https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

Hydrology and Earth System Sciences
Discussions

Estimation of groundwater age distributions from hydrochemistry: Comparison of two metamodelling algorithms in the Heretaunga Plains aquifer system, New Zealand

Conny Tschritter[1], Christopher J. Daughney[2], Sapthala Karalliyadda[3], Brioch Hemmings[1], Uwe Morgenstern[3], Catherine Moore[3]

[1]GNS Science, Taupo, New Zealand
[2]National Institute of Water and Atmospheric Research, Wellington, New Zealand
[3]GNS Science, Lower Hutt, New Zealand

Correspondence to: Conny Tschritter (c.tschritter@gns.cri.nz)

Abstract

Groundwater age or residence time is important for identifying flow and contaminant pathways through groundwater systems. Typically, groundwater age and age distributions are inferred via lumped parameter models based on measured age tracer concentrations. However, due to cost and time constraints, age tracers are usually only sampled at a small percentage of the wells in a catchment. This paper describes and compares two methods to increase the number of groundwater age data points and assist with validating age distributions inferred from lumped parameter models. Two machine learning techniques with different strengths were applied to develop two independent metamodels that each aim to establish relationships between the hydrochemical parameters and the modelled groundwater age distributions in one test catchment. Ensemble medians from the best model realisations per age distribution percentile were used for comparison with the results from traditional lumped parameter models based on age tracers. Results show that both metamodelling techniques generally work well for predicting groundwater age distributions from hydrochemistry. Therefore, these techniques can be used to assist with the interpretation of lumped parameter models where age tracers have been sampled, and they can also be applied to predict groundwater age distributions for wells that have hydrochemistry data available, but no age tracer data.

1. Introduction

Groundwater age describes the residence time of a parcel of water within the aquifer system, i.e., the time elapsed since recharge. Water from different flow pathways converges at sampling points such as wells and springs. Thus, each sample is a mixture of different groundwater with varying sources and ages (Maloszewski and Zuber, 1996). Understanding the age of water in the groundwater system is key to determining flow paths, recharge rates and recharge sources, as well as understanding the sustainability of groundwater abstraction, the movement of contaminants in water and the impacts of land use on water quality (Ginn et al., 2009; Daughney et al., 2010; Massoudieh et al., 2012; Morgenstern and Daughney, 2012).

The groundwater age distribution and the mean age, or mean residence time, can be inferred using two main methods. Commonly, groundwater age is inferred from age tracer concentrations, in combination with lumped parameter models (LPMs), to reflect mixing of water from different flow paths. LPMs are quick to use and allow the representation of different flow and mixing models, which are matched to the measured concentrations of environmental tracers, like tritium, sulphur hexafluoride (SF_6) and chlorofluorocarbons (CFCs), in the groundwater sample (Maloszewski and Zuber, 1996; Morgenstern and Daughney, 2012). A disadvantage is that LPMs can only be fitted to locations at which tracer concentrations have been measured. As an alternative to LPMs, physically based numerical flow and transport models can also be used to assess groundwater age and transit time distributions. Some such investigations have focused on the mathematical descriptions of groundwater age and its dynamics (Ginn et al., 2009; Cornaton, 2012; Engdahl, 2017), whereas other investigations have evaluated the role of groundwater age and environmental tracer data in model calibration, alongside other data sources such

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

as hydraulic heads and stream flow measurements (Portniaguine and Solomon; 1998; Zhu, 2000; Massoudieh et al., 2012). An advantage of the numerical modelling studies is that they can evaluate age distributions spatially and temporally across the entire model domain, and account for age distributions with more complex shapes than can be represented by simple LPMs. A key disadvantage is that the development of numerical models typically requires much more time and effort compared to the simpler LPMs, even after accounting for the time and costs of measuring the environmental tracer concentrations at the sites of interest. Additionally, as outlined in Knowling et al. (2020), numerical models require appropriate structure and parameterisation to ensure that the information from age tracers can be robustly assimilated by the model.

In recognition of the limitations of the above-listed methods, various less time and cost-intensive methods have previously been trialled to increase the amount of available groundwater age data in areas where no age tracers have been sampled and analysed, and where a numerical flow and transport model is not available. Typically, these alternative methods for estimating groundwater age rely on groundwater chemistry data, hydrogeological information (e.g. bore construction parameters, recharge, geology, etc.), or a combination thereof (Edmunds and Smedley, 2000; Daughney et al., 2010; Beyer et al., 2016; Marçais et al., 2018), linking groundwater chemistry and hydrogeological parameters to groundwater age and transit time of water through the aquifer. Most such previous studies have relied on statistical data analysis methods, i.e. discriminant analysis, principal component analysis, regression analysis etc., that were used independently or in combinations with each other to identify and model relationships between groundwater chemistry and age data (Daughney et al., 2010; Beyer et al., 2016). These methods have been shown to be reasonably successful in deriving mean groundwater age, either as an age category or absolute age, but did not provide estimates of the full groundwater age distributions, which are more meaningful for contaminant transport and drinking water security than mean age (Beyer et al., 2016; Weissmann et al., 2002; Suckow, 2014).

This study builds on previous investigations of the use of groundwater chemistry as a proxy to infer groundwater age, with the aim of using metamodels to assess the full age distribution instead of just the mean age. Metamodels (also known as 'surrogate' or 'data driven' models) are statistical or machine learning-based 'models of models', which are trained on observed data and can be used to extrapolate relationships to enable predictions to be made at unsampled locations or times (Fienen et al., 2015, 2016, 2018; Asher et al., 2015; Starn and Belitz, 2018; Starn et al. 2021). Therefore, they provide a cost-efficient alternative to both physically based distributed numerical models or LPMs, whenever sufficient training data exists (Razavi et al., 2012). Alternatively, in more data-sparse contexts, they may be used in combination with numerical modelling efforts (Koch et al., 2019; Reichstein et al., 2019). Metamodels can make relatively rapid predictions of system behaviour or characteristics based on the relationships that are established with observed data. Although metamodel predictions will typically have a higher uncertainty than numerical model predictions, they can be made more rapidly while efficiently dealing with high parameter dimensionality (Fienen et al., 2016). Metamodels have been developed for various hydrogeological applications (Fienen et al., 2018; Nolan et al., 2018; Starn and Belitz, 2018; Asher et al., 2015), including the prediction of groundwater age distributions from hydrogeographic and bore-specific observations and features, or numerical flow model outputs (Fienen et al., 2016; Starn and Belitz, 2018). However, none of these metamodeling studies have investigated the use of hydrochemistry for the predictions of groundwater age.

Specifically, this study evaluates and compares the performance of two ensemble machine learning techniques with the goal of estimating groundwater age distributions from groundwater chemical composition. Symbolic regression (SR) is a machine learning technique that attempts to identify explicit mathematical expressions in an input dataset. It is initiated by developing a population of naïve random mathematical expressions that conform to *a priori* selected grammar rules. The initial mathematical expressions are then combined and evolved through an approach such as genetic programming, in order to develop a set of formulas that describe the relationship(s) of interest with sufficient accuracy (Gomes et al., 2019). Gradient boosted regression (GBR) is a machine learning method that aims to minimize the prediction error through a regression tree model - a sequence of regression trees. Each sequential addition of the new regression tree will minimize the prediction error

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022


Hydrology and Earth System Sciences
Discussions

made by the previous tree and thus decrease the overall prediction error. Whilst there are numerous possible machine learning methods that can be used for this purpose (e.g. Random Forest, Bagged decision trees, Neural Networks), we selected the Symbolic Regression (SR) and Gradient Boosted Regression (GBR) techniques based on the amount of available data, ease of use and adaptability, and/or proven potential in similar research. For example, unlike most other machine learning methods, SR provides the actual equation for the resulting model. This means that the user can directly see the calculation that is being performed, which in turn helps to check on the physical basis of the equation and also helps with transferring the model into other software like Microsoft Excel, making it more accessible to a wider user group. GBR, on the other hand, is a highly adaptive strong predictive model and has previously successfully been used in other studies to predict groundwater age from hydrophysical parameters (Fienen et al., 2018, 2016).

The GBR and SR approaches were implemented to estimate selected percentiles in the LPM groundwater age distribution based on measured groundwater chemistry on a per-sample basis in a test catchment, the Heretaunga Plains, New Zealand. Although this study uses LPM-derived age distributions as the metamodel prediction targets, we note that our approach would also be applicable to the use of groundwater chemistry to predict the age distributions derived from a physically based numerical model.

2. Study Area

The Heretaunga Plains is a 300km^2 SW-NE trending fault-bounded depression located on the east coast of New Zealand's North Island (Fig. 1). The Ngaruroro, Tutaekuri and Tukituki are the three main rivers that traverse the Heretaunga Plains, which have long-term median flows near the coast of 19.9, 8.5 and 21.8 m^3/s, respectively (Waldron et al., 2019). The area has a temperate climate with average temperatures of 17°C in summer and 10°C in winter, and average annual rainfall of approximately 800 mm near the coast (Dravid and Brown, 1997). Land cover in the western portion of the catchment is comprised primarily of native forest, scrub, and tussock, whereas the eastern portion is primarily exotic grassland (mostly used for grazing sheep and beef cattle) with lesser areas of orchard, vineyard, and short-rotation cropland, along with urban areas located near the coast (Smith et al., 2020).

2.1. Geology and hydrogeology

Starting in the Miocene, tectonic activity associated with the Hikurangi Trough, which is part of the Australian-Pacific plate boundary, resulted in the development of an actively subsiding syncline ('Napier Syncline') in the area of the Heretaunga Plains (Fig. 1). The axis of this syncline is oriented subparallel to the orientation of the lengths of the plains, and the resulting depression has since been infilled by marine and alluvial deposits representing several glacial – interglacial cycles and associated sea level fluctuations (Lee et al., 2014). The total depth of this depression is uncertain, but it has been estimated to be between 900 m (Dravid and Brown, 1997) and 1,600 m (Beanland et al., 1998).

The main aquifers of the Heretaunga Plains are composed of highly transmissive, gravel-dominated fluvial deposits from the late Pleistocene (Maraekakaho Formation) and Holocene (Heretaunga Formation), deposited by the three major rivers in the plains (Dravid and Brown, 1997). Lee et al. (2014) analysed 4051 lithological well logs provided by regional authorities and found that most of the primarily 20 – 50 m deep bores terminate in gravels deposited during the last glaciation (71 000–12 000 years ago). Towards the coast, these gravel deposits are overlain by silt- and clay-dominated marine sediments, deposited during the Holocene marine transgression, which thicken towards the coast and act as a confining layer. Smaller gravel aquifers also occur at the coast. Further inland, Holocene terrestrial deposits, i.e., gravel, sand, clay, and silt, interfinger with the marine deposits, resulting in an interconnected confined-unconfined aquifer system of aquifers (Dravid and Brown, 1997). Thick Holocene gravel fans, associated with the Ngaruroro, Tukituki and Tutaekuri rivers, which have been mapped from bore logs, are likely hydraulically connected to the underlying last glacial gravels (Lee et al., 2014; Begg et al., 2022).

