sensors

MDPI

Article

Laboratory Calibration and Field Validation of Soil Water Content and Salinity Measurements Using the 5TE Sensor

Nessrine Zemni [1,2,*], Fethi Bouksila [1], Magnus Persson [3], Fairouz Slama [2], Ronny Berndtsson [3,4] and Rachida Bouhlila [2]

[1] National Institute for Research in Rural Engineering, Water, and Forestry, Box 10, Ariana 2080, Tunisia; bouksila.fethi@iresa.agrinet.tn

[2] Laboratory of Modelling in Hydraulics and Environment, National Engineering School of Tunis, University of Tunis El Manar (ENIT), Box 37, Le Belvédère Tunis 1002, Tunisia; fairouz.slama@enit.utm.tn (F.S.); rachida.bouhlila@enit.utm.tn (R.B.)

[3] Department of Water Resources Engineering, Lund University, Box 118, SE-221 00 Lund, Sweden; magnus.persson@tvrl.lth.se (M.P.); ronny.berndtsson@tvrl.lth.se (R.B.)

[4] Centre for Middle Eastern Studies, Lund University, Box 201, SE-221 00 Lund, Sweden

* Correspondence: nessrine.zemni@enit.utm.tn; Tel.: +216-28-083-156

Received: 8 October 2019; Accepted: 27 November 2019; Published: 29 November 2019

Abstract: Capacitance sensors are widely used in agriculture for irrigation and soil management purposes. However, their use under saline conditions is a major challenge, especially for sensors operating with low frequency. Their dielectric readings are often biased by high soil electrical conductivity. New calculation approaches for soil water content (θ) and pore water electrical conductivity (ECp), in which apparent soil electrical conductivity (ECa) is included, have been suggested in recent research. However, these methods have neither been tested with low-cost capacitance probes such as the 5TE (70 MHz, Decagon Devices, Pullman, WA, USA) nor for field conditions. Thus, it is important to determine the performance of these approaches and to test the application range using the 5TE sensor for irrigated soils. For this purpose, sandy soil was collected from the Jemna oasis in southern Tunisia and four 5TE sensors were installed in the field at four soil depths. Measurements of apparent dielectric permittivity (Ka), ECa, and soil temperature were taken under different electrical conductivity of soil moisture solutions. Results show that, under field conditions, 5TE accuracy for θ estimation increased when considering the ECa effect. Field calibrated models gave better θ estimation (root mean square error (RMSE) = 0.03 m^3 m^{-3}) as compared to laboratory experiments (RMSE = 0.06 m^3 m^{-3}). For ECp prediction, two corrections of the Hilhorst model were investigated. The first approach, which considers the ECa effect on K′ reading, failed to improve the Hilhorst model for ECp > 3 dS m^{-1} for both laboratory and field conditions. However, the second approach, which considers the effect of ECa on the soil parameter K_0, increased the performance of the Hilhorst model and gave accurate measurements of ECp using the 5TE sensor for irrigated soil.

Keywords: soil salinity; soil water content; FDR sensor; soil pore water electrical conductivity; sensor calibration and validation; real time monitoring

1. Introduction

In arid and semiarid countries, such as Tunisia, irrigation is necessary for improved agricultural production. Water resources with good quality are limited, resulting in the use of low-quality irrigation water. This can induce soil salinization, leading to crop yield reduction, decreasing the agricultural productivity, and causing general income loss [1,2]. Thus, accurate monitoring of soil salinity in time

and space is of great importance for precision irrigation scheduling to save water and avoid soil degradation. Over the last decades, soil dielectric sensors have been developed to measure apparent electrical conductivity (ECa) from which real soil salinity, the soil pore electrical conductivity (ECp), can be estimated [3]. Time domain reflectometry (TDR) has been established as the most accurate dielectric technique to estimate both volumetric water content (θ) and ECp in soils providing automatic, simultaneous, and continuous readings [4]. The efficiency of the TDR method has led to development of other techniques based on similar principles, such as capacitance methods. Some examples are the WET (Delta-T Devices Ltd., Cambridge, UK) and the 5TE (Decagon Devices Inc., Pullman, WA, USA) sensors, both based on frequency domain reflectometry (FDR). Compared to TDR, FDR sensors use a fixed frequency wave instead of a broad-band signal that makes them cheaper and smaller [5]. Dielectric methods are based on determination of apparent soil electrical conductivity (ECa) and soil apparent dielectric permittivity (Ka) [6]. Many models for the relationships between Ka and θ [4,7], ECa-θ, and ECa-ECp-Ka have been proposed in recent research [3,8–10]. However, dielectric properties are affected by physical and chemical soil properties. For example, high ECa affects the wave propagation, leading to errors in the estimation of Ka [11,12]. Thus, it is important to improve θ and ECp prediction models.

Hilhorst [8] presented a theoretical model describing a linear relationship between ECa and Ka to predict ECp. This linear model can be used in a wide range of soil types without soil-specific calibration. Persson [13] evaluated the Hilhorst model using TDR in three sandy soils and confirmed the accuracy of the linear model with significant dependency on soil type. Many researchers [14–17] have tested the Hilhorst model using the WET sensor and showed that it can be improved with soil specific calibration. Using the WET sensor, improved correction of the Hilhorst model was proposed by Bouksila et al. [18], using loamy sand soil with about 65% gypsum. They found that the accuracy of ECp prediction is very poor when using standard soil parameters (K_0). Thus, they proposed a correction by introducing a third-order polynomial fitted to the K_0–ECa relationship instead of using the default K_0. Kargas et al. [6] introduced a linear permittivity corrected model, proposed by Robinson et al. [5], in the Hilhorst relationship. They found that the correction depends on soil characteristics and that it is valid for ECa close to 2 dS m^{-1}. These approaches consider the ECa effect on the prediction of ECp. However, research has not been performed using simultaneous controlled laboratory and field-scale experiments where effects of heterogeneity, root density, insect burrowing, etc., affect the observations [19]. Ideally, sensor calibration should be performed in structured soils due to its importance for pore size distribution and associated matrix potential [20]. Research has shown that calibration in repacked soil columns differs from calibration in disturbed soil used in laboratory experiments [21]. In addition, intrinsic soil factors such as soil temperature, presence of gravel, and microorganisms affect the soil structure and porosity contributing to the variability in ECa and Ka measurements under field conditions as compared to measurements in the laboratory [19].

Nowadays, farmers are embracing precision agriculture using sensors with high accuracy and low cost to increase yields and maintain the sustainability of irrigated land. The 5TE dielectric soil sensor, which also uses the Hilhorst model for ECp estimation, was introduced in 2007 and it is much cheaper than the WET sensor [22]. Several recent studies have investigated the 5TE probe in agricultural applications [2,23,24]. The 5TE sensor has electrodes at the end of the probe that are influenced by soil density making them sensitive to any variation in soil structure and θ content [25]. Despite this fact, most studies on the 5TE sensor performance [16,26,27] have been carried out under laboratory conditions. Thus, almost no research has been done in the field for testing its performance for ECp estimation, neither with the most used linear Hilhorst model nor with the more recent ECp approach proposed in literature. Another important practical aspect is to determine the application range of these sensors for irrigated soils under saline conditions. For example, it is important to determine at what ECa threshold the dielectric losses are no longer negligible and need to be corrected for. Furthermore, there is a lack of understanding of how laboratory calibration can be translated into field conditions.

Thus, the sensors must be calibrated and validated under both conditions in order to assess the errors associated with translating one to the other [28].

In view of the above, the objective of the present study was to assess the performance of the 5TE sensor to estimate soil water content and soil pore electrical conductivity for a representative sandy soil used for cultivation of date palms. Both standard models and a novel approach using corrected models to compensate for high electrical conductivity were used. Results from both field and laboratory experiments were compared. The location of the field experiments was the Jemna oasis, southern Tunisia.

2. Materials and Methods

Soil parameter acronyms, data source, sensor specification and models used in the present work were presented in Appendix A.

2.1. Theoretical Considerations

Any porous medium, such as soils, can be characterized by its permittivity, which is a complex quantity (K) composed of a real part (K′) describing energy storage, and an imaginary part (K″) describing energy loss:

$$K = K' - j\,K'' \quad \text{with } j = \sqrt{-1} \tag{1}$$

For soils with low salinity, it is often assumed that the polarization and conductivity effects can be neglected [4]. Under such conditions, the effect of K″ is eliminated and K′ becomes equal to K, represented by Ka as the apparent dielectric constant [4]. Under saline conditions, the imaginary part of the dielectric permittivity increases with ECa, leading to error in the permittivity measurement. This problem becomes important for frequencies lower than 200 MHz [6]. According to Campbell [29], for a frequency range of 1–50 MHz, conductivity is the most important mechanism related to energy loss. However, using the hydra impedance probe, Kelleners and Verma [30] found that, in general, the total energy loss is related to relaxation loss except for fine sandy soil, where it is equal to zero at 50 MHz.

2.1.1. Permittivity-Corrected Linear Model

Many researchers [5,17,31,32] have studied how well low-frequency capacitance sensors measure Ka and to what degree it is affected by K″. In general, it has been shown that the most important factor to consider is the conductivity effect on Ka, whereas the effect of relaxation losses appears to be small [4,6]. Thus, it is possible to correct the Ka reading by introducing a term for the ECa effect. Based on the work of Whalley [32], Robinson et al. [5] proposed a permittivity-corrected linear model where the theoretical permittivity can be considered equivalent to the refractive index of measurements by the TDR. Robinson et al. [5] conducted experiments using TDR and capacitance dielectric sensor in sandy soils with high ECa levels (up to 2.5 dS m^{-1}) and they proposed a linear model that includes the ECa effect on the Ka prediction according to:

$$\sqrt{K'} = \sqrt{Ka} - 0.628\ ECa \tag{2}$$

From this equation, we notice that the increase of ECa (dS m^{-1}) leads to an increase in Ka. Using Equation (2), a corrected permittivity K′ can be determined eliminating the ECa effect [6].

2.1.2. Water Content Model

The dielectric constant is about 80 for water (at 20 °C), 2 to 5 for dry soil, and 1 for air. Therefore, Ka is highly dependent on θ. Various equations for the Ka vs. θ relationship have been published. The

most used θ-model is a third-order polynomial [4]. However, Ledieu et al. [7] showed that there is a simpler linear relationship for the θ prediction with only two empirical parameters, of the form:

$$\theta = a\sqrt{Ka} + b \quad (3)$$

where a and b are fitting parameters.

Figure 1 shows a schematic of calibration and validation possibilities for θ estimations that were used in the present study. The calibration consisted of fitting of parameters in different models (Figure 1). Optimal values for a and b, vs. a' and b' were determined by linear regression in the relationship $\sqrt{Ka}$-θ_m denoted as the CAL-Ka model (Figure 1, Step-A.1) and $\sqrt{K'}$-θ_m denoted as CAL-Kar model (Figure 1, Step-A.2), respectively. The θ_m was measured in experiments for different salinity levels. The standard Ledieu et al. [7] model (Figure 1) was used for comparison purposes as it is the simplest known model for mineral soil. The different steps (A.1 and A.2) were first completed using laboratory experiments (laboratory calibration) and then using field data (field calibration). The laboratory and field calibrated models were then compared with each other (Figure 1, Step-A.3). Finally, we used field data (step B.1, B.2, and B.3) to validate the laboratory experiments (laboratory model validation).

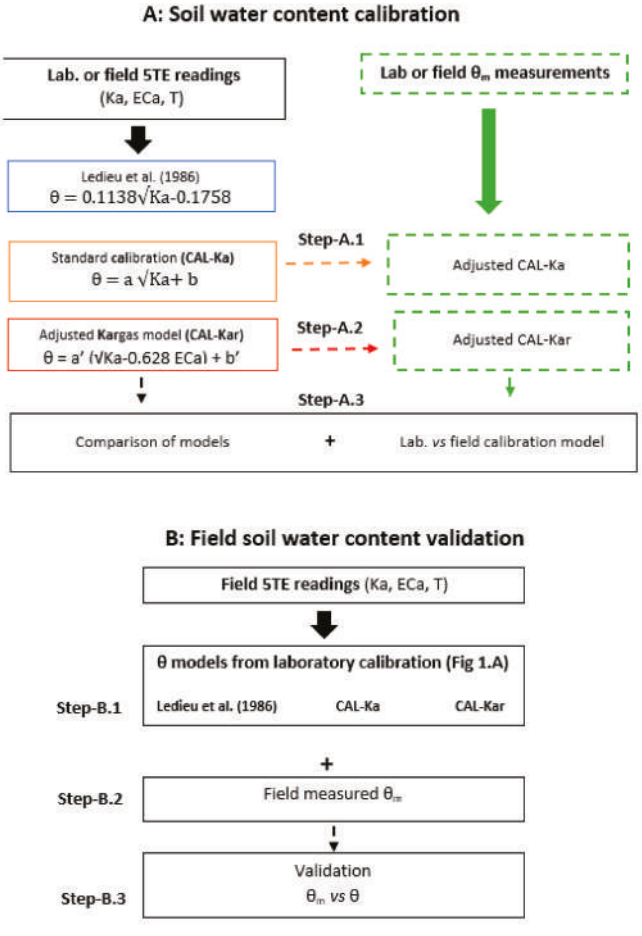


Figure 1. Schematic of θ calibration and validation possibilities investigated in the present study.

2.1.3. Pore Water Electrical Conductivity Model

Different studies [33,34] have shown that ECa depends on both θ and ECp. Malicki et al. [35] and Malicki and Walczak [9] found that for Ka > 6 and when ECp is constant, the relationship between Ka and ECa is linear. An empirical ECp–ECa–Ka model has, thus, been proposed. Based on their results, Hilhorst [8] presented the following equation applicable when $\theta \geq 0.10\ m^3\ m^{-3}$:

$$ECp = \left(\frac{K_w}{(Ka - K_0)}\right) \times ECa \quad (4)$$

where K_w is the dielectric constant of the pore water (equal to 80.3) and K_0 is a soil parameter equal to Ka when ECa = 0 (see [8], for details). According to Hilhorst [8], the K_0 parameter depends on soil texture but is independent of ECa. He found the range of K_0 to be between 1.9 and 7.6. For best results, this should be determined experimentally for each soil type. For most soils, a value of 4.1 has been recommended. One should notice, that in the Hilhorst model (Equation (4)), the Ka, K_w, and K_0 represent the real part of the dielectric constant only. From the linear relationship ECp = f (ECa), the slope that is inversely proportional to ECp and intercept K_0 can be determined.

In the present study, the Hilhorst model (Figure 2, Step-C.1) was tested using varying K_0 soil parameters (4.1, 6 and 3.3). The K_0 = 4.1 is the default value recommended by Hilhorst, K_0 = 6 is the recommended value in the 5TE manual [36] while K_0 = 3.3 is the value measured with distilled water according to the WET sensor manual [37].

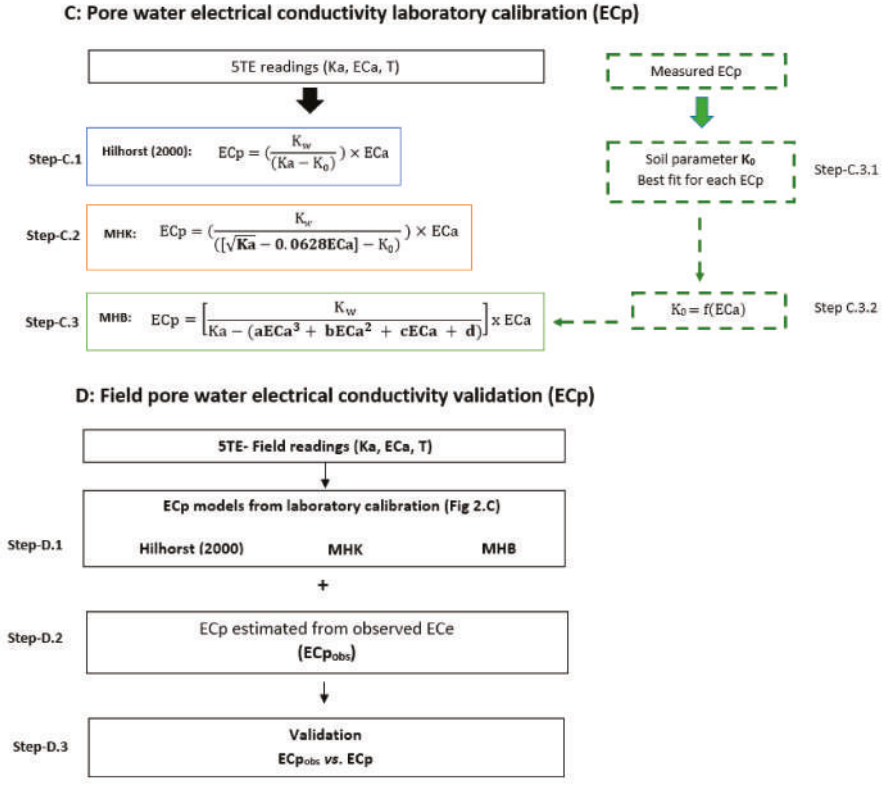


Figure 2. Schematic of electrical conductivity (ECp) calibration and validation used in the present paper.

Inspired by Bouksila et al. [18] and Kargas et al. [6], a modification of the Hilhorst model was investigated. Accordingly, a permittivity-corrected linear equation (Equation (2)) can be introduced in the Hilhorst model (Figure 2, Step-C.2) and ECp is predicted with two different K_0 values (K_0 = 4.1 and K_0 = 3.3). Beside this, the soil fit parameter K_0 is calculated for each salinity level by minimizing the mean square error (MSE) of the estimated ECp in the Hilhorst model (Step-C.3.1). The best fit K_0 parameters are then plotted against ECa for the seven different ECp and a third-order polynomial function is determined (Step-C.3.2), and introduced in the Hilhorst model (Step-C.3). Finally, we used field data (step D.1, D.2, and D.3) to validate the laboratory experiments (laboratory model validation).

The temperature is an important factor influencing the electrical conductivity measurements; indeed, all ECa reading were adjusted in the present work using Equation (5). Besides, during experiments the temperature effect on K_w parameter was considered using the recommended temperature correction equation in the 5TE manual [36].

$$ECa_{25} = ECa\ [1 - ((T - 25)\ \times\ 0.02)] \quad (5)$$

Measured Ka, ECa, and T in laboratory and field experiments are converted to ECp using the Hilhorst [8] model (Step-C.1), Kargas et al. [6] approach (Step-C.2), and Bouksila et al. [18] approach (Step-C.3), denoted as H, MHK, and MHB, respectively.

The different approaches in Figures 1 and 2 have not been tested before using the 5TE sensor. The approaches CAL-Kar, MHK and MHB have previously only been tested once under controlled laboratory condition using the WET sensor. The novelty of the present work is to validate these approaches under field condition using the low cost capacitance sensor 5TE. In addition, the MHB approach developed by Bouksila et al. [18], used an experimentally determined $K_0 = f(ECa)$ relationship. Our new approach instead uses a K_0 derived from best-fit parameter for each ECp level, which make the application of MHB approach much easier since there is no need for the K_0 laboratory experiment.

Model performance for θ and ECp, was evaluated using both the root mean square error (RMSE) and coefficient of determination (R^2). In addition, mean relative error (MRE) and coefficient of variation (CV) were used for ECp and θ, respectively.

2.2. Study Area

The field study was conducted in the Jemna oasis (33°36'15."N, 9°00'39."E), belonging to the Agricultural Extension and Training Agency (AVFA) located in the Kebeli Governorate, southern Tunisia. The oasis is equipped with a micro-irrigation system. The main crop is adult date-palm trees. The climate is arid with an annual rainfall of less than 100 mm, which is insufficient to sustain agriculture. The annual potential evapotranspiration is about 2000 mm [38]. Groundwater, situated at 17 m soil depth, with an electrical conductivity (ECiw) of about 3.5 dS m^{-1}, is used for irrigation. The pH of groundwater is 7.8 and the geochemical facies is sodium chloride. Soil samples were collected from the top soil at 0–0.5 m depth. The soil was leached with distilled water in order to remove soluble salts and oven dried (105 °C) for 24 h. Then, the soil was passed through a 2 mm sieve. Soil particle size distribution was determined using the sedimentation method (pipette and hydrometer) and the electrical conductivity of saturated soil paste extract (ECe) was measured according to the United States Department of Agriculture (USDA) [39]. A summary of soil properties is presented in Table 1.

Table 1. Particle size percentage, pH and electrical conductivity of saturated soil paste extract (ECe) of investigated soil samples.

Depth (m)	Clay (%)	Fine Silt (%)	Coarse Silt (%)	Fine Sand (%)	Coarse Sand (%)	pH	ECe (dS m^{-1})
0–0.5	5	3	4	22	65	8.5	1.8

2.3. Laboratory Experiments

Seven NaCl solutions with different electrical conductivity (0.02, 0.2, 0.5, 3.6, 5.3, 7.2, and 8.2 dS m^{-1}) were prepared for the infiltration experiments. The soil was initially mixed with a small amount (about 0.05 m^3 m^{-3}) of the same water as used in the infiltration experiments to prevent water repellency. The soil was repacked into a plexiglas soil columns, 0.12 m in diameter and 0.15 m long (Soil Measurement System, Tucson, Arizona), to the average dry bulk density encountered in the field (about 1450 kg m^{-3}).

The 5TE sensor was used for observations [23]. It is a multifunctional sensor measuring Ka, ECa, and T (for more details, see Appendix A). The measuring frequency is 70 MHz and it is a three-rod type sensor with 0.052 m long prongs and 0.01 m spacing between adjacent prongs [23,40]. The 5TE probe was inserted vertically in the center of the column. Upward infiltration experiments were carried out by stepwise pumping a known volume of a NaCl solution (45 mL) with a precise syringe pump from the bottom of the column. Twenty minutes after each injection, three measurements of Ka, ECa, and temperature were taken and averaged. This procedure was repeated until saturation (0.40 m^3 m^{-3}) was reached. Four hours after reaching saturation, measurements were again taken and pore water was extracted from the bottom of the column with a manual vacuum pump. Electrical conductivity of extracted pore water ECp_m was measured with a conductivity meter. In total, seven upward infiltration experiments were conducted, one for every NaCl solution.

2.4. Field Measurements

Four 5TE sensors were installed between date-palm trees at four soil depths (0.10, 0.15, 0.30, and 0.45 m). The 5TE probes were connected to a Decagon Em50 data logger. The DataTrac3 software version 3.15 [23] was used to download collected data from the Em50. Volumetric soil water content and pore electrical conductivity were estimated using standard parameters of the Ledieu et al. [7] and Hilhorst [8] models, respectively. In addition, soil samples were taken by hand auger at the same depth of sensor installation on 24 April and 3 October 2018. Gravimetric water content θ_m and electrical conductivity of saturated soil paste extract (ECe) were measured in laboratory according to USDA standards. The soil dry bulk density (Bd) was measured in the field using the cylinder method at five soil depths (0.1 m depth intervals to 0.5 m). During April 2018, the average soil Bd was equal to 1.43 g cm^{-3} and varied from 1.3 to 1.6 g cm^{-3}.

3. Results

3.1. Soil Water Content

Figure 3 presents the relationship between Ka and observed θ_m with different salinity levels (ECp, dS m^{-1}) measured during the upward infiltration experiments. For largest ECp, ECa did not exceed 2.5 dS m^{-1}. It is seen that ECa considerably affects the Ka readings, especially for high ECp. This can lead to significant errors for both Ka and ECa, indicating that 5TE probe readings need to be corrected when used in saline soils. The overestimation of Ka as ECa increases has been described by several authors (e.g., [19,27]).

In Figure 4, Ka and K′ (corrected with Equation (2)) for two ECp levels (3 and 9.8 dS m^{-1}) are plotted against measured θ_m. K′ values are very close to Ka when ECp ≤ 3 dS m^{-1}, especially at low θ ($\theta \leq 0.15$ m^3 m^{-3} and ECa ≤ 0.43 dS m^{-1}). However, for ECp = 9.8 dS m^{-1}, the difference between Ka and K′ is more pronounced, especially for $\theta \geq 0.15$ m^3 m^{-3} and ECa ≥ 0.75 dS m^{-1}.

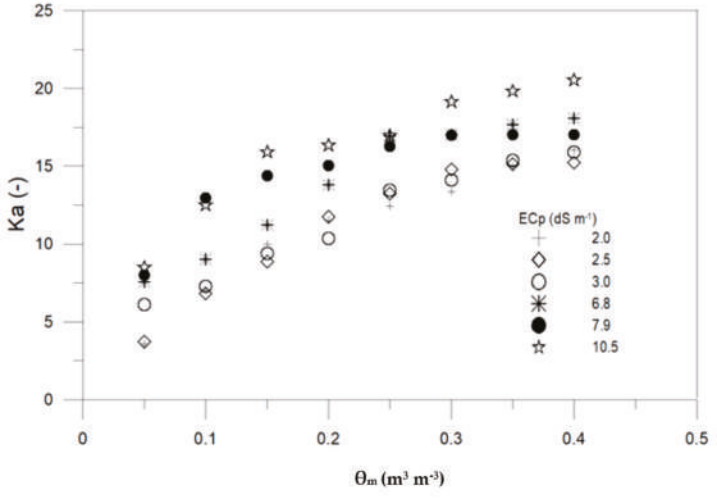


Figure 3. Apparent dielectric permittivity (Ka) vs. measured volumetric water content (θ_m) for various pore electrical conductivity (ECp) levels (dS m^{-1}).

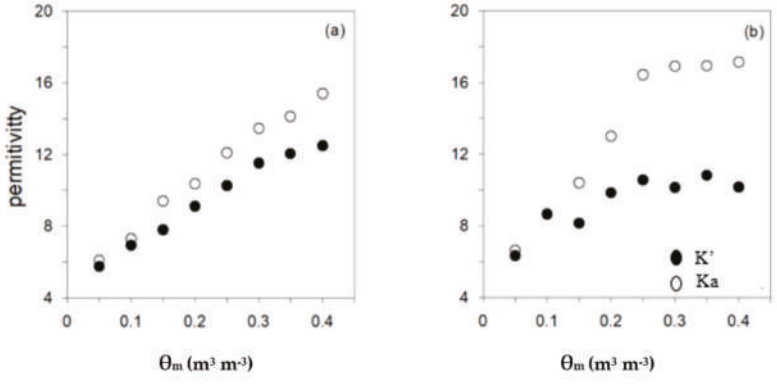


Figure 4. Relationship Ka-θ_m (open circles) and K′-θ_m (filled circles) using the 5TE sensor for ECp = 3 dS m^{-1} (**a**) and ECp = 9.8 dS m^{-1} (**b**).

The calibrated parameters using laboratory data for CAL-Ka and CAL-Kar approaches are presented in Table 2. For all models tested under laboratory conditions, RMSE increased with ECp. Soil water content from CAL-Kar approach matched well measured θ_m for ECp ≤ 3 dS m^{-1} (ECa < 0.7 dS m^{-1}) and gave the best θ estimation compared to the Ledieu et al. [13] model and the soil-specific calibration CAL-Ka. However, for ECp ≥ 6.8 dS m^{-1}, the CAL-Ka approach gave lower RMSE compared to the CAL-Kar model. For high ECp (≥ 6.8 dS m^{-1}), the performance of the CAL-Kar model deteriorated.

Table 2. Root mean square error (RMSE, m^3 m^3), determination coefficient (R^2) and coefficient of variation (CV,%) of estimated soil water content using Ledieu et al. [7], standard calibration (CAL-Ka) and permittivity corrected model (CAL-Kar) for different water pore electrical conductivity (ECp).

Laboratory Calibration				
ECp (dS m^{-1})		**Ledieu et al. (1986)**	**CAL-Ka**	**CAL-Kar**
	Fit	Equation (4)	$\theta = 0.16\sqrt{Ka^1} - 0.30$	$\theta = 0.18\sqrt{K'^2} - 0.33$
ECp ≤ 3	RMSE	0.06	0.05	0.04
	R^2	0.93	0.95	0.95
ECp = 6.8	RMSE	0.08	0.06	0.10
	R^2	0.73	0.87	0.50
6.8 < ECp ≤ 10.5	RMSE	0.09	0.07	0.13
	R^2	0.77	0.85	0.39
Mean RMSE		0.08	0.06	0.09
Mean R^2		0.8	0.9	0.6
CV (%)		26.5	20	19.8
Field calibration				
	Fit		$\theta = 0.15\sqrt{Ka} - 0.26$	$\theta = 0.20\sqrt{K'} - 0.37$
ECa[3] ≤ 0.7 and 1.7 ≤ ECe[4] ≤ 4.1	RMSE (m^3 m^{-3})	-	0.04	0.03
	R^2	-	0.94	0.97
	CV (%)	-	23	24
Field validation				
ECa ≤ 0.7 and 1.7 ≤ ECe ≤ 4.1	RMSE (m^3 m^{-3})	0.1	0.060	0.060
	R^2	0.80	0.88	0.97
	CV (%)	27	21	24

[1] Apparent soil permittivity, [2] Corrected apparent soil permittivity, [3] Soil apparent electrical conductivity, [4] Electrical conductivity of saturated soil paste extract.

3.2. Field Validation of Soil Water Content Models

During field experiments, Ka measured by the four 5TE probes varied from 6.5 to 11, ECa from 0.17 to 0.75 dS m^{-1}, and measured soil moisture (θ_m) from 0.10 to 0.24 m^3 m^{-3}. According to R^2 of field validation results (Table 2), the best model to predict θ under field conditions is CAL-Kar followed by CAL-Ka. However, RMSE analysis indicates that there is no significant difference between observed and estimated θ using both approaches, implying that both predicted θ accurately for ECa ≤ 0.7 dS m^{-1}.

From Figure 5, a slight underestimation of the different models is observed and this is more pronounced for the Ledieu et al. [7] model. The underestimation can be related to adsorbed water, resulting in a lower amount of mobile water in the soil, thus reducing the Ka readings (detection) by the 5TE sensor and eventually resulting in underestimation of Ka [41,42]. The difference between observed and predicted θ may also be attributed to variability in soil structure, bulk density, presence of stones, roots, and other inert material in the core samples. The difference may also be linked to the spatial variability of θ between sampled and monitored soils. Similar findings have been reported for mineral soils using the 5TE sensor [41], for Luvisol using the 5TM capacitance sensor [42], and using the ECH2O sensor in sandy soil [43]. The success of CAL-Ka and CAL-Kar models to calculate θ at field conditions is closely linked to the low range of ECa data measured by the 5TE sensor, below 0.7 dS m^{-1}, during the period of investigation.

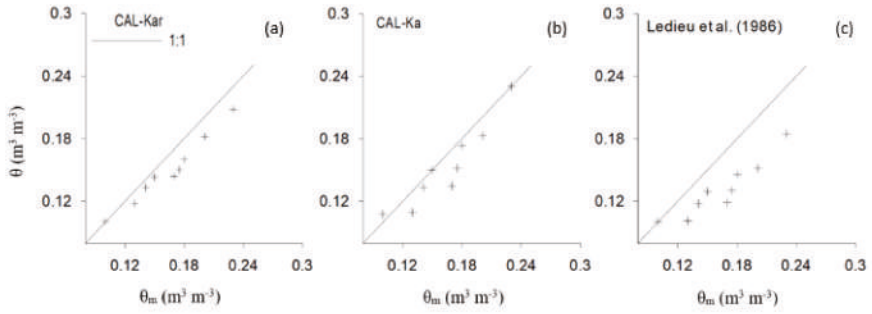


Figure 5. Estimated soil water content (θ) vs. measured (θ_m) using CAL-Kar approach (**a**), CAL-Ka approach (**b**) and Ledieu et al. [13] model (**c**) under field conditions, solid line gives the 1:1 relationship.

For the same range of soil salinity, RMSE was higher for the field as compared to laboratory data. For laboratory experiments, soil was crushed, washed, and passed through a 2 mm sieve. This means that its structure was changed as well as the pore size distribution, and some of the organic matter may have been removed. This allows more mobile water compared to field conditions [44]. As well, for field conditions, observed Bd profiles are not uniform and may vary with time. In contrast to the controlled laboratory experiments (e.g., constant Bd), the field Bd spatial and temporal variation will induce an additional error when laboratory models are used to estimate θ.

We used the field data to calibrate the CAL-Ka and CAL-Kar models, the calibrated parameters for the models are presented in Table 2 (Field calibration). The RMSE decreased from 0.06 to 0.04 m^3m^{-3} and from 0.06 to 0.03 m^3 m^{-3} for CAL-Ka and CAL-Kar, respectively. Thus, the CAL-Kar approach gave better field predictions of θ. Similarly, Kinzli et al. [45] reported that field calibration was most successful for sandy soils. According to this finding, we may support the earlier conclusion that the permittivity corrected (CAL-Kar) model is recommended under field conditions if ECa is below 0.75 dS m^{-1}. However, the Ledieu et al. [7] model cannot be used safely under field conditions in the case when soil specific calibration is not available.

3.3. Soil Pore Electrical Conductivity (ECp)

3.3.1. ECp Laboratory Calibration

Table 3 presents the RMSE for the different models. All models showed good performance in the 0–3 dS m^{-1} range, except Hilhorst with (K_0 = 6) and MHK with K_0 = 4.1. Moreover, RMSE results (Table 3), showed an increase of the range of default H model validity until ECp = 6.8 dS m^{-1}. This finding can be linked to the higher operating frequency of 5TE (70 MHz) compared to the capacitance sensor used by Hilhorst (30 Mhz). Hilhorst reported that the model assumption ceases to be accurate at higher salinity as ECp significantly deviates from that of free water.

Table 3. Root mean square error (RMSE, dS m^{-1}) of estimated pore electrical conductivity (ECp) using Hilhorst (K_0 = 4.1, 3.3, and 6), modified Hilhorst according to Kargas et al. [6] (MHK) (K_0= 4.1 and 3.3), and modified Hilhorst according to Bouksila et al. [18] (MHB) models.

ECp (dS m^{-1})	**Hilhorst (2000)**			**MHK**		**MHB**
Soil Parameter-K_0	**K_0 = 4.1**	**K_0 = 3.3** [1]	**K_0 = 6**	**K_0 = 4.1**	**K_0 = 3.3** [1]	**Best Fit K_0 = f (ECa^2)**
ECp ≤ 3	0.29	0.14	0.83	0.88	0.34	0.044
ECp = 6.8	0.57	0.21	1.7	6.3	3.8	0.050
6.8 < ECp ≤ 10.5	1.48	0.99	3.06	-	-	0.054

[1] K_0 soil parameter determined experimentally according to the method in the Wet sensor manual using distilled water. [2] Soil apparent electrical conductivity.

From the results presented in Table 3, the ECp limit for accurate measurements seems to be 6.8 dS m^{-1}. Similar results were reported by Scudiero et al. [40], using the 5TE sensor and ECp limit <10 dS m^{-1} with RMSE equal to 0.68 dS m^{-1}. Using the H model with K_0 value recommended in the Decagons manual (K_0 = 6) showed a larger RMSE for all salinity levels compared the default parameter (K_0 = 4.1). The H model with K_0 = 3.3 (determined experimentally according to the WET manual) gave better results for the three salinity ranges. Persson [13] stated that the H model using a fitted soil parameter gives ECp values statistically similar to other model results (e.g., [3,10,46]).

Focusing on the modified Hilhorst model using the MHK approach with K_0 = 4.1, one can observe that the RMSE is at maximum, especially for ECp ≥ 6.8 dS m^{-1}. Kargas el al. [6] validated this approach using a lower salinity level (ECp ≤ 6 dS m^{-1}). According to our results (Figure 7), an overestimation of the H model, especially at ECp ≥ 3 dS m^{-1}, is observed. Similarly, Visconti et al. [19] showed an overestimation of ECp in the range of 0–10 dS m^{-1} and Scudiero et al. [40] showed an overestimation of ECp in the range 3–10 dS m^{-1}, both working with the 5TE sensor and the H model. In the present study, the H model overestimated ECp, thus using the MHK approach will not improve results.

The observed overestimation by the H model might be due to K_0, which was assumed to be equal to 4.1. In addition, one should note that the H model does not consider solid particle surface conductivity, which could contribute to the ECp error [17]. From Table 3, decreasing K_0 from 4.1 to 3.3 for both the H and MHK model leads to a significant decrease of RMSE, two times lower than the default. The H model seems to be more dependent on the soil parameter K_0 than on Ka and ECa.

K_0 estimated from the best fit approach for the different salinity levels is plotted against ECa in Figure 6. The K_0 range varied between 1.29 and 3.2 with a mean of 3.0, which is similar to the K_0 determined experimentally using distilled water (K_0 = 3.3).

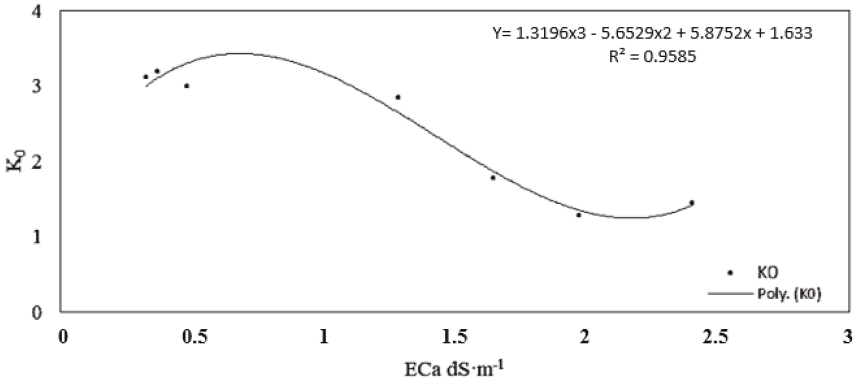


Figure 6. Best fit soil parameter (K_0) vs. bulk soil electrical conductivity (ECa).

At saturation, ECa was equal to 0.32 dS m^{-1} and 2.4 dS m^{-1} and Ka was equal to 15 and 19 for the lowest (2 dS m^{-1}) and the highest (10.5 dS m^{-1}) observed ECp, respectively. According to Figure 6, K_0 decreases with increasing salinity. Similar to [18], our results showed that K_0 is not constant, but depends on ECa and that a third-order polynomial fitted the K_0–ECa relationship rather well (R^2 ≥ 0.95). K_0 = f (ECa) in Figure 6, was used in the H model to predict ECp. Compared to the H model, for the individual ECp levels, using the MHB model, RMSE decreased significantly.

Figure 7 shows observed and predicted ECp using the H model with three different K_0 and the MHK and MHB approaches, respectively. All model performances, are approximately the same for ECp ≤ 3 dS m^{-1}, except when using K_0 = 6 and K_0 = 4.1 for H and MHK models, respectively.

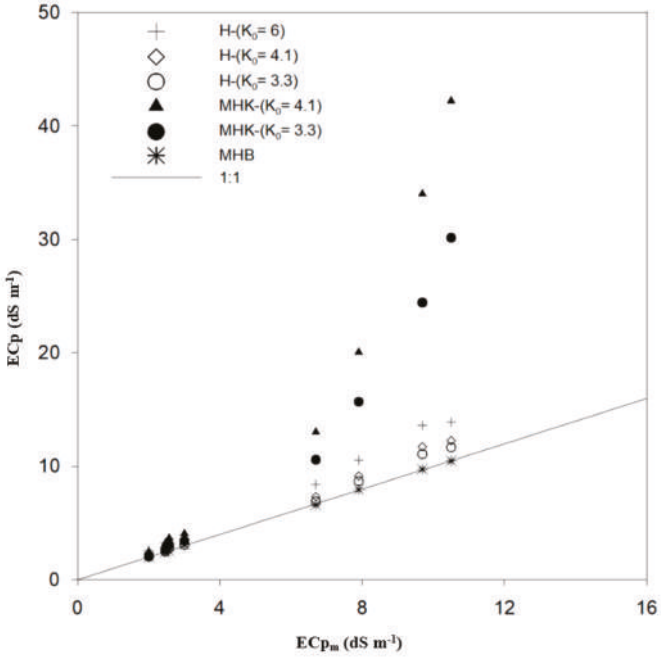


Figure 7. Estimated pore electrical conductivity (ECp) vs. measured for different model tested for laboratory conditions.

Based on the laboratory results, the MHB approach improved the H model and gave accurate estimation of ECp with R^2 = 0.99 for all salinity levels. Thus, for high soil salinity (6.8 dS m^{-1} ≤ ECp ≤ 10.5 dS m^{-1}), the MHB approach is recommended for achieving optimal accuracy of ECp measurements. For lower ECp (≤3 dS m^{-1}), the standard H model is sufficient. For high ECp, the MHK approach failed to reproduce the observed ECp correctly and the approach is not recommended based on the results of our study. Further studies for different soil types are needed so that this combined approach in predicting ECp can be validated.

3.3.2. Field Validation of ECp Models

Unfortunately, we do not have field observed ECp to validate and statistically compare the different models. Instead, we determined a linear relationship (ECp = f (ECe)) for different calculated ECp, using the H, MHK, and MHB models and 5TE measurements, with observed field ECe. Several researchers have studied relationships between ECe and ECp, e.g., [3], showing that the relationship is strongly linear. The relationship (ECp = f (ECe)) with the highest R^2 = 0.9 was chosen to predict the field ECp values (ECp_{obs}). During the investigation period, ECe was determined from soil samples, according to the USDA standard (collected at the same depth as the location of the 5TE sensors), ranging between 1.7 and 4.1 dS m^{-1}. The relatively low soil salinity is due to a rainfall observed in the field one day before soil sampling.

The observed ECp_{obs} obtained from the best fit relationship is plotted against the estimated ECp for the different models in Figure 8. The H model with K_0 = 6.6 was not included in the figure since it gave out of range values. The ECp estimation with MHB approach appears uniformly scattered about the 1:1 line. On the other hand, the H model with K_0 = 3.3 shows a cloud of points near the 1:1 line.

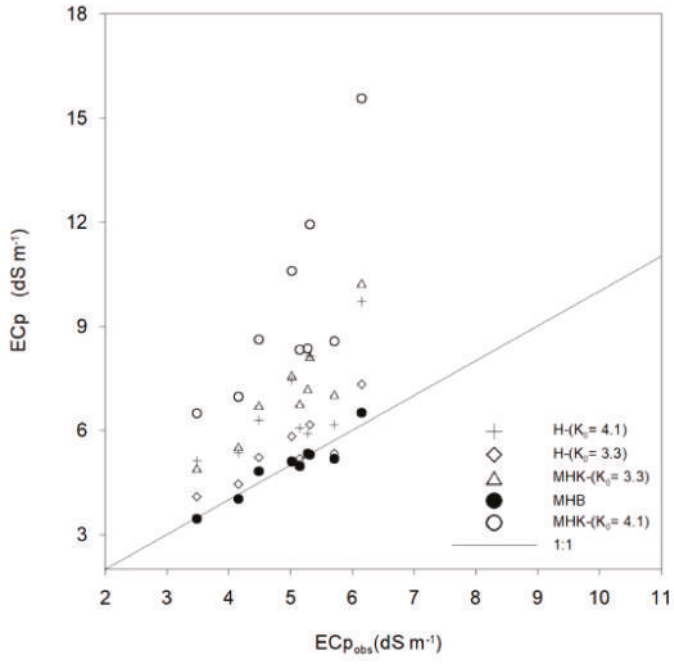


Figure 8. Estimated ECp vs. observed under field conditions.

Compared to laboratory results, for the same ECa range (ECa ≤ 0.7 dS m^{-1}) (Table 4), observed errors are higher for the field validation. The RMSE increased for all models. Errors are mainly related to a number of factors absent in the laboratory but present under field conditions. Due to this reason, a methodological approach composed by laboratory calibration and field validation is optimal.

Table 4. Root mean square error (RMSE, dS m^{-1}) and determination coefficient (R^2) of Hilhorst (K_0 = 4.1, 3.3, and 6), modified Hilhorst according to Kargas et al. [6] (MHK) (K_0 = 4.1 and 3.3) and modified Hilhorst according to Bouksila et al. [18] (MHB) models field validation.

	Hilhorst (2000)		**MHK**		**MHB**	
ECa2 ≤ 0.7 and 1.7 ≤ ECe3 ≤ 4.1	**K_0 = 4.1**	**K_0 = 3.3[1]**	**K_0 = 6**	**K_0 = 4.1**	**K_0 = 3.3[1]**	**Best Fit K_0 = f (ECa)**
RMSE (dS m−1)	0.82	0.70	10	1.8	1.34	0.30
R^2	0.53	0.73	0.26	0.56	0.77	0.90

[1.] K_0 soil parameter determined experimentally according to the method in the Wet sensor manual using distilled water. [3] Soil apparent electrical conductivity, [4] Electrical conductivity of saturated soil paste extract.

The MHB approach presents a significant improvement of the H model, especially at high ECp (Table 4). The H and MHK model fit is acceptable for field and laboratory conditions only for ECp ≤ 3dS m^{-1} while the MHB approach is acceptable for field conditions and it can be safely used for sandy soil and ECp ≤ 7 dS m^{-1}.

Since variation and uncertainties in the field are higher, it is recommended to validate the calibrated models with field data. According to our results, the H model with K_0 = 6 is not recommended either with laboratory nor field data. However, the reduction of K_0 to 3.3 increased the performance of the model and it can be safely used for ECp < 3 dS m^{-1}. For ECp > 3 dS m^{-1}, the MHK approach did not

improve the H model with RMSE more than 1 dS m^{-1} and it is not recommended. Thus, for achieving optimal accuracy of ECp measurements, the MHB approach is recommended for ECp ≤ 7dS m^{-1}.

4. Conclusions

In this study, the 5TE sensor performance for volumetric soil water content (θ) and soil pore electrical conductivity (ECp) estimation was investigated under laboratory and field conditions. First, two procedures for θ estimation based on a linear relationship of √Ka-θ_m (CAL-Ka approach) and √K'-θ_m (CAL-Kar approach) were investigated. Using the CAL-Kar approach, the effect of soil apparent electrical conductivity (ECa) on the real part of the complex dielectric permittivity (K') was considered. In addition, the Ledieu et al. [7] relationship was used for comparison purposes. A site-specific validation of CAL-Ka and CAL-Kar models using 5TE field subset data and θ from soil samples at different depth was performed. Secondly, 5TE performance for soil salinity assessment was investigated using the H linear model according to correction proposed by Kargas et al. [6] (MHK model), and Bouksila et al. [17] (MHB model). The default value of soil parameter K_0 = 4.1 and K_0 = 6 recommended in the 5TE manual was used for comparison.

