DATA VISUALISATION FOR IMPROVED UNDERSTANDING OF LARGE WATER QUALITY DATASETS

Lisa Mace (Beca Ltd), Andrew Watson (Beca Ltd) and Bridget O'Brien (Christchurch City Council)

ABSTRACT

Christchurch City Council (CCC) and Environment Canterbury had a large amount of groundwater data but found that it was difficult to track changes in groundwater quality in the underground aquifers when data was stored and analysed in spreadsheets. Standard graphing techniques could be used to identify trends in a single bore, however, each bore was not monitored consistently over the entire data set. In many cases, one bore was monitored for a number of years and then monitoring swapped to another nearby bore. This meant that the trend in water quality for the area could not be easily identified.

CCC commissioned Beca to provide a data visualisation tool. The aim of this project was to compile and present available groundwater quality data to aid in early warning of any contamination risks to the drinking-water supply that may gradually develop over time. Microsoft Power BI was used to present two data bases (CCC and Environment Canterbury data) with a total of 12,474 data points dating from 1954 to 2017. Six parameters that are key to drinking water quality were selected for the visualisation tool: E. coli, total coliforms, nitrate, ammonia, chloride and dissolved reactive phosphorous.

A tool was developed that presents the data geospatially to clearly illustrate what is happening with the water quality in and around Christchurch over time. Maximum acceptable values and guideline values from the Drinking Water Standards for New Zealand were used to define the warning colour system so that significant data points could be identified. It is also possible to interrogate individual data points through the Microsoft Power BI platform. This project resulted in a tool that is valuable in informing an understanding of long terms trends in water quality, alerting CCC to groundwater quality changes that could have impacts on the water supply, informing future investment considerations including location of new wells, and whether treatment or other mitigation measures may be required in the future.

KEYWORDS

Data analysis, visualisation, groundwater, drinking water

PRESENTER PROFILE

Lisa Mace is a Senior Process Engineer with six years' experience in the Water Industry mostly in drinking-water treatment. She is particularly interested in ways that water suppliers can improve their understanding of source water quality and therefore reduce risks.

1 INTRODUCTION

Modern analytical tools give us the ability to view data in more valuable ways. This paper examines one such example where Microsoft Power BI was used to visually display water quality data and therefore give the user greater insight into what the results may indicate. This paper also looks at other applications for similar tools.

The quality of the drinking water that we consume can affect significant portions of our population. Deteriorating water quality is an issue throughout New Zealand and particularly in Canterbury. In recent years, there has been concern around the increase in contaminants, such as nitrates, due to intensified land use in both urban and rural contexts. The ability to predict groundwater quality issues before a drinking-water supply is affected would be a valuable tool that would allow drinking-water suppliers to proactively manage potential issues before the community is affected.

Christchurch City Council (CCC) sources its drinking water from 139 wells across the city and pumps this water directly into the reticulation via 53 pump stations. The wells pump water from five aquifers beneath the city. The water in these aquifers is usually decades old and has its original sources in the Waimakariri River and rainfall on the Canterbury plains to the northwest of the city. Generally, the water is not treated and so understanding the trends in groundwater quality is vital.

Both CCC and Environment Canterbury (ECan) had a large amount of groundwater data but found that it was difficult to track changes in groundwater quality in the underground aquifers when data was stored and analysed in spreadsheets. Standard graphing techniques could be used to identify trends in a single bore, however, each bore was not monitored consistently over the entire data set. In many cases, one bore was monitored for a number of years and then monitoring swapped to another nearby bore. This meant that the trend in water quality for the area could not be easily identified.

This created an opportunity to build a tool that would allow CCC to monitor and view long term trends in water quality and predict likely risks. The aim of this is to inform future investment considerations including location of new wells, and the need for treatment processes or other mitigation measures. In 2016 CCC commissioned Beca to develop a data visualisation tool.

2 BUILDING THE TOOL

2.1 DATA CONSOLIDATION

A large amount of groundwater quality data is collected by CCC and ECan across Christchurch City and the Canterbury Plains. CCC store water quality data using the Water Outlook software, which was relatively easy to access and compile.

The ECan water quality data was significantly more difficult to compile. The data was stored in several different spreadsheets that were "owned" by different individuals and projects across ECan. Once the data was acquired, it was stored in various formats and needed to be "cleaned" and reformatted. The ECan groundwater monitoring data is from monitoring wells that are not used to supply drinking water.

These data sets were compiled into a single database of over 12,400 groundwater samples. The data set included bores on the western side of Christchurch City across the Canterbury Plains to the foothills (Figure 1). The bore water data covers a long time period, from 1954 through to 2017.

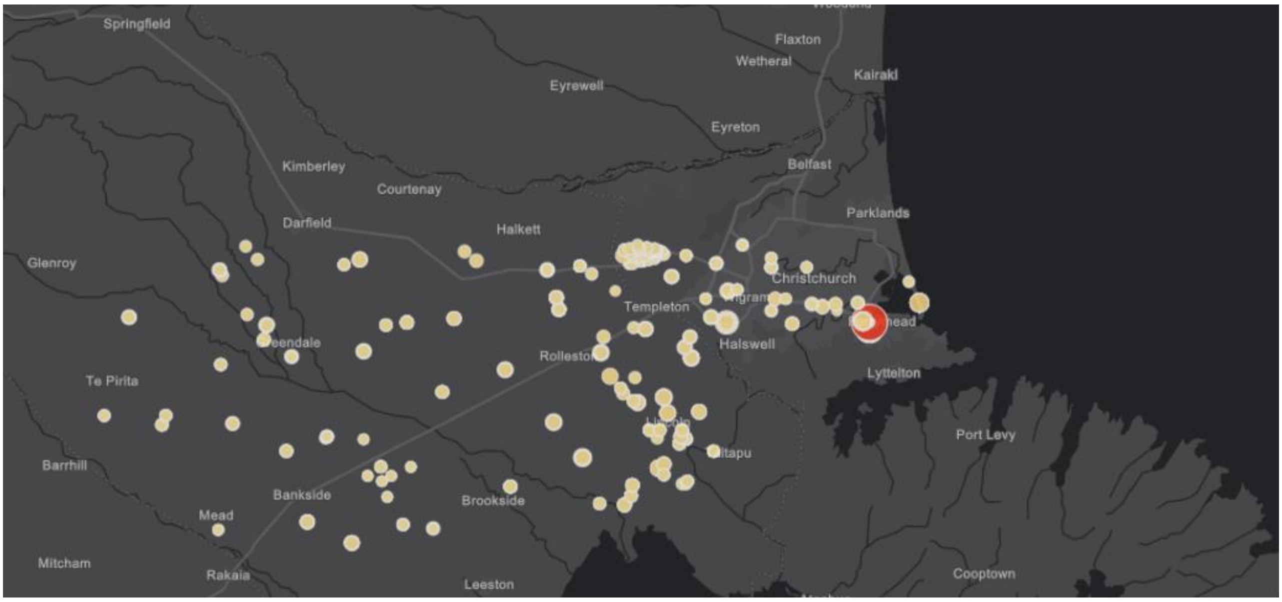

Figure 1: Geographical spread of bores included in the bore water data set

2.2 KEY WATER QUALITY PARAMETERS

The bore water data collected covered nearly 80 different parameters. Some parameters have been monitored for a short period as part of a specific study while other parameters have been monitored continuously.

