Groundwater Knowledge Integration System (GKIS):

Probabilistic groundwater prospectivity mapping with iterative updating of conceptualisation

Luk JM Peeters, Tao Cui, Trevor Pickett, Mat Gilfedder,
Dirk Mallants, Andrew Taylor, Zhenjiao Jiang,
Brady Flinchum, Kevin Cahill, Tim Munday

Goyder Institute for Water Research
Technical Report Series No. 20/01

www.goyderinstitute.org

Goyder Institute for Water Research Technical Report Series ISSN: 1839-2725

The Goyder Institute for Water Research is a partnership between the South Australian Government through the Department for Environment and Water, CSIRO, Flinders University, the University of Adelaide, the University of South Australia and the International Centre of Excellence in Water Resource Management. The Institute enhances the South Australian Government's capacity to develop and deliver science-based policy solutions in water management. It brings together the best scientists and researchers across Australia to provide expert and independent scientific advice to inform good government water policy and identify future threats and opportunities to water security.

Enquires should be addressed to: Goyder Institute for Water Research
Level 4, 33 King William Street
Adelaide, SA 5000
tel: 08 8236 5200
e-mail: enquiries@goyderinstitute.org

Citation

Peeters, L.J.M., Cui, T., Pickett, T., Gilfedder, M., Mallants, D., Taylor, A., Jiang Z., Flinchum, B., Cahill, K., Munday, T. (2020) *Groundwater Knowledge Integration System (GKIS): Probabilistic groundwater prospectivity mapping with iterative updating of conceptualisation.* Technical Report Series 20/01, Goyder Institute for Water Research, Adelaide.

Disclaimer

The CSIRO as the project partner, advise that the information contained in this publication comprises general statements based on scientific research and does not warrant or represent the completeness of any information or material in this publication. The project partners do not warrant or make any representation regarding the use, or results of the use, of the information contained herein about to its correctness, accuracy, reliability, currency or otherwise and expressly disclaim all liability or responsibility to any person using the information or advice. Information contained in this document is, to the knowledge of the project partners, correct at the time of writing.

Contents

Figures

Tables

Executive Summary

Ensuring water supply is a crucial aspect of social, cultural and economic development in remote, arid regions in Australia.

One of the great challenges in sustainable groundwater management in the Australian outback is siting bores that can yield the necessary volume of water with acceptable quality, for a predefined period of time, without compromising other uses, economic, ecological or social, of the groundwater resource. Especially in data poor areas, siting a bore has always an associated risk of failure of not finding a sustainable water resource. Even if water of suitable quality is found during drilling and initial yields are promising, the bore may not be able to sustain the required pumping rate and may run dry after a couple of years.

This report presents the Groundwater Knowledge Information System (GKIS), which aims to provide an assessment of the probability of successfully siting a sustainable bore. Initially, based on information available from a desktop study, these probabilities will generally be low. By integrating additional information, these probabilities will be refined spatially, i.e. probabilities will decrease in areas where suitable water resources are not likely to be found, while in more promising areas, probabilities will increase. There are many factors that affect the probability of success, including aquifer geometry, hydraulic properties and salinity but also the conditions for sustainable development, such as the volume of water required, the timeframe and the maximally allowable salinity. The GKIS is designed to integrate additional data, regardless of scale or extent, as it becomes available to update the probability of success. As more data and information becomes available, this can be incorporated in an iterative succession to further improve the conceptualisation.

Finding a potential water supply that meets the defined salinity thresholds and pumping sustainability criteria depends on aquifer thickness, hydraulic conductivity (*K*), porosity and salinity, and on the thresholds of required pumping rate, duration and salinity. A global sensitivity analysis highlights that in the Anangu Pitjantjatjara Yankunytjatjara (APY) Lands, aquifer salinity is the most important variable, followed by *K*. The salinity threshold is shown to be more important than pumping rate or duration in determining acceptable groundwater resource conditions.

In the Musgrave Province of South Australia, the deep palaeovalley systems are most prospective for groundwater development, with higher probability for sustainable development in the Lindsay West Palaeochannel than in the Lindsay East Palaeochannel. The fractured bedrock aquifer has lower probability for sustainable development, but is still prospective, especially when looking for low pumping rates and low salinities.

Through the implementation of the three conceptualisations, the probability in the palaeovalley is increased as more information becomes available. The current method is based on four key hydraulic parameters including hydraulic conductivity, specific yield (porosity), saturated aquifer thickness, and salinity. To improve the estimation, efforts should be made to improve the estimate/measurement of these controlling parameters.

Acknowledgments

This study was carried out as part of the G-Flows Stage-3 Project, with funding through the Goyder Institute of Water Research.

We acknowledge the traditional owners of the Anangu Pitjantjatjara Yankunytjatjara (APY) Lands, the Pitjantjatjara, Yankunytjatjara and Ngaanyatjarra people.

We would like to thank the APY General Manager and the entire APY Executive Board who were supportive of the G-Flows Stage-3 project.

1 Background

The Goyder Institute's Facilitating Long-Term Outback Water Solutions Stage 3 (G-FLOWS 3) Project is primarily concerned with the validation of the regional hydrogeological framework methodology developed in projects G-FLOWS 1 and 2. It also seeks to extend that framework to provide a regional groundwater resource assessment in a data poor area through the integration of regional geophysical and geological data, coupled with targeted hydrogeological data acquisition (principally through drilling) and interpretation.

Research outcomes for GFLOWS 3 include workflows for defining groundwater resources, conceptual models of aquifer systems and groundwater quality for community, industry and environment in remote areas characterised by a paucity of data. These workflows will have application in other parts of the South Australia, and elsewhere in Australia where knowledge of the groundwater resources is limited. In the Anangu Pitjantjatjara Yankunytjatjara (APY) Lands it will also deliver a regional-scale resource assessment, providing a framework for determining sustainable groundwater use across a substantive part of the region.

Part of the G-FLOWS 3 project will extend the scientific approaches undertaken in the preceding stages 1 and 2, where in the former, legacy airborne electromagnetic (AEM) data acquired by mineral explorers were re-interpreted to develop a hydrogeological framework to identify groundwater resource potential of the region (see Munday *et al.* 2013). The Geological Survey of South Australia (GSSA) and the Department for Environment and Water (DEW) have invested in the acquisition and processing of new AEM data which covers the Musgrave Province in South Australia, with the aim of identify potential palaeovalley hosted groundwater resources which could help secure water supplies for remote communities and support potential mineral resource developments as well as development opportunities such as the pastoral industry.

2 Introduction

One of the great challenges in sustainable groundwater management in the Australian outback is siting bores that can yield the necessary volume of water with acceptable quality, for a predefined period, without compromising other uses, economic, ecological or social, of the groundwater resource. Especially in data poor areas, siting a bore has an associated risk of failure of not finding a sustainable water resource. Even if water of suitable quality is found during drilling and initial yields are promising, the bore may not be able to sustain the required pumping rate and may run dry after a couple of years.

The objective of the Groundwater Knowledge Information System (GKIS) is to provide an assessment of the risk of failure or, to put it in a more positive frame, the probability of successfully siting a sustainable bore. Initially, based on information available from a desktop study, these probabilities will generally be low. By integrating additional information, these probabilities will be refined spatially, i.e. probabilities will decrease in areas where suitable water resources are not likely to be found, while in more promising areas, probabilities will increase. There are many factors that affect the probability of success, including aquifer geometry, hydraulic properties and salinity but also the conditions for sustainable development, such as the volume of water required, the timeframe and the maximally allowable salinity. The GKIS is designed to integrate additional data, regardless of scale or extent, as it becomes available to update probability of success.

The Methods section details the theory behind the probabilistic mapping. This approach has been applied in the remote and arid Anangu Pitjantjatjara Yankunytjatjara (APY) Lands (part of the Musgrave Province) in South Australia, based on information contained in the G-FLOWS 1 report (Munday et al. 2013), the data from the recent G-Flows Stage-3 drilling program (Keppel et al. 2019, Costar et al. 2019), and the historically available and the new acquired AEM data (Soerensen et al. 2018). The discussion section uses the assumption-hunting approach of Peeters (2017) to list and analyse the various implicit and explicit assumptions embedded in this approach. The final section summarises the work, provides conclusions and outlines potential further work.

3 Study area

The Anangu Pitjantjara Yankunytjatjara (APY) Lands are situated in central Australia, in the north west of South Australia. The area has an arid climate with a low and unreliable rainfall, averaging less than 300 mm/yr and high potential evaporation, in excess of 3000 mm/yr. The area is underlain by the Proterozoic Musgrave Province, which forms the topographically elevated Musgrave Ranges in the north (Figure 1). The Proterozoic granites and gneisses are intensely weathered and covered with a Cenozoic palaeovalley network and Quaternary sand dunes.

There are no permanent water features in the landscape, only a small number of ephemeral streams. Groundwater supply for local communities is sourced from bores both in the weathered bedrock and in the palaeovalley sediments. Appendix A provides a summary of the available borehole information in terms of lithology, yield, water level and salinity. Bores are generally less than 100 m deep and half of the bores have standing water levels less than 12 m below ground surface. Yield is highly variable with very few bores producing more than 100 L/s. Salinity is also variable across several orders of magnitude, ranging from tens to tens of thousands mg/L.

A comprehensive desktop study in 2013 produced a regional scale hydrogeological framework (Figure 2) (Munday et al. 2013), which was followed in 2016 by a large scale airborne electromagnetic survey (Soerensen et al. 2018) and a drilling campaign in 2018 (Keppel et al. 2019; Krapf et al. 2019; Costar et al. 2019).

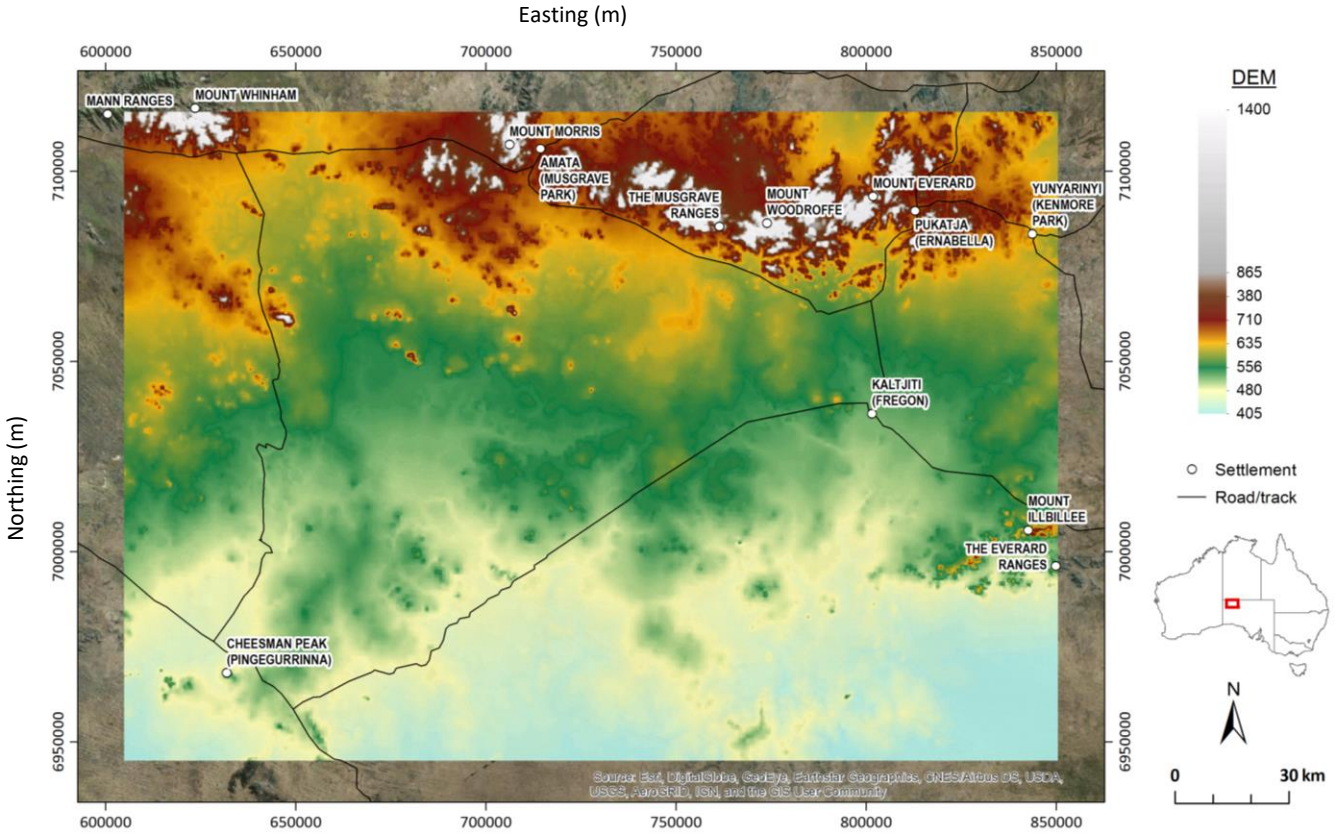


Figure 1. Land surface topography of the study area, overlain with road network and main community and peak features. Note that the coloured box shows the case-study area which is used in subsequent figures in this report.

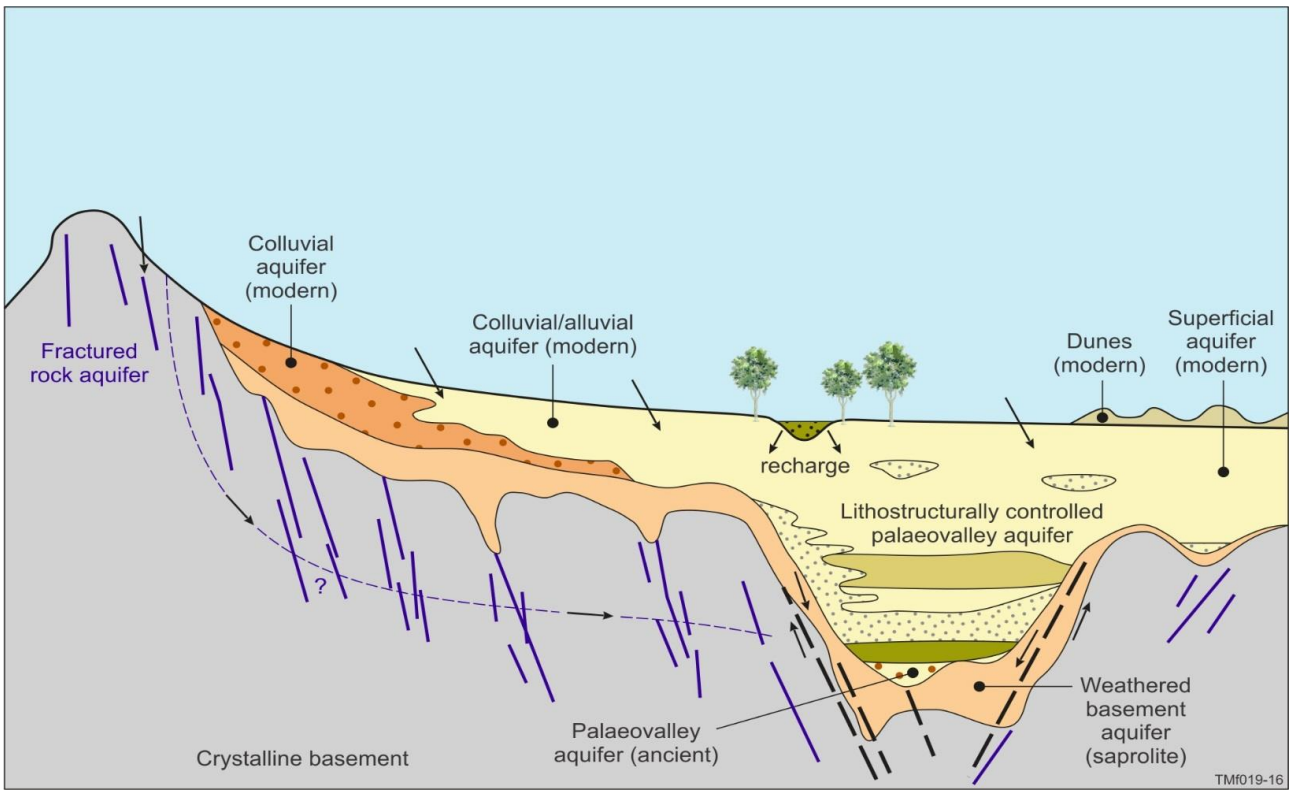


Figure 2. Hydrogeological conceptualisation for the Musgrave Province (after Munday et al. 2013 and Gogoll 2016).

4 Methodology

Probabilistic methods have become an important part of earth science research (Scheidt et al. 2018) and in recent years they have increasingly been applied in practical groundwater modelling studies (e.g. Peeters et al. 2018). Probabilistic approaches in groundwater studies tend to focus on relatively well-understood, data rich areas and be applied for water resource management, impact assessment or contaminant transport (Bode et al. 2018). Probabilistic approaches in groundwater exploration, characterisation and mapping mainly focus on GIS and data mining techniques to establish an empirical relationship between physico-chemical attributes and successful bores (see e.g. Naghibi et al. 2017). Such empirical relationships can be highly efficient but require a large database to infer such relationships. Even if these relationships can be inferred, they are often very region-specific.

Rather than relying on a data-driven approach, the GKIS method starts from first principles by characterising the potential for groundwater production. The largest drawback however is that most groundwater exploration efforts aim at characterising the potential for groundwater production, without explicitly considering the sustainability of the resource. As outlined in Bredehoeft et al. (1982), efficient water resource management requires to explicitly define how much water needs to be produced and what the constraints are with regards to quality and economic, social and environmental impact of water production. The probabilistic groundwater mapping methodology therefore starts with defining relevant targets and thresholds. The second step is then to compute the probability of meeting each threshold based on the available data and understanding of groundwater flow processes. The final step is to compute the joint probability of simultaneously meeting all targets and thresholds.

4.1 Targets and thresholds

Efficient water resource management requires explicitly defined targets for water production and constraints on quality and economic, social and environmental impact (Bredehoeft et al. 1982). Within the context of potential sustainable groundwater development in the study area, the main objectives for groundwater management are to find a location where a pumping rate Q_s can be sustained over a period t_s, producing water with a salinity C less than the threshold value C_s.

Sustaining the pumping rate Q_s implies, at the very least, that the drawdown over period t_s does not cause the bore to run dry. The maximum potential pumping rate Q_{max} can be defined as the pumping rate that would cause the groundwater level close to the bore to drop below the top of the bore screen. In addition to that, the volume of water available in the aquifer, V, needs to larger than the volume of water that will be pumped during the development time horizon. Note that this requires an explicit definition of the volume of aquifer under consideration in the probability calculation.

The probability of developing a sustainable bore, P_s, is the joint probability of meeting all thresholds, which can be written as:

$$P_s = P(Q_{max} > Q_s, V > Q_s\, t_s, C < C_s) \qquad \text{Eq. 1}$$

The targets and thresholds captured in Eq. 1 need to be tailored to the specific management objectives for sustainable groundwater management, by for instance introducing a drawdown threshold at existing bores or a flow threshold at springs.

Estimation the probability of threshold exceedance requires probability distributions for each of the quantities in Eq. 1 (Q_{max}, V and C). These probability distributions need to be estimated from the available data and system knowledge, either directly, such as salinity, or derived from aquifer geometry and properties, such as maximum pumping rate and available water volume.

4.2 Maximum pumping rate Q_{max}

Drawdown in an aquifer due to pumping can be described as (Theis, 1935):

$$\Delta h = \frac{Q}{4\pi Kd} W\left(\frac{Sr^2}{4Kdt}\right) \quad \text{Eq. 2}$$

with Δh the drawdown (m) at distance r (m) from the bore in an aquifer with hydraulic conductivity K (m/d), thickness d (m) and storativity S (-) due to pumping with pumping rate Q (m^3/d) over time t (d). $W()$ is the exponential integral or well function. This equation assumes a fully penetrating well and that the hydraulic properties do not change during pumping. Storativity is the volume of water released from storage per unit decline in hydraulic head in the aquifer, per unit area of the aquifer. In a confined aquifer storativity is equal to the specific storage (S_s) multiplied with the saturated thickness. In an unconfined aquifer, storativity is equal to specific yield (S_y) which is often approximated by the effective porosity (n_e).

In section 2.1, we defined Q_{max} as the maximum pumping rate that will not cause the bore to fall dry. This can be made more explicit as the pumping rate that will cause groundwater levels to drop to the top of the well screen. As the position and length of the well screen of future bores is not known, we redefine Q_{max} as the pumping rate that, at 1 m from the bore, causes a drawdown $\Delta h = \alpha d$. With α equal to 0.05, this corresponds to a drawdown not in excess of 5% of the aquifer thickness. Q_{max} can then be written as:

$$Q_{max} = (\alpha d)\frac{4\pi Kd}{W\left(\frac{S}{4Kdt}\right)} \quad \text{Eq. 3}$$

4.3 Aquifer volume (V)

The volume of water available in the aquifer V can be written as:

$$V = SAd \quad \text{Eq. 4}$$

with A the area of the aquifer.

The available volume of water changes through time in response to regional recharge and discharge fluxes. For the case study in the APY lands, these are not taken into account, because these fluxes are estimated to be relatively small (Leaney et al. 2013). There are no permanent groundwater discharge features in the APY landscape. By not accounting for recharge or discharge the volume of water is underestimated. This provides a conservative estimate of potential water availability which fits in a risk averse water management framework in arid regions.