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

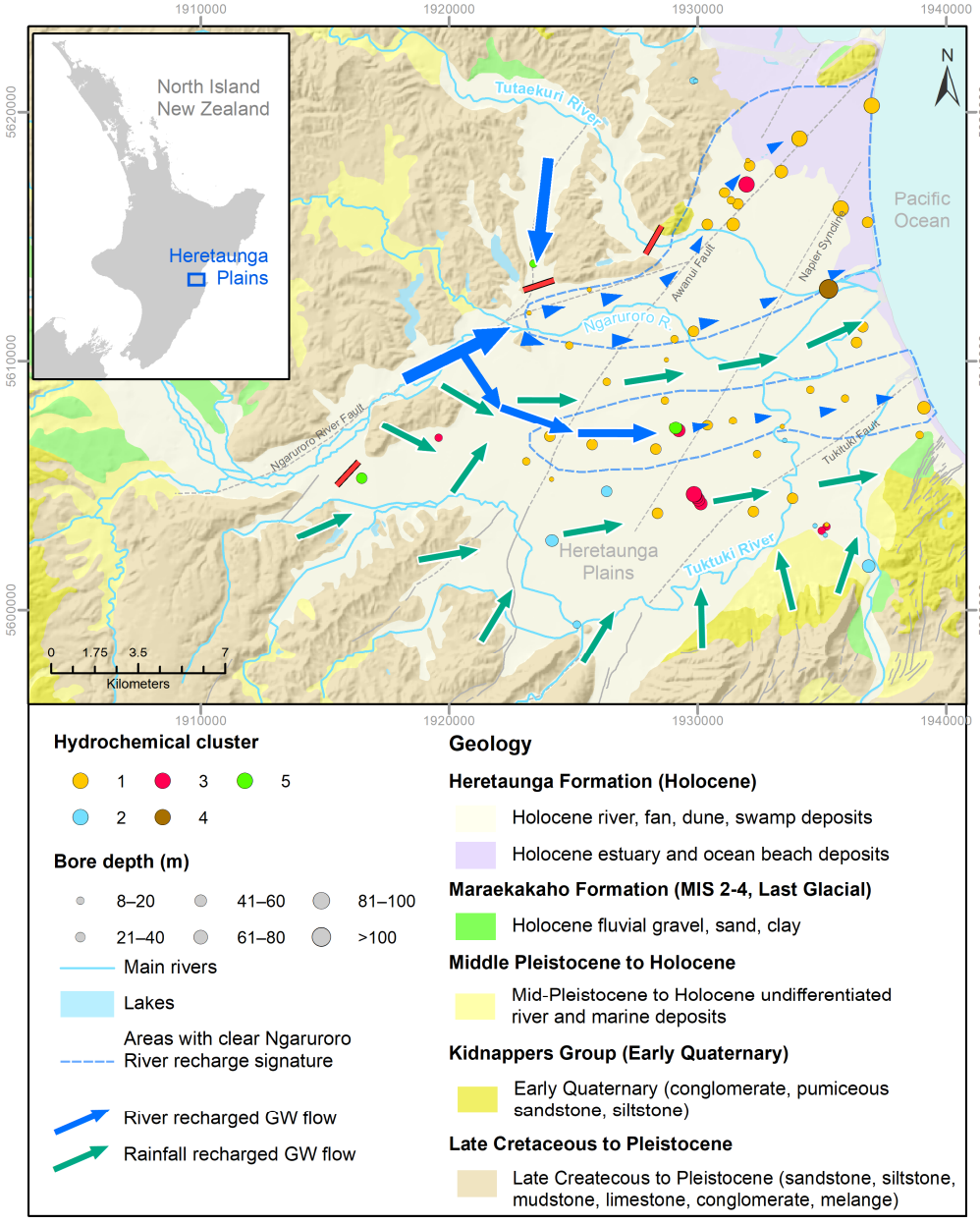


Figure 1: Bore locations and depth, hydrochemical cluster, geology, and inferred groundwater flow dynamics in the Heretaunga Plains (hydrochemical cluster and groundwater flow dynamics from Morgenstern et al. (2018); geology from Heron, 2020; rivers and lakes from LINZ, 2022). The length of the arrows is proportional to the estimated flow rate.

Underlying the Pleistocene gravel deposits are Late Cretaceous to Pleistocene marine and terrestrial deposits (mudstone, melange and mudstone, sandstone, siltstone, limestone, and conglomerates). Based on their mapped occurrence outside of the Heretaunga Plains and seismic reflection data from within the plains, these deposits are expected to underlie the study area at

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022


Hydrology and Earth System Sciences
Discussions

depth. However, none of the groundwater bores reach these deposits, and the only bore data available is from a small number of petroleum exploration bores (Dravid and Brown, 1997; Lee et al., 2014).

2.2. Groundwater flows

Sources of groundwater recharge into the Heretaunga Plains aquifers have been inferred from river flow gauging surveys (Wilding, 2018), groundwater level monitoring (Smith et al., 2020), numerical modelling (Rakowski, 2018; Rakowski and Knowling, 2018), and assessments of water chemistry, stable isotopes, and age tracers (Morgenstern et al., 2018). These methods collectively indicate that losses from the main rivers occur in limited areas but contribute about two thirds of the total volume of groundwater recharge to the aquifer system, with the remainder of recharge sourced from rainfall percolation through the soil zone across a wider area of the Heretaunga Plains (Fig. 1).

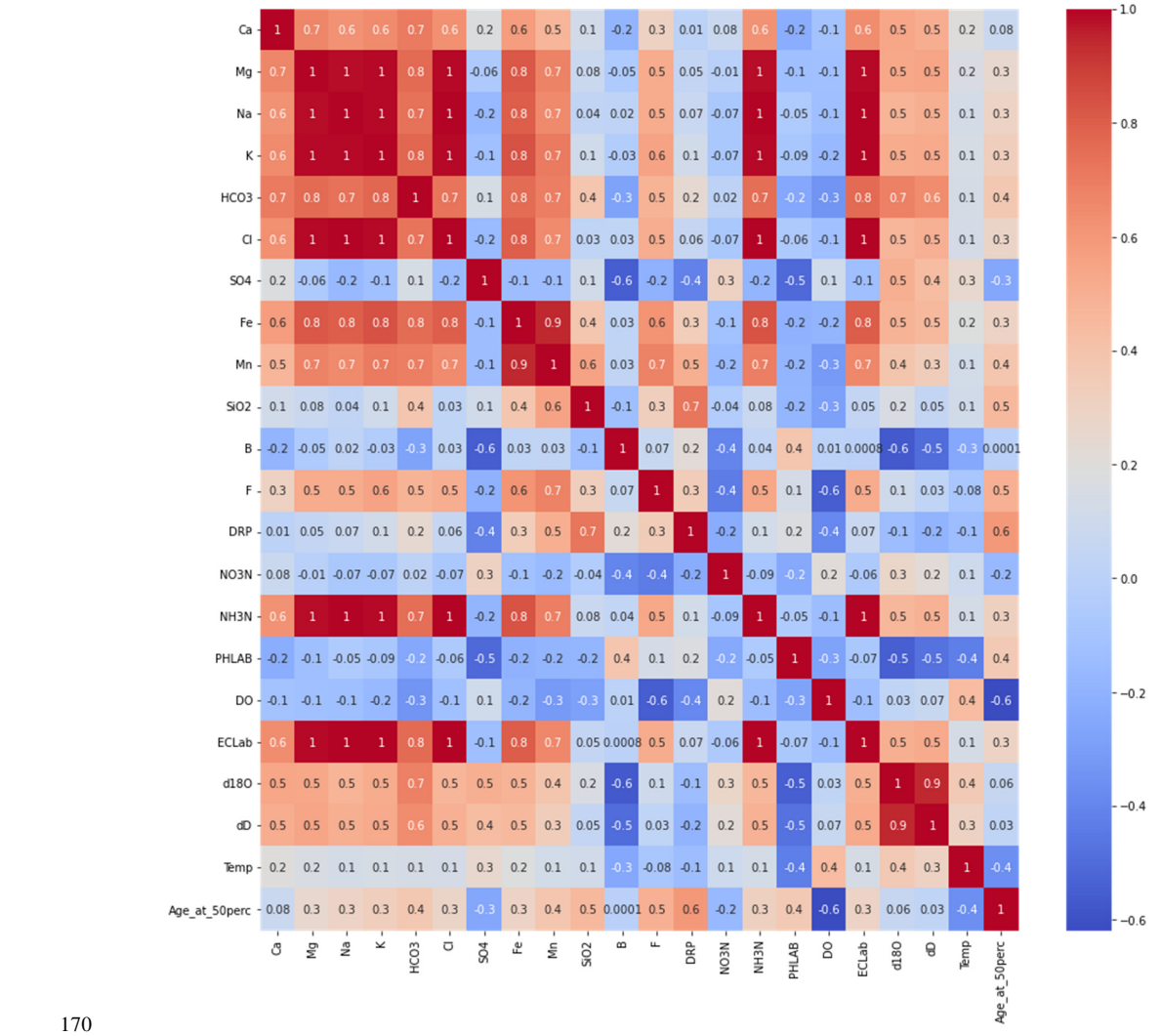
The dominant groundwater flow direction is from west to east, following the topographic gradient towards the coast (Fig. 1). Artesian and sub-artesian conditions are observed in bores in the confined aquifer zone near the coast (Dravid and Brown, 1997). Age tracer measurements indicate relatively rapid horizontal groundwater velocities of ca. 3-5 km/year in some parts of the Heretaunga Plains aquifer system, particularly in proximity to losing reaches of the main rivers (Morgenstern et al., 2018). Bores as deep as 75 m below ground surface can have tritium concentrations similar to modern rainfall, indicating that vertical groundwater flow can also be relatively rapid in some areas. In contrast, the older groundwaters and slower flow velocities of ca. 0.1-0.2 km/year are inferred nearer the coast, which could result from widening of the aquifer cross section and/or decreasing hydraulic conductivity, e.g., reflecting the presence of finer-grained sediments of marine origin (Morgenstern et al., 2018).

Approximately 40% of the discharge from the aquifer system is estimated to occur via seepage into streams and springs, with the remaining discharge evenly split between abstraction and flows across the coastal boundary (Rakowski and Knowling, 2018). Total abstraction has approximately doubled in the last 30 years, with an average annual increase of approximately 3.5%, due primarily to increases in irrigation and industrial use of groundwater (Rakowski and Knowling, 2018). This increase in total abstraction is inferred to be the cause of long-term declines of summer groundwater levels (average rate ca. 5 cm/year between 1989 and 2018), which are observed in some unconfined parts of the aquifer system (Smith et al., 2020).

2.3. Hydrochemistry

Groundwaters in the Heretaunga Plains have a range of hydrochemistry (Fig. 1), arising from the spatially variable processes of human impact and natural geochemical evolution, as observed elsewhere in New Zealand (Daughney et al., 2012; Morgenstern and Daughney, 2012). Generally, natural geochemical evolution is expected to affect the redox state, with younger groundwaters more likely to be oxic than anoxic, thereby affecting the concentrations of redox-sensitive substances such as dissolved oxygen (DO), NO_3-N, NH_3-N, Fe, Mn and SO_4 (Tesoriero and Puckett, 2011; Daughney et al., 2010). Natural water-rock-interaction also typically causes the concentrations of the major ions to increase with time and distance along a groundwater flow path (Morgenstern and Daughney, 2012). Human influence on groundwater chemistry in New Zealand is primarily indicated by elevated concentrations of NO_3-N, sometimes co-occurring with elevated concentrations of Na, K, Mg and/or Cl (Daughney et al., 2012; Morgenstern and Daughney, 2012). The dominant recharge source also influences hydrochemistry, with groundwaters sourced primarily from rainfall seepage through the soil zone tending to have higher total dissolved solids (TDS) and higher concentrations of the parameters associated with human activity compared to groundwaters sourced from river seepage (Morgenstern and Daughney, 2012). These general drivers of hydrochemistry can lead to reasonably strong correlations among the levels of several parameters, as is observed for the groundwaters in the Heretaunga Plains (Fig. 2).

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022



170

Figure 2: Pearson R correlation matrix among hydrochemical parameters and groundwater age, estimated via the LPM at the 50th percentile of the age distribution, across all Heretaunga Plains groundwater samples used in this study (n=76).