An assessment was undertaken of some key parameters relevant to drinking water quality, based on the Drinking Water Standards for New Zealand 2005 (amended 2008) (DWSNZ). As part of the initial development of the tool, six key parameters were selected as indicators of water quality and used as a basis for assessing the water quality data. In addition, dissolved reactive phosphorus was included as an indicator of the effects of farming in the source catchment. These parameters are:

- **Escherichia coli (E. coli)** - an indicator of microbial contamination from faecal material of the water which poses a direct risk to public health.
- **Total coliforms** - an indicator of microbial contamination of the water which poses a direct risk to public health. Trends in total coliforms may indicate an increased likelihood of pathogens.
- **Nitrate nitrogen** - an indicator of contamination from human activities such as farming or effluent discharge, which pose a direct risk to public health.
- **Ammonia nitrogen** - an indicator of contamination from human activities such as farming or effluent discharge.
- **Chloride** - an indicator of contamination from human activities such as farming, or possible sea water ingress to the aquifer for coastal bores. Does not pose a direct risk to public health but can exacerbate corrosion.
- **Dissolved reactive phosphorous (DRP)** - an indicator of possible contamination from human activities such as farming or effluent discharge. Does not pose a direct risk to public health.

2.3 WATER QUALITY PARAMETER WARNINGS

The main value in the tool was identified to be the ability to track trends in water quality parameters over time. However, it is also of value to display warnings so that the user also understands the trigger points for concern.

The DWSNZ has Maximum Acceptable Values (MAV) and the Guideline Values (GV) for key water quality parameters.

- The MAV is defined as "the concentration of a determinand, below which the presence of the determinand does not result in any significant risk to a consumer over a lifetime of consumption"
- The GV is defined as "the value for aesthetic determinand that, if exceeded, may render the water unattractive to consumers"

In addition to the MAV and GV parameters from the DWSNZ, early warning values for each parameter were identified. These early warning values provide an alert that further action may be needed, such as increased monitoring. For the initial development of the tool, the early warning values were arbitrarily defined as 50% of the MAV or GV. This can be adjusted.

Table 1 summarises the parameter early warning values selected.

Table 1: Parameter early warning values

Parameter	MAV or GV	Early warning
E. coli (cfu / 100 mL)	MAV <1	-
Total Coliforms (cfu / 100 mL)[2]	-	-
Nitrate Nitrogen (mg/L)	MAV = 50	25
Ammonia Nitrogen (mg/L)	GV = 1.5	0.75
Chloride (mg/L)	GV = 250	125
Dissolved Reactive Phosphorus (mg/L P)[1]	-	-

1. There are no guideline values for DRP. An increasing DRP level over time would indicate contamination from human activities such as farming or effluent discharge.

2. There are no guideline values for total coliforms. Similar to E. coli, a presence of total coliforms would indicate microbial contamination however it can provide an earlier warning than E. coli.

2.4 SOFTWARE

Microsoft Power BI was selected as the best software for building the tool. This is because it is a relatively low cost tool that integrates in well with other Microsoft products. It has good functionality for compiling multiple sets of data in different forms as was the case for this project. This can include pulling in 'live' data. It is also possible to publish the tool onto a webpage to allow the public to view the output.

3 RESULTS

Figure 6 to 7 show snips of the data for each parameter. The data is more clearly shown using the visualisation tool as the 'play' button can be used to display the data over time. The bar graphs at the bottom of each figure show the average concentrations for the year. They are shown in an expanded view in Appendix A.

Note that:

- Each dot on each map represents a sample taken. The larger and darker the dot, the higher the concentration. Multiple dots in one location indicate multiple samples were taken.
- The slider beneath each graph shows the time frame for data being visualised.

- The graph beneath the slider shows the early warning value (labelled as Priority 2), GV or MAV, and how the annual average concentration has changed over time.

