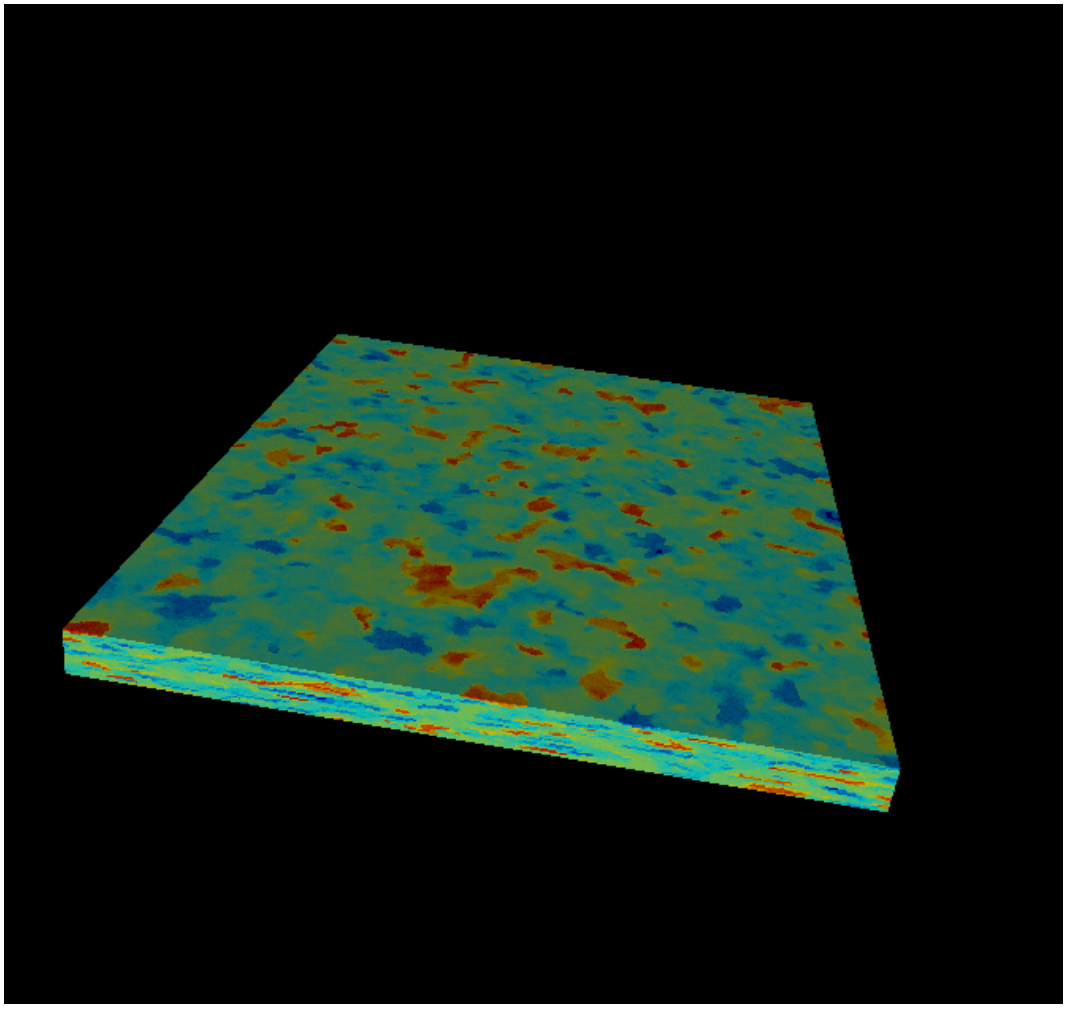
Quantifying Groundwater Ages in Heterogeneous Environments



James L. McCallum

BSc. EnvSci. (Honours).

As a requirement in full for the degree of Doctor of Philosophy in the School of the Environment, Flinders University, South Australia

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

James L. McCallum

Co-authorship

James McCallum is the primary author on all manuscripts in this thesis. On all submitted papers, the co-authors provided intellectual supervision and editorial content.

Acknowledgments

I would like to acknowledge the funding from the Australian Research Council and the National Water Commission via the National Centre for Groundwater Research and Training. Without this funding the work undertaken in this thesis would not have been possible.

I would also to like to thank my supervisor Craig Simmons and my co-supervisor Peter Cook, for their contributions to scientific ideas and editorial comment. Also to my other co-supervisor Adrian Werner for editorial assistance. I would also like to thank Nick Engdahl, Tim Ginn and Graham Fogg (and members of the Fogg/McKay lab) for hosting me at UC Davis for 2 months of my PhD. This time, whilst brief, contributed greatly to my understanding and gave me new insights into areas I had previously not considered. Also thanks to Peter Cook, Craig Simmons, Adrian Werner, Daan Herckenrath, Nick Engdahl and Tim Ginn for their scientific and editorial contribution to the manuscripts presented in this thesis. Additionally, I would like to thank editors and reviewers of manuscripts where feedback has been received.

On a personal level, thanks to my wife Nicole for her support (and editing) over my PhD. You have helped to make the last few years a success. I would also like to thank Cameron, Saskia and Dylan for making uni an enjoyable place to be.

Summary

The age of groundwater is of great interest as it infers the timescales of groundwater flow and solute transport. As such, a number of methods exist for determining the age of groundwater in simple groundwater systems where flow and transport properties are constant in space. These methods include the use of naturally occurring and anthropogenic compounds (from this point collectively referred to as environmental tracers) to estimate groundwater age, and the use of numerical simulation techniques. The applicability of these methods outside of these contexts have been demonstrated to be limited. Hence, there is a need to assess how these methods may be applied in non-ideal contexts.

This body of work addresses the quantification of groundwater ages in complex environments where flow and transport properties are highly spatially variable. Specifically, this work investigates: (1) The bias of traditional "apparent age" estimates in heterogeneous environments. (2) The accuracy of correction schemes to correct for errors in age estimates encountered in complex environments. (3) How the choice of a geostatistical model can impact on the estimates of numerically simulated groundwater age distributions in heterogeneous environments. (4) A new method for estimating groundwater age distributions without the assumption of a prior model.

In the first part of this research, we used numerical simulations of synthetic two-dimensional aquifers to investigate the bias of age dating techniques. We simulated synthetic aquifers by varying both the range of hydraulic conductivity values and the structure of the hydraulic conductivity fields. Numerical flow simulations were undertaken and environmental tracer concentrations were simulated. We hypothesised that errors in apparent ages may behave as a correctable bias rather than as a random error. These biases are due to non-linear temporal variations in the concentrations of the compounds used for dating, and the mixing arising from the variations of flow-paths in these systems. The findings of this study suggest that using multiple tracers with differing biases may allow for mean ages to be determined.

The second part of this we assessed the accuracy of techniques that account for biases. These techniques include the use of correction factors based on aquifer structures, the use of mean transit times derived from lumped parameter models, and techniques using ages from multiple tracers. The study was also implemented in a numerical context allowing comparisons to be made. Simulations comprised of four aquifers - a homogeneous aquifer, an aquifer with high conductivity lenses and two aquifer/aquitard systems. Environmental tracer concentrations were then used in conjunction with correction schemes. The research highlighted some of the limitations regarding the use of environmental tracers to infer groundwater "age". The correction schemes require some knowledge of the aquifers limiting the usefulness of adding such data to studies. Additionally, many of the correction schemes are not applicable outside of the context of their explicit assumptions. This work has implications for the use of environmental tracers outside of ideal contexts, in that using apparent ages in conjunction with models of mean or advective ages will produce erroneous results.

The third study investigated the numerical simulation of groundwater age distributions in heterogeneous aquifers. The study involved testing the ability of a number of geostatistical techniques to re-create the groundwater age distribution of a two-dimensional synthetic aquifer with varying levels of hydraulic conductivity data. Generally, in practice, values of hydraulic conductivity are only known at discrete locations. To simulate an entire K-field, a simulation technique is required. The study demonstrated the importance of the geostatistical model when estimating residence time distributions. Generally, a larger amount of conditioning data and a method able to recreate multiple scales of features will improve the estimate.

The final study proposed a new method for estimating groundwater age distributions with environmental tracer data. The use of lumped parameter models to estimate groundwater age distributions is limited by the simple assumptions of the model. Numerical simulation techniques, whilst able to simulate more complex systems, are limited by the required detail of spatially variable parameters. In this approach we assumed that that the relationship between historic concentrations and measured groundwater concentrations could be fully explained by the convolution relationship. We used this to assess how various levels of concentration data are able to inform the groundwater residence time distribution. We demonstrate that even with large amounts of environmental tracer data, estimates of residence time distributions are highly non-unique. This has implications for the use of lumped parameter models, as the ability of a model to fit environmental tracer concentration data does not necessarily validate it's choice.

Contents

List of Figures

List of Tables

1 Introduction

1.1 The research problem

Groundwater constitutes approximately 95% of available fresh water resources (Freeze and Cherry, 1979). The time frames of groundwater movement are on the order of weeks to tens of thousands of years (Freeze and Cherry, 1979). Given the range of such time frames, the study and quantification of groundwater ages is important. This is because these time frames relate directly to the time taken for resources to be replenished and the timescales of contaminant protection. Of these groundwater systems, heterogeneous environments, where properties vary spatially, remain some of the most challenging environments to characterise. An understanding of when this heterogeneity is important to consider is of continued interest (de Marsily et al., 2005).

Given the importance of time frames of flow and transport, the quantification of groundwater age is of considerable interest. Studies in groundwater age can take the form of measurement with environmental tracers (Busenberg and Plummer, 1992; Cook and Solomon, 1995; Dunkle et al., 1993; Corcho Alvardo et al., 2007, and others) or through numerical simulation (Goode, 1996; Varni and Carrera, 1998). Ages determined from environmental tracer concentrations may be useful in quantifying recharge rates (Vogel, 1967; McMahon et al., 2011), determining flow rates (Larocque et al., 2009) and calibrating groundwater models (Reilly et al., 1994; Sanford et al., 2004; Michael and Voss, 2009). Numerical simulation techniques in the absence of environmental tracer concentrations can be used in a theoretical context for exploring ages in systems (Cornaton and Perrochet, 2006a; Cardenas, 2007; Post et al., 2013). As such, these two techniques are able to give valuable insight into the time frames of processes in groundwater systems.

Currently, the use of apparent ages, groundwater age distributions and numerical modelling techniques are common tools used in groundwater studies. Methods for determining apparent ages use temporal variations in chemical compounds to set a clock for when groundwater was recharged (Cook and Bohlke, 2000). These methods generally assume that the water in a sample is made up of single age. The applicability of apparent age methods have been validated in the context of recharge areas of homogeneous aquifers (Schlosser et al., 1989; Solomon and Sudicky, 1991; Ekwurzel et al., 1994) and when using differencing techniques in 2D synthetic heterogeneous aquifers (Larocque et al., 2009). However these techniques have been showed to be biased by diffusion and dispersion (Cornaton et al., 2011; Castro and Goblet, 2005; Sanford, 1997; Sudicky and Frind, 1981) and heterogeneity (Weissmann et al., 2002). Additional studies have demonstrated that these biases are a result of sample being a mixture of ages, rather than a single age (Varni and Carrera, 1998; Bethke and Johnson, 2002a; Park et al., 2002; Weissmann et al., 2002; Waugh et al., 2003).

The simulation of advective ages, mean ages and age distributions can be undertaken with numerical models. These numerical simulation techniques can be used to calibrate groundwater models with apparent ages (Reilly et al., 1994; Sanford and Buapeng, 1996; Portniaguine and Solomon, 1998; Sanford et al.,

2004; Castro and Goblet, 2005; Michael and Voss, 2009) and define well protection time frames (Molson and Frind, 2012). Comparisons between simulated ages and apparent ages are limited based on the accuracies of apparent ages discussed above. The simulation of groundwater age distributions can be used to demonstrate time frames of solute transport in simple aquifers (Cardenas and Jiang, 2010; Etcheverry and Perrochet, 2000; Cornaton and Perrochet, 2006a) and in complex flow systems (Weissmann et al., 2002; Cornaton and Perrochet, 2006b; Engdahl et al., 2012). Using simulated distributions in conjunction with environmental tracer concentrations generally results in a hypothesis testing scenario whereby distributions that explain concentrations represent valid hypotheses (Weissmann et al., 2002; Troldborg et al., 2007; Leray et al., 2012).

The quantification of groundwater age distributions traditionally uses assumed models of the age distribution (Maloszewski and Zuber, 1982). These models are based on simple geometries and mixing relationships. The ability of these models to represent the systems they are applied to is generally not tested (the notable exceptions being Eberts et al., 2012; Massoudieh et al., 2012). These models have only a few parameters and can be constrained with a time series of environmental tracer data (Stolp et al., 2010) or multiple tracers obtained at one time (Solomon et al., 2010). Recent studies have suggested that these assumed models can be well constrained with data from multiple tracers (Massoudieh et al., 2012). However, the existing set of models are limited to specific environments and may not provide useful information outside of these contexts.

Despite some of the issues identified with these specific techniques, groundwater age may still provide vital information in the context of groundwater studies. This can occur through the correction of existing techniques, or the development of new techniques. Generally, this thesis focuses on the adaption of techniques for predicting groundwater ages in complex environments. It specifically addresses four knowledge gaps:

Firstly, the accuracy of environmental tracers in heterogeneous environments has received some attention. Weissmann et al. (2002) explored the relationship between apparent CFC ages and mean ages in a highly heterogeneous environment. The authors found that the apparent CFC ages consistently under-estimated the mean age. This was due to the ages being a distribution rather than a single value. The biasing of apparent ages due to mixing has been investigated previously (Varni and Carrera, 1998; Park et al., 2002; Bethke and Johnson, 2002a; Waugh et al., 2003; Bethke and Johnson, 2008). These studies suggest that the errors may not be random, but rather a bias based on the degree of mixing and the nature of temporal variations in tracer concentrations. For this reason, it is important to understand how heterogeneity impacts the level of mixing of ages, and how this mixing biases environmental tracer concentrations and the apparent ages derived from them. This understanding may allow for a framework for correcting ages in heterogeneous environments, and a method for using age biasing to determine the degree of heterogeneity.

