

## Article

# Formal Assessment of Agreement and Similarity between an Open-Source and a Reference Industrial Device with an Application to a Low-Cost pH Logger

Evmorfia P. Bataka <sup>1</sup>, Persefoni Maletsika <sup>2</sup> and Christos T. Nakas <sup>1,3,\*</sup>

<sup>1</sup> Laboratory of Biometry, Department of Agriculture, Crop Production and Rural Environment, University of Thessaly, 384 46 Volos, Greece; bataka@uth.gr

<sup>2</sup> Laboratory of Pomology, Department of Agriculture, Crop Production and Rural Environment, University of Thessaly, 384 46 Volos, Greece; pmalets@uth.gr

<sup>3</sup> University Institute of Clinical Chemistry, Inselspital—Bern University Hospital, University of Bern, 3010 Bern, Switzerland

\* Correspondence: cnakas@uth.gr

**Abstract:** Open-source devices are nowadays used in a vast number of research fields like medicine, education, agriculture, and sports, among others. In this work, an open-source, portable, low-cost pH logger, appropriate for in situ measurements, was designed and developed to assist in experiments on agricultural produce manufacturing. The device was calibrated manually using pH buffers for values of 4.01 and 7.01. Then, it was tested by manually measuring the pH from the juice of citrus fruits. A waterproof temperature sensor was added to the device for temperature compensation when measuring the pH. A formal method comparison process between the open-source device and a Hanna HI9024 Waterproof pH Meter was designed to assess their agreement. We derived indices of agreement and graphical assessment tools using mixed-effects models. The advantages and disadvantages of interpreting agreement through the proposed procedure are discussed. In our illustration, the indices reported mediocre agreement and the subsequent similarity analysis revealed a fixed bias of 0.22 pH units. After recalibration, agreement between the devices improved to excellent levels. The process can be followed in general to avoid misleading or over-simplistic results of studies reporting solely correlation coefficients for formal comparison purposes.

**Keywords:** open-source logger; open-source software; agreement; similarity; pH



**Citation:** Bataka, E.P.; Maletsika, P.; Nakas, C.T. Formal Assessment of Agreement and Similarity between an Open-Source and a Reference Industrial Device with an Application to a Low-Cost pH Logger. *Sensors* **2024**, *24*, 490. <https://doi.org/10.3390/s24020490>

Academic Editors: Takaharu Kameoka, Kazuki Kobayashi, Wei Guo and Kotaro Takayama

Received: 14 December 2023

Revised: 5 January 2024

Accepted: 11 January 2024

Published: 12 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Open-source devices are becoming very popular and even essential in an increased number of fields, such as education [1], agriculture [2,3] medicine [4] and biology [5], among others. The Maker Movement [6] that unfolded after the resurgence of the participatory Web 2.0 [7], the interfusion of open source, the decreased cost of electronic parts and other social influences are a few of the examples that contributed to the flourishing of prototype development. Furthermore, this phenomenon was boosted after the launch of development boards like Arduino [8,9] and Raspberry-pi [5] that tend to simplify intricate electronic assemblies by using basic software and programming. Openly accessible tutorials simplify the technical parts, provide visual aids for the wiring, issue the code in each case, and allow users with basic or no experience in electronics and coding to replicate or customize projects according to their needs. Open-source software and hardware solutions can be used in research and industry. Low acquisition cost and easy customization are two of the most important advantages of using such devices compared to industrial ones.

Oellermann et al. [10] highlighted three points where open-source electronics aided the scientific community. First, open-source devices help individual researchers by increasing the customization, the efficiency and scalability of the experiments, while increasing

data quantity and improving data quality. Second, they assist institutes since the open access to customizable high-end technologies increased the interdisciplinary collaborative network potential. Third, they succor the scientific community by improving transparency and reproducibility. Also, they help detach research capacity from funding and escalate innovation. Most of the labs worldwide do not have access to vital funding to keep up with the state-of-the-art lab equipment. Open-source devices contribute to the rapid supply of equipment for the labs with a low cost and high level of customizability.

Quality assessment of an in-house built open-source device must be validated by comparing measurements to these of a reference device such as validated industrial equipment of a reference standard device, in terms of measurement agreement [11]. The comparison is not limited to open-source devices compared to industrial devices but extends to the comparison of methods in general. “Agreement” measures the “closeness” between readings. Thus, agreement is a comprehensive term that contains both accuracy and precision. Typically, one of the devices/methods is treated as the reference, then agreement concerns a method or measurement comparison study (MCS) of the tested device versus the reference one.

This article presents the development of an open-source device that measures the pH of citrus fruit juice and describes the analytical procedure for a method comparison study between the open-source device and its corresponding industrial. The acidity of fleshy fruit, as measured by titratable acidity (TA) and/or pH, is an important component of fruit organoleptic quality. Fruit acidity is associated with the presence of organic acids, with malic and citric acids as the most abundant in most ripe fruits [12]. There is an interrelated relationship between pH and TA. Titratable acidity is determined by neutralizing the acid present in a known quantity of food sample using a standard base, while the endpoint for titration is usually a target pH (or the color change in a pH-sensitive dye). In addition, the TA of fruits is used, along with the total soluble solid (TSS) content (sweetness), as a maturity index (TSS/TA) [13].

Citrus is one of the most important commercial fruit crops in the world that includes important crops such as oranges, mandarins, lemons, grapefruits and others [14]. Fruit weight, size, shape, external color, TSS, TA and TSS/TA, juice content, chemical and nutritional composition are important quality traits for fresh citrus consumption and marketability [14,15]. In citrus fruit, the content of sugars and organic acids varies according to the species, varieties, environmental conditions and horticultural practices [15]. Moreover, sugars and organic acids in the fruit contribute to the perceived flavor, while aroma depends on many volatile organic compounds determining consumer preference [16]. The TSS/TA ratio has been used worldwide as the main commercial maturity index of citrus fruit internal quality. A TSS/TA ratio of at least 6 or higher is acceptable for commercial marketability; however, important differences may exist depending on the citrus species and varieties, as well as also on the growing regions. In particular, ratios acceptable for marketability usually range from 7–9:1 for oranges and mandarins to 5–7:1 for grapefruits [14]. Thus, TA and pH measurements of citrus fruit juice are of high importance for the evaluation of fruit organoleptic quality and maturity.

The benefits of using an open-source device interchangeably with a corresponding industrial device are mostly based on the lower cost and configurability of the former. Thus, a method comparison study between the constructed open-source device and a reference industrial one was designed and their agreement and similarity were formally assessed. Measurement ranges where the difference between the two devices are acceptable are discussed. When using recalibration methods, the agreement increases.

The paper is organized as follows. The design of the device and its key components are introduced in Section 2.1. Section 2.2 presents the reference device. Section 2.3 describes the five steps to implement a formal statistical method comparison study. The application is discussed in Section 3 and follows the format of Section 2. We end with a discussion and a conclusion. Methodological and further details are given in the Appendices A and B.

## 2. Materials and Methods

### 2.1. Design of the Device

#### 2.1.1. Hardware

The open-source logger is equipped with two sensors: a pH sensor [17] from Seeed studio and a temperature sensor DS18B20 [18]. A 16-bit analog to digital converter is added to the design to improve the precision of the voltage reading since the output of the sensor is analog. The development board for this device is the Adafruit feather proto 32u4. An Adafruit Featherwing logger [19] was added for instant capture of the measurement in a microSD card embedded with a timestamp by pressing a push button. A Nokia 5110 LCD monitor [20] was added to display the values of voltage, pH, temperature, and battery. A 1200 mAH LiPo battery is the main power source of the device, which can be charged via micro-USB to USB-A while the device is operating.

#### The pH sensor

SEN0169 (Figure 1) is an analog pH meter, specifically designed for Arduino and Arduino-compatible microcontrollers. The electrode is considered industrial. The sensor has a long life (>0.5 years in long-term monitoring mode), is highly accurate ( $\pm 0.1$  pH at 25 °C), it has fast response ( $\leq 1$  min), has a measuring range from 0 to 14 pH and includes a gain adjustment potentiometer (Appendix A, Table A1). The output voltage of the electrode is linear (Appendix A, Table A5) and is capable of long-term monitoring. The sensor is built for industrial use and is equipped with a BNC connector and PH2.0 sensor interface. Appendix A (Table A1) summarizes the technical specifications of the probe. The communication between the sensor and the Microcontroller Unit (MCU) is one-way since the sensor transmits data using an analog MCU pin. Since the 32u4 MCU uses a 10-bit analogue to digital converter (ADC), an ADC1115 16bit ADC and gain amplifier is added to increase the sensor's precision.



**Figure 1.** DFRobot PH meter (SEN0169) (source: DFRobot official website).

#### The temperature sensor

DS18B20 is a waterproof digital temperature sensor designed for Arduino or Arduino-compatible microcontrollers. According to the manufacturer, since the sensor's signal is digital, no signal degradation is present even if the distances between the MCU and the sensor are very long. The sensor provides 9-to-12-bit resolution temperature readings (configurable via software). The communication protocol between the MCU and the sensor is 1-Wire. Multiple DS18B20 sensors can connect on the same 1-Wire bus since they are produced with a unique silicon serial number. Table A2 (Appendix A) summarizes the technical specifications of the sensor.

### ADS1115 16bit ADC with gain amplifier

This module is a precision module (ADC) with 16 bits of resolution. The first 15 bits are used for the value and the last bit is used for the sign of the value. It is equipped with a voltage reference and an oscillator. It uses the I2C communication protocol to interact with the MCU. Four different slave addresses can be selected allowing four different ADS1115 [21] modules to be connected in the same bus. Its operating voltage ranges from 2 to 5.5 Volts. Furthermore, it can converge signals at rates up to 860 samples per second. Its second functionality includes a programmable gain amplifier that provides input ranges from inputs to as low as  $\pm 256$  mV with increments of 0.0078125 mV, thus measuring both small and large signals with high resolution. Moreover, it offers an input multiplexer, which provides two differential or four single-ended inputs. Last, the module operates in continuous conversion mode or a single-shot mode. This means that it automatically powers down in single-shot mode, reducing the power consumption during the measuring periods. To avoid damaging the module, the gain should be set more than or equal to the input voltage of the channel.

#### 2.1.2. Software

DFRobot provides a library for the SEN0169 via GitHub [22]. The library includes a calibration mode. However, the calibration was performed manually due to the MCU's incompatibility. Furthermore, the code was developed without using the library.

The code functionality is described as follows. First, the MCU reads the signal of the pH sensor via the ADS1115 in continuous mode using a single input channel. Second, in case an instantaneous measurement needs to be taken and stored, the user will press the push button and the measurement embedded with a time stamp will be stored in the microSD card. The function *button()* and *store()* provide these functionalities. After the calibration procedure, the equation is stored in the sketch and the *measure()* function returns the pH measurement after inserting the input voltage. The function *measure()* returns the proper calibration line, depending on the temperature of the liquid. The sketch is available on GitHub [23].

#### 2.1.3. Calibration Method

The calibration procedure was performed using two pH buffers. Eight measurements were taken. The first two were taken from 4.01 and 7.01 pH buffers when the liquid's temperature was 7.5 °C. The same procedure followed for temperatures of 12.5 °C, 17.5 °C and 22.5 °C. The probe was removed from the solution 1 min after its insertion to reach the response time according to the sensor's datasheet. The temperature of the buffer solution was measured using the DS18B20 temperature sensor. Table A3 summarizes the voltage and their corresponding pH values. Each temperature interval uses the calibration equation of the corresponding midpoint temperatures. For example, the first Equation (1) will be used for the range between 5 °C and 10 °C. Figure A1 presents the four calibration lines per temperature range.

$$y_{7.5^{\circ}\text{C}} = -6.27615x + 16 \quad (1)$$

$$y_{12.5^{\circ}\text{C}} = -6.1349695x + 16.2576$$

$$y_{17.5^{\circ}\text{C}} = -6.0241x + 16.09036$$

$$y_{22.5^{\circ}\text{C}} = -6.04351x + 16.1196615$$

Variable  $y$  represents the pH measurements and variable  $x$  represents the output of the sensor (in Volts). The subscript in variable  $y$  represents the temperature of the pH buffer. The equations were added to the Arduino sketch in the `measure()` function. The pH was automatically calculated depending on the temperature measurement.

#### 2.1.4. Cost of the Device

Table A4 summarizes the cost per component and the total cost. The cost can be reduced if parts like the development board and the ADS1115 can be replaced by cheaper equivalents from other brands. Furthermore, the final product does not usually use development boards, removing the inessential parts. Thus, the cost and the device's footprint are reduced, especially when the PCB is designed and printed with Surface Mounted Discrete (SMD) electronic parts replacing the through hole equivalents.

#### 2.2. The Reference Device

The reference device is a Hanna Instruments HI9024 Waterproof pH Meter [24] (Figure 2b). It is a heavy-duty pH meter designed for laboratory use and its accuracy is sustained even under harsh industrial conditions. It can easily be calibrated and has three memorized buffer values (4.01, 7.01 and 10.01). The device has automatic buffer recognition, thus avoiding errors during the calibration procedure. Moreover, it is equipped with a temperature compensation function. The temperature can be measured using a temperature sensor probe or can be entered manually. Since there was no temperature sensor available, the temperature was set manually using the DS18B20 sensor, which was embedded in the open-source logger. Thanks to its waterproof cylindrical case, the temperature sensor was inserted in the solution that was intended to be measured during the experiment. The specific pH meter model is not available in the market since it is considered obsolete. An equivalent but contemporary model is HANNA HI 99171. Its late 2023 cost in local vendors is around EUR 585 including shipping costs.