For soil water content, calibration considering the ECa effect on K' increased the performance of the 5TE sensor under field conditions for ECa ≤ 0.75 dS m^{-1} (R^2 = 0.97, RMSE = 0.06 m^3 m^{-3}). However, the error in predicting θ was highest (0.10 m^3 m^{-3}) when the Ledieu et al. [7] model was used. Indeed, this model cannot be safely used under field conditions. Thus, we conclude that field calibration of the 5TE sensor is recommended for accurate soil water content estimation. Soil pore electrical conductivity calibration results, show that the 5TE sensor limit using the default H model is equal to 6.8 dS m^{-1} with RMSE = 0.57 dS m^{-1} and MRE = 9%. The 5TE sensor manual value (K_0 = 6) is not recommended. However, K_0 = 3.3 increases model performance over the investigated salinity range. The MHK approach, introducing the permittivity correction in the H model, failed to reproduce the observed ECp correctly and it is not recommended. In the next step, considering the effect of ECa on the K_0 soil parameter in the H model (MHB approach), it was found that the standard model improves and gives accurate estimation of ECp with R^2 equal to 0.99 for all salinity levels. Under field conditions, the MHB approach gives the best results for sandy soils.

It is a challenge to perform real-time monitoring of irrigated land under high-saline conditions to provide sustainable agriculture and farmer income increase. Using θ and ECp observations, it was shown that a methodological approach composed of a laboratory calibration and field validation is necessary. Further studies, for different soil types, are needed to validate this combined approach in predicting ECp.

Author Contributions: N.Z. was the main author executing the experiments, data curation, formal analysis and writing. F.B. assisted in the execution of experiments. F.B. and M.P. contributed in data curation, formal analysis and writing original draft. Investigation was carried by N.Z., F.B., and F.S. R.B. (Ronny Berndtsson), F.S. and R.B. (Rachida Bouhlila) provided advice and assisted in reviewing and editing the final document. Funding acquisition and resources were made by F.B., M.P. and R.B. (Ronny Berndtsson). F.B., M.P. and R.B. (Rachida Bouhlila) supervised this work. All authors provided assistance in reviewing and editing the manuscript. All authors contributed to the conceptualization, methodology and validation of the work.

Funding: This research was funded by the Tunisian Institution of Agricultural Research and Higher Education (IRESA) through the SALTFREE project (ARIMNET2/0005/2015, grant agreement N° 618127) and the European Union Horizon 2020 program, under Faster project, grant agreement N° [810812].

Acknowledgments: The authors acknowledge support received from Soil department (DGACTA, Tunisia).

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data, in the writing the manuscript, or in the decision to publish the results.

Appendix A

Table A1. Soil parameter acronyms, data source, sensor specification and models used in the present work.

Soil Parameter	Acronym	Data Source	Sensor/Method
Soil dry bulk density	Bd	Measured	Cylinder method- United States Department of Agriculture (USDA)
Soil pH	pH	Measured	pH-meter
Apparent soil permittivity	Ka	Measured	5TE-probe
Soil parameter	K_0	Estimated	5TE-probe
Dielectric constant of pore water	K_w	Estimated	5TE-probe
Corrected apparent soil permittivity	K′	Estimated	5TE-probe
Soil temperature	T	Measured	5TE-probe
Electrical conductivity of saturated soil paste extract	ECe	Measured	EC-meter/USDA method
Soil apparent electrical conductivity	ECa	Measured	5TE-probe
Irrigation water electrical conductivity	ECiw	Measured	EC-meter
Measured soil water content	θ_m	Measured	Gravimetric method-USDA
Estimated volumetric water content	θ	Estimated	θ –Models (see Figure 1)
Laboratory measured pore water electrical conductivity	ECp_m	Measured	EC-meter
Field observed pore water electrical conductivity	ECp_{obs}	Measured	ECp_{obs} = a ECe + b (see Figure 2)
Pore water electrical conductivity	ECp	Estimated	ECp-Models (see Figure 2)
5TE sensor specification			
Type	Specifics		
Sensor type	FDR (Frequency Domain Reflectometry)		
Power supply	+3.6 to +15 V		
Frequency	70 MHz		
Size	Length 10.9 cm (4.3 in) Width 3.4 cm (1.3 in) Height 1.0 cm (0.4 in)		
Measurement volume	300 cm^3		
Direct output data	Ka, ECa, and T		
Indirect output data	θ and ECp		
Range (Ka, ECa)	1–80, 0–7 dS m^{-1}		
Resolution (Ka, ECa)	0.1, 0.01 dS m^{-1}		
Accuracy (Ka, ECa)	±3%, ±10%		
Models			
CAL-Ka (see Figure 1)	Calibration of soil water content model without permittivity correction		
CAL-Kar (see Figure 1)	Calibration of soil water content model with permittivity correction according to Kargas et al. (2017)		
H (see Figure 2)	Standard Hilhorst (2000) model for ECp prediction		
MHK (see Figure 2)	Modified Hilhorst model according to Kargas et al. (2017) for ECp prediction		
MHB (see Figure 2)	Modified Hilhorst model according to Bouksila et al. (2008) for ECp prediction		
Model performance statistic tool			
RMSE	Root Mean Square Error		
R^2	Coefficient of determination		
MRE	Mean Relative Error		
CV	Coefficient of Variation		

References

1. Selim, T.; Bouksila, F.; Berndtsson, R.; Persson, M. Soil Water and Salinity Distribution under Different Treatments of Drip Irrigation. *Soil Sci. Soc. Am. J.* **2013**, *77*, 1144–1156. [CrossRef]
2. Slama, F.; Zemni, N.; Bouksila, F.; De Mascellis, R.; Bouhlila, R. Modelling the Impact on Root Water Uptake and Solute Return Flow of Different Drip Irrigation Regimes with Brackish Water. *Water* **2019**, *11*, 425. [CrossRef]
3. Rhoades, J.D.; Manteghi, N.A.; Shouse, P.J.; Alves, W.J. Soil Electrical Conductivity and Soil Salinity: New Formulations and Calibrations. *Soil Sci. Soc. Am. J.* **1989**, *53*, 433–439. [CrossRef]
4. Topp, G.C.; Davis, J.L.; Annan, A.P. Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. *Water Resour. Res.* **1980**, *16*, 574–582. [CrossRef]
5. Robinson, D.A.; Gardner, C.M.K.; Cooper, J.D. Measurement of relative permittivity in sandy soils using TDR, capacitance and theta probes: Comparison, including the effects of bulk soil electrical conductivity. *J. Hydrol.* **1999**, *223*, 198–211. [CrossRef]
6. Kargas, G.; Persson, M.; Kanelis, G.; Markopoulou, I.; Kerkides, P. Prediction of Soil Solution Electrical Conductivity by the Permittivity Corrected Linear Model Using a Dielectric Sensor. *J. Irrig. Drain. Eng.* **2017**, *143*, 04017030. [CrossRef]
7. Ledieu, J.; Ridder, P.D.; Clerck, P.D.; Dautrebande, S. A method of measuring soil moisture by time-domain reflectometry. *J. Hydrol.* **1986**, *88*, 319–328. [CrossRef]
8. Hilhorst, M.A. A Pore Water Conductivity Sensor. *Soil Sci. Soc. Am. J.* **2000**, *64*, 1922–1925. [CrossRef]
9. Malicki, M.A.; Walczak, R.T. Evaluating soil salinity status from bulk electrical conductivity and permittivity. *Eur. J. Soil Sci.* **1999**, *50*, 505–514. [CrossRef]
10. Mualem, Y.; Friedman, S.P. Theoretical Prediction of Electrical Conductivity in Saturated and Unsaturated Soil. *Water Resour. Res.* **1991**, *27*, 2771–2777. [CrossRef]
11. Dalton, F.N. Development of Time-Domain Reflectometry for Measuring Soil Water Content and Bulk Soil Electrical Conductivity. In *Advances in Measurement of Soil Physical Properties: Bringing Theory into Practice*; Topp, G.C., Reynolds, W.D., Green, R.E., Eds.; Soil Science Society of America: Madison, WI, USA, 1992; pp. 143–167. [CrossRef]
12. Nadler, A.; Gamliel, A.; Peretz, I. Practical Aspects of Salinity Effect on TDR-Measured Water Content A Field Study Contribution from the Agricultural Research Organization, Volcani Center, Bet Dagan, 50-250, Israel; No 611/98 1998 series. *Soil Sci. Soc. Am. J.* **1999**, *63*, 1070–1076. [CrossRef]
13. Persson, M. Evaluating the linear dielectric constant-electrical conductivity model using time-domain reflectometry. *Hydrol. Sci. J.* **2000**, *47*, 269–277. [CrossRef]
14. Hamed, Y.; Samy, G.; Persson, M. Evaluation of the WET sensor compared to time domain reflectometry. *Hydrol. Sci. J.* **2006**, *51*, 671–681. [CrossRef]
15. Inoue, M.; Ould Ahmed, B.A.; Saito, T.; Irshad, M.; Uzoma, K.C. Comparison of three dielectric moisture sensors for measurement of water in saline sandy soil. *Soil Use Manag.* **2008**, *24*, 156–162. [CrossRef]
16. Kargas, G.; Soulis, K.X. Performance Analysis and Calibration of a New Low-Cost Capacitance Soil Moisture Sensor. *J. Irrig. Drain. Eng.* **2012**, *138*, 632–641. [CrossRef]
17. Regalado, C.M.; Ritter, A.; Rodríguez-González, R.M. Performance of the Commercial WET Capacitance Sensor as Compared with Time Domain Reflectometry in Volcanic Soils. *Vadose Zone J.* **2007**, *6*, 244–254. [CrossRef]
18. Bouksila, F.; Persson, M.; Berndtsson, R.; Bahri, A. Soil water content and salinity determination using different dielectric methods in saline gypsiferous soil. *Hydrol. Sci. J.* **2008**, *53*, 253–265. [CrossRef]
19. Visconti, F.; de Paz, J.M.; MartÃnez, D.; Molina, M.J. Laboratory and field assessment of the capacitance sensors Decagon 10HS and 5TE for estimating the water content of irrigated soils. *Agric. Water Manag.* **2014**, *132*, 111–119. [CrossRef]
20. Nimmo, J.R. Porosity and Pore Size Distribution. *Encycl. Soils Environ.* **2004**, *3*, 295–303.
21. Iwata, Y.; Miyamoto, T.; Kameyama, K.; Nishiya, M. Effect of sensor installation on the accurate measurement of soil water content. *Eur. J. Soil Sci.* **2017**, *68*, 817–828. [CrossRef]
22. Pardossi, A.; Incrocci, L.; Incrocci, G.; Malorgio, F.; Battista, P.; Bacci, L.; Rapi, B.; Marzialetti, P.; Hemming, J.; Balendonck, J. Root Zone Sensors for Irrigation Management in Intensive Agriculture. *Sensors* **2009**, *9*, 2809–2835. [CrossRef] [PubMed]

23. Baram, S.; Couvreurd, V.; Harter, T.; Read, M.; Brown, P.H.; Kandelous, M.; Smart, D.R.; Hopmans, J.W. Estimating Nitrate Leaching to Groundwater from Orchards: Comparing Crop Nitrogen Excess, Deep Vadose Zone Data-Driven Estimates, and HYDRUS Modeling. *Vadose Zone J.* **2016**, 15. [CrossRef]
24. Gamage, D.N.V.; Biswas, A.; Strachan, I.B. Field Water Balance Closure with Actively Heated Fiber-Optics and Point-Based Soil Water Sensors. *Water* **2019**, *11*, 135. [CrossRef]
25. Evett, S.R.; Tolk, J.A.; Howell, T.A. Soil Profile Water Content Determination. *Vadose Zone J.* **2006**, *5*, 894–907. [CrossRef]
26. Schwartz, R.C.; Casanova, J.J.; Pelletier, M.G.; Evett, S.R.; Baumhardt, R.L. Soil Permittivity Response to Bulk Electrical Conductivity for Selected Soil Water Sensors. *Vadose Zone J.* **2013**, 12. [CrossRef]
27. Varble, J.L.; Chavez, J.L. Performance evaluation and calibration of soil water content and potential sensors for agricultural soils in eastern Colorado. *Agric. Water Manag.* **2011**, *101*, 93–106. [CrossRef]
28. Jae-Kwon, S.; Won-Tae, S.; Jae-Young, C. Laboratory and Field Assessment of the Decagon 5TE and GS3 Sensors for Estimating Soil Water Content in Saline-Alkali Reclaimed Soils. *Commun. Soil Sci. Plant Anal.* **2017**, *48*, 2268–2279.
29. Campbell, J.E. Dielectric Properties and Influence of Conductivity in Soils at One to Fifty Megahertz. *Soil Sci. Soc. Am. J.* **1990**, *54*, 332–341. [CrossRef]
30. Kelleners, T.J.; Verma, A.K. Measured and Modeled Dielectric Properties of Soils at 50 Megahertz. *Soil Sci. Soc. Am. J.* **2010**, *74*, 744–752. [CrossRef]
31. Jones, S.B.; Blonquist, J.M.J.; Robinson, D.A.; Philip Rasmussen, V.; Or, D. Standardizing Characterization of Electromagnetic Water Content Sensors: Part 1. Methodology. *Vadose Zone J.* **2005**, *4*, 1048–1058. [CrossRef]
32. Whalley, W.R. Considerations on the use of time-domain reflectometry (TDR) for measuring soil water content. *J. Soil Sci.* **1993**, *44*, 1–9. [CrossRef]
33. Persson, M. Soil Solution Electrical Conductivity Measurements under Transient Conditions Using Time Domain Reflectometry. *Soil Sci. Soc. Am. J.* **1997**, *61*, 997–1003. [CrossRef]
34. Rhoades, J.D.; van Schilfgaarde, J. An Electrical Conductivity Probe for Determining Soil Salinity. *Soil Sci. Soc. Am. J.* **1976**, *40*, 647–651. [CrossRef]
35. Malicki, M.W.R.; Walczak, R.; Koch, S.; Fluhler, H. Determining soil salinity from simultaneous readings of its electrical conductivity and permittivity using TDR. In Proceedings of the Time Domain Reflectometry in Environmental, Infrastructure, and Mining Applications United States Department of Interior Bureau of Mines, the Time Domain Reflectometry in Environmental, Infrastructure, and Mining Applications United States Department of Interior Bureau of Mines, 7–9 September 1994; Northwestern University: Evanston, IL, USA; pp. 328–336.
36. *5TE. Sensor Manual. 5TE- Water Content, Electrical Conductivity (EC) and Temperature sensor*; Decagon Devices Inc.: Pullman, WA, USA, 2016; Available online: https://www.decagon.com/ (accessed on 15 March 2016).
37. WET. *Sensor Manual-UTM-1.6*; Delta-T Devices Ltd.: Cambridge, UK, 2019; Available online: https://www.delta-t.co.uk/ (accessed on 10 July 2019).
38. Zammouri, M.; Siegfried, T.; El-Fahem, T.; Kriaca, S.; Kinzelbach, W. Salinization of groundwater in the Nefzawa oases region, Tunisia: Results of a regional-scale hydrogeologic approach. *Hydrogeol. J.* **2007**, *15*, 1357–1375. [CrossRef]
39. USDA. Diagnostic and improvement of saline and alkali soil. In *Agriculture Handbook No. 60*; US Department of Agriculture: Washington, DC, USA, 1954. Available online: https://www.ars.usda.gov (accessed on 20 September 2019).
40. Scudiero, E.; Berti, A.; Teatini, P.; Morari, F. Simultaneous Monitoring of Soil Water Content and Salinity with a Low-Cost Capacitance-Resistance Probe. *Sensors* **2012**, *12*, 17588–17607. [CrossRef]
41. Bircher, S.; Demontoux, F.O.; Razafindratsima, S.; Zakharova, E.; Drusch, M.; Wigneron, J.-P.; Kerr, Y. L-Band relative permittivity of organic soil surface layers: A new dataset of resonant cavity measurements and model evaluation. *Remote Sens.* **2016**, *8*, 1–17. [CrossRef]
42. Parvin, N.; Degra, A. Soil-specific calibration of capacitance sensors considering clay content and bulk density. *Soil Res.* **2016**, *54*, 111–119. [CrossRef]
43. Cardenas-Lailhacar, B.; Dukes, M.D. Precision of soil moisture sensor irrigation controllers under field conditions. *Agric. Water Manag.* **2010**, *97*, 666–672. [CrossRef]
44. Kassaye, K.; Boulange, J.; Saito, H.; Watanabe, H. Calibration of capacitance sensor for Andosol under field and laboratory conditions in the temperate monsoon climate. *Soil Tillage Res.* **2019**, *189*, 52–63. [CrossRef]

45. Kinzli, K.-D.; Manana, N.; Oad, R. Comparison of Laboratory and Field Calibration of a Soil-Moisture Capacitance Probe for Various Soils. *J. Irrig. Drain. Eng.* **2012**, *138*, 310–321. [CrossRef]
46. Heimovaara, T.J.; Focke, A.G.; Bouten, W.; Verstraten, J.M. Assessing Temporal Variations in Soil Water Composition with Time Domain Reflectometry. *Soil Sci. Soc. Am. J.* **1995**, *59*, 689–698. [CrossRef]

sensors

MDPI

Article

Quantitative Analysis of Elements in Fertilizer Using Laser-Induced Breakdown Spectroscopy Coupled with Support Vector Regression Model

Wen Sha [1], Jiangtao Li [1], Wubing Xiao [1], Pengpeng Ling [1] and Cuiping Lu [2,*]

[1] Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, School of Electric Engineering and Automation, Anhui University, Hefei 230061, China

[2] Laboratory of Intelligent Decision, Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China

* Correspondence: cplu@iim.ac.cn; Tel.: +86-551-6559-5025

Received: 21 June 2019; Accepted: 22 July 2019; Published: 25 July 2019

Abstract: The rapid detection of the elements nitrogen (N), phosphorus (P), and potassium (K) is beneficial to the control of the compound fertilizer production process, and it is of great significance in the fertilizer industry. The aim of this work was to compare the detection ability of laser-induced breakdown spectroscopy (LIBS) coupled with support vector regression (SVR) and obtain an accurate and reliable method for the rapid detection of all three elements. A total of 58 fertilizer samples were provided by Anhui Huilong Group. The collection of samples was divided into a calibration set (43 samples) and a prediction set (15 samples) by the Kennard–Stone (KS) method. Four different parameter optimization methods were used to construct the SVR calibration models by element concentration and the intensity of characteristic line variables, namely the traditional grid search method (GSM), genetic algorithm (GA), particle swarm optimization (PSO), and least squares (LS). The training time, determination coefficient, and the root-mean-square error for all parameter optimization methods were analyzed. The results indicated that the LIBS technique coupled with the least squares–support vector regression (LS-SVR) method could be a reliable and accurate method in the quantitative determination of N, P, and K elements in complex matrix like compound fertilizers.

Keywords: fertilizer; support vector regression; laser-induced breakdown spectroscopy; grid method; genetic algorithm; particle swarm optimization; least squares

1. Introduction

The foundation of precise fertilization is accurately obtaining the content of elements in compound fertilizers to maximize their benefits. At present, the main sensing methods used by fertilizer manufacturers are national standard methods [1], inductively coupled plasma–atomic emission spectroscopy (ICP-AES) [2], flame atomic absorption spectrometry (FAAS) [3], atomic absorption spectroscopy (AAS) [4], near infrared reflectance spectroscopy (NIRS) [5], etc. These sensing methods need field samples and pretreatment before laboratory analysis, which are time-consuming, labor-intensive, and expensive requirements. Meanwhile, due to the contents of N, P, and K in compound fertilizer being typically high, fertilizer must be diluted several times before measuring, which leads to increase systematic errors. Further, these sensing methods cannot provide real-time information during the production process. At present, the pass rate of compound fertilizer products is only 97% [6], and non-qualifying products need to be returned to the factory for reprocessing, causing huge economic losses. Therefore, a technology that can quickly and accurately detect the elemental content in compound fertilizers is urgently needed.

Laser-induced breakdown spectroscopy (LIBS) is an ideal laser high-temperature ablation spectroscopy technique because it provides fast (in the order of milliseconds), insitu (smaller sample

size), non-destructive, safe, environmentally friendly (no secondary pollution), multi-element analysis, and direct analysis of any state of matter. It has been successfully used in applications relating to water pollution [7], coal combustion [8], agriculture [9], space exploration [10], etc. In recent years, some studies have used the LIBS technique to detect the components of fertilizer. Andrade et al. turned liquid fertilizer into a solid state and detected its content using the LIBS technique. The levels of Cu, K, Mg, Mn, Zn, As, Cd, Cr and Pb in the fertilizer were analyzed, and the detection error range was ~0.02%–0.06%, which demonstrated that this method provides accurate measurement of liquid fertilizer [11]. Nicolodelli et al. used single-pulse and dual-pulse LIBS technology to measure phosphate rock and organic phosphate fertilizer. The samples were identified by principal component analysis (PCA) and partial least squares regression (PLSR), and the recognition result was able to reach a 95% confidence level, which showed that LIBS technology can be used to rapidly classify phosphate fertilizers in situ [12]. Yao et al. analyzed phosphorus and potassium elements in compound fertilizers. A quantitative analysis model was established by using the partial least squares (PLS) method in Unscrambler software. The obtained results were superior to those obtained using traditional methods, but the accuracy of detection still needed improvement [13]. Marangoni et al. used LIBS technology to analyze phosphorus in 26 different organic and inorganic fertilizers. Baseline correction and peak intensity normalization were used to pre-process the spectrum. The correlation between the measured value of LIBS and the true value was improved, but the absolute error of the two verification samples was close to 5% [14]. Andradeet al. optimized the LIBS system parameters and directly analyzed the levels of Cd, Cr, Pb, B, Cu, Mn, Na, Zn, Ca, and Mg elements in solid compound fertilizer. The quantitative analysis results were compared to those obtained using ICP-AES, and the correlation was good. The detection limit of the above elements was determined to be ~2 ppm–1%. These results demonstrated the ability of LIBS to be used for rapid analysis of fertilizer [15]. Liao et al. used LIBS to analyze the phosphorus content in compound fertilizer. The correlation coefficient increased from 0.83 to 0.98 when considering the influence of the oxygen characteristic line, and the relative error was only ~0.38%–1.70%. However, the number of samples was too small, so further modeling should be completed if the method is to be applied to actual field detection [16].

The above studies all showed that LIBS technology can be used to detect the types and contents of elements in chemical fertilizers. However, the accuracy and reliability of the quantitative analysisis still a shortcoming of LIBS, which greatly limits its practical application. Thus, there are many problems to be solved before it can be used in actual applications in rapid sensing.

Support vector regression (SVR) is a machine learning method based on statistical learning theory. It uses interval maximization to carry out model training by mapping difficult problems in the original space to a higher-dimensional space and seeking interval maximization to calculate the optimal linear hyper plane. The number of compound fertilizer samples collected in this experimental method was small, which is suitable for the statistical analysis of small samples [17]. In small samples, nonlinear pattern recognition has certain advantages. Zhang et al. employed a LIBS technique coupled with SVR and PLS methods to perform quantitative and classification analysis of 20 slag samples. The results showed that the SVR model could eliminate the influence of nonlinear factors due to self-absorption in the plasma and provide a better predictive result. It has been confirmed that the LIBS technique coupled with the SVR method is a promising approach to achieving the online analysis and process control of slag [18]. Shi et al. compared PLSR and SVR methods for quantitative analysis of the concentrations of five main elements (Si, Ca, Mg, Fe, and Al) in sedimentary rock samples. The parameter optimization method used was genetic algorithm (GA). The results demonstrated that the SVR model performed better, with more satisfactory accuracy and precision under the optimized conditions [19]. He et al. employed single-pulse and double-pulse LIBS to analyze nutrient elements in soil. Good performance was obtained using the PLSR and LS-SVR calibration model, with R^2 greater than 0.95 in both the calibration and prediction sets for all nutrient elements. The results indicated that LIBS combined with PLSR and LS-SVR could be a good method for detecting nutrient elements in soil [20]. Liu et al. also used the PLSR and LS-SVR quantitative analysis method to detect the Cd

content in soil. The results showed that the LS-SVR model under an Ar atmosphere obtained the best performance. The root-mean-square error for calibration (RMSEC) and the root-mean square error for prediction (RMSEP) were only 0.026 and 0.034, respectively, which demonstrated the ability of LIBS for the accurate quantitative detection of Cd in soil [21]. To the best of our knowledge, the simultaneous quantitative detection of N, P, and K elements based on LIBS coupled with SVR models by different parameter optimization methods has not been investigated.

The purpose of this paper was to explore the detection ability of LIBS for N, P, and K elements in fertilizer, and to find a fast and accurate quantitative analysis method. A selection of 58 fertilizer samples were provided by Huilong Chemical Fertilizer Plant, Anhui, China. The contents of all three elements in compound fertilizer were determined by ICP-AES. Four different parameter optimization methods were employed to establish the SVR model for quantitative analysis of all elements. The accuracy levels of the four SVR models were compared based on the performance of each model.

2. Materials and Methods

2.1. Sample Preparation

In this study, 58 compound fertilizer samples were collected and placed into sealed plastic bags to avoid contamination. As the compound fertilizer products had been pelletized, they had to be broken into powder before spectral scanning. All fertilizer powders were then sieved through a 60 mesh screen. A total 2 g of powder from each of the 58 fertilizer samples was weighed and pressed into tablets 30 mm in diameter and 2 mm in thickness, using 5 MPa force for 60 s (769YP-40C, KQ, Tianjin, China). The reference concentrations of N, P, and K in these samples were analyzed by ICP-AES. The statistics of the N, P, and K concentrations in the compound fertilizer samples are listed in Table 1.

Table 1. Statistics of the effective constituents of compound fertilizer samples.

Properties	Total Nitrogen (TN/%)	P_2O_5 (%)	K_2O (%)
Minimum value	13.60	14.50	14.40
Maximum value	15.60	16.70	16.40
Mean value	14.42	15.79	15.39
Standard deviation values	2.86	2.97	3.29

2.2. Experimental Setup

The self-built LIBS system used in this experiment is shown in Figure 1. A Q-switched Nd: YAG pulsed laser (ICE450, 1064 nm, 6 ns pulse duration, Big Sky Laser Technologies, Morgan Hill, CA, USA; note that the company has changed its name to Quantel Laser) was used to generate the plasma on the compound fertilizer pellet. The pulse laser energy was 100 mJ, and it was focused with a 2.54 cm diameter, 4.5 cm focal length convex lens onto the fertilizer sample. The spot diameter size of the pulsed laser was approximately 0.5 mm, and the peak power density on the compound fertilizer sample was able to reach 2.2 GW/cm^2. The light emitted from the plasma was collected via a quartz lens with 3.5 cm focal length and transmitted via an optical fiber with a diameter of 200 μm to a spectrometer (Avantes-ULS2048-USB2, Avantes, Apeldoorn, The Netherlands). The spectrometer used has four channels containing separate gratings and a charge-coupled device array, and all spectra were taken simultaneously in the wavelength ranges of 190–510 and 690–890 nm. The resolution of the spectrometer was approximately 0.1 nm. The spectrometer was triggered by the laser Q-switch output, and it had a digital delay generator which can control the gate delay. In this experiment, the delay time and the integration time were set for spectra acquisition at 1.28 μs and 1.05 ms (spectrometer minimum integration time), respectively [22]. Sample tablets were placed on the X-Y rotary stage, the speed of which can be adjusted by stepper motor, and the laser beam was adjusted to focus 3 mm below the sample surface to acquire LIBS spectra [23]. Argon was then passed through the cylinder, draining the air and forming an Ar atmosphere on the surface of the compound fertilizer sample, as shown in

Figure 1. Thus, the sample was immersed in an Ar atmosphere. In order to eliminate shot-to-shot fluctuation, each sample was measured eight times, and each spectrum was collected with an average of 20 laser shots.

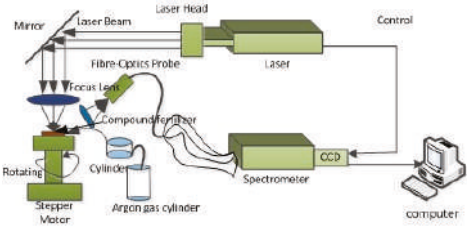


Figure 1. Schematic diagram of the laser-induced breakdown spectroscopy (LIBS) system for fertilizer samples.

2.3. SVR Algorithm Model Establishment

Under local thermal equilibrium (LTE) conditions and ignoring self-absorption effects, the measured characteristic atomic spectrum of ionic line intensity in LIBS spectroscopy can be expressed as

$$I_{k,i} = FC_s \frac{A_{k,i} g_k}{U_s(T)} \exp(-E_k / k_B T) \quad (1)$$

where the subscripts k and I indicate the upper and lower energy levels of the transition line, respectively; F is an instrumental constant for fixed experimental conditions; Cs is the atomic or ionic number density of the specific element; g, A, and $U(T)$, are the statistical weight, transition probability and partition function at temperature T, respectively; k_B and E represent the Boltzmann constant and excitation energy, respectively; and I is the spectrally integrated line intensity. When the plasma is in the local thermal equilibrium state, the plasma temperature can be approximated as a constant. Equation (1) can be simplified as,

$$C_S = AI_{k,i} \quad (2)$$

The concentration of the element to be tested in the sample can be calculated according to Equation (2). However, due to the influence of the matrix effect, parameter A is difficult to determine experimentally. Moreover, when the concentration of the element increases, as self-absorption effect occurs, and the relationship between C_s and $I_{k,i}$ can be expressed as,

$$C_S = K_b (I_{k,i})^b \quad (3)$$

where K_b is the proportionality factor and b is the absorption coefficient. We simplified $I_{k,i}$, to I, and substituted it as a variable to support vector machine regression objective function,

$$C_S = \Sigma_{i \in v} \partial_i k_{libs}(I_i, I) + b \quad (4)$$

where v is the set of support vectors; ∂_i is the Lagrange multiplier; k_{libs} $(I_i,\ I)$ is the kernel function; and b is the constant.

A hybrid kernel function for LIBS is obtained by considering the relationship between the element concentration and spectrum line intensity (Equations (2) and (3)):

$$k_{libs}(I_i, I) = cII_i + (1-c)\exp\left(\frac{-\| I - I_i \|}{2g^2}\right) \quad (5)$$

The mixed kernel function consists of two parts; the former is a linear kernel function cII_i, and the latter is a radial basis kernel function. The support vector machine kernel function is often used to solve nonlinear mapping problems in data. A large number of experiments and data have shown that the

radial basis kernel function has high fitting and prediction accuracy, so it is usually selected as a kernel function for research. The adjustment of the parameters in the SVR largely determines the regression effect. When the radial basis kernel function is selected as the kernel function, the penalty coefficient c and the kernel parameter g are mainly optimized in the quantitative model of the compound fertilizer element analysis [24].

2.4. Parameter Optimization Methods

In related research, SVR optimization methods have mainly been summarized as non-heuristic and heuristic. The traditional grid search method (GSM) and experimental method are non-heuristic, but the experimental method uses several experiments to compare parameters and determine the optimal ones. It is time-consuming and not easy to find the optimal parameters, so this research did not use the experimental method.

In the GSM method, X.L. Liu et al. proposed that the parameters c and g should be divided into an equal grid within a certain spatial range [25]. Every grid node then represents a set of parameters. In the optimization process, the optimal parameters were found by gradual approximation of all the nodes in the grid, and the c and g parameters with the smallest regression mean square error were taken as the optimal parameters. A large number of experimental studies have shown that the parameters c and g have interval sensitivity. If the parameter optimization interval can first be roughly determined, then an accurate search can be performed to reduce unnecessary calculations. First, a large step size should be used to perform a rough search in a large range and to select a set of c, g values for the minimum regression mean square error. If multiple sets of c and g values correspond to the minimum regression mean square error in the parameter selection process, then the group of c and g values with the smallest parameter c should be selected as the best parameters. Too high a value of parameter c would lead to an over-learning state, that is, a state where the RMSEC is small, but that of the prediction set is large, and the generalization ability of the SVR is reduced. After finding the local optimal parameters, we selected a cell in the vicinity of this group parameters and used a small step size to perform the second, finer search to find the final optimal parameters. Chen P W et al. proposed a global probability search algorithm based on biological mechanisms such as natural selection and genetic variation [26]. As with other heuristic search methods, the evolutionary mechanisms of organisms are simulated during evolutionary computation, starting from a set of solutions and evaluating their performance. Hybridization and mutational gene manipulations are then performed to generate a group of next-generation solutions with better performance metrics, until the final search for the global optimal solution. The particle swarm optimization (PSO) algorithm is proposed by J. Kennedy and R.C. Eberhart et al. and based on the study of the predation behavior of birds [27]. The solution of each problem is regarded as a bird in the search space, denoted as particles. In each iteration, the particle will track two "extreme values" to update itself; one is the optimal solution found by the particle itself, and the other is the current optimal solution found by the entire population. This extreme value is the global extreme.

In 1999, Suyken et al. added the squared error term to the standard SVR objective function and proposed the LS-SVR method [28]. In this method, the observed value is the sample value, and the theoretical value is the assumed fitting function. The fitting function model is then obtained when the objective function is the smallest. We set the objective function as

$$h_\theta(x_1, x_2, \ldots x_n) = \theta_0 + \theta_1 x_1 + \ldots + \theta_{n-1} x_{n-1} \tag{6}$$

and the loss function as,

$$J(\theta) = \frac{1}{2}(X\theta - Y)^T (X\theta - Y) \tag{7}$$

After derivation and sorting out the parameters as,

$$\theta = (X^T X)^{-1} X^T Y \tag{8}$$

It can be seen that the LS algorithm is simple and efficient. The constraint avoids the quadratic programming in the objective function and solves the problems of robustness, sparseness, and large-scale operation, which greatly shortens the optimization time. However, there are also some limitations: (1) when the inverse matrix of X^TX does not exist, the LS algorithm is no longer applicable, and data processing needs to remove redundant features; (2) the fitting function must be a linear function, which must be converted before use; and (3) when the sample feature number N is large, it takes a lot of time to calculate the inverse matrix, and it may not be able to be calculated. In this study, the number of compound fertilizer samples was small, and the characteristic dimension of SVR was declining. Thus, the LS algorithm could be used for parameter optimization.

3. Results

3.1. Spectral Analysis

Due to the existence of a large number of matrix element emission lines in the compound fertilizer, many characteristic lines interfered with each other. When selecting the characteristic line of an element, an unsaturated line with a high signal-to-noise ratio should be selected. According to the National Institute of Standards and Technology (NIST) database, the characteristic lines of elemental phosphorus are 213.5 nm, 214.9 nm, 215.4 nm, 253.4 nm, 253.6 nm, 255.3 nm, and 255.5 nm; the characteristic lines of elemental K are 404.7 nm, 766.5nm, and 769.9 nm; and the characteristic lines of elemental N are 742.4 nm, 744.2 nm, 746.8 nm, 856.7nm, 859.4 nm, 862.9 nm, 870.3 nm, 871.2 nm, and 871.8 nm. Figure 2 shows the spectrum of a compound fertilizer sample (Sample No.1) in the ranges of 210–405 nm and 740–890 nm. Although the two characteristic lines of P at 253.4 nm and 253.6 nm were strong, they were easily interfered with by the characteristic line of iron (Fe). In addition, the two lines at 255.3 nm and 255.5 nm were too close to distinguish. The characteristic lines at the wavelengths of 213.5 nm, 214.9 nm, and 215.4 nm were not interfered by other elements. It can be seen from Figure 2 that the characteristic lines of N were observed easily without any interference by other spectral lines, and the intensity at 746.8 nm was the strongest. The characteristic lines of elemental K in the compound fertilizer were not rich. Only three characteristic lines at 404.4 nm, 766.5 nm, and 769.9 nm were observed. The line at 404.4 nm was possibly interfered with by the characteristic line of Fe at 404.8 nm. However, the lines at 766.5 nm and 769.9 nm were too strong to be due to self-absorption. Elemental oxygen is also one of the main ingredients of compound fertilizer; the characteristic lines of O are 777.2 nm, 844.6 nm, and 882.0 nm.

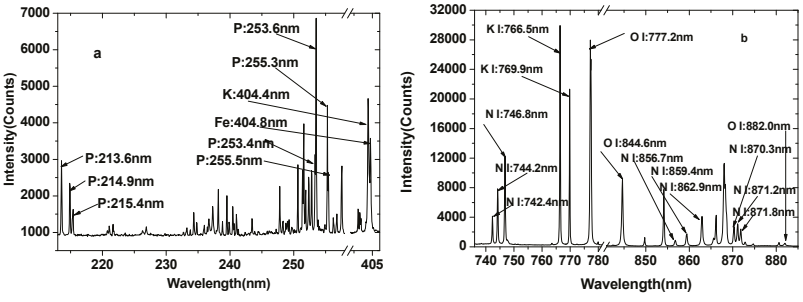


Figure 2. LIBS spectra of a compound fertilizer sample in the ranges of (**a**) 210–405 nm and (**b**) 740–890 nm.

3.2. Univariate Analysis

Before modeling, the 58 fertilizer samples were divided into a calibration set (43 samples) and a prediction set (15 samples) using the Kennard–Stone (KS) method. The univariate calibration models were constructed using the line intensities (the height of Lorentz fits) of N at 746.8 nm, P at 213.6 nm, and K at 404.4 nm versus the corresponding contents. Figure 3a–c shows the calibration curves of N, P,

and K elements, respectively. Figure 3a indicates the linear trend between the N line intensity and content, with a coefficient of correlation of 0.809. For P, the coefficient of correlation is 0.909, while it is 0.857 for K. For all three elements, the correlation coefficients cannot meet practical measurement needs and should be improved for further quantitative analysis.

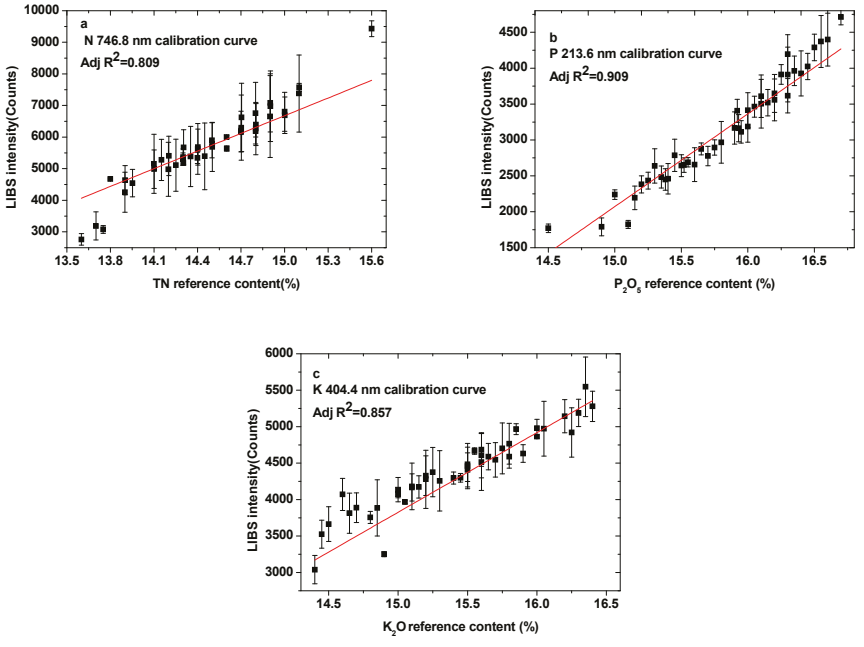


Figure 3. Calibration curves of the elemental spectral lines: (**a**) N: 746.8 nm, (**b**) P: 213.6 nm, (**c**) K: 404.4 nm.

3.3. SVR Analysis Models of Compound Fertilizer

SVR is a multivariate analytical technique which can make full use of spectral information and improve the accuracy of quantitative analysis by reducing the matrix effect. Calibration sets were used to construct the SVR model using the contents of N, P, and K elements and the LIBS spectral signal. These correlations were used to predict the contents of the prediction set. For each element, a proper spectral range was selected for modeling analysis in order to avoid over-fitting of the model. The reduced spectral ranges of 740–890 nm for N, 210–260 nm and 770–885 nm for P, and 400–410 nm and 770–885 nm for K were used to obtain the calibration model. MATLAB software was used for SVR model construction. The statistical parameters that determine the capacity of the regression model are the training time, the determination coefficients of the calibration set (R^2_C) and prediction set (R^2_P), and the RMSEC and RMSEP, which are given in this paper.

3.3.1. Particle Swarm Optimization

Figure 4a–f shows the calibration and prediction results of SVR models using the PSO algorithm for N, P, and K elements, respectively. All of the parameters for both the calibration and prediction sets are presented in Table 2. The training times for N, P, and K were 2.98 s, 3.31 s, and 4.32 s, respectively. The determination coefficients for the calibration sets (R^2_C) were 0.930 for N, 0.980 for P, and 0.979 for K. Those for the prediction sets (R^2_P) were 0.923, 0.964, and 0.952. Meanwhile, for N, P, and K, respectively, the values of the RMSEC were 0.0996, 0.0701, and 0.0894, and those of the RMSEP were 0.0952, 0.0677, and 0.0921. The PSO-SVR optimization data showed that there was little difference between the three optimization times. The correlation coefficient between the N element calibration set

and the prediction set was small and the error was large. The determination coefficients R^2_C and R^2_P for N indicated that the correlation needed to be improved. Meanwhile, the RMSEC and RMSEP for elements N and K were large.

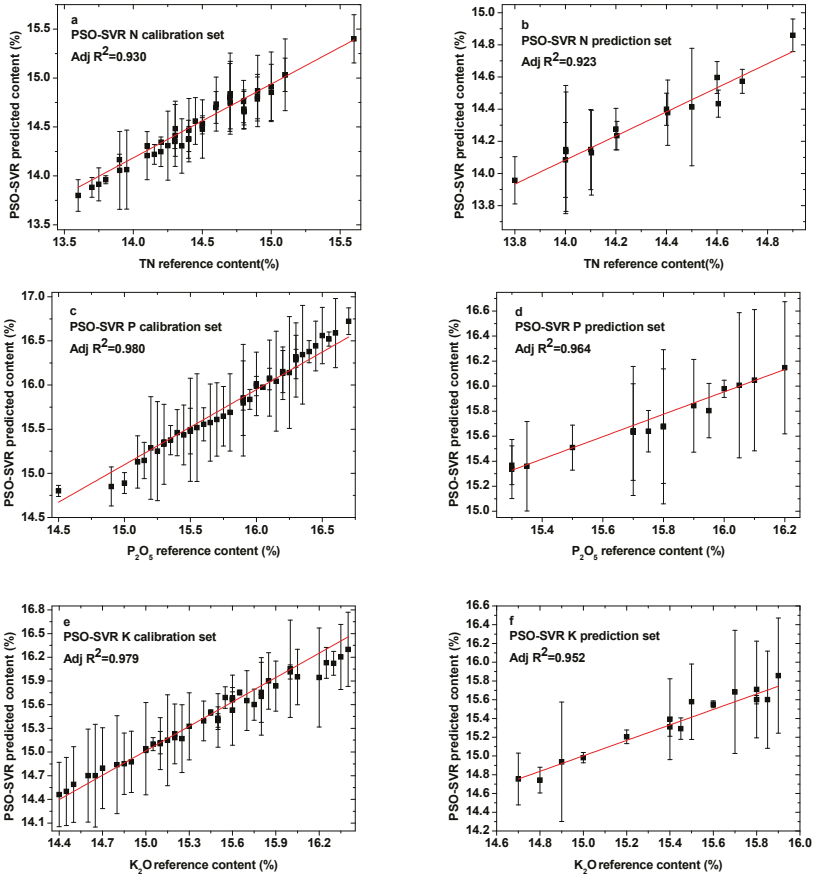


Figure 4. Comparison between PSO-SVR predicted content and reference content present in the (**a**) N calibration set; (**b**) N prediction set; (**c**) P calibration set; (**d**) P prediction set; (**e**) K calibration set; and (**f**) K prediction set.

Table 2. The results of the particle swarm optimization–support vector regression (PSO-SVR) model for elements N, P, K.

Element	t/s	R^2_C	RMSEC	R^2_P	RMSEP
N	2.98	0.930	0.0996	0.923	0.0952
P	3.31	0.980	0.0701	0.964	0.0677
K	4.32	0.979	0.0894	0.952	0.0921

3.3.2. Genetic Algorithm

The optimal calibration and prediction results of N, P, and K elements based on parameters obtained using the genetic algorithm are shown in Figure 5a–f. Table 3 presents the detailed parameter results. The training times increased dramatically for all three elements. They were 5.67 s, 5.09 s, and 12.37 s, for N, P, and K, respectively. The R^2_C of N increased from 0.930 to 0.948, but the R^2_P

changed from 0.923 to 0.936. The values of RMSEC and RMSEP for N were reduced to 0.0688 and 0.0694, which was better than the PSO algorithm. However, for elements P and K elements, the parameter optimization results of PSO and GA were basically the same, with only the R^2_P value of P increased, from 0.964 to 0.985. This indicates that GA and PSO had many similarities, but GA was less inefficient due to random variation.