4.4 Probability of developing a sustainable bore ($\boldsymbol{P_s}$)

The probability of developing a sustainable bore is calculated on a raster. Probability distributions of the hydraulic properties (K,S), geometry (A, d) and salinity (C) are estimated from available data in a way that is consistent with the hydrogeological conceptualisation. By generating a large number of randomly chosen samples from these distributions and evaluating these with the equations above allows to numerically approximate the probabilities that each individual constraint is met and the joint probability of meeting all constraints simultaneously.

The hydrogeological conceptualisation is expected to evolve as new information becomes available. As the conceptualisation changes, the probability distributions will need to be updated. We iteratively compute the probability of developing a sustainable bore in the APY three times. In each iteration more information becomes available which merits a revisit and update of the conceptualisation.

We only include hydrogeological factors in the calculation of the probability of a successful bore. This probability will also be affected by logistical factors, such as the accessibility and steepness of the terrain or the distance to an existing water distribution network and cultural factors, such as permissions for drilling issued by the traditional owners and custodians of the land. While some physical attributes, such as steepness of the terrain, are straightforward to obtain, formalising cultural factors is not trivial. The probability of a developing a sustainable bore as calculated here should therefore be considered as a contribution to the suite of information required to inform an investment decision.

5 Results

The above presented GKIS methodology can be used to formally and systematically integrate hydrogeological information to spatially estimate the probability that a predefined pumping rate can be sustained over a given period. In this section, the methodology was implemented for three different abstractions of the hydrogeological conceptualisation shown below (Figure 3). The Jupyter Notebook version of the implementation is also be provided in the supplementary materials.

The three different abstractions are:

1. Three hydrostratigraphic units
2. Single aquifer with equivalent properties
3. Two aquifer system

The GKIS is a grid-based probability estimate method; the current implementation was based on the grid with 500 m resolution.

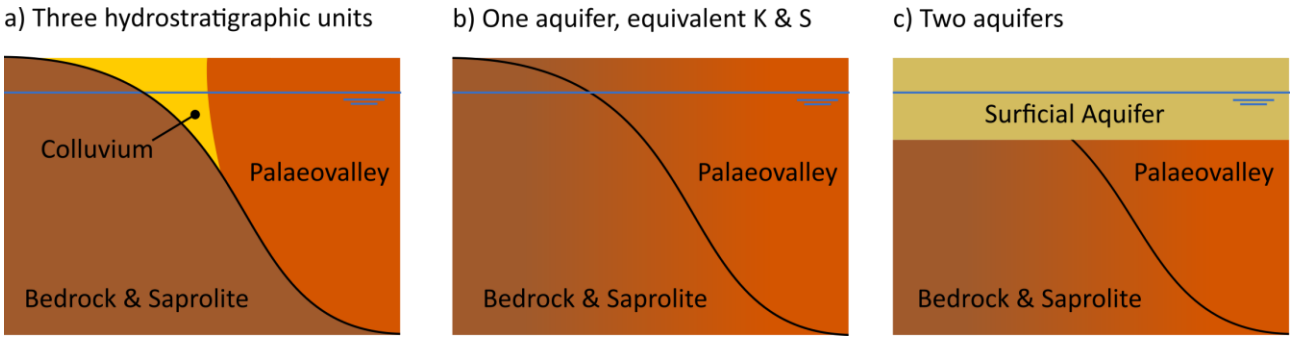


Figure 3. Abstractions of hydrogeological conceptualisation.

5.1 Three hydrostratigraphic units

5.1.1 INTRODUCTION

The hydrogeological framework established in Munday et al. (2013) distinguishes three major aquifer types in the region:

1. Palaeovalley: alluvial and calcrete aquifers with potential for buried palaeovalley aquifers underneath
2. Colluvium: aquifers in mainly colluvial sediments
3. Bedrock: aquifers in saprolite or weathered and fractured rock

To evaluate the probability that a pumping rate can be sustained, grids of aquifer area, depth, porosity/storage and hydraulic conductivity are needed. Figure 4 displays a rasterised version of the hydrogeological framework over the study area.

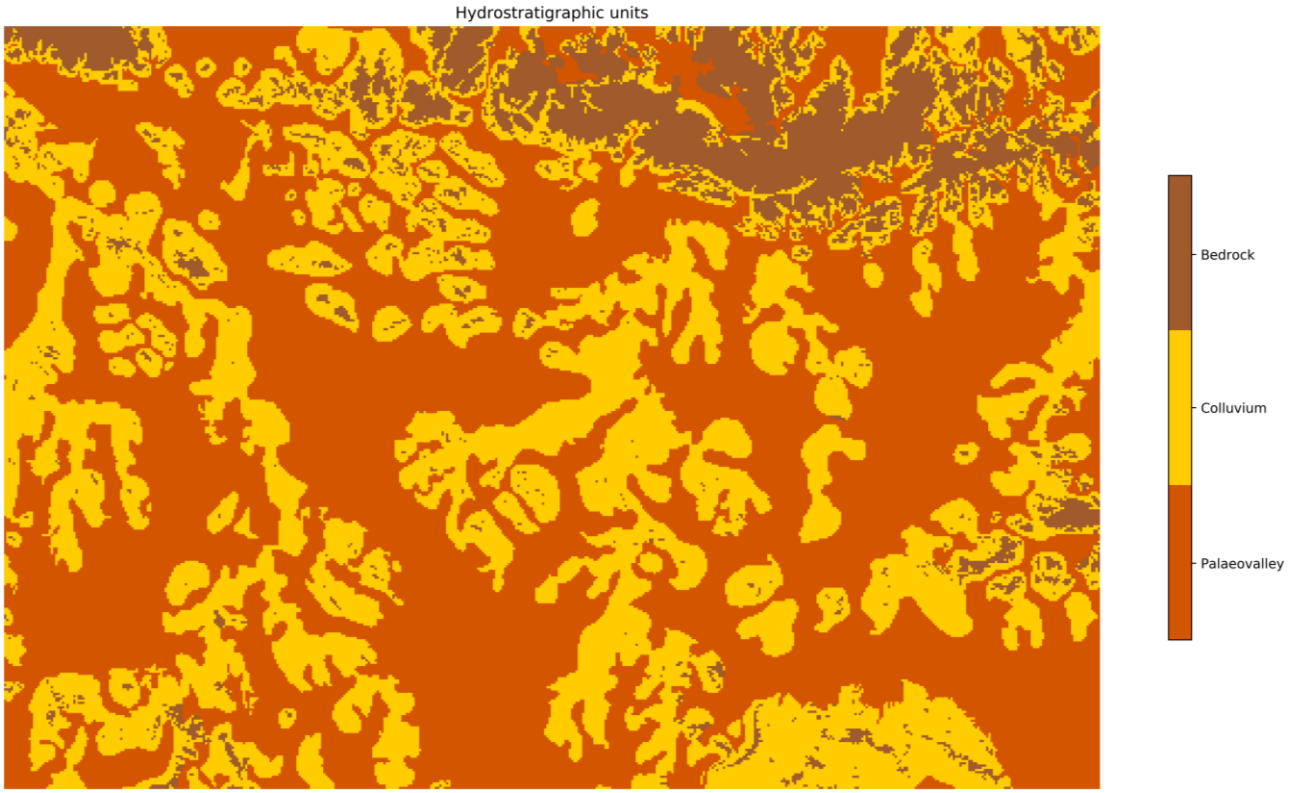


Figure 4. The distribution of three hydrogeologic units based on Munday et al. (2013) (note that location of this area is shown in Figure 1).

5.1.2 PRIOR PROBABILITY DISTRIBUTION FUNCTIONS

The prior probability distribution functions (pdf) capture the belief, in a Bayesian sense, of what the range and distribution of a parameter is, informed by the current knowledge of the aquifer system. We chose to use beta distributions to describe these pdfs. These analytic functions describe pdfs with only two shape parameters (alpha and beta) and two scale parameters (minimum and maximum). They are very flexible and are bounded. For every hydrostratigraphic units (HSU) (palaeovalley, colluvium, bedrock) each variable (hydraulic conductivity, specific yield, saturated thickness and salinity) is described with these four parameters.

The beta distribution parameters need to reflect the current knowledge of the system. As no reliable measurements are available to validate or check these distributions, the narrative that guided the choice of parameter values is provided in Table 1. The resulting distribution is shown in Figure 5.

Table 1: Narrative that guided the selection of parameters that control the beta distribution for different variables and hydrogeologic units.

PROPERTY	HYDROSTRATIGRAPHIC UNITS (HSU)	NARRATIVE
Log10 hydraulic conductivity (*K*)	Palaeovalley	An almost symmetrical distribution between 0.001 m/d and 1 m/d, with a median close to 0.5 m/d. It represents a range of unconsolidated sediments such as silt and fine sands. There is little evidence from borehole logs for extensive deposits of coarse sands or gravels. A slight asymmetry on the lower end towards lower values captures the presence of consolidated calcrete lenses or more clayey lenses/layers in palaeovalley settings.
Log10 hydraulic conductivity (*K*)	Colluvium	A symmetrical distribution between 0.01 m/d and 1 m/d, with a median close to 0.5 m/d. It represents a similar range of lithologies as the Palaeovalley HSU, but with less prevalent clay lenses. There is therefore no asymmetry towards lower values such as for the Palaeovalley HSU.
Log10 hydraulic conductivity (*K*)	Bedrock	A strongly left skewed distribution between 0.0001 m/d and 10 m/d, with a median close to 0.001 m/d. Bedrock with low fracture density will have a low K, while bedrock with very high fracture density will locally have high K values. High density fracture zones are generally associated with areas where faults or lineaments intersect (Gogoll et al. 2018).
Specific yield (Sy)	Palaeovalley	A slightly left skewed distribution between 0.01 and 0.3, with median close to 0.1. It represents unconsolidated sediments, with especially deeper in the palaeovalleys' coarser material.
Specific yield (Sy)	Colluvium	Left skewed distribution between 0.01 and 0.3, with median close to 0.06. Median is less than Palaeovalley HSU to reflect less prevalent coarser material.
Specific yield (Sy)	Bedrock	Strongly left skewed distribution between 0.01 and 0.3, with a median close to 0.03. This reflects a very low porosity in bedrock with low fracture density and high porosity in densely fractured rock.
Thickness	Palaeovalley	Left skewed distribution between 10 m and 100 m, with a median close to 30 m. Where the Palaeovalley HSU is mapped, a thickness of at least 10 m is assumed. Most bores in this HSU are less than 30 m deep, but locally bores have been drilled that encountered palaeovalley sediments to depths of more than 80 m.
Thickness	Colluvium	Left skewed distribution between 0 m and 75 m, with a median close to 15 m. The mapped colluvium can be very variable, but is expected to be less thick than the mapped Palaeovalley HSU.
Thickness	Bedrock	Strongly left skewed distribution between 0 m and 50 m, with a median close to 7 m. Areas with low fracture density have low saturated thickness. In areas with high fracture density, faults and fractures can contribute to flow at greater depth.
Salinity	Palaeovalley	Left skewed distribution between 500 mg/L and 25,000 mg/L with a median close to 5000 mg/L. Low diffuse recharge and high ET are expected to result in high salinity. At greater depths in the aquifer older, less saline water can be found, especially when clay lenses limit mixing with younger, more saline water. The older, less saline water in this setting corresponds to water infiltrated under different climatic conditions, with higher precipitation and less evaporation.
Salinity	Colluvium	Left skewed distribution between 500 mg/L and 25,000 mg/L with a median close to 2500 mg/L. Similar to Palaeovalley HSU, but lower median to reflect the Colluvial HSUs are closer to localised recharge areas at the edge of large outcrops.
Salinity	Bedrock	Strongly left skewed distribution between 500 mg/L and 10,000 mg/L, with a median close to 1000 mg/L. Bedrock HSUs are expected to have a thicker unsaturated zone and are thus less prone to salinity increase due to evapotranspiration. Hydrogeologically isolated fracture zones may have retained older, less saline groundwaters.

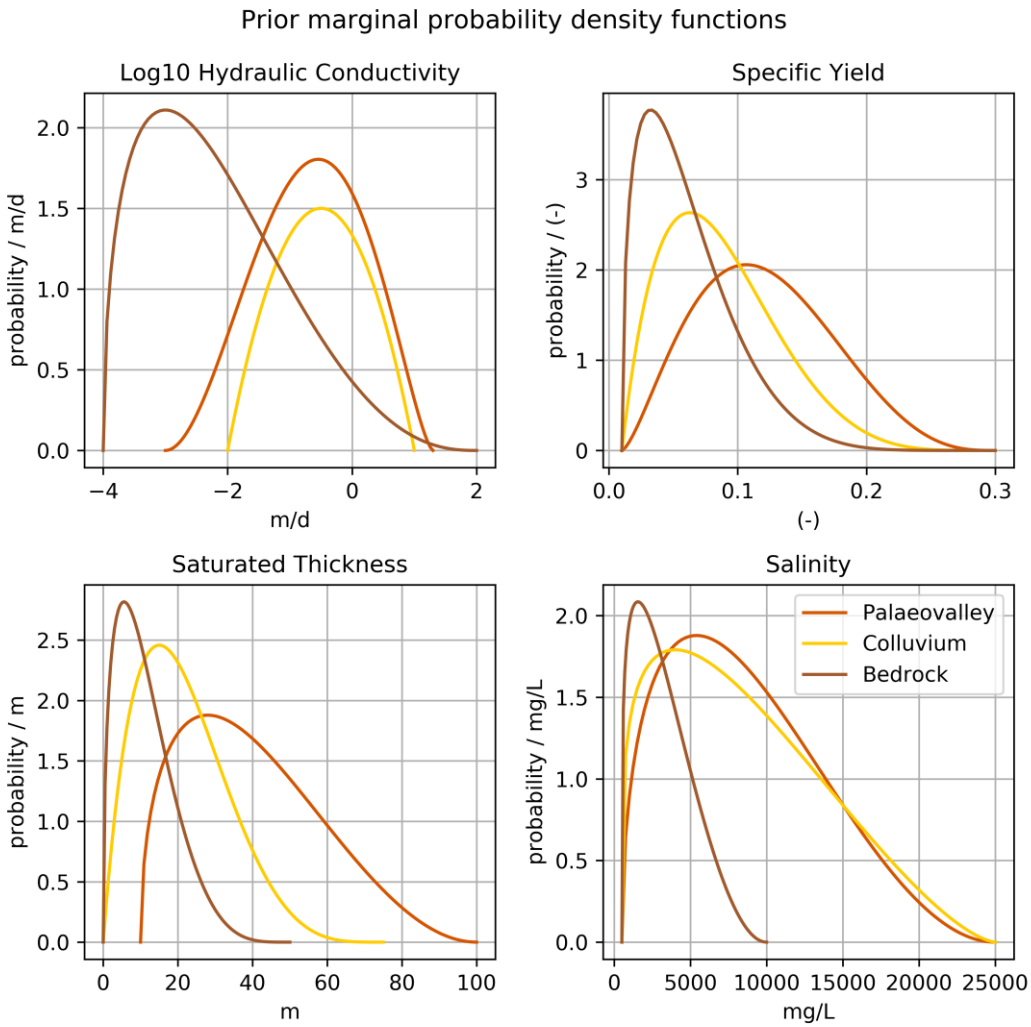


Figure 5. Prior β probability distributions for Log10 hydraulic conductivity, specific yield, saturated thickness and salinity for the alluvial (Palaeovalley), Colluvium, and basement (Bedrock) hydrogeological units based on data and interpretation in Varma (2012) and Munday et al. (2013).

The narrative table shows that these properties are not independent. For the Bedrock HSU for instance, fracture density influences conductivity, specific yield and saturated thickness. In the Palaeovalley HSU, a thick aquifer is likely to be a palaeovalley filled with coarser material and hence a higher permeability and porosity. We use Gaussian copulas to capture this covariance between properties. For each HSU, we specify a covariance in a standard multivariate normal distribution (zero mean and unit variance). The off-diagonal terms in the covariance matrix are used to express the degree of correlation between properties. A large number of random samples are generated from the multivariate normal distribution with covariance according to the covariance matrix. Based on the cumulative distribution function for the normal distribution, the quantile of each sample is computed. This normalises all values in the random sample in the [0,1] interval. The inverse cumulative distribution function of the marginal beta distribution describing the property is subsequently used to generate the value.

We specify a single level of correlation between hydraulic conductivity, specific yield and thickness for each HSU. The correlation is specified to be strong for the Bedrock HSU (0.75) as fracture density will have a great control on the properties. The correlation is specified to be weaker for Colluvium and Palaeovalley HSUs (0.5) as lithology is believed not to exert as large an influence. The resulted correlations are demonstrated in Figure 6. Salinity is currently not specified to be correlated with lithology.

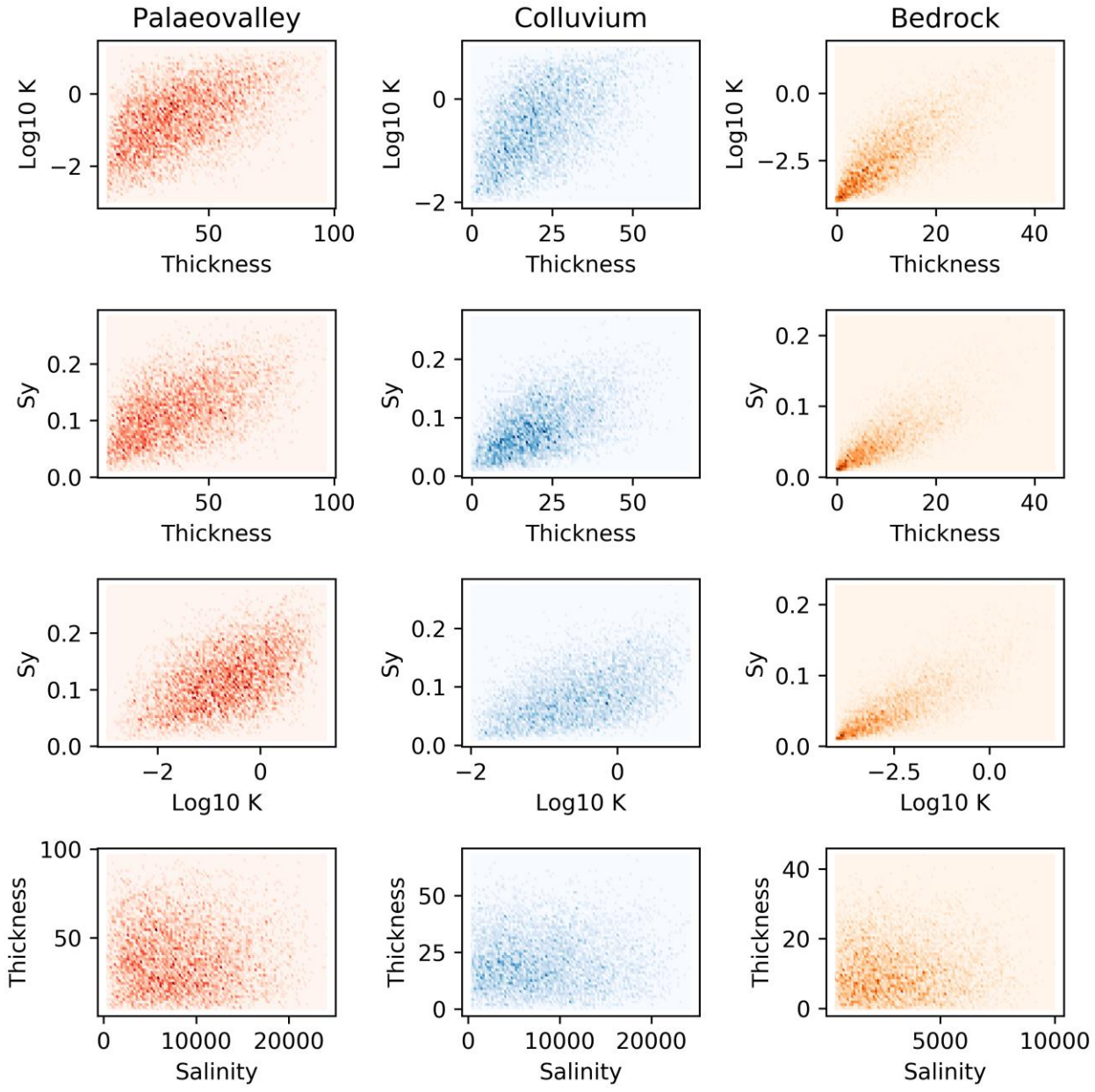


Figure 6. Correlation between different hydrogeological variables (Log10 hydraulic conductivity (Log10K), specific yield (Sy), thickness and salinity, for each of the three hydrostratigraphic units (HSU): Palaeovalley, Colluvium, Bedrock.

5.1.3 MAXIMUM PUMPING RATE *QMAX*

The maximum pumping rate is defined as the pumping rate that will cause a drawdown of $\alpha = 0.05$ of the saturated thickness at 1 m from the bore. This is computed through a straight forward application of the Theis (1935) equation:

$$Q_{max} = \frac{4\pi\alpha d^2 \mathrm{K}}{\mathrm{W}\left(\frac{S}{4Kdt}\right)} \qquad \text{Eq. 5}$$

The exponential integral (*W ()*) is approximated using equation 4 from Srivastava & Guzman-Guzman (1998). The maximum pumping rate is calculated for an alpha value of 0.05 (drawdown 5% of saturated thickness at 1 m from the pumping well) and for a pumping period of 10 years (Figure 7).

