

## Research Article

# Quantification of Uncertainty in Mathematical Models: The Statistical Relationship between Field and Laboratory pH Measurements

Kurt K. Benke<sup>1,2</sup> and Nathan J. Robinson<sup>3,4</sup>

<sup>1</sup>School of Engineering, University of Melbourne, Parkville, VIC, Australia

<sup>2</sup>Department of Economic Development, Jobs, Transport and Resources (DEDJTR), Parkville Centre, 32 Lincoln Square North, Parkville, VIC, Australia

<sup>3</sup>Department of Economic Development, Jobs, Transport and Resources (DEDJTR), Bendigo Centre, Cnr Midland Hwy and Taylor Street, Epsom, VIC, Australia

<sup>4</sup>Faculty of Science and Technology, Federation University, University Drive, Mount Helen, VIC, Australia

Correspondence should be addressed to Kurt K. Benke; Kurt.Benke@ecodev.vic.gov.au

Received 20 February 2017; Accepted 5 June 2017; Published 23 August 2017

Academic Editor: Balwant Singh

Copyright © 2017 Kurt K. Benke and Nathan J. Robinson. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The measurement of soil pH using a field portable test kit represents a fast and inexpensive method to assess pH. Field based pH methods have been used extensively for agricultural advisory services and soil survey and now for citizen soil science projects. In the absence of laboratory measurements, there is a practical need to model the laboratory pH as a function of the field pH to increase the density of data for soil research studies and Digital Soil Mapping. The accuracy and uncertainty in pH field measurements were investigated for soil samples from regional Victoria in Australia using both linear and sigmoidal models. For samples in water and CaCl<sub>2</sub> at 1:5 dilutions, sigmoidal models provided improved accuracy over the full range of field pH values in comparison to linear models (i.e., pH < 5 or pH > 9). The uncertainty in the field results was quantified by the 95% confidence interval (CI) and 95% prediction interval (PI) for the models, with 95% CI < 0.25 pH units and 95% PI = ±1.3 pH units, respectively. It was found that the Pearson criterion for robust regression analysis can be considered as an alternative to the orthodox least-squares modelling approach because it is more effective in addressing outliers in legacy data.

## 1. Introduction

The assessment of soil pH provides information for diagnosis of soil condition and contributes to routine field surveys of soil properties [1]. The pH is probably the most commonly measured soil property [2, 3], yielding information on “plant nutrient availability, aluminium and heavy metal availability, organic matter decomposition, liming requirements, and microbial activity.” The measurement of soil pH can vary markedly depending upon sample collection and treatment procedures including soil-to-solution ratio and addition of indicators or reagents.

The pH of soil samples taken in the field is often measured in situ using a colorimetric indicator method that is based

on subjective assessment between a standard pH colour chart and the colour of the soil in response to indicator reagents [4] or flocculating agents [1]. Field pH measurements are cost-effective, accessible and convenient, and safe, require minimal training, and are almost instantly available. Although colour matching is a fast and inexpensive method for field pH assessment (pH<sub>F</sub>), there is uncertainty in the prediction accuracy and precision, or error, in the results. Laboratory pH is routinely measured in a 1:5 soil-to-water suspension (pH<sub>W</sub>) with possible addition of a salt solution (typically CaCl<sub>2</sub> or KCl) to account for seasonal variations or management interventions [5]. The lab pH measurement is used as the reference, given that its relative accuracy and precision is typically ±0.1 units [6].

In the absence of laboratory measurements there has been high dependence on field pH measurements for screening and soil classification purposes. The rapid uptake of Digital Soil Mapping (DSM) [7, 8] combined with citizen science has the potential to increase the geographic spread and density of observations to improve mapping and predictions of soil properties [9]. There are opportunities to exploit the thousands of legacy field pH measurements in government databases in combination with contemporary measurements from citizens to better understand the distribution and changes in soil acidity. Methods and mathematical models will be required to transform these field measurements into laboratory equivalents for soil mapping and science purposes.

The relationship between laboratory and field pH using a colorimetric indicator method has been the subject of studies in the quest for reliable field methods. Wherry [10] led initial discussions on the implementation of field assessment methods to measure pH. Numerous commercially available pH kits were developed over the following two decades but questions remained as to the reliability and relative merits of using these kits in preference to laboratory analysis [11]. Some kits achieved comparable accuracies but the simple and rapid operation of a kit was found to be important also. In Australia, Raupach [12] pioneered an indicator method that satisfied these requirements of accuracy and simplicity. Further refinement and testing by Raupach and Tucker [4] established that a single determination of pH using the indicator method resulted in a standard deviation of 0.42 units from the laboratory reference method. Farr [1] also found that the standard deviation between a field colorimetric method and laboratory method was of a similar magnitude.

Assessment of pH with colour cards from soil test kits can be affected by a range of errors associated with manual application of the test kit in operational conditions. Steinhardt and Mengel [13] identified that there were potential sources of uncertainty leading to a variation between field and laboratory methods including poorly adjusted indicators, seasonal variation of pH, sample variation, and overlap in pH intervals for the chemicals used in the indicator solution. Baker et al. [14] provide an example where, with one operator performing all field pH assessments, there can be strong agreement with laboratory pH measurements ( $r^2 = 0.91$ ,  $n = 288$ ).

Uncertainty in  $pH_F$  interpretation is introduced by such factors as user experience with colorimetric indicator kits, visual deficiencies, daylight spectral content, and variability in test kits. Recognition of the possible sources of error in the assessment of field pH will help to reduce the dispersion of results for  $pH_F$  through the implementation of concise and practical quality assurance procedures (see discussions on risk and uncertainty [15–17]). This will ultimately benefit the soil science community including those in DSM that can make use of this “soft” data resource [18].

There are numerous examples where laboratory pH methods have been harmonised using datasets of various sizes for different regions (see Minasny et al. [3]). However, we are unaware of any attempts to harmonise field pH measurements with laboratory methods using extensive datasets that potentially contain a vast range of error sources. This has been identified as a challenge in the integration of citizen

supplied measurements with those that are more precise [9].

The objectives of this investigation were to (a) establish a functional relationship between field pH and lab pH with an extensive dataset from regional Victoria, (b) quantify accuracy and uncertainty of the field pH measurements, and (c) identify possible sources of error in field pH data including visual interpretation of the colour card measurement. In establishing a functional relationship between field and laboratory pH measurements, an experiment was designed to compare (i) linear and sigmoidal models, (ii) least-squares and Pearson performance indices, and (iii) quantised data and randomised data.

## 2. Materials and Methods

**2.1. Study Area and Soil Data.** The dataset used for analysis is derived from the *Victorian Soil Information System* (VSIS) which contains field and laboratory pH observations from soil profiles collected across regional Victoria.

In practice,  $n = 2436$  soil samples were used to establish a mathematical relationship for field pH with the corresponding laboratory pH measurements (Figure 1). Laboratory measurements were undertaken with a 1:5 soil-to-water suspension (Method 4A1) and a 0.01 M  $CaCl_2$  extract ( $pH_C$ ; Method 4B1) [6]. Field pH measurements were taken according to Method 4G1, the field determination method [4] as described in Rayment and Lyons [6].

**2.2. Field Measurements of pH.** In the field determination technique, a colour card with predefined pH levels is used to make a subjective assessment against a paste composed of indicator solution (phenol red, bromocresol purple, and bromocresol green) mixed with soil and dusted with barium sulphate powder. The reagents used in the field pH technique have remained the same since method development including the quantity of reagents in the indicator solution. The colour chart was developed with the combination of reagents in mind to “find the best range of colours” [4]. The current version of the colour chart features 16 colour intervals to represent soil pH values between  $pH = 2$  and  $pH = 10$ , as shown in Figure 2 (Inoculo Laboratories, Australia, 2014). The variability in pH measurements is shown in Figure 3.