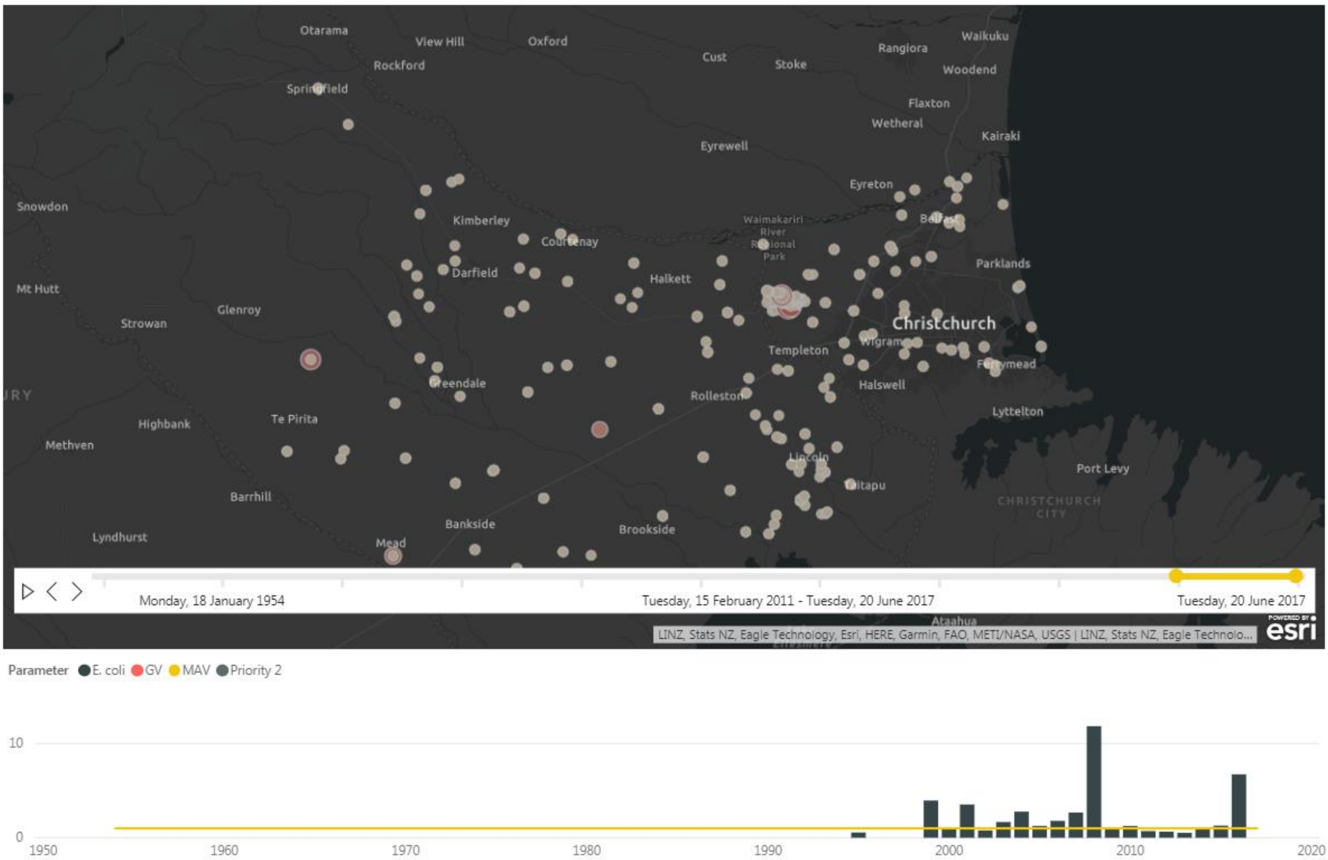

Figure 2: Variations in E. coli concentration

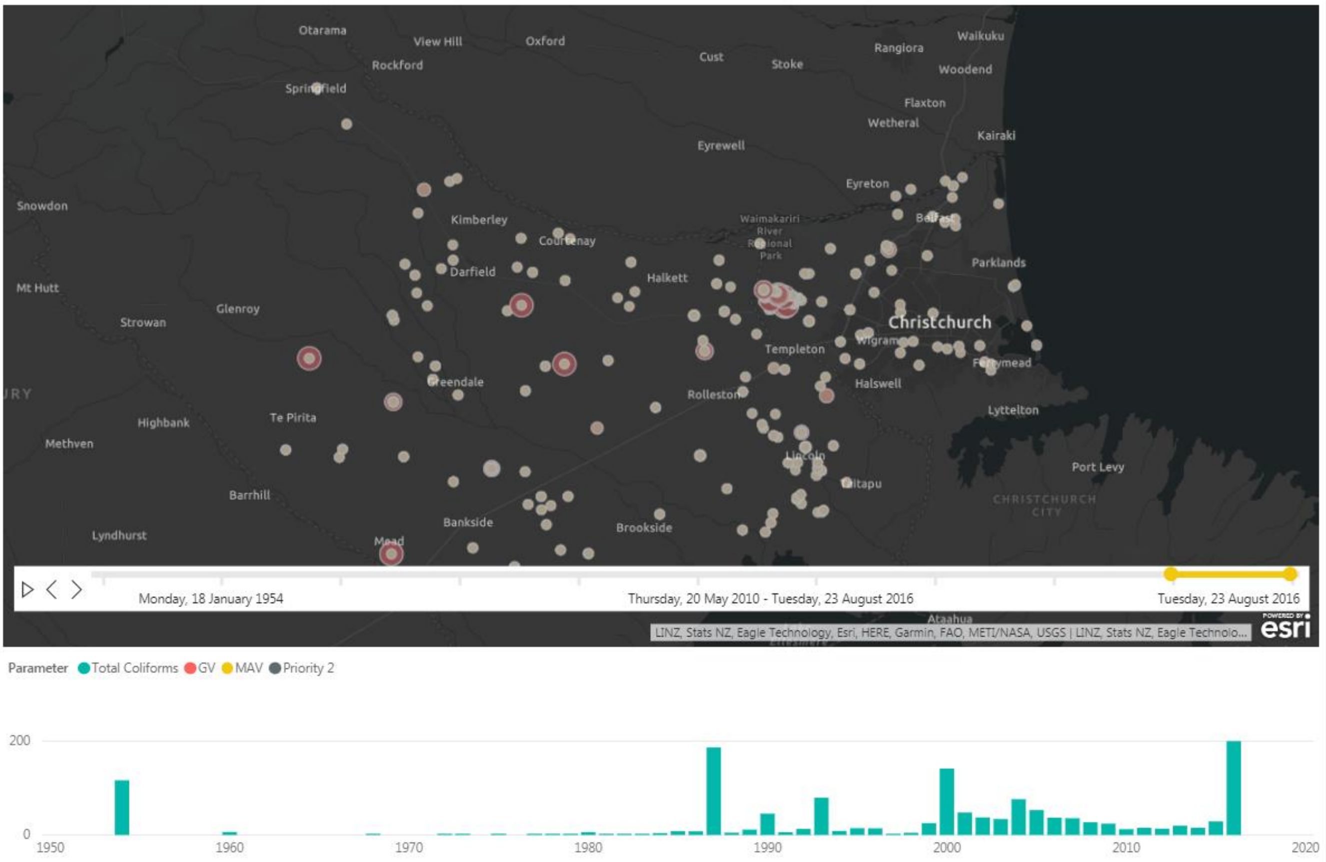

Figure 3: Variations in total coliforms concentration

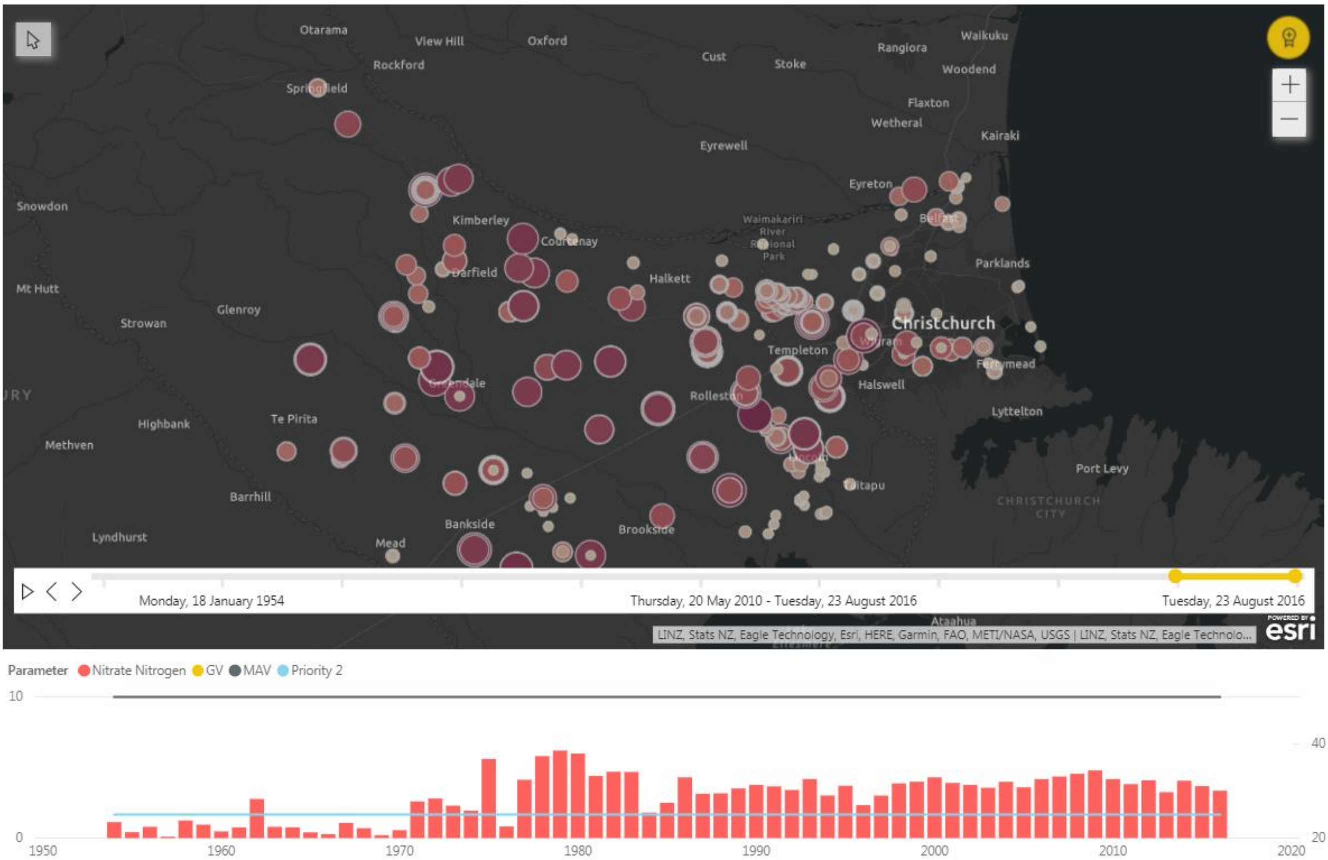

Figure 4: Variations in nitrate nitrogen concentration

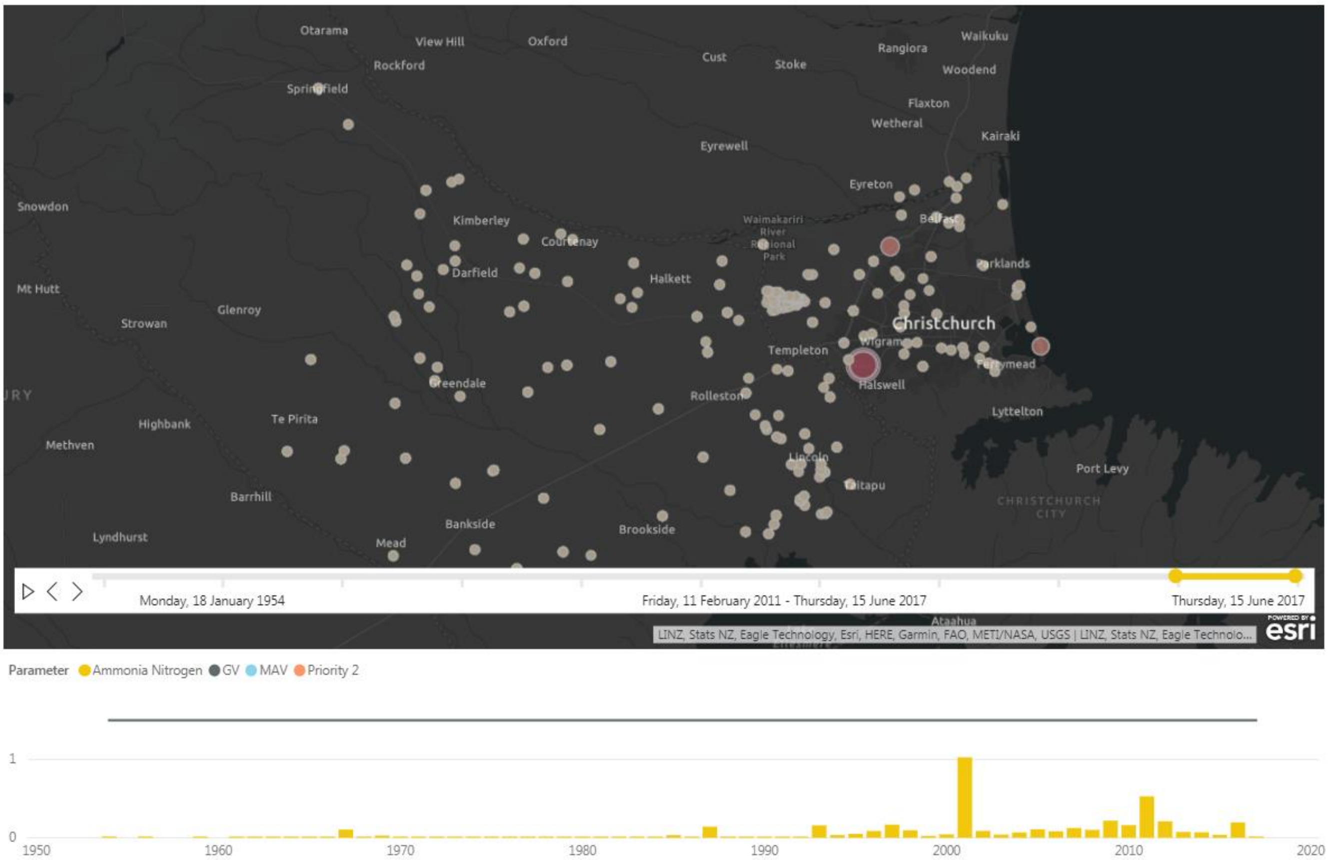

Figure 5: Variations in ammonia nitrogen concentration

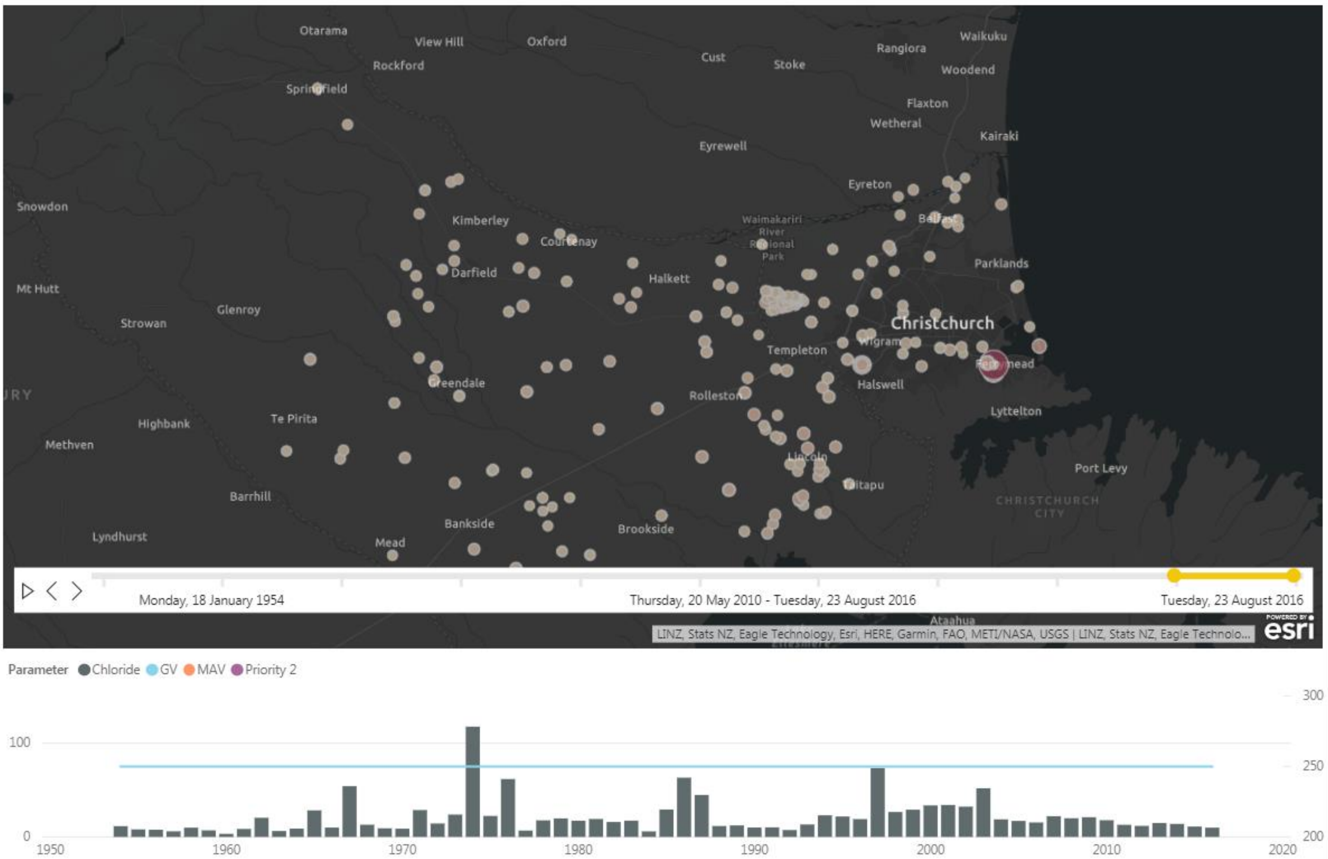

Figure 6: Variations in chloride concentration

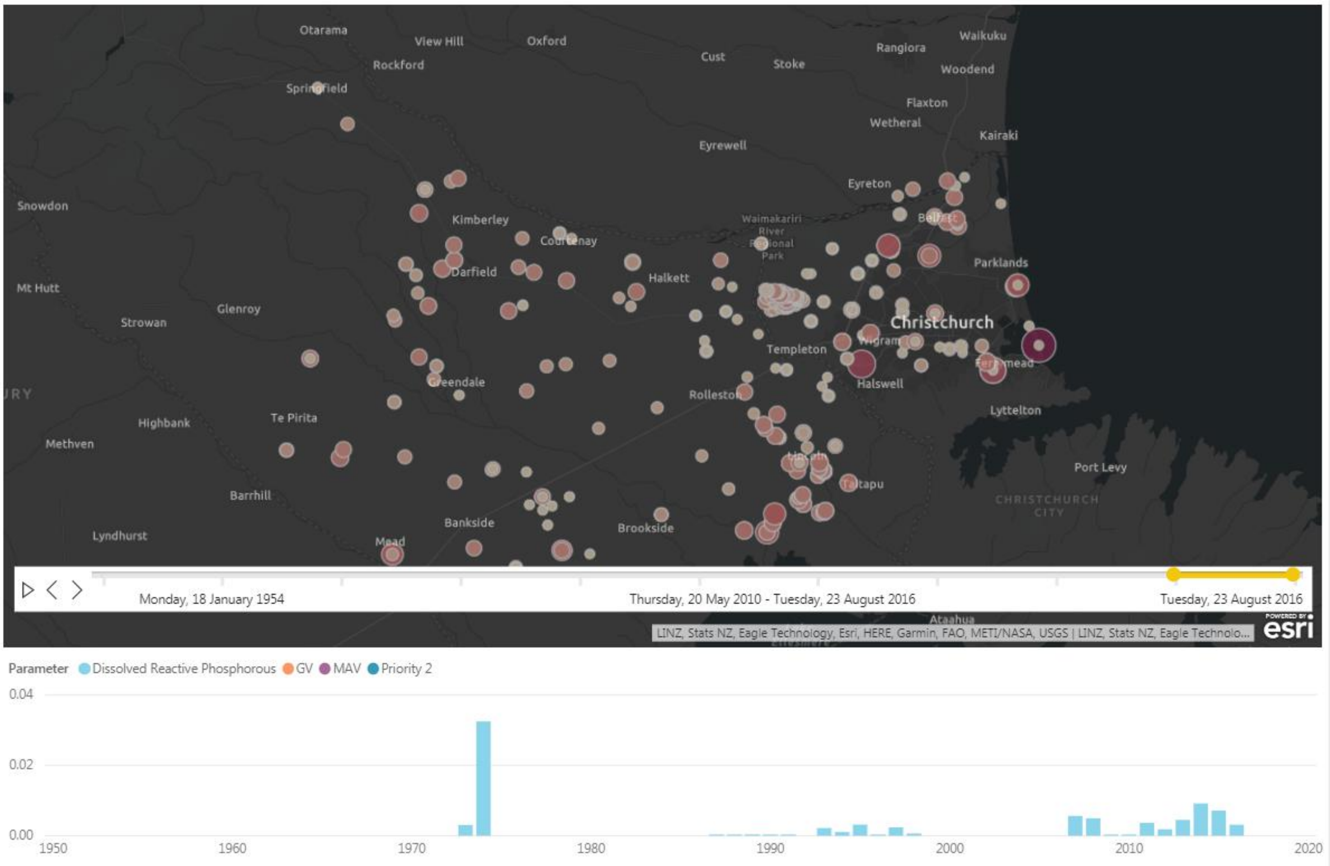

Figure 7: Variations in dissolved reactive phosphorus concentration

This tool has shown a few interesting trends in the data:

- **E. coli** - E. coli concentrations start to exceed the MAV from 2008 onwards. Exceedances are generally seen to the south-west of the city. Note that these exceedances are only seen in shallow ECan monitoring bores, not in CCC bores used for drinking water supply.
- **Total coliforms** - from the late 1970s onwards there was an increase in sampling. There are pockets of increased total coliforms levels from the late 1980s onwards. The location of the bores with increased concentration varies but they are generally south-west of the city. Total coliforms do not have a MAV or GAV and are mostly used to understand trends.
- **Nitrate nitrogen** - from the late 1970s onwards there was an increase in the amount of sampling and higher concentrations were seen south west of the city. Over the last 25 years, the maximum recorded values have been well below half of the MAV (50 mg/L).
- **Ammonia nitrogen** - the number of bores sampled has also increased. From the late 1980s onwards there has been an increase in concentrations in the south-west side of the city (Halswell/Wigram area) with smaller increases seen on the north and the east.
- **Chloride** - monitoring has increased significantly and in recent years there has been an increase in concentrations in the Ferrymead area which likely indicates saline intrusion. This well is no longer in use for water supply. Dissolved reactive phosphorus does not have a MAV or GV but is an indicator of saline intrusion.
- **Dissolved reactive phosphorus** - sampling locations and frequencies have varied, and the areas of high concentration seem to fluctuate. However, from 2004

