In-situ tryptophan-like fluorescence: A real-time indicator of faecal contamination in drinking water supplies

J.P.R. Sorenson, D.J. Lapworth, B.P. Marchant, D.C.W. Nkhuwa, S. Pedley, M.E. Stuart, R.A. Bell, M. Chirwa, J. Kabika, M. Liemisa, M. Chibesa

British Geological Survey, Maclean Building, Wallingford, OX10 8BB, UK
British Geological Survey, Nicker Hill, Keyworth, Nottingham, NG12 5GG, UK
University of Zambia, Great East Road Campus, P.O. Box 32179, Lusaka, Zambia
Robens Centre for Public and Environmental Health, University of Surrey, Guildford, GU2 7XH, UK
Lukanga Water and Sewerage Company Limited, P.O. Box 81745, Kabwe, Zambia

Abstract

Enteric pathogens are typically inferred from the presence of surrogate indicator organisms such as thermostolerant (faecal) coliforms (TTCs). The analysis of TTCs requires time-consuming incubation in suitable laboratories, which can limit sampling resolution, particularly during critical pollution events. Here, we demonstrate the use of in-situ fluorimeters targeting tryptophan-like compounds as a rapid, reagentless indicator of TTCs in groundwater-derived potable water supplies in Africa. A range of other common indicators of TTCs were also determined including nitrate, turbidity, and sanitary risk survey scores. Sampling was conducted during both the dry and wet seasons to investigate seasonality. Tryptophan-like fluorescence was the most effective predictor of both presence/absence and number of TTCs during both seasons. Seasonal changes in tryptophan-like fluorescence in deeper supplies suggest it is transported more efficiently through the aquifer than TTCs. Moreover, the perennial elevated concentrations in some wells suggest it is more resilient than TTCs in groundwater. Therefore tryptophan-like fluorescence could also be a better indicator of some smaller, more easily transported, and long-lived, pathogenic enteric viruses. These sensors have the potential to be included in real-time pollution alert systems for drinking water supplies throughout the world, as well as for mapping enteric pathogen risks in developing regions.

© 2015 Published by Elsevier Ltd.

1. Introduction

Many pathogens transmitted through drinking water are of faecal origin and these continue to pose a threat to human health globally (Ashbolt, 2004). In the USA, up to half of all groundwater supplies have had evidence of faecal contamination resulting in 750,000 to 5.9 million cases of waterborne illnesses per year (Savichcheva and Okabe, 2006). Currently, the WHO Guidelines for Drinking Water Quality, adopted as standards in many countries, use the indicator group thermostolerant (faecal) coliforms (TTC), or specifically Escherichia coli, as a measure of the safety of drinking water supplies. Analysis for these organisms requires well-trained operators working with sterile equipment and reagents in laboratory conditions, which are not always easily accessible. Furthermore, sampling can be a significant hindrance in remote areas.
Technological advances now mean that portable fluorimeters targeting the TLF peak are commercially available as possible in-situ or roaming field sensors. Cumberland et al. (2012) demonstrated their potential in controlled laboratory conditions: observing tentative positive correlations with a range of indicator organisms, including E. coli, in river and effluent samples. Their use in groundwater is likely to be particularly promising given the general low turbidity and reasonably stable temperature, which can both significantly impact detectable fluorescent signals (Baker, 2005; Downing et al., 2012; Khamis et al., 2015).

This study is the first to pilot the use of in-situ TLF fluorimeters for the rapid assessment of the microbiological quality of drinking water supplies, with a focus on groundwater in the developing world. The aims of this work were to investigate the application of a portable fluorimeter to indicate i) whether a groundwater supply was faecally contaminated; ii) the extent of any contamination; and iii) compare TLF as an indicator of contamination against other common indicators such as NO₃, CI, turbidity, and sanitary risk scores (SRS). Faecal contamination was determined through the presence of TTCs.

