



# Deep drainage estimates using multiple linear regression with percent clay content and rainfall

D. L. Wohling<sup>1,2</sup>, F. W. Leaney<sup>3</sup>, and R. S. Crosbie<sup>3</sup>

<sup>1</sup>School of the Environment, Flinders University of South Australia, Bedford Park, P.O. Box 2100, Adelaide, SA 5001, Australia

<sup>2</sup>Department for Water, Government of South Australia, Adelaide, Level 11, 25 Grenfell Street, Adelaide, SA 5000, Australia

<sup>3</sup>Water for a Healthy Country National Research Flagship, CSIRO Land and Water, Glen Osmond Waite Road – Gate 4, Glen Osmond, SA 5064, Australia

Correspondence to: D. L. Wohling (daniel.wohling@sa.gov.au), F. W. Leaney (fred.leaney@csiro.au), R. S. Crosbie (russell.crosbie@csiro.au)

Received: 4 April 2011 – Published in Hydrol. Earth Syst. Sci. Discuss.: 5 May 2011

Revised: 23 January 2012 – Accepted: 20 February 2012 – Published: 24 February 2012

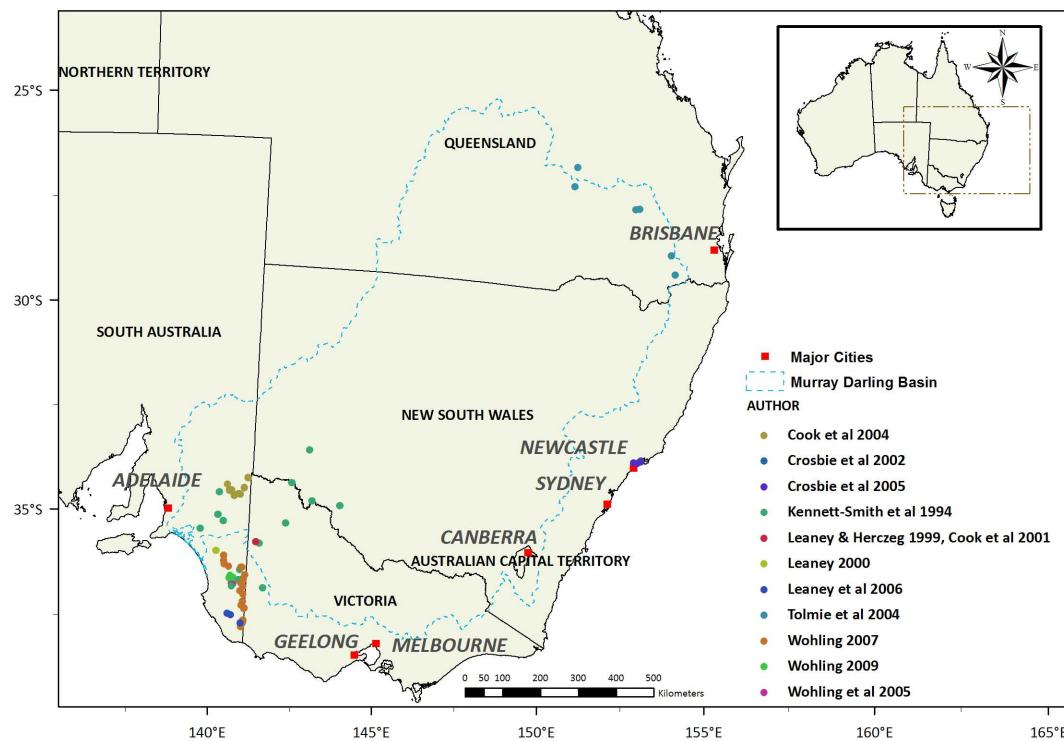
**Abstract.** Deep drainage estimates are required for effective management of water resources. However, field measurements are time consuming and costly so simple empirical relationships are often used. Relationships developed between clay content of the surface soil and deep drainage have been used extensively in Australia to provide regional estimates of deep drainage but these relationships have been poorly justified and did not include rainfall in the relationships. Here we present a rigorous appraisal of clay content of soils and rainfall as predictors of deep drainage using an extensive database of field observations from across Australia. This study found that annual average rainfall and the average clay content of the top 2 m of the soil are statistically significant predictors of point scale deep drainage. Relationships have been defined for annual, perennial and tree type vegetation as a line of best fit along with 95 % confidence intervals. This allows the uncertainty in these deep drainage estimates to be assessed for the first time.

## 1 Introduction

The capacity of effective groundwater management practices are expanded as our understanding of the driving processes, timescales and spatial distribution of groundwater recharge develop. Rainfall is an important predictor of groundwater recharge and as such there is value in understanding the spatial and temporal nature of rainfall. Petheram et al. (2002) established that rainfall variation accounted for 60 % of observed recharge variation for annual vegetation on sand soils

and 23 % of observed recharge variation for annual vegetation on non-sand soils across Australia. However, accurate measurement of recharge may be difficult to attain due to the small deep drainage fluxes that occur below the root zone. Invariably, groundwater recharge is associated interchangeably with deep drainage or *potential* recharge albeit a distinct difference defines the two terms. Simply, recharge is the flux of water that contributes to the saturated groundwater reservoir whereas deep drainage is the flux of water that moves through the unsaturated zone past the root zone. Deep drainage may become recharge after a lag time if there are no impeding layers that would create lateral flow. Assuming one-dimensional flow and that deep drainage will eventually recharge the unconfined aquifer, unsaturated soil zone profiles can provide long-term mean annual recharge estimates for both steady state and transient environments (Walker, 1998). Point scale methods are used to estimate low deep drainage fluxes in a transient environment however scaling up is often not practical because of the costs involved to obtain the required data. Hence, it is necessary to use a surrogate measurement that is more readily available as a proxy for deep drainage.

