environments

*Article*

# Epistemic Uncertainties in the Assessment of Regional Soil Acidification

**Kurt Benke [1,2], Nathan Robinson [3,*], Sorn Norng [2], David Rees [2] and Garry O'Leary [4,5]**

1 School of Engineering, University of Melbourne, Parkville, VIC 3010, Australia; kbenke@unimelb.edu.au
2 Centre for AgriBioscience, AgriBio, State Government of Victoria, Bundoora, VIC 3083, Australia; sorn.norng@agriculture.vic.gov.au (S.N.); david.rees@agriculture.vic.gov.au (D.R.)
3 Centre for eResearch and Digital Innovation, Federation University Australia, University Drive, Mount Helen, VIC 3350, Australia
4 Grains Innovation Park, Agriculture Victoria, 110 Natimuk Road, Horsham, VIC 3401, Australia; garry.oleary@agriculture.vic.gov.au
5 Centre for Agricultural Innovation, University of Melbourne, Parkville, VIC 3010, Australia
* Correspondence: n.robinson@federation.edu.au

**Abstract:** The increasing acidification of soil due to pollution and agricultural management practices is a growing problem worldwide, where food production is already under threat by climate change, more frequent droughts, and soil nutrient depletion. Soil acidification is quantified by pH measurements and is a primary metric for soil health. High soil acidity is a constraint on the production of grains and other crops because it decreases the bioavailability of important plant nutrients while increasing soil toxicity arising from an imbalance of essential soil elements. Field pH can be estimated by colour test kits which are very cost-effective and particularly suitable for developing countries where laboratory services are not available or fail to provide timely results. Because the pH test kit is based on visual colour matching between a colour card scale and a soil sample in solution, there are epistemic uncertainties, such as variability in expert opinion, differences in colour vision, measurement error, instrumentation, and changes in daylight spectral content. In this study, expert human observers were compared in experiments conducted using a standard pH test kit under a range of environmental conditions. A significant difference in uncertainty in colour discrimination was evident between male and female experts, whereas changes in daylight conditions had lower impact on the variance of pH estimates. In a group of subject matter experts, the male standard error (0.35 pH) was 57% higher on average over the range of pH values (pH = 4 → 10) compared to females (0.22 pH). This error was largest (70%) in the low pH 4 to 6.5 range, which is a critical range for successful amelioration of soil acidification. The results suggest that historical database measurements may have hitherto unrecognised uncertainties that affect confidence intervals for experimental data that in turn will have an impact on predictive models and policy development.

**Keywords:** agricultural practices; food production; sustainability; soil health; climate change

check for updates

**Citation:** Benke, K.; Robinson, N.; Norng, S.; Rees, D.; O'Leary, G. Epistemic Uncertainties in the Assessment of Regional Soil Acidification. *Environments* **2022**, *9*, 97. https://doi.org/10.3390/environments9080097

Academic Editors: Giannantonio Petruzzelli and Ioannis K. Kalavrouziotis

Received: 23 May 2022
Accepted: 28 July 2022
Published: 3 August 2022



## 1. Introduction

Increasing levels of soil acidity due to pollution and agricultural management practices has accelerated land degradation and is a threat to global food security, climate, and environmental sustainability [1]. Soil acidity is recognised as one of the key indicators of soil health and its control is required for sustainable management practices used by agricultural industries, such as horticulture and field crop production. National soil information systems are based on large datasets of soil observations that include measurements of soil acidity [2]. In Australia, online systems are delivered by state and territory governments, including the Victorian Soil Information System (VSIS) database [3], which supports mapping, modelling, data mining, and predictive analytics. The database supports big data analytics as part of the emerging culture of digital agriculture. Confidence in the data quality of soil assessments can be improved by uncertainty analysis and error quantification.

The field measurement of soil acidity is possible by inexpensive and readily available pH test kits, which have been used extensively for soil surveys and advisory services in many parts of the world. The pH test kits are particularly useful in the developing world, where laboratory services are not available or fail to provide timely results. There is, however, an issue with the accuracy and reliability of the results produced by the test kits. The dominant historical theme in uncertainty analysis has been statistical variability in experimental measurements and the impact on model calibration [4,5]. More recently, epistemic uncertainties have also been recognised as significant sources of error [6]. Early progress has been reported to unify the treatment of uncertainties into a single approach under the proposed generalised representation of uncertainty in the modelling process (GRUMP) [7,8].

Epistemic uncertainty is associated with the lack of information or system knowledge, as distinct from aleatory uncertainty (statistical variability), as described in recent studies [8,9]. Uncertainty can be divided into either aleatory or epistemic uncertainty, and the latter can be further subdivided into Type I and Type II epistemic uncertainty, as defined in [8]. Type I epistemic uncertainty relates to known unknowns, such as data entry errors and expert opinion, while Type II epistemic uncertainty relates to unknown unknowns [8]. Unknown unknowns are more serious in nature as they may challenge prior assumptions in either theory or modelling, but also introduce unexpected risks that can affect predictions or decision making, leading to 'black swan' events [8,10].