2. Materials and methods

2.1. Study site

Kabwe is located in Zambia’s Central Province approximately 150 km north of the capital Lusaka. It has a population of over 200,000, with a high proportion residing within informal settlements on the outskirts of the city such as Makululu — regarded as one of the largest slums in southern Africa with an estimated 46,000 inhabitants (LgWSC, 2014). The city is predominantly underlain by several hundred metres of either the Lower Roan Group (quartzite, schist and pelite) or Upper Roan Group (dolomite). The bedrock is concealed beneath continuous saprolite and laterite superficial deposits that are typically 5–20 m thick (Houston, 1982). Groundwater is generally encountered 5–10 m below ground level, with the superficials typically in hydraulic contact with the deeper aquifer within the karstic bedrock. The local climate is sub-tropical with rainfall exhibiting strong seasonality: 95% falls between mid-November and mid-April (Nkuwu et al., 2006). Natural surface waters are absent, as rainfall rapidly infiltrates into the subsurface.

Groundwater is the major source of drinking water supply for the city. The centralised supply system abstracts groundwater from deep boreholes within peri-urban wellfields, which is then treated and piped directly to properties, or dispersed via communal taps and water kiosks within the informal settlements. Households frequently also self-supply groundwater to some extent as the centralised supply can be unreliable and is charged on a per volume basis. Within informal settlements this is generally restricted to vulnerable shallow hand-dug wells and illegal connections to pre-treated water within the centralised supply network. In more numerous, but less hazardous, schemes, the raw water is discharged into stream beds or swamps. This may sometimes begin to encroach into the wellfield areas, with concerns over the potential threats to the city's groundwater resources in the medium to long-term.

2.2. Groundwater sampling and analysis

A total of 117 groundwater samples were obtained from a mixture of supplies that were distributed across the city (Fig. 1). These were composed of 55 samples in the dry season (September 2013) and 62 in the subsequent wet season (January 2014), of which 45 were obtained from the same supplies. The dry season sampling included 25 boreholes and 30 shallow wells, whereas in the wet season 26 boreholes and 36 shallow wells were investigated. These supplies included the city wellfields (K1-12), a mixture of private supplies within both higher and lower cost residential areas, as well as those in the industrial zone (K26-28).

Groundwater samples were obtained once field measurements of pH, specific electrical conductance (SEC), Eh, dissolved oxygen (DO) and temperature had stabilised during pumping. Turbidity was also measured on an agitated pumped sample, as settling was rapid. In-situ TLF was undertaken by sampling 5 L of groundwater and immersing a portable UviLux Fluorimeter (λ_em = 280 nm, λ_ex = 360 nm) (Chelsea Technologies Group Ltd, UK) in the dark. The sensor utilises a UV light emitting diode (LED) light source and photomultiplier allowing a high level of sensitivity. The minimum detection limit for these sensors is 0.17–0.19 µg/L (Khamis et al., 2015) and repeatability is within ± 0.12–0.29 µg/L up to a concentration of 50 µg/L. (Table S1).

The factory calibration was used which showed a good agreement with a range of ten synthetic standards and a bench Varian Cary Eclipse fluorescence spectrophotometer for the same wavelength pair at 20 °C (Fig. 2). Standards were produced by dissolving laboratory grade L-tryptophan (Acros Organics, USA) in ultrapure water. An excellent linear relationship between the portable (in µg/L) and bench (in Raman units) fluorimeters (R² = 0.9952) can be used to convert all concentrations on the portable device to Raman units by multiplication by 0.0024 at 20 °C.

Microbiological samples were collected in sterile 60 ml brown-glass bottles. Brief interviews with the well owners, or other informed persons, confirmed that the supplies sampled had not been recently chlorinated, or were chlorinated at a point further down the distribution system. Samples were stored in a cool box and transported back to the laboratory for analysis. All samples were processed within seven hours of collection.

