with 95% of the area between 2.51 and 4.00.

Validation of the regression kriging model based on 446 withheld data indicated a RMSE of 0.90, meaning that, the predictions of the drainage index on average deviate approximately 0.9 away from the observed value. The concordance between the observed and predicted values was a reasonable 0.49. The plot in Figure 3 show the observations and corresponding predictions with respect to the 1:1 relationship (draw as a red dashed line). Predictions appear to be strongest around drainage index values between 2.5 and 4.0 From visual inspection of the plot it is clear there does not seem to be any bias, such as prevalence for over or under predictions (mean error was calculated as -0.08). A linear model fitted to the observed and

fitted data resulted in a co-efficient of determination of 31%, indicating a reasonable correlative relationship (green solid line).

[Insert Figure 3]

Figure 3. Observed and fitted plot of drainage index based on regression kriging predictions for the validation dataset. Red dashed line indicates a line of concordance (1:1 relationship). Green solid line is a regression line for the linear relationship between the observed and predicted drainage index.

Discussion

The continuous index of soil drainage proposed in this study requires little information other than tacit knowledge of soil drainage spatial variability, and observed soil matrix colour descriptions. The fundamental limitation of this is that we have assumed there is a direct correlation between soil colour and soil drainage. There is good physical evidence though to support this relationship (e.g. Bouma 1983; Evans and Franzmeier 1988). It is likely, since we have been able to establish a correlative relationship between our index of soil drainage and some topographical variables, that some validation in terms of soil water measurements or measurements of water table proximity is warranted in further studies.

There is significant value from the land management perspective in quantifying soil drainage as a spatially continuous variable across the landscape. In this study, we have avoided mapping the well-known and established drainage classes. While there may be value in doing this, the grading between classes is qualitative and in the absence of direct measurement, allocation to a particular class is a subjective designation. Nevertheless, it was by necessity (due to the data used) that we had to develop our own model, which by default treats soil drainage as a continuous variable. Subsequently, mapping the continuously varying drainage index across the HWCPID revealed spatial patterns more attuned to what one would observe in the landscape; that is, continuously varying rather than discreetly apportioned. In the HWCPID, it is believed that discreet variations of soil drainage are the exception rather than the rule.

On the basis of using digital soil mapping methods, we have been able to validate quantitatively the spatial model of soil drainage. Comparatively with other digital soil mapping studies, the results found in this study are acceptable (Grunwald 2009). Other studies that have examined the relationship between soil drainage and environmental information have reported stronger correlations than that reported in this study (e.g. Campling et al. 2002; Chaplot et al. 2004). It is suspected that scale may be one cause for this discrepancy. For example we have attempted to describe the variations of soil drainage

468 over a much larger area of land which we know to be rather complex (topographically *and*
469 lithologically). While these results may be improved upon, the map, from a soil surveyor's
470 perspective, adequately coincides with the knowledge we have developed over the years of
471 survey in the HWCPID.
472 In terms of the spatial prediction model, we used the Cubist models as an attempt to
473 mirror what a soil surveyor would observe in the landscape. That is, given particular
474 combinations of features or characteristics of the landscape, a particular soil or characteristic
475 of the soil will behave similarly. The quantitative interpretation of this and what was found
476 in this study, was that vertical distance to a channel network was a divisive and important
477 physical attribute determining the estimation of soil drainage. Such that, given a certain
478 threshold, different predictive models were applied. More generally, we have found that
479 using such rule-based spatial prediction functions makes them more interpretable (from the
480 soil survey perspective) and particularly useful for digital soil mapping.
481 Correcting the deviation between what was observed (from the data) and what was
482 predicted using a spatial model is a worthwhile pursuit. What was clear in this study is that
483 the observed variations of soil colour described something much more complex than what
484 the spatial model was able to describe. There could be many reasons and explanations for
485 this. One of them is that soil colour alone and attribution thereof can have significant
486 influence on the interpretation of soil processes. While fuzzy set theory is embedded within
487 our model, which by definition embraces the subjectivity around soil colour attribution, our
488 model is by no means immune to poorly attributed soil colour descriptions. Ultimately, this
489 can have flow-on effects when soil colour is then used in some sort of quantitative model
490 e.g. soil drainage (Chaplot et al. 2004). O'Donnell et al. (2010) have proposed a
491 standardised procedure of soil colour attribution based on image processing. Or perhaps
492 usage of a colorimeter would be enough to make standardised assessments of soil colour.
493 Currently, standardised assessments of soil colour are not made during soil survey around
494 the world, so it is likely it will be some time before we can test the applicability of our
495 method using such assessments.
496 The spatial model of soil drainage in this study principally used topographical variables
497 as predictive covariate information. We also incorporated a regression kriging model with
498 the intention of further modelling spatial trend that was not detectable from the topographic
499 information. Regression kriging made some improvement of the prediction in comparison to
500 just using a deterministic model. Nevertheless, due to a limitation in the availability of
501 additional sources of predictive information, we were unable to explore more complex