In pedology, colour is a property that is often used for identification and classification 'as an indicator of soil physical, chemical and biological properties as well as of the occurrence of soil processes' [11]. For example, a brown colour may indicate the presence of iron oxide, or a white colour may indicate the presence of calcites [12]. Colour can also be used in other contexts, such as using a pH field test kit for matching a colour card with soil colour in solution, or to discriminate between soil types in a colour map.

This study extended research on epistemic uncertainties relating to field measurements of soil acidity using pH colour test strips under daylight illumination [13]. Epistemic uncertainties are not normally considered in traditional error analysis because they are often unknown and difficult to quantify for statistical analysis and remain hidden within the soil information systems used for land-use planning. The analysis in this study involved assessment of soil acidity using field pH test kits with colour strips that are subject to visual colour discrimination by expert opinion under daylight illumination.

### *1.1. Test Kits for Field pH*

Field pH is determined using a colorimetric indicator method based on visual comparison of a pH colour chart against the colour of the soil in solution, as shown in Figure 1 [14,15]. The pH value provides information on nutrients, heavy metal availability and toxicity, liming requirements, and microbial activity [16,17]. Although the measurement of field pH using a portable pH test kit is less accurate than laboratory measurements, it is a fast and inexpensive method for prescreening, soil surveys, where and when expensive laboratory tests are required, and matching legacy data to laboratory measurements [18]. The kits have the potential for widespread use in the developing world.


**Figure 1.** Example of soil pH colour indicator card and sample pot for field pH determination.

In Australia, the field pH can be readily measured in situ using a soil pH test kit developed by the CSIRO (Commonwealth Scientific and Industrial Research Organisation) and manufactured by the Australian company Inoculo (Moorabbin, Australia) [15]. Aside from its use in agriculture, the Inoculo test kit is widely used for cost-effective regular monitoring and maintenance in nurseries, market gardens, in home gardening, by potting mix manufacturers, and by lawn suppliers. In the case of agriculture, the test kit enables fast and cost-effective prescreening of soil pH and provides information for data harmonisation if more reliable laboratory results are not available, or required, due to access time and expense.

### 1.2. Experimental Uncertainties

The soil pH test kit used for the assessment of soil acidity requires visual matching between the colour on the test card and the colour of the soil sample in solution, a procedure that is subject to ambient environmental conditions and expert opinion. For example, population studies have revealed that at least 10% of the male population is deficient in colour discrimination [19–21]. Furthermore, there is variability within and between male and female populations. Major sources of uncertainty that may affect interpretation of the pH test results include (a) colour vision differences, (b) colour card reflectance properties, (c) daylight spectral content, (d) time-of-day, (e) atmospheric scattering of light, and (f) random choice errors. For example, the difference between the colours burgundy and brown might be difficult to discern in some lighting conditions, while poor colour vision may result in a random choice error, such as between the colours red and green for males [19].

### 1.3. Null Hypothesis

The aim of the experiment was to determine whether soil acidity assessment using pH field test results based on visual colour discrimination is subject to epistemic uncertainties due to differences in colour perception and environmental effects. In particular, the statistical null hypothesis $H_0$ was that there was no significant difference in performance between soil scientists engaged in soil sampling, especially between male or female expert opinion. The results are relevant to (a) improving the accuracy of pH field assessments, (b) improving visual interpretation of colour-coded soil maps [22], and (c) improving visual classification of physical soil samples by their colour [11]. All three reasons are important to the assessment of soil health condition.

## 2. Sources of Variability in Colour Discrimination

### 2.1. Colour Vision Deficiencies

At least 10% of males are classified as colour blind or colour deficient so that there is a 1:10 random chance of an erroneous result in colour matching, which directly affects performance of the pH test strip [23]. Most colour deficiency relates to uncertainty and confusion in the red-green hues due to defects on the X chromosome in males and this can be detected by the Ishihara Colour Chart (also used for automotive, aviation, and marine licence testing). Reasons for defective colour vision may be congenital (but mostly stable over a lifetime), injury or disease related (for example photoreceptor damage in the case of welders), or lesions in the cerebral cortex [21]. In the case of farmers, vision problems may be even greater due to a lifetime of outdoor work without adequate eye protection.

Other reasons for defective colour vision include ageing eyes and retinal degeneration due to excessive exposure to UV radiation from artificial lighting indoors and direct sun exposure outdoors [21,24]. Retinal disease can also reduce the quality of colour perception in different lighting conditions. In particular, the side effects of diabetic retinopathy and age-related macular degeneration, which are caused by genetic and lifestyle factors, and have an incidence that increases with increasing age. In this respect, the incidence of defective colour vision may well greatly exceed 10% for ageing adult males.