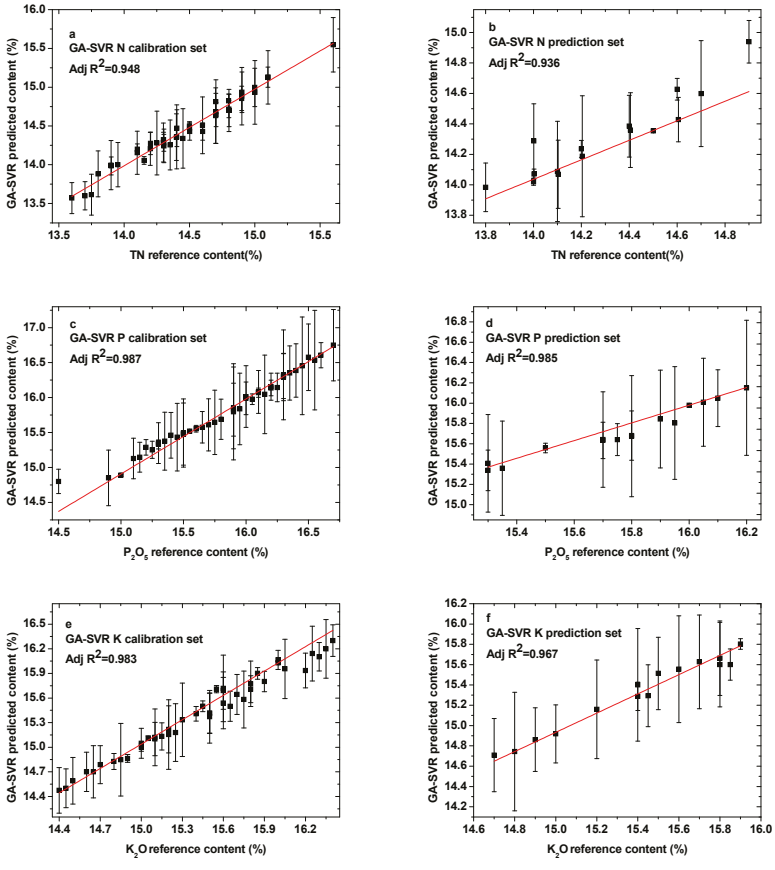


Figure 5. Comparison between GA-SVR predicted content and reference content present in the (**a**) N calibration set; (**b**) N prediction set; (**c**) P calibration set; (**d**) P prediction set; (**e**) K calibration set; and (**f**) K prediction set.

Table 3. The results of the genetic algorithm–support vector regression (GA-SVR) model for N, P, and K.

Element	t/s	R^2_C	RMSEC	R^2_P	RMSEP
N	5.67	0.948	0.0688	0.936	0.0694
P	5.09	0.987	0.0692	0.985	0.0680
K	12.37	0.983	0.0775	0.967	0.1007

3.3.3. Grid Search Method

A quantitative analysis SVR model of elements N, P, and K in the compound fertilizer was established based on parameter optimization by GSM. Figure 6a–f shows the calibration and prediction results of GSM parameter optimization for the elements N, P, and K. All of the results are presented in

Table 4. For N, the training times, RMSEC, and RMSEP were almost identical to the results of the GA algorithm, but the R^2_C and R^2_P increased greatly. Among the three parameter optimization methods, for the P element, the results obtained by the GSM method were the best, and the training time was only 1.76 s. However, the best results for the K element were obtained by using the GA algorithm.

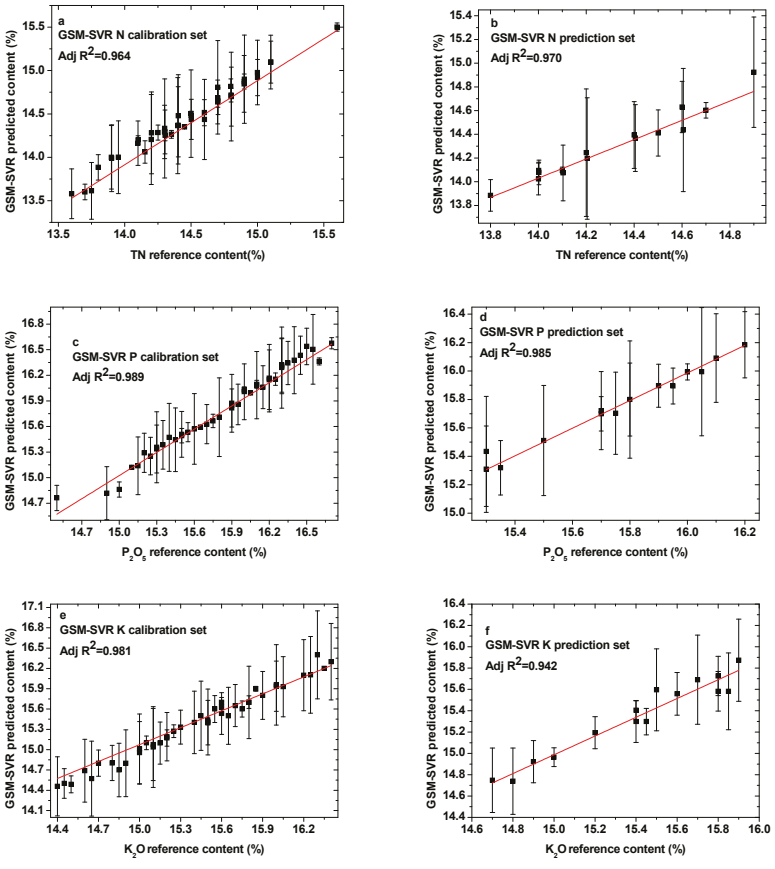


Figure 6. Comparison between GSM-SVR predicted content and reference content present in the (**a**) N calibration set; (**b**) N prediction set; (**c**) P calibration set; (**d**) P prediction set; (**e**) K calibration set; and (**f**) K prediction set.

Table 4. The results of grid search method–support vector regression (GSM-SVR) model for elements N, P, and K.

Element	t/s	R^2_C	RMSEC	R^2_P	RMSEP
N	4.89	0.964	0.0685	0.970	0.0712
P	1.76	0.989	0.0632	0.985	0.0576
K	4.21	0.981	0.0942	0.942	0.0969

It can be seen that the quantitative analysis of the elements N, P, and K in compound fertilizer from the above three parameter optimization results was better than that by traditional methods [23], but it should be further improved.

3.3.4. Least Squares

The calibration and prediction results of the LS parameter optimization models for all analyzed elements are provided in Figure 7a–f. All of the specific parameter optimization results are stated in Table 5. It can be observed in Figure 6 that the calibration and prediction data points fitted well, indicating that the LS-SVR model had reliable prediction power for the quantitative analysis of compound fertilizer. In addition, the R^2 values of the calibration and prediction sets for all elements were obviously improved. All the values of R^2 were greater than 0.99. Meanwhile, for all elements, the values of RMSEC and RMSEP were significantly reduced. The values of RMSEC were reduced to 0.0240, 0.0258, and 0.0248 for elements N, P, and K, respectively, and the values of RMSEP were only 0.0218 for N, 0.0261 for P, and 0.0248 for K. The training time was significantly reduced; all the values of t were smaller than 0.3 s, which is more suitable for the rapid quantitative analysis of elements in compound fertilizer. Thus, it was demonstrated that the LS-SVR model can be developed to predict the content of unknown samples.

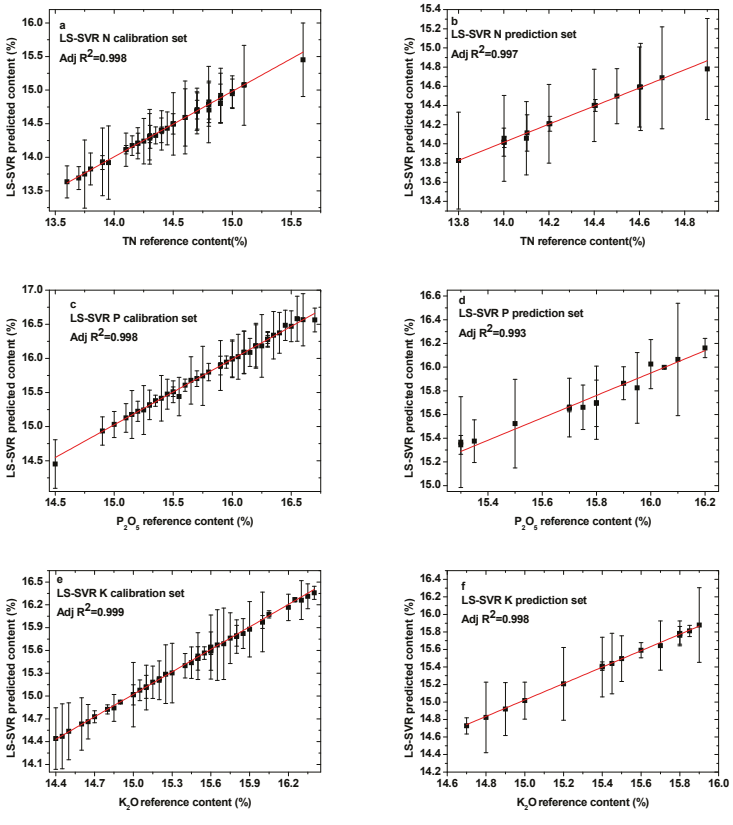


Figure 7. Comparison between the LS-SVR predicted content and reference content present in the (**a**) N calibration set; (**b**) N prediction set; (**c**) P calibration set; (**d**) P prediction set; (**e**) K calibration set; and (**f**) K prediction set.

Table 5. The results of the least squares–support vector regression (LS-SVR) model for elements N, P, and K.

Element	*t/s*	R^2_C	RMSEC	R^2_P	RMSEP
N	0.23	0.998	0.0240	0.997	0.0218
P	0.02	0.998	0.0258	0.993	0.0261
K	0.02	0.999	0.0239	0.998	0.0248

4. Conclusions

In summary, we demonstrated that LIBS coupling with the SVR method can provide a robust and accurate technology for the analysis of compound fertilizers. Four parameter optimization SVR models—the PSO model, the GA model, the GSM model, and the LS model were employed to quantitatively analyze elements N, P, and K in fertilizer. In general, the complex element composition of fertilizer causes difficulty for traditional calibration methods. For the conventional PSO, GA, and GSM parameter optimization methods, the determination coefficients for all three elements were greater than 0.92, and the root-mean-square errors were less than 0.101. However, the best parameter optimization model was the GSM method for N, GSM for P, and GA for K. A parameter optimization method suitable for quantitative analysis of all three elements was still needed. A LS-SVR model was then used to establish a quantitative analysis model for the three elements. From the results of the LS-SVR model, calibration and prediction models were obtained for the three elements with determination coefficients close to 1. For elements N, P, and K, respectively, the values of RMSEC were 0.0240, 0.0258, and 0.0239, and those of RMSEP were 0.0218, 0.0261, and 0.0248. After considering the evaluation indicators of the model comprehensively, the LS-SVR model is regarded as the most suitable for quantitative analysis of the three elements, with robust and satisfactory modeling performance. This model could therefore provide a basis for real-time analysis of N, P, and K elements in compound fertilizers. Furthermore, methods for improving the accuracy of the LIBS technique in the rapid detection of compound fertilizer on production lines will be the focus of future work.

Author Contributions: J.L., C.L. and W.S. conceived and designed the experiment; P.L. performed the experiments; J.L. and W.X. contributed to data analysis; C.L. and W.S. analyzed the data and wrote the whole paper; all authors reviewed the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (61505001) and Science Technology Service for Regional Key Projects of Chinese Academy of Sciences (CAS) (KFJ-STS-QYZD-106).

Acknowledgments: The fertilizer samples tested in this research were provided by the compound fertilizer production line of Bengbu Huilong Chemical Fertilizer Plant, Anhui, China.

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

References

1. Cui, S.L. Effective phosphorus detection in compound fertilizers by weight method. *Mod. Agric. Sci. Technol.* **2016**, *5*, 231–232.
2. Wang, M.R.; Yuan, Y.M.; Tao, N.L. Determination of available phosphate content in calcium magnesium phosphate by ICP-AES. *Mod. Agric. Sci. Technol.* **2012**, *7*, 20–21.
3. Yuan, C.; Lv, G.L. Determination of chromium content in compound fertilizer by flame atomic absorption spectrometry. *Chem. Fertil. Ind.* **2018**, *45*, 17–18.
4. Xiao, Z.M.; Liang, H.J. Improvement of rapid determination method of available magnesium content in compound fertilizer. *Chem. Fertil. Ind.* **2015**, *42*, 12–15.
5. Song, L.; Zhang, H.; Ni, X.-Y.; Wu, L.; Liu, B.-M.; Yu, L.-X.; Wang, Q.; Wu, Y.-J. Quantitative analysis of contents in compound fertilizer and application research using near infrared reflectance spectroscopy. *Guang Pu Xue Yu Guang Pu Fen Xi* **2014**, *34*, 73–77. [PubMed]
6. Cai, Y.Q.; Wang, J.; Wang, Z.Q.; Meng, F.M.; Liu, J.X. A production method of granular urea formaldehyde slow-release compound fertilizer. *Phosphate Compd. Fertil.* **2017**, *32*, 18–19.
7. Zhang, D.C.; Hu, Z.Q.; Su, Y.B.; Hai, B.; Zhu, X.L.; Zhu, J.F.; Ma, X. Simple method for liquid analysis by laser-induced breakdown spectroscopy (LIBS). *Opt. Express* **2018**, *26*, 18794–18802. [CrossRef]

8. Yan, C.; Qi, J.; Liang, J.; Zhang, T.; Li, H. Determination of coal properties using laser-induced breakdown spectroscopy combined with kernel extreme learning machine and variable selection. *J. Anal. At. Spectrom.* **2018**, *33*, 2089–2097. [CrossRef]
9. Jull, H.; Künnemeyer, R.; Schaare, P. Nutrient quantification in fresh and dried mixtures of ryegrass and clover leaves using laser-induced breakdown spectroscopy. *Precis. Agric.* **2018**, *19*, 823–839. [CrossRef]
10. Cho, Y.; Horiuchi, M.; Shibasaki, K.; Kameda, S.; Sugita, S. Quantitative Potassium Measurements with Laser-Induced Breakdown Spectroscopy Using Low-Energy Lasers: Application to In Situ K-Ar Geochronology for Planetary Exploration. *Appl. Spectrosc.* **2017**, *71*, 1969–1981. [CrossRef]
11. Andrade, D.F.; Sperança, M.A.; Pereira-Filho, E.R. Different sample preparation methods for the analysis of suspension fertilizers combining LIBS and liquid-to-solid matrix conversion: Determination of essential and toxic elements. *Anal. Methods* **2017**, *9*, 5156–5164. [CrossRef]
12. Nicolodelli, G.; Senesi, G.S.; Perazzoli, I.L.D.O.; Marangoni, B.S.; Benites, V.D.M.; Milori, D.M.B.P. Double pulse laser induced breakdown spectroscopy: A potential tool for the analysis of contaminants and macro/micronutrients in organic mineral fertilizers. *Sci. Total. Environ.* **2016**, *565*, 1116–1123. [CrossRef] [PubMed]
13. Yao, S.; Lu, J.; Li, J.; Chen, K.; Li, J.; Dong, M. Multi-elemental analysis of fertilizer using laser-induced breakdown spectroscopy coupled with partial least squares regression. *J. Anal. At. Spectrom.* **2010**, *25*, 1733. [CrossRef]
14. Marangoni, B.S.; Silva, K.S.G.; Nicolodelli, G.; Senesi, G.S.; Cabral, J.S.; Villas-Boas, P.R.; Nogueira, A.R.A.; Teixeira, P.C.; Benites, V.M.; Milori, D.M.B.P. Phosphorus quantification in fertilizers using laser induced breakdown spectroscopy (LIBS): A methodology of analysis to correct physical matrix effects. *Anal. Methods* **2016**, *8*, 78–82. [CrossRef]
15. Andrade, D.F.; Pereira-Filho, E.R. Direct Determination of Contaminants and Major and Minor Nutrients in Solid Fertilizers Using Laser-Induced Breakdown Spectroscopy (LIBS). *J. Agric. Food Chem.* **2016**, *64*, 7890–7898. [CrossRef] [PubMed]
16. Liao, S.Y.; Wu, X.L.; Li, G.H.; Wei, M.; Zhang, M. Multi-element nonlinear quantitative analysis of phosphorus in compound fertilizer by laser induced breakdown spectroscopy. *Spectrosc. Spectr. Anal.* **2018**, *38*, 271–275.
17. Chen, L. PSO-SVM Learning Algorithm and Its Application in Spatial Data Analysis. Master's Thesis, Xi'an Polytechnic University, Xi'an, China, 2012.
18. Zhang, T.; Wu, S.; Dong, J.; Wei, J.; Wang, K.; Tang, H.; Yang, X.; Li, H. Quantitative and classification analysis of slag samples by laser induced breakdown spectroscopy (LIBS) coupled with support vector machine (SVM) and partial least square (PLS) methods. *J. Anal. At. Spectrom.* **2015**, *30*, 368–374. [CrossRef]
19. Niu, G.; Lin, Q.; Shi, Q.; Xu, T.; Li, F.; Duan, Y. Quantitative analysis of sedimentary rocks using laser-induced breakdown spectroscopy: Comparison of support vector regression and partial least squares regression chemometric methods. *J. Anal. At. Spectrom.* **2015**, *30*, 2384–2393.
20. He, Y.; Liu, X.; Lv, Y.; Liu, F.; Peng, J.; Shen, T.; Zhao, Y.; Tang, Y.; Luo, S. Quantitative Analysis of Nutrient Elements in Soil Using Single and Double-Pulse Laser-Induced Breakdown Spectroscopy. *Sensors* **2018**, *18*, 1526. [CrossRef]
21. Liu, X.; Liu, F.; Huang, W.; Peng, J.; Shen, T.; He, Y. Quantitative Determination of Cd in Soil Using Laser-Induced Breakdown Spectroscopy in Air and Ar Conditions. *Molecules* **2018**, *23*, 2492. [CrossRef]
22. Sha, W.; Niu, P.; Zhen, C.; Lu, C.; Jiang, Y. Analysis of Phosphorus in Fertilizer Using Laser-Induced Breakdown Spectroscopy. *J. Appl. Spectrosc.* **2018**, *85*, 653–658. [CrossRef]
23. Zhang, B.; Ling, P.; Sha, W.; Jiang, Y.; Cui, Z. Univariate and Multivariate Analysis of Phosphorus Element in Fertilizers Using Laser-Induced Breakdown Spectroscopy. *Sensors* **2019**, *19*, 1727. [CrossRef] [PubMed]
24. Qu, J.; Chen, H.Y.; Liu, W.Zh.; Li, Zh.B.; Zhang, B.; Ying, Y.H. Application of support vector machine based on improved grid search in quantitative analysis of gas. *Chin. J. Sens. Actuators* **2015**, *28*, 774–778.
25. Liu, X.L.; Jia, D.X.; Li, H. Research on kernel parameter optimization of support vector machine in speaker recognition. *Sci. Technol. Eng.* **2010**, *10*, 1669–1673.
26. Chen, P.W.; Wang, J.Y.; Lee, H. Model selection of SVMs using GA approach. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks, Budapest, Hungary, 25–29 July 2004.
27. Lu, J.H. Research on Particle Swarm Optimization Algorithm for Solving Automatic Cotton Matching Problem. Ph.D. Thesis, Zhejiang University, Hangzhou, China, 2011.

28. Castro-Garcia, R.; Agudelo, O.M.; Tiels, K.; Suykens, J.A.K. Hammerstein system identification using LS-SVM and steady state time response. In Proceedings of the 2016 European Control Conference (ECC), Ålborg, Denmark, 29 June–1 July 2016; pp. 1063–1068.

sensors

MDPI

Article

Univariate and Multivariate Analysis of Phosphorus Element in Fertilizers Using Laser-Induced Breakdown Spectroscopy

Baohua Zhang [1,*], Pengpeng Ling [2], Wen Sha [2], Yongcheng Jiang [2] and Zhifeng Cui [3]

[1] School of Electronics and Information Engineering, Anhui University, Hefei 230061, China
[2] School of Electric Engineering and Automation, Anhui University, Hefei 230601, China; ppling@yeah.net (P.L.); ahu001@163.com (W.S.); ycjiang@126.com (Y.J.)
[3] Institute of Atomic and Molecular Physics, Anhui Normal University, Wuhu 241000, China; zfcui@mail.ahnu.edu.cn
* Correspondence: qinji1983@126.com; Tel.: +86-551-6386-1237

Received: 5 March 2019; Accepted: 9 April 2019; Published: 11 April 2019

Abstract: Rapid detection of phosphorus (P) element is beneficial to the control of compound fertilizer production process and is of great significance in the fertilizer industry. The aim of this work was to compare the univariate and multivariate analysis of phosphorus element in compound fertilizers and obtain a reliable and accurate method for rapid detection of phosphorus element. A total of 47 fertilizer samples were collected from the production line; 36 samples were used as a calibration set, and 11 samples were used as a prediction set. The univariate calibration curve was constructed by the intensity of characteristic line and the concentration of P. The linear correlation coefficient was 0.854 as the existence of the matrix effect. In order to eliminate the matrix effect, the internal standardization as the appropriate methodology was used to increase the accuracy. Using silicon (Si) element as an internal element, a linear correlation coefficient of 0.932 was obtained. Furthermore, the chemometrics model of partial least-squares regression (PLSR) was used to analysis the concentration of P in fertilizer. The correlation coefficient was 0.977 and 0.976 for the calibration set and prediction set, respectively. The results indicated that the LIBS technique coupled with PLSR could be a reliable and accurate method in the quantitative determination of P element in complex matrices like compound fertilizers.

Keywords: fertilizer; phosphorus element; laser-induced breakdown spectroscopy; chemometrics

1. Introduction

The use of compound fertilizers in agriculture to improve soil quality is very common. China's compound fertilizer use ranks first in the world according to statistics. Fertilization in some areas is extremely unreasonable, causing serious environmental pollution [1]. Phosphorus(P) element is a major nutrient element for crops and is very important in agriculture. Quality control is very important for compound fertilizer manufacturers, and can help guarantee the quality of products. At present, the real-time sensing rapid detection of compound fertilizer production has been mainly manual sampling, sample preparation and laboratory testing. The traditional sensing method of P in compound fertilizers is the phosphomolybdate quinoline gravimetric method [2], which is mature and has high accuracy. However, this sensing detection technique commonly requires the dissolution of the solid sample, which involves the use of high temperatures and strong oxidants. With the development of sensing analysis technology, optical detection methods are increasingly used for the detection of compound fertilizer components, such as flame atomic absorption spectrometry (FAAS), and inductively coupled plasma-mass spectrometry (ICP-MS) [3,4], but, because of a typically high

concentration in compound fertilizer, the sensing measurement of P needs to be diluted several times. The aforementioned sensing detection methods are time-consuming, labor intensive, and expensive. As a consequence, the use of the traditional detection techniques increases the systematic errors besides producing large volumes of chemical residues [5].

Laser-induced breakdown spectroscopy (LIBS) is an emerging analytical technique in the current spectroscopic field. The LIBS technique has many advantages, such as in situ detection, real-time, remote sensing capability, multi-elemental analysis, minimal sample preparation, and direct analysis of any state of matter [6]. It has been successfully used in chemical and biological testing [7], water pollution [8], coal combustion [9], agriculture [10], artifacts and jewelry identification [11], space exploration [12], etc. Some works have used the LIBS technique for the analysis of the component of compound fertilizer. Farooq et al. determined P, Mg, and Mn in the fertilizer using LIBS [13]. Quantitative LIBS analysis of phosphorus in 26 different organic and inorganic fertilizers has been reported by Bruno S. Marangoni et al., however, the absolute error of the measurement for the two verification samples is close to 5% [14]. The elements of Cu, K, Mg, Mn, Zn, As, Cd, Cr and Pb in liquid fertilizers were analyzed with LIBS technique by Daniel Fernandes Andrade et al. [15]. S.C. Yao et al. detected phosphorus and potassium elements in the compound fertilizer using LIBS, and the PLS quantitative analysis model was established by using Unscrambler software [16]. Daniel Fandrade et al. have reported an application of LIBS for quantification of the metal elements in solid compound fertilizers [17]. However, quantitative aspects have generally been considered a shortcoming of LIBS, which greatly limits its application. Thus, there are still many problems to be solved prior to routine practical applications.

The univariate calibration considers the emission intensities of excited element and its concentration. However, the fertilizer was a complex sample, which contains many elements of Fe, Si, Mg, Al, and O. All of these elements may produce matrix effects, and also, due to the fluctuations observed in LIBS technique associated with the instruments and sample non-uniformity, many strategies are used for the calibration methods, such as different spectral preprocessing and multivariate calibration models. Internal standardization is a common method used to minimize fluctuations in LIBS technique, which consists of normalizing the analytical signal by an internal signal. Usually, the internal element concentration must be nearly constant [18]. However, the internal element concentration may slightly change from sample to sample, thus, the accuracy of quantitative analysis results still needs to be improved. In chemometrics, partial least-squares regression (PLSR) is one of the multivariate analytical techniques. It is very crucial to reduce the matrix effect when dealing with complex sample [19].

In this work, 47 fertilizer samples provided by the compound fertilizer production line of Hefei Hongsifang Chemical Fertilizer Plant, Anhui, China were used as the testing samples. The concentration of phosphorus in compound fertilizerswas determined by inductively coupled plasma (ICP). The univariate calibration was established by the LIBS intensity and the concentration of P. Then, the internal standardization method and PLSR were used to quantitatively analyze the phosphorus concentration. The main goal of this research is to prove that LIBS technique can be used for on-line rapid detection of phosphorus element in compound fertilizer.

2. Materials and Methods

2.1. Sample Preparation

In this study, 47 compound fertilizer samples were collected and placed into sealed plastic bags so as to avoid contamination by manufacturer. Since these samples had been pelletized, these samples were smashed by using a grinder and sieved through a 60-mesh screen. 2 g powders from each of 47 fertilizer samples was weighed. All fertilizer powders were pressed into tablets with 25 mm diameter and 5 mm in thickness, using 5 MPa pressure for 1 min (769YP-40C, KQ, Tianjin, China).

The actual concentration of P in these samples were analyzed by inductively coupled plasma (ICP). The statistics of the P concentrations in compound fertilizer samples was listed in Table 1.

Table 1. Statistics of the Effective Constituents of Compound Fertilizer Samples.

Properties	P_2O_5 (%)
Minimum value	46.27
Maximum value	49.22
Mean value	47.731
Standard deviation values	0.617

2.2. Experimental Setup

Figure 1 gave out the spectral acquisition system used in this experiment. The laser pulses were generated by using a Q-switched Nd: YAG laser (ICE450, 1064 nm, 6 ns pulse duration, Big Sky Laser Technologies, Morgan Hill, CA, USA; Note that the company has changed its name to Quantel Laser). The laser pulse energy was 100 mJ and focused onto the sample through a lens of focal length 5 cm. The spot size of beam was approximately 0.5 mm and the peak power density on the surface of the compound fertilizer sample reached 2.2 GW/cm^2.When the plasma generated from the fertilizer sample, a quartz lens with 3.5 cm focal length, which was connected to a four-channel spectrometer (Avantes-ULS2048-USB2, Avantes, Apeldoorn, The Netherlands) via a 200-μm diameter optical fiber, was used to collect the spectra from the plasma. The spectrograph signal was integrated with a charge-coupled device detector. This spectrometer can simultaneously take all spectra in the wavelength ranges of 190–510 and 690–890 nm, and the resolution of the spectrometer was approximately 0.1 nm. The laser Q-switch output was used to trigger the spectrometer, and the spectrometer has a digital delay generator, which can control the gate delay. Here, the Q-switched delay time selected for spectra acquisition was 1.28 μs and the integration time was 1.05ms (spectrometer minimum integration time) [20]. A rotary platform on which the fertilizer sample was placed was rotated uniformly to avoid continuous ablation of the same spot. During the experiment, each sample was measured eight times, and each spectrum collected was an average of 20 laser spot, and during each measurement, the fertilizer sample rotated once by adjusting the speed of the stepper motor.

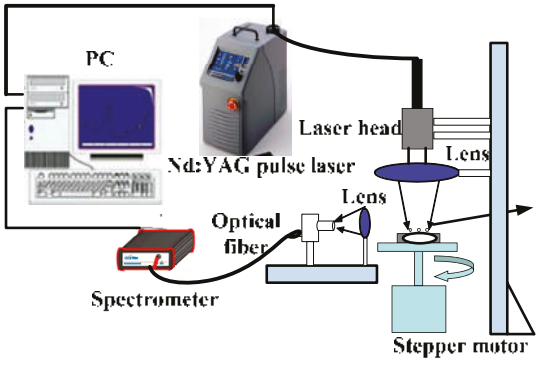


Figure 1. Schematic of LIBS experimental system.

2.3. Chemometrics Methods

In LIBS technique, the calibration curve method, also called univariate analysis, is a traditional quantitative analysis method. The element characteristic line intensity is proportional to the concentration in the sample, when there is no self-absorption [21]. Therefore, the calibration curve can be established by the element concentration and the intensity of LIBS signal. Then, the concentration of

an unknown sample can be calculated according to the calibration curve. However, due to the matrix effect, this method is not suitable for quantitative analysis of complex sample. Internal standardization is usually used for quantitative analysis in LIBS technique, which can improve the accuracy of LIBS technique and reduce the fluctuations observed in LIBS measurements, but, there are some principles for selecting the internal standard element [22]. The concentration of the internal standard element should be approximately constant, and the wavelength of the internal standard element should be close to the analytical element. In addition, the excitation potential should be similar for the internal standard element and the analytical element. The coefficient of determination and the root mean squared error, for the calibration set and validation set, were adopted to evaluate the performance of the internal standardization model.

PLSR is widely used for quantitative analysis of LIBS spectra in recent years [23,24]. This method performs quantitative spectral analysis by selecting latent variables [25,26]. Therefore, it is very important to select the latent variables, which directly determined the predictive performance of the calibration model. The PLSR model was established by the LIBS signal intensity and the concentration of P for fertilizer samples. In order to avoid the overfitting of the PLSR model, and also to obtain a reliable and robust PLSR model, full cross-validation was applied. The number of latent variables was determined when the mean squared error was minimum. Furthermore, the statistic parameters for evaluating the performance of PLSR model include the determination coefficient for calibration (R_C^2) and prediction (R_P^2), the root mean square error for calibration (RMSEC) and prediction (RMSEP), and residual predictive deviation (RPD) [27]. All data processing procedures were compiled with MATLAB.

3. Results

3.1. Spectral Analysis

The LIBS spectrum of the compound fertilizer pellet (No.1 sample) in the ranges of 210–220 and 250–260 nm is shown in Figure 2, which includes the emission lines of silicon (Si) and P. Compound fertilizer production enterprises generally use phosphate ore as raw material. Thus, Si is one of the main ingredients of compound fertilizers, and the characteristic lines of Si are 212.4 nm, 221.1 nm, and 221.7 nm according to the National Institute of Standards and Technology (NIST) database. It can be seen from Figure 2 that the compound fertilizer sample contains abundant characteristic lines of P element. The feature spectral lines of P element detected by LIBS were 213.6 nm, 214.9 nm, 215.4 nm, 253.4 nm, 253.6 nm, 255.3 nm, and 255.5 nm. The characteristic lines of 253.4 nm and 253.6 nm are interfered by the characteristic line of Fe element in the compound fertilizer. The adjacent peaks of the two characteristic lines of 255.3 nm and 255.5 nm can be clearly distinguished, and Lorentz double peak fitting is needed when fitting the line intensity. The characteristic lines of 213.6 nm, 214.9 nm and 215.4 nm are not disrupted by other elements.

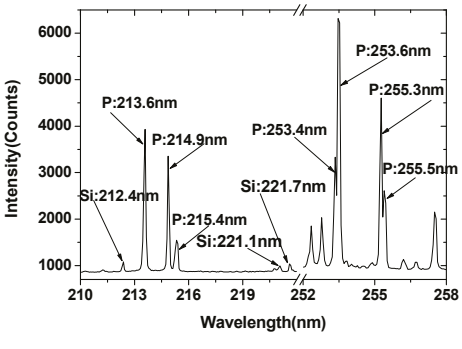


Figure 2. LIBS spectrum of compound fertilizer sample in the ranges of 210–222 and 252–258 nm (n = 20).

In order to obtain a stable signal, the focus of laser beam was adjusted at the position of the fertilizer sample. When the laser focus located on the surface of the fertilizer sample, the distance was recorded as d = 0 mm. Then, the laser focusing system was adjusted, each time moving the laser focus to the surface of the fertilizer sample 1 mm, up to 8mm. The P: 213.6 nm was selected as the analytical line, the relationship of the line intensity and the signal-to-background ratio (SBR) with the laser focus position was shown in Figure 3. The maximum value of the line intensity and the SBR of P were all located at 3 mm below the surface of the fertilizer sample. When the focus of the laser beam gradually moved downward from the surface of the fertilizer sample, the laser pulse energy was more absorbed by the sample, so that the ablation amount was gradually increased, more atoms and ions were in an excited state. But, as the distance between the laser pulse focus and the surface of the fertilizer sample further increased, the radiant power of the laser pulse on the surface of the fertilizer sample gradually decreased, so that the ablation amount of the composite fertilizer sample decreased. However, when the focus of the laser pulse moved down to a certain distance, it was basically difficult to break down the fertilizer sample, so the line intensity and the SBR of the phosphorus element tended to be stable.

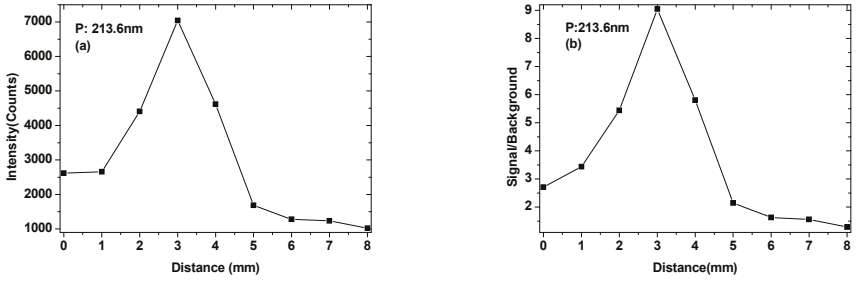


Figure 3. The line intensity of P 213.6 nm (**a**) and signal to background ratio (**b**) as a function of detection distance.

3.2. Univariate Analysis

Before modeling, 47 fertilizer samples were split into a calibration set (36 samples) and a prediction set (11 samples) based on the K-S method. The univariate calibration model was constructed by the line intensity (the height of Lorentz fits) of P versus the corresponding concentration [28]. The content of P is in the range of 46.27–49.22%. Figure 4a–c show the calibration curve of three characteristic lines of P. The spectral line (P: 213.6 nm) that obtained the best modeling result was applied for subsequent analysis. Figure 4a indicates the linear trend between line intensity and concentration with a coefficient of correlationof 0.854. For the prediction set, 11 fertilizer samples were used to estimate the prediction accuracy of the LIBS technology. The predicted content of samples can be obtained by taking the line intensity into the calibration fitting curve. The relation between the reference content and LIBS-predicted content for P is shown in Figure 4d, with an R^2 value of 0.923. Although the relative error of the prediction set is not very large, but the correlation coefficient should be improved for further quantitative analysis.

3.3. Multivariate Analysis

According the principles for selecting the internal standard element, the characteristic line of Si 212.4 nm was selected as the internal standard line. The internal standard curve was constructed by calculating the lineintensity ratio of the analytical element and the internal standard element. The ratio of P line intensity (213.6 nm) to that of Si (212.4 nm) versus the concentration of P in fertilizers was used for calibration. Figure 5a shows the internal standard curve for P element. The value of R^2 was obviously improved to 0.932 from 0.854. Similar to the univariate calibration method, the sample content of prediction set were calculated. Figure 5b shows the relation between the

reference concentration and LIBS predictedconcentration for P, with R^2 changing from 0.923 to 0.946. The range of relative error was 0.04–0.65%, which was improved with the univariate calibration method. These results indicate that the internal standard method can partly eliminate the instability of the LIBS signal, but the detection sensitivity and the prediction accuracy still need to be improved.

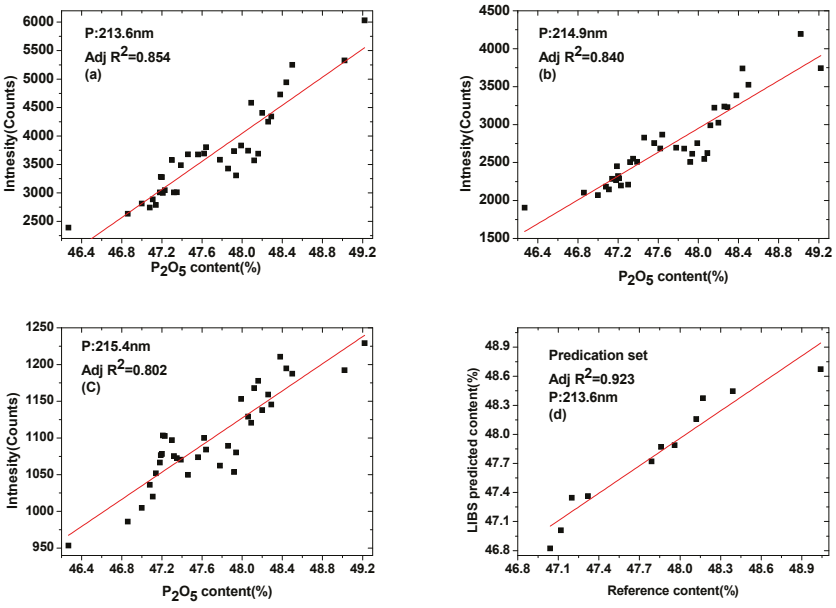


Figure 4. Calibration curves of P spectral line: (**a**) P: 213.6 nm, (**b**) P: 214.9 nm, (**c**) P: 215.4 nm, and (**d**) the relation of LIBS predicted value and reference value for the prediction set.

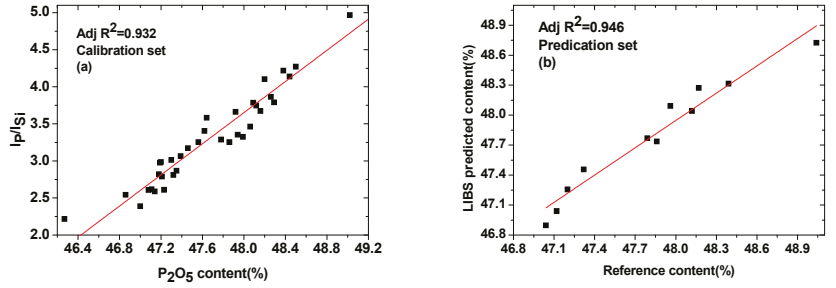


Figure 5. Internal standard method: (**a**) calibration curve using Si as an internal standardelement and (**b**) comparison of P content predicted by LIBS and the referencevalue (ICP).

Because of the complex fertilizer matrix, the analysis focusing only on one line intensity of an element might result in the loss of valid information, which cannot meet the requirements of the quantitative analysis of LIBS. PLSR is one of the multivariate analytical techniques, which can make full use of the spectral information, reduce the matrix effect and improve the accuracy of quantitative analysis. Sample sets were the same as those of the above univariate model and internal standard model. Calibration set was used to construct a model correlating the LIBS signal and the concentration of P; this correlation can later be used to predict concentrations of prediction set. In order to improve the processing speed and avoid overfitting of the model, a proper spectral range was selected for modeling analysis. A reduced spectral range from the full spectrum, which included most of the strong

lines of P, was taken into account for PLSR model. The reduced wavelength ranges from 210 to 260 nm for P was used to obtain calibration model. The MATLAB software was used for PLSR. Seven principal components are used to construct the PLSR model. Figure 6a–b shows the calibration and prediction results of PLSR model for P, respectively.

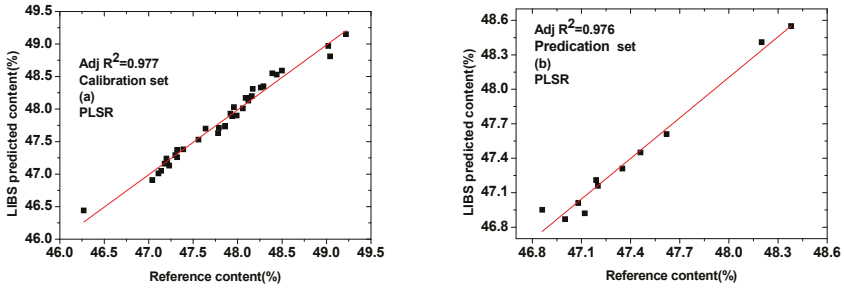


Figure 6. Comparison between LIBS predicted value (PLSR model) and reference value presented in (**a**) thirty-six calibration samples and (**b**) eleven prediction samples.

It can be seen from Figure 6 that most of the calibration and prediction data points were distributed around the fitting curve, which indicating that the PLSR model performed well in predicting of P content. The major statistic parameters that determine capacity of the regression model are R_C^2, R_P^2, RMSEC, RMSEP and RPD. All of the parameters for both calibration and prediction sets are presented in Table 2. The determination coefficient for calibration set (R_C^2) was changed from 0.932 to 0.977, while for prediction (R_P^2) set was improved to 0.976 from 0.946. The slope for both the calibration and prediction sets was close to one, which showed a strong correlation between predicted and reference values. Meanwhile, the value of RMSEC and RMSEP were 0.117 and 0.113, respectively, which was better than the results reported by S.C. Yao et al. [16]. In [16], the RMSEC was 0.234, and in this case, RPD value exceeded 5, which suggesting that the established prediction models can be employed for robust quantitative analysis. In real agricultural applications, the measured value of P content in compound fertilizer is 1.5% plus or minus the standard value, and the absolute difference between different laboratory measurements is not more than 0.5% (ISO 5315: 1984, MOD). In this paper, the difference between the predicted value of LIBS and the reference value isin the range of 0.02%–0.32%, which can fit the requirement. The total relative error for calibration and prediction set was 5.83% and 2.08%, respectively, which was better than the results reported by Bruno S. Marangoni et al. [14]. In [14], the average error of 15% found in cross-validation of LIBS quantification appeared feasible for P quantification in fertilizers. It is demonstrated that PLSR model can be developed to predict the concentrations of unknown samples. Thus, the measurement accuracy of this result can meet the measurement requirements [29].

Table 2. R^2 Regression Coefficients, RMSEC, RMSEP, RPD for Calibration and Prediction Curve.

Parameters	R_C^2	RMSEC	R_P^2	RMSEP	RPD
Values	0.977	0.117	0.976	0.113	5.31

4. Discussion

In this paper, the LIBS technique was used for univariate and multivariate analyses of P element in compound fertilizers. Forty-seven samples were provided by Hongsifang production. The calibration curve was established based on the three selected emission lines and the concentration of P. The results showed that the characteristic line of 213.6 nm was most suitable for establishing calibration curves. Due to the occurrence of matrix effects, the prediction accuracy of the method could not be achieved by applying the univariate calibration method, which only using I_P as the variable. The internal

standardization method based on Si was naturally present in the samples, which showed the proper correction of the P signal. The internal standard method was found to be better than the calibration method because the correlation coefficient for the calibration set was changed from 0.854 to 0.932. Moreover, the range of relative errors are 0.04–0.65%. Thus, the internal standard method can improve the accuracy of the measurements in some extent. Then, PLSR was used as a multivariate analytical technique for analysis of compound fertilizers in pellet form. From the results of PLSR regression, calibration and prediction models were obtained for P element with very good correlation coefficients. The values of RMSEC and RMSEP were 0.117 and 0.113, with RPD value of 5.31. All of these results demonstrated that the PLSR regression method can improve the accuracy of LIBS measurement, and the results in this study can provide the basis of real-time analysis of P in compound fertilizer.

Author Contributions: B.Z., Z.C. and W.S. conceived and designed the experiment; P.L. performed the experiments; Y.J. and W.S. contributed to data analysis; B.Z. and W.S. analyzed the data and wrote the whole paper; all authors reviewed the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (61505001).

Acknowledgments: The fertilizer samples tested in this research was provided by the compound fertilizer production line of Hefei Hongsifang Chemical Fertilizer Plant, Anhui, China.

