Predicting fruit and vegetable processing wash-water quality

Gurvinder S. Mundi, Richard G. Zytner, Keith Warriner and Bahram Gharabaghi

ABSTRACT

Wastewaters from the fresh produce processing industry are high in solids and organic matter requiring adequate treatment prior to disposal or recycling. Characterization of the processing wastewater, also referred to as wash-water is challenging, as the quality is a function of the produce. Analysis of water quality parameters, such as total suspended solids, total solids, total dissolved solids, chemical oxygen demand, biochemical oxygen demand, total nitrogen, total phosphorus, ammonia, and electrical conductivity from different fruit and vegetable operations were analyzed to develop the innovative power function models and ranking system to estimate wash-water quality. The developed models take the form of $Y = a(x)^b$, where Y, a, x, and b are estimate, scale, rank, and location parameters, respectively. The location and rank range from −0.65 to −3.18 and 0.05 (worst water quality) to 1, respectively, while the scale parameters are highly variable. Average and standard deviation estimation models show a very good fit for washing only ($R^2 > 73\%$) and washing with processing ($R^2 > 79\%$). The models and ranks highlight the degree of treatment required to address protection of surface and ground water and make the water quality conform to regulatory standards, benefiting watershed managers, government agencies, consultants, farmers, producers, processors and technology providers.

Key words | environmental protection, fresh produce, management, model, treatment, wash-water, wastewater

Gurvinder S. Mundi
Richard G. Zytner (corresponding author)
Bahram Gharabaghi
School of Engineering,
University of Guelph,
50 Stone Road East, Guelph, Ontario,
Canada N1G 2W1
E-mail: *rzytner@uoguelph.ca*

Keith Warriner
Department of Food Science,
University of Guelph,
50 Stone Road East, Guelph, Ontario,
Canada N1G 2W1

INTRODUCTION

The fresh produce industry is a heavy user of water, and also produces large quantities of wastewater of varying characteristics with respect to solids content, biochemical oxygen demand (BOD), nutrients, and pathogens (Gil *et al.* 2009; Lehto *et al.* 2009, 2014; Castro-Ibáñez *et al.* 2017). The untreated wastewater can be detrimental to soils, rivers, and lakes if not treated adequately, due to the high levels of organic matter and solids it contains (Kern *et al.* 2006; Meneses *et al.* 2017). There is a diverse range of fresh produce types and it follows that the wastewater characteristics are highly variable (Lehto *et al.* 2011; Mundi *et al.* 2017). Some wastewaters are compatible with biological treatments while others require chemical and/or physical treatments. Many challenges exist in determining suitable treatment technologies and the degree of treatment required, as one treatment solution does not fit all (Lehto *et al.* 2014; Mundi & Zytner 2015; Mundi *et al.* 2017). Literature shows limited information on detailed characteristics of fresh produce wastewaters, also referred to as wash-water, from diverse fruit and vegetable washing and processing operations.

Post-harvest washing and processing of fruits and vegetables requires large volumes of water, which produce an equivalent amount of wash-water (Casani *et al.* 2005). Simple washing in comparison to washing with peeling of root vegetables use up to 3.0 m^3 and 5.0 m^3 of water per tonne of product, respectively (Lehto *et al.* 2014). In comparison, this metric varies with facilities in Ontario, as two main types of operations are encountered. One is continuous washing (continuous flows producing large volumes) while the other is dunk tank washing (batch process with small volumes). In addition, operation type and sizing, unit processes, and water efficiency practices have a major impact on water usage and wash-water production. As such, water usage can range from 2.7 to 20 m^3 per tonne of product processed. In comparison, the dunk tank

doi: 10.2166/wst.2018.109

washing requires as little as 0.6 m^3 per tonne of product. The different unit processes can consist of mechanical soil removal, washing, and processing (cutting, budding, and peeling), in addition to cooling and packaging (Lehto *et al.* 2009). The different unit processes employed at each facility depend on many factors, such as the type of product being washed, fruit or vegetable, and whether it is a root vegetable or leafy green. Water is also used for in-place cleaning of the machinery and floor area.

Food safety and quality are essential in fresh produce processing and can impact the product while in storage, in transit, or on store shelves (Gil *et al.* 2009). The process water has the potential to carry over pathogens, spoilage microbes, and pesticides during washing processes if not sanitized adequately (Griffith *et al.* 2015; van Haute *et al.* 2015). The generated wash-water is a function of processing and product type (fruit or vegetable). Wash-water characteristics include heavy loads of solids, organic matter, oxygen demand, nutrients, pathogens, and spoilage microbes (Casani *et al.* 2005; Kern *et al.* 2006; Mundi 2013; Lehto *et al.* 2014; Mundi *et al.* 2017). High organic matter and solids make it a challenge to treat the wash-water, especially from peeling of root vegetables (Lehto *et al.* 2009). Thus, treatment of wash-water is necessary to reduce contaminants and allow wash-water to meet effluent requirements as set by government agencies to safeguard watersheds, environments, and ecosystems (Michalak *et al.* 2013; Schoen *et al.* 2017).

Increased demand in combination with changing climate require resilient management of water resources, such as wash-water treatment and reuse (Bhupathiraju & Tucker 2011; Manzocco *et al.* 2015; Li *et al.* 2016). The focus of recent studies has been to improve water management by reducing process water and assess effective treatment of wash-water for potential reuse applications, while maintaining the highest food safety (Ivey & Miller 2013; Gómez-López *et al.* 2017; Lehto *et al.* 2018). The key drivers behind these studies are the more stringent standards coupled with an increase in fines, including closure due to non-compliance. Castro-Ibáñez *et al.* (2017), Gil *et al.* (2009) and Lehto *et al.* (2009, 2011, 2014, 2018) completed studies which provide excellent detail on characterization, in addition to treatment feasibility. This study will build on existing research and extend the characteristics of wash-waters to tree fruit and above-ground vegetables, while summarizing industry wide wash-water quality information into simple models. The goal of the study is to develop power regression models that are easy to understand, implement, and are insightful. These models allow stakeholders to understand the level of contamination in the wash-water to addresses treatment, disposal, or reuse.

MATERIALS AND METHODS

Data collection

Wash-waters from different fruit and vegetable washing and processing operations were collected as outlined in Table 1.