Oxic groundwaters inferred to be recharged from rivers are found across much of the study area (denoted as Cluster 1 in Fig. 1). These groundwaters typically have Ca and HCO_3 as the dominant cation and anion, with concentrations of ca. 20-30 mg/L and 50-100 mg/L, respectively (Morgenstern et al., 2018). Due to their redox status, such groundwaters have concentrations of DO, NO_3-N and SO_4 above their respective analytical detection limits, but concentrations of Fe, Mn and NH_3-N are usually below detection. These groundwaters display relatively little indication of land-use impacts: concentrations of NO_3-N are typically below 1 mg/L, and microbial pathogens and pesticides are generally not detected (Smith et al., 2020). In some locations, particularly near the margins of the plains, these river-recharged groundwaters can display concentrations of Ca and HCO_3 that are 2-3 times higher than elsewhere, likely due to the influence of carbonate-rich geologies in the surrounding hills (denoted as Cluster 2 in Fig. 1).

Oxic groundwaters inferred to be recharged from rainfall occur in a small number of areas of the plains (denoted as Cluster 3 in Fig. 1). These groundwaters also typically have HCO_3 as the dominant anion but can have either Ca or Na as the dominant cation (Morgenstern et al., 2018). Otherwise, these groundwaters are generally hydrochemically similar to the oxic river-

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

Hydrology and Earth System Sciences
Discussions

recharged groundwaters described above, except for having slightly higher concentrations of NO_3-N, typically in the range 2-2.5 mg/L, as a result of modest land use impacts, along with slightly higher concentrations of Ca, Mg, Na, K and/or SiO_2 due to their accumulation during passage of recharge water through the soil zone (see Daughney and Morgenstern, 2012).

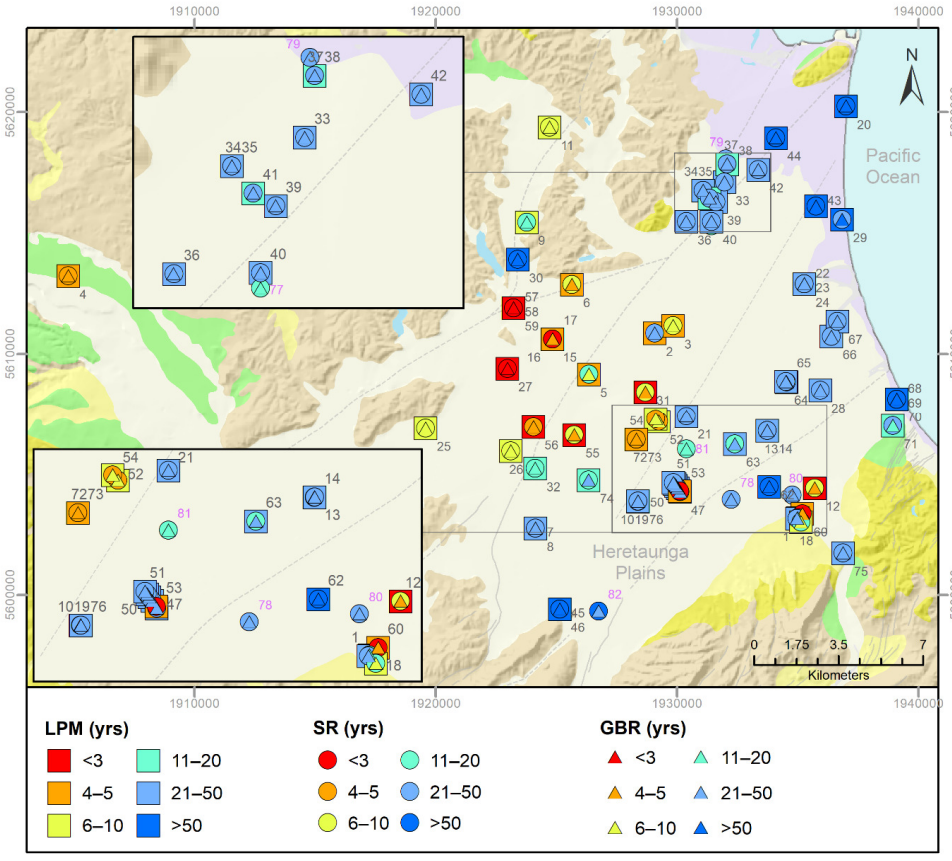
Anoxic groundwaters occur in a small number of wells (denoted as Cluster 4 in Fig. 1). Depending on their redox state, these groundwaters typically have detectable concentrations of Fe, Mn and/or NH_3-N but low or non-detectable concentrations of DO, NO_3-N and/or SO_4. Concentrations of PO_4-P are also observed to be higher in anoxic than oxic groundwaters, likely due to solubilisation associated with reductive dissolution of iron oxide minerals in the aquifer (Langmuir, 1997). A small number of wells have NH_3-N concentrations roughly twice as high as elsewhere, which may indicate contamination by wastewater (denoted as Cluster 5 in Fig. 1).

Groundwater chemistry shows complex relationships to groundwater age in the Heretaunga Plains (Morgenstern et al., 2018). Generally, shallow groundwaters within the Holocene unconfined gravels are estimated to have mean residence time (MRT) of 0-10 years, with a progressive increase in MRT to the range 40-80 years for deeper groundwaters near the coast (Fig. 3).

Younger groundwaters are more likely to be oxic, whereas deeper groundwaters are more likely to be anoxic, which affects the location and depth profiles of DO, NO_3-N, NH_3-N, Fe, Mn, SO_4 and PO_4-P. However, these relationships between age, location, depth, and groundwater chemistry are complex, for example because there are locations where young groundwaters are found at depth and older groundwaters are found near the surface, as a result of the complex flow paths in the Heretaunga Plains aquifer system. Accordingly, Morgenstern et al. (2018) did not report any predictive relationships between groundwater chemistry and groundwater age.

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

Hydrology and Earth System Sciences
Discussions
Open Access
EGU



205 **Figure 3: Locations of sites used for the model development and predictions showing the LPM, SR, and GBR modelled ages for the 50th percentile. The geological formations shown on the background map (Heron, 2020) are explained in Fig. 1. Labels correspond to mapIDs for each site: Grey: sites used in the development of the metamodels (mapIDs 1-76); Pink: sites without LPM data, used for predictions (mapIDs 77-82).**

3. Methods

3.1. Data

210 This study used hydrochemical data from Morgenstern et al. (2018) (Supplementary Material Table S2). The dataset is comprised of 76 groundwater samples collected from 69 sites in the Heretaunga Plains (Fig. 1). Bore depths ranged from 8 to 147 metres below ground surface (m bgs) (25th, 50th and 75th percentiles were 30, 46 and 71 m bgs, respectively). These hydrochemical samples were mainly collected during sampling campaigns in 2014, 2016 and 2019. Most samples (75%) were collected in the period April to June, with approximately even proportions of the remaining samples collected in the periods
215 January to March or November to December. All sites were sampled according to standard protocols involving purging of bores and stabilisation of pH, DO, electrical conductivity (EC) and temperature (T) as measured in the field using portable meters prior to sample collection (Daughney et al., 2007). Censored results for any parameter in any sample were replaced with the corresponding analytical detection limit in order to allow application of the machine learning methods in this study.

This study used age tracer data from Morgenstern et al. (2018) (Supplementary Material Table S2). The age tracers tritium
220 (^{3}H), CFCs and SF_6 were selected for their appropriateness for the relatively young groundwaters found in many New Zealand aquifer systems (Stewart and Morgenstern, 2001). One set of age tracer samples was collected from each site at the same times

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

as the above-mentioned samples that were analysed for hydrochemistry. Additionally, at 39 of the 76 sites, between two and 12 additional sets of age tracer samples had been collected for other investigations extending back as early as 1995. We point out that, in New Zealand and the Southern Hemisphere in general, bomb tritium has now fully dispersed, removing any ambiguity from fitting LPMs and increasing the reliability of tritium-based age interpretations. This is not yet the case for the Northern Hemisphere, where bomb tritium is still present in significant amounts within the groundwater systems still causing ambiguity in the age interpretations (Stewart et al., 2021).

All age tracer analyses were performed at the GNS Science Water Dating Lab. Tritium was analysed in a 1 L unfiltered, unpreserved sample using 95-fold electrolytic enrichment followed by ultra-low-level liquid scintillation spectrometry (Morgenstern and Taylor, 2009). The detection limit was 0.02 tritium units (TU), and the reproducibility of a standard enrichment was 1% via deuterium calibration. Samples for analysis of CFCs and SF_6 were collected in strict isolation from the atmosphere, as described by Daughney et al. (2007), using 125 mL and 1 L bottles, respectively. Concentrations of CFCs (CFC-11 and CFC-12) and SF_6 were analysed at GNS Science by gas chromatography (GC) using an electron capture detector as described by Busenberg and Plummer (1992) and van der Raaij (2003). Detection limits were 3×10^{-15} mol kg^{-1} for CFCs and 2×10^{-17} mol kg^{-1} for SF_6. Dissolved argon and nitrogen concentrations were measured simultaneously with CFCs by GC using a thermal conductivity detector (analytical accuracy is 1% and 3%, respectively). The argon and nitrogen concentrations were used to estimate the temperature at the time of recharge and the excess air concentration as described by Heaton and Vogel (1981), which allowed calculation of the atmospheric partial pressure (ppt) of CFCs and SF_6 at the time of recharge.

This study made use of all available age tracer data to constrain the LPM for the relevant site. Evaluation of the groundwater age distribution involved fitting of a LPM to the age tracer data using the TracerLPM workbook (Jurgens et al., 2012), following the approaches of Daughney et al. (2010) and Morgenstern et al. (2015). This involved use of the convolution integral to compare the measured tracer concentration at the sampling point (C_{out}) with its concentration in rainfall at the time of recharge (C_{in}), calculated following Eq. (1):

$$C_{out}(t) = \int_0^\infty C_{in}(t-\tau)\, e^{-lt} g(\tau) d\tau \quad (1)$$

where t is the time of observation, τ is the transit time (groundwater age), $e^{-\lambda\tau}$ is the decay term with $\lambda = \ln(2)/T_{1/2}$ (i.e. radioactive decay term for tritium with half-life $T_{1/2}$ = 12.32 years) and $g(\tau)$ is the system response function (Zuber et al., 2005). The time-series C_{in} for tritium input via rainfall was based on concentrations measured monthly at Kaitoke, near Wellington, New Zealand, since the 1960s (Morgenstern and Taylor, 2009), whereas the time-series for inputs of CFCs and SF_6 were based on measured and reconstructed data from Cape Grim, Australia, and other southern hemisphere sites (Cunnold et al., 1997; Maiss and Brenninkmeijer, 1998; Prinn et al., 2000; Thompson et al., 2004). The system response function defines the shape of the distribution of ages within the water sample, for example, as arising from convergence and mixing of groundwater flow paths at the well during sampling. System response functions comprised by a singular or binary exponential piston flow model (EPM) have been shown to provide good matches to time-series age tracer data for a wide range of New Zealand groundwater systems (Daughney et al., 2010; Morgenstern and Daughney, 2012; Morgenstern et al., 2015). A singular EPM has two unknowns, T and f:

$$g = 0 \text{ for } \tau < T(1-f) \quad (2)$$

$$g = \frac{1}{Tf} e^{\left(-\frac{\tau}{Tf}+\frac{1}{f}-1\right)} \text{ for } \tau \geq T(1-f) \quad (3)$$

where T is the mean residence time (MRT) and f is the ratio of the volume of exponential flow to the total flow volume at the groundwater discharge point, with T(1-f) being the time it takes for groundwater to flow through the piston flow section of the aquifer. A binary EPM combines two singular EPMs and hence has five unknowns: T_1 and f_1 for the first EPM, T_2 and f_2 for

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

the second EPM, and r, the ratio of the two single EPMs in an overall system response function used to model the final water age distribution. For six of the sites considered in this study, the age tracer data could not be fitted by a single LPM, indicating that the age distribution had changed over time, and hence different age distributions were retained and treated as separate samples in the input dataset.