### 2.2. Daylight Spectral Content

The colour of ambient illumination and the spectral reflectance of the pH colour card will affect the visual colour matching process in pH test kits. Measurements of the daylight spectral distribution in Australia have been taken in the past due to its importance to colour rendition and appearance in the case of paint, textiles, film, dyes, and fashion. It was reported in a survey of spectral distribution of daylight that 'photographers have often commented that transparencies photographed in Australia have observable differences in contrast and colour balance from transparencies photographed in the Northern Hemisphere . . . .', and this is attributed to the different spectral distribution of Australian daylight [25].

Australian measurements have shown essential agreement with Northern Hemisphere measurements but, with a higher level of irradiance in the ultraviolet spectral region side of the full radiator locus [25]. This effect has unknown impacts on colour card matching and pH assessment over the course of the day, or seasonally, with a possible bias to a higher pH reading on a standard colour card.

### 2.3. Atmospheric Light Scattering

A source of error that appears to have been neglected in the past is the impact of atmospheric light scattering on the spectral content of viewing conditions for the pH test card. The colour of an object in the field, such as a test strip used for pH estimation, depends on the colour of the ambient light and the spectral reflectance of the object. A change in the wavelength distribution of daylight illumination will affect the perceived colour of the pH test card. The transmittance of a beam of light through the atmosphere is attenuated according to the Beer–Lambert Law. At a specified wavelength, $\lambda$, the transmittance $T_\lambda$ along a straight line is given by

$$T_\lambda = \left[\frac{I_i}{I_0}\right]_\lambda = exp(-\sigma_{ext} c x) \text{ where } \sigma_{ext} = \sigma_{abs} + \sigma_{sca} \quad (1)$$

where $I_i$ is the transmitted intensity, $I_0$ is the incident intensity, $c$ is the concentration of particles, $x$ is the path length, $\sigma_{ext}$ is the extinction cross-section, $\sigma_{abs}$ is the absorption cross-section, and $\sigma_{sca}$ is the scattering cross-section. The light in the forward direction of the beam is either absorbed or scattered out of the beam before it is detected by the observer or sensor. Due to light scattering, the apparent colour of a pH test card varies under daylight illumination throughout the day as the sun moves through the sky (continuously changing the transmittance path length from the sun to the observer). Sunlight is scattered preferentially in the atmosphere by particles in the air that have diameters significantly smaller than the wavelength of visible light ($\lambda$~400–700 nm). This phenomenon is referred to as Rayleigh scattering and results in diffuse sky radiation with its characteristic blue colour [26,27]. In Rayleigh scattering, the cross-section for scattering, $\sigma_{sca}$, has the following relationship with refractive index, $n$, particle diameter, $d$, and wavelength, $\lambda$,

$$\sigma_{sca} = \frac{2\pi^5 d^6}{3\lambda^4}\left(\frac{n^2-1}{n^2+2}\right)^2 \tag{2}$$

The scattering cross-section, therefore, varies inversely as the fourth power of the wavelength of incident unpolarised light, i.e.,

$$\sigma_{sca} \approx \frac{1}{\lambda^4} \tag{3}$$

The strong inverse wavelength dependence on scattered intensity means that the blue wavelengths are scattered much more strongly out of the forward direction of the beam compared with the longer wavelengths (at the red end of the spectrum). Near sunset, the incident rays are nearly tangential to the atmosphere, and the path length travelled is so great that nearly all the blue wavelength radiation has been scattered out of the beam. The visible light observed then has the characteristic red colour of dusk.

Because of atmospheric light scattering, colour matching using a test kit is influenced by time-of-day, especially comparing midday to sunset. Note that, with a typical colour matching card [15], the increasing red appearance corresponds to increasing acidity, especially for pH < 3. This means that, as the day progresses, there will be an increasing red bias and therefore increasing error in the measurement of acidity (see Figure 1). A similar effect is also due to seasonal differences, comparing midsummer to midwinter daylight conditions. Moreover, there may be reflections from high-vis shirts (yellow or orange).

### *2.4. Reflection Density of Colour Cards*

Colour cards for pH estimation can be subject to variable performance due to the quality of liquid dye mixing and consistency, batch-to-batch differences, and colour reproduction quality of the printed card. The spectral characteristics of the incident and reflected radiant fluxes will affect the reflection density, DR, of the print. The measured reflection density as a function of wavelength is described as follows [28]:

$$D_R = \log_{10}\frac{\int S_\lambda O_\lambda P_{0\lambda} d\lambda}{\int S_\lambda O_\lambda P_{s\lambda} d\lambda} \tag{4}$$

where, at each wavelength, $\lambda$, $S_\lambda$ is the spectral sensitivity of the paper, $O_\lambda$ is the transmittance of the optical system without the sample, $P_{0\lambda}$ is the radiant flux incident on the sample, and $P_{s\lambda}$ is the radiant flux reflected by the sample, i.e., the pH colour card. Different batches of dye solution may produce different results, due to different purity, manufacturer, or date of production. The colour card should be replaced with each batch of dye solution used and both items should be supplied by the same manufacturer.