onwards, that data shows increased concentrations throughout the region. Dissolved reactive phosphorus does not have a MAV or GV but is an indicator of increased farming activities.

4 EXTENSIONS OF THE TOOL

The main value in this data visualisation tool will be in monitoring water quality trends over time. This requires the continuous addition of data, which has been made easier as there is now a way to pull live data from Water Outlook directly into the tool. The tool can also inform adjustments to the water quality monitoring regime. It may be worthwhile extending the tool to show aquifer depths, bore types (e.g. water supply vs monitoring), land use changes over time, the relationship between parameters and to use an algorithm to highlight changes in concentration and therefore potential areas of concern.

It also would be of value to add groundwater data from neighbouring Waimakariri and Selwyn District Councils.

It may also be valuable to allow the public to access this data, or at least some portion of it.

5 OTHER APPLICATIONS

Data visualisation has a wide range of application throughout the country. Similar to Christchurch, there are many locations that would benefit from a geospatial tool to summarise their data. A number of applications applicable to the water industry include:

- **Bore level tracking** (as done by Kāpiti Coast District Council) allows residents to track water levels in surrounding monitoring bores to understand when groundwater levels are low. For Kāpiti Coast District Council this data is updated hourly and so residents can see almost real time information.

- **Spatial viewing of photos** (either standard photos of 360° photos) can be improved by showing the user where the photos were taken. This may be useful if a series of photos were collected along a pipe length, at a pump station or within a treatment plant. A tool could easily be set up to show these photos spatially. It could also show progression of photos over time which may be useful for asset or construction monitoring. This can also be a low cost way of building a model that allows people to do site walkovers remotely.