**2.3. Laboratory Measurements of pH.** Measurements for  $pH_w$  and  $pH_C$  since 2010 are acquired automatically using a *Radiometer Analytical SAS* titration system comprising a PHM92 pH meter, a CDM240 conductivity meter, and a SAC950 sample changer. These instruments are calibrated according to the methods described by Rayment and Lyons [6]. A similar automated system from the same company was used between 1992 and 2010 according to the methodology described by Rayment and Higginson [19]. The  $pH_w$  and  $pH_C$  results were determined manually using earlier models of the same manufacturer’s equipment prior to 1992. Instruments were calibrated according to the manufacturer’s specifications. The error in laboratory pH measurements is reported as  $\pm 0.1$  pH units [6].

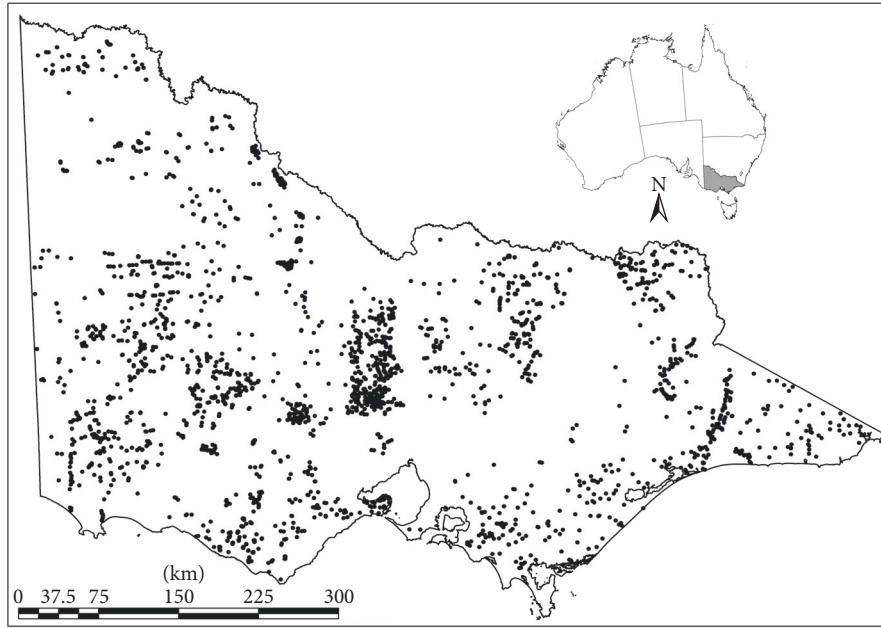


FIGURE 1: Locations for soil pH measurements in Victoria using a soil pH indicator test kit.

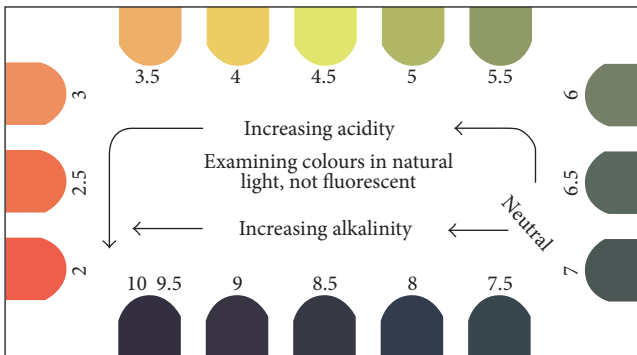


FIGURE 2: Example soil pH colour indicator card (Inoculo Laboratories, Australia, 2014).

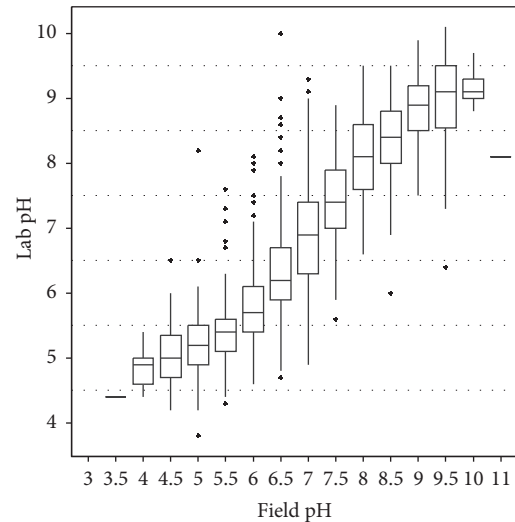


FIGURE 3: Box plot shows variability in field pH measurements.

**2.4. Data Processing.** The field pH measurements and corresponding laboratory pH measurements were used as inputs to evaluate a range of proposed mathematical models used to predict lab pH from field pH. Model fitting was carried out using SYSTAT Software TableCurve 2D V5.01. More than 3000 different mathematical models were fitted using an automated process covering a variety of functional types: for example, polynomial types, transition functions, exponential types, linear types, and mixed models. Indications of goodness of fit were provided by various statistical criteria, including coefficient of determination ( $r^2$ ), the ANOVA  $F$ -statistic, and sum-of-squares errors (SSE). The best fits based on the foregoing criteria (and presented in this paper) were associated with linear and sigmoidal functions, depending on the input range selected in field pH measurements.

**2.5. Data Randomisation.** A field measurement of soil pH is determined by treating the sample with an indicator solution with barium sulphate and then comparing the resultant colour with colours from a card with pH values quantised according to a colour step wedge (Figure 2). The available discrete colour palette ( $N = 16$ ) is associated with a series of increasing pH values, with 0.50 increment; that is,  $x = 2, 2.5, 3, 3.5, \dots, 10$ . The possible error in the step wedge reading is therefore 0.50 pH units due to misclassification to an adjacent colour.

The transformation used for randomisation, in order to compare results with the fixed discrete palette, is as follows.

For each value of the field pH,  $x$ , at a specified *fixed* increment along the  $x$ -axis, replace  $x$  with a random value  $x_{\text{rand}}$  over the interval  $[x_{\text{min}}, x_{\text{max}}]$  according to

$$x_{\text{rand}} = x_{\text{min}} + R [x_{\text{max}} - x_{\text{min}}], \quad (1)$$

where  $R$  is a real random number which is then scaled to occupy the defined interval  $[x_{\text{min}}, x_{\text{max}}]$ . A set of contiguous intervals of this width covers the full domain.

Therefore, in the step wedge, each value of  $x_{\text{rand}}$  is

$$x_{\text{rand}} = f(x_{\text{min}}, x_{\text{max}}, R), \quad (2)$$

where  $x_{\text{min}} = x - 0.25$ ,  $x_{\text{max}} = x + 0.25$ ,  $R = \text{rand}[0, 1]$ .

By a process of *inductive reasoning* in mathematics, this procedure can be generalised for an arbitrary dataset. For each value  $x$  in the original dataset, replace it with  $x_{\text{rand}}$  in the new dataset with the following transformation:

$$x_{\text{rand}} = \left(x - \frac{w}{2}\right) + r \left[\left(x + \frac{w}{2}\right) - \left(x - \frac{w}{2}\right)\right], \quad (3)$$

where  $x, x_{\text{rand}} \in \mathfrak{R}$ ,

where  $r$  is the random variable (in this case, the uniform probability distribution) and  $w$  is the width of the class interval about the quantised variable  $x$  in the step wedge. This operation produces a random scatter plot suitable for regression analysis. The operation is consistent with the theory of quantisation error in digital-to-analog conversion [20] and the uniform prior distribution in Bayesian theory [21]. It reflects the uncertainty in the input of  $x$  due to the quantised nature of the colour card.