### *2.5. Colour Pigment Quality*

Colour test cards may be printed using either pigment-based inks or traditional dye-based inks (as used in inkjet printers). Pigment-based inks consist of encapsulated particles that are not absorbed into the substrate (paper), as occurs with dye-based inks. Pigment-based inks provide superior performance as they have much greater colour stability and archival properties. The cheaper dye-based inks, however, have a greater range in brightness and colour, which affects accuracy in determination of acidity. They are also much less prone to the problem of metamerism (colour shifts under different illumination).

### *2.6. Liquid Solvents for Soil Samples*

The quality of liquid solvents used for soil samples and the print quality of associated colour cards from different technologies may introduce inaccuracy and uncertainty in pH test kits. The reliability of indicator test kits may be affected if solvents have deteriorated with age, shelf-life, or heat and have impurities present. If so, there may be a small but detectable shift in colour imagery recorded on transparencies, i.e., emulsions or indicator solutions. A small shift towards the ultraviolet end of the spectrum suggests an experimental error that affects colour matching using the pH test kit and therefore pH estimates. This effect has not been quantified and would benefit from further research.

### *2.7. Random Errors in Colour Choice*

Random errors are possible in the case of choosing the colour on the pH test card. For example, if colour is represented as a random variable $\{X_1, X_2, \ldots, X_n\}$, the multivariate probability density function for random colour choice is given by the multinomial probability distribution:

$$P(X_1 = x_1, \ldots, X_n = x_n) = \frac{N!}{\prod_{i=1}^{n} x_i} \prod_{i=1}^{n} \theta_i^{x_i} \tag{5}$$

where $x_i$ are positive integers $Z* = \{0\} \cup Z^+$s.t. $\sum_{i=1}^{n} x_i = N$ and $\sum_{i=1}^{n} \theta_i = 1, \theta_i > 0$. where $N$ = number of trials, $n$ = number of possible outcomes, $x_i$ = is the number of occurrences of outcome $i$, and $\theta_i$ = is the probability of observing outcome $I$ [29,30]. More specifically, the joint distribution of $X_1$, $X_2$, ... , $X_n$ relates to mutually exclusive events that follow the multinomial probability distribution, which is an extension of the binomial distribution, so that the probability that $X_1$ occurs $x_1$ times, and $X_n$ occurs $x_n$ times, can be described as [29,30]:

$$P_N(x_1, x_2, \ldots, x_n) = \frac{N!}{x_1! \cdots x_n!} \theta_1^{x_1} \cdots \theta_n^{x_n} \tag{6}$$

The variance (uncertainty) in $X_i$ is given as follows:

$$\sigma_i^2 = N\theta_i(1 - \theta_i) \tag{7}$$

### *2.8. Expert Opinion and Experience*

A source of variability not accounted for in recorded results in the soil database is entirely subjective and due to judgement from expert opinion, which may be affected by biological sex and the level of experience. The effect due to the sex of the analyst is probably due to colour vision deficiencies associated with males during discrimination between certain colours.

### *2.9. Error Propagation through the System*

Error propagation through the system produces uncertainty in results that accumulates from many sources, such as those outlined previously. For example, assuming normally distributed errors, the total variance at the output of the system may be expressed as follows [31]:

$$\sigma_{total}^2 = \sum_{i=1}^{n} w_i \sigma_i^2 \tag{8}$$

where $i = 1, \ldots, n$ denotes sources of variance, $\sigma^2{}_i$ is representing uncertainty (see possible sources listed in Sections 2.1–2.8 above), and $w_i$ are weighting factors. In this study, examples of uncertainties affecting colour determination using test kits include sex, colour vision, observer variability, and light spectral composition.

## 3. Materials and Methods

Three experiments of increasing complexity were conducted to test for variability in expert opinion using the pH test kits. In the first experiment, the variability in colour discrimination was investigated for a mixed group of scientists from Agriculture Victoria Research (a state government agency). Colour classification was tested using the Ishihara Colour Test Chart [19]. The results for 20 male and female staff were subject to comparison and analysis. The subject was asked to read the hidden number (YES/NO response).

In the second experiment, a subgroup analysis was based on an experienced scientific cohort of 12 soil scientists for comparison between male and female staff by using the online Farnsworth–Munsell 100 Hue Colour Vision Test [20,32].

In the third and final experiment, visual assessment of field pH, by matching soil sample to a colour card using the pH test kit, was carried out in a controlled study with replications using male and female assessors engaged in soil research. A model was tested incorporating sex, light quality, experience, and their interactions. The pH test kit used as a reference was manufactured by Inoculo [15]. The experimental method followed standard procedure for psychophysical experiments recording perceptual judgements against a psychometric scale [33]. Thirteen soil samples with pH levels chosen to represent those in field pH test kits, e.g., 4.5, 5.0, 5.5, etc. were selected from 1800 soil samples submitted for laboratory analysis (1:5 soil-water dilution) at the Agriculture Victoria Research Macleod laboratory [13].