The first effort to introduce a surrogate clay measurement for predicting deep drainage was by Kennett-Smith et al. (1994) who defined a log-linear relationship between deep drainage and clay content for the 0–2 m soil depth interval. The relationship was based on 89 deep drainage estimates made using the chloride front displacement (CFD) method (Walker et al., 1991) in the 300–400 mm rainfall areas of South Australia and Victoria in south-eastern Australia



**Fig. 1.** Location of studies used to develop the multiple linear relationships.

(Fig. 1). The average clay content over the 0–2 m depth interval (rather than 0–0.5 m or 0–1 m) was chosen as it provided the highest correlation for deep drainage and clay content using the Kennett-Smith et al. (1994) data set (SKM, 2002).

More recent investigations (Leaney et al., 2004; Wohling, 2007) revised the relationship given by Kennett-Smith et al. (1994) for higher rainfall areas based on local recharge investigations; however minimal statistical rigour was employed in developing the revised correlations. Personal judgement defined each new relationship rather than a statistical correlation.

Comparison of the log-linear relationship between deep drainage and average clay content (0–2 m) data from Kennett-Smith et al. (1994), Leaney et al. (2004) and Wohling (2007) provided evidence for a trend attributable to rainfall amount (Fig. 2). This can be seen in the relationship of Wohling (2007) which gives the highest deep drainage throughout the range of clay content using data from field sites with a rainfall range of 470–570 mm yr<sup>-1</sup>, compared to the 390–500 mm yr<sup>-1</sup> range of Leaney et al. (2004) and 300–400 mm yr<sup>-1</sup> range of Kennett-Smith et al. (1994).

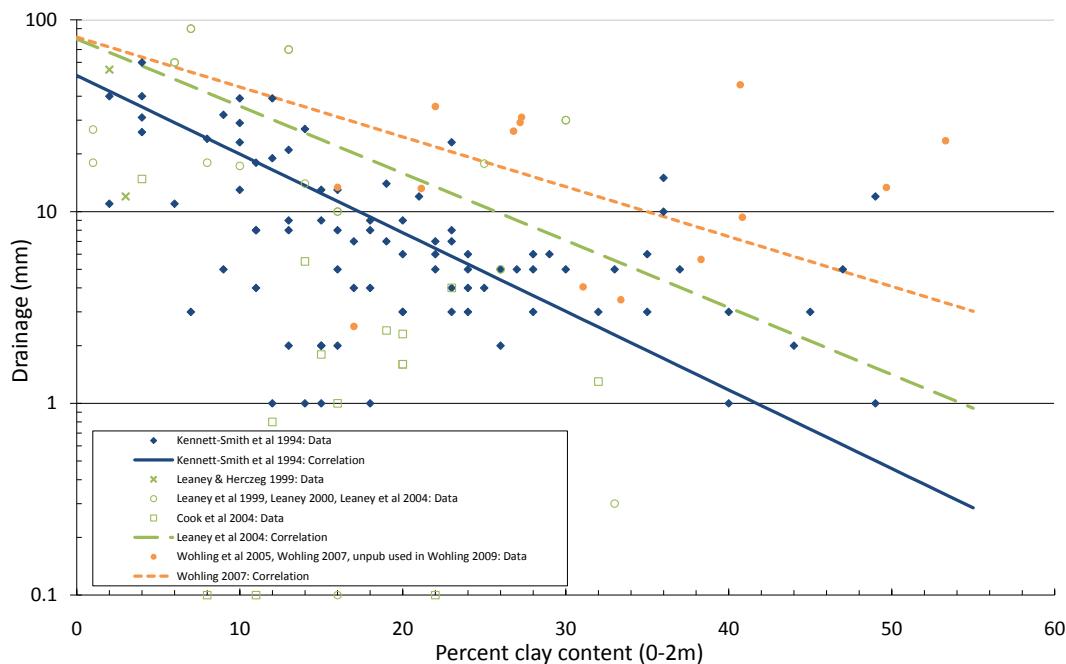
These relationships between deep drainage and average clay content (0–2 m) have been widely used as a proxy for determining deep drainage rate estimates beneath annual crops and pastures on land previously cleared of native vegetation (Kennett-Smith et al., 1994; Leaney et al., 1999, 2004; Leaney and Herczeg, 1999; Leaney, 2000; Cook et al., 2001,

2004; Wohling et al., 2005; Wohling, 2007) in southeastern Australia.

Crosbie et al. (2010a) recently compiled and reviewed a database of 4386 recharge and deep drainage estimates over 172 studies across dry land areas of Australia. This included a review of several previous Australian recharge review papers (Kennett-Smith et al., 1994; Petheram et al., 2000, 2002; Cook et al., 2001; Tolmie and Silburn, 2003; Silberstein, 2010). Crosbie et al. (2010b) then used this database to develop a tool for estimating the relationships between recharge and rainfall, soil type and land use in data poor areas. The soil type in these relationships were classified into five discrete groups, however the soil properties relevant to deep drainage are known to exist as a continuum rather than discrete classes (Kennett-Smith et al., 1994).

This paper tests the finding of SKM (2002) that the average clay content of the top 2 m of the soil profile is the best proxy for estimating deep drainage by examining a more substantial database (Crosbie et al., 2010a) in terms of the data quantity and spatial variability (including soil and rainfall variability) to provide the best unsaturated zone proxy for estimating deep drainage.