A metric of uncertainty for random data in the interval  $[x_{\text{min}}, x_{\text{max}}]$  is the variance,  $\sigma^2$ , and for a uniform distribution,  $S$ , is given by

$$S = U(x_{\text{rand}}, \sigma^2), \quad (4)$$

where

$$\sigma^2 = \frac{1}{12} (x_{\text{max}} - x_{\text{min}})^2. \quad (5)$$

The uncertainty in the result due to quantisation of data is associated with a fixed increment step wedge and is equivalent to the error in measurement from an analog-to-digital (A/D) converter.

**2.6. Model for Functional Relationship.** In the case of a perfect match between field pH and lab pH, the 1:1 plot of the functional relationship would follow a straight line ( $y_{\text{Lab pH}} = a + by_{\text{Field pH}}$  with  $a = 0$  and  $b = 1$ ) indicating one-to-one correspondence and zero average bias. This approach is associated with a ‘‘calibration’’ exercise. It is prudent, however, to also consider a functional relationship that is nonlinear. This is because the field pH and lab pH are measured under different physical and methodological conditions and are likely to show some divergence at the extremes of the pH range. Data processing and experimental results described

later suggest the two model types that fitted the data best (linear and sigmoidal) were range dependent.

The models were tested on the experimental data and compared under different modelling conditions. First, the linear model with two parameters ( $a$  and  $b$ ):

$$y' = a + bx. \quad (6)$$

The second model was a generic S-shaped curve, the so-called sigmoidal model with 4 parameters  $a$ ,  $b$ ,  $c$ , and  $d$ :

$$y'' = a + \frac{b}{(1 + \exp(-((x - c)/d)))}. \quad (7)$$

Given the pH domain ( $\text{pH}_F = 2-10$ ), it may be that a piecewise approximation of functions provides the best relationship over the full range; that is,

$$y = \begin{cases} y', & \text{if } \text{pH}_F > \alpha \text{ or } \text{pH}_F < \beta, \\ y'', & \text{if } \text{pH}_F \leq \alpha \text{ or } \text{pH}_F \geq \beta, \end{cases} \quad (8)$$

where  $\alpha$  and  $\beta$  are  $\text{pH}_F$  values bounding the linear approximation, near the toe and shoulder of the curve. Concatenation of regression-based models, however, introduces problems of continuity at the interface and a single model covering the full domain is more desirable.

Two error minimisation schemes were tested, with the first being the standard least-squares regression analysis,  $\delta_1$ , where

$$\text{Minimise } \delta_1 = \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (9)$$

There is a problem with least-squares fitting based on minimisation of the sum-of-squares residuals. In the presence of significant outliers, the square of the residuals may shift the fitted curve away from the main data in some subregions. An ad hoc approach often used is to *exclude* data beyond three standard deviations of the mean, but such deletion changes the sample set used by a ‘‘cookie cutter’’ approach and was not considered.

A more robust approach is the Pearson index,  $\delta_2$ , for curve fitting. This index is much less sensitive to outliers than the least-squares criterion and is given by

$$\text{Minimise } \delta_2 = \sum_{i=1}^n \ln \left( \sqrt{1 + |y_i - \hat{y}_i|^2} \right). \quad (10)$$

Outliers in this case have much less impact on the fitted curve due to the logarithmic transformation, which compresses the difference between measured and predicted values. The operation is effective for large differences that are due to isolated outliers. The Pearson index is robust compared with least-squares fitting and is appropriate for the case where unpredictable random errors may appear in the original dataset. This situation can occur in the case of field pH measurements where quality control (QC) is not observed, in contrast to laboratory pH measurements.



Finally, a corollary note on uncertainty in the linear model: the well-known expression for error propagation in a multivariate predictive model,  $y = f(x_i)$ , is given by differential error analysis, often referred to as the *chain rule* or *delta rule* [22]:

$$\sigma^2(y) = \sum_{i=1}^n \left[ \frac{\partial f(\mathbf{x}_0)}{\partial x_i} \right]^2 \sigma^2(x_i). \quad (11)$$

In the case of a simple univariate linear model with unity slope and zero average bias, the last expression can be simplified to the following (for (6)):

$$\sigma^2(y) \approx \sigma^2(x). \quad (12)$$

That is, the error or uncertainty in the  $x$  value is transmitted directly to the output value  $y$  without expansion or compression. A linear model represents an ideal “calibration” plot between field pH measurements and lab pH measurements. A linear approximation, however, may not necessarily be appropriate at very low or very high pH values due to the physics of the measurement apparatus and soil properties.

**2.7. Goodness of Fit.** The performance of all models was evaluated by various criteria to confirm goodness of fit to the data. This included coefficient of determination ( $r^2$ ) for explained variation by the model, sum-of-squares errors (SSE), also known as fit standard error (FitStdErr), and the  $F$ -statistic for significance of the regression from ANOVA. For the linear model, prediction accuracy and dispersion (uncertainty) were both important, as well as bias (i.e., proximity to optimum values:  $a = 0$  and  $b = 1$ ). Also computed were the *uncertainty intervals*, that is, the confidence interval (95% CI) and prediction interval (95% PI). The CI and PI are computed as follows.

The CI for the *average* value of  $y$  from the model, denoted by  $\hat{y}$ , is specified for a given  $x^*$ , with  $s$  for standard deviation of residuals, and, using the  $t$ -distribution, is given by [23]

$$\hat{y} = \pm t_{n-2}^* s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{(n-1)s_x^2}}. \quad (13)$$

Also, the PI for a *single* future value of  $y$  from the model, denoted by  $\hat{y}$ , specified for a given  $x^*$ , is broader by a factor of unity in the square root sign:

$$\hat{y} = \pm t_{n-2}^* s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{(n-1)s_x^2}}. \quad (14)$$

The CI accounts for the uncertainty in estimation of the mean value  $\hat{y}$  whilst the PI covers fluctuations and is therefore wider in span.

### 3. Results

**3.1. Field pH versus Lab pH in Water (pH<sub>F</sub> versus pH<sub>W</sub>).** A factorial experiment ( $2 \times 2 \times 2$ ) was implemented for modelling the relationship between field pH and lab pH

TABLE 1: Nested factorial design for linear model: field pH versus lab pH in 1:5 water (pH<sub>F</sub> versus pH<sub>W</sub>).

Model	Linear			
Data type	Original data		Randomised	
Performance index	Least squares	Pearson criterion	Least squares	Pearson criterion
Figure	Figure 4(a)	Figure 4(b)	Figure 5(a)	Figure 5(b)

TABLE 2: Nested factorial design for sigmoidal model: field pH versus lab pH in 1:5 water (pH<sub>F</sub> versus pH<sub>W</sub>).

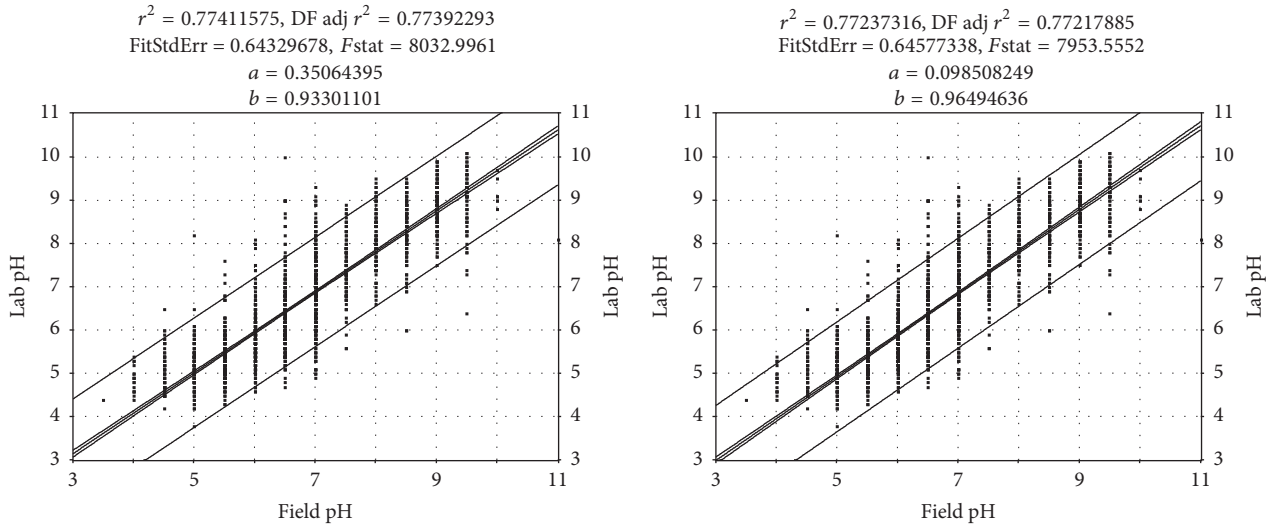
Model:	Sigmoidal			
Data type	Original data		Randomised	
Performance index	Least squares	Pearson criterion	Least squares	Pearson criterion
Figure	Figure 6(a)	Figure 6(b)	Figure 7(a)	Figure 7(b)