Secondly, the concept of groundwater age as a distribution is poorly understood. This understanding extends to how this distribution impacts on environmental tracer concentrations. These concentrations are generally interpreted as an apparent age that assumes the water is of a single age. This age also differs from the mean

and advective age that is often generated by analytic and numeric models for comparison to apparent ages (Varni and Carrera, 1998). However, these apparent ages are often directly compared to mean or advective ages when estimating recharge (Vogel, 1967; Solomon et al., 1993; McMahon et al., 2011) or calibrating groundwater models (Sanford and Buapeng, 1996; Sanford et al., 2004; Janssen et al., 2008; Michael and Voss, 2009). It is not well understood when and where this approach will yield accurate or poor results. Additionally, a number of methods exist for correcting tracer ages. These methods may yield better results for comparison to mean or advective ages.

Thirdly, groundwater residence time distributions are highly sensitive to aquifer heterogeneity (Cornaton and Perrochet, 2006b; Weissmann et al., 2002; Gomez-Hernandez and Wen, 1998). Numerical modelling techniques are able to predict these RTDs but require a high level of parametrization, including the hydraulic conductivity field. Generally in application, only conductivity values at discrete locations are known, and as such a model is required to estimate the greater K-field (Renard and Allard, 2013). These methods include multi-Gaussian techniques which preserve only the spatial correlation of values and multi-point methods, which preserve the structure of the K-field, however require greater knowledge of the aquifer. For this reason, it is important to understand how the choice of a geostatistical model impacts the estimated RTD and the key features of the K-field that control this. This is especially important in environments where connection exists between extreme values (Gomez-Hernandez and Wen, 1998).

Finally, groundwater age distributions can provide important information about aquifer structures and solute transport properties (Engdahl et al., 2012; Troldborg et al., 2007). Traditional estimation of RTDs requires an assumed model (Maloszewski and Zuber, 1982; Corcho Alvardo et al., 2007; Solomon et al., 2010; Stolp et al., 2010). These models have few parameters and can easily be calibrated with environmental tracer data (Massoudieh et al., 2012). Unfortunately, these assumed RTDs are based on relatively simple conceptual models that limit the information that can be gained about the groundwater system. Conversely, numerical groundwater models require a high level of parametrization and are limited by structural and parameter uncertainties. There is a need for methods for assessing groundwater age distributions without the limitations of simple assumptions or the need for over-parametrization.

1.2 Research aims

This PhD broadly addresses the quantification of groundwater ages in complex environments. The themes covered in this thesis include correcting apparent ages for mixing biases, how geostatistical techniques impact the numerical prediction of groundwater age distributions and the development of new tools for determining groundwater age distributions. This body specifically addresses:

i. how groundwater age distributions are impacted by various levels of heterogeneity. Additionally, what impacts these distributions have on apparent ages and if the use of apparent ages from multiple tracers can be used to infer mean ages.

ii. the errors between measured apparent ages and mean or advective ages. This study also assesses methods for accounting for mixing biases. The aim is to better understand the errors of these techniques so that such errors can be propagated through predictive models and form part of an uncertainty assessment.

iii. how various geostatistical models are able to recreate key hydraulic properties that influence the prediction of residence time distributions in a synthetic heterogeneous aquifer where flow and transport are controlled by highly connected high-K features.

iv. a new technique for estimating groundwater age distributions using environmental tracer data without assuming the form of the distribution a-priori. This model was proposed with an additional uncertainty technique and tested against known RTDs.

The specific knowledge gaps have been the focuses of four manuscripts presented in Chapters 2 through 5. One of the manuscripts is now in press and three are at various stages of the review process.

1.3 Contribution of this PhD

This PhD explores the estimation of groundwater ages is heterogeneous aquifers. These issues are explored through a number of controlled numerical studies. The advantage of these being that the true answer is known making comparisons achievable. This research examines several techniques for estimating groundwater ages in heterogeneous environments. It does so by specifically examining: (i) the biasing of apparent age estimates due to mixing in heterogeneous environments. Additionally exploring the potential for correcting such a bias using multiple tracers. Previous research has indicated that these ages are incorrect in such environments however the sources of these errors have not been attributed to bias. (ii) what methods exist for overcoming the errors in age measurements outside of ideal environments. Many applications of apparent ages derived from environmental tracer concentrations directly compare these ages to mean or advective ages. Numerous studies have demonstrated that these ages are not representative of either mean or advective ages. We explore and test some methods for overcoming these errors. (iii) the ability of various geostatistical techniques to re-create properties of K-fields important to the estimation of numerical groundwater age distributions. In general practice K values are only known at discrete locations so the choice of model is important when predicting these distributions. (iv) how can information be obtained about the mixtures of age distributions in the absence of an assumed model. Current practice relies on a set of LPMs with limited assumptions to infer groundwater age distributions from environmental tracer concentrations. This limits the information that can be obtained from environmental tracer concentrations.

These findings will contribute to the accuracy of groundwater age predictions in heterogeneous environments. This will lead to a better understanding of these environments.

2 Manuscript 1: Bias of apparent tracer ages in heterogeneous environments

Accepted in Ground Water: McCallum, J.L., P.G. Cook, C.T. Simmons and A.D. Werner (in press), Bias of apparent ages in heterogeneous environments, *Ground Water*, doi:10.1111/gwat.12052.

Authors: James L. McCallum, Peter G. Cook, Craig T. Simmons and Adrian D. Werner

Abstract

The interpretation of apparent ages often assumes that a water sample is composed of a single age. In heterogeneous aquifers, apparent ages estimated with environmental tracer methods do not reflect mean water ages due to the mixing of waters from many flow paths with different ages. This is due to non-linear variations in atmospheric concentrations of the tracer with time resulting in biases of mixed concentrations used to determine apparent ages. The bias of these methods is rarely reported and has not been systematically evaluated in heterogeneous settings. We simulate residence time distributions (RTDs) and environmental tracers CFCs, SF_6, ^{85}Kr and ^{39}Ar in synthetic heterogeneous confined aquifers and compare apparent ages to mean ages. Heterogeneity was simulated as both K-field variance (σ^2) and structure. We demonstrate that an increase in heterogeneity (increase in σ^2 or structure) results in an increase in the width of the RTD. In low heterogeneity cases, widths were generally of the order of ten years and biases generally less than 10%. In high heterogeneity cases, widths can reach 100s of years and biases can reach up to 100%. In cases where the temporal variations of atmospheric concentration of individual tracers vary, different patterns of bias are observed for the same mean age. We show that CFC-12 and CFC-113 ages may be used to correct for the mean age if analytical errors are small.

2.1 Introduction

Knowledge of groundwater age allows for the timescales of groundwater flow to be directly determined. These timescales can be used to infer recharge rates (Vogel, 1967), estimate flow rates (Larocque et al., 2009), determine timescales of contaminant transport (Bohlke and Denver, 1995), constrain predictions made with groundwater flow models (Janssen et al., 2008), and define well protection areas (Molson and Frind, 2012). A number of definitions of groundwater age exist. The kinematic age (or advective age) refers directly to the time required for a particle to move by means of advection from the recharge location to the point where the aquifer is sampled (Varni and Carrera, 1998). Groundwater ages can also be defined as a distribution, whereby the molecules that make up a sample are not all of a single age but of many ages due to mixing within the aquifer (Varni and Carrera, 1998; Ginn, 1999; Cornaton and Perrochet, 2006a,b). This distribution is known as the residence time distribution but is commonly referred to as the age distribution. Mean age is the mean of this distribution (Goode, 1996). Mean age and kinematic age can be significantly different when mixing or mass transfer by diffusion are important (Castro and Goblet, 2005). The age of

interest is most likely dependant on the application. When determining recharge rates or flow rates, the kinematic age is most useful because it relates directly to hydraulic properties. However, when assessing contamination susceptibility, age definitions that reflect the mixing processes within the aquifer are more desirable, hence mean age or the entire age distribution are likely to be more meaningful as these will reflect solute transport properties.

Apparent ages are ages determined from environmental tracers. The relationships between apparent ages and advective or mean ages received much attention. In general when variations in tracer concentration with time are non-linear, processes that lead to spreading and mixing will result in biases in apparent age estimates when compared to mean or advective ages (Bethke and Johnson, 2002a; Waugh et al., 2003). The sources of spreading and mixing include diffusion and dispersion (Sudicky and Frind, 1981; Ekwurzel et al., 1994; Sanford, 1997; Bethke and Johnson, 2002a; Castro and Goblet, 2005). The bias due to non-linear variations in tracer concentration with time has been demonstrated for radioactive decay (Varni and Carrera, 1998; Park et al., 2002; Bethke and Johnson, 2008) and for variations in recharge concentrations of anthropogenic tracers (Ekwurzel et al., 1994; Waugh et al., 2003).

In heterogeneous aquifers, the accuracy of environmental tracers when estimating mean or advective ages has been shown to be variable. Weissmann et al. (2002) found that apparent ages determined using CFCs in a highly heterogeneous alluvial fan aquifer were much younger than mean ages. This was due to large components of the age distribution falling outside the range of waters where CFCs were present in recharge, due to non-linear atmospheric concentrations. In a theoretical study, using radioactive tracers with ideal half-lives, Larocque et al. (2009) found that flow rates were estimated more accurately with age dating methods than pump tests and hydraulic gradients. This method used an age differencing method to estimate flow rates. Although the tracers used were theoretical, half-lives were selected to relate to available tracers. The authors found that the best tracer to use was a tracer where the half-life was similar to the travel time.

In this paper, we systematically investigate how biases in apparent ages estimated from environmental tracers are affected by aquifer heterogeneity. The tracers CFC-11, CFC-12, CFC-113, SF_6, ^{85}Kr and ^{39}Ar are analysed using synthetic heterogeneous fields. Firstly, we determine the bias of the methods. Secondly, we explore the causes of bias, and thirdly we investigate whether bias correction is possible using multiple tracers.

2.2 Theory

Using environmental tracers to date groundwater relies on variations in the concentration of chemical species with time. These variations may be due to changes in recharge concentration of a tracer, radioactive decay of a tracer within the subsurface or production of a tracer within the subsurface (e.g. Cook and Bohlke, 2000). Although important to the dating of groundwater, these variations can produce biases when waters consist of a mixture of ages. Figure 2.1 presents the role of these temporal variations of concentration

in the biases of samples where waters are of mixed ages. Where temporal variations of tracer concentrations are linear over time, the mixing of waters of different ages will result in a concentration consistent with the tracer concentration corresponding to the mean age of the mixed sample (Figure 2.1A). In this case, no bias is observed. In a case where temporal variations of a tracer concentration result in an increase in the slope of the concentration with time, the concentration of mixed waters will result in an apparent age younger than the mean age (Figure 2.1B). It is also of note that the magnitude of the bias is related to the difference in individual ages. The bias increases as the difference in individual ages increases. In the case where the temporal variation of tracer results in decrease in the slope of the atmospheric concentration with time, apparent ages derived for mixed waters will result in a bias towards older ages (Figure 2.1C).

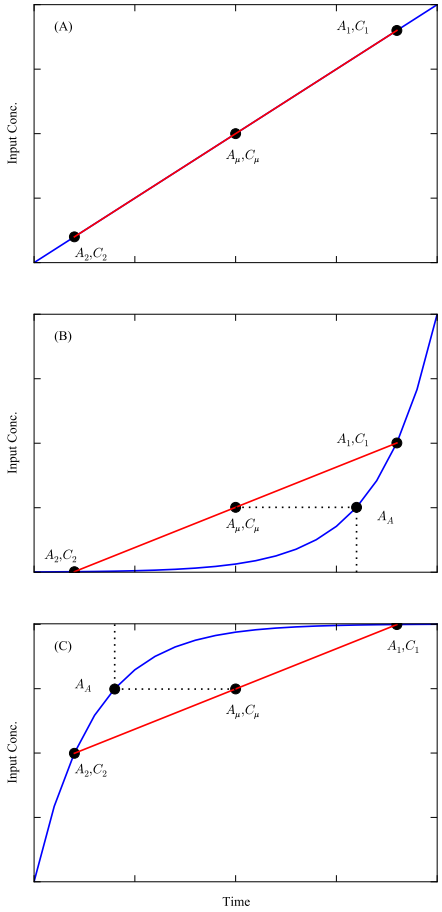


Figure 2.1: Age biasing for the mixture of two distinct waters due to variations in tracer concentration with time. The two waters have ages A_1 and A_2 and equivalent input concentrations C_1 and C_2. A_μ and C_μ correspond to the mean age and mean equivalent input concentration and A_A corresponds to the apparent age. For mixing where the variation in tracer concentration with time is linear, the apparent age corresponds to the mean age (Figure 2.1A). When the slope of the variation in tracer concentration increases with time, apparent ages are biased to younger ages (Figure 2.1B). When the slope of the variation in tracer concentration decreases with time, the apparent ages are biased to older ages (Figure 2.1C).

Although the bias of apparent ages is presented in Figure 2.1 for two waters of distinct ages, the concept also applies to water where a sample is made up of a distribution of ages (Waugh et al., 2003). The authors demonstrated that an increase in the width of the age distribution increased the bias of apparent age determined from environmental tracers. For this reason, the increased mixing and spreading in an aquifer will

result in a larger bias in apparent ages.

2.3 Methods

2.3.1 Conceptual model

A synthetic confined aquifer was used to investigate the effects of heterogeneity on tracer techniques (Figure 2.2A). The model domain was 500 m (x-direction) by 200 m (y-direction) and of unit depth, simulating two-dimensional, horizontal flow. A two-dimensional model was chosen over a three-dimensional model to allow for a larger number of scenarios to be implemented. A Darcy flux of 0.365 m/y was applied to the upstream boundary. This was chosen over a specified head boundary so that flow conditions were consistent in all scenarios. Water was removed from the domain using a specified head of 1.5 m at the downstream boundary. For all scenarios, a geometric mean hydraulic conductivity of 365 m/y and a porosity of 0.3 were used (chosen to be representative of a sand; Domenico and Schwartz, 1998). For solute transport simulations, a value of 0.04 was used for both longitudinal and transverse dispersivities and a tortuosity (defined as distance travelled over linear distance) of 1.7 was implemented (values chosen after Weissmann et al., 2002). Low values of dispersivity were chosen as heterogeneity was modelled at a small scale. Observation points were employed at varying distances from the upstream boundary (Figure 2.2A).