A literature search of comparable studies found that no other work had used actual soil clay content data as a proxy for deep drainage or recharge in correlation with rainfall as field clay content is typically not reported or reported inconsistently. Sophocleous (1992) used multi-linear regression



**Fig. 2.** Summary of deep drainage and percent clay content (0–2 m) correlations.

analysis of total annual precipitation, average maximum spring soil profile water storage, average shallowest spring depth to water table and average spring precipitation rate to assess recharge. Delin et al. (2000) applied multi-linear regression models to assess the effects of topography and soil properties on recharge on lowland and upland agricultural field sites near Princeton, Minnesota; and used precipitation (amount of water applied), average rainfall or irrigation water application intensity and antecedent moisture content for the entire unsaturated zone as the independent variables. Keese et al. (2005) evaluated controls on recharge by simulating deep drainage in a five metre thick one-dimensional unsaturated flow code using climate data, vegetation and soil coverages from online sources for Texas and simulated 30-year mean annual recharge. Vegetation and soil textural variability both resulted in a large range of recharge rates within each region; however Keese et al. (2005) also found that the spatially weighted long-term recharge rates were much less variable and positively correlated with mean annual rainfall. For non-vegetated, texturally variable simulations, Keese et al. (2005) explained that multi-linear regressions using log-log data showed that including clay content (profile average) with precipitation explains 80 % of the variation in recharge and that recharge varied over 1 to 2 orders of magnitude locally, because of textural variability among soil profiles. The modelled results presented by Keese et al. (2005), specifically the order of magnitude over which recharge varied, are comparable with the measured results presented in this paper.

This paper builds upon previous research by using multiple linear regression (MLR) to develop relationships capable

of predicting deep drainage from measured soil clay content and mean annual rainfall, where previous studies have not used both rainfall and measured soil clay content to estimate deep drainage. Specifically the aims of this paper are to:

1. determine if a statistically significant relationship can be developed to predict deep drainage using previously published field data;
2. determine the best metric of the clay content of the soil for predicting deep drainage; and
3. estimate the uncertainty in the deep drainage estimates made using MLR.

## 2 Methods

### 2.1 Data collation of previous studies

In order to improve the statistical correlation between deep drainage, clay content and rainfall, as developed in this paper; the database of Crosbie et al. (2010a) was filtered to include only those recharge and deep drainage estimates that reported rainfall amount and clay content of the soil. Less than 5 % (202 recharge/deep drainage estimates) of the original database (Crosbie et al., 2010a) fitted these criteria (Fig. 1). It was from this filtered set of data, which includes all data presented in Fig. 2, that the MLR was performed. The data was re-queried and re-classified to assess how soil structure may affect deep drainage rates. The average clay content was determined over various intervals (0 to 0.5 m,

0 to 1 m and 0 to 2 m) to determine the best depth for averaging to use as a predictor of deep drainage. Also, a classification based on the highest clay percentage in a soil profile (rather than an average over an interval) was used to investigate whether or not the use of impeding layers could improve the correlation.

The data was categorized into annuals, perennials and tree type vegetation as differing vegetation types were found to give differing rates of deep drainage for a particular soil type and rainfall (Crosbie et al., 2010b). An analysis was also performed on a combined perennial and tree category due to a scarcity of data for both groups. Of those 202 recharge and deep drainage estimates having soil textual data, 177 are estimates beneath annuals, 7 beneath perennials and 18 beneath trees. The mean annual rainfall ranged between 260 and 731 mm and clay content between zero and 60 % for those annual vegetation sites; while for perennial vegetation, mean rainfall ranged from 587 mm to 1265 mm and soil clay content between zero and 60 %; and for tree vegetation annual rainfall varied from 460 mm to 731 mm with clay content between 0.7 and 60 %.

Data used in this paper are heavily biased in the south-western Murray Darling Basin with the remaining data coming from the mid coast of New South Wales (NSW) and south eastern Queensland (Fig. 1). The studies that have generated most of the data are summarised as follows. Kennett-Smith et al. (1994) compiled a review of studies in the south-western Murray Darling Basin, primarily the Mallee and Wimmera regions of north-western Victoria, eastern South Australia and south-western NSW. Kennett-Smith et al. (1994) used the chloride front displacement (CFD), water balance, and water table fluctuation (WTF) methods to determine deep drainage and recharge associated with the clearing of *Eucalyptus* mallee woodland (in the form of shrub land, scrub or heath) between 1900–1914 and 1920–1930 for rotational cereal cropping, sheep and beef cattle grazing. Leaney and Herczeg (1999) and Cook et al. (2001) studied the Mallee region of the south-west Murray Darling Basin including the Little Desert and Big Desert using soil-water chloride concentration profiles, groundwater chloride, stable isotopes of water and carbon-14 to determine recharge processes and provide deep drainage estimates beneath Mallee vegetation and cleared pastures and cropping. Similarly, Leaney (2000) investigated recharge processes and provided deep drainage estimates for the Tintinara area of South Australia using the CFD and steady state soil chloride mass balance (CMB) methods for mallee vegetation and cleared dry land cropping and pastures. Cook et al. (2004) investigated deep drainage and aquifer recharge in the north-east Mallee region of South Australia and determined that clearing of native vegetation (*Eucalyptus* mallee) and replacement with dryland cropping and pastures lead to increased salt loads to the Murray River. Cook et al. (2004) used the CFD method for estimating point estimates of deep drainage and developed equations for estimating the time lag between an increase in deep drainage

and increase in aquifer recharge. They then spatially extrapolated the data to provide a deep drainage map and predicted time delays for transport through the unsaturated zone.

Wohling et al. (2005) used the CFD method to estimate deep drainage beneath pastures in the Padthaway region of south east South Australia. Leaney et al. (2006) used the CMB method under native vegetation, dry land pasture, and tree plantations in the lower south-east of South Australia. In the Border Designated Area and Keith region of south-east South Australia, Wohling (2007) used the water balance, daily soil water balance, CMB, CFD and LEACHM (Leaching Estimation and Chemistry Model, Hutson 2003) to determine recharge and deep drainage rates beneath native vegetation, pastures and crops. Unpublished data used in Wohling (2009) followed on from the work of Wohling et al. (2005) to estimate additional deep drainage rates under pastures using the CFD in the Padthaway region of south-east South Australia.