using (a) the linear and sigmoidal models, (b) least-squares and Pearson performance indices, and (c) fixed and data randomisation regimes, as shown in Tables 1 and 2. The results are illustrated in Figures 4(a)–7(b) with the 95% CI and 95% PI presented and in greater detail in the following sections.

The dataset for field pH is plotted against laboratory measurements in Figure 4(a), where field pH appears in increments of 0.5, as vertical arrays of data at fixed horizontal increments, which is referred to as *quantisation* in the field pH measurements. This vertical linear structure in the field data, although used in the past by others [12], may have introduced problems with respect to assumptions underlying model fitting by statistical methods (regression), which is normally applied to random scatter plots. For the purpose of experimental comparison, random values were generated to replace discrete vertical columns of data points, as depicted in Figure 5(a).

**3.1.1. Linear Model: Original Data for pH<sub>F</sub> versus pH<sub>W</sub>.** A plot of lab pH versus field pH for the experimental dataset is presented in Figure 4(a) with the linear model fitted by least-squares. The corresponding linear model fitted with the Pearson criterion rather than least-squares is shown in Figure 4(b). These model parameters (Table 3) show that 77% percent of the variation is explained by linear regression in both cases but the result for the Pearson fit to the data is slightly less significant. The Pearson criterion is less sensitive to outliers and is more robust but conservative, whereas the least-squares criterion may have been affected by some outliers. Both models are highly significant and acceptable, according to the  $F$ -statistic, with the Pearson model having an advantage of following the 1:1 line more closely in a “calibration” sense.

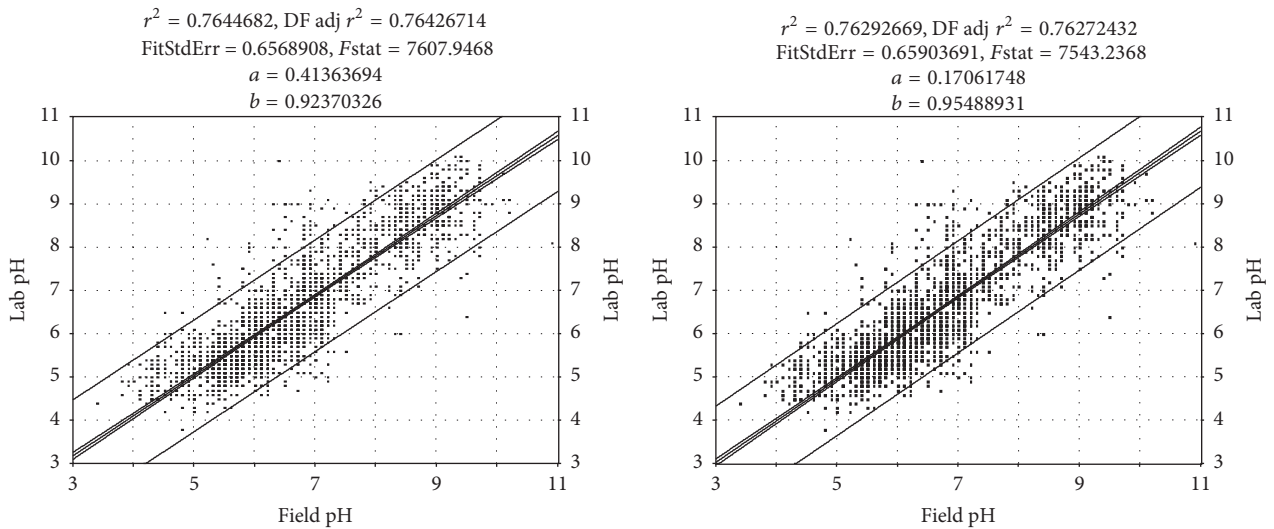
**3.1.2. Linear Model: Randomised Data for pH<sub>F</sub> versus pH<sub>W</sub>.** In the case of the randomised data, in Figure 5(a) the linear



(a) Linear model (original data) for field  $pH_F$  versus  $pH_W$  with least-squares minimisation and 95% CI (narrow band) and 95% PI (broad band). Note the degree of variation possible at different pH levels (cf. Figure 3, Steinhardt and Mengel [13])

(b) Linear model (original data) for field  $pH_F$  versus  $pH_W$  with Pearson minimisation and 95% CI (narrow band) and 95% PI (broad band)

FIGURE 4



(a) Linear model (randomised data) for field  $pH_F$  versus  $pH_W$  with least-squares minimisation and 95% CI (narrow band) and 95% PI (broad band)

(b) Linear model (randomised data) for field  $pH_F$  versus  $pH_W$  with Pearson minimisation and 95% CI (narrow band) and 95% PI (broad band)

FIGURE 5

model was fitted by least-squares estimation, and the corresponding linear model fitted with the Pearson criterion rather than least-squares is shown in Figure 5(b). The difference between ordered pairs of quantised and randomised data for field pH in Figures 4(a) and 5(a) was tested for significance by using the parametric paired  $t$ -test for observations. With null hypothesis  $H_0: \mu_1 - \mu_2 = \mu_D$ , a two-sided test produced

no significant difference for  $\mu_D$  ( $p < 0.01$ ). This means that randomisation did not introduce any significant difference or bias to the data apart from less significance in the model fitted but slightly wider uncertainty bands, as expected.

In the case of least-squares fitting in Figure 5(a), note that the intercept or average bias of 0.414 is positive (Table 3), which is nearly 0.50, that is, the step increment in the colour