- **Inspection scheduling** can be carried out using data visualisation tools to optimise inspection visits. This is particularly useful when there are large teams inspect assets throughout a region such as reservoirs, bores, chemical handling, etc. The tool can display inspection locations and status (inspected, incomplete). The visual aspect of this allows user to more effectively schedule a series of visits based on their proximity to each other.

- **Predictive maintenance** including using available asset data to predict future failures. Pipe material, age, condition and past failure data has been used to predict where future failures are most likely. This is a valuable tool for asset planning.

- **Electricity usage monitoring** can be carried out using a map to show live usage. This can be a valuable tool for allowing asset owners to optimise treatment plant or pump station use.

- **Water network integrity mapping** – it could be used as a diagnostic tool to help determine where a network breach may have occurred by plotting E. coli and total coliforms in space and time. A modern technology extension to the way mapping deaths from cholera was what led to the understanding that disease can spread through a water supply in 1854 in London, thereby laying the foundations for public health and epidemiology.

- **Community information/education** could be improved with publicly available data visualisation tools. These can be published online and available for view without software. These tools are becoming more and more common. CCC publish a tool called SmartView which allows the public to access all kinds of information. These tools allow the community to be better informed and allow them to have a greater understanding of their environment.

6 CONCLUSIONS

This project resulted in a tool that is valuable in informing an understanding of long term trends in groundwater quality, providing CCC with an early warning system to water quality changes that could have impacts on the water supply, informing future investment considerations including location of new wells, and whether treatment processes or other mitigation measures may be needed in the future.

There are many other applications where large data sets could be better understood when the graphical context is brought to life.

