Carroll, Steven Paige and Dawes, Les A. and Hargreaves, Megan and Goonetilleke, Ashantha (2009) ***Faecal pollution source identification in an urbanising catchment using antibiotic resistance profiling, discriminant analysis and partial least squares regression***. Water Research, 43(5). pp. 1237-1246.

Author version of paper published as:
Carroll, Steven P., Dawes, Les, Hargreaves, Megan and Goonetilleke, Ashantha (2009) Faecal pollution source identification in an urbanising catchment using antibiotic resistance profiling, discriminant analysis and partial least squares regression.
Water Research 43(5): pp. 1237-1246

Faecal pollution source identification in an urbanising catchment using antibiotic resistance profiling, discriminant analysis and partial least squares regression

Carroll, S.P.[1], Les Dawes[2], Megan Hargreaves[3], Goonetilleke, A.[2]

[1]South Burnett Regional Council, Kingaroy, Queensland, Australia
[2]School of Urban Development, Queensland University of Technology, Brisbane, Australia
[3]School of Life Sciences, Queensland University of Technology, Brisbane, Australia

ABSTRACT

Increasing urbanisation and changes in land use leads to adverse impacts on the quality of natural water resources. The specific sources of contamination are often difficult to identify using conventional water quality monitoring techniques. This acts as a significant constraint to the development of appropriate management techniques to protect natural water resources. Consequently, alternative means of identifying pollutant sources and their locality are necessary. In this study, Antibiotic Resistance Patterns (ARP) were established for a library of 1005 known *E. coli* source isolates obtained from human and non-human (domesticated animals, livestock and wild) sources in an urbanising catchment in Queensland State, Australia. Discriminant Analysis (DA) was used to differentiate between the ARP of source isolates and to identify the sources of faecal contamination. Partial Least Square (PLS) regression was then utilised on identified human source isolates to correlate their locality with specified sampling locations within the catchment. The resulting ARP DA indicated that a majority of the faecal contamination in the rural areas was non-human. However, the percentage of human isolates increased significantly in urbanised areas using onsite systems for wastewater treatment. The PLS regression was able to develop predictive models which indicated a high correlation of human source isolates from the urban area. The study results confirms the feasibility of using ARP for source tracking faecal contamination in surface waters, as well as predicting their point of origin.

Keywords: Onsite Systems, *E. coli*, Antibiotic Resistance Analysis, Discriminant Analysis, Partial Least Squares Regression

INTRODUCTION

Increasing urbanisation and other land use changes in the southeast region of Queensland State, Australia has resulted in adverse impacts on the quality of natural water resources. With urbanisation representing the dominant land use change, the resulting non-point sources of contamination can have a significant impact on surface water quality. However, using conventional water quality monitoring techniques, the specific sources of contamination are often hard to identify in order to develop appropriate mitigation management strategies.

One of the most commonly suspected sources of faecal contamination of water resources are onsite wastewater treatment systems (OWTS), particularly septic tank-soil adsorption systems. Increased urbanisation in the fringes of metropolitan areas has led to the reliance on onsite wastewater treatment systems for the treatment and dispersal or sewage effluent. Numerous studies have found that inadequate soil properties, inappropriate site location and poor management and maintenance techniques can lead to numerous scenarios of failing systems. This can result in the contamination of ground and surface water resources due to the percolation of inadequately treated sewage effluent from soil based effluent disposal areas (Harris 1995, Paul et al 1997, Young and Thackston 1999, Paul et al 2000, Lipp et al 2001, Pang et al 2003). Microbiological contamination of water resources are of critical concern due to public health risks (Hagedorn et al 1999, Wiggins et al 1999). However, due to the numerous possible sources of faecal bacteria, it has until recently been difficult to isolate onsite systems as a significant source of faecal pollution.

In order to effectively manage the inherent risks resulting from sewage effluent contamination, not only is the identification of the different sources of contamination crucial, but also predicting their locality. The most recent methods for identifying sources of faecal contamination are based on the use of bacterial source tracking (BST) techniques. Several BST methods have been trialled in recent years with limited success (Hagedorn et al 1999, Meays et al 2004). These include: calculating the ratio of faecal coliform to faecal streptococci (Pourcher et al 1991, Howell et al 1996); determining proportions of thermotolerant coliforms to faecal sterols (coprostanol and 24-ethylcoprostanol) (Leeming et al 1998); and species differentiation of faecal streptococci amongst various animals (Devries et al 1993). More recent BST methods have employed molecular methods such as genetic makeup profiles of specific bacteria isolates, including random amplified polymorphic DNA or rep-PCR DNA extraction methods (Parveen et al 1999, Dombeck et al 2000). Additionally, the physiological characteristics used in biochemical BST techniques, such as Antibiotic Resistance Patterns (ARP) of different sources of faecal bacteria have also been used (Wiggins 1996, Hagedorn et al 1999, Whitlock et al 2002,). In this study, the use of Antibiotic Resistance Pattern (ARP) analysis was employed.

ARP essentially utilises the resistance of selected faecal bacteria isolates, in this case *Escherichia coli (E. coli),* to several antibiotics at varying concentrations to obtain their resistance profiles. The underlying assumption in the ARP technique is that due to the increased use of antibiotics by humans and domesticated animals, isolated *E. coli* bacteria from these host sources will have a relatively higher resistance than that of wild animals (Wiggins 1996). The ARP technique requires a library of known *E. coli* isolates, from human and non-human sources, to be tested for their respective ARP. These are then analysed statistically using multivariate analytical techniques such as discriminant analysis in order to separate the respective patterns into source groups. Appropriate validation of the library of source isolates and statistical methods used for analysis is essential to ensure that unknown isolates are correctly classified. This was achieved by undertaking a cross-validation procedure for all known source isolates in the developed library. This procedure randomly removes isolates from the known source library and treats them as an unknown source to test the classification ability (Harwood et al 2000).Once the source library has been developed, *E. coli* from the investigated water samples are tested for their ARP and compared to the source library and categorised according to the respective grouping of known source isolates with similar ARPs.

However, although the use of these methods, in particular ARP, are useful for identifying the particular host source of the bacteria being investigated, most studies have not been able to demonstrate the actual locality of sources themselves. This is difficult due the high number of environmental variables and flow conditions that are involved in the transport of microorganisms from the source to the point of monitoring. However, by applying additional multivariate statistical methods, such as Partial Least Squares (PLS) regression to the collated data from studies undertaken, not only can the bacteria sources be identified, but also their locality can be predicted based on the position of targeted monitoring points. Targeted sampling of surface waters allows for variability in faecal bacteria concentrations during differential flow conditions (e.g. rainfall events and tidal fluctuations) to be accommodated in the analysis, thereby allowing identified faecal pollution point sources to be more successfully modelled (McDonald et al 2006, Hartel et al 2005, Kuntz et al 2003).