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

References

1. Niu, X.Y. Research on Variable-Rate Fertilizer Application Technology and Its Applying System in Precision Agriculture. Ph.D. Thesis, Hebei Agriculture University, Baoding, Hebei, China, 2005.
2. Cui, S.L. Effective Phosphorus Detection in Compound Fertilizers by Weight Method. *Mod. Agric. Technol.* **2016**, *5*, 231–232.
3. Yang, Q.Q.; Zhou, C.H. The Test of Potassium Oxide on Mixed and Compound Fertilizer by Flame Atomic Absorption Spectrometry. *Contemp. Chem. Ind.* **2007**, *36*, 209–211.
4. Wang, M.-R.; Yuan, Y.-M.; Tao, N.-L. Determination of Available Phosphate Content in Calcium Magnesium Phosphate by ICP-AES. *Mod. Agric. Sci. Technol.* **2012**, *7*, 20–21.
5. Anastas, P.T. Green chemistry and the role of analytical methodology development. *Crit. Rev. Anal. Chem.* **1999**, *29*, 167–175. [CrossRef]
6. Peng, J.; He, Y.; Ye, L.; Shen, T.; Liu, F.; Kong, W.; Liu, X.; Zhao, Y. Moisture influence reducing method for heavy metals detection in plant materials using laser-induced breakdown spectroscopy: A case study for chromium content detection in rice leaves. *Anal. Chem.* **2017**, *89*, 7593–7600. [CrossRef] [PubMed]
7. Gottfried, J.L. Discrimination of biological and chemical threat simulants in residue mixtures on multiple substrates. *Anal. Bioanal. Chem.* **2011**, *400*, 3289–3301. [CrossRef]
8. Zhang, D.C.; Hu, Z.Q.; Su, Y.B.; Hai, B.; Zhu, X.L.; Ma, X. Simplemethod for liquid analysis by laser-induced breakdown spectroscopy (LIBS). *Opt. Express* **2018**, *26*, 18794–18802. [CrossRef]
9. Yan, C.H.; Qi, J.; Liang, J.; Zhang, T.L.; Li, H. Determination of coal properties using laser-induced breakdown spectroscopy combined with kernel extreme learning machine and variable selection. *J. Anal. At. Spectrom.* **2018**, *33*, 2089–2097. [CrossRef]
10. Jull, H.; Kunnemeyer, R.; Schaare, P. Nutrient quantification in fresh and dried mixtures of ryegrass and clover leaves using laser-induced breakdown spectroscopy. *Precis. Agric.* **2018**, *19*, 823–839. [CrossRef]
11. Anglos, D.; Detalle, V. *Cultural Heritage Applications of LIBS in Laser-Induced Breakdown Spectroscopy*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 531–554.
12. Cho, Y.; Horiuchi, M.; Shibasaki, K.; Kameda, S.; Sugita, S. Quantitative Potassium Measurements with Laser-Induced Breakdown Spectroscopy Using Low-Energy Lasers: Application to In Situ K-Ar Geochronology for Planetary Exploration. *Appl. Spectrosc.* **2017**, *71*, 1969–1981. [CrossRef]
13. Farooq, W.A.; Al-Mutairi, F.N.; Khater, A.E.M.; Al-Dwayyan, A.S.; AlSalhi, M.S.; Atif, M. Elemental analysis of fertilizer using laser induced breakdown spectroscopy. *Opt. Spectrosc.* **2012**, *112*, 874–880. [CrossRef]
14. Marangoni, B.S.; Silva, K.S.G.; Nicolodelli, G.; Senesi, G.S.; Cabral, J.S.; Villas-Boas, P.R.; Silva, C.S.; Teixeira, P.C.; Nogueira, A.R.A.; Benites, V.M.; et al. Phosphorus quantification in fertilizers using laser

induced breakdown spectroscopy (LIBS): A methodology of analysis to correct physical matrix effects. *Anal. Methods* **2016**, *8*, 78–82. [CrossRef]
15. Andrade, D.F.; Pereira-Filho, E.R. Direct determination of contaminants and major and minor nutrients in solid fertilizers using laser-induced breakdown spectroscopy (LIBS). *J. Agric. Food Chem.* **2016**, *64*, 7890–7898. [CrossRef]
16. Yao, S.C.; Lu, J.D.; Li, J.Y.; Chen, K.; Li, J.; Dong, M.R. Multi-elemental analysis of fertilizer using laser-induced breakdown spectroscopy coupled with partial least squares regression. *J. Anal. At. Spectrom.* **2010**, *25*, 1733–1738. [CrossRef]
17. Andrade, D.F.; Sperança, M.A.; Pereira-Filho, E.R. Different sample preparation methods for the analysis of suspension fertilizers combining LIBS and liquid-to-solid matrix conversion: Determination of essential and toxic elements. *Anal. Methods* **2017**, *9*, 5156–5164. [CrossRef]
18. Bechlin, M.A.; Ferreira, E.C.; Neto, J.A.G.; Ramos, J.C.; Borges, D.L.G. Contributions on the use of bismuth as internal standard for lead determinations using ICPbased techniques. *J. Braz. Chem. Soc.* **2015**, *26*, 1879–1886.
19. Clegg, S.M.; Sklute, E.; Dyar, M.D.; Barefield, J.E.; Wiens, R.C. Multivariate analysis of remote laser-induced breakdown spectroscopy spectra using partial least squares, principal component analysis, and related techniques. *Spectrochim. Acta B* **2009**, *64*, 79–88. [CrossRef]
20. Sha, W.; Niu, P.; Zhen, C.; Lu, C.; Jiang, Y. Analysis of phosphorus in fertilizer using laser-induced breakdown spectroscopy. *J. Appl. Spectrosc.* **2018**, *84*, 653–658. [CrossRef]
21. Miziolek, A.W.; Palleschi, V.; Schechter, I. *Laser-Induced Breakdown Spectroscopy (LIBS): Fundamentals and Applications*; Cambridge University Press: New York, NY, USA, 2006.
22. Cui, Z.; Zhang, X.; Yao, G.; Wang, X.; Xu, X.; Zheng, X.; Feng, E.; Ji, X. Quantitative analysis of the trace element in Cu-Pb alloy by the LIBS. *Acta Phys. Sin.* **2006**, *55*, 4506–4513.
23. Nie, P.; Dong, T.; He, Y.; Xiao, S. Research on the effects of drying temperature on nitrogen detection of different soil types by near infrared sensors. *Sensors* **2018**, *18*, 391. [CrossRef]
24. He, Y.; Xiao, S.; Nie, P.; Dong, T.; Qu, F.; Lin, L. Research on the optimum water content of detecting soil nitrogen using near infrared sensor. *Sensors* **2017**, *17*, 2045. [CrossRef]
25. Martín, M.E.; Hernández, O.M.; Jiménez, A.I.; Arias, J.J.; Jiménez, F. Partial least-squares method in analysis by differential pulse polarography. Simultaneous determination of amiloride and hydrochlorothiazide in pharmaceutical preparations. *Anal. Chim. Acta* **1999**, *381*, 247–256. [CrossRef]
26. Sjöström, M.; Wold, S.; Lindberg, W.; Persson, J.Å.; Martens, H. A multivariate calibration problem in analytical chemistry solved by partial least-squares models in latent variables. *Anal. Chim. Acta* **1983**, *150*, 61–70. [CrossRef]
27. Rossel, R.A.V.; McGlynn, R.N.; McBratney, A.B. Determining the composition of mineral-organic mixes using UV-vis–NIR diffuse reflectance spectroscopy. *Geoderma* **2006**, *137*, 70–82. [CrossRef]
28. Amamou, H.; Bois, A.; Ferhat, B.; Redon, R.; Rossetto, B.; Ripert, M. Correction of the self-absorption for reversed spectral lines: Application to two resonance lines of neutral aluminium. *J. Quant. Spectrosc. Radiat. Transf.* **2003**, *77*, 365–372. [CrossRef]
29. Sirven, J.B.; Bousquet, B.; Canioni, L.; Sarger, L. Laser-induced Breakdown Spectroscopy of CompositeSamples: Comparison of Advanced Chemometrics Methods. *Anal. Chem.* **2006**, *78*, 1462–1469. [CrossRef]

sensors

MDPI

Article

The Efficiency of Color Space Channels to Quantify Color and Color Intensity Change in Liquids, pH Strips, and Lateral Flow Assays with Smartphones

Joost Laurus Dinant Nelis [1,*], Laszlo Bura [2], Yunfeng Zhao [1,3], Konstantin M. Burkin [4], Karen Rafferty [3], Christopher T. Elliott [1] and Katrina Campbell [1,*]

1 Institute for Global Food Security, School of Biological Sciences, Queen's University of Belfast, 19 Chlorine Gardens, Belfast BT9 5DL, UK; Y.Zhao@qub.ac.uk (Y.Z.); chris.elliott@qub.ac.uk (C.T.E.)
2 Department of Food and Drug, University of Parma, Parco Area delle Scienze 27/A, 43124 Parma, Italy; laszlo.bura@studenti.unipr.it
3 School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, 125 Stranmillis Road, Belfast BT9 5AH, UK; K.Rafferty@ee.qub.ac.uk
4 Faculty of Chemistry, Lomonosov Moscow State University, 1-3 Leninskiye Gory, GSP-1, Moscow 119991, Russia; burkin-kost@yandex.ru
* Correspondence: J.Nelis@qub.ac.uk (J.L.D.N.); katrina.campbell@qub.ac.uk (K.C.)

Received: 25 October 2019; Accepted: 19 November 2019; Published: 21 November 2019

Abstract: Bottom-up, end-user based feed, and food analysis through smartphone quantification of lateral flow assays (LFA) has the potential to cause a paradigm shift in testing capabilities. However, most developed devices do not test the presence of and implications of inter-phone variation. Much discussion remains regarding optimum color space for smartphone colorimetric analyses and, an in-depth comparison of color space performance is missing. Moreover, a light-shielding box is often used to avoid variations caused by background illumination while the use of such a bulky add-on may be avoidable through image background correction. Here, quantification performance of individual channels of RGB, HSV, and LAB color space and ΔRGB was determined for color and color intensity variation using pH strips, filter paper with dropped nanoparticles, and colored solutions. LAB and HSV color space channels never outperformed the best RGB channels in any test. Background correction avoided measurement variation if no direct sunlight was used and functioned more efficiently outside a light-shielding box (prediction errors < 5%/35% for color/color intensity change). The system was validated using various phones for quantification of major allergens (i.e., gluten in buffer, bovine milk in goat milk and goat cheese), and, pH in soil extracts with commercial pH strips and LFA. Inter-phone variation was significant for LFA quantification but low using pH strips (prediction errors < 10% for all six phones compared). Thus, assays based on color change hold the strongest promise for end-user adapted smartphone diagnostics.

Keywords: smartphone colorimetrics; lateral flow assay quantification; color space; image correction; food contaminant screening; allergens; background correction; point of site analyses

1. Introduction

1.1. General Introduction

Detecting, quantifying, and mitigating against contamination in the food supply chain is paramount to global food security. High-end laboratory equipment such as mass spectrometry is often used for this purpose [1]. Unfortunately, such equipment is often unavailable in the developing world [2]. Moreover, contamination and fraud often go undetected due to a lack of surveillance. A recent report from the European Rapid Alert System for Food and Feed (RASFF) showed that alert notifications

in 2017 increased 26% compared to 2016 and reported on various outbreaks (with pathogens and mycotoxins being most prominent) [3]. The main increase in alerts were follow up alerts representing additional testing on products that might have been placed on the market in another country [3]. This increase in follow up alerts might indicate that the system suffers from limited traceability in the industrialized global market, as has been reported previously [4,5]. Use of on-site detection methods (performed by primary producers, supermarkets, or even consumers) can complement the current systems and provide a means for developing countries to enhance food security. Indeed, a plethora of smartphone-based devices has been developed to complement laboratory-based analyses in several sectors including the food sector. Some examples are the fluorescent detection of antibodies against recombinant bovine growth hormone in milk [6,7], fluorescent detection of *Escherichia coli* in yoghurt and egg [8], colorimetric detection of marine toxins (okadaic acid and saxitoxin) in shellfish [9], aflatoxin B1 in maize [10], hazelnut allergen in cookies [11], and peanut allergen in cookies [12]. Such systems have great potential to influence the future market of food quality analyses particularly if analyses are rapid and straightforward allowing uptake at the consumer level. To this end, especially paper-based colorimetric assays such as the lateral flow assays have great potential since test results can be quantified rapidly with a smartphone [13]. A good example is the system reported by Ross et al., which enabled smartphone-based hazelnut allergen quantification within 2 min after applying the extract to the assay [11]. Other systems that may become rapid after optimization are liquid-based assays such as ELISA for which several examples exist of commercial diagnostic tests, such as various assays for mycotoxin analyses, that can be performed in under 10 min [14]. This being said, the food sector is relatively behind in implementation of smartphone-based technology for rapid/real time detection of contaminants when compared to other sectors such as medicine and environmental contaminant detection and it may be interesting to piggyback on such systems for food contaminant analyses [15].

1.2. Hyphenating Lateral Flow and Other Colorimetric Assays with Smartphones

A LFA is a paper-based platform consisting of a sample pad, a conjugate release pad, a membrane with a test and control line and an absorbent pad, all attached to a backing card. If a liquid sample containing a target is loaded onto the sample pad, it will run over the conjugate release pad and membrane to the absorbent pad by capillary force. In this process, the target can bind/form a complex with a labelled immunoreagent (often gold-nanoparticle, carbon black, or latex bead conjugated antibody) present in the conjugate release pad of the LFA. When this complex arrives at the test line other immunoreagents immobilized there (often antibodies in a sandwich assay set-up) can catch the complex which causes a colored line to appear. A control line is equally formed by a similar immunoreaction of other immunoreagents present in the conjugation pad which can complex specifically with other immunoreagents immobilized on the control line to ensure the assay functioned properly [16]. Thus, LFAs are rapid colorimetric tests with color intensity variation in relation to the target concentration and results can generally be read within 5–10 min [14]. Simply photographing such a test and quantifying the color by image analyses on a user friendly app with a smartphone without compromising the functionality of the phone is attractive [13]. Moreover, this combination has merit since smartphones are ubiquitous, allow for user independent quantification of the LFA and enable real-time and place stamped reporting of the results using the smartphone's wireless connectivity and build-in GPS. Moreover, many LFA have already been commercialized for on-site, non-expert use for food contaminant screening [14]. However, only one commercial smartphone-based LFA reader (RIDA Smart app; R-Biopharm) was identified for food contaminant analyses from the > 300 commercial assays included in the mentioned database [14]. This assay quantifies various LFAs testing for mycotoxins but only works for those specific assays and on a few android-based smartphones. Evidently, consumer friendly rapid on-site analyses would greatly benefit from an app that is compatible with a large variety of phone-models and allows rapid quantification of a larger variety of (commercially available) colorimetric assays with phone A by interpolating on a calibration curve made in the laboratory with phone B. To this end, a universal approach for color quantification

is needed. Color quantification with a smartphone using red green blue (RGB), hue saturation value (HSV), or lightness and chromatic axes A/B (LAB) color spaces have been reported for this purpose and were reviewed recently in the context of smartphone-based biosensors [17], smartphone-based food diagnostics [18], and quantitative LFA [19]. Roda et al. [17] outlines LAB and HSV color spaces as superior to RGB space for measuring small changes in color. However, which of the individual channels should be used or why HSV or LAB color spaces are better than RGB are not mentioned. Moreover, superior performance of the L channel of LAB is equally reported for the smartphone-based detection of hazelnut allergen with LFAs using carbon black based labeling [11] although no comparison with other channels was presented. In another work [20] paper-based detection of the mycotoxin producing black mold (*Stachybotrys chartarum*) is detected by using the R channel. However, no comparison with the B or G channel or other color spaces was reported. In these previous works [18,19] no recommendations on color space use were given although it has been reported that the R channel instead of the combined weighed RGB values lead to background reduction [21]. In another study, the performance of a ΔRGB or ΔLAB system was compared and ΔRGB outperformed ΔLAB for the quantification of plasmonic-ELISA assays [22] and found to perform well in a colorimetric paper-based assay [23]. To obtain Δ values a resultant of R, G, B or L, A, B vectors relative to control values was calculated [22]. However, no comparison between the single R, G, B or L, A, B channels with ΔRGB was performed in either study. Thus, there currently is no consensus regarding which color space/channel should be used for the smartphone image analyses needed for optimal functioning of smartphone hyphenated colorimetric assays such as ELISA, LFA, and other paper-based systems for food contaminant analyses. However, color space/channel choice clearly effects the performance of smartphone-based assays since several of the studies mentioned above reported substantial differences in assay performance in function of the color space/channel used. As a result, conversions to color spaces other than RGB are perhaps executed unnecessarily or suboptimal channels/color spaces may be chosen as a starting point for smartphone-based image analyses based on incomplete recommendations found in the literature. Other issues with current systems are (i) background illumination variation and (ii) inter-phone channel values variation. A light-shielding box is often used to tackle variation in background illumination [9,17–19], however, this detracts from the opportuneness of smartphone-based analyses if add-on items are required. Inter-phone channel value variation is especially important since it is undesirable to develop any system to be used by consumers if bespoke calibration is required for each phone. Unfortunately, most systems described previously were only tested with a single phone [9,11,17–20,24]. Overall, in the scientific literature there appears to be large variations in opinions regarding the use of color spaces and channel combinations. Moreover, no extensive comparison of the individual channels of RGB, LAB, HSV color space, and ΔRGB values with various phones has been identified.

1.3. Workflow Reported in This Study

In the present study, prediction accuracy of individual channels of RGB, HSV and LAB color space, and ΔRGB was determined for the quantification of color variation, using pH strips, and color intensity variation, using filter paper with dropped nanoparticles often used for LFA (i.e., gold, latex or carbon black nanoparticles) and nanoparticle and oxidized tetramethylbenzidine (TMB) solutions in ELISA wells. Background correction was shown to avoid measurement variation in the absence of direct sunlight more efficiently than a light-shielding box. Inter-phone (n = 6) variation was limited for color change quantification, permitting the quantification of color with phone A using a calibration curve constructed with phones B–F. The optimized system was validated using various phones with model food application exemplars in the quantification of gluten in buffer, bovine milk in goat milk and goat cheese, and pH determination in soil extracts with commercial pH strips and LFA.

2. Materials and Methods

2.1. Materials

Dibasic and monobasic sodium phosphate, sodium carbonate/bicarbonate, HCl (37%), 3,3′,5,5′-tetramethylbenzidine (TMB), NaOH, $HAuCl_4$, sodium citrate, $AgNO_3$, L-ascorbic acid, blue latex beads (LNPs), horseradish peroxidase (HRP), gluten from wheat (crude, ≥ 75% protein), and grade 1 Whatman filter paper were purchased from Sigma-Aldrich (Irvine, UK). Carbon black (N220) (CB) was obtained from Cabot Corporation (Ravenna, Italy), ZEU Proteon Gluten Express ZE/PR/GL25 and ZEU IC-BOVINO lateral flow assays for gluten and cow milk protein detection and ZEU Proteon ELISA were obtained from Zeulab (Zaragoza, Spain). Cow milk and goat cheese were purchased in the local market. Pure goat milk was home produced. DUS alkaline pH strips (5.00–8.50) were purchased from DFI (Gimhae, Korea). Soil was collected in Queen's University garden (Belfast, UK). UV-Vis measurements were performed using a Tecan Safire IIplate reader. Smartphone measurements were performed with a Huawei P8 Lite (12 megapixels (MP)), iPhone 7 (12 MP), Samsung Galaxy Tab E (5 MP), Xiaomi mi5 (16 MP), HTC One M7 (4 MP) and a Samsung Galaxy J7 (13 MP), and a Smartscope Xscience (Ravensburger, The Netherlands) smartphone loop with 3D printed lenses for magnification where mentioned.

2.2. Nanoparticle Synthesis

Glassware was cleaned with piranha solution and aqua regia to remove all residues. The Turkevich method [25] was used for gold nanoparticle (GNP) synthesis. Briefly, 500 μL of 100 mM $HAuCl_4$ together with 194.5 mL MQ was brought to boil, while stirring, in a round bottom flask equipped with a condenser. At boiling point, 5 ml sodium citrate solution (1% (w/v)), was added and the mixture was left boiling for 30 min then cooled down gradually. Stock GNP concentration (2.8 nM) was estimated using a protocol detailed in [26] and size estimations were reported elsewhere [27].

2.3. pH System (Color Change)

Citrate-phosphate buffers (0.1 M; pH 5.0, 5.5), phosphate buffer (0.1 M; pH 6.0–7.5), and carbonate/bicarbonate buffers (0.1 M; pH 8.0, 8.5) were used for calibration (pH 5.0, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5) and prediction (pH 5.5, 6.25, 6.75, 7.25, 7.75, 8.25) curves. For buffered soil extracts, 5 g of soil was added to 25 mL of buffer and vortexed. The pH was adjusted to match the buffer pH described above. Samples were allowed to settle 30 min and directly used for pH measurements. Images taken with the Huawei of pH strips dipped in various buffers are shown in Figure 1.

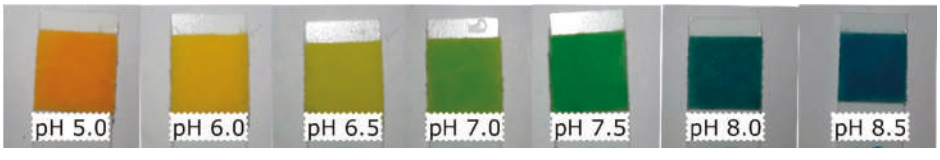


Figure 1. Exemplary images taken with the Huawei of pH strips that were used to build calibration curves.

2.4. Nanoparticle Suspensions and Filter Paper Preparation

A 40X concentration was obtained from a stock GNP solution by centrifuging (13,000 RCF; 30 min; 20 °C). Concentrates were used to prepare a 2/3 dilution series from 37.5 up to 2.195 nM to construct calibration curves and a 2/3 dilution series from 31.25 up to 2.74 nM for predictions. For CB, a 10 mg/mL dispersion was sonicated 30 min then diluted to a 2 mg/mL dispersion and sonicated for another 10 min. Serial dilutions (2X) were made until 0.0078125 mg/mL. For predictions, a 1.5 mg/ml CB solution was made and diluted in 2X steps until 0.0117 mg/mL. LNP stock concentration (2.5%) was 2X diluted until

0.00977% to construct calibration curves. For the prediction set a 1.875% solution was diluted in 2X steps until 0.00732422%. Images taken with the Huawei of the nanomaterials on filter paper are shown in Figure 2.

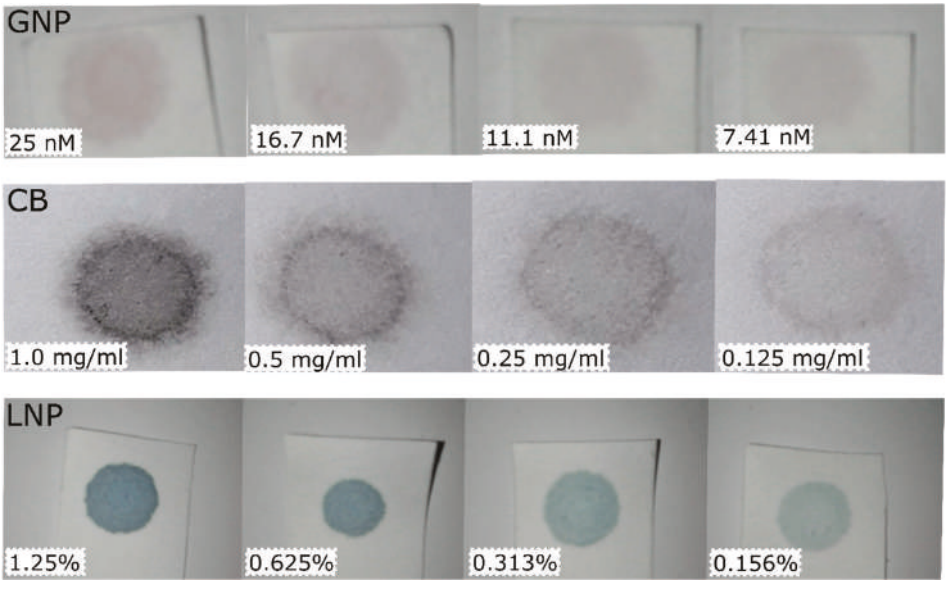


Figure 2. Exemplary images taken with the Huawei of the filter papers with nanomaterials on them at various concentrations. Top row are gold nanoparticles (GNP). Middle row are carbon black nanoparticles (CB). Bottom row are latex nanoparticles (LNP). Particle concentration of the solutions used is indicated in the left bottom corner of each image.

2.5. Liquid Assays Preparation

Colloid GNPs (200 µL) at varying concentration (8 step, 2X serial dilution from 84 nM for calibration and 7 step 2X dilution from 42 nM for prediction) or 150 µL of TMB with HRP at varying concentration (8 step 2X serial dilution from 60 pM for calibration and 6 step 2X dilution from 40 pM for prediction) were pipetted into transparent 96-well plates. TMB enzymatic reaction was stopped after 30 min with 4N H_2SO_4 (50 µL/well). Absorbance was read at 450 nm for plates with HRP and 513 nm (plasmon peak) for plates with colloid GNPs. Smartphone pictures were taken thereafter. An exemplary image taken with the Xiaomi of the colloid GNPs is shown in Figure 3.

2.6. Sample Preparation and Picture Capturing

From GNP, CB, and LNP dilutions 5 µL was dropped on filter paper (n = 3), dried, and photographed. pH strips were immersed in buffer, dried, and photographed after 40 s. LFA test strips were photographed 10 min after exposure to extracts. For liquid assays, a phone with white screen was placed under the 96-well plate to provide counter illumination and avoid reflections in the images. All pictures were taken from 5 cm distance with the flashlight on (Figure 4). For background illumination experiments the following light changes were tried: Dark background in a closed windowless laboratory, normal room light (TL light), indirect sunlight in a windowsill, direct sunlight in a windowsill. A black cardboard box (11 × 11 × 5 cm) was made for the Huawei (Figure 4). The box had a hole precisely in the center for the camera and flash. Prediction images taken under varying background illumination were interpolated on calibration curves constructed at room light conditions to test robustness of the background correction applied. All images taken in the box for prediction

were interpolated on a calibration curve constructed with images taken in the box at identical light conditions (room light).

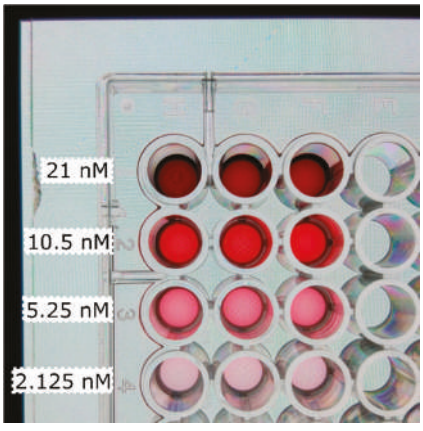


Figure 3. Image taken with the Xiaomi of colloid gold nanoparticles in a 96-well plate. Concentrations vary per row (as indicated). Each row (1–4) contains three replicas (H–F) used to construct the calibration curves. Backlight is provided using the white screen of the Huawei placed under the 96- well plate.

Figure 4. Left; an image of the Samsung taking an image of a LFA (the ZEU Proteon Gluten Express) under room-light conditions with the flashlight on at 5 cm distance. Right; an image of the light-shielding box used. The dimensions of the box are 11 × 11 × 5 cm to maintain the standard 5 cm as a distance for the photo capturing. The box had a hole in the center precisely for the camera and flash. The phone displayed on the box is the Huawei.

2.7. Commercial Assays

LFA for gluten, bovine milk, and cheese detection were used according to the instructions of the manufacturer. Briefly, (for estimation of gluten with the commercial LFA assay; ZEU Proteon Gluten Express ZE/PR/GL25), 1 g of gluten was extracted with 10 mL of the given extraction solution, vortexed and centrifuged for 10 min at 3500 RCF. The supernatant was diluted with analysis buffer to obtain 10, 5, 2.5, 1.25, 0.75, 0.5, 0.025, 0.01, and 0.005 ppm of gluten for calibration curve construction and 2, 0.5, 0.3, 0.15, 0.02, and 0.0075 ppm for predictions. The test strip was immersed in 250 µL of these dilutions for 10 min and photographed. For the estimation of cow milk/cheese in goat milk/cheese the ZEU-IC-BOVINO LFA assay was used. Cow milk was spiked into pure, home produced goat milk at 5%, 2.5%, 1.25%, 0.625%, 0.3125%, 0.15625%, 0.078125%, 0.04%, and 0.02% for calibration curve

construction and at 3.75%, 1.875%, 0.46875%, 0.234375%, and 0.117188% for predictions. LFAs were immersed in two drops of these extracts diluted with three drops dilution solution. Purity of the goat cheese used was tested using the RC-BOVINO ELISA kit specific for the IgG of cow milk following the manufacturer's instructions. Next extracts of cow cheese and pure goat cheese were prepared by adding 5 g of homogenized cheese to 10 ml H_2O and vortexing. Extracts were centrifuged (10 min; 3000 RCF) and goat cheese supernatant was spiked with cow cheese supernatant (final concentrations 8%, 4%, 2%, 1%, 0.75%, 0.5%, and 0.1% of cow cheese) and used for calibration curve generation. Final concentrations of 6%, 3%, 1.5%, and 0.375% cow cheese were used for predictions. LFAs were immersed in one drop of extract diluted with two drops of the given dilution solution and photographed after 10 min. For all commercial assays duplicates were used. Figure 5 shows an image of a set of LFAs used to build a calibration curve for gluten quantification.

Figure 5. A set of LFAs used to build a calibration curve for gluten quantification with various smartphones. The concentrations of gluten used are indicated.

2.8. Scoring System

A scoring system was used to rate the performance of the channels which allowed to generate calibration curves with an $R^2 > 0.80$ (R, G, B, V, L, H, and V). B of LAB (called C of LAC from hereon to avoid confusion with the B channel of RGB) was not considered since it performed only for pH and even there was outperformed by all other channels. R^2 values on the regression functions fitted to the predicted concentrations were ranked from highest to lowest. Lowest being 1 point, highest 7 points. The same scoring system was used for the slopes of the regression fits. The channel with a slope closest to 1 got 7 points the next in line 6, etc. If no regression line could be fitted through the predictions the score was 0 points both for prediction and slope. Slope and R^2 scores were summed up to get a final score for each channel.

2.9. Software and Data Treatment

Standard RGB values were detected with a free app from Google Play (RGB Android) (n = 3 per image) and converted to HSV and Cielab with the open source library of colormine.org (last accessed 11 August 2019). S values were rescaled to have a 1–100 range (instead of 0–1) to match RGB scale for visualization. For A and B values 128 points were added to allow a scale from 0–256 and avoid negative numbers complicating the background correction applied.

ΔRGB was calculated as specified in [22]. The applied image background correction was as follows:

$$\text{Corrected channel values} = (\text{Raw signal value})/(\text{Raw background value}) \times \text{Raw signal value} \quad (1)$$

For paper-based assays white paper was used as the background except for the LFA where control lines were used instead. For liquid assays, wells filled with water adjacent to the tested wells were used as background. All channel values mentioned in the results are corrected channel values. Data analyses was performed with GraphPad Prism 6 Software. For post-hoc analyses, a one-way ANOVA Tukey's multiple comparison test was used. For two-way ANOVA Sidak's multiple comparison test was used. *p*-values were corrected for multiplicity. For calibration curves, a four-parameter dose-response curve was used unless mentioned otherwise. Mean average error percentile (MAE) was calculated on the totality of predictions over the concentration ranges mentioned above. LOD, IC_{50}, and linear range were obtained by interpolating 90%, 50%, and 20%–80% signal values from fitted normalized curves.

3. Results and Discussion

3.1. Comparing Channel Performance on a Huawei P8

Various concentrations (n = 3) of GNP, CB, and LNP nanoparticles were dropped on filter paper and photographed. R, G, B, values of all pictures were extracted and used to calculate L, A, B, H, S, V and ΔRGB values. The values were fitted to either a four-parameter dose-response curve or a two-phase decay function, which ever resulted in the best fit in terms of R^2 value (Table 1). The fitted calibration curves (with an $R^2 > 0.8$) are depicted in Figure 6, left column. Of the LAB color space (called LAC to avoid confusion of B channels with RGB); only L generated adequate calibration curves in all color systems. A curve in all color systems had an $R^2 < 0.8$ and were not used for predictions. C allowed generating a curve for pH determination only and symmetry was observed which made it only possible to predict pH below pH 7. Thus, C was not used for predictions. However, combining A and C might enable color change quantification, as shown previously [28]. From the HSV color space S never reached R^2 values above 0.8 and was not used. H was effective in generating calibration curves for pH (as previously shown [29]) and CB, although errors were large in the latter color system with a mean average error (MAE) on all predictions (n = 18) of 90% ± 113%. The V channel worked well, producing calibration curves with $R^2 > 0.95$ in all color systems except for GNP where R^2 was 0.88. Prediction images of various pH, GNP, CB, and LNP concentrations were taken and interpolated using the fitted calibration curves of Figure 6. Predicted values are plotted in a scatter plot (Figure 6; right column). The R^2 and slope values of the regressions are shown Table 2. For pH predictions, most channels shown in Figure 6 worked for predictions between pH 8.25 and 6.5. At the lowest pH tested (5.5) only R and ΔRGB allowed good predictions. For variation in nanoparticle concentrations low concentrations caused higher variance in all channels. H performed particularly badly in CB predictions. Larger variation in prediction was observed for the color intensity variations as for color change (MAE for the best channels were typically around 25%–35% ± 20%–40% and 1%–2.5% ± 1%–1.5%, respectively). See Section 3.2 for more detail and MAEs for various phones. This might be explained partly by the greater variance in the hand-made replicas for the latter system and partly by the potentially greater effects of background illumination variation on these tests (which was investigated in Section 3.3). To tease apart differences a scoring system (see Section 2.8) was adapted to rank the performance of the channels for each color system individually as well as for all color systems tested (Tables 2 and 3). B, R, and ΔRGB scored the highest overall score, followed by V, L, G, and finally H. H underperformed since that channel was only effective in pH predictions. Moreover, in three out of four color systems either R, B, or ΔRGB had a top score (albeit shared with L in some cases, Table 2). For LNP predictions, the V channel scored highest. However, the R^2 values and slopes of V, L, G, and B were close to each other (Table 2) and differences may only reflect random variation. In any case, this comparison shows that conversion to HSV and LAC color space did not increase performance compared to RGB in all color systems tested.

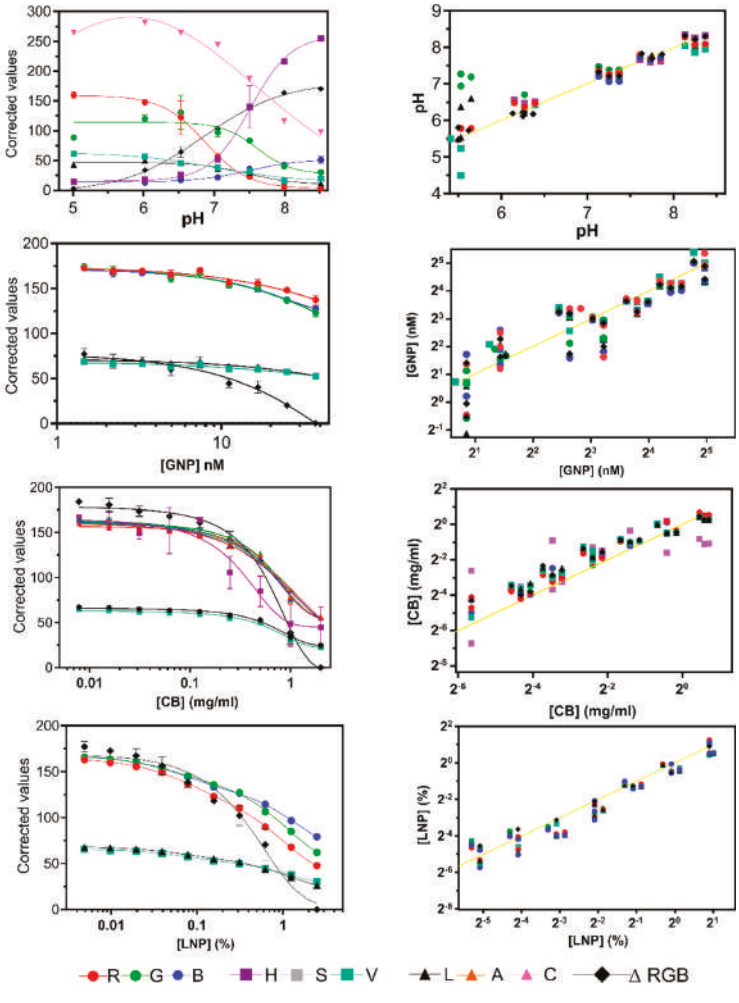


Figure 6. The left column shows calibration curves (with an $R^2 > 0.8$) obtained from background corrected channel values from the RGB, HSV, and LAB (Called LAC to avoid confusion of the B channels of RGB and LAB) color spaces for various pH values, GNP, CB, and LNP concentrations. The right column shows scatter plots of predictions obtained with those calibration curves. The yellow line represents a perfect correlation (slope = 1). Color and symbol codes are indicated.

Table 1. Function types used to fit data points of the papers with varying GNP, CB, LNP concentration, and pH strips measured in various buffers. R^2 values for each fit are mentioned. The fits with $R^2 > 0.80$ were used as calibration curves in Figure 6. C stands for the B channel of LAB which is called LAC throughout the manuscript to avoid confusion of the B channel of LAB with the B channel of RGB. Sigmoidal stands for four-parameter dose-response.

	GNP		CB		LNP		pH	
Channel	**R^2**	**Function**	**R^2**	**Function**	**R^2**	**Function**	**R^2**	**Function**
R	0.8871	Sigmoidal	0.9856	Sigmoidal	0.9847	Two phase decay	0.9795	Sigmoidal
G	0.9365	Sigmoidal	0.9859	Sigmoidal	0.9830	Two phase decay	0.8350	Sigmoidal
B	0.9383	Sigmoidal	0.9877		0.9818	Two phase decay	0.9531	Sigmoidal
ΔRGB	0.9374	Sigmoidal	0.9870	Sigmoidal	0.9867	Two phase decay	0.9958	Sigmoidal
H	-	-	0.8862	Sigmoidal	0.6042	Two phase decay	0.9844	Sigmoidal
S	0.6895	Sigmoidal	-	-	0.7707	Sigmoidal	0.1878	Sigmoidal
V	0.8794	Sigmoidal	0.9836	Sigmoidal	0.9803	Two phase decay	0.9600	Sigmoidal
L	0.9330	Sigmoidal	0.9861	Sigmoidal	0.9811	Two phase decay	0.9610	Sigmoidal
A	-	-	-	-	-	-	0.7979	Sigmoidal
C	-	-	-	-	-	-	0.9941	Sigmoidal

Table 2. Slopes and R^2 scores of log–log regression functions fitted to predictions of GNP, CB, LNP concentrations, and pH values shown in the scatter plots of Figure 6. A scoring system was used to attribute scores to R^2 and slope values for each channel (see Section 2.8). Scores (Table 3) determined ranks for each color system, as well as total scores per channel. Δ is ΔRGB. Ch is channel.

	GNP			CB			Latex			pH			Total Scores
Rank	**Ch**	**R^2**	**Slope**	**Ch**	**R^2**	**Slope**	**Ch**	**R^2**	**Slope**	**Ch**	**R^2**	**Slope**	**Ch**	**Score**
1	Δ	0.9173	1.008	B	0.979	0.801	V	0.951	1,012	Δ	0.987	0.974	B	37
2	L	0.9008	0.996	R	0.977	0.814	L	0.949	1,049	R	0.996	0.860	R	36
3	G	0.9164	0.972	G	0.973	0.795	G	0.947	1.028	H	0.977	0.874	Δ	35
4	B	0.8856	0.996	L	0.974	0.658	B	0.944	1.051	V	0.927	1.043	V	34
5	R	0.895	1.168	Δ	0.971	0.677	R	0.927	1.074	B	0.973	1.239	L	34
6	V	0.8678	1.058	V	0.972	0.676	Δ	0.762	1.178	L	0.898	0.603	G	32
7	H	-	-	H	0.462	0.354	H	-	-	G	0.712	0.307	H	12

Table 3. Scores of the individual channels in each color system, as well as overall scores for all color systems. Shared scores are in bold.

GNP		CB		LNP		pH		Total	
Channel	**Score**	**Channel**	**Score**	**Channel**	**Score**	**Channel**	**Score**	**Channel**	**Score**
ΔRGB	**12**	B	**13**	V	14	ΔRGB	13	B	37
L	**12**	R	**13**	L	**11**	R	11	R	36
G	10	G	9	G	**11**	H	10	ΔRGB	35
B	9	L	7	B	8	V	9	V	**34**
R	6	ΔRGB	6	R	6	B	7	L	**34**
V	5	V	6	ΔRGB	4	L	4	G	32
H	0	H	2	H	0	G	2	H	12

3.2. Comparing Channel Performance between Phones

The experiments detailed in Section 3.1 were repeated on a tablet and iPhone to test if channel performance behaved similarly over various digital camera devices. Figure 7 shows a comparison of the total mean average percentile error (MAE) for the predictions (pH n = 15; GNP, LNP, and CB n = 18) over various concentrations. MAEs for pH calculations were low and stayed below 2.5% for R and ΔRGB for all phone models. For color intensity, low-end MAEs were typically 20%–30% ± 20%–30% for GNP and LNP predictions. For CB higher MAEs were observed although best functioning channels (B and ΔRGB) showed MAEs below 50%. Two-way ANOVAs (Table 4) show that for pH the variance in MAE caused by channels was significant ($p = 0.0002$), as well as interaction between channels and phone models ($p < 0.006$) although phone models alone did not cause significant variance in MAE ($p = 0.24$). For (LNP), (GNP), and (CB) predictions the phone model caused significant differences ($p < 0.0001$) as well as channel and interaction ($p < 0.0001$) in all but the GNP dataset, where only the phone model effects on MAEs was significant ($p < 0.0001$). Thus, choosing any specific channel as universally ideal for smartphone colorimetric analyses or even for a color specific system seems challenging, especially for color intensity change. Nonetheless, Sidak's post-hoc simple effects within rows of multiple comparisons showed that most of the different effects between phones was limited to one or another channel of the RGB color system, or the H channel (for CB predictions). For ΔRGB no significant effects were observed in this test and thus shows that an error in one of the RGB channels can be compensated in this model, which makes it an interesting option. L and V equally show little variation between phones. However, variance on the MAE as well as absolute errors in the L and V values, although not significant, was larger when compared to the best functioning channel of RGB or ΔRGB in each color system (for L in pH predictions the highest MAE is 3.8% ± 8.4% versus 1.8% ± 1.7% for ΔRGB and for CB predictions the MAE was approximately 40%–50% ± 30% for B and ΔRGB and 50%–80% ± 30%–70% for L and V). Overall, H channel performed poorly in the color intensity experiments and did not outperform R, G, B, V, or L channels for color change predictions. Thus, H was no longer used in the following experiments.

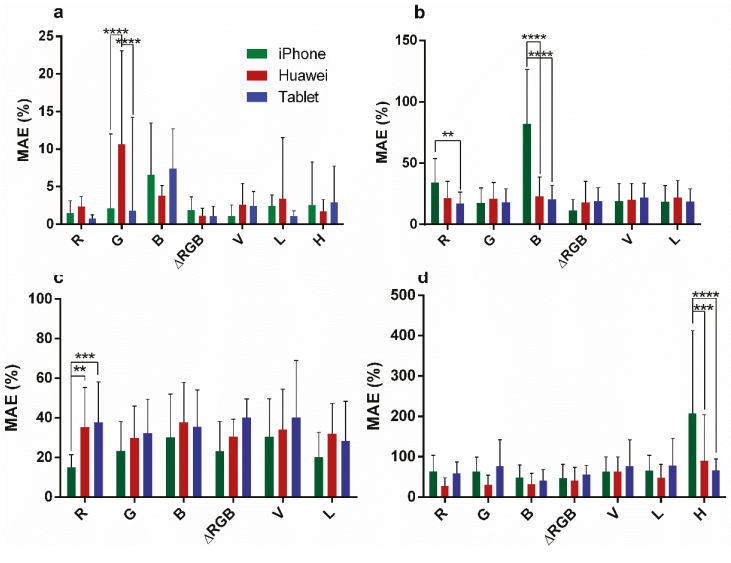


Figure 7. Mean average percentile errors (MAEs) for pH (**a**), (LNP) (**b**), (GNP) (**c**), and (CB) (**d**) predictions using an iPhone (dark green), Huawei (dark red), or Tablet (blue). Significant post-hoc Sidak multi comparisons of two-way ANOVAs are indicated. Stars indicate *p*-values with *p*-value correction for multiplicity. ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$.

Table 4. Two-way ANOVA on phone model and channel for GNP, CB, LNP, and pH predictions.

	GNP		CB		LNP		pH Original	
Source	**Var**	***p*-Value**	**Var**	***p*-Value**	**Var**	***p*-Value**	**Var**	***p*-Value**
Interaction	2.8	0.45	8.7	0.0001	25.7	<0.0001	8.1	0.0058
Channel	2.6	0.10	7.1	<0.0001	15.0	<0.0001	7.8	0.0002
Phone	7.9	<0.0001	5.7	<0.0001	5.2	<0.0001	0.8	0.24

3.3. Comparing Box/No-Box Effects on Predictions in Various Background Settings

The effectiveness of the internal background correction was tested, with a Huawei, by comparing MAEs for pH and [GNP] prediction under varying background illumination with predictions when a light shielding box was used. Channels R, G, B, ΔRGB, V, and L were used since H has proven only functional for pH measurements and even for that application, performance was suboptimal. Little variation in MAEs was observed for pH estimation using the channels R and ΔRGB, even at illumination in direct sunlight (Figure 8). Moreover, two-way ANOVA was only significant for channels and not for interaction or background illumination (Table 5). Interestingly, for R and ΔRGB predictions variation was bigger using the box compared to all background illumination conditions. This may be explained by extensive scatter and unequal light distribution within the box compared to when no box was used. If the background varies between two images but stays equal throughout the individual images then internal background correction should largely correct for it. However, the error can be introduced if background illumination within a picture is patchy. This may also explain the observation by Masawat et al. that a larger box produces less error on predictions as a small one [24]. For [GNP] predictions, no difference was observed between box and no box over the conditions dark, room light, and indirect sunlight for all channels. However, in direct sunlight the background correction applied ceases to function and two-way ANOVA and post-hoc multiple comparisons were highly significant ($p < 0.0001$) for background illumination and direct sunlight, respectively (Table 5). At all other background illumination conditions, the MAE and variation on the MAE were similar in all conditions, including using the box. Thus, the simple background illumination correction applied eliminates the necessity to use a box for all color systems tested if measuring in direct sunlight is avoided.

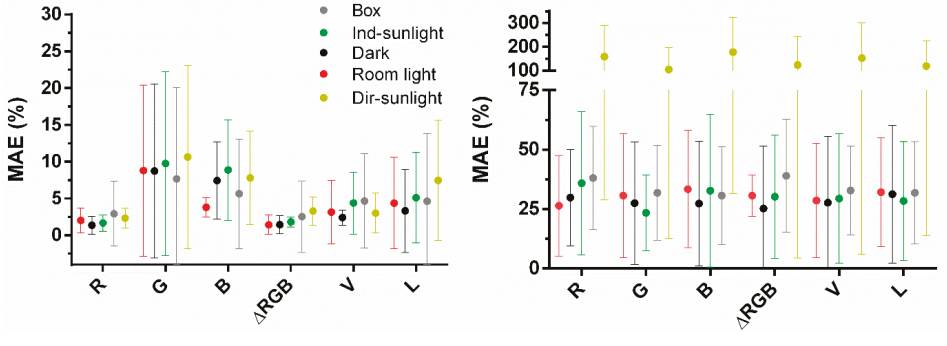


Figure 8. Left, MAEs for pH predictions (n = 15), right, MAEs for [GNP] predictions (n = 18), at various illumination conditions for R, G, B, ΔRGB, V, and L channels without use of a box and in a box at room illumination conditions. Images under dark (black balls) background were taken in a dark room; room background (red balls) in a windowless laboratory illuminated with a tube light bulb; indirect sunlight (green balls) in a windowsill at a cloudy day; direct sunlight (dark yellow balls) in a windowsill in full sunlight.