ACKNOWLEDGEMENTS

The author’s would like to acknowledge Environment Canterbury who provided data and input into this project. We would also like to acknowledge Becky Macdonald and Justin Fail who led the CCC Data Visualisation project from Beca’s side.

APPENDIX A – EXPANDED BAR CHARTS

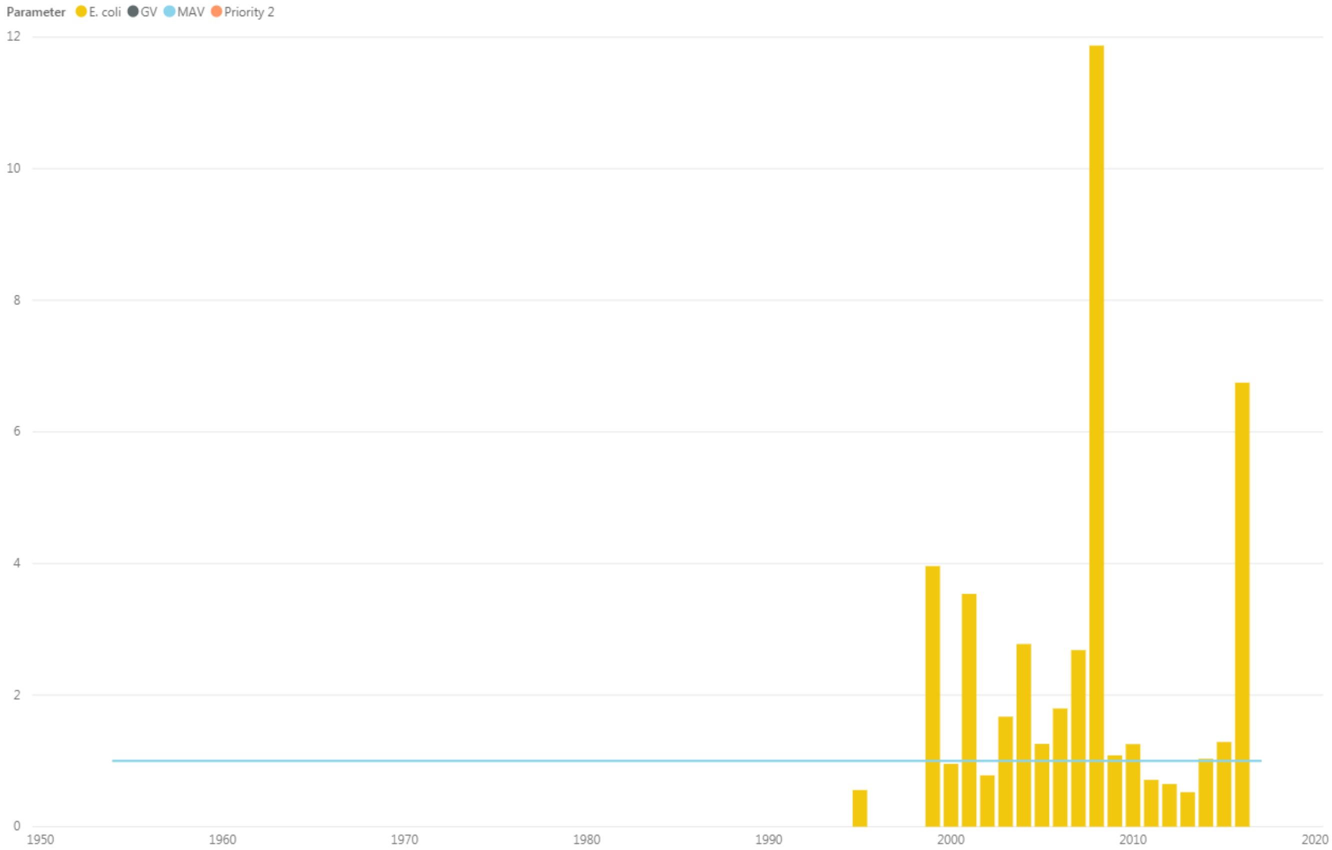

