A novel approach for *Campylobacter* spp. detection: could spectrophotometry help early prediction?

C. Schang[1], R. Henry[1], K. Zhang[2] & D.T. McCarthy[1]*

[1]*Environmental and Public Health Microbiology Laboratory (EPHM Lab), Department of Civil Engineering, Monash University, Wellington Rd, Clayton, Victoria, 3810, Australia*
[2] *Water Research Centre (WRC), School of Civil and Environmental Engineering, UNSW Sydney, High street, Kensington, NSW 2052, Australia.*

**Corresponding author email: david.mccarthy@monash.edu*

Highlights

- First investigation of the potential of spectrophotometry for the detection of *Campylobacter*
- Genetic programming showed some promising results when using larger volumes data.
- Matrix specific effects contribute to detection variability

Introduction

Campylobacteriosis is a zoonosis spread into the environment through the release of faecal material. Current WHO figures suggest that *Campylobacter* are the leading cause of diarrheal disease in industrialized nations with annually more than 60, 000 and 17, 000 confirmed cases reported respectively in the United Kingdom (UK) and Australia alone (Corvisy, 2013; Hughes and Gorton, 2013). Enumeration of *Campylobacter* from complex source samples can be difficult due to the fastidiousness and fragility of the organism. Furthermore, isolation from urban waters is problematic, as they are usually present at low concentrations (Koenraad et al., 1997). Culture-based methods for the enumeration and isolation of *Campylobacter* from waters have become the international standard (Standardization ISO, 2005). However, culture-based methods are time-consuming and expensive, requiring filtration, selective enrichment, isolation and biochemical confirmation (~9 days to report). A modified MPN-PCR method described in Henry et al. (2016), evaluated by analysing 147 estuarine samples collected over a 2 year period, demonstrated that the intra-laboratory performance of the MPN-PCR was superior to that of AS/NZS ($\sigma = 0.7912$, $P < 0.001$; $\kappa = 0.701$, $P < 0.001$) with an overall diagnostic accuracy of ~94%. This method reduced in lab analysis times to 4 days instead of the traditional 9 days to report. However, both the traditional culture based method and the modified MPN-PCR method remain costly (in both resources and staffing) thus, the search for cheaper and technically more accessible method remains ongoing.

Campylobacter growth occurs under microaerophilic conditions (85% N_2, 10% CO_2, and 5% O_2). Visual observation of standard culture media (Preston broth) after 48 hrs at 42°C demonstrated a range of coloured appearances with anecdotal evidence suggesting positive cultures to have a "red" colouration. Therefore, this study investigated the potential of predicting the presence of absence of *Campylobacter* after the initial 48 hrs incubation of the Preston broth using spectrophotometry. The absorbance spectrum results were used to generate models to predict the presence/absence of *Campylobacter* in each tube using various wavelength.

Methodology

Sampling method

Water, sewer and sediment samples were collected at various location around Melbourne, Australia. Samples were analysed for *Campylobacter* spp. using the MPN-PCR method described in Henry et al. (2016). In addition, some Preston tubes were also plated on blood-free charcoal agar plates and subsequently on Horse Blood Agar (HBA) plates for confirmation. During preliminary testing (16/07/2018 and 02/12/2018), 1mL of Preston broth from the 3 top volumes applied to the MPN method, as well as 1mL of the positive and negative controls, were tested for absorbance spectrum using a 1mL plastic cuvette and a spectrophotometer (DR500, HACH). In the second stage of the study (28/02/2019 and 09/02/2020), 100µL of each Preston broth tubes for each of the source samples were transferred into a tissue culture 96 wells microplate (Falcon) and

tested using a Multiskan Sky spectrophotometer (Thermo Fisher Scientific). The absorbance spectrum of each well was scanned for wavelength between 220nm and 850nm (visual spectrum).

Modelling and Data analysis.

