data

MDPI

Data Descriptor

Taguchi Orthogonal Array Dataset for the Effect of Water Chemistry on Aggregation of ZnO Nanoparticles

Rizwan Khan, Muhammad Ali Inam, Du Ri Park, Saba Zam Zam and Ick Tae Yeom *

Graduate School of Water Resources, Sungkyunkwan University (SKKU) 2066, Suwon 16419, Korea; rizwankhan@skku.edu (R.K.); aliinam@skku.edu (M.A.I.); enfl8709@skku.edu (D.R.P.); sabazamzam@skku.edu (S.Z.Z.)

* Correspondence: yeom@skku.edu; Tel.: +82-31-299-6699

Received: 29 May 2018; Accepted: 13 June 2018; Published: 15 June 2018

Abstract: The dynamic nature of engineered nanoparticle (ENP) aggregation behavior and kinetics are of paramount importance in the field of toxicological and environmental nanotechnology. The Taguchi orthogonal array (OA) $L_{27}(3^{13})$ matrix based on a fractional factorials design was applied to systematically evaluate the contribution and significance of water chemistry parameters (pH, temperature, electrolyte, natural organic matter (NOM), content and type) and their interactions in the aggregation behavior of zinc oxide nanoparticles (ZnO NPs). The NPs were dispersed into the solution using a probe-sonicator cell crusher (Bio-safer, 1200-90, Nanjing, China). The data were obtained from UV–Vis spectroscopy (Optizen 2120 UV, Mecasys, Daejeon, Korea), Fourier Transform Infrared Spectrometery (FT-IR 4700, spectroscopy, a JASCO Analytical Instruments, Easton, Pennsylvania, USA) and particle electrophoresis (NanoZS, Zetasizer, Malvern Instruments Ltd., Worcestershire, UK). The dataset revealed that Taguchi OA matrix is an efficient approach to study the main and interactive effects of environmental parameters on the aggregation of ZnO NPs. In addition, the aggregation profile of ZnO NPs was significantly influenced by divalent cations and NOM. The result of the FT–IR data presents a possible mechanism of ZnO NP stabilization in the presence of different NOM. This data may be helpful to predict the aggregation behavior of ZnO NPs in environmental and ecotoxicological contexts.

Dataset: Available as the supplementary file.

Keywords: aggregation; Fourier transform infrared spectrometer; Taguchi orthogonal array; UV-Vis spectrophotometer; water chemistry; zinc oxide nanoparticles

1. Summary

With the development of nanotechnology, production, and utilization of engineered nanoparticles (ENPs) has increased in consumer products and commercial applications. Among the diverse ENPs, zinc oxide nanoparticles (ZnO NPs) are the third most frequently used metal-containing nanomaterials due to their unique structural properties [1]. The estimated annual global production is currently between 570–33,400 t/year, and is anticipated to reach 58,000 t/year by 2020 [2]. However, such broad applications and production of ZnO NPs have generated concern to both the scientific and public communities due to their environmental release, which may increase the bioavailability and NPs' toxicity in an aqueous environment [3]. The potential adverse effects of ZnO NPs on phytoplankton, plants, and even human cell lines are well-known [4]. However, limited knowledge is available

regarding the nanoparticle–liquid interaction and aggregation among NPs. Thus, understanding the aggregation phenomena of ZnO NPs under various parameters of water chemistry is important to properly assess their risk.

Several researchers have studied the ZnO NP fate and mobility in an aqueous environment [5–7]. ZnO NP aggregation phenomena are highly influenced by physicochemical properties and solution chemistry [8]. NP properties such as size, shape, and coating may cause substantial variation in their aggregation kinetics [9]. Few studies [10,11] have reported that large size particles tend to aggregate, however, tiny size particles remain suspended in solution and increase the risk of harm to aquatic species. The variation in solution pH and temperature affect the surface potential and solubility of ZnO NPs which may also influence their agglomeration phenomena [7,12]. Fast aggregation has been reported in high-salt content solutions due to compression of the electrical double layer (EDL) [9,13]. The studies also showed that ubiquitous natural organic matter (NOM) stabilizers kept ZnO NPs suspended in water due to steric hindrance and potential charge reversal [14,15]. Furthermore, few studies have demonstrated that the presence of divalent cations i.e., Ca^{2+}, Mg^{2+} may form intermolecular bridging between NOM and ZnO NPs, to promote aggregation [16]. Additionally, the presence of trivalent electrolytes such as aluminum sulfate (Al2(SO4)3) with the strong potential to form a bridge between metal NPs and aggregates, may enhance the hetero-aggregation due to the formation of aluminum complexes [17,18]. These results further suggest that different electrolytes play a significant role in the aggregation of ENPs in aqueous matrices.

In previous studies [19–22], the influence of various environmental parameters on agglomeration phenomena of ZnO NPs has generally been studied using individual factors and has been limited to pure water. Additionally, the interaction of multiple parameters affecting this phenomenon has been rarely investigated by environmental scholars. Therefore, it is important to evaluate the aggregation phenomena of ZnO NPs together with a combination of different parameters that may affect their function and toxicity in an ecosystem.

A recent study has shown a significant influence of several environmental parameters on sedimentation and aggregation of ZnO NPs [23]. Numerous datasets on such phenomena of other ENPs i.e., copper oxide (CuO) and titanium dioxide (TiO_2) have been reported [22,24]. However, to the knowledge of the authors, there are no datasets available regarding the effect of combined multiple environmental parameters on ZnO NP aggregation. In light of this, ZnO NP aggregation phenomena in single and multiple environmental parameters were investigated. The fractional factorial method based on Taguchi orthogonal array (OA) L_{27} (3^{13}) design matrix was used to measure the main and interactive effects of multiple factors. This study quantitatively presents a ZnO NPs aggregation dataset obtained from UV–Vis spectrophotometry, Fourier transform infrared spectrometry (FT–IR), zeta potential analysis and Optizen view 4.2.5 PC interference software.

The dataset obtained from the UV/Vis spectrophotometer presents the absorbance data of ZnO NPs at various environmental conditions such as pH, electrolyte type and concentration, and NOM type and concentration. Earlier, the zeta potential and FT–IR spectra of ZnO–NOM complexes were elucidated under different aqueous matrices [23]. This data may contribute to a better understanding of the agglomeration phenomena and fate of ZnO NPs in complex heterogeneous environments.

2. Data Description

Here, we present a dataset of ZnO NP aggregation efficiency and its kinetics in an Excel (Microsoft Corp., Redmond, WA, USA) spreadsheet file (see supplementary material). Tables S1–S12 show the ZnO NP data obtained after a bench scale aggregation experiment which are listed and described in Tables 1 and 2 at 25 °C ± 0.5. Tables S1 and S2 show the absorbance data across various wavelengths and concentrations. This data is used to determine the concentration from a quantitative standard linear curve based on the Beer–Lambert law. Table S3 shows turbidity data at various time and power levels of sonicator to obtain the optimum sonication condition for further aggregation experiments. The data in Table S4 contains the residual absorbance and zeta potential at various pH ranges. This data shows

that the solubility of ZnO NPs increases sharply at pH below 7 or above 11. Considering the dissolution of ZnO NPs at such conditions, further experiments were conducted in the pH range 7–9.5 to minimize this effect and to obtain systematic data of ZnO NP aggregation. The data in Tables S5 and S6 shows the residual absorbances and corresponding zeta potentials at various electrolytes, NOM types and concentrations after 6 h. This can be used to predict the effect of interfering ions on the aggregation phenomena of ZnO NPs in an aqueous environment. Table S7 shows the time-resolved optical absorbance measured at 370 nm across various solution matrices. The aggregation rate k can be estimated from this data using Stokes sedimentation theory; it also indicates that aggregation kinetics may vary substantially in the presence of high electrolyte and NOM concentrations. Table S8 shows the data derived from FT/IR spectra of pristine ZnO NPs before and after interaction with high concentration of various NOM i.e., humic acid, salicylic acid and citric acid. It shows the percent transmittance against the wave numbers (500–4000 cm^{-1}). This data is helpful in providing the information about modification in the surface of ZnO NPs after interaction with NOM.

Table 1. Type of experiment and condition of single parameter experiments.

Experiment	Parameter	Unit	Condition
A	pH	-	3 to 12
B	KCl, Na_2SO_4	mM	0.01 to 100
	$MgCl_2$	mM	0.01 to 25
C	HA, SA, CA	$mg{\cdot}L^{-1}$	0.5 to 100
D	Synthetic Environmental water samples		

Where HA, SA and CA corresponds to humic acid, salicylic acid and citric acid respectively.

Table 2. The input parameter and their corresponding levels for the Taguchi L_{27} (3^{13}) experiment.

Parameter	Units	Level		
		1	2	3
pH	-	7.0	8.0	9.5
Electrolyte Concentration	mM	25	50	100
NOM Concentration *	mgL^{-1}	5	10	25
Electrolyte type	-	KCl	Na_2SO_4	$MgCl_2$
NOM type	-	HA	SA	CA
Temperature	°C	15	25	35

* NOM corresponds to natural organic matter.

Furthermore, Tables S9–S11 show the data of residual absorbance and corresponding aggregation efficiencies, variance and range analysis obtained from Taguchi OA L_{27} (3^{13}) matrix for multi-parameters experiments. This data is important to determine the analysis of variance (ANOVA) and, consequently, the dominant parameters influencing the ZnO NPs aggregation in a real environment. In addition, the data in Tables S12 and S13 shows the residual absorbance and aggregation efficiencies obtained across various environmentally tested waters. This data is helpful in providing ZnO NPs aggregation phenomena in real environmental conditions.

3. Methods

3.1. Solution and Synthetic Waters Preparation

The nanopowder of ZnO NPs assay 97%; humic acid sodium salt (HA), salicylic acid (SA) and citric acid (CA) were purchased from the Sigma-Aldrich (St. Louis, MO, USA) and used without further purification. While inorganic salts i.e., potassium chloride (KCl), sodium sulfate (Na_2SO_4), magnesium chloride ($MgCl_2.6H_2O$), nitric acid (HNO_3), hydrochloric acid (HCl), sodium hydroxide (NaOH) used in this study were obtained from local suppliers. According to the vendor, NPs have particle size <50 nm with a specific surface area 15–25 m^2/g and it contain 6% Al as dopant. The 100 mg/L of ZnO NPs solution was prepared by adding 10 mg of ZnO nanopowder in 100 mL of nanopure

water. The solution was subjected to probe-sonication using an ultrasonicator (ultrasonic cell crusher, Bio-safer, 1200-90, Nanjing, China) for 5–40 min and 100–600 W power to obtain a homogenous NP dispersion. Moreover, the solution turbidity was also measured using Turbidimeter (Hach Benchtop 2100N, Loveland, CO, USA). The natural organic matter (NOM) stock solutions were prepared by dissolving 1 g powder of each NOM in 1 L of deionized (DI) water. The pH of HA was adjusted to 10 using NaOH to ensure complete dissolution of HA. The solution was stirred at 600 rpm for overnight to improve further stability [25]. All the stock solutions' pH values were adjusted to 7 prior to storage in the dark at 4 °C. The stock solutions of inorganic salts with 0.1 M ionic strength (IS) were prepared by adding 7.45 g KCl, 14.2 g Na_2SO_4 and 20.32 g $MgCl_2$.6H2O in 1 L solution respectively. Four synthetic waters were prepared in the laboratory, while tap water and wastewater were collected from Sungkyunkwan University, Suwon and metal processing industry located at Ulsan, Korea respectively. The detailed characteristics of all water samples are already presented in our previous study [23]. The IS (*Ie*) was calculated from the electrical conductively (*EC*) [26] according to the following equation.

$$Ie = 0.01442 \times EC^{1.009} \quad (1)$$

3.2. Standard Calibration Curve Analysis

Prior to the aggregation experiments, optical absorption analysis was conducted using a UV/Vis spectrophotometer (Optizen 2120 UV, Mecasys, Daejeon, Korea) using a 100 mm path length quartz cuvette. The device was equipped with high-quality Tungsten–Halogen lamp and a Deuterium lamp which allows the analysis of ZnO NPs at relatively low concentrations (100 to 1000 µg/L). The absorption spectra of the colloids were collected between 200–800 nm at 25 °C. The background was set up with the proper solvent and then, each colloid sample was examined. Since UV–Vis spectra fit well to Beer–Lamber law; thus, a standard calibration curve was collected at 370 nm and used as a quantitative criterion of ZnO NPs.

3.3. Aggregation Efficiency of Nanoparticles

The ZnO NPs concentration and aggregation efficiencies were estimated using fitted curve equation having R^2 of 0.9993 as shown in Equations (2) and (3). In current study, initial concentration of ZnO NPs in vial was maintained at 100 mg/L.

$$Y = 0.0086x - 0.0116 \quad (2)$$

where Y represents the concentration of ZnO NPs and x indicates residual absorbance [24].

$$A = \frac{Ci - Ct}{Ci} \times 100 \quad (3)$$

where A corresponds to aggregation efficiency, Ci and Ct are concentration of ZnO NPs before and after aggregation experiments.

3.4. Aggregation and Sedimentation Experiments

The experiments were conducted in standard sedimentation glass tubes (Fisher Scientific Co., Fair Lawn, NJ, USA) with the dimensions (height; 30 cm, radius; 2 cm) for 6 h. The predetermined concentration of inorganic salts and NOM were spiked into 100 mg /L of ZnO NPs suspension. The pH was adjusted using 0.1 M HCl and 0.1 M NaOH solution. The suspension was subjected to ultrasonication to fully disperse the NPs in solution. However, in the case of NOM interactions, the ZnO NPs dispersions were ultrasonicated twice for a cycle of 15 min. During the interval of two sonications, NOM was added and mixed with a magnetic stirrer for 5 min to homogenize the dispersion and then experiments were performed. After completion of aggregation experiments, 1 mL aliquots were collected from each suspension tube at $\frac{1}{4}$ height (7.5 cm from top of tube) to measure

the absorbance of residual ZnO NPs. Prior to aliquot analysis, the equipment was calibrated using a blank nanopure water solution. Following this approach, single- and multiple-parameter experiments were performed as shown in Tables 1 and 2. In the case of kinetic experiments, the instrument was specifically tuned in UV/Vis kinetic mode to optimize the wavelength at 370 nm and time setting in the selection window. Afterwards, the sample was placed in a quartz cuvette and the absorbance was continuously measured after every 3 min. The settling rate and velocity of the NPs were calculated using first-order kinetics and Stokes's equation [27] respectively.

$$\ln Ct = \ln Ci - kt \tag{4}$$

where t represents the time; C_i and Ct are ZnO NPs concentrations at initial and t time points, respectively; and k is the sedimentation rate constant in (h^{-1}). Moreover, the settling velocity V_t of NPs under controlled conditions i.e., pH 7, 100 mg/L ZnO NPs having primary particle size 50 nm and density 5.60 g/cm^3 at various temperature levels (15–35 °C) were calculated using Stoke's law equation (Table 3).

$$V_t = \frac{gd^2(\rho_\rho - \rho_m)}{18\mu} \tag{5}$$

where g represents acceleration of gravity (m/s^2), d particle diameter (m), ρ_ρ density of particle (g/cm^3), ρ_m density of medium (g/cm^3), μ viscosity of medium (g/m·s) and V_t settling velocity of particles respectively.

Table 3. Calculated Vt of ZnO NPs at various water temperature under controlled conditions.

Temperature (°C)	15	25	35
V_t ($\times 10^{-3}$ cm/h)	1.978	2.533	3.105

3.5. Sample Preparation for FTIR Measurement

The Fourier transform infrared spectra (FT/IR-4700, spectroscopy, JASCO Analytical Instruments, Easton, PA, USA) of ZnO NPs before and after interaction with various NOM were recorded. Spectroscopy has been used as a useful tool for the identification of structures and functional groups. The extracted aliquots of ZnO NOM complexes after aggregation experiment were collected and NPs were separated by centrifugation (Hettich Centrifuger Universal 320R, Tuttlingen, Germany) at 3000 RPM for 15 min. The resultant ZnO NP pellets were subjected to vacuum drying for 24 h to remove any moisture content. The high sensitive Mercury–Cadmium–Telluride (MCT) focal-plane array detector was cooled for 30 min with liquid nitrogen (LIN). The dried samples were inserted into the sample platform to characterize the attachment of functional groups in the range of wave number 500–4000 cm^{-1}. The Thermo Scientific OMNIC FTIR software was used to analyze the data and exported in the text file format for further processing.

3.6. Zeta Potential Analysis

After the aggregation experiments, an aliquot was removed from the suspension and placed into the zeta potential cell. The cell was filled up with the colloids avoiding the formation of bubbles. The electrophoretic mobility of the particles was measured using a Zetasizer Nano-ZS (Malvern Instrument Ltd., Worcestershire, UK). The temperature in the Zetasizer was set to 25 ± 0.1 °C. Three electrophoretic mobility measurements were made for each sample. The value of electrophoretic mobility was directly converted by default function of equipment. The default Smoluchowski equation is given [28] below.

$$\zeta = 4\pi\eta\left(\frac{\mathrm{v}}{\mathrm{E}}\right)/\varepsilon \tag{6}$$

where ζ represents zeta potential (mV), $(\frac{V}{E})$ electrophoretic mobility (microns/sec × volt/cm), η solvent viscosity (poise) and ε is dielectric constant.

3.7. Characterization of Water Samples

The aliquot obtained from tap and industrial wastewater were filtered through a 0.45-µm glass fiber filter; digested with 99% pure HNO3 prior to analysis [29]. The Inductively coupled plasma mass spectrometry (ICP-MS, 7700x, Agilent Technologies, Palo Alto, CA, USA) was used to analyze the concentration of common element i.e., potassium (K^+), sodium (Na^+), copper (Cu^{2+}), iron (Fe), magnesium (Mg^{2+}), arsenic (As), calcium (Ca^{2+}) and antimony (Sb). The industrial wastewater filtered samples were further diluted in 1:10 ratio to measure the concentration of common anions ($NO_3{}^-$, Cl^- and $SO_4{}^{2-}$) using ion chromatography with metrosep A Supp5 column in (861 Advanced Compact IC, Metrohm, Herisau, Switzerland).

3.8. Taguchi Orthogonal Array Experimental Design

In recent years, many design of experiments (DOE) including full factorial and fractional factorial design have been developed and widely used in many engineering fields to optimize and design experiments [30–32]. A full factorial design may become onerous to be completed with the increase of number of parameters. Therefore, the concept of Taguchi orthogonal array (OA) based on fractional factorial design was developed by Dr. Genichi Taguchi to study a large number of variables with a small number of experimental trials [33]. Using the OA approach significantly reduce the number of experimental configurations to be studied and provides the main and interactive effects. The parameters and levels are selected according to the experimental design. The term parameters correspond to the number of experimental variables while the term levels show the number of different values a variable can assume according to its discretization.