Table 1 | Fruit and vegetable wash-waters assessed to formulate Power-Rank models for predicting water quality parameters

Product type and name	Washing	Washing and processing	Grand total
Above-ground			
Broccoli	1		1
Melons	2		2
Mushroom	7		7
Peppers	4		4
Snap Beans	3		3
Squash/sweet potato	2		2
Sweet corn	1		1
Tomato	7		7
Zucchini	2		2
Above-ground total	**29**		**29**
Leafy green			
Boston lettuce	2		2
Spinach	3		3
Leafy green total	**5**		**5**
Root vegetable			
Carrot	28	50	78
Ginseng	11		11
Ginseng seed	4		4
Green onion	1		1
Mixed	3	9	12
Potato	35	3	38
Pumpkin/gourd	9		9
Sweet potato	25		31
Root vegetable total	**116**	**62**	**184**
Tree fruit			
Apple	12	13	25
Sour cherry	2		2
Tree fruit total	**14**	**13**	**27**
Grand total	**164**	**75**	**239**

The collected samples were assessed for nine water quality parameters: total solids (TS), total suspended solids (TSS), total dissolved solids (TDS), electrical conductivity (EC), chemical oxygen demand (COD), biochemical oxygen demand (BOD), total phosphorus (TP), total nitrogen (TN) and ammonia (NH_4-N). All water quality parameters were measured in milligrams per litre [mg/L] except for EC, which was measured in microsiemens per centimeter [μS/cm]. Statistical analysis of water quality parameters was completed to derive statistical inference properties (mean, standard deviations and variance) to assess probability functions (distribution fits) and to derive equations for predicting raw wash-water quality parameters mentioned above.

Table 1 shows that a total of 239 samples were obtained, 164 on just washing and 75 on washing and processing. All wash-water samples were categorized as either (1) washing, just removing the soil or washing the product to ensure microbiological safety of the food for consumption, or (2) washing and processing wash-waters, where facilities partake in both, washing the soils off, and cutting and peeling the product, i.e. carrots and potatoes. These samples were analyzed to assess water quality parameters of the washing and processing wash-waters. In total, testing was completed on wash-waters from 21 different types of fruits and vegetables, see Tables 2 and 4. The full dataset consists of four subsets, (1) Dataset#1–2013 to 2015 from Mundi *et al.* (2017) (most recent), (2) Dataset#2–2011 to 2013 from Mundi *et al.* (2017), (3) Dataset#3–2012 to 2013 from Ontario Ministry of Agriculture Finance and Rural Affairs (OMAFRA) and (4) Dataset#4–2004 to 2005 from OMAFRA. Wash-water data on water quality parameters were concatenated into one dataset from the four datasets mentioned above. Standard methods for conducting water quality testing were employed to test wash-water samples, Method 2540 (TSS and TS) and Method 5210 (BOD) (APHA/AWWA/WEF 2012). Additional testing was done by Hach instrumentation (DR5000-03) using Hach water quality testing kits for COD, TN, TP, and NH_4-N (TNT821,826,843,830) (Hamilton *et al.* 2006). The EC was measured using Zetasizer Nano-Z apparatus (Malvern, ZEN2600, UK).

Data analysis

Discrepancies and outliers in the data were assessed by calculating mean, minimum and maximum of each water quality parameter. Pivot tables were prepared for average and standard deviation values of different products (apple, carrot, etc.) for the different wash-water types to help develop Power-Rank models. Statistical and engineering analysis was required to understand the distribution fits (normal, log, etc.) of the water quality data. This was useful in determining the level of treatment required to meet effluent regulatory compliance at a certain confidence interval (CI), i.e. 95% or lower. Figure 1 shows the process flow used to understand the level of treatment needed for each water quality parameter.

The first step was to assess for probability distribution fit, for example, of all TSS values for potato wash-water. TSS data were imported into EasyFit distribution fitting software (Mathwave Technologies) to assess the fit under one of the 57 available probability functions within the software. The Kolmogorov-Smirnov (KS) goodness-of-fit test was used to rank the 57 probability functions to determine which one best describes the different water quality parameters. The KS goodness-of-fit test compares two continuous distributions of the water quality parameters, e.g. TSS, and ranks a distribution based on the distance between the expected (the given probability function) and the observed distribution (TSS water quality parameter) (Massey 1951; Thompson *et al.* 2016). The two-parameter lognormal distribution was selected as the probability function to represent the water quality parameters as it consistently ranked in the top 21 distributions. For example, TSS and BOD for potato washing wash-waters were selected to demonstrate the fitting of the lognormal distribution function, see Figure 2(a) and 2(b). Both suspended (TSS) and dissolved (BOD) constituencies were a good fit for the log normal distribution as shown by TSS and BOD probability of exceedance curves. This was also the case with the washing and processing wash-waters for potatoes. However, these were not highlighted in Figure 2.

The statistical parameters of the lognormal distribution, location (μ) and scale (σ) were calculated for the different water quality parameters for this study, using the developed power function relationships. The relationship between the mean (m), variance (v), location (μ), scale (σ), and confidence interval (CI) for the lognormal distributions are presented in Equations (1)–(4) (Atieh *et al.* 2017). The confidence interval at 95% were calculated for each water quality parameters, giving an indication of the lower and upper limits expected. Equation (4) is used to calculate the confidence interval, based on the modified Cox method, where n is the number of samples and z is the percentile of standard normal distribution (Zhou & Gao 1997). This is a naïve approach, which utilizes z as the multiplier. However, when t-distribution is used as the multiplier, it is referred to

Table 2 | Average and standard deviation of water quality parameters of different washing type wash-waters

AVG and STD	TSS		TS		TDS		TP		TN		COD		BOD		EC		NH_4-N	
Product	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**
Apple	60	50	600	170	950	470	87.3	85.4	3.2	1.0	300	290	120	120	1,550	580	0.1	0.1
Boston lettuce	140	20	560	20	420	10	3.8	0.1	3.9	0.3	140	80	90	30	500	50[a]	0.3	0.1
Broccoli	40[a]	20[a]	680[a]	70[a]	360[a]	50[a]	0.3	0.1[a]	2.0	0.4[a]	450[a]	110[a]	320	100[a]	370	40[a]	5.1	2.9[a]
Carrot	2,690	1,330	3,780	1,920	1,570	500	3.8	0.2	43.7	144.6[a]	1,440	350	320	100	2,150	440	1.2	1.0[a]
Ginseng	1,120	900	1,600	750	720	500	22.5	35.7	4.0	3.2	320	150	50	90	830	170	1.3	2.9
Ginseng seed	270[a]	40[a]	2,000	470[a]	880[a]	290[a]	49.3	53.5	17.0	2.8	730[a]	200[a]	450	430	1,600	1,490	23.9	27.7
Green onion	50[a]	20[a]	2,500	670[a]	470[a]	90[a]	0.5	0.2[a]	9.2[a]	0.9[a]	140[a]	50[a]	40	20[a]	720	30[a]	0.2[a]	0.2[a]
Melons	320	10	920	20	600	30	10.4	0.1	24.0	0.8	300	20	80	20	670	20	0.4	0.1
Mixed[b]	470	170	1,030	130	650	240	8.9	7.2	22.6	0.8	150	33	50	60	800	30	0.2	0.2
Mushroom	280	240	1,230	890	900	800	2.0	1.8	18.5	23.9	2,880	2,410	130	140	1,850	1,700	11.8	15.9
Peppers	40	20	750	150	710	170	0.3	0.2	17.6	12.8	140	80	40	20	940	190	0.2	0.1[a]
Potato	8,370	9,970	5,340	2,680	1,060	520	26.2	28.9	39.7	36.7	4,340	4,110	610	690	1,510	750	39.8	40.6
Pumpkin/gourd	50	50	1,130[a]	220[a]	3,760[a]	4,560[a]	22.0	14.1	8.1[a]	0.8[a]	390	120	10	10[a]	3,560	990	0.1[a]	0.1[a]
Snap beans	50	20	680	90	630	100	0.7	0.1	7.8	0.4	160	40	50	40	810	240	0.3	0.2
Sour cherry	70[a]	30[a]	1,230	320	340[a]	50[a]	1.3	0.9	7.0	1.3[a]	2,720[a]	1,920[a]	1,030	600	350	70	1.1	0.6[a]
Spinach	100	100	540	90	440	50	2.6	1.8	3.0	0.4	310	270	330	250	420	50[a]	0.6	0.2
Squash/sweet potato	580	250	680	120	100	130	1.8	0.5	6.2	0.7	200	80[a]	42	20	330	80	1.0	0.5
Sweet corn	70	80[a]	490	50[a]	420	50[a]	3.6	0.4[a]	3.9[a]	0.6[a]	220	220	160	80[a]	640	60[a]	0.4[a]	0.1[a]
Sweet potato	1,000	860	5,610	7,170	490	430	6.2	4.6	6.4	5.1	10[a]	40[a]	10	20	320	160	0.1	0.2
Tomato	60	60	1,590	1,110	1,720	1,130	1.7	2.7	1.7	0.8	90	60	20	10	1,470	1,030	0.2	0.1
Zucchini	40	10	570	20	530	10	0.2	0.02	17.4	0.5	50	20	10	10	780	10	0.2	0.1[a]