The main focus of the study discussed in this paper was to apply the ARP technique for determining both, the potential sources of faecal contamination in a mixed landuse catchment as well as the location of the sources. The study was undertaken in Ningi Creek catchment, Caboolture Shire, Queensland State, Australia.

MATERIALS AND METHODS

Study Catchment and Location of Monitoring Sites

Ningi Creek catchment covers 72km^2 and consists of mixed land use including urban, agricultural, pine forestry and natural bushland. At present, the catchment is experiencing significant urban development. The urbanised areas in Ningi Creek catchment are all serviced with OWTS, and their cumulative effect has become a major concern for the region's local government due to increased pollution of the waterway.

Seven surface water monitoring sites (SW1-SW7) were established for determining the level of faecal pollution in Ningi Creek, and for the collection of *E. coli* isolates for source discrimination. Five groundwater monitoring wells (GW1-GW5) were established in an urban residential area to assess the level of faecal pollution in groundwater directly below residential area which uses OWTS for sewage treatment. Water samples were collected on a monthly basis over a twelve month period. Figure 1 shows the locations of the monitoring sites and the corresponding catchment details.

Figure 1.

Sample Collection

A total of 84 surface water samples and 60 groundwater samples were collected on a monthly basis over a twelve month period from each of the surface water (SW1-SW7) and groundwater (GW1-GW5) monitoring locations. This sampling period was selected to allow the collection of samples during both the drier winter period following into the spring wet season. Water samples were collected at regular monthly intervals both during base flow conditions and after rainfall events as they occurred. Generally, faecal bacteria concentrations increased following rainfall events which can significantly alter concentrations in surface waters. However, by adopting the targeted sampling approach, the influence of flow variations were negated. Additionally, samples were collected following the peak of high tide to limit influence on samples caused by tidal flows. This ensured that the direction of flow during sampling was consistently in the same direction from upstream (SW7) to downstream (SW1). Sites were sampled in chronological order (SW1 to SW7) to avoid sediment disturbance impacting on the samples collected. Water samples were collected in sterilised glass bottles, stored and transported in crushed ice until analysis could be undertaken. All samples were analysed within 8 hours of collection.

Development of Source Library for Discriminant Analysis (DA)

To discriminate between the different sources of faecal bacteria in collected water samples, a source library of antibiotic resistance patterns of known source isolates was required. Faecal samples were collected from human and the primary non-human sources of faecal matter within the catchment. Samples used for the development of the source library were collected during each groundwater and surface water sampling episode. However, in order to maximise the source library, faecal samples were additionally collected randomly throughout the twelve month sampling period. The collection of random samples provided for extra variability in the source data. Five faecal samples were collected directly from human subjects in order to ensure that known human *E. coli* isolates were obtained. Eight additional human faecal samples were also collected from onsite wastewater treatment systems within the catchment, as well as from a local municipal wastewater treatment plant. The main reason for collecting faecal samples directly from humans as well as from sewage treatment facilities was to compare the accuracy of the methodology adopted for source identification. Even though the majority of *E. coli* isolates collected from the onsite wastewater treatment facilities would be of human origin, there is a possibility of cross-contamination with non-human *E. coli* isolates, such as from birds and rodents.

Major non-human faecal sources were identified and collected throughout the sampling phase, including livestock, domestic and wild animal sources observed near monitoring locations. Nineteen faecal samples were collected representing the three major sources of domesticated animals in both catchments, including dogs, cats and poultry. Additionally, fourteen livestock faecal samples representing beef and dairy cows, horses and goats were obtained from agricultural farms within the catchment. All livestock animals within the catchment are grass fed, with faecal samples collected from fresh manure piles dispersed throughout the grazing pastures. Fifteen faecal samples representing five wild animal sources were collected to obtain a random representation for the whole of the catchment. Sources included kangaroo, wallaby, koala, possum, and waterfowl. All these sources were observed in the catchment, with faecal samples collected from observed resting or roosting sites.

Microbiological and ARP Analysis

Collected sewage and faecal samples from known sources for developing the source library were tested using membrane filtration techniques. Isolation of *E. coli* from faecal samples obtained from known sources was achieved by adding 1.0g of faecal matter or 1.0mL of effluent sample to 100mL of sterile buffered dilution water (0.0425g L^{-1} KH_2PO_4 and 0.4055 g L^{-1} $MgCl_2$ in 100 ml distilled water) and vortexing for one minute (APHA 1999). Serial dilutions of 10^{-2} and 10^{-4} were prepared in buffered dilution water, and 1mL, 10mL and 90mL of the 10^{-4} dilution were filtered for analysis. For collected water samples, volumes ranging from 0.1mL to 100mL were filtered to permit isolated colonies on each plate.

Filtration was performed for both faecal and water samples, using 0.45µm, 47mm sterile gridded filter membranes (Millipore Corporation, Bedford, MA). Following filtration of each sample, the membranes were aseptically transferred to petri-pads soaked in M-Endo medium (Millipore Corporation, Bedford, MA) and incubated at 30°C for 24 hours. The filter funnel apparatus was treated with 70% Ethanol between uses, and then washed thoroughly with sterile distilled water. To validate the filtration technique, sample blanks using sterile distilled water as the sample were prepared during each sampling episode to ensure cross contamination did not occur. Where sample blanks indicated cross contamination, all samples collected during that episode were re-tested within a 24 hour timeframe. The Filter funnels were autoclaved between sampling episodes. Following 18 - 24 hours incubation, plates with isolated colonies were selected for use in isolation of putative *E. coli*. Colonies with a metallic sheen were taken to indicate putative *E. coli.* These colonies were sub-cultured onto Nutrient agar plates, and then further tested for Indole reaction, (Growth in Tryptone water at 37°C for 24 hours followed by addition of Kovac's Indole Reagent) and for growth plus gas production at 44.5°C in Brilliant Green Lactose Bile Broth (BGLBB) (Eijkmann test). In the case of a large number of sheened colonies being present, the number of colonies selected for isolation was taken as equal to the square-root of the number of colonies present. Those isolates with a positive reaction to both tests were recorded as confirmed thermotolerant *E. coli.*

ARP analysis was used to identify the different sources of faecal contamination in ground and surface water, with the main aim of identifying human from non-human sources. The process used for determining the respective ARP of *E. coli* followed the procedure outlined by Harwood et al (2000) and Whitlock et al (2002). Antibiotic stock solutions were prepared from available commercial antibiotics (Sigma Chemical Co. St Louis) and applied to sterile trypticase soy agar (TSA) prior to pouring into 150 mm sterile petri dishes. Each petri dish contained one specific concentration of each antibiotic. The antibiotics used and their respective concentrations are as follows; Amoxicillin (5, 10, 15 and 20μg L^{-1}); Cephalothin (10, 25, 50 and 100μg L^{-1}); Erythromycin (20, 50, 100 and 200μg L^{-1}); Gentamicin (20, 40, 60 and 80μg L^{-1}); Ofloxacin (5, 10, 15, and 20μg L^{-1}); Chlortetracycline (20, 40, 60 and 80μg L^{-1}); Tetracycline (20, 40, 60 and 80μg L^{-1}); and Moxalactam (5, 10, 15 and 20μg L^{-1}). The choice of antibiotics used in this study was based on their common use in humans and domesticated animals.