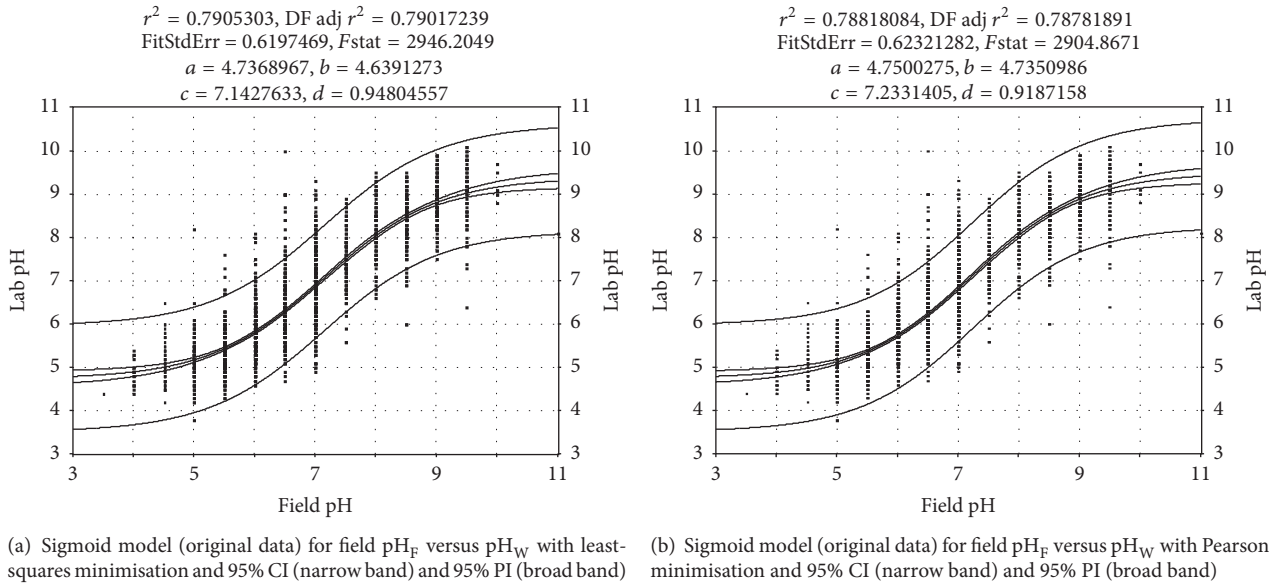


FIGURE 6

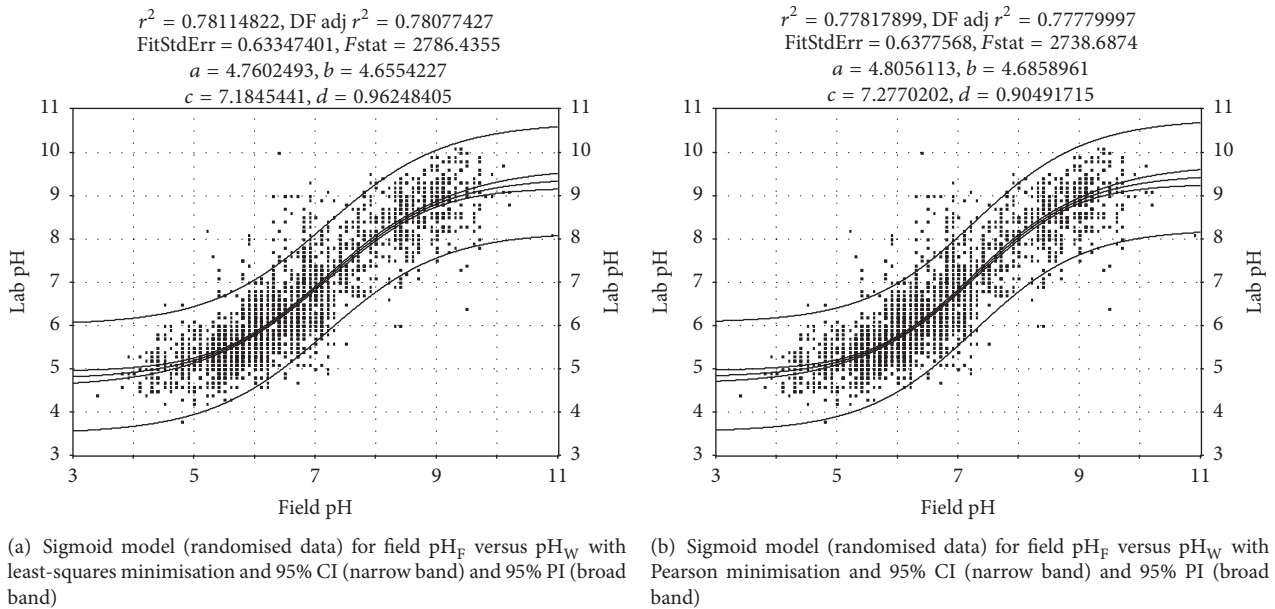


FIGURE 7

card. Note also that the slope is 0.924, which is nearly unity. This result captures the uncertainty in the input value of field pH and also provides a linear relationship with lab pH.

3.1.3. Sigmoidal Model: Original Data for  $pH_F$  versus  $pH_W$ . The four-parameter sigmoidal model fitted to the original data is shown in Figure 6(a) and the corresponding sigmoidal model fitted with the Pearson criterion rather than least-squares is shown in Figure 6(b). That is, 79% percent of the variation is explained by sigmoidal regression in both cases

but the result for the Pearson fit to the data is slightly less significant (Table 4). Whilst the global fit for the sigmoidal model is slightly improved with respect to explained variation ( $r^2$ ), the significance of the fit ( $F$ -value) is much less, despite having more degrees of freedom with respect to the number of parameters. The sigmoidal model is a better fit in the extremes; that is, field  $pH < 5$  and field  $pH > 9.5$ . This is clearly evident by visual inspection of Figure 4(a) versus Figure 6(a) at  $pH_F = 4$  which shows that nearly all scatter points lie above the mean prediction in the linear case, whereas, in the

TABLE 3: Model parameters (rounded to three significant figures) for the linear model of field pH against lab pH using both the least squares and Pearson index for the quantised and randomised data.

Model fit	Quantised		Randomised	
	Least squares	Pearson index	Least squares	Pearson index
$r^2$	0.774	0.772	0.764	0.763
$F_{1,2435}$	8033	7954	7608	7543
$p$	<0.00001	<0.00001	<0.00001	<0.00001
Fit s.e.	0.643	0.646	0.657	0.659
$a$	0.351	0.099	0.414	0.171
$b$	0.933	0.965	0.924	0.955

TABLE 4: Model parameters for the sigmoidal model of field pH against lab pH with least squares and Pearson index fit for the quantised and randomised data.

Model fit	Quantised		Randomised	
	Least squares	Pearson index	Least squares	Pearson index
$r^2$	0.791	0.788	0.781	0.778
$F_{1,2435}$	2946	2905	2786	2739
$p$	<0.00001	<0.00001	<0.00001	<0.00001
Fit s.e.	0.620	0.623	0.635	0.638
$a$	4.737	4.750	4.760	4.806
$b$	4.639	4.735	4.655	4.686
$c$	7.143	7.233	7.185	7.277
$d$	0.948	0.919	0.962	0.905

sigmoidal case, the mean prediction nearly bisects the scatter data. The toe of the curve is fairly constant for lab pH < 5.

**3.1.4. Sigmoidal Model: Randomised Data for  $pH_F$  versus  $pH_W$ .** The four-parameter sigmoidal model was also fitted to the original data, as shown in Figure 7(a). The corresponding sigmoidal model fitted with the Pearson criterion is shown in Figure 7(b). Both models using sigmoidal regression explain around 78% percent of the variation but the result for the Pearson fit to the data is slightly less significant (Table 4). Overall results are similar but slightly worse than the linear models for the quantised data. This is expected, due to accommodation for the uncertainty in the step wedge increments in the colour card.

**3.2. Field pH versus lab pH in  $CaCl_2$  ( $pH_F$  versus  $pH_C$ ).** A plot of  $pH_F$  versus  $pH_C$  for the quantised data is given in Figure 8(a) with a linear model fitted by least-squares estimation, where  $r^2 = 0.751$ ,  $F_{1,2435} = 4441$  ( $p < 0.00001$ ), and Fit s.e. = 0.689. The least-squares intercept is  $a = 0.890$  and slope is  $b = 0.987$ .

A plot of  $pH_F$  versus  $pH_C$  for the quantised data with the sigmoidal model fitted by least-squares estimation is given in

Figure 8(b), where  $r^2 = 0.766$ ,  $F_{1,2435} = 1607.3$  ( $p < 0.00001$ ), and Fit s.e. = 0.668. The least-squares parameters for the sigmoidal model are  $a = 3.952$ ,  $b = 4.711$ ,  $c = 7.233$ , and  $d = 0.9030$ . In Figure 8(c), the quantised data is plotted with the Pearson criterion, where  $r^2 = 0.755$ ,  $F_{1,2435} = 1510.1$  ( $p < 0.00001$ ), and Fit s.e. = 0.684. The least-squares parameters for the Pearson sigmoidal model are  $a = 4.050$ ,  $b = 4.752$ ,  $c = 7.413$ , and  $d = 0.8044$ .