[a]Estimated using Power-Rank models.

[b]Mixture of carrot, spinach, and beetroot.

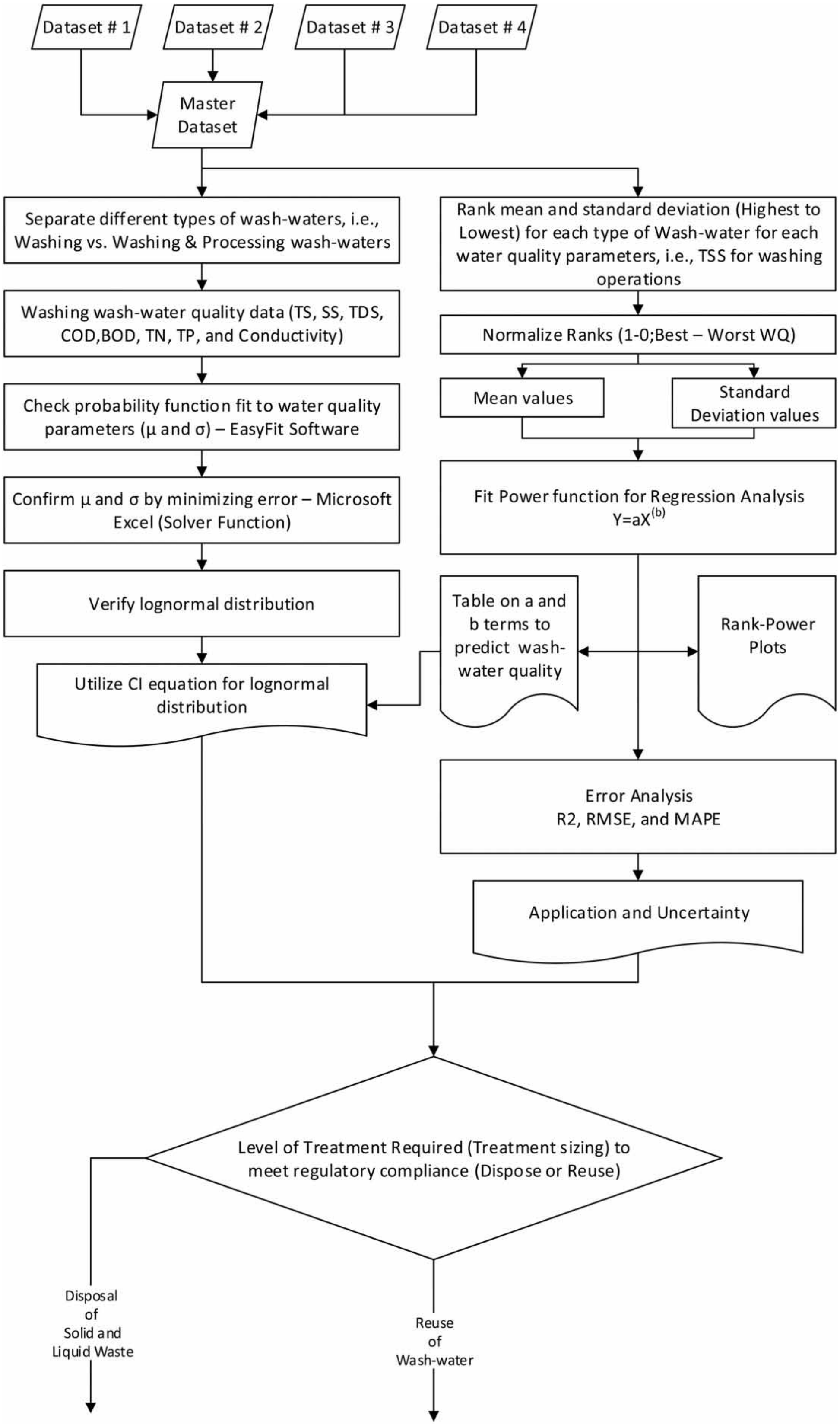

Figure 1 | Process flow diagram of the analysis performed.