Psychophysical measurements were recorded in the form of colour discrimination using a pH colour card with 16-step scale for matching the colour card against the soil samples for field pH (using the test kit). Ten subjects in the analysis were males and females with 20/20 vision wearing their normal correction. Age range was 35 to 60 years. The experiment was a subanalysis of a factorial study with degrees of freedom (df) = 280 that consisted of pH (13 levels), sex (2 levels), experience (2 levels), and ambient light level (2 levels). The software package GENSTAT was used for Restricted Maximum Likelihood (REML) estimation for variance components analysis [34]. The primary predictors and interactions are expressed symbolically as follows:

$$y \sim x_1 + x_2 + x_3 + x_1x_2 + x_2x_4 + x_1x_3 + x_2x_3 + x_1x_2x_4 + x_1x_2x_3 + x_2x_3x_4 + \varepsilon \quad (9)$$

where $y$ = pH colour (from card), with predictors $x_1$ = pH (laboratory), $x_2$ = sex (male or female), $x_3$ = light quality, high (near midday, 1 pm) or low (near sunset, 5 pm), $x_4$ = experience level, and $\varepsilon$ = error term. Multiplicative terms represent interactions. A sparse algorithm with AI (average information) optimisation was used. Variables with missing data values were excluded.

The predictor $x_3$ is a proxy for the effect of daylight spectral distribution, atmospheric light scattering, and the reflective properties of the pH test chart. Experiments were conducted in the Southern Hemisphere in midautumn on a sunny day with clear blue skies.

## 4. Results

### *4.1. Experiment 1: Ishihara Test*

In the case of the Ishihara Colour Test, the 20 subjects correctly discriminated the Arabic numerals embedded in the test patterns (by a Yes/No Count Test), except for one male who could not discriminate a number in one pattern (i.e., 7% failure rate in a small sample). Although all females could discriminate the numbers from the background immediately, 10 of 14 males (71%) required some time, in the order of seconds, for at least one pattern (including six who needed more time for two or more patterns) whilst only one female of the six showed any difficulty. The results for this group of technical staff engaged in soil science suggested that females performed better on the Ishihara test for red-green colour discrimination. It is noteworthy that colour pH test cards and some maps have adjacent regions with red-green areas. The male interpretations of field pH would therefore be subject to greater uncertainty.

*4.2. Experiment 2: Munsell Test*

In the case of the Munsell test, the average score for the whole group was G = 34.3, where a lower score is superior (std dev = 20.6). The male score was M = 28.8 (std dev = 23.9) and the female score was F = 38.3 (std dev = 18.7). This difference was not significant statistically when using the t-test between means with unequal variances (at l.o.s. $\alpha = 0.01$). When the two best male scorers (who were highly experienced) were removed from the male sample, the male score was M = 41.3 (std dev = 23.4), which was a similar mean score to that for females. The standard deviation for males did not change. This suggests that the males and females were similar in performance on average for prediction accuracy but that males were subject to greater variability, that was 25% larger for the Munsell test (where the metric for uncertainty was the standard deviation). The difference in colour discrimination was more statistically significant between less experienced males and females.

The dispersion for male and female groups at each level of pH is shown in Figure 2 with error bars ($n = 12$, $n_F = 7$, $n_M = 5$, at each pH level). The fitted model in the scatter plot shows strong linear correlation with pH test kit colours for both males and females, indicating similar performance ($r^2 > 0.91$). The standard error (s.e.) at each pH level was computed for the mean of both male and female groups and is shown in Table 1.

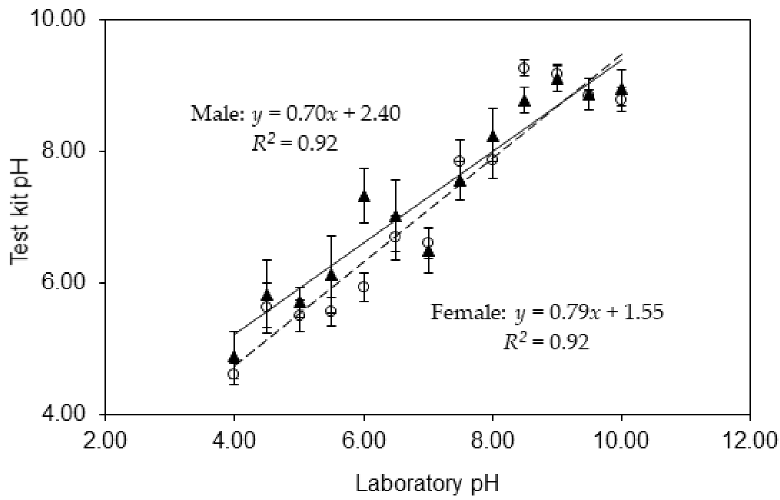


**Figure 2.** Estimated pH from test kit against laboratory derived pH for male (▲) and female (○) groups showing strong linear correlation but divergent relationship at low pH. Error bars show +/− standard error and 1:1 relationship shown by the dashed line.