Isolates selected as having sheened colonies on m-Endo, and both Indole and Eijkmann positive, were included for ARP profiling. The isolates were inoculated into nutrient broth and incubated for 18 hours at 37°C. Subsequent broths were diluted to 0.5 MacFarland Standard in fresh nutrient broth. The diluted isolates were placed in multipoint inoculator cups (Denley Multipoint Inoculator A400) for inoculation onto a series of 32 antibiotic plates (8 antibiotics, 4 different concentrations), plus one TSA medium blank. The plates were incubated at 37°C for 24 hours.

After incubation, each plate of isolates was inspected and the relative growth for each antibiotic and concentration was recorded. Four different ratings (1 to 4) were utilised to distinguish respective ARPs. An isolate received a rating of (1) for no growth; (2) for filmous growth; (3) for restricted growth of colonies (growth of a few colonies); and (4) for full growth of colonies. The main reason for using the four ratings was to include more variability into the patterns than would be achieved through the use of two values (for example, 1 for no growth and 2 for full growth). These ARP ratings were utilised for discriminating between the respective source isolates.

DA and PLS of Antibiotic Resistance Patterns

Antibiotic resistance patterns for each of the sources and unknown *E. coli* isolates (based on the 1-4 scale for growth) were input into a spreadsheet and analysed using Discriminant Analysis (DA) with StatisiXL ver1.4 software (Roberts and Withers 2004). DA is a multivariate statistical analysis technique where a data set containing *X* variables is separated into a number of pre-defined groups using linear combinations of analysed variables. This allows analysis of their spatial relationships and identification of the respective discriminative variables for each group (Wilson 2002). Objects that retain similar variances in the analysed parameters will have similar discriminant scores, and therefore when plotted, will group together. Also relationships between variables can be easily identified by the respective coefficients. Strongly correlated variables will generally have the same magnitude and orientation when plotted, whilst uncorrelated variables are typically orthogonal to each other.

There are two main functions for which DA is commonly employed, and is most beneficial for ARP analysis. Firstly, it can be used to analyse the differences between two or more groups of multivariate data using one or more discriminant functions in order to maximally separate the identified groups. Secondly, DA can be employed to obtain linear mathematical functions which can be used to classify the original data, or new, unclassified data, into the respective groups (Brereton 1990). This classification procedure can be used to calculate the percentages of misclassified isolates and determine the average rate of correct classification (ARCC) of isolates in their respective categories (Wiggins 1996). To provide a more rigorous predictive capability for the source library, a cross-validation procedure (also referred to as *hold-out analysis* or *jack-knifing*) was undertaken. This procedure randomly removes isolates from the known source library and treats them as an unknown source to test the classification ability of the library (Harwood et al 2000). The process utilised in this study followed similar procedures to the pulled-sample cross-validation process described by Wiggins et al (2003). Any identified clonal source isolates in the developed library were removed prior to DA and validation procedures being undertaken. As multiple isolates from the same sample may have similar resistance profiles, the library may appear to be more representative due to this profile similarity. To overcome this issue, all isolates from the same sample were removed during the pulled-sample cross-validation procedure, and reclassified according to the resistance profiles of the remaining isolates. For the human versus non-human pooled analysis, five random samples from the human category and ten from the non-human category were individually pulled out and reclassified. Once the classification of source isolates was completed and isolates identified as human or non-human, the ARP of confirmed human isolates were submitted to a PLS regression analysis to predict their place of origin.

Antibiotic resistance patterns were submitted to Partial Least Squares (PLS) regression using MATLAB Release 13. PLS regression is a method for comparing two data sets; X (predictor variables) and Y (matrix of response variables) by a linear multivariate model (Wold et al 2001). The use of PLS models in multivariate data analysis is of particular value. Unlike more common methods such as multiple linear regression (MLR), PLS can analyse data sets that are strongly collinear, may be noisy (in the case of environmental data sets due to sudden changes in concentrations) and numerous X-variables (Wold et al 2001) that are difficult to assess on their own merits. Essentially, the use of PLS allows a model to be developed for the data set of interest whereby one or a number of dependant variables are utilised where the selected response variable is modelled from multiple predictor variables.

In most environmental data sets, the responses are commonly associated with water quality parameters, such as pH or dissolved oxygen, whilst the predictors are the sample locations.

However, in the case of this study, to successfully model the locality of origin of identified human source isolates, the response variables (Y) were modelled as the sampling locations, with the measured resistance to the different concentrations of antibiotics used as the predictor variables. To compare the correlation of human source isolates to specific localities, a simple PLS model was also developed to check the correlation of non-human isolates to those identified as non-human from upstream monitoring locations.

RESULTS

Discriminant Analysis (DA) of *E. coli* Antibiotic Resistance Patterns

From the 61 faecal samples collected from known sources, a total of 1003 *E. coli* isolates were enumerated, and their patterns of antibiotic resistance determined. Of these isolates, 175 were human isolates, which in turn were separated on the basis of 101 being directly human, 39 from OWTS and 35 from the sewage treatment plant.

DA for the pooled human versus non-human isolates performed exceptionally well with both human and non-human categories showing clear discrimination between isolates, as shown in Figure 2. To assess whether the source libraries retained sufficient isolates to correctly classify the unknown sources, pulled-sample cross-validations were conducted on the pooled human and non-human isolates. The overall ARCC for the libraries used to re-classify randomly pulled human samples was 88.5% as shown in Table 1. For re-classifying randomly pulled non-human source samples, the ARCC for the source libraries was 80.4%.

Figure 2

Table 1

The correct classification rates were similar to those derived in other studies which achieved ARCC of >80% for human versus non-human pooled categories (Wiggins et al 1999, Harwood et al 2000, Whitlock et al 2002, Booth et al 2003). Hence, the ARCC's confirmed that the library was sufficiently large enough to provide adequate discrimination between human and non-human sources. Pulled non-human source samples had slightly lower correct classification rates mostly due to the relationship between the domestic and human categories in respect of antibiotic usage.

Classification of Unknown Source Isolates

From the 144 water samples collected from the twelve monitored surface and ground water sampling locations, 199 unknown isolates were selected for ARP analysis. Applying DA to the unknown source isolates and utilising the human versus non-human source library, the percentage of human isolates contained in the collected water samples were obtained. Table 2 provides the percentages of human and non-human isolates from the respective sampling points. From the DA analysis of samples obtained from Ningi Creek, a majority of the unknown source isolates were classified as non-human, particularly isolates obtained from upstream sampling locations. However, the percentage of human source isolates increased in Ningi Creek after passing the urbanised areas (SW1-SW3) and the consequent increase in OWTS, as evident in Table 2.