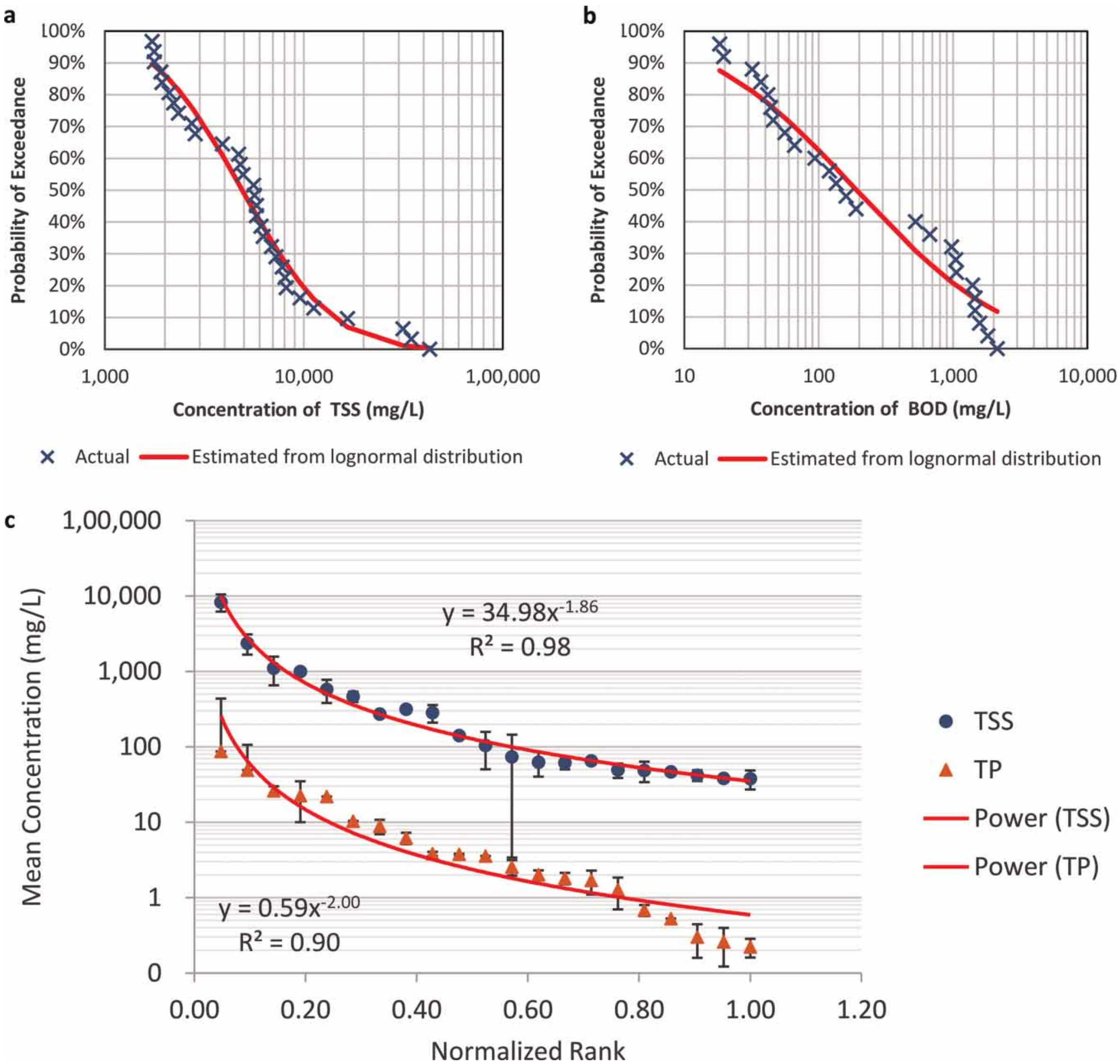

Figure 2 | Probability of exceedance, estimated vs. predicted using lognormal distribution for potato wash-water. (a) TSS and (b) BOD, and (c) normalized rank vs. mean concentration curves and corresponding power functions for fitting data for TSS and TP.

as the modified Cox method.

$$\mathrm{m} = \exp\left(\mu + \frac{\sigma^2}{2}\right) \tag{1}$$

$$\mu = \ln\left(\frac{\mathrm{m}}{\sqrt{1 + \frac{\nu}{\mathrm{m}^2}}}\right) \tag{2}$$

$$\sigma = \sqrt{\ln\left(1 + \frac{\nu}{\mathrm{m}^2}\right)} \tag{3}$$

$$CI = \exp\left(\left[\left(\mu + \frac{\sigma^2}{2}\right) \mp \left(z\sqrt{\left(\frac{\sigma^2}{n} + \frac{\sigma^4}{2(n-1)}\right)}\right)\right]\right) \tag{4}$$

The EasyFit software was also employed to calculate the mean and standard deviation values for each water quality parameter. The values for μ and σ were confirmed and optimized using the solver function in Excel software to minimize root mean square error (RMSE). The purpose of this exercise was to validate values calculated by EasyFit. Verification of location and scale values (μ and σ) ensured statistically correct distribution were applied in obtaining the mean and CI, which are the lower and upper limits of water quality parameters, such as for TSS. The mean and CI values are valuable in determining percent reduction required by treatment process to meet regulatory effluent compliance for TSS and other monitored water quality parameters.

Development of water quality prediction model (Power-Rank models)

Developing models to predict water quality parameters of wash-water requires the use of power functions, which helps to correlate water quality and normalized rank of the products. The power function is represented as $y = f(x) = ax^b$. Parameter a serves as a scaling factor (σ), moving the values of x^b up or down, leads to increase or decrease, respectively. The parameter b, called the power, determines the function's rates of growth or decay, the location factor (μ). While parameter y is the predicted value, i.e. TSS, and parameter x is the normalized rank, defined below. Figure 2(c) highlights the transformed data with corresponding power function parameters for mean concentration of TSS and TP vs. normalized rank. The power function is used to recognize a power trend, which might be difficult to see, but a simple log transformation of dataset can reveal a pattern that is unique to power functions.

Thus, water quality data were log transformed to the y-axis and utilized the power function. In addition, a ranking of average and standard deviation values (x-axis) from high to low was used to define the magnitude of each water quality parameter. This was done by first sorting the average values from high to low, where the highest average value was ranked as 1 and lowest was ranked as the last number. The rankings were further normalized, from 0–1, where zero was the highest magnitude, while one was the lowest magnitude. The formula for normalized rank is as follows, rank divided by n, where n is a number of products in a water quality parameter, i.e. TSS has 17 mean concentrations corresponding to different products. The normalization of ranks was important since not all water quality parameters have the same number of products. Normalized ranks vs. mean concentration plots were derived for all water quality parameters, both from averages and standard deviations values.

Some water quality data were unavailable for each sample. For example, some samples had missing TSS values but collected TP or vice versa. There were also missing BOD, COD while TDS was available within the same sample. In order to complete the dataset, missing ranks were added based on the corresponding similar water quality parameters, for example TSS with TP. The assumption was later verified by confirming the exponent values, b, of TSS and TP, where non-significant different values meant the assumption was correct for analysis.

Model performance evaluation

Statistical and graphical methods were used to assess the performance and validity of the Power-Rank models developed for wash-water quality prediction of the location and scale terms. The coefficient of determination (R^2) was used to understand the amount of observed variance within the models, Equation (5). The RMSE and mean absolute percent error (MAPE) were used to understand model accuracy and precision, Equations (6) and (7). RMSE show differences between the observed and predicted values in the units of the variable of study. In Equations (5) through (7), variables, O, and P are the observed and model predicted values, respectively, and n is the number of observations. A bar over the letter O or P, represents the average value, i.e., $\bar{O}$ is the average value of the observed series. The noted statistical measures were utilized to understand unique properties of the model performance.

$$R^2 = \left[\frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2} * \sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2}}\right]^2 \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{n}} \quad (6)$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{(O_i - P_i)}{O_i}\right| \quad (7)$$

RESULTS AND DISCUSSION

The wash-water quality parameters follow a lognormal distribution, as verified with the sampled data. This concept was demonstrated using TSS and BOD from washing wash-waters. Figure 2(a) and 2(b) show suspended (TSS) and dissolved (BOD) constituents as a good fit for the log normal distribution, as shown by the probability of exceedance curves. The RMSE for TSS and BOD, in Figure 2(a) and 2(b), are 5% and 6%, respectively. These error values are very low, suggesting a very good fit. Twenty-one different produce wash-water types were investigated for the nine water quality parameters, consisting of two different types of wash-water streams; first washing and, second, both

washing and processing (cutting and peeling) (Lehto *et al.* 2009; Mundi *et al.* 2017). Wash-water quality parameters and their corresponding mean and standard deviations are highlighted in Table 2 for wash-waters from washing operations, and, similarly, Table 4 for wash-waters from washing and processing operations. The missing data, noted by the asterisk in Tables 2 and 4, were calculated using the Power-Rank models. These tables depict average and standard deviation values corresponding to sampled and collected data, which consist of various facilities, operations, and wash-water types.