**Table 1.** Mean and s.e. results for pH values (4.00 to 10.00) for female and male groups.

| pH | Female Mean | Female s.e.Mean | Male Mean | Male s.e.Mean |
|---|---|---|---|---|
| 4.00 | 4.61 | 0.17 | 4.89 | 0.37 |
| 4.50 | 5.62 | 0.38 | 5.82 | 0.51 |
| 5.00 | 5.50 | 0.24 | 5.70 | 0.23 |
| 5.50 | 5.56 | 0.21 | 6.13 | 0.58 |
| 6.00 | 5.93 | 0.22 | 7.32 | 0.41 |
| 6.50 | 6.68 | 0.33 | 7.02 | 0.54 |
| 7.00 | 6.60 | 0.24 | 6.49 | 0.33 |
| 7.50 | 7.85 | 0.32 | 7.55 | 0.30 |
| 8.00 | 7.87 | 0.28 | 8.25 | 0.41 |
| 8.50 | 9.26 | 0.12 | 8.78 | 0.19 |
| 9.00 | 9.18 | 0.14 | 9.11 | 0.19 |
| 9.50 | 8.85 | 0.08 | 8.87 | 0.25 |
| 10.00 | 8.79 | 0.18 | 8.96 | 0.27 |
| Average | 7.10 | 0.22 | 7.30 | 0.35 |

The results show male s.e. is higher over the range of pH values, i.e., male results are subject to greater uncertainty. For pH = 5, the standard error for the male group is about three times higher than for the female group.

Results show male standard error is 70% higher on average over the range of pH values (pH = 4 $\rightarrow$ 10), i.e., male results are subject to slightly higher uncertainty (the result is statistically significant using the t-test at l.o.s. $\alpha = 0.01$).

The higher standard error for males indicates that there is a greater probability of error in the male pH results. Reliability is inversely proportional to standard error, which means that males may be less reliable than females because of the greater error margin recorded in experiments. Overall, from a broad land management perspective, these indicative results show the pH field test performed quite well (see also Figure 2).

### *4.3. Experiment 3: Field pH Test Kit*

The effect of pH, sex, and lighting conditions on colour discrimination shows strong primary and interactive effects (Table 2). The experiment was a subanalysis of a factorial study design (see [13]), with df = 280 and factors pH (13 levels), sex (2 levels), experience (2 levels), light quality (2 levels).

**Table 2.** Experimental results for the model represented by Equation (9), showing significant effects due to sex and its interaction with experience level.

| Model Fixed Terms | F Statistic | *p*-Value |
|---|---|---|
| pH (laboratory) | 2.92 | <0.001 |
| Sex | 9.81 | 0.002 |
| Light Quality | 1.03 | 0.311 |
| pH × Sex | 1.11 | 0.353 |
| Sex × Experience | 10.09 | <0.001 |
| pH × Light Quality | 1.11 | 0.349 |
| Sex × Light Quality | 2.24 | 0.136 |
| pH × Sex × Experience | 1.35 | 0.129 |
| pH × Sex × Light Quality | 0.86 | 0.587 |
| Sex × Experience × Light Quality | 2.67 | 0.071 |

In Table 2, the most significant effects were sex, and sex interaction with experience. The effect of light quality was not significant. The three-way interaction of sex, experience, and light quality was marginally significant.

## 5. Discussion

The pH test kits widely used in the field assessment of soil acidification provide inputs for landscape and environmental mapping, modelling, and soil health analysis. The experimental results revealed that uncertainties relating to human colour perception and environmental effects introduce errors in pH results using the test kits. The results of three different colour vision tests suggest that there exists significant variability between assessors in a cohort of experts, and within the male and female subgroups. The experiments revealed that males have a larger error at most pH levels and represent a significant source of difference in results (due to the minority with colour vision deficits). In this study, the statistical null hypothesis $H_0$ was rejected that there is no significant difference in performance between the soil scientists engaged in soil pH assessments. Factors contributing to pH test errors include (a) vision differences, (b) print card reflectance properties, (c) daylight spectral content, (d) atmospheric scattering of light, and (e) random choice errors. The greatest source of difference in results is due to differences in colour vision between males and females.

Historical records in a soil database may, therefore, have uncertainties in measurements from a variety of sources. This has implications for using field pH assessments in modelling in soil science and hydrology, as well as for mapping and monitoring changes in soil acidity.

Further improvements in accuracy and precision in data collection could be targeted at decreasing uncertainty in field pH measurements between pH = 5 and 6, where the differences between males and females are greatest, and where soil pH is important for crops and cultivars that are sensitive to acidic conditions.

Future research could improve on this pilot study by increasing the sample size, using more experts, and greater sampling in the range pH = 5 to 6 to quantify the effect of colour vision differences on the accuracy and dispersion. The issues involved in colour interpretation may be summarised as follows, (a) population variability and sex differences in colour vision appear to be sources of uncertainty in pH assessments, (b) when using expert opinion, female scores in colour discrimination have lower statistical variance and are therefore more reliable for pH assessments, and (c) potential environmental sources of colour variability include light intensity level, time-of-day, atmospheric scattering causing colour shifts in illumination, and print quality of test materials. A significant result is that males using pH test kits on a routine basis are advised to have an eye test to check for colour vision deficiencies.