Table 2

Partial Least Square (PLS) Analysis of ARP

After using DA to determine the percentage of different sources at each of the sampling locations, the resistance patterns obtained were analysed using PLS regression to model their correlation between sampling localities. Four PLS models were developed in order to ascertain the locality of classified human and non-human source isolates established through the initial DA analysis. The models developed included (1) a predictive model using source isolates from groundwater samples GW1-GW5 extracted from the urban development for predictors to model the response at sampling site SW1 (Human - X:GW1-5 predictors, Y: SW1 response); (2) a model using source isolates from groundwater samples GW1-GW5 for predictors to model the response at sampling site SW3 (Human X:GW1-5 predictors, Y: SW3 response); (3) a model using source isolates from groundwater samples GW1-GW5 for predictors to model the response at sampling site SW2 (Human X:GW1-5 predictors, Y: SW2 response); and (4) a predictive model using source isolates from surface water sampling sites SW4-SW5 upstream in Ningi Creek for predictors to model the response at sampling site SW2 (Non-Human X:SW4-7 predictors, Y: SW3 response). Model 4 was developed to predict the response of non-human faecal source isolates from upstream predictors at site SW2.

In developing the PLS regression models, an appropriate number of latent variables (LV) required to provide a more precise model were determined. In assessing the identified human source isolates, it was noted from the root mean square (RMS), cross validation (CV) and reduced eigenvalue (RE) tests that only 3 latent variables were necessary for models 1, 3 and 4. However, modelling non-human source isolates for their locality required a higher number of latent variables to provide a more precise model, and as such, 5 LV's were required for model 2. The main reason for this is the additional noise or variation obtained from non-human sources that consist of isolates obtained from numerous wild and domesticated animals, all of which have variations in their antibiotic resistance patterns.

The results of the PLS models including measures of model performance are provided in Table 3 and Figures 3a to 3d. Although all models showed reasonable correlation between antibiotic resistance patterns, there is a distinct separation between SW1 and SW2 and SW3. Figures 3a and 3b show the observed vs predicted plots for the models using ARP patterns for isolates obtained from groundwater samples. PLS model 1 has the statistics of LV = 3, R^2X = 0.8812, R^2Y = 0.9979, Q^2 = 0.9994, and root mean square error of prediction (RMSEP) = 0.6255. This indicates that the model performed reasonably well, although increasing the number of LV did not significantly improve the model performance. The regression line in the observed versus predicted plot (Figure 3a) also indicated good model performance with a high r^2 value and low bias and SEP. A similar result was obtained for model 2 which performed slightly better overall, as indicated by the model statistics of LV = 3, R^2X = 0.9804, R^2Y = 0.9997, Q^2 = 0.9999, and RMSEP= 0.5340. The observed versus predicted plot (Figure 3b) also showed that the model performed exceptionally well. Models 3 and 4 however did not perform as well, as indicated by the statistics of LV = 5, R^2X = 0.7344, R^2Y = 0.9944, Q^2 = 0.9997 and RMSEP = 1.4760 for model 3 and LV = 3, R^2X = 0.9737, R^2Y = 0.9838, Q^2 = 0.9991 and RMSEP = 1.2460 for model 4. The observed versus predicted plots, Figures 3c and 3d respectively, also showed poorer model performances with lower r^2 values, and more bias included in the PLS models. The standard errors of prediction (SEP) are also higher for these models.

Figure 3a-d

Although all four models retained high r^2 values (>0.90), models 3 and 4 retained more bias and a higher SEP than models 1 and 2. Models with a larger bias and higher SEP would retain predictors that are highly influential on the model, causing a lower predictive capability. This is a result of similar (but not identical or clonal) ARP profiles of isolates in the library from different sources (eg wild animals).Variations inherent in the ARP patterns that would be associated with differences in source isolates (ie non-human) create unwanted noise or variation in the data matrix. This is a common feature in environmental data which can vary significantly based on weather conditions and natural river flows. Tidal fluctuations also exert a major influence on the natural water flow in Ningi Creek, and consequently variability in the number and source of faecal isolates can exist as a result. From the faecal samples analysed, a direct relationship between the residential areas and Ningi creek was identified between sampling sites SW1 and SW3, with SW2 found to be influenced more by faecal sources located upstream as identified from sampling sites SW4-SW7. This relationship was assessed by undertaking an additional model development (Model 4) using identified non-human source isolates as the predictors. As indicated in Table 3, the model retained higher errors of prediction than that found in using the human isolates (Models 1 and 2) indicating that the data used in the model were not appropriate for predicting the locations of source isolates as that used in models 1 and 2. The results for model 4 do not indicate or ultimately prove that no influence from the nearby urban development exists, as some human source isolates were identified at this location (Table 2). However, a higher proportion of isolates were classified as non-human at this location, suggesting that the site is more influenced by other point sources upstream.

Table 3

DISCUSSION

The increasing use of OWTS in rapidly urbanising areas without centralised sewage treatment facilities can cause detrimental environmental and public health impacts. However, the ability to assess sewage contamination of water resources in areas of high densities of OWTS has been difficult, as no reliable means of identifying the various sources of faecal pollution has been available until recently. The main purpose of this study was two-fold; (i) to establish an ARP library and use DA to distinguish between human and non-human source isolates collected from surface water and groundwater sampling points in Ningi Creek catchment, and (ii) apply PLS regression to predict the point of locality of classified human source isolates. The use of ARP for identifying the various sources of faecal contamination within Ningi Creek catchment has shown good results, and its use for linking this contamination to OWTS in the study area has been successful. However, in order to ensure that the source library is maintained and continues to maintain adequate predictive ability on isolate patterns, new known source isolates need to be collected on an ongoing basis to keep the library up to date and to ensure suitable classification of water samples. Developed source libraries can have a limited lifespan in relation to their ongoing predictive capabilities. Factors such as the age of source isolates (bacteria will continually evolve to accommodate changes in the surrounding environment and changes in antibiotic resistance), changes to host organism food and locations, urban and agricultural practices, as well as environmental variables can all influence the accuracy and classification ability of the developed database. Several studies have investigated factors that influence source library development (Wiggins et al 2003, Harwood et al 2000, Graves et al 2007). Although the source library was sufficient for this study, future analysis will require further sampling and source library development in addition to the existing library to encapsulate the existing conditions and variations in the data at the time of sampling.

The results of the DA undertaken on the known source *E. coli* isolates indicated that applying ARP for the identification of human vs non-human sources of faecal contamination was successful. To correctly classify the sources of selected isolates, libraries created must contain sufficient isolates to ensure they are adequately representative to provide satisfactory discrimination between known source isolates (Wiggins 1996). It is generally recommended that a few hundred isolates for each identified source may be necessary for providing adequate discrimination between source isolates (Hagedorn et al 1999, Wiggins et al 2003). However, it was found that a smaller source library was sufficient for obtaining the desired outcomes, mostly due to the need to discriminate between human and non-human sources.