**Figure 2.** (a) The hand squeezing procedure. (b) Hanna pH meter. (c) DFRobot pH meter probe.

#### 2.3. Designing a Method Comparison Study

To evaluate the open-source device validity, its measurements need to be formally compared with a reference. In other words, a method comparison study needs to be designed to assess the novel device's agreement with the reference device. Five steps can be defined for such studies:



1. Establishment of the experimental design.
2. Exploratory analysis.
3. Assessment of the agreement and similarity between the two devices.
4. Identification of possible sources of disagreement using similarity and repeatability assessment for each device.
5. Recalibration of the novel device to improve the agreement.

### 2.3.1. Experimental Design

Proper experimental design is of utmost importance for valid results and adequate reproducibility. Repeated (towards intra-variability estimation) and replicate (towards inter-variability estimation) measurements are both multiple response measurements taken at the same combination of factor settings. However, repeated measurements are taken during the same experimental run or consecutive runs, while replicate measurements are taken during identical-conditioned but different experimental runs, which are often randomized. Their differences affect the structure of the dataset and the statistical analysis applied to process the data. In many situations, researchers mistakenly take for granted the sample's independence even though they sample from the same subject. This occurs when the experimental unit is not defined properly and instead of replicates, the researchers provide repeated measurements (pseudo-replications).

There are two possible categories of repeated measurements that the present experiment's data fall into: unlinked and linked data. Following Carstensen et al. [25], unlinked data refer to repeated measurements that are not paired in the sense that the measurements of the two methods are obtained separately. Thus, unlinked data are not necessarily measured concurrently. There is no need for the methods to have the same number of repeated measurements. However, linked data, in which each subject may experience consecutive measurements over time, are paired. Unlike the unlinked data, the devices/methods need to have an equal number of paired repeated measurements per one subject but may vary between different subjects. The true value does not need to stay constant over time but there is no systematic effect of time on the paired trajectories beyond the dependence induced in them by the common measurement time.

A well-designed experiment must include a proper definition of the experimental design, the type and number of repeated measurements, the sample size calculation/consideration, and a list of possible covariates. The described methods include covariate information handling.

### 2.3.2. Exploratory Analysis

A Bland–Altman plot [26] is typically used to assess the data for heteroscedasticity, dependency of the difference from the measurement range, outliers and a linear trend that indicates a correlation between differences and averages. Moreover, a scatterplot can be used as a supplementary plot to investigate the relationship between the two methods. Furthermore, a trellis plot is useful for the visualization of the spread of the repeated values and possible biases of the two methods. A trellis plot [11] is constructed by using the  $x$ -axis as the measurement range and the  $y$ -axis as the subjects' id. The two methods are differentiated by using two different symbols for each measurement per subject. Interaction plots between subjects and methods (devices), and subjects and time are useful for the researcher to graphically assess the category of repeated measurement. In case there is significant subject  $\times$  method interaction, an extra term should be added during the modeling process. In case there is significant subject  $\times$  time interaction, there is a possibility that the data are linked. This can be verified formally using criteria such as AIC, BIC and log-likelihood to assess the model quality.

### 2.3.3. Statistical Tools to Assess Agreement and Similarity

Mixed-effects and measurement-error models can be fit to the data and their estimated coefficients and variance components are used to produce agreement and similarity indices. These methods go beyond the assessment via standard correlation coefficients given the capacity of handling repeated measures and covariate information. Furthermore, correlation does not imply agreement, which is the cornerstone in method comparison studies [11]. Mixed-effects models are a special case of measurement-error models; specifically if there is evidence in the exploratory analysis that the proportional bias significantly deviates from 1, measurement-error models must be used instead of their mixed-effects counterparts.

The extended Bland–Altman plot can be used during the exploratory analysis step to assess this assumption. If there is a linear trend, then there is evidence of violation of the equal proportional bias assumption of the mixed-effects model. However, this trend might be due to different precisions of the two methods. In any case, the extended Bland–Altman plot can be evaluated using the *bland\_altman\_plot()* function from “*MethodCompare*” [27] package.

The methodology to fit mixed-effects models to the data is described in Appendix B.1.1, which also covers cases when the data are heteroscedastic and when covariates are added. All the steps to prepare the data and implement the models along with their diagnostics are available in an in-house-built R-script [23], which is based on [28].

The methodology to fit measurement-error models to the data is described in Appendix B.1.2, which also covers cases when the data are heteroscedastic but does not include covariates. The R-package “*MethodCompare*” [27] can be used to implement the relevant methodology [29–31]. The data must be in wide format. The output includes a list with the estimated bias (differential (fixed) and proportional) including 95% confidence intervals. Moreover, a list of models along with various variables needed for the estimation is returned.

### Indices and methods to assess agreement and similarity

Indices can quantify the agreement and similarity between two or more devices. There are two categories of indices: the absolute (or unscaled) and the relative (or scaled) indices [32]. A detailed review about agreement indices can be found in [33].

Absolute indices report measures according to the magnitudes of the actual data. They are unscaled and independent of between-sample variation.

The total deviation index (TDI) is used here for the evaluation of the agreement and similarity between two method/devices (inter-agreement). Specifically, TDI is an index that captures a predefined proportion ( $p$ ) of data within a boundary ( $\delta$ ) from target values, defined by  $TDI(p) < \delta$ .

Two measurement methods may be considered to have sufficient agreement if a large proportion of their differences is small. Thus, we define  $p$  as the proportion of their differences and  $\delta$  as the sufficient difference. Its estimate can be evaluated using (A7).

TDI can be also used for the evaluation of the intra-agreement for each device separately. The estimates can be evaluated using (A14, A18).

Relative indices are scaled values on a predefined range and usually lie between  $-1$  and  $1$ . The concordance correlation coefficient (CCC) is the most popular index for assessing agreement between quantitative measurements (inter-agreement). There is perfect agreement when  $CCC = 1$ , no agreement when  $CCC = 0$  and perfect negative agreement when  $CCC = -1$ . Its estimate can be evaluated using formulas (A8, A9). CCC can be also used for the evaluation of the intra-agreement for each device separately. The estimates can be evaluated using formulas (A15, A19).

Moreover, the 95% limits of agreement produce an interval within which 95% of differences between measurements made by the methods/devices are expected to lie [26].

An in-house-built R-script [23] implementing the relevant methods [28] may be used to evaluate CCC and TDI along with their corresponding bounds. Moreover, TDI evaluation and its upper bound, based on an alternative formulation of mixed-effects models [34], can be implemented [35]. CCC evaluation and confidence intervals for inference, instead of a lower bound, using an alternative formulation of mixed-effects models [36,37] can be implemented using the “*cccrm*” package [38]. The limits of agreement can be evaluated and presented graphically by the “*methodCompare*” package via the *bland\_altman\_plot()* function along with the corresponding extended Bland–Altman plot. A wide data format to implement *bland\_altman\_plot()* is needed. The package “*blandr*” [39] can be used to evaluate the limits of agreement along with their confidence intervals, superimposed on a Bland–Altman plot.

Moreover, the bias plot (*bias\_plot()* function [27]) from the “*methodCompare*” package evaluates the differential (fixed) and proportional bias (described in Appendix B.1.2) and offers a useful display that quantifies systematic bias (fixed and proportional) along the measurement range.

#### 2.3.4. Investigating Possible Sources of Disagreement

##### Assessing Similarity

Early research assessing similarity measures was focused on paired data [40]. Precision and accuracy via the quantification of fixed and proportional bias, along with the precision ratio, were proposed as measures of similarity [41].

For mixed-effect models, only the fixed bias can be evaluated since proportional bias is assumed to be equal to 1. To implement the standard methodology to evaluate the fixed bias and precision ratio [11], an R-script is available online [23]. The formal methodology for similarity assessment can be found in Appendix B.2.2.

For measurement-error models, similarity can be evaluated using a bias plot (discussed in Section 2.3.3) and a precision plot [29–31]. The precision plot can be implemented using the “*methodCompare*” [27] package via the *precision\_plot()* function.

##### Assessing Repeatability

The evaluation of repeatability is essential and can be used to identify possible sources of disagreement. It is considered as intra-method agreement and is an essential part of the agreement study. When a method/device has low intra-method agreement it will most probably have low inter-method agreement suggesting poor overall agreement of the methods or devices.

For mixed-effect models, CCC, TDI and corresponding 95% limits of agreement can be used to assess intra-method agreement. These are evaluated for each method/device separately and assess the agreement between repeated measurements with the same device. Implementation of relevant methods [11,28] is possible using an online R-script developed by the first author [23].

For measurement-error models, repeatability can be assessed graphically via a bias plot (discussed in Section 2.3.3) by investigating the spread of the measurements of each method/device. Repeatability can also be assessed using a trellis plot.

#### 2.3.5. Recalibration Methods

For the mixed-effects model, a recalibration procedure is performed by subtracting the fixed-bias. The relevant methodology [28] can be implemented using the R-script available online [23].

For the measurement-error model, a recalibration procedure is described in [29–31] and the function *compare\_plot()* from the “*MethodCompare*” package can be used to implement it. Appendix B.4 describes the procedure.

Figure 3 summarizes the workflow of a method comparison study.

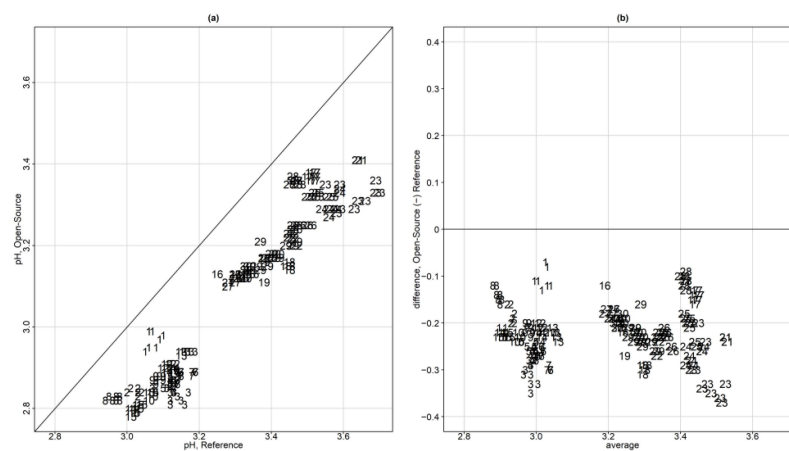




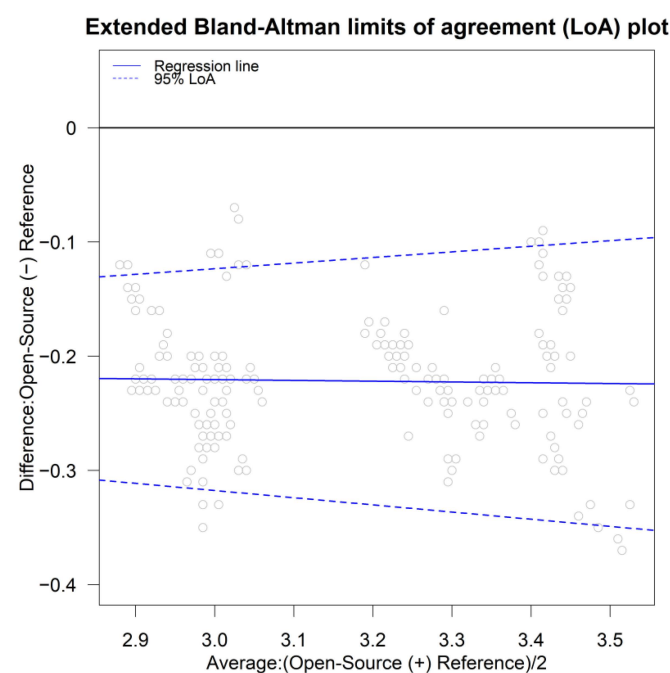
There are two clusters formed in the data. The lower left corresponds to the grapefruit pH while the upper right corresponds to the orange juice.

Figure 4b displays a Bland–Altman plot (averages vs. differences) without the limits of agreement. For higher values of pH, the differences seem to have slightly higher spread compared to the lower values of pH. This is a sign of possible heteroscedastic errors. There is no obvious trend in the Bland–Altman plot suggesting a common scale for the assays, verifying the common scale assumption for the mixed-effects model. This is also obvious in the extended Bland–Altman plot, which was produced using the “MethodCompare” package (Figure 5).

Figure 6 displays a trellis plot. The vertical axis is divided into rows and each row displays all the repeated measurements for one subject and both devices using method-specific colors. Blue color represents the measurements for the reference device while yellow represents the measurements for the open-source device.