Table 5. Two-way ANOVA on background illumination and channel for GNP, and pH predictions.

	pH		GNP	
Source	**Var**	***p*-Value**	**Var**	***p*-Value**
Interaction	1.652	0.9893	2.177	0.6248
Channels	13.24	<0.0001	0.6425	0.4002
Light conditions	0.9718	0.3046	39.7	<0.0001

3.4. Channel Performance for Colored Liquids

Channel performance to quantify color intensity in liquid solutions (as typically done for ELISA) was investigated using various concentrations of colloid GNPs (mimicking plasmonic ELISA) and various amounts of HRP to oxidize TMB (mimicking standard HRP-based ELISA). Color change was measured using a benchtop spectrometer and a smartphone (Xiaomi) using R, G, B, ΔRGB, V, and L channels. All four-parameter dose response curves showed good fits ($R^2 > 0.8$) and were used to predict GNP and HRP concentrations. Predictions and linear regressions are shown in Figure 9a,b. Slopes and R^2 values of the calibration curves and R^2 and slope values of the linear regression functions are shown in Table 6. For [HRP] prediction, the spectrometer, B, G, and ΔRGB channels were acceptable. For [GNP] prediction the spectrometer and G, B, ΔRGB, and L channels were acceptable (Figure 9a; Table 6). For comparison, the calibration curves of these channels and the spectrometer were normalized (Figure 9c,d). LOD, linear range, and IC_{50} for both assays are shown in Table 6. Interestingly, G, B, and ΔRGB channels for colloid [GNP] had a ~3X lower LOD as the absorption curve of the spectrometer (~1.5 versus ~7.5 nM, respectively). For the [HRP] curves, B and ΔRGB channels gave lower LODs as the spectrometer (~2.3 versus ~3.4 pM, respectively). Thus, the smartphone-based system was slightly more sensitive than the spectrometer. However, the linear range was reduced compared to the spectrometer. MAEs were calculated for G, B, ΔRGB, and spectrometer [HRP] predictions. For G and B only three concentrations ($n = 3 \times 3$) were used since the linear range was reduced. For ΔRGB and spectrometer four concentrations ($n = 4 \times 3$) were used. One-way ANOVA showed that MAEs for [HRP] prediction using G and B channels did not significantly differ from the MAEs for [HRP] prediction when the spectrometer was used (although with reduced range). ΔRGB had slightly higher MAEs ($p < 0.05$ Tukey post-hoc; Figure 9c inset). This is probably due to the error introduced into ΔRGB from the R channel which varied significantly and could not be used to build a calibration curve. For [GNP] predictions ΔRGB and L channels had significantly higher MAEs as G and B and the spectrometer ($p < 0.01$; Tukey post-hoc; Figure 9d inset). Again, the R channel could not be used to build a calibration curve.

3.5. Channel Performance Comparison Using Commercial LFA for Milk Allergen Detection

Commercial LFA for the quantification of bovine milk in goat milk and cow cheese in goat cheese (Figure 10a,b) were used to further challenge the smartphone-based quantification using the Huawei. For the LFA used to quantify bovine milk in goat milk (Figure 10a) G, B, L, and V channels showed promising calibration curves ($R^2 > 0.9$). However, the curve determined with the V channel values was flat (min–max difference about 25 corrected value units). Equally, the L channel was quite flat and had slightly lower R^2 values (0.91) as B and G (0.93 and 0.94, respectively). ΔRGB did not allow construction of a calibration curve, which was most likely caused by variations observed in the R channel values. Thus, only B, G, and L curves were used for predictions. Linear regressions on predictions were good for each channel ($R^2 > 0.95$). However, the L channel did not predict at 0.23% cow milk (Figure 11). Goat cheese extract was spiked with cow cheese and used to test the applicability of the assay for the identification of cow protein in goat cheese. Here, only the B channel gave a satisfactory calibration curve ($R^2 = 0.951$; Figure 10b). Predictions showed excellent linear regression ($R^2 > 0.98$) (Figure 10b, inset).

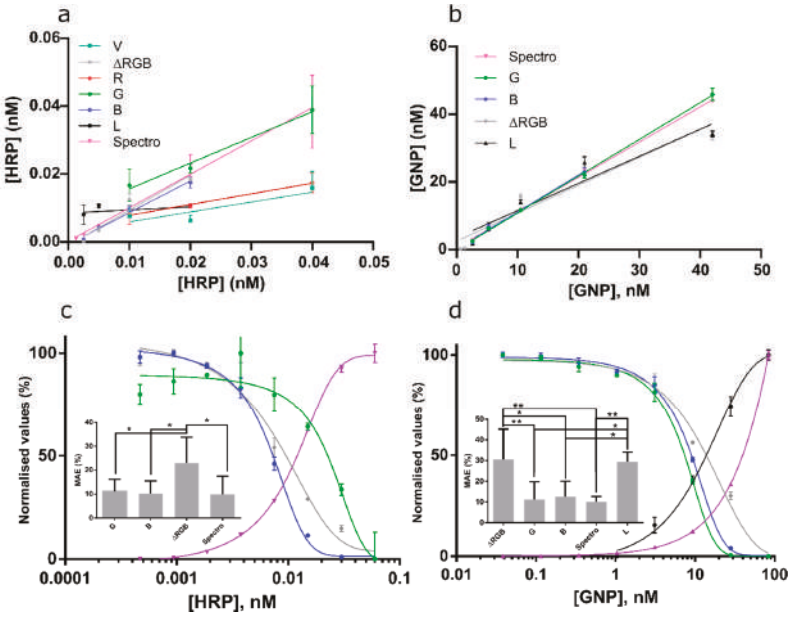


Figure 9. (**a**) Linear regression functions for oxidation of tetramethylbenzidine (TMB) with horseradish peroxidase (HRP) in ELISA wells for the smartphone analyses using various channels and a benchtop ELISA plate spectrometer. (**b**) Linear regression functions for colloid GNP at various concentrations in ELISA wells for the smartphone analyses using various channels and a benchtop ELISA plate spectrometer. Color codes and channels used are indicated. (**c**,**d**) normalized four-parameter dose-response curve fits for HRP oxidation of TMB (*a*) and colloid GNP (*b*) in ELISA wells. Green and blue circles stand for G and B channels, grey diamonds ΔRGB, magenta triangles absorption values measured by a spectrometer. (**c**,**d** inset) MAEs calculated from the predictions shown in (**a**,**b**). Stars indicate *p*-values from post-hoc analyses with *p*-value correction for multiplicity. * = $p < 0.05$, ** = $p < 0.01$.

Table 6. Analytical parameters of prediction and calibration curves for GNP and TMB solution ELISA. R^2 and slopes given for prediction curves are from linear regression lines. All calibration curves were prepared using four-parameter dose-response functions. LOD, linear range, and IC_{50} values were interpolated from the normalized calibration curves at 90%, 80%–20%, and 50% for R, G, B, and ΔRGB channels and 10%, 20%–80%, and 50% for spectrometer values, respectively.

		Prediction Curves		Calibration Curves (nM)			
Solution Type	**Channel**	**R^2**	**Slope**	**R^2**	**LOD**	**Linear Range**	**IC_{50}**
GNP	G	0.9966	1.086	0.9969	1.5	3.1-12.8	7.4
GNP	B	0.9890	1.089	0.9973	1.8	3.7–16.2	9.0
GNP	ΔRGB	0.9094	0.824	0.9890	1.6	4.1–32.7	13.8
GNP	L	0.913	0.8009	0.8947	-	-	-
GNP	spectro	0.9989	1.039	0.9996	7.5	15.3–65.7	39.7
HRP	B	0.9742	0.932	0.9957	0.0025	0.0038–0.012	0.0073
HRP	R	0.6087	0.601	0.829	NA	0.010–0.038	0.027
HRP	G	0.8365	0.506	0.9377	NA	0.0083–0.036	0.023
HRP	ΔRGB	0.9099	0.734	0.9804	0.0023	0.0039–0.020	0.010
HRP	L	0.07789	0.08	0.9307	-	-	-
	V	0.532	0.29	0.8992	-	-	-
HRP	spectro	0.9752	1.153	0.9982	0.0034	0.0057–0.021	0.012

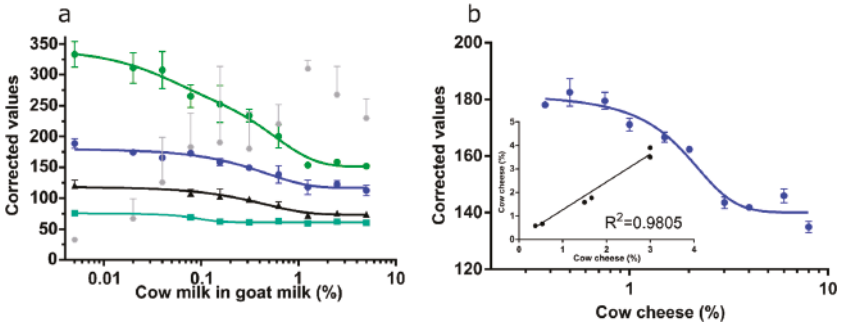


Figure 10. (**a**) G (green balls), B (blue balls), L (black triangles), and V (turquoise squares) channel values were fitted to a calibration curve for LFA quantification of cow milk spiked into pure goat milk. ΔRGB values (grey diamonds) did not allow to fit to a curve. (**b**) B channel values fitted to a calibration curve and, (inset), linear regression on predictions for cow milk in goat milk at three concentrations (n = 6).

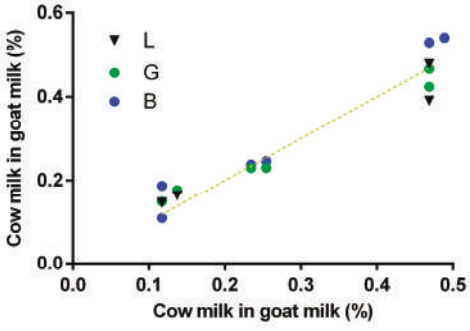


Figure 11. Predictions of cow milk (*x*-axis) in goat milk using L, G, and B channels. Each replica (n = 2) is indicated. Dashed line represents a curve at 45 degrees.

3.6. Channel Performance Comparison Using Commercial LFA for Gluten and pH Strips for Soil pH Prediction

Next, the ability to use various phones to quantify gluten with a commercial LFA, as well as pH strip quantification in buffered soil was tested (Figure 12). B values performed again optimally for the LFA quantification and were used to construct calibration curves for the quantification of gluten in a buffer using four different phones (Xiaomi, Huawei, iPhone, and Tablet) (Figure 12a). All phones enabled the construction of calibration curves (R^2 = 0.93; 0.88; 0.94; 0.86 for Xiaomi, Huawei, iPhone, and Tablet, respectively). Unfortunately, the curves did not overlap, except for the Tablet and Huawei, and cannot be directly used by end-users without calibration of individual phones. Magnification of LFA strips was attempted to correct inter-phone variation. LFA were photographed under the lens of a low-cost instrument (the smartscope; Ravensburger). Here, calibration curves with G channel values showed the highest R^2 values (0.87, 0.95, 0.93, and 0.92 for Xiaomi, Huawei, iPhone, and Tablet, respectively) (Figure 12b). Although the distance between the curves was narrower as in the previous set-up, the overlap of all curves was insufficient for the use of various phones to predict values on the same calibration curve. Thus, for color intensity variation using LFA, it seems that two- point calibration or camera calibration is necessary. Some suggested methods exist to obtain such calibration and improve sensitivity [30–33]. One interesting concept is to use a color reference chart in order to stabilize color variations caused by built-in automatic image correction operations [30]. Another option would be adjusting the white balance of the phones to a standardized value. This may be done by using a reference grey card or by locking the exposure and gain while selecting a preset white balance

in a control window that is illuminated with an external phone-independent constant light source [32]. Setting the white balance, gain, and exposure can be done in various manual camera applications (e.g., Open Camera, ProShot) and using such a system in combination with a greycard or external constant light source (flashlights of various phones can be different in light output) may decrease inter-phone variation. Moreover, higher sensitivity may be reached by adjusting the exposure time of the assay in such manual camera applications as was previously shown for other smartphone-based LFA quantification for the detection of bacterial fruit blotch [33]. However, the four-parameter dose-response curves did overlap for all four phones for pH determination of buffered soil samples using ΔRGB (Figure 12c) and R (Figure 12d) using the Xiaomi, Huawei, iPhone, and Tablet ($R^2 > 0.99$) showing that the proposed method worked without additional adjustments for this assay. Overlap was slightly less when using ΔRGB values, thus, the R channel was used from hereon for this application. Another two phone models (HTC and Samsung) were included to check the universality of the system. Four-parameter dose-response curves showed again excellent fits ($R^2 > 0.99$) and overlap with the other curves (Figure 12d). Finally, MAEs of predictions (n = 15) were calculated for each phone using its own calibration curve or a calibration curve constructed with all phone models except the phone model used to take the images for the predictions (Figure 12d; inset). Two-way ANOVA analyses of this data was significant for the type of curve ($p < 0.001$). MAEs increased significantly when predictions were performed on phone A with calibration curves prepared using phones B–F. However, for all phones except the tablet this increase in MAE was below 3% which explains why post-hoc simple effects within phone models and multiple comparisons was only significant for the tablet ($p < 0.001$). Thus, direct quantification of color change by end-users with the pH strips without the need of phone calibration seems possible if slightly increased MAEs are acceptable.

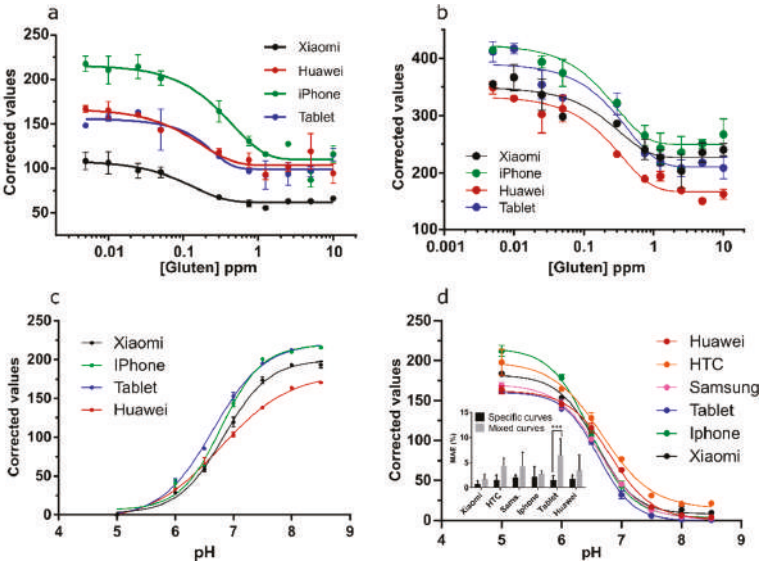


Figure 12. (**a**) B channel values fitted to four calibration curves for LFA quantification of gluten spiked into buffer with four phones (models indicated). (**b**) Calibration curves for LFA quantification using the smartscope and corrected G values. (**c**) Calibration curves for pH in buffered soil extract using corrected ΔRGB values. (**d**) R channel values fitted to calibration curves for pH estimation of buffered soil extracts with six phones (models indicated). Inset, the results of post-hoc analyses on the MAEs for pH predictions (n = 15) using the phone's specific calibration curve or a mixed curve of all phone models except the one used for prediction. Stars indicate *p*-values with *p*-value correction for multiplicity. *** = $p < 0.001$. All calibration curves were made using a four-parameter dose-response function.

4. Conclusions

The ability of all channels of the RGB, HSV, and LAB (named LAC here) color spaces to quantify color intensity and color change variations was tested in paper based and liquid assays. Channel performance was compared under varying background illumination using various phones. A channel of LAC proved not suitable for color quantification and C channel performance was suboptimal. L channel showed good performance in most systems tested (as did V) but both never outperformed the best RGB channel for a specific test. S performed suboptimal in all systems tested. H channel performed satisfactorily for color change but not for color intensity quantification. Moreover, H channel never outperformed the best RGB channel for a specific test. Overall, R functioned best for color change while B and G worked best for color intensity variation. ΔRGB values did show some robustness towards errors in an individual channel of RGB thus providing more universal applicability. However, ΔRGB stops to perform if an error in an individual channel of the RGB color space is too high. Thus, all RGB channels and ΔRGB should be initially plotted for assay development to determine the optimum channel. However, conversions to LAC or HSV color space are unnecessary. These results are specific for colorimetric assays and the implications might not hold for fluorescence or chemiluminescence-based assays where using luminance values seems a more logical choice. This being said, for such assays the light measured will equally pass the Bayer filter of the phone and is converted to RGB values. Mathematically converting these RGB values to LAB values may lead to similar results as reported here. However, a detailed characterization of the performance of RGB, LAB, and HSV color spaces/channels of such luminance-based assays with various phones and fluorophores is needed to test this assumption. Using a light-shielding box to prevent the error caused by background variation was less efficient than internal background correction if images were not taken in direct sunlight. Thus, use of a box is superfluous for colorimetric analysis. For color intensity variation it was shown that images taken with phone A could not be used on a calibration curve taken with phone B. Magnification did improve this situation but did not completely resolve the problem. Thus, camera calibration, white balance adjustment, and exposure time adjustment should be considered for LFA quantification. For color change quantification (pH determination of buffered soil extracts) calibration curves of the six phone models tested overlapped significantly. Thus, color change quantification by end-users without a light-shielding box or phone specific calibration seems feasible. Future research will focus on reducing prediction error and inter-phone variation for color change and intensity-based assays by combining channels from RGB, HSV, and CieLAB color spaces with machine learning algorithms.

Author Contributions: J.L.D.N. was the main author executing the experiments, data curation, and writing. L.B. assisted in the execution of experiments and data curation. Y.Z. and K.M.B. assisted in executing the experiments. C.T.E. and K.R. provided advice and assisted in reviewing and editing the final document. K.C. contributed to the conceptualization, supervised this work, and provided assistance in reviewing and editing the manuscript.

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 720325.

Acknowledgments: We further want to thank Master Sparrow Nelis for providing his smartscope for image magnification and Ilyana Nelis for providing home produced pure goat milk.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Malik, A.K.; Blasco, C.; Picó, Y. Liquid chromatography-mass spectrometry in food safety. *J. Chromatogr. A* **2010**, *1217*, 4018–4040. [CrossRef] [PubMed]
2. Matumba, L.; Van Poucke, C.; Njumbe Ediage, E.; De Saeger, S. Keeping mycotoxins away from the food: Does the existence of regulations have any impact in Africa? *Crit. Rev. Food Sci. Nutr.* **2017**, *57*, 1584–1592. [CrossRef] [PubMed]

3. The Rapid Alert System for Food and Feed (2017 Annual Report). Available online: https://ec.europa.eu/food/sites/food/files/safety/docs/rasff_annual_report_2017.pdf (accessed on 20 November 2019).
4. Sivadasan, S.; Efstathiou, J.; Calinescu, A.; Huatuco, L.H. Advances on measuring the operational complexity of supplier-customer systems. *Eur. J. Oper. Res.* **2006**, *171*, 208–226. [CrossRef]
5. Knowles, T.; Moody, R.; McEachern, M. European food scares and their impact on EU food policy. *Br. Food J.* **2007**, *109*, 43–67. [CrossRef]
6. Ludwig, S.K.; Tokarski, C.; Lang, S.N.; van Ginkel, L.A.; Zhu, H. Calling Biomarkers in Milk Using a Protein Microarray on Your Smartphone. *PLoS One* **2015**, *10*, e0134360. [CrossRef] [PubMed]
7. Ludwig, S.K.J.; Zhu, H.; Phillips, S.; Shiledar, A.; Feng, S.; Tseng, D.; van Ginkel, L.A.; Nielen, M.W.F.; Ozcan, A. Cellphone-based detection platform for rbST biomarker analysis in milk extracts using a microsphere fluorescence immunoassay. *Anal. Bioanal. Chem.* **2014**, *406*, 6857–6866. [CrossRef]
8. Zeinhom, M.M.A.; Wang, Y.; Song, Y.; Zhu, M.-J.; Lin, Y.; Du, D. A portable smart-phone device for rapid and sensitive detection of *E. coli* O157:H7 in Yoghurt and Egg. *Biosens. Bioelectron.* **2018**, *99*, 479–485. [CrossRef]
9. Fang, J.; Qiu, X.; Wan, Z.; Zou, Q.; Su, K.; Hu, N.; Wang, P. A sensing smartphone and its portable accessory for on-site rapid biochemical detection of marine toxins. *Anal. Methods* **2016**, *8*, 6895–6902. [CrossRef]
10. Lee, S.; Kim, G.; Moon, J. Performance improvement of the one-dot lateral flow immunoassay for aflatoxin b1 by using a smartphone-based reading system. *Sensors* **2013**, *13*, 5109–5116. [CrossRef]
11. Ross, G.M.S.; Bremer, M.G.E.G.; Wichers, J.H.; Van Amerongen, A.; Nielen, M.W.F. Rapid antibody selection using surface plasmon resonance for high-speed and sensitive hazelnut lateral flow prototypes. *Biosensors* **2018**, *8*, 130. [CrossRef]
12. Coskun, A.F.; Wong, J.; Khodadadi, D.; Nagi, R.; Tey, A.; Ozcan, A. A personalized food allergen testing platform on a cellphone. *Lab Chip* **2013**, *13*, 636–640. [CrossRef] [PubMed]
13. Tsagkaris, A.S.; Nelis, J.L.D.; Ross, G.M.S.; Jafari, S.; Guercetti, J.; Kopper, K.; Zhao, Y.; Rafferty, K.; Salvador, J.P.; Migliorelli, D.; et al. Critical assessment of recent trends related to screening and confirmatory analytical methods for selected food contaminants and allergens. *TrAC Trends Anal. Chem.* **2019**, *121*, 115688. [CrossRef]
14. Nelis, J.L.D.; Tsagkaris, A.S.; Zhao, Y.; Lou-Franco, J.; Nolan, P.; Zhou, H.; Cao, C.; Rafferty, K.; Hajslova, J.; Elliott, C.T.; et al. The end user sensor tree: An end-user friendly sensor database. *Biosens. Bioelectron.* **2019**, *130*, 245–253. [CrossRef] [PubMed]
15. Nelis, J.; Elliott, C.; Campbell, K. "The Smartphone's Guide to the Galaxy": In Situ Analysis in Space. *Biosensors* **2018**, *8*, 96. [CrossRef]
16. Koczula, K.M.; Gallotta, A. Lateral flow assays. *Essays Biochem.* **2016**, *60*, 111–120.
17. Roda, A.; Michelini, E.; Zangheri, M.; Di Fusco, M.; Calabria, D.; Simoni, P. Smartphone-based biosensors: A critical review and perspectives. *TrAC Trends Anal. Chem.* **2016**, *79*, 317–325. [CrossRef]
18. Rateni, G.; Dario, P.; Cavallo, F. Smartphone-Based Food Diagnostic Technologies: A Review. *Sensors* **2017**, *17*, 1453. [CrossRef]
19. Urusov, A.E.; Zherdev, A.V.; Dzantiev, B.B. Towards Lateral Flow Quantitative Assays: Detection Approaches. *Biosensors* **2019**, *9*, 89. [CrossRef]
20. Suaifan, G.A.R.Y.; Zourob, M. Portable paper-based colorimetric nanoprobe for the detection of Stachybotrys chartarum using peptide labeled magnetic nanoparticles. *Microchim. Acta* **2019**, *186*, 230. [CrossRef]
21. Zhang, X.; Zhi, X.; Zhang, C.; Wang, K.; Cui, D.; Li, D.; Wang, C. A CCD-based reader combined quantum dots-labeled lateral flow strips for ultrasensitive quantitative detection of anti-HBs antibody. *J. Biomed. Nanotechnol.* **2012**, *8*, 372–379. [CrossRef]
22. Murdock, R.C.; Shen, L.; Griffin, D.K.; Kelley-Loughnane, N.; Papautsky, I.; Hagen, J.A. Optimization of a paper-based ELISA for a human performance biomarker. *Anal. Chem.* **2013**, *85*, 11634–11642. [CrossRef] [PubMed]
23. Kong, T.; You, J.B.; Zhang, B.; Nguyen, B.; Tarlan, F.; Jarvi, K.; Sinton, D. Accessory-free quantitative smartphone imaging of colorimetric paper-based assays. *Lab Chip* **2019**, *19*, 1991–1999. [CrossRef] [PubMed]
24. Masawat, P.; Harfield, A.; Namwong, A. An iPhone-based digital image colorimeter for detecting tetracycline in milk. *Food Chem.* **2015**, *184*, 23–29. [CrossRef] [PubMed]
25. Turkevich, J. Colloidal gold. Part I. *Gold Bull.* **1985**, *18*, 125–131. [CrossRef]
26. Haiss, W.; Thanh, N.T.K.; Aveyard, J.; Fernig, D.G. Determination of Size and Concentration of Gold Nanoparticles from UV. *Anal. Chem.* **2007**, *79*, 4215–4221. [CrossRef]

27. Mcvey, C.; Logan, N.; Thanh, N.T.K.; Elliott, C.; Cao, C. Unusual switchable peroxidase-mimicking nanozyme for the deter- mination of proteolytic biomarker. *Nano Res.* **2019**, *12*, 1–8. [CrossRef]
28. Shen, L.; Hagen, J.A.; Papautsky, I. Point-of-care colorimetric detection with a smartphone. *Lab Chip* **2012**, *12*, 4240–4243. [CrossRef]
29. Cantrell, K.; Erenas, M.M.; de Orbe-Payá, I.; Capitán-Vallvey, L.F. Use of the Hue Parameter of the Hue, Saturation, Value Color Space as a Quantitative Analytical Parameter for Bitonal Optical Sensors. *Anal. Chem.* **2010**, *82*, 531–542. [CrossRef]
30. Kim, S.D.; Koo, Y.; Yun, Y. A smartphone-based automatic measurement method for colorimetric pH detection using a color adaptation algorithm. *Sensors* **2017**, *17*, 1604. [CrossRef]
31. Zhao, Y.; Choi, S.Y.; Nelis, J.L.D.; Zhou, H.; Cao, C.; Campbell, K.; Elliott, C.; Rafferty, K. Smartphone Modulated Colorimetric Reader with Color Subtraction. In Proceedings of the IEEE Sensors 2019 Conference, Montreal, QC, Canada, 27–30 October 2019.
32. Skandarajah, A.; Reber, C.D.; Switz, N.A.; Fletcher, D.A. Quantitative imaging with a mobile phone microscope. *PLoS ONE* **2014**, *9*, e96906. [CrossRef]
33. Saisin, L.; Amarit, R.; Somboonkaew, A.; Gajanandana, O.; Himananto, O.; Sutapun, B. Significant Sensitivity Improvement for Camera-Based Lateral Flow Immunoassay Readers. *Sensors* **2018**, *18*, 4026. [CrossRef] [PubMed]

sensors

MDPI

Article

An Innovative Ultrasonic Apparatus and Technology for Diagnosis of Freeze-Drying Process

Chin-Chi Cheng [1,*], Yen-Hsiang Tseng [2] and Shih-Chang Huang [1]

[1] Department of Energy and Refrigerating Air-Conditioning Engineering, National Taipei University of Technology, Taipei 10608, Taiwan; albertschuang@gmail.com
[2] Tai Yiaeh Enterprise Co., Ltd., New Taipei City 23942, Taiwan; yhtntut@gmail.com
* Correspondence: newmanch@ntut.edu.tw; Tel.: +886-2-27712171 (ext. 3527)

Received: 17 April 2019; Accepted: 8 May 2019; Published: 11 May 2019

Abstract: The freeze-drying process removes water from a product through freezing, sublimation and desorption procedures. However, the extreme conditions of the freeze-drying environment, such as the limited space, vacuum and freezing temperatures of as much as −50 °C, may block the ability to use certain diagnostic sensors. In this paper, an ultrasonic transducer (UT) is integrated onto the bottom of a specially designed frozen bottle for the purpose of observing the freeze-drying process of water at varying amounts. The temperatures and visual observations made with a camera are then compared with the corresponding ultrasonic signatures. Among all of the diagnostic tools and technologies available, only ultrasonic and visual records are able to analyze the entire progression of the freeze-drying process of water. Compared with typical experiment settings, the indication of drying point for water by the amplitude variations of ultrasonic L^3 echo could reduce the process period and energy consumption. This study demonstrates how an innovative frozen bottle, an integrated ultrasonic sensor and diagnostic methods used to measure and optimize the freeze-drying process of water can save energy.

Keywords: freeze-drying process diagnosis; ultrasonic transducer (UT); freezing/drying point; drying period

1. Introduction

Dehydration extends food's usage period longer than that of fresh food by preserving it in a stable and safe condition [1,2]. The conventional methods of drying include solar drying, air drying, spray drying [3], microwave drying [4], infrared drying, fluidized bed drying [5], spouted bed drying, vacuum drying and freeze-drying [6]. Drying methods can be separated into natural and artificial categories. Artificial methods are more advantageous than the natural methods [7] because they can remove large amounts of moisture efficiently by being able to control the different parameters involved such as the temperature, drying air flux and time of drying and so forth. [8]. For long-term storage of food, drug and biopharmaceutical products, most manufacturers utilize the freeze drying process, due to its advantages including better stability, easy handling and storage, as well as better overall product quality [9,10].

The freeze-drying process is comprised of the freezing, primary drying and then secondary drying stages [11] to remove the water from a product. The water contained in food is cooled down and becomes ice during the freezing stage. This stage governs the sublimation and desorption rates and the quality of the lyophilized product [12]. Then, in the primary drying stage, the air in the vacuum chamber is exhausted and the chamber pressure is reduced below the vapor pressure of ice. Meanwhile, the shelf temperature increases gradually to sublimate the ice. The residual water inside the food will be desorbed thoroughly in the secondary drying stage [13].

In the freezing stage, the accuracy of the freezing point data is related to the water activity, frozen water [14], freezing and the thawing of frozen food. The freezing point is also important in order to estimate the freezing time, the end point of freezing and the fraction of unfrozen water in food. Then, in the primary drying stage, the chamber pressure and food temperature [15] decides the sublimation rate of the ice, primary drying period, dried pore structure and product quality. A lower product temperature and the corresponding lower vapor pressure of the ice can result in longer primary drying times and higher manufacturing costs [16]. However, if the product temperature increases above the "critical formulation temperature," this may lead to losing the pore structure, to shrinkage or to fully collapsing. In the secondary drying stage, the product temperature controls the rate of desorption and the obtainable moisture level. Further reduction of the chamber pressure is not necessary in this stage [17]. Secondary drying times are usually designed to achieve a reduction of moisture content within the dried product to less than 1% [18]. Due to the importance of these parameters and its wider applications to the freeze drying process, the estimated freezing/drying points and the ability to model these properties is crucial in food processing (freezing and drying) and food stability during storage [10,19].

Being able to perform real-time process diagnosis will be beneficial to understanding the complex interplay between the different elements during the freeze-drying process to better enhance quality control procedures. However, the extreme conditions of the freeze-drying environment, such as the limited space, vacuum and freezing temperatures as low as −50 °C, may block the ability to use certain diagnostic sensors [6,20]. The typically used diagnostic tools for real-time measurement of the freeze-drying process are temperature and pressure sensors [20,21]. In order to provide temperature information, a temperature sensor is put in direct contact with a frozen sample [21]. However, inconveniences arise because it is often not easy to remove the sensor after the freeze-drying process. Additionally, using this diagnostic method on drug and biopharmaceutical products may cause undesired contamination due to a sensor being in direct contact with the sample. The chamber pressure is measured by the pressure sensor or vacuum gauge. The measured chamber pressure value cannot directly indicate the physical variation of frozen product real-time during the freeze-drying process [20,22]. In this paper, an idea for diagnosing the physical property variations of a product in the frozen bottle to provide additional information for a machine system is presented, as shown in Figure 1.

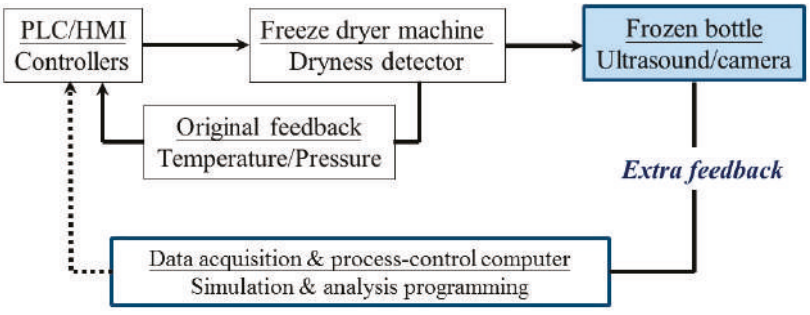


Figure 1. Controlling structure of the freeze-drying process by using the innovative frozen bottle method.

In order to circumvent the limitations of the mentioned sensors and provide real-time process diagnosis, the ultrasonic technique is one of the most widely known non-destructive and non-intrusive methods [23]. The fundamental properties of ultrasonic signals include reflection and transmission coefficients, velocity, attenuation and scatter signals given off from the materials. These signatures have specific relationships with properties of materials [24], process variations [25] and sample quality [26]. Ultrasonic echoes can also detect the temperature [27,28] and pressure [29] of a material. Passot et al. utilized ultrasonic technology to decrease the sublimation time during the primary drying stage by

controlling the nucleation temperature [30]. The authors presented an innovatively designed frozen bottle with an ultrasonic transducer and ultrasonic technology to be utilized for the diagnosis of the freezing point during the frozen process in our previous paper [22]. In this study, the ultrasonic technique will be applied for real-time diagnosis of the freeze-drying process. An ultrasonic transducer (UT) is integrated into the bottom of a frozen bottle. During the freeze-drying process at various amounts of water, the process is analyzed by using ultrasonic technology to evaluate the freezing/drying points and the drying period to optimize the process and save energy.

2. Design of Frozen Bottle with Ultrasonic Transducer

In order to solve the aforementioned limitations regarding the transmission of signals in a vacuum, as well as the limitations of being unable to view the process using the steel freezing bottle apparatus or the opaque samples, an innovative freezing bottle apparatus was designed as shown in Figure 2. The innovatively designed container used for freezing is comprised of two parts: a steel plate as the bottom holder and a transparent polymethylmethacrylate (PMMA) tube as the side wall of the container. Figure 2a shows the side view of the frozen bottle. A steel plate with a diameter of 25 mm and a height of 6mm was designed for better heat transmission with the heating/cooling shelf. A transparent PMMA tube measuring 31 mm high, with a diameter of 21 mm and a thickness of 2 mm was designed for better visibility. The steel plate and PMMA tube were then glued together. Figure 2b shows the top view of the apparatus. Figure 2c shows the bottom view of the apparatus. The UT is comprised of $Pb(Zr_xTi_{1-x})O_3$ (PZT) material and integrated into the cavity of steel plate using the sol-gel spray technique, as described in previous publications [22,24,31]. The operational temperature range of the UT is between −100 and 400 °C.

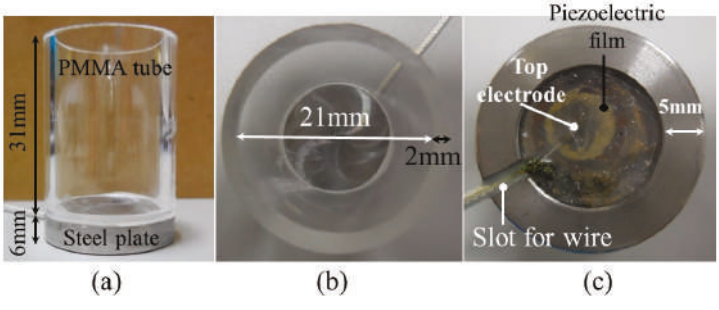


Figure 2. Photographs of the designed apparatus from (**a**) the side, (**b**) the top and (**c**) the bottom view.

To understand the transmission paths of the ultrasonic signals within the apparatus, a schematic view of the apparatus placed upright is displayed in Figure 3. A temperature sensor (Type T thermocouple) was installed in the middle of the apparatus for measuring the water/ice temperature. The temperature can be recorded and then compared with the ultrasonic signals received during the freeze-drying process. As shown in Figure 3a, an ultrasonic transducer (UT) was integrated onto the bottom of the steel plate. When electric pulses are applied to the piezoelectric film, the ultrasonic signals are transmitted into the steel plate. L^n (n = 1, 2, . . .) denotes the nth round trip of longitudinal ultrasonic echoes reflected from the interface of the steel plate/water or ice and L_w is the echo reflected from the water/air interface. The L^3 and L_w echoes are used to monitor the freeze-drying process and water/ice state. The height of the water level and the thickness of the bottom of the apparatus are denoted as H and D, respectively.

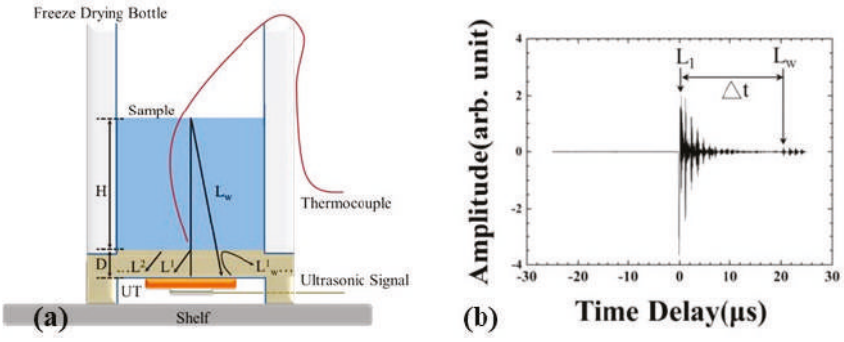


Figure 3. (**a**) Schematic drawing of cross-section of the designed apparatus with an ultrasonic transducer (UT) and thermocouple and (**b**) shows the typical ultrasonic echoes acquired by the UT.

When the water was fed into the apparatus, the typical ultrasonic signals acquired by the UT are presented in Figure 3b. It was observed that the L^1 echo, which reflected from the steel plate/water or ice interface, appeared as 0.81μs and remained so during the entire process. The operating frequency of the ultrasonic transducer was 9.13 MHz and the −6 dB bandwidth was 8.0 MHz. The signal to noise ratio (SNR) for the ultrasonic L^1 echo was 38.5 dB. The L_w echo, which propagated in the water and reflected from the water/air interface, was observed at 20.39 μs. Following the L^1 and L_w echoes, there were several echoes reflected from the steel plate/water or ice interface. The desired ultrasonic echoes were detected by the acquiring windows of the acquiring system. The time delay difference between the L^1 and L_w echoes is denoted as Δt.

3. Experiments

In order to confirm the hypothesis that the apparatus was able to perform improved diagnostics of the freeze-drying process in a vacuum, a 4 L air-cooled split-type shelf freeze-drying machine, equipped with vacuum chamber, shelf, control units, refrigeration system and vacuum pump, was used to carry out all experiments, as shown in Figure 4a. The vacuum chamber was used for freezing and drying the samples under temperature conditions ranging from +50 to −40 °C and pressures from 760 to 0.05 Torr. The apparatus was set on the shelf of vacuum chamber for the freeze-drying process. The control unit was composed of a programmable logic controller (PLC) and a human-machine interface (HMI) for controlling the freeze-drying process. The measured temperature data from the thermocouple were recorded by the PLC control unit every second during the freeze-drying process.

The ultrasonic signal is triggered and received by the pulse/receive unit (5072PR, Olympus, Japanese), as shown in Figure 4b. It is equipped with a broadband negative spike pulser and receiver, which can be operated in reflection or transmission mode. All the experiments in this study were carried out using the ultrasonic pulse/receiver unit's pulse-echo mode. The ultrasonic signals were acquired every second during the entire cycle. The ultrasonic signals received during the freeze-drying process were monitored and recorded with the digital storage USB oscilloscope (DSO-U2400, Perytech, Taipei, Taiwan), as shown in Figure 4c.

To compare the temperatures and ultrasonic signals, a charge-coupled device (CCD) camera, digital video recorder (DVR) and PC were used to capture dynamic images of the apparatus during the freeze-drying process, as shown in Figure 4d–e. The CCD camera (CV-M10BX, JAI, Japan) is a progressive scan camera with a standard interlaced video output at a resolution compatible with VGA or SVGA formats. The camera was equipped with a 35 mm lens for better resolution. The DVR (08KD, Kingnet, New Taipei City, Taiwan) is an 8-channel recorder that uses the H.264 image compression format.

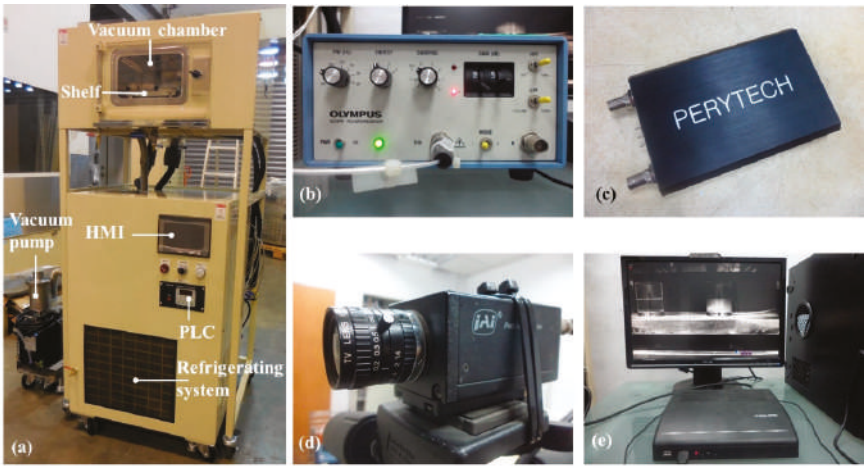


Figure 4. Photographs of (**a**) the split-type freeze-drying machine with vacuum chamber, controller unit, refrigeration unit and vacuum pump, (**b**) the ultrasonic pulse/receive unit, (**c**) a PC-based digital oscilloscope, (**d**) Charge-coupled device (CCD) camera, (**e**) digital video recorder (DVR) and PC.

In this study, different amounts of water at levels of 5, 10 and 15 mm were put into the apparatus for the duration of the freeze-drying process. The settings of a typical freeze-drying procedure are outlined in Table 1. The shelf temperature was set at −30 °C to freeze the water and the freezing period was 120 min. During the freezing period, the water in the apparatus/container was cooled down and it began to freeze. After enough cooling and freezing, the chamber was exhausted and the chamber pressure was reduced from 760 to 0.17 Torr within several seconds. Then the primary and secondary drying processes were started and the ice sublimated into gas continuously. The shelf temperature was incrementally increased by 10 degree intervals from −30 to 10 °C at heating and pausing periods of 60 and 120 min, respectively. The pausing period of 0 °C is 300 min to sublimate ice further. During the sublimation of the ice, the chamber pressure was reduced further to 0.036 Torr.

Table 1. The typical procedure used for the freeze-drying process.

Freezing Temp. (°C)	Freezing Period (min)	Chamber Pressure (Torr)	Drying Temp. (°C)	Heating Period (min)	Pausing Period (min)
−30	120	760 | 0.17	−20	60	120
			−10	60	120
			0	60	300
			10	60	120

4. Results and Discussions

For exploring the physical phenomena that occur during the freeze-drying process and searching for the indicators of freezing or drying completeness, the measured temperature and pressure in the vacuum chamber, ultrasonic echoes and velocity of the water/ice and visual observation during the freeze-drying process are analyzed in the following sections.

4.1. Temperature and Pressure Variations in the Vacuum Chamber during the Freeze-Drying Process

Temperature and pressure variations during the freeze-drying process have a close relationship with the state of the water/ice and both affect the size of the ice crystal nuclei and the freeze-dried product's quality. In the experiment, the measured room, shelf, cold trap, water/ice temperatures and pressures for the freezing and drying periods of the freeze-drying procedure of the water at a level of

15 mm are shown in Figure 5a,b, respectively. In Figure 5a, the room temperature was 32 °C during the entire process and the chamber pressure was kept at 760 Torr for the experiment's duration of 120 min. The shelf and cold trap temperatures were set at −30 and −40 °C and reached the desired values at 40 and 15 min, respectively. The water temperature was 32.1 °C in the beginning and was cooled down to a supercooled state of −0.4 °C, 16.1 min into the experiment. After the latent heat of the water was released completely, the phase change from water to ice was completed at 43.7 min and the ice temperature was −16.5 °C. Therefore, the cooling (ΔP_{CT}) and freezing (ΔP_{FT}) period determined by the sample temperature could be defined as the period from the start of cooling to the supercooled state and from the supercooled state to the phase change end, respectively.

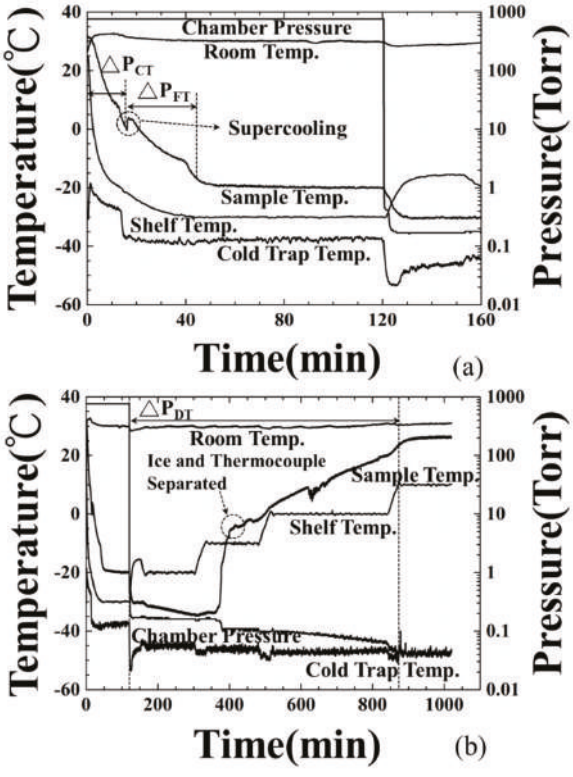


Figure 5. Temperature (room, shelf, sample, cold trap) and pressure (chamber) variations during freeze-drying process at (**a**) the freezing stage and (**b**) drying stage.

At the 120-min mark of the experiment, the drying procedure started and the chamber pressure was reduced from 760 to 0.17 Torr. The shelf temperature was increased in 10 °C increments, from −30 to 10 °C, according to the designed schedule in Table 1. The cold trap temperature was cooled down further to condense the vapor exhausted from the chamber. The chamber pressure was reduced further to 0.036 Torr during the drying process. At the 370-min mark, the ice temperature increased suddenly from −36 to −10 °C due to the sublimation of the ice and gradually increased to 23 °C by the end of the experiment. This abnormal situation was due to the detachment of the thermocouple from the ice and will be confirmed in the following section. Consequently, the ice temperatures during the second drying period is not available. The drying end could not be estimated from the temperature information and the typical experiment procedure would end at the 1020-min mark of the experiment.