Classification of the unknown *E. coli* isolates collected from the monitored surface water and groundwater sampling sites provided two significant findings. Firstly, higher percentages of human *E.coli* source isolates were identified in areas surrounding the residential developments, namely SW1-SW3, relying on onsite systems for the treatment and dispersal of wastewater. Isolates obtained from groundwater samples also indicated high percentages of human *E.coli* isolates, notably a result of the dispersal of sewage effluent from OWTS and the shallow groundwater conditions. Higher percentages of non-human source isolates were identified in the less developed upstream segments of Ningi Creek, with increasing percentages of human source isolates as the creek meandered past the urbanised areas.

PLS regression modelling was undertaken to correlate human source ARP from sampling points SW1 to SW3 to antibiotic resistance patterns obtained from groundwater monitoring sites located within the urban development (GW1-GW5) and from sampling sites (SW4-SW7) located upstream in the catchment in order to locate their point of origin.

The results of the PLS modelling indicated that there is a distinct correlation between human source isolates derived from groundwater samples (GW1-GW5) and those obtained from surface sampling sites SW1 and SW3. Although SW2 indicated some correlation with the ARP from source isolates obtained from the groundwater samples in the urban development, a stronger correlation with source isolates collected from upstream sampling sites in Ningi Creek was evident. To further assess the relationship of SW2 with upstream source isolates, a predictive model using non-human source isolates as the predictors for SW2 was developed. This PLS model provided a higher correlation between SW2 and the upstream non-human source isolates to that compared to the groundwater samples. The main reason for this is attributed to the location of these sites on opposite sides of a mud flat that divides the flow in Ningi Creek. Previous research had demonstrated that surface water channels that can drain or flush easily, such as canals and short flow-through waterways retained relatively better microbial water quality (Griffin et al 1999). This is notably due to the flushing and removal of contaminated water by changing tides, followed by the inflow of better quality water unaffected by the surrounding localised contaminant sources. Additionally, studies have also shown that *E.coli* generally enters surface waters during wet conditions and that during high tide events the source location is more pronounced (Solo-Gabriele et al. 2000). Additional measurements of surface and groundwater sources during observed tidal conditions found that groundwater contributions are strongest during low tides. This increases the input of *E.coli* through groundwater-surface water interaction, where onsite wastewater treatment systems can be a source of *E. coli* contamination. Taking into consideration this phenomenon, SW1 and SW3 located in the longer section of Ningi Creek and bypassing the island would retain water for longer periods of time and be under a greater influence from local nearby contaminant sources such as the urban development.

SW1 and SW3 would receive isolates from the same upstream sources, but would also be greatly influenced by isolates originating from the residential development through groundwater infiltration into Ningi Creek. Model results for SW2 however indicate that this site may not directly be influenced by groundwater flows from the residential area, thereby having no distinct correlation with the ARP patterns from this locality. Further research on the interaction between groundwater and surface water flows and sampling site SW2 is necessary to clearly define the relationship between the urban development as a contaminant point source at this location.

The benefits from the use of multivariate statistical techniques, in particular DA and PLS regression to identify the source and point of origin of human faecal contamination within the investigated catchment has been significant. The analysis undertaken on collected ARP profiles have successfully identified the main sources of faecal contamination and was also able to predict their point of locality. This was advantageous in that with the isolate sources being identified and the ability to predict the locality of these sources with the PLS models, more appropriate mitigation management strategies can be implemented to protect water resources from microbiological contamination.

Conclusions

1. The study results confirmed that the use of Antibiotic Resistance Pattern (ARP) analysis together with Discriminant Analysis (ARP) is a robust method for determining the different sources of faecal pollution in a catchment with widely different land uses. However, identification of the specific locality of the pollution source requires additional assessment.
2. The Discriminant Anaylsis undertaken confirmed that the nature of possible faecal pollution in a waterway can be directly correlated to the surrounding land use. In the study, the majority of the *E. coli* isolates collected were from non-human sources in the upstream segments of the catchment where the predominant land use is natural bushland and agriculture. Increasing human source isolates were identified in the downstream

urbanised areas where onsite systems are used for the treatment and dispersal of wastewater.

3. The Partial Least Squares (PLS) Regression modelling undertaken confirmed that the pollution of a surface water resource will have a direct impact on the underlying groundwater. In the study, the identified source isolates indicated a high correlation between human source isolates at the surface water sampling sites, SW1 and SW3 with source isolates collected from the groundwater sampling sites, GW1-GW5 which were located in the urban area.

ACKNOWLEDGEMENTS

The authors would like to thank Caboolture Shire Council and Queensland University of Technology for funding this research project.

REFERENCES

APHA (1999) Standard Methods for the Examination of Water and Wastewater - 20th Edition. Washington, DC: American Public Health Association, American Water Works Association and Water Environment Federation.

Booth, A.M., Hagedorn, C., Graves, A.K., Hagedorn, S.C., and Mentz, K.H. (2003) Sources of Fecal Pollution in Virginia's Blackwater River. *J Environ Eng* **129:** 547-552.

Brereton, R.G. (1990) Chemometrics: Applications of mathematics and statistics to laboratory systems. Ellis Horwood, New York.

Devriese, L.A., Pot, B., and Collins, M.D. (1993) Phenotypic identification of the genus Enterococcus and differentiation of phylogenetically distinct eneterococcal species and species groups. *J Appl Bacteriol* **75:** 399-408.

Dombeck, P.E., Jonhnson, L.K., Zimmerley, S.T., and Sadowsky, M.J. (2000) Use of repetitive DNA sequences and the PCR to differentiate Escherichia coli from human and animal sources. *Appl Environ Microbiol* **66:** 2572-2577.

Graves, A.K., Hagedorn, C., Brooks, A., Hagedorn, R.L. and Martin, E. (2007) Microbial source tracking in a rural watershed dominated by cattle. *Water Research* **41**:3729–3739

Griffin, D.W., Gibson III, C.W., Lipp, E.K., Riley, K., Paul III, J.H. and Rose, J.B. (1999) Detection of Viral Pathogens by Reverse Transcriptase PCR and of Microbial Indicators by Standard Methods in the Canals of the Florida Keys. *Appl Environ Microbiol* **65 (9):** 4118-4125.

Hagedorn, S.C., Robinson, S.L., Filtz, J.R., Grubbs, S.M., Angier, T.A., and Reneau Jr, R.B. 1999. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. *Applied and Environmental Microbiology* **65(12):** 5522-5531.

Harwood, V.J., Whitlock, J., and Withington, V. 2000. Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: Use in predicting the source of fecal contamination in subtropical waters. *Applied and Environmental Microbiology* **66(9):** 3698-3704.