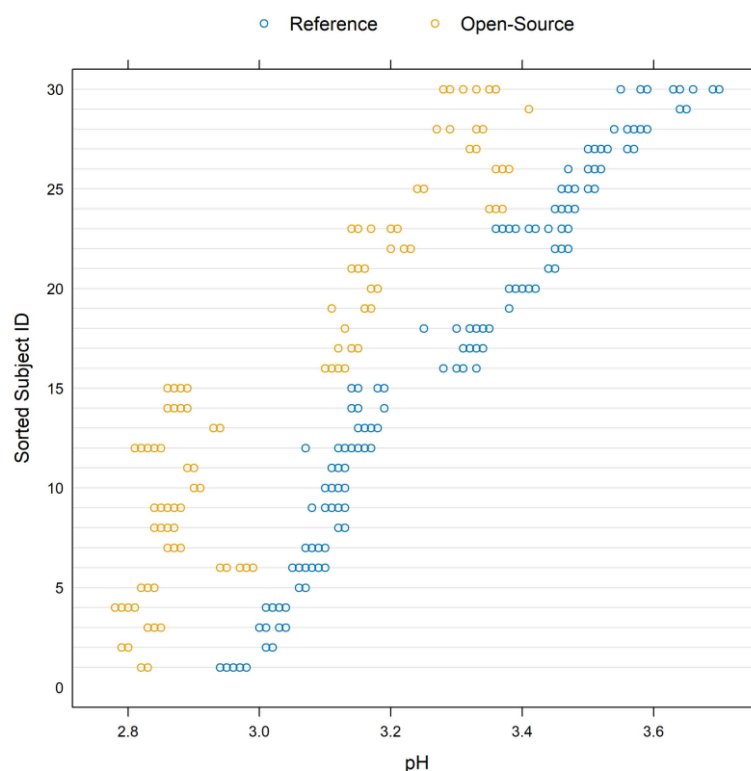


**Figure 4.** (a) Scatterplot for the pH measurements of the reference versus the open–source device. Each subject is represented by a unique ID number and the repeated values share the same ID subject number symbol. (b) Bland–Altman plot (averages vs. differences) without the limits of agreement.



**Figure 5.** Extended Bland–Altman plot and LOA. There is slight evidence of heteroskedastic errors. No trend is apparent; thus, a common scale is assumed for the assays.

Since the repeated measurements are plotted in one row, within-subject variability is visible and easy to compare with the between-subject variability. The open-source device shows slightly less within-subject variation compared to the reference. The between-subject variation ranges between 2.78 and 3.7 and a summary is presented in Table 2. A consistent bias is also visible in the graph, suggesting a constant fixed bias throughout the measurement range. The open-source device underestimates the pH measurements by approximately 0.22 units.



**Figure 6.** Trellis plot for the pH measurements. The vertical axis is divided into rows and each row displays all the repeated measurements for one subject using method-specific colors. Blue represents the measurements for the reference while yellow represents the measurements for the open-source device.

**Table 2.** The minimum, 1st quartile, median, mean 3rd quartile, and maximum pH values per device.

|             | Minimum | Q1    | Median | Mean  | Q3    | Max   |
|-------------|---------|-------|--------|-------|-------|-------|
| Open-Source | 2.780   | 2.868 | 3.045  | 3.053 | 3.220 | 3.410 |
| Reference   | 2.040   | 3.110 | 3.220  | 3.275 | 3.460 | 3.700 |

Figure A2a,b display the interaction plots for the subject x method and subject x time, respectively. For the subject x method interaction plot, the average per subject for every measurement is plotted on the vertical axis and each method on the horizontal (Figure A2a). There is evidence of a significant subject x method interaction since the lines intersect. Figure A2b displays the subject x time interaction. The repeated measurements are averaged per method for each subject (vertical axis) and the time points are displayed on the horizontal axis. Some lines intersect, providing evidence of possible interaction between the subjects and time.

### 3.3. Statistical Tools to Assess Agreement and Similarity

Initially, the data were fit with the homoscedastic model with no covariates for linked data (A2) and then the corresponding heteroscedastic. However, the additional computa-

tional burden provided by the subject to occasion interaction for the linked data hindered the procedure to calculate the confidence bounds for the indices. Thus, the unlinked homoscedastic and heteroscedastic models were chosen to proceed with the analysis. There is no obvious sign of a fan shape. AIC and BIC were subsequently calculated, and the heteroscedastic model was chosen (Table 3).

**Table 3.** AIC, BIC, log-likelihood and degrees of freedom for the homoscedastic and the heteroscedastic model.

|                 | AIC       | BIC       | Log-Likelihood | Degrees of Freedom |
|-----------------|-----------|-----------|----------------|--------------------|
| Homoscedastic   | −2945.889 | −2919.507 | 1478.945       | 6                  |
| Heteroscedastic | −2999.128 | −2963.952 | 1507.564       | 8                  |

At a subsequent stage, model (A4) was used to fit the data, which includes the covariates without interactions. According to the AIC and BIC criteria, the model without covariates was preferred. Table 4 displays AIC, BIC, log-likelihood and degrees of freedom for the heteroscedastic models with and without covariates.

**Table 4.** AIC, BIC, log-likelihood and degrees of freedom for model selection.

| Covariates | AIC       | BIC       | Log-Likelihood | Degrees of Freedom |
|------------|-----------|-----------|----------------|--------------------|
| No         | −2999.128 | −2963.952 | 1507.564       | 8                  |
| Yes        | −2641.498 | −2597.528 | 1330.749       | 10                 |

To account for heteroscedasticity, a sequence of 20 values starting from 2.78, which is the minimum value for the average values of the two methods, and 3.7, which is the maximum value for the average of the two methods, was created. Then, the variance function was defined as  $g(u_i, \delta) = |u_i|^\delta$ , where  $\tilde{u}_i = h(\bar{y}_{i1}, \bar{y}_{i2}) = \frac{\bar{y}_{i1} + \bar{y}_{i2}}{2}$ . The variance function parameter  $\tilde{u}_i$  can also be chosen as the average values per subject of the reference device. No significant changes were reported regardless of the choice of  $\tilde{u}_i$ . The following values were recorded: parameter  $\delta_1 = 4.07$  for the reference and  $\delta_1 = 3.28$  for the open-source device. The model's counterparts are displayed in Table 5.

**Table 5.** Model counterparts.

| Parameter                        | Estimate | SE   | 95% CI           |
|----------------------------------|----------|------|------------------|
| $\beta_0$                        | −0.22    | 0.01 | [−0.24, −0.20]   |
| $\mu_b$                          | 3.27     | 0.04 | [3.20, 3.35]     |
| $\log(\sigma_b^2)$               | −3.24    | 0.26 | [−3.75, −2.72]   |
| $\log(\psi^2)$                   | −6.67    | 0.26 | [−7.18, −6.15]   |
| $\log(\sigma_{\varepsilon_1}^2)$ | −17.46   | 1.61 | [−20.61, −14.31] |
| $\log(\sigma_{\varepsilon_2}^2)$ | −16.42   | 1.61 | [−19.58, −13.27] |
| $\delta_1$                       | 4.07     | 0.70 | [2.71, 5.44]     |
| $\delta_2$                       | 3.28     | 0.70 | [1.91, 4.65]     |

Diagnostics for the optimal model (Figure A3, Appendix A) display the standardized residuals on the horizontal axis vs. the quantiles of the standard normal distribution. The plot reveals a slight deviation from the normal distribution. The standard errors for the estimates are reasonable; thus, the agreement and similarity indices' evaluation proceeds using this model.

Substituting the ML estimates from Table 5 in (A3) to obtain the fitted distribution ( $Y_1, Y_2$ ) given the pH level  $\tilde{u}$  yields

$$N_2 \left( \begin{pmatrix} -0.22 \\ 3.05 \end{pmatrix}, \begin{pmatrix} 0.04062325 + 2.615606 \times 10^{-8} \tilde{u}^{4.07} & 0.03934918 \\ 0.03934918 & 0.04062325 + 7.374885 \times 10^{-8} \tilde{u}^{3.28} \end{pmatrix} \right),$$

while for  $D$ , given  $\tilde{u}$ :

$$D|\tilde{u} \sim N_1\left(-0.22, 0.002548149 + 2.615606 \times 10^{-8}\tilde{u}^{4.07} + 7.374885 \times 10^{-8}\tilde{u}^{3.28}\right)$$

The intra-method difference distribution given  $\tilde{u}$  is produced by substituting the parameters from Table 5 in (A13):

$$D_1|\tilde{u} \sim N_1\left(0, 5.231212 \times 10^{-8}\tilde{u}^{4.07}\right),$$

$$D_2|\tilde{u} \sim N_1\left(0, 1.474977 \times 10^{-8}\tilde{u}^{3.28}\right).$$

$D_1$  denotes the differences for the reference and  $D_2$  denotes the differences for the open-source device.

### Assessment of agreement

Using formula (A6) to calculate the limits of agreement substituting the model's counterparts in Table 5 and the variance function, the inter- and intra-device agreement is displayed in Figure 7. For the inter-agreement of the devices, Table 6 summarizes the ranges for the 95% limits of agreement for pH data as a function of the magnitude of measurements. The inter-method limits, based on the distribution of  $D$ , are centered at  $-0.22$ . For lower pH values, the LOAs are narrower compared to the higher pH values and are in the range of  $[-0.3464, -0.3237]$  for lower LOA and  $[-0.1193, -0.0966]$  for upper LOA. The intervals reveal a systematic underestimation of the pH measurements from the open-source device. Figure 8 illustrates the Bland–Altman plot and limits of agreement along with their corresponding confidence intervals. The “*Blandr*” R package was used to produce the plot.

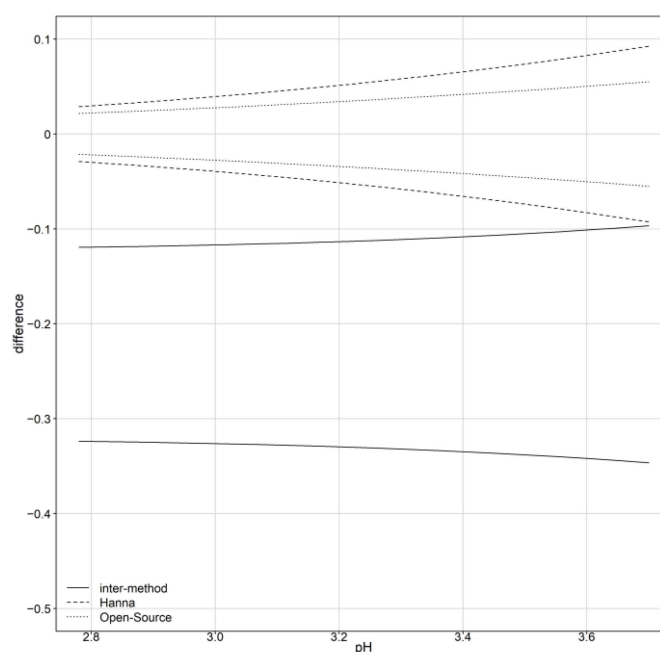
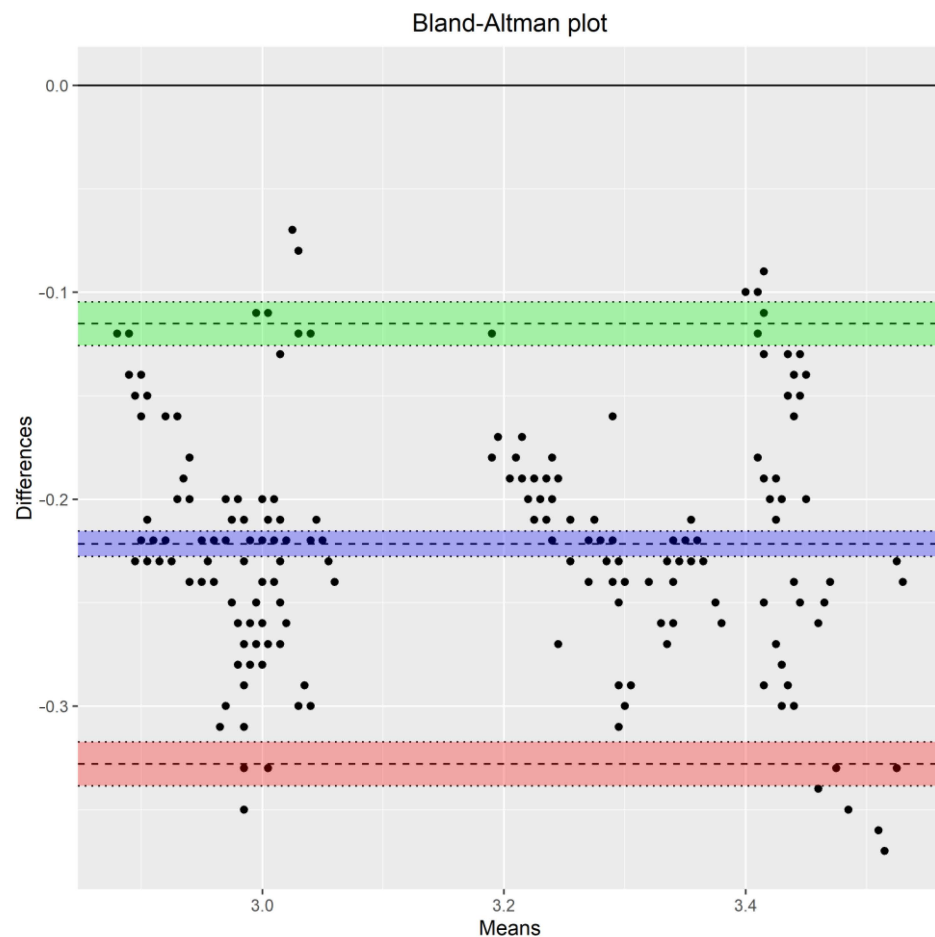


Figure 7. The 95% limits of inter- and intra-method agreement.