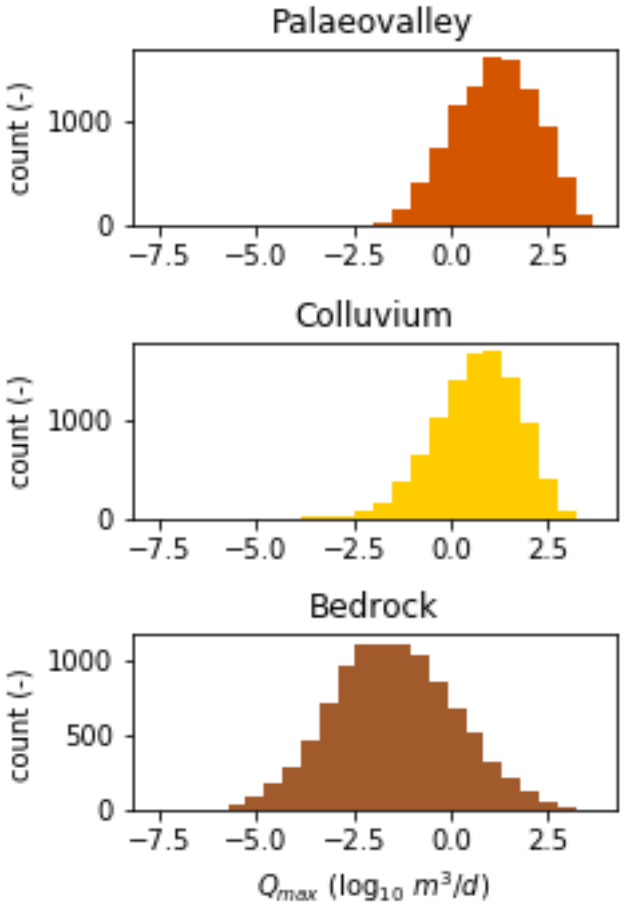


Figure 7. Histograms of maximum pumping rates (Q_{max}) for the three hydrogeological units.

5.1.4 AQUIFER VOLUME

The volume of water available in the aquifer (V) can be written as:

$$V = S\,A\,d \qquad \text{Eq. 6}$$

(note *A* is the area of the grid cell, which is 500 m x 500 m). The distributions of the three hydrogeological units are presented in Figure 8 based on 10,000 samples. Similar to the Q_{max} distribution, statistically the palaeovalley hosts more water than colluvium and bedrock, although not much more than colluvium.

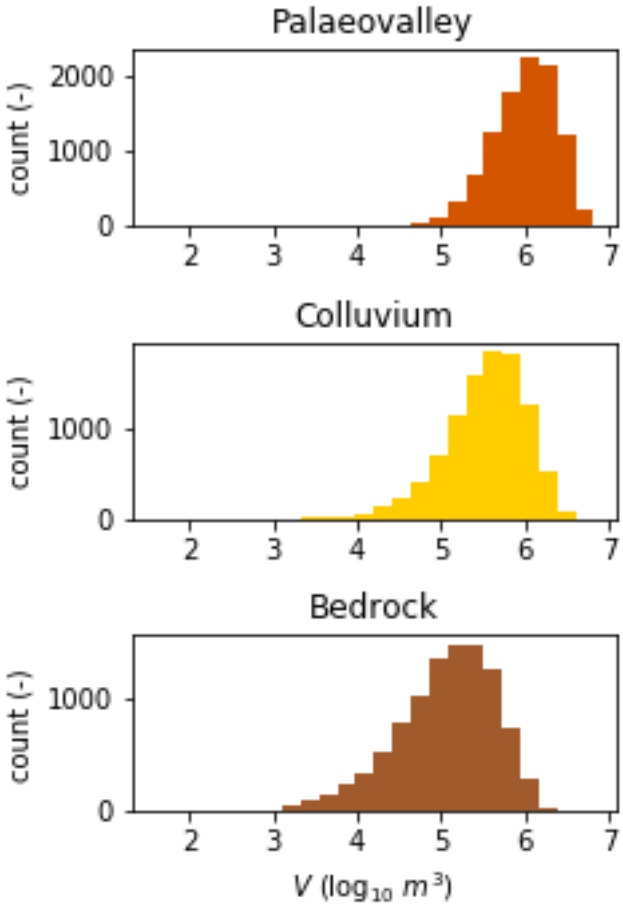


Figure 8. Histograms of cell-based aquifer volumes for the three hydrogeological units.

5.1.5 PROBABILITY OF Q>30 m^3/d FOR 10 YEARS WITH SALINITY < 5000 mg/L

The probability of developing a sustainable bore, i.e. sustaining a pumping rate of 30 m^3/d for 10 years with a salinity of less than 5000 mg/L, is calculated as the fraction of samples in the ensemble that simultaneous meet these constraints (Figure 9). By examining the effects of thresholds on the probability to drill a sustainable well, salinity is considered the most limiting factor (see Appendix B), while the time horizon does not exert much impact on the probability. The overall probabilities for the palaeovalley, colluvium and bedrock are 0.11, 0.07 and 0.03, respectively.

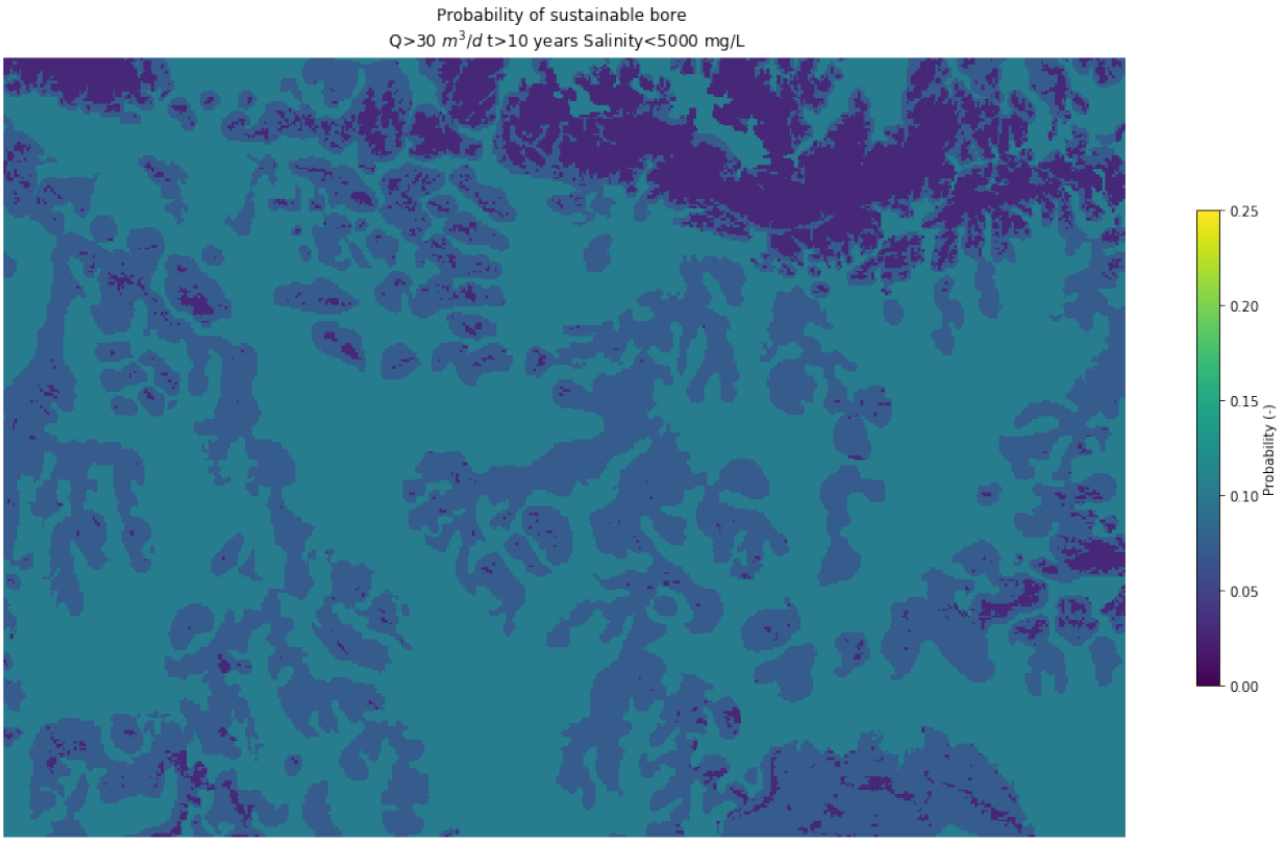


Figure 9. Probability of finding a groundwater bore that meets the defined criteria (note that location of this area is shown in Figure 1).

5.2 Single layer with equivalent properties

5.2.1 INTRODUCTION

In the previous conceptualisation, every hydrogeological unit is considered homogeneous within itself. In this section, the spatial variability is introduced.

A regional scale airborne electromagnetic (AEM) survey was flown over the APY Lands. The three-dimensional electrical conductivity field resulting from the inversion of that dataset was combined with bore hole data and the regional hydrogeological framework to derive a probabilistic surface of the top of saprolite (Figure 10). This deriving process is documented in another notebook: 'Probabilistic data assimilation for cover thickness mapping in the APY Lands' (Peeters and Visser 2019). The average depth of this surface is used to generate a water table map of the area (Figure 11), which is documented in notebook 'Bayesian Data Fusion for water table mapping in APY'.

The AEM data allows to map out in great spatial detail the distribution of palaeovalley systems. The conceptualisation is updated. Rather than three distinct hydrostratigraphic units, the system is considered as a single, continuous, unconfined aquifer system. The base of the system is chosen to be at 150 m depth. This value is chosen to be larger than the maximal depth to top of saprolite from all the ensembles. For each grid cell, an equivalent transmissivity and storativity is calculated as the thickness-weighted average of the properties assigned to the bedrock and palaeovalley HSUs. Salinity is also calculated as the thickness-weighted average, as salinity is specified to change with HSU.

Like for the previous conceptualisation, we start with establishing the prior probabilities. Unlike the three HSU conceptualisation, these are spatially variable, linked to the probabilistic estimate of top of saprolite (Figure 10).

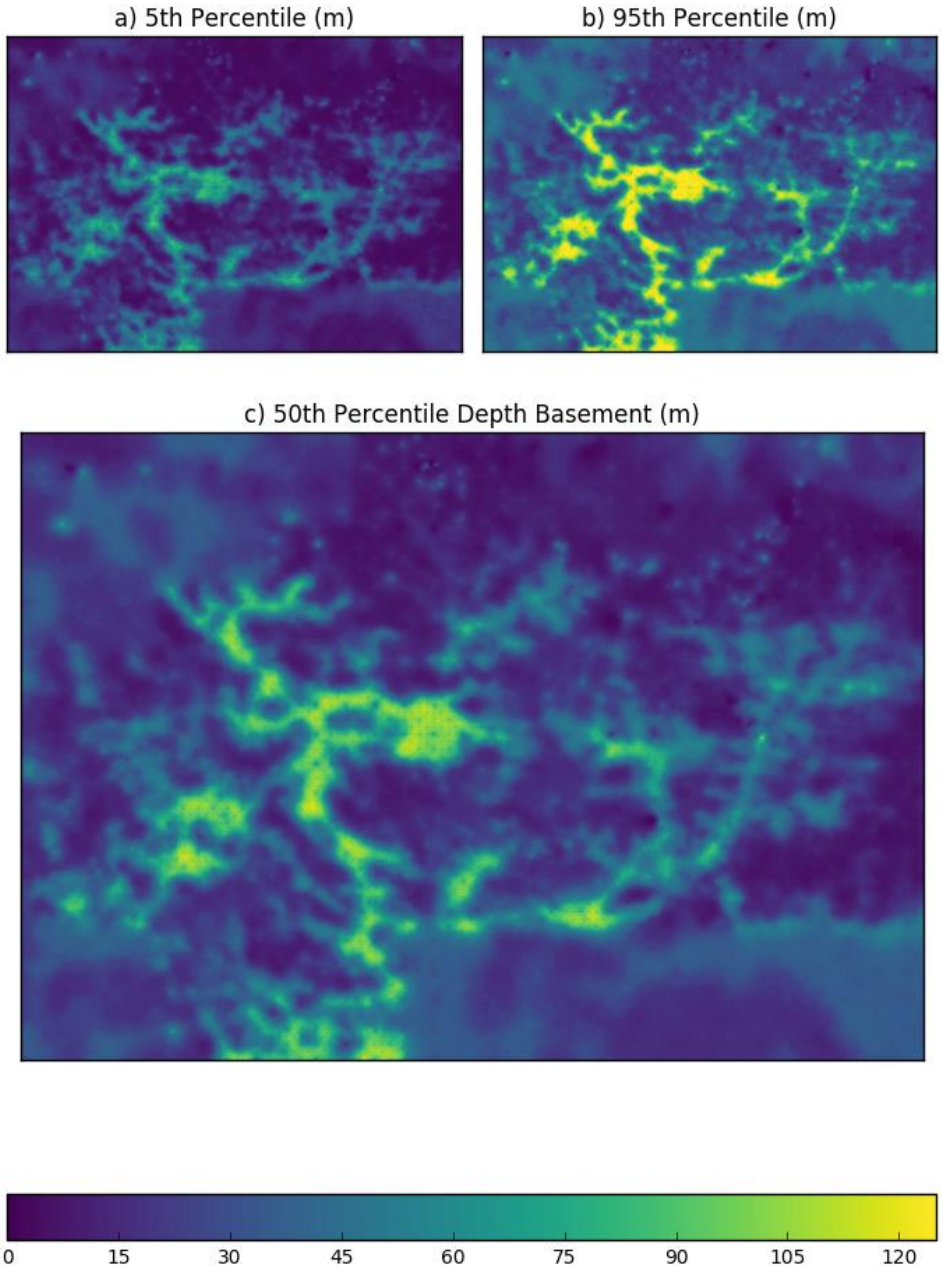


Figure 10. Depth of the saprolite upper boundary (note that location of this area is shown in Figure 1).

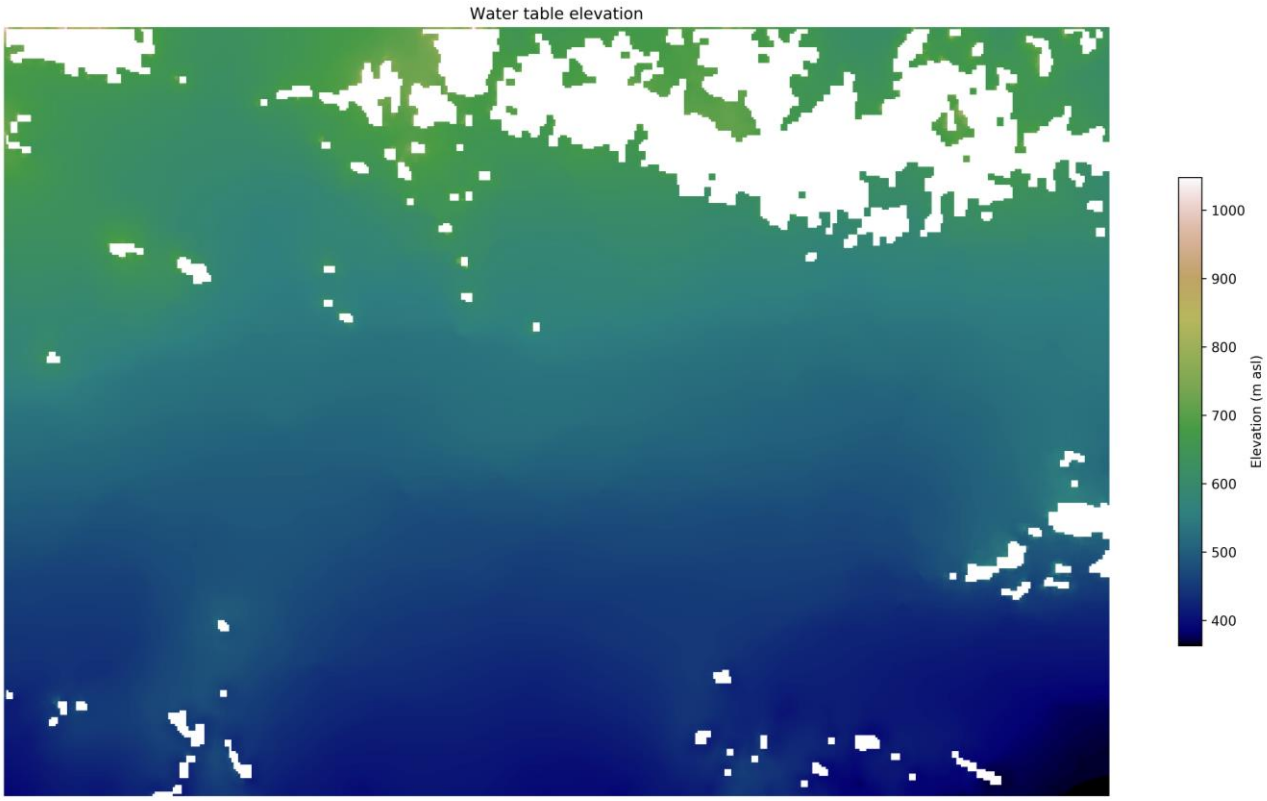


Figure 11. Water table elevation (note that location of this area is shown in Figure 1).

In this conceptualisation, the saturated thickness (d_{sat}) of the continuous aquifer with a nominal, maximum thickness of 150 m can be computed as:

$$d_{sat} = 150 - (z - h) \quad \text{Eq. 7}$$

with *z* topography and *h* the water table elevation. The minimal saturated thickness is 1 m. One representative aquifer thickness is shown in Figure 12 based on this equation.

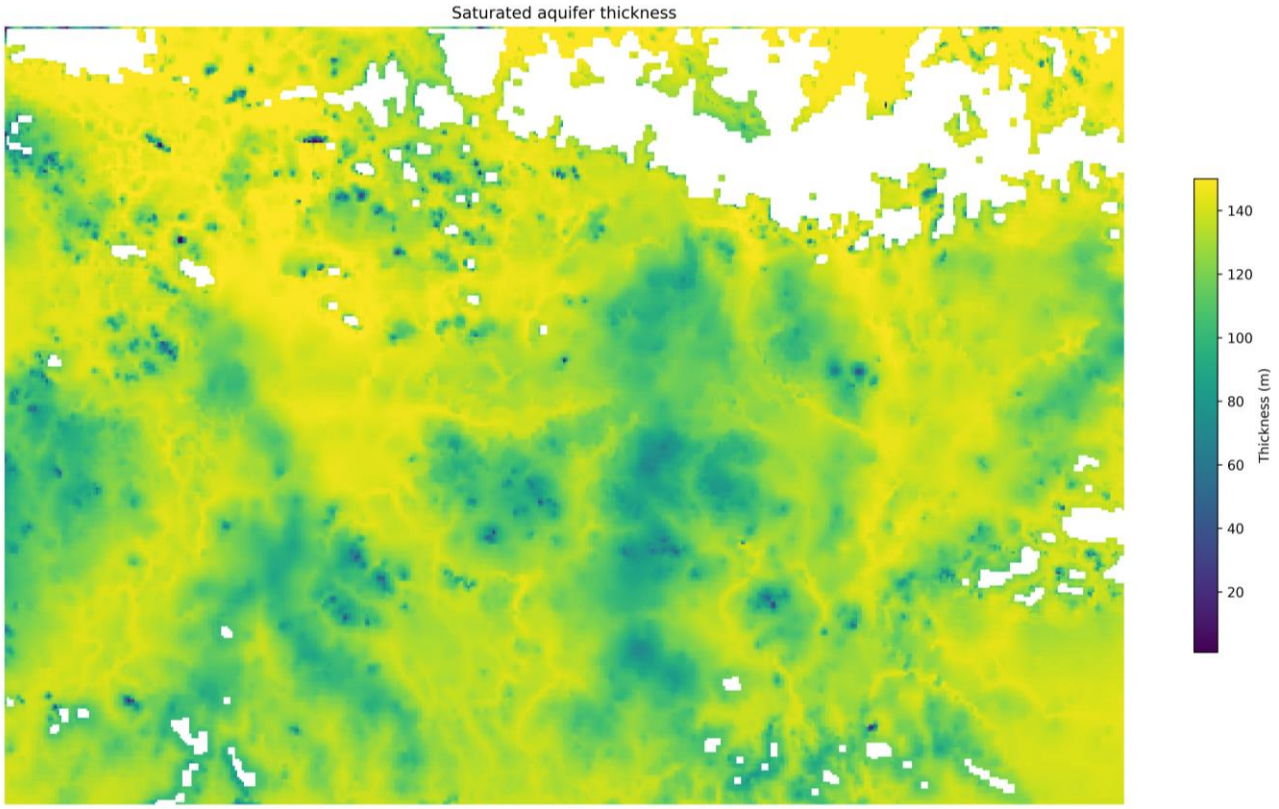


Figure 12. A representative saturated aquifer thickness map based on one realisation for the top boundary of the saprolite (note that location of this area is shown in Figure 1).

The equivalent properties for the aquifer are based on the fraction of bedrock in each grid cell. The fraction of bedrock in the saturated aquifer thickness can then be expressed as:

$$f_B = \min\left(1, \frac{150 - d_{sap}}{d_{sat}}\right). \qquad \text{Eq. 8}$$

The results for a selected realisation are presented in Figure 13.