Thermotolerant coliforms (TTCs) were isolated and enumerated using the membrane filtration method and Membrane Lauryl Sulfate Broth (MLSB, Oxoid Ltd) as the selective medium. Typically 50 ml (giving a limit of detection of 2 c.f.u/100 ml), or an appropriate dilution of the sample, were filtered through a 0.45 µm nitrocellulose membrane (PALL Gelman). The membrane was transferred on to an adsorbent pad saturated with MLSB, pre-incubated for approximately one hour at ambient temperature, then placed on to a rapid plate reader (Tecan).
2.3. Sanitary risk score

A sanitary risk assessment was undertaken at every supply during the wet and dry season surveys. The assessment was carried out using the methods published by the WHO and the sanitary risk assessment forms appropriate to the supply type (WHO, 1997). The supplies and surrounding areas were assessed for the presence of defined hazards, such as latrines and waste dumps, and pathways for contaminants to reach the sample point, using pre-set questions. The sanitary risk score (SRS) was calculated from the number of positive responses.

2.4. Statistical analyses

Statistical models and tests were performed with MATLAB v14.1 to determine the extent to which TLF concentration and other predictors indicate the presence of TTC contamination and the accuracy with which they could be used to model TTC counts. The TTC count in water supply $i$ was denoted $y_i$, and the binary variable, $y_i$, indicated whether or not the TTC exceeded the detection limit of 600000 microorganisms per litre.
require the predictors to conform to a particular statistical distribution. Therefore, it was preferred to the T-test which requires the assumption of a Gaussian distribution. In each test the null hypothesis was that the expected value of the predictor was the same for contaminated and uncontaminated supplies.

2.4.2. Predicting the presence/absence of TTCs

The probability that a water supply was contaminated with TTCs was represented by a logistic regression model (Dobson, 2001). If \( p_i \) is the probability that there is contamination at supply \( i \), then the logistic model is written:

\[
\ln \left( \frac{p_i}{1 - p_i} \right) = \beta_0 + \sum_{j=1}^{X_i} \beta_j x_{ij}
\]

where \( x_{ij} \) is the value of predictor \( j \) at water supply \( i \) and the \( \beta_j \) are coefficients which, in this study, were estimated by least squares (Draper and Smith, 1981). A stepwise regression algorithm (Draper and Smith, 1981) was used to decide which of the available predictors should be included in the model. This approach guards against too many predictors being included. In such circumstances the model is said to be overfitted, meaning that it is overly suited to the intricacies of the data set to which it has been fitted but will not achieve the same accuracy on independent validation data.

The stepwise algorithm adds one predictor to the model at a time. It considers all of the available predictors not yet included in the model in turn and selects the one which causes the largest decrease in the mean squared residuals. If this decrease is deemed significant for a \( p \)-value of 0.05, according to a F-test (Draper and Smith, 1981), this parameter is added to the model. The iterative process continues with the remaining parameters until none of them lead to a significant improvement to the model. The stepwise algorithm was implemented using the MATLAB function stepwiselm. The observed values of TLF were highly skewed (Figure S1). Therefore these data were log transformed prior to the estimation of regression models. Note that a positive shift of 0.1 was required to ensure that the logarithm was always defined.

A classifier of TTC contamination was then formed by designating all water supplies with \( p_i \) greater than some threshold \( \hat{p} \) as contaminated. There is uncertainty associated with such a classification. False positive errors occur when the classifier erroneously suggests that the supply is contaminated whereas false negative errors occur when the supply is erroneously classified as uncontaminated. Similarly, true positive and true negative results refer to correctly classified contaminated and uncontaminated supplies.

The effectiveness of a classifier can be assessed in terms of the receiver operating characteristic curve (Hanley and McNeil, 1983). This is a plot of the proportion of true positive results against the proportion of false positive results as \( \hat{p} \) is varied. The area under this curve is a measure of the effectiveness of the classifier. It will be 1 for a perfect classifier and 0.5 if the classifier is performing no better than a random choice.

estimated for TTC counts, the residuals were highly skewed, contrary to the assumptions required by the model. Therefore a model was instead estimated for the logarithm of the TTC counts.