**Table 6.** Estimates of 95% limits of agreement for the inter-method agreement for pH measurements as a function of their magnitude.

| Limits of Agreement (Inter-Method) |         |         |         |         |         |         |
|------------------------------------|---------|---------|---------|---------|---------|---------|
|                                    | Minimum | Q1      | Median  | Mean    | Q3      | Max     |
| Lower                              | −0.3464 | −0.3369 | −0.3305 | −0.3321 | −0.3263 | −0.3237 |
| Upper                              | −0.1193 | −0.1167 | −0.1125 | −0.1125 | −0.1060 | −0.0966 |



**Figure 8.** Bland–Altman plot using the “blandr” package in R. Apart from the limits of agreement and the mean difference, their corresponding confidence intervals are plotted. The green and red confidence intervals correspond to the upper and lower limits of agreement respectively. The purple confidence interval corresponds to the mean difference.

Table 7 presents CCC and TDI estimates, and lower and upper confidence bounds, respectively, before and after recalibration. Before recalibration, the estimates for CCC range between 0.5970 and 0.6032, while the corresponding lower confidence bounds range between 0.4776 and 0.4839 throughout the pH measurement range. TDI (0.9) estimates range between 0.2883 and 0.3031 and their corresponding upper confidence bounds range between 0.3095 and 0.3232 throughout the pH measurement range.

**Table 7.** CCC and TDI estimates with their corresponding lower and upper confidence bounds throughout the pH measurement range before and after recalibration.

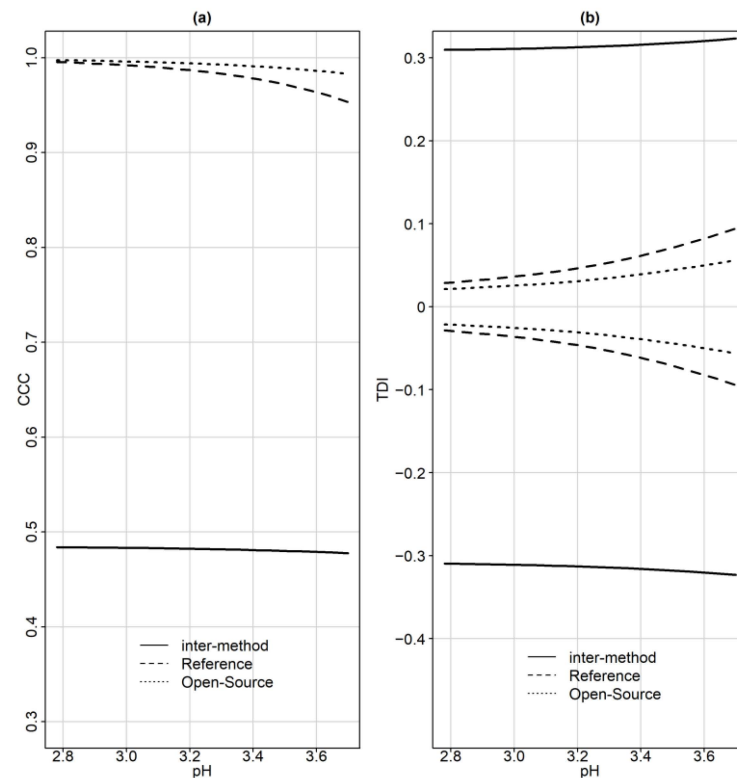
| Before Recalibration |        |                        |
|----------------------|--------|------------------------|
| Grid                 | CCC    | Lower Confidence Bound |
| 2.78                 | 0.6032 | 0.4839                 |
| 3.7                  | 0.5970 | 0.4776                 |
| Grid                 | TDI    | Upper Confidence Bound |
| 2.78                 | 0.2883 | 0.3095                 |
| 3.7                  | 0.3031 | 0.3232                 |
| After Recalibration  |        |                        |
| Grid                 | CCC    | Lower Confidence Bound |
| 2.78                 | 0.9666 | 0.9407                 |
| 3.7                  | 0.9509 | 0.9194                 |
| Grid                 | TDI    | Upper Confidence Bound |
| 2.78                 | 0.0857 | 0.1052                 |
| 3.7                  | 0.1048 | 0.1215                 |

Figure 9a presents one-sided 95% pointwise confidence bands for CCC as a function of the magnitude of the measurements. The solid line represents a lower CCC confidence bound for the inter-method agreement and ranges between 0.4776 and 0.4839. The lower CCC band decreases as the pH level increases. Thus, the agreement becomes progressively worse but only by a small amount. The inter-method agreement is not considered to be satisfactory. Figure 9b presents the one-sided 95% pointwise upper confidence bands for inter- and intra-method versions of TDI (0.9) and their reflections over the horizontal line at zero. For the inter-method agreement TDI (0.9), which is represented by the solid line, upper confidence bound ranges between 0.3095 and 0.3232. As the pH level increases from 2.78 to 3.7, TDI increases. The bound of 0.3232 shows that 90% of differences in measurements from the devices fall within  $\pm 0.3232$  when the true value is 3.7. Such a difference is unacceptably large for many applications. The bounds of 0.3095 and 0.3232 are, in proportional terms, 8.36 and 8.74% of the true value, respectively. A non-significant difference appears for the inter-method agreement throughout the pH measurement range. The similarity evaluation reveals that a difference in the means of the devices is a contributor to disagreement. TDI and CCC improve after recalibration.

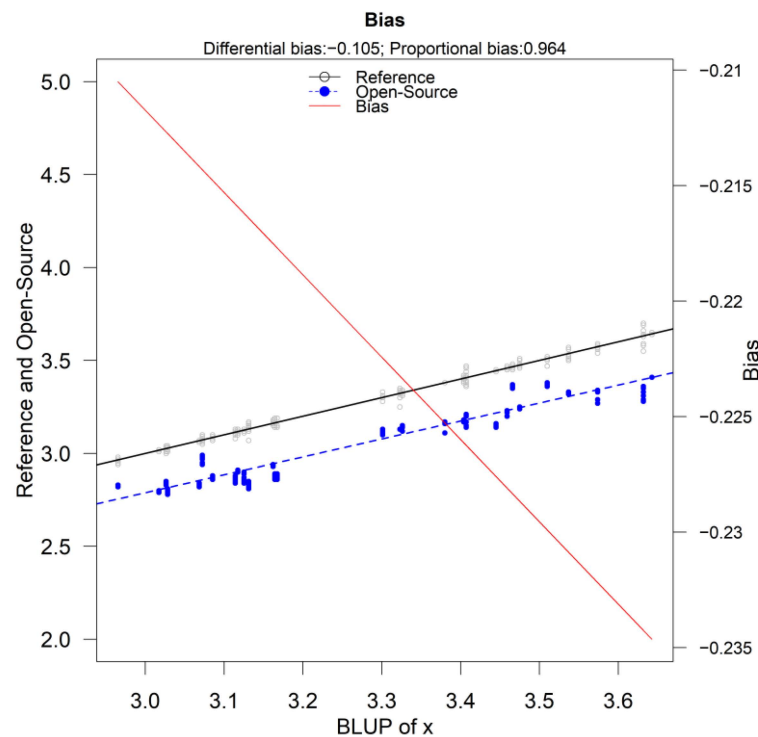
Overall, as the magnitude increases, TDI increases and CCC decreases. This means that the inter-method agreement becomes worse as the magnitude increases.

Following an alternative approach [35] to calculate TDI (0.9), the estimates are similar with equivalent conclusions before and after recalibration. The same applies for CCC [36].

Figure 10 displays the bias plot. The proportional bias is 0.965 (95% CI [0.9352, 0.9938]) and the fixed bias is  $-0.1052$  (95% CI  $[-0.2013, -0.091]$ ). The fixed bias estimate is different compared to the standard estimate [11] because the parameter estimation method is different. However, the red solid line that corresponds to the total bias confirms the findings that follow, presented in Table 8, Section 3.4, since the total bias is in the range of  $[-0.235, -0.21]$ .



**Figure 9.** (a) One-sided 95% pointwise confidence bands for CCC as a function of the magnitude of the measurements. The solid line represents a lower CCC confidence bound for the inter-method agreement. (b) One-sided 95% pointwise upper confidence bands for intra-method versions of TDI (0.9) and their reflections over the horizontal line at zero.

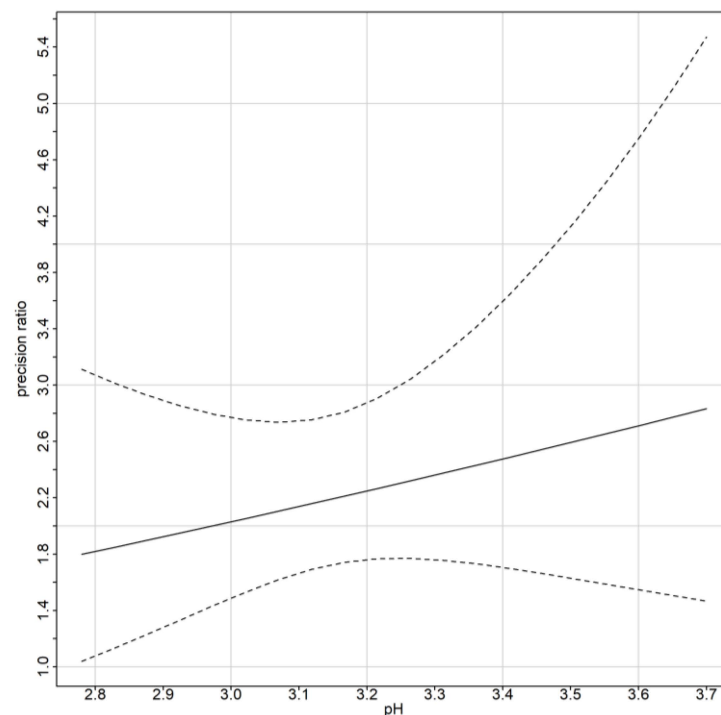


**Figure 10.** Bias plot of the reference versus the open-source device. The proportional bias is 0.965 (95% CI [0.9352, 0.9938]) and the fixed bias is estimated at  $-0.1052$  (95% CI  $[-0.2013, -0.091]$ ).

### 3.4. Investigating Possible Sources of Disagreement

#### Similarity Assessment

For the assessment of similarity, fixed bias and precision ratios are estimated. Fixed bias represents the difference in means of the two devices under the equal scale assumption. Since the errors are heteroscedastic and the precision is defined as the error variance of the reference over the error variance of the open-source, the precision as a function of magnitude of measurement is displayed in Figure 11. Table 8 summarizes the two indices. The open-source device is twice to three times more precise than the reference. The fixed bias is  $-0.22$  units for the open-source device compared to the reference. The open-source device underestimates the pH measurement by  $0.22$  units since the entire interval is below zero. The open-source device can be considered as of higher precision. These findings are consistent with the exploratory analysis.



**Figure 11.** Precision ratio along with corresponding 95% confidence intervals.

**Table 8.** Precision ratio estimates as a function of the magnitude of measurement. The fixed bias is  $-0.22$  units for the open-source device compared to the reference. The open-source device underestimates the pH measurement by  $0.22$  units.

| Similarity Assessment |      |                           |  |
|-----------------------|------|---------------------------|--|
| Grid                  | 2.78 | Lambda Estimate<br>1.7987 | Confidence Interval<br>[1.03969, 2.737423]           |
|                       | 3.7  | 2.832176                  | [1.770244, 5.471250]                                 |
| Fixed Bias            |      | Estimate<br>$-0.221497$   | Confidence Interval<br>[ $-0.239736$ , $-0.203257$ ] |

#### Evaluation of Repeatability

CCC, TDI and the limits of agreement are calculated for the intra-agreement of each device separately. Figure 7 displays the limits of agreement as a function of the magnitude of measurement. The limits of agreement for the open-source device (dotted lines) are included in the reference's LOA (dashed lines). Table 9 summarizes the ranges for the 95%

limits of agreement for pH data as a function of the magnitude of measurement. The open-source device LOAs are narrower compared to the reference, suggesting better repeatability. Based on the distributions of  $D_1$  and  $D_2$ , the intra-method limits are centered at zero. In Figure 9a, the CCC index is presented for inter- and intra-method agreement. The dashed and dotted lines represent the intra-method agreement for the reference and open-source device, respectively. For the reference device, the upper bound ranges between 0.9534 and 0.9955 and for the open-source, it ranges between 0.9830 and 0.9975. The intra-method agreement for both devices is considered excellent. However, the open-source device has higher intra-method agreement. This conclusion is expected since the similarity assessment reported smaller error variation for the open-source device.