In the current study, the variation of all input parameters was systematically studied at three levels as shown in Table 2. We applied the standard OA $L_{27}(3^{13})$ (Figure 1a) of Taguchi method using various parameters of water chemistry to investigate ZnO NP aggregation behavior. The standard OA was modified (Figure 1b) using a line separation method to assign the parameters and interactions to various columns. The ZnO NPs aggregation phenomena were evaluated under the control of six parameters i.e., pH (A), electrolyte concentration (B), NOM concentration (C), electrolyte type (D), NOM type (E) and temperature (F) with three levels and two potential interactions i.e., A × B and B × C. The OA $L_{27}(3^{13})$ which has 27 rows corresponds to a number of trials with 13 columns at 3 levels as shown in Table S9. The plan of the experiment was as follows. The pH (A) was assigned in the first column, electrolyte concentration (B) in the second column, NOM concentration (C) in the fifth column, electrolyte type (D) in ninth column, NOM type (E) in tenth column and temperature (F) in the twelfth column. The possible two parameter interactions $(A \times B)_1$ and $(A \times B)_2$ were assigned to third and fourth columns while eighth and eleventh were assigned to $(B \times C)_1$ and $(B \times C)_2$ interactions. The remaining sixth, seventh and thirteenth columns were left empty in order to calibrate the standard matrix. The experiments were conducted for each combination of parameters as per selected OA in triplicate and average values were reported.

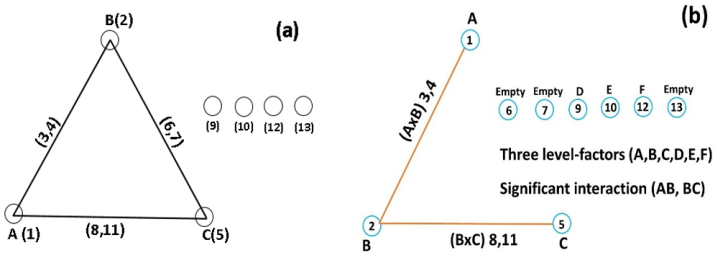


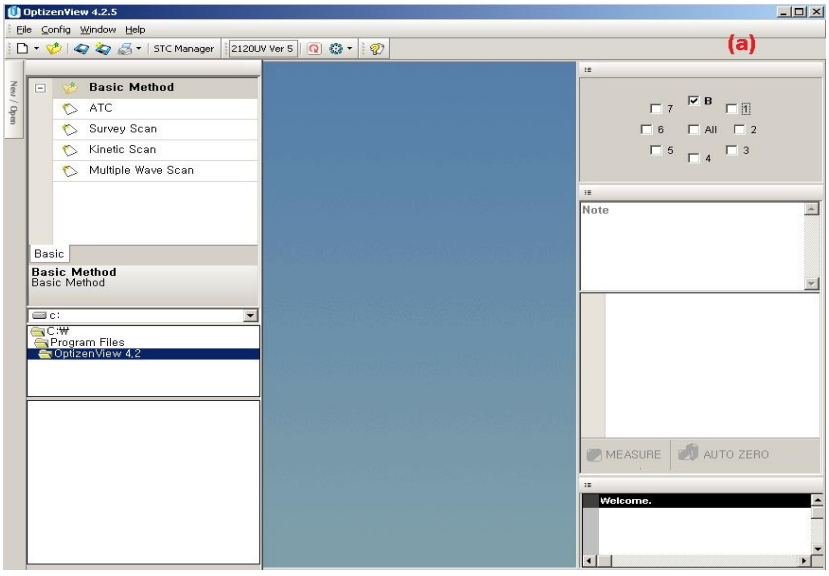
Figure 1. Standard linear graph for L_{27} (3^{13}) array (**a**); and modified linear graph for column assignment (**b**).

3.9. Statistical Analysis

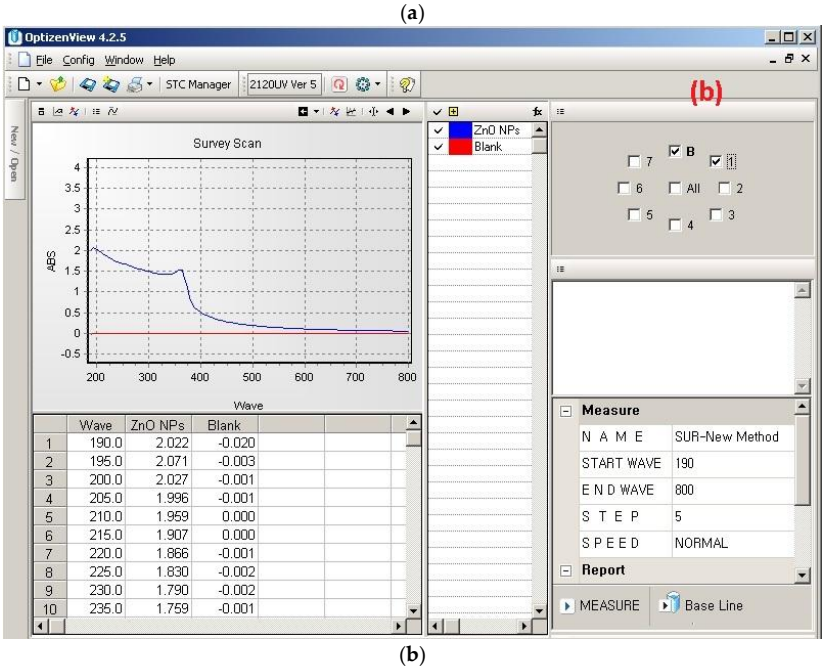
The experimental data obtained by following the Taguchi OA matrix procedure were analyzed by the general linear model (GLM) with Tukey's post Hoc test using the software MINITAB18.0 (Minitab, Inc., State College, PA, USA). The MINITAB was also used for range analysis and analysis of variance (ANOVA) and predicting the dominant parameters for ZnO NPs aggregation. Probability p-value less than 0.05 were considered to be significant. The normality and homogeneity of data were considered before conducting the ANOVA. Each experiment was performed in triplicate and average values were reported.

3.10. Optizen View 4.2.5

A UV–Vis spectrophotometer was connected with a PC interference software "Optizen view 4.2.5" to obtain the absorbance and real-time kinetics data of ZnO NPs in solution. The method adopted for obtaining sample survey scans and kinetic data is illustrated in Figure 2a–d.



(**a**)



	Wave	ZnO NPs	Blank
1	190.0	2.022	-0.020
2	195.0	2.071	-0.003
3	200.0	2.027	-0.001
4	205.0	1.996	-0.001
5	210.0	1.959	0.000
6	215.0	1.907	0.000
7	220.0	1.866	-0.001
8	225.0	1.830	-0.002
9	230.0	1.790	-0.002
10	235.0	1.759	-0.001

(**b**)

Figure 2. *Cont.*

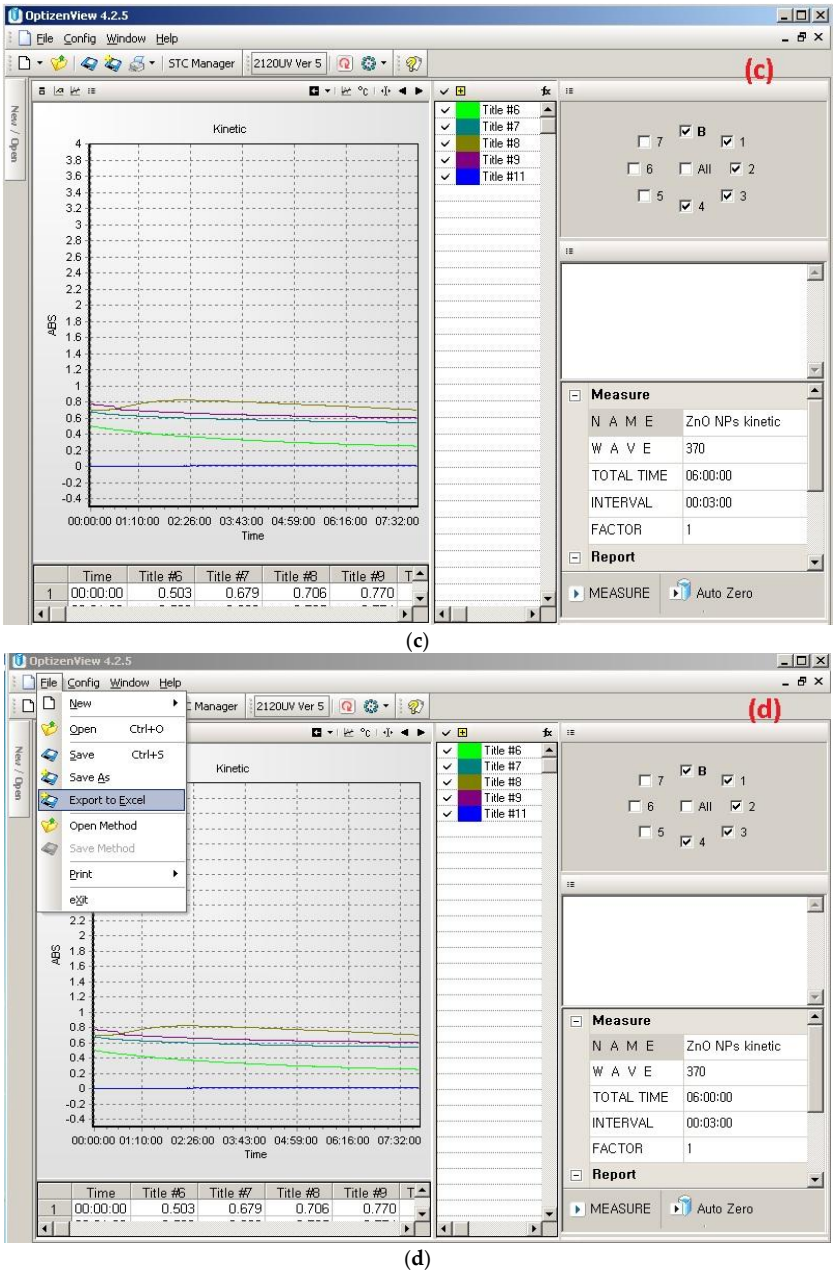


Figure 2. An illustration of Optizen view 4.2.5 software showing (**a**) main screen of user interference; (**b**) survey scan for determination of ZnO NPs wavelength; (**c**) time-resolved kinetic study of ZnO NPs under various solution chemistry; (**d**) procedure to export data to Excel.

4. User Note

Our dataset provides quality information about the sedimentation and aggregation phenomena of ZnO NPs that might be useful for researchers and the scientific community. The approach described here may help researchers who intend to increase the number of parameters and their interactions, in the range of our dataset. In addition, it may add to toxicological studies about ENPs, given that many parameters in our dataset are commonly found in natural water bodies. In the view of a global scale, one may use the dataset to incorporate more information of water chemistry parameters present in the local environments to assess other parameters that were not covered in the current matrix data.

Supplementary Materials: The following are available online at http://www.mdpi.com/2306-5729/3/2/21/s1. Table S1: Survey scan of ZnO NPs suspension in nanopure water, Table S2: Standard linear Curve of ZnO

NPs at 365 nm, Table S3: Turbidity of the solution at various time and power levels of sonicator, Table S4: Absorbance and zeta potential of ZnO NPs at various pH range, Table S5: Absorbance of ZnO NPs at different electrolyte concentration of KCl, Na_2SO_4 and $MgCl_2$, Table S6: Absorbance and corresponding zeta potential of ZnO NPs at various NOM concentrations of humic, salicyclic and citric acid respectively, Table S7: Normalized absorbance of ZnO NPs at various experimental conditions, Table S8: FT-IR spectra of ZnO NPs before and after interaction with 100 mg/L concentration of various NOM, Table S9: Column assignment for various parameters and two interactions in Taguchi L_{27} (3^{13}) orthogonal array and experimental absorbance, Table S10: The detail of the range analysis, Table S11: The detail result of the variance analysis, Table S12: Aggregation kinetic of Environmental Water samples, Table S13: Characterization of various environmental water samples and corresponding residual absorbance.

Author Contributions: R.K. and I.T.Y. conceived and designed the study. R.K. and M.A.I. conducted the experiment. S.Z.Z. and D.R.P. were responsible for experimental setup arrangements. R.K. and M.A.I. analyzed and revised the dataset. R.K. led the writing of the manuscript.

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

References

1. Keller, A.A.; McFerran, S.; Lazareva, A.; Suh, S. Global life cycle releases of engineered nanomaterials. *J. Nanopart. Res.* **2013**, *15*. [CrossRef]
2. Handy, R.D.; Von Der Kammer, F.; Lead, J.R.; Hassellöv, M.; Owen, R.; Crane, M. The ecotoxicology and chemistry of manufactured nanoparticles. *Ecotoxicology* **2008**, *17*, 287–314. [CrossRef] [PubMed]
3. Xia, T.; Kovochich, M.; Liong, M.; Mädler, L.; Gilbert, B.; Shi, H.; Yeh, J.I.; Zink, J.I.; Nel, A.E. Comparison of the Mechanism of toxicity of Zinc and Cerium Oxide Nanoparticles Based on Dissolution and Oxidative Stress Properties. *Am. Chem. Soc. Nano* **2008**, *2*, 2121–2134. [CrossRef] [PubMed]
4. Brunner, T.J.; Wick, P.; Manser, P.; Spohn, P.; Grass, R.N.; Limbach, L.K.; Bruinink, A.; Stark, W.J. In vitro cytotoxicity of oxide nanoparticles: Comparison to asbestos, silica, and the effect of particle solubility. *Environ. Sci. Technol.* **2006**, *40*, 4374–4381. [CrossRef] [PubMed]
5. Majedi, S.M.; Kelly, B.C.; Lee, H.K. Combined effects of water temperature and chemistry on the environmental fate and behavior of nanosized zinc oxide. *Sci. Total Environ.* **2014**, *496*, 585–593. [CrossRef] [PubMed]
6. Baalousha, M. Aggregation and disaggregation of iron oxide nanoparticles: Influence of particle concentration, pH and natural organic matter. *Sci. Total Environ.* **2009**, *407*, 2093–2101. [CrossRef] [PubMed]
7. Zhang, Y.; Chen, Y.; Westerhoff, P.; Hristovski, K.; Crittenden, J.C. Stability of commercial metal oxide nanoparticles in water. *Water Res.* **2008**, *42*, 2204–2212. [CrossRef] [PubMed]
8. Degen, A.; Kosec, M. Effect of pH and impurities on the surface charge of zinc oxide in aqueous solution. *J. Eur. Ceram. Soc.* **2000**, *20*, 667–673. [CrossRef]
9. Zhou, D.; Keller, A.A. Role of morphology in the aggregation kinetics of ZnO nanoparticles. *Water Res.* **2010**, *44*, 2948–2956. [CrossRef] [PubMed]
10. Mudunkotuwa, I.A.; Rupasinghe, T.; Wu, C.-M.; Grassian, V.H. Dissolution of ZnO Nanoparticles at Circumneutral pH: A Study of Size Effects in the Presence and Absence of Citric Acid. *Langmuir* **2012**, *28*, 396–403. [CrossRef] [PubMed]
11. Liu, W.S.; Peng, Y.H.; Shiung, C.E.; Shih, Y.H. The effect of cations on the aggregation of commercial ZnO nanoparticle suspension. *J. Nanopart. Res.* **2012**, *14*. [CrossRef]
12. Bian, S.W.; Mudunkotuwa, I.A.; Rupasinghe, T.; Grassian, V.H. Aggregation and dissolution of 4 nm ZnO nanoparticles in aqueous environments: Influence of pH, ionic strength, size, and adsorption of humic acid. *Langmuir* **2011**, *27*, 6059–6068. [CrossRef] [PubMed]
13. Van Hoecke, K.; De Schamphelaere, K.A.C.; Van Der Meeren, P.; Smagghe, G.; Janssen, C.R. Aggregation and ecotoxicity of CeO_2 nanoparticles in synthetic and natural waters with variable pH, organic matter concentration and ionic strength. *Environ. Pollut.* **2011**, *159*, 970–976. [CrossRef] [PubMed]
14. Philippe, A.; Schaumann, G.E. Interactions of dissolved organic matter with natural and engineered inorganic colloids: A review. *Environ. Sci. Technol.* **2014**, *48*, 8946–8962. [CrossRef] [PubMed]
15. Zhu, M.; Wang, H.; Keller, A.A.; Wang, T.; Li, F. The effect of humic acid on the aggregation of titanium dioxide nanoparticles under different pH and ionic strengths. *Sci. Total Environ.* **2014**, *487*, 375–380. [CrossRef] [PubMed]

16. Keller, A.A.; Wang, H.; Zhou, D.; Lenihan, H.S.; Cherr, G.; Cardinale, B.J.; Miller, R.; Zhaoxia, J.I. Stability and aggregation of metal oxide nanoparticles in natural aqueous matrices. *Environ. Sci. Technol.* **2010**, *44*, 1962–1967. [CrossRef] [PubMed]
17. Zhang, J.; Dong, Q.; Liu, Y.; Zhou, X.; Shi, H. Response to shock load of engineered nanoparticles in an activated sludge treatment system: Insight into microbial community succession. *Chemosphere* **2016**, *144*, 1837–1844. [CrossRef] [PubMed]
18. Zhou, X.; Huang, B.-C.; Zhou, T.; Liu, Y.C.; Shi, H. Aggregation behavior of engineered nanoparticles and their impact on activated sludge in wastewater treatment. *Chemosphere* **2015**, *119*, 568–576. [CrossRef] [PubMed]
19. Li, X.; Lenhart, J.J.; Walker, H.W. Aggregation kinetics and dissolution of coated silver nanoparticles. *Langmuir* **2012**, *28*, 1095–1104. [CrossRef] [PubMed]
20. Zhang, Y.; Chen, Y.; Westerhoff, P.; Crittenden, J. Impact of natural organic matter and divalent cations on the stability of aqueous nanoparticles. *Water Res.* **2009**, *43*, 4249–4257. [CrossRef] [PubMed]
21. Peng, Y.H.; Tsai, Y.C.; Hsiung, C.E.; Lin, Y.H.; hsin Shih, Y. Influence of water chemistry on the environmental behaviors of commercial ZnO nanoparticles in various water and wastewater samples. *J. Hazard. Mater.* **2017**, *322*, 348–356. [CrossRef] [PubMed]
22. Peng, C.; Shen, C.; Zheng, S.; Yang, W.; Hu, H.; Liu, J.; Shi, J. Transformation of CuO Nanoparticles in the Aquatic Environment: Influence of pH, Electrolytes and Natural Organic Matter. *Nanomaterials* **2017**, *7*. [CrossRef] [PubMed]
23. Khan, R.; Inam, M.A.; Zam, S.Z.; Park, D.R.; Yeom, I.T. Assessment of Key Environmental Factors Influencing the Sedimentation and Aggregation Behavior of Zinc Oxide Nanoparticles in Aquatic Environment. *Water* **2018**, *10*. [CrossRef]
24. He, G.; Chen, R.; Lu, S.; Jiang, C.; Liu, H.; Wang, C.; He, G.; Chen, R.; Lu, S.; Jiang, C.; et al. Dominating Role of Ionic Strength in the Sedimentation of Nano-TiO_2 in Aquatic Environments. *J. Nanomater.* **2015**, *2015*, 1–10. [CrossRef]
25. Dong, Y.; Li, X.; Huang, Y.; Wang, H.; Li, F. Coagulation and dissolution of zinc oxide nanoparticles in the presence of humic acid under different pH values. *Environ. Eng. Sci.* **2016**, *33*, 347–353. [CrossRef]
26. Marion, G.M.; Babcock, K.L. Predicting specific conductance and salt concentration in dilute aqueous solutions. *Soil Sci.* **1976**, *122*, 181–187. [CrossRef]
27. Kato, H.; Fujita, K.; Horie, M.; Suzuki, M.; Nakamura, A.; Endoh, S.; Yoshida, Y.; Iwahashi, H.; Takahashi, K.; Kinugasa, S. Dispersion characteristics of various metal oxide secondary nanoparticles in culture medium for in vitro toxicology assessment. *Toxicol. Vitr.* **2010**, *24*, 1009–1018. [CrossRef] [PubMed]
28. Smoluchowski, M. *Handbuch der Elektrizitat und Magnetisums 1914*; Graetz, L., Ed.; Barth-Verlag: Leipzig, Germany, 1921.
29. Fang, L.; Xiao, X.; Kang, R.; Ren, Z.; Yu, H.; Pavlostathis, S.G.; Luo, J.; Luo, X. Highly Selective Adsorption of Antimonite by Novel Imprinted Polymer with Microdomain Confinement Effect. *J. Chem. Eng. Data* **2018**, *63*, 1513–1523. [CrossRef]
30. Phadke, M.S. *Quality Engineering Using Robust Design*; Prentice Hall: Upper Saddle River, NJ, USA, 1989; ISBN 0137451679.
31. Lan, W.G.; Wongt, M.K.; Chen, N.; Sin, Y.M. Orthogonal array design as a chemometric method for the optimization of analytical procedures. Part 5. Three-level design and its application in microwave dissolution of biological samples. *Analyst* **1995**, *120*, 1115–1124. [CrossRef] [PubMed]
32. Chidambara, R.; Quan, H.Q. Advanced oxidation processes for wastewater treatment: Optimization of UV/H2O2 process through a statistical technique. *Chem. Eng. Sci.* **2005**, *60*, 5305–5311. [CrossRef]
33. Taguchi, G.; Konishi, S.; Konishi, S. *Taguchi Methods: Orthogonal Arrays and Linear Graphs: Tools for Quality Engineering*; American Supplier Institute: Dearborn, MI, USA, 1987; ISBN 094124301X.