A regression model such as Equation (2) determines the entire probability density function for the TTC count at a water supply from the values of the predictor variables from that supply. Therefore, the model can be used to determine the probability that the TTC count falls within particular intervals. In this case, the five risk classes outlined by the WHO (1997) to prioritise interventions in areas where microbiological contamination of water supplies is widespread were used.

3. Results

3.1. Characteristics of TTC contaminated water supplies

The results of the Mann-Whitney rank Sum tests suggested that the mean TLF concentration was significantly (\( p < 0.001 \)) larger in water supplies contaminated with TTCs as opposed to water supplies where they were absent (Figure 3). The relationship remained significant at the \( p < 0.001 \) level when the supplies from each season were considered separately and when only boreholes were considered. No significant difference was evident when only shallow wells were considered but, since there were only five shallow wells where TTCs were absent, this finding is likely to be the result of insufficient data. Overall, the median TLF concentrations were 0.9 and 9.2 \( \mu \)g/L in uncontaminated and contaminated supplies, respectively.

Other variables showing a highly significant difference (\( p < 0.001 \)) included SRS, turbidity, depth to water, Cl and SO\(_4\) (Figure 3). Phosphate species or DOC were not appreciably elevated in contaminated water supplies, although NH\(_4\) was significant \( (p = 0.019) \). Temperature, pH and turbidity typically ranged between 23.8 and 27.7 °C, 6.4-7.4, and 1-20 NTU, respectively, across the dataset. All variables correlated with faecal contamination were also significantly correlated with TLF concentration (notably SRS, NO\(_3\), and NH\(_4\)) with no relationship observed with DOC, pH, DO, or temperature (Figure S2).

3.2. Predicting the presence/absence of TTCs

Transformed (and raw) TLF was a significant predictor (\( p < 0.001 \)) of the presence/absence of TTC according to the logistic regression model (Equation [1]). Other indicators were significantly correlated with the probability of TTC contamination (Table 1), including Cl and the SRS. However, transformed TLF was the first predictor to be included in the logistic regression model and the results of the F-tests demonstrate that the addition of other predictors did not significantly improve the model (Table S2). The findings were the same when data from the wet and dry seasons were considered individually, as well as when boreholes and shallow wells were considered separately. Therefore, for the predictions the data were combined.
Fig. 3. Box plots of transformed tryptophan-like fluorescence and other variables for the presence or absence of thermotolerant coliforms. P-values are denoted as * <0.05, ** <0.01, and *** <0.001. Dots indicate 5th and 95th percentiles.

Table 1
Estimated coefficients and p-values for the hypothesis that the coefficients are zero for single-predictor models of probability of the presence and number of thermotolerant coliforms. Areas under the receiver operator characteristic curves for single predictor classifiers are also shown. Classifiers are only formed for predictors that are significantly (p < 0.05) related to contamination.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Logistic regression</th>
<th>Linear regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Area</td>
<td>β</td>
</tr>
<tr>
<td>Transformed TLF</td>
<td>0.92</td>
<td>2.12</td>
</tr>
<tr>
<td>SRS</td>
<td>0.90</td>
<td>0.61</td>
</tr>
<tr>
<td>Turbidity</td>
<td>0.76</td>
<td>0.48</td>
</tr>
<tr>
<td>Depth to water</td>
<td>0.80</td>
<td>2.86</td>
</tr>
<tr>
<td>CI</td>
<td>0.86</td>
<td>1.63</td>
</tr>
<tr>
<td>NO₃</td>
<td>0.73</td>
<td>0.35</td>
</tr>
<tr>
<td>NO₃-Cl</td>
<td>NA</td>
<td>-0.34</td>
</tr>
<tr>
<td>SO₄₂⁻</td>
<td>0.71</td>
<td>0.31</td>
</tr>
<tr>
<td>TP</td>
<td>NA</td>
<td>0.25</td>
</tr>
<tr>
<td>SRP</td>
<td>0.63</td>
<td>-0.41</td>
</tr>
<tr>
<td>NH₄</td>
<td>NA</td>
<td>2.88</td>
</tr>
<tr>
<td>DOC</td>
<td>NA</td>
<td>-0.74</td>
</tr>
<tr>
<td>SEC</td>
<td>NA</td>
<td>0.00</td>
</tr>
<tr>
<td>DO</td>
<td>NA</td>
<td>-0.17</td>
</tr>
</tbody>
</table>