4.2. Amplitude Variation of Ultrasonic Signals during the Freeze-Drying Process

The application of freeze-drying technology can guarantee the product's quality. However, the process diagnosis and control will benefit by guaranteeing product quality and energy conservation. The conventional tools for monitoring this process may be limited to visual observation and temperature measurement. According to the authors' knowledge, utilizing ultrasonic technology to monitor the freeze-drying process is rare. In order to investigate the correlation between the observed ultrasonic signals and the freeze-drying process of water, the amplitudes of the L^3 and L_w echoes in Figure 3b with respect to the process time were acquired and presented in Figure 6. The details of freeze-drying process measured by the amplitude variations of ultrasonic L^3 and L_w echoes were demonstrated as follows:

(1) At the 3.0-min mark: Water was poured into the container. At this moment, the amplitude of ultrasonic L^3 echo decreased and L_w echo appeared, due to the fact that a part of the ultrasonic energy transmitted into the water through the steel plate/water interface. Water started to cool down.

(2) At the 19.2-min mark: The amplitudes of the L^3 and L_w echoes decreased further, due to the variation of water acoustic impedance. In this supercooling period, this ultrasonic phenomenon may indicate the appearance of ice crystals. The alteration of water property will cause electrical impedance changing [32]. Hence, the amplitude decreasing point of the L^3 and L_w echoes is defined as the freezing point of water and the cooling period (ΔP_{CUT}) determined by the amplitude of the L^3 echo is defined as the period from water-in to the freezing point.

(3) At the 43.0-min mark: The amplitude of the L^3 echo increased to the relative maximum value and that of the L_w echo increased to a stable level, due to the stable ultrasonic impedance. This ultrasonic phenomenon may indicate phase change end of water and a flat ice surface. Hence, the freezing period (ΔP_{FUT}) determined by the ultrasonic L^3 echo is defined as the period from the freezing point to the phase change end.

(4) At the 125.0-min mark: The chamber was exhausted and the ice started to sublimate at the 120-min mark. The amplitude of L^3 echo decreased a little and that of L_w echo disappeared gradually, due to the sublimation of ice and the rough ice surface.

(5) At the 370.0-min mark: The amplitude of the L^3 echo increased suddenly from the bottom line and reach a stable value, due to the fact that the electromechanical coupling factor varied and the ultrasonic energy transmitting into the ice through the steel/ice interface reduced [33]. This ultrasonic phenomenon indicated the reduction of contact surface between ice and steel plate in the drying stage. At this moment, the sublimation of the ice reached a certain level and the thermocouple detached from the ice, as shown in Figure 5b.

(6) At the 885.6-min mark: The amplitude of the L^3 echo increased suddenly from the reducing tendency and reached a stable value at this moment, due to the fact that the electromechanical coupling factor varied and ultrasonic energy transmitting into the ice through the steel/ice interface reduced further. This ultrasonic phenomenon indicated the complete sublimation of the ice and only few minerals remained on the steel plate surface. This moment is defined as the drying end point. The drying period (ΔP_{DUT}) determined by the ultrasonic L^3 echo is defined as the period from the exhaust of chamber to the complete sublimation of the ice, that is, the drying end point. Compared with the typical experiment settings, the indication of drying ends for water by the amplitude variations of ultrasonic L^3 echo could reduce the processing period of the 134.4 min/cycle and save 13% of consumed electricity.

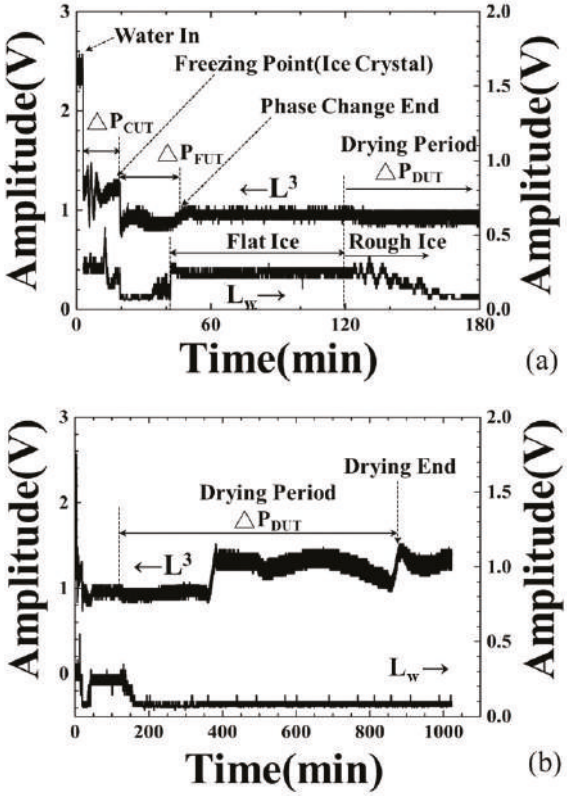


Figure 6. Amplitude variations of ultrasonic L^3 and L_w echoes during freeze-drying process at (**a**) freezing stage and (**b**) drying stage, indicating the water-in, freezing point (ice crystal), phase change end and drying end.

4.3. Variation of Ultrasonic Velocity during the Freeze-Drying Process

During the freezing process, the slow cooling speed caused bigger ice crystal nuclei than did the fast cooling time. Ultrasonic velocity has a close relationship with temperature and it is one of the indicators that shows the cooling speed of water, the phase change end and the start of sublimation. The ultrasonic velocity in the water is calculated according to the following equation:

$$v_w = 2H/\Delta t \quad (1)$$

where H is the height of the water level in Figure 3a and Δt is the time delay between the ultrasonic L^1 and L_w echoes in Figure 3b. The result is shown in Figure 7.

In Figure 7, when the water was poured into the container, the ultrasonic velocity was first measured at the 3.0-min mark under the water temperature of 27.2 °C. At that moment, the ultrasonic velocity in the water was 1588.5 m/s. From the 3.0 to the 17.4-min mark, when the water temperature decreased to 1.0 °C, the ultrasonic velocity decreased to 1434.5 m/s, indicating that the temperature of the water had decreased during the cooling process. Comparing Figure 7 with Figure 5a, the relationship between ultrasonic velocity and water temperature is expressed by the following equation.

$$v_w(T_w) = 1.42418 \times 10^3 + 1.03190 \times 10 \times T_w - 0.15453 \times T_w^2 \quad (2)$$

where v_w (m/s) is the ultrasonic velocity in water and T_w is the water temperature. The ultrasonic velocity decreased with the reducing water temperature. The ultrasonic velocity presented by authors in this study is higher than that presented by Bilanuik and Wang [34], due to some minerals contained in the water. Therefore, the cooling speed of water can be determined by measuring the ultrasonic velocity in the water. From the 17.4 to 38.3-min mark of the experiment, the ultrasonic velocity disappeared and the amplitude of the L_w echo also decreased in noise level due to the formation of ice crystal nuclei. After the 38.3-min mark, the ultrasonic L_w echo appeared again and the ultrasonic velocity gradually increased from 3586.4 to 3597.2 m/s, due to the phase change end of the water and the completion of the freezing process. From 43.0 to 120.0-min mark, the ultrasonic velocity maintained a steady level of 3597.2 m/s for the pausing period. Therefore, ultrasonic velocity can be used to indicate the cooling speed, the phase change end and completion of the freezing process of water.

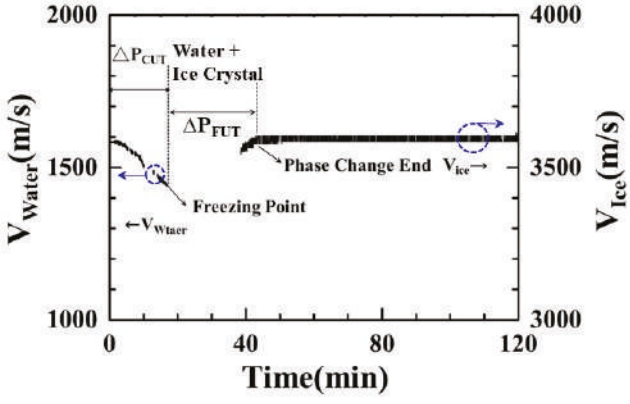


Figure 7. Ultrasonic velocity variation in the water during the freeze-drying process, which indicates the water-in, the freezing point (ice crystallization) and the phase change end.

4.4. Visual Observation of the Container during the Freeze-Drying Process

The dynamic physical phenomena that can be observed in the container shows the phase change from water to ice and the sublimation detail of the ice during the freeze-drying process in great detail. These phenomena could also be evidence of other measured parameters. Figure 8a–l are the photographs of the container with water/ice during the freeze-drying process at the 0–888-min mark, respectively. The freeze-drying process observed in Figure 8a–l is described as follows:

(1) Figure 8a: Water was poured into the container that had a thermocouple in the middle and then the container was set on the shelf of the freeze-drying machine.
(2) Figure 8b–d: Most of the water was still in a liquid state. However, the water close to the bottom of the container started to become opaque and some ice crystals began to form beside the thermocouple. The ice crystals increased gradually from the bottom to the top of the container. This was an indication that the water started to freeze.
(3) Figure 8e: Half of the water became very slushy and the visibility became worse. Some air bubbles also appeared within the slush.
(4) Figure 8f: Most of the water had become frozen into ice. Some air bubbles were pushed up from the bottom to the top. The ice surface was flat, which indicated that the stress on the ice during the freezing process was reduced.
(5) Figure 8g–h: The chamber was exhausted and a vacuum was created. The ice sublimated from the top progressing downwards towards the bottom of the container and from the exterior inwards towards the interior of the container. The ice surface was composed of porous structure and appeared rough.

(6) Figure 8i–j: The ice shrank more due to further sublimation. The contact surface between the ice and the steel plate was also reduced. The thermocouple seemed to detach from the ice at the 480-min mark of the process. This could explain why the sample temperature increased suddenly from −36 to −10 °C at 390.5-min mark. At the 540-min mark, the contact surface between the ice and the steel plate was reduced even further, which may diminish the conduction of heat being transferred from the shelf to the ice.

(7) Figure 8k–l: There was tiny amount of residual ice that remained on the surface of the steel plate at the 840-min mark. Finally, at the 888-min mark, only a little of ice crystals remained on the surface of the steel plate.

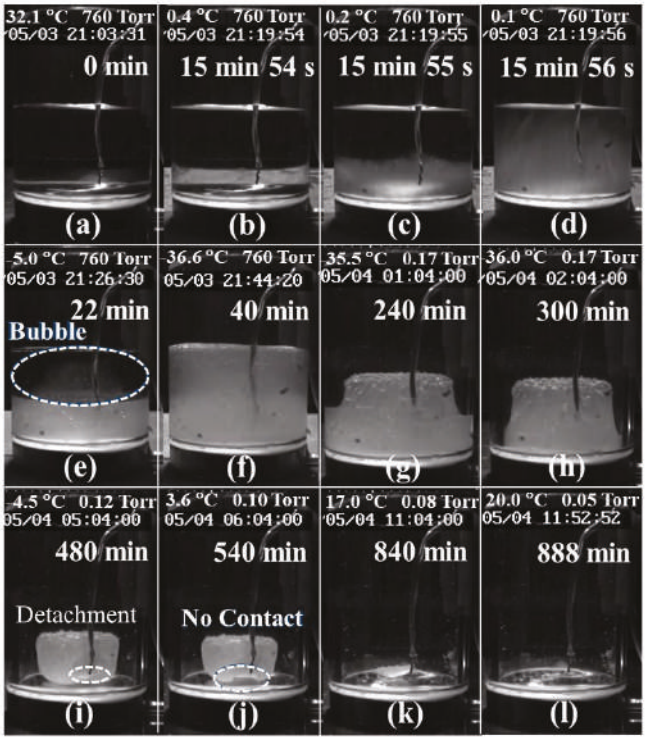


Figure 8. Photographs of the water/ice in the container during the freeze-drying process along with the corresponding process times of (**a**) 0 min, (**b**–**d**) 15 min 54~56 s, (**e**) 22 min, (**f**) 40 min, (**g**) 240 min, (**h**) 300 min, (**i**) 480 min, (**j**) 540 min, (**k**) 840 min and (**l**) 888 min.

The freeze-drying process can be observed clearly in Figure 8. These results can provide additional evidence for the previous measuring methods. However, this observing method is only suitable for the transparent container and plain solution. For the unobservable container and the opaque material, such as metal container and colored samples, the temperature and ultrasonic techniques are more capable to diagnose the mentioned phenomena. From the measured results, the thermocouple detached from the ice and further temperature information of the ice was not available. In this viewpoint, the ultrasonic technique is a more reliable and capable method for diagnosing the freeze-drying process.

4.5. Freeze-Drying Processes of Various Water Levels

Before applying the ultrasonic technology to the freeze-drying process, the linear relationship between ultrasonic signatures and the water level is the fundamental requirement. To verify this

relationship, the containers, which were filled with water in the height of 5, 10, 15 mm, passed through the freeze-drying process under the shelf temperature setting of −30 °C. In this research, the variation during the drying period will be evaluated. During the drying stage, the thermocouple cannot measure the samples' temperature profiles due to the detachment of thermocouple from the sample. The ultrasonic signals are able to diagnose the drying phenomena. The experimental results of the amplitude of the ultrasonic L^3 echo are shown in Figure 9. Figure 9a–c are the amplitude variations of ultrasonic L^3 echo in water levels of 5, 10, 15 mm, respectively, during the drying stage. The timings of the drying end indicated by ultrasonic L^3 echo are 329.8, 526.7 and 885.6-min mark for the water levels of 5, 10 and 15 mm, respectively. The drying periods of various water levels are also indicated. The corresponding timings of the drying end indicated by ultrasonic L^3 echo are illustrated in Table 2. It seems that there is a linear relationship between the timings of drying end and the water level.

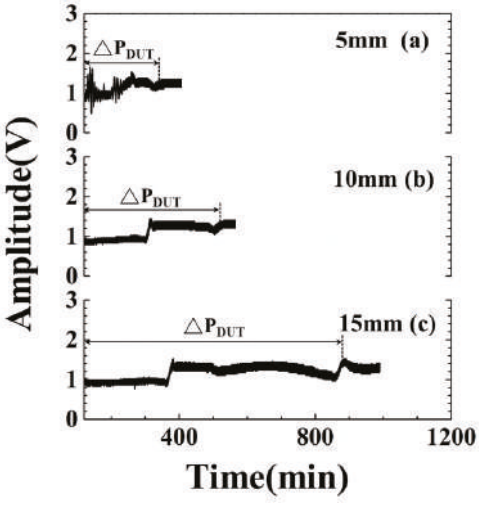


Figure 9. Amplitude variations of ultrasonic L^3 echo with respect to various water levels at drying stage. Water level: (**a**) 5, (**b**) 10 and (**c**) 15 mm.

Table 2. Timings and temperatures of water in, freezing point, phase change end and drying end indicated by temperature and ultrasonic L^3 echo for water levels of 5, 10, 15 mm.

Water Level (mm)	Items	Watering Point	Freezing Point	Phase Change End	Drying End
5	$Time_{Temp}$ (min)	3.0	16.3	24.4	N/A
	$Time_{UT}$ (min)	3.0	15.8	20.6	329.8
	Temperature (°C)	29.8	−2.8	−10.8	N/A
10	$Time_{Temp}$ (min)	3.0	14.0	28.9	N/A
	$Time_{UT}$ (min)	3.0	13.7	27.1	526.7
	Temperature (°C)	30.4	−2.4	−14.9	N/A
15	$Time_{Temp}$ (min)	3.0	19.2	46.7	N/A
	$Time_{UT}$ (min)	3.0	19.5	45.8	885.6
	Temperature (°C)	32.1	−0.4	−16.5	N/A

The measured drying periods are compared with the water level for evaluating the mentioned linear relationship. The results are shown in Figure 10. The estimated error of drying period under the

experimental conditions is less than 5%. In the water level range from 5 to 15 mm, the average drying periods indicated by ultrasonic L^3 echo increase from 185.1 to 742.8 min. The slope of the fitting line is 55.8 min/mm. The drying period is expressed as:

$$\Delta P_{DUT} = -112.8 + 55.8 * H \quad (3)$$

where ΔP_{DUT} is the drying period in Figure 9 and H is the water level in Figure 3. Even though there are no temperature and visual information, one can still estimate the required drying period according to the filled water level based on Equation (3). In the future, this ultrasonic sensor and evaluating technology would be installed into freeze drying machine to detect the phase change and feedback the information to machine control system, as shown in Figure 1. Comparing with the typical experiment settings of 1020 min, the ultrasonic technique can clearly indicate the dynamic phenomena and completion of cooling/freezing/drying stages at each water level before the end of the experiment for reducing the freeze-drying process and saving energy.

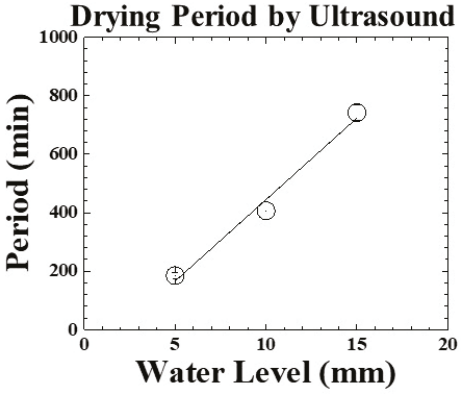


Figure 10. Drying periods measured by ultrasonic signal with respect to various water levels. Water level: 5, 10, 15 mm.

5. Conclusions

In this paper, an ultrasonic transducer (UT) is integrated onto the bottom of a specially designed container to analyze the freeze-drying process of water at various water amounts. The measured ultrasonic signatures are compared with the temperature and visual records. Among these three measured methods, ultrasonic and visual records are able to document the entire progression of the freeze-drying process, including the water-in, freezing/drying points, the phase change end of the water and the cooling/freezing/drying periods. The ultrasonic velocity in the water also indicates the cooling tendency of the water. The drying period increases with the water level. The increase rate, which is evaluated by the amplitude of the ultrasonic L^3 echo, is 55.8 min/mm. During the drying stage, the thermocouple cannot measure the entire temperature profile of the ice due to the detachment of the thermocouple from the ice. Only the ultrasonic signals and visual records are utilized for diagnosing the drying phenomena. However, the observing method is only suitable for the transparent container and plain solution. Comparing with the typical experiment settings, the indication of drying ends for water at level of 15 mm by the amplitude variations of ultrasonic L^3 echo could reduce the process period of 134.4 min/cycle and save 13% of consumed electricity. Therefore, this study demonstrates the use of a specially designed container, integrated ultrasonic sensor and technology for analyzing and optimizing the freeze-drying process of water for saving the process cost and energy.

Author Contributions: C.-C. Cheng initiated the idea, provided the draft and completed the revising of the paper. Y.-H. Tseng prepared the integrated sensor and practiced the experiments of freeze-drying process. S.-C. Huang analyzed the data and assisted to revise paper.

Funding: This work is funded by Ministry of Science and Technology of Taiwan, ROC under Contracts No: MOST 104-2221-E-027-121, MOST-105-2221-E-027-103 and MOST 107-3113-E-008-003.

Acknowledgments: The authors would like to provide special thanks to Hong-Ping Cheng (Department of Energy and Refrigerating Air-Conditioning Engineering, National Taipei University of Technology) for providing the useful discussions, Rei-Zhu Hsueh (Tai Yiaeh Enterprise Co., Ltd., New Taipei City, Taiwan) for providing useful maintaining assistance of freeze drying machine.

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

References

1. Zhao, J.H.; Liu, F.; Wen, X.; Xiao, H.W.; Ni, Y.Y. State diagram for freeze-dried mango: Freezing curve, glass transition line and maximal-freeze-concentration condition. *J. Food Eng.* **2015**, *157*, 49–56. [CrossRef]
2. Chizoba Ekezie, F.G.; Sun, D.W.; Han, Z.; Cheng, J.H. Microwave-assisted food processing technologies for enhancing product quality and process efficiency: A review of recent developments. *Trends Food Sci. Technol.* **2017**, *67*, 58–69. [CrossRef]
3. Ishwarya, S.P.; Anandharamakrishnan, C. Spray-Freeze-Drying approach for soluble coffee processing and its effect on quality characteristics. *J. Food Eng.* **2015**, *149*, 171–180. [CrossRef]
4. Monteiro, R.L.; Link, J.V.; Tribuzi, G.; Carciofi, B.A.M.; Laurindo, J.B. Microwave vacuum drying and multi-flash drying of pumpkin slices. *J. Food Eng.* **2018**, *232*, 1–10. [CrossRef]
5. Senadeera, W.; Alves-Filho, O.; Eikevik, T. Influence of drying conditions on the moisture diffusion and fluidization quality during multi-stage fluidized bed drying of bovine intestine for pet food. *Food Bioprod. Process.* **2013**, *91*, 549–557. [CrossRef]
6. Maisnam, D.; Rasane, P.; Dey, A.; Kaur, S.; Sarma, C. Recent advances in conventional drying of foods. *J. Food Technol. Preserv.* **2016**, *1*, 25–34.
7. Ahrens, D.C.; Villela, F.A.; Doni Filho, L. Physiological and industrial quality of white-oat (Avena sativa L.) seeds in intermittent drying. *Rev. Bras. Sementes* **2000**, *22*, 12–20. [CrossRef]
8. Toshniwal, U.; Karal, S. A review paper on Solar Dryer. *Int. J. Eng. Res. Appl.* **2013**, *3*, 896–902.
9. Kasper, J.C.; Friess, W. The freezing step in lyophilization: Physico-chemical fundamentals, freezing methods and consequences on process performance and quality attributes of biopharmaceuticals. *Eur. J. Pharm. Biopharm.* **2011**, *78*, 248–263. [CrossRef]
10. Siew, A. Freeze-drying process optimization. *Pharm. Technol.* **2018**, *42*, 18–23.
11. Ratti, C. Hot air and freeze-drying of high-value foods: A review. *J. Food Eng.* **2001**, *49*, 311–319. [CrossRef]
12. Rahman, M.S.; Guizani, N.; Al-Khaseibi, M.; Al-Hinai, S.A.; Al-Maskri, S.S.; Al-Hamhami, K. Analysis of cooling curve to determine the end point of freezing. *Food Hydrocoll.* **2002**, *16*, 653–659. [CrossRef]
13. Mellor, J.D.; Bell, G.A. Freeze-Drying—The Basic Process. In *Encyclopedia of Food Sciences and Nutrition*; Elsevier: Amsterdam, The Netherlands, 2003; pp. 2697–2701. ISBN 9780122270550.
14. Wang, D.Q.; Hey, J.M.; Nail, S.L. Effect of Collapse on the Stability of Freeze-Dried Recombinant Factor VIII and α-Amylase. *J. Pharm. Sci.* **2004**, *93*, 1253–1263. [CrossRef]
15. Tang, X.; Nail, S.L.; Pikal, M.J. Freeze-drying process design by manometric temperature measurement: Design of a smart freeze-dryer. *Pharm. Res.* **2005**, *22*, 685–700. [CrossRef]
16. Chang, B.S.; Patro, S.Y. Freeze-drying Process Development for Protein Pharmaceuticals. In *Lyophilization of Biopharmaceuticals*; American Association of Pharmaceutical Scientists: Arlington, VA, USA, 2004; pp. 113–138. ISBN 978-0-9711767-6-8.
17. Pikal, M.J.; Shah, S.; Roy, M.L.; Putman, R. The secondary drying stage of freeze drying: Drying kinetics as a function of temperature and chamber pressure. *Int. J. Pharm.* **1990**, *60*, 203–207. [CrossRef]
18. Breen, E.D.; Curley, J.G.; Overcashier, D.E.; Hsu, C.C.; Shire, S.J. Effect of moisture on the stability of a lyophilized humanized monoclonal antibody formulation. *Pharm. Res.* **2001**, *18*, 1345–1353. [CrossRef]
19. Rahman, M.S. *Handbook of Food Preservation*, 2nd ed.; Rahman, M.S., Ed.; Taylor & Francis Group: Milton Park, UK, 2007; ISBN 9781845697587.

20. Nail, S.; Tchessalov, S.; Shalaev, E.; Ganguly, A.; Renzi, E.; Dimarco, F.; Wegiel, L.; Ferris, S.; Kessler, W.; Pikal, M.; et al. Recommended Best Practices for Process Monitoring Instrumentation in Pharmaceutical Freeze Drying—2017. *AAPS PharmSciTech* **2017**, *18*, 2379–2393. [CrossRef]
21. Malik, N.; Gouseti, O.; Bakalis, S. Effect of freezing with temperature fluctuations on microstructure and dissolution behavior of freeze-dried high solid systems. *Energy Procedia* **2017**, *123*, 2–9. [CrossRef]
22. Tseng, Y.H.; Cheng, C.C.; Cheng, H.P.; Lee, D. Novel Real-Time Diagnosis of the Freezing Process Using an Ultrasonic Transducer. *Sensors* **2015**, *15*, 10332–10349. [CrossRef]
23. Hellier, C.J. *Handbook of Nondestructive Evaluation*; McGraw-Hill Companies, Inc.: New York, NY, USA, 2003; ISBN 007139947X.
24. Cheng, C.C.; Yang, S.Y.; Lee, D. Novel real-time temperature diagnosis of conventional hot-embossing process using an ultrasonic transducer. *Sens. Switz.* **2014**, *14*, 19493–19506. [CrossRef] [PubMed]
25. Wu, Y.L.; Cheng, C.C.; Kobayashi, M.; Yang, C.H. Novel design of extension nozzle and its application on real-time injection molding process diagnosed by ultrasound. *Sens. Actuators A Phys.* **2017**, *263*, 430–438. [CrossRef]
26. Cheng, C.-C.; Wu, C.-L.; Wu, K.-T.; Yang, S.-Y. Real-time diagnosis of gas-assisted hot embossing process by ultrasound. *Polym. Eng. Sci.* **2013**, *53*, 2175–2182. [CrossRef]
27. Kažys, R.; Voleišis, A.; Voleišienė, B. High temperature ultrasonic transducers: Review. *Ultrasound* **2008**, *63*, 7–17.
28. Takahashi, M.; Ihara, I. Ultrasonic monitoring of internal temperature distribution in a heated material. *Jpn. J. Appl. Phys.* **2008**, *47*, 3894–3898. [CrossRef]
29. Cheng, C.C. Micromolding of polymer nanocomposites diagnosed by ultrasound. *J. Polym. Eng.* **2010**, *30*, 95–108. [CrossRef]
30. Passot, S.; Tréléa, I.C.; Marin, M.; Galan, M.; Morris, G.J.; Fonseca, F. Effect of controlled ice nucleation on primary drying stage and protein recovery in vials cooled in a modified freeze-dryer. *J. Biomech. Eng.* **2009**, *131*, 074511. [CrossRef] [PubMed]
31. Cheng, C.-C.; Young, S.-L.; Chen, H.-Z.; Yang, S.-Y. Substrate Effect on Characteristics of PbZrxTi1−xO3 (PZT) Film. *Integr. Ferroelectr.* **2014**, *150*, 51–58. [CrossRef]
32. Shuyu, L. Load characteristics of high power sandwich piezoelectric ultrasonic transducers. *Ultrasonics* **2005**, *43*, 365–373. [CrossRef]
33. Arnold, F.J.; Roger, L.L.B.; Gonçalves, M.S.; Mühlen, S.S. Electrical Impedance of Piezoelectric Ceramics under Acoustic Loads. *Ecti Trans. Electr. Eng. Electron. Commun.* **2014**, *12*, 48–54.
34. Bilaniuk, N.; Wong, G.S.K. Speed of sound in pure water as a function of temperature. *J. Acoust. Soc. Am.* **1993**, *93*, 1609–1612. [CrossRef]

sensors

MDPI

Article

Temperature and Strain Correlation of Bridge Parallel Structure Based on Vibrating Wire Strain Sensor

Lu Peng [1,*], Genqiang Jing [1], Zhu Luo [1], Xin Yuan [2], Yixu Wang [1] and Bing Zhang [1]

[1] National Center of Metrization for Equipments of Roads and Bridges, Research Institute of Highway Ministry of Transport, Beijing 100088, China; gq.jing@rioh.cn (G.J.); ZhuLuo@ncmerb.com (Z.L.); wyx@ncmerb.com (Y.W.); b.zhang@rioh.cn (B.Z.)

[2] School of Electrical and Information Engineering, Wuhan Institute of Technology, Wuhan 430205, China; qcmhssk@163.com

* Correspondence: lu.peng@rioh.cn

Received: 29 October 2019; Accepted: 21 January 2020; Published: 24 January 2020

Abstract: Deformation is a ubiquitous phenomenon in nature. This process usually refers to the change in shape, size, and position of an object in the time and spatial domain under various loads. Under normal circumstances, during engineering construction, technicians are generally required to monitor the safe operation of structural facilities in the transportation field and the health of bridge, because monitoring in the engineering process plays an important role in construction safety. Considering the reliability risk of sensors after a long-time work period, such as signal drift, accurate measurement of strain gauges is inseparable from the value traceability system of high-precision strain gauges. In this study, two vibrating wire strain gauges with the same working principle were measured using the parallel method at similar positions. First, based on the principle of time series, the experiment used high-frequency dynamic acquisition to measure the thermometer strain of two vibrating wire strain gauges. Second, this experiment analyzed the correlation between strain and temperature measured separately. Under the condition of different prestress, this experiment studied the influencing relationship of temperature corresponding variable. In this experiment, the measurement repetitiveness was analyzed using the meteorology knowledge of single sensor data, focused on researching the influence of temperature and prestress effect on sensors by analyzing differences of their measurement results in a specified situation. Then, the reliability and stability of dynamic vibrating wire strain gauge were verified in the experiment. The final conclusion of the experiment is the actual engineering in the later stage. Onsite online meteorology in the application provides support.

Keywords: parallel position; bridge structure; temperature; vibrating wire strain sensor

1. Introduction

Bridge health monitoring and diagnostic discriminant models have always been a key challenge for the transportation sector worldwide. In a previous study, on a regular and irregular basis, construction workers used different monitoring instruments to test some components of structure, analyzed the data, and finally evaluated the performance of structure. In 2018, Mao by monitoring dynamic characteristics of Sutong cable-stayed bridge (SCB), including acceleration and strain responses, as well as modal frequencies, are investigated through one-year continuous monitoring data under operating conditions by the structural health monitoring system. One-year continuous modal frequencies of SCB are identified using the Hilbert–Huang transform method. Variability analysis of the structural modal frequencies due to environmental temperature and operational traffic is then conducted. Results show that temperature is the most important environmental factor for vertical and torsional modal frequencies. The traffic load is the second critical factor especially for the fundamental vertical frequency

of SCB [1]. In 2019, Wang reported that the real-time monitoring data collected from a long-span cable-stayed bridge is utilized to demonstrate the feasibility of the improved BDLM-based method. In particular, the present BDLM-based method allows for probabilistic forecasts, offering substantial information about the target TIS response, such as mean and confidence interval. Results show that the improved BDLM is capable of capturing the relationship between temperature and TIS. Compared to the AR model, multiple linear regression (MLR) model, and BDLM without the AR component, the improved BDLM shows better forecasting performance in modeling and forecasting the TIS of a long-span bridge [2]. Such measurement methods have the drawback of large errors and discontinuities. The ideal health monitoring system should accurately reflect the change in grassroot structure under the influence of factors such as temperature, humidity, and other environmental factors, installation and deployment methods, and the sensor's own error, and establish an effective health assessment and prediction model. For example, Mao proposed according to one-year continuous monitoring of strain data recorded by the structural health monitoring system of the Sutong cable-stayed bridge, the lifetime fatigue reliability of three welded details of the orthotropic steel deck was investigated, detailed analysis of the separated components of the raw strain data was first conducted, and included the slow-varying trend and the dynamic component. The strain dynamic component was mainly induced by the local vehicle axle loads. Rainflow counting was used to obtain the stress range histograms, which were then used to calculate the equivalent stress range according to the lognormal-fitting method. Finally, a time-dependent fatigue reliability evaluation of the described three welded details was conducted using one-year monitoring strain data. Results showed that the fatigue performance of the two welded details, RTDD and DTD, remained satisfactory after 100 years of operation because the failure probabilities were both lower than 10^{-5}. The designed cutout of the diaphragm was applied to the RTD weld at the welded connection between the U rib and diaphragm. This cutout was validated as a means to help achieve better fatigue resistance for the RTD weld [3]. For bridge strain monitoring sensors, most experimental procedures use resistive strain gauges, vibrating wire strain gauges, fiber grating strain gauges, etc. Although the development of fiber grating technology in recent years has led to a new round of equipment upgrades for structural monitoring methods, vibrating wire strain gauges are most widely used in stage bridge health monitoring systems.

In recent years, studies on the heat of vibrating wire sensors at home and abroad have continued unabated, and domestic and foreign experts and scholars have conducted a lot of research to improve the sensor performance. The research on sensors is gradually extended towards high precision, large scale, small volume, and multiple applications. During the improvement in the performance of the sensor, higher requirements are also imposed on the measurement accuracy. To improve accuracy, in 2010, He proposed a low-voltage excitation with the feedback method. By pre-excitations, the vibration frequency of the sensor can be used as the output of the driver. The feedback signal frequency is very close to the real frequency of the sensor, so the wire can reach resonance state quickly. The optimal excitation strategy was verified by the new designed detection circuit. The data perform as short-time excitations with large resonance amplitude, therefore the anti-interference ability got enhanced at lower cost of signal processing circuit, finally increasing the measurement time and improving the accuracy of measured frequency [4]. In 2010, Wen proposed a method for the frequency measurement of vibrating wire sensors with LM3S6965 as the control core. An equal precision measurement was used to effectively improve the measurement accuracy of the system [5]. In 2016, Tian et al. proposed a design method of nonlinear compensation that is provided for the nonlinear relation between the measured force and the frequency of vibrating wire sensor. Data density is increased by interpolation based on the principle of cubic spline interpolation. The F-f curve of vibrating wire sensor is fitted and revised by MATLAB based on the least-square method, more accurately the smooth curve according to engineering practice is obtained and a reasonable conclusion is obtained. Experiments show that the method achieves the accurate fitting to the F-f curve of vibrating wire sensor with the limited data. The F-f function of vibrating wire sensor which is long buried can be rapidly determined by this procedure [6]. In 2017, Chen et al. proposed an

adaptive vibration measurement method based on fast Fourier transform, which uses digital Fourier filtering to automatically filter noise interference and then uses Quinn algorithm for high-precision frequency calculation. Based on the STM32 processor platform, the method was used in vibrating a wire sensor frequency measurement system. The test results show that the relative error of system frequency measurement is less than 0.01% in the case of no noise interference; in the case of severe white noise interference (signal to noise ratio is −20 dB), the system's frequency measurement relative error is less than 0.3%. Compared to other frequency measurement methods, it was found that the method has better noise suppression and frequency measurement accuracy [7]. In terms of static metrology and calibration, the experiment proved that a vibrating wire strain gauge can be widely used in the field of geotechnical engineering, and the accuracy requirement for variable measurement is relatively low. However, in the application of traffic engineering, the measurement accuracy is correspondingly improved, accompanied by the needs of metrology and calibration. Many scholars used a strain sensor calibration device consisting of a calibration frame and digital display to calibrate certain metering characteristics of strain gauge. For example, in 2009, Xu invented a vibrating wire strain sensor calibration device that uses a dial gauge as a gauge for deformation, with a deformation range of up to 3 mm and a strain measurement resolution of about 0.1%. The results show that the strain in the middle of the sensitive grid of strain gauge is the largest and gradually decreases to zero at both ends when measuring. The force has no effect on the sensitive gate base layer and bonding layer at both ends and there is a strain transition zone between the base layer and the sensitive gate, the bonding layer, and the base layer. The longer the horizontal width of bonding layer is the thinner the thickness and the larger the shear modulus will be, and the strain transfer will be more efficient if the active zone of the sensitive grid is longer [8]. In 2011, Zhang et al. designed a vibrating wire strain sensor calibration device. They used a grating scale as the measuring standard for deformation. The maximum calibration distance was 300 mm, and the measurement resolution was 2 μm. The degree is about 0.2% [9]. In 2016, Mai et al. invented a vibrating wire strain gauge calibration device. A dial gauge was used as the measurement standard for deformation amount, so that the resolution was 0.01 mm [10]. During engineering application, many in-depth studies have been conducted on the dynamic calibration of strain gauges. In 2016, Bai et al. affixed FBG sensors on standard beams. Loads were applied to the standard beams to deform them, and the beams were measured. The surface strain was used to calibrate the sensitivity coefficient of FBG strain gauge, and the measured strain range could be analyzed up to 2000 με [11]. In 2017, Zhang proposed a method for the indirect calibration of fiber grating strain sensors using a fiber grating temperature sensor. The lossless calibration of strain sensors was achieved [12]. During the detection, monitoring, and metering of vibrating wire strain gauges, the temperature deformation caused by temperature changes has been the research direction of many scientists. Chen et al. monitored data in engineering applications based on the working principle of vibrating wire strain gauges. The relationship between temperature and strain was analyzed using the relevant data of bridge strain monitoring, and the relationship between temperature influence and strain of the string itself and the temperature field of structural section was evaluated [13]. Bai et al. analyzed that when the temperature changes significantly, a variety of mathematical models were used to fit and calibrate the monitoring data, eliminating the temperature drift of strain gauge, and the experimental results reflect the original characteristics of deformation [14]. Agostiono investigation of the sensing features of the long-period fiber gratings (LPGs) fabricated in hollow core photonic crystal fibers (HC-PCFs) by the pressure assisted electric arc discharge (EAD) technique. In particular, the characterization of the LPG in terms of shift in resonant wavelengths and changes in attenuation band depth to the environmental parameters: Strain, temperature, curvature, refractive index, and pressure is presented. Results show that LPGs in HC-PCFs represent a novel high performance sensing platform for measurements of different physical parameters including strain, temperature and, especially, for measurements of environmental pressure. The pressure sensitivity enhancement is about four times greater if comparing LPGs in HC and standard fibers. Moreover, differently from LPGs in standard fibers, these LPGs realized in innovative fibers, i.e., the HC-PCFs, are not sensitive to the

surrounding refractive index. During the online calibration study of strain sensors [15], offline removal and reinstallation of sensing elements pose a risk to the continuity and consistency of monitoring data. At present, the load test method is mainly used, but the load test method cannot eliminate the instability of structure such as the stiffness and strength of the structure itself [16,17]. Therefore, a more effective method at this stage is to install a traceable high-frequency dynamic monitoring sensor in parallel to the sensor to perform parallel measurement, thus achieving the online calibration of a long-term monitoring sensor. At present stage, the measurement calibration of the strain sensor system in various industries only adopts static calibration before installation or carries out a static and dynamic load test for calibration, and does not carry out online calibration during use. The static calibration cannot determine the influence of the error and noise, temperature change, and prestress effect on the strain value. Static and dynamic load tests cost a lot of manpower and materials, and cannot provide reference for sensor calibration under complicated passive excitation. In the online calibration research of strain sensors, because the offline disassembly and reinstallation of the sensing elements pose a risk to the continuity and consistency of the monitoring data, the load test method is mainly used at this stage, but the load test cannot eliminate the rigidity and strength of the structure itself.

In the research process, we carried out a number of relevant tests. For example, the calibration of fixed excitation by a simply supported beam model is shown in Figure 1. The strain correlation test of a small simply supported beam at constant temperature in a high and low temperature box is shown in Figure 2. Fatigue test verification is shown in Figure 3. In order to further verify the test, a test was carried out on Jiujiang Bridge to verify Figure 4. Most of the above studies adopted common source excitation schemes, mainly to verify the calibration of the strain monitoring system under common source excitation by parallel measurement methods, which provided important verification support for the coupling relationship between temperature, prestress, and strain output in the online calibration of dynamic strain measurement in this study.

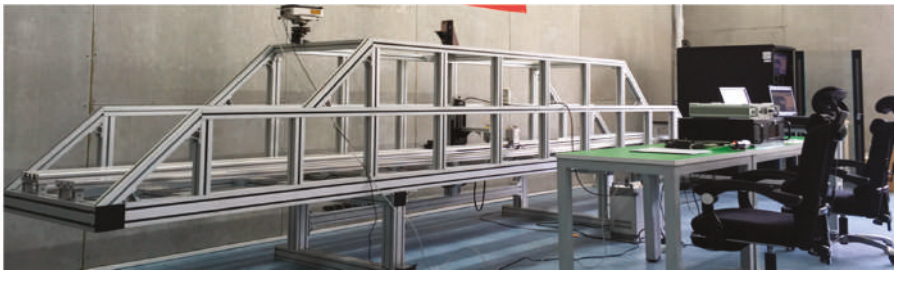
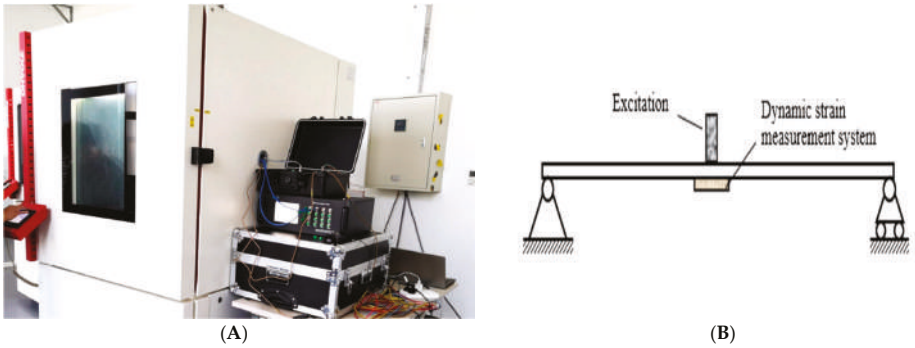


Figure 1. Schematic diagram of simply supported beam test model.



(**A**) (**B**)

Figure 2. High and low temperature dynamic strain test. (**A**)Test equipment, (**B**) Assembly drawing of experiment beam.

Figure 3. Verified by fatigue tests on bridge model.

(**A**) (**B**)

Figure 4. Test verification of Jiujiang Bridge. (**A**) Jiujiang Bridge, (**B**) Test process.

During this study, two vibrating wire strain gauges with the same working principle were measured using the parallel method at similar positions. The feasibility of a parallel measurement scheme was verified [18]. Under the condition of different prestresses, this experiment studied the influencing relationship of temperature corresponding variable [19,20]. We focused on researching the influence of temperature and prestress effect on sensors by analyzing differences of their measurement results in a specified situation. The measurement repetitiveness was analyzed using the meteorology knowledge of single sensor data, and then the reliability and stability of a dynamic vibrating wire strain gauge were verified in the experiment. The relevant measurement performance of a dynamic vibrating wire strain gauge was demonstrated.

2. Working Principle of Vibrating Wire Strain Gauge

A vibrating wire sensor is tested by steel string vibration. An experiment is conducted to characterize the force according to the variation in vibration frequency. In an actual output frequency signal, there is no strain gauge; it must be field calibration, signal drift, long distance transmission, and long time. The problem of poor durability was used, and the robustness was good [21]. This experiment solves the shortcoming of unstable strain gauges for a long time, and this product can be widely used in bridge monitoring.

A vibrating wire sensor has good measurement characteristics; it can achieve nonlinear characteristics of less than 0.1%, sensitivity of 0.05%, and less than 0.1%/10 °C temperature error.

After the strain gauge is manufactured, its steel string has a certain initial tensile force T_0 and thus has an initial frequency F_0. When the strain gauge is installed, the tensile force of vibrating wire changes with deformation, and the strain can be measured by the tensile force change of vibrating wire. Set the tension of the vibrating wire to T and the natural frequency to f. The relationship between tension and frequency can be expressed as Equation (1):

$$T = \mathrm{K}f^2 \tag{1}$$

where K is related to the length of string, and the mass per unit length can be expressed using Equation (2):

$$\Delta T = T - T_0 = K(f^2 - f_0^2) \tag{2}$$

Assuming that the strain increment of strain gauge can be set to the strain increment of vibrating wire, Equation (3) can be derived as follows:

$$\varepsilon_h = \varepsilon_g = \frac{\Delta K}{EA} \tag{3}$$

When EA is the axial stiffness of steel string, it can be derived from Equation (4) as follows:

$$\varepsilon_h = \frac{K}{EA}(f^2 - f_0^2) = k_h(f^2 - f_0^2) \tag{4}$$

A mathematical model of the vibrating wire sensor can be expressed using Equations (5) and (6) as follows:

$$\mathrm{F} = K(f^2 - f_0^2) \tag{5}$$

$$\mathrm{F} = A(f^2 - f_0^2) + B(f - f_0) \tag{6}$$

When the length is such that the fine string of mass m is subjected to tension F (Figure 5), the natural frequency f can be expressed as Equations (7)–(9) as follows:

$$f = \frac{1}{2}\sqrt{\frac{F}{ml}} \tag{7}$$

$$f = \frac{1}{2l}\sqrt{\frac{ES\Delta l}{\rho l}} = \frac{1}{2l}\sqrt{\frac{E\Delta l}{\rho_v l}} \tag{8}$$

$$f = \varphi(F) \tag{9}$$

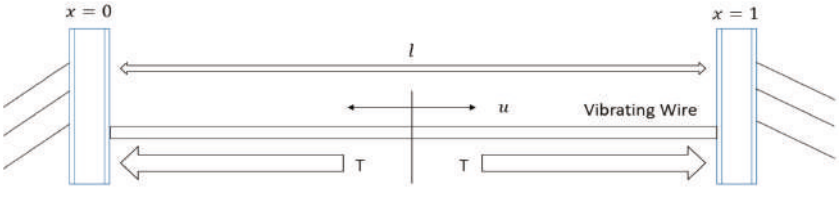


Figure 5. Working principle diagram of vibrating wire strain gauge.

