Probabilistic modelling of groundwater salinity using borehole and airborne electromagnetics (AEM) data

Neil Symington
Geoscience Australia
Symonston, ACT
neil.symington@ga.gov.au

Anandaroop Ray
Geoscience Australia
Symonston, ACT
anandaroop.ray@ga.gov.au

Chris Harris-Pascal
Geoscience Australia
Symonston, ACT
chris.harrispascal@ga.gov.au

Kok Piang Tan
Geoscience Australia
Symonston, ACT
kokpiang.tan@ga.gov.au

Richard Taylor
Geoscience Australia
Symonston, ACT
richard.taylor@ga.gov.au

Yusen Ley-Cooper
Geoscience Australia
Symonston, ACT
yusen.leycooper@ga.gov.au

Ross C Brodie
Geoscience Australia
Symonston, ACT
ross.brodie@ga.gov.au

SUMMARY

Groundwater is a critical resource for supporting human consumption, stock water, agricultural use, and mineral or energy extraction as well as the environment. However, the quality of groundwater varies enormously from potable to hyper-saline, particularly in the Australian context. To evaluate the suitability of a groundwater resource, the spatial distribution of salinity within an aquifer is typically estimated by measuring the electrical conductivity (EC) of groundwater sampled from boreholes. However, drilling is a logistically and economically challenging task, and hydrogeologists are usually left with a sparse set of measurements from which to infer groundwater salinity over large spatial extents.

Airborne electromagnetic (AEM) surveying is a geophysical technique for estimating the bulk electrical conductivity of the near-surface. Where AEM bulk conductivity is well correlated with groundwater salinity in aquifers, AEM is a useful tool for modelling salinity in the data sparse areas between boreholes. We present a probabilistic method for modelling groundwater salinity and a case study from the Keep River Plains in the Northern Territory. Co-located probabilistic AEM inversions and EC measurements on pore fluids at coincident locations were fused to calculate an empirical joint probability density function. This function allowed us to estimate salinity away from the bores by sampling the ensemble of AEM conductivities. Unlike deterministic methods that provide a single estimate of salinity, our method generates an ensemble of estimates, which can be used to quantify predictive uncertainty. The results provided by our method can feed into decision making while accounting for uncertainty, allowing responsible management of land and water resources.

Key words: groundwater salinity, AEM, probabilistic modelling, inversion

INTRODUCTION

Securing adequate quantities of acceptable quality water for sustaining human livelihoods, well-being and ecosystems is a major challenge of the 21st century (Bakker, 2012). Groundwater will continue to provide a significant fraction of human and ecological water needs, particularly in many areas of Australia with scarce or unreliable surface water resources. However, groundwater varies in quality and a significant groundwater is unsuitable for human use due to high levels of dissolved salts. Measuring or modelling the quality of a groundwater resource is typically challenging due to sparse borehole measurements or hydrogeological heterogeneity.

The traditional approach of estimating the quality of groundwater is to measure the electrical conductivity (EC) or total dissolved solids (TDS) of water sampled from boreholes that intersect aquifers. These measurements can be interpolated using geostatistical techniques to produce a map of groundwater quality. However, in areas where borehole data are sparse these models may have large uncertainties away from boreholes. A number of geostatistical and machine learning approaches have been developed for modelling sparse point data by exploiting correlated variables with high lateral resolution. Airborne electromagnetics (AEM) is a geophysical technique for broad-scale mapping of the near surface. The electrical bulk conductivity profile of the sub-surface can be inferred from AEM data using geophysical inversion (Menke, 2018). Notably in our study, bulk conductivity is correlated with the groundwater EC in saturated, clastic sedimentary aquifers allowing us to model salinity between the bores. Previous works have derived empirical regression functions for calculating salinity from bulk conductivity (Paine 2003, Lawrie et al 2012).

In this study we used a probabilistic approach for inferring groundwater salinity from borehole and AEM data in the Keep River Plains, on the northern border of Western Australia and the Northern Territory. Instead of producing a single model of groundwater salinity, we produce an ensemble of models from which we can calculate statistics for quantifying uncertainty, which are essential for assessing risk for groundwater management. The AEM data were inverted using a probabilistic inversion algorithm to estimate the posterior distribution of bulk conductivities that are able to fit the AEM data. Analytical probability distribution functions for groundwater salinity were estimated by binning high resolution measurements from pore fluid samples extracted from aquifer material recovered using sonic drilling. These probability distributions were propagated through our model.