Fig. 4. (A) Receiver operator characteristic curve for classifier of the presence or absence of TTC using transformed TLF as the only predictor. (B) Error rates for this classifier.
0.67 and the final model is written:

\[
\ln y_i = -0.34 + 1.11 \ln(TLF + 0.1) + 0.35 \text{NO}_3^+ + 0.25 \text{SRS}
\]  

(4)

The addition of other predictors did not significantly improve the model (Table S2). This might indicate that the predictors included in Equation (4) accounted for the observed significant relationships.

Additionally, there is an increase in the median transformed TLF concentration with elevated WHO risk class, although the levels of scatter produce overlap between all classes (Fig. 5a). Nevertheless, the single-predictor linear regression model using TLF can be used to estimate the probability of a water supply belonging to each risk class (Fig. 5b). This highlights that as TLF concentration increases there is a greater probability of the water supply being within higher risk groups.

3.4. Risk classes for water supplies across Kabwe estimated by TTC count or modelled TLF

Deeper boreholes are considered less at-risk than shallow wells across the city using either modelled TLF or TTC count (Fig. 6). Boreholes are most at-risk within the central districts, with these risks elevated in the wet season, and higher when allocated using modelled TLF compared to the TTC count. There are also notable seasonal changes in risk in the city’s centralised peri-urban water supply boreholes (K1-12) using modelled TLF which are not apparent from the TTC data (Fig. 6). In the dry season, there are no risks by TTC count or modelled TLF in any of the boreholes, with the exception of K3 and K11 being considered as low risk by modelled TLF. Following the onset of the rains, modelled TLF suggests increases in risk at the majority of these boreholes, including some located beyond the city limits (K2, K6, and K12). Furthermore, high risks are indicated at those sites closest to the urban area (K8, K9, and K11). On the other hand, TTC data only suggest elevated medium and low wet season risks at K8 and K9, respectively.

Shallow wells are typically at medium or greater risk perennially in the central districts of Kabwe by modelled TLF (Fig. 6). However, TTC counts suggest many of these wells have no risk during the dry season, but medium or greater during the wet season. Nevertheless, wells within Makululu generally remained at least high risk during both seasons using either modelled TLF or TTC count. Elsewhere, other wells also considered at a minimum of high risk by both TTC count and modelled TLF in the dry season, predominantly remain so in the wet season.

4. Discussion

4.1. Uncertainty in TTC counts and TLF concentrations

Doubts have been expressed about the reliability of TTCs to indicate faecal contamination in water. Although the TTC group includes the species E. coli, which is generally considered to be specific for faecal contamination, it also includes other genera such as Klebsiella and Citrobacter which are not necessarily of faecal origin and can emanate from alternative organic sources such as decaying plant materials and soils (WHO, 2011). Nevertheless, studies have shown that greater than 90% of thermotolerant coliforms are E. coli (Dufour, 1997 cited in Leclerc et al. (2001)) and as high as 99% in groundwater impacted by poor environmental sanitation in Africa (Howard et al., 2003). Therefore this is not considered a significant source of uncertainty.

The uncertainty of measurement of the microbiological analysis was calculated using the RSDRC, which is the average relative standard deviation of reproducibility of a series of duplicate analyses. The RSDRC calculated for the duration of the sampling programme was 0.1433. Using a coverage factor of 2.056, which was determined from the number of duplicate analyses used to calculate RSDRC, the 95% confidence intervals for any result could be calculated. For example, a result of 100 c.f.u./100 ml would have an upper and lower limit of 197 and 51 c.f.u./100 ml respectively. This figure is slightly higher than might be anticipated in a laboratory accredited to International Standard ISO17025:2005, but is not unusual for microbiological analysis.