In this study, the Genetic Programming (GP) was selected to generate the quantitative relationships between presence/absence of *Campylobacter* and the absorbance spectrum results. GP is an evolutionary approach that extends genetic algorithms to allow the exploration of the space of problems. The GP model works firstly by randomly generating an initial set of populations (in this case algorithms consisting numbers and mathematical operators to calculate *Campylobacter* presence / absence from spectrum results) (Fig. 1). All the initial populations (*i.e.,* algorithms) are then assessed for a fitness value (*e.g.,* correlation coefficient, mean square error, R^2 goodness of fit, etc). If the quality criterion is met, then the optimum algorithm will be output; otherwise, the initial populations with high fitness are used to generate their offspring / new generations through by mimicking the basic principles of Darwinian evolution, *e.g.,* copy, crossover and mutation. The process is iterative, allowing for the generation of optimum candidate algorithms. It has already been widely used in many areas of water resource engineering to generated algorithms that help to understand complex and highly non-linear behaviour of many problems, such as stormwater quality modelling (Zhang *et al.*, 2019) and groundwater quality estimation (Aryafar *et al.*, 2019).

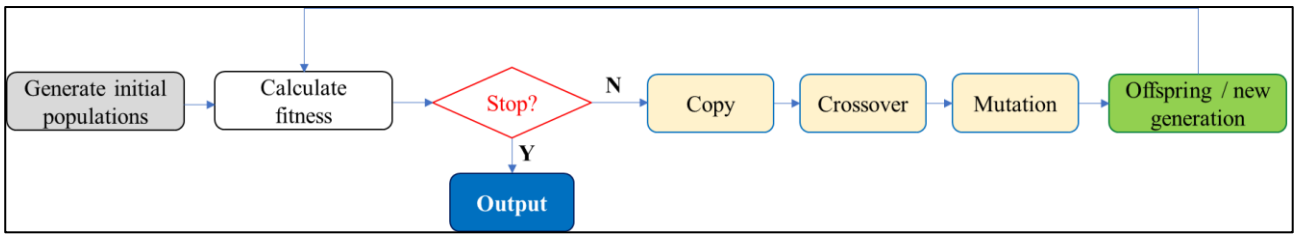

Figure 1. General steps of the genetic programming

The GP software *Eureqa*, developed by Schmidt and Lipson (2009; 2013) was applied to this study. As *Eurequa* only deals with numerical data, but the modelling outcomes of this study are binary values, i.e., presence (value 1) or absence (value 0) of *Campylobacter*. Therefore, *Eureqa* searched for an algorithm that goes inside a build-in step function. E.g., Outcome=step [f (A_{340}, A_{221}, A_{223}…, A_{850})], where outcome is a binary value that can be applied to the model, A_{340}, A_{341} … A_{850} are the absorbance at corresponding wavelengths from 340 to 850, and f (A_{340}, A_{221}, A_{223}…, A_{850}) is the algorithm that the software attempts to find following GP approach. The operators included in finding the algorithm in the step function were *constant, multiplication, addition, subtraction, division*, *exponential*, *natural logarithm*, *power* and *square root*. Absolute error was selected as the fitness to order the output algorithms but all other fitting parameters (*e.g.,* correlation coefficient, R^2) were also calculated. As *Eureqa* could generate algorithms very high complexity to fit the data, therefore Akaike Information Criterion (AIC – which penalizes complex models more heavily (Bozdogan, 1987)) was also applied to generate the candidate algorithms that have high fit but also relative less complicated structures.

Preliminary phase absorbance results were directly compared to the PCR results were applied to the GP model. In contrast, second phase absorbance results were standardize using the Skanit software (Thermo Fisher Scientific, version 5.0) where a blank substraction was performed. The corrected results were then compared to the presence and absence results of the MPN-PCR method as well as the confirmed presence/absence of *Campylobacter* in the samples tested using the traditional culture based method (AS/NZS, 2001). Further, GP was applied to define differences associated with factors including sample matrix and sample location.

Results and discussion

Ninety-five water samples where included in the preliminary phase of testing (Table 1). For the next phase, 2006 wells were analysed for both absorbance spectrum and *Campylobacter* PCR. Of these 2006, 754 also had corresponding results from the AS culture-based method.

Table 1. Summary of samples analysed using each method during the preliminary testing and the full testing.