The different fruit and vegetable types can be categorized into: root-vegetable (carrot, ginseng, ginseng seed, green onion, potato, and sweet potato), tree fruit (apple and sour cherry), leafy greens (Boston lettuce and spinach), mixed (carrot, spinach, and beetroot) and above ground (broccoli, melons, mushroom, peppers, snap beans, pumpkin, sweet corn, tomato, and zucchini). It was evident that many root vegetables show high levels of solids as indicated by the TSS and TS values in both types of waste streams. These results suggest the need for physical treatment (settling pond) to reduce solids loading. Conversely, tree fruit and above-ground commodities show a low level of TSS and TS, which might be easier to treat for suspended solids. However, this was not the case for TP, TN, COD, and BOD, suggesting the requirement for a biological and/or chemical treatment to handle the high loads of contamination. For example, one of the tree fruit wash-water samples had very high TN and TP values due to the use of dunk tank for washing, with a 5-day water change cycle. Some facilities utilize dunk tanks, which serve as batch processes to clean the commodity. Unfortunately, these dunk tanks can accumulate contaminates, such as solids, TN, TP, BOD and pathogens. In addition, considerable differences exist in flows/volumes for small scale seasonal operations as compared to large operations, which use continuous water flow for washing and processing, diluting the contaminants. TP was approximately 10 mg/L for most products except for root vegetables. In general, TN values were much higher in comparison to TP. Higher BOD, COD and EC values were noticed for root vegetables in comparison to others. Highlighting that physical treatment (settling pond) alone will not suffice for root-vegetables, the addition of aeration, biological, and/or chemical treatment will be required. Different types of produce, based on their ranks, which can be similar for a category of wash-water (i.e., root vegetable) can highlight contaminants of concern to target for treatment. These trends and results are in line with previous studies by Mundi & Zytner (2015), Mundi *et al.* (2017), Kern *et al.* (2006), Lehto *et al.* (2009, 2014) and Casani *et al.* (2005).

Estimated values for average and standard deviation serve as a valid range to assess regulatory compliance, or determining the degree of treatment/technology required. However, some of the predicted values seem to be over or understated. This is due to estimation of normalized ranks, which introduce small errors into the system, and impact the prediction accuracy of the Power-Rank models. Figure 2(c) shows higher ranks for the TP model (curve) overestimating the TP values. Similarly, washing and processing operation wash-waters were summarized in Table 4. Table 4 shows that much higher levels of solids and oxygen demand, via TS, COD, and BOD. This was expected, as Table 4 represents operations which part-take in cutting, budding, and peeling stages, increasing the organic fraction of wash-waters (Kern *et al.* 2006; Lehto *et al.* 2018).

The normalized ranks for the Power-Rank model are derived in Table 3 for washing operations, and Table 4, bottom, for washing and processing operations. While most are based on sampled values, some are estimated, as indicated by the asterisk marks. The normalized ranks in Tables 3 and 4 may represent a source of error, as these were estimated based on the concept that similar wastewater constituents have similar levels of contamination. Small inaccuracies are inherent to estimating normalized rank, which can be magnified when using the Power-Rank models to predict water quality parameters. In addition, it is important to note that if the estimated normalized rank value lies at either end of the predicted curve of the Power-Rank models, then it will have a larger error associated with it due to the fitting of the curve.

Power-Rank model parameters, a and b, in addition to performance indicators, such as R^2, RMSE, and MAPE are highlighted in Table 5. Parameters a and b are reported to two decimal places to allow the model to predict to a greater accuracy, while the performance indicators are reported to zero decimal place for ease of interpretation. The scalar parameter, b, for the Power-Rank models are comparable and show hidden trends. Upon close inspection of Table 5, one can see that TSS, TP, and NH_4-N all have values around 2. Similarly, BOD and COD have values around 1.6, and TS, TDS, TN and EC have values of about 0.8. These parameters with similar Power-Rank model curves show that they are correlated. These correlations, however, were quite different for washing and processing wash-waters. The parameter b values for TSS was standalone around 3, TDS, TN and NH_4-N have values around 1.6, while TS, TP, COD, BOD and EC all have values around 1. These

Table 3 | Normalized ranks of water quality parameters of different washing type wash-waters (washing only)

Normalized rank (0.05–1) product	TSS		TS		TDS		TP		TN		COD		BOD		EC		NH_4-N	
	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD
Apple	0.67	0.62	0.81	0.52	0.24	0.24	0.05	0.05	0.81	0.38	0.52	0.24	0.43	0.38	0.24	0.29	0.95	0.71
Boston lettuce	0.48	0.86	0.90	0.86	0.81	0.81	0.48	0.95	0.76	1.00	0.81	0.67	0.48	0.62	0.76	0.76[a]	0.62	0.86
Broccoli	0.95[a]	0.90[a]	0.76[a]	0.76[a]	0.86[a]	0.86[a]	0.95	0.86[a]	0.95	0.86[a]	0.33	0.48	0.24	0.29[a]	0.86	0.86[a]	0.19	0.19[a]
Carrot	0.10	0.10	0.14	0.14	0.14	0.14	0.43	0.71	0.05	0.05[a]	0.19	0.19	0.29	0.43	0.10	0.33	0.29	0.29[a]
Ginseng	0.14	0.14	0.29	0.38	0.38	0.38	0.19	0.14	0.71	0.29	0.38	0.43	0.67	0.48	0.43	0.48	0.24	0.24
Ginseng seed	0.33[a]	0.57[a]	0.24	0.29[a]	0.29[a]	0.29[a]	0.10	0.10	0.48	0.33	0.24	0.33	0.14	0.14	0.19	0.10	0.10	0.10
Green onion	0.86[a]	0.81[a]	0.19	0.24[a]	0.62[a]	0.62[a]	0.86	0.76	0.38[a]	0.57[a]	0.76	0.76	0.76	0.76[a]	0.62	0.95[a]	0.71[a]	0.57[a]
Melons	0.38	0.95	0.57	0.95	0.57	0.57	0.29	0.90	0.14	0.52	0.48	1.00	0.52	0.67	0.67	0.90	0.48	0.81
Mixed[b]	0.29	0.33	0.48	0.62	0.48	0.48	0.33	0.29	0.19	0.43	0.71	0.90	0.57	0.52	0.52	0.67	0.76	0.48
Mushroom	0.43	0.29	0.43	0.33	0.33	0.33	0.62	0.48	0.24	0.14	0.14	0.14	0.38	0.24	0.14	0.05	0.14	0.14
Peppers	1.00	0.71	0.62	0.57	0.43	0.43	0.90	0.67	0.29	0.19	0.86	0.62	0.81	0.81	0.38	0.43	0.81	0.90
Potato	0.05	0.05	0.10	0.10	0.19	0.19	0.14	0.19	0.10	0.10	0.05	0.05	0.10	0.05	0.29	0.24	0.05	0.05
Pumpkin/gourd	0.81	0.52	0.43[a]	0.43[a]	0.05[a]	0.05[a]	0.24	0.24	0.43[a]	0.62[a]	0.29	0.52	1.00	1.00	0.05	0.19	1.00	1.00
Snap beans	0.76	0.76	0.67	0.71	0.52	0.52	0.81	0.81	0.52	0.90	0.67	0.81	0.62	0.57	0.48	0.38	0.57	0.62
Sour cherry	0.71[a]	0.67[a]	0.38	0.48	0.90[a]	0.90[a]	0.76	0.52	0.57	0.48[a]	0.10	0.10	0.05	0.10	0.90	0.62	0.33	0.33[a]
Spinach	0.52	0.38	0.95	0.81	0.76	0.76	0.57	0.43	0.90	0.95	0.43	0.29	0.19	0.19	0.81	0.81[a]	0.43	0.52
Squash/sweet potato	0.24	0.24	0.71	0.67	1.00	1.00	0.67	0.57	0.67	0.76	0.62	0.57	0.71	0.71	0.95	0.57[a]	0.38	0.38
Sweet corn	0.57	0.43[a]	1.00	0.90[a]	0.95	0.95[a]	0.52	0.62	0.86[a]	0.71[a]	0.57	0.38	0.33	0.33[a]	0.71	0.71[a]	0.52[a]	0.76[a]
Sweet potato	0.19	0.19	0.05	0.05	0.71	0.71	0.38	0.33	0.62	0.24	0.95	0.86	0.90	0.86	1.00	0.52	0.90	0.43
Tomato	0.62	0.48	0.33	0.19	0.10	0.10	0.71	0.38	1.00	0.67	0.90	0.71	0.86	0.90	0.33	0.14	0.67	0.67
Zucchini	0.90	1.00	0.86	1.00	0.67	0.67	1.00	1.00	0.33	0.81	1.00	0.95	0.95	0.95	0.57	1.00	0.86	0.95[a]