**Table 9.** The 95% limits of agreement for intra-method agreement for pH data as a function of the magnitude of measurement. The open-source device has narrower LOA compared to the reference pH meter. Thus, the open-source device has better repeatability. Based on the distributions of  $D_1$  and  $D_2$ , the intra-method limits are centered at zero.

| Limits of Agreement (Intra-Method) |         |         |         |         |         |         |
|------------------------------------|---------|---------|---------|---------|---------|---------|
| Reference                          | Minimum | Q1      | Median  | Mean    | Q3      | Max     |
| Lower                              | −0.0926 | −0.0713 | −0.0540 | −0.0564 | −0.0400 | −0.0289 |
| Upper                              | 0.0289  | −0.0400 | 0.0540  | 0.0564  | 0.0713  | 0.0926  |
| Open-Source                        | Minimum | Q1      | Median  | Mean    | Q3      | Max     |
| Lower                              | −0.0550 | −0.0446 | −0.0356 | −0.0366 | −0.0280 | −0.0216 |
| Upper                              | 0.0216  | 0.0280  | 0.0356  | 0.0366  | 0.0446  | 0.0550  |

Figure 9b illustrates TDI (0.9). For the open-source device, which is represented by the dotted line, the TDI (0.9) lower bound ranges between 0.0213 and 0.0562, while for the reference, the dashed line, it ranges between 0.028 and 0.0945. The interpretation for TDI (0.9) is as follows: the bound of 0.0213 implies that 90% of the time, the difference between two replications of the open-source device on the same subject falls within  $\pm 0.0213$  when the true pH value is 2.78. The TDI bounds for both devices are only 0.76–1.03% of the magnitude of measurement, indicating a high degree on intra-method agreement. Table 10 displays CCC and TDI (0.9) along with their corresponding bounds for the minimum and maximum range of the measurements. The high intra-method agreement of CCC values reflect that the within-subject variations for both assays are very small compared to the between-subject variation.

### 3.5. Recalibrating the Open-Source Device

The similarity evaluation reveals that the fixed bias (difference in the means) contributes to the disagreement between the two devices. Recalibration of the open-source devices by subtracting  $-0.22$  from its measurements makes the mean difference zero and improves the extend of agreement substantially. Table 7 reports CCC and TDI estimates and confidence bounds after recalibration. CCC improves significantly. The lower confidence bands range from 0.9194 to 0.9407, revealing excellent agreement throughout the measurement range. TDI also improves and ranges from 0.1052 to 0.1215 throughout the measurement range. The agreement for this case study is considered acceptable. TDI (0.9) and CCC were also calculated after recalibration following the work of Escaramis et al. [23] and Carrasco [36], respectively. They are both close to Table 7 values, with TDI (0.9) and CCC being lower compared to Table 7 values.



**Table 10.** Summary for CCC and TDI (0.9) for the intra-method agreement.

| Grid | CCC for Repeatability:<br>Reference   | Lower Confidence Bound |
|------|---------------------------------------|------------------------|
| 2.78 | 0.9973                                | 0.9955                 |
| 3.7  | 0.9732                                | 0.9534                 |
| Grid | CCC for repeatability:<br>Open-Source | Lower Confidence Bound |
| 2.78 | 0.9985                                | 0.9975                 |
| 3.7  | 0.9904                                | 0.9830                 |
| Grid | TDI for repeatability:<br>Reference   | Upper Confidence Bound |
| 2.78 | 0.0243                                | 0.0286                 |
| 3.7  | 0.0777                                | 0.0945                 |
| Grid | TDI for repeatability:<br>Open-Source | Upper Confidence Bound |
| 2.78 | 0.0181                                | 0.0213                 |
| 3.7  | 0.0462                                | 0.0562                 |

#### 4. Discussion

A portable open-source device that measures the pH of the juice of grapefruits and oranges was designed and constructed for laboratory experiments and in situ measurements. To evaluate its functionality, a method comparison study between the open-source device and a corresponding industrial device was designed. The statistical analysis to assess their agreement was based on indices and graphical methods using mixed-effects models. The agreement indices evaluated were the Concordance Correlation Coefficient (CCC) and the Total Deviation Index (TDI). TDI estimates and confidence bounds were evaluated using (A1) and the methodology described in Section 2.3 [11]. There were small differences between the two methods probably due to the different formulations of the mixed-effects models.

Overall, agreement between the two devices is not satisfactory but improves to excellent levels after recalibration since the main source of disagreement is the fixed bias (0.22 pH units).

Further experiments can be conducted to investigate the agreement for an extended range of measurements and a greater variety of fruits or other applications that include soil pH or substrate pH in soilless cultivations. The ionic strength of the solution can also be included in the list of covariates in case it is suspected that it might affect the device measurements in a different way for each device. An R-Script, schematics and Arduino code for researchers to follow the proposed methodology and develop the open-source device are available [23].

#### 5. Conclusions

This paper highlights the assessment of open-source devices, regarding their functionality and the validity of their measurements. The most effective way to validate the measurements of a novel device is to compare them to established commercial/industrial devices. The official and reliable process to accomplish this task involves the design and application of a method comparison study that includes proper experimental design and statistical tools to assess the agreement and similarity between the two devices. This methodology is applied mostly in medical research [42,43] but not limited to it.

Parts of the proposed guide are described in the literature [44], but restricted to the Bland–Altman plot and ICC. The current research proposes a step-by-step procedure to validate open-source devices, including the experimental design, descriptive statistics and a variety of formal statistical assessment and encourages the development of a protocol applied to this highly blooming field.

The incentive behind the design of the present method comparison study is based on the novel device's low cost and configurability compared to the reference device and the possibility of interchangeable use. Specifically, the open-source device cost is around four times cheaper compared to the reference device (cost of the reference device discussed in Section 2.2. and the open-source device in Appendix A, Table A4). The accuracy of the open-source device is  $\pm 0.1$  pH (at 25 °C) and the measuring pH interval is between 0 and 14 pH units as per the manufacturer's statement.

**Author Contributions:** Conceptualization, E.P.B. and C.T.N.; methodology, E.P.B., P.M. and C.T.N.; software, E.P.B.; validation, E.P.B.; formal analysis, E.P.B.; investigation, E.P.B., P.M. and C.T.N.; resources, E.P.B., P.M. and C.T.N.; data curation, E.P.B.; writing—original draft preparation, E.P.B.; writing—review and editing, E.P.B., P.M. and C.T.N.; visualization, E.P.B.; supervision, C.T.N.; project administration, E.P.B. and C.T.N.; funding acquisition, E.P.B. and C.T.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the European Social Fund (5113934).

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data available online: <https://github.com/kersee112358/Chapter-5---Ph-logger-sensor> (accessed on 15 November 2023).

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

## Appendix A

**Table A1.** Technical Specifications of pH sensor SEN0169.

| Specifications        |                      |
|-----------------------|----------------------|
| Module Power          | 3.3 or 5 Volts       |
| Measuring Range       | 0–14 pH              |
| Measuring Temperature | 0–60 °C              |
| Accuracy              | $\pm 0.1$ pH (25 °C) |
| Response Time         | $\leq 1$ min         |
| Interface             | Analog Output        |

**Table A2.** Waterproof DS18B20 digital temperature sensor specifications.

| Specifications        |                                   |
|-----------------------|-----------------------------------|
| Module Power          | 3.3 or 5 Volts                    |
| Measuring Range       | −55 °C to 125 °C                  |
| Accuracy              | $\pm 0.5$ °C from −10 °C to 85 °C |
| Resolution            | 9 to 12 bits ADC                  |
| Interface             | 1-Wire                            |
| Steel Tube Dimensions | 6 mm diameter by 35 mm long       |

**Table A3.** Temperature of the pH buffer during the calibration procedure. Eight measurements were taken: four for 4.01 pH buffer and four for 7.01 pH buffer.

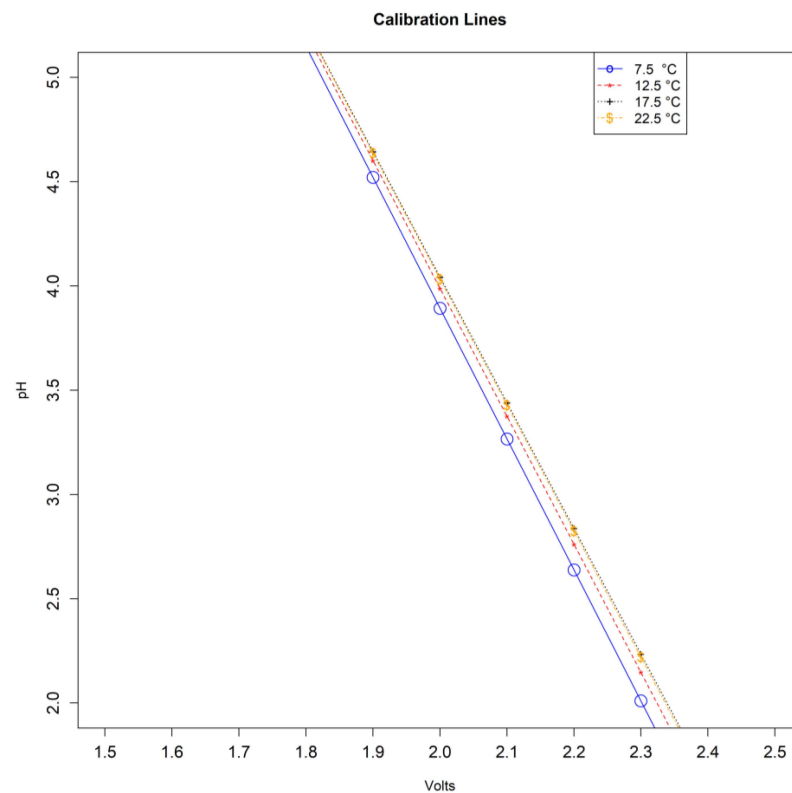
| Temperature | Temperature Range (°C) | Voltage | pH   |
|-------------|------------------------|---------|------|
| 7.5 °C      | [5, 10]                | 1.983   | 4.01 |
| 12.5 °C     | (10, 15]               | 1.505   | 7.01 |
| 17.5 °C     | (15, 20]               | 1.998   | 4.01 |
| 22.5 °C     | (20, 25]               | 1.509   | 7.01 |
| 7.5 °C      | [5, 10]                | 2.007   | 4.01 |
| 12.5 °C     | (10, 15]               | 1.509   | 7.01 |
| 17.5 °C     | (15, 20]               | 2.010   | 4.01 |
| 22.5 °C     | (20, 25]               | 1.514   | 7.01 |

**Table A4.** Device cost (EUR) in late 2023 according to local vendor prices and international vendor Mouser.

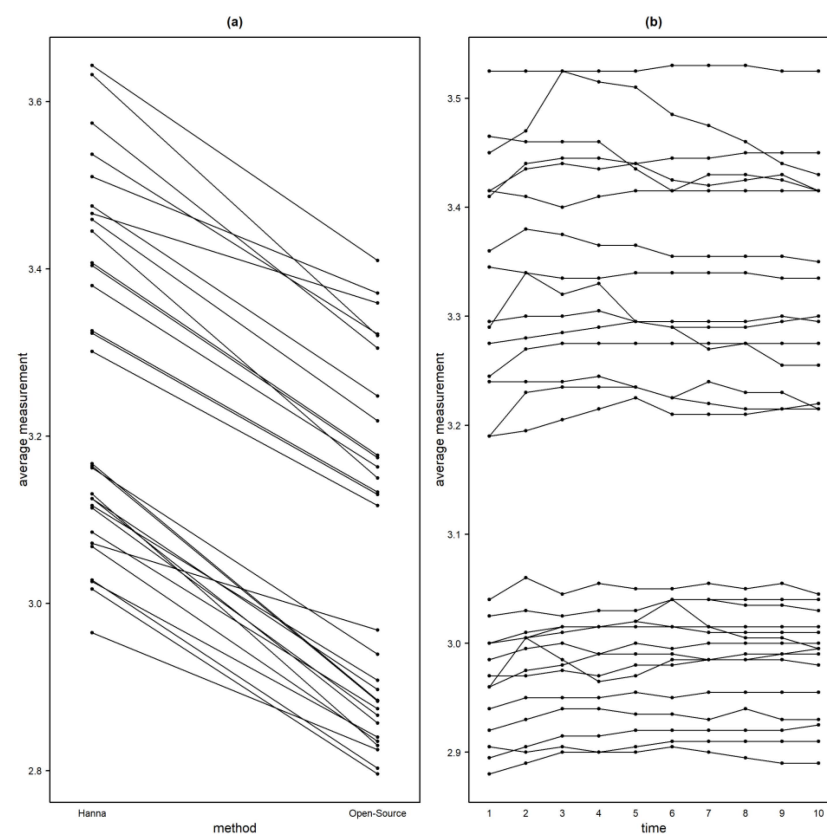
| Item                               | Cost (EUR) |
|------------------------------------|------------|
| DFRobot PH meter Pro Kit (SEN0619) | 70         |
| Temperature Sensor DS18B20         | 3          |
| Adafruit Feather proto 32u4        | 18.75      |
| Adafruit Adalogger FeatherWing     | 8.41       |
| Adafruit ADS1115 ADC 16bit         | 18.4       |
| microSD Card                       | 3          |
| Coin Cell Battery                  | 1          |
| IP66 Enclosure Box                 | 5          |
| PCB                                | 1.5        |
| Nokia 5110 module                  | 4.68       |
| Others (Wires, Solder)             | 2          |
| Total                              | 127.63     |

**Table A5.** The output of the pH electrode in Millivolts, and the linear relationship between pH value and output voltage (25 °C) according to the manufacturer (DFRobot).