data

MDPI

Data Descriptor

Data for an Importance-Performance Analysis (IPA) of a Public Green Infrastructure and Urban Nature Space in Perth, Western Australia

Greg D. Simpson [1,*] and Jackie Parker [2]

[1] College of Science, Health, Engineering, and Education-Environmental and Conservation Sciences, Murdoch University, Perth 6150, WA, Australia
[2] School of Design and Built Environment, Curtin University, Perth 6102, WA, Australia; 17966131@student.curtin.edu.au
* Correspondence: G.Simpson@murdoch.edu.au

Received: 23 November 2018; Accepted: 12 December 2018; Published: 17 December 2018

Abstract: This Data Descriptor shares the dataset generated by a visitor satisfaction survey of users of a mixed-use public green infrastructure (PGI) space in Perth, Western Australia, that incorporates remnant and reintroduced urban nature (UN). Conducted in the Austral summer of 2016–2017, the survey (n = 393) utilized the technique of Importance-Performance Analysis (IPA) to elucidate perceptions of PGI users regarding performance of the amenity and facilities at the study site. There is a growing body of research that reports the innate, inbuilt affinity of humans to natural systems and living things. As humankind has grown exponentially over the past 50 years, humanity, as a species, is living an increasingly urbanized lifestyle, resulting in spreading urban footprints and increased population densities that are causing humans to become increasingly disconnected from nature. These conflicting phenomena are driving research to understand the contribution that PGI and UN can make to enhancing the quality of life of urban residents. With diminishing opportunities to acquire or create new PGI spaces within ever-more-densely populated urban centers, understanding, efficiently managing, and continuously improving existing PGI spaces is crucial to access the benefits and services that PGI and UN provide. The IPA technique can provide the data necessary to inform an evidenced-based approach to managing and resourcing PGI and UN spaces.

Dataset: The dataset has been submitted for publication as a supplement to this Data Descriptor

Keywords: biophilic design; green infrastructure; Importance-Performance Analysis; IPA; public amenity; public open space; renaturing cities; urban nature; urban planning; visitor satisfaction

1. Summary

In addition to the rapid growth of humankind over the past 50 years, humanity, as a species, is becoming increasingly urbanized [1–3]. The Biophilic Hypothesis proposed by Wilson [4] states that humans have an innate, inbuilt affinity to natural systems and living things; however, the increase in the urban footprint and population density is causing human populations to become increasingly disconnected from nature [3,5,6]. These phenomena are driving research into the ways that public green infrastructure (PGI) and urban nature (UN) can enhance the quality of life of urban residents (e.g., [3,5,7–13]). With diminishing opportunities to acquire and/or create new PGI spaces within ever-more-densely populated urban centers, understanding, efficiently managing, and continuously

improving existing PGI spaces is crucial to access the benefits and services that PGI and UN provide for humankind [10,12].

This Data Descriptor shares the dataset for an Importance-Performance Analysis (IPA) survey conducted at a PGI space in Perth, Western Australia. While underutilized in the management of PGI spaces [11,13], IPA techniques provide a relatively simple and straightforward method for quantitatively assessing the performance of PGI and UN spaces. The analysis and insights arising from this dataset are reported in the Research Article by Parker and Simpson [12] published in the *Landscape Urbanism and Green Infrastructure* special issue of the MDPI journal *Land*.

Informed by the review article of Parker and Simpson [11,13] and the IPA research of Newsome et al. [14], Soldić Frleta [15], and Taplin [16], a pen-and-paper-based self-report questionnaire was developed in order to survey visitors to the Lake Claremont PGI space. The questionnaire was designed to gather data regarding the demographic profile of PGI users, and their perceptions regarding the performance of 19 attributes of quality PGI spaces identified from the literature and 3 site-specific attributes (see Section 2). A convenience intercept survey was carried out at the Lake Claremont PGI space in the 2016–2017 Austral summer, coinciding with the peak summer holiday and recreation period [3,17,18].

The demographic and IPA data collected during that survey are shared via the comma separated variable (.csv) file attached to this Data Descriptor as Supplementary Materials. Publication of this data has the potential to benefit others who are researching, planning, and managing urban PGI and UN with the goals of contributing to better PGI, enhancing the protection and renaturing of UN, and creating healthier and more liveable urban environments.

As reported above, the research associated with the dataset shared in this Data Descriptor has produced a systematic quantitative literature review article [11], the associated Data Descriptor [13], and contributed to the publication of an IPA-focused research article [12].

2. Data Description

The data extracted from the 393 questionnaires completed by PGI users were captured in Microsoft Excel and are provided as a .csv file with this Data Descriptor. The de-identified demographic data recorded for each participant is described in Tables 1–3.

Table 1. The metadata specification for gender data.

Question 1: How Do You Describe Yourself?			
Identifier	**Descriptor**	**Data Type**	**Data Values**
D1	Gender	Categorical	1 = Female 2 = Male 3 = Other 4 = Prefer not to disclose 5 = No Response

Table 2. The metadata specification for age data.

Question 2: Which Age Group Do You Belong to?			
Identifier	**Descriptor**	**Data Type**	**Data Values**
D2	Age	Categorical	1 = 18 to 24 Years of Age 2 = 25 to 34 Years of Age 3 = 35 to 44 Years of Age 4 = 45 to 54 Years of Age 5 = 55 to 64 Years of Age 6 = 65+ Years of Age 7 = No Response

Table 3. The metadata specification for usual place of residence data.

Question 3: Where do you live? Please Tick the Box Most Relevant to You and STATE the Suburb/Town/City/Country of Residence			
Identifier	**Descriptor**	**Data Type**	**Data Values**
D3	Place of Residence	Categorical	1 = Surrounding Suburbs (<5 km) 2 = Other Metropolitan Suburbs 3 = Regional Western Australia 4 = Other Australian States 5 = International 6 = No Response
D4	Place of Residence	Text	Open Question NR = No Response

Data relating to the perceptions of visitors to the Lake Claremont PGI space regarding the importance and performance of 22 attributes of quality PGI were gathered using the question shown in Figure 1.

How important are the following features of Lake Claremont to you and how satisfied are you with their management?

Item	Column A Importance					Column B Satisfaction					
For each feature listed below, please indicate: **1) The Importance of the feature by circling the most relevant the number in Column A.** **2) Your Satisfaction with that feature by circling the appropriate number in Column B.** **If you have not experienced a listed feature during today's visit, then please place an X under 'Unable to Report' at far right of Column B (Satisfaction).**	Not at all important	Not very important	Somewhat important	Very important	Extremely important	Not at all satisfied	Not very satisfied	Somewhat satisfied	Very satisfied	Extremely satisfied	Unable to report
Availability of shade (trees or structures)	1	2	3	4	5	1	2	3	4	5	
Bird watching infrastructure (observation deck, rotunda)	1	2	3	4	5	1	2	3	4	5	
Children's playground(s)	1	2	3	4	5	1	2	3	4	5	
Directional signs within the park	1	2	3	4	5	1	2	3	4	5	
Dog exercise area	1	2	3	4	5	1	2	3	4	5	
Ease of access to and around the site	1	2	3	4	5	1	2	3	4	5	

Figure 1. An extract from the survey question used to gather data regarding visitor perceptions of the Importance of attributes of the Lake Claremont public green infrastructure (PGI) and urban nature (UN) spaces and the Performance of those attributes in meeting visitor expectations (measured as visitor satisfaction).

The 22 attributes of quality PGI spaces used to assess the perceptions of PGI users regarding the Lake Claremont PGI space are reported in Table 4, and the possible responses for the perceived importance and performance of those attributes are reported in Tables 5 and 6.

Table 4. The metadata specification for identifiers of Importance and Performance ranking data.

Attribute	Importance Ranking Identifier	Performance Ranking Identifier	References Reporting PGI Attribute
1. Availability of shade (trees or structures)	IA1	PA1	[19–21]
2. Bird watching infrastructure (observation deck, rotunda)	IA2	PA2	[22–25]
3. Children's playground(s)	IA3	PA3	[19,24,26–30]
4. Directional signs within the park	IA4	PA4	[19,22,31–33]
5. Dog exercise area	IA5	PA5	[19,34]
6. Ease of access to and around the site	IA6	PA6	[25,30,35–38]
7. Fencing	IA7	PA7	[19]
8. High quality European/English-themed spaces and areas	IA8	PA8	[21,34,39–43]
9. High-quality infrastructure (paths, lights, toilets, barbeque (BBQ), benches)	IA9	PA9	[31,33,42–48]
10. High-quality lake water body	IA10	PA10	[3,49–51]
11. High-quality nature spaces and areas	IA11	PA11	[23,26,30,43,46,52–54]
12. High-quality services (café, gym, golf club)	IA12	PA12	[24,31,33,42,44–48]
13. High-quality turf	IA13	PA13	[19,33,34]
14. Interpretive information and signs	IA14	PA14	[19]
15. Native fauna presence and activity	IA15	PA15	[3,22–24]
16. Off-leash dog exercise	IA16	PA16	[19,34]
17. On-leash dog walking	IA17	PA17	[19,34]
18. Other sporting installations (Aquatic Centre, Cricket/Hockey Oval, and/or Tennis Club)	IA18	PA18	Site-Specific
19. Par 3 Golf Course	IA19	PA19	Site-Specific
20. Park exercise equipment	IA20	PA20	Site-Specific
21. Personal safety	IA21	PA21	[32,33,43]
22. Tree management	IA22	PA22	[20–22,30,42,45,55]

Table 5. The metadata specification for Importance of Attributes ranking data.

Importance Ranking Identifier	Data Type	Data Values
IA1 to IA22	Categorical	1 = Not at all important 2 = Not very important 3 = Somewhat important 4 = Very important 5 = Extremely important

Table 6. The metadata specification for Performance of Attributes ranking data.

Performance Ranking Identifier	Data Type	Data Values
PA1 to PA22	Categorical	1 = Not at all satisfied 2 = Not very satisfied 3 = Somewhat satisfied 4 = Very satisfied 5 = Extremely satisfied 0 = Unable to report

3. Methods

3.1. Study Site

The visitor satisfaction data reported in this Data Descriptor were collected at the Lake Claremont PGI and UN space in Perth, Western Australia (31.9738° S, 115.7771° E). Additional information regarding the location, land-use history, current condition, and utilization of the mixed-use Lake Claremont PGI and UN spaces are provided in Parker [10], Parker and Simpson [12], and Simpson and Newsome [3]. Location and surrounding land-use maps are provided in Simpson and Newsome [3] and Parker and Simpson [12], respectively. The following paragraph provides a short summary of the geomorphology and vegetation of the study site.

Lake Claremont is located on the Tamala Limestone zone of the Swan Coastal Plain at the boundary of the Quinadlaup Dune System and the older Spearwood Dune System [3,56]. Under the Koppen climate classification, the Southwest of Western Australia experiences a Mediterranean climate with hot dry summers and cooler wetter winters [18,57,58]. The traditional custodians of the

land, the people of the Noongar nation, identify six seasons for this region [58,59]. Prior to European colonization, the indigenous vegetation of the region was a mix of *Agonis* and Tuart Woodlands and *Banksia* Woodlands [3,60–62]. As described by Simpson and Newsome [3], the site was heavily modified and degraded as a result of European colonization. Today, the vegetation present at the site is a mix of remnant and renatured indigenous vegetation in the UN spaces to the west and north of the lake, while the eastern and southern sides of the lake consist primarily of grassed areas with a mix of trees that are exotic, local native, and out-of-area 'native' species [3]. Lake Claremont is listed as a Conservation Category Wetland in the Government of Western Australia Geomorphic Wetlands: Swan Coastal Plain dataset [3,63]. The remnant and renatured indigenous vegetation of Lake Claremont and the surrounding PGI and UN spaces has a level of protection under the Bush Forever Site 220 and Environmentally Sensitive Area classifications of the Government of Western Australia and through the Government of Australia's classification of the remnant Banksia Woodland of the Swan Coastal Plain as an Endangered Ecological Community [3,64–66].

3.2. Survey

To inform the development of the survey questionnaire, relevant literature was consulted [11,13] guided by Pickering and Byrne [67] and the PRISMA method of Moher et al. [68]. The sourced literature revealed a number of universally recognized PGI features, such as access paths, open turf areas, seating, infrastructure, and playgrounds. These universally recognized PGI features were assessed in terms of their presence at the Lake Claremont study site and suitability for inclusion in the IPA question of the survey. The questionnaire asked three tick-box categorical demographic questions (Tables 1–3) and one IPA question (Figure 1). The demographic question regarding a usual place of residence of the participant included an open-ended aspect that allowed participants to share the suburb, town, Australian state, and/or country that represented their usual place of residence. The IPA question assessed the perceptions of PGI users regarding the importance and performance of 22 attributes of quality PGI spaces (Tables 4–6) that were rated on a five-point Likert scale for importance and a six-point Likert scale, which incorporated an *Unable to Report* option, for performance.

A power analysis was undertaken prior to surveying to ensure that sufficient participant numbers would be achieved to allow for valid inferences to be drawn from the results [69]. It was determined that 259 participants were required to be 90% confident that detected differences between the performance of attributes were a valid effect at an $\alpha = 0.05$ level of statistical significance assuming a correlation of $\rho = 0.2$ between the Importance and Performance rankings of survey participants. The responses from 393 analyzable questionnaires are shared via the dataset connected to this Data Descriptor.

Participants were recruited through a convenience intercept approach to surveying PGI users. Several survey events were scheduled at differing times of the day and across all days of the week during December 2016 and January 2017 to limit the potential for response bias and cognitive bias (i.e., an elevated response of participants during the festive season or weekends in comparison to day-to-day life).

Once completed, participants immediately returned their anonymous, self-reported, pen-and-paper-based questionnaires to the researchers who secured the questionnaire for later transcription of the data. After each field survey event, de-identified responses of participants were recorded in an Excel workbook for analysis and storage.

3.3. Limitations and Learnings

The survey reported in this Data Descriptor and the research article of Parker and Simpson [12] was the first time that either author had utilized the IPA technique. For that reason, the authors adapted the questionnaire developed and tested by others to conduct the IPA study reported in Newsome et al. [14] for their PGI research. Unfortunately, only after the data collection phase of the PGI survey was completed did it become apparent that there was a non-fatal flaw in the design of the

IPA question from the survey of Newsome et al. [14] that had consequently been carried over to the PGI survey.

Two important, but rarely considered, assumptions that underpin the Likert scale are that the ordinal categories have the same *span* or *intensity* and that the mid-point of the scale is a neutral inflection point between the negative response categories and positive response categories [70–72]. Those assumptions are generally met by constructing a Likert scale with an equal number of matched positive and negative categories. The assumption of a neutral mid-point is met either by providing an explicitly stated neutral mid-point on a Likert scale with an odd number of categories, or implicitly through the use of a *forced-choice* Likert scale that has an even number of categories [70,73]. As its name suggests, a forced-choice Likert scale requires survey participants to express either a positive or a negative view either side of the unstated neutral mid-point. The combination of these two assumptions provides the opportunity for a linear relationship to exist between the importance and performance of the assessed attributes that is implicit in a Martilla and James [74] IPA Matrix [10,16,75].

It will now be self-evident that the IPA question presented in Figure 1 utilities a forced-choice Likert scale with three positive response categories and just two negative response categories. While not fatal for the IPAs presented in Newsome et al. [14] and Parker and Simpson [12], it does mean that the negative ordinal categories of the dataset shared in this Data Descriptor span a slightly wider range than the span for the positive categories, and the mid-point of the scale has a value of 2.5 rather than the value of 3 as would normally apply for a 5-point Likert scale.

Supplementary Materials: Lake Claremont IPA Dataset.csv is available online at http://www.mdpi.com/2306-5729/3/4/69/s1.

Author Contributions: G.S. and J.P. made equal contributions to this paper and as such are co-first authors.

Funding: This research received no external funding.

Acknowledgments: We thank our colleague David Newsome for his guidance of J.P.'s research and comments on the associated thesis. We thank Dianne Parker for her efforts in transcribing the anonymous survey data from the self-report questionnaires to an Excel spreadsheet. This research was undertaken under Murdoch University Human Ethics Committee Approval 2016/213.

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

References

1. Beatley, T. *Biophilic Cities: Integrating Nature into Urban Design and Planning*; Island Press: Washington, DC, USA, 2011; ISBN 978-1-59-726715-1.
2. Kopecká, M.; Szatmári, D.; Rosina, K. Analysis of urban green spaces based on Sentinel-2A: Case studies from Slovakia. *Land* **2017**, *6*, 25. [CrossRef]
3. Simpson, G.; Newsome, D. Environmental history of an urban wetland: From degraded colonial resource to nature conservation area. *Geo Geogr. Environ.* **2017**, *4*, e00030. [CrossRef]
4. Wilson, E.O. *Biophilia*; Harvard University Press: Cambridge, MA, USA, 1984; ISBN 978-0-67-407442-2.
5. Miller, J.R. Biodiversity conservation and the extinction of experience. *Trends Ecol. Evol.* **2005**, *20*, 430–434. [CrossRef] [PubMed]
6. Neuman, M. The compact city fallacy. *J. Plan. Educ. Res.* **2005**, *25*, 11–26. [CrossRef]
7. Jones, C.; Newsome, D. Perth (Australia) as one of the world's most liveable cities: A perspective on society, sustainability and environment. *Int. J. Tour. Cities* **2015**, *1*, 18–35. [CrossRef]
8. Newton, P.W. Liveable and sustainable? Socio-technical challenged for the twenty-first century cities. *J. Urban Technol.* **2012**, *19*, 81–102. [CrossRef]
9. Okulicz-Kozaryn, A. City life: Rankings (liveability) versus perceptions (satisfaction). *Soc. Indic. Res.* **2011**, *110*, 433–451. [CrossRef]
10. Parker, J.A. Survey of Park User Perception in the Context of Green Space and City Liveability: Lake Claremont, Western Australia. Master's Thesis, Murdoch University, Perth, Australia, 2017. Available online: http://researchrepository.murdoch.edu.au/id/eprint/40856/ (accessed on 8 October 2018).