Harris, P.J. (1995) Water quality impacts from onsite waste disposal systems to coastal areas through groundwater discharge. *Environ Geol* **26:** 262-268.

Howell, J.M., Coyne, M.S., and Cornelius, P.L. (1996) Effect of sediment particle size and temperature on fecal bacteria mortality rates and the fecal coliform/fecal streptococci ratio. *J Environ Qual* **25:** 1216-1220.

Hartel, P., Gates, K., Payne, K., McDonald, J., Rodgers, K., Hemmings, S., Fisher, J. and Gentit. L. (2005) Targeted sampling to determine sources of fecal contamination. S*tormwater* **6**:46–53.

Kuntz, R.L., Hartel, P.G., Godfrey, D.G., McDonald, J.L., Gates, K.W. and Segars. W.I. (2003) Targeted sampling protocol with *Enterococcus faecalis* for bacterial source tracking. *J. Environ. Qual.* **32**:2311–2318.

Leeming, R., Bate, N., Hewlett, R., and Nichols, P.D. (1998) Discriminating fecal pollution: a case study of stormwater entering Port Phillip Bay, Australia. *Water Sci Technol* **38:** 15-22.

Lipp, E.K., Farrah, S.A., and Rose, J.B. (2001) Assessment and impact of microbial fecal pollution and human enteric pathogens in a coastal community. *Mar Pollut Bull* **42:** 286-293.

Meays, C.L., Broersma, K., Nordin, R., and Mazumder, A. 2004. Source tracking fecal bacteria in water: A critical review of current methods. *Journal of Environmental Management* **73(1):** 71-79.

McDonald, J.L., Hartel, P.G., Gentit, L.C., Belcher, C.N., Gates, K.W., Rodgers, K., Fisher, J.A., Smith, K.A. and Payne, K.A. (2006) Identifying Sources of Fecal Contamination Inexpensively with Targeted Sampling and Bacterial Source Tracking. *J. Environ. Qual.* **35**:889–897

Pang, L., Close, M., Goltz, M., Sinton, L., Davies, H., Hall, C., and Stanton, G. (2003) Estimation of setback distances based on transport of E. coli and F-RNA phages. *Environ Int* **29:** 907-921.

Parveen, S., Portier, K.M., Robinson, K., Edminston, L., and Tamplin, M.L. (1999) Discriminant analysis of ribotype profiles of Escherichia coli for differentiating human and nonhuman sources of fecal pollution. *Appl Environ Microbiol* **65:** 3142-3117.

Paul, J.H., McLaughlin, M.R., Griffin, D.W., Lipp, E.K., Stokes, R., and Rose, J.B. (2000) Rapid movement of wastewater from onsite disposal systems into surface waters in the lower Florida Keys. *Estuaries* **23:** 662-668.

Paul, J.H., Rose, J.B., Jiang, S.C., Zhou, X., Cochran, P., Kellogg, C. et al. (1997) Evidence for groundwater and surface marine water contamination by waste disposal wells in the Florida Keys. *Water Res* **31:** 1448-1454.

Pourcher, A.M., Devriese, L.A., Hernandez, J.F., and Delattre, J.M. (1991) Enumeration by a miniaturized method of Escherichia coil, Streptococcus bovis and Enterococci as indicators of the origin of fecal pollution of waters. *J Appl Bacteriol* **70:** 525-530

Roberts, A. and Withers, P. (2004). StatistiXL Version 1.4. Kalamunda, Western Australia. www.statistiXL.com.

Solo-Gabriele, H.M., Wolfert, M.A., Desmarais, T.R. And Palmer, C.J. (2000) Sources of *Escherichia coli* in a Coastal Subtropical Environment. *Appl Environ Microbiol* **66(1):** 230-237.

Whitlock, J.E., Jones, D.T., and Harwood, V.J. (2002) Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis. *Water Res* **36:** 4273-4282.

Wiggins, B.A. (1996) Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters. *Appl Environ Microbiol* **62:** 3997-4002.

Wiggins, B.A., Andrews, R.W., Conway, R.A., Corr, C.L., Dobratz, E.J., Dougherty, D.P. et al. (1999) Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. *Appl Environ Microbiol* **65:** 3483-3486.

Wiggins, B.A., Cash, P.W., Creamer, W.S., Dart, S.E., Garcia, P.P., Gerecke, T.M. et al. (2003) Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries. *Appl Environ Microbiol* **69:** 3399-3405.

Wilson, D.I. (2002) Derivation of the chalk superficial deposits of the North Downs, England: an application of discriminant analysis. *Geomorphology* **42:** 343-364.

Wold, S., Trygg, J., Berglund, A., and Antti, H. (2001) Some recent developments in PLS modelling. *Chemometrics and Intellengent Laboratory Systems* **58**: 131-150

Young, K.D., and Thackston, E.L. (1999) Housing density and bacterial loading in urban streams. *J Environ. Eng* **125:** 1177-1180.