3.2. Symbolic Regression and Gradient Boosted Regression Models

SR models were developed using HeuristicsLab version 3.3.16.17186 (Wagner et al., 2014). SR settings allowed a maximum tree depth and length of 15 and 150, respectively, based on a multi-symbolic expression crossover with internal crossover point probability of 90%. SR grammar rules permitted arithmetic, exponential and logarithmic functions; permission of conditionals (e.g. if/then statements) was also assessed in terms of ability to improve model fits.

GBR models were developed using the GBR package that is available with the open-source scikit-learn library in Python (Pedregosa et al., 2011). The hyperparameters were tuned to find the optimal parameters (tree depth = 4, sample split = 2 and learning rate = 0.05) to assist model convergence. A stopping criterion was applied to determine the number of estimators (regression trees) required (if the model score was not improved by at least 0.01 in the last 50 iterations then the model was considered to have converged) and, in most of the cases, the models achieved their optimal solution at around 50 –75 estimators (boosting iterations).

The first stage in developing the SR and GBR models was to generate an ensemble of independent models for each of nine selected percentiles (5^{th}, 10^{th}, 20^{th}, 33^{rd}, 50^{th}, 66^{th}, 80^{th}, 90^{th}, 95^{th}) in the LPM-derived water age distributions. Hereafter these are referred to as 'unchained models' to differentiated them from the 'chained models' described below. The input dataset for the unchained models consisted of the sample-specific values for 21 hydrochemical parameters: Ca, Mg, Na, K, HCO_3, Cl, SO_4, Fe, Mn, SiO_2, NO_3-N, NH_3-N, PO_4-P, pH, EC, δ^2H and $\delta^{18}O$ (all measured in the lab), along with T and DO (measured in the field).

For both the SR and GBR methods, for each age percentile, ten data split realizations were generated by dividing the input data into testing and training subsets. One hundred repeat models were constructed for each split realization. For the SR method, each split realization was constructed through independent and random division of ten input data duplicates with a test/train split ratio of 33/66. For the GBR method, the input data was divided into ten folds in a 10-fold cross-validation procedure with a test/train split ratio of 10/90. In the cross-validation procedure, each fold was sequentially "held-out" in the testing data with the remaining nine folds comprising the training data set; this was repeated 100 times with some shuffling of the data between folds, for each repeat. Then, for both SR and GBR, from the total of 100 models produced at each split realization or fold, we selected the four best performing models. The best performing models were indicated by the highest Pearson R^2, as long as $R^2_{Training} \geq 0.7$ and $ABS(R^2_{Training} - R^2_{Testing}) \leq 0.2$, ensuring adequately and similarly high goodness of fit across the training and testing datasets, analogously to the Akaike Information Criterion (Gomes et al., 2019). We note that these criteria do not discuss model performance beyond model-to-measurement fits, and that selection of such criteria may result in biased model rankings, as discussed by Schöniger et al. (2014); however, these criteria are commonly adopted and appropriate in most contexts that data-driven modelling is used.

Overall, for both SR and GBR, this approach produced a final group of 40 independent models for each of the nine above-listed age percentiles (Fig. 4). These resulting model ensembles were summarized using the average, median, median absolute deviation (MAD), and standard deviation (SD) of the predictions for each of the nine percentiles in the age distributions. The SR and GBR methods also automatically determined the influence of each of the above-listed input variables with respect to model predictions for each age percentile, providing a quantification of the relative importance of each input variable for the prediction of groundwater age distributions.

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

Hydrology and Earth System Sciences
Discussions

The second stage in developing the SR and GBR models was to implement a chaining approach that connected the models for the unchained percentiles in the age distributions. This was done to ensure that the separately simulated percentiles had an appropriate relationship to each other, e.g., that the value for the 10th percentile in the age distribution for any sample had to be greater than or equal to the 5th percentile in the age distribution at the same sample. The implementation of the chaining approaches for the SR and GBR models varied slightly. For the SR method, independent models were first developed for each of the nine percentiles in the age distribution as described above, then the model for each individual percentile was re-modelled based on the ensemble median value from all age percentiles; for example, the chained model estimate for the 5th percentile in the age distribution was based on the unchained models for all nine percentiles. For the GBR method, the first step was to use the hydrochemical data to develop an unchained model to simulate the 5th percentile in the age distribution across all samples, as described above. Then this model for the 5th percentile in the age distribution was subsequently used as input, along with the hydrochemical data, to develop a second model to simulate the 10th percentile in the age distribution across all samples, which in turn was used in conjunction with the hydrochemical data to develop a third model to simulate the 20th percentiles across all sites, and so on. For both the SR and the GBR approaches, the chained model development followed the same split and validation procedure as were used for the development of the unchained models.

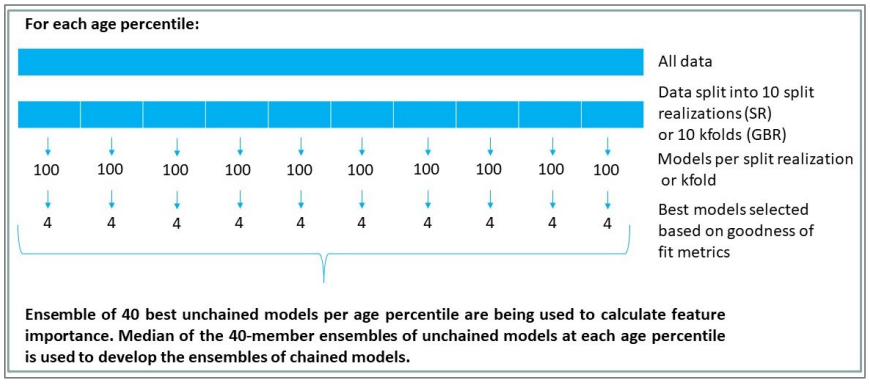


Figure 4: Schematic of workflow used for SR and GBR modelling.

4. Results and Discussion

4.1. SR and GBR Model Performance

The range in performance of the individual unchained models is illustrated for selected sites and percentiles in Fig. 5. Recall that each of the investigated percentiles was evaluated using its own suite of 40 individual models, each of which was applied to all samples in the input dataset. For all nine of the investigated percentiles and for the majority of sites, the individual unchained models produced a normal or near-normal distribution of ages with relative standard deviation of approximately 45%. However, a multi-modal distribution of age estimates was produced by model ensembles for some sites and/or percentiles, e.g. as seen with the SR for the 10th percentile at the Waipatu site (mapID 36) in Fig. 5. Moreover, for a small number of sites within the test dataset, a few of the unchained models yielded very high and inaccurate age estimates that strongly biased the ensemble average, as shown by mean ¹ median, even though the individual model's overall R^2 remained high. To avoid this biasing for the few cases where it occurred, we characterised the central tendency and width of each 40-member ensemble using the median and MAD instead of the average and SD.

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

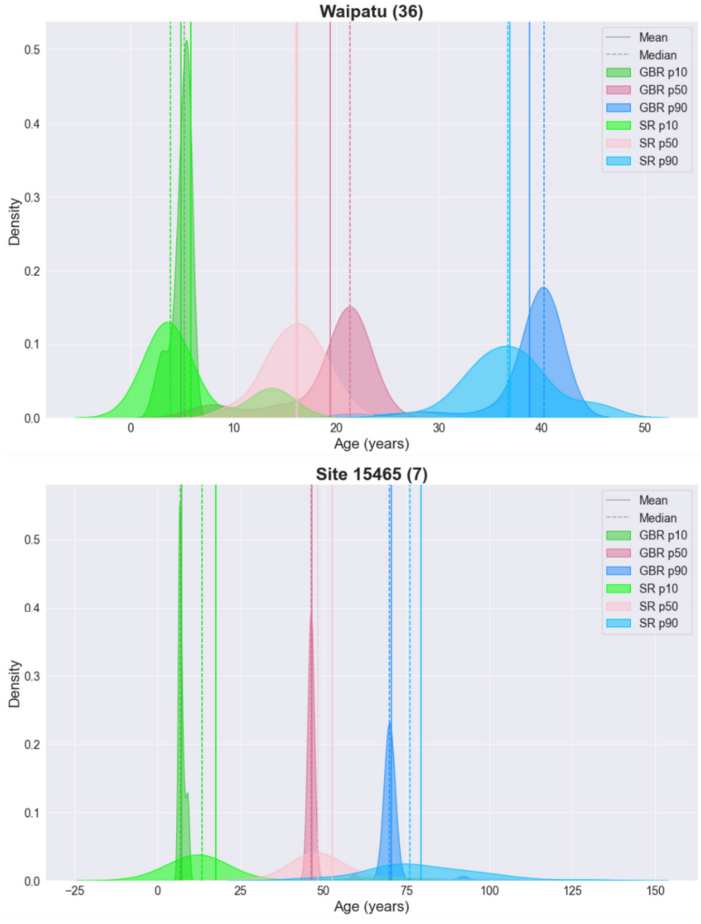


Figure 5: Range in predictions from ensembles of 40 individual unchained SR and GBR models for three different percentiles (10th, 50th and 90th) and two selected sites. Solid and dashed vertical lines indicate ensemble means and medians, respectively. MapIDs in brackets link to the location of the sites on the map in Fig. 3.

In general, the machine learning models provided good matches to the LPM age distributions (Fig. 6, Supplementary Table S1, and Supplementary figures S1-S4). Across all samples and all nine percentiles in the age distribution, the ensembles of unchained SR models had average R^2 of 0.83 and MAE of 7.5 years, and the ensembles of GBR models had average R^2 of 0.98 and MAE of 1.16 years. The chaining procedure provided improved fits for both the SR and GBR algorithms. The ensembles of chained SR models had average R^2 of 0.94 and MAE of 4.4 years, and the ensembles of GBR models had average R^2 of 0.95and MAE of 2.16 years, across all samples and all nine percentiles in the age distribution. The goodness of model fits obtained indicates that the hydrochemical dataset is well suited to estimation of the selected percentiles in the LPM-derived groundwater age distributions in this study area.

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

Hydrology and Earth System Sciences
Discussions

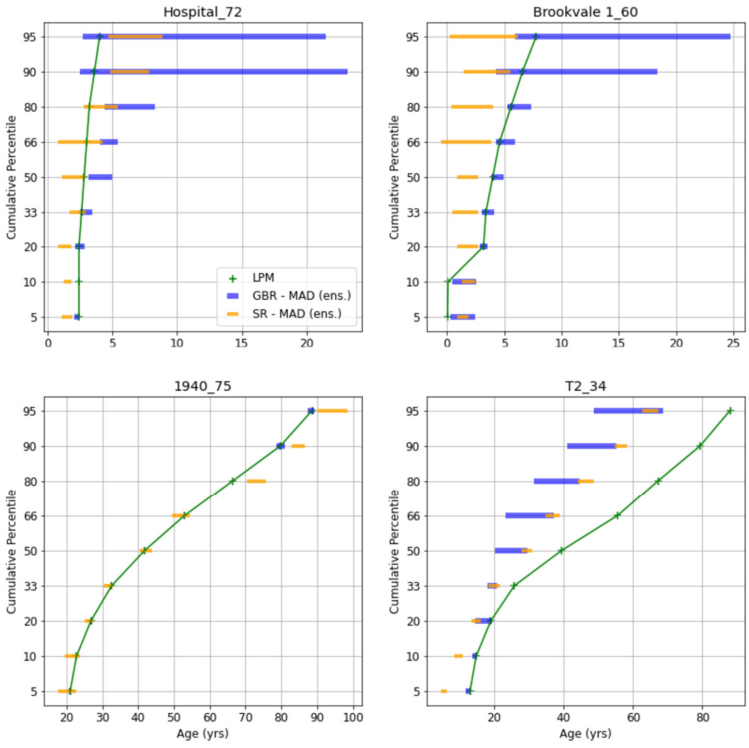


Figure 6: Age distributions for selected samples based on LPMs (Morgenstern et al. 2018) compared to the chained SR and GBR models developed in this study (bars represent ensemble MAD). The legend in the upper left sub-figure is representative for all sub-figures. The map ID following the underscore links with the location of the site on Fig. 3.