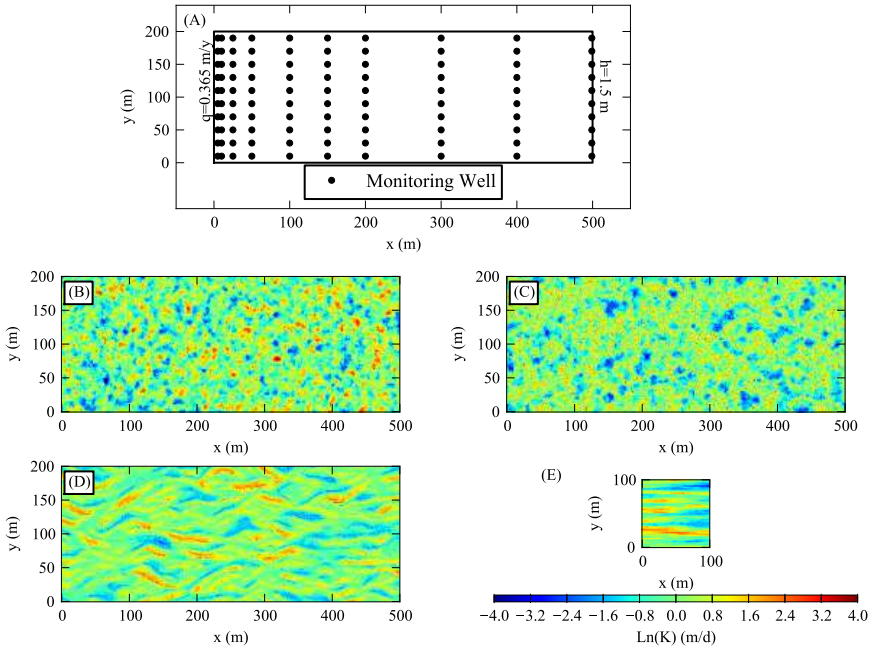


Figure 2.2: (A) conceptual model set up showing the location of wells and examples of (B) a multi-Gaussian field, (C) a connected field based on the method of Zinn and Harvey (2003) and (D) a field simulated by the direct sampling method using the training image presented in (E) and allowing rotations of up to 45 degrees.

2.3.2 Environmental tracers

Six environmental tracers were investigated; CFC-11, CFC-12, CFC-113, SF_6, ^{85}Kr and ^{39}Ar. CFCs are produced for use as refrigerants, aerosol propellants and solvents. Production of CFCs began in the 1930s which led to their accumulation in the atmosphere. CFCs enter the water cycle by dissolving in water during recharge. If the solubility is known, the water concentration can be related to an equivalent atmospheric concentration. Groundwater dating is achieved by relating the equivalent atmospheric concentration of the sample to the corresponding historic concentration. CFC-11, CFC-12 and CFC-113 are able to date water recharged after 1947, 1941 and 1955, respectively (Busenberg and Plummer, 1992).

SF_6 is used as an insulator in electrical switches. Production of this compound began in 1953 and the atmospheric concentration has increased steeply due to a long atmospheric residence time. SF_6 enters the water cycle similarly to CFCs and dating is achieved in a similar manner. SF_6 is able to date groundwater recharged after 1960 (Busenberg and Plummer, 2000).

Krypton-85 is a radioactive isotope with a half-life of 10.76 years. Krypton-85 is produced by nuclear power plants. The increase in the number of these nuclear plants since 1955 has resulted in an increase in the atmospheric concentration. Measured ^{85}Kr activity in water can be related to atmospheric concentrations after accounting for radioactive decay. It is suitable for dating groundwater recharged after 1960 (Loosli et al., 2000).

Argon-39 is produced by the interaction of cosmic rays with argon and potassium nuclides in the atmosphere. The concentration of ^{39}Ar in the atmosphere is constant as it is unaffected by anthropogenic sources. Dating is achieved by assuming that the reduction in concentrations of groundwaters is due to radioactive decay. Argon-39 has a half-life of 269 y making it suitable for dating waters in the range of 100 to 1000 years (Loosli et al., 2000).

Table 2.1: List of tracer decay constants and diffusion coefficients. Data obtained from Cook and Herczeg (2000) accept where specified.

Tracer	Symbol	Decay Constant (/y)	Diffusion Coefficient (m^2/y)
Water (RTD)	RTD	0.0	3.65×10^{-2} $^{(a)}$
Argon 39	^{39}Ar	2.58×10^{-3}	2.68×10^{-2} $^{(a)}$
CFC-11	CFC-11	0.0	3.19×10^{-2}
CFC-12	CFC-12	0.0	3.41×10^{-2}
CFC-113	CFC-113	0.0	3.00×10^{-2} $^{(c)}$
Krypton 85	^{85}Kr	6.44×10^{-2}	5.81×10^{-2}
Sulfur Hexafluoride	SF_6	0.0	3.79×10^{-2}

$^{(a)}$- After Maharajh and Walkley (1973).
$^{(b)}$- After Solomon and Sudicky (1991).
$^{(c)}$- After Cook et al. (2005).

For modelling environmental tracer concentrations, it was assumed that water entering the upstream boundary was consistent with atmospheric concentrations at the time of recharge. The recharge concentrations are presented in Figure 2.3. In all cases, it was assumed that the observation points were sampled in 2011. Individual tracers were modelled using appropriate free water diffusion and accounting for decay of input

concentrations (Table 2.1). These were coupled with the mechanical dispersion and tortuosity properties of the aquifer.

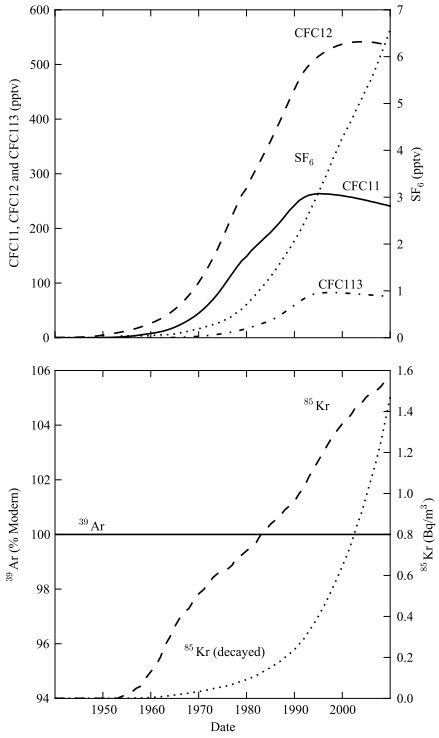


Figure 2.3: Atmospheric concentrations of modelled environmental tracers.

2.3.3 Hydraulic conductivity fields

To simulate the variations in K, three types of field structures were implemented – multi-Gaussian, Connected and Structured Fields. In a similar approach to what has been used by Larocque et al. (2009), multiple K field structures were simulated to assess how varying structure and connectivity impacted environmental tracer methods. Flow and solute transport properties of the porous media are dependent not only on the variance of the K field but also the connectivity structure (Renard and Allard, 2013). Multi-Gaussian fields were simulated using a sequential Gaussian simulator Deutsch and Journel (1998) with an isotropic correlation length of 10 m (Figure 2.2B). Connected fields were simulated after the method of Zinn and Harvey (2003) with an isotropic correlation length of 10 m (Figure 2.2C). This method transforms a multi-Gaussian field such that high values of K are connected and low values of K are isolated, whilst maintaining a correlation structure similar to a multi-Gaussian field. The third type of K field was simulated using the direct sampling method (Mariethoz et al., 2010) with the use of transform invariant distances (Mariethoz and Kelly, 2011) (Figure 2.2D). In this method, a basic training image was created using the sequential Gaussian simulator with correlation lengths of 200 m and 10 m in the x and y directions respectively (Figure 2.2E). The method uses informed neighbourhood points of known distances from an uninformed point on the simulation grid to search the training image and return a value for the uninformed point. The distances of these informed points obtained from the simulation grid are subject to transformations. The training image is then searched

at random with these transformed grids. In the case here, the transformation implemented was the use of rotations up to 45 degrees.

For each K structure, four values of log K variance ($\sigma^2_{ln(K)}$) were simulated: 0.3, 1.0, 2.5 and 4.0. These structure parameters are the same as those used by Larocque et al. (2009). Reported values of $\sigma^2_{ln(K)}$ in sedimentary aquifers range from 0.24 at the Cape Cod site (Hess et al., 1992) to 4.5 at the MADE site (Rehfeldt et al., 1992). The values chosen lie within these ranges. For each combination of structure and variance, 30 replicates were generated.

2.3.4 Numerical implementation

The problem domain was discretised to a 1 m by 1 m grid. A steady-state flow solution was obtained using the finite-element model HydroGeoSphere (Therrien et al., 2006). The finite-element approximation of the saturated steady-state simulation forms a system of linear equations. These equations were solved iteratively with a residual value of 10^{-9} m.

Residence time distributions (RTDs) and environmental tracer concentrations were simulated using a random walk particle tracking technique. The advantages of this technique in heterogeneous media include the elimination of numerical dispersion when using small dispersivities and a reduction in computation times (Salamon et al., 2006). The method used was the Ito-Fokker-Plank Interpolation method (LaBolle et al., 1996). The accuracy of this method has been demonstrated in both multi-Gaussian and non-Gaussian K fields (Salamon et al., 2006). In all cases, 2000 back-tracked particles were used to determine RTDs and concentrations at each of the 100 observation points in Figure 2.2.

Estimates of the RTD were made by constructing a cumulative distribution of the 2000 particles at each well (Weissmann et al., 2002). The concentration at an individual point was determined by:

$$c = \frac{1}{np} \sum c_0(t - A_p) \tag{2.1}$$

where c is the tracer concentration (ML^{-3}), np is the number of particles, t is the time of sampling (T), A_P is the time since the particle was recharged (T) and $c_0(t - A_p)$ is the atmospheric concentration of the tracer at the time when the particle was recharged (ML^{-3}) (corrected for radioactive decay in the case of ^{39}Ar and ^{85}Kr).

2.3.5 Apparent age determination

In the case of CFCs and SF_6, apparent ages can be determined directly by matching groundwater concentrations to equivalent atmospheric mixing ratios. Due to declines in concentration, CFC-11 and CFC-113 are only able to date water recharged prior to 1990. CFC-12 is only able to date groundwater recharged prior to 1997. Dating using ^{85}Kr is undertaken in a similar way once decay of the atmospheric concentration has

been accounted for. For ^{39}Ar, the apparent age (A_A) of a sample can be determined as:

$$A_A = -\lambda^{-1} \ln\left(\frac{c}{c_0}\right) \tag{2.2}$$

where λ is the radioactive decay constant (T^{-1}).

2.3.6 Calculation of bias and bias error

The error of each tracer method was assessed with two metrics – bias and bias error. Initially, apparent age estimates for each tracer were grouped into five-year intervals. For each of these five year intervals, the bias was calculated as the average value of the error between the mean age and apparent age. If the conditions of the K field were known, and these patterns of bias were known, a correction may be feasible. The bias error was calculated as standard deviation of the error for all apparent ages in each five year interval. This is a random component of the error and cannot be corrected. A correction for bias requires that this random error is small relative to the bias.

2.4 Results

2.4.1 RTDs

Figure 2.4 presents residence time distributions from a single well located at x = 50 m and y = 90 m for selected modelled heterogeneity scenarios. At low values of $\sigma^2_{ln(K)}$, water samples show very narrow residence time distributions. This suggests that the mixture of waters does not cover a wide range of ages. In addition, the differences between multiple realisations of the same K-field structures are relatively small. As the value of $\sigma^2_{ln(K)}$ increases, the width of the residence time distributions increases significantly. It is also apparent that the width of the connected and structured fields is greater than the width of the multi-Gaussian fields. There is also an increasing variation between individual realisations. This variation occurs not only in widths but also for central values. Further, there is evidence of multimodal distributions in all structures for values of $\sigma^2_{ln(K)} \leq 2.5$. These are shown by the multiple points of inflexion in the cumulative distributions.

2.4.2 Age bias and bias error

Figure 2.5 presents the relationship between the bias and bias error for two scenarios. Figure 2.5A presents the bias of apparent CFC-12 ages in a multi-Gaussian field for $\sigma^2_{ln(K)}$ = 0.3. Only a small error is observed between apparent ages and mean ages over the whole range of apparent recharge dates. Additionally, there is little scatter in this error with the exception of waters with apparent recharge dates less than 1950. Figure 2.5B presents the bias of apparent CFC-12 ages in a structured field for $\sigma^2_{ln(K)}$ = 4.0. Apparent ages

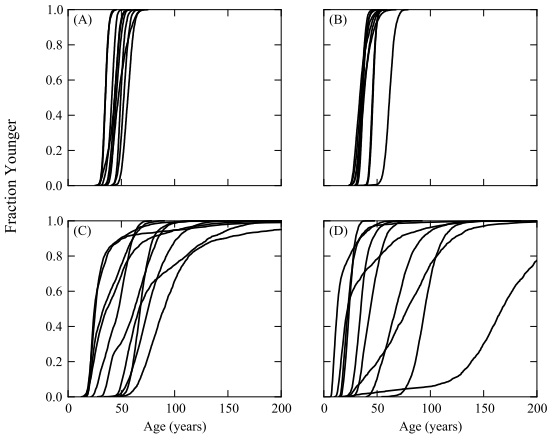


Figure 2.4: Residence time distributions (RTDs) for 10 realisations of (A) multi-Gaussian fields with $\sigma^2_{ln(K)}$ = 0.3, (B) structured fields with $\sigma^2_{ln(K)}$ = 0.3, (C) multi-Gaussian fields with $\sigma^2_{ln(K)}$ = 4.0 and (D) structured fields with $\sigma^2_{ln(K)}$ = 4.0. RTDs presented were obtained at an observation point located at x = 50 m and y = 90 m (Figure 2.2).

overestimate mean ages for apparent recharge dates after 1980 and underestimate mean ages for apparent recharge dates prior to 1970. There is also considerable scatter of the error, especially for older waters. Figure 2.5C presents the corresponding average bias for the simulations presented in Figure 2.5A. Some bias is observed for water recharged prior to 1960. The error of this bias is small, due to minimal variation in the bias of apparent ages (Figure 2.5E). Figure 2.5D presents the average bias corresponding to the simulations presented in Figure 2.5B. Here a strong bias of up to 70 years is observed in the extreme case. Bias errors also become very large for apparent recharge dates prior to 1980 (Figure 2.5F). In the extreme case, the bias error is almost as great as the bias.