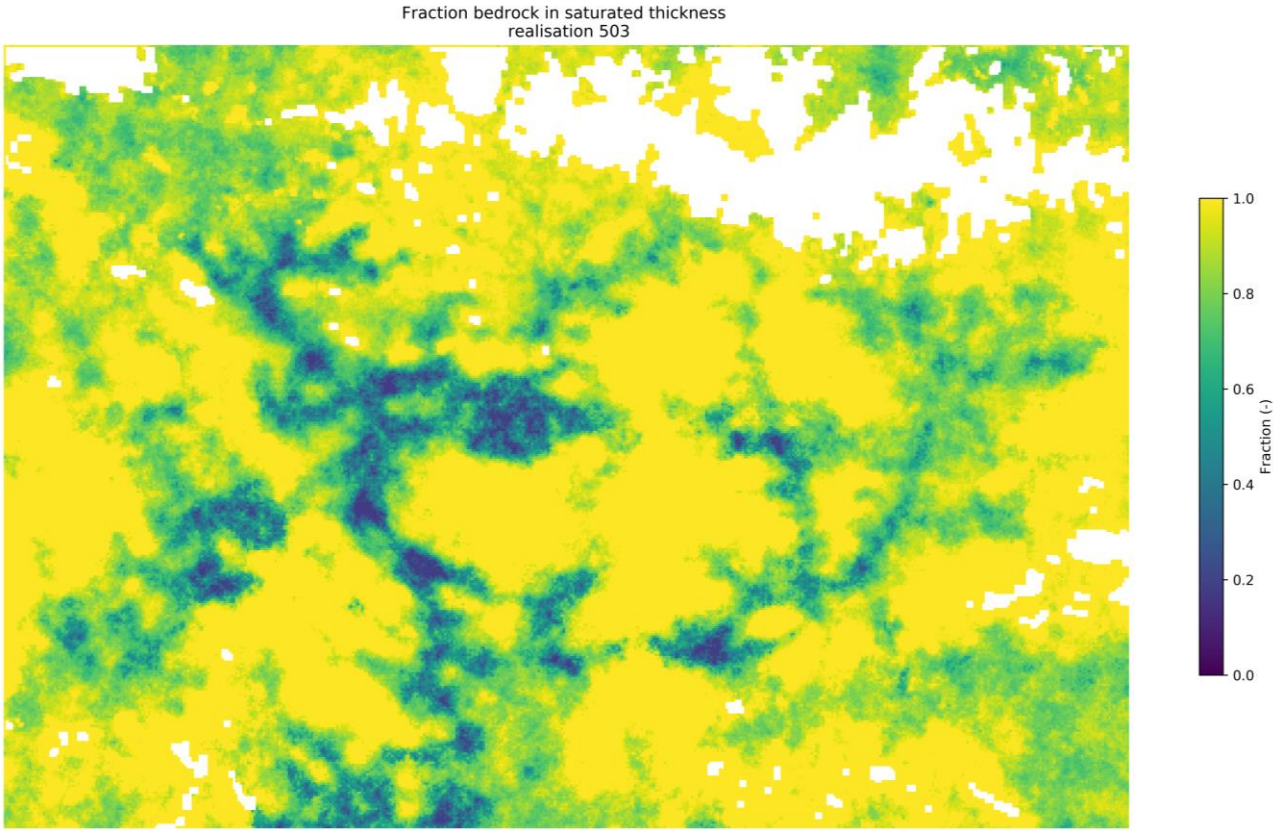


Figure 13. Fraction of bedrock in saturated thickness (note that location of this area is shown in Figure 1).

The equivalent properties (K, S and salinity) are computed as:

$$X_{eq} = f_b * X_b + (1\text{-}f_B) * X_p \qquad \text{Eq. 9}$$

In this equation, X_B, is a random sample of the property, from the prior probability distributions defined earlier, for the bedrock HSU. X_P is a random sample of the same property, but for the palaeovalley HSU. The calculation of maximum pumping rate and the aquifer volume is the same as for the three HSU conceptualisation. However, spatial variability is allowed for the current conceptualisation.

5.2.2 PROBABILITY OF Q>30 m^3/d FOR 10 YEARS WITH SALINITY <5000 mg/L

For the single aquifer conceptualisation, the probability of developing a sustainable bore, i.e. sustaining a pumping rate of 30 m^3/d for 10 years with a salinity of less than 5000 mg/L, is, like for the three HSU conceptualisation, calculated as the fraction of samples in the ensemble that simultaneously meet these constraints (Figure 14).

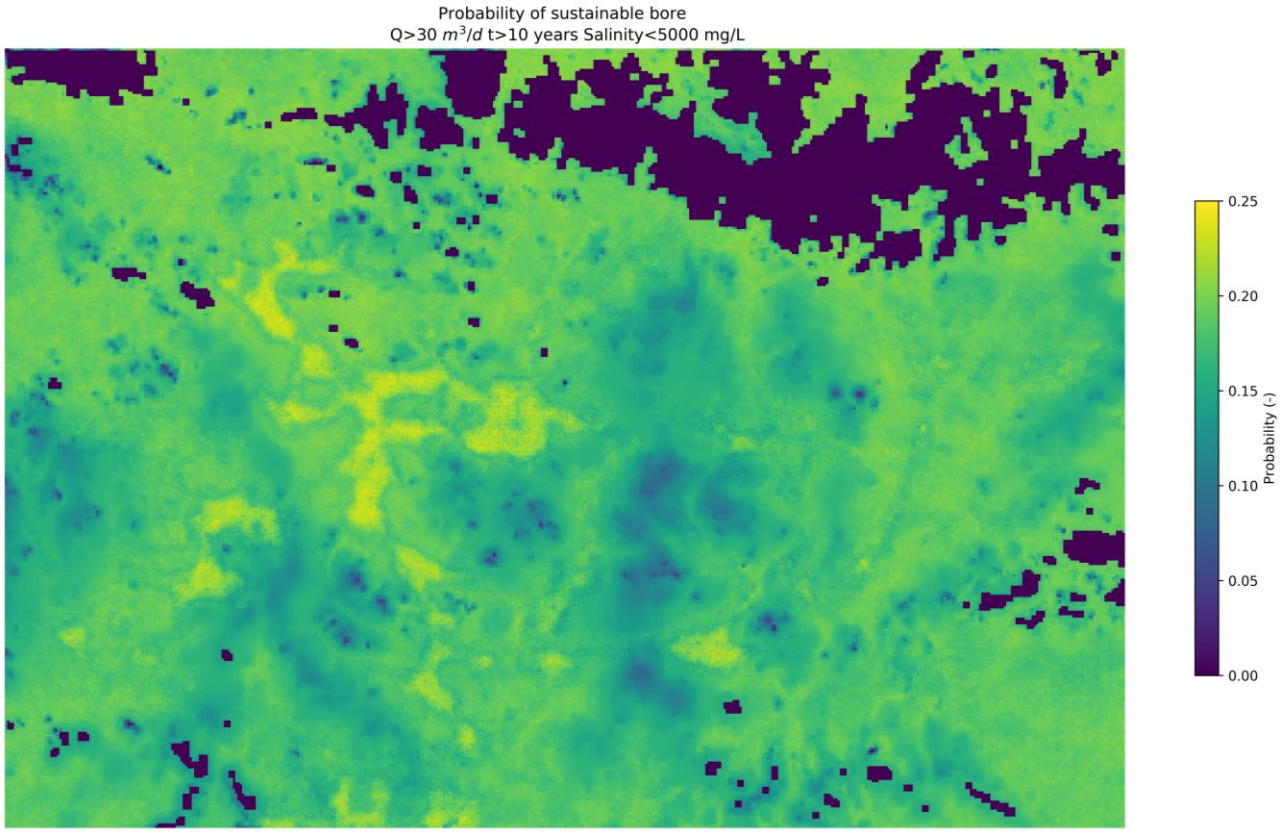


Figure 14. Probability of finding a groundwater bore that meets the defined criteria (note that location of this area is shown in Figure 1).

This now allows to compute the change in probability due to changing the conceptualisation (Figure 15):

$$\Delta P = P_{SingleAquifer} - P_{ThreeHSU} \quad \text{Eq. 10}$$

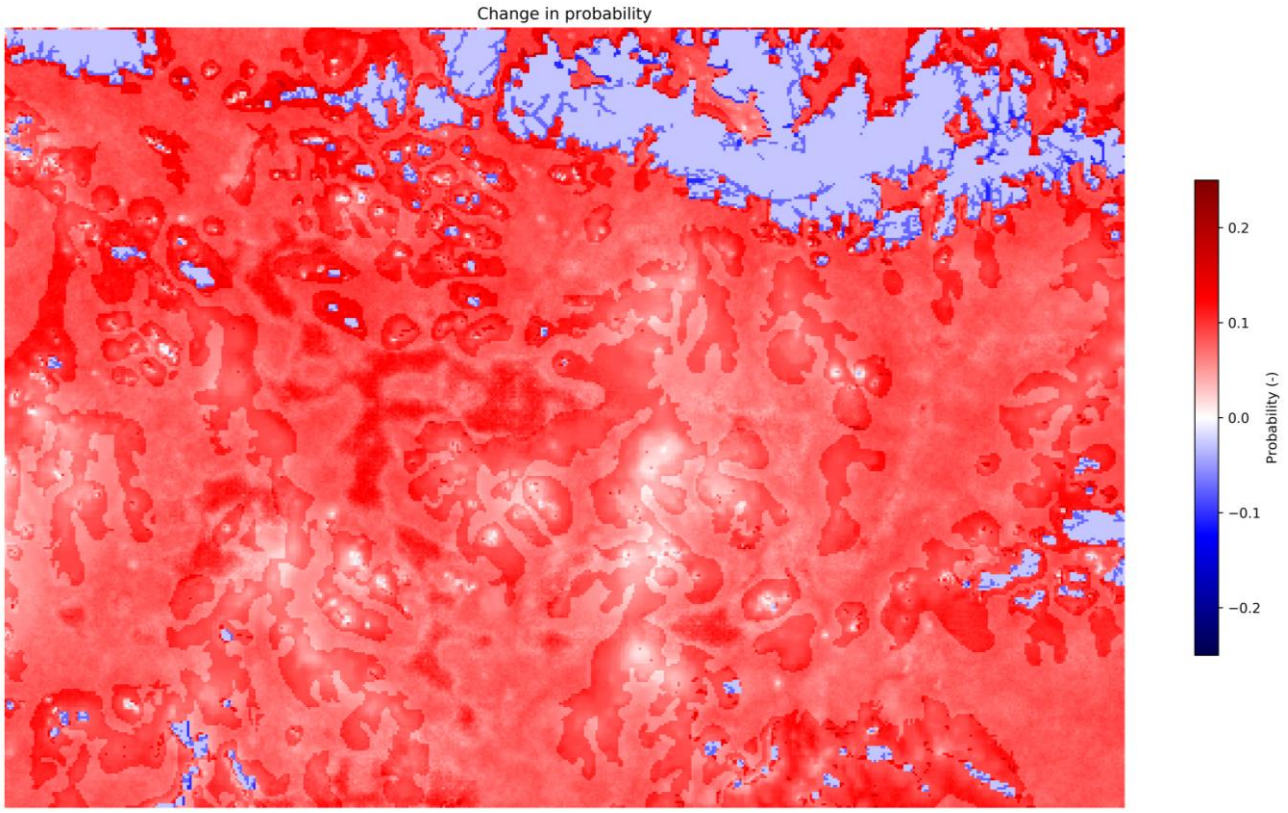


Figure 15. Difference of probability between the single aquifer conceptualisation and the three-hydrostratigrap hic unit (HSU) conceptualisation (note that location of this area is shown in Figure 1).

The figure above shows that:

1. Probability decreases markedly in areas of large outcrops as no water table is computed for these areas
2. Probability increases for the bedrock HSU outside large outcrops. This is due to the increase in saturated thickness
3. Probability decreases slightly or remains the same for colluvial HSU.
4. Probability increases markedly in the deeper parts of palaeovalleys.

5.3 Two aquifers

5.3.1 INTRODUCTION

Detailed analysis of the borehole logs and core retrieved from borehole DH1 in the Lindsay East palaeochannel are described in Krapf et al (2019) and reproduced below (Figure 16).

The first 25 m is characterised as fluvio-aeolian sandplain deposits with pedogenic calcrete that are spatially extensive. Below that unit, from 25 m to 63 m, the sediments consist mostly of sand and they are described as fluvial deposits or palaeovalley fill. The next 20 m are fluvial-lacustrine brackish to estuarine/marginal marine deposits. The borehole did not reach bedrock, the final 10 m are described as older fluvial deposits of palaeovalley fill.

Fig 12e in Krapf et al (2019) (reproduced below Figure 17) is the basis for the third conceptualisation; a continuous superficial aquifer underlain either by weathered bedrock or palaeovalley sediments. There is no extensive low permeable unit separating the superficial unit from the deeper water bearing units. The entire system is therefore considered to behave as an unconfined aquifer system. The deeper marine deposits potentially form an aquitard, separating a top palaeovalley aquifer from a deep palaeovalley aquifer in contact with saprolite and bedrock. As the deposits are only identified in a single borehole, the extent and thickness of this clayey unit is difficult to map. At this stage of the analysis, this level of detail is not integrated in the probabilistic mapping.

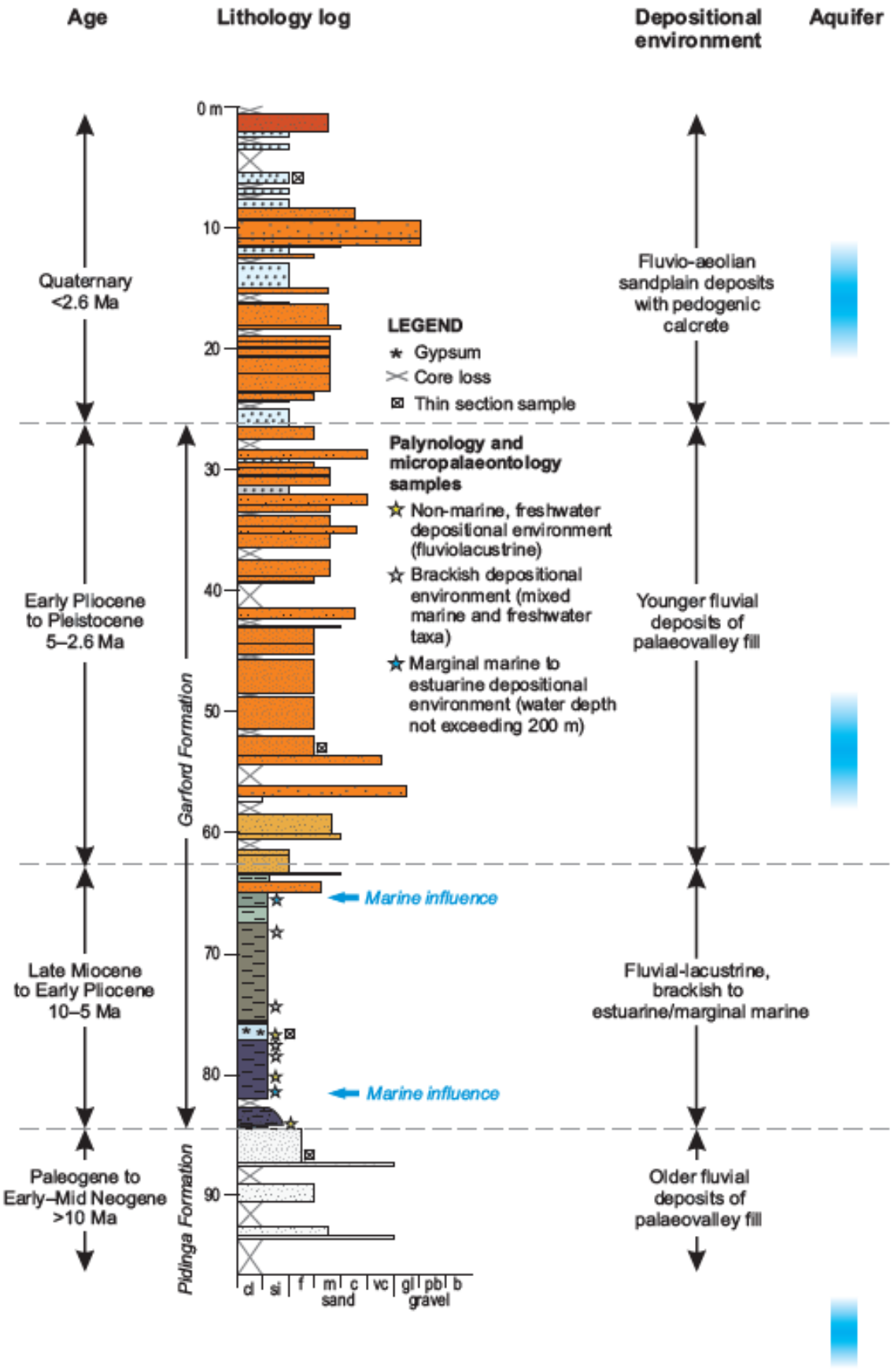


Figure 16. Borehole log DH1 (after Krapf et al. 2019).

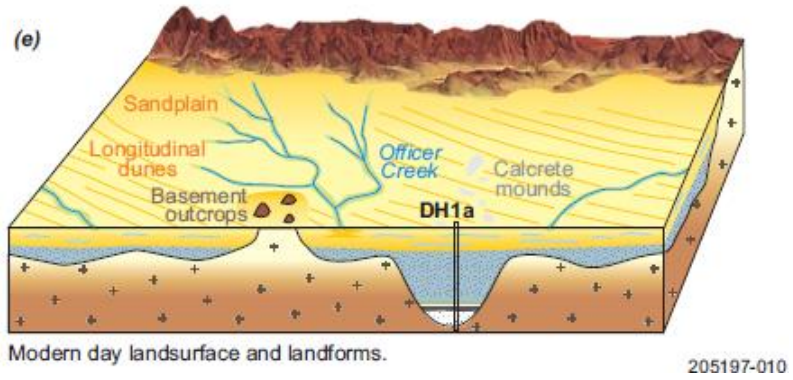


Figure 17. Krapf conceptual model (after Krapf et al. 2019).

The probability of drilling a sustainable bore is calculated by combining the probability for the superficial aquifer and for the deeper aquifer system. The deeper aquifer system is conceptualised similarly as for the single aquifer system (continuous aquifer), but with an updated saturated thickness.

5.3.2 SUPERFICIAL AQUIFER

The saturated thickness of the superficial aquifer is at most 30 m, based on the interpretation of Krapf et al. (2019) and borehole analysis. Before making the probability calculation spatially explicit, we calculate the probability of drilling a sustainable bore as a function of saturated thickness for the three HSUs. This will allow to evaluate the minimal saturated thickness required to have a probability higher than 0.01.

A saturated thickness of at least 10 m is required to have a probability of drilling a sustainable bore that is greater than 0.01 (Figure 18). It also shows that the prior probability distributions for the colluvial HSU result for higher probabilities at increased saturated thickness. This is a reflection of the narrative used in defining the prior probabilities, in which the palaeovalley HSU is considered to have shallow clay lenses and calcretes with lower permeability.

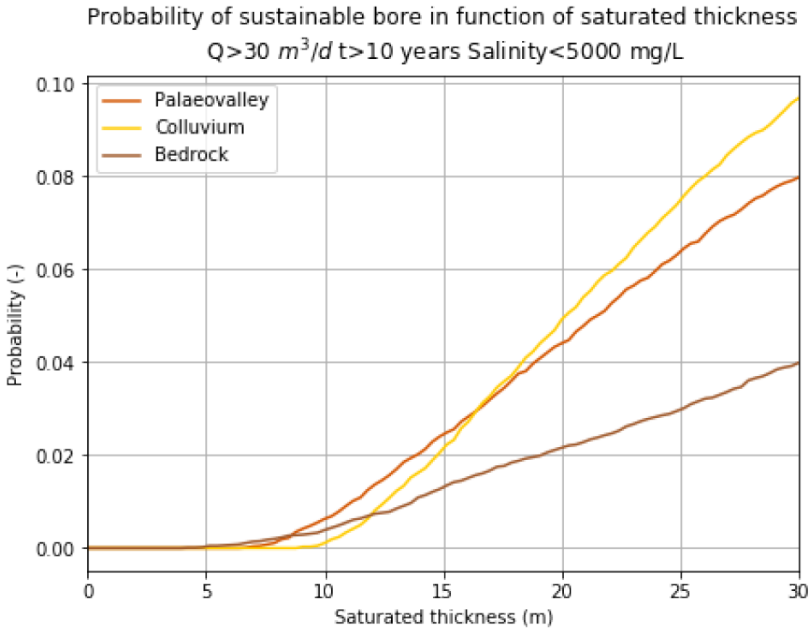


Figure 18. Probability of sustainable bore in function of saturated thickness.

Rather than redoing the probability calculations, we consider the superficial aquifer to have the same distribution of hydraulic parameters and salinity as previously assigned to the colluvial HSU. The probability of drilling a sustainable bore in the surficial aquifer can then be computed using the relationship between saturated thickness and probability for the colluvial HSU.

In the north east of the study area, the mapped water tables are shallower close to the edge of the large outcrops (Musgrave Ranges), which leads to greater saturated thickness and therefore increased probability (Figure 19). Do note however that the maximum probability to drill a sustainable bore that is shallower than 30 m is less than 10%.