Within their overall performance, the SR and GBR models had slight variations in the goodness of fit across the nine modelled percentiles in the age distributions (Supplementary Table S1). For the unchained models, the ensemble mean R^2 values for the SR and GBR algorithms are highest for the 50th percentile and decrease slightly towards both the lowest and highest percentiles, which represent the youngest and oldest water in the age distributions, respectively. The chained models also displayed this relationship between goodness of fit and the percentile being modelled, though to a less pronounced degree than for the unchained models. This finding may reflect that the youngest and oldest age fractions are most likely to have censored hydrochemical results for certain parameters, i.e. concentrations reported as being below the analytical detection limit. For example, as discussed in Section 2.3, young groundwaters are more likely to be oxic and hence contain near- or below-detection concentrations of Fe, Mn and NH_3-N, whereas older groundwaters are more likely to be anoxic and therefore contain near- or below-detection concentrations of DO and NO_3-N (Daughney et al., 2010; Morgenstern and Daughney, 2012). Thus, the approach taken in this study of replacing all censored concentrations with their corresponding analytical detection limits may have impacted the ability of the metamodelling methods to discriminate or simulate the lowest and highest percentiles in the age distributions. Overall, the slightly poorer model fits produced by both the SR and GBR algorithms at the extremes of the age distribution suggest that caution should be exercised when using hydrochemistry-age relationships to evaluate the potential for the presence of contaminants such as pathogens, which tend to occur in the youngest age fraction of a water sample, or geogenic substances such as Fe or Mn, which are more likely to occur in oldest age fractions.

The goodness of the SR and GBR fits also varied across the study area in relation to sampling location and bore depth (Fig. 7).

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

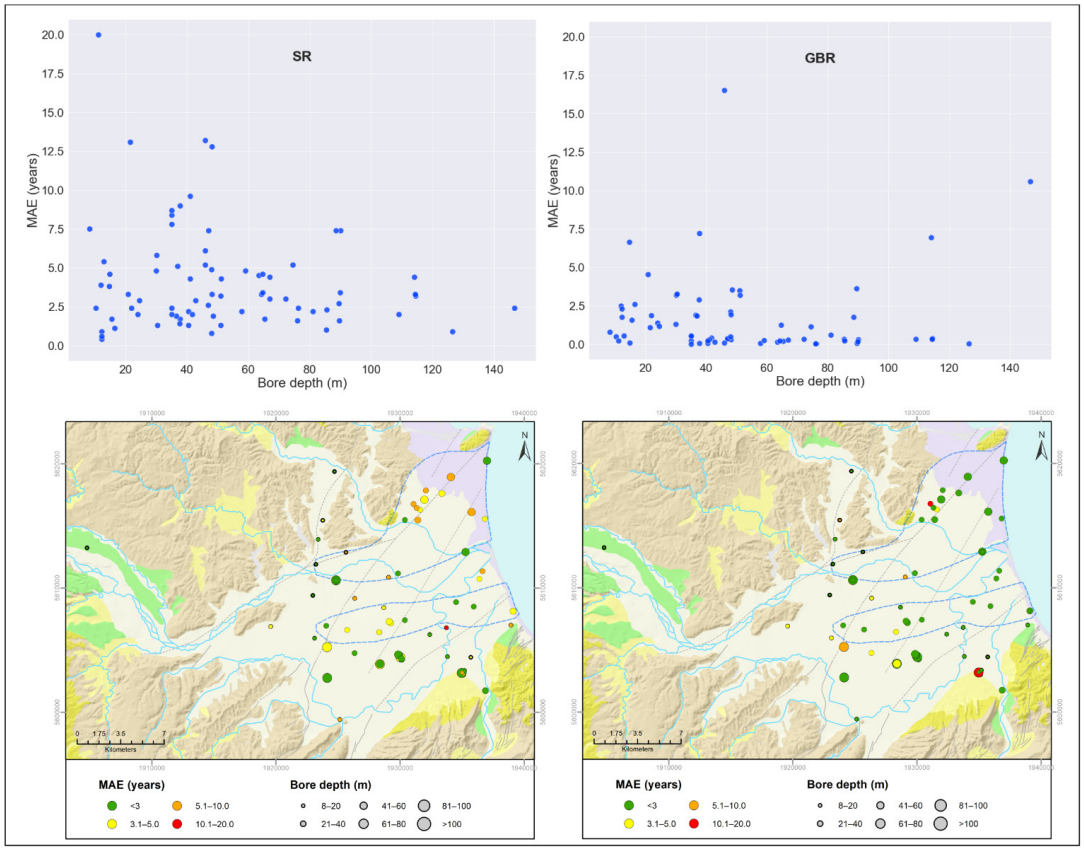


Figure 7: Ensemble average MAE for each site across all nine modelled percentiles vs. bore depth (top graphs) and site location (bottom figures) for chained SR models (left figures) and GBR models (right figures). On the bottom maps, colours represent MAE and symbol size is scaled to bore depth.

For the SR model, the poorer fits (MAE > 7 year) were confined to samples from bore depths < 70m. This relationship was not evident for GBR predictions. The model fits were also related to site location in a way that could not be entirely explained by the spatial variations in bore depth (e.g. deeper bores typically located nearer the coast). Site 4362 (mapID 4), which is located upstream of the plains shows good fits of MAE < 3 years with both SR and GBR. This is of interest as this site is the westernmost site used in this study, with the highest distance from all other sites. However, it is a shallow bore located in an area where last glacial gravel deposits have been mapped at the ground surface. No other bore is located where these deposits have been mapped on the surface, however, the last glacial gravels represent the main productive aquifer unit in the plains, and the majority of bores in this study source their water from this aquifer. These aspects of spatial and depth bias in model fit are inferred to arise from spatial variations in hydrochemistry caused by groundwater-surface water interaction and groundwater flow paths through the aquifer.

Based on the results above, the SR and GBR methods are seen to produce equivalently good fits for the application in this study. This is in agreement with studies that successfully predicted groundwater age from hydrophysical data using metamodels (Fienen et al., 2016, 2018; Starn and Belitz, 2018; Starn et al., 2021), in particular with Fienen et al. (2016), who produced comparable results predicting groundwater age using three different machine learning approaches with the same input dataset. Both, SR and GBR have their advantages and disadvantages with regard to model construction, transparency,

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

Hydrology and Earth System Sciences
Discussions

adaptability to new parameters or applications, etc. For example, SR provides an explicit model equation as output, which is more transparent and straightforward to apply in other applications but may be subject to injection of bias by the modeller's decisions for the allowable SR grammar rules. GBR is less transparent but may be less subject to the injection of modeller biases. Both SR and GBR can be adapted to new parameters and modelling settings rapidly. Performance and accuracy of the two methods are very similar, and the results show that either metamodelling method can successfully be used to predict groundwater age distributions from hydrochemistry.

4.2. Relationships between Hydrochemistry and Groundwater Age Distribution

The hydrochemical parameters with most influence on the SR and GBR models were identified by scaling the relative variable weights for each model from 0 to 1, then determining the median and MAD of these weightings within each 40-member ensemble at each of the nine modelled percentiles in the age distribution. Sensitivity analysis was undertaken with the SR models, whereby the value for a given hydrochemical variable was increased or decreased by 10% while the values for all other variables were held constant. Variable weightings and sensitivity analysis for the unchained models were similar to those of the chained models, so the following discussion focusses on the chained models.

PO_4-P (DRP), NH_3-N, DO and T (groundwater temperature) were found to be the parameters with greatest overall influence on the unchained models, having median weights across all age percentiles of 0.74, 0.55, 0.46 and 0.48 respectively, for the SR, and 0.37, 0.98, 1.7 and 0.20 for the GBR, respectively. All other hydrochemical parameters had median weights of less than 0.4 and most had median weights of less than 0.2 for the SR. Median weights for all other hydrochemical parameters for the GBR were below 0.2 (Fig. 8). Sensitivity analysis showed that, when the values of all other hydrochemical variables were held constant, a 10% increase in PO_4-P or NH_3-N resulted in a median increase in the estimated age of 3% and 1%, respectively, across all simulated age percentiles, whereas a 10% increase in DO caused a 1% decrease in the simulated age, but only for the 5th to 20th percentiles. Temperature was the most sensitive parameter, with a 10% increase in T causing the estimated age to decrease by approximately 20% across all percentiles. Sensitivity analysis showed that, for all other variables, a 10% increase in value resulted in a change of less than 1% in the estimated age for any percentile.

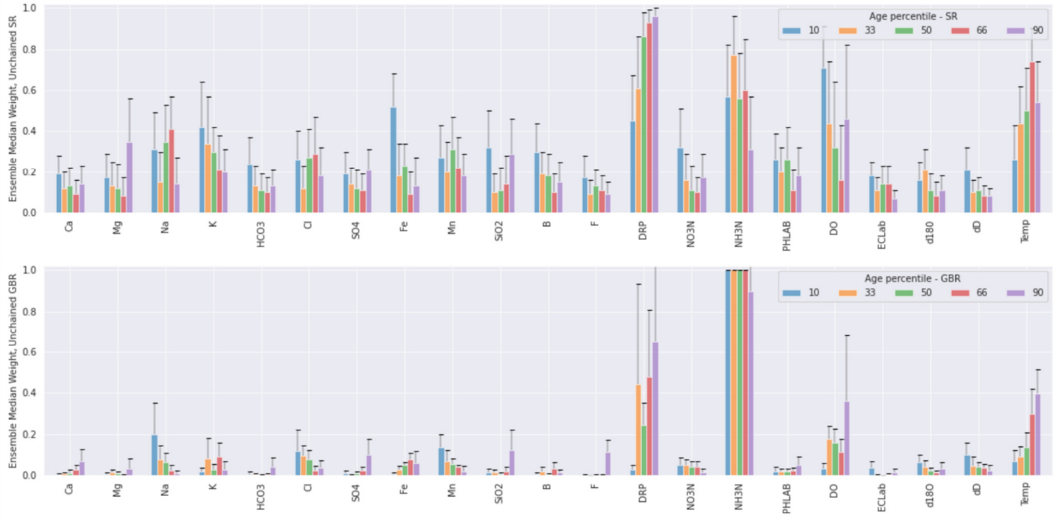


Figure 8: Ensemble median variable weights (scaled from 0 to 1) from unchained SR and GBR models, for selected percentiles in the age distribution (10th, 33rd, 50th in red, 66th and 90th, shown from left to right for each variable). Error bars represent the ensemble median absolute deviation.

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022


Hydrology and Earth System Sciences
Discussions

The importance of most hydrochemical parameters varied substantially within the individual models of each 40-member ensemble, as shown by relative MADs (MAD/median) of up to 60% (Fig. 8). This shows that, for a given age percentile, a particular hydrochemical variable could have high weighting in some models but low weightings for other models. This result likely reflects the high correlation among some hydrochemical variables (Fig. 2). Due to this correlation, if a particular variable is randomly selected for inclusion in the initiation of the SR or GBR algorithm, those variables to which it is correlated provide relatively little improvement in model predictions so tend to be excluded. For example, a model that includes K would be relatively unlikely to include Na, due to the correlation between them. However, because the SR and GBR algorithms are seeded randomly, models that include K (but likely not Na) as well as models that include Na (but likely not K) can be produced within a single 40-member ensemble – resulting in an overall higher MAD and lower median weighting for such variables across the ensemble. This interpretation is supported by the fact that the variables such as PO_4-P, NH_3-N, DO and T have relatively low correlations to other variables (Fig 2).