Sensitivity can be derived using Equation (10) as follows:

$$f^2 = \frac{1}{4l^2}\frac{E\Delta l}{\rho_v l} = K\varepsilon \tag{10}$$

After differentiation, Equations (11) and (12) can be deduced as follows:

$$2fdf = Kd\varepsilon \tag{11}$$

$$\mathrm{k} = \frac{df}{d\varepsilon} = \frac{K}{2f} \tag{12}$$

The material coefficient can be calculated using Equation (13):

$$\begin{cases} K = \frac{1}{4l^2}\frac{E}{\rho_v} = \frac{ES}{4l^2\rho} \\ \varepsilon = \frac{\Delta l}{l} \end{cases} \tag{13}$$

The above formula shows that sensitivity k is directly proportional to material coefficient K and inversely proportional to the vibration frequency of string.

After many experiments, the initial frequency is f_0 when the measured tension is F_0, and the vibration frequency is f_1 when the measured tension is $F_1 = f_0 + f$. The nonlinear error of vibrating wire strain gauge can be expressed using Equations (14) and (15) as follows:

$$f_1 = \frac{1}{2}\sqrt{\frac{F_0 + \Delta F}{ml}} = \frac{1}{2}\sqrt{\frac{F_0}{ml}}\sqrt{1 + \frac{\Delta F}{F_0}} = f_0\sqrt{1 + \frac{\Delta F}{F_0}} = f_0\sqrt{1 + \varepsilon_F} = f_0\,(1 + \varepsilon_F)^{\frac{1}{2}} \tag{14}$$

$$f_1 = f_0(1 + \frac{1}{2}\varepsilon_F - \frac{1}{8}\varepsilon_F^2 + \frac{1}{16}\varepsilon_F^3 - \cdots) \tag{15}$$

At that time $F_2 = F_0 - \Delta F$, leading to Equation (16):

$$f_2 = f_0(1 + \frac{1}{2}\varepsilon_F - \frac{1}{8}\varepsilon_F^2 + \frac{1}{16}\varepsilon_F^3 - \cdots) \tag{16}$$

Its quadratic nonlinearity error can be expressed as Equation (17):

$$\frac{\left|\frac{1}{8}f_0\varepsilon_F^2\right|}{\frac{1}{2}f_0\varepsilon_F} = \frac{1}{4}\varepsilon_F \tag{17}$$

The above formula shows that the larger ε_F, the larger δ_m. At the same time, the ambient temperature mainly influences the frequency stability. The bulk density ρv and Δl caused by F do not change with the ambient temperature, so the frequency stability can be expressed using Equation (18):

$$\gamma_f = \frac{df}{f} = \frac{dE}{2}E - \frac{3}{2}\frac{dl}{l} \tag{18}$$

When the ambient temperature changes, the vibrating wire strain gauge material and the measured structural material have different linear expansion coefficients, and the sensor is subjected to additional stretching or compression. The additional strain can be expressed as Equation (19):

$$\varepsilon\mathrm{T} = (\alpha - \beta) \bullet \Delta\mathrm{T} \tag{19}$$

where:

εT: Additional strain caused by temperature effect;
α: Linear expansion coefficient ($^{\circ}\mathrm{C}^{-1}$) of the structural material to be tested;
B: The coefficient of linear expansion ($^{\circ}\mathrm{C}^{-1}$) of a steel string of a vibrating wire strain gauge;
ΔT: Temperature change amount.

In actual application, the vibrating wire strain sensing system generally adopts software compensation. After the thermistor is set to collect the working environment temperature in the electromagnetic coil, the optimized temperature compensation algorithm is combined with the software to compensate in the demodulation instrument. The temperature strain compensation of the vibrating wire strain gauge mostly utilizes the two-dimensional regression method, polynomial fitting method,

and neural network method. After the test results of the vibrating wire strain gauge are linearly fitted, the strain calculation method can be expressed as Equation (20):

$$\varepsilon = \mathrm{a} \times (f_i^4 - f_0^4) + \mathrm{b}(f_i^2 - f_0^2) + \mathrm{k}_T(T_i - T_0) \quad (20)$$

where:

ε: The dependent variable of the current time relative to the initial position (10^{-6});
k: The steel string strain gauge minimum reading 10^{-6}/(kHz^2);
f_i^2: The Steel string strain gauge current output modulus kHz^2;
f_0^2: The Steel string strain gauge initial output frequency modulus kHz^2;
k_T: The steel string strain gauge temperature correction factor 10^{-6} °C;
T_i: The steel string strain gauge current time temperature value (°C);
T_0: Temperature value when measuring f_0 (°C).

When using polynomial fitting, the coefficients a and b were calculated using the least squares fitting method, and the strain calculation method for integrating temperature changes is shown in Equation (21):

$$\varepsilon = \mathrm{a} \times (f_i^4 - f_0^4) + \mathrm{b}(f_i^2 - f_0^2) + \mathrm{k}_T(T_i - T_0) \quad (21)$$

The abovementioned various theoretical calculation methods and vibrating wire strain gauge measurement characteristics can better measure the strain, eliminate the strain generated by environmental influence, more accurately understand the mechanical strain of structure, and analyze the stress state of structural facility.

3. Test Plan

Two sensors were installed in parallel on the 45th steel tooling for free acquisition at the same frequency. The device (Figure 6) shows the material properties of No. 45 steel (Table 1). The acquisition device is equipped with a high-frequency dynamic acquisition device, which is a dynamic vibrating wire automatic acquisition system. This system uses a nonsweeping technology scheme to prevent the steel string vibration from attenuating, and an embedded mirror oscillation circuit to ensure excitation frequency. By matching the phase and true motion of steel cord and simultaneously detecting the resonant frequency of steel string through several waveform periods, noise immunity and resolution of measurement have largely improved compared with the flat-domain periodic averaging method.

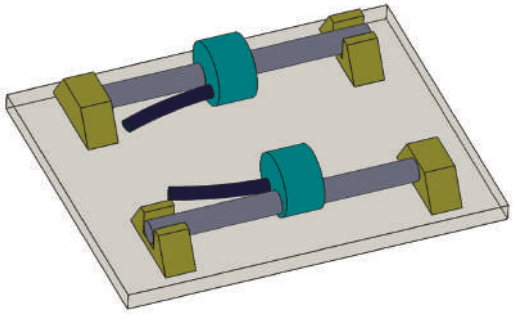


Figure 6. Caption for figure schematic diagram of the test fixture.

Table 1. Forty-five steel material characteristics.

Content	Value and Unit
Density	7.85 g/cm^3
Modulus of elasticity	210 GPa
Poisson ratio	0.269
Tensile strength	600 MPa
Yield strength	355 MPa
Elongation	16%
Section shrinkage	40%
Impact work	39 J

The measurement module used the patented 8,671,758 products, and the dynamic measurement rate is 20 to 333 Hz. At the same time of dynamic measurement, the module also performs auxiliary measurement, performs static measurement at 1 Hz, provides finer measurement resolution, and better anti-interference to external noise source performance. The thermistor input signal of each vibrating string acquisition channel was measured at a high resolution of 24 bits at 1 Hz. The performance of thermistor parameters is shown in Table 2. The excitation module has a resolution of 26 mV and a dynamic measurement rate of 20, 50, 100, 200, and 333.33 bHz. The range of sensor resonance frequency is shown in Table 3. The measurement frequency accuracy is ± (0.005% reading + measurement resolution). The noise level corresponding to the measurement resolution b at different sampling rates is shown in Tables 2–4.

Table 2. Thermistor performance parameters.

Content	Value and Unit
Half bridge arm	0.1% accurate resistance is 4.99 KΩ
Excitation voltage	1.5 V
Resolution	0.002 Ω RMS @ 5 KΩ thermistor
Accuracy (−55–85 °C)	0.15% of reading
Measurement rate	1 Hz

Table 3. Range of sensor resonance frequencies.

Sample Rate (Hz)	Minimum Sensor Frequency (Hz)	Maximum Sensor Frequency (Hz)
20	290	6000
50	290	6000
100	580	6000
200b	1150	6000
333b	2300	6000

Table 4. Measurement resolution b (typical value for a 2.5 kHz resonant frequency sensor).

Sample Rate (Hz)	Noise Level (Hz RMS)
1	0.005
20	0.008
50	0.015
100	0.035
200C	0.11
333C	0.45

In the measurement, both steel plate and strain gauge will produce different deformations due to temperature changes. To minimize the acquisition error, a correlation curve between temperature and strain was analyzed. During the test, welding was used. To solder the two sensor holders on the tooling and for better verification, the effect of temperature response, and correlation, when collecting

the zero point, one sensor was set to the free state, one sensor passed the fixture. The force was applied such that the sensor zero point acquisition and the first sensor's zero point acquisition strain difference is about 140 $\mu\varepsilon$.

4. Analysis of Test Data

The relationship between temperature and strain measurement shows that the strain changes correspondingly when the temperature changes. The strain reversely sways as the temperature changes. We got a good corresponding relationship. The temperature and strain data collected using the two sensors A and B were compared, and the correlation between the two data was analyzed. During the analysis, there are usually two ways of correlation: Autocorrelation and cross-correlation. The autocorrelation function is known as the autocorrelation equation and used to describe the correlation between the correlation functions of related data at different times as shown in Equation (22):

$$R_f(\tau) = f(\tau) * f^*(-\tau) = \int_{-\infty}^{\infty} f(t+\tau) f^*(t) dt = \int_{-\infty}^{\infty} f(t) f^*(t-\tau) dt \tag{22}$$

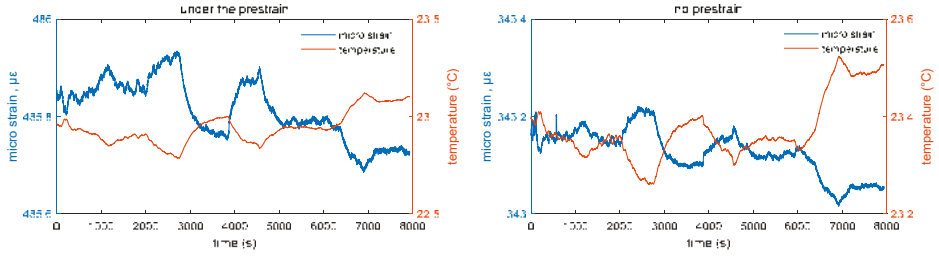
At the same time, cross-correlation or cross-covariance were also used to represent a measure of similarity between two signals. Cross-correlation mainly analyzes the degree of correlation between two time series. Cross-correlation is essentially similar to the convolution of two functions. For discrete functions f_i and g_i, it can be defined as Equation (23):

$$(f * g)_i \equiv \sum_j f_j^* g_{i+j} \tag{23}$$

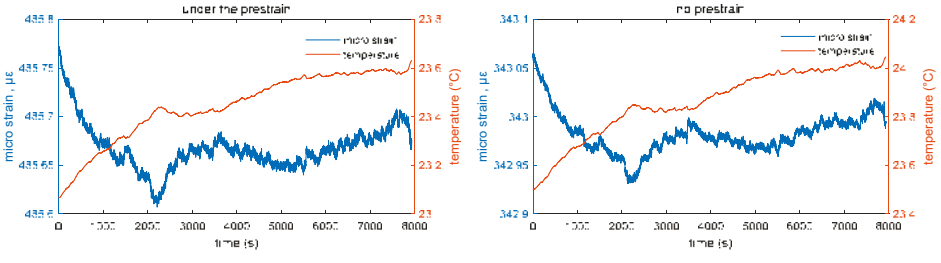
If the continuous signal is set to two sets of $f(x)$ and $g(x)$, then the cross-correlation is defined as Equation (24):

$$(f * g)(x) \equiv \int f^*(t) g(x+t) dt \tag{24}$$

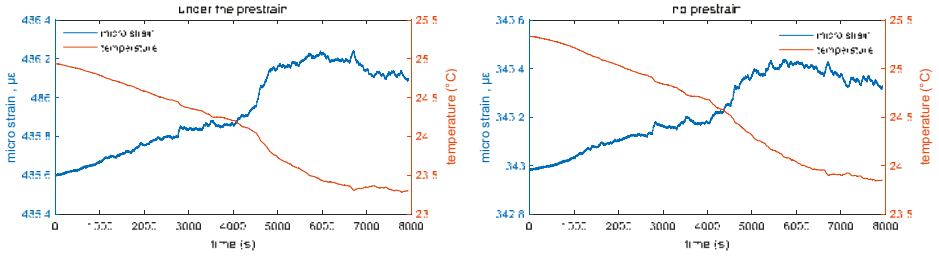
During this test analysis, the data collected using the two sensors A and B were compared, and the correlation between the two sets of data was analyzed. Their similarities were derived to verify the accuracy and reliability of dynamic acquisition of data obtained using sensors A and B. The correlation coefficient between the temperature data of sensors A and B was calculated to be 0.9983, and the correlation degree is highly correlated. The correlation coefficient between the strain data of sensors A and B is 0.9895, and the correlation degree is highly correlated. Cross-calculation can be obtained. The correlation between temperature and strain of sensor A is −0.6683, and the correlation between temperature and strain of sensor B is −0.5573. The experimental results show that the values are significantly correlated (Figure 7).



(**A**) Temperature change of 0.2 °C



(**B**) Temperature change of 0.6 °C



(**C**) Temperature change of 1.8 °C

Figure 7. Free state strain values of strain sensors A and B in a temperature change of 0.2, 0.6, and 1.8 °C.

During the test, to better use the knowledge of metrology to analyze the dynamic RMS change output, verify the reliability of linear relationship between temperature and strain monitoring, and repeat the calculation of temperature measurement and strain measurement, the experiment measured the comprehensive reflection of various random influencing factors, including the instability of tooling materials used, random error caused by the laying process, instability of sensing instrument, environmental conditions, and other factors, as well as the actual measured randomness. The measured object also affects the dispersion of measured values, especially when the random variation in the measured object during the bridge monitoring process is large. Therefore, the dispersion of experimentally measured values is typically slightly greater than the dispersion introduced by the sensor static calibration standard itself. To be less affected by outliers, all the deformations and the corresponding temperature were measured. During the calculation of repeatability of the experiment, the arithmetic mean value of measured value should be calculated first, and the calculation formula can be expressed as Equation (25):

$$\overline{F} = \frac{\sum_{i=1}^{n} F_i}{n} \tag{25}$$

where:

F_i—The strain measurement indication of the i-th measurement, $\mu\varepsilon$;
n—The number of measurements.

The measurement repeatability can be quantitatively expressed using the experimental standard deviation $\mathrm{Sr(y)}$, and the calculation formula can be expressed as Equation (26):

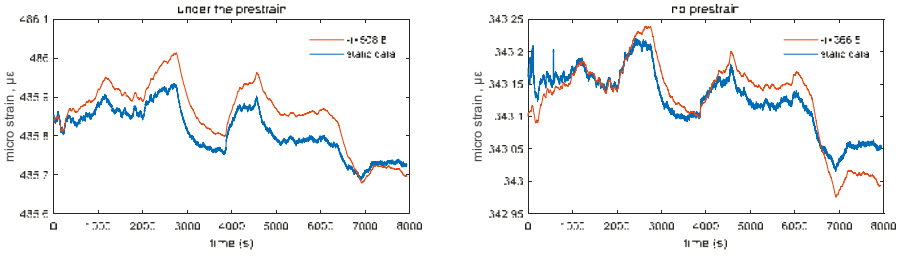
$$\mathrm{Sr\ (y)} = \sqrt{\frac{\sum_{i=1}^{n} (F_i - \overline{F})^2}{n-1}} \tag{26}$$

where:

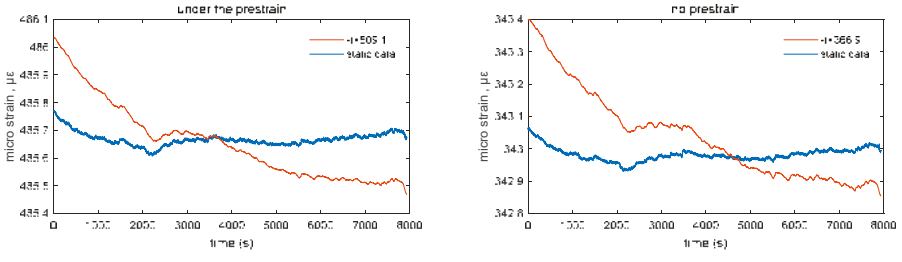
F_i—The measured strain measured at the i-th measurement, $\mu\varepsilon$;
$\overline{F}$—The arithmetic mean of indications of strain measurement, $\mu\varepsilon$;
n—number of measurements.

Using the above formula, under the synchronous acquisition test conditions, sensors A and B were repetitively used for the temperature and strain measurements, respectively. The repeatability of sensor A strain indication is 0.16071, and the temperature measurement repeatability is 0.09971. The repeatability of strain measurement of sensor B is 0.11743, and the repeatability of temperature measurement is 0.08209. The repeatability shows that the stability of sensor B is better, and the experiment can be performed under the condition that the initial state of sensor is small. The measured values are more stable and more clearly characterized by the corresponding relationship between temperature and strain.

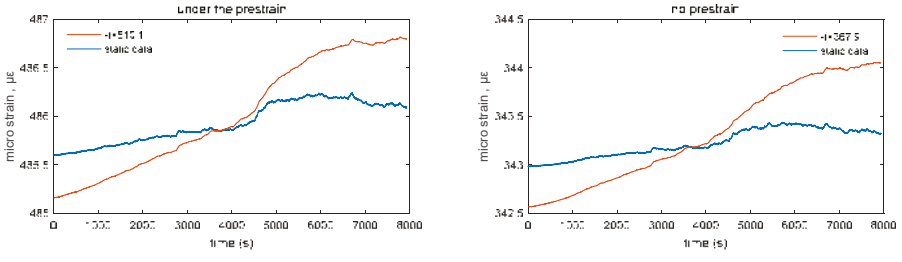
At the same time, the relationship between strain and temperature was analyzed during a temperature variation of 0.2, 0.6, and 1.8 °C. Figures 4–6 show that the correlation did not change with the change in temperature. A stable negative correlation curve relationship was maintained. Among them, sensor A has a relatively weak induction of temperature and deformation under the prestrain of 140 $\mu\varepsilon$, which can be characterized as shown in the figure. In the absence of external excitation, the change in experimental numerical temperature is less than the perception of sensor B. At the same time, the corresponding strain produces a small change in the output. Sensor A temperature and the strain relationship map shows that in the free state of acquisition device, the experimental temperature will also have a corresponding change during small changes. However, both sensors A and B can consistently respond to the corresponding output temperature and strain signals (Figure 8).



(**A**) Strain-temperature relationship at 0.2 °C



(**B**) Strain-temperature relationship at 0.6 °C.



(**C**) Strain-temperature relationship at 1.8 °C.

Figure 8. Strain-temperature relationship between 0.2, 0.6, and 1.8 °C intervals for strain gauges A and B.

After calculating different temperature changes, the experimental values show the correlation between strain and temperature. By analyzing the strain-temperature correlation function, it can be found that the temperature change is less than that of sensor B when there is no external excitation. At the same time, the change in output corresponding to the strain is small. In different temperature variation ranges, the experimental data were used to analyze the relationship between the temperature and strain of sensor A. It was observed that the acquisition device is in a free state, and the temperature is small. The effect of corresponding change will be produced. According to statistical analysis, the strain-temperature correlation can be expressed as $\Delta\mu\varepsilon = \Delta T + 510$, and the strain measurement result has a negative correlation with the temperature change. From the relationship between the temperature and corresponding strain curves of sensor B in the free state, it can be observed that the strain-temperature correlation can be expressed as $\Delta\mu\varepsilon = \Delta T$ in the state where sensor B is more free under the same conditions as sensor A. $\Delta\mu\varepsilon = \Delta T + 367$; the strain measurement results have a stable negative correlation with temperature change. During data acquisition, it was found that strain gauge A is larger at the same temperature due to prestressing, the strain output is larger than the output of sensor B, and the correlation between temperature, peak, and valley is more prominent. It was assumed

that 0.2 is a one-fold change in temperature. During the three-fold change and nine-fold change, the steady-state strain curve is stable, and the correlation function is not fluctuating. However, in the state where the prestrain of sensor A is 140 $\mu\varepsilon$, the induction of temperature and deformation is relatively weak, and the difference is about 3–6 $\mu\varepsilon$, which is about 2–5% of the prestress, as shown in Figures 7 and 8. In the later test, more theoretical and experimental verifications are needed. Combined with the artificial intelligence model training method, the expansion effect caused by the material itself after the vibration of vibrating string was analyzed, and the environmental impact error was maximized. The tooling surface was optimized for temperature conductivity. Since this test is the calibration test for the maximum fitting, instrumentation error, etc., the welding process was used, so that the above effects can be neglected. At the same time, referring to Table 4, the noise of the device is 9.775×10^{-9} $\mu\varepsilon$, which can be ignored. Through comprehensive evaluation, it was found that the method of synchronously collecting sensors at similar positions can achieve the online comparison of the corresponding variable measuring sensors. The later test can be calibrated by high-frequency dynamic acquisition to calibrate the low-frequency acquisition and combined with the relevant optimization algorithm for calculation. The maximum optimization restores the original deformation magnitude, and at the same time reduces the instability of constitutive performance parameters such as the stiffness and strength of structure. The experimental results show the calibration of the corresponding variable monitoring sensor.

5. Conclusions

In engineering, the strain test method is usually used to determine the actual stress state of structure and monitor the structural deformation increment, and the structural strain test is an important method to solve the structural strength problem. During strain detection and monitoring, the calibration of strain gauge plays a significant role in the accuracy of data. To ensure the accuracy of traceability of magnitude, we should study the corresponding relationship between the temperature and strain of strain gauge and the online calibration test method, which is important for the online metrological traceability system of strain monitoring sensor. We focused on researching the influence of temperature and prestress effect on sensors by analyzing differences of their measurement results in specified situation. In this study, a dynamic vibrating wire measurement module was used to dynamically collect the temperature and strain of two vibrating wire sensors in the free state, and the correlation between temperature and strain was analyzed. The correlation coefficient between the temperature data of sensors A and B were statistically analyzed. The degree of relevance 0.9983 is highly correlated. The experimental results show that the correlation coefficient between the strain data of sensors A and B is 0.9895, and the correlation degree is highly correlated. After cross-calculation, a significant correlation was observed between temperature and strain indications. An experiment was carried out to analyze the measurement repeatability by using the knowledge of metrology. After analysis, it was found that the repeatability of temperature and strain values is within 0.2, and the consistency between the measured results of the same measurement is good. At the same time, the relationship between strain and temperature in different temperature variation intervals was analyzed. It was found that the temperature change has a small effect on the strain in the prestressed state, but the relationship between temperature and strain is relatively stable. Finally, the experiment verified the feasibility of parallel measurement scheme corresponding to the online calibration of variable monitoring sensor and analyzed the effect of temperature corresponding to the variable measurement under different prestress conditions, providing theoretical and experimental verification support for establishing an online metrological verification standard. This study can also better confirm the accuracy of bridge health monitoring data, effectively reduce the monitoring error caused by temperature drift, effectively eliminate the impact of environmental load, and better feedback the essential deformation of bridge structure. Considering the reliability risk of sensors after a long-time work period, such as signal drift, the approach proposed in this paper can be used as a substitute to such strain monitoring system in order to determine whether these sensors should be replaced. We expect our approach to be expandable

to various applications. It can provide important support for the effective monitoring data extraction of the bridge health monitoring system.

Author Contributions: Conceptualization, L.P., G.J., Z.L., Y.W., and B.Z.; Data curation, L.P., X.Y., and Y.W.; Formal analysis, Y.W.; Funding acquisition, L.P.; Writing of original draft, L.P.; Writing—review and editing, L.P. All authors have read and agree to the published version of the manuscript.

Funding: This research was funded by the Project of National Key R & D Program of China grant number 2017YFF0206305 and The Key Projects of Public Scientific Research Institutes (2018-9031).

Acknowledgments: The authors would like to acknowledge the support of National Key R & D Program of China under grant number 2017YFF0206305 and the Key Projects of Public Scientific Research Institutes (2018-9031).

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

References

1. Mao, J.X.; Wang, H.; Feng, D.M.; Tao, T.Y.; Zheng, W.Z. Investigation of Dynamic Properties of Long-Span Cable-Stayed Bridges Based on One-Year Monitoring Data Under Normal Operating Condition. *Struct. Control Health Monit.* **2018**, *25*, e2146. [CrossRef]
2. Wang, H.; Zhang, Y.M.; Mao, J.X.; Wan, H.P.; Tao, T.Y.; Zhu, Q.X. Modeling and Forecasting of Temperature-induced Strain of a Long-span Bridge using an Improved Bayesian Dynamic Linear Model. *Eng. Struct.* **2019**, *192*, 220–232. [CrossRef]
3. Mao, J.X.; Wang, H.; Li, J. Fatigue Reliability Assessment of a Long-Span Cable-Stayed Bridge Based on One-Year Monitoring Strain Data. *J. Bridge Eng.* **2019**, *24*, 05018015. [CrossRef]
4. He, H.; Wang, W.; Tian, D.; Sun, J.; Xiong, C. Optimization of Vibrating Wire Sensor Excitation Strategy. *Chin. J. Sens. Actuators* **2010**, *23*, 74–77. [CrossRef]
5. Wen, Z.; Xia, Z.; Li, F. Research on Frequency Measurement Technology of String Vibration Sensor. *J. Xi'an Poly. Univ.* **2012**, *1*, 75–79.
6. Tian, Z.; Huang, Z. Nonlinear Compensation of Vibrating Wire Sensor Based on Spline Interpolation and Least-Squares Method. *Meas. Control Technol.* **2016**, *35*, 7–10.
7. Chen, N.; Li, H.; He, H.; Xie, K. An Self-Adaptive Vibrating Pick-up Method for Vibrating Wire Sensor. *J. Guangxi Univ. (Nat. Sci. Ed.)* **2017**, *42*, 1145–1150.
8. Xu, Y.; Yang, X.; Wei, T.; Yao, J. Analysis of Strain Transfer Influence Factors of Resistance Strain Sensor. *China Meas. Test* **2018**, *1*, 136–142.
9. Zhang, Y.; Liang, P.; Zhang, Y.; Chen, W.; Huang, M. Design of Calibration Device for Vibrating Wire strain Transducer. *Ind. Instrum. Autom.* **2011**, *4*, 52–55.
10. Guangzhou Guangcai Testing Instrument Co., Ltd. *Vibrating Wire Strain Gauge Calibration Instrument: China*; Guangzhou Guangcai Testing Instrument Co., Ltd.: Guangzhou, China, 2016.
11. Bai, S.; Xiao, Y.; Huang, B.; Liu, G. Research on Strain Calibration Method of Fiber Bragg Grating Sensor. *J. Vib. Mea. Diagn.* **2016**, *36*, 321–324.
12. Zhang, H.; Nie, F.; Fan, D. Calibration Method, Device and System of Fiber Bragg Grating Sensor. CN106871810A, 20 June 2017.
13. Chen, C.; Yan, D.; Chen, Z.; Tu, G.; Tian, Z. Technique Research of Vibrational Chord Strain Gauge to Concrete. *China J. Highw. Transp.* **2004**, *17*, 29–33.
14. Bai, T.; Deng, T.; Xie, J.; Hu, F.P. Accurate Mathematical Model of Vibrating Wire Sensor and Its Application. *Chin. J. Rock Mech. Eng.* **2005**, *24*, 5965–5969.
15. Iadicicco, A.; Campopiano, S. Sensing Features of Long Period Gratings in Hollow Core Fibers. *Sensors* **2015**, *15*, 8009–8019. [CrossRef] [PubMed]
16. Wang, Y.B.; Zhao, R.D.; Chen, L.; Xu, Y.; Xie, H.Q. Temperature Correction Test of Vibrating Wire Strain Sensor. *J. Archit. Civ. Eng.* **2017**, *34*, 68–75.
17. Lee, H.M.; Park, H.S. Measurement of Maximum Strain of Steel Beam Structures Based on Average; Strains from Vibrating Wire Strain Gages. *Exp. Technol.* **2013**, *37*, 23–29. [CrossRef]
18. Choi, S.W.; Kwon, E.; Kim, Y.; Hong, K.; Park, H.S. A Practical Data Recovery Technique for Long-Term Strain Monitoring of Mega Columns during Construction. *Sensors* **2013**, *13*, 10931–10943. [CrossRef] [PubMed]

19. Park, H.; Lee, H.; Choi, S.; Kim, Y. A Practical Monitoring System for the Structural Safety of Mega-Trusses Using Wireless Vibrating Wire Strain Gauges. *Sensors* **2013**, *13*, 17346–17361. [CrossRef] [PubMed]
20. Barot, D.; Wang, G.; Duan, L. High-Resolution Dynamic Strain Sensor Using a Polarization-Maintaining Fiber Bragg Grating. *IEEE Photonics Technol. Lett.* **2019**, *31*, 709–712. [CrossRef]
21. Jin, X.; Sun, C.; Duan, S.; Liu, W.; Li, G.; Zhang, S.; Chen, X.; Zhao, L.; Lu, C.; Yang, X.; et al. High Strain Sensitivity Temperature Sensor Based on a Secondary Modulated Tapered Long Period Fiber Grating. *IEEE Photonics J.* **2019**, *11*, 1–8. [CrossRef]

sensors

MDPI

Article

Train Hunting Related Fast Degradation of a Railway Crossing—Condition Monitoring and Numerical Verification

Xiangming Liu * and Valéri L. Markine *

Department of Engineering Structures, Delft University of Technology, 2628 CN Delft, The Netherlands
* Correspondence: Xiangming.Liu@tudelft.nl (X.L.); V.L.Markine@tudelft.nl (V.L.M.)

Received: 4 March 2020; Accepted: 15 April 2020; Published: 17 April 2020

Abstract: This paper presents the investigation of the root causes of the fast degradation of a railway crossing. The dynamic performance of the crossing was assessed using the sensor-based crossing instrumentation, and the measurement results were verified using the multi-body system (MBS) vehicle-crossing model. Together with the field inspections, the measurement and simulation results indicate that the fast crossing degradation was caused by the high wheel-rail impact forces related to the hunting motion of the passing trains. Additionally, it was shown that the train hunting was activated by the track geometry misalignment in front of the crossing. The obtained results have not only explained the extreme values in the measured responses, but also shown that crossing degradation is not always caused by the problems in the crossing itself, but can also be caused by problems in the adjacent track structures. The findings of this study were implemented in the condition monitoring system for railway crossings, using which timely and correctly aimed maintenance actions can be performed.

Keywords: railway crossing; wheel-rail impact; train hunting; numerical verification; railway track maintenance

1. Introduction

In the railway track system, turnouts (switches and crossings) are essential components that allow trains to pass from one track to another. A standard railway turnout is composed of three main parts: switch panel, closure panel, and crossing panel, as shown in Figure 1. In a railway turnout, the crossing panel is featured to provide the flexibility for trains to pass in different routes.

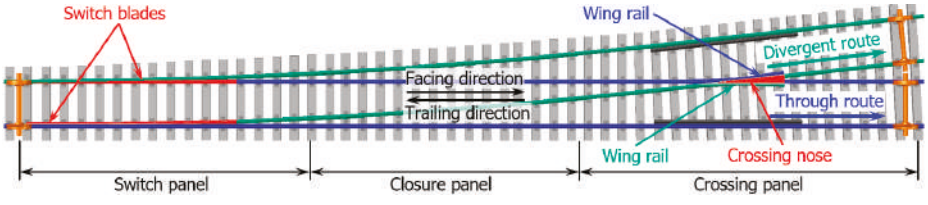


Figure 1. Standard left-hand railway turnout and the definition of the passing routes.

For rigid crossings that are commonly used in conventional railway lines, the gap between the wing rail and the nose rail usually results in high wheel-rail impacts in the transition region where the wheel load transits from the wing rail to the nose rail (vice versa, Figure 2), which makes the crossing a vulnerable spot in the railway track. In the case of crossings that are mainly used for the through route traffic (e.g., crossings in the crossover), there is no specific speed limit [1] and trains can pass through

the crossings with a high velocity of up to 140 km/h. The high train velocity makes the wheel-rail impact more serious. In the Dutch railway system, around 100 crossings are urgently replaced every year [2] due to unexpected fatal defects, which not only result in substantial maintenance efforts, but also lead to traffic disruption and can even affect traffic safety.

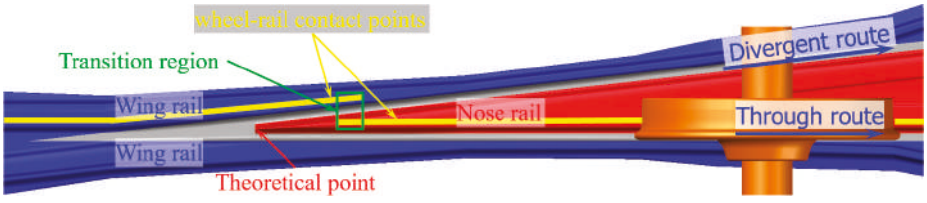


Figure 2. Wheel-rail interaction in the railway crossing for through route traffic.

In contrast to a switch panel, wherein sensors are instrumented for condition monitoring [3,4] and remaining useful life prediction [5], monitoring in a crossing panel is usually absent. As a result, the real-time information on the condition of railway crossings is limited. The present maintenance activities are mainly reactive and based on the experience of the contractors. In this case, the root causes of the crossing degradation are not always resolved by the maintenance actions, and the crossings are likely to be operated in a degraded condition. To improve this situation, necessary guidance for maintenance actions is highly required.

Proper crossing maintenance usually relies on condition assessment and degradation detection, which can be realized through field monitoring. In recent years, condition monitoring techniques have been frequently applied in the railway industry. Aside from the above-mentioned instrumentation on the turnout switches, vehicle-based monitoring systems have been applied in track stiffness measurement [6] and estimation [7], track alignment estimation [8], hanging sleepers detection [9], and track fault detection [10], etc. Compared with the normal track, the current studies on railway crossings are mainly based on numerical simulation. Typical contributions include wheel-rail interaction analysis [11–21], damage analysis [16,17,22,23], and prediction [18,24,25] as well as crossing geometry and track stiffness optimization for better dynamic performance [16,26]. Field measurements are mainly used for the validation of numerical models. The monitoring of railway crossings for condition assessment and degraded component detection is still limited.

In the previous study, key indicators for the crossing condition assessment based on the field measurement were proposed [27,28]. Additionally, a numerical vehicle-crossing model was developed using a multi-body system (MBS) method to provide the fundamental basis for the condition indicators [29]. In this study, the condition indicators, as well as the MBS model, were applied in the condition monitoring of a fast degraded railway crossing. The main goals of this study were to investigate the root causes of the crossing degradation as well as to assess the effectiveness of the current maintenance actions.

Based on the objectives, this paper is presented in the following order. The experimental and numerical tools, including the crossing condition indicators, are briefly introduced in Section 2. The measurement results and the crossing degradation analysis as well as the effectiveness of the current maintenance actions are presented in Sections 3 and 4. Based on the measurement results and field inspections, the root causes for the fast crossing degradation were investigated with the assistance of the MBS model, as presented in Section 5. In Section 6, the verification of the effectiveness of the maintenance actions is given. Finally, in Section 7, major conclusions are provided.

2. Methodology

In this section, the experimental tools for the crossing condition monitoring, as well as the indicators for the crossing condition assessment, are briefly introduced. The MBS vehicle-crossing model for the verification of the experimental findings is also presented.

2.1. Experimental Tools

The experimental tools mainly consisted of the in-site instrumentation system modified from ESAH-M (Elektronische System Analyse Herzstückbereich-Mobil) and the video gauge system (VGS) for wayside monitoring, as briefly described below. Both tools have already been introduced and actively applied in previous studies. Detailed information regarding the installation and data processing can be found in [27,30].

2.1.1. Crossing Instrumentation

The main components of the crossing instrumentation are an accelerometer attached to the crossing nose rail for 3-D acceleration measurement, a pair of inductive sensors attached in the closure panel for train detection as well as train velocity calculation, and the main unit for data collection. An overview of the instrumented crossing is shown in Figure 3.

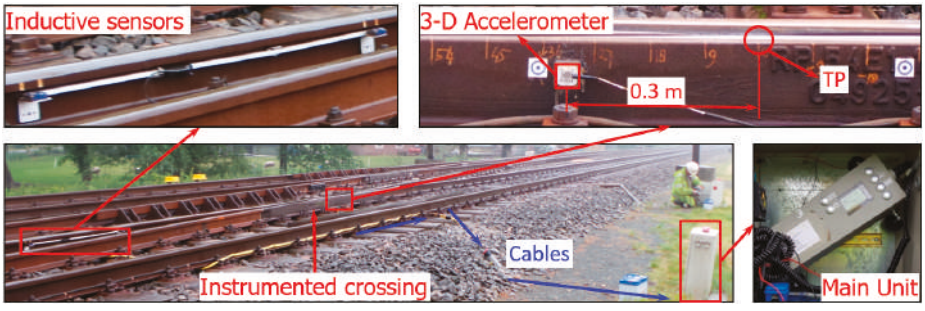


Figure 3. Crossing instrumentation based on ESAH-M.

The main outputs of the crossing instrumentation were the dynamic responses of the crossing nose, including the wheel-rail impact accelerations and locations, etc. All these responses were calculated within the transition region, which can be obtained through field inspection [29]. Based on these measured responses and the correlation analysis between the responses [28], two critical condition indicators related to the wheel impact and fatigue area, respectively, were proposed.

The wheel impact is reflected by the vertical accelerations, which were obtained from the crossing and processed through statistical analysis. This indicator is mainly based on the magnitude of the impacts due to each passing wheel (Figure 4a), and the changes in time indicate the different condition stages of the crossing (Figure 4b).

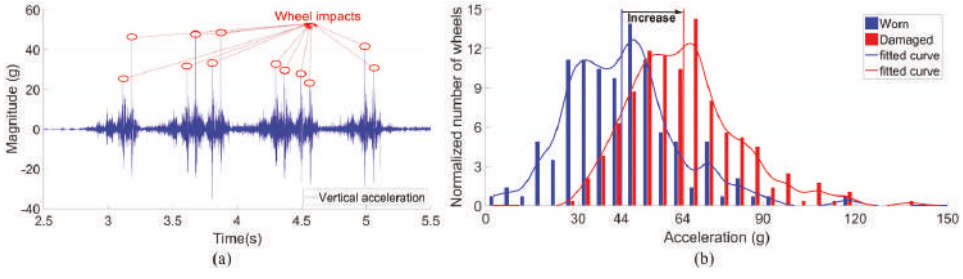


Figure 4. Indicator for the wheel impact. (**a**) Procedure for the obtainment of wheel impacts. (**b**) Example of the variation of the wheel impacts in different condition stages.

The fatigue area is defined as the region where the majority of wheel impacts are located on the crossing, and where ultimately the crack initiates (Figure 5a). In practice, the fatigue area can

be simplified as the confidence interval of $[a - \sigma, a + \sigma]$, where a is the mean value of the wheel-rail impact locations, and σ is the standard deviation. The location and size of the fatigue area are critical values for the assessment of crossing wear and plastic deformation. A wide fatigue area usually represents well-maintained rail geometry. As demonstrated in Figure 5b, when the crossing condition was degraded from "Worn" to "Damaged", the fatigue area was dramatically narrowed and shifted further from the theoretical point (TP) of the crossing. More information about the fatigue area can be found in the previous study [27].

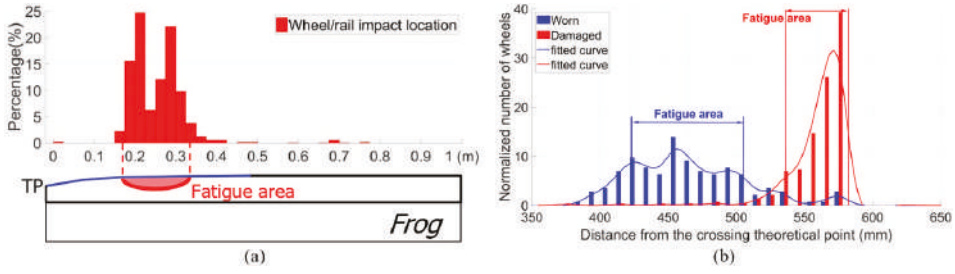


Figure 5. Demonstration of the crossing fatigue area detection. (**a**) Definition of the fatigue area. (**b**) Example of the fatigue area changes in different crossing condition stages.

2.1.2. Wayside Monitoring System

The VGS for wayside monitoring is a remote measurement device based on digital image correlation (DIC). It uses high-speed digital cameras to measure the dynamic movements of the selected targets in the track. The system, set up together with the targets installed on the crossing rail next to the instrumented accelerometer, is shown in Figure 6a, and the demo of the displacement measurement is shown in Figure 6b. The main outputs are the vertical displacements of the tracked targets with a stable sampling frequency of up to 200 Hz.

Figure 6. Wayside monitoring. (**a**) System setup. (**b**) The screen of displacement measurements.

Due to the limitation of the experimental conditions, the wayside monitoring system is usually set up close by the side of the track, which will introduce extra noise in the measured displacement results. To improve the accuracy of the measurement, the noise part needs to be eliminated. The noise mainly comes from the ground-activated camera vibration, which can be manually activated by hammering the ground near the camera. The measured camera vibrations in both the time and frequency domains are given in Figure 7.

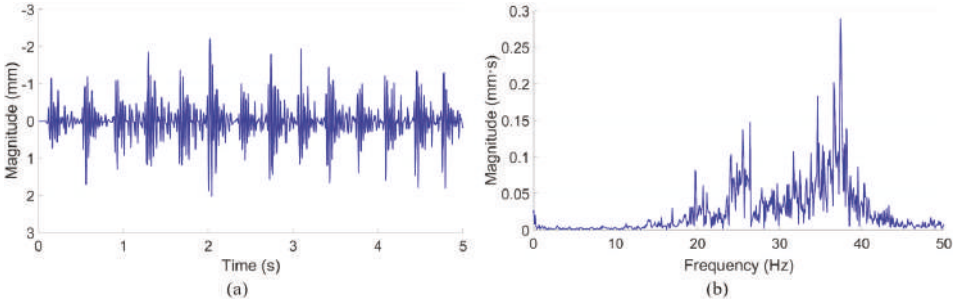


Figure 7. Ground activated camera vibration. (**a**) Time domain signal. (**b**) Frequency domain responses.

Despite the differences in the displacement responses in the two monitored crossings, the main resonance of the camera vibration was around 15–45 Hz. In the previous study [30], the main components in the displacement signal were elaborated. The train-track components related to displacement responses are mainly distributed below 10 Hz, which do not overlap with the camera vibration introduced noise. The noise part due to camera vibration can then be reduced through low-pass filtering, as shown in Figure 8.

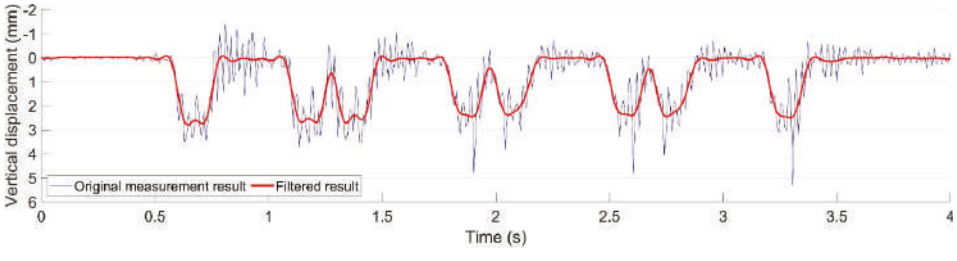


Figure 8. Examples of the measured rail vertical displacement.

The magnitude of the dynamic vertical displacement of the rail directly reflects the intensity of the track movement due to the passing trains. By comparing the measured rail displacement with the reference level, which can be obtained from numerical simulation using the parameters in the designed condition, the ballast settlement level of the monitored location can be estimated. The MBS model for the crossing performance analysis is described later in this section.

2.2. Multi-Body System (MBS) Vehicle-Crossing Model

The numerical model for the crossing performance analysis was developed using the MBS method VI-Rail (Figure 9a). The rail pads, clips, and ballast were simulated as spring and damping elements (rail busing and base busing, Figure 9b). In the vehicle model, the car body, bogie frames and the wheelsets were modeled as rigid bodies with both the primary suspension and secondary suspension taken into account (Figure 9b). The track model was a straight line with the crossing panel (Figure 9c) situated in the middle of the track. The rail element for the acceleration and displacement extraction was the lumped rail mass located 0.3 m from the TP of the crossing (Figure 9d), which is consistent with the setup of the field measurements (Figures 3 and 6a).

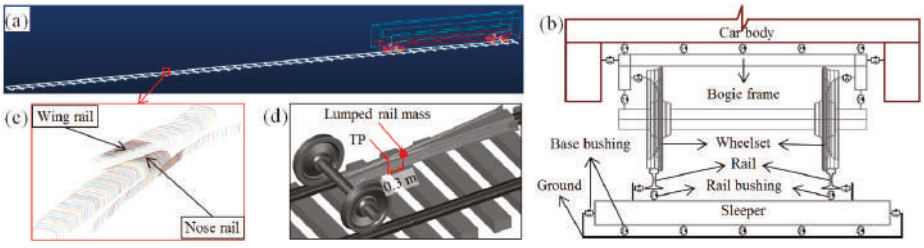


Figure 9. Multi-body system (MBS) model. (**a**) Vehicle-track model. (**b**) Flexible connections in the model. (**c**) Crossing profiles. (**d**) Rail element for acceleration extraction.

The detailed model development, experimental validation, and numerical verification can be found in the previous study [29]. Corresponding to the condition indicators, the main outputs of the MBS model are the wheel impact acceleration, transition region and wheel-rail contact forces. Using the MBS model, the condition of the monitored crossing, as well as the detected track degradations, can be verified.

3. Field Measurements and Analysis

The monitored crossing was a cast manganese crossing with an angle of 1:9. As part of a crossover, trains mainly pass the crossing in the facing through route (Figure 2) with a velocity of around 140 km/h. The on-site view of the crossing is shown in Figure 10a. According to the maintenance record, this crossing was suffering from fast degradation with the service life of only around three years (18 years on average [2]). At the beginning of the condition monitoring, the damaged crossing was completely renovated.