To demonstrate this workflow, we used our ensemble of groundwater salinity models to calculate the probability that groundwater from the top ten metres of the saturated zone falls within four salinity classes, based on the Australian and New Zealand guidelines for fresh and marine water quality (ANZECC and ARMCANZ, 2000). This study demonstrates how probabilistic methods can assist groundwater management decision making. It does not however, provide any advice about groundwater use on the Keep River Plains.

METHOD AND RESULTS

Data acquisition

Data acquisition, processing and analysis were funded by Exploring for the Future (EFTF), an Australian Government funded program focused on better understanding potential mineral, energy and groundwater resources in northern Australia. In the Keep River Plains area, 11 bores were drilled using sonic drilling and aquifer material was recovered (Figure 1). Pore fluid samples were extracted from aquifer material at decimetre scale sampling density and where possible, EC measurements were recorded. The depth to the water table was measured at each bore and interpolated to derive the water table elevation map shown in Figure 1 using ordinary kriging with a Gaussian variogram with a range of 12,000 m.

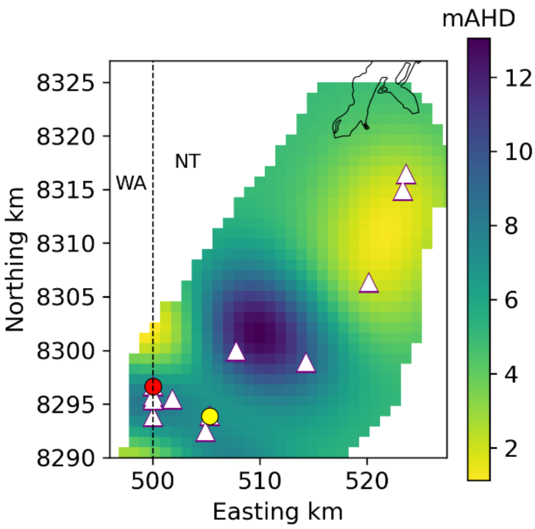

Figure 1. Location of the 11 boreholes where pore fluid EC measurements were collected on a water table elevation map. Boreholes 635735 (yellow circle) and 635740 (red circle) are used to evaluate the model.

AEM data were acquired using the SkyTEM312 system, flown at 200 metre line spacing (see Ray et al., 2019 for technical detail on the Ord-Keep AEM survey). We inverted the AEM data using a Bayesian rjMCMC (reversible jump Markov-chain Monte Carlo) algorithm (Green, 1995) implemented in Geoscience Australia's GA-AEM software (Brodie and Reid, 2013). A notable development used in this work has been the addition of parallel tempering for improved rjMCMC convergence, as implemented in Blatter et al. (2018). The rjMCMC yields an ensemble of samples from a distribution of layered-earth models that fit the data within noise (Figure 2a). The "reversible-jump" formulation of MCMC allows the ensemble to contain conductivity models with differing numbers of layers (Figure 2c). This allows the model complexity to vary and provides the best fit to the data while being parsimonious in the number of layers. We have recently modified the software to also provide Bayesian estimates of "nuisance" parameters, such as the magnitude of the noise in the data and errors in the reported height of the transmitter, allowing the output of the inversion to more accurately reflect the credible range of conductivity models given the collected dataset (e.g., Minsley 2011).