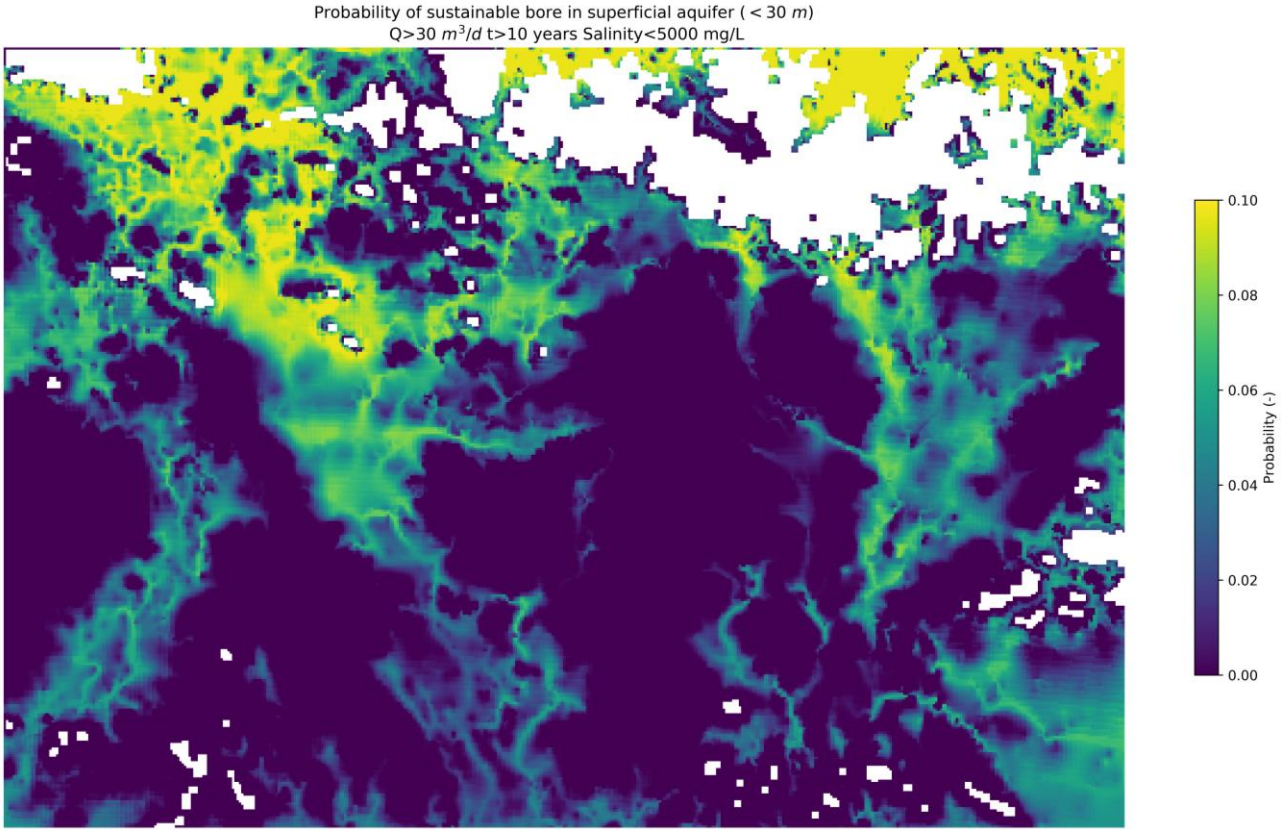


Figure 19. Probability of siting a sustainable bore in superficial aquifer (note that location of this area is shown in Figure 1).

5.3.3 DEEPER, CONFINED AQUIFER

The deeper confined aquifer is similar to the single aquifer with equivalent properties defined earlier. The main point of difference is a smaller saturated thickness, as the top 30 m is considered to be part of the surficial aquifer. As the top 30 m is no longer part of the aquifer, the saturated thickness and fraction of bedrock needs to be updated. The revised fraction of bedrock in the saturated zone is used to calculate equivalent properties again. The same equations are used as above, with specific yield replaced with specific storage. With all these aspects taken into account, the updated probability map is shown in Figure 20. The probability of drilling a sustainable bore is only higher than 0.20 where the palaeovalley sediments are thick, which are in isolated patches in the Lindsay East Palaeochannel and more extensive areas in the Lindsay West Palaeochannel.

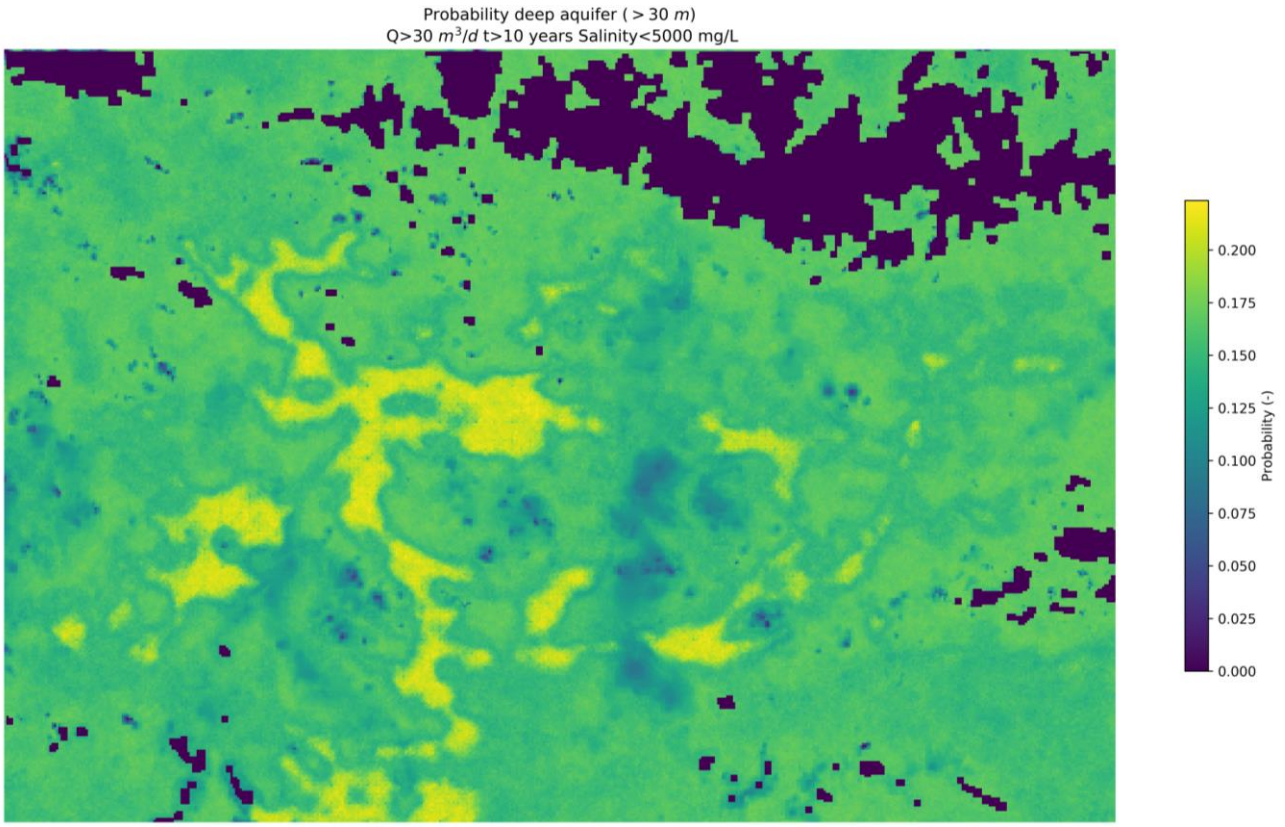


Figure 20. Probability of siting a sustainable bore in the deep and confined aquifer for the defined criteria (note that location of this area is shown in Figure 1).

5.3.4 COMBINED PROBABILITY

The combined probability, i.e. the probability of drilling a sustainable bore in either the surficial or the deep aquifer, can be calculated as:

$$P_c = 1 - (1-P_{surf})\,(1-P_{deep}). \qquad \text{Eq. 11}$$

The combined probability is shown in Figure 21.

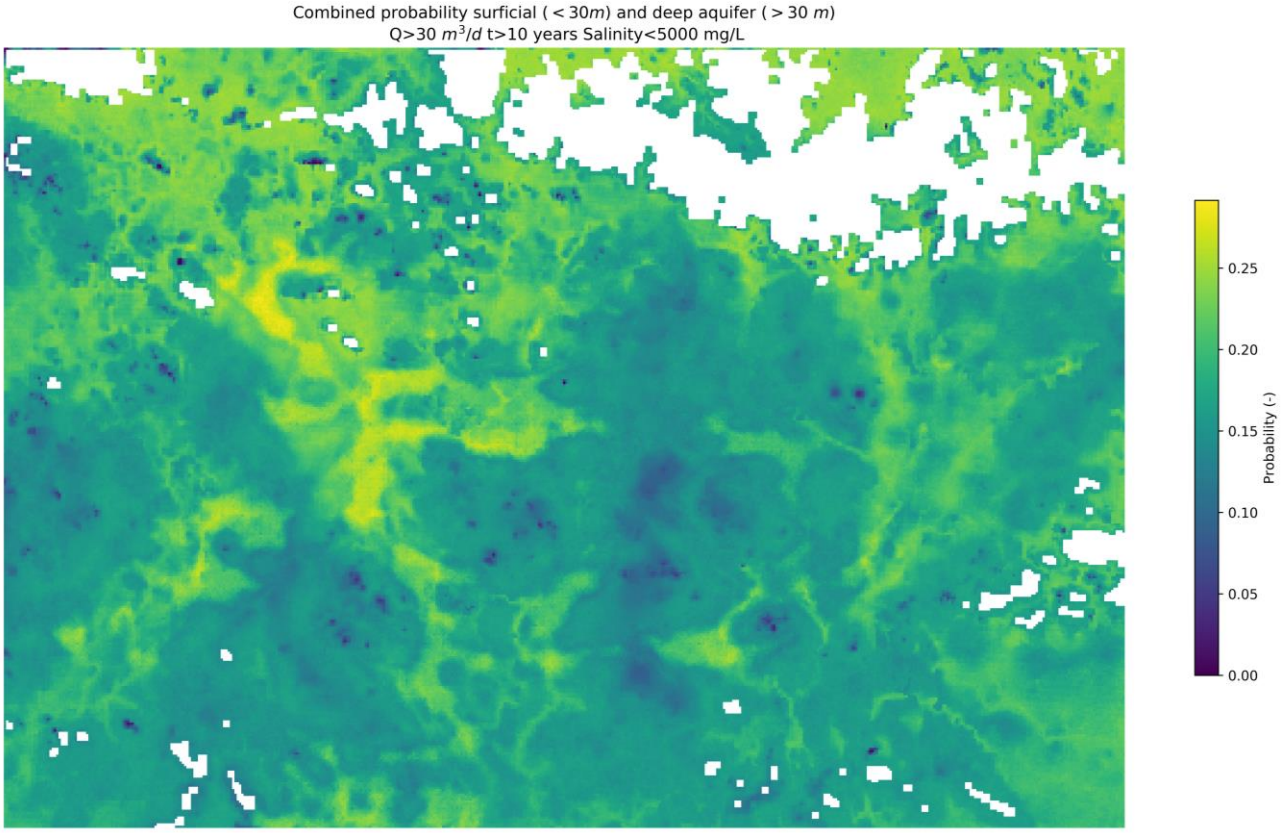


Figure 21. Combined probability for the two-aquifer conceptualisation for the defined criteria (note that location of this area is shown in Figure 1).

The difference between the combined probability and the probability using the single aquifer conceptualisation is shown below Figure 22. This map shows that the two aquifer conceptualisation results in increased probability in the palaeovalleys and a decrease at the edges of the palaeovalleys. In the single aquifer conceptualisation, the latter areas have a considerable fraction of palaeovalley sediment, increasing the equivalent properties. The saturated thickness in the surficial aquifer is however small, which results in an overall decrease in probability. In Lindsay East the greatest increase in probability occurs just outside the palaeovalley, where there is a combination of a relatively low probability shallow system and low probability fractured rock, which however yield moderate probability when combined. Drilling just outside palaeovalleys is locally interesting, provided bores are targeting fractured bedrock.

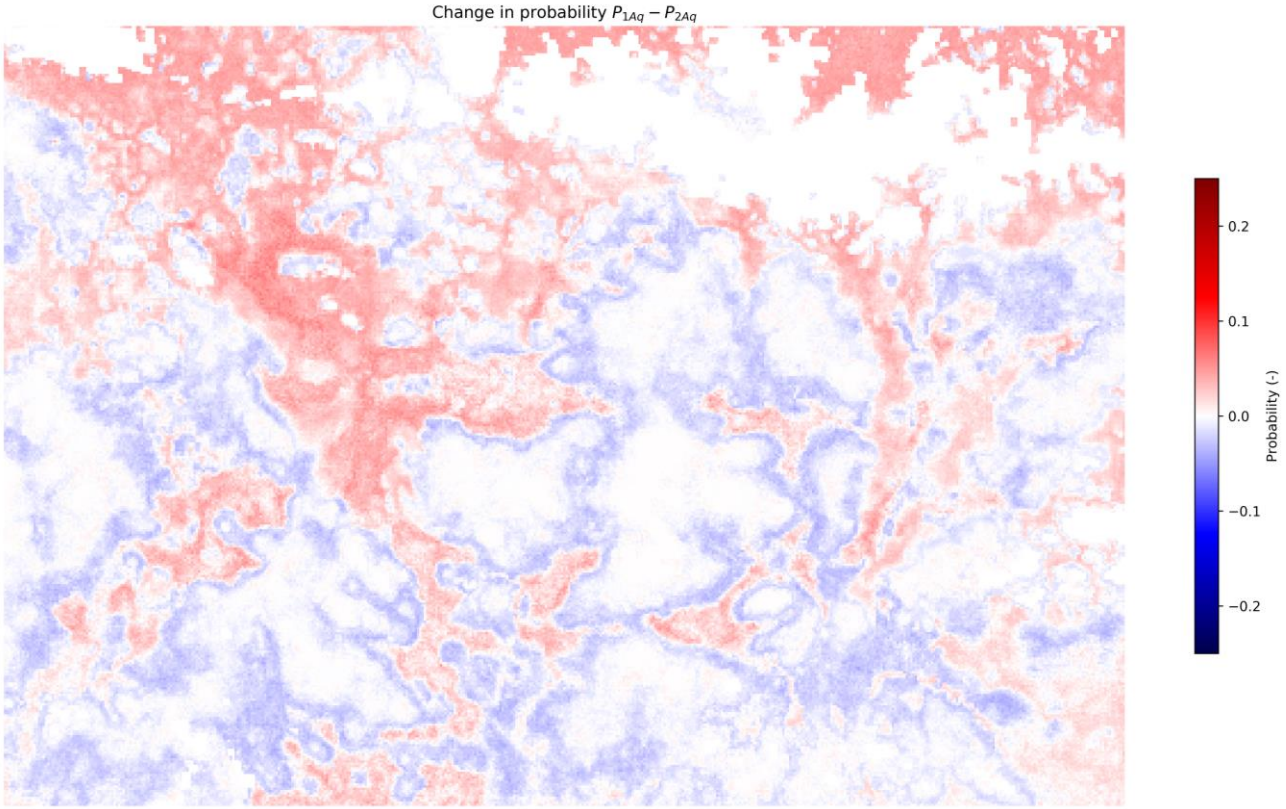


Figure 22. The difference between the combined probability (two aquifers) and the probability using the single aquifer conceptualisation (note that location of this area is shown in Figure 1).

6 Discussion

The probabilistic groundwater mapping presented in this study relies on a number of assumptions and choices. To systematically discuss the effect of these assumptions on probability estimates, we adopt the assumption hunting methodology outlined in Peeters (2017).

The methodology lists all assumptions and scores for each of the assumptions (high, medium or low), based on these four questions:

1. What is the likelihood that I would have made the same choice if I had more or different *data*? (metric D in Table 2)
2. What is the likelihood that I would have made the same choice if I had more *time and budget*? (metric R in Table 2)
3. What is the likelihood that I would have made the same choice if I had a better *model/software*? (metric T in Table 2)
4. What is the likelihood that the model *predictions* are very different if I change the assumption? (metric P in Table 2)

The scoring is undertaken by the authors and is a personal and therefore subjective interpretation. Table 2 provides and overview of the assumptions and their scoring. Each assumption and its scoring (importance for or impact on questions 1 to 4) is discussed in greater detail below.

Table 2: List of assumptions and their scoring according to importance of Data (D), Resources (R), Technical limitations (T), and impact on Predictions (P). Red colour corresponds to a high, amber to a medium and green to a low score.

NUMBER	ASSUMPTION	D	R	T	P
1	The exploration question is relevant for groundwater management				
2	Grid cells are independent (no spatial correlation)				
3	Theis-based analytic solution to determine Q				
4	Grid-cell-based water balance without recharge and discharge				
5	Beta distributions capture prior knowledge				

6.1 Exploration question is relevant

A narrow interpretation of this assumption is that the targets set for groundwater exploration (i.e. 30 m^3/d with salinity less than 5000 mg/L for 10 years) are relevant for groundwater management. A broader interpretation is whether the probability of developing a successful bore is relevant to metrics for groundwater management.

The data and technical attribute are scored low on this assumption, as this assumption does not depend on data availability or technical limitations. Resources are scored medium as more time and budget would have allowed to engage with water resource managers more to test and refine the groundwater exploration questions that are relevant and which targets are realistic.

The effect on predictions is scored high, as any change in the metrics or targets will yield very different results.

6.2 Grid cells are independent (no spatial correlation)

The current methodology relies on the assumption that it is sufficient to drill one bore per grid cell to characterise the entire grid cell and that each grid cell is independent. This lack of accounting for spatial correlation implies that information for a grid cell does not inform the probability distributions of its neighbours.

Establishing such spatial correlation, requires a large database of information and knowledge on the aquifer properties and dynamics. These are not available, hence the scoring 'high' on this attribute. Even when these data were available, the mathematics behind the methodology would require a complete overhaul, which is a technical challenge. The technical attribute is therefore scored 'high' as well. Both the data and technical limitations can in theory be overcome, but it would require a large amount of time and budget, so the resources attribute is also scored 'high'.

The effect on predictions is however scored medium. Accounting for spatial correlation would locally change probabilities, but at a regional scale the results will remain of the same order of magnitude.

6.3 Pumping rate (Q) based on Theis analytic solution

The Theis analytic solution for time varying drawdown is straight-forward to implement but is inflexible when it comes to represent more realistic boundary conditions, confinement status and heterogeneity.

The data attribute is scored low. More data would allow to better characterise the flow system, but it would not be able to be taken into account due to the limitations of the Theis solution. A numerical groundwater model can take this data into account to compute in more detail what pumping rate is required to cause no more than five percent drawdown at one meter from the bore. This capability is available, so the technical attribute is scored 'low'. Implementing such an optimisation scheme with numerical modelling would however require an enormous amount of resources, both in developing time and computational load. The resources attribute is therefore scored 'high'.

The effect on predictions is scored 'medium'. A numerical model would allow to locally refine the results, but at a regional scale the results will remain of the same order of magnitude.

6.4 Water balance without recharge and discharge

The available volume of groundwater is computed as if there is no connection between grid cells and that there is no recharge or discharge.

The data attribute is scored 'medium' as the data and understanding of regional recharge and discharge mechanisms and rates is still incomplete. Accounting for recharge and discharge mechanisms in a water balance approach is not technically challenging, nor is calculating aquifer volumes from geological models. The technical attribute is therefore scored 'low'. Developing stochastic geological models, updating the algorithms and including recharge and discharge mechanisms is however time-consuming. The resources attribute is therefore scored 'high'.

The prediction attribute is again scored 'low'. Not accounting for recharge and discharge and neighbouring cells will lead to an underestimate of the volume available.

6.5 Beta distributions capture expert knowledge

The prior probability distributions of the aquifer properties and geometry are assumed to adequately capture the current understanding of the aquifer system.

The data attribute is scored ‘high’ as more data will allow to empirically establish these distributions. A greater investment in establishing these prior distributions, for instance through formal expert elicitation, will increase robustness of these estimates. The resources attribute is therefore scored ‘medium’. Implementing different prior distributions is a trivial task, which is why the technical attribute is scored ‘low’.

The effect on predictions is scored ‘high’. The comparison with the bore yield data showed that the priors used are realistic, but it highlighted how difficult it is to validate these distributions. The probability results showed that even subtle differences in priors, like between the alluvial and colluvial units, can yield very different results.

7 Conclusion

We developed and applied an iterative framework for probabilistic groundwater mapping to the Musgrave Province of South Australia. The groundwater prospectivity question is formulated as a quantity of interest that explicitly captures the pumping rate, pumping duration and salinity constraints water supply needs to satisfy. The workflow calculates the probability of meeting these constraints based on: (i) different conceptualisations; (ii) governing groundwater flow equations; and (iii) informed priors of the relevant parameters.

The stochastic, grid-based approach is flexible which allows conceptualisations and associated parameter distributions to be updated as more information and data becomes available or the groundwater system understanding develops. The method is based on closed-form analytic solutions to the groundwater flow equations, which makes the method computationally efficient. The constraints for water supply can be easily changed, which makes this method amenable to interactive interface.