relationships of soil drainage with other environmental variables. For example, parent material or underlying geology has been shown to be a useful variable (Bell et al. 1992). Intuitively, different lithologies will impart differing soil physical characteristics, such that the drainage characteristics of a soil developed from limestone will be different from those developed on siltstone or sandstone etc. The best available geological survey of the HWCPID (1:100000; Hawley et al. 1995) informs us that while siltstones are the most predominant lithology, there are also sandstones and silty sandstone parent materials. A limitation of our mental model of soil drainage is that it has been refined where the lithology is predominantly siltstone—as most of the soil sampling has been conducted on this lithology. There is potential bias regarding estimation of soil drainage to contend with where other lithologies are found. However, the question is whether the current geological survey could be used to refine our model of soil drainage? It is unlikely that it would, because while instructive, it is neither comprehensive or of the appropriate scale. Furthermore, soil processes such as colluviation and alluviation have often created soil profiles of complex and mixed lithology that is near impossible to disentangle from geological survey maps. We envisage that in the future, gamma-radiometric survey will provide us with information regarding the lithology and lithological processes at the necessary detail to be included within our soil drainage model. Gamma radiometry refers to the measurement of naturally occurring gamma radiation which is emitted from the ground surface (Cook et al. 1996). Such information has been shown to describe the distribution of soil-forming materials and weathering processes over large areas (Wilford 2012)

In the absence of detailed lithological information, a pragmatic solution may be to examine whether the digital mapping of soil texture grades (or soil variables derived from them such as bulk density etc.) are useful for interpreting variations of soil drainage. In the HWCPID, where soil textures are recorded predominantly as that derived from hand bolusing, we need to explore methods of how to incorporate these data within a digital soil mapping framework.

Conclusions

By necessity of the data available, we have developed an index of soil drainage which incorporates tacit knowledge of the soil surveyor and observed soil matrix colour. Soil drainage is evaluated as a whole-profile, weighted combination of the soil colour at each generic soil horizon. Fuzzy set theory is built into the drainage index model as a means to

dampen the subjectivity of soil colour attribution. We believe the approach can be generalised to other areas once the unique soil colour and soil drainage relationships have been defined by an expert.

In our study, we found that the topographical variables most strongly correlated with soil drainage are topographic wetness index, mid-slope position, multi-resolution valley bottom flatness index and vertical distance above channel network. Cubist models were used to model the relationship of the drainage index with a suite of topographic variables with the dual purpose of understanding the spatial variation of soil drainage and to validate our mental model of soil drainage developed over the years from successive field surveys. Validation of the spatial model of soil drainage was adequate in consideration of the scale of mapping and nature of the data. The associated map corresponds meaningfully to what we have generally observed in the field. The incorporation of new information specifically from gamma-radiometry or soil texture may be useful solutions in improving our understanding of soil drainage in the HWCPID.

References

Aitkenhead, M.J., Coull, M., Towers, W., Hudson, G., Black, H.I.J., 2013. Prediction of soil characteristics and colour using data from the National Soils Inventory of Scotland. Geoderma 200-201, 99-107.

Bell, J.C., Cunningham, R.L., Havens, M.W., 1992. Calibration and validation of a soil landscape model for predicting soil drainage class. Soil Science Society of America Journal 56, 1860-1866.

Blavet, D., Leprun, J.C., Mathe, E., Pansu, M., 2002. Soil colour variables as simple indicators of the duration of soil waterlogging in a West African catena, Proceedings of the 17th World Congress of Soil Science, Thailand, pp. 331-1311.