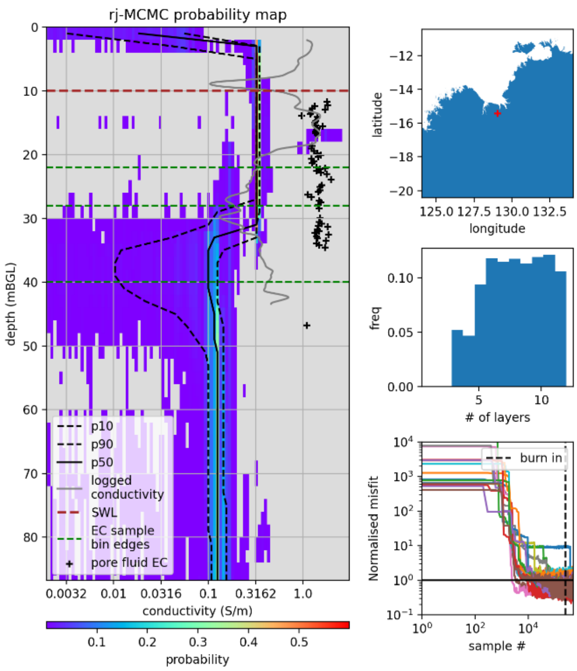

Figure 2. A) A conductivity-depth histogram illustrating the posterior distribution of bulk conductivities in the model ensemble for the AEM fiducial nearest to borehole 635735. EC measurements and the conductivity log from the borehole are plotted over the top. For this study EC data were manually binned and the edges of the EC bins are plotted as green horizontal lines. B) The geographic location of the borehole. C) A histogram showing the frequency number of layers used by the rj-MCMC inversion model. D) A plot showing the RMS misfit for each chain in the parallel tempering inversion. A misfit approaching one reveals that inversion converged and the model ensemble fit the data acceptably.

Deriving relationship between borehole EC and AEM conductivity

Methods for modelling EC with AEM bulk conductivity can only be used if these two variables are correlated. To investigate this relationship, we inverted the nearest AEM fiducial to each of the boreholes. At each borehole the EC measurements were binned into approximately 10m interval and the mean and standard error for each bin was calculated and used to generate an analytical probability density function (PDF) (e.g. Figure 2a). This allowed us to simulate a joined PDF for each interval using the AEM bulk conductivity ensemble.

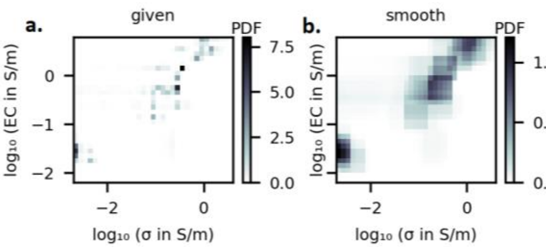

Figure 3a Joint probability density function (PDF) of AEM conductivity σ and groundwater EC derived from joint sampling of bores from all locations. **b.** The joint PDF that has been smoothed to account for the sparse training data. It is this joint PDF that we sample for estimating EC away from the bores.

By sampling the distributions of EC and bulk conductivity we generated a joint probability distribution describing covariance given the uncertainties of both (Figure 3a). Box-car smoothing was applied to account for sparse sampling and make the relationship more general (Figure 3b). The joint PDF reveals that EC and bulk conductivity are positively correlated, with two distinct clusters, which we interpret as two groundwater populations; a very resistive population and a moderately to very conductive population (Figure 3b)

Modelling EC using AEM conductivity

From the joint PDF in Figure 3b, we are able to predict the 'conditional' distribution of EC away from boreholes. The smoothed joint PDF was sampled 10,000 times using bulk conductivity samples, which were themselves sampled from the PDF generated from the probabilistic inversions. The sampling was done using AEM bulk conductivities at ~450 pilot points using values from the top 10 m of saturated thickness as estimated using our water table map. In Figure 4 we display the mean, median (p50), 10th percentile (p10) and 90th percentile (p90) of the modelled EC distribution, plotted at each point. Features in the map that persist across the p10, p50 and p90 distributions, such as high salinity water in the NE, can be confidently interpreted.

Figure 4 suggests there are at least three spatially distinct populations of groundwater salinity that can be interpreted, high salinity in the NE, moderate salinity through the centre of the region and low salinity in the NW and SE margins.