## 6. Conclusions

Soil acidity is recognised as one of the key indicators of soil condition and its measurement is necessary to support sustainable management practices in the environment and agriculture. Increasing levels of soil acidification have accelerated the decline in food production and represent a threat to global food security, while adding to the adverse impacts of climate change. The monitoring of soil acidification by soil pH has high priority because soil acidity is regarded as the single most important indicator of soil health.

The soil pH is often measured in the field using a soil sample in solution that is compared visually with a colour test card, with different colours corresponding to different pH levels. The field pH test kit enables broad interpretations for many land management decisions, including identification of acid soils and alkaline soils. The test also identifies areas for sample collection where higher resolution laboratory pH values are needed.

Potential sources of uncertainty in field pH estimates include variability in expert opinion, male and female colour vision differences, and variability in daylight spectral content. The results suggest the effect of light quality was not very significant.

The pH measurements of several soil samples were recorded by experienced field scientists using a standard field pH test kit as a reference. The effect modifiers were investigated, including light quality (intensity and time-of-day), sex (for colour vision differences), and experience level. The colour classification performance of both male and female assessors was quantified in three psychophysical experiments using (a) the Ishihara colour scale, (b) the Munsell colour scale, and (c) an experimental design using a reduced maximum likelihood method with interaction terms. There was significant variability in performance within and between males and females based on standard deviation and standard error in test scores.

The large errors in the male score for colour discrimination showed that there is increased uncertainty and reduced reliability in field pH classification by males. The results also suggest that there may be greater uncertainties in historical data corresponding to regional pH assessments than previously assumed because it was not recorded whether the analyst was male or female. It is also noted that the spectral content of light due to atmospheric light scattering may affect pH field measurements. Compensation for these effect modifiers is important because pH field tests in agriculture and in the developing world are likely to be based on pH colour test kits for the foreseeable future. It is recommended that males be tested for colour vision deficiencies if they are involved in routine use of pH test kits used in field measurements.

**Author Contributions:** Conceptualization, K.B., N.R. and D.R.; methodology, K.B., N.R. and S.N.; formal analysis, K.B., G.O., S.N., N.R. and D.R.; writing—original draft preparation, K.B., N.R. and G.O.; writing—review and editing, K.B., G.O. and N.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data used in this paper is from the Victorian Soil Information System (VSIS), Victorian Department of Jobs, Precincts and Regions.

**Acknowledgments:** We would like to thank Kerry Stott, Ivanah Oliver, Matt Kitching, Mark Imhof, Lucy Finger, Christie Ho, and Michele Davey for participating in this research, and the Land Knowledge Foundations Project.