11. Parker, J.; Simpson, G.D. Public Green Infrastructure Contributes to City Liveability: A Systematic Quantitative Review. *Land* **2018**, *7*. in press.
12. Parker, J.; Simpson, G. Visitor Satisfaction with a Public Green Infrastructure and Urban Nature Space in Perth, Western Australia. *Land* **2018**, *7*. in press.
13. Simpson, G.D.; Parker, J. Data on Peer Reviewed Papers about Green Infrastructure, Urban Nature, and City Liveability. *Data* **2018**, *3*, 51. [CrossRef]
14. Newsome, D.; Rodger, K.; Pearce, J.; Chan, K.L.J. Visitor satisfaction with a key wildlife tourism destination within the context of a damaged landscape. *Curr. Issues Tour.* **2017**. [CrossRef]
15. Soldić Frleta, D. Island destinations' tourism offer-tourists' vs. residents' attitudes. *Tour. Hosp. Manag.* **2014**, *20*, 1–14. Available online: https://hrcak.srce.hr/123774 (accessed on 15 December 2018).
16. Taplin, R.H. Competitive importance-performance analysis of an Australian wildlife park. *Tour. Manag.* **2012**, *33*, 29–37. [CrossRef]
17. Patroni, J.; Day, A.; Lee, D.; Chan, J.K.L.; Kerr, D.; Newsome, D.; Simpson, G.D. Looking for evidence that place of residence influenced visitor attitudes to feeding wild dolphins. *Tour. Hosp. Manag.* **2018**, *24*, 87–105. [CrossRef]
18. Simpson, G.; Newsome, D.; Day, A. Data from a survey to determine visitor attitudes and knowledge about the provisioning of wild dolphins at a marine tourism destination. *Data Brief* **2016**, *9*, 940–945. [CrossRef] [PubMed]
19. Crawford, D.; Timperio, A.; Giles-Corti, B.; Ball, K.; Hume, C.; Roberts, R.; Andrianopoulos, N.; Salmon, J. Do features of public open spaces vary according to neighbourhood socio-economic status? *Health Place* **2008**, *14*, 889–893. [CrossRef] [PubMed]
20. Norton, B.A.; Coutts, A.M.; Livesley, S.J.; Harris, R.J.; Hunter, A.M.; Williams, N.S. Planning for cooler cities: A framework to prioritise green infrastructure to mitigate high temperatures in urban landscapes. *Landsc. Urban Plan.* **2015**, *134*, 127–138. [CrossRef]
21. Shackleton, S.; Chinyimba, A.; Hebinck, P.; Shackleton, C.; Kaoma, H. Multiple benefits and values of trees in urban landscapes in two towns in northern South Africa. *Landsc. Urban Plan.* **2015**, *136*, 76–86. [CrossRef]
22. Barth, B.J.; FitzGibbon, S.I.; Wilson, R.S. New urban developments that retain more remnant trees have greater bird diversity. *Landsc. Urban Plan.* **2015**, *136*, 122–129. [CrossRef]
23. Dallimer, M.; Irvine, K.N.; Skinner, A.M.; Davies, Z.G.; Rouquette, J.R.; Maltby, L.L.; Warren, P.H.; Armsworth, P.R.; Gaston, K.J. Biodiversity and the feel-good factor: Understanding associations between self-reported human well-being and species richness. *BioScience* **2012**, *62*, 47–55. [CrossRef]
24. Irvine, K.N.; Devine-Wright, P.; Payne, S.R.; Fuller, R.A.; Painter, B.; Gaston, K.J. Green space, soundscape and urban sustainability: An interdisciplinary, empirical study. *Local Environ.* **2009**, *14*, 155–172. [CrossRef]
25. Sushinsky, J.R.; Rhodes, J.R.; Possingham, H.P.; Gill, T.K.; Fuller, R.A. How should we grow cities to minimize their biodiversity impacts? *Glob. Chang. Biol.* **2012**, *19*, 401–410. [CrossRef] [PubMed]
26. Bratman, G.N.; Hamilton, P.; Daily, G.C. The impacts of nature experience on human cognitive function and mental health. *N. Y. Acad. Sci.* **2012**, *1249*, 118–136. [CrossRef] [PubMed]
27. Hausmann, A.; Slotow, R.O.B.; Burns, J.K.; Di Minin, E. The ecosystem service of sense of place: Benefits for human well-being and biodiversity conservation. *Environ. Conserv.* **2016**, *43*, 117–127. [CrossRef]
28. Raquel, C.D.S.M.; Montalto, F.A.; Palmer, M.I. Potential climate change impacts on green infrastructure vegetation. *Urban For. Urban Green.* **2016**, *20*, 128–139. [CrossRef]
29. Salata, K.; Yiannakou, A. Green Infrastructure and climate change adaptation. *TeMA J. Land Use Mobil. Environ.* **2016**, *9*, 7–24.
30. Thompson, C.W. Urban open space in the 21st century. *Landsc. Urban Plan.* **2002**, *60*, 59–72. [CrossRef]
31. Cattell, V.; Dines, N.; Gesler, W.; Curtis, S. Mingling, observing, and lingering: Everyday public spaces and their implications for well-being and social relations. *Health Place* **2008**, *14*, 544–561. [CrossRef]
32. Francis, J.; Giles-Corti, B.; Wood, L.; Knuiman, M. Creating sense of community: The role of public space. *J. Environ. Psychol.* **2012**, *32*, 401–409. [CrossRef]
33. Giles-Corti, B.; Broomhall, M.H.; Knuiman, M.; Collins, C.; Douglas, K.; Ng, K.; Lange, A.; Donovan, R.J. Increasing walking: How important is distance to attractiveness, and size of public open space? *Am. J. Prev. Med.* **2005**, *28*, 169–176. [CrossRef]
34. Sugiyama, T.; Gunn, L.D.; Christian, H.; Francis, J.; Foster, S.; Hooper, P.; Owen, N.; Giles-Corti, B. Quality of public open spaces and recreational walking. *Am. J. Public Health* **2015**, *105*, 2490–2495. [CrossRef] [PubMed]

35. Appiah-Opoku, S. Using protected areas as a tool for biodiversity conservation and ecotourism: A case study of Kakum National Park in Ghana. *Soc. Nat. Resour.* **2011**, *24*, 500–510. [CrossRef]
36. Hillsdon, M.; Panter, J.; Foster, C.; Jones, A. The relationship between access and quality of urban green space with population physical activity. *Public Health* **2006**, *120*, 1127–1132. [CrossRef] [PubMed]
37. Keniger, L.E.; Gaston, K.J.; Irvine, K.N.; Fuller, R.A. What are the benefits of interacting with nature? *Int. J. Environ. Res. Public Health* **2013**, *10*, 913–935. [CrossRef] [PubMed]
38. Van Herzele, A.; Wiedemann, T. A monitoring tool for the provision of accessible and attractive urban green spaces. *Landsc. Urban Plan.* **2003**, *63*, 109–126. [CrossRef]
39. Balding, M.; Williams, K.J. Plant blindness and the implications for plant conservation. *Conserv. Biol.* **2016**, *30*, 1192–1199. [CrossRef]
40. Chen, B.; Adimo, O.A.; Bao, Z. Assessment of aesthetic quality and multiple functions of urban green space from the users' perspective: The case of Hangzhou Flower Garden, China. *Landsc. Urban Plan.* **2009**, *93*, 76–82. [CrossRef]
41. Green, T.L.; Kronenberg, J.; Andersson, E.; Elmqvist, T.; Gómez-Baggethun, E. Insurance value of green infrastructure in and around cities. *Ecosystems* **2016**, *19*, 1051–1063. [CrossRef]
42. Grose, M.J. Changing relationships in public open space and private open space in suburbs in south-western Australia. *Landsc. Urban Plan.* **2009**, *92*, 53–63. [CrossRef]
43. Ikin, K.; Le Roux, D.S.; Rayner, L.; Villaseñor, N.R.; Eyles, K.; Gibbons, P.; Manning, A.D.; Lindenmayer, D.B. Key lessons for achieving biodiversity-sensitive cities and towns. *Ecol. Manag. Restor.* **2015**, *16*, 206–214. [CrossRef]
44. Antognelli, S.; Vizzari, M. Landscape liveability spatial assessment integrating ecosystem and urban services with their perceived importance by stakeholders. *Ecol. Indic.* **2017**, *72*, 703–725. [CrossRef]
45. Shanahan, D.F.; Lin, B.B.; Bush, R.; Gaston, K.J.; Dean, J.H.; Barber, E.; Fuller, R.A. Toward improved public health outcomes from urban nature. *Am. J. Public Health* **2015**, *105*, 470–477. [CrossRef] [PubMed]
46. Soga, M.; Yamaura, Y.; Aikoh, T.; Shoji, Y.; Kubo, T.; Gaston, K.J. Reducing the extinction of experience: Association between urban form and recreational use of public greenspace. *Landsc. Urban Plan.* **2015**, *143*, 69–75. [CrossRef]
47. Taylor, B.T.; Fernando, P.; Bauman, A.E.; Williamson, A.; Craig, J.C.; Redman, S. Measuring the quality of public open space using Google Earth. *Am. J. Prev. Med.* **2011**, *40*, 105–112. [CrossRef] [PubMed]
48. Villanueva, K.; Badland, H.; Hooper, P.; Koohsari, M.J.; Mavoa, S.; Davern, M.; Roberts, R.; Goldfeld, S.; Giles-Corti, B. Developing indicators of public open space to promote health and wellbeing in communities. *Appl. Geogr.* **2015**, *57*, 112–119. [CrossRef]
49. Dale, P.E.R.; Connelly, R. Wetlands and human health: An overview. *Wetl. Ecol. Manag.* **2012**, *20*, 165–171. [CrossRef]
50. Do, Y.; Kim, S.B.; Kim, J.Y.; Joo, G.J. Wetland-based tourism in South Korea: Who, When, and Why. *Wetl. Ecol. Manag.* **2015**, *23*, 779–787. [CrossRef]
51. Revell, G.; Anda, M. Sustainable urban biophilia: The case of greenskins for urban density. *Sustainability* **2014**, *6*, 5423–5438. [CrossRef]
52. Andersson, E.; Barthel, S.; Borgström, S.; Colding, J.; Elmqvist, T.; Folke, C.; Gren, Å. Reconnecting Cities to the Biosphere: Stewardship of Green Infrastructure and Urban Ecosystem Services. *Ambio* **2014**, *43*, 445–453. [CrossRef]
53. Battisti, C. Experiential key species for the nature-disconnected generation. *Anim. Conserv.* **2016**, *19*, 485–487. [CrossRef]
54. Van den Berg, A.E.; Hartig, T.; Staats, H. Preference for nature in urbanized societies: Stress, restoration, and the pursuit of sustainability. *J. Soc. Issues* **2007**, *63*, 79–96. [CrossRef]
55. Hartig, T.; Evans, G.W.; Jamner, L.D.; Davis, D.S.; Gärling, T. Tracking restoration in natural and urban field settings. *J. Environ. Psychol.* **2003**, *23*, 109–123. [CrossRef]
56. Tapsell, P.; Newsome, D.; Bastian, L. Origin of yellow sand from Tamala limestone on the Swan Coastal Plain, Western Australia. *Aust. J. Earth Sci.* **2003**, *50*, 331–342. [CrossRef]
57. Met Office. *Helping You Understand Weather and Climate*; Met Office, Government of the UK: Exeter, UK, 2012; pp. 2–3.

58. Simpson, G.D. Cracking the Niche: An Investigation into the Impact of Climatic Variables on Germination of the Rare Shrub *Verticordia staminosa* Subspecies *staminosa* (Myrtaceae). Honours Thesis, Murdoch University, Perth, Australia, 2011. Available online: http://researchrepository.murdoch.edu.au/id/eprint/8485/ (accessed on 25 October 2018).
59. South West Aboriginal Land & Sea Council. Kaartdijin in Noongar—Noongar Knowledge: Sharing Noongar Culture. Available online: https://www.noongarculture.org.au/ (accessed on 27 October 2018).
60. Beard, J.S. Definition and locations of the Banksia woodlands. *J. R. Soc. West. Aust.* **1989**, *71*, 85–86.
61. Fowler, W. Soil Seed Bank Dynamics in Transferred Topsoil: Evaluating Restoration Potentials. Honours Thesis, Murdoch University, Perth, Australia, 2012. Available online: http://researchrepository.murdoch.edu.au/id/eprint/13389/ (accessed on 25 October 2018).
62. Ritchie, A.; Sinclair, E.; Stevens, J.; Commander, L.; Davis, R.; Fowler, W. EcoCheck: Perth's Banksia Woodlands Are in the Path of the Sprawling City. The Conversation 2016. Available online: https://theconversation.com/ecocheck-perths-banksia-woodlands-are-in-the-path-of-the-sprawling-city-59911 (accessed on 27 October 2018).
63. Government of Western Australia. Geomorphic Wetlands, Swan Coastal Plain (DBCA-019). Available online: https://catalogue.data.wa.gov.au/dataset/geomorphic-wetlands-swan-coastal-plain (accessed on 22 November 2018).
64. Government of Western Australia: Department of Water Environmental Regulation. Environmentally Sensitive Areas. Available online: https://www.der.wa.gov.au/your-environment/environmentally-sensitive-areas (accessed on 22 November 2018).
65. Government of Western Australia: Department of Planning, Lands, and Heritage. Bush Forever: Volume 2: Directory of Bush Forever. Available online: https://www.planning.wa.gov.au/publications/5911.aspx (accessed on 22 November 2018).
66. Australian Government. Species Profile and Threats Database: Banksia Woodlands of the Swan Coastal Plain Ecological Community. Available online: http://www.environment.gov.au/cgi-bin/sprat/public/publicshowcommunity.pl?id=131&status=Endangered (accessed on 22 November 2018).
67. Pickering, C.M.; Byrne, J. The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early career researchers. *High. Educ. Res. Dev.* **2013**, *33*, 534–548. [CrossRef]
68. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Ann. Intern. Med.* **2009**, *151*, 264–269. [CrossRef] [PubMed]
69. Francis, G.; Garing, A. *Foundations of Statistics*; Pearson Australia: Sydney, Australia, 2013; ISBN 978-1-48-861343-2.
70. Albaum, G. The Likert scale revisited. *Mark. Res. Soc. J.* **1997**, *39*, 1–21. [CrossRef]
71. Babbie, E.R. *The Practice of Social Research*, 6th ed.; Wadsworth Publishing Company: Belmont, CA, USA, 1992; pp. 180–181. ISBN 0-534-15576-6.
72. Sarantakos, S. *Social Research*, 2nd ed.; Macmillan Education Australia Pty. Ltd.: South Yarra, Australia, 1998; pp. 89–90. ISBN 0-7329-4577-1.
73. Jennings, G. *Tourism Research*; John Wiley & Sons: Milton, Australia, 2001; ISBN 978-0471342557.
74. Martilla, J.A.; James, J.C. Importance-Performance Analysis. *J. Mark.* **1977**, *41*, 77–79. [CrossRef]
75. Oh, H. Revisiting importance–performance analysis. *Tour. Manag.* **2001**, *22*, 617–627. [CrossRef]

data

MDPI

Data Descriptor

Dataset for SERS Plasmonic Array: Width, Spacing, and Thin Film Oxide Thickness Optimization

Christopher M. Klenke [1,2], Zachary T. Brawley [1,3], Stephen J. Bauman [1], Ahmad A. Darweesh [1,4], Desalegn T. Debu [3] and Joseph B. Herzog [1,3,5,*]

1 Microelectronics-Photonics Graduate Program, University of Arkansas, 731 W. Dickson St., Fayetteville, Arkansas, AR 72701, USA; klenke11@live.missouristate.edu (C.M.K.); ztbrawle@email.uark.edu (Z.T.B.); sjbauman@email.uark.edu (S.J.B.); aadarwee@email.uark.edu (A.A.D.)
2 Physics, Astronomy, and Materials Science Department, Missouri State University, 921 S. John Q Hammons Pkwy, Springfield, MO 65897, USA
3 Department of Physics, University of Arkansas, 825 W. Dickson St., Fayetteville, Arkansas, AR 72701, USA; dtdebu@email.uark.edu
4 Department of Physics, College of Science, AL-Nahrain University, Baghdad 10081, Iraq
5 R.B. Annis School of Engineering, University of Indianapolis, Indianapolis, IN 46227, USA
* Correspondence: jbherzog@uark.edu; Tel.: +1-317-781-5398

Received: 22 August 2018; Accepted: 18 September 2018; Published: 19 September 2018

Abstract: Surface-enhanced Raman spectroscopy (SERS) improves the scope and power of Raman spectroscopy by taking advantage of plasmonic nanostructures, which have the potential to enhance Raman signal strength by several orders of magnitude, which can allow for the detection of analyte molecules. The dataset presented provides results of a computational study that used a finite element method (FEM) to model gold nanowires on a silicon dioxide substrate. The survey calculated the surface average of optical surface enhancement due to plasmonic effects across the entire model and studied various geometric parameters regarding the width of the nanowires, spacing between the nanowires, and thickness of the silicon dioxide substrate. From this data, enhancement values were found to have a periodicity due to the thickness of the silicon dioxide. Additionally, strong plasmonic enhancement for smaller distances between nanowires were found, as expected; however, additional surface enhancement at greater gap distances were observed, which were not anticipated, possibly due to resonance with periodic dimensions and the frequency of the light. This data presentation will benefit future SERS studies by probing further into the computational and mathematical material presented previously.

Dataset: Available online: https://osf.io/6vb2w/?view_only=7629f06f081541968dbe986951167c4e

Keywords: plasmonics; thin film; SERS; computational electromagnetics; nanowires; nano-optics; grating; array

1. Introduction

Due to advances in nanotechnology engineering and manufacturing in recent years, there has been an increase in improvements to surfaced-enhanced Raman spectroscopy (SERS). SERS is important for both its versatility and effectiveness at fingerprinting analyte species. It can detect analyte molecules down to the individual molecule level [1–3] as well as analyze different material phases [4,5] in a variety of fields such as experimental chemical sensing (both organic and inorganic) [4], biomedical technologies [4,5], and substance detection [6–8] based on detecting the vibrational modes of analyte

molecules. A Raman (inelastic) scattering signal forms a unique spectrum corresponding to energy shifts in the molecule compared to the Rayleigh (elastic) scattering. Unfortunately, this method is limited by the amount of light scattered in this way by molecules, which often produces a weak signal that can be difficult to detect.

To improve the limit of detection, the method of SERS can be employed. Plasmonic structures on the substrate surface can improve the Raman signal by many orders of magnitude [9–11] and can be tuned by taking advantage of specific nanostructure geometries capable of enhancing the electric near-field surrounding the nanostructures. One example of such nanostructures is a plasmonic nanograting; a structure that assumes a periodic assembly of plasmonic elements with nanoscale dimensions in at least one direction. By changing the space between the elements, one can create a resonance condition and maximize the local electric field enhancement, which is typically considered as the ratio of the magnitudes of the local electric field and the incident light $|\boldsymbol{E}/\boldsymbol{E_0}|$ [12,13].