List of Figures

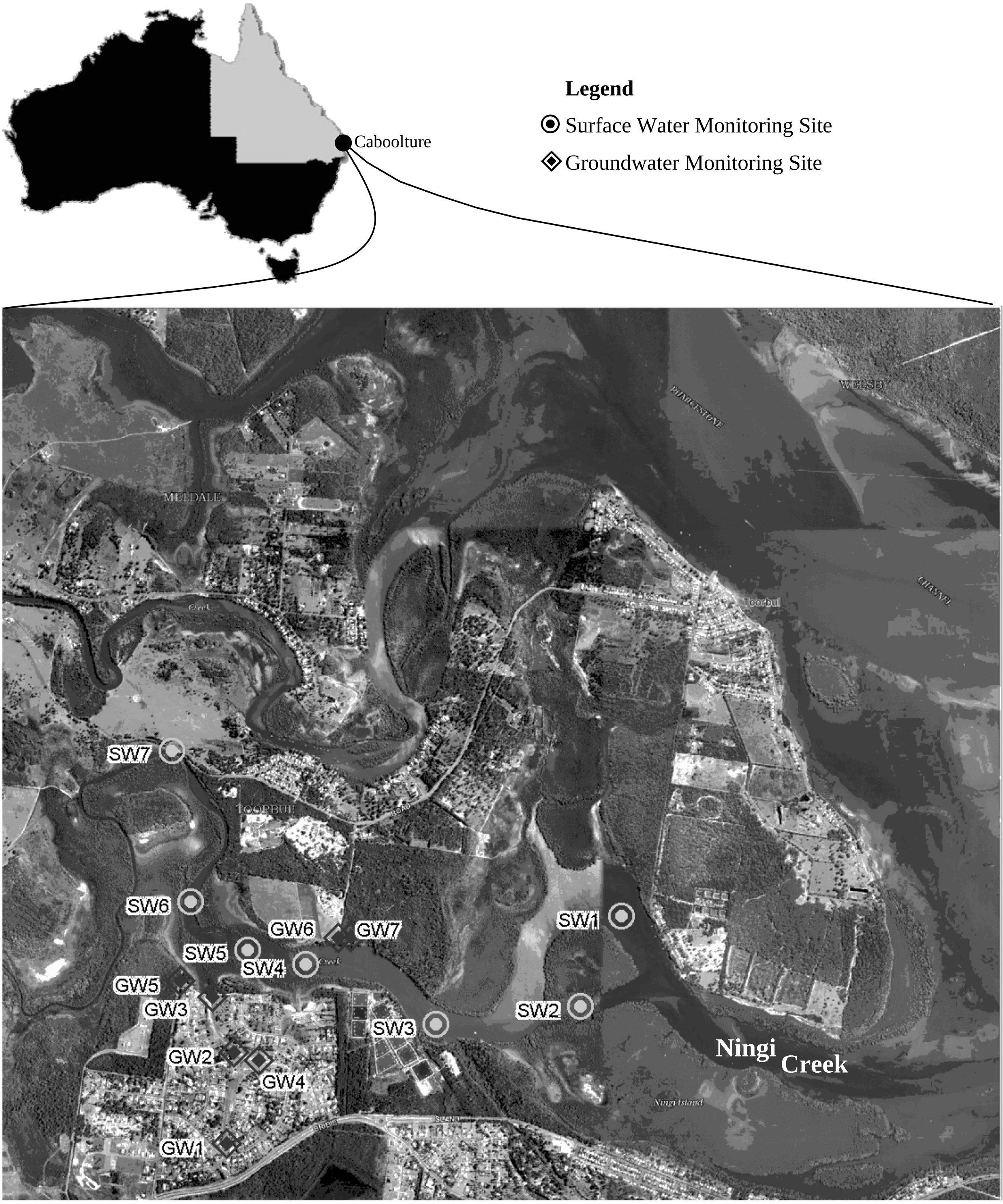

Figure 1: Ningi Creek catchment and established monitoring sites

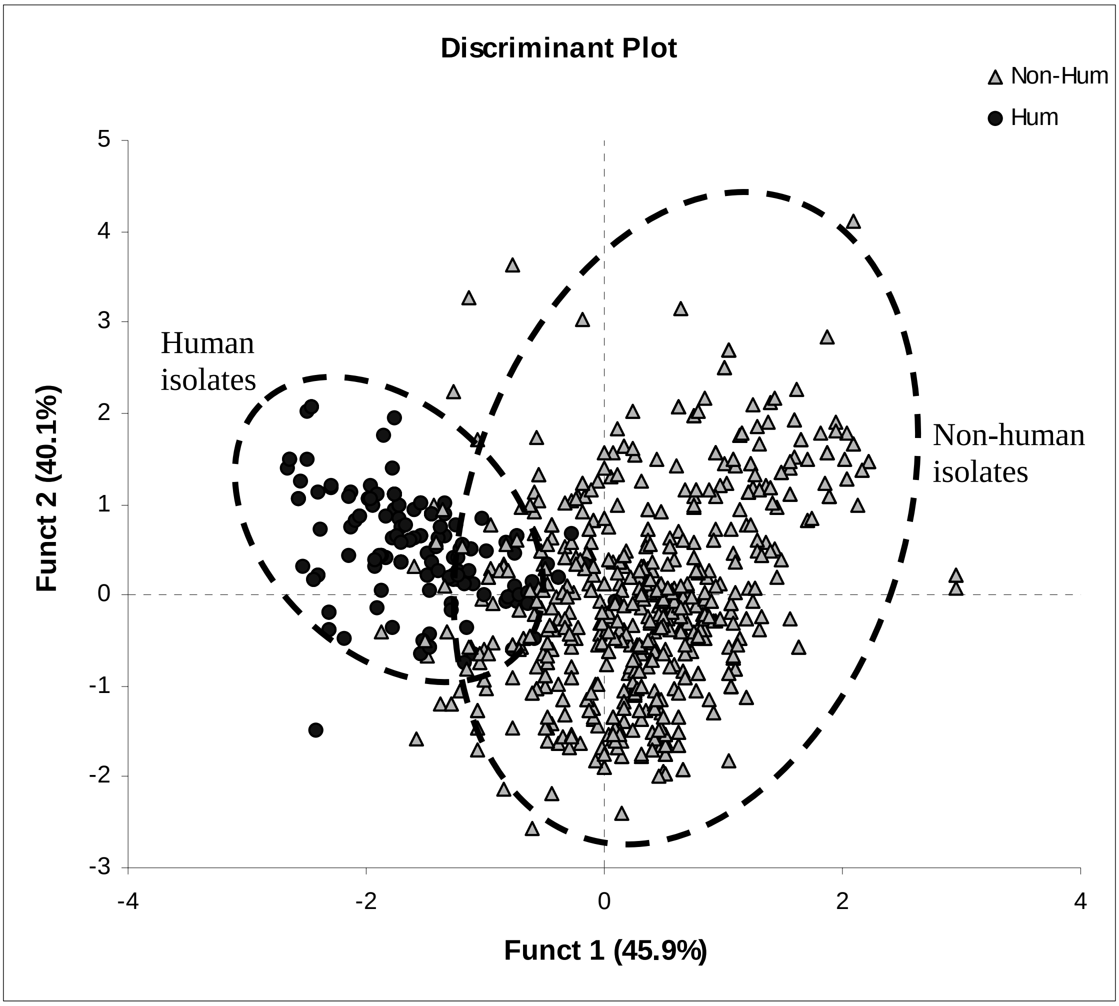

Figure 2: Discriminant analysis plot of source library isolates for pooled human versus non-human categories