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

## References

1. ITPS. *Status of the World's Soil Resources*; Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils: Rome, Italy, 2015.
2. Keller, A. National Soil Information Systems to Support Spatial Planning Procedures. *disP-Plan. Rev.* **2013**, *49*, 68–71. [CrossRef]
3. Hunter, D.; Williams, S.; Robinson, N. VSIS an new system for Victorian soil data. In Proceedings of the 19th World Congress of Soil Science, Brisbane, Australia, 1–6 August 2010; Gilkes, R., Ed.; Australian Society of Soil Science Inc.: Brisbane, Australia, 2010.
4. McBratney, A.B. On variation, uncertainty and informatics in environmental soil management. *Aust. J. Soil Res.* **1992**, *30*, 913–935. [CrossRef]
5. Minasny, B.; McBratney, A.B.; Mendonça-Santos, M.L.; Odeh, I.O.A.; Guyon, B. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. *Soil Res.* **2006**, *44*, 233–244. [CrossRef]
6. Refsgaard, J.C.; van der Sluijs, J.P.; Højberg, A.L.; Vanrolleghem, P.A. Uncertainty in the environmental modelling process—A framework and guidance. *Environ. Model. Softw.* **2007**, *22*, 1543–1556. [CrossRef]
7. Robinson, N.J.; Benke, K.; Hopley, J.; MacEwan, R.J.; Clark, R.; Rees, D.B.; Kitching, M.; Imhof, M.P.; Bardos, D. Multi-source data integration and identification of uncertainties affecting production of a digital soil map. In *GlobalSoilMap: Basis of the Global Spatial Soil Information System*; Arrouays, D., McKenzie, N., Hempel, J., deForges, A.C.R., McBratney, A., Eds.; CRC Press: London, UK, 2014; pp. 353–358.
8. Robinson, N.; Benke, K.; Norng, S. Identification and interpretation of sources of uncertainty in soils change in a global systems-based modelling process. *Soil Res.* **2015**, *53*, 592–604. [CrossRef]
9. Arslan, J.; Benke, K.K. Progression of Geographic Atrophy: Epistemic Uncertainties Affecting Mathematical Models and Machine Learning. *Transl. Vis. Sci. Technol.* **2021**, *10*, 3. [CrossRef] [PubMed]
10. Seely, H. *Pieces of Intelligence: The Existential Poetry of Donald H. Rumsfeld*; Simon and Schuster: New York, NY, USA, 2005.
11. Viscarra Rossel, R.; Minasny, B.; Roudier, P.; McBratney, A.B. Colour space models for soil science. *Geoderma* **2006**, *133*, 320–337. [CrossRef]
12. Brady, N.C.; Weil, R.R. *Elements of the Nature and Properties of Soils*; Pearson Prentice Hall: Hoboken, NJ, USA, 2010.
13. Robinson, N.; Norng, S.; Rees, D.B.; Benke, K.K.; Davey, M. Assessment of Error Sources in Measurements of Field pH: Effect of Operator Experience, Test Kit Differences, and Time-Of-Day. *Commun. Soil Sci. Plant Anal.* **2018**, *49*, 269–285. [CrossRef]
14. Raupach, M.; Tucker, B.M. The field determination of soil reaction. *J. Aust. Inst. Agric. Sci.* **1959**, *25*, 129–133.
15. Inoculo. *Inoculo Laboratories*; Inoculo: Moorabbin, VIC, Australia, 2021; Available online: http://www.inoculo.com.au/ (accessed on 18 October 2021).
16. Slattery, W.J.; Conyers, M.K.; Aitken, R.L. Soil pH, aluminium, manganese and lime requirement. In *Soil Analysis: An Interpretation Manual*; Sparrow, L.A., Peverill, K.I., Reuter, D.J., Eds.; CSIRO Publishing: Melbourne, Australia, 1999; pp. 103–128.
17. Minasny, B.; McBratney, A.B.; Brough, D.M.; Jacquier, D. Models relating soil pH measurements in water and calcium chloride that incorporate electrolyte concentration. *Eur. J. Soil Sci.* **2011**, *62*, 728–732. [CrossRef]
18. Farr, E. An investigation on laboratory and field methods of determining the pH of soil suspensions. *J. Sci. Food Agric.* **1972**, *23*, 1089–1097. [CrossRef]
19. Ishihara, S. *Ishihara's Test Chart for Colour Deficiency*; Kanehara Trading Inc.: Tokyo, Japan, 2004.
20. Paramei, G.V. Color discrimination across four life decades assessed by the Cambridge Colour Test. *J. Opt. Soc. Am. A* **2012**, *29*, A290–A297. [CrossRef] [PubMed]
21. Self, E. Color Vision Deficiencies. 2014. Available online: http://psych.fullerton.edu/eriko/research/ColorVision.html (accessed on 20 June 2014).
22. Isbell, R.F. *The Australian Soil Classification*; Revised Edition; CSIRO Publishing: Collingwood, Melbourne, Australia, 2002.
23. Pettijohn, P.F. *Psychology: A ConnecText*; Dushkin Publishing Group, Inc.: Guilford, CT, USA, 1998.
24. DeMarco, P.; Pokorny, J.; Smith, V.C. Full-spectrum cone sensitivity functions for X-chromosome-linked anomalous trichromats. *J. Opt. Soc. Am. A* **1992**, *9*, 1465–1476. [CrossRef] [PubMed]
25. Dixon, E.R. Spectral distribution of Australian daylight. *J. Opt. Soc. Am.* **1978**, *68*, 437–450. [CrossRef]
26. Morgan, J. *Introduction to Geometrical and Physical Optics*; McGraw-Hill: New York, NY, USA, 1953.

27. Weidner, R.T.; Sells, R.L. *Elementary Classical Physics*; Allyn and Bacon: Boston, MA, USA, 1981; Volume 2.
28. James, T.H. *Theory of the Photographic Process*; Macmillan: New York, NY, USA, 1977.
29. Papoulis, A. *Probability, Random Variables, and Stochastic Processes*, 2nd ed.; McGraw-Hill: New York, NY, USA, 1984.
30. Weisstein, E.W. *Multinomial Distribution*. From MathWorld—A Wolfram Web Resource. Available online: http://mathworld.wolfram.com/MultinomialDistribution.html (accessed on 4 December 2015).
31. Benke, K.K.; Norng, S.; Robinson, N.J.; Benke, L.R.; Peterson, T.J. Error propagation in computer models: Analytic approaches, advantages, disadvantages and constraints. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 2971–2985. [CrossRef]
32. Farnsworth, D. The Farnsworth-Munsell 100-Hue and Dichotomous Tests for Color Vision. *J. Opt. Soc. Am.* **1943**, *33*, 568–574. [CrossRef]
33. Benke, K.K.; Skinner, D.R.; Woodruff, C.J. Convolution operators as a basis for objective correlates of texture perception. *IEEE Trans. Syst. Man Cybern.* **1988**, *18*, 158–163. [CrossRef]
34. Ripley, B.D.; Thompson, M. Regression techniques for the detection of analytical bias. *Analyst* **1987**, *112*, 377–383. [CrossRef]