The application of the GKIS framework to the APY Lands delivered regional grids of top of saprolite and a water table elevations, which in turn informed regional scale prospectivity maps. The probability of drilling a bore that can sustain pumping at 30 m^3/d with salinity less than 5000 mg/L for 10 years is at most 30%. The most prospective areas are the palaeovalleys. Outside the palaeovalley system prospectivity is only elevated if the surficial aquifer has a shallow water table and is underlain by a fractured bedrock. The most limiting factor in groundwater exploration is the groundwater salinity. Understanding the spatial distribution of salinity is one of the key knowledge gaps in groundwater prospectivity.

References

Bode F, Ferré T, Zigelli N, Emmert M and Nowak W (2018) Reconnecting Stochastic Methods With Hydrogeological Applications: A Utilitarian Uncertainty Analysis and Risk Assessment Approach for the Design of Optimal Monitoring Networks. *Water Resources Research* 54: 2270–2287. doi:10.1002/2017wr020919

Bredehoeft JD, Papadopulos SS and Cooper HH (1982) Groundwater: the water-budget myth. Chapter 4 in: Scientific Basis of Water-Resource Management, Studies in Geophysics. National Academy press: Washington, DC.

Costar A, Love A, Krapf C, Keppel M, Munday T, Inverarity K, Wallis I, Soerensen C (2019) Hidden water in remote areas – using innovative exploration to uncover the past in the Anangu Pitjantjatjara Yankunytjatjara Lands. *Mesa Journal* 90(2), 23–35.

Gogoll M (2016) Challenges targeting potable groundwater in the APY Lands, SA. Presentation at SA-NRM Science Conference 2016, University of Adelaide 13-15 April 2016. Day 3 Session 8.

Gogoll M, Howles S, Dennis K (2017) The desert doesn't have to be dry: Potable groundwater and infrastructure in remote areas. OZWater 2017 Conference, Sydney.

Herman J and Usher W (2017) SALib: An open-source Python library for sensitivity analysis. *Journal of Open Source Software* 2(9). 97, doi:10.21105/joss.00097

Hintze JL, Nelson RD (1998) Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician* 52(2), 17.

Keppel M, Costar A, Love A, Krapf C (2019) G-FLOWS Stage 3 APY Lands Drilling Program, north-western South Australia. Technical Report Series 19/39, Goyder Institute for Water Research, Adelaide.

Krapf C, Costar A, Stoian L, Keppel M, Gordon G, Inverarity K, Love A, Munday T (2019) A sniff of the ocean in the Miocene at the foothills of the Musgrave Ranges – unravelling the evolution of the Lindsay East Palaeovalley. *MESA Journal* 90(2), 4–22.

Leaney FW, Taylor AR, Jolly ID and Davies PJ (2013) Facilitating Long Term Out-Back Water Solutions (G-FLOWS) Task 6: Groundwater recharge characteristics across key priority areas. Goyder Institute for Water Research Technical Report Series No. 13/6, Adelaide. Retrieved from http://www.goyderinstitute.org/_r178/media/system/attrib/file/169/13_6%20Leaney%20et%20al%20GFLOWS%20-%20HIGH%20QUALITY.pdf

Munday TJ (2013) The role of airborne geophysics in facilitating long-term outback water solutions to support mining in South Australia: 23rd Geophysical Conference, ASEG, Extended Abstracts, http://dx.doi.org/10.1071/ASEG2013ab189.

Munday T, Abdat T, Ley-Cooper Y and Gilfedder M (2013) Facilitating Long-term Outback Water Solutions (G-FLOWS Stage-1): Hydrogeological Framework. Goyder Institute for Water Research Technical Report Series No. 13/12.

Naghibi SA, Moghaddam DD, Kalantar B, Pradhan B and Kisi O (2017) A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. *Journal of Hydrology* 548: 471–483. doi:10.1016/j.jhydrol.2017.03.020

Peeters LJM (2017) Assumption Hunting in Groundwater Modeling: Find Assumptions Before They Find You. *Groundwater* 55: 665–669. doi:10.1111/gwat.12565

Peeters LJM, Visser G (2019) Probabilistic data assimilation for cover thickness mapping in the APY Lands. Jupyter Notebook. CSIRO.

Peeters LJM, Crosbie RS, Henderson BL, Holland K, Lewis S, Post DA and Schmidt RK (2018) The importance of being uncertain. *Water e-Journal* 3: 10. doi:10.21139/wej.2018.018

Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning. MIT Press.

Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M and Tarantola S (2010) Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications* 181: 259–270. doi:10.1016/j.cpc.2009.09.018

Scheidt C, Li L and Caers J (2018) Quantifying Uncertainty in Subsurface Systems. Wiley.

Soerensen C, Munday T, Krapf C, Love A, Costar A, Inverarity K, Gogoll M, Gilfedder M (2018) Uncovering the Musgrave Province in South Australia using airborne EM: ASEG Extended Abstracts 2018, 1-5. DOI: 10.1071/ASEG2018abT5_3H

Srivastava R, Guzman-Guzman A (1998) Practical Approximations of the Well Function. *GroundWater* 36(5): 844–848. https://doi.org/10.1111/j.1745-6584.1998.tb02203.x

Theis CV (1935) The relation between the lowering of the piezometric surface and the rate and duration of discharge of a well using groundwater storage. *Am. Geophys. Union Trans* 16: 519–524.

Varma S (2012) Hydrogeological review of the Musgrave Province, South Australia. Goyder Institute for Water Research Technical Report Series No. 12/8, Adelaide. http://www.goyderinstitute.org/_r183/media/system/attrib/file/174/12_8%20G-FLOWS%20Musgrave_hydrogeology%20Varma_Nov2012.pdf

Appendix A Summary of available bore information

The South Australian Government database WaterConnect was accessed in January 2019 and the data within the project area downloaded (Figure 23). Within the reporting area there are 532 database entries that correspond to bores with a driller's log available. This information, downloaded as zip file containing several .csv files, is stored and summarised in spreadsheets.

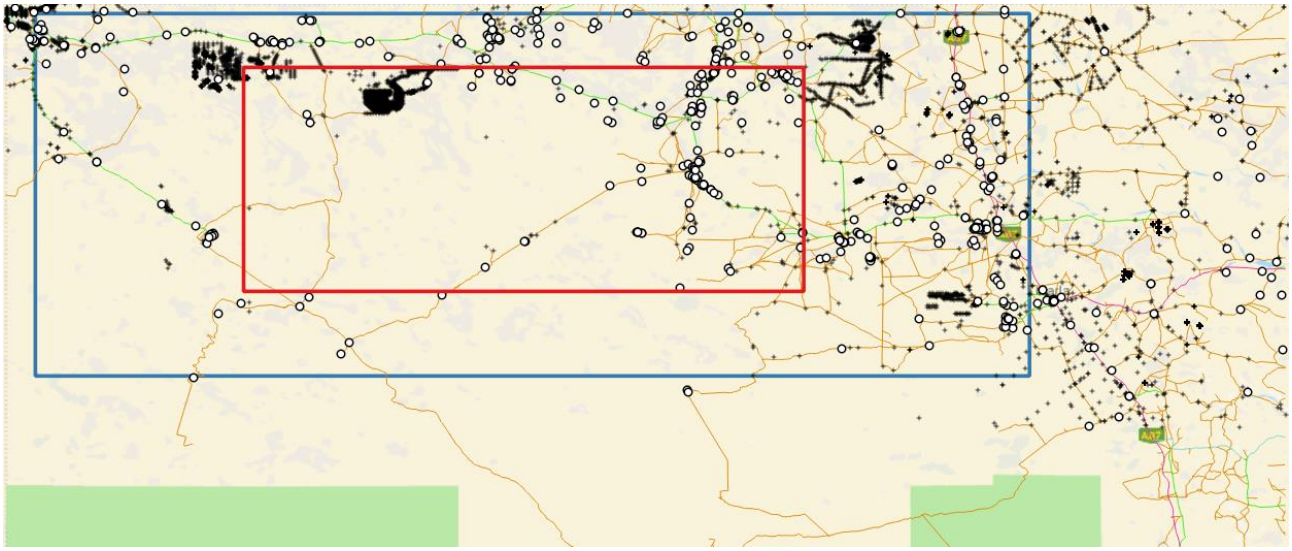


Figure 23. WaterConnect query showing: Bores (+), Bores with drillers log (o), reporting boundary (blue), GKIS area (red).

Table 3 describes this spreadsheet and the various data sources. All data in the database is linked through the unique identifier borehole identifier: DHNO.

The sheet 'Drillers Log' contains an interpretation of the drillers' log. It records the total depth of the bore, the depth of basement if reached, the lithology of the basement and the nature of the cover. The cover is classified as:

0. No or thin cover (< ~5 m)
1. Thick cover (>~15 m), indication of fluvial sediment transport
2. Thin cover (~5 m < cover > ~15 m), no or little indication of fluvial sediment transport

Basement is identified as the transition from predominantly unconsolidated or weathered lithologies to unweathered magmatic and metamorphic lithologies or unweathered, consolidated metasediments.

The drillers' log often lack the detail to enable an unambiguous identification of the cover-basement interface. It is therefore expected that the interpretation is variable and has considerable uncertainty.

Table 3: Description of spreadsheet GFLOWSDATA_LP.xlsx.

NAME	SOURCE	DESCRIPTION
Elevation	elevation.csv	Elevation as recorded in WaterConnect
Drillers Log	drillerslog.csv	Interpretation of drillers log
Coordinates	Raiber & Cui	coordinates in MGA 52 and elevation from 1s DEM
Salinity	salinity.csv	WaterConnect salinity
Salinity_Pivot	Salinity	pivot table of sheet Salinity, average TDS, EC and pH per DHNO
Well Summary	well_summary.csv	WaterConnect well_summary
Construction_Summary	construction_summary.csv	WaterConnect construction summary
WaterCut	watercut.csv	WaterConnect watercut
Waterlevel	waterlevel.csv	pivot table of WaterConnect waterlevel, average swl per DHNO
SA_TMI_RTP_1VD	\\osm-15-cdc\OSM_CBR_MR_GFLOWS3_work\GIS\Data_Raster\Magnetics\SA_TMI_RTP_1VD.tif	1st vertical derivative of Total Magnetic Intensity per DHNO
Thickness_50ohm_AEM	\\osm-15-cdc\OSM_CBR_MR_GFLOWS3_work\AEM\WB_Grids_tiffs_delivery_05102017\For_Nicole_PepinNini_24112017\50ohmm_cutoff_cover_thickness	Thickness of regolith per DHNO, based on merged AEM contour 50 Ohm
SkyTEM_smart	\\osm-15-cdc\OSM_CBR_MR_GFLOWS3_work\AEM\WB_Grids_tiffs_delivery_05102017\For_Nicole_PepinNini_24112017\Smart_interp_SkyTEM_cover_thickness	Thickness of regolith per DHNO, based on smart interpolation of SkyTEM AEM
AEM	\\osm-15-cdc\OSM_CBR_MR_GFLOWS3_work\AEM\WB_Grids_tiffs_delivery_05102017\For_Nicole_PepinNini_24112017\Conductivity_depth_intervals_merged_SkyTEM_TEMPEST	AEM conductivity per depth interval, per DHNO
EDA		combination of all sheets, using VLOOKUP

In the database, there are 532 bores with a drillers' log amenable for interpretation.

From those 532 bores, 323 reached basement. In 238 bores of these 320, unweathered basement was identified, in 84 only weathered basement was identified.

The cover sediments were classified as no or thin cover in 172 bores, thin cover (~15 m) or not transported in 66 bores and thick cover (~15 m) or transported in 294 bores.

Basement lithology was grouped in magmatic (224), metamorphic (30) and sedimentary (34). In 32 bores that reach basement, basement lithology could not be classified.

All bores have a TDS measurement, 514 have an entry for standing water level and 469 have bore yield recorded.

There are 283 bores in the AEM survey, the majority in the SkyTEM survey (272 bores) and only 11 bores in the TEMPEST survey.

A.1 Objective

The sensitivity analysis and value of information research indicated that the most influential factors in the exploration for sustainable groundwater resources in the APY lands are salinity, hydraulic conductivity, storage and aquifer thickness. The spatial density of observations of these factors is very low. This means we cannot directly interpolate the available data to generate grids for the Groundwater Knowledge Integration System.

The goal of this exploratory data analysis is to find proxy data that correlate to the factors we are interested in so that we can use them as covariates in spatial interpolation.

The factors we examine are:

1. Depth to basement / regolith thickness
   Proxy for base of sedimentary and weathered zone aquifer
2. Water level
   In combination with base of aquifer yields aquifer thickness
3. Total Dissolved Solids
   Salinity
4. Bore Yield
   Proxy for hydraulic conductivity and storage

The covariates we consider are:

1. Cover type from drillers log
2. Basement depth from drillers log
3. Basement depth from AEM
4. 1st vertical derivative of Total Magnetic Intensity

A.2 Methods

For this dataset, we chose to rely on violin plots, a graphical exploratory data analysis technique (Hintze & Nelson, 1998), due to the large variability and noise in the dataset.

Violin plots can be thought of as a combination of a histogram and a boxplot Figure 24. A histogram shows the detail of a distribution, but it is not straightforward to compare different groups. A boxplot is better suited to compare groups, but the distribution is only represented by a select number of percentiles (0^{th}, 5^{th}, 25^{th}, 50^{th}, 75^{th}, 95^{th}, 100^{th}). A violin plot combines a boxplot with a kernel density estimate of the distribution. This allows to preserve the detail in the distribution and compare different groups.

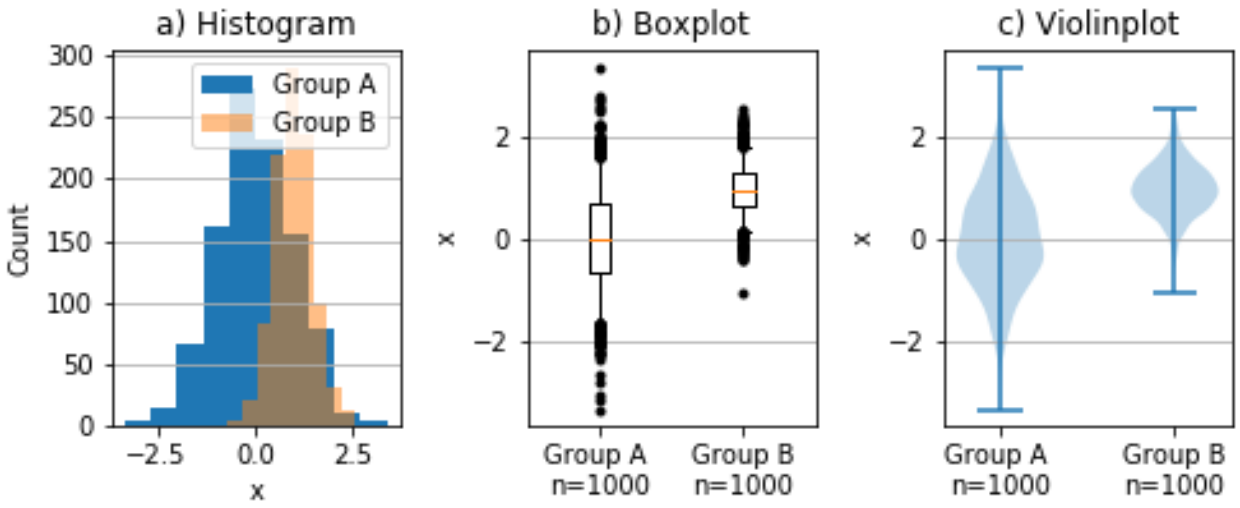


Figure 24. Example of (a) histogram, (b) boxplot, and (c) violinplot, to visualise a dataset with two groups with a different distribution. Group A consists of 1.000 random samples from a normal distribution with mean 0 and standard deviation 1, while Group B has 1.000 random samples from a normal distribution with mean 1 and standard deviation 0.5.

A.3 Results

A.3.1 DEPTH TO BASEMENT

There are 323 bores that reached basement. The distribution of depths is shown in the histogram in Figure 25a. The depth to basement is very skewed, with more than 120 bores where basement is reached within 5 m. The maximum depth to basement is 105 m. Figure 25b shows the histogram of the minimum depth to basement. Minimum depth to basement is the depth to basement if the basement is reached and the total bore depth otherwise.

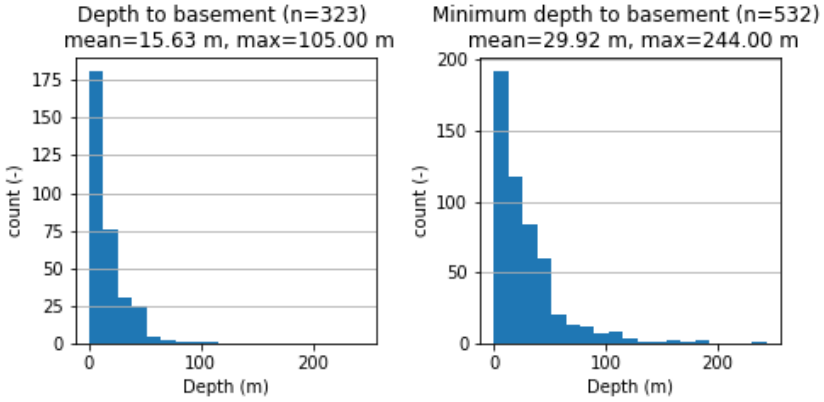


Figure 25. Histogram of (a) depth to basement from bore data and (b) minimum depth to basement.

The interpretation of drillers' logs in cover type is evaluated against the depth to basement in Figure 26. The large outlier in the 'No Cover' group is due to an intense weathering profile on top of sedimentary rocks. The range of thicknesses in the 'Thin Cover' group reflects a number of bores with a deep weathering profile, but no clear indication of transported sediment. In the 'Thick Cover' group, depth to basement is also skewed, with the median close to 20 m.

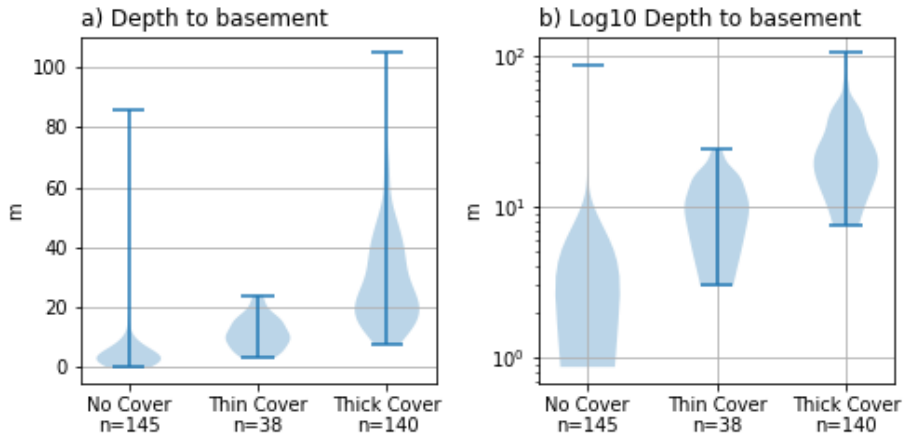


Figure 26. Violinplot of depth to basement (a) and log10 depth to basement (b), grouped by cover type.

The depth to basement is also inferred from the airborne electromagnetic data from the TEMPEST and SkyTEM surveys (Soerensen et al., 2018). Figure 27a shows how the depth of the 50 ohm contour of electric resistivity from the merged TEMPEST and SkyTEM AEM data compares to the depth of basement from the bores. Figure 27b shows the depth to basement from a smart interpolation of the SkyTEM data against the depth to basement from the bores.