Blavet, D., Mathe, E., Leprun, J.C., 2000. Relations between soil colour and waterlogging duration in a representative hillside of the West African granito-gneissic bedrock. Catena 39, 187-210.

Bouma, J., 1983. Hydrology and soil genesis of soils with aquic moisture regimes. In: L.P. Wilding, N.E. Smeck, G.F. Hall (Eds.), Pedogenesis and Soil Taxonomy. 1. Concepts and Interactions. Elsevier Science Publishers, The Netherlands, pp. 253-281.

Campling, P., Gobin, A., Feyen, J., 2002. Logistic modeling to spatially predict the probability of soil drainage classes. Soil Science Society of America Journal 66, 1390-1401.

Chaplot, V., Walter, C., Curmi, P., 2000. Improving soil hydromorphy prediction according to DEM resolution and available pedological data. Geoderma 97, 405-422.

Chaplot, V., Walter, C., Curmi, P., Lagacherie, P., King, D., 2004. Using the topography of the saprolite upper boundary to improve the spatial prediction of the soil hydromorphic index. Geoderma 123, 343-354.

Cialella, A., Dubayah, R., Lawrence, W., Levine, E., 1997. Predicting soil drainage class using remotely sensed and digital elevation data. Photogrammetric Engineering and Remote Sensing 62, 171-177.

CIE, International Commission on Illumination., 1978. Recommendations on Uniform Color Spaces, Color-Difference Equations, and Psychometric Color Terms, Supplement No. 2 to Publication CIE No. 15(E-1.3.1)1971, /(TC-1.3)1978. Bureau Central de la CIE, Paris.

Cook, S.E., Corner, R.J., Groves, P.R., Grealish, G.J., 1996. Use of airborne gamma radiometric data for soil mapping. Australian Journal of Soil Research 34, 183-194.

Evans, C.V., Franzmeier, D.P., 1988. Color index values to represent wetness and aeration in some Indiana soils. Geoderma 41, 353-368.

FAO, Food and Agriculture Organisation., 1998. World Reference Base for Soil Resources. Food and Agriculture Organization of the United Nations, Rome.

Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resources Research 39, doi:10.1029/2002WR001426, 12 .

Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma 152, 195-207.

Hawley, S.P., Glen, R.A., Baker, C.J., 1995. Newcastle Coalfield Regional Geology 1:100 000, 1st Edition. Geological Survey of New South Wales, Sydney, Australia.

Isbell, R.F., 1996. The Australian Soil Classification. CSIRO Publishing, Collingwood, Australia.

Kidd, D.B., Malone, B.P., McBratney, A.B., Minasny, B., Webb, M.A., 2014. Digital mapping of a soil drainage index for irrigated enterprise suitability in Tasmania, Australia. Soil Research 52, 107-119.