4.2.1. Organic matter oxidation

Consistent with a decrease in redox potential over time, the SR and GBR models identify that the concentrations of NH_3-N, Fe and Mn all tend to increase with groundwater age, whereas concentrations of DO and NO_3-N tend to decrease. These patterns are anticipated, based on observations of other New Zealand groundwater systems (Daughney et al., 2012; Morgenstern and Daughney, 2012) and the known sequence of energetics in the oxidation of organic matter in aquifers (McMahon and Chapelle, 2008). The strong positive weightings of PO_4-P in the age models (Fig. 8) are inferred to reflect its release into solution concomitant with reductive dissolution of iron and/or manganese oxide minerals (e.g. Hongve, 1997; Johnson and Loeppert, 2006). Of note, PO_4-P is retained as a predictor variable in over 90% of the individual SR models developed across all percentiles, indicating that the geochemical processes that control its concentration are omnipresent across the study area. The SR and GBR models do not identify a strong negative correlation between SO_4 concentration and groundwater age (Fig. 8), suggesting that redox potential has not declined to sulphate-reducing conditions at a sufficient number of sites for this relationship to be prevalent in the dataset. Likewise, the SR and GBR models detect only weak relationships between groundwater age and HCO_3 or pH, suggesting that the concentrations of these variables are not exclusively controlled by previously reported relationships between organic matter oxidation, alkalinity and acidity (Scott and Morgan, 1990; Sverdrup et al., 2019).

The results of this study provide semi-quantitative insights into the geochemical kinetics of organic matter oxidation in the Heretaunga Plains aquifer system, based on the rate of decline in redox potential as indicated by changing concentrations of the above-listed redox-sensitive parameters. Evaluation of the concentrations of DO, NO_3-N, Fe and Mn indicates that the oxidative capacity of the Heretaunga Plains groundwater is dominated by DO (cf. Scott and Morgan, 1990). As noted above, the models in this study indicate that a 10% increase in DO concentration corresponds to a median decrease of approximately 1% in the estimated groundwater age. Noting that different types of organic matter may oxidise at different rates (Westrich and Berner, 1984), for simplicity we assume that all organic matter in the study area is equally reactive (Middelburg, 1989) and that its oxidation is described by first-order kinetics (Tarutis Jr, 1993). With these assumptions, the models developed in this study indicate an average rate constant of k (1/y) = -0.6, which is comparable to the values reported by Westrich and Berner (1984) and Middelburg (1989), albeit for marine sediments instead of for aquifers. Greater insight into the rate of organic matter oxidation in the Heretaunga Plains could be gained if future studies make measurements of the concentrations and reactivities of dissolved and solid-phase organic matter.

4.2.2. Water-rock interaction

Aside from reductive dissolution of iron and manganese oxides discussed in the previous section, water-rock interaction is expected to increase the dissolved concentrations of mineral-forming elements such as Ca, Mg, Na, K or SiO_2 (Sverdrup et al.,

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022


Hydrology and Earth System Sciences
Discussions

2019). These five hydrochemical parameters have low median weightings in the SR and GBR models developed in this study (Fig. 8), suggesting that they are not important predictors of groundwater age at the majority of sites. However, these same five parameters were retained in close to half of all models across all percentiles, suggesting that they are important predictors of age for at least some sites.

We infer that the concentrations of Ca, Mg, Na, K and SiO_2 are retained in the SR and GBR models primarily as a means of differentiating rainfall-recharged groundwaters from river-recharged groundwaters, which then allows the algorithms to apply appropriate age estimations depending on the relevant recharge source. As noted in Section 2.3, river-recharged groundwaters typically have slightly lower concentrations of Ca, Mg, Na, K and/or SiO_2 compared to rainfall-recharged groundwaters, which results from the relatively faster accumulation of these substances during the passage of water through the soil zone, likely due to mineral dissolution (e.g. of carbonates), ion exchange and evaporation (Morgenstern et al., 2018). This inference is supported by the fact that the age estimates for the subset of rainfall-recharged sites are generally more sensitive to the concentrations of Ca, Mg, Na, K and SiO_2 used as input to the models. This inference would be usefully tested through future investigations that evaluate the respective roles of silicate mineral weathering, ion exchange and evaporation during the passage of recharge through the soil zone in the Heretaunga Plains.

4.2.3. Human impacts

The SR and GBR models do not identify strong relationships between groundwater age and any of the commonly analysed indicators of human impact on groundwater quality. In New Zealand, human impacts on groundwater quality are most readily identified by elevated concentrations of NO_3-N, sometimes co-occurring with elevated concentrations of Na, K, Mg and/or Cl (Daughney et al., 2012; Morgenstern and Daughney, 2012). That NO_3-N is not a strong predictor of groundwater age in the Heretaunga Plains likely reflects that many of the sites are recharged from rivers, which have lower NO_3-N concentrations compared to groundwaters that are recharged from rainfall (Section 2.2), and/or that the degree of impact evident in the recharge water has not changed substantially over time. Elevated concentrations of PO_4-P in New Zealand groundwater can arise from dairying land use, especially over gravel or sand aquifers (McDowell et al., 2015), but such land use is not common in the Heretaunga Plains (Smith et al., 2020), and hence PO_4-P concentrations are instead inferred to reflect geogenic origin (Section 4.2.1). Concentrations of pesticides, emerging contaminants or microbial pathogens can also reveal human impact on groundwater quality but were not analysed in this study.

4.2.4. Temperature

The SR and GBR models reveal a strong inverse relationship between T and estimated groundwater age, i.e. as T increases, the modelled groundwater age decreases. Particularly for the higher age percentiles, T is among the variables with the highest median weightings across the model ensembles (Fig. 8).

The relationship between T, reaction rates and estimated groundwater age can be semi-quantitatively evaluated using the Arrhenius expression (Eq. 4):

$$\frac{d \log k}{dT} = \frac{E_a}{2.303RT^2} \quad (4)$$

where k is the rate constant, E_a is the activation energy, R is the gas constant, and T is expressed in the Kelvin scale. For the comparison of reaction rates k_1 and k_2 at two temperatures T_1 and T_2, the above expression can be arranged as Eq. (5):

$$\log \frac{k_1}{k_2} = \frac{E_a}{2.303R}\left[\frac{1}{T_2} - \frac{1}{T_1}\right] \quad (5)$$

The above equation suggests that the average E_a, aggregated across all reaction types, is approximately 25 kcal/mol, derived from T_1 = 15°C = 288K (the median across all samples), T_2 = 16.5°C (a 10% increase above the median) = 289.5K, and k_1/k_2

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022


Hydrology and Earth System Sciences
Discussions

= 0.8 (because the models developed in this study indicate that a 10% increase in T causes a median decrease of approximately 20% in the estimated age). This estimated value for E_a is in the range expected for mineral dissolution reactions (8-36 kcal/mol) and ion exchange (>20 kcal/mol) (Lasaga, 2018) and for organic matter decomposition (ca. 20-30 kcal/mol) (Leifeld and von Lützow, 2014). The correspondence between these previously published values of E_a and the estimate derived in this study suggests that the effect of T on modelled groundwater age may indeed be driven by increases in the rates of reactions such as organic matter oxidation and water-rock interaction as discussed above. Accordingly, we surmise that T is retained in the models as an important modifier of the effects of such reactions on hydrochemistry. However, further research is required to rigorously test this possibility.

We acknowledge that T may affect the estimated groundwater ages through hydrologic factors rather than geochemical kinetics as described above. For example, one possibility is that there is a significant difference in the temperature of slower-moving groundwaters that are recharged from rainfall compared to faster-moving groundwaters that are recharged primarily from river seepage (see Fig. 1). This hypothesis is not supported by the measured values of T, which show no significant differences arising from the inferred groundwater recharge source. Moreover, it is probable that river-recharged groundwaters, being sourced from higher altitude precipitation, would be cooler than rainfall-recharged groundwaters, which would lead to a relationship between temperature and age that is opposite to observed in this study. However, this study is limited by a relatively small number of samples, so further investigation with collection of samples across a wider range of seasons and recharge conditions would be beneficial to elucidate any hydrological controls on the observed relationships between T and groundwater age.

4.3. Applications

4.3.1. Estimation of groundwater age distributions without age tracer data

The SR and GBR models developed in this study can be used to estimate the values for the nine specific percentiles in the age distributions solely on the basis of groundwater chemistry, i.e. for sites and samples for which age tracer data are not available. To illustrate this application, we make use of data collected through the Hawke's Bay Regional Council groundwater quality monitoring programme (Supplementary Material Table S3). These samples are collected using the same protocols and analysed with very similar procedures for the same variables as described in Section 3.1, so are considered suitable for use with the SR and GBR models developed in this study. The only exception is for B, F, δ^2H and $\delta^{18}O$, which are not routinely measured by the regional council; however, these four variables all have low influence in the SR and GBR models (Fig. 8), and so we conclude that their exclusion from the input dataset has little influence for this application.

The spatial variations in groundwater age for sites without available age tracer data are shown in (Fig. 3). In some areas, the SR and GBR estimates provide infilling of modelled age in between the locations where LPMs are currently available. For example, the SR and GBR age estimates can improve the understanding of the demarcation between the zones of younger river-recharged groundwater in contrast to older rainfall-recharged groundwaters (Fig. 1). The SR and GBR models can also be used for spatial extrapolation. For example, the SR and GBR models can estimate groundwater age quite distant from the nearest sites having available LPMs, thereby providing useful information for groundwater management where such information was previously lacking.

The temporal variations in groundwater age can also be assessed using the SR and GBR models at sites for which time series groundwater chemistry data are available (Fig. 9). Application in this study suggests that the temporal shifts in the groundwater age distribution at a single site can be substantially larger than the ensemble MAD for a single sampling date. This suggests that temporal or seasonal variations in the groundwater age distribution can be reasonably large at some sites. This inference is supported by the age tracer results for the few sites that had been sampled on more than one occasion. For these sites, there were cases where the LPM age distributions were inferred to vary temporally or seasonally, based on observed shifts in the

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

concentrations of the age tracers; these sites also displayed temporally variable hydrochemistry. Thus, it is reasonable to anticipate that seasonal or longer-term variations in recharge and/or abstraction on groundwater flows that are known to affect groundwater age distributions (Engdahl et al., 2016; Toews et al., 2016; Yang et al., 2018) may be indicated by temporal shifts in groundwater chemistry. The SR and GBR models developed in this study indicate such shifts in groundwater age distribution. Moreover, different percentiles in the age distribution at a single site can display quite different temporal patterns, as shown by a relatively constant 50th percentile but more variable 10th percentile at Site 413 (mapID 20), or the opposite pattern at Site 611 (mapID 77) (Fig. 9).

We acknowledge that the SR and GBR models developed in this study were based primarily on samples collected in the period April to June and the years 2014-2019 (Section 3.1), hence caution must be exercised for their application to other seasons or time periods. We also acknowledge that the meta-models described in this paper are specific to the training region. While the same meta modelling approaches may be used elsewhere where there is sufficient groundwater chemistry data, the same meta model hyperparameters are unlikely to apply in other regions (Doherty and Moore, 2021). Therefore, age tracer training data sets would be required also for other regions. However, the metamodeling approach enables a space-for-time substitution while preserving the key processes that relate groundwater chemistry to groundwater age. We therefore conclude that the SR and GBR models can offer useful insights to spatial and temporal patterns in groundwater age distribution based on chemistry and therefore assist sustainable groundwater management if age tracer data are not available.