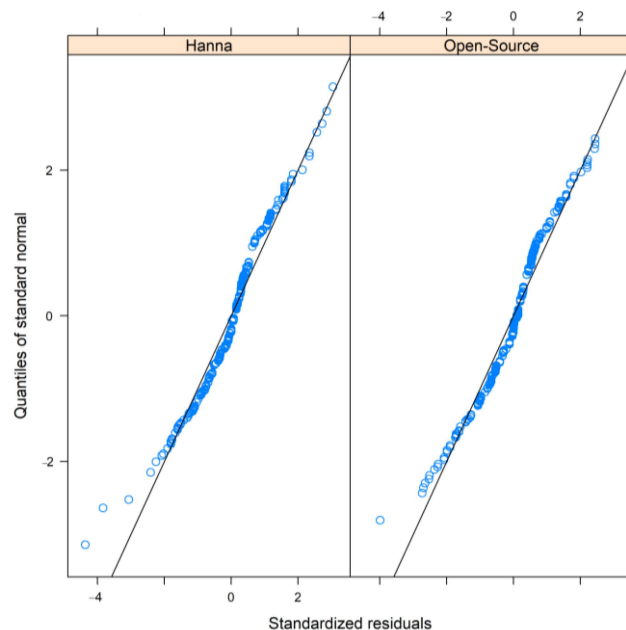
| Voltage (mV) | pH Value |
|--------------|----------|
| 414.12       | 0.00     |
| 354.96       | 1.00     |
| 295.80       | 2.00     |
| 236.64       | 3.00     |
| 177.48       | 4.00     |
| 118.32       | 5.00     |
| 59.16        | 6.00     |
| 0            | 7.00     |
| −59.16       | 8.00     |
| −118.32      | 9.00     |
| −177.48      | 10.00    |
| −236.64      | 11.00    |
| −295.80      | 12.00    |
| −354.96      | 13.00    |
| −414.12      | 14.00    |



**Figure A1.** Calibration curves for 7.5 °C, 12.5 °C, 17.5 °C and 22.5 °C.



**Figure A2.** (a) Interaction plot between method and subjects. There is evidence of subject  $\times$  method interaction since the lines intersect. (b) Interaction plot between subjects and time. A few of the lines intersect, providing evidence of possible, but not strong interaction, between the subjects and time.



**Figure A3.** Standardized residuals on the horizontal axis vs. the quantiles of the standard normal distribution. The plot reveals a slight deviation from the normal distribution.

## Appendix B

### Appendix B.1. Models

#### Appendix B.1.1. Mixed-Effects Models

Mixed effects model for unlinked data are described as follows [11]:

$$Y_{i1k} = b_i + b_{i1} + e_{i1k}, \quad Y_{i2k} = \beta_0 + b_i + b_{i2} + e_{i2k} \quad (\text{A1})$$

- $k = 1, \dots, m_{ij}$  are the repeated measurements.
- $i = 1, \dots, n$  is the subject's number.
- $j = 1, 2$  is the method's number.
- $\beta_0$  is a fixed effect and represents the difference in the fixed biases of the methods.
- $b_{ij}$  follow independent  $N_1(0, \psi^2)$  distributions. This is an interaction term. One interpretation for  $b_{ij}$  is the effect of method  $j$  on subject  $i$ . These interactions are subject-specific biases of the methods. They are a characteristic of the method-subject combination that remains stable during the measurement period.
- $b_i$  follows independent  $N_1(\mu_b, \sigma_b^2)$  distributions.
- $e_{ijk}$  follow independent  $N_1(0, \sigma_{e_j}^2)$  distributions.
- $b_i, b_{ij}$  and  $e_{ijk}$  are mutually independent.

To examine the measures of similarity and agreement, we must retrieve the parameters of the assumed model (A1), which produces a bivariate distribution for  $(Y_1, Y_2)$ . By dropping the subscripts for the sake of simplicity, we have:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N_2 \left( \begin{pmatrix} \mu_b \\ \beta_0 + \mu_b \end{pmatrix}, \begin{pmatrix} \sigma_b^2 + \psi^2 + \sigma_{e_1}^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \psi^2 + \sigma_{e_2}^2 \end{pmatrix} \right)$$

Thus,  $D = Y_2 - Y_1 \sim N_1(\beta_0, 2\psi^2 + \sigma_{e_1}^2 + \sigma_{e_2}^2)$ .

Then, the model has a total of six unknown parameters  $(\beta_0, \mu_b, \sigma_b^2, \psi^2, \sigma_{e_1}^2, \sigma_{e_2}^2)$ .

Linked data are modeled as in model (A1), except for the addition of the term  $b_{ik}^*$ , which represents the random effect of the common time  $k$  on the measurements.

$$Y_{i1k} = b_i + b_{i1} + b_{ik}^* + e_{i1k}, \quad Y_{i2k} = \beta_0 + b_i + b_{i2} + b_{ik}^* + e_{i2k} \quad (\text{A2})$$



$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N_2 \left( \begin{pmatrix} \mu_b \\ \beta_0 + \mu_b \end{pmatrix}, \begin{pmatrix} \sigma_b^2 + \psi^2 + \sigma_{b^*}^2 + \sigma_{e_1}^2 & \sigma_b^2 + \sigma_{b^*}^2 \\ \sigma_b^2 + \sigma_{b^*}^2 & \sigma_b^2 + \psi^2 + \sigma_{b^*}^2 + \sigma_{e_2}^2 \end{pmatrix} \right)$$

Thus,  $D = Y_2 - Y_1 \sim N_1(\beta_0, 2\psi^2 + \sigma_{e_1}^2 + \sigma_{e_2}^2)$ .

Then, the model has a total of seven unknown parameters  $(\beta_0, \mu_b, \sigma_b^2, \psi^2, \sigma_{b^*}^2, \sigma_{e_1}^2, \sigma_{e_2}^2)$ .

When the errors of models (A1) and (A2) are heteroscedastic,  $\sigma_{e_1}^2$  and  $\sigma_{e_2}^2$  are replaced with  $\sigma_{e_1}^2 g_1^2(u, \delta_1)$  and  $\sigma_{e_2}^2 g_2^2(u, \delta_2)$ . For a given  $u_i$  denoted as  $\tilde{u}_i$  ( $\tilde{u}$  for subject  $i$ ), function  $g$  is the variance function and  $\delta$  is a vector of heteroscedasticity parameters, where for  $\delta = 0$  corresponds to homoscedasticity. Variance covariate  $u$  is defined in advance ( $\tilde{u}$ ) and accounts for heteroscedasticity. Choudhary and Nagaraja [11] set  $\tilde{u}_i = h(\bar{y}_{i1}, \bar{y}_{i2}) = \bar{y}_{i1}$  if method 1 is the reference and  $\tilde{u}_i = h(\bar{y}_{i1}, \bar{y}_{i2}) = \frac{\bar{y}_{i1} + \bar{y}_{i2}}{2}$  otherwise. For the variance function  $g$ , two simple models are introduced: first, the power model, where  $g(u, \delta) = |u|^\delta$ ; second, the exponential model,  $g(u, \delta) = \exp(u\delta)$ . The parameters  $\delta_j$  can be estimated while fitting the model using ML and the “nlme” package [45]. More details on the choice of the variance function  $g$  can be found in [46]. AIC and BIC can be compared to distinguish between different candidate models for the variance functions.

The distribution of  $(Y_1, Y_2)$  for unlinked data is the following:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} | \tilde{u} \sim N_2 \left( \begin{pmatrix} \mu_b \\ \beta_0 + \mu_b \end{pmatrix}, \begin{pmatrix} \sigma_b^2 + \psi^2 + \sigma_{e_1}^2 g_1^2(\tilde{u}, \delta_1) & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \psi^2 + \sigma_{e_2}^2 g_2^2(\tilde{u}, \delta_2) \end{pmatrix} \right) \quad (A3)$$

For the linked data:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} | \tilde{u} \sim N_2 \left( \begin{pmatrix} \mu_b \\ \beta_0 + \mu_b \end{pmatrix}, \begin{pmatrix} \sigma_b^2 + \psi^2 + \sigma_{b^*}^2 + \sigma_{e_1}^2 g_1^2(\tilde{u}, \delta_1) & \sigma_b^2 + \sigma_{b^*}^2 \\ \sigma_b^2 + \sigma_{b^*}^2 & \sigma_b^2 + \psi^2 + \sigma_{b^*}^2 + \sigma_{e_2}^2 g_2^2(\tilde{u}, \delta_2) \end{pmatrix} \right)$$

Based on the model parameters the heteroscedastic difference distribution is the following both for the unlinked and linked data:

$$D | \tilde{u} \sim N_1(\beta_0, 2\psi^2 + \sigma_{e_1}^2 g_1^2(\tilde{u}, \delta_1) + \sigma_{e_2}^2 g_2^2(\tilde{u}, \delta_2))$$

### Models with covariates

Other factors might affect the agreement between the two methods. Covariates might affect the means of the methods (mean covariates), explaining part of the variability in the measurements. Covariates might also interact with the method or affect the error variance (variance covariates). In any case, the extent of the agreement is affected by the covariates. The mixed-effects models (A1) and (A2) can be extended as follows:

For unlinked data:

$$Y_{ijk} = \mu_j(x_{1i}, \dots, x_{ri}) + v_i + b_{ij} + e_{ijk} \quad (A4)$$

For linked data:

$$Y_{ijk} = \mu_j(x_{1i}, \dots, x_{ri}) + v_i + b_{ik}^* + b_{ij} + e_{ijk}$$

- $x_1, \dots, x_r$  are the mean covariates.
- $v_i \sim$  independent  $N_1(0, \sigma_b^2)$  is defined as  $v_i = b_i - \mu_b$
- $e_{ijk} \sim$  independent  $N_1(0, \sigma_{e_j}^2 g_j^2(u, \delta_j))$  accounts for possible heteroscedasticity
- $b_{ij} \sim$  independent  $N_1(0, \psi^2)$
- $b_{ik}^* \sim$  independent  $N_1(0, \sigma_{b^*}^2)$

Choudhary and Nagaraja [11] describe the detailed methodology for defining mean and variance-specific covariates.

### Appendix B.1.2. Measurement-Error Models

Taffé [29,30] and Taffé et al. [31] proposed various graphical tools to assess the bias and precision of measurement methods in method comparison studies by following the measurement error model proposed by [47] but using an alternative estimation procedure based on an empirical Bayes approach. The model is as follows:

$$Y_{1ij} = a_1 + \beta_1 x_{ij} + \varepsilon_{1ij}, \varepsilon_{1ij} | x_{ij} \sim N(0, \sigma_{\varepsilon_1}^2(x_{ij}, \theta_1)) \quad (A5)$$

$$Y_{2ij} = x_{ij} + \varepsilon_{2ij}, \varepsilon_{2ij} | x_{ij} \sim N(0, \sigma_{\varepsilon_2}^2(x_{ij}, \theta_2))$$

$$x_{ij} \sim f_x(\mu_x, \sigma_x^2)$$

- $Y_{kij}$  is the  $j^{th}$  replicate measurement by method  $k$  using individual  $i$ ,  $j = 1, \dots, n_i$ ,  $i = 1, \dots, N$ ,  $k = 1, 2$ .
- $n_i$  denotes the number of repeated measurements per subject.
- $x_{ij}$  is a latent trait variable with density  $f_x$  representing the true latent trait.
- $\varepsilon_{kij}$  represents measurement errors by method  $k$ .
- The variances of these methods  $\sigma_{\varepsilon_k}^2(x_{ij}, \theta_k)$  are heteroscedastic and increase with the level of the true latent trait  $x_{ij}$ , which depends on the vectors of unknown parameters  $\theta_k$  [29].
- The mean value of the latent trait is  $\mu_x$  and the variance is  $\sigma_x^2$ .
- It is assumed that the latent variable represents the true unknown but constant value of the trait for individual  $i$  and therefore  $x_{ij} \equiv x_i$ .
- The parameters  $a_1$  and  $\beta_1$  are considered fixed and the error produced by them is called systematic. Their values depend on the measurement method. Parameter  $a_1$  is defined as the fixed bias (or differential bias) of the method, while  $\beta_1$  is defined as the proportional bias. Fixed bias is the added constant that the measurement method adds to the true value for every measurement. Proportional bias is based on the measured quantity and is the slope of Equation (A5). The true value is multiplied by the proportional bias and is interpreted as the amount of change in the measurement method if the true value changes by 1 unit.

This modification of the classical measurement error model considers that heteroscedasticity depends on the latent trait and not on the observed average, compared to Choudhary's and Nagaraja's methodology [11]). The model is assumed to be linear even though non-linear functions of  $x_i$  can be used and easily interpreted. To visually assess the plausibility of the straight-line model, a graphical representation of  $|\varepsilon_{2ij}^*|$  versus  $\hat{x}_i$  provides a good start. Term  $|\varepsilon_{2ij}^*|$  is the regression of the absolute values of the residuals  $\varepsilon_{2ij}^*$  from the linear regression model  $y_{2ij} = \alpha_2^* + \beta_2^* \hat{x} + \varepsilon_{2ij}^*$  on  $\hat{x}_i$  (the estimate of the latent trait) by ordinary least squares.