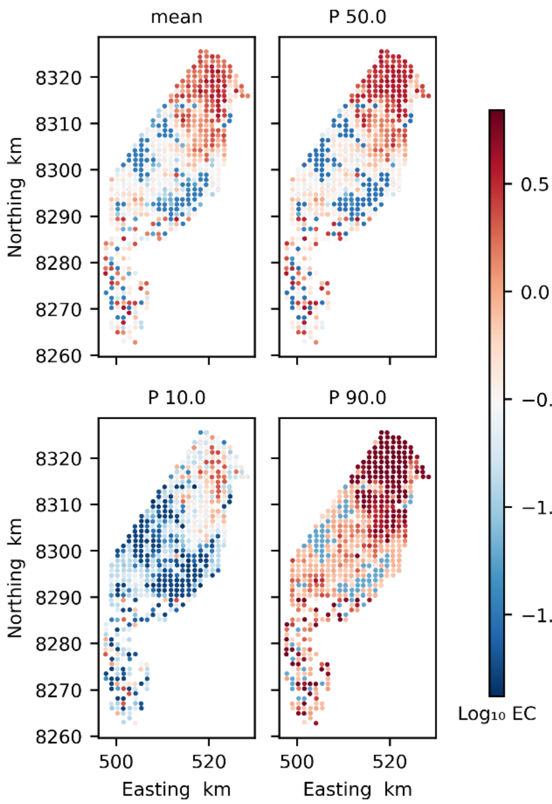

Figure 4. The mean, median (p50), 10th percentile (p10) and 90th percentile (p90) predicted EC value for each point in our study area. The underlying distribution was generated by sampling the joint PDF function in Figure 3b using the sampled value of bulk conductivity from ~450 pilot points.

To evaluate the workflow, we withheld boreholes 635734 and 635740 from the modelling process and compared our predicted salinity with the distribution of EC measurements at these locations (Figure 1). The results reveal that the distribution of EC falls within the high probability regions of our modelled PDF (Figure 5). This suggests that our model is performing well on the validation set.

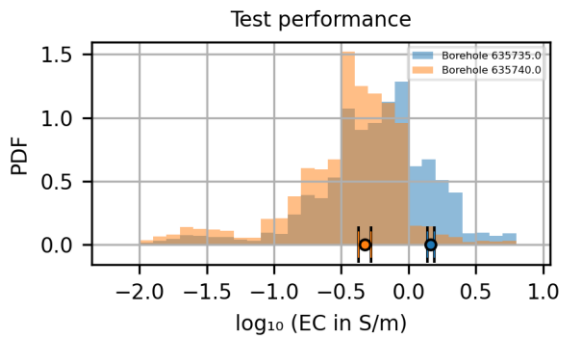

Figure 5. This figure demonstrates the validation of model performance at two boreholes withheld during the model training. The box and whisker plots towards the bottom of the diagram show the mean EC value with error bars to two standard deviations. The histogram represents the ensemble of modelled EC generated using the joint PDF in Figure 3b and bulk conductivities from the nearest AEM sounding. As the groundwater EC distribution falls within the high probability region of the modelled distribution, we are satisfied with model performance.

Probabilistic classification of groundwater salinity

A major benefit of undertaking a thorough uncertainty analysis is that it allows our scientific findings to inform risk analysis. To demonstrate how this work could inform decision making, we have used the results to produce maps of the probability that groundwater within the top 10 metres of saturated aquifer falls within the salinity classes defined in Table 1.

	Class 1: human consumption	Class 2: agricultural use	Class 3: stock water/ process use	Class 4: saline
TDS mg/L	<600	600-1920	1920-16640	>16640
EC S/m	<0.094	0.094-0.3	0.3-2.6	>2.6

Table 1. Classes of groundwater salinity in EC and TDS based on the Australian and New Zealand guidelines for fresh and marine water quality (ANZECC and ARMCANZ, 2000). Note that for reasons of simplicity, salinity classes described here are exclusive. However, groundwater suitability is not exclusive and we are not suggesting that fresh water is unsuitable for agricultural use or stock water.

The results show that there is a clear spatial distribution of high probabilities that groundwater belongs within the classes listed above. The water that is most likely acceptable for human consumption is generally located along the NW and SE margins. The groundwater within the centre and SW of the study show a greater than 50% probability of being classified as suitable for stock water and processing. As expected, the groundwater in the NE, in the vicinity of the estuary, show the highest likelihood of being saline. The advantage of a map such as that presented in Figure 6, is that it allows a groundwater manager to discover the probability that groundwater from the top 10 metres of the saturated zone at any point in the study area has an acceptable level of salinity for some specified purpose. This workflow is easily adapted to answer other specific questions about groundwater salinity.