False-negatives and underestimates of TTCs may occur where the environment conditions within certain groundwater supplies are unfavourable for the survival of bacteria, but TLF is less affected. There were three supplies where TTCs were absent but TLF exceeded 10 μg/L. At two of these sites the water differed from a typically neutral pH; being 5.1 and 10.5 in supplies K40 and K55, respectively. Moreover, K40 also contained high concentrations of heavy metals which have a toxic influence on bacteria (Poppen and Schijven, 2006), yet also quench TLF fluorescence to a degree (Tabak et al., 1989). These two sites actually account for two of the five shallow wells where TTCs were absent.

It was confirmed that none of the supplies had recently been chlorinated before sampling. However, a parallel study confirmed the presence of trihalomethanes (by-products of chlorination) at K40, K45, and K14, although this
(2015) demonstrated a 0.65 and 1.14 µg/L reduction in TLF for every 1 °C increase in temperature in 10 and 20 µg/L solutions, respectively. This equates to uncertainty of 22–25% across the typical temperature, and it should be re-iterated that there was no observed inverse relationship between TLF and temperature (Figure S2). pH values are within the 5–8 bracket of minimal risks with the exception of highly alkaline and acidic soils.  

It is also likely to remove a proportion of TLF, given its direct association with bacteria (Elliott et al., 2006a). However, a fraction of the fluorescence is likely to be in a free dissolved form (Baker et al., 2007), which could mean it is transported more easily and rapidly through porous media. There is evidence for this in Kabwe where increases in modelled TLF risks were greater than TTC risks with the changes epidemiologically linked to the source of the contamination.
for 16–45 days (Taylor et al., 2004). These rates of inactivation are similar for many pathogens but enteric viruses generally inactivate at a slower rate (John and Rose, 2005). Therefore, elevated TLF could be a better indicator of pathogens which are more long-lived in the environment, thus providing a more precautionary estimate of pathogen risk.

It is unsurprising that TLF is more resilient than TTCs in groundwater as it is contained within organisms. Therefore, as TTCs and other organisms die-off following their release into groundwater (an unfavourable habitat) many cells will rupture and release TLF in dissolved and more complex phases into the environment. It may remain present in groundwater in these forms as groundwater ecosystems are less active in comparison to surface waters, or be taken up by other microorganisms. Either way, it will continue to contribute to a TLF signal. Furthermore, it has been demonstrated that under nutrient limited conditions E. coli release dissolved tryptophan, as they transit from a culturable to a dormant viable but nonculturable state (Arana et al., 2004).

5. Conclusions

This is the first study to investigate the use of in-situ tryptophan-like fluorescence (TLF) for the rapid assessment of the biological quality of drinking water supplies. TLF was significantly elevated in supplies where thermotolerant coliforms (TTCs) were detected, alongside many other traditional indicators, but was demonstrated to be the most effective indicator of TTC presence/absence. TLF was also the most significant indicator of the number of TTCs, although the linear regression model could be improved through the inclusion of NO₃ and sanitary risk scores (SRS). A single-predictor linear regression model using TLF was used to estimate the probability of a water supply belonging to each WHO risk class. This highlights that as TLF concentration increases there is a greater probability of a water supply being within the higher risk groups.

The use of in-situ TLF sensors has multiple methodology advantages over bacterial indicators for inferring the presence of enteric pathogens. TLF is likely to be more mobile and resilient in groundwater and, as such, a more precautionary indicator of enteric pathogens in groundwater, including where bacterial indicators are absent. Importantly, TLF has the added advantage of potentially detecting the presence of TTC in viable but nonculturable states. Furthermore, the sensors require no reagents and provide instantaneous readings. This could facilitate their inclusion in real-time pollution alert systems for drinking water supplies throughout the world, for the rapid mapping of enteric pathogen risks in developing regions, and as an initial screening tool to inform and complement further water quality investigations.

Acknowledgements

We are grateful to Makondo Wenycvslow, Managing Director of
basin rivers. Water Res. 36 (10), 2571–2581.