Figure 2.6 presents the bias of apparent age estimates for CFC-11, CFC-12, CFC-113, SF_6 and ^{85}Kr. At low heterogeneities ($\sigma^2_{ln(K)}$ = 0.3), the bias observed for all tracers is less than 10 years in all structures. In highly heterogeneous cases ($\sigma^2_{ln(K)}$ = 4.0), this bias becomes quite large reaching 70 years in some cases. It is also evident that the structure of the K-field is important with structured and connected fields demonstrating a higher degree of bias than multi-Gaussian fields. Also of interest is that as the degree of heterogeneity is increased, the agreement between individual tracers is reduced.

Figure 2.7 presents the errors of the biases presented in Figure 2.6. For the low heterogeneous cases ($\sigma^2_{ln(K)}$ = 0.3) the variation of the bias is quite low. This variation increases at higher levels of heterogeneity ($\sigma^2_{ln(K)}$ = 4.0). As with the mean bias, the variation of the bias is related to structure. Structured and connected fields show higher variability than multi-Gaussian fields. In most cases, the error of the bias is similar to the actual bias (Figure 2.6). In an extreme case, apparent recharge dates of around 1940 derived from CFC-12 have a bias of more than 70 years with a bias error of almost 50 years.

Figure 2.8 presents the bias and bias error of apparent recharge estimates made using ^{39}Ar. In all cases, apparent ages are younger than mean ages. This bias of a radioactive tracer has been demonstrated previously

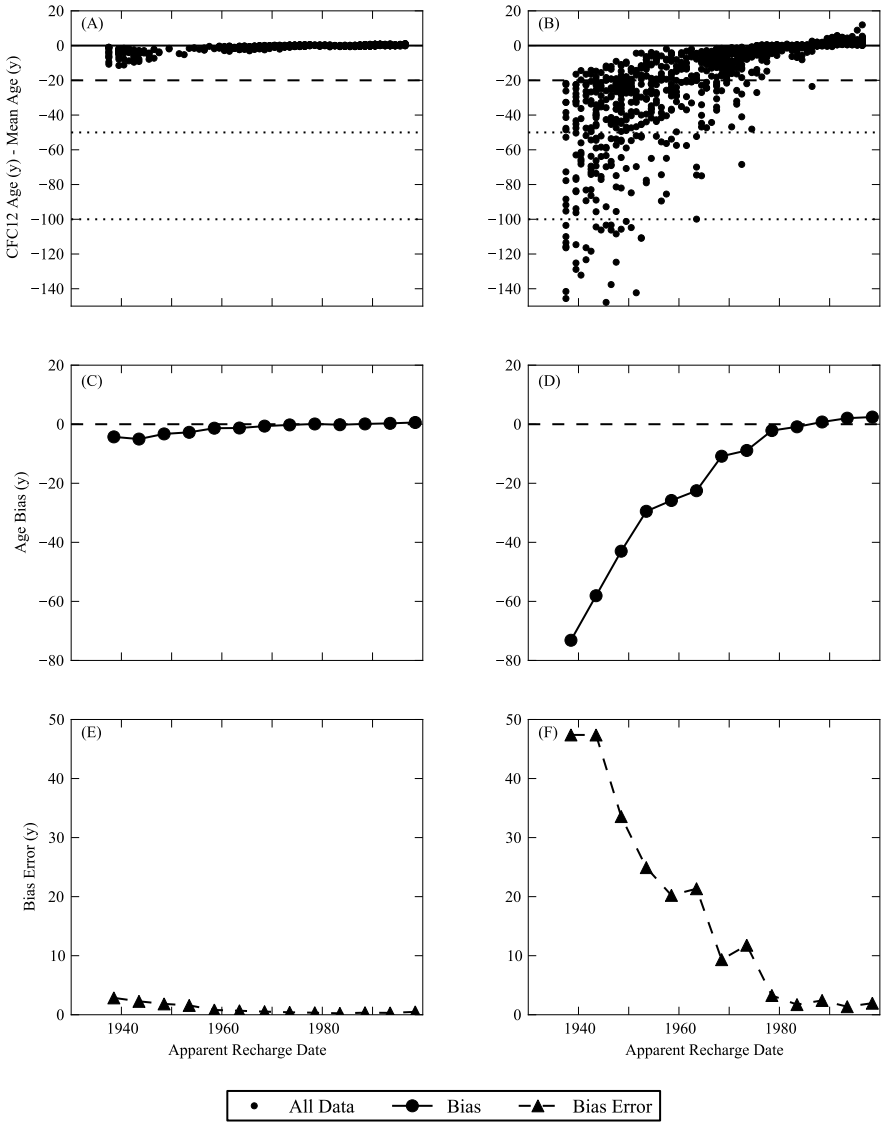


Figure 2.5: Example of error for all simulations from (A) multi-Gaussian fields with $\sigma^2_{ln(K)}$=0.3 and (B) Structured fields with $\sigma^2_{ln(K)}$ = 4.0 for 100 observation points within 30 realisations. Figures (C) and (D) represent the average bias for Figures (A) and (B) respectively. Figures (E) and (F) represent the standard deviation of the relationship. If a sample were taken from a structured field with $\sigma^2_{ln(K)}$ = 4.0 and yielded an apparent CFC-12 recharge date of 1960, this could be corrected to an apparent recharge date of 1938 $\pm$ 20 years based on a bias of -22 years and a bias error of 20 years.

by Varni and Carrera (1998). This bias is less than 10 years for $\sigma^2_{ln(K)}$ values of 0.3 and 1.0 for the ages investigated here (up to 500 years). Biases are increased for $\sigma^2_{ln(K)}$ values of 2.5 and 4.0. The structure of the K field is also important, with larger biases occurring in connected and structured fields than in multi-Gaussian fields. The bias error of apparent age estimates is also increased with increasing values of $\sigma^2_{ln(K)}$. Bias errors of less than 10 years are observed for $\sigma^2_{ln(K)}$ values of 0.3 and 1.0. For higher values of $\sigma^2_{ln(K)}$, bias error can be in excess of 40 years with bias error increasing with age.

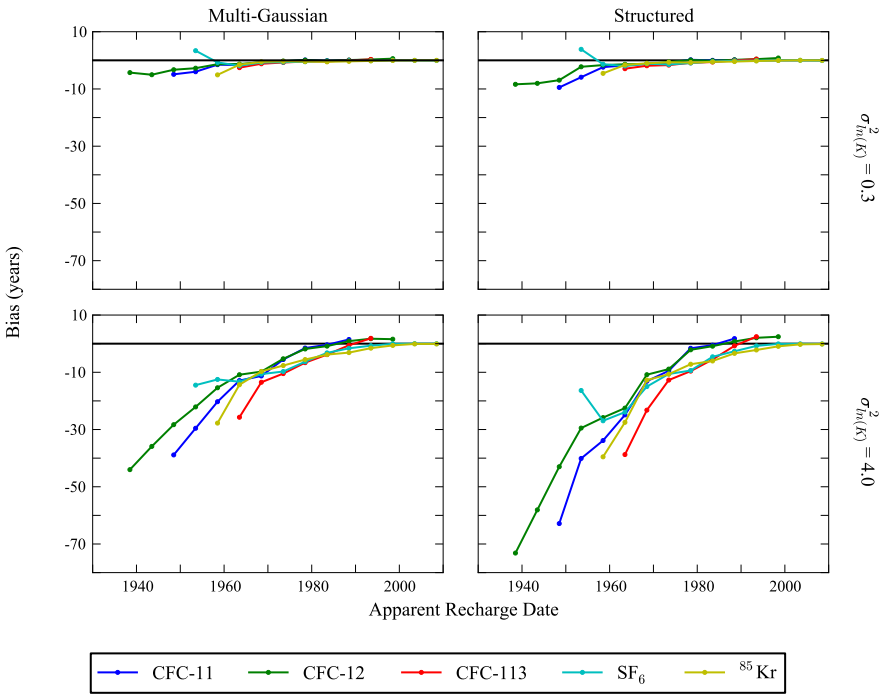


Figure 2.6: Age bias of young tracers CFC-11, CFC12, CFC-113, SF_6 and ^{85}Kr for all 100 observation points and 30 realisations of selected K fields presented in Figure 2.2. Bias is determined for 5-year groupings of apparent ages.

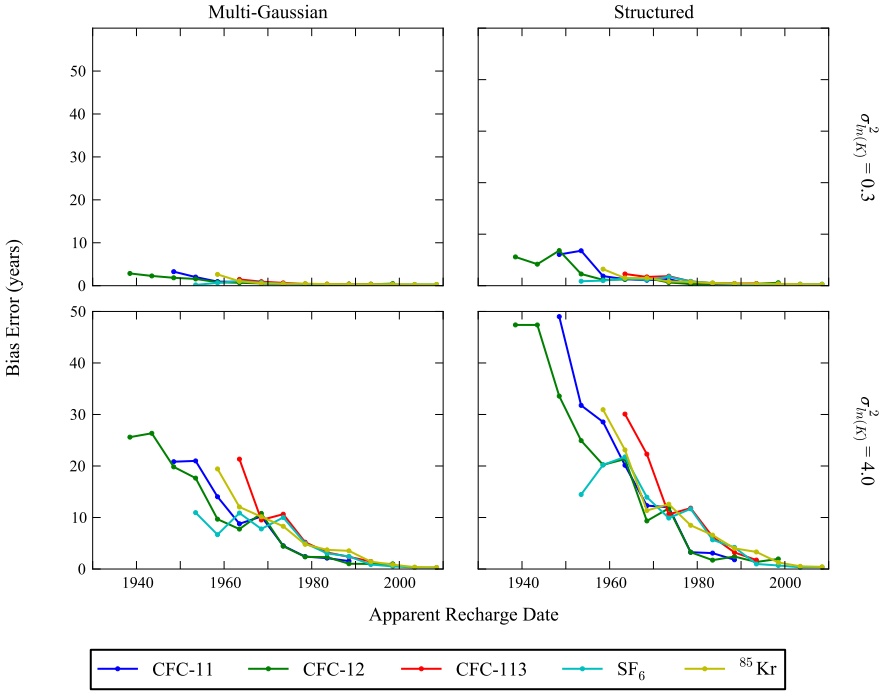


Figure 2.7: Bias error of young tracers CFC-11, CFC12, CFC-113, SF_6 and ^{85}Kr for all 100 observation points and 30 realisations of selected K fields presented in Figure 2.2. Bias error is determined for 5-year groupings of apparent ages.

2.4.3 Correcting for bias

Given that some tracers show different patterns of bias, comparing different tracers may give insight into the magnitude of bias. CFC-113 and CFC-12 demonstrate different patterns of bias especially at high values of

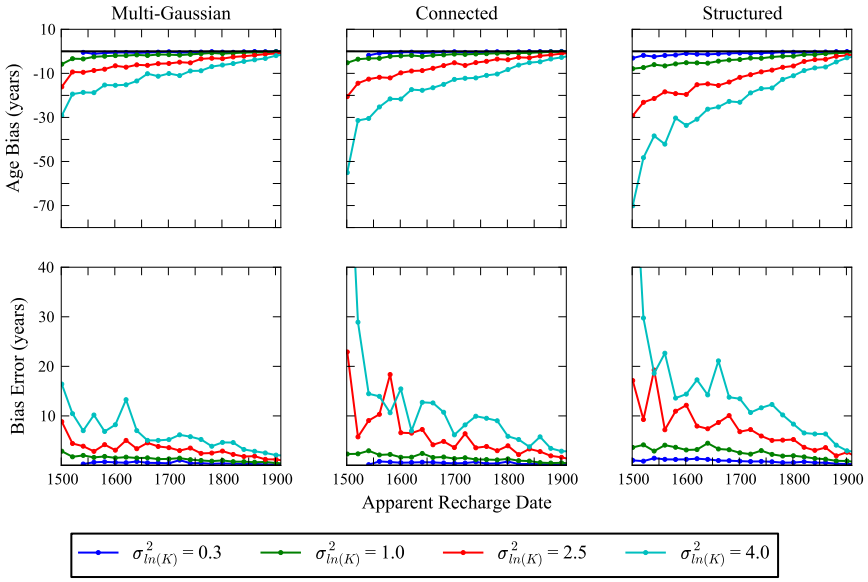


Figure 2.8: Bias and bias error of ^{39}Ar for all 100 observation points and 30 realisations of each K field presented in Figure 2.2. Bias and bias error are determined for 20 year groupings of apparent ages.

$\sigma^2_{ln(K)}$ (Figure 2.6). Figure 2.9 shows the relationship between apparent CFC-12 age, the difference between CFC-12 and CFC-113 apparent ages and the mean age. If only apparent CFC-12 age was used, multiple mean ages may be represented by an apparent age. This is shown by the non-vertical contours of mean ages. By comparing this with CFC-113 ages, the range of possible mean ages can be constrained. This relationship is well constrained for mean ages of less than 55 years however a relationship is unclear for mean ages greater than this. These relationships apply to any tracers that show different patterns of bias. For example, apparent ages between 1970 and 1990 for CFC-11 and CFC-12 show similar patterns of bias to each other however different patterns than all other tracers. For apparent ages between 1960 and 1970, CFC-113 shows different patterns of bias to all other tracers. For apparent recharge dates prior to 1960 CFC-11 has a larger bias than CFC-12.

Tracers that apply to different age ranges may also be compared. For example, ^{39}Ar ages can only date waters older than 90 years and CFC-12 ages are only considered reliable at ages less than 60 years. Figure 2.9 also shows the relationship for observations where both these tracers reported ages within these ranges. This occurred for 102 observations mostly in highly heterogeneous fields. For CFC-12 ages close to 60 years, ^{39}Ar ages are only slightly biased compared to mean ages. For younger CFC-12 ages, a larger bias is observed in ^{39}Ar ages. This suggests that using CFC-12 ages and ^{39}Ar ages can help to constrain mean ages.