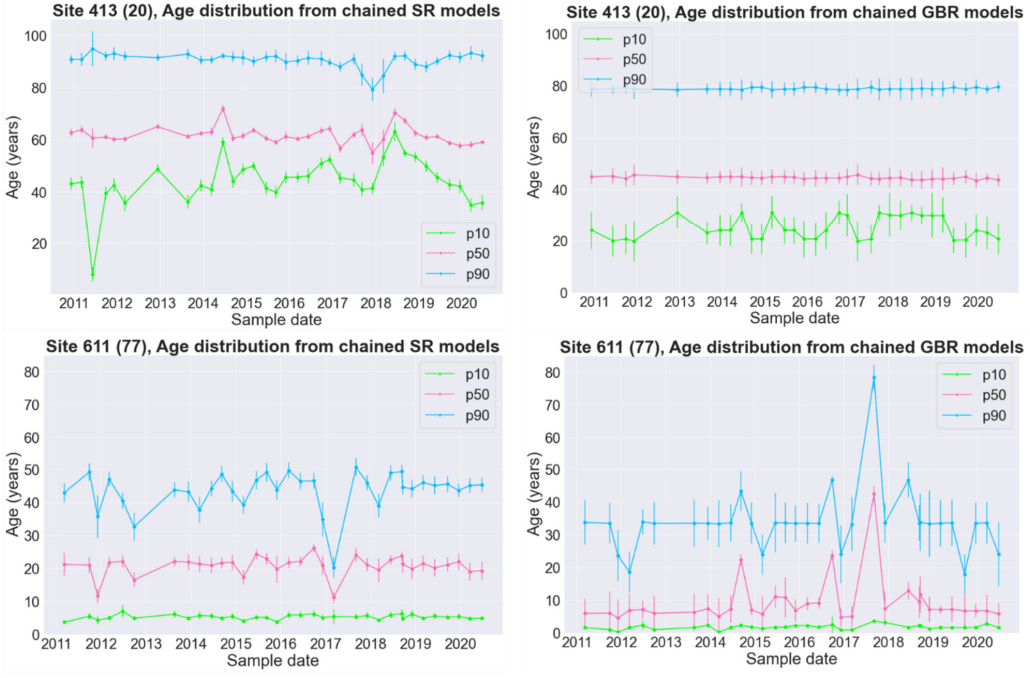


Figure 9: Temporal patterns in the estimated values for selected age percentiles at two sites based on SR models (left) and GBR models (right). Points and error bars represent the median and MAD for the relevant 40-member model ensemble, respectively. MapIDs in brackets link to the location of the sites on the map in Fig. 3.

4.3.2. Constraining LPMs

The approach taken in this study is to treat the LPM-based age distribution as truth, which the SR and GBR models are subsequently developed to reproduce. The SR and GBR models are therefore acting to generalise the relationships between

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

chemistry and the LPM age percentiles based on a set of independent samples from a range of sites. That the SR and GBR models perform well across the Heretaunga Plains dataset indicates that there are generalisable relationships between groundwater chemistry and LPM age. Thus, samples for which the SR and GBR models perform poorly may indicate that the LPM age distribution is in fact in error and could be better constrained if outputs from the SR and GBR models were taken into account.

The SR and GBR models may assist the choice of mixing model to be used when fitting an LPM to age tracer data. In the absence of any other information, the general practice is to select the simplest LPM mixing model that provides an adequate fit to the available age tracer data, in accordance with the principle of parsimony. Thus, an EPM (having just two optimisable parameters) is typically the mixing model applied for sites that have been sampled for age tracers on only one occasion, because a more complex BMM (with five optimisable parameters) would be under-constrained and therefore unjustifiably complex. However, the age distribution inferred from groundwater chemistry may indicate sites for which a BMM is more appropriate, even if the available age tracer data can be adequately fitted by an EPM. This case is illustrated for Site 1940 (mapID 75), which has only been sampled for age tracers on one occasion, but for which the SR and GBR models imply a more complex age distribution more consistent with a BMM than an EPM (Fig. 6). While we recommend that the fitting of the LPM should be based on the age tracer data, we suggest that the SR and GBR models can usefully guide which LPM mixing model(s) may be appropriate.

The SR and GBR models may also help to constrain the LPM parameter values, in particular in the Northern Hemisphere, where tritium age interpretations are still ambiguous. Depending on groundwater age, and time and frequency of age tracer sampling, for some sites, the age distribution parameters can robustly be constrained. These could be used for training the meta models which could then be used to help estimate the parameters for the sites where age tracers are insufficient to constrain all model parameters. Specific applications could also be to use the metamodels to constrain the second LPM model parameter where only one age tracer measurement in time is available, and to demonstrate variable ages in a well where lots of hydrochemistry data is available (e.g. quarterly monitoring) but only few age tracer samples.

5. Conclusions and Future Work

Overall, this study has shown that SR and GBR are useful approaches for codifying the relationships between hydrochemistry and groundwater age at the aquifer scale. This finding is consistent with previous studies that have identified statistical or first-principles relationships between groundwater chemistry and age (e.g. Daughney et al., 2012; Morgenstern et al., 2012; Beyer et al., 2016, Sverdrup et al., 2019). The key advance in this study is to extend these hydrochemistry-age relationships to specific percentiles in the age distribution, thereby providing greater insights for groundwater management, such as the potential for occurrence of young or old groundwater fractions that may be associated with specific types of contaminants.

We identify three avenues for extension of this study. First, the SR and GBR approaches could be extended to the estimations of groundwater age distributions from hydrochemistry in other aquifers, or even at the national and international scales. Such work would serve to identify, or otherwise, the universality and transferability of the age-hydrochemistry relationships and the models that encode them. A second opportunity for further work is to apply the age-hydrochemistry relationships to improve or calibrate kinetics models of water-rock interaction or biogeochemical processes (e.g. Sverdrup et al., 2019). A third opportunity for further work is to apply the age-hydrochemistry relationships in real-world groundwater management contexts. This could include assessments of the security of groundwater supplies in a changing climate, the occurrence and transport of contaminants, and the moderation of groundwater abstraction regimes.

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

Hydrology and Earth System Sciences
Discussions

Data availability

595 The datasets relevant to this study are available from the corresponding author on request.

Author contribution

Conceptualisation and funding acquisition: CM, UM, CD. Supervision: BH, CM. Data curation: CT, CD, SK. Investigation & methodology: CD, SK, CT, BH. Writing and visualisations original draft: CT, CD, SK. Writing review & editing: CT, CD,
600 SK, BH, CM, UM

Competing interests

At the time of writing, Chris Daughney held a half-time secondment as Chief Science Advisor for Te Uru Kahika Regional and Unitary Councils Aotearoa. The other authors declare that they have no competing interests.

605

Acknowledgements

This paper represents a collaboration between NIWA and GNS Science, in a GNS-lead programme called Te Whakaheke o te Wai, funded by the New Zealand Ministry of Business Innovation and Employment (MBIE, grant number C05X1803). Co-funding was provided by GNS′ Groundwater research programme (MBIE, Strategic Science Investment Fund, grant number
610 C05X1702). The authors would also like to thank Hawke's Bay Regional Council, and Hastings and Napier district councils for providing data. Thanks also go to the Te Whakaheke o te Wai project partners and Science Advisory Panel for advice and support.

References

615 Asher, M. J., Croke, B. F. W., Jakeman, A. J., and Peeters, L. J. M.: A review of surrogate models and their application to groundwater modeling, Water Resour. Res., 51, 5957–5973, https://doi.org/10.1002/2015WR016967, 2015.

Beanland, S., Melhuish, A., Nicol, A., and Ravens, J.: Structure and deformational history of the inner forearc region, Hikurangi subduction margin, New Zealand., New Zealand J. Geol. Geophys., 41, 325–342, https://doi.org/10.1080/00288306.1998.9514814, 1998.

620 Begg, J. G., Jones, K. E., Lee, J. M., and Tschritter, C.: 3D geological model of the Napier-Hastings urban area, GNS Science, Lower Hutt, N.Z., 30 pp., 2022.

Beyer, M., Jackson, B., Daughney, C., Morgenstern, U., and Norton, K.: Use of hydrochemistry as a standalone and complementary groundwater age tracer, J. Hydrol., 543, 127–144, https://doi.org/10.1016/j.jhydrol.2016.05.062, 2016.

Busenberg, E. and Plummer, L. N.: Use of chlorofluorocarbons (CCl3F and CCl2F2) as hydrologic tracers and age-dating
625 tools: The alluvium and terrace system of central Oklahoma, Water Resour. Res., 28, 2257–2283, 1992.

Cornaton, F. J.: Transient water age distributions in environmental flow systems: The time-marching Laplace transform solution technique, Water Resour. Res. 48, W03524, 2012.

Cunnold, D. M., Weiss, R. F., Prinn, R. G., Hartley, D., Simmonds, P. G., Fraser, P. J., Miller, B., Alyea, F. N., and Porter, L.: GAGE/AGAGE measurements indicating reductions in global emissions of CCl3F and CCl2F2 in 1992–1994, J. Geophys.
630 Res. Atmos., 102, 1259–1269, 1997.

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

Daughney, C. J., Jones, A., Baker, T., Hanson, C., Davidson, P., Reeves, R. R., Zemansky, G. M., and Thomson, M.: A national protocol for state of the environment groundwater sampling in New Zealand, Institute of Geological & Nuclear Sciences, Wellington, N.Z., 2007.

Daughney, C. J., Morgenstern, U., van der Raaij, R., and Reeves, R. R.: Discriminant analysis for estimation of groundwater age from hydrochemistry and well construction: application to New Zealand aquifers, Hydrogeol J, 18, 417–428, https://doi.org/10.1007/s10040-009-0479-2, 2010.

Daughney, C. J., Raiber, M., Moreau-Fournier, M., Morgenstern, U., and van der Raaij, R.: Use of hierarchical cluster analysis to assess the representativeness of a baseline groundwater quality monitoring network: comparison of New Zealand's national and regional groundwater monitoring programs, Hydrogeol J, 20, 185–200, https://doi.org/10.1007/s10040-011-0786-2, 2012.

Doherty, J. and Moore, C.: Decision Support Modelling Viewed through the Lens of Model Complexity, National Centre for Groundwater Research and Training, Flinders University, South Australia, 91 pp., 2021.

Dravid, P. N. and Brown, L. J.: Heretaunga Plains Groundwater Study. Volume 1: Findings, Hawke's Bay Regional Council, Napier, 278 pp., 1997.

Edmunds, W. M. and Smedley, P. L.: Residence time indicators in groundwater: the East Midlands Triassic sandstone aquifer, Applied Geochemistry, 15, 737–752, https://doi.org/10.1016/S0883-2927(99)00079-7, 2000.

Engdahl, N. B.: Transient effects on confined groundwater age distributions: Considering the necessity of time-dependent simulations, Water Resour. Res. 53, 7332–7348, 2017,

Engdahl, N. B., McCallum, J. L., and Massoudieh, A.: Transient age distributions in subsurface hydrologic systems, J. Hydrol., 543, 88–100, 2016.

Fienen, M. N., Nolan, B. T., Feinstein, D. T., and Starn, J. J.: Metamodels to bridge the gap between modeling and decision support, USGS Pub. Res., 2, 2015.

Fienen, M. N., Nolan, B. T., and Feinstein, D. T.: Evaluating the sources of water to wells: Three techniques for metamodeling of a groundwater flow model, Environ. Model. Softw., 77, 95–107, https://doi.org/10.1016/j.envsoft.2015.11.023, 2016.

Fienen, M. N., Nolan, B. T., Kauffman, L. J., and Feinstein, D. T.: Metamodeling for Groundwater Age Forecasting in the Lake Michigan Basin, Water Resour. Res., 54, 4750–4766, https://doi.org/10.1029/2017WR022387, 2018.

Ginn, T. R., Haeri, H, Massoudieh, A., and Foglia, L.: Notes on groundwater age in forward and inverse modeling, Transp. Porous Med., 79, 117–134, 2009.

Gomes, F. M., Pereira, F. M., Silva, A. F., and Silva, M. B.: Multiple response optimization: Analysis of genetic programming for symbolic regression and assessment of desirability functions, Knowledge-Based Sys., 179, 21–33, 2019.

Heaton, T. H. E. and Vogel, J. C.: "Excess air" in groundwater, J. Hydrol., 50, 201–216, 1981.

Heron, D. W. (Ed.): Geological Map of New Zealand 1:250 000, GNS Science, Lower Hutt, 2020.