Tolmie et al. (2004) used a transient CMB method in south-east Queensland to determine recharge beneath native vegetation, cropping and pastures, and perennial lucerne. Crosbie et al. (2005) investigated recharge in the Tomago Sand Beds near Newcastle, NSW. The WTF method was used to predict gross recharge under perennial vegetation in heath and mining revegetation areas.

## 2.2 Statistical methods

Regression analysis explains movements in the dependant variable ( $Y$ ) as a function of the independent variables ( $X$ ) through the quantification of an equation. The simplest measure of the quality of fit is the coefficient of determination ( $R^2$ ), which is the proportion of variation of  $Y$  about its mean ( $\bar{Y}$ ) that is explained by the model. The higher the  $R^2$  the closer the estimated regression fits the sample data.

The level of significance (significant  $F$ ) tests whether the hypothesis, that the data forms a relationship, is null, specifically testing whether the data has no relationship. If the hypothesis is null it can be concluded that the equation does have a significant overall fit. For example, when significant  $F$  is less than 0.05 (5 %), then a relationship is likely. If significant  $F$  is more than 0.05, then a relationship is unlikely.

The P-value provides a measure of how much evidence we have against the null hypothesis, i.e. we test the probability of getting statistics approaching the actual observed data, assuming a hypothesis that is null. The lower the P-value, the more significant the result is as it is unlikely the null hypothesis is true.

Average annual deep drainage estimates ( $d$ ,  $\text{mm yr}^{-1}$ ) were set as the dependant  $Y$  variable and clay content (either average or highest percentage, %) and average annual rainfall (mm) were set as the independent  $X$  variables.  $R^2$ , significant  $F$ , P-value and regression coefficients were presented as outputs, along with the 95 % prediction intervals around

**Table 1.** Multiple linear regression results under annual vegetation.

Data treatment, ANNUALS	Average clay % (0–0.5 m)	Average clay % (0–1 m)	Average clay % (0–2 m)	Highest clay % (0–2 m)
$R^2$	0.4323	0.4890	0.3521	0.3233
Observations	79	83	177	51
Significant $F$	4.54E-10	2.17E-12	3.99E-17	8.51E-05
P-value (clay variable)	0.867	0.837	3.7E-05	0.608
P-value (rainfall variable)	1.16E-10	2.84E-13	5.68E-18	2.65E-05
Coefficient – Intercept ( $y_0$ ) multiple regression	-0.493	-0.757	-0.039	-1.006
Coefficient – $b$ (clay) multiple regression	-6.20E-04	7.87E-04	-1.19E-02	-2.75E-03
Coefficient – $a$ (rainfall) multiple regression	3.52E-03	3.98E-03	3.03E-03	4.38E-03
Coefficient – intercept interval – ( $y_0$ ) 95 % prediction – lower			-0.993	
Coefficient – $b$ (clay) 95 % prediction interval – lower			-1.19E-02	
Coefficient – $a$ (rainfall) 95 % prediction interval – lower			2.99E-03	
Coefficient – intercept ( $y_0$ ) 95 % prediction interval – higher			0.915	
Coefficient – $b$ (clay) 95 % prediction interval – higher			-1.19E-02	
Coefficient – $a$ (rainfall) 95 % prediction interval – higher			3.06E-03	

the best fit for the best model for each vegetation type. All regression equations were given in the form:

$$\log d = y_0 + (b \cdot \text{clay}) + (a \cdot \text{rainfall}) \quad (1)$$

where  $y_0$ ,  $b$  and  $a$  were fitting parameters.

### 3 Results

#### 3.1 Determine if a statistically significant relationship can be developed to predict deep drainage using previously published field data

Within each vegetation category, the data was interrogated to summarise average clay contents over various depth intervals or the highest clay content across a depth interval. For each interrogated data set, a MLR analysis was performed for both deep drainage and log deep drainage terms. Consistently the log deep drainage regression gave a superior  $R^2$ , and therefore we only provide those results here. Table 1 summarises results from the MLR for annual vegetation and emphasizes that statistically significant relationships between the average clay content over the 0–0.5 m, 0–1 m and 0–2 m intervals and highest clay content over the 0–2 m interval can be developed to predict deep drainage. For each case stated in Table 1, the Significant  $F$  term is less than 0.05, indicating that a relationship between log deep drainage, rainfall and each clay content variable are likely to be true.

Results from the MLR analysis for deep drainage under trees and perennials are summarised in Table 2 and again highlight statistically significant relationships with Significant  $F$  less than 0.05 in each case. Limited soil textural information was available in conjunction with deep drainage or

recharge estimates for trees and perennials around Australia and as such a combined category was assessed.

#### 3.2 Determine the best metric of the clay content of the soil for predicting deep drainage

As described previously, the simplest measure of fit,  $R^2$ , illustrates how well the MLR model fits the sampled data. For the case of deep drainage below annuals (post-clearing) data, the strongest correlation was given by the average clay content (0–1 m),  $R^2 = 0.489$ . However when we investigated the statistical fit more thoroughly we found that even though there was an overall significance to the MLR for average clay content (0–1 m), (Significant  $F = 2.17E-12$ ), there was no correlation between log deep drainage and clay (P-value = 0.837). It was similarly true for the average clay content (0–0.5 m) and highest clay content (0–2 m) MLR results. Despite having a lower degree of fit to the data ( $R^2 = 0.3521$ ), the MLR statistical results for average clay content (0–2 m) showed a relationship was very likely to exist between average clay content (0–2 m), rainfall and log deep drainage (Significant  $F = 3.99E-17$ ). Moreover, there was evidence for correlations between log deep drainage and clay (P-value = 3.7E-05); and log deep drainage and rainfall (P-value = 5.68E-18). From the evidence provided in Table 1, we conclude that the best metric of clay content to use as a surrogate for determining deep drainage is the average between zero and two metres.