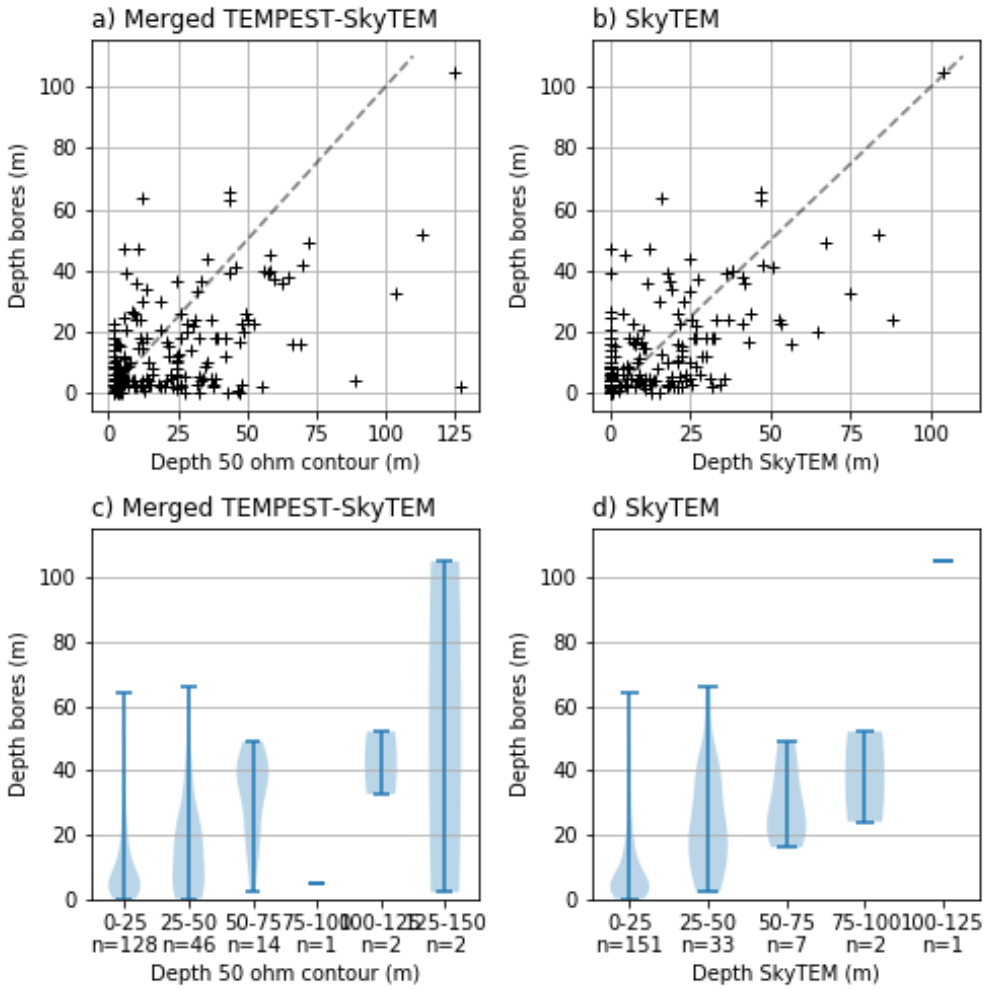


Figure 27. Depth to basement from bores compared to depth to basement from a) the 50 ohm contour from the merged TEMPEST-SkyTEM data and b) smart interpolation of the SkyTEM data. The same information is presented in violinplots for c) the merged TEMPEST-SkyTEM data and d) the smart interpolation of the SkyTEM data. The airborne electromagnetic derived basement depth is grouped in 25 m bins.

From these figures it is apparent that there is considerable scatter in the relationship between depth to basement from bore data and depth to basement inferred from the AEM data. The violinplots in Figure 27c and Figure 27d help in visualising the trends in the data. Both AEM-derived basement depths overestimate the depth to basement from borehole data interpretation. This can be attributed to the definition of the "basement interface" defined in drill chips/core which commonly tend to underestimate the true interface between weathered and unweathered bedrock. The interface pick often overlooks the presence of subtle weathering effects and is commonly defined where recognisable minerals/textures relating to hard rock are determined. The other factor requiring consideration is the sampling resolution of an AEM system versus drilling. The latter resolves a boundary on the scale of centimetres. The AEM, being a diffusive technology, defines an average depth as the lateral resolution of an AEM system is hundred plus metres in dimension. There are also large uncertainties in the positional accuracy of the borehole data.

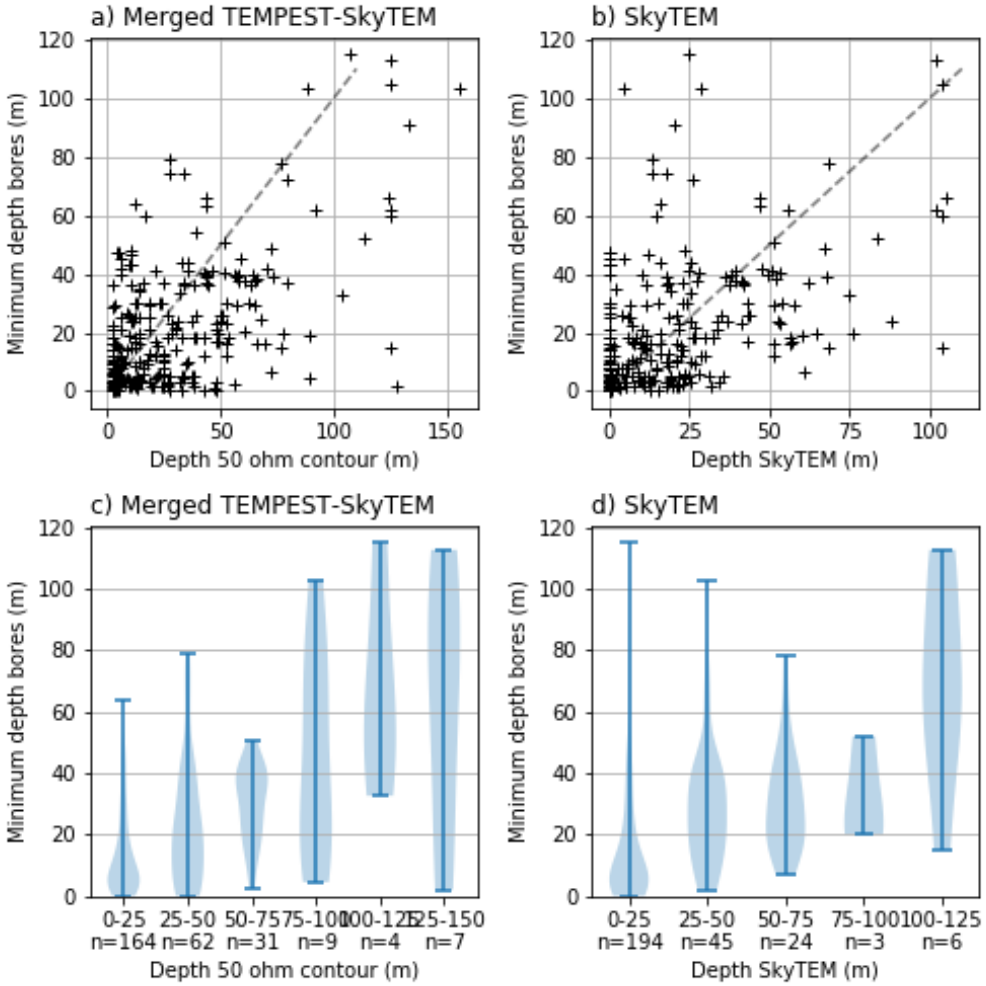


Figure 28. Minimum depth to basement from bores compared to depth to basement from a) the 50 ohm contour from the merged TEMPEST-SkyTEM data and b) smart interpolation of the SkyTEM data. The same information is presented in violinplots for c) the merged TEMPEST-SkyTEM data and d) the smart interpolation of the SkyTEM data. The airborne electromagnetic derived basement depth is grouped in 25 m bins.

Figure 28 compares the AEM derived depth to basement with the minimum depth to basement from the bore data. It highlights the sampling bias in areas where basement is deeper. The AEM derived values correspond better to the minimum depth estimates than the depth estimates, but especially the depth based on the 50 ohm contour overestimates the depth to basement. This is partly due to the smoothed nature of the AEM resistivity maps which hamper direct comparison with borehole data.

The exploratory data analysis shows that the depth to basement derived from the 50 ohm contour of the merged TEMPEST-SkyTEM is a good proxy for depth to basement.

A.3.2 TOTAL DISSOLVED SOLIDS

All the bores in the database have at least one value for total dissolved solids (TDS). Figure 29 shows the distribution of TDS.

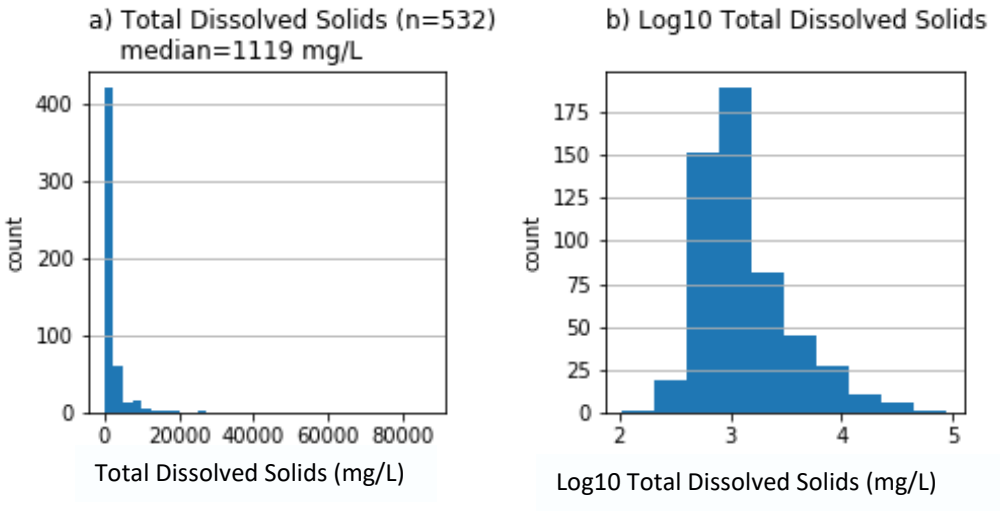


Figure 29. Histogram of a) total dissolved solids and b) log10 of total dissolved solids.

The dataset is very skewed, even after log10 transform. The minimum TDS value is 105 mg/L, the maximum reported is 88,910 mg/L.

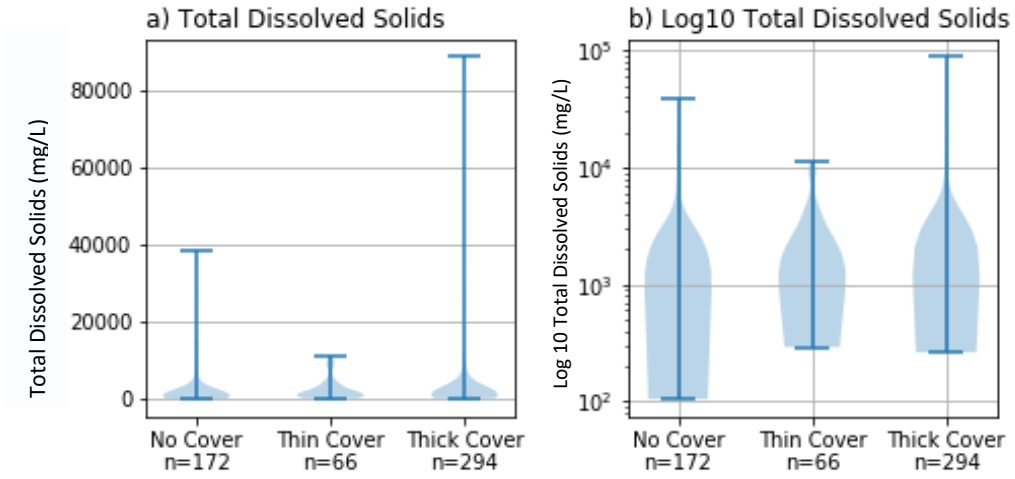


Figure 30. Violinplot of a) total dissolved solids and b) log10 of total dissolved solids, grouped by cover type

Figure 30 shows the distribution of TDS grouped per cover type. There is no clear trend, although the lowest values are associated with bores classified as 'No Cover'.

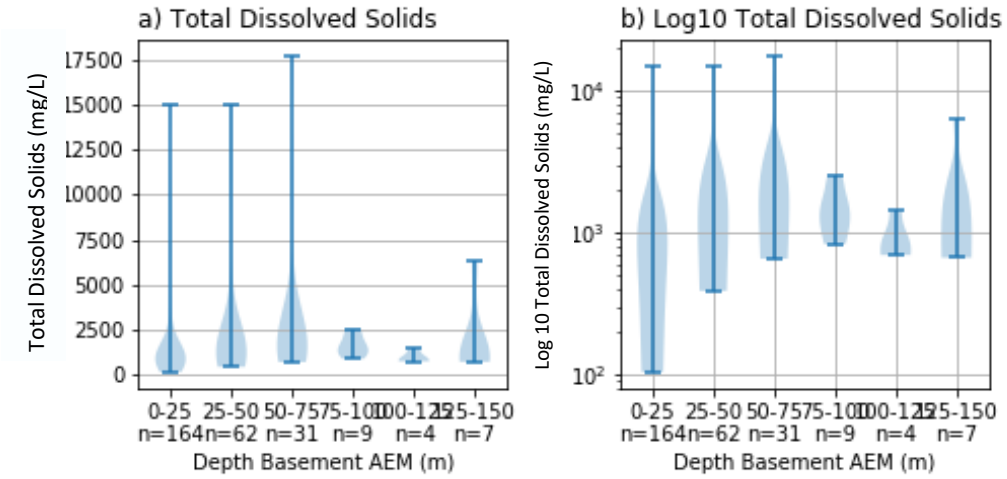


Figure 31. Violinplot of a) total dissolved solids and b) log10 of total dissolved solids, grouped by depth interval inferred from the 50 ohm contour of the merged TEMPEST and SkyTEM data.

Figure 31 shows an increasing trend in salinity with thicker cover, up to thicknesses of 75 m. For bores where basement is deeper than 75 m, salinity appears to decrease slightly. This decreasing trend is probably not significant, due to the small sample size.

The final violinplot examines the relationship between fault and fracture density and TDS. We use the first vertical derivative of the total magnetic intensity (1st VD TMI) as proxy for fault and fracture density, where values less than 125 are associated with higher fracture and fault density.

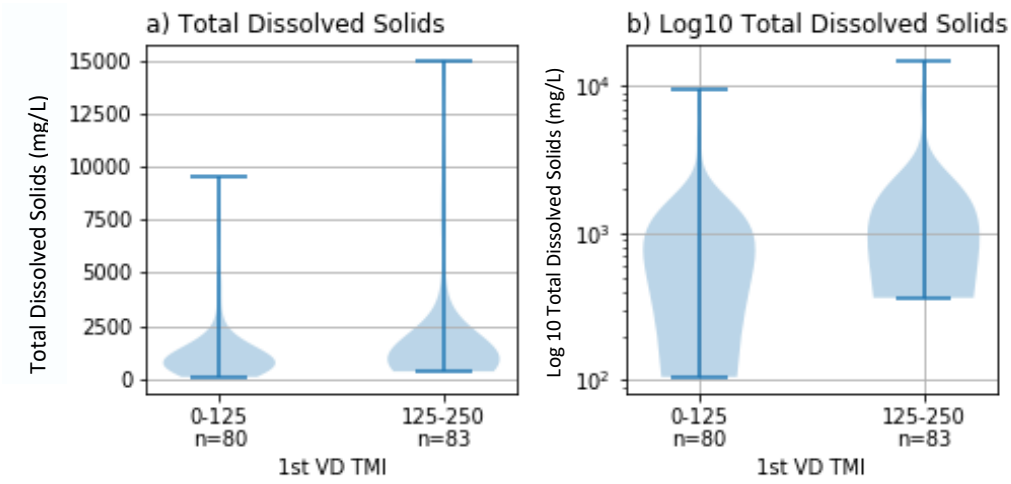


Figure 32. Violinplot of a) TDS and b) Log10 TDS at bores with AEM derived depth to basement less than 25 m, grouped according to the first vertical derivative of total magnetic intensity.

Figure 32 only contains the bores where the depth to basement is less than 25 m. It appears that salinity tends to be less in bores in areas with higher probability of increased fracture density.

A.3.3 BORE YIELD

Bore yield is a proxy for hydraulic conductivity and storage. There are 469 bores with bore yield data, 88 of these are equal to zero. It is not clear if a zero yield means the bore is dry or if the yield was not measured. Zero yield data are excluded from the log10 plots.

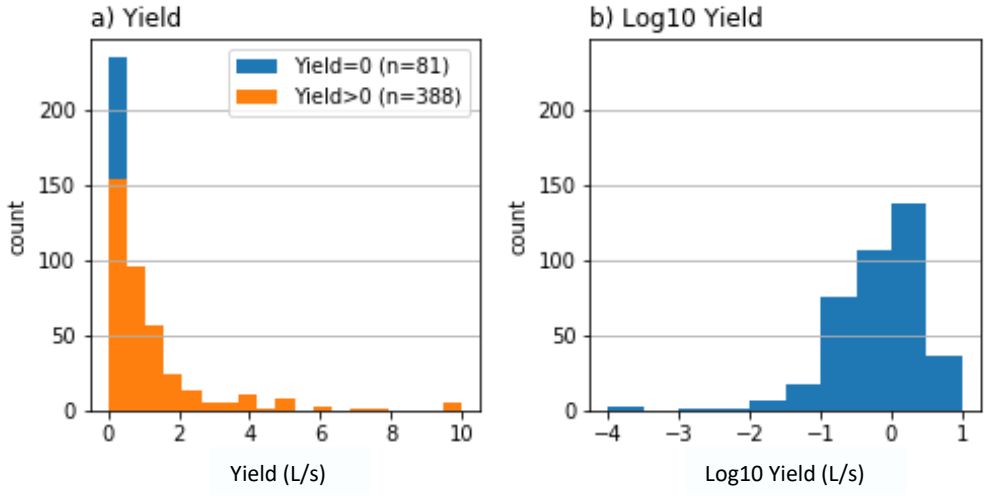


Figure 33. Histogram of a) bore yield and b) log10 bore yield.

Figure 33 shows that yield has a skewed distribution, but that log10 is more symmetric. The median yield is 0.52 L/s. The median excluding yield values equal to zero is 0.76 L/s.

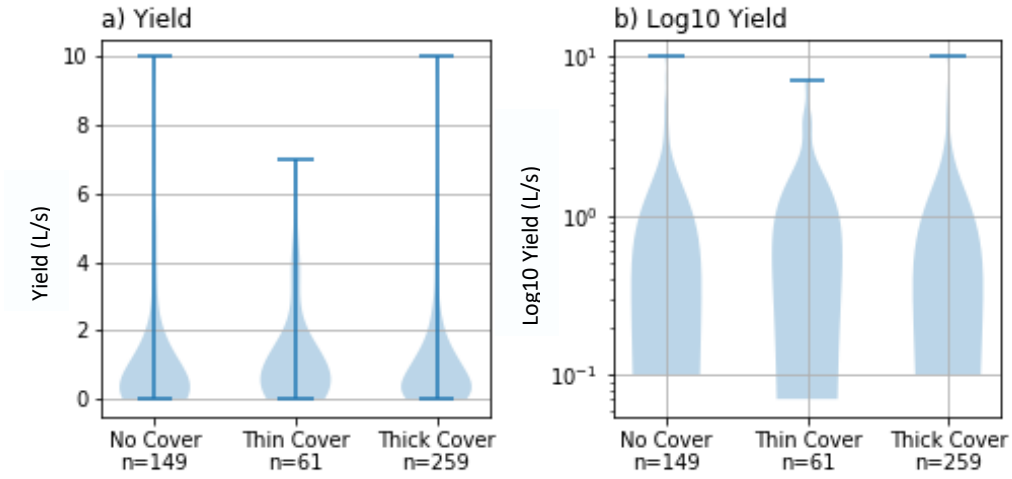


Figure 34. Violinplot of a) bore yield and b) log10 bore yield, grouped by cover type.

Figure 34 shows no apparent trend in bore yield based on cover type.

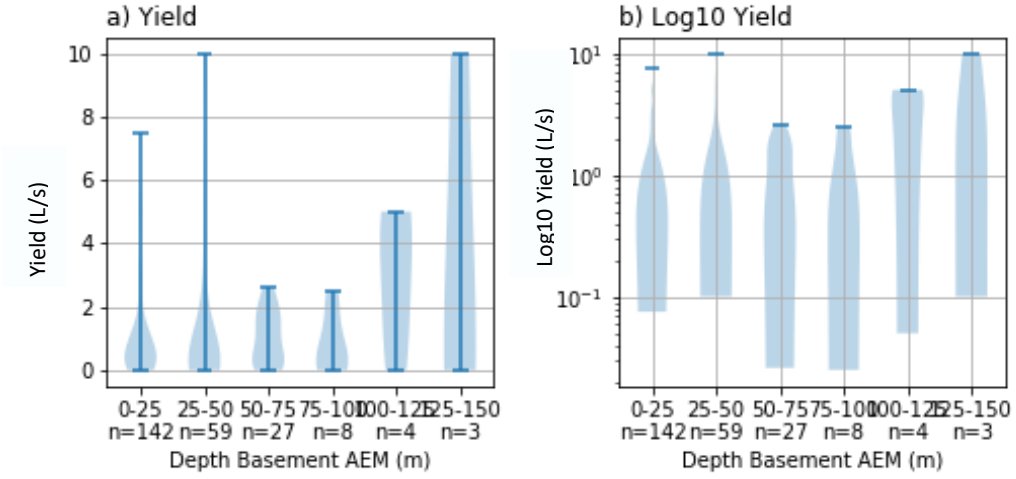


Figure 35. Violinplot of a) bore yield and b) log10 bore yield, grouped by depth to basement interval derived from the 50 ohm contour of the TEMPEST-SkyTEM data.

Figure 35 confirms that there is no apparent trend in yield with cover thickness, but the range of yield values increases with increasing depth to basement.