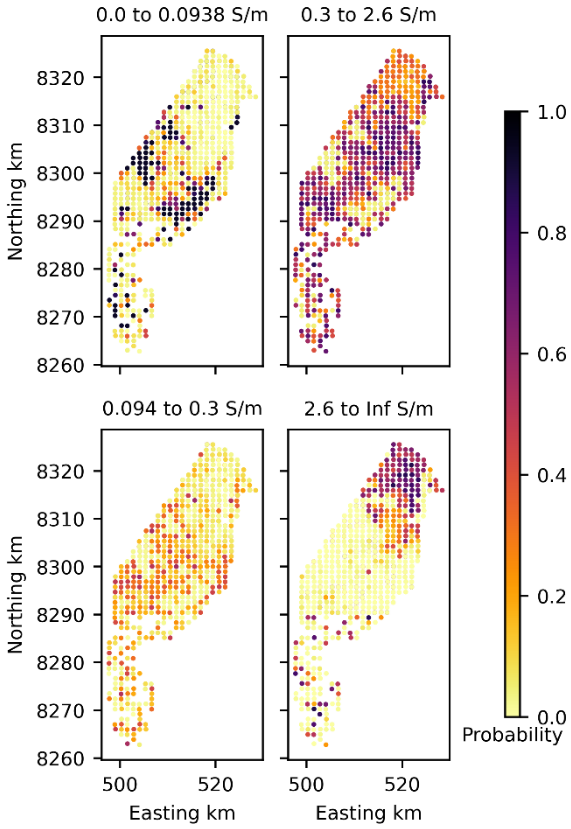

Figure 6. A map showing the probability that groundwater From the top 10 metres of the saturated aquifer falls within each of the four salinity classes described in Table 1. This map provides an accessible way to understand groundwater salinity and uncertainty in the context of specifed groundwater salinity classes.

CONCLUSIONS

In this study, we have demonstrated a workflow for probabilistic modelling of groundwater salinity using borehole data and AEM. The workflow was applied to a hypothetical use case within the Keep River Plains in the Northern Territory, Australia. Our approach considers the uncertainty inherent in AEM inversion and in the sampling of pore fluid. We have also shown how the modelling workflow can be used to support decision making that takes these uncertainties into account.

ACKNOWLEDGMENTS

We thank Sam Buchanan, Anna Haiblen and Baskaran Sundaram for their constructive reviews of this abstract. We thank all the staff from the groundwater team and the Geoscience Australia laboratories for acquiring the borehole data used in this study. This abstract is published with the permission of the CEO of Geoscience Australia.

REFERENCES

ANZECC (Australian and New Zealand Environment and Conservation Council) & ARMCANZ (Agriculture and Resource Management Council of Australia and New Zealand), 2000. Australian and New Zealand guidelines for fresh and marine water quality, ANZECC & ARMCANZ, Canberra.

Bakker, K. 2012. Water security: research challenges and opportunities. Science, 337(6097), 914-915.

Blatter, D., Key, K., Ray, A., Foley, N., Tulaczyk, S., Auken, E., 2018, Trans-Dimensional Bayesian Inversion of Airborne Transient EM Data From Taylor Glacier, Antarctica. Geophysical Journal International, Volume 214, 1919–1936

Brodie, R. C., and Reid, J. 2013. Monte Carlo Inversion of SkyTEM(TM) AEM data from Lake Thetis, Western Australia, in proceedings of 23rd International Geophysical Conference and Exhibition, Melbourne.

Lawrie, K.C., Brodie, R.S., Tan, K.P., Gibson, D., Magee, J., Clarke, J.D.A., Halas, L., Gow, L., Somerville, P., Apps, H.E., Christensen, N.B., Brodie, R.C., Abraham, J., Smith, M., Page, D., Dillon, P., Vanderzalm, J., Miotlinski, K., Hostetler, S., Davis, A., Ley-Cooper, A.Y., Schoning, G., Barry, K. andLevett, K. 2012. BHMAR Project: Data acquisition, processing, analysis and interpretation methods. Geoscience Australia Record 2012/11. 824p

Green, P. J., 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination, Biometrika, 82(4), 711-732.

Leaney, W.S., 1990. Parametric wavefield decomposition and applications: 60th Meeting, SEG, San Francisco, Expanded Abstracts, 1097-1100.

Menke W., 2018. Geophysical data analysis: discrete inverse theory, Academic Press.

Minsley B. J., 2011. A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data. Geophysical Journal International 187:252–72.

Paine, J.G., 2003. Determining salinization extent, identifying salinity sources, and estimating chloride mass using surface, borehole, and airborne electromagnetic induction methods. Water Resources Research, 39(3).

Ray A., Symington N. & Steklova K., 2019. Ord–Keep Rivers AEM: SkyTEM® airborne electromagnetic data, Geoscience Australia, Canberra.