598 Kovac, M., Lawrie, J.M., 1990. Soil Landscapes of the Singleton 1:250 000 Sheet. Soil
599 Conservation Service of NSW, Sydney, Australia.
600 Kravchenko, A.N., Bollero, G.A., Omonode, R.A., Bullock, D.G., 2002. Quantitative
601 mapping of soil drainage classes using topographical data and soil electrical
602 conductivity. Soil Science Society of America Journal 66, 235-243.
603 Kuhn, M., Weston, S., Keefer, C., Coulter, N., C code for Cubist by Ross Quinlan., 2016.
604 Cubist: Rule- and Instance-Based Regression Modeling. R package version 0.0.12.,
605 http://CRAN.R-project.org/package=Cubist.
606 Lin, L.I., 1989. A concordance correlation-coefficient to evaluate reproducibility. Biometrics
607 45(1), 255–268.
608 Malone, B.P., de Gruijter, J.J., McBratney, A.B., Minasny, B., Brus, D.J., 2011. Using
609 additional criteria for measuring the quality of predictions and their uncertainties in a
610 digital soil mapping framework. Soil Science Society of America Journal 75, 1032-
611 1043.
612 McBratney, A.B., Mendonca-Santos, M.L., Minasny, B., 2003. On digital soil mapping.
613 Geoderma 117, 3-52.
614 Minasny, B., McBratney, A.B., Whelan, B.M., 2005. VESPER version 1.62. Australian
615 Centre for Precision Agriculture, McMillan Building A05, The University of Sydney,
616 NSW 2006, http://sydney.edu.au/agriculture/pal/software/vesper.shtml.
617 Northcote, K.H., 1979. A Factual Key for the Recognition of Australian Soils. 4th Edition.
618 Rellim Technical Publications, Glenside, South Australia.
619 Odgers, N.P., McBratney, A.B., Minasny, B., 2011. Bottom-up digital soil mapping. I. Soil
620 layer classes. Geoderma 163, 38-44.
621 O'Donnell, T.K., Goyne, K.W., Miles, R.J., Baffaut, C., Anderson, S.H., Sudduth, K.A.,
622 2010. Identification and quantification of soil redoximorphic features by digital image
623 processing. Geoderma 157, 86-96.
624 O'Loughlin, E.M., 1986. Prediction of surface saturation zones in natural catchments by
625 topographic analysis. Water Resources Research 22, 794-804.
626 Peng, W., Wheeler, D.B., Bell, J.C., Krusemark, M.G., 2003. Delineating patterns of soil
627 drainage class on bare soils using remote sensing analyses. Geoderma 115, 261-279.
628 Pickering, E.W., Veneman, P.L.M., 1984. Moisture regimes and morphological
629 characteristics in a hydrosequence in Central Massachusetts. Soil Science Society of
630 America Journal 48, 113-118.
631 Pretorius, M.L., Van Huyssteen, C.W., Brown, L.R., 2017. Soil color indicates carbon and
632 wetlands: developing a color-proxy for soil organic carbon and wetland boundaries on
633 sandy coastal plains in South Africa. Environmental Monitoring and Assessment 189,
634 556-566.
635 Quinlan, R., 1992. Learning with continuous classes. , Proceedings of the 5th Australian Joint
636 Conference On Artificial Intelligence Hobart, Tasmania, pp. 343-348.
637 Robertson, A.R., 1977. The CIE 1976 color-difference formulae. Color Research and
638 Application 2, 7-11.
639 Schaetzl, R., Krist, F., Stanley, K., Hupy, C., 2009. The natural soil drainage index: An
640 ordinal estimate of long-term soil wetness. Physical Geography 30, 383-409.
641 Schulze, D.G., Nagel, J.L., Vanscoyoc, G.E., Henderson, T.L., Baumgardner, M.F., Stott,
642 D.E., 1993. Significance of organic matter in determining soil colours. In: J.M.
643 Bigham, E.J. Ciolkosz (Eds.), Soil Color. SSSA Special Publications, pp. 71-90.
644 Schwertmann, U., Taylor, R.M., 1977. Iron Oxides. In: J.B. Dixon, S.B. Weed (Eds.),
645 Minerals in Soil Environments. Soil Science Society of America, Madison, WI.

646 Simonson, G.H., Boersma, L., 1972. Soil morphology and water table relations. 2.
647 Correlation between annual water table fluctuations and profile features. Soil Science
648 Society of America Proceedings 36, 649-653.
649 R Core Team, 2015. R: A language and environment for statistical computing. R Foundation
650 for Statistical Computing, Vienna, Austria.
651 The National Committee on Soil and Terrain, 2009. Australian Soil and Land Survey Field
652 Handbook. Third Edition. CSIRO Publishing, Collingwood, Australia.
653 Torrent, J., Schwertmann, U., Fechter, H., Alferez, F., 1983. Quantitative relationships
654 between soil color and hematite content. Soil Science 136, 354-358.
655 Vepraskas, M.J., Wilding, L.P., 1983. Aquic moisture regimes in soils with and without low
656 chroma colors. Soil Science Society of America Journal 47, 280-285.
657 Viscarra Rossel, R.A., Minasny, B., Roudier, P., McBratney, A.B., 2006. Colour space
658 models for soil science. Geoderma 133, 320-337.
659 Wilford, J., 2012. A weathering intensity index for the Australian continent using airborne
660 gamma-ray spectrometry and digital terrain analysis. Geoderma 183, 124-142.
661 Wischmeier, W.H., Smith, D., 1978. Predicting rainfall erosion losses: a guide to
662 conservation planning. USDA-ARS Agriculture Handbook Nos. 537, Washington
663 DC.
664 Zadeh, L.A., 1965. Fuzzy sets. Information and Control 8, 338-353.
665

666

667

668