Figure 8: Variations in average E. coli concentration – bar graph

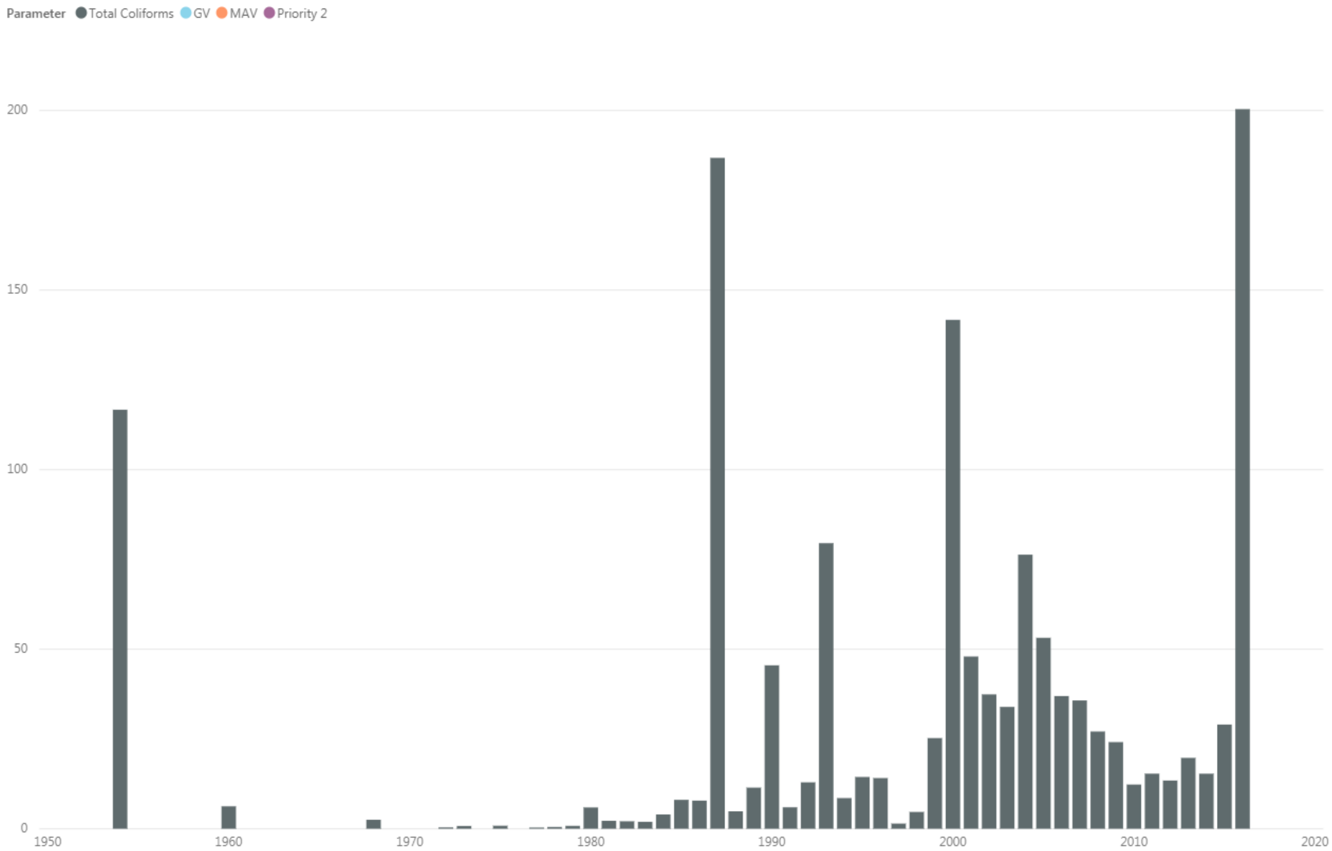

Figure 9: Variations in average total coliforms concentration – bar graph

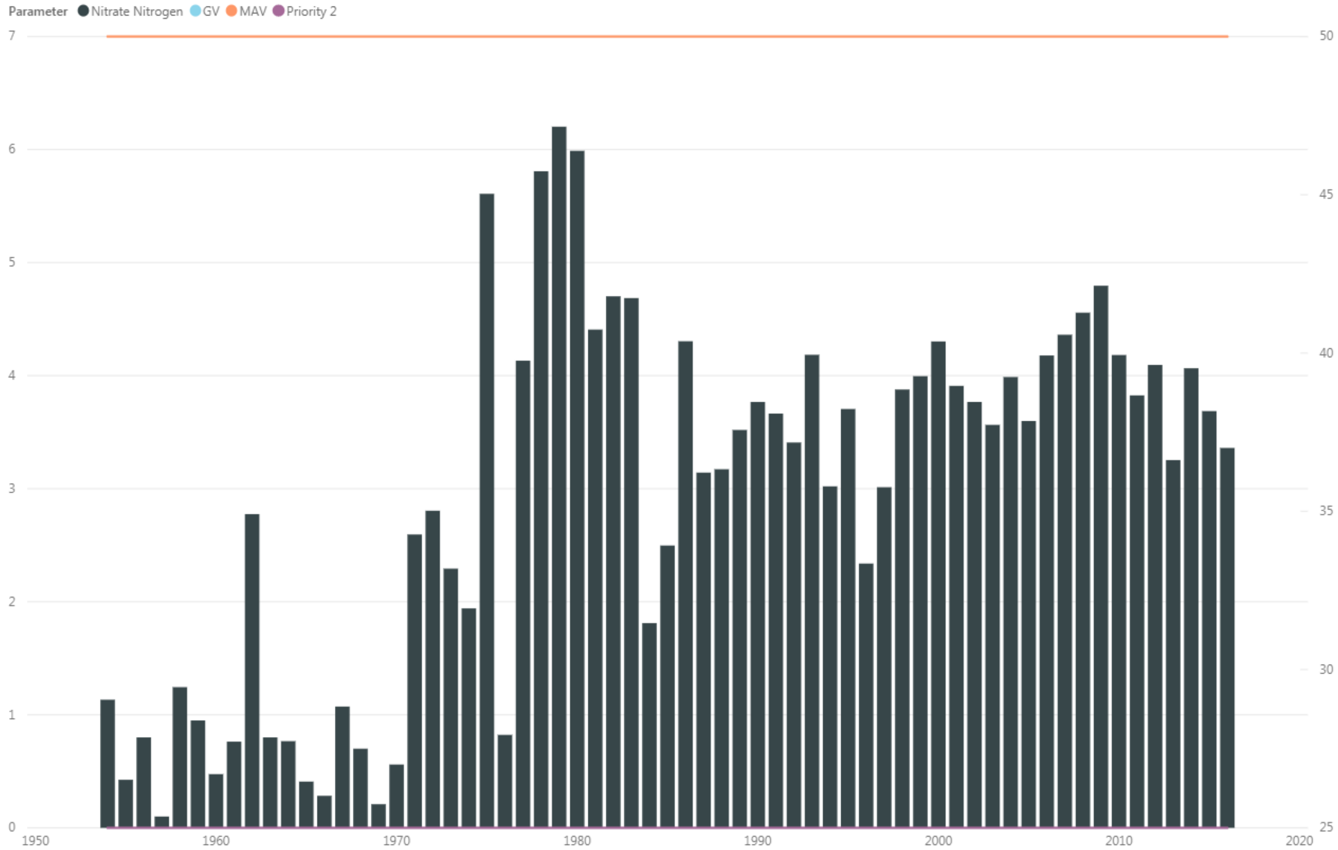

Figure 10: Variations in average nitrate nitrogen concentration – bar graph

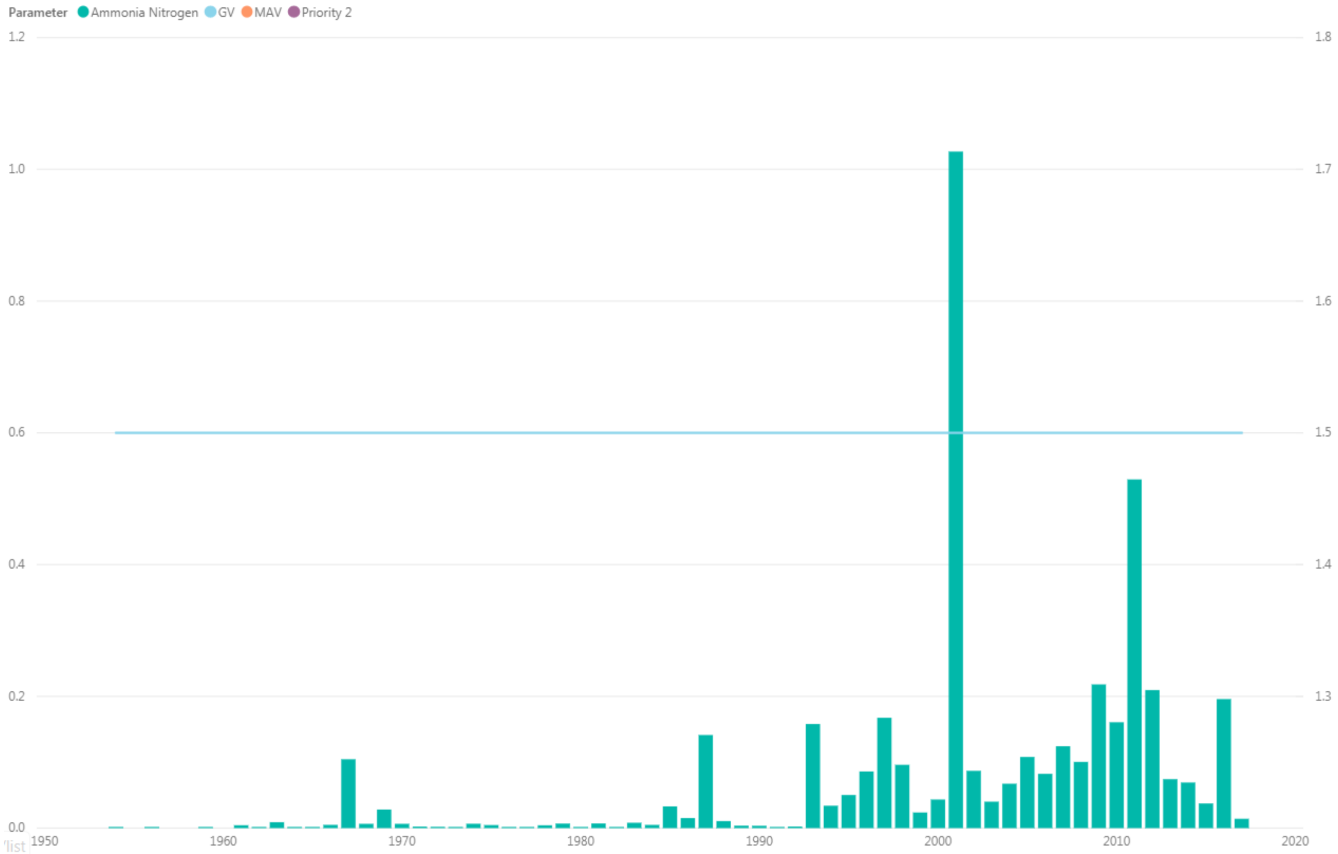