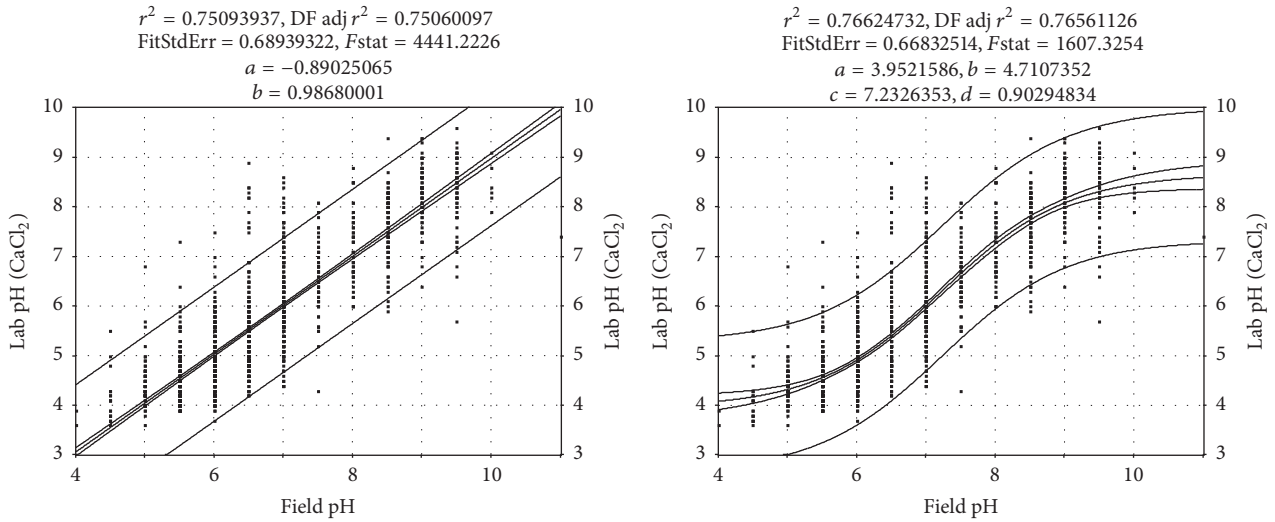
## 4. Discussion

**4.1. Comparison between Field pH versus Lab pH in Water ( $pH_F$  versus  $pH_W$ ).** The original data with the linear model and least-squares fitting produced a good fit with average bias of  $a = 0.35$  as can be seen in Figure 4(a). The fit was not as good below  $pH_F = 5$  and above  $pH_F = 9$ , where a number of points fell outside the 95% CI, but mostly within the 95% PI. Replacing the least-squares criterion with the Pearson criterion for robust fitting, in order to minimise the effect of outliers or spurious results, was more effective, with the coefficients producing an excellent linear relationship (average bias of  $a = 0.10$  and slope  $b = 0.96$ ), as can be seen in Figure 4(b). Both the 95% CI and 95% PI were only slightly broader, despite less weighting given to outliers. This model, using the Pearson criterion, was perhaps the most useful linear relationship because of the fit with the 1:1 line, with nearly zero bias and unity slope. As the logarithmic transformation compresses the range that limits the weighting given to outliers, this has effectively reduced the impact of outliers (with potential high error due to interpretation error in the field). In contrast, the least-squares approach amplifies the effect of outliers appearing in the sum-of-squares difference expression comparing predictions and measurements. The uncertainty interval in the field pH was about  $\pm 1.3$  units with reference to lab pH @  $pH_F = 7$  for the 95% PI, with this error band being similar across the full domain.

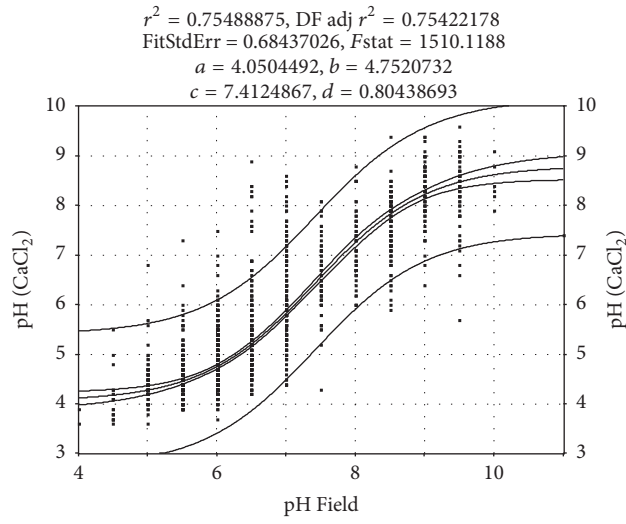
When the quantised data were randomised, which was equivalent to an injection of uniform random noise to account for the uncertainty due to step width on the colour card, the model fit was still very good but the average bias had increased to  $a = 0.41$ , as evident in Figure 5(a). The uncertainty intervals were similar in width. This model with randomised quantised predictions arguably reflected the uncertainty due to the colour matching best because the average bias was 0.41, which is nearly the width of the step wedge increment of 0.5  $pH_F$  units. This assumes, however, that the colour card was the main source of uncertainty, which has not been confirmed. Finally, the uncertainty interval in the field pH was about  $\pm 1.3$  units with reference to lab pH at  $pH_F = 7$ , with this error band being similar across the full domain.

A combination of randomised data and Pearson criterion produced the result depicted in Figure 5(b), which was a pragmatic real-world result for the linear model. The average bias was only 0.17 and slope 0.96. Overall, the regression statistics were very significant for all four linear models, which could all be used to represent the relationship between field pH and lab pH, with the main difference being the values of the model coefficients for bias and slope.





(a) Linear model (original data) for field  $pH_F$  versus  $pH_C$  with least-squares minimisation and 95% CI (narrow band) and 95% PI (broad band) (b) Sigmoidal model (original data) for field  $pH_F$  versus  $pH_C$  with least-squares minimisation and 95% CI (narrow band) and 95% PI (broad band)



(c) Sigmoidal model (original data) for field  $pH_F$  versus  $pH_C$  with Pearson minimisation and 95% CI (narrow band) and 95% PI (broad band)

FIGURE 8

The sigmoidal model was also fitted under the specified conditions, that is, with (a) original data with least-squares, (b) original data with Pearson minimisation, (c) randomised data with least-squares, and (d) randomised data with Pearson minimisation. Adding white noise by randomisation to account for step wedge uncertainty produced visible but small broadening in the uncertainty intervals on the plots. The most significant result was that the Pearson criterion and randomisation did not lead to appreciable increases in the 95% CI and 95% PI.

Although the linear model was sufficient for most of the pH range and is suitable for general use, the sigmoidal model produced a better fit to the data than the linear model below  $pH_F = 5$  and above  $pH_F = 9$ , which is clearly visible on all plots for the 95% PI (where the mean value lies at the

centre of the scatter data in contrast to the linear model). This is consistent with models derived by Henderson and Bui [24] and Minasny et al. [3] in the prediction of  $pH_C$  where a nonlinear model (sigmoidal in our example) performs better than a linear model for the highly buffered soil at extremely low and high pH values. The results provided by Figure 6(a) for the original data produced the best overall relationship between the field pH and lab pH. The result for Figure 6(b) using the Pearson index indicates no additional advantage from using this approach.