[a]Estimated using Power-Rank models.
[b]Mixture of carrot, spinach, and beetroot.

Table 4 | Average, standard deviations and normalized ranks of different washing and processing type wash-waters (washing and processing)

	TSS		TS		TDS		TP		TN		COD		BOD		EC		NH_4-N	
Product	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**
AVG and STD																		
Apple	80	54	2,760	2,370	5,190	5,020	9.3	13.6	10.1	14.5	1,160	1,670	690	990	1,150	860	1.0	0.74
Carrot	1,450	1,110	6,470	2,560	4,340	1,670	2.2	1.5	2.6	0.30	2,340	3,000	750	1,150	670	190	0.6	1.8
Mixed[a]	530	130	1,630	440	1,100	330	4.0[b]	3.5[b]	45.0	1.00[b]	14,700[b]	5,570[b]	1,830	5,200 [b]	620[b]	80[b]	5.6	2.8[b]
Potato	5,220	3,930	6,220	4,130	1,000	720	33.3	17.0	27.0	23.1	4,070	2,560	570	280	1,520[b]	1,670[b]	7.2	8.5
Normalized rank (0.05–1)																		
Apple	1.00	1.00	0.75	0.75	0.25	0.25	0.50	0.50	0.75	0.50	1.00	1.00	0.75	0.50	0.50	0.50	0.75	1.00
Carrot	0.50	0.50	0.25	0.50	0.50	0.50	1.00	1.00	1.00	0.75	0.75	0.50	0.50	0.75	0.75	0.75	1.00	0.75
Mixed[a]	0.75	0.75	1.00	1.00	0.75	1.00	0.75[b]	0.75[b]	0.25	1.00	0.25[b]	0.25[b]	0.25	0.25[b]	1.00	1.00[b]	0.50	0.50[b]
Potato	0.25	0.25	0.50	0.25	1.00	0.75	0.25	0.25	0.50	0.25	0.50	0.75	1.00	1.00	0.25	0.25[b]	0.25	0.25

[a]Mixture of iceberg lettuce, broccoli, carrots, green peppers, celery, and apple.
[b]Estimated using Power-Rank models.

Table 5 | Power-rank model parameters and performance parameters

	TSS		TS		TDS		TP		TN		COD		BOD		EC		NH_4-N	
	AVG	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**	**AVG**	**STD**
Washing type wash-water																		
a	34.99	9.83	527.80	43.52	315.47	41.48	0.59	0.09	3.29	0.30	91.50	28.90	17.26	12.00	0.42	0.03	0.11	0.03
b	−1.87	−2.33	−0.90	−1.91	−0.81	−1.54	−2.00	−2.90	−1.07	−2.03	−1.44	−1.79	−1.69	−1.66	−0.79	−1.75	−2.12	−2.68
R^2 (%)	98	97	97	88	73	49	90	90	85	95	93	94	80	86	90	80	97	96
RMSE (units)	470	487	631	1,619	125	207	38	114	10	2	740	638	431	267	0.3	1	6	17
MAPE (%)	15	25	9	42	20	84	56	85	27	21	19	22	78	41	17	55	39	25
Washing and processing type wash-water																		
a	146.36	63.42	2,044	831.59	1,023.5	379.64	2.26	2.09	4.27	0.37	1,244.8	1,789.8	522.98	364.46	0.62	0.08	0.72	0.88
b	−2.78	−3.18	−0.99	−1.33	−1.32	−1.93	−1.96	−1.75	−1.93	−3.34	−1.70	−0.82	−0.83	−1.92	−0.65	−2.19	−1.89	−1.67
R^2 (%)	92	95	79	67	82	98	99	81	85	74	97	88	92	75	91	91	84	95
RMSE (units)	888	702	1,353	856	1,086	271	1	5	10	11	175	180	130	216	0.2	0.25	2	0.3
MAPE (%)	45	31	22	45	25	12	3	32	42	46	6	6	11	21	7	20	34	11

are useful in reducing the Power-Rank models further, making them more robust and convenient. The rank along with the characteristics of the power function (a and b) will determine the level of wash-water quality, which then can be used for assessing the level of treatments required (physical, biological, and/or chemical).

The Power-Rank model equations show very good fit for the majority of the water quality parameters as indicated by R^2. RMSE values highlight the standard errors in terms of units, which are useful when applying models for practical applications. Given the variability of water quality parameters within wash-waters, variability is also expected in the RMSE. In addition, water quality parameters with large number of data points show lower RMSE, while fewer samples lead to higher RMSE. However, the overall trend showed that the waters derived from different facilities processing the same produce type show similar characteristics for some water quality parameters, as indicated by RMSE. This highlights that there was little variation with respect to wastewater composition derived from different facilities processing the same product type. This may be because similar products are processed in a similar manner. More importantly, a distinction was made in this study between washing versus washing and processing waste streams. This implication allows the extension of the research to be applied to other geographical locations around the world utilizing similar processes and unit operations. MAPE also highlights the error in terms of percentage, which also varies in range. Error and performance parameters do not show concerning trends or extreme values, validating the Power-Rank models for use in determining the different levels of treatment needed for implementation, and to meet effluent requirements post treatment.