Computational electromagnetic methods are often used to study theoretical SERS responses of different nanostructures, making use of techniques such as finite-difference time domain (FDTD), finite element method (FEM), and discrete dipole approximation (DDA) [14–16]. Prior studies have primarily investigated the nanogaps between the structures and not the complete surface of the gratings [17,18]. This gap-field-only method presents an idealized data set that only accounts for light through the strongest of enhanced electrical fields and does not give a full representation of enhancement across the entire grating pattern. For practical purposes, the nanograting must improve optical enhancement not only in the space between nanowires, but across the whole surface. The data discussed in [19] and provided fully in this paper (see supplemental file) offers a more complete evaluation of the enhanced field along a substrate's surface and its efficacy in SERS.

The data presented and discussed in this paper show the results of computational modeling of plasmonic nanostructures. A FEM was used in this study, and line averages across the surface of a simulated 2D cross-section of a grating and substrate were analyzed to improve the understanding of the electric field properties surrounding the structure.

2. Methods

To study plasmonic electric field enhancements from Au nanostructures, we used the FEM technique via the software COMSOL Multiphysics [20]. The enhancement of the local electric field intensity, $(|\boldsymbol{E}|/|\boldsymbol{E_0}|)^2$, was calculated. The software then calculated a line average using the values at each mesh point across the entire 2D cross-sectional structure surface, from which enhancement values were derived for the entire model.

The base model accuracy was verified using convergence tests and analytical techniques so that it may be used for the desired parametric studies. Multiple sweeps were then simulated across three parameters: wire width (w), electrode spacing (s), and SiO_2 thickness (t_{SiO2}), in 10 nm steps. A schematic of the simulated structure and diagram of the models with their parameters are shown in Figure 1. The path over which the line average enhancement was calculated follows the boundary between the air region and the adjacent material regions below. It thus follows along the SiO_2 surface until reaching the Au structure where it follows the air-metal boundary until the other side of the metallic structure where it again follows the air-SiO_2 interface.

Components of the model held constant during all sweeps include the height of the air above the structure, which was held at 500 nm, the thickness of the Si layer beneath the SiO_2 layer, which was held at 200 nm, and the Ti adhesion layer between each Au grating and the SiO_2 layers, which was held at 1 nm. The Ti layer was kept at this small thickness due to the adhesive properties of Ti between Au and SiO_2 [21] in conjunction with its inherent mitigation of plasmonic effects [21]. The Au grating thickness was held at 15 nm to represent current nanofabrication [22] capabilities.

In all simulations, the lengths of the nanowires were simulated to be infinite, as was the orthogonal structure periodicity ($P = w + s$) by applying periodic boundary conditions. The air and Si layers were simulated as infinitely thick via port boundary conditions, representing perfect light absorption at the boundaries. Ports on the top and bottom of the simulation space were used to introduce wave excitation at the top and eliminate backscattering at the bottom boundary. The goal of the model was to simulate surface optical enhancement for improved SERS to compare with our lab's experimental Raman spectroscopy capability, which uses a 785 nm diode laser to avoid sample photoluminescence at more energetic incident wavelengths while increasing the scattering signal compared to that resulting from longer wavelength incident light. To achieve this, a plane wave of wavelength $\lambda_0 = 785$ nm was simulated as being excited at the top port, incident in the negative y-direction and polarized in the x-direction, as depicted in Figure 1a. The wave parameters were held constant through all parametric sweeps. The complex optical material properties of Si, SiO_2, Ti, Au, and air were used as per commonly cited experimental results [23–25]. Lastly, the top corners of the wire structures were beveled with a radius of 4 nm to better represent the experimental results of structure fabrication [22].

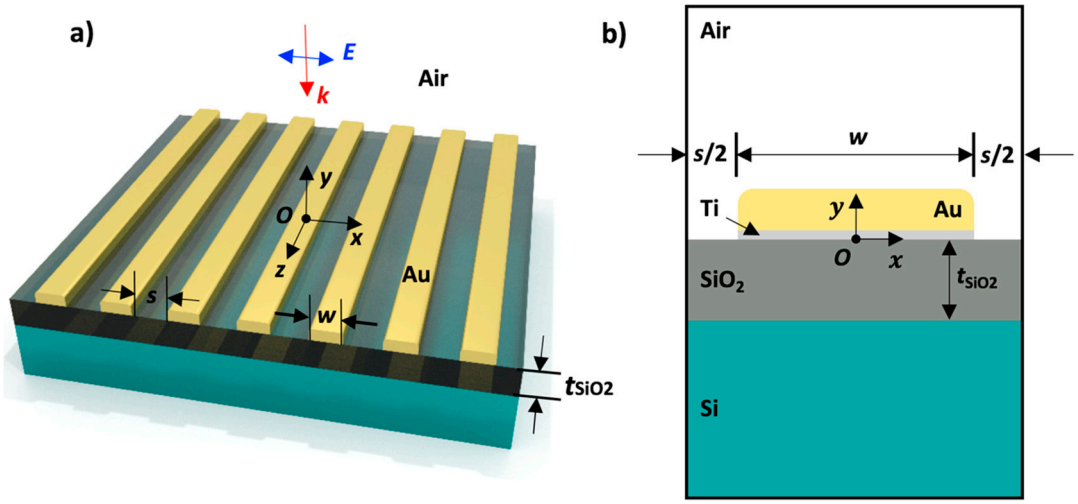


Figure 1. Schematics of the FEM model. (**a**) 3D schematic of the SERS substrate structure which is modeled using a (**b**) 2D cross-section simulation. Structure consisting of a Si substrate with a SiO_2 thin film on top and Au nanowires bonded with a Ti adhesion to layer the SiO_2 thin film layer. Variables swept include the gold nanowire width (w), the spacing between electrodes (s), and the thickness of the SiO_2 thin film (t_{SiO2}). The incident light direction and polarization is shown. Note: Ti adhesion layer is not visible in (**a**).

Figure 2 depicts the optical field enhancement produced at specific values for all three variable parameters: $s = 10$ nm, $w = 70$ nm, and $t_{SiO2} = 330$ nm. The color map describes the variation of optical enhancement across the entire surface. These values are not optimized to the combination that would allow for the greatest amount of surface enhancement and yet still produce very strong electric field augmentation.

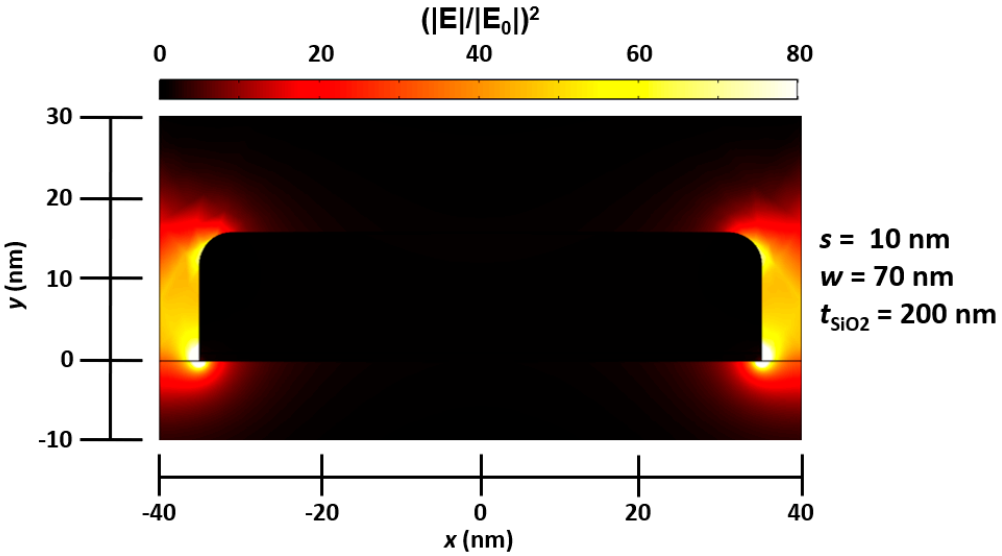


Figure 2. Graphical output of FEM plotting the electric field distribution of the optical enhancement of the simulation space of a periodic SERS substrate with the following variables: w = 10 nm, s = 70 nm, and t_{SiO2} = 200 nm. Please note that this plot shows the full width of the model, but it does not show the full height.

3. Data Description

The datasets accompanying this paper as supplementary files provide the data compiled during the computational studies in [21]. The primary file is separated into subfolders that are denoted by the figures in [21], with an additional file being the base model used to create the data presented. In these subfolders are files of three different types: text files containing the data exported by the completed models, Microsoft Excel files containing the line average calculations of models of specific parameters, and visual figures produced by MATLAB (The Mathworks, Natick, MA, USA). This section will explain the organization of the files as well as the contents of each file set.

3.1. COMSOL Base Models

The dataset presented in the supplementary file was found using a FEM computational model [19]. The model simulates the incident light upon the surface of the entire structure, and then calculates the plasmonic electric field enhancement upon the light. By using the mesh technique, it approximates the electric field in the free space surrounding the model.

To calculate the electric fields in this model, the frequency domain partial differential equation describing the electric field of a wave propagating in a medium was used:

$$\nabla \times {\mu_r}^{-1} (\nabla \times \boldsymbol{E}) = {k_0}^2 (\varepsilon_r - j\sigma/\omega\varepsilon_0)\boldsymbol{E}, \quad (1)$$

in which ∇ is the gradient operator, μ_r is the permeability constant for the material in which the electromagnetic wave is studied, $\boldsymbol{E}$ represents the electric field, k_0 is the wave number, ε_r is the permittivity constant of each material studied, j is the current density, σ is the material's electrical conductivity, ω is the wave's angular frequency, and ε_0 is the permittivity of free space.

To find the vector component $\boldsymbol{E}$, the wave equation is solved while a polarized, background field of

$$\boldsymbol{E}\,(x,z) = \boldsymbol{E}_0(x,z)e^{-ik_0 y}, \quad (2)$$

is applied throughout the space. As the model is time-independent, the fields are considered time-averaged.

The file includes the geometry design illustrated in Figure 1, and the file can also be modified for new geometries as well. COMSOL v5.3a is needed to open the model files attached to the dataset, but also the data results are exported in text files if one does not have access to COMSOL.

3.2. Exported Data Text Files

The text files hold the data produced by the model simulation results. They describe the electric field, both real and imaginary, for the substrate model and calculated in each direction at different mesh points. They are separated by folders denoting a specific parameter which is constant for all files in the folder. For example, in the folder *Fig 1 Brawley Materials*, there is the specific folder labelled *s10_t330*. This designates that all text files within the folder have a constant s (spacing between nanowires) value of 10 nm and t_{SiO2} (thickness of the SiO_2) value of 330 nm. Within this folder are three text files with names similarly formatted; however, the categorizing of the names is not necessarily in the same order. Consider *fig_1b_w50*, for example. This file gives the electric field with both the real and imaginary parts for each direction of x, y, and z for a fixed nanowire width of w = 50 nm, s = 10 nm, and t_{SiO2} = 330 nm.

In these files, the formatting of the text is dictated by the software's export feature. In this format, the "%" symbols denote descriptor lines for the data given. The first line describes the name of the base model used to create the data, but this line of the file does not necessarily correspond with the exact figure being addressed. The second line describes the software version COMSOL 5.3.1.229. The third line gives the date on which the exported file was created. The fourth line describes the number of dimensions the model is simulating, which is always two, as the study is a two-dimensional analysis. The fifth line shows the number of nodes, or mesh points, being calculated by the model. The sixth line describes the number of expressions determined by the model, which is always 6 due to the calculation of both the real and imaginary parts of the electric field for all the x, y, and z-directions. The seventh line gives a short description of the columns below it, beginning in the 10th row, starting with the third column. The eighth line gives the units used for the x and y coordinates, and in all cases should be in units of nm. The ninth line quantitatively describes all the columns below it. It begins with x and y, describing the coordinates of each mesh point. For example, *fig_1b_w50* begins with many values for the x-coordinate being $x = -30$, for which it runs through the y-values for all mesh points that have an x-coordinate of -30. Due to the previously mentioned formatting from the export feature, it additionally lists the electrode spacing, s, the wire width, w, and the wave frequency for each individual column. The wire spacing, and width, are consistent within each individual file, while the frequency of the light is constant through all files at 3.189×10^{14} Hz due to our light wavelength being 785 nm. Thus, the third and fourth columns describe the real and imaginary values of the electric field, respectively, in the x-direction. The fifth and sixth columns describe the real and imaginary values of the electric field, respectively, in the y-direction, while the seventh and eighth columns describe the real and imaginary values, respectively, of the values of the electric field in the z-direction. Table 1 gives a sample of the file arrangement beginning at row 10 for file *fig_1b_w50*.

Table 1. Data for the file *fig_1b_w50* as it appears in the rows beginning on line 10. Please note that values in the figure are restricted to first four decimals for brevity.

x	y	Real (E_x)	Imag (E_x)	Real (E_y)	Imag (E_y)	Real (E_z)	Imag (E_z)
−30	−530	−0.0799	0.3404	-6.6255×10^{-4}	0.0017	0	0
−30	−511.8181	−0.2425	0.2516	-2.5247×10^{-4}	1.9505×10^{-4}	-1.9377×10^{-18}	-2.4711×10^{-17}
−30	−494.4976	−0.3350	0.1007	1.9317×10^{-5}	-7.0334×10^{-5}	-3.2628×10^{-18}	-4.1869×10^{-17}

3.3. Excel Files

The Excel files are in the folders pertaining to their corresponding figures and labelled with the title of the figure represented by each. For example, in the folder *Fig 3 Brawley Materials* is the Excel file *Fig3a* regarding Figure 3a in [19]. As with the text files, the beginning lines are reserved for

descriptions of the data within the file and are also denoted by "%" symbols at the beginning of each line. The first line describes the name of the model file used to generate the data. The second line lists the version of the software used for this specific file and in all cases is COMSOL 5.3.0.316 [20]. The third line lists the date of the file export. The fourth line gives the title of the table within a given model from which the data is being exported. Finally, the fifth line gives the title for each column below it. The first column says ElectrodeSpacing which represents s, the second column says WireWidth1, the w value, and the third column describes the formula for the average optical field enhancement emw.normE^2, being measured across the entire surface for each set of parameters. Beginning in the sixth row, each column is specific to the title in the fifth row. Each number in the first and second columns gives the values for s and w, respectively. The third column gives the value for the surface average electric field enhancement across the entire model. As s and w are varied from 10 to 600, this gives line average values for 3600 combinations of s and w parameters, per t_{SiO2} value.

3.4. Matlab Images

Lastly are the MATLAB image files which end with ".fig." These files are labelled like the Excel files, in that they are titled directly after their corresponding figure. For example, in folder *Fig 3 Brawley Materials*, the MATLAB figure file Figure 3a matches Figure 3a in [19]. These files are created from the data used to create the corresponding figure in [19]; however, their coloration has been changed and matches the figures present here.

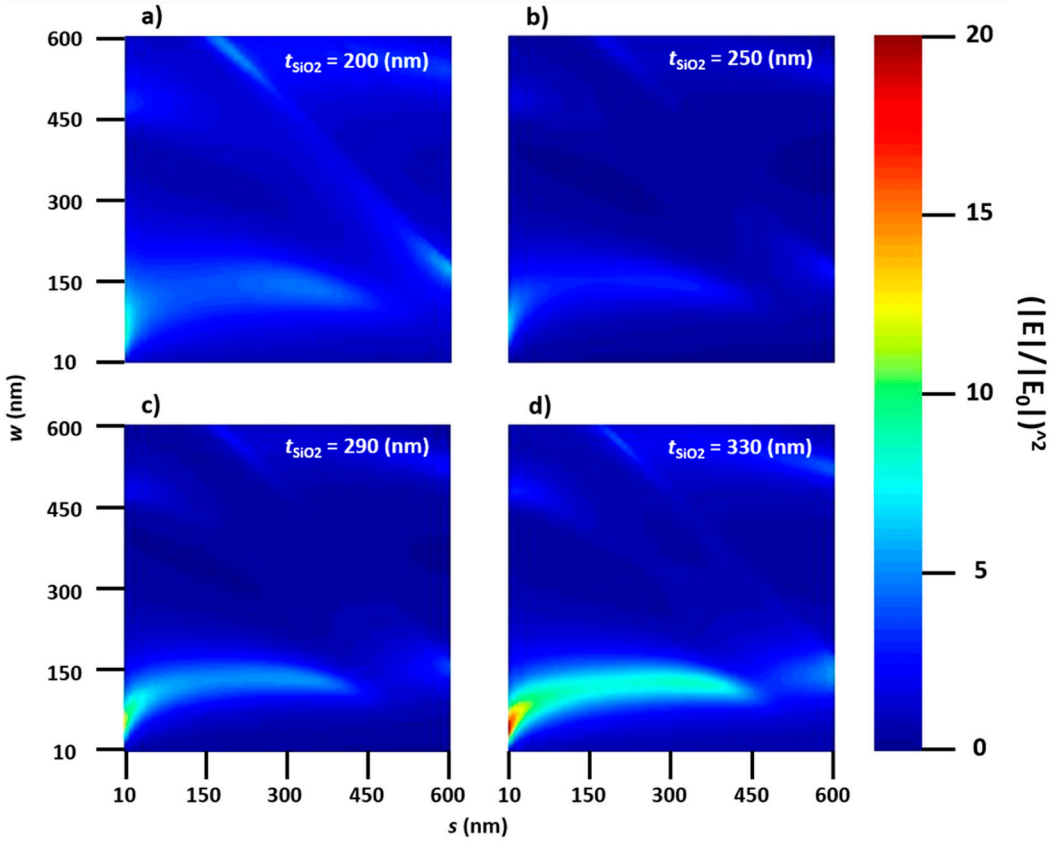


Figure 3. Graphical renderings of average optical surface enhancement of *s* against *w* for varying t_{SiO2} values: (**a**) t_{SiO2} = 200 nm; (**b**) t_{SiO2} = 250 nm; (**c**) t_{SiO2} =290 nm; and (**d**) t_{SiO2} = 330 nm.

4. Conclusions

The data presented was produced in [19], and the findings regarding possible SERS improvements are discussed. The data was generated using the FEM via COMSOL Multiphysics modeling software.

The data specifically chosen to be published correlates with the figures originally published in [19] and has related figures in this paper as well.

There are many potential benefits to SERS in plasmonics; however, most studies have been concerned with distances between plasmonic structures below 100 nm. In [19], we see that there are modes of plasmonic interference at some widths and spacings significantly greater than 100 nm that also induce strong electric fields. While not as strong as optimal widths and spacings, they can provide optical enhancement up to 11 times that of a similar surface with no plasmonic enhancement [19]. Due to the difficult nature of fabricating structures on the nanometer scale, these findings could be valuable to fabrication techniques in which it is easier to make larger structures.

The intent of this paper is to increase the reproducibility of the work achieved in [19]. We hope to elucidate our methods and computational models to create a framework which other computational plasmonics researchers can follow to build on or validate our work. The data is published in a text file format to make it accessible to researchers who may not have access to or experience with the COMSOL Multiphysics software package. The base COMSOL model is also made available for other work that may want to use and modify them for new geometries and studies. We hope that other computational methods, such as FDTD, can produce similar results and thus encourage more research in the field of plasmonic SERS substrates.