Although heterogeneity cannot be directly implied from two tracers, the magnitude of the difference between apparent ages obtained from two tracers can be indicative of the level of mixing. As stated above, the bias differences between tracers become more apparent at higher heterogeneities. Although this is not a direct correlation, the magnitude of error may give insight into aquifer mixing processes. As increased heterogeneity increases the mixing, apparent age differences may be a proxy for heterogeneity.

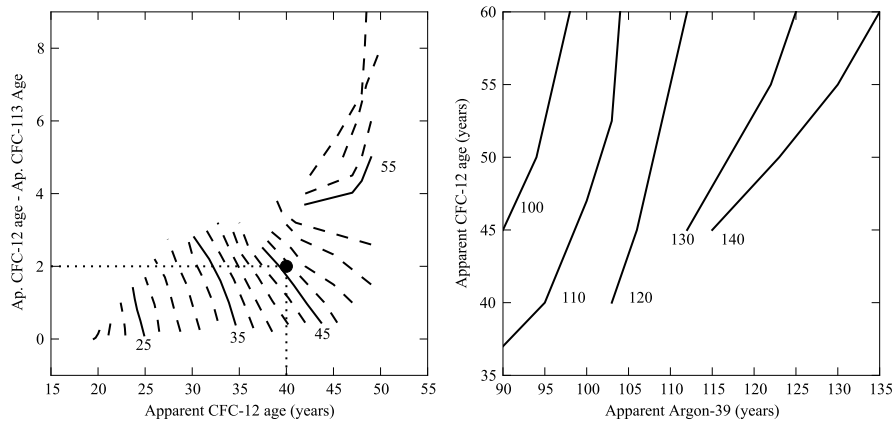


Figure 2.9: Relationships between apparent ages from two tracers and mean ages for all K field structures and variances. Lines represent contours of equal mean ages. Thus for example if a sample reported a CFC-12 apparent age of 40 years, and a CFC-113 apparent age of 38 years, the actual mean age would be approximately 47 years.CFC-113 and CFC-12 relationships are based on all observations where apparent ages were reported for both tracers. Argon-39 and CFC-12 relationships are based on 102 observations where CFC-12 apparent ages were less than 60 years and ^{39}Ar ages were greater than 90 years.

2.5 Discussion

This paper has demonstrated the relationships between apparent ages obtained from environmental tracers and mean ages in heterogeneous systems. In general, we demonstrate that increasing heterogeneity leads to an increase in the errors of apparent age estimates. We also observed that as the heterogeneity increases, the consistency of apparent ages of different tracers is reduced. We have demonstrated that as these patterns of bias are related to temporal variations in atmospheric concentrations, differences may be indicative of mean ages. In specific examples we demonstrated that relationships between CFC-12 and CFC-113, and CFC-12 and ^{39}Ar can give insight into mean ages.

If differences between two tracers are used to correct mean ages, the biases due to heterogeneity must be greater than analytical errors of the tracer techniques. The apparent age error from analysis is related to the slope of the temporal variation of concentration. If the slope is high, then small analytical errors may not result in large apparent age errors. Cook and Solomon (1997) suggested that the analytical error of CFCs is better than 4 years for a purely advective system. This accounts for uncertainty of recharge temperature and estimates of excess air. This range is of a similar magnitude as the bias observed between CFC-12 and CFC-113 due to heterogeneity and hence these relationships may not be distinguishable in reality. Similarly, for ^{39}Ar, reported errors of waters with apparent ages of approximately 90 years are relatively high. Corcho Alvardo et al. (2007) reported an apparent age of 91 years $\pm$ 35 years and Sultenfuss et al. (2011) reported an age of 90 years $\pm$ 50 years. Given the high error of such measurements, it may be difficult to distinguish the effects of heterogeneity.

Although our method has investigated the role of heterogeneity in the error of mean ages derived from environmental tracers, other sources of error exist. For CFCs, errors may occur due to sorption, degradation and the incorrect assumption of recharge temperatures (Ekwurzel et al., 1994). Further errors between

tracers may be caused by the thickness of the unsaturated zone and the varying effects this has on tracers (Cook and Solomon, 1995). In addition, sampling of some tracers requires large amounts of water. Krypton-85 for example requires 120 litres of water (Ekwurzel et al. 1994) and ^{39}Ar requires two m^3 of water (Loosli et al., 2000) which may result in further errors due to sampling of greater portions of the aquifer.

The flow field used here was very simple. Previous examples in the literature have investigated how flow systems may influence RTDs. Cardenas (2007) demonstrated how topographically driven flows can result in power law RTDs at discharge features. It is likely that in the presence of discharge features such as streams, a large number of flow paths may converge on the same location. Reilly et al. (1994) found that near a discharge feature, apparent tracer ages did not correspond well to particle tracked ages. Pint et al. (2003) found that a large range of ages occurred at discharge features. If the analysis here were repeated for an unconfined aquifer, where flows were driven by vertical recharge, it is likely that the flow paths created would lead to wider distributions. This was not investigated here because our aim was to determine the role of K fields in confined aquifers.

Although heterogeneities in the two dimensional flow fields are shown to cause wide residence time distributions, in field situations three-dimensional flow will occur. Pohll et al. (2000) found that when comparing 2D and 3D fields, solute arrival times were earlier in 3D fields than in 2D fields. This was due to a lower number of connections between high K zones in the 2D fields. It is likely that the use of a 3D field would have led to further spreading of the RTDs. This would create a greater level of bias between apparent ages and mean ages.

The choice of baseline parameters for flow and solute transport may also have impacted our results. Sudicky and Frind (1981) related the errors in ages to the ratio of advective to diffusive fluxes. An increase in diffusive fluxes will increase the error of apparent ages. This flux would increase by decreasing tortuosity. The choice of porosity would increase the advective fluxes and would have reduced the bias.

Additional errors exist when inferring kinematic ages using mean ages. Mean ages can become much larger than kinematic ages due to diffusive exchange with low conductivity areas. A similar phenomena is observed due to diffusion from aquitards (Bethke and Johnson, 2002b), diffusion in fractured rock systems (Cook et al., 2005), and vertical dispersion (Castro and Goblet, 2005). These processes result in an underestimation of flow rates when assuming that mean age or apparent age is equivalent to the kinematic age. Ages obtained at a point in the aquifer may also not be representative of all points in the aquifer at a given distance. This may lead to error when estimating aquifer flow rates (Larocque et al., 2009).

Our findings have also highlighted the importance of K field structure when assessing heterogeneity. Multi-Gaussian structures are often implemented, as they require limited information about the structure of the aquifer. These methods however can fail to recreate patterns of solute transport observed in field applications. This is why the connected fields used in this study were implemented (Zinn and Harvey, 2003). This was to represent field scale mixing phenomena with a two dimensional model. The third structure implemented was intended to represent a heterogeneity based on a braided river system. The mixing observed

was greater in well-connected and more structured fields than multi-Gaussian fields. In the cases of apparent ages, the assumption of a multi-Gaussian field structure would have resulted in an underestimation of the age bias. For example, if variance was obtained from detailed hydraulic conductivity testing at a site, the level of bias could not be assumed without knowledge of the structure of the heterogeneity. This would require knowledge of the connected structure of the field, as this is equally important in determining the flow and solute transport properties of a field. These patterns would be dependent on site specific applications.

2.6 Conclusions

In this study, we investigated the bias of apparent age estimates made using the tracers CFC-11, CFC-12, CFC-113, SF_6, ^{85}Kr and ^{39}Ar in heterogeneous aquifers. It was found that an increase in K variance increases the bias of age tracer estimates compared to mean ages. In the extreme cases, biases of 100 % were observed due to heterogeneous fields. The level of this bias is dependent not only on the variance of K fields but also on the structure. Although a bias exists for different structures and levels of variance, age errors cannot be directly corrected for, as the variability of the bias is high relative to the magnitude. The bias is a result of two things. Firstly, the shape of the atmospheric concentration of a tracer will determine the direction of bias. Ideally, tracers that show a linear increase over time will have no bias, however this is only observed in available tracers over small ranges of possible dating periods. Secondly, the magnitude of the error is dependent on the width of age distributions. Over the ranges where the bias patterns of individual tracers vary, apparent ages from two tracers may give insight to the error of apparent ages and allow for the correction of mean age. In the case here, we demonstrated how apparent ages obtained from CFC-12 and CFC-113 may be used to determine mean ages.

Acknowledgments

The Authors wish to thank Gregoire Mariethoz for his assistance with geoststistical methods, Frank Schwartz, Gary Weissmann and an anonymous reviewer for comments that strengthened this manuscript. Funding for this research was provided by the Australian Research Council and the National Water Comission.

3 Manuscript 2: Limitations of the use of environmental tracers to infer groundwater age

Submitted to Ground Water.

Authors - James L. McCallum, Peter G. Cook and Craig T. Simmons

Abstract

The use of groundwater age to estimate recharge and flow rates, as well as calibrate groundwater models is common practice. Unfortunately, the relationships between apparent ages measured using environmental tracers and the advective and mean groundwater ages they are compared to are subject to some limitations. In this paper, we explore the accuracy of these relationships and demonstrate their conditional nature. Additionally, we examine some correction schemes as ways to overcome the errors in these relationships. We show that relationships between measured apparent ages and mean or advective ages are quite easily biased resulting in limited usefulness when making comparisons. However, we found that correction schemes that accounted for mixing biases were able to improve the accuracy of mean ages using multiple tracers. Corrections for advective ages are limited as the relationship between the movement of a solute tracer and the advective age requires specific knowledge of aquifer properties. Our work suggests that with correction, environmental tracers are able to inform mean ages, however comparison to advective ages may result in large errors. Hence, mean ages provide better calibration targets than advective ages.

3.1 Introduction

Knowing the age of groundwater at a location in an aquifer is of great interest. This is because groundwater ages can inform recharge rates (Vogel, 1967; McMahon et al., 2011), flow rates (Larocque et al., 2009), and well protection timeframes (Molson and Frind, 2012). These are all important issues that confront groundwater scientists. Additionally, groundwater age data is seen as useful in calibrating groundwater models (Reilly et al., 1994; Sanford and Buapeng, 1996; Sanford et al., 2004; Janssen et al., 2008; Ginn et al., 2009; Sanford, 2011). Thus, the characterisation of groundwater ages through measurement and simulation continues to be of interest.

Groundwater age is frequently measured through the use of naturally occurring and anthropogenic compounds (from this point forward collectively referred to as environmental tracers) (Cook and Herczeg, 2000). In the simplest form, this approach assumes that the water making up a sample is of a single age (Cook and Bohlke, 2000). In more detailed studies, a distribution of ages may be assumed and fitted to concentrations of one or multiple compounds and a mean transit time may be inferred (Maloszewski and Zuber, 1982; Stolp et al., 2010). These two concepts, apparent ages and age distributions, may also differ from the advective age (sometimes called kinematic or true water age) or the mean ages of groundwater

at the location in the aquifer where the sample comes from (Varni and Carrera, 1998). Ultimately, the age which is of most interest will be dependant on the interpretive model or application.

Given that some confusion exists in the definitions of age and how measured ages relate to these, we aim to present some of these concepts clearly. In this paper, we describe the use of environmental tracers to infer groundwater age. We first look at some basic concepts of groundwater age – including some definitions. We then investigate why environmental tracers prove useful in inferring groundwater ages. We also present how various definitions of groundwater age relate to one another. We then investigate how and where comparisons between apparent ages and mean or advective ages yield positive or negative results, and present some known correction schemes to infer mean or advective ages from apparent ages. In doing so we hope to provide a clear picture of what environmental tracers can infer about groundwater age.

3.2 Definitions of groundwater age

3.2.1 Advective age

The advective age (also known as true groundwater age and kinematic age) is simply the time taken for a water particle to move from a recharge zone to the location of sampling purely via advection (i.e. the velocity of groundwater). This age can be used to directly infer recharge rates (Vogel, 1967; Solomon and Sudicky, 1991; McMahon et al., 2011), estimate groundwater flow rates (Larocque et al., 2009) and calibrate numerical groundwater models (Sanford et al., 2004; Michael and Voss, 2009; Janssen et al., 2008). The usefulness of this age is that it reflects only the movement of groundwater due to hydraulic forcing – recharge, discharge, hydraulic gradients, hydraulic conductivity and porosity. These ages can easily be incorporated into models via advective particle tracking schemes.

3.2.2 Mean age

The mean age refers to the average age of the water molecules at a location in the aquifer (Goode, 1996). The multiple ages that contribute to the age are due to groundwater age being a distribution (discussed below) rather than a single value. Differences between mean and advective ages occur as groundwater age is subject to dispersion and diffusion. This may lead to exchange of waters between slow moving and fast moving flow paths resulting in ages that do not represent the advective ages (Bethke and Johnson, 2002b). These ages are more likely to represent timeframes of contaminant flow in aquifers than advective ages.

3.2.3 Age distributions

Age as a distribution has its roots in reservoir theory. It is based on the observation that when the composition of water flowing into a system is suddenly changed, the breakthrough of the changed water is not

instantaneous but rather gradual. This is because the travel times within a system do not represent a single value but many (Danckwerts, 1952). In the context of groundwater age studies, the samples collected with which to infer groundwater ages represent a distribution of ages. This is because samples represent multiple water molecules that have travelled different paths over different time frames. These multiple flow paths occur due to diffusion, dispersion, long well screens, heterogeneity or recharge/discharge relationships. This concept of age as a distribution has been defined using simple models such as the exponential model or dispersion model (Maloszewski and Zuber, 1982) and more recently with numerical simulation techniques (Varni and Carrera, 1998; Ginn, 1999; Cornaton and Perrochet, 2006a; Cornaton, 2012; Gomez and Wilson, 2013). Importantly, these distributions represent the equivalent breakthrough of a conservative tracer hence they may offer some understanding of solute transport processes in an aquifer (Engdahl et al., 2012). The mean age relates to the mean of this distribution. Age distributions may become wide in environments where adjacent flow paths have significantly different ages (e.g. in heterogeneous aquifers) or where sampling conditions encounter a number of flow paths (e.g. long well screens).