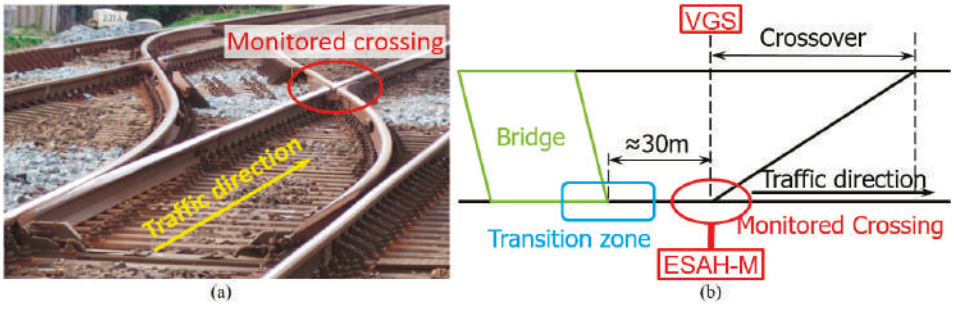


Figure 10. Overview of the monitored crossing. (**a**) On-site view. (**b**) Sketch view.

Figure 10b gives a sketch view of the crossing, including the setup of the monitoring devices and the layout of the adjacent structures, especially the small bridge in front of the crossing. Considering that the bridge is located quite close to the monitored crossing, the performance of the crossing might be affected by the bridge, which will be discussed later.

The measurement results from the crossing instrumentation were based on multiple train passages in one monitoring day. For the wayside monitoring, one sufficient train passage is enough to estimate the ballast condition. To maximally reduce the influence of the vehicle-related variables, the selected results were restricted to the commonly operated VIRM trains with velocities of around 140 km/h.

3.1. Wheel Impacts

Based on the estimated transition regions, the wheel impact accelerations were calculated. The distribution of the wheel impacts due to multiple wheel passages is shown in Figure 11a.

The sample size, in this case, was 78 passing wheels. It can be seen that the wheel impacts presented a bimodal distribution. Around 80% of the wheel impacts were below 50 g, while the remaining 20% of the wheel impacts were extremely high with a mean value of around 350 g. Such a polarized distribution of impacts indicates the highly unstable wheel-rail interaction in this crossing. It was demonstrated in a previous study [29] that for this type of railway crossing, the average level of the wheel impact is around 50 g, meaning that the 20% of high impacts of the monitored crossing are already more than seven times higher than the average impact level. It can be imagined that such high impacts will dramatically accelerate the degradation procedure of the crossing.

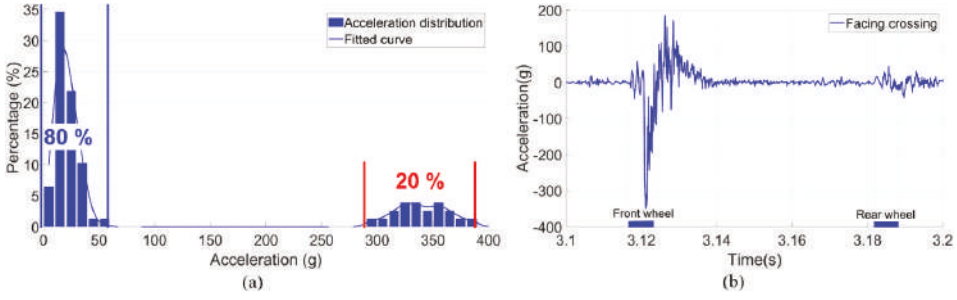


Figure 11. Vertical acceleration responses of the monitored crossings. (**a**) Distribution based on multiple train passages in one day. (**b**) Example of impacts due to one bogie.

An example of the impact acceleration response in the time-domain due to the first bogie of a VIRM train is shown in Figure 11b. It can be seen that for the two passing wheels from the same bogie, the impacts can be quite different. The impact due to the front wheel was up to 350 g, while the rear wheel activated vertical acceleration was only 20 g. It has to be noted that the high impacts were not always introduced by the front wheel, but appeared to have random occurrences. Such results further confirmed the instability of wheel-rail interaction at this crossing.

3.2. Fatigue Area

The measured fatigue area of the monitored crossing is presented in Figure 12. It can be seen that the wheel impacts were widely distributed at 0.22–0.38 m from the TP with the fatigue area size of 0.16 m. According to the previous study [28], the transition region (Figure 2) for this type of crossing is around 0.15–0.4 m. The fatigue area widely covered 64% of the transition region, which can be considered to be in line with the expectation of a new crossing profile. Such results further confirmed that the crossing rail was not worn or deformed.

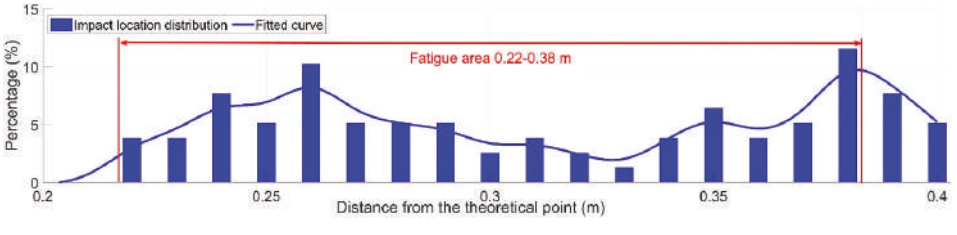


Figure 12. Measured fatigue area of the monitored crossing.

It has to be noted that the fatigue area does not conform to the normal distribution (referring to the "Worn" stage demonstrated in Figure 5b). Combined with the results of the wheel impacts such a fatigue area further confirmed the instability of the wheel-rail contact in the monitored crossing.

In a previous study [27], it was found that the crossing degradation was accompanied by the increase of wheel-rail impacts and the reduction in the fatigue area. The large number of extremely high wheel-rail impacts and relatively wide fatigue area clearly indicate the abnormal performance of the monitored crossing. Finding the root causes of such abnormality is the key to improving the dynamic performance of the crossing.

3.3. Ballast Settlement

The measured vertical displacement of the crossing rail is presented in Figure 13. It can be seen that the vertical rail displacement was around 4 mm. The measured displacement result can be considered to have two main parts: the elastic deformation and the gap between the sleeper and ballast. Considering that the ballast settlement is the accumulated effect due to multiple wheel passages, the plastic deformation caused by each passing train can be neglected. Due to the high impacts in the crossing panel, the ballast is usually settled unevenly, which results in hanging sleepers. Using the validated MBS model, it was calculated that the rail displacement in the reference condition was 1.4 mm (Figure 13), which only consisted of the elastic deformation part. By comparing these two results, it could be calculated that the gap between the sleeper and ballast was 2.6 mm, which can be estimated as the settlement of ballast. It was observed that the rail displacement obtained from the MBS simulation was much higher than that in a normal track (less than 1 mm [27,31]), which indicates the vulnerability of the ballast in the railway crossings.

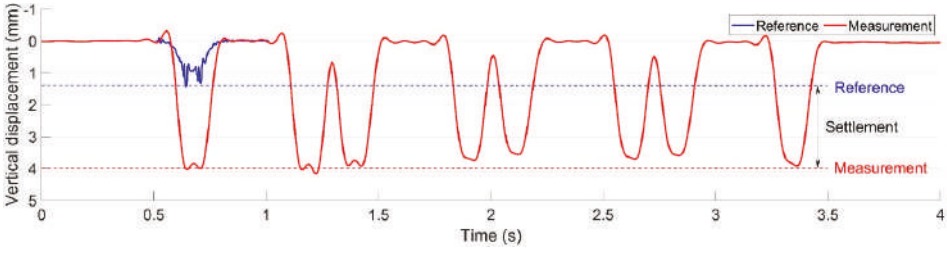


Figure 13. Ballast settlement in the monitored crossing.

In a previous study [27], it was found that track irregularities such as rail joints and turnout crossings can lead to the fast deterioration of the ballast, and the ballast settlement will in turn accelerate the degradation procedure of other related track components. In this study, the 2.6 mm ballast settlement was already higher than those in the previously monitored welded joints (≈1.5 mm) and movable crossings (≈2 mm), which revealed the seriously deteriorated ballast condition.

It can be concluded that the monitored crossing was suffering from rapidly occurring, extremely high wheel-rail impacts and severe ballast settlement. For a recently renovated crossing, such performance is quite abnormal.

4. Effectiveness Analysis of the Maintenance Actions

The constantly occurring extremely high wheel-rail impacts as well as serious ballast settlement clearly indicate the degraded condition of the crossing. In order to improve such a situation, various maintenance actions were implemented in this location including ballast tamping, fastening system renovation, etc. In this section, the effectiveness of the maintenance actions are briefly discussed, as presented below.

4.1. Ballast Tamping

Considering that the crossing rail was lately renovated with limited wear or plastic deformation, the severe ballast settlement was suspected to be the main cause for the high wheel-rail impacts.

Therefore, ballast tamping actions were frequently performed in this location by the local contractor. However, due to the lack of maintenance facilities, the tamping actions were mainly performed using the squeezing machine (Figure 14a) without track geometry correction. It can be imagined that the settled ballast cannot be fully recovered with such tamping action. As shown in Figure 14b, after tamping, the rail displacement was not dramatically reduced.

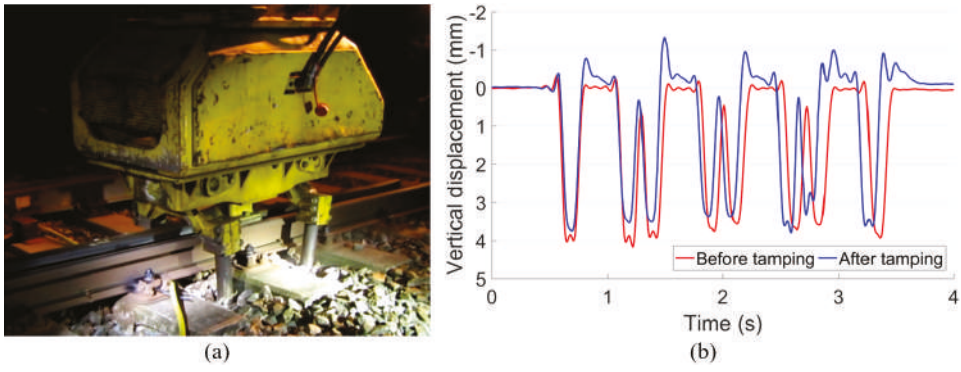


Figure 14. (**a**) Squeezing machine used for ballast tamping in the monitored crossing. (**b**) Measured rail displacement before and after ballast tamping.

The development of the wheel-rail impacts before and after tamping are presented in Figure 15. In this figure, each point represents the mean value of the impact accelerations based on multiple wheel passages in one monitoring day. It was discussed in a previous study [28] that the fluctuation of the wheel impacts was highly affected by external disturbances such as the weather. Still, it can be seen that the regression values before and after tamping were both around 100 g.

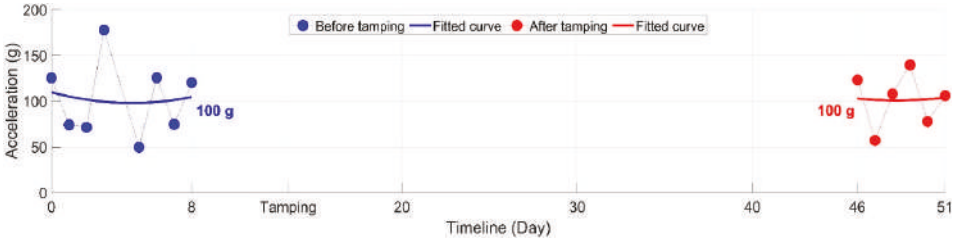


Figure 15. Development of the wheel-rail impacts before and after ballast tamping.

It can be concluded that such frequently implemented ballast tamping had no improvement in either the ballast condition or the dynamic performance of the monitored crossing. Without figuring out the root causes for the fast crossing degradation, such ineffective ballast tamping should be suspended.

4.2. Fastening System Renovation

During the monitoring period, the fastening system was found to be degraded with some broken bolts. Such degradation affected the lateral stability of the track. Therefore, the fastening system, mainly the bolts in the guard rails and the clips, was renovated, as shown in Figure 16.

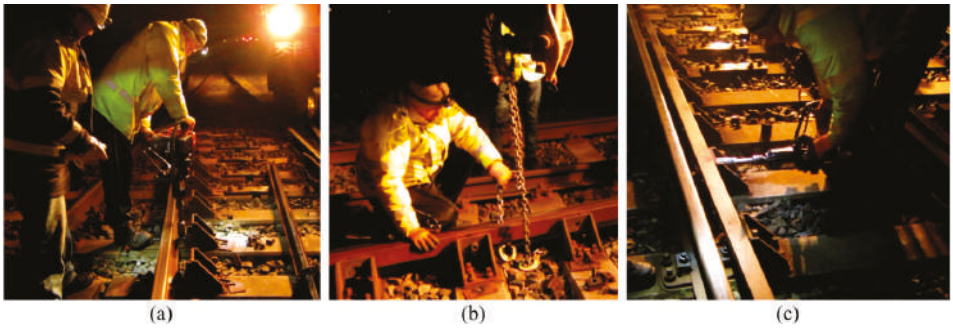


Figure 16. Fastening system renovation. (**a**) Remove the broken bolts. (**b**) Reposition the guard rail. (**c**) Install new bolts.

The development of the wheel-rail impacts before and after renovation is shown in Figure 17. The upper figure is the development of the mean value, and the lower figure gives the ratio of different impact levels in each monitoring day, corresponding to the value in the upper figure.

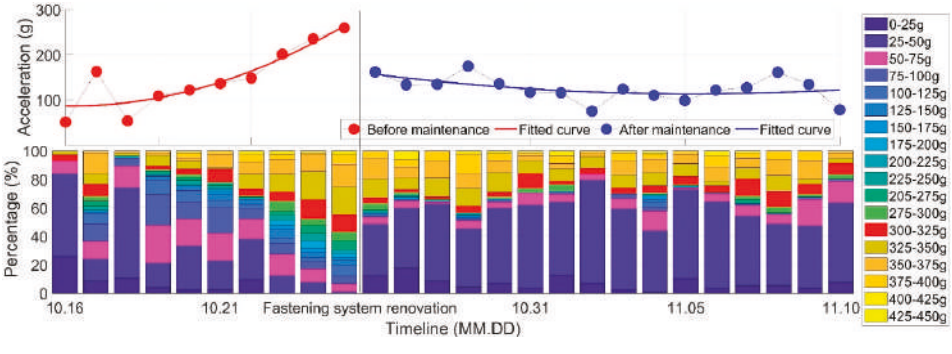


Figure 17. Effect of fastening system renovation on the dynamic performance of the crossing.

It can be seen from Figure 17 that before the renovation, the wheel-rail impact showed a clear increasing trend with the impact values widely distributed from 0 to 450 g. Such a degradation trend indicates that maintenance is urgently required due to the defects of the fastening system. After the renovation, the wheel-rail impacts were dramatically reduced in terms of the mean value and separated into two distribution modes, which is similar to those shown in Figure 11a. Such improvement is due to the enhancement in the track integrity. However, the wheel-rail impacts above 300 g were still a large proportion after maintenance, which means that the sources for such high wheel-rail impacts were not found.

In practice, ballast tamping is currently one of the few options for contractors to maintain the track. However, the unimproved crossing performance clearly indicates the ineffectiveness of tamping. The fastening system renovation was a forced action to repair damaged components. Although the crossing performance was improved, the extremely high wheel-rail impacts were not reduced, thus the sources for the fast crossing degradation were not eliminated. To figure out the root causes for the crossing damage, the track inspection was extended to the bridge in front of the crossing (Figure 10b). The results for the track inspection, as well as the numerical verification using the MBS model, are presented in the next section.

5. Damage Sources Investigation

In this section, the track inspection, including the whole turnout and the adjacent bridge, is presented. The inspected degradations will be input into the MBS model to verify the influence on the crossing performance. As a reference, the dynamic responses in the designed condition with no track degradations were also simulated and compared with those in degraded conditions. The verification results, followed by the analysis, are also presented.

5.1. Track Inspection

In the field inspection, it was found that the bridge was not well aligned in the track, but deviated around 15 cm, as shown in Figure 18a. Such deviation introduced a curve into the track, which was likely to be out of design since no elevation was set up in the outer rail. It can be imagined that the passing trains could not pass the track along the central line but tended to have wheel flange contact with the outer rail, which eventually leads to the severe wear in the switch blade (Figure 18b).

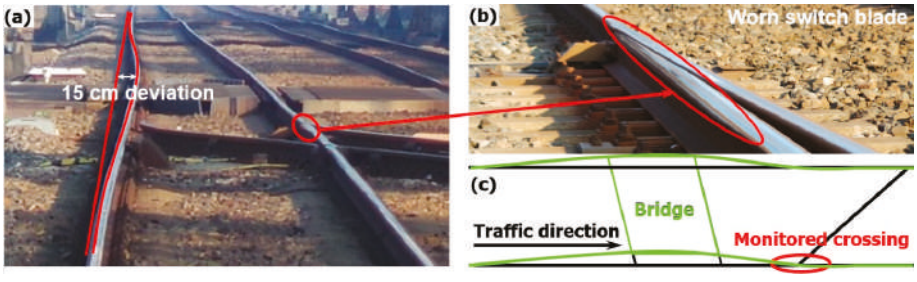
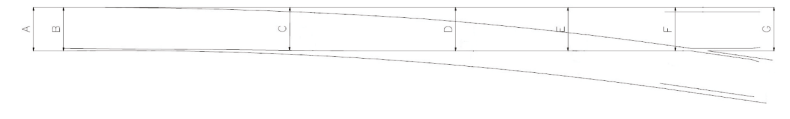


Figure 18. Track deviation in front of the crossing. (**a**) Inspected curve introduced by the bridge. (**b**) Worn switch rail. (**c**) Demonstration of the bridge deviation.

The accumulated effect of the track deviation was also reflected in the variated track gauge. It was shown in the measurement results that the gauge variations along the whole turnout were up to 3 mm, as presented in Table 1. Considering that the monitored crossing is located quite close to the bridge (Figure 18c), such track misalignment, including the track deviation in the bridge and track gauge variation along the turnout, may affect the wheel-rail interaction in the crossing.

Table 1. Track gauge measurement results in the critical sections along the turnout.



Location	A	B	C	D	E	F	G
Deviation (mm)	+2	+3	−2	−2	+2	+3	0

5.2. Numerical Verification and Analysis

In order to verify the effect of the track lateral misalignment on the performance of the crossing, both the bridge-introduced curve and the track gauge variation were input into the MBS vehicle-crossing model (Figure 9). The equivalent track lateral irregularities as the model input are shown in Figure 19.

In the MBS model, the crossing type is the same as the monitored 1:9 crossing with the rail type of UIC54 E1. The vehicle model is consistent with the recorded VIRM train with the wheel profile of S1002. The initial track parameters of Dutch railways [32] applied in the model are given in Table 2.

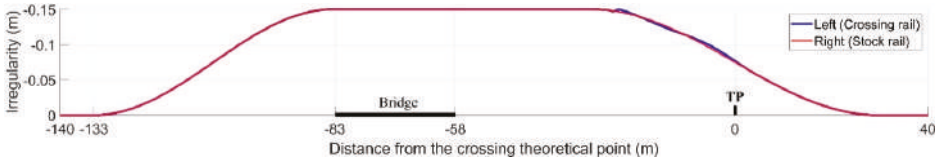


Figure 19. Equivalent lateral irregularities in the track.

Table 2. Track parameters.

Track Components		Stiffness, MN/m	Damping, kN·s/m
Rail pad/Clips	Vertical	1300	45
	Lateral	280	580
	Roll	360	390
Ballast	Vertical & lateral	45	32

With the track misalignment taken into account, the crossing condition was considered as degraded. The simulation results of both wheels in the bogie, including the wheel impact accelerations and transition regions, were compared with the results in the designed condition [29], as shown in Figure 20.

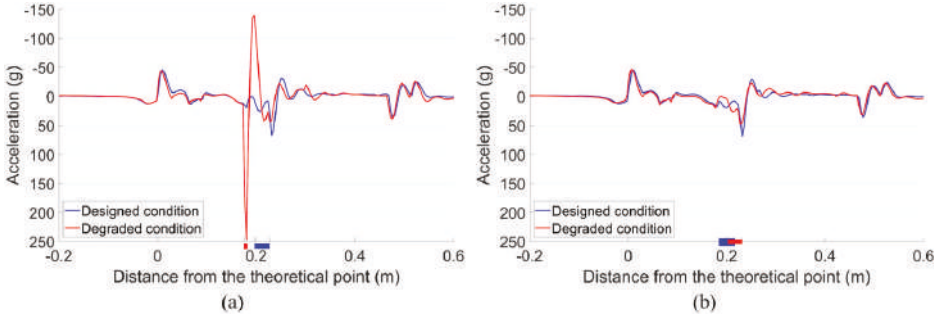


Figure 20. Vertical impact acceleration responses and transition regions. (**a**) Front wheel. (**b**) Rear wheel.

It can be seen from Figure 20a that with the lateral irregularity taken into account, the impact of the front wheel was dramatically increased to 247 g, which was 4 times higher than the reference value (around 62 g) in the designed condition. While for the rear wheel from the same bogie, the impact was 48 g, which was even lower than the reference value. Despite the slight difference in the absolute values, the simulation results were consistent with the measurement results (Figure 11). Meanwhile, the transition region of the front wheel was 0.176–0.182 m from the TP with a size of only 0.006 m. Compared with the reference level (0.196–0.217 m with a size of 0.031 m, [29]), it was much narrower and closer to the TP, indicating earlier wheel impact and much sharper wheel load transition in the crossing. For the rear wheel, although the transition region was located farther from the TP, the size was almost the same as the reference value.

Such results clearly show that the curve and lateral track misalignment in front of the crossing can lead to unstable wheel-rail contact in the crossing and sometimes result in extremely high impacts. Additionally, the front and rear wheels pass through the crossing quite differently, which indicates that the performance of the rear wheel is not independent, but is affected by the front wheel.

For the wheel-rail contact forces, the tendency was similar to the acceleration responses, as shown in Figure 21. With the degraded track condition, the maximum contact force of the front wheel in the degraded condition was 468 kN, which was twice as high as that in the designed condition (235 kN). While for the rear wheel, the difference between the degraded condition and the designed condition was limited.

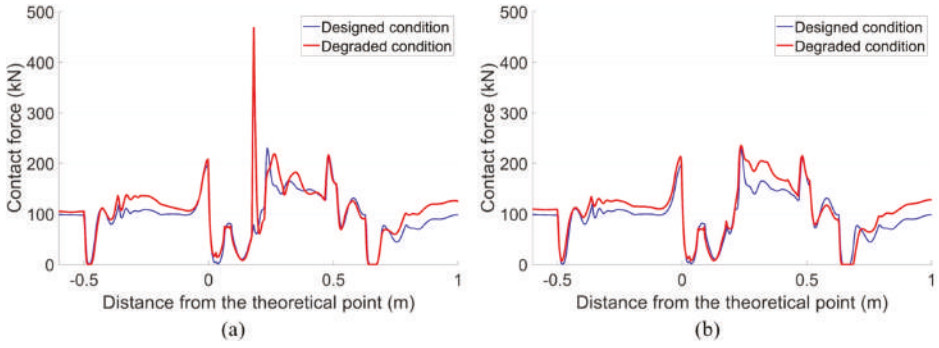


Figure 21. Vertical wheel-rail contact responses of the facing crossing. (**a**) Front wheel. (**b**) Rear wheel.

To understand how the track misalignment affects the wheel-rail interaction in the crossing, the relationship between the wheel lateral displacements and wheel-rail contact forces were analyzed. Before that, the wheel lateral displacement in the designed condition is presented in Figure 22. When the train enters the crossing panel, the variated rail geometry will lead to the lateral movement of the wheel. The maximum lateral displacement was around 0.7 mm.

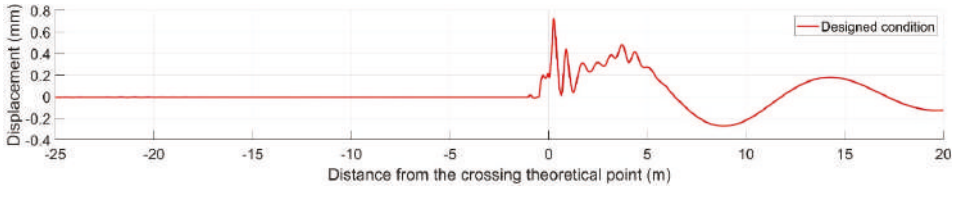


Figure 22. Wheel lateral displacement in the designed condition.

In the degraded condition with track lateral irregularities, the lateral displacements of the wheels were dramatically changed, as shown in Figure 23. It can be seen that both the front wheel and the rear wheel showed activated hunting oscillation before and after passing through the crossing, but the trajectories were quite different. For the front wheel, the lateral movement was more intense and ran toward the crossing nose rail near the TP. The maximum lateral displacement corresponding to the position with the highest contact force was 2.3 mm, which means that compared with that in the designed condition, the wheel flange was around 1.6 mm closer to the nose rail. Comparatively speaking, such displacement of the rear wheel was only 0.3 mm. Such results indicate that the wheel-rail impact was profoundly affected by the movement of the wheel. When the wheel approaches closer to the crossing nose, the wheel-rail impact is likely to be increased. It can be concluded that the train hunting activated by the lateral track misalignment in front of the crossing is the main cause of the extremely high wheel-rail impacts.

The train hunting effect also explains the unstable wheel-rail impacts. For the rear wheel, the lateral movement was affected not only by the track misalignment but also by the front wheel from the same bogie. As a result, these two wheels led to quite different wheel trajectories. It can be imagined that in the real-life situation, there are much more factors that may affect the hunting motion of each passing wheelset such as the initial position of the wheel when entering the misaligned track section, the mutual interaction between the adjacent wheelsets, the lateral resistance of the track, and even the weather condition [28], etc. The combined effect of all these factors ultimately resulted in the polarized distribution of the impact acceleration responses (Figure 11a).

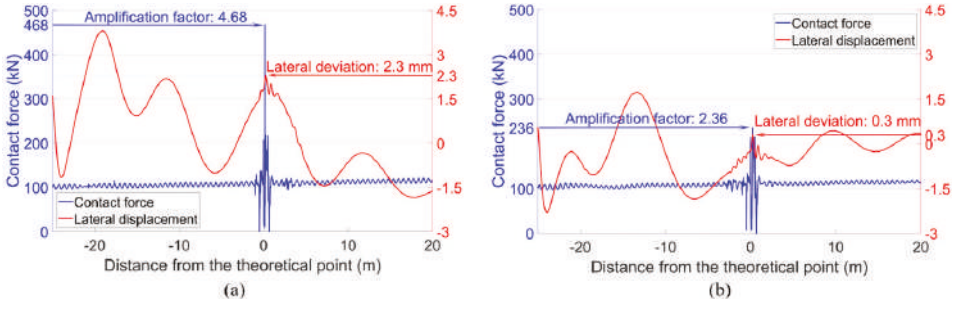


Figure 23. Wheel-rail contact forces and wheel lateral displacements. (**a**) Front wheel. (**b**) Rear wheel.

5.3. Respective Effect of Lateral Curve or Track Gauge Deviation

It can be noticed that in the previous analysis, the input track misalignment consisted of two parts: the lateral curve introduced by the bridge and the track gauge deviation. In order to understand the effect of each part in the wheel-rail interaction, these two parts were further analyzed, and the results are presented below.

Considering the bridge-introduced lateral curve, the wheel-rail contact forces and the lateral wheel displacements were calculated, as presented in Figure 24. It can be seen that in the front wheel, the bridge-introduced curve mainly resulted in the lateral shift of the wheel trajectory due to the centripetal force. Such a shift was only 0.5 mm near the crossing nose when compared with the designed condition, and the effect on the wheel impact was limited. For the rear wheel, the combined effect of the curve and the motion of the front wheel resulted in the lateral deviation of 0.9 mm, which was quite close to that in the designed condition and had no significant influence on the wheel-rail impact.

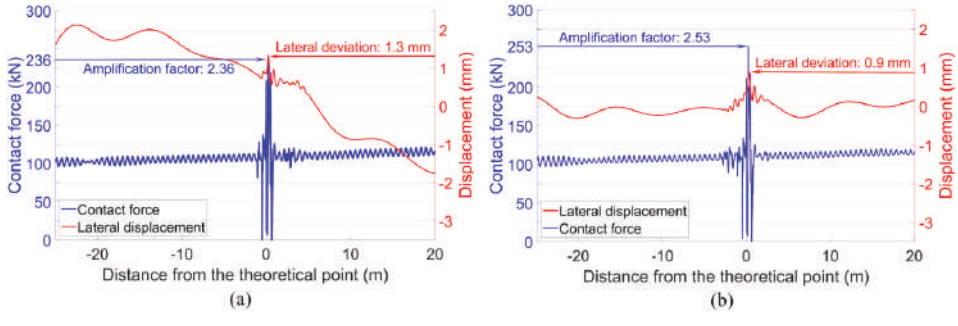


Figure 24. Wheel-rail contact forces and lateral wheel displacements. (**a**) Front wheel. (**b**) Rear wheel.

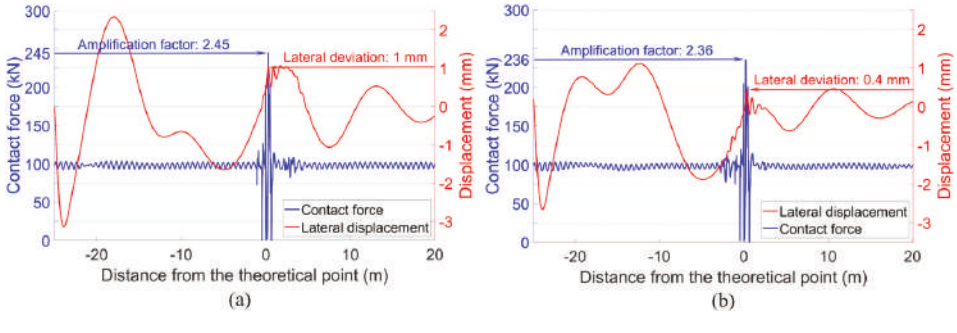


Figure 25. Wheel-rail contact forces and wheel lateral displacements. (**a**) Front wheel. (**b**) Rear wheel.

The effect of the track gauge deviation on the wheel-rail interaction is demonstrated in Figure 25. Different from the effect of the bridge-introduced curve, the deviated track gauge activated the hunting motion of the passing wheels. Still, the resulted lateral wheel displacements were not large enough to amplify the wheel-rail impact. The maximum displacements corresponding to the wheel impacts were 1 mm in the front wheel and 0.4 mm in the rear wheel, respectively.

5.4. Summary

Based on the above analysis, it can be concluded that the extremely high wheel-rail impacts in the monitored crossing were caused by the hunting oscillation of the passing trains. Such train hunting was the combined effect of the bridge-introduced curve in front of the crossing and the deviated track gauge along the turnout. When the maximum wheel lateral displacement reaches a certain level (e.g., 2.3 mm), the wheel-rail impact will be dramatically amplified.

It has to be noted that although the curve in front of the crossing did not directly activate train hunting, the activated lateral shift of the passing wheels resulted in the wear in the switch blade (Figure 18b) and contributed to the track gauge deviation. Therefore, such a curve can be considered as the root cause of the fast degradation of the monitored crossing. To improve the performance of the crossing, this curve has to be first eliminated.

In the previous study [28], it was proven that high rail temperature due to the long duration of sunshine would amplify the existing track geometry deviation in turnout and lead to the increase in the wheel-rail impacts. The train hunting activated by the track gauge deviation in this study further confirmed these results.

6. Effect of Maintenance-Related Degradation

According to the measurement results, the monitored crossing also suffered from ballast settlement and broken clips. In order to better simulate the real-life situation, these track defects were respectively added to the degraded MBS model developed in Section 5.2. The combined effects were simulated and analyzed, as presented below.

6.1. Effect of Ballast Settlement

It is shown in Figure 13 that the detected ballast settlement was around 2.6 mm. To simplify the problem, a vertical irregularity was introduced in the MBS model to simulate the ballast settlement, as shown in Figure 26. In this irregularity function, the amplitude was 1.3 mm, and the wavelength was 10 m. The trough of the wave was located 0.3 m from the TP of the crossing, which was consistent with the instrumented accelerometer and the installed displacement target.

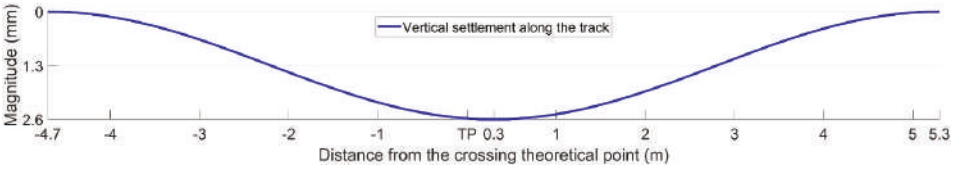


Figure 26. Ballast settlement introduced in the MBS model.

With the ballast settlement taken into account in the MBS model, the dynamic performance of the crossing was simulated. The representative results are shown in Figure 27. It can be seen that the simulation results were almost the same as those without ballast settlement (Figure 23), despite the slightly increased impact force of the front wheel (from 468 kN to 487 kN). It can be concluded that the existence of ballast settlement had a limited influence on the dynamic performance of the crossing. From another point of view, the ballast settlement was more likely to be the accumulated effect of the

high wheel-rail impacts. Such results further explain the ineffectiveness of the frequently performed ballast tamping since ballast settlement is not the main cause of the extremely high wheel-rail impacts.

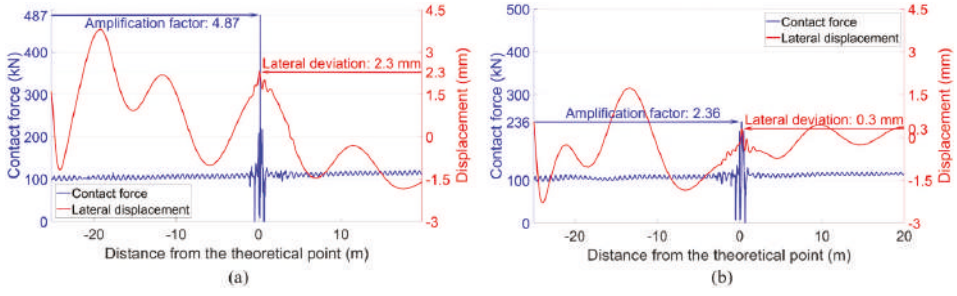


Figure 27. Wheel-rail contact forces and lateral wheel displacements. (**a**) Front wheel. (**b**) Rear wheel.

6.2. Influence of Reduced Lateral Support

It is shown in Figure 16 that the defects of the fastening system can increase the instability of the wheel-rail impact in the crossing. Combined with the maintenance action and the simulation results in Section 5, it can be inferred that this effect was caused by the reduced lateral track resistance. To verify this inference in the degraded model (Section 5.2), the input lateral stiffness of the clips in the crossing panel was reduced from 280 MN/m (Table 2) to 2.8 N/m, and the corresponded damping was reduced from 580 kN·s/m to 5.8 N·s/m. Based on these inputs, the wheel-rail contact forces and the lateral wheel displacements were calculated, as presented in Figure 28.

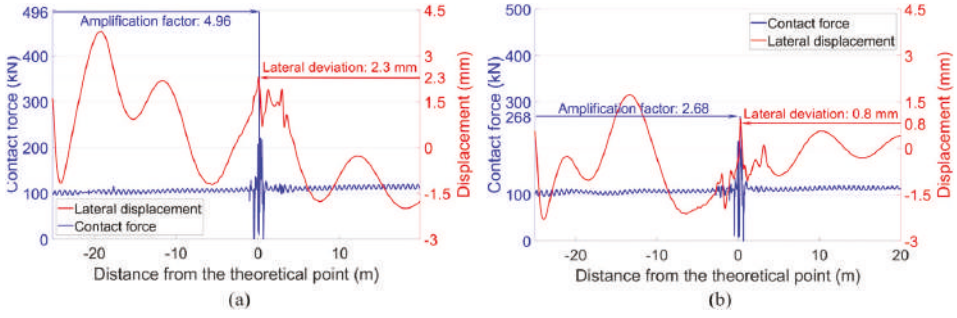


Figure 28. Wheel-rail contact forces and lateral wheel displacements. (**a**) Front wheel. (**b**) Rear wheel. Note: Ballast settlement was not taken into account.

It can be seen from Figure 28 that with the reduced lateral stiffness and damping of the clips, the impacts of both the front wheel and the rear wheel were slightly increased (compared with the results in Figure 23). Moreover, the hunting motion of wheels in the crossing panel was more intense. As a result, the lateral deviation of the rear wheel increased from 0.3 mm to 0.8 mm. It can be imagined that with the impacts of the passing trains, the track alignment will continuously be changing due to the reduced structural integrity. The changed track alignment will, in return, act on the wheel-rail interaction and eventually lead to more unstable wheel impacts in the crossing (Figure 17). From this point of view, renovating the defected fastening system is necessary for a monitored crossing. Enough track lateral resistance can help to maintain better crossing performance.

7. Conclusions

In this study, the root cause of the fast degradation of a 1:9 crossing in the Dutch railway system was investigated. The effectiveness of some typical track maintenance actions was also assessed and verified. Based on the measurement and simulation results, the following conclusions can be drawn:

- The fast crossing degradation was directly caused by the extremely high wheel-rail impacts, and the root cause for such high impacts was the hunting of the passing trains that were activated by the track lateral misalignment in front of the crossing. When the lateral deviation of the passing wheel exceeds a certain extent (e.g., 2.3 mm), the wheel-rail contact situation will change and the wheel impacts will be dramatically increased. To improve the current situation, such track misalignment needs to be eliminated;
- Ballast settlement is likely to be the accumulated effect of the high wheel-rail impacts. The influence on the crossing performance is somewhat limited. Ballast tamping, especially with only the squeezing machine, cannot improve the dynamic performance of the crossing. In the case of not knowing the sources of damage, it is better to take no action, rather than implement ballast tamping;
- Fastening system renovation helped improved the crossing performance by providing better lateral support in the track but was not targeted to the fundamental problem. Therefore, such damage repair action is useful, but not enough for an improvement in the crossing performance.

This study further verified the effectiveness of the previously proposed condition indicators in the investigation of the damage sources of the crossing. Since the root causes for the fast degradation were the deviated track in front of the crossing, this means that the degradation detection is not only restricted to the crossing itself but can also take the adjacent structures into account.

The activated train hunting reasonably explained the instability of wheel-rail interaction in the crossing, which pointed out a possible direction to maintain the problematic crossings in the Dutch railway network. As part of the Structural Health Monitoring System for railway crossings developed in TU Delft, the findings in this study will help improve the current maintenance philosophy from "failure reactive" to "failure proactive", and eventually lead to sustainable railway crossings.

Author Contributions: This article is written by X.L. and supervised by V.L.M. All authors have read and agreed to the published version of the manuscript.

Funding: The field measurements in this research were funded by ProRail and performed within the framework of the joint ProRail and TU Delft project "Long-Term Monitoring of Railway Turnouts".

Acknowledgments: The authors would like to thank the support from I.Y. Shevtsov from ProRail and technical support from G.H. Lambert and other colleagues from ID^2. Furthermore, the fruitful cooperation with Strukton Rail is highly appreciated.

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

References

1. *ProRail, Ontwerpvoorschrift (Baan en Bovenbouw) Deel 6.1: Wissels en Kruisingen Alignement, Tussenafstanden, Overgangsvrije Zones en Bedieningsapparatuur, V006*; ProRail: Utrecht, The Netherlands, 2015; p. 12.
2. Shevtsov, I. Rolling Contact Fatigue Problems at Railway Turnouts. Available online: http://www.sim-flanders.be/sites/default/files/events/Meeting_Materials_Nov2013/mm_13112013_ivan_shevtsov_prorail.pdf (accessed on 13 November 2013).
3. Cornish, A.T. Life-Time Monitoring of in Service Switches and Crossings through Field Experimentation. Ph.D. Thesis, Imperial College London, London, UK, May 2014.
4. Molina, L.F.; Resendiz, E.; Edwards, J.R.; Hart, J.M.; Barkan, C.P.L.; Ahuja, N. Condition Monitoring of Railway Turnouts and Other Track Components Using Machine Vision. In Proceedings of the Transportation Research Board 90th Annual Meeting, Washington, DC, USA, 23–27 January 2011.
5. Chen, C.; Xu, T.; Wang, G.; Li, B. Railway turnout system RUL prediction based on feature fusion and genetic programming. *Measurement* **2020**, *151*, 107162. [CrossRef]

6. Wang, P.; Wang, L.; Chen, R.; Xu, J.; Gao, M. Overview and outlook on railway track stiffness measurement. *J. Mod. Transp.* **2016**, *24*, 89–102. [CrossRef]
7. Yeo, G.J. Monitoring Railway Track Condition Using Inertial Sensors on an in-Service Vehicle. Ph.D. Thesis, University of Birmingham, Birmingham, UK, June 2017.
8. De Rosa, A.; Alfi, S.; Bruni, S. Estimation of lateral and cross alignment in a railway track based on vehicle dynamics measurements. *Mech. Syst. Signal Process.* **2019**, *116*, 606–623. [CrossRef]
9. Balouchi, F.; Bevan, A.; Formston, R. Detecting Railway Under-Track Voids Using Multi-Train in-Service Vehicle Accelerometer. In Proceedings of the 7th IET Conference on Railway Condition Monitoring, Birmingham, UK, 27–28 September 2016.
10. Tsunashima, H. Condition Monitoring of Railway Tracks from Car-Body Vibration Using a Machine Learning Technique. *Appl. Sci.* **2019**, *9*, 2734. [CrossRef]
11. Kassa, E.; Nielsen, J.C.O. Stochastic analysis of dynamic interaction between train and railway turnout. *Veh. Syst. Dyn.* **2008**, *46*, 429–449. [CrossRef]
12. Alfi, S.; Bruni, S. Mathematical modelling of train-turnout interaction. *Veh. Syst. Dyn.* **2009**, *47*, 551–574. [CrossRef]
13. Ren, Z.; Sun, S.; Xie, G. A method to determine the two-point contact zone and transfer of wheel rail forces in a turnout. *Veh. Syst. Dyn.* **2010**, *48*, 1115–1133. [CrossRef]
14. Pletz, M.; Daves, W.; Ossberger, H. A wheel set/crossing model regarding impact, sliding and deformation explicit finite element approach. *Wear* **2012**, *294*, 446–456. [CrossRef]
15. Anyakwo, A.; Pislaru, C.; Ball, A. A New Method for Modelling and Simulation of the Dynamic Behaviour of the Wheel-rail contact. *Int. J. Autom. Comput.* **2012**, *9*, 237–247. [CrossRef]
16. Pålsson, B.A. Optimisation of Railway Switches and Crossings. Ph.D. Thesis, Chalmers University of Technology, Göteborg, Sweden, February 2014.
17. Wei, Z.; Chen, S.; Li, Z.; Dollevoet, R. Wheel-Rail Impact at Crossings-Relating Dynamic Frictional Contact to Degradation. *J. Comput. Nonlinear Dyn.* **2017**, *12*, 1–11. [CrossRef]
18. Xin, L. Long-Term Behaviour of Railway Crossings Wheel-Rail Interaction and Rail Fatigue Life Prediction. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, June 2017.
19. Chiou, S.; Yen, J. Modeling of railway turnout geometry in the frog area with the vehicle wheel trajectory. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2018**, *232*, 1598–1614. [CrossRef]
20. Skrypnyk, R.; Nielsen, J.C.O.; Ekh, M.; Pålsson, B.A. Metamodelling of wheel-rail normal contact in railway crossings with elasto-plastic material behavior. *Eng. Comput.* **2019**, *35*, 139–155. [CrossRef]
21. Torstensson, P.T.; Squicciarini, G.; Krüger, M.; Pålsson, B.A.; Nielsen, J.C.O.; Thompson, D.J. Wheel-rail impact loads and noise generated at railway crossings influence of vehicle speed and crossing dip angle. *J. Sound Vib.* **2019**, *456*, 119–136. [CrossRef]
22. Wiest, M.; Daves, W.; Fischer, F.D.; Ossberger, H. Deformation and damage of a crossing nose due to wheel passages. *Wear* **2008**, *265*, 1431–1438. [CrossRef]
23. Johansson, A.; Pålsson, B.; Ekh, M.; Nielsen, J.; Ander, M.; Brouzoulis, J.; Kassa, E. Simulation of wheel-rail contact and damage in switches & crossings. *Wear* **2010**, *271*, 472–481.
24. Nielsen, J.C.O.; Li, X. Railway track geometry degradation due to differential settlement of ballast/subgrade Numerical prediction by an iterative procedure. *J. Sound Vib.* **2018**, *412*, 441–456. [CrossRef]
25. Skrypnyk, R.; Ekh, M.; Nielsen, J.C.O.; Pålsson, B.A. Prediction of plastic deformation and wear in railway crossings—Comparing the performance of two rail steel grades. *Wear* **2019**, *428–429*, 302–314. [CrossRef]
26. Wan, C. Optimisation of Vehicle-Track Interaction at Railway Crossings. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, September 2016.
27. Liu, X.; Markine, V.L.; Wang, H.; Shevtsov, I.Y. Experimental tools for railway crossing condition monitoring. *Measurement* **2018**, *129*, 424–435. [CrossRef]
28. Liu, X.; Markine, V.L. Correlation Analysis and Application in the Railway Crossing Condition Monitoring. *Sensors* **2019**, *19*, 4175. [CrossRef]
29. Liu, X.; Markine, V.L. Validation and Verification of the MBS Models for the Dynamic Performance Study of Railway Crossings. *Eng. Struct.* **2019**, submitted.
30. Wang, H.; Markine, V.L.; Liu, X. Experimental analysis of railway track settlement in transition zones. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2018**, *232*, 1774–1789. [CrossRef] [PubMed]

31. Xu, L.; Yu, Z.; Shi, C. A matrix coupled model for vehicle-slab track-subgrade interactions at 3-D space. *Soil Dyn. Earthq. Eng.* **2020**, *128*, 105894. [CrossRef]
32. Hiensch, M.; Nielsen, J.C.O.; Verheijen, E. Rail corrugation in The Netherlands measurements and simulations. *Wear* **2002**, *253*, 140–149. [CrossRef]