Supplementary Materials: The following are available online at: https://osf.io/6vb2w/?view_only=7629f06f081541968dbe986951167c4e (the dataset accompanying this paper will be available on the website).

Author Contributions: J.B.H. and C.M.K. decided to share data via open access, C.M.K. exported and organized data from model simulations, Z.T.B., J.B.H., S.J.B., and A.A.D. created the models, Z.T.B., S.J.B., and A.A.D. completed the experiments, C.M.K., Z.T.B., and D.T.D., analyzed the data, J.B.H. performed supervision, C.M.K. created original manuscript, J.B.K., Z.T.B., and S.J.B. provided revisions for manuscript.

Funding: This research was funded by NSF/EEC grant number 1757979, as well as partly by the Arkansas Biosciences Institute. S.J. Bauman received funding from the Doctoral Academy Fellowship via the University of Arkansas Graduate School in addition to the SPIE Optics and Photonics Education Scholarship. A. Darweesh had funding due to the Iraqi Ministry of Higher Education and Scientific Research (MoHESR). The University of Arkansas through the Department of Physics, the Fulbright College of Arts and Sciences, and the Office of Research & Innovation also provided funding and general research assets.

Acknowledgments: The Micro Electronics and Photonics Research Experience for Undergraduates, as well as the NSF, deserve acknowledgment for the support and opportunity that made this possible.

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

References

1. Wang, A.X.; Kong, X. Review of recent progress of plasmonic materials and nano-structures for surface-enhanced Raman scattering. *Materials* **2015**, *8*, 3024–3052. [CrossRef] [PubMed]
2. Kneipp, J.; Kneipp, H.; Kneipp, K. SERS—A single-molecule and nanoscale tool for bioanalytics. *Chem. Soc. Rev.* **2008**, *37*, 1052–1060. [CrossRef] [PubMed]
3. Long, J.; Yang, T. Observation of single molecule dynamic behaviors with SERS: Desorption and conformation switching. In Proceedings of the Conference on Lasers and Electro-Optics (CLEO), San Jose, CA, USA, 5–10 June 2016; pp. 1–2.
4. Sharma, B.; Frontiera, R.R.; Henry, A.-I.; Ringe, E.; van Duyne, R.P. SERS: Materials, applications, and the future. *Mater. Today* **2012**, *15*, 16–25. [CrossRef]
5. Vo-Dinh, T. SERS chemical sensors and biosensors: New tools for environmental and biological analysis. *Sens. Actuators B Chem.* **1995**, *29*, 183–189. [CrossRef]
6. Motz, J.T.; Hunter, M.; Galindo, L.H.; Gardecki, J.A.; Kramer, J.R.; Dasari, R.R.; Feld, M.S. Optical fiber probe for biomedical Raman spectroscopy. *Appl. Opt.* **2004**, *43*, 542–554. [CrossRef] [PubMed]
7. Faulds, K.; Smith, W.E.; Graham, D.; Lacey, R.J. Assessment of silver and gold substrates for the detection of amphetamine sulfate by surface enhanced Raman scattering (SERS). *Analyst* **2002**, *127*, 282–286. [CrossRef] [PubMed]

8. Barbillon, G.; Sandana, V.E.; Humbert, C.; Bélier, B.; Rogers, D.J.; Teherani, F.H.; Bove, P.; McClintock, R.; Razeghi, M. Study of Au coated ZnO nanoarrays for surface enhanced Raman scattering chemical sensing. *J. Mater. Chem. C* **2017**, *5*, 3528–3535. [CrossRef]
9. Gramotnev, D.K.; Bozhevolnyi, S.I. Plasmonics beyond the diffraction limit. *Nat. Photonics* **2010**, *4*, 83–91. [CrossRef]
10. Schuller, J.A.; Barnard, E.S.; Cai, W.; Jun, Y.C.; White, J.S.; Brongersma, M.L. Plasmonics for extreme light concentration and manipulation. *Nat. Mater.* **2010**, *9*, 193–204. [CrossRef] [PubMed]
11. Dai, D.; He, S. A silicon-based hybrid plasmonic waveguide with a metal cap for a nano-scale light confinement. *Opt. Express* **2009**, *17*, 16646–16653. [CrossRef] [PubMed]
12. Lin, Q.-Y.; Li, Z.; Brown, K.A.; O'Brien, M.N.; Ross, M.B.; Zhou, Y.; Butun, S.; Chen, P.-C.; Schatz, G.C.; Dravid, V.P.; et al. Strong coupling between plasmonic gap modes and photonic lattice modes in DNA-assembled gold nanocube arrays. *Nano Lett.* **2015**, *15*, 4699–4703. [CrossRef] [PubMed]
13. Kalachyova, Y.; Mares, D.; Lyutakov, O.; Kostejn, M.; Lapcak, L.; Švorčík, V. Surface Plasmon Polaritons on Silver Gratings for Optimal SERS Response. *J. Phys. Chem. C* **2015**, *119*, 9506–9512. [CrossRef]
14. Ding, S.; You, E.-M.; Moskovits, M. Electromagnetic theories of surface-enhanced Raman spectroscopy. *Chem. Soc. Rev.* **2017**, *46*, 4042–4076. [CrossRef] [PubMed]
15. Zhao, D.; Pinchuk, A.O.; McMahon, J.; Li, S.; Ausman, L.; Atkinson, A.; Schatz, G. Methods for Describing the Electromagnetic Properties of Silver and Gold Nanoparticles. *Acc. Chem. Res.* **2008**, *41*, 1710–1720. [CrossRef] [PubMed]
16. Bohren, C.F.; Huffman, D.R. *Absorption and Scattering of Light by Small Particles*; Wiley: Hoboken, NJ, USA, 1983.
17. Bauman, S.J.; Brawley, Z.T.; Darweesh, A.A.; Herzog, J.B. Substrate oxide layer thickness optimization for a dual-width plasmonic grating for surface-enhanced Raman spectroscopy (SERS) biosensor applications. *Sensors* **2017**, *17*, 1530. [CrossRef] [PubMed]
18. Lyvers, D.P.; Moon, J.-M.; Kildishev, A.V.; Shalaev, V.M.; Wei, A. Gold nanorod arrays as plasmonic cavity resonators. *ACS Nano* **2008**, *2*, 2569–2576. [CrossRef] [PubMed]
19. Brawley, Z.T.; Bauman, S.J.; Darweesh, A.A.; Debu, D.T.; Tork Ladani, F.; Herzog, J.B. Plasmonic Au Array SERS Substrate with Optimized Thin Film Oxide Substrate Layer. *Materials* **2018**, *11*, 942. [CrossRef] [PubMed]
20. COMSOL, Inc. COMSOL Multiphysics Version 5.3a: RF Module. Available online: https://www.comsol.com/ (accessed on 25 July 2018).
21. Brawley, Z.T.; Bauman, S.J.; Abbey, G.P.; Darweesh, A.A.; Nusir, A.I.; Manasreh, O.; Herzog, J.B. Modeling and optimization of Au-GaAs plasmonic nanoslit array structures for enhanced near-infrared photodetector applications. *J. Nanophotonics* **2017**, *11*, 016017. [CrossRef]
22. Hatab, N.A.; Hsueh, C.-H.; Gaddis, A.L.; Retterer, S.T.; Li, J.-H.; Eres, G.; Zhang, Z.; Gu, B. Free-standing optical gold bowtie nanoantenna with variable gap size for enhanced Raman spectroscopy. *Nano Lett.* **2010**, *10*, 4952–4955. [CrossRef] [PubMed]
23. Johnson, P.B.; Christy, R.W. Optical constants of the noble metals. *Phys. Rev. B* **1972**, *6*, 4370–4379. [CrossRef]
24. Green, M.A.; Keevers, M.J. Optical properties of intrinsic silicon at 300 K. *Prog. Photovolt. Res. Appl.* **2007**, *3*, 189–192. [CrossRef]
25. Gao, L.; Lemarchand, F.; Lequime, M. Refractive index determination of SiO_2 layer in the UV/Vis/NIR range: Spectrophotometric reverse engineering on single and bi-layer designs. *J. Eur. Opt. Soc. Rapid Publ.* **2013**, *8*, 13010. [CrossRef]

data

MDPI

Data Descriptor

A Data Set of Portuguese Traditional Recipes Based on Published Cookery Books

Alexandra Soveral Dias and Luís Silva Dias *

Department of Biology, University of Évora, Ap. 94, 7000-554 Évora, Portugal; alxandra@uevora.pt
* Correspondence: lsdias@uevora.pt; Tel.: +351-266-760-881

Received: 27 December 2017; Accepted: 5 March 2018; Published: 8 March 2018

Abstract: This paper presents a data set resulting from the abstraction of books of traditional recipes for Portuguese cuisine. Only starters, main courses, side dishes, and soups were considered. Desserts, cakes, sweets, puddings, and pastries were not included. Recipes were characterized by the province and ingredients regardless of quantities or preparation. An exploratory characterization of recipes and ingredients is presented. Results show that Portuguese traditional recipes organize differently among the eleven provinces considered, setting up the basis for more detailed analyses of the 1382 recipes and 421 ingredients inventoried.

Dataset: available as www.mdpi.com/2306-5729/3/1/9/s1.

Keywords: food; Portugal; traditional recipes

1. Summary

Eating is one of the obligatory activities of human beings, but unlike many other activities, eating usually requires active and deliberate effort. In humans, this occurs hopefully once or twice a day, every day during our lifetime, requiring a more or less complex, extended, and permanent web of activities starting with production, through storage, transportation, delivery, and transformation, before foods are consumed.

In addition to being an obligatory activity, eating is also a pleasure, especially as we distance from survival alone. As stated almost two centuries ago by Brillat-Savarin in the seventh introductory aphorism to his *magnus opus*, "*the pleasure of the table is enjoyed at all ages, conditions, countries and days, can be associated with all other pleasures and is the last we enjoy, giving us solace for the loss of all others*" [1] (p. ix).

However, cultural habits and preferences create food tastes and choices as well as inhibitions and aversions, complicating the introduction of new foods and composite dishes while simultaneously, reinforcing the use of those already established [2] (p. 481). A good example of the former are the words of caution and simultaneously of seduction that an enthusiastic convert to a new cuisine offer to less knowledgeable and less fortunate [3] (pp. 16–18). An example of the latter is the persistence of recipes transmitted by word-of-mouth that are, by example and imitation, at the core of the so-called traditional cuisine. In this data set, we report the less-well-known traditional cuisine of Portugal.

Portugal is a relatively small country, located in the western-most part of the Iberian Peninsula which comprises also two Atlantic archipelagoes, Madeira and the Azores islands. In all, it has slightly more than 92,000 km^2 (89,000 km^2 in mainland with maximum N–S length of 577 km and W–E length of 286 km, plus about 801 km^2 in Madeira and 2322 km^2 in the Azores). According to the 2011 census, the resident population is about 10,000,000 inhabitants in the mainland plus about 500,000 in the two islands, slightly more in Madeira than in Azores [4]. However, in a number of aspects, Portugal has little internal coherence and homogeneity for its size; there is a possible division of mainland

Portugal (the islands being separate cases) between the more mountainous North and the flatter South. Alternately, between the Atlantic Northwest and the more continental Northeast, both contrasting with the Mediterranean-type South, with Atlantic and Mediterranean inroads in the South and North, respectively. Nevertheless, these climatic and orographic contrasts have been counteracted by long, intense, and homogenizing human activities [5] (pp. 144–167).

This heterogeneity was noted with admiration nearly two centuries ago, especially in relation to the climate and vegetation [6] (pp. 89–99, 140–143). Later, it was stated without exaggeration that "*few countries of its size present such high level of natural differences, so that someone instantly transported from the middle of Minho* (the northern half of Entre Douro e Minho; Figure 1) *to the middle of Alentejo would think to have travelled not about 80 leagues* (around 325 km) *that separates them but thousands of them*" [7] (p. xi). For a short but insightful presentation of the strong territorial heterogeneity, see also [8] (pp. 1–4).

Figure 1. Approximate limits of the Portuguese division in provinces adopted. Insets are of the Atlantic islands of Madeira and Azores.

Altogether, this patchwork of climates and landscapes explains the various attempts to define and delimit provinces as administrative surrogates of the so-called natural regions of Portugal. The last such attempt was in the 1930s. In 1933's Constitution, provinces were considered the upper level in the territorial and administrative organization of Portugal, their exact delimitation being done only in 1936 [9], largely based upon Girão's proposals [10]. However, only eleven provinces were considered in the mainland, instead of the proposed thirteen. In 1959, provinces were lawfully extinguished. Nevertheless, they have been recognized and used in everyday life until today and are—some more than others—a very important part of the territorial identity of Portuguese people.

We hypothesized that provinces—being climatic and environmentally natural regions that became identity-generating entities—would somehow also correspond to separate and recognizable sub-cuisines with identifiable and segregating traits. For example, in the diachronic use of herbs, spices, and other condiments; in the way alien ingredients disseminated throughout Portuguese cuisine or, in a more encompassing framework, the reproducibility at the small but highly variable scale of the so-called "Darwinian gastronomy" hypothesis put forward to explain the worldwide pattern of the

use of spices [11,12]. As a prerequisite to investigate these hypotheses, we set out to create a data set of recipes described by their ingredients and the province where they were traditionally cooked and eaten. For strictly practical reasons, derived from the provincial organization adopted in the cookery books, we further reduced the number of provinces, merging Minho and Douro Litoral under the more ancient designation of Entre Douro e Minho. Alto and Baixo Alentejo were also merged under the umbrella designation of Alentejo (Figure 1).

2. Data Description

Sources for the data set were recipe books explicitly dealing with or aimed at presenting Portuguese traditional cooking. To be included in the data set, the origin of recipes had to be known. Therefore, to be selected and abstracted, cookery books had to allow (1) all or at least most of the recipes to be assigned to any one of the provinces shown in Figure 1, and (2) no province was omitted, deliberately or otherwise. As a consequence, some published materials could not be used, namely the classic Olleboma's book first published in 1936 [13] or the near encyclopedic series of Portuguese cookery books authored or co-authored by A. Saramago, left incomplete because of his death in 2008 when only six provinces had been published [14–19].

Thus, six cookery books were used [20–26]. Dates of first publication ranged from 1981 to 2006, but in only two [25,26] was the first edition after 2000. Data were derived from the cookery books by one researcher and independently double-checked by another, thus ensuring that all decisions were reached by consensus. The greatest care was taken in building this inventory, but we cannot reject that errors might have gone unnoticed. If errors are found or suspected, please let us know so they can be corrected.

Data were organized in tabular form in an MS Excel® 2010 spreadsheet, which is available in the Supplementary Materials (Supplementary S1) as a non-proprietary comma-separated value (CSV) format file. Each line in the data set presents information for a single recipe in columnar form. The first column, headed "CODE", displays alphanumeric codes for recipes and is composed of a four-letter acronym of the province name and a three-digit identifier for the recipe within the province (AÇOR stands for Azores, ALEN for Alentejo, ALGA for Algarve, BALT for Beira Alta, BBAI for Beira Baixa, BLIT for Beira Litoral, EDMI for Entre Douro e Minho, ESTR for Estremadura, MADE for Madeira, RIBA for Ribatejo, and TMAD for Trás-os-Montes e Alto Douro). Recipes are displayed in alphabetical order by province and alphabetic order by recipe name in Portuguese, and if necessary by ascending order of number of ingredients. Precise location of the recipes is not shown because such information is mostly absent. Undetermined origin ranged from 32% in Ribatejo (second minimum of 38% in Beira Baixa) to 93% in Azores and Madeira (second maximum of 68% in Alentejo), with an average of 63% when the recipes were pooled together.

The second column, "RECIPE (Portuguese)", presents the name of the recipe in Portuguese, and the third, "RECIPE (English translation)", the translation of the name in English. Translation essentially followed the English edition of Modesto's cookery book published in 1989 [20,27]. The fourth and fifth columns ("REF" and "PAGE") provide reference numbers and pages where the full recipe can be found. Whenever the recipe was abstracted from Modesto's cookery book [20], two page numbers are displayed in "PAGE" separated by a slash. The first number refers to the Portuguese edition and the second to the English translation [27], unless the recipe occurs on the same page in the two editions. The full list of references used is presented at the end.

The remaining 421 columns display individual ingredients in the recipes in alphabetical order by Portuguese names (1 if present in the recipe, blank if absent). Each column is headed by an alphanumeric code composed of the letter "I" (for ingredient) and a three-digit identifier for the ingredient, followed by its name in Portuguese and its translation to English in parentheses, and whenever it existed by the EFSA FoodEx2 food code [28].

Translations of ingredient names to English relied heavily on the English edition of Modesto's cookery book [27], but other useful sources were also used [3,29,30]. Finally, some of the English

translations were reviewed by colleagues with in-depth expertise in specific terminologies, namely alcoholic beverages [31]; seafood and fresh water foods, including fish, shellfish, and mollusks [32]; and game, livestock, and poultry, including their parts [33]. For ease of use, we also provide an English–Portuguese glossary of ingredient names in the Supplementary Materials (Supplementary S2). It goes without saying that responsibility for any errors in the data set and the glossary are ours.

Despite our efforts, translation into English was not possible for a number of recipes, and especially certain ingredients, largely because such ingredients are not used in English-speaking culinary areas. In all, seventeen ingredients could not be translated, most being pork-based dry-cured sausages (53%) and wines (24%).

3. Methods

3.1. Ingredients

Salt was not included in the data set as an ingredient because it was always included in the recipe or, in relatively few cases where no mention of it was made, its presence is implicit through one or more ingredients, usually *bacalhau* (Atlantic cod, salted and dried; I035). In Portuguese cuisine, *bacalhau* is always the dry, highly-salted form that almost always has to be soaked in water for one or more days, the water being changed a couple of times so that most of the salt is removed. It is also worth remarking that *coentro* (coriander; I121) is widely consumed throughout the world in a variety of forms (mostly seeds) [34] but in Portuguese cuisine, only the leaves or sprigs composed of stem and leaves (preferably fresh) are used.

Decisions had to be taken in relation to ingredients that are mixtures of ingredients, which it is assumed everyone knows how to prepare without further instructions, specifically side dishes. In these cases, we considered a minimum set of components and included them in the recipe. Thus, potato puree was described as butter, milk, and potato (e.g., in *Cavalas recheadas*, Stuffed mackerel, AÇOR024 or *Coelho à caçadora*, Jugged rabbit, ESTR071), regardless of other ingredients that might be used, depending on the cook or preferences (e.g., white pepper or nutmeg). *Esparregado*, etymologically related to asparagus but usually done using spinach or other vegetables, traditionally contains bayleaf, flour, garlic, and olive oil (e.g., in *Galinholas à alentejana*, Alentejo style woodcocks, ALEN104). Also, *arroz branco* (white rice) is assumed to require only butter and rice (e.g., in *Lampreia*, Lamprey, BBAI045 or *Tripas à moda do Porto*, Oporto style tripes, EDMI175); ingredients adopted for white rice agreed with the description in one recipe from Madeira (*Atum assado*, Grilled tuna, MADE005). Finally, fried potatoes (chips, French fries) were listed as potatoes and oil for frying.

Conversely, dishes not prepared "in the moment" were taken as a single ingredient, regardless of their composition (e.g., *enchidos*, dry-cured sausages, or *molho inglês*, Worcester sauce, I252).