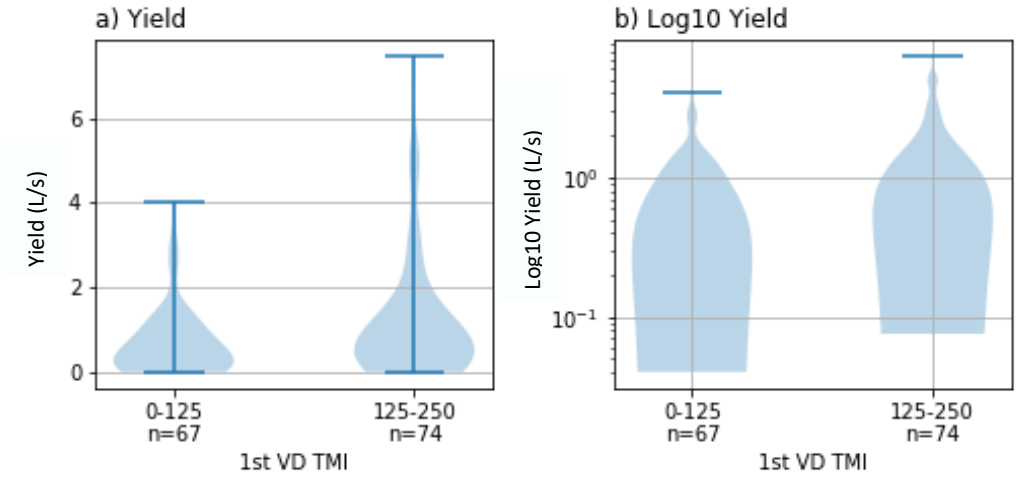


Figure 36. Violinplot of a) bore yield and b) log10 bore yield at bores with an airborne electromagnetic derived depth to basement less than 25 m, grouped by the 1st vertical derivative of total magnetic intensity.

Figure 36 appears to indicate that smaller yield is associated with areas of high fracture and fault probability. The difference is however smaller than for salinity (Figure 32) and this correlation is likely spurious.

A.3.4 STANDING WATER LEVEL

Standing water level measurements are available in 514 bores, while 469 bores have a record of water cut.

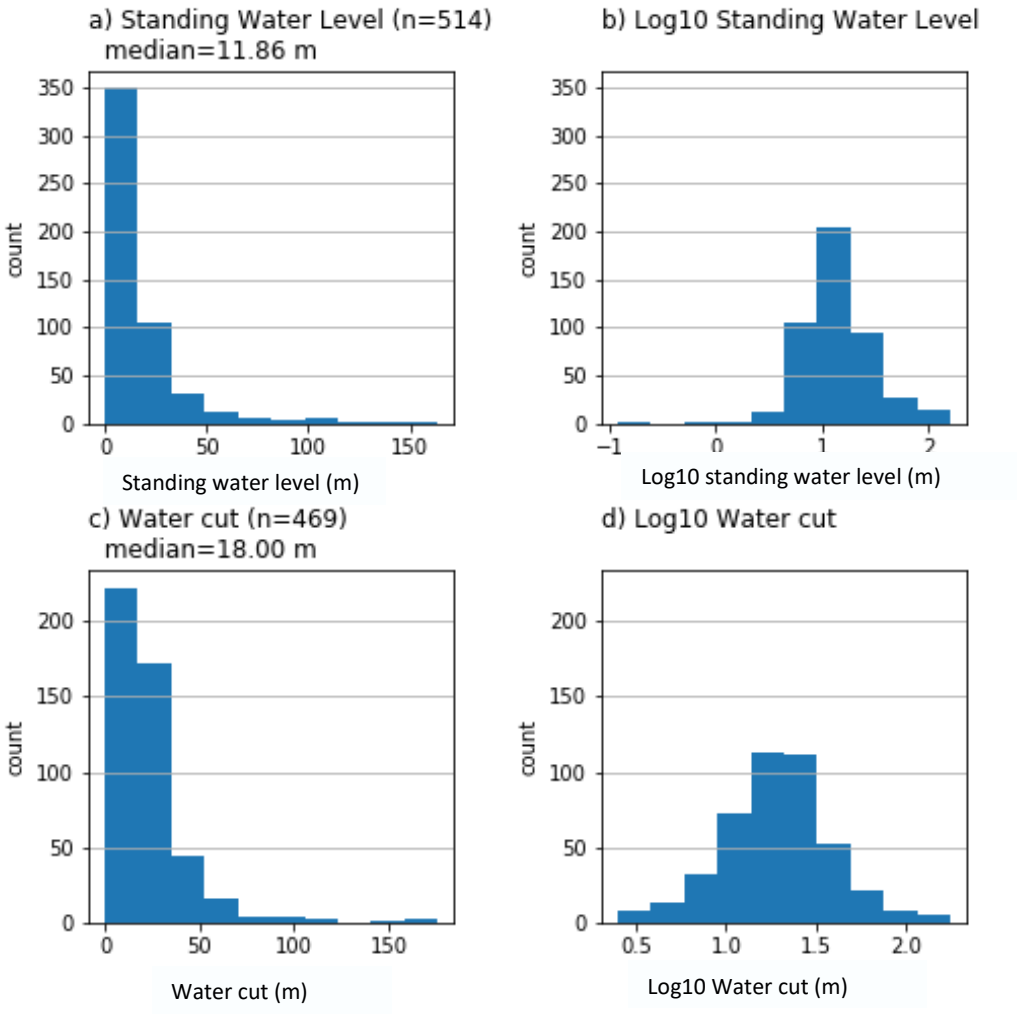


Figure 37. Histogram of a) standing water level, b) log10 standing water level, c) water cut and d) log10 water cut

Figure 37 shows that both standing water level and water cut have skewed distributions, but are symmetric after log10 transform. In half of the bores the standing water level is less than 12 m deep, while water is encountered in half of the bores at depths shallower than 18 m.

If the standing water level is much shallower than the water cut, it may indicate confined conditions.

Table 4: Confinement status of bores based on standing water level and water cut

AIRBORNE ELECTROMAGNETIC DEPTH TO BASEMENT INTERVAL	BORE DEPTH 0-25 m	BORE DEPTH 25-50 m	BORE DEPTH 50-75 m	BORE DEPTH 75-100 m	BORE DEPTH 100-125 m	BORE DEPTH 125-150 m
SWL<Watercut-2 m	81	38	14	3	3	2
SWL>Watercut-2 m	83	24	17	6	1	5

Table 4 cross-tabulates confinement status against AEM derived depth to basement interval. No apparent trends can be identified.

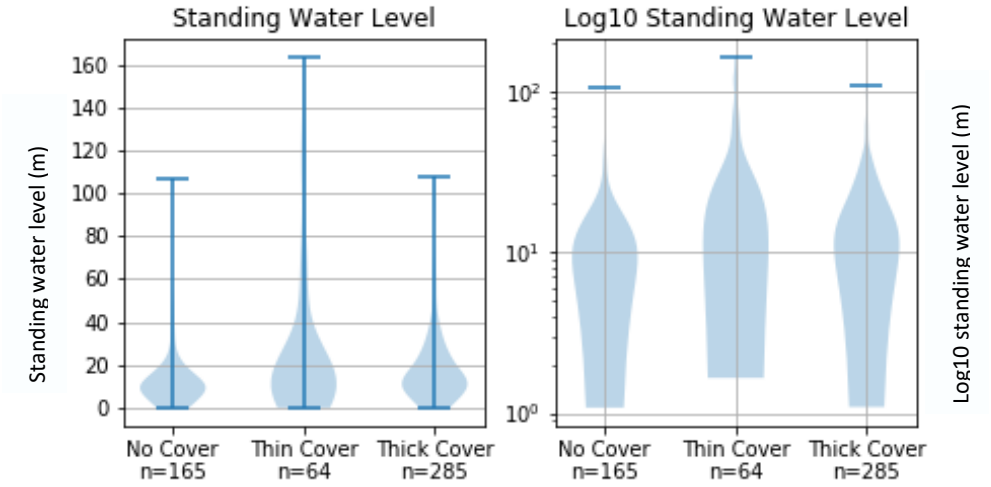


Figure 38. Violinplot of a) standing water level and b) log10 standing water level, grouped by cover type.

Figure 38 shows the standing water level, grouped by cover type. There is an apparent trend of standing water levels being deeper in bores classified 'Thin Cover'.

Figure 39 examines the relationship between standing water level and ground level elevation. At the lowest elevations, water levels are deepest.

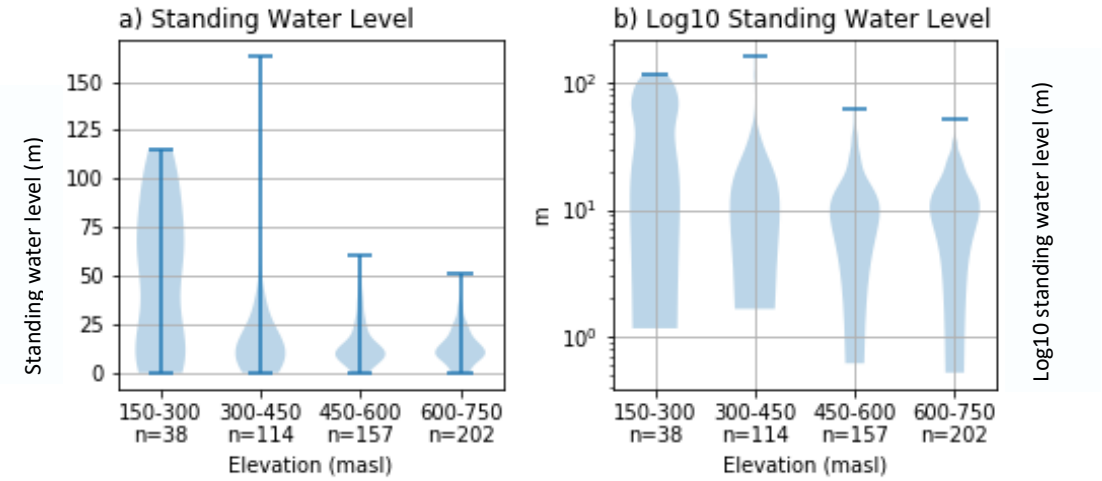


Figure 39. Violinplot of a) standing water level and b) log10 standing water level, grouped by ground level elevation interval.

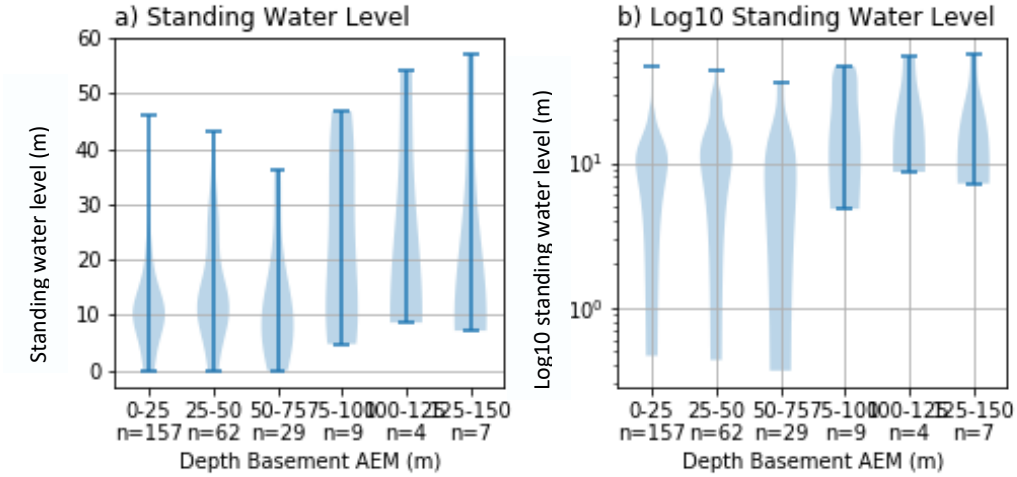


Figure 40. Violinplot of a) standing water level and b) log10 standing water level, grouped by grouped by depth to basement interval derived from the 50 ohm contour of the TEMPEST-SkyTEM data.

Figure 40 shows that standing water levels increase with increasing depth to basement, with a marked change at depths to basement in excess of 75 m.

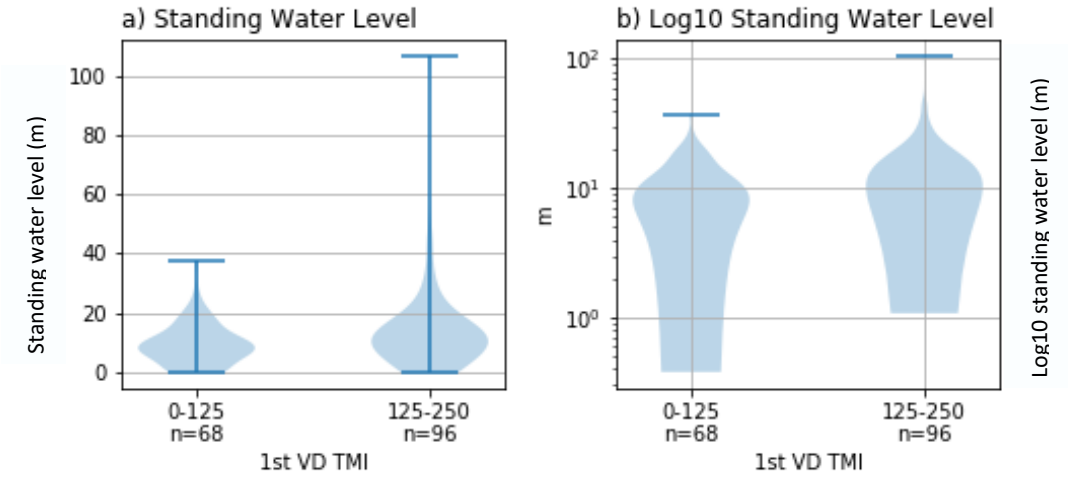


Figure 41. Violinplot of a) standing water level and b) log10 standing water level at bores with an airborne electromagnetic derived depth to basement less than 25 m, grouped by the 1st vertical derivative of total magnetic intensity.

Figure 41 indicates that standing water levels are shallower in areas with a higher probability of high fault density.

Appendix B Sensitivity Analysis

B.1 Methodology

The probability of developing a successful bore, and by extension the maximum number of bores to drill, is a function of the hydraulic properties (K, n_e), geometry (d) and salinity (S) of the aquifer as bore as the thresholds and constraints (Q_t, T_t, and S_t):

$$P_w = f(K, n_e, d, S, Q_t, T_t, S_t) \quad \text{Eq. B-1}$$

Determining which parameter in this equation is most influential towards P_w is not trivial. For instance, the maximum pumping rate is a function of hydraulic conductivity and thickness. As the range of K values is likely to be larger than the range of thickness values, K seems the more important parameter. However, thickness also appears in the volume calculation. This non-linearity makes it difficult to estimate which parameter is of most importance *a priori*. The thresholds add another level of complexity; if the salinity threshold value is high, salinity may not be an important parameter. If the threshold value is low, however, salinity may be the dominant feature.

In order to estimate the sensitivity of P_w to these parameters, a Sobol sensitivity analysis is carried out (Saltelli et al. 2010). Sobol sensitivity analysis is a global variance-based technique that not only estimates the individual effect of a parameter but also its total effect, accounting for interactions between parameters.

The method relies on randomly sampling parameter values and processing the results to calculate a sensitivity index that scales from 0 (not important) to 1 (very important). Calculating P_w however requires a range of values, not a single parameter combination. To address this issue, a random parameter combination is sampled from the parameter probability distributions assigned to each grid cell. For the sensitivity analysis, P_w is then calculated using a uniform distribution (U) for each of the parameters in Eq. B-2, with maximum and minimum value of the distribution, U(min, max), defined by plus and minus ten percent of the range of the parameter value y, respectively:

$$U\big(y_i - 0.1(max(y) - \min(y)), y_i + 0.1(max(y) - \min(y))\big) \quad \text{Eq. B-2}$$

B.2 Results

The probability of success and the corresponding maximum number of bores to allow for in the budget are controlled by the physical properties of the system (hydraulic conductivity, porosity, aquifer geometry and salinity) but also by the water resource management targets of pumping rate, duration of pumping and salinity. The non-linear nature of the probability equation makes determining the controlling factor a non-trivial task. As described in the methods section, sensitivity analysis provides a systematic way of determining which variable has the largest influence on the probability.

The Sobol sensitivity analysis as implemented in the Python SALib package (Herman and Usher, 2017) requires 20,000 model evaluations to estimate the sensitivity indices for the individual and total effect of the nine parameters, for each of the three hydrogeological units. Figure 42 summarises the results of the sensitivity analysis.

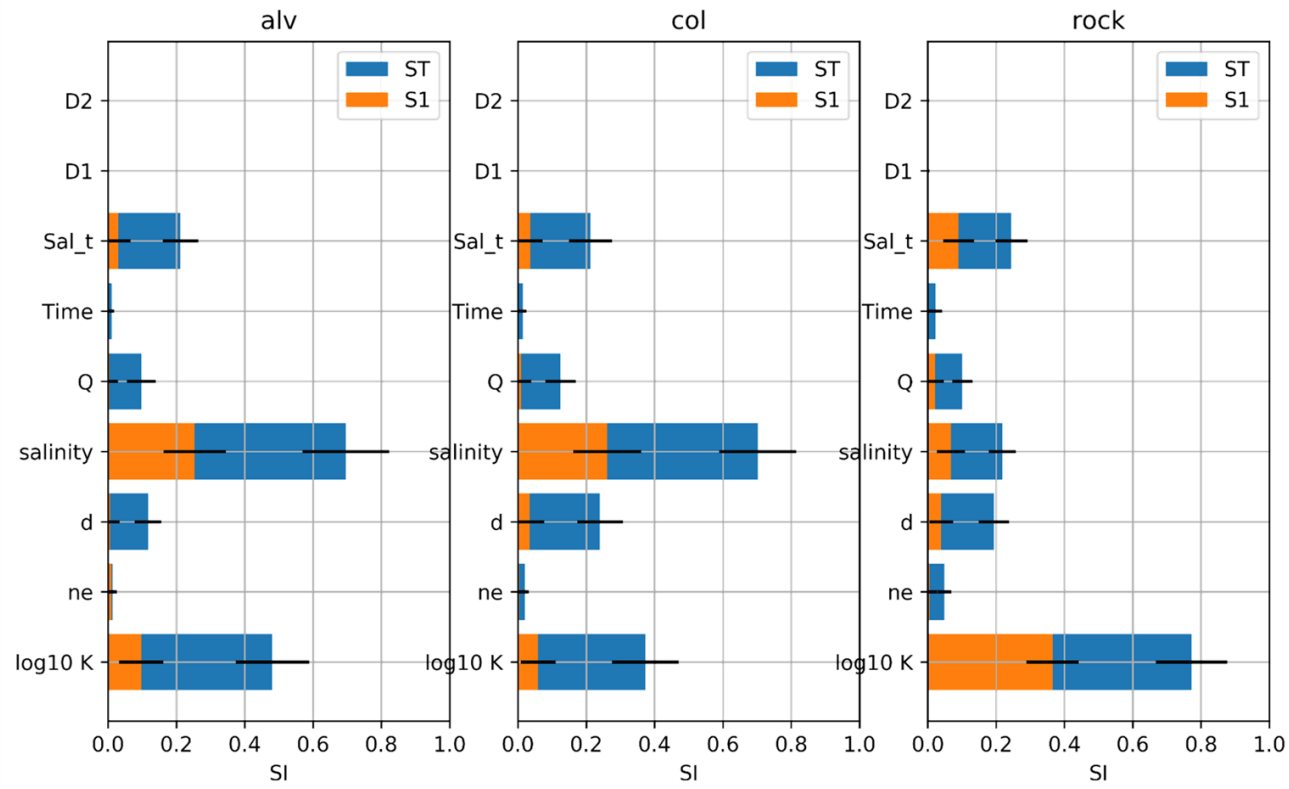


Figure 42. Sobol sensitivity indices (SI) for total (ST, blue) and individual (S1, orange) effect of each variable for the three hydrogeological units. The black bars indicate 95 percent confidence intervals of the sensitivity indices obtained through bootstrapping. Variables D1 and D2 are dummy variables, random variables that do not affect the probability, which can be used as reference.

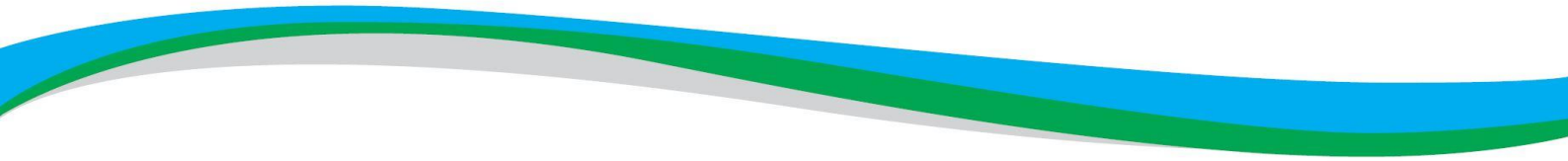
It is apparent from Figure 42**Error! Reference source not found.** that the dominant variable for both the alluvial and colluvial unit is the salinity, while in the basement unit hydraulic conductivity is more important. Aquifer thickness is an important secondary variable across all units. The most important threshold across the three units is the salinity threshold.

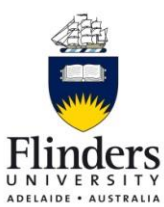
The total effect of variables is considerably greater than the individual effect, which highlights the importance of accounting for parameter interactions in analysing the potential of successfully developing water supply bores.

B.3 References

Herman J and Usher W (2017) SALib: An open-source Python library for sensitivity analysis. *Journal of Open Source Software* 2(9). 97, doi:10.21105/joss.00097

Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M and Tarantola S (2010) Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications* 181: 259–270. doi:10.1016/j.cpc.2009.09.018





The Goyder Institute for Water Research is a partnership between the South Australian Government through the Department for Environment and Water, CSIRO, Flinders University, the University of Adelaide, the University of South Australia, and the International Centre of Excellence in Water Resource Management.