Previous studies from Kern *et al.* (2006), Lehto *et al.* (2009, 2014) and Casani *et al.* (2005) highlighted characteristics and did not show such a wide and diverse set of data, as given by this study with the Power-Rank models. The use of the Power-Rank model can be demonstrated as follows. Say TSS from a ginseng washing only operation is to be predicted. The average TSS value can be determined using the Power-Rank coefficients from Table 5, which is equal to Ax^b or $34.99*x^{-1.87}$ where x is the normalized rank. The value of x can be obtained from Table 3 (washing only), which is equal to 0.14. Therefore, the average TSS value is equal to $34.99*(0.14)^{-1.87}$, or 1,383 mg/L. Likewise, the Power-Rank model for standard deviation is calculated as follows, $9.83*(0.14)^{-2.33}$, where coincidently the rank for standard deviation is also 0.14 (Table 3). Hence the standard deviation is equal to 960 mg/L. As a result, the TSS value for a wash-water from a ginseng washing operation from the developed models is 1,380 ± 960 mg/L (Average ± Standard Deviation). These levels can be used in determining effluent values for different treatment processes while evaluating various wash-water treatment options, including capacity using the flows generated by the system in question.

The equations listed in Table 5 were derived from a large number of facilities with varying operating and management conditions, and can be used to predict water quality parameters for average set of conditions. Prior to this study, the ranges and values of wash-water quality were unknown, making it difficult to identify potential treatment options. However, it should be noted that the current models do not account for the type of soil attached to a produce. Wash-water that contains mineral soils versus organic soils will exhibit different characteristics. For example, wash-water with mineral soils might be better suited for settling technologies, while wash-water contaminated by organic based soils are best suited for a technology like dissolved air flotation (Mundi & Zytner 2015). Similar challenges could exist for a parameter like nutrients. Additional wash-water data are needed to determine the impact the various parameters have on the Power-Rank models and are being considered for future studies.

Having the Power-Rank models shown in this paper provides researchers, technology providers and government officials the information needed to better understand the potential magnitude of the impact untreated wash-waters can have on the environment. The wash-water water quality parameters can be utilized for watershed studies to estimate impact on sensitive waterbodies, such as lakes and streams from agricultural operations involved in washing and processing fruits and vegetables. They also can provide insight on whether the wash-water will meet sewer discharge limits if the producer is serviced with a municipal sewer. In addition, the models will provide the information needed to design and evaluate wastewater treatment processes to achieve treatment objectives that may be set by producers or government agencies. The predicted raw wash-water quality parameters coupled with operating parameters, such as flow, allow consultants to determine the required amount of settling area needed for solids, or the amount of aeration/aerators required for biological breakdown of organic matter.

Wastewater discharge requirements are listed in Table 6 for various Canadian jurisdictions in comparison to German limits (specific to fruit and vegetable industry wastewaters). Using the predicted effluent quality from untreated wash-

Table 6 | Effluent requirements for wastewater discharge in Canada and Germany, water quality parameters measured in mg/L

	Target concentration for drinking water[a]	Target concentration for sanitary and combined sewer discharge[b]	Provincial water quality objectives[c]	German regulation limits for fruit and vegetable wastewater[d]
TSS	–	350	25[e]	–
TS	–	–	–	–
TDS	500	–	–	–
TP	0.01	10	0.02	2
TN	–	–	–	18
NH_4-N	0.02[f]	–	–	10
BOD	–	300	20[e]	25
COD	–	–	–	110

[a]Data obtained from Supporting Document for Ontario Drinking Water Quality Standards, Objectives and Guidelines, Tables 1, 2, and 4.
[b]Data obtained from City of Toronto Sewer Discharge and Storm Water Discharge Limits, Table 1.
[c]Data obtained from PWQO for Surface Water, some parameters are subjected to additional conditions.
[d]Federal Law Gazzete, The Ordinance on Requirements for the Discharge of Waste Water into Waters.
[e]Limits for effluent discharged to receiving waters; Guidelines for Effluent Quality and Wastewater Treatment at Federal Establishments.
[f]Additional requirements related to pH.

water via the Power-Rank models shows how much treatment is required prior to release. The target water quality is indicated by the Provincial Water Quality Objectives (PWQO) (see Table 6), which allows the release to surface water. The water quality parameters to meet and their corresponding limits are listed as 25 mg/L for TSS, 0.02 mg/L for TP, and 0.02 mg/L for NH4-N (MOECC 2016).

The current guidelines highlight three different types of legally enforceable requirements, which are based on either achievable treatment technology, the 75th percentile effluent water quality to meet the set PWQO, and/or site-specific receiving water quality requirements based on the assimilative capacity of the receiving watercourse (Trenouth *et al.* 2018). Emerging and new affordable technologies can make it easier to meet achievable/desired treatment requirements, while implementation/enforcement of the receiving water quality requirements can be very expensive/impractical. Thus, it is suggested that the 75th percentile confidence interval (calculated easily using the models presented in this study) be adopted as the effluent limit for fruit and vegetable wash-water, treated or untreated.

CONCLUSIONS

The study employs novel regression techniques for comprehensive fruit and vegetable washing and processing wash-waters for many different commodities in Southern Ontario to develop functional tools for the prediction of wash-water quality. A total of 239 samples were selected based on availability, with the collected data representing the fruit and vegetable washing and processing industry. Key water quality parameters include TSS, TDS, TS, TN, TP, EC, NH_4-N, COD and BOD, which follow a lognormal distribution. This study demonstrated how Power-Rank models derived using regression analysis can be used to predict the wash-water quality for fruit and vegetable processing wash-water. The models derived from the two different types of operations encountered, washing versus washing and processing, performed very well as indicated by the R^2 values. The RMSE and MAPE performance parameters indicate the expected error boundaries, which is in the range of acceptable. The datasets were an amalgamation of the industry wide samples. Variability is inherent to the developed models and may be improved by increasing sample size for various product types. Specifically, for produce with a low number of samples, such as broccoli, sweet corn, and green onions, which only had one sample.

The models developed in this study compliment previous characterization studies which show the range of water quality expected from different wash-waters. The advantages of this study are that it includes additional data and Power-Rank models for water quality parameters predictions. Some limitation exists, such as over and underestimating the mean and standard deviation of products for which the rank values were estimated. The high variability of water quality parameters seen in wash-waters means high levels of treatment may be required to treat and dispose of wash-water into the environment (surface, lakes, and rivers). Therefore, the prediction of water quality levels based on product, and the corresponding rank can yield very good estimates of expected water quality from

washing and processing facilities. This information is valuable for farmers, government agencies, consulting firms, technology providers, and other stakeholders interested in determining wash-water rudimentary treatment and sizing from predicted levels of water quality parameters. The study was successful in capturing a wide variety of information and concatenating into a useful tool for making an important decision on treatment of wash-waters.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge both financial and technical support received from our research partner OMAFRA (Deanna Nemeth, Rebecca Shortt, and Tim Brook) and the many vegetable and fruit organizations, cooperators, producers, and processors who participated in the studies accumulating this paper. The authors greatly appreciate the assistance of the many coop and summer students for their hard work: Deanne Durward, Ryan Good, Shreya Ghose, Laurissa Christie, Lindsay Johnston and Duncan Hartwick.