### Appendix B.2. Indices to Assess Agreement and Similarity

#### Appendix B.2.1. Assessing Agreement

To evaluate the agreement and similarity between the reference and the open-source device, CCC and TDI were calculated using both (A1) and (A2) [11] and methodologies provided by [34,36].

The limits of agreement [26,44,48] for both unlinked and linked data were calculated using the parameters in models (A1) and (A2) via the following formulas provided by [11]:

$$LOA = \beta_0 \pm 1.96 \cdot \sqrt{2\hat{\psi}^2 + \hat{\sigma}_{\varepsilon_1}^2 + \hat{\sigma}_{\varepsilon_2}^2}, j = 1, 2 \quad (A6)$$

For heteroscedastic data,  $\sigma_{\varepsilon_1}^2$  and  $\sigma_{\varepsilon_2}^2$  are replaced with  $\sigma_{\varepsilon_1}^2 g_1^2(u, \delta_1)$  and  $\sigma_{\varepsilon_2}^2 g_2^2(u, \delta_2)$ .

For unlinked and linked data, using models (A1, A2) and following the approach described by [11], TDI can be calculated using the following formula:

$$TDI(p) = \sqrt{\left\{ (2\psi^2 + \sigma_{e_1}^2 + \sigma_{e_2}^2) \chi_{1,p}^2 \left( \frac{\beta_0^2}{\{2\psi^2 + \sigma_{e_1}^2 + \sigma_{e_2}^2\}} \right) \right\}} \quad (A7)$$

For unlinked data, under model (A1), [11] proposed the following formula to calculate CCC:

$$CCC = \frac{2\sigma_b^2}{\beta_0^2 + 2(\sigma_b^2 + \psi^2) + \sigma_{e_1}^2 + \sigma_{e_2}^2} \quad (A8)$$

While for the linked data under model (A2):

$$CCC = \frac{2(\sigma_b^2 + \sigma_{b^*}^2)}{\beta_0^2 + 2(\sigma_b^2 + \psi^2 + \sigma_{b^*}^2) + \sigma_{e_1}^2 + \sigma_{e_2}^2} \quad (A9)$$

### Inference for TDI

Escaramis et al. [34] proposed tolerance intervals for TDI. The value  $k_p$  is obtained by replacing  $\mu_D$  and  $\sigma_D$  by their REML estimate counterparts derived from mixed-effects models (available in [34]) in the expression  $\hat{k}_p = \hat{\mu}_D + z_{p_1} \hat{\sigma}_D$ . For inference, a one-sided tolerance interval is computed that covers the  $p_1$ -percent of the population from D with a stated confidence.

Let T be the studentized variable of  $\hat{\mu}_D + z_{p_1} \hat{\sigma}_D$ . T follows a non-central Student-t distribution with the non-centrality parameter  $z_{p_1} \sqrt{N}$ ,  $T \sim t_\nu(z_{p_1} \sqrt{N})$ , where  $N = 2 \cdot n \cdot m$  is the total possible paired-measurement differences between the two method/devices and the degrees of freedom  $\nu$  are derived from the residual degrees of freedom. For the case of using individual-device interaction or discarding it, [34] described different cases for obtaining  $\nu$ . An upper bound for the TDI estimate can be constructed by using the following  $(1 - \alpha) \cdot 100\%$  one-sided tolerance interval, where  $\alpha$  is the type I error rate:

$$UB(1 - \alpha) \cdot 100\%(\hat{k}_p) = \hat{\mu}_D + t_\nu(1 - \alpha, z_{p_1} \sqrt{N}) \frac{\hat{\sigma}_D}{\sqrt{N}}$$

It corresponds to the exact one-sided tolerance interval for at least the  $p_1$  proportion of the population [49,50].

To perform a hypothesis test if the interest is to ensure that at least  $p$ -percent of the absolute differences between paired measurements are less than a predefined constant  $k_0$ , Lin's form of hypothesis can be followed [32].

$$H_0 : k_p \geq k_0, H_1 : k_p < k_0$$

$H_0$  is rejected at level  $\alpha$  if:

$$UB(1 - \alpha) \cdot 100\%(\hat{k}_p) = \hat{\mu}_D + t_\nu(1 - \alpha, z_{p_1} \sqrt{N}) \frac{\hat{\sigma}_D}{\sqrt{N}} < k_0$$

Choudhary and Nagaraja [11] use the large-sample theory of ML estimators to compute standard errors, confidence bounds and tolerance intervals. When the sample is not large enough, bootstrap confidence intervals are produced. The estimators of the model-based counterparts are obtained from models (A1, A2), depending on the nature of the data (unlinked or linked, respectively). To compute simultaneous confidence intervals and bounds, the percentiles of appropriate functions of multivariate normally distributed pivots are needed. Following [51], the R-package "multcomp" [52] can be used. Choudhary and Nagaraja [11] proposed a Bootstrap-t UCB and a modified nearly unbiased estima-

tor (MNUT approximation) for computing the critical value, the  $p$ -value and the upper confidence bound (UCB) [11].

### Inference for CCC

Asymptotic distribution of the estimated CCCs can be used for inference if the data are modeled via a large sample size [53]. Choudhary and Nagaraja [11] use an asymptotic distribution of the estimated CCCs to produce an upper confidence bound when the sample is large and bootstrap methods when the sample is small. Since the concordance correlation coefficient is related to the intraclass correlation coefficient (ICC), inference methods for ICC can be used for CCC [53].

### Appendix B.2.2. Assessing Similarity

Following Choudhary and Nagaraja [11], to evaluate the similarity, the marginal distributions of  $Y_1$  and  $Y_2$  are examined via estimates and two-sided confidence intervals. Their distributions are given by Equations (A1) and (A2) for unlinked and linked data. The fixed bias and the precision ratio are the two measures of similarity that will be evaluated using mixed-effects model. Last, fixed bias, proportional bias and precisions are evaluated under measurement-error models.

Fixed bias will be estimated via the model's counterparts. According to models (A1, A2), the fixed bias is estimated using  $\mu_1 - \mu_2$  for unlinked and linked data.

The precision ratio is evaluated in two different cases.

First, for models that ignore subject  $\times$  method interactions:

$$\lambda = \frac{\sigma_{e_1}^2}{\sigma_{e_2}^2}$$

Second, for models that include subject  $\times$  method interactions:

$$\lambda = \frac{\sigma_{e_1}^2 + \psi^2}{\sigma_{e_2}^2 + \psi^2}$$

The precision ratios are assumed to be estimated when the errors are homoscedastic. For heteroscedastic data, we replace  $\sigma_{e_1}^2$  and  $\sigma_{e_2}^2$  with  $\sigma_{e_1}^2 \sigma_1^2(u_i, \delta_1)$  and  $\sigma_{e_2}^2 \sigma_2^2(u_i, \delta_2)$ ; thus, the fixed bias remains the same, but the precision ratio is given by:

$$\lambda = \frac{\sigma_{e_1}^2 \sigma_1^2(u_i, \delta_1)}{\sigma_{e_2}^2 \sigma_2^2(u_i, \delta_2)}$$

For inference, the method described by [47] is used for heteroscedastic data. Specifically, if  $\theta$  is a vector of the model's counterparts, then the measure of similarity is a function of  $\theta$ . Denoting the measure of similarity as  $\varphi$ , and  $b^*$  a value in the measurement range, then  $\varphi(b^*)$  is any measure of similarity in a specific value (the measure is assumed to be scalar). Substituting  $\theta$  with its corresponding ML estimate,  $\hat{\theta}$ , in its expression gives its ML estimator  $\hat{\varphi}(b^*)$ . Using the delta method [54], when the sample size is large,  $\hat{\varphi}(b^*) \sim N_1(\varphi(b^*), G'(b^*)I^{-1}G(b^*))$ , where  $G(b^*) = \left(\frac{\partial}{\partial \theta}\right)\varphi(b^*)|_{\theta=\hat{\theta}}$ , can be computed numerically. Thus, an approximate  $100(1 - \alpha)\%$  two-sided pointwise confidence interval for  $\varphi(b^*)$  on a grid of values of the measurement range can be computed as

$$\hat{\varphi}(b^*) \pm z_{1-\frac{\alpha}{2}} \left\{ G'(b^*)I^{-1}G(b^*) \right\}^{\frac{1}{2}}.$$

### Appendix B.3. Assessing Repeatability

Following [11] and mixed-effect models, for unlinked data, repeated measurements are replications of the same underlying measurement. Instead of using the bivariate distributions  $(Y_1, Y_2)$  for measurements of the two methods on a randomly selected subject

from a population,  $Y_j^*$  is defined as a replication of  $Y_j$ , where  $j = 1, 2$  denote the two methods/devices. By definition,  $Y_j$  and  $Y_j^*$  have the same distribution. CCC and TDI are modified and are calculated. By dropping the subscripts for model (A1), for unlinked data:

$$Y_1^* = b + b_1 + e_1^*, Y_2^* = b + b_2 + e_2^* \quad (A10)$$

is induced, similar to (A1) by dropping the subscripts.

$$\text{Then, for method 1: } \begin{pmatrix} Y_1 \\ Y_1^* \end{pmatrix} \sim N_2 \left( \begin{pmatrix} \mu_b \\ \mu_b \end{pmatrix}, \begin{pmatrix} \sigma_b^2 + \psi^2 + \sigma_{e_1}^2 & \sigma_b^2 + \psi^2 \\ \sigma_b^2 + \psi^2 & \sigma_b^2 + \psi^2 + \sigma_{e_1}^2 \end{pmatrix} \right) \quad (A11)$$

$$\text{For method 2: } \begin{pmatrix} Y_2 \\ Y_2^* \end{pmatrix} \sim N_2 \left( \begin{pmatrix} \beta_0 + \mu_b \\ \beta_0 + \mu_b \end{pmatrix}, \begin{pmatrix} \sigma_b^2 + \psi^2 + \sigma_{e_2}^2 & \sigma_b^2 + \psi^2 \\ \sigma_b^2 + \psi^2 & \sigma_b^2 + \psi^2 + \sigma_{e_2}^2 \end{pmatrix} \right) \quad (A12)$$

where  $e_1^*$  and  $e_2^*$  are independent copies of  $e_1$  and  $e_2$  as defined in (A1).

In addition,  $D_j = Y_j - Y_j^*$  can be defined as the difference in two replications of method  $j$ . From Equations (A10) and (A11), it can be calculated that:

$$D_j \sim N_1(0, 2\sigma_{e_j}^2), j = 1, 2. \quad (A13)$$

Thus,

$$CCC_j = \frac{\sigma_b^2 + \psi^2}{\sigma_b^2 + \psi^2 + \sigma_{e_j}^2} \quad (A14)$$

$$TDI_j = \left\{ 2\sigma_{e_j}^2 \chi_{1,p}^2(0) \right\}^{\frac{1}{2}} \text{ for } j = 1, 2. \quad (A15)$$

By dropping the subscripts for model (A2) for linked data:

$$Y_1^* = b + b_1 + b^{**} + e_1^*, Y_2^* = b + b_2 + b^{**} + e_2^*$$

$$\text{For method 1: } \begin{pmatrix} Y_1 \\ Y_1^* \end{pmatrix} \sim N_2 \left( \begin{pmatrix} \mu_b \\ \mu_b \end{pmatrix}, \begin{pmatrix} \sigma_b^2 + \psi^2 + \sigma_{b^{**}}^2 + \sigma_{e_1}^2 & \sigma_b^2 + \psi^2 \\ \sigma_b^2 + \psi^2 & \sigma_b^2 + \psi^2 + \sigma_{b^{**}}^2 + \sigma_{e_1}^2 \end{pmatrix} \right) \quad (A16)$$

$$\text{For method 2: } \begin{pmatrix} Y_2 \\ Y_2^* \end{pmatrix} \sim N_2 \left( \begin{pmatrix} \beta_0 + \mu_b \\ \beta_0 + \mu_b \end{pmatrix}, \begin{pmatrix} \sigma_b^2 + \psi^2 + \sigma_{b^{**}}^2 + \sigma_{e_2}^2 & \sigma_b^2 + \psi^2 \\ \sigma_b^2 + \psi^2 & \sigma_b^2 + \psi^2 + \sigma_{b^{**}}^2 + \sigma_{e_2}^2 \end{pmatrix} \right) \quad (A17)$$

In the above expression,  $b^{**}$ ,  $e_1^*$  and  $e_2^*$  are independent copies of  $b^*$ ,  $e_1$  and  $e_2$  as defined in (A2).