3.2.4 Moments of the age distribution

In addition to mean ages, higher order moments of the age distribution can be estimated. The second moment – variance, which is a measure of the width of the distribution, has received some attention (Varni and Carrera, 1998; Waugh et al., 2003). This is a basic measure of the range/width of ages that might be expected at a location.

3.3 Tracer ages

In this section, we particularly discuss the concepts of age dating, with regard to the assumption that the water in a sample is made up of a single age. Hence, we do not explicitly discuss the errors associated with sampling techniques, assumptions of recharge conditions, etc. A large amount of literature has been devoted to these topics (e.g., Cook and Solomon, 1997)

The concept behind the use of environmental tracers to determine aspects of groundwater age is that there is some variation in a chemical compound over time. This may be due to variations in recharge concentrations, radioactive decay, and production of a compound in an aquifer system or a combination of these factors. By knowing how these concentrations vary with time and hence age, a measured concentration of such a compound can be related to groundwater age.

In the simplest definitions, groundwater age is defined as an apparent age, which refers to the time at which the water sample was last in equilibrium with the atmosphere. This often assumes that the sample consists of water of a single age. This can be determined by relating the concentration in the aquifer to an atmospheric concentration after accounting for decay or production. As an example, radiocarbon dating assigns an age based on the decay occurring since the water was last in equilibrium with the atmosphere. By attributing

the change in concentration to radioactive decay, an age dependant process, the difference in concentrations yields an apparent age.

3.4 Relationships between ages

3.4.1 Summary of previous knowledge

An excellent comprehensive comparison of different definitions of age is provided by Varni and Carrera (1998). Here, in addition to providing simulation techniques for all age definitions mentioned above, the authors undertook simulations to allow for comparison. The authors found that age distributions arose by mixing along and between flow paths. Mixing between flow paths meant that advective ages could lie anywhere within or outside of the age distribution. With regard to apparent ages, the authors demonstrated that radiometric ages (apparent ages obtained from radioactively decaying tracers) were sensitive to the mean of the distribution and the second moment – variance (a measure of width). Hence, an apparent age is not representative of a mean age unless the width of the distribution is small and only representative of an advective age if mixing between streamlines is negligible and the width of the distribution is small. As an example, if Carbon-14 was used to obtain a date on a sample with a mean age of 5000 years, variances of 8.4×10^6 and 1.8×10^7 years would correspond to errors of 10% and 20% respectively. Their work suggested that if multiple tracers with differing half-lives are measured, then the mean and variance of the distribution may be determined.

Concepts of bias of apparent ages for simple mixtures have also been demonstrated (Park et al., 2002; Bethke and Johnson, 2002a, 2008; McCallum et al., In Press). In general, these concepts deal with the issues of non-linear variations in chemical species with time. This concept suggests that the average concentration of a mixture of two waters with distinct age will only be equivalent to the mean if concentration variations are linear. Hence, a bias will exist if using an apparent age derived from a compound that has non-linear variations with age. This temporal variation may be due to variations in atmospheric concentrations, radioactive decay or production within the aquifer. As an example, if Carbon-14 was used to analyse a sample with equal parts of 1000 and 3000 year old water, an apparent age of 1940 years would be obtained (compared to a mean of 2000 years). If the distinct components had ages of 500 and 3500 years, the apparent age would be 1864 years. Figure 3.1 highlights these concepts – non-linear variations result in biases and the magnitude of the bias is determined by the difference in the ages. Importantly, this bias also applies to continuous mixtures that arise naturally in groundwater systems.

Additional studies have looked at the relationship between radiometric ages and advective ages (Sudicky and Frind, 1981; Sanford, 1997; Cornaton et al., 2011). These systematic studies focus on how ages are biased due to the existence of mobile and immobile components of the aquifer. In these cases if something is known about these zones, corrections can be made to account for the exchanges between fast and slow flow paths. These studies investigate the petitioning of water between mobile and immobile zones. Further

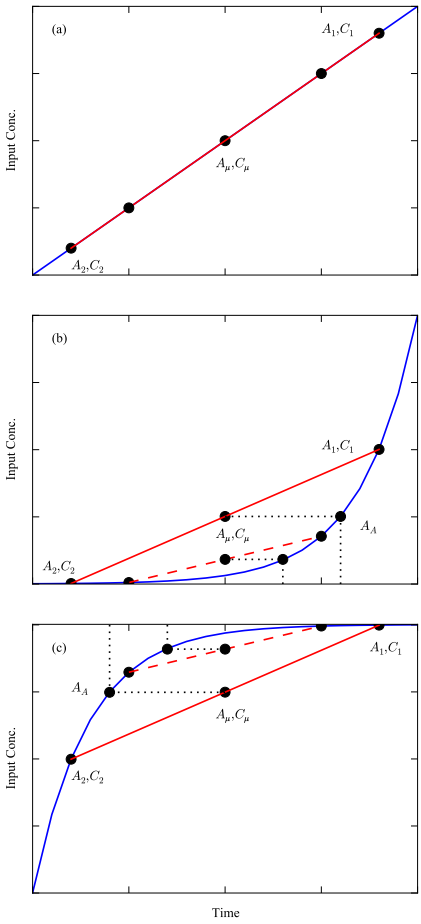


Figure 3.1: Concept of mixing bias of two samples of distinct age (modified from Park et al., 2002). Figure 3.1A demonstrates that mixing where concentration varies linearly with age result in no bias. Where concentrations increase their rate of change with age, a bias to younger ages is observed (Figure 3.1B). Where concentrations reduce their rate of variation with time a bias to older ages is observed (Figure 3.1C). The magnitude of the bias is determined by the magnitude of the difference between ages (See relative errors between solid and dashed lines).

correction is then needed for bias due to variations in chemical species (i.e. radioactive decay). As an example, Sanford (1997) found that in the Bangkok basin, carbon-14 ages within the aquifer were on average 10,000 years older than advective ages due to diffusion between water flowing in the aquifer and immobile (and hence older) water stored within aquitards.

Additional studies other than that of Varni and Carrera (1998) have also investigated the relationship between age distributions and apparent ages determined from environmental tracer concentrations. Waugh et al. (2003) used an inverse-Gaussian model of age distributions to interpret environmental tracer concentrations. The study assessed "young" tracers (tritium, CFCs, SF6) in addition to radiometric tracers. The authors found that apparent ages derived from CFCs and tritium/helium differed from mean ages by 9-12 years when the width of distributions reached 20 years. Interestingly, they also found that environmental tracers could be used to estimate the mean and a measure of the width of the age distribution consistent with conclusions from radiometric tracers (Varni and Carrera, 1998). McCallum et al. (In Press) used this concept to suggest a way to correct mean ages using apparent ages obtained from multiple tracers with

different biases. This method assumed that different tracers presented a different bias for the same mixture of ages due to differences in their input concentrations (Figure 3.1). Comparing the bias of two different tracers can give insight to the mean age.

When estimating whole distributions, Massoudieh and Ginn (2011) showed that using a radioactive tracer could allow for the estimation of an assumed inverse-Gaussian distribution. Although this assumption is limited, Waugh et al. (2003) demonstrated that means and widths estimated with the same distribution were accurate when estimating multi-modal distributions. In a study where distributions were estimated without the specifying the shape of the distribution a-priori, McCallum et al. (In Press) found that even with large amounts of environmental tracer concentrations data, the shape of distributions remained non-unique. However, the authors found that mean ages could be estimated accurately. The findings of Waugh et al. (2003) and McCallum et al. (In Press) are consistent in that they suggest that while Environmental tracer concentrations can be explained by many distributions, mean ages can potentially be determined independent of an exact distribution.

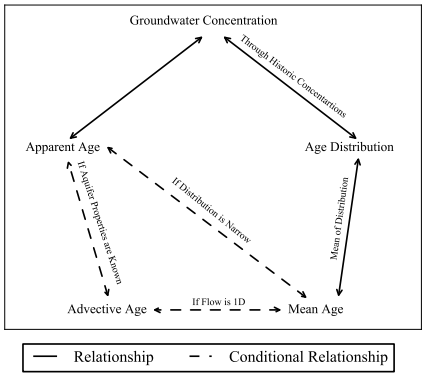


Figure 3.2: Schematic of relationship between environmental tracer concentrations and the different definitions of groundwater age.

In general, studies have suggested that the age of water that makes up a sample from an aquifer is composed of a distribution of ages. If this distribution is due to mixing along a single flow path then the mean age will represent the advective age. In cases where exchange occurs between flow paths then the mean age will no longer represent the advective age. Conversely, environmental tracer concentrations yield information about mean ages when the widths of age distributions are negligible. Therefore, apparent ages can only inform advective ages when water is from a single flow path where dispersion is negligible or if knowledge of the system allows for a correction (i.e. knowledge of mobile/immobile zones). A schematic of these relationships is presented in Figure 3.2.

3.4.2 How wide can distributions be and how wrong can ages be?

In this section, we look at some simple flow systems and demonstrate how ages are biased.

Figure 3.3 presents some simple aquifers used to demonstrate the issues with groundwater ages. All aquifers were simulated to be 5000 m (x – direction) by 50 m (z-direction). Figure 3.3A represents an aquifer

receiving constant recharge along the top boundary with discharge occurring through the top of the model between 4990 and 5000 m. A uniform hydraulic conductivity of 1 m/day and a constant porosity of 0.1 were assumed. Figure 3.3B represents the same configuration with high conductivity lenses located throughout the aquifer. These lenses have a hydraulic conductivity of 10 m/day. Figure 3.3C represents horizontal flow through a layered aquifer. The aquifer consists of bottom and top layers with hydraulic conductivities of 0.1 m/day and a central layer with a conductivity of 10 m/day. Figure 3.3D represents a similar setup, however with sloping layers and recharge above the high-K layers outcrop.

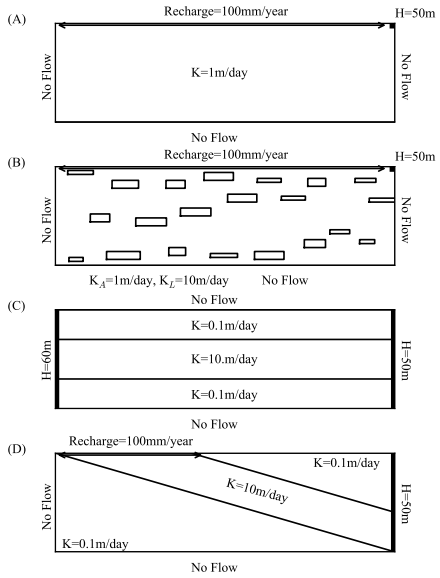


Figure 3.3: : Selected aquifers to perform simulations of environmental tracer concentrations, advective ages, mean ages and groundwater age distributions.

Flow in all aquifers was simulated using HydroGeoSphere (Therrien et al., 2006). Here the problem domain was discretised to a 1m by 1m grid. Fixed head boundaries were applied as per Figure 3.3, and recharge was applied through specified flux boundaries. The saturated steady state flow equations were solved iteratively until a residual error of 10^{-9} m had been achieved.

Simulations of kinematic ages were undertaken using an advective backward particle tracking method with linear interpolation of the velocity field. Groundwater age distributions and environmental tracer concentrations (with the exception of tritium and helium) were simulated using a random walk particle tracking scheme (Weissmann et al., 2002; McCallum et al., In Press). This scheme implemented was the Ito-Fokker-Planck method with linear velocity interpolation (LaBolle et al., 1996). Mean ages were calculated as the mean travel time of all particles (Weissmann et al., 2002). In the case of tritium/ helium simulations, the need to simulated the different diffusion coefficients of the two species was not achievable with the random walk scheme. In this case, solute transport simulations were carried out using HydroGeoSphere. The scheme implemented the use of the van Leer flux limiter option, equivalent to a TVD scheme. For simulations, values of 1.0 m and 0.1m were selected for longitudinal dispersivity and transverse dispersivity respectively. Diffusion coefficients were different between species and effective diffusion coefficients were

obtained by multiplying these by the porosity.

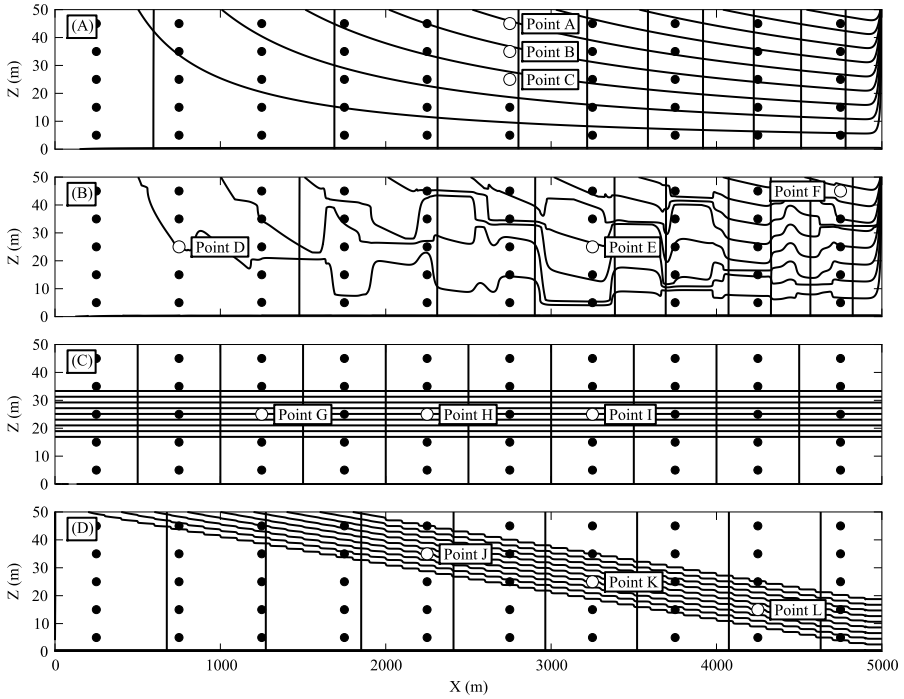


Figure 3.4: Results of simulations of hydraulic heads and streamlines for corresponding aquifers in Figure 3.3. Dots represent locations where age simulations were undertaken (labelled points will be explored in further detail).