Hongve, D.: Cycling of iron, manganese, and phosphate in a meromictic lake, Limnol. Oceanogr., 42, 635–647, 1997.

Johnson, S. E. and Loeppert, R. H.: Role of organic acids in phosphate mobilization from iron oxide, Soil Sci Soc Am J, 70, 222–234, 2006.

Jurgens, B. C., Boehlke, J. K., and Eberts, S. M.: TracerLPM (Version 1): An Excel® Workbook for Interpreting Groundwater Age Distributions from Environmental Tracer Data Dispersion, 2012.

Knowling, M.J., White, J.T., Moore, C.R, Rakowski, P., Hayley, K. On the assimilation of environmental tracer observations for model-based decision support. Hydrol. Earth Syst. Sci., 24, 1677–1689, 2020.

Koch, J., Berger, H., Henriksen, H. J., and Sonnenborg, T. O.: Modelling of the shallow water table at high spatial resolution using random forests, Hydrol. Earth Syst. Sci., 23, 4603–4619, 2019.

Langmuir, D.: Aqueous environmental Geochemistry, 600, 1997.

Lasaga, A. C.: Transition state theory, in: Kinetics of geochemical processes, De Gruyter, 135–170, 2018.

Lee, J. M., Bland, K. J., Townsend, D. B., and Kamp, P. J. J.: Geology of the Hawke's Bay area, Institute of Geological and Nuclear Sciences 1: 250,000 geological map 8.1 sheet+ 93 p, 2011.

Lee, J. M., Tschritter, C., and Begg, J. G.: A 3D geological model of the greater Heretaunga/Ahuriri Groundwater Management Zone, Hawke's Bay, 31 pp., 2014.

Lee, J. M., Begg, J. G., and Bland, K. J.: Geological map of the Napier-Hastings urban area, GNS Science, Lower Hutt, N.Z., 2020.

Leifeld, J. and von Lützow, M.: Chemical and microbial activation energies of soil organic matter decomposition, Biol. Fertil. Soils, 50, 147–153, 2014.

LINZ: Topo250 map series, LINZ Data Service, https://data.linz.govt.nz, Last accessed 25/08/2022.

Maiss, M. and Brenninkmeijer, C. A.: Atmospheric SF6: trends, sources, and prospects, Environ. Sci. Technol., 32, 3077–3086, 1998.

Maloszewski, P. and Zuber, A.: Lumped parameter models for the interpretation of environmental tracer data, in: Manual on mathmatical models in isotope hydrogeology, edited by: International Atomic Energy Agency, Vienna, Austria, 9–58, 1996.

Marçais, J., Gauvain, A., Labasque, T., Abbott, B. W., Pinay, G., Aquilina, L., Chabaux, F., Viville, D., and de Dreuzy, J.-R.: Dating groundwater with dissolved silica and CFC concentrations in crystalline aquifers, Sci. Total Environ., 636, 260–272, https://doi.org/10.1016/j.scitotenv.2018.04.196, 2018.

Massoudieh, A., Sharifi, S., and Solomon, K. D.: Bayesian evaluation of groundwater age distribution using radioactive tracers and anthropogenic chemicals, Water Resour. Res., 48, W09529, doi:10.1029/2012WR011815, 2012.

McDowell, R. W., Cox, N., Daughney, C. J., Wheeler, D., and Moreau, M.: A national assessment of the potential linkage between soil, and surface and groundwater concentrations of phosphorus, J. Am. Water Resour. Assoc., 51, 992–1002, 2015.

McMahon, P. B. and Chapelle, F. H.: Redox processes and water quality of selected principal aquifer systems, Groundwater, 46, 259–271, 2008.

Middelburg, J. J.: A simple rate model for organic matter decomposition in marine sediments, Geochim. Cosmochim. Acta, 53, 1577–1581, 1989.

Morgenstern, U. and Daughney, C. J.: Groundwater age for identification of baseline groundwater quality and impacts of land-use intensification – The National Groundwater Monitoring Programme of New Zealand, J. Hydrol., 456–457, 79–93, https://doi.org/10.1016/j.jhydrol.2012.06.010, 2012.

Morgenstern, U. and Taylor, C. B.: Ultra low-level tritium measurement using electrolytic enrichment and LSC, Isotopes Environ Health Stud, 45, 96–117, 2009.

Morgenstern, U., Daughney, C. J., Leonard, G., Gordon, D., Donath, F. M., and Reeves, R.: Using groundwater age and hydrochemistry to understand sources and dynamics of nutrient contamination through the catchment into Lake Rotorua, New Zealand, Hydrol Earth Syst Sci, 19, 803–822, https://doi.org/10.5194/hess-19-803-2015, 2015.

Morgenstern, U., Begg, J. G., Van der Raaij, R., Moreau, M., Martindale, H., Daughney, C. J., Franzblau, R. E., Stewart, M. K., Knowling, M. J., Toews, M. W., Trompetter, V., Kaiser, J., and Gordon, D. A.: Heretaunga Plains aquifers: groundwater dynamics, source and hydrochemical processes as inferred from age, chemistry, and stable isotope tracer data, GNS Science, 92 pp., doi:10.21420/G2Q92G, 2018.

Nolan, B. T., Green, C. T., Juckem, P. F., Liao, L., and Reddy, J. E.: Metamodeling and mapping of nitrate flux in the unsaturated zone and groundwater, Wisconsin, USA, J. Hydrol., 559, 428–441, https://doi.org/10.1016/j.jhydrol.2018.02.029, 2018.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V.: Scikit-learn: Machine learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.

Portniaguine, O., and Solomon, D. K.: Parameter estimation using groundwater age and head data, Water Resour. Res. 34, 637–645, 1998.

Prinn, R. G., Weiss, R. F., Fraser, P. J., Simmonds, P. G., Cunnold, D. M., Alyea, F. N., O'doherty, S., Salameh, P., Miller, B. R., and Huang, J.: A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE, J. Geophys. Res. Atmos., 105, 17751–17792, 2000.

Rakowski, P.: Heretaunga Plains aquifer model: Scenarios report, Napier (NZ), 2018.

Rakowski, P. and Knowling, M. J.: Heretaunga aquifer groundwater model: development report, 182 pp., 2018.

Razavi, S., Tolson, B. A., and Burn, D. H.: Review of surrogate modeling in water resources, Water Resour. Res., 48, 2012.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., and Carvalhais, N.: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, 2019.

Schöniger, A., Wöhling, T., Samaniego, L., and Nowak, W.: Model selection on solid ground: Rigorous comparison of nine ways to evaluate Bayesian model evidence, Water Resour. Res., 50, 9484–9513, 2014.

Scott, M. J. and Morgan, J. J.: Energetics and conservative properties of redox systems, in: Chemical modelling of aqueous systems II, edited by: Melchior, D. C. and Basset, R.L., ACS Publications, Washington DC, 368–378, 1990.

Smith, J., Barber, J., Ellmers, J., Fake, D., Haidekker, S., Hicks, A., Kozyniak, K., Lynch, B., Maradasz-Smith, A., Norris, T., Rushworth, G., Sessions, L., Shanahan, B., Waldron, R., and Wilding, T. K.: Our Hawke's Bay Environment: Key issues report 2013-2018, Napier (NZ), 137 pp., 2020.

Starn, J. J. and Belitz, K.: Regionalization of Groundwater Residence Time Using Metamodeling, Water Resour. Res., 54, 6357–6373, https://doi.org/10.1029/2017WR021531, 2018.

Starn, J. J., Kauffman, L. J., Carlson, C. S., Reddy, J. E., and Fienen, M. N.: Three-dimensional distribution of groundwater residence time metrics in the glaciated United States using metamodels trained on general numerical simulation models, Water Resour. Res., 57, e2020WR027335, 2021.

Stewart, M. and Morgenstern, U.: Age and source of groundwater from isotope tracers., in: Groundwaters of New Zealand, edited by: Rosen, M. R. and White, P. A., New Zealand Hydrological Society, 161–183, 2001.

Stewart, M. K., Morgenstern, U., and Cartwright, I.: Comment on "A comparison of catchment travel times and storage deduced from deuterium and tritium tracers using StorAge Selection functions" by Rodriguez et al. (2021), Hydrol. Earth Syst. Sci., 25, 6333–6338, https://doi.org/10.5194/hess-25-6333-2021, 2021.

Suckow, A.: The Age of Groundwater – definitions, models and why we do not need this term, Appl. Geochemistry, 50, https://doi.org/10.1016/j.apgeochem.2014.04.016, 2014.

Sverdrup, H. U., Oelkers, E., Erlandsson Lampa, M., Belyazid, S., Kurz, D., and Akselsson, C.: Reviews and syntheses: Weathering of silicate minerals in soils and watersheds: Parameterization of the weathering kinetics module in the PROFILE and ForSAFE models, Biogeosci. Discuss., 58, https://doi.org/10.5194/bg-2019-38, 2019.

Tarutis Jr, W. J.: On the equivalence of the power and reactive continuum models of organic matter diagenesis, Geochim. Cosmochim. Acta, 57, 1349–1350, 1993.

Tesoriero, A. J. and Puckett, L. J.: O2 reduction and denitrification rates in shallow aquifers, Water Resour. Res., 47, https://doi.org/10.1029/2011WR010471, 2011.

Thompson, T. M., Butler, J. H., Daube, B. C., Dutton, G. S., Elkins, J. W., Hall, B. D., Hurst, D. F., King, D. B., Kline, E. S., and Lafleur, B. G.: Halocarbons and other atmospheric trace species, Climate Monitoring and Diagnostics Laboratory Summary Report No. 27, 5, 115–135, 2004.

Toews, M. W., Daughney, C. J., Cornaton, F. J., Morgenstern, U., Evison, R. D., Jackson, B. M., Petrus, K., and Mzila, D.: Numerical simulation of transient groundwater age distributions assisting land and water management in the Middle Wairarapa Valley, New Zealand, Water Resour. Res., 52, 9430–9451, 2016.

Van der Raaij, R.: Age-dating of New Zealand groundwaters using sulphur hexafluoride, Victoria University of Wellington, Wellington, New Zealand, 2003.

https://doi.org/10.5194/hess-2022-258
Preprint. Discussion started: 31 August 2022

Wagner, S., Kronberger, G., Beham, A., Kommenda, M., Scheibenpflug, A., Pitzer, E., Vonolfen, S., Kofler, M., Winkler, S., and Dorfer, V.: Architecture and design of the HeuristicLab optimization environment, in: Advanced methods and applications in computational intelligence, edited by: Klembous, R., Nikodem, J., Jacak, W., and Chaczko, Z., Springer, Heidelberg, 197–261, 2014.

Waldron, R., Kozyniak, K., and Wilding, T. K.: Hawke's Bay rainfall and river flow 2013-2018 state of the environment technical report, Napier (NZ), 45 pp., 2019.

Weissmann, G. S., Zhang, Y., LaBolle, E. M., and Fogg, G. E.: Dispersion of groundwater age in an alluvial aquifer system, Water Resour. Res., 38, 16-1-16–13, https://doi.org/10.1029/2001WR000907, 2002.

Westrich, J. T. and Berner, R. A.: The role of sedimentary organic matter in bacterial sulfate reduction: The G model tested 1, Limnol. Oceanogr., 29, 236–249, 1984.

Wilding, T. K.: Heretaunga Springs: Gains and losses of stream flow to groundwater on the Heretaunga Plains, 92 pp., 2018.

Yang, J., Heidbüchel, I., Musolff, A., Reinstorf, F., and Fleckenstein, J. H.: Exploring the dynamics of transit times and subsurface mixing in a small agricultural catchment, Water Resour. Res., 54, 2317–2335, 2018.

Zhu, C.: Estimate of recharge from radiocarbon dating of groundwater and numerical flow and transport modeling, Water Resour. Res., 36, 2607–2620.