	Number of wells tested		
Sample category	**Spectrum tested**	**Tested via PCR**	**Tested AS**
2019 (preliminary testing)			
Water	95	95	95
2020 (new spectrophotometer)			
Water	1440	1440	474
Sediments	62	62	58
Sewage	144	144	42
Animal Faeces	42	42	14
Positive controls	42	42	39
Negative controls	127	127	127
Blank	149	149	0
TOTAL	**2006**	**2006**	**754**

Results from the preliminary phase evaluation on the Yarra water samples provided early evidence of a significant correlation between measured absorbance and confirmed *Campylobacter* growth(R^2 = 0.391) (Table 2). However, the introduction of a larger number of samples, which included the 11 tubes of the MPN method rather than just three out of 11 tubes, showed less significant results for the water samples (R^2 = 0.292). The correlations were reduced when adding water samples from other locations than the Yarra River (0.214). These were lowered again when complex sample matrices such as sewage, sediment or animal faeces were included within analysis (0.106). Further, evaluation of this phenomenon and the investigation of different absorbance spectrum wavelengths and other models (e.g. neural networks) will be investigated in the future.

Table 2. Summary of modelling results

Datasets	**Candidate algorithm**	**R^2**	**Correlation coefficient**	**Mean Square Err.**
All data collected	Outcome = step (A_{664} * A_{665} - A_{641}*A_{832} - A_{832}^3)	0.106	0.558	0.197
	Outcome = step (A_{662} – sqrt (A_{641}* A_{741}))	0.097	0.574	0.199
All water samples	Outcome = step (1.195 * A_{665} *A_{803} - A_{652} *A_{848} - A_{840} * A_{848}^2)	0.214	0.606	0.177
	Outcome = step (A_{574} * A_{777} - A_{540} *A_{848})	0.194	0.595	0.181
Yarra water samples (2020)	Outcome = step (0.717*A_{537} *A_{723} - A_{469} *A_{828})	0.298	0.649	0.162
	Outcome = step (1.394*A_{586} - A_{472})	0.292	0.652	0.164
Yarra Water Samples (2019)	Outcome = step (A_{451} + A_{381} * A_{555} - A_{472} – exp (A_{366} *A_{602}))	0.391	0.700	0.124
	Outcome = step (A_{657} + A_{381} * A_{569} – exp (A_{366} *A_{603}))	0.361	0.680	0.130

Conclusions and future work

The results from this first stage study focusing on the larger volumes of the MPN method showed some promising result using the GP models. However, the inclusion of samples from a range of source waters as well as additional sample matrices reduced the significance of the observed correlations. Future work will include the impact of other water quality parameters, a deeper investigation of the impactful wavelength to outcome as well as other data-driven approaches to models (e.g. neural networks).

References

Aryafar, A. , Khosravi, v. Zarepourfard, H. and Rooki, R. (2019). Evolving genetic programming and other Al-based models for estimating groundwater quality parameters of Khezri plain, easter Iran. Environmentl Earth science 78(3),69.

Corvisy, R. (2013). NNDSS Annual Report Writing Group. Australia's notifiable disease Status, 2011: Annual Report of the National Notifiable Diseases Surveillance System. Communicable Diseases Intelligence.

Hughes, G. J., and Gorton, R. (2013). An evaluation of SaTScan for the prospective detection of space-time *Campylobacter* clusters in the North East of England. Epidemiol. Infect. 141, 2354–2364. doi: 10.1017/S0950268812003135

Henry, R. M., Schang, C. M-P., Chandrasena, K. K. G. I., Deletic, A., Edmunds, M., Jovanovic, D., ... McCarthy, D. T. (2015). Environmental monitoring of waterborne *Campylobacter*: evaluation of the Australian standard and a hybrid extraction-free MPN-PCR method. *Frontiers in Microbiology*, *6*(FEB), 1 - 13. https://doi.org/10.3389/fmicb.2015.00074

Koenraad, P. M. F. J., Rombouts, F. M., and Notermans, S. H. W. (1997). Epidemiological aspects of thermophilic *Campylobacter* in water-related environments: a review. Water Environ. Res. 69, 52–63. doi: 10.2175/106143097X125182

Schmidt, M. and Lipson, H. (2009). Distilling Free-Form Natural Laws from Experimental Data. Science 324(5923,81.

Schmidt, M. and Lipson, H. (2013). Eureqa (version 0.99.5 beta) [software]. [available at www.eureqa.com].

Zhang, k., Deletic, A., Bach, P., Shi, B., Hathaway, J.M. and McCarthy, D.T. (2019). Testing of new stormwater pollution build-up algorithms by a genetic programming appach. Journal of Environmental Management 241, 12-21.