Figure 3.4 represents the head and streamline results of the theoretical aquifers. In Figure 3.4A, flow is controlled by recharge and discharge features. In Figure 3.4B, we see that flow lines are highly refracted towards high K lenses. These lenses bring together waters from different flow paths of different ages. In Figures 3.4C and 3.4D, flow occurs predominantly through the high K layers.

Figure 3.5 presents groundwater age distributions, advective ages and mean ages for selected points identified in Figure 3.4. For points in the homogeneous aquifer (points A, B and C), distributions of ages are relatively narrow and mean and advective ages show good agreement. In the case of wells from the aquifer with high K lenses (points D, E and F), distributions become multimodal (as evidenced by multiple inflexion points), due to the different flow paths drawn to the high K lenses. For points D and E, advective and mean ages show good agreement. However in the case of point F, this difference becomes greater. In the two cases where layered flow is simulated (Points G through L), significant tailing is observed. This is due to exchanges between the relatively mobile and relatively immobile parts of the flow system. These situations represent the largest errors between mean and advective ages.

Figure 3.6 presents the relationships between some apparent ages and advective ages. For the homogenous aquifer (Figure 3.3A), apparent CFC-12 ages are over-estimated for advective ages less than 40 years and underestimated for advective ages older than this. Apparent tritium/helium ages have a similar bias for

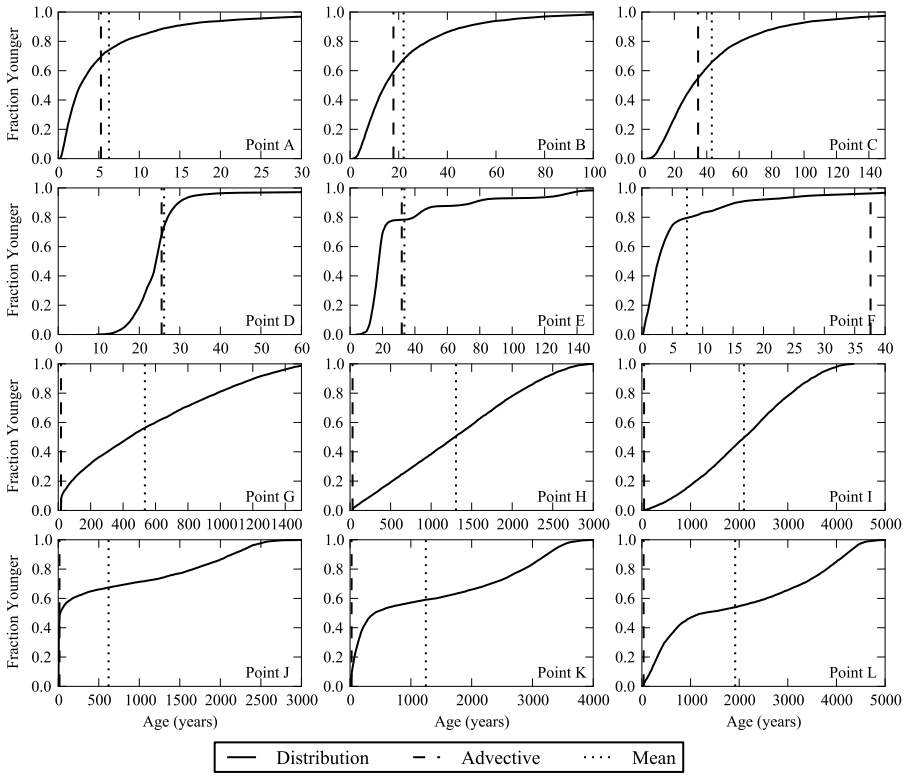


Figure 3.5: Advective ages, mean ages and age distributions for points presented in Figure 3.4.

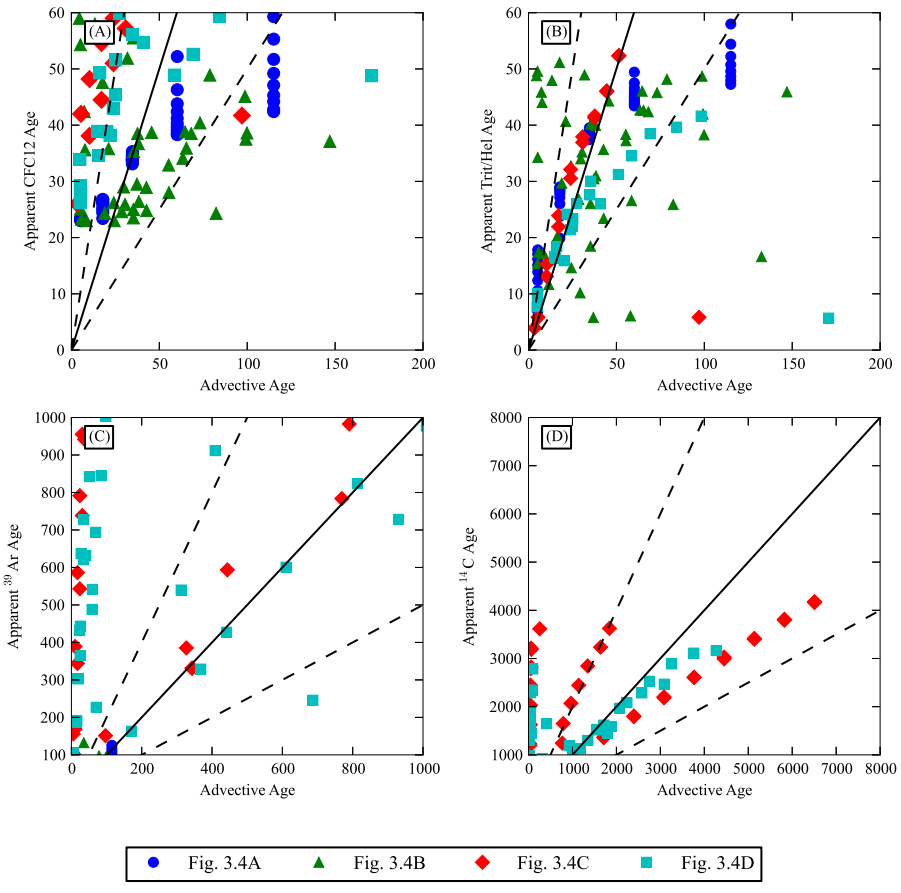


Figure 3.6: Some apparent ages plotted against advective ages for all points shown in Figure 3.4. Solid lines represent a 1:1 line and dashed lines represent 50% errors.

older advective ages. As demonstrated in Figures 3.5A-C, distributions at these points are narrow and mean ages show good agreement to advective ages. In this case, these biases relate to temporal variations of the environmental tracer input concentrations. Comparisons between tritium/helium ages have been demonstrated in one dimensional studies (Schlosser et al., 1989; Ekwurzel et al., 1994). The higher bias observed here is due to transverse dispersion. Although these errors are larger, Solomon and Sudicky (1991) found the estimates of recharge made using points near the water-table in two-dimensional simulations still yielded favourable results. In the case of the theoretical aquifer with high K lenses (Figure 3.3B), apparent CFC12 and tritium/helium ages show poorer agreement and in some cases the errors are greater than 50%.

For scenarios where layered flow is observed (Figures 3.3C and 3.3D), a large portion of apparent CFC12 ages are overestimated compared to advective ages. These locations relate to mobile zones where diffusion slows the movement of the species used to estimate apparent ages. Points where ages are underestimated are located in immobile zones, where diffusive exchange occurs with faster flow paths. Interestingly, over the same range, apparent tritium/ helium ages show good agreement to advective ages, with the exception of points from the aquifer in 3.3D for ages greater than 40 years. When comparing ^{39}Ar and ^{14}C apparent ages to advective ages, poor agreement is observed. Argon-39 ages are generally overestimated however a small number are underestimated. Carbon-14 apparent ages show three distinct groups. These are related to the position within the mobile or immobile components of the aquifer. As the points are not screened across the entire mobile portion of the aquifer, points located closer the aquitard experience a higher level diffusive exchange, despite the flow configuration resulting in the same advective age. The importance of diffusion in the estimation of ages has been the motivation of previous studies (Sudicky and Frind, 1981; Sanford, 1997).

Figure 3.7 presents the relationships between mean ages and apparent ages for the theoretical aquifers presented in Figure 3.3. For the homogenous aquifer (Figure 3.3A), apparent ages agree better with mean ages than advective ages. In almost all cases where CFC12 and tritium/helium are used to estimate apparent ages, estimates are within 50%. In the case of an aquifer with high K lenses, agreement between mean and apparent ages appears similar to that of advective and apparent ages. In the cases where layered flow is simulated (Figures 3.3C and 3.3D) apparent ^{39}Ar ages show a bias in under predicting mean ages. In the case of the sloping aquifer, some of these errors are greater than 50%. For the same scenarios, ^{14}C ages are in excellent agreement to mean ages. This is due to the half-life being large relative to the travel times. This is consistent with the theory presented by Varni and Carrera (1998) and the results demonstrated by Larocque et al. (2009) with regard to the relationship between accuracy and half life.

Table 3.1 presents the errors of the estimates of mean and advective ages made using apparent ages. These estimates were calculated as mean absolute relative errors:

$$Error = \frac{1}{n}\sum \frac{|obs - act|}{obs} \tag{3.1}$$

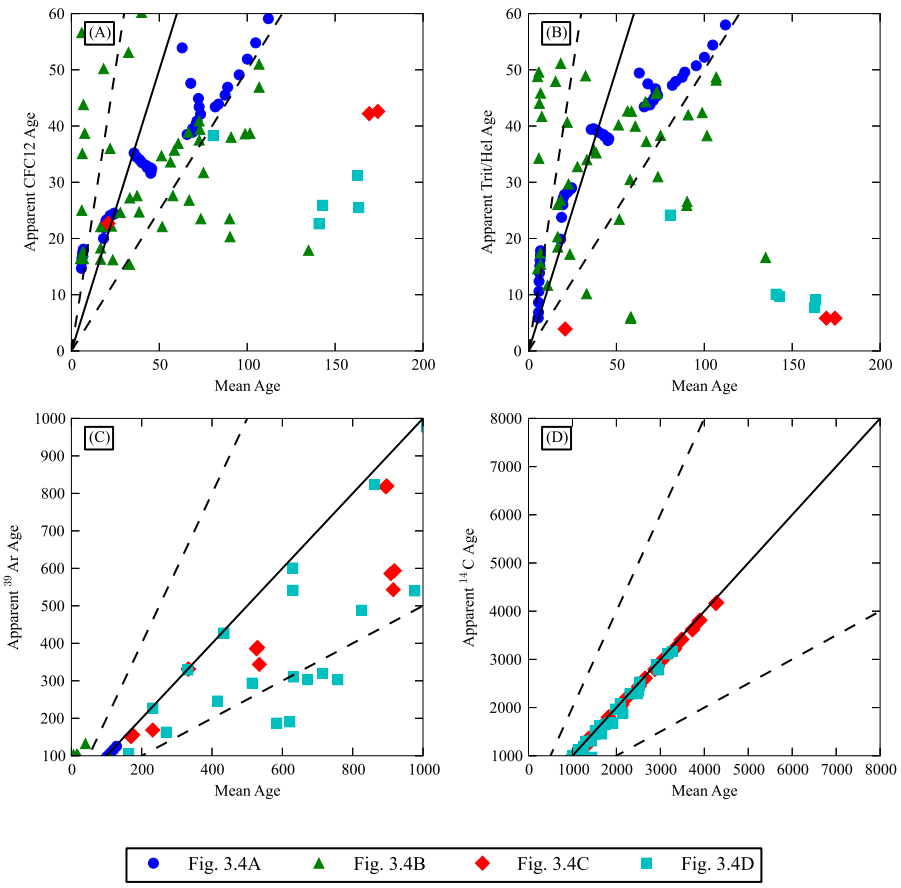


Figure 3.7: Some apparent ages plotted against mean ages for all points shown in Figure 3.4. Solid lines represent a 1:1 line and dashed lines represent 50% errors.

where n is the number of samples, *obs* is the apparent age and *act* is the mean or advective age. The errors were reported relative to the observed age, as this is more practical in the application of errors.

Errors are quite high for both mean and advective ages when estimating apparent ages using young tracers in all scenarios (CFCs, SF_6, and tritium/helium, Table 3.1). Errors are larger for lens and layered flow scenarios consistent with Figures 3.6 and 3.7. For tritium helium estimates, errors in the case with high K lenses are of the same order as the homogenous case. However, much larger errors than for other young tracers are observed in the aquitard scenarios. This is due to the varying diffusion coefficients of the two compounds. This impacts ages not reported on the scales of Figures 3.6 and 3.7. Interestingly, for all scenarios outside of the homogeneous case, advective ages are better approximated than mean ages. However, these errors can reach 100% or higher. In the case of radioactive tracers (^{39}Ar and ^{14}C), apparent ages are in better agreement to mean ages. Carbon-14 ages are more accurate where available due to the larger half-life relative to the width of distributions.

In general, this section has demonstrated a few things. Dispersion, heterogeneities and diffusion with aquitards cause age to become a distribution. These distributions impact the accuracy of apparent ages obtained using environmental tracers. In the cases presented here, apparent ages agree better with mean ages than advective ages for the case of the homogenous aquifer and the aquifer with layered flow when

Table 3.1: Summary of age errors for apparent and corrected ages. Values are in mean absolute relative error as calculated using Equation 3.1.