Figure 11: Variations in average ammonia nitrogen concentration – bar graph

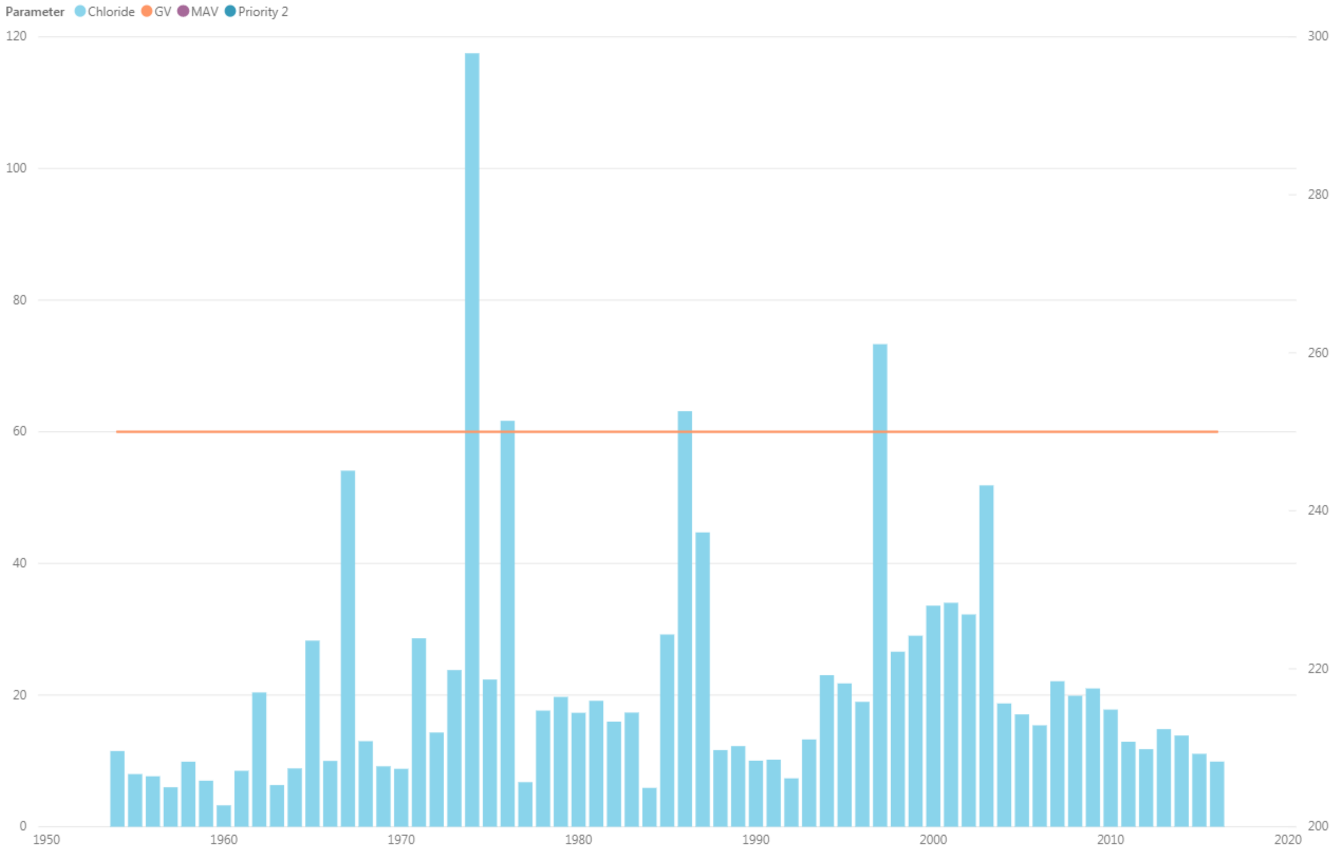

Figure 12: Variations in average chloride concentration – bar graph

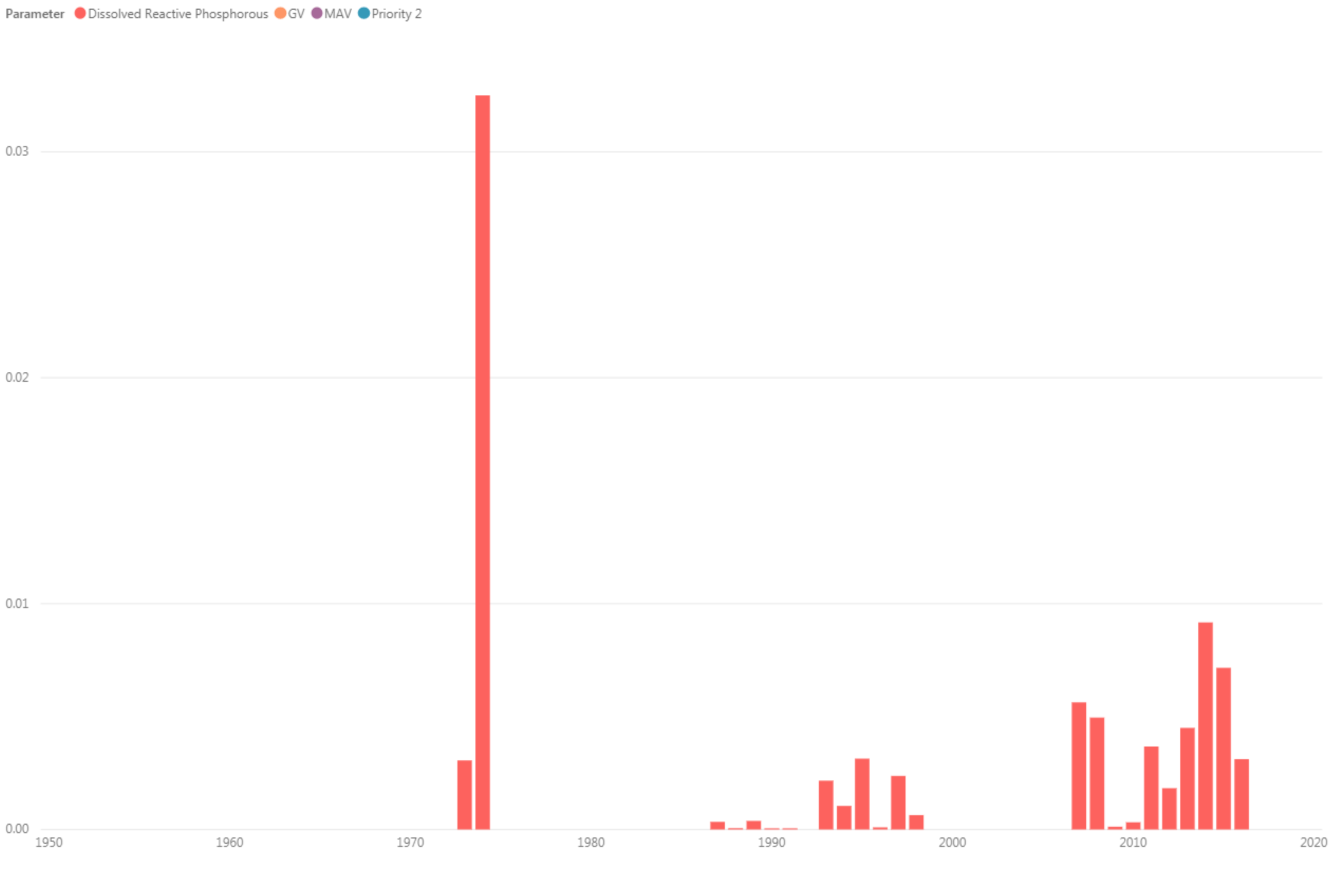

Figure 13: Variations in average dissolved reactive phosphorus concentration – bar graph