In a few cases, linguistic differences in Portuguese had to be accounted for, the most important involving *segurelha* (savory, I384), which is an ingredient in *Feijão verde à alentejana* (Alentejo style string beans, ALEN097). The name *segurelha* is also used in the Madeira Islands, except that in Madeira it refers to thyme [20] (p. 296), [27] (p. 314), [35] (p. 283). Another case was *pimenta da Jamaica* (allspice, I018), which is only used in recipes from the Azores and Madeira. In the latter, it is almost never referred to as *pimenta da Jamaica* but as *alcepás* (more rarely as *acepás*), clearly a corruption of its name in English.

In Portuguese, true saffron (*Crocus sativus* L.) is *açafrão*, but this term is also applied to the much less expensive turmeric (*Curcuma longa* L.) and safflower (*Carthamus tinctorius* L.). In addition to *açafrão*, the names *açaflor* and *açafroa* are also mentioned, the latter representing—in Portuguese—the grammatical feminine of *açafrão*. Throughout the data set, *açaflor* or *açafroa* are referred to as ingredients twice, in AÇOR019 and AÇOR038, respectively, while *açafrão* is referred to 26 times. In two of them, it is explicitly stated that true saffron should be used (EDMI009 and EDMI064). In the other 24, no clarification is made and we assumed it meant turmeric, less frequently safflower, but never saffron.

Throughout the cookery books, there is reference to several peppers, namely *pimenta preta* (black pepper, I314), *pimenta rosa* (rose pepper, I315), *pimenta verde* (green pepper, I316), *pimenta vermelha* (red pepper, I317), and *pimenta da Jamaica*, already mentioned. Singly or jointly, they are present in 54 recipes. Additionally, in about 780 recipes, there is mention of *pimenta* without any other quality attached; we assumed this always meant white pepper (I313).

3.2. Recipes

Recipes in the data set include starters, main courses, side dishes, and soups. Desserts, cakes, sweets, puddings, and pastries were not considered. Recipes are described by the ingredients they require in terms of presence or absence, completely disregarding quantities because presence might be expected to be a relatively stable characteristic—at least more stable than quantities, which strongly depend either on the availability of ingredients or on the taste and likings of those for whom meals are prepared. On the importance of availability and taste in this framework, see for example [36] (pp. 48–49).

Traditional cuisine is supposed, at the very least, to involve a large number of recipes eaten in individual households where if someone will dislike a given ingredient, vinegar for example, it is reduced or eliminated; the opposite also being true where an ingredient is favored. Obviously, one ingredient may be replaced with another (e.g., vinegar by lemon), which is taken to be more or less equivalent. We have seen this happen with, for example, turnips being replaced by potatoes in *Coelho à Capitão-Mor* (Rabbit "à Capitão-Mor", BALT036). This may help explain the frequency of instructions like "*use this or that ingredient*" or "*such and such ingredient can also be used*". An example of the former can be found in *Caldo-verde à Minhota* (Minho Style shredded cabbage soup, EDMI077) which indicates the use of *salpicão* (I359) or *chouriço* (I116). Of the latter, in *Cachola de porco* (Pork "cachola", ESTR046), it is stated that *água-pé* (I102) can be replaced by ordinary white wine although *Cachola* apparently "*tastes infinitely better with água-pé*" [20] (p. 209), [27] (p. 222). Whenever such instructions appeared, we retained only the first ingredient (i.e., we only listed *salpicão*) and disregarded ingredients that "*could also be used*" (i.e., we only listed *água-pé*), except when these were explicitly considered as a separate variant of the recipe.

When we finished abstracting the six cookery books [20–26], the data set totaled 1644 recipes and a precautionary check on repeated recipes was performed. Pairwise comparisons between each recipe and all others was done, and a likeness-value (*LI*) calculated as:

$$LI = N_{A,B}/\max(N_A,N_B), \quad (1)$$

where $N_{A,B}$ is the number of ingredients present simultaneously in recipes A and B, and $\max(N_A,N_B)$ is the number of ingredients in the recipe with the greater number (or the number of ingredients of recipe A or B if $N_A = N_B$), with *LI* ranging from 0 to 1. When $LI = 0$, there were no common ingredients between recipes. When $LI = 1$, recipes have the same number of ingredients and all ingredients of one recipe occur in the other. $LI = 0.5$ was obtained whenever the number of ingredients common to two recipes was half the maximum number of ingredients of the richest recipe. However, recipes could have $LI = 1$ and still be different, because of the way the dish is composed and cooked. Therefore, we checked ways of cooking for every pair of recipes with $LI = 1$; the result being that 262 where true repetitions, for the most part recipes from Alentejo (40), Algarve (34), Trás-os-Montes e Alto Douro (30), and Entre Douro e Minho (28). In addition, 19 recipes were found to have $LI = 1$ when compared with another, but were cooked in different ways, most involving recipes from Algarve (8) and Trás-os-Montes e Alto Douro (4). True repetitions were eliminated from the data set, which was reduced to 1382 recipes. Entries that were eliminated belonged to later publications.

3.3. Evaluating Bias in Cookery Books Examined

As explained above, the data set was based on six cookery books by different authors, which had different aims and methods of collecting and selecting recipes. Therefore, it is conceivable that authors—and thus cookery books and recipes—might be biased. The impact of putative biases is naturally dependent on the share each cookery book made to the total number of recipes.

Three books [20,24,25] out of six provided 85% of recipes, and always more than 80% of recipes in each province, with the exception of the Azores, and were selected to investigate for bias in their portrait of Portuguese traditional cuisine. The underlying rationale for the method used was that if bias exists, then from the occurrence of one or a combination of ingredients in a recipe it would be possible to predict the book from where that recipe came. We adopted a non-parametric tree-structured classification method [37,38], which is preferable to other methods such as discriminant analysis or logistic regression because of its robustness with respect to outliers and misclassified points, its applicability to any data structure, and its automatic stepwise selection and complexity reduction. In addition, it easily accommodated interactions between variables without prior selection of variables and gave estimates for incorrect identifications. Finally, it results in decision rules which are easy to understand and apply [37] (pp. 56–58).

Binary trees were generated using SPAD Data Mining & Text Mining, v. 6.5.0 (SPAD, Paris, France). Only recipes from the three books [20,24,25] were used. The relative cost of misidentification was constant and unitary, and 25 independent runs were done, randomly assigning 66% of recipes to the learning group and 33% to the validation group. Overall, misidentification of cookery books in the validation group was the major criterion for selection.

Optimal partitions were always obtained, and 21 out of 25 independent runs resulted in the same binary tree with only two ingredients: white pepper and *manteiga* (butter, I231). The same ingredients plus *safio/congro* (conger, I355) and *vinho branco* (white wine, I412) were present in two additional trees. The same four ingredients plus *azeite* (olive oil, I032) were present in two additional trees. Misidentification by the shorter tree ranged from 43.7% to 44.7% in the validation group, with recipes from Modesto [20] being generally better identified (between 93.0% and 94.1% of correct identification). Recipes from Guedes [24] were generally misidentified (31.2–36.0% of correct identification), while recipes from Valente [25] were always misidentified. Because longer trees did not noticeably reduce the percentage of misidentification, we kept the shorter and most frequent binary tree. Recipes were identified as belonging to Modesto [20] either when they lacked white pepper or when they had white pepper but lacked butter. Recipes were identified as belonging to Guedes [24] only when they had white pepper and butter. Recipes were never identified as belonging to Valente [25]. Therefore, some minor bias around white pepper and butter might be present, and it may be wise to exert some caution as to whether or not the conclusions of analyses appear to be dependent on these ingredients. Possible courses of action might be to perform analyses with and without white pepper and butter, or the two ingredients included as supplementary to evaluate the stability of conclusions.

3.4. Exploratory Characterization of Recipes and Ingredients

All statistics in this section were done with Statgraphics 4.2 (STSC, Inc., Rockville, MD, USA) or with MS Excel® 2010. We adopted a type I error rate $\alpha = 0.001$ as a threshold reference for strong evidence against the null hypothesis [39]. In general, data are presented as mean $\pm$ standard error (SE) and sample size (n).

3.4.1. Recipes

Alentejo clearly topped the rank for numbers of recipes with 209, followed by Trás-os-Montes e Alto Douro, Estremadura, and Entre Douro e Minho with 186, 182, and 180 recipes, respectively. These provinces were well above the average number of recipes, which was approximately 126. Additionally, with more than the average number of recipes was Algarve, with 147 recipes. Conversely, the other six

provinces (Azores, the three Beiras, Madeira, and Ribatejo) had totals well below the average, ranging from 70 to 89 (Figure 2).

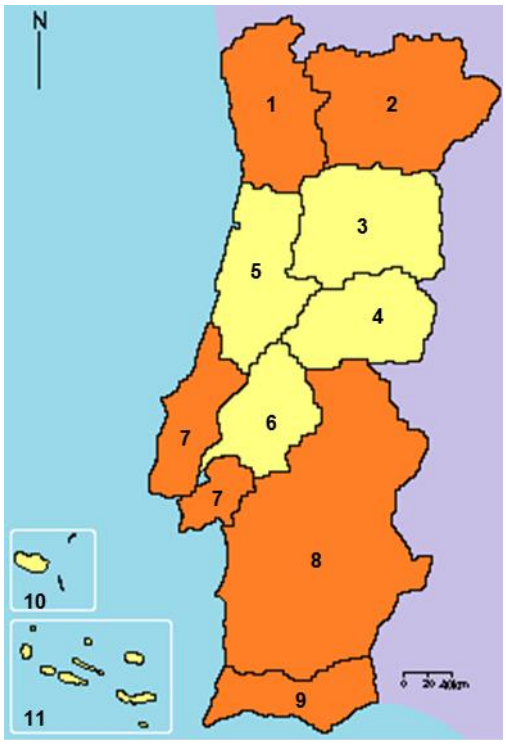


Figure 2. Over-representation (orange) and under-representation (light yellow) of number of recipes in provinces of Portugal. 1 Entre Douro e Minho, 2 Trás-os-Montes e Alto Douro, 3 Beira Alta, 4 Beira Baixa, 5 Beira Litoral, 6 Ribatejo, 7 Estremadura, 8 Alentejo, 9 Algarve, 10 Madeira, 11 Azores.

Considering all 1382 recipes, the number of ingredients per recipe ranged from one in two recipes (*Sável fumado*, Smoked shad, EDMI158; *Cracas*, Barnacles, MADE047), the ingredient being allis shad or barnacles, to 23 also in two recipes (*Cabrito com arroz à moda de Monção*, Monção style goatling with rice, BBAI019, and *Feijoada à transmontana*, Trás-os-Montes Style bean stew, TMAD108). The mean number of ingredients was (mean ± SE and sample size n) 9.5 ± 0.1, n = 1382 (median of 9), which is larger than the mean number of ingredients per recipe (nine ingredients including salt) found in combined worldwide inventories of recipes [40]. The coefficient of skewness g_1 = 0.647 was highly significant ($p = 7.9 \times 10^{-23}$, two-tailed test), meaning the distribution of numbers of ingredients was skewed to the right. When skewness was tested separately for each province, there was strong evidence of highly skewed distributions for the numbers of ingredients to the right only in the two northernmost provinces of Entre Douro e Minho and Trás-os-Montes e Alto Douro (Figure 1) which had g_1-values of 0.721 and 0.826, and p-values 6.9×10^{-5} and 3.5×10^{-6}, respectively.

The mean numbers of ingredients in each province was compared with the mean numbers of ingredients of all provinces to assess over-, average-, and under-representation. For this, we determined the probability of a t-Student value necessary for inclusion of the means for each province in the confidence interval of all 1382 Portuguese recipes. Significantly over-represented provinces (mean number of recipes from 10.0 ± 0.4, n = 70 to 10.4 ± 0.3, n = 182, one-tailed p-values less than 3.2×10^{-8}), by increasing order of means, were Madeira, Entre Douro e Minho, Azores, and Estremadura. Significantly under-represented provinces (mean number of recipes from 9.0 ± 0.4, n = 82 to 7.8 ± 0.4, n = 79, one-tailed p-values less than 2.1×10^{-8}), by decreasing order of means, were Ribatejo, Algarve, Beira Baixa, and Beira Alta. Average representation (mean numbers of recipes from 9.2 ± 0.2, n = 209 to

9.6 ± 0.3, n = 86, one-tailed p-values from 0.006 to 0.424) was found for the remaining three provinces, Alentejo, Trás-os-Montes e Alto Douro, and Beira Litoral (Figure 3).

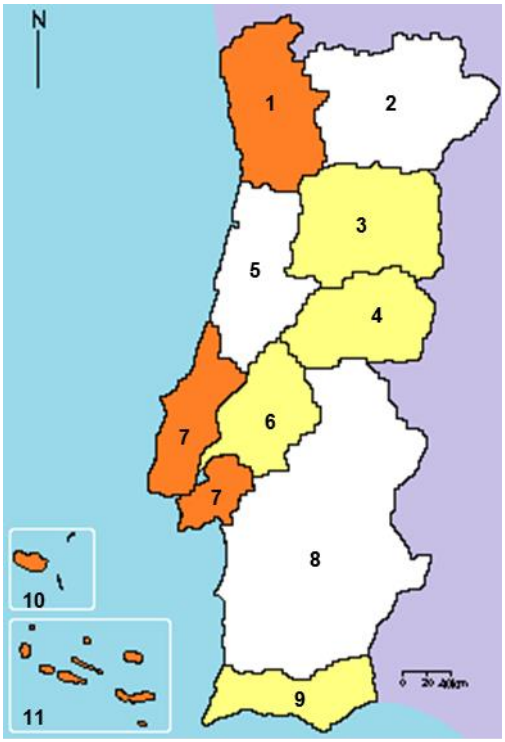


Figure 3. Over-representation (orange), average representation (white), and under-representation (light yellow) of mean number of ingredients of recipes in provinces of Portugal. 1 Entre Douro e Minho, 2 Trás-os-Montes e Alto Douro, 3 Beira Alta, 4 Beira Baixa, 5 Beira Litoral, 6 Ribatejo, 7 Estremadura, 8 Alentejo, 9 Algarve, 10 Madeira, 11 Azores.

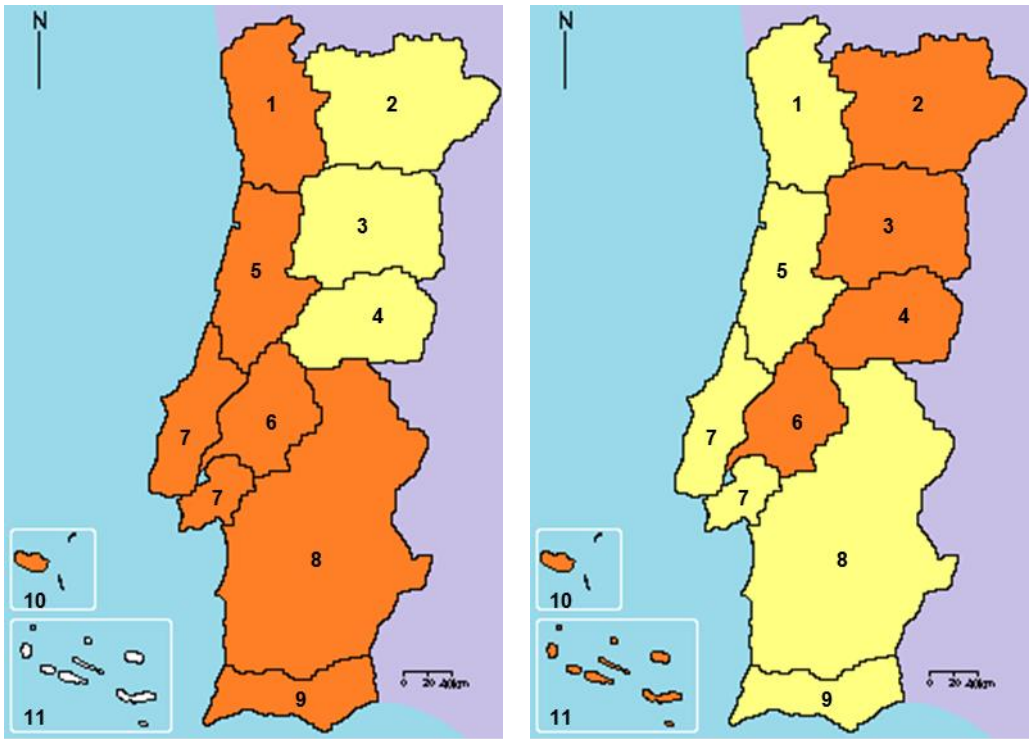
Comparisons like those described where parts (provinces) are compared with the whole to which they belong (country) are not completely independent, which might help explain the lack of clarity in patterns of differences among provinces. The only exception was the group of over-representation that included provinces that were either islands (Azores and Madeira) or included the largest and wealthiest cities of Portugal, Lisbon in Estremadura and Oporto in Entre Douro e Minho. An alternative approach might be to test the mean numbers of ingredients for all provinces. Simultaneous comparisons were done using a least squares linear regression approach with dummy variables to prevent the ambiguity resulting from lack of "transitivity", which frequently arises in simultaneous test procedures. For example, mean A is not significantly different from mean B, mean B is not significantly different from mean C either, but means A and C are significantly different [41,42].

Forward stepwise selection with replication was used, and the complete candidate models included only qualitative independent variables (the provinces), binary coded as 0, 1. An experiment-wise type I error rate was adopted for the coefficients of regression and calculated using the Dunn–Šidák method [43]. Coefficients of determination (R^2) are presented as proportions of the maximum R^2 possible [44] (p. 246). The significant heteroscedasticity detected in untransformed data using the two-tailed F distribution ($p = 5.1 \times 10^{-6}$) was strongly reduced (p = 0.002), which was deemed acceptable, when the numbers of ingredients were transformed using natural logarithms.

A significant three-term polynomial could be fitted to logarithmically-transformed data. Significance levels of the coefficients were $p \leq 0.020$; $R^2 = 0.673$; lack of fit $F_{8,1371} = 1.787$, corresponding to p = 0.075, and thereby, significant differences among provinces were found. After solving the equation for the values 0, 1 of the binary variables, provinces could be separated as three groups.

Group 1, was composed of recipes from Beira Alta only, comparatively poor in ingredients (7.8 ± 0.4, $n = 79$); group 2 was composed of recipes from Estremadura only, comparatively rich in ingredients (10.4 ± 0.3, $n = 182$); and group 3 was composed of recipes from the remaining nine provinces pooled together, with an intermediate number of ingredients per recipe (9.5 ± 0.1, $n = 1121$).

Complementary to the previous approach was the analysis of likeness-values (Equation (1)). Considering all 954,271 *LI*-values, the most notable feature was the low likeness of recipes. *LI*-values range from 0 to 1, with a mean of 0.218 ± 0.0001, $n = 954{,}271$, and a median of 0.2. Skewness $g_1 = 0.424$ was highly significant ($p \approx 0$, two-tailed test). A detailed analysis of *LI*-values is beyond the scope of this paper, but a few points are worthy of examination. Firstly, comparing mean *LI*-values within provinces (main diagonal elements in Table 1 below) with the mean *LI*-value of Portugal allowed the identification of three groups. The first, with under-represented *LI*-values, was composed of Beira Alta, Beira Baixa, and Trás-os-Montes e Alto Douro. The second, of near-average representation, was composed of the Azores. The third, over-representation, was composed of the remaining seven provinces (Figure 4a). *LI*-values within provinces might mirror the intensity of interactions and common use of recipes among inhabitants of each province. Alternatively, they might mirror the heterogeneity of environmental conditions, and thus of food supply (or both). Either way, increased interactions and common use, and decreased heterogeneity of the food supply, would likely increase *LI*-values. Therefore, frequency and distribution within and between provinces of complete unlikeness ($LI = 0$) is worth analyzing.