4.2. Comparison between Field pH and Lab pH in  $CaCl_2$  ( $pH_F$  versus  $pH_C$ ). Modelling field pH against lab pH in  $CaCl_2$  solution using the original data was completed first with the linear model. Although the linear model produced statistical

significance in the regression analysis and provided a useful model ( $r^2 = 0.75$ ), it revealed an average bias of  $-0.89$  and slope of  $0.99$  with less accurate curve fitting in the toe and shoulder of the plot. This result is similar to Slattery et al. [2], where a difference of  $0.84$  was determined. In contrast, the sigmoidal plot produced an improved fit to the data below  $\text{pH}_F = 5$  and above  $\text{pH}_F = 9$ , as evident in Figure 8(b), where  $r^2 = 0.77$ . This represents an improvement over the linear approximation. Application of the Pearson criterion as illustrated in Figure 8(c) resulted in slight broadening of the uncertainty interval with this error band being also similar across the full range, that is, about  $\pm 1.3$   $\text{pH}_w$  units for the 95% PI with reference to lab pH @  $\text{pH}_F = 7$  in the sigmoidal plot, and similar across the full range. Randomisation was not applied to the sigmoidal plot as it was found from the previous results that a marginal increase in the uncertainty intervals (95% CI and 95% PI) added no significant advantage to the models with quantised field pH observations.

Models fitted for samples in water ( $\text{pH}_w$ ) were superior on statistical metrics to models fitted in  $\text{CaCl}_2$  ( $\text{pH}_C$ ), for both standard least-squares and Pearson error minimisation approaches (cf. Figure 4(a) versus Figure 8(a) and also Figure 6(a) versus Figure 8(b)).

**4.3. Sources of Error in Field pH Determination.** Using soil pH test kits introduces a number of factors that increase uncertainty but are not greatly appreciated by users. There are four primary factors in field pH testing which relate to human vision and the interpretation of the colour cards in the test kits including

- (i) colour vision deficiencies;
- (ii) changes in daylight spectral content;
- (iii) atmosphere light scattering;
- (iv) variability in pH test kits.

Detailed discussion on these error sources can be found in the literature (see, e.g., DeMarco et al. [25], K. K. Benke and K. E. Benke [26], and Self [27]).

It is recommended that more attention be given in future to

- (a) quality control of test kits to monitor manufacturing variability;
- (b) initial eye testing for new operators of test kits to check for colour vision deficiencies;
- (c) field application of test kits to reflect time-of-day effects on colour rendition, for example, noon versus dusk;
- (d) warnings of possible colour misinterpretation due to wearing sunglasses, or indoor use of test kits.

**4.3.1. Seasonal Variability in pH Effects.** Change in soil pH during a year can be in the order of  $0.5$  pH units. This temporal trend can be cyclical or dominated by periods of extreme seasonal conditions such as high rainfall and temperature differences that impact the soil water content [28]. Whilst temporal changes have been observed in laboratory

pH measurements due to these factors, it is unclear if changes will likewise occur in field pH measurement. Samples for field pH assessment should be in a field moist state; however, this is sometimes impractical due to logistical constraints that may constrain the time of year in which sampling took place and the environmental conditions leading up to and at the time of sampling.

**4.3.2. Sample Variability.** The field determination method recommends the use of a small quantity of field moist soil (about half a gram) for assessment against the colour card. Although field pH kits recommend replication of such pH assessments, this rarely occurs in field situations. As a consequence there is the potential to inadvertently assess a subsample that is “atypical” of the sampled depth or horizon. An example is where extreme differences have been recorded between field and laboratory measurements. From 18 records in the dataset for this study, where lab pH was  $>1.5$  pH units above field pH, 15 were from northwest Victoria and exhibited strong to very strong effervescence with the addition of 1 M HCl to the sample. This suggests that the subsample used for field pH assessment contained concentrated carbonates that were unrepresentative of the total homogenised sample.

A further consideration is where a gradation in pH exists for peds of the subsample. There is a possibility that the field pH result may vary with the lab pH measurement on a homogenised sample which has been prepared with all coarse material ( $>2$  mm) removed.

## 5. Conclusion

**5.1. General Comments.** Characterisation of the field pH dataset for Victoria was accomplished and documented using a range of fitted models under different constraints and modelling conditions. In particular, the linear and sigmoidal models fitted to the data provided statistically significant relationships between field and laboratory pH measurements. The results suggest that a portable test kit with a colour card is a rapid and inexpensive approach to soil pH estimation because the results can be readily calibrated with laboratory measurements. There is, however, greater uncertainty in the field results as quantified by the 95% CI and 95% PI.

Although the linear model represents a good general model, the sigmoidal relationship provided a better fit across the full range of pH measurements, especially for extreme values. With respect to the field pH observations overall, the field pH uncertainty is represented by 95% CI ( $<0.25$  pH units) and 95% PI ( $\pm 1.3$  pH units), with reference to laboratory  $\text{pH}_w$  versus  $\text{pH}_C$  (which are both measured to within a tolerance of  $\pm 0.1$  pH units).

This study has quantified the level of uncertainty in field pH measurements with reference to laboratory pH, supporting the use of field pH data from citizens for mapping and monitoring purposes. Field pH observations can be of value where accurate and precise laboratory measurements are absent in the space-time inventory of soils. Harmonising  $\text{pH}_F$  with  $\text{pH}_w$  versus  $\text{pH}_C$  via predictive models will assist the application of legacy data and contemporary citizen-sourced measurements in Digital Soil Mapping applications.

Uncertainty introduced by interpretation of the colour card arises from its discrete nature (16 levels) with a notional classification error of 0.50 pH units. Possible environmental sources of error in the colour card readings were noted, including misclassification due to variability in ambient lighting, colour vision deficiencies, quality of test kits, and temporal and sample variability issues. These are subjects requiring further research. The current dataset was based on a pooled population of observers and there is scope for future refinement to identify individual variations in performance between experienced and inexperienced operators.

The extensive statistical analysis of different models and their application with respect to the Victorian soil database, under a variety of conditions, provides detailed information for modellers and researchers in soil science, landscape studies, geography, environmental engineering, and hydrology.

### 5.2. Specific Comments

- (1) More than 3000 different mathematical models were fitted to the experimental data using an automated process covering a variety of functional types, for example, polynomial types, transition functions, exponential types, linear types, and mixed models. Indications of goodness of fit were provided by various statistical criteria, including coefficient of determination ( $r^2$ ), ANOVA  $F$ -statistic, and sum-of-squared errors (SSE). The best fits based on statistical metrics were found to be linear and sigmoidal models.
- (2) A factorial experiment ( $2 \times 2 \times 2$ ) was implemented for modelling the relationship between field pH and lab pH using (a) linear and sigmoidal models, (b) least-squares and Pearson performance indices for error minimisation, and (c) fixed and data randomisation regimes
- (3) A data randomisation scheme was introduced for the purpose of uncertainty analysis; that is, the class intervals in the pH data were treated as uniform priors in the Bayesian interpretation.
- (4) The error minimisation scheme used for the regression models was subject to comparison between standard least-squares and the Pearson criterion for treatment of scatterplots with significant outliers (which can skew results for standard regression analysis).
- (5) Models fitted for samples in water ( $\text{pH}_w$ ) were superior on statistical metrics to models fitted in  $\text{CaCl}_2$  ( $\text{pH}_c$ ) solution, for example, linear and sigmoidal models, using both standard least-squares and Pearson error minimisation approaches (cf. Figure 4(a) versus Figure 8(a) and also Figure 6(a) versus Figure 8(b)).

### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Acknowledgments

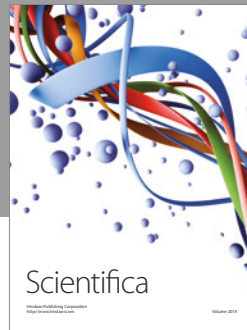
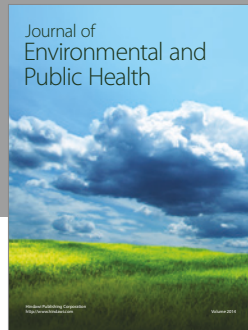
The authors thank Matt Kitching and Doug Crawford for many thoughtful discussions and insights on pH measurements. This paper was delivered with support of the Land Knowledge Foundations Project within DEDJTR.