A MLR using the average clay content (0–2 m) for trees gave a good fit against the data,  $R^2 = 0.7665$ . A MLR was performed on average clay contents (0–0.5 m) and (0–1 m) beneath trees, however the results are not given here. Whilst a relationship was likely between log deep drainage,

**Table 2.** Multiple linear regression results under trees and perennial vegetation types.

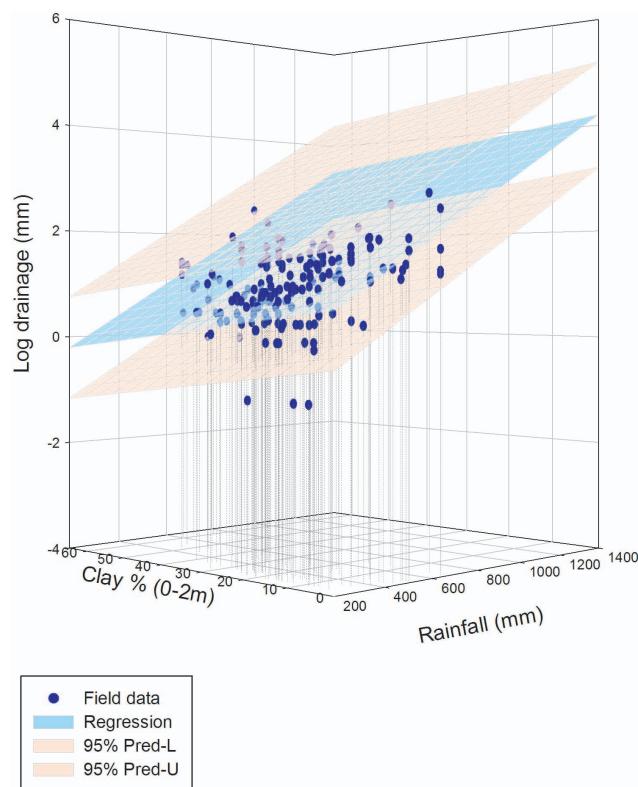
	TREES	PERENNIALS	TREES & PERENNIALS
Data treatment, TREES & PERENNIALS	Average clay % (0–2 m)	Average clay % (0–2 m)	Average clay % (0–2 m)
$R^2$	0.7665	0.9373	0.8696
Observations	18	7	25
Significant $F$	1.83E-05	3.93E-03	1.86E-10
P-value (clay variable)	4.92E-06	0.727	1.62E-06
P-value (rainfall variable)	5.89E-03	0.039	2.27E-07
Coefficient – intercept ( $y_0$ ) multiple regression	-0.986	-1.846	-0.723
Coefficient – $b$ (clay) multiple regression	-3.35E-02	-4.77E-03	-2.74E-02
Coefficient – $a$ (rainfall) multiple regression	3.62E-03	3.91E-03	2.99E-03
Coefficient – intercept ( $y_0$ ) 95 % prediction interval – lower			-1.671
Coefficient – $b$ (clay) 95 % prediction interval – lower			-2.76E-02
Coefficient – $a$ (rainfall) 95 % prediction interval – lower			2.92E-03
Coefficient – intercept ( $y_0$ ) 95 % prediction interval – higher			0.225
Coefficient – $b$ (clay) 95 % prediction interval – higher			-2.72E-02
Coefficient – $a$ (rainfall) 95 % prediction interval – higher			3.08E-03

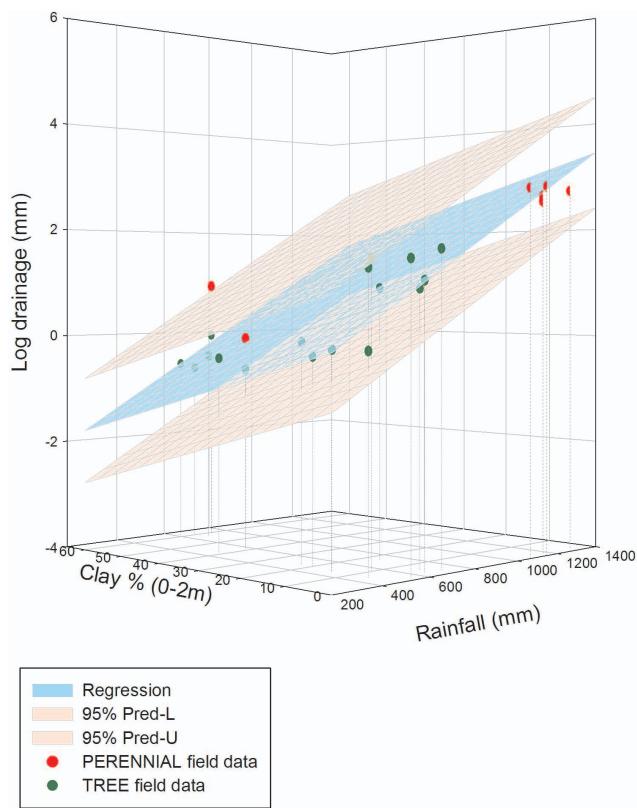
rainfall and average clay content (0–0.5 m and 0–1 m), significant  $F < 0.05$ , in each case a relationship was not likely between log deep drainage and clay; and log deep drainage and rainfall, P-value  $> 0.05$ . The database included deep drainage, rainfall and average clay content (0–2 m) information for seven perennial vegetation sites that provided a very good statistical fit ( $R^2 = 0.9373$ ) and showed that a relationship was likely to exist (Significant  $F = 3.93E-03$ ). However, it was unlikely that a correlation existed between log deep drainage and average clay content (0–2 m), P-value = 0.727.