In addition,  $D_j = Y_j - Y_j^*$  can be defined as the difference in two replications of method  $j$ . From Equations (A16) and (A17), it can be calculated that:

$$D_j \sim N_1(0, 2(\sigma_{b^{**}}^2 + \sigma_{e_j}^2)), j = 1, 2.$$

Thus,

$$CCC_j = \frac{\sigma_b^2 + \psi^2}{\sigma_b^2 + \psi^2 + \sigma_{b^{**}}^2 + \sigma_{e_j}^2} \quad (A18)$$

$$TDI_j = \left\{ 2(\sigma_{e_j}^2 + \sigma_{b^{**}}^2) \chi_{1,p}^2(0) \right\}^{\frac{1}{2}} \text{ for } j = 1, 2. \quad (A19)$$

#### Appendix B.4. Recalibration Methods

For measurement-error models, a recalibration procedure is performed by computing:

$$y_{1ij}^* = \frac{Y_{1ij} - \hat{a}_1^*}{\hat{\beta}_1^*}$$



where  $\hat{\alpha}_1^*$  is the estimate of the proportional bias and  $\hat{\beta}_1^*$  is the estimate of the fixed bias and  $Y_{ij}^*$  is the recalibrated value. The method performs well, according to simulations, with a sample size of 100 subjects and 10 to 15 repeated measurements per individual from the reference method and only 1 from the new. It is possible that after the recalibration procedure, the novel method turns out to be more precise than the reference. The recalibration procedure can be implemented using the `compare_plot()` function from the “*methodCompare*” package [27].

## References

- Heradio, R.; Chacon, J.; Vargas, H.; Galan, D.; Saenz, J.; De La Torre, L.; Dormido, S. Open-Source Hardware in Education: A Systematic Mapping Study. *IEEE Access* **2018**, *6*, 72094–72103. [CrossRef]
- Coleman, G.; Salter, W.; Walsh, M. OpenWeedLocator (OWL): An Open-Source, Low-Cost Device for Fallow Weed Detection. *Sci. Rep.* **2022**, *12*, 170. [CrossRef] [PubMed]
- Mesas-Carrascosa, F.J.; Verdú Santano, D.; Meroño, J.E.; Sánchez de la Orden, M.; García-Ferrer, A. Open Source Hardware to Monitor Environmental Parameters in Precision Agriculture. *Biosyst. Eng.* **2015**, *137*, 73–83. [CrossRef]
- Niezen, G.; Eslambolchilar, P.; Thimbleby, H. Open-Source Hardware for Medical Devices. *BMJ Innov.* **2016**, *2*, 78–83. [CrossRef] [PubMed]
- Jolles, J.W. Broad-scale Applications of the Raspberry Pi: A Review and Guide for Biologists. *Methods Ecol. Evol.* **2021**, *12*, 1562–1579. [CrossRef]
- Papavlasopoulou, S.; Giannakos, M.N.; Jaccheri, L. Empirical Studies on the Maker Movement, a Promising Approach to Learning: A Literature Review. *Entertain. Comput.* **2017**, *18*, 57–78. [CrossRef]
- Atkinson, R.K.; Sabo, K.; Conley, Q. The Participatory Web. In *Handbook of Technology in Psychology, Psychiatry and Neurology: Theory, Research, and Practice*; Nova Science Publishers, Inc.: Hauppauge, NY, USA, 2012; pp. 91–120. ISBN 9781621000044.
- Louis, L. Working Principle of Arduino and Using It as a Tool for Study and Research. *Int. J. Control Autom. Commun. Syst.* **2016**, *1*, 21–29. [CrossRef]
- Arduino-Introduction. Available online: <https://www.arduino.cc/en/Guide/Introduction> (accessed on 10 May 2023).
- Oellermann, M.; Jolles, J.W.; Ortiz, D.; Seabra, R.; Wenzel, T.; Wilson, H.; Tanner, R.L. Open Hardware in Science: The Benefits of Open Electronics. *Integr. Comp. Biol.* **2022**, *62*, 1061–1075. [CrossRef]
- Choudhary, P.K.; Nagaraja, H.N. *Measuring Agreement*; Wiley: Hoboken, NJ, USA, 2017; ISBN 9781118078587.
- Etienne, A.; Génard, M.; Lobit, P.; Mbéguié-A-Mbéguié, D.; Bugaud, C. What Controls Fleshy Fruit Acidity? A Review of Malate and Citrate Accumulation in Fruit Cells. *J. Exp. Bot.* **2013**, *64*, 1451–1469. [CrossRef]
- Catrin, T.; Sadler, G.D. PH and Titratable Acidity. In *Food Analysis*; Nielsen, S.S., Ed.; Springer International Publishing: Cham, Switzerland, 2017; pp. 389–406. ISBN 978-3-319-45776-5.
- Lado, J.; Rodrigo, M.; Zacarías, L. Maturity Indicators and Citrus Fruit Quality. *Stewart Postharvest Rev.* **2014**, *10*, 1–6.
- Bermejo, A.; Cano, A. Analysis of Nutritional Constituents in Twenty Citrus Cultivars from the Mediterranean Area at Different Stages of Ripening. *Food Nutr. Sci.* **2012**, *3*, 639–650. [CrossRef]
- Zheng, H.; Zhang, Q.; Quan, J.; Zheng, Q.; Xi, W. Determination of Sugars, Organic Acids, Aroma Components, and Carotenoids in Grapefruit Pulps. *Food Chem.* **2016**, *205*, 112–121. [CrossRef]
- DFRobot SEN0161-V2. Available online: [https://wiki.dfrobot.com/Gravity\\_\\_Analog\\_pH\\_Sensor\\_Meter\\_Kit\\_V2\\_SKU\\_SEN0161-V2](https://wiki.dfrobot.com/Gravity__Analog_pH_Sensor_Meter_Kit_V2_SKU_SEN0161-V2) (accessed on 10 May 2023).
- Maxim Integrated Products DS18B20. Available online: <https://www.analog.com/media/en/technical-documentation/data-sheets/DS18B20.pdf> (accessed on 10 February 2022).
- Adafruit Adalogger Featherwing Overview. Available online: <https://learn.adafruit.com/adafruit-adalogger-featherwing/overview> (accessed on 15 August 2023).
- Philips Sparkfun: Nokia 5110 Lcd. Available online: <https://www.sparkfun.com/products/10168> (accessed on 5 February 2021).
- ADS1115 Ultra-Small, Low-Power, 16-Bit, Analog-to-Digital Converter with Internal Reference. Available online: <https://learn.adafruit.com/adafruit-4-channel-adc-breakouts> (accessed on 18 August 2023).
- DFRobot\_PH Library. Available online: [https://github.com/DFRobot/DFRobot\\_PH](https://github.com/DFRobot/DFRobot_PH) (accessed on 17 March 2023).
- Statistical Methods and Experimental Design for Agreement and Similarity Studies between Open-Source Software and Hardware Devices and Their Corresponding Commercials for Applications in Agriculture—Scripts and Documentation. Available online: <https://github.com/kersee112358/Chapter-5---Ph-logger-sensor> (accessed on 15 November 2023).
- Hanna Instruments Transcat. Available online: <https://www.transcat.com/media/pdf/HI9024.pdf> (accessed on 20 May 2023).
- Carstensen, B.; Simpson, J.; Gurrin, L.C. Statistical Models for Assessing Agreement in Method Comparison Studies with Replicate Measurements. *Int. J. Biostat.* **2008**, *4*, 16. [CrossRef] [PubMed]
- Altman, D.G.; Blandt, J.M. Measurement in Medicine: The Analysis of Method Comparison Studies. *J. R. Stat. Soc. Ser. D Stat.* **1983**, *32*, 307–317. [CrossRef]
- Taffé, P.; Peng, M.; Stagg, V.; Williamson, T. MethodCompare: An R package to assess bias and precision in method comparison studies. *Stat. Methods Med. Res.* **2019**, *28*, 2557–2565. [CrossRef] [PubMed]

28. Choudhary, P.K.; Nagaraja, H.N. Measuring Agreement: Models, Methods, and Applications. Available online: [https://personal.utdallas.edu/~pankaj/agreement\\_book/](https://personal.utdallas.edu/~pankaj/agreement_book/) (accessed on 14 May 2023).
29. Taffé, P. Effective Plots to Assess Bias and Precision in Method Comparison Studies. *Stat. Methods Med. Res.* **2018**, *27*, 1650–1660. [CrossRef]
30. Taffé, P. Assessing Bias, Precision, and Agreement in Method Comparison Studies. *Stat. Methods Med. Res.* **2020**, *29*, 778–796. [CrossRef]
31. Taffé, P.; Halfon, P.; Halfon, M. A New Statistical Methodology Overcame the Defects of the Bland–Altman Method. *J. Clin. Epidemiol.* **2020**, *124*, 1–7. [CrossRef]
32. Lin, L.; Hedayat, A.S.; Wu, W. *Statistical Tools for Measuring Agreement*; Springer: New York, NY, USA, 2012; ISBN 978-1-4614-0561-0.
33. Barnhart, H.X.; Haber, M.J.; Lin, L.I. An Overview on Assessing Agreement with Continuous Measurements. *J. Biopharm. Stat.* **2007**, *17*, 529–569. [CrossRef]
34. Escaramís, G.; Ascaso, C.; Carrasco, J.L. The Total Deviation Index Estimated by Tolerance Intervals to Evaluate the Concordance of Measurement Devices. *BMC Med. Res. Methodol.* **2010**, *10*, 31. [CrossRef]
35. Escaramís, G.; Ascaso, C.; Carrasco, J.L. The Total Deviation Index Estimated by Tolerance Intervals to Evaluate the Concordance of Measurement Devices, R-Script. Available online: [https://static-content.springer.com/esm/art:10.1186/1471-2288-10-31/MediaObjects/12874\\_2009\\_434\\_MOESM1\\_ESM.PDF](https://static-content.springer.com/esm/art:10.1186/1471-2288-10-31/MediaObjects/12874_2009_434_MOESM1_ESM.PDF) (accessed on 10 May 2023).
36. Carrasco, J.L.; Philips, B.R.; Puig-Martinez, J.; King, T.S.; Chinchili, V.M. Estimation of the concordance correlation coefficient for repeated measures using SAS and R. *Comput. Methods Programs Biomed.* **2013**, *109*, 293–304. [CrossRef]
37. Fleiss, J.L. *The Design and Analysis of Clinical Experiments*, 1st ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1986.
38. Carrasco, J.L.; Martinez, J.P. cccrm: Coconcordance Correlation Coefficient. R Package Version 2.1.0. Available online: <https://cran.r-project.org/web/packages/cccrm/index.html> (accessed on 10 December 2023).
39. Datta, D. Blandr: Bland-Altman Method Comparison. 0.5.1. Available online: <https://cran.r-project.org/web/packages/blandr/vignettes/introduction.html> (accessed on 10 December 2023).
40. Grubbs, F.E. On Estimating Precision of Measuring Instruments and Product Variability. *J. Am. Stat. Assoc.* **1948**, *43*, 243. [CrossRef]
41. Mandel, J.; Stiehler, R.D. Sensitivity—A Criterion for the Comparison of Methods of Test. *J. Res. Natl. Bur. Stand. (1934)* **1954**, *53*, 155. [CrossRef]
42. Khan, S.; Saultry, B.; Adams, S.; Kouzani, A.Z.; Decker, K.; Digby, R.; Bucknall, T. Comparative accuracy testing of non-contact infrared thermometers and temporal artery thermometers in an adult hospital setting. *Am. J. Infect. Control.* **2021**, *49*, 597–602. [CrossRef] [PubMed]
43. Bland, J.M.; Altman, D.G. Measuring Agreement in Method Comparison Studies. *Stat. Methods Med. Res.* **1999**, *8*, 135–160. [CrossRef] [PubMed]
44. Bongers, C.C.W.G.; Daanen, H.A.M.; Bogerd, C.P.; Hopman, M.T.E.; Eijsvogels, T.M.H. Validity, Reliability, and Inertia of Four Different Temperature Capsule Systems. *Med. Sci. Sports Exerc.* **2018**, *50*, 169–175. [CrossRef] [PubMed]
45. Pinheiro, J.; Bates, D.; R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. R Package Version 3.1-162. 2022. Available online: <https://cran.r-project.org/web/packages/nlme/nlme.pdf> (accessed on 10 December 2023).
46. José, C.; Pinheiro, D.M.B. *Mixed-Effects Models in S and S-PLUS*; Springer: New York, NY, USA, 2000; ISBN 0-387-98957-9.
47. Nawarathna, L.S.; Choudhary, P.K. A Heteroscedastic Measurement Error Model for Method Comparison Data with Replicate Measurements. *Stat. Med.* **2015**, *34*, 1242–1258. [CrossRef] [PubMed]
48. Bland, J.M.; Altman, D.G. Agreement between Methods of Measurement with Multiple Observations per Individual. *J. Biopharm. Stat.* **2007**, *17*, 571–582. [CrossRef]
49. Hahn, G.J. Statistical Intervals for a Normal Population Part I. Tables, Examples and Applications. *J. Qual. Technol.* **1970**, *2*, 115–125. [CrossRef]
50. Hahn, G.J.; Meeker, W.Q. *Statistical Intervals*; Wiley: Hoboken, NJ, USA, 1991; ISBN 9780471887690.
51. Hothorn, T.; Bretz, F.; Westfall, P. Simultaneous Inference in General Parametric Models. *Biom. J.* **2008**, *50*, 346–363. [CrossRef]
52. Multcomp. R Package. Available online: <https://cran.r-project.org/web/packages/multcomp/index.html> (accessed on 13 December 2023).
53. Carrasco, J.L.; Jover, L. Estimating the Generalized Concordance Correlation Coefficient through Variance Components. *Biometrics* **2003**, *59*, 849–858. [CrossRef] [PubMed]
54. Lehmann, E.L.; Casella, G. *Theory of Point Estimation*; Springer: New York, NY, USA, 1998; ISBN 0-387-98502-6.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.