### References

- [1] E. Farr, "An investigation on laboratory and field methods of determining the pH of soil suspensions," *Journal of the Science of Food and Agriculture*, vol. 23, no. 9, pp. 1089–1097, 1972.
- [2] W. J. Slattery, M. K. Conyers, and R. L. Aitken, "Soil pH, aluminium, manganese and lime requirement," in *Soil Analysis: An Interpretation Manual*, L. A. Sparrow, K. I. Peverill, and D. J. Reuter, Eds., pp. 103–128, CSIRO Publishing, Melbourne, Australia, 1999.
- [3] B. Minasny, A. B. Mcbratney, D. M. Brough, and D. Jacquier, "Models relating soil pH measurements in water and calcium chloride that incorporate electrolyte concentration," *European Journal of Soil Science*, vol. 62, no. 5, pp. 728–732, 2011.
- [4] M. Raupach and B. M. Tucker, "The field determination of soil reaction," *Journal of the Australian Institute of Agricultural Science*, vol. 25, no. 2, pp. 129–133, 1959.
- [5] R. E. White, "On the measurement of soil pH," *The Journal of the Australasian Institute of Agricultural Science*, pp. 3–14, 1969.
- [6] G. E. Rayment and D. J. Lyons, *Soil Chemical Methods: Australasia*, CSIRO Publishing, Melbourne, Australia, 2011.
- [7] A. McBratney, M. Mendonça Santos, and B. Minasny, "On digital soil mapping," *Geoderma*, vol. 117, no. 1–2, pp. 3–52, 2003.
- [8] B. Minasny and A. B. McBratney, "Digital soil mapping: a brief history and some lessons," *Geoderma*, vol. 264, pp. 301–311, 2016.
- [9] D. G. Rossiter, J. Liu, S. Carlisle, and A.-X. Zhu, "Can citizen science assist digital soil mapping?" *Geoderma*, vol. 259–260, pp. 71–80, 2015.
- [10] E. T. Wherry, "Soil acidity and a field method for its measurement," *Ecology*, vol. 1, no. 3, pp. 160–173, 1920.
- [11] D. D. Mason and S. S. Obenshain, "A comparison of methods for the determination of soil reaction," *Soil Science Society of America Journal*, vol. 3, pp. 129–137, 1939.
- [12] M. Raupach, "The errors involved in ph determination in soils," *Australian Journal of Agricultural Research*, vol. 5, no. 4, pp. 716–729, 1954.
- [13] G. C. Steinhardt and D. B. Mengel, "Comparison of ph measurement by field and laboratory methods," *Communications in Soil Science and Plant Analysis*, vol. 12, no. 1, pp. 71–78, 1981.
- [14] D. E. Baker, C. E. Rayment, and R. E. Reid, "Predictive relationships between ph and sodicity in soils of tropical queensland," *Communications in Soil Science and Plant Analysis*, vol. 14, no. 11, pp. 1063–1073, 1983.
- [15] T. Aven, *Front Matter*, John Wiley & Sons Ltd, 2003.
- [16] K. K. Benke, K. E. Lowell, and A. J. Hamilton, "Uncertainty analysis and risk assessment in the management of environmental resources," *Australasian Journal of Environmental Management*, vol. 14, no. 4, pp. 243–249, 2007.
- [17] N. J. Robinson, K. K. Benke, and S. Norng, "Identification and interpretation of sources of uncertainty in soils change in a global systems-based modelling process," *Soil Research*, vol. 53, no. 6, pp. 592–604, 2015.

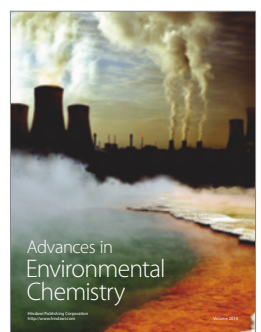
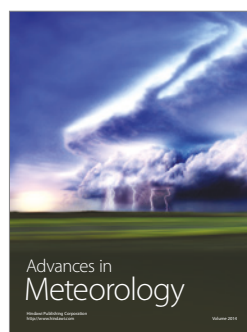
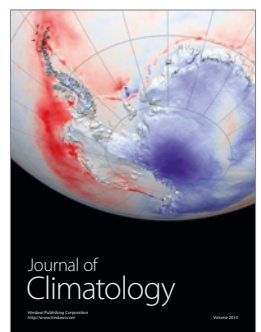
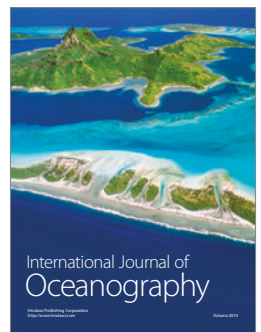
- [18] D. D'Or, P. Bogaert, and G. Christakos, "Application of the BME approach to soil texture mapping," *Stochastic Environmental Research and Risk Assessment*, vol. 15, no. 1, pp. 87–100, 2001.
- [19] G. E. Rayment and F. R. Higginson, *Australian Laboratory Handbook of Soil and Water Chemical Methods*, Inkata Press Pty Ltd, Melbourne, Australia, 1992.
- [20] M. J. M. Pelgrom, *Analog-to-Digital Conversion*, Springer, Dordrecht, Netherlands, 2010.
- [21] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer Science & Business Media, New York, NY, USA, 2013.
- [22] J. R. Taylor, *An Introduction to Error Analysis*, Univ. Science Books, Sausalito, Calif, USA, 1997.
- [23] M. Çetinkaya-Rundel, "Confidence and Prediction intervals and transformations," Statistics 101, Duke University, 2013, <https://stat.duke.edu/courses/Fall13/sta101/slides/unit7lec3H.pdf>.
- [24] B. L. Henderson and E. N. Bui, "An improved calibration curve between soil pH measured in water and CaCl<sub>2</sub>," *Australian Journal of Soil Research*, vol. 40, no. 8, pp. 1399–1405, 2002.
- [25] P. DeMarco, J. Pokorny, and V. C. Smith, "Full-spectrum cone sensitivity functions for X-chromosome-linked anomalous trichromats," *Journal of the Optical Society of America A*, vol. 9, no. 9, pp. 1465–1476, 1992.
- [26] K. K. Benke and K. E. Benke, "Uncertainty in health risks from artificial lighting due to disruption of circadian rhythm and melatonin secretion: a review," *Human and Ecological Risk Assessment*, vol. 19, no. 4, pp. 916–929, 2013.
- [27] E. Self, "Color vision deficiencies," 2014, <http://psych.fullerton.edu/eriko/research/ColorVision.html>.
- [28] M. K. Conyers, N. C. Uren, K. R. Helyar, G. J. Poile, and B. R. Cullis, "Temporal variation in soil acidity," *Australian Journal of Soil Research*, vol. 35, no. 5, pp. 1115–1129, 1997.





**Hindawi**

Submit your manuscripts at  
<https://www.hindawi.com>







**Minerva Access is the Institutional Repository of The University of Melbourne**

**Author/s:**

Benke, K; Robinson, NJ

**Title:**

Quantification of Uncertainty in Mathematical Models: The Statistical Relationship between Field and Laboratory pH Measurements

**Date:**

2017

**Citation:**

Benke, K. & Robinson, N. J. (2017). Quantification of Uncertainty in Mathematical Models: The Statistical Relationship between Field and Laboratory pH Measurements. Applied and Environmental Soil Science, 2017, <https://doi.org/10.1155/2017/5857139>.

**Persistent Link:**

<http://hdl.handle.net/11343/191807>

**File Description:**

Published version