A review of the recharge and deep drainage database (Crosbie et al., 2010a) provided limited soil textural data for tree and perennial land use types and as such a merged MLR analysis was performed to test whether an increased sample size would improve the data fit and to test the level of significance of the data set. The merged tree and perennial MLR provided a better fit of the data compared to trees alone,  $R^2 = 0.8696$ , and illustrated that a relationship between average clay (0–2 m), rainfall and log deep drainage for the merged data was likely to exist, Significant  $F = 1.86E-10$ .

### 3.3 Estimate the uncertainty in the deep drainage estimates made using MLR

Given that a relationship between average clay content (0–2 m), rainfall and log deep drainage was very likely to exist (Significant  $F = 3.99E-17$ ), we were warranted in producing 95 % prediction intervals for the annual vegetation log deep drainage, rainfall and average clay content (0–2 m) correlation (Table 1, Fig. 3). The 95 % prediction interval described a range within which 95 % of data was expected to occur. Data points that fell outside of this range were for estimates

**Fig. 3.** Plot of multi linear regression equation developed for annual vegetation type also showing 95 % prediction intervals.



**Fig. 4.** Plot of multi linear regression equation developed for combined tree and perennial vegetation types also showing 95 % prediction intervals.

of deep drainage in very low rainfall zones ( $<300\text{ m}$ ) that have very low deep drainage ( $<1\text{ mm yr}^{-1}$ ).

Figure 4 illustrates the MLR analysis for the merged trees and perennials data set and includes 95 % prediction intervals to indicate the level of uncertainty surrounding the analysis. Perennials (red circles) and trees (green circles) are highlighted in different colours to distinguish between the two data sources. Two clusters of perennial vegetation data are evident in Fig. 4; (1) those having high rainfall, zero clay content and high recharge; and (2) those having lower rainfall, high clay content (42–60 %) and minimal recharge. The lack of perennial vegetation recharge data prevented a meaningful relationship being developed and hence use of the merged perennial and tree data set has been provided.

#### 4 Discussion

SKM (2002) summarised a compilation of work undertaken during the 1990's by various authors (Kennett-Smith et al., 1992a,b, 1993; HydroTechnology, 1994; O'Connell et al., 1997) and provided an assessment of recharge dependence on the average clay content for several intervals in the unsaturated zone (0–0.5 m, 0–1 m and 0–2 m) beneath non-cleared

and cleared land in the Wimmera-Mallee region of southern Australia. With limited data available in terms of clay content, and encompassing a narrow rainfall range, SKM (2002) suggested that clay content within the 0–2 m interval provided the best fit as a proxy measurement for determining deep drainage. Now, with the aid of a larger database of recharge and deep drainage estimates across Australia detailing clay content in the unsaturated zone over a larger rainfall range, we have been able to develop a more reliable approach for predicting deep drainage and can now provide an uncertainty range about those deep drainage estimates. Using this larger data set, we have provided statistical evidence that

- the average clay content rather than the highest clay content for a soil profile was a more appropriate proxy measure;
- the average clay content (0–2 m) as opposed to 0–0.5 m and 0–1 m still provided the best predictor of deep drainage;
- in conjunction with rainfall data improved the significance of such a correlation with deep drainage that had previously not been specified; and
- have an improved confidence in using the average clay content (0–2 m) of the unsaturated zone as a reliable proxy for estimating deep drainage rates across Australia.

It must be noted that all deep drainage estimates presented were point-scale measurements and as such any uncertainty range relates to that point-scale. Crosbie et al. (2010b) explains that if relationships developed to predict deep drainage at smaller than paddock scale were then used to estimate deep drainage at regional scales, the uncertainty would be overstated.

Many of the deep drainage data analysed here have been estimated using the CFD method; assumptions to these estimates include the date of native vegetation clearance. The clearance date has a large effect on deep drainage rate predictions using the CFD method. In some cases, we have taken broad scale information rather than paddock scale and subsequently have used a very uncertain date rather than the exact date. If a site was not cleared or cleared sometime after we had originally assumed, then there is a probability that we have been looking at a historical rather than a displaced chloride profile, particularly in very low rainfall zones like the Mallee region.

A larger sample size can imply that there will be a reduced likelihood of a few data points skewing the result. For example, the average clay content (0–1 m) annual data set had three data points with low deep drainage rates that had very low clay contents and tended to bias the data set, therefore giving the MLR a negative relationship between clay and deep drainage. This was a relationship that was opposite to what we would have expected.

The 95% prediction intervals given in Fig. 3 ( $\log d$  vs. average clay content 0–2 m vs. rainfall) display an approximate two order of magnitude spread of data. The method of Crosbie et al. (2010b) developed to predict recharge in data poor areas presented 95 % prediction intervals having greater than two orders of magnitude threshold. Our 95 % prediction intervals were realised via the development of regression equations made using field measurements of soil clay content, as compared to Crosbie et al. (2010b) who assigns an Australia-wide coverage of integrated vegetation and soil types to correlate mean annual rainfall with recharge estimates. More importantly, our work provided an indication of the uncertainty in the deep drainage vs. percent clay correlation that has been used in studies for the last two decades without any such consideration or quantification of the level of uncertainty involved.

Numerous factors affect diffuse recharge to unconfined aquifers including vegetation type and density, climate variability, soil texture, land use and land management practices. As a general rule, recharge decreases as soil texture becomes heavier i.e. percent clay content is elevated (Kennett-Smith et al., 1994). Heavy textured soils have lower hydraulic conductivities and higher water holding capacities, whereas sandy soils have high permeability and low water holding capacity and as such deep drainage through sandier textured soils is likely to be higher than through heavier (clay) soils (SKM, 2002). Although under very high clay contents the correlation between deep drainage and soil structure becomes less apparent as cracks and other preferential flow paths can increase deep drainage (SKM, 2002).