REFERENCES

APHA/AWWA/WEF 2012 *Standard Methods for the Examination of Water and Wastewater*, 22nd edn. American Public Health Association/American Water Works Association/Water Environment Federation, Washington, DC, USA.

Atieh, M., Taylor, G., Sattar, A. M. & Gharabaghi, B. 2017 Prediction of flow duration curves for ungauged basins. *Journal of Hydrology* **545**, 383–394.

Bhupathiraju, S. N. & Tucker, K. L. 2011 Coronary heart disease prevention: nutrients, foods, and dietary patterns. *Clinica Chimica Acta* **412** (17), 1493–1514.

Casani, S., Rouhany, M. & Knøchel, S. 2005 A discussion paper on challenges and limitations to water reuse and hygiene in the food industry. *Water Research* **39** (6), 1134–1146.

Castro-Ibáñez, I., Gil, M. I. & Allende, A. 2017 Ready-to-eat vegetables: current problems and potential solutions to reduce microbial risk in the production chain. *LWT-Food Science and Technology* **85**, 284–292.

Gil, M. I., Selma, M. V., López-Gálvez, F. & Allende, A. 2009 Fresh-cut product sanitation and wash water disinfection: problems and solutions. *International Journal of Food Microbiology* **134** (1), 37–45.

Gómez-López, V. M., Gil, M. I. & Allende, A. 2017 A novel electrochemical device as a disinfection system to maintain water quality during washing of ready to eat fresh produce. *Food Control* **71**, 242–247.

Griffith, A., Neethirajan, S. & Warriner, K. 2015 Development and evaluation of silver zeolite antifouling coatings on stainless steel for food contact surfaces. *Journal of Food Safety* **35**, 345–354.

Hamilton, A. J., Mebalds, M. I., Aldaoud, R. & Heath, M. 2006 Physical, chemical and microbial characteristics of wastewater from carrot washing in Australia: implications for reuse and environmental discharge. *Journal of Vegetable Science* **11** (3), 57–72.

Ivey, M. L. & Miller, S. A. 2013 Assessing the efficacy of pre-harvest, chlorine-based sanitizers against human pathogen indicator microorganisms and *Phytophthora capsici* in non-recycled surface irrigation water. *Water Research* **47** (13), 4639–4651.

Kern, J., Reimann, W. & Schlüter, O. 2006 Treatment of recycled carrot washing water. *Environmental Technology* **27** (4), 459–466.

Lehto, M., Sipilä, I., Sorvala, S., Hellstedt, M., Kymäläinen, H. R. & Sjöberg, A. M. 2009 Evaluation of on-farm biological treatment processes for wastewaters from vegetable peeling. *Environmental Technology* **30** (1), 3–10.

Lehto, M., Kuisma, R., Määttä, J., Kymäläinen, H. R. & Mäki, M. 2011 Hygienic level and surface contamination in fresh-cut vegetable production plants. *Food Control* **22** (3), 469–475.

Lehto, M., Sipilä, I., Alakukku, L. & Kymäläinen, H. R. 2014 Water consumption and wastewaters in fresh-cut vegetable production. *Agricultural and Food Science* **23** (4), 246–256.

Lehto, M., Kuisma, R., Kymäläinen, H. R. & Mäki, M. 2018 Neutral electrolyzed water (NEW), chlorine dioxide, organic acid based product, and ultraviolet-C for inactivation of microbes in fresh-cut vegetable washing waters. *Journal of Food Processing and Preservation* **42** (1).

Li, Z., Huang, G., Wang, X., Han, J. & Fan, Y. 2016 Impacts of future climate change on river discharge based on hydrological inference: a case study of the Grand River Watershed in Ontario, Canada. *Science of The Total Environment* **548**, 198–210.

Manzocco, L., Ignat, A., Anese, M., Bot, F., Calligaris, S., Valoppi, F. & Nicoli, M. C. 2015 Efficient management of the water resource in the fresh-cut industry: current status and perspectives. *Trends in Food Science & Technology* **46** (2), 286–294.

Massey Jr., F. J. 1951 The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association* **46** (253), 68–78.

Meneses, Y. E., Stratton, J. & Flores, R. A. 2017 Water reconditioning and reuse in the food processing industry: current situation and challenges. *Trends in Food Science & Technology* **61**, 72–79.

Michalak, A. M., Anderson, E. J., Beletsky, D., Boland, S., Bosch, N. S., Bridgeman, T. B. & DePinto, J. V. 2013 Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proceedings of the National Academy of Sciences* **110** (16), 6448–6452.

Ministry of the Environment and Climate Change (MOECC) 2016 *B-1-5 Deriving Receiving Water Based Point Source Effluent Requirements for Ontario Waters*. https://www.ontario.ca/page/b-1-5-deriving-receiving-water-based-point-source-effluent-requirements-ontario-waters#section-8 (accessed 8 September 2017).

Mundi, G. S. 2013 *Assessment of effective solids removal technologies to determine potential for vegetable washwater reuse*. MASc Thesis, School of Engineering, University of Guelph, Guelph, ON, Canada.

Mundi, G. S. & Zytner, R. G. 2015 Effective solid removal technologies for wash-water treatment to allow water reuse in the fresh-cut fruit and vegetable industry. *Journal of Agricultural Science and Technology A* **5**, 396–407.

Mundi, G. S., Zytner, R. G. & Warriner, K. 2017 Fruit and vegetable wash-water characterization, treatment feasibility study and decision matrices. *Canadian Journal of Civil Engineering* **44** (11), 971–983.

Schoen, M. E., Xue, X., Wood, A., Hawkins, T. R., Garland, J. & Ashbolt, N. J. 2017 Cost, energy, global warming, eutrophication and local human health impacts of community water and sanitation service options. *Water Research* **109**, 186–195.

Thompson, J., Sattar, A. M., Gharabaghi, B. & Warner, R. C. 2016 Event-based total suspended sediment particle size distribution model. *Journal of Hydrology* **536**, 236–246.

Trenouth, W. R., Gharabaghi, B. & Farghaly, H. 2018 Enhanced roadside drainage system for environmentally sensitive areas. *Science of The Total Environment* **610**, 613–622.

van Haute, S., Sampers, I., Jacxsens, L. & Uyttendaele, M. 2015 Selection criteria for water disinfection techniques in agricultural practices. *Critical Reviews in Food Science and Nutrition* **55** (11), 1529–1551.

Zhou, X. H. & Gao, S. 1997 Confidence intervals for the log-normal mean. *Statistics in Medicine* **16** (7), 783–790.

First received 8 December 2017; accepted in revised form 21 February 2018. Available online 8 March 2018