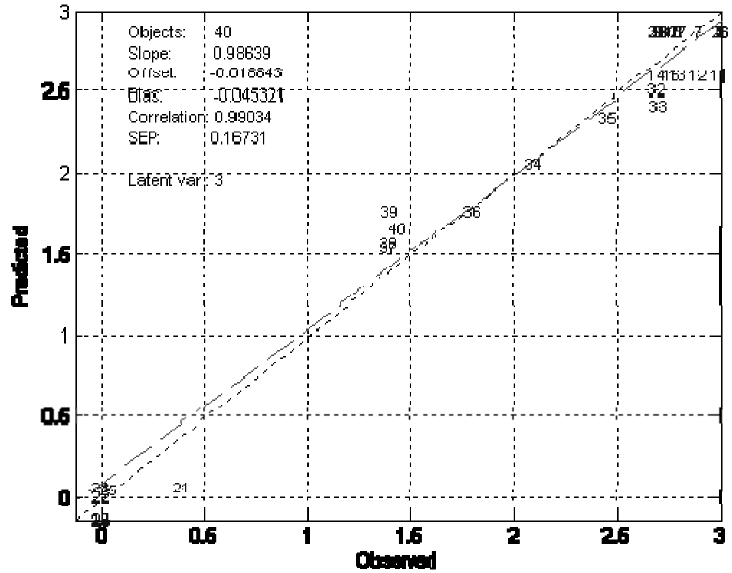

Figure 3a – PLS Model 1

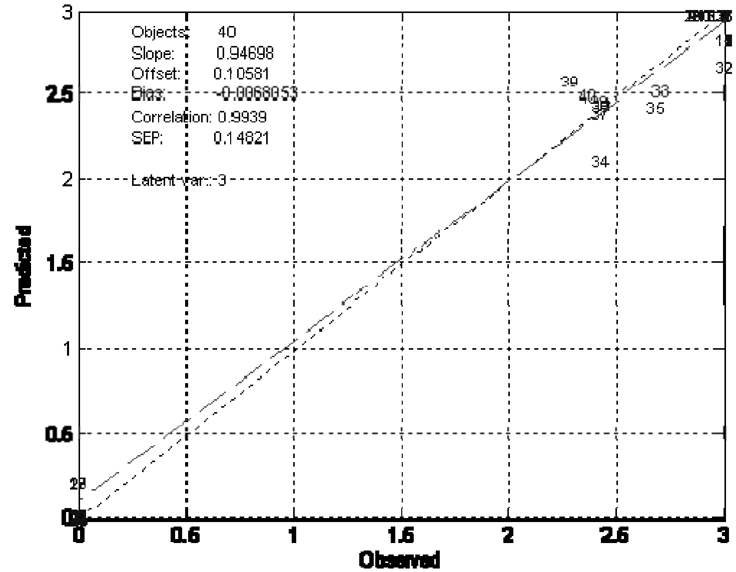

Figure 3b – PLS Model 2

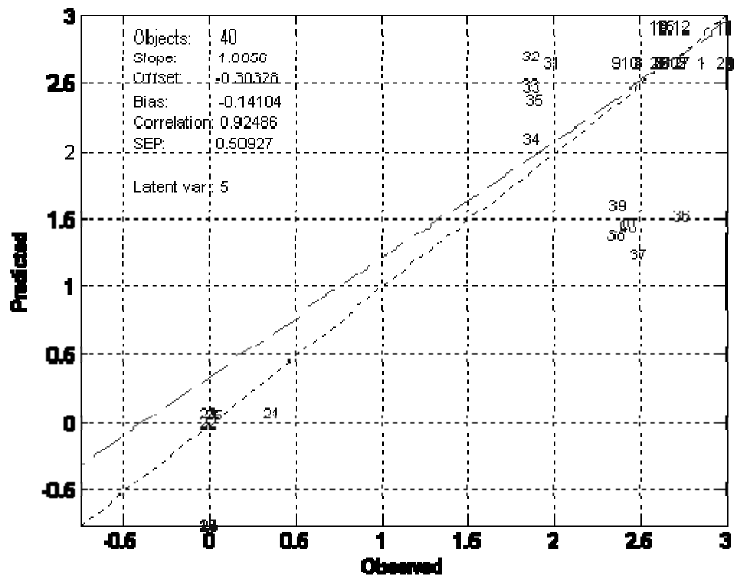

Figure 3c – PLS Model 3

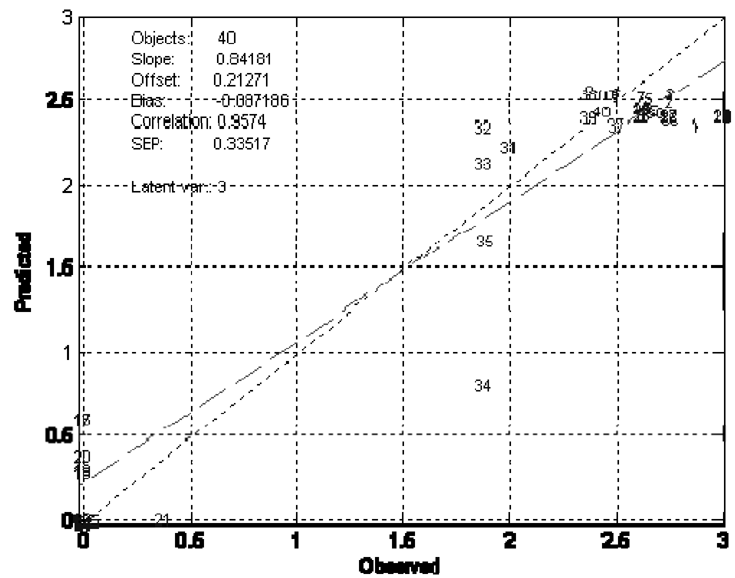

Figure 3d – PLS Model 4

Figure 3: Observed versus predicted plots of faecal source isolates for PLS

Table 1: Classification rates and ARCC for cross validation analysis for human and non-human source isolates

Source	Classification rates for pulled human isolates			Classification rates for pulled non-human isolates		
	Non-Human	Human	Correctly Classified	Non-Human	Human	Correctly Classified
Non-Human (*n = 828*)	706	122	85.3%	638	190	77.1%
Human (*n = 175*)	15	160	91.7%	29	146	83.7%
Average Rates Correct Classification		(ARCC)	88.5%		(ARCC)	80.4%

Table 2: Source identification of unknown isolates from monitored sites

Monitoring Site	*E. coli*[a] cfu/100mL			No. Isolates[b]	Source Identification (%) of unknown source isolates	
	Mean	Min	Max		Human[c]	Non-Human[c]
Ningi Creek				*(n = 129)*		
SW1	1,751	23	3,600	*17*	87	13
SW2	315	10	660	16	21	79
SW3	1,165	250	6,305	21	68	32
SW4	474	47	1,121	22	5	95
SW5	540	231	1,152	16	6	94
SW6	357	29	1,063	20	14	86
SW7	660	40	1,216	18	7	93
Urban Development				*(n = 70)*		
GW1	424	1	2,512	16	60	40
GW2	328	1	2,450	14	100	0
GW3	49	10	270	15	100	0
GW4	1,274	1	2,459	12	67	33
GW5	318	3	1,210	13	95	5

[a] General statistical of E .coli concentrations in collected samples

[b] Unknown isolates collected from monitored sites over twelve months sampling period selected for ARA

[c] Pooled source categories for human vs non-human isolate DA

Table 3: PLS regression model performance characteristics

PLS Model	LV	R2X	R2Y	Q2	RMSEP
1. Human X:GW1-5 predictors, Y: SW1 response	3	0.8812	0.9979	0.9994	0.6255
2. Human X:GW1-5 predictors, Y: SW3 response	3	0.9804	0.9997	0.9999	0.5340
3. Human X:GW1-5 predictors, Y: SW2 response	5	0.7344	0.9944	0.9997	1.4760
4. Non-Human X:SW4-7 predictors, Y: SW2 response	3	0.9737	0.9838	0.9991	1.2460