The frequency and distribution of heavy rainfall events, particularly in more arid areas, may provide a greater dependence on localised recharge (e.g. flooding, stream bed recharge) as opposed to diffuse recharge mechanisms. In semi-arid to arid landscapes where low water fluxes are common, residence times in unsaturated zones may be long, and as such the potential for spatial and temporal variability of deep drainage can be significant. Moreover, the unpredictability of recharge due to climatic variability (rainfall and evapotranspiration amount, duration and intensity) combined with spatial variability of soil texture presents difficulties when estimating the spatial and temporal variability of recharge. While the spatial variability within a land use class may present uncertainties when assessing site specific recharge rates in a broad ranging correlation. For example, annuals encompass a considerable diversity of annual crops and pastures and can represent a variety of farming management practices, including cropping and fallow rotations, direct drill or conventional tillage, and winter crops or summer crops.

Given the range of variables that impact on deep drainage estimates and the resultant level of uncertainty of any such estimate, it would be natural to pose questions over the applicability of scaling up point deep drainage measurements to management scale estimates of deep drainage and associated

uncertainty predictions. Up scaling requires mapping of percentage clay content over large areas. Consequently, the accuracy of deep drainage estimates at management scale boundaries are dependent upon the correlation between the measured soil profile clay content at each investigation site compared to an estimate of clay content from a Geographical Information System (GIS) layer, for example the Soil Landscape Unit (SLU) coverage (DWLBC, 2010). A regression through measured soil textural data and SLU soil textural estimates from Wohling (2007) (data not provided here) gave evidence that use of such a relationship for up scaling was not reliable. Consequently up scaling using SLU estimates of clay content (0–2 m), which will ultimately depend on a correlation between field and SLU estimates of clay content, should be exercised with caution.

An over-riding question is whether there was enough data to give indicative deep drainage information about perennial (clay content 0–2 m, 7 sites) and tree (clay content 0–2 m, 18 sites) vegetation types? When we combined the trees (clay content 0–2 m, 18 sites) and perennials (clay content 0–2 m, giving a total of 25 data points), the MLR correlation was more significant than either of the individual trees or perennials datasets. It also provided a better regression coefficient ( $R^2$ ) than either of the tree datasets. Until such time as the database can be expanded with more trees and perennials information, use of a combined correlation is prudent.

## 5 Conclusions

By incorporating a multi-linear regression approach, we have been able to establish significant relationships to predict deep drainage beneath annual and, tree and perennial vegetation types using measured and published soil clay content field data and mean annual rainfall data. Furthermore, we have statistically established that the best metric of the clay content of the soil is the average clay content from 0 to 2 m for both annual vegetation and a combined tree and perennials data set. Previous studies had not used both mean annual rainfall and measured soil textural data to estimate deep drainage. Furthermore, given the significance of the correlations between deep drainage, percent clay content and rainfall, we provided 95 % prediction intervals to illustrate the uncertainty of those deep drainage estimates.

As in the future, deep drainage rates are estimated at new sites concurrently with the collection of soil texture data, we will be able to improve the relationships between rainfall, clay content and deep drainage. We expect that the order of magnitude that our 95 % prediction intervals encompass will diminish as the uncertainty of deep drainage estimates improves with the addition of further data, that encompasses a greater range of clay content and annual rainfall data.

Improving our methodology for estimating deep drainage rates in areas where data capture has traditionally been difficult, or of low priority, enables predictions of groundwater

flow to be developed and modelled with greater reliability. However it is the prediction of uncertainty surrounding those estimates that will ultimately facilitate our capacity to better manage the groundwater resource in times of increasing demand.

**Acknowledgements.** This paper builds on several projects funded by the Australian Government. Daniel Wohling's contribution forms part of the "Minimising Salt Accession in the South East of South Australia" project funded by the National Action Plan for Salinity and Water Quality. Fred Leaney and Russell Crosbie's contribution was funded by the "A Consistent Approach to Groundwater Recharge Determination in Data-Poor Areas" project funded by the National Water Commission. Ian Jolly and Steve Barnett provided constructive and valuable comments that benefited the manuscript considerably.