(**a**) (**b**)

Figure 4. Over-representation (orange), average representation (white), and under-representation (light yellow) of (**a**) mean likeness-values within provinces of Portugal, (**b**) frequency of recipes of each province involved in comparisons with likeness-values $LI = 0$ weighted by the number of recipes of the province. 1 Entre Douro e Minho, 2 Trás-os-Montes e Alto Douro, 3 Beira Alta, 4 Beira Baixa, 5 Beira Litoral, 6 Ribatejo, 7 Estremadura, 8 Alentejo, 9 Algarve, 10 Madeira, 11 Azores.

Table 1. Means of likeness-values (*LI*) between recipes of a province and recipes of the same (main diagonal) or other province. Numbers in bold indicate that *LI*-values within a province were larger than mean *LI*-values between that province and all others.

	Alentejo	**Algarve**	**Azores**	**Beira Alta**	**Beira Baixa**	**Beira Litoral**	**Entre Douro e Minho**	**Estremadura**	**Madeira**	**Ribatejo**	**Trás-os-Montes e Alto Douro**
Alentejo	**0.258**										
Algarve	0.242	**0.264**									
Azores	0.212	0.218	**0.220**								
Beira Alta	0.200	0.211	0.183	0.194							
Beira Baixa	0.208	0.210	0.180	0.184	0.197						
Beira Litoral	0.245	0.253	0.217	0.220	0.212	**0.271**					
Entre Douro e Minho	0.226	0.242	0.208	0.213	0.208	0.250	0.250				
Estremadura	0.223	0.244	0.207	0.200	0.200	0.242	0.234	0.243			
Madeira	0.207	0.226	0.206	0.187	0.188	0.220	0.213	0.219	**0.248**		
Ribatejo	0.229	0.234	0.200	0.197	0.198	0.239	0.227	0.221	0.204	0.229	
Trás-os-Montes e Alto Douro	0.200	0.210	0.183	0.191	0.186	0.212	0.216	0.200	0.186	0.195	0.200

About 12% of *LI*-values were null. However, their impact on the mean *LI*-value of Portugal was negligible, because when null *LI*-values were removed, the mean increased slightly from 0.218 ± 0.0001, n = 954,271 to 0.248 ± 0.0001, n = 841,726, the median increased from 0.200 to 0.231, and skewness g_1 = 0.703 remained highly significant ($p \approx 0$, two-tailed test). Because it is likely that provinces with more recipes also had more recipes involved in the comparisons, all frequencies were weighted against numbers of recipes. Considering all provinces pooled together, the frequency of comparisons with *LI* = 0 weighted by numbers of recipes involved was 162.9. Using this value as the reference to compare each province, two groups were identified. The group of over-representation involved in comparisons having *LI* = 0 comprised Azores, Beira Alta, Beira Baixa, Madeira, Ribatejo, and Trás-os-Montes e Alto Douro (weighted frequencies from 173.5 to 217.8). The group of under-representation comprised Alentejo, Algarve, Beira Litoral, Entre Douro e Minho, and Estremadura (weighted frequencies from 129.9 to 155.0), which is almost the opposite of the distribution examined above (Figure 4b) with Madeira and Ribatejo moving between groups.

Despite the need for further investigation, our analyses suggests that *LI*-values might relate to characteristics of provinces, regardless of what these may be, but such characteristics support a hypothesis of individuality. This would imply that the more "unique" a province is (cookery-wise), the larger the likeness will be among its recipes, and the smaller these values will be when compared with recipes from other provinces. Data summarized in Table 1 provide the first insight into this hypothesis.

Alentejo, Algarve, Azores, Beira Litoral, and Madeira seemed to fit in a group of "unique" and highly distinctive provinces. Thus, besides Beira Litoral (which ranks second in relation to Alentejo and Algarve), this group includes the Atlantic islands and Southern Portugal. Entre Douro e Minho and Estremadura seem to be as alike as they are alike to Beira Litoral and Algarve, respectively. Conversely, the remaining provinces are more alike other provinces than among themselves. Beira Alta, Beira Baixa, and Ribatejo are more alike with Beira Litoral (an apparent attractor); Trás-os-Montes e Alto Douro is more alike to neighboring Entre Douro e Minho followed by Beira Litoral.

3.4.2. Use of Ingredients

As seen in Section 2, 421 different ingredients were found in 1382 recipes. The top three ingredients (in frequency but not necessarily in amounts consumed) were olive oil, *cebola* (onion, I105) and *alho* (garlic, I022), which were present in 66%, 64%, and 59% of recipes, respectively. The top five also included white pepper and *salsa* (parsley, I360), present in 56% and 41% of recipes. When provinces were considered, the five top ingredients in Algarve, Beira Alta, Beira Litoral, Entre Douro e Minho, and Estremadura were exactly the same, even if their rank might vary. Conversely, in Azores, Beira Alta, Madeira, Ribatejo, and Trás-os-Montes e Alto Douro, parsley was always replaced by *batata* (potato, I046) in the top five ingredients, while in Alentejo, parsley was replaced by *pão* (bread, I277), which in Portugal is made almost exclusively of wheat flour. *Broa/pão de milho* (corn bread, I060) and *pão de centeio* (rye bread, I278) were recorded as ingredients in far fewer recipes, localized to Beira Litoral and Entre Douro e Minho (corn bread) and Trás-os-Montes e Alto Douro (rye bread). However, the comparison of individual ingredients can be misleading because the same basic sources of food can be used in different ways. For example, pork appears in the data set in 66 different ways, including as *enchidos*, suckling pig and fats. Various fish, including shellfish and mollusks but excluding Atlantic cod, salted and dried, appeared in 85 different ways; beef appeared in 36 ways; lamb in 24; chicken and goat in 14 ways; hare in four; turkey in three; and duck and rabbit in two different ways each. In addition, wines appear in nine different ways and peppers appear in six. Therefore, we combined data from these ingredients and recalculated the frequencies.

As before, olive oil and onion were the most frequently used, but garlic was replaced by peppers, present in 61% of recipes. In fourth place was garlic, and the top five was completed by pork, present in one way or another in more than half (53%) the recipes, replacing parsley. However, when provinces were examined, a more fragmented pattern emerged; now, it was Beira Litoral, Entre Douro e Minho,

Ribatejo, and Trás-os-Montes e Alto Douro where the five top ingredients were the same, even if the rank of the ingredients might vary. Conversely, in Alentejo and Beira Baixa, peppers were replaced by bread and parsley, respectively; in Algarve and Estremadura, pork by fish and parsley; in the Azores, olive oil by wines; in Beira Alta, onion by potato; and in Madeira, garlic and pork by potato and fish.

Incidentally, Atlantic cod, salted and dried, appeared in only 129 recipes, ranking 23rd (in provinces 13th to 38th, Ribatejo and Azores, respectively)—well under the 1001 different ways to cook Atlantic cod claimed by the Portuguese [3] (p. 24) or the more modest 365 ways, one for each day of the year [29] (p. 37), [45].

3.5. Final Remarks

Clearly, Portuguese traditional recipes organize differently across provinces. In the exploratory characterization, Entre Douro e Minho and Estremadura always occurred together, as did Beira Alta and Beira Baixa. Alentejo and Algarve almost always appeared together with the first group, while Ribatejo and Azores occurred together with the second of hinterland Beiras. The remaining two, Trás-os-Montes e Alto Douro and Beira Litoral, appeared with Madeira. It is also worth noting that Beira Litoral had an "attractor" role in terms of likeness-values for ingredients. Detailed analyses could confirm or inform the tentative grouping presented, providing a clearer picture of Portuguese traditional cuisine and hopefully pointing to possible explanations.

The data set presented here represents a comprehensive review that brings together an otherwise dispersed body of published knowledge on Portuguese traditional recipes. As such, it may constitute a reference inventory for investigating and testing a number of hypotheses on the distribution and pattern of use of ingredients within and among natural regions of Portugal. For example, how do provinces group in relation to the presence of different types of ingredients, namely herbs and spices, or vegetables, meats, fishes, and shellfish? Further, what might be the relation, if any, of the groups formed and factors like climatic, demographic, economic, and historic characteristics or events?

The data set might also be useful to investigate the extent and reality of tradition and traditional use attributed to the recipes. When building the data set, and in this paper, we accepted as good the classification of recipes as traditional without attempting to define or characterize this term. For the moment, we have to assume that the recipes fall somewhere between two opposing poles, where one is truly traditional (i.e., part of an aggregate of customs and practices that give continuity to a culture and to a social group [46] (p. 84)), and the other comprised of "invented traditions", more or less loosely based on past customs that once had the capacity of change and evolution, but are now frozen in various ideals or ritualized ways [47].

Supplementary Materials: The following are available online at www.mdpi.com/2306-5729/3/1/9/s1, Supplementary S1: A data set of Portuguese traditional recipes; Supplementary S2: English-Portuguese glossary of ingredient names.

Acknowledgments: We thank Maria João Cabrita (Departamento de Fitotecnia, Universidade de Évora, Évora, Portugal), P. R. Almeida (Departamento de Biologia, Universidade de Évora, Évora, Portugal), and Miguel Elias (Departamento de Fitotecnia, Universidade de Évora, Évora, Portugal) for their generous help in the translation to English of names of ingredients. Names of alcoholic beverages (M.J.C.); names of sea and fresh water foods, including fish, shellfish and mollusks (P.R.A.); names of game, livestock and poultry including their parts (M.E.). We also thank two anonymous reviewers for comments, suggestions and corrections made to earlier versions of this paper.

Author Contributions: A.S.D. conceived and designed the data set; A.S.D. and L.S.D. collected data and double-checked the data set; L.S.D. performed numerical and statistical analyses; A.S.D. and L.S.D. wrote the paper.

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

References

1. Brillat-Savarin, J.-A. *Physiologie du Gout, ou Méditations de Gastronomie Transcendante; Ouvrage Théorique, Historique et a l'Ordre du Jour, Dédié aux Gastronomes Parisiens, par un Professeur, Membre de Plusieurs Societés*

Littéraires et Savantes; A. Sautelet et Cie Libraires: Paris, France, 1826; Volume 1, p. 390. Available online: http://catalogue.bnf.fr/ark:/12148/cb30161657n.public (accessed on 23 October 2017).
2. Helen Hemingway. Macropedia. In *The New Enciclopædia Britannica*, 15th ed.; Helen Hemingway: Chicago, IL, USA, 1974; Volume 7, p. 1135.
3. Sarvis, S. *A Taste of Portugal*; Charles Scribner's Sons: New York, NY, USA, 1967; p. 192.
4. Instituto Nacional de Estatística. Statistics Portugal. Available online: https://www.ine.pt/xportal/xmain?xpgid=ine_main&xpid=INE (accessed on 15 December 2017).
5. Ribeiro, O. *Portugal. O Mediterrâneo e o Atlântico*, 5th ed.; Revised and Augmented; Sá da Costa: Lisboa, Portugal, 1987; p. 189.
6. Balbi, A. *Essai Statistique sur le Royaume de Portugal et d'Algarve, Comparé aux Autres États de l'Europe, et Suivi d'un Coup d'Oeil sur l'État Actuel des Sciences, des Lettres et des Beaux-Arts Parmi les Portugais des Deux Hémisphères*; Rey et Gravier, Libraires: Paris, France, 1822; Volume 1, p. 481.
7. Costa, B.C.C.; Castro, L. *Le Portugal du Point de Vue Agricole*; Imprimerie Nationale: Lisbonne, Portugal, 1900; p. 965. Available online: https://ia601700.us.archive.org/20/items/leportugalaupoin00cost/leportugalaupoin00cost.pdf (accessed on 4 December 2017).
8. Disney, A.R. *A History of Portugal and the Portuguese Empire*; Cambridge University Press: Cambridge, UK, 2009; Volume 1, p. 386.
9. *Código Administrativo (Aprovado pelo Decreto-Lei n.° 27-424 de 31 de Dezembro de 1936) Contendo as Leis n.os 1:940 de 3 de Abril de 1936, e 1:945 e 1:1946 de 21 de Dezembro de 1936 que Promulgam as Bases da Organização Administrativa e Dão Nova Redacção a Alguns Artigos da Constituição, e Inserindo Todas as Tabelas e Mapas Anexos ao Código*; Emprêsa Jurídica Editora: Coimbra, Portugal, 1937; p. 248.
10. Girão, A.A. *Esbôço Duma Carta Regional de Portugal*, 2nd ed.; Rewritten and Augmented; Imprensa da Universidade: Coimbra, Portugal, 1933; p. 224.
11. Billing, J.; Sherman, P.W. Antimicrobial functions of spices: Why some like it hot. *Q. Rev. Biol.* **1998**, *73*, 3–49. [CrossRef] [PubMed]
12. Sherman, P.W.; Billing, J. Darwinian gastronomy: Why we use spices. *BioScience* **1999**, *49*, 453–463. [CrossRef]
13. Bello, A.M.O. *Culinária Portuguesa*; Assírio & Alvim: Lisboa, Portugal, 1994; p. 454. ISBN 972-37-0370-X.
14. Saramago, A.; Fialho, M. *Cozinha Alentejana*; Assírio & Alvim: Lisboa, Portugal, 1998; p. 271. ISBN 972-37-0494-3.
15. Saramago, A. *Cozinha Transmontana: Enquadramento Histórico e Receitas*; Assírio & Alvim: Lisboa, Portugal, 1999; p. 275.
16. Saramago, A. *Cozinha do Minho: Enquadramento Histórico e Receitas*; Assírio & Alvim: Lisboa, Portugal, 2000; p. 254.
17. Saramago, A. *Cozinha Algarvia: Enquadramento Histórico e Receitas*; Assírio & Alvim: Lisboa, Portugal, 2001; p. 228.
18. Saramago, A. *Cozinha da Beira Interior: Enquadramento Histórico e Receitas*; Assírio & Alvim: Lisboa, Portugal, 2002; p. 254.
19. Saramago, A. *Cozinha da Beira Litoral: Enquadramento Histórico e Receitas*; Assírio & Alvim: Lisboa, Portugal, 2003; p. 238.
20. Modesto, M.L. *Cozinha Tradicional Portuguesa*, 2nd ed.; Editorial Verbo: Lisboa, Portugal, 1982; p. 335.
21. Modesto, M.L.; Praça, A. *Festas e Comeres do Povo Português*, 2nd ed.; Editorial Verbo: Lisboa, Portugal, 2000; Volume 1, p. 214.
22. Modesto, M.L.; Praça, A. *Festas e Comeres do Povo Português*; Editorial Verbo: Lisboa, Portugal, 1999; Volume 2, p. 218.
23. Bourguette, V.; Mourato, D. *Gastronomia Regional Portuguesa*, 5th ed.; Impala Editores: Ranholas, Portugal, 2005; p. 240.
24. Guedes, F. *Receitas Portuguesas. Os Pratos Típicos das Regiões*, 3rd ed.; D. Quixote: Lisboa, Portugal, 2005; p. 301.
25. Valente, M.O.C. *O Melhor da Cozinha Regional Portuguesa*; Círculo de Leitores: Rio de Mouro, 2004; p. 248.
26. Ribeiro, M. *Do Minho ao Algarve. Açores e Madeira*; Planeta Editora: Lisboa, Portugal, 2006; p. 370.
27. Modesto, M.L. *Traditional Portuguese Cooking*, 6th ed.; Verbo: Lisboa, Portugal, 2014; p. 359.
28. European Food Safety Authority. Foodex2 Browsing Tool (Updated: 18 August 2017). Available online: https://www.efsa.europa.eu/en/data/data-standardisation (accessed on 21 January 2018).

29. Anderson, J. *The Food of Portugal*; William Morrow and Company: New York, NY, USA, 1986; p. 304.
30. Carter-Tripp, M. Guide to Food in Portugal. Available online: http://www.xacobeo.fr/ZF1.07.lex.en-pt_food-term.pdf (accessed on 24 June 2016).
31. Cabrita, M.J.; University of Évora, Évora, Portugal. Personal communication, 2017.
32. Almeida, P.R.; University of Évora, Évora, Portugal. Personal communication, 2017.
33. Elias, M.; University of Évora, Évora, Portugal. Personal communication, 2017.
34. Diederichsen, A. *Coriander (Coriandrum sativum L.)*; IPK; IPGRI: Rome, Italy; Gatersleben, Germany, 1996; p. 83.
35. Silva, F.A.; Menezes, C.A. *Elucidário Madeirense. Volume Terceiro O–Z*, 3rd ed.; Regional Secretariat for Tourism and Culture, Regional Directorate for Cultural Affairs: Funchal, Portugal, 1966; p. 413.
36. Turmo, I.G. *200 Años de Cocina. Historia y Antropología de la Alimentación*; Cultiva Libros: Madrid, Spain, 2013; p. 375.
37. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984; p. 358.
38. Guegen, A.; Nakache, J.P. Méthode de discrimination basée sur la construction d'un arbre de decision binaire. *Rev. Stat. Appl.* **1998**, *36*, 19–38. Available online: http://www.numdam.org/item?id=RSA_1988__36_1_19_0 (accessed on 17 December 2017).
39. Sterne, J.A.C.; Smith, J.D. Sifting the evidence—What's wrong with significance tests? *Br. Med. J.* **2001**, *322*, 226–231. [CrossRef]
40. Chatterjee, U.; Kumar, V.; Madalli, D.P. Formalizing food ingredients for data analysis and knowledge organization. *Collnet. J. Scientometr. Inf. Manag.* **2016**, *10*, 289–309. [CrossRef]
41. Chew, V. Comparing treatment means: A compendium. *HortScience* **1976**, *11*, 348–357.
42. Penas, A.C.; Dias, L.S.; Mota, M.M. Precision and selection of extraction methods of aphelenchid nematodes from maritime pinewood, *Pinus pinaster* L. *J. Nematol.* **2002**, *34*, 62–65. Available online: http://journals.fcla.edu/jon/article/view/69396 (accessed on 12 December 2017). [PubMed]
43. Ury, H.K. A comparison of four procedures for multiple comparisons among means (pairwise contrasts) for arbitrary sample sizes. *Technometrics* **1976**, *18*, 89–97. Available online: http://amstat.tandfonline.com/doi/abs/10.1080/00401706.1976.10489405#.WjBXJKNLHiy (accessed on 12 December 2017). [CrossRef]
44. Draper, N.R.; Smith, H. *Applied Regression Analysis*, 3rd ed.; John Wiley: New York, NY, USA, 1998; p. 706.
45. Costa, A.P. Portugal. A dialogue of cultures. In *Culinary Cultures of Europe. Identity, Diversity and Dialogue*; Goldstein, D., Merkle, K., Eds.; Council of Europe Publishing: Strasbourg, France, 2005; pp. 347–356.
46. Helen Hemingway. Micropedia. In *The New Enciclopædia Britannica*, 15th ed.; Helen Hemingway: Chicago, IL, USA, 1974; Volume X, p. 1031.
47. Hobsbawm, E. Introduction: Inventing traditions. In *The Invention of Tradition*; Hobsbawm, E., Ranger, T., Eds.; Cambridge University Press: Cambridge, UK, 1983; pp. 1–14.