Edited by: S. Attinger

## References

- Cook, P. G., Leaney, F. W., and Jolly, I. D.: Groundwater Recharge in the Mallee Region, and Salinity Implications for the Murray River – A Review, CSIRO Land and Water, Technical Report No. 45/01, 2001.
- Cook, P. G., Leaney, F. W., and Miles, M.: Groundwater Recharge in the North-East Mallee Region, South Australia, CSIRO Land and Water Technical Report 25/04, 2004.
- Crosbie, R. S., Binning, P., and Kalma, J. D.: A time series approach to inferring groundwater recharge using the water table fluctuation method, *Water Resour. Res.*, 41, W01008, doi:10.1029/2004WR003077, 2005.
- Crosbie, R., Jolly, I., Leaney, F., Petheram, C., and Wohling, D.: Review of Australian Groundwater Recharge Studies, CSIRO: Water for a Healthy Country National Research Flagship, 2010a.
- Crosbie, R. S., Jolly, I. D., Leaney, F. W., and Petheram, C.: Can the dataset of field based recharge estimates in Australia be used to predict recharge in data-poor areas?, *Hydrol. Earth Syst. Sci.*, 14, 2023–2038, doi:10.5194/hess-14-2023-2010, 2010b.
- Delin, G. N., Healy, R. W., Landon, M. K., and Bohlke, J. K.: Effects of topography and soil properties on recharge at two sites in an agricultural field, *J. Am. Water Resour. Assoc.*, 36, 1401–1416, 2000.
- DWLBC: Soil Landscapes – Analysis Data, Government of South Australia, Department of Water, Land and Biodiversity Conservation, Soil and Land Program, 2010.
- Hutson, J. L.: Leaching Estimation and Chemistry Model, LEACHM, Model description and user's guide, School of Chemistry, Physics and Earth Sciences, The Flinders University of South Australia, 2003.
- HydroTechnology: Estimation of Groundwater Recharge in the Mallee and Wimmera (south-western Murray Basin): A Land Systems Approach, Draft Report prepared for MDBC/CSIRO, May 1994.
- Keese, K. E., Scanlon, B. R., and Reedy, R. C.: Assessing controls on diffuse groundwater recharge using unsaturated flow modelling, *Water Resour. Res.*, 41, W06010, doi:10.1029/2004WR003841, 2005.
- Kennett-Smith, A. K., O'Leary, G. J., and Thorne, R.: Estimation from soil water profiles under dryland agriculture, eucalypt plantation, and natural woodland in the Wimmera district of Victoria, CSIRO, Division of Water Resources, Divisional Report 92/2, 1992a.
- Kennett-Smith, A. K., Cook, P. G., and Thorne, R.: Comparison of recharge under native vegetation and dryland agriculture in the Big Desert region, Victoria, CSIRO Centre for Groundwater Studies Report No. 46, 1992b.
- Kennett-Smith, A. K., Thorne, R., and Walker, G. R.: Comparison of recharge under native vegetation and dryland agriculture near Goroke, Victoria, CSIRO Centre for Groundwater Studies Report No. 49, 1993.
- Kennett-Smith, A. K., Cook, P. G., and Walker, G. R.: Factors affecting groundwater recharge following clearing in the south western Murray Basin, *J. Hydrol.*, 154, 85–105, 1994.
- Leaney, F.: Groundwater Salinisation in the Tintinara area of SA: Results of field investigations, April 2000, CSIRO Land and Water, Technical Report 34/00, 2000.
- Leaney, F., Walker, G., Knight, J., Dawes, W., Bradford, A., Barnett, S., and Stadter, F.: Potential for Groundwater Salinisation in the Tintinara area of SA: Impacts of planned irrigation allocations, Centre for Groundwater Studies, Technical Report No 33/99, 1999.
- Leaney, F., Barnett, S., Davies, P., Maschmedt, D., Munday, T., and Tan, K.: Groundwater Salinisation in the Tintinara Highland Area of SA: Revised estimates using spatial variation for clay content in the unsaturated zone. CSIRO Land and Water Technical Report 24/04, 2004.
- Leaney, F. W., Mustafa, S., and Lawson, J.: Salt Accumulation and Water Balance Under Different Land Use in Bakers Range Area, CSIRO Land and Water Science Report 05/06, 2006.
- Leaney, F. W. J. and Herczeg, A. L.: The Origin of Fresh Groundwater in the SW Murray Basin and its Potential of Salinisation, CSIRO Land and Water, Technical Report 7/99, 1999.
- O'Connell, M., Thorne, R., O'Leary, G., Mason, G., and Hoxley, G.: Water movement beneath cleared and native vegetation in the Victorian Mallee and Wimmera Region, in: Murray Darling 1997 Workshop Extended Abstracts, Toowoomba, Queensland, 36–41, 1997.
- Petheram, C., Walker, G., Grayson, R., Thierfelder, T., and Zhang, L.: Towards a framework for predicting impacts of land-use on recharge: 1. A review of recharge studies in Australia, *Aust. J. Soil Res.*, 40, 397–417, 2002.
- Petheram, C., Zhang, L., Walker, G., and Grayson, R.: Towards a framework for predicting impacts of land-use on recharge: 1. A review of recharge studies in Australia, CSIRO Land and Water Technical Report 28/00, Canberra, 2000.
- Silberstein, R. P.: Review of Recharge and Water Use Studies of Vegetation on Gnangara Mound, CSIRO Land and Water, Science Report, Perth, 2010.
- SKM: Recharge in the Mallee, An Estimation of Groundwater Recharge in the Mallee and Wimmera (South-western Murray Basin): A Land System Approach, Sinclair Knight Merz Pty Ltd., 2002.
- Sophocleous, M.: Groundwater recharge estimation and regionalization: the Great Bend Prairie of central Kansas and its recharge statistics, *J. Hydrol.*, 137, 113–140, 1992.

- Tolmie, P. E. and Silburn, D. M.: Estimating Deep Drainage in the Queensland Murray-Darling Basin: Review of Past Research, Queensland Department of Natural Resources and Mines Report QNRM04100, Toowoomba, 2003.
- Tolmie, P. E., Silburn, D. M., and Biggs, A. J. W.: Estimating deep drainage in the Queensland Murray-Darling Basin using soil chloride, Department of Natural Resources and Mines, Coorparoo, Queensland, QNRM03020, 2004.
- Walker, G. R.: Using Soil Water Tracers to Estimate Recharge, Studies in Catchment Hydrology: The Basics of Recharge and Discharge, CSIRO Australia, 1998.
- Walker, G. R., Jolly, I. D., and Cook, P. G.: A new chloride leaching approach to the estimation of diffuse recharge following a change in land use, *J. Hydrol.*, 128, 49–67, 1991.
- Wohling, D.: Minimising Salt Accession to the South East of South Australia. The Border Designated Area and Hundred of Stirling Salt Accession Projects, Volume 2 – Analytical Techniques, Results and Management Implications, DWLBC Report 2008/23, Government of South Australia, through Department of Water, Land and Biodiversity Conservation, Adelaide, 2007.
- Wohling, D.: Investigations to enhance the Padthaway groundwater model, South Australia, Department of Water, Land and Biodiversity Conservation, DWLBC Report 2009/32, 2009.
- Wohling, D., Leaney, F., Davies, P., and Harrington, N.: Groundwater Salinisation in the Naracoorte Ranges Portion of the Padthaway Prescribed Wells Area, South Australia, Department of Water, Land and Biodiversity Conservation, DWLBC Report 2005/27, 2005.