		Advective				Mean			
		Fig 3.4A	Fig 3.4B	Fig 3.4C	Fig 3.4D	Fig 3.4A	Fig 3.4B	Fig 3.4C	Fig 3.4D
Apparent age	CFC-11 [1]	52%	68%	157%	720%	50%	82%	606%	1183%
	CFC-12 [1]	47%	89%	88%	62%	46%	104%	438%	1287%
	CFC-113 [2]	57%	94%	95%	53%	59%	116%	573%	1609%
	SF_6 [2]	34%	191%	94%	39%	35%	212%	617%	2073%
	$^3H/^3He$ [3]	47%	44%	706%	4277%	41%	45%	3837%	5604%
	^{39}Ar [4]	7%	85%	59%	107%	3%	83%	44%	109%
	^{14}C [5]	-	-	64%	42%	-	-	3%	5%
Sanford (1997)	^{39}Ar	-	-	87%	85%	-	-	-	-
	^{14}C	-	-	81%	78%	-	-	-	-
Bethke and Johnson (2002a)	^{39}Ar	-	-	86%	83%	-	-	-	-
	^{14}C	-	-	92%	92%	-	-	-	-
Dispersion Model	CFCs and SF_6	-	-	-	-	8%	128%	156%	229%
	$+^3H$	-	-	-	-	6%	134%	177%	263%
	$+^{39}Ar$	-	-	-	-	6%	134%	187%	256%
Multiple Ages	CFC-12 and CFC-113	-	-	-	-	3%	22%	-	-
	CFC-12 and SF_6	-	-	-	-	4%	20%	-	-
	^{39}Ar and ^{14}C [6]	-	-	-	-	-	-	3%	6%

[1] – calculated for apparent ages less than 50 years
[2] – calculated for apparent ages less than 40 years
[3] – calculated for apparent ages less than 60 years
[4] – calculated for apparent ages greater than 100 years and less than 1000 years
[5] – calculated for apparent ages > 1000 years
[6] – calculated for apparent ^{14}C ages greater than 800 years and apparent ^{39}Ar ages less than 1200 years.

using radioactive tracers. In the case where flow was controlled by high K lenses, apparent ages appeared more representative of advective ages.

3.5 Proposed correction schemes

In this section, we look at some previously proposed correction schemes for apparent ages that may yield accurate estimates of advective or mean ages.

3.5.1 Physically based models

As mentioned above, a number of processes influence the mixing of waters resulting in differences between apparent ages and true groundwater ages. Correction can be achieved using models that account for such mixing. These correction factors are more easily derived for radioactive tracers due to the nature of the temporal variations. Often these correct for diffusive exchange with aquitards and immobile zones (Sudicky and Frind, 1981; Sanford, 1997; Bethke and Johnson, 2002b; Cornaton et al., 2011). The draw back of these correction schemes is that they require some knowledge of aquifer parameters, and that they often require the use of simplified assumptions of temporal variations in concentrations. Quite often, the correction factors will be highly sensitive to these assumed values.

We can apply these correction schemes to the ages interpreted from our synthetic cases. The two corrections chosen were those of Sanford (1997) and Bethke and Johnson (2002b). These two methods are implemented as they rely on knowledge of the relative thickness of mobile and immobile zones, which can be obtained from bore logs. The Sanford (1997) corrected age (t_c) is defined as:

$$t_c = t_u \left(\frac{\lambda}{\lambda + \lambda'} \right) \tag{3.2}$$

where t_u is the uncorrected age, λ is the radioactive decay constant and λ' is the ^{14}C loss due to diffusion defined as:

$$\lambda' = \frac{2w}{\theta a} (\lambda D)^{1/2} \tag{3.3}$$

where w is the width factor:

$$w = \tanh \left[(b/2) (\lambda / D) \right]^{1/2} \tag{3.4}$$

and θ is porosity, a is the thickness of the flow zone, D is the bulk diffusion and b is the thickness of the stagnent zone.

The model of Bethke and Johnson (2002b) compares mean ages to advective ages. Here, we have assumed that the biases from radioactive decay are smaller than those from exchanges between mobile and immobile

zones. Based on Table 3.1, this is more likely accurate for ^{14}C ages than ^{39}Ar ages. In this case:

$$t_c = t_u \left(\frac{a}{a+b} \right) \tag{3.5}$$

Note that this equation has been modified in that we have removed the use of a half width and porosity as the 1/2 in the division cancels, and porosity is constant in our case. The parameters used to estimate the correction factors for the theoretical aquifers presented in Figures 3.3C and 3.3D are presented in Table 3.2.

Table 3.2: Paramaters used for correction schemes.

Parameter	Units	Value
Decay Constant	year^{-1}	1.21×10^{-4}, 2.58×10^{-3}
Flow zone thickness (a)	m	20.0
Stagnant zone thickness (b)	m	30.0
Effective Diffusion Coefficient (D)	m^{2}year^{-1}	0.00365
Porosity (θ)	-	0.1

Figure 3.8 presents the results of the application of some correction schemes to apparent ^{39}Ar and ^{14}C to the layered theoretical aquifers (Figures 3.3C and 3.3D). Here we see that the correction factors result in a general overestimation of kinematic ages. This may be due to sample points not being screened over the entire mobile domain. Table 3.1 presents the comparison of the relative errors. In most cases, the corrections result in larger errors for comparisons to advective ages. Figure 3.8 suggests that a bias exists, in that older ages are predicted. This may be due to assumptions not being explicitly met in the sampled points.

3.5.2 Lumped parameter models

A method for estimating mean ages is using lumped parameter models. In these cases, assumed distributions of groundwater age are used to fit observed concentration data when temporal variations in concentration are known using the convolution integral. These models are based on assumptions of mixing occurring upstream of the sampling location. Common models include the Exponential Model and the Dispersion Model (Maloszewski and Zuber, 1982) and combinations of these models (Stolp et al., 2010). The exponential model assumes uniform mixing across an aquifer receiving constant recharge. The Dispersion model assumes mixing based on the ratio of dispersive to advective forcing upstream. These assumed mixing patterns represent ideal situations that may not be met in the field and are often not validated. The mean transit time obtained from these models is equivalent to a mean age.

The drawback of this method is that there can be a high level of correlation between the parameters of the models used preventing unique estimates from being obtained (Corcho Alvardo et al., 2007; Massoudieh et al., 2012). Additionally, the assumptions of the models are highly limited and are not representative of discretely screened wells.

The relationship between historic concentrations (c_{in}), the groundwater age distribution (g) and measured

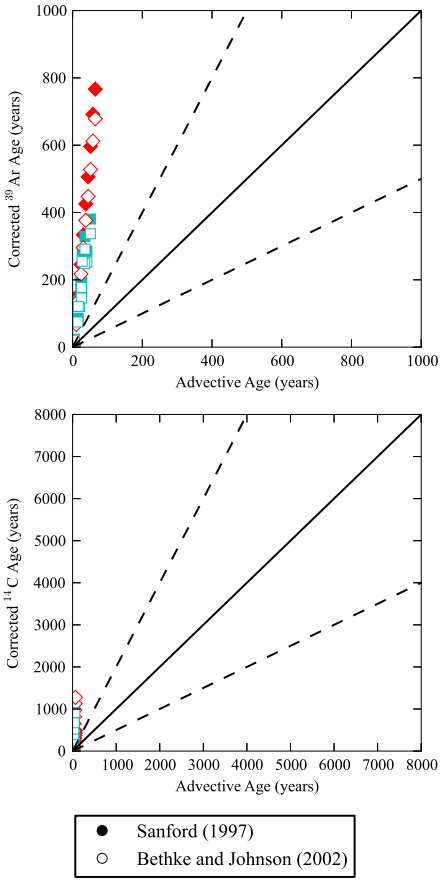


Figure 3.8: Comparison of correction schemes based on aquifer assumptions for points located in high permeability zones (Figure 3.4C and 3.4D). The model of Sanford (1997) corrects for decay and dual domain effects. When using the model of Bethke and Johnson (2002a) only a correction for dual domain effects is considered.

groundwater concentrations (c_{out}) is defined as:

$$c_{out} = \int_0^t c_{in}(t-\tau)g(\tau)\cdot d\tau \tag{3.6}$$

The shape of the groundwater age distribution is obtained from a predefined model. For this example we have used the dispersion model (Maloszewski and Zuber, 1982). This was chosen as it represents at inverse-Gaussian distribution which has been demonstrated to provide good estimates of mean ages regardless of the actual shape of the distribution (Waugh et al., 2003). In this model the age frequency distribution is defined as:

$$g(t) = \frac{1}{t\sqrt{(4\pi P_D t/T_m)}} \exp\left[-\frac{(1-t/T_m)^2}{(4P_D t/T_m)}\right] \tag{3.7}$$

where P_D is the dispersion parameter and T_m is the mean age.

The use of multiple environmental tracer concentrations can allow for the two parameter model to be calibrated. In this case we used the least squares fitting program PEST to calibrate the dispersion model. We implemented three levels of data from the theoretical aquifer simulations: CFCs and SF_6; CFCs, SF_6 and tritium; and CFCs, SF_6, tritium and ^{39}Ar. In application, only locations where CFC, SF_6 and tritium

concentrations were reported were used. The data were weighted by the inverse of the values, giving each observation an equivalent weight. Only estimates where the objective function was less than 5 were considered good fits.

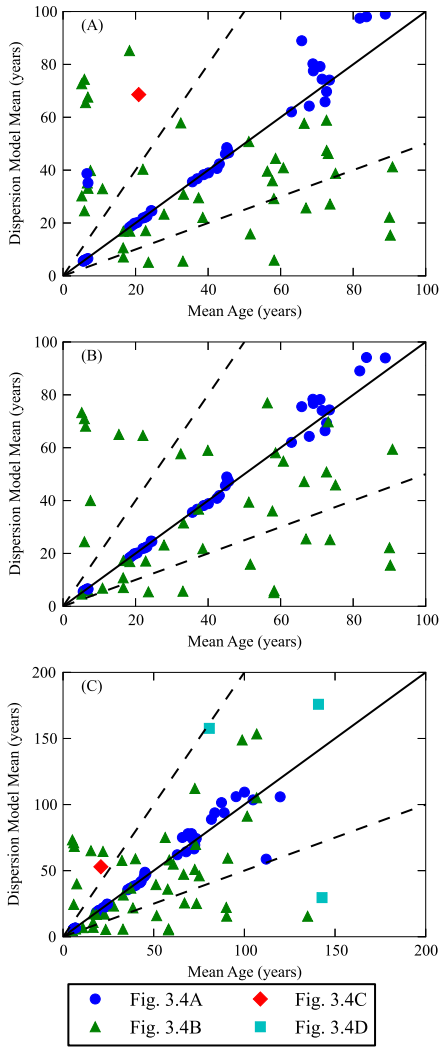


Figure 3.9: Estimates of mean transit time from fitting environmental tracer data with the dispersion model plotted against mean ages for points located in Figure 3.4. Figure 3.9A represents estimates made using CFCs and SF_6, Figure 3.9B represents estimates made using the same plus tritium and Figure 3.9C also used Argon-39. In each case, fits were achieved using a minimisation function.

Figure 3.9 presents comparisons between mean ages and estimates of mean transit time obtained from fitting the dispersion model to environmental tracer concentrations. In the case of the homogenous aquifer, estimates of mean transit time show good agreement to mean ages. In the case of the aquifer with high K lenses, ages do not agree as well. Table 3.1 presents the mean absolute relative errors of the estimates of MTT compared to mean age. Errors are significantly reduced for the homogenous aquifer. In the case of the three heterogeneous aquifers, errors remain high. This may suggest that the tailing observed in these aquifers is un-reconcilable with the assumed model.

3.5.3 Multiple tracer ages

A number of methods for correction exist that require apparent ages obtained from environmental tracer measurements. These corrections can only be for a mean age and not an advective age. Varni and Carrera (1998) presented a correction whereby the use of two radioactive tracers could be used to determine the first two moments of the age distribution. The difficulty of this method however is finding tracers with significantly different half-lives that give accurate readings in the same water sample. Other authors have proposed the use of relationships between tracer concentrations to determine mean groundwater ages (Waugh et al., 2003; McCallum et al., In Press). The drawback to these methods is that errors in estimated concentrations result in differences of a similar magnitude as changes in the mixture of ages. This then makes multiple tracer age methods limited in their range of applicability.

The methods suggested by Waugh et al. (2003) and McCallum et al. (In Press) require a graphical approach. In the case here, an inverse-Gaussian model was used in conjunction with the convolution integral (Equation 3.6). The concentrations arising from this relationship were used to determine apparent ages. Contours of mean ages (known from the inverse-Gaussian model) were then plotted from relationships between apparent CFC-12 and CFC-113 ages and apparent CFC12 and SF_6 ages. Apparent ages from theoretical simulations were then plotted over these relationships and mean ages were determined. The relationship between the apparent age obtained from a radioactive tracer, and the mean and variance of the age distribution, was described by Varni and Carrera (1998) as:

$$A_R \approx A_\mu - \frac{1}{\lambda} \ln\left(1 + \frac{\lambda^2 \sigma^2}{2}\right) \tag{3.8}$$

where A_R is the radiometric age, A_μ is the mean age and σ^2 is the variance of the age distribution. In this case, the apparent age and the radioactive decay constant are known but the mean and variance are unknown. Thus, if two apparent ages from two tracers with different decay constants are known, the mean and variance can be determined. In the theoretical case presented here, both an apparent ^{39}Ar age (A_{R1}) and an apparent ^{14}C age (A_{R2}) are known so we can implement the function:

$$f\left(A_\mu, \sigma^2\right) = \sum_{i=1}^{2} \left[A_{Ri} - A_u + \frac{1}{\lambda} \ln\left(1 + \frac{\lambda_i^2 \sigma^2}{2}\right)\right]^2 \tag{3.9}$$

where λ_1 and λ_2 are the radioactive decay constants of ^{39}Ar and ^{14}C respectively. In this case, the equation was minimised by adjusting the values of A_μ and σ^2 using the BFGS method within the "minimize" function in SciPy (Jones et al., 2001–).

Figure 3.10 presents some estimates of mean age based on differences between multiple tracer estimates of apparent ages. Figures 3.10A and 3.10B show considerable scatter in the application of the methods proposed by Waugh et al. (2003) and McCallum et al. (In Press). However, there is still a general trend, particularly for the homogenous aquifer. Figures 3.10C and 3.10D present estimates of the mean and

variance of age distributions based on theory presented in Varni and Carrera (1998). Here, it is shown that using two apparent ages from tracers with different half-lives, we can estimate the mean of the age distribution. Table 3.1 shows that the use of graphical methods offers improvement over the direct use of apparent age when determining mean age. Interestingly, although the use of two radioactive tracers reduces errors when compared to ^{39}Ar apparent ages, the magnitude of errors is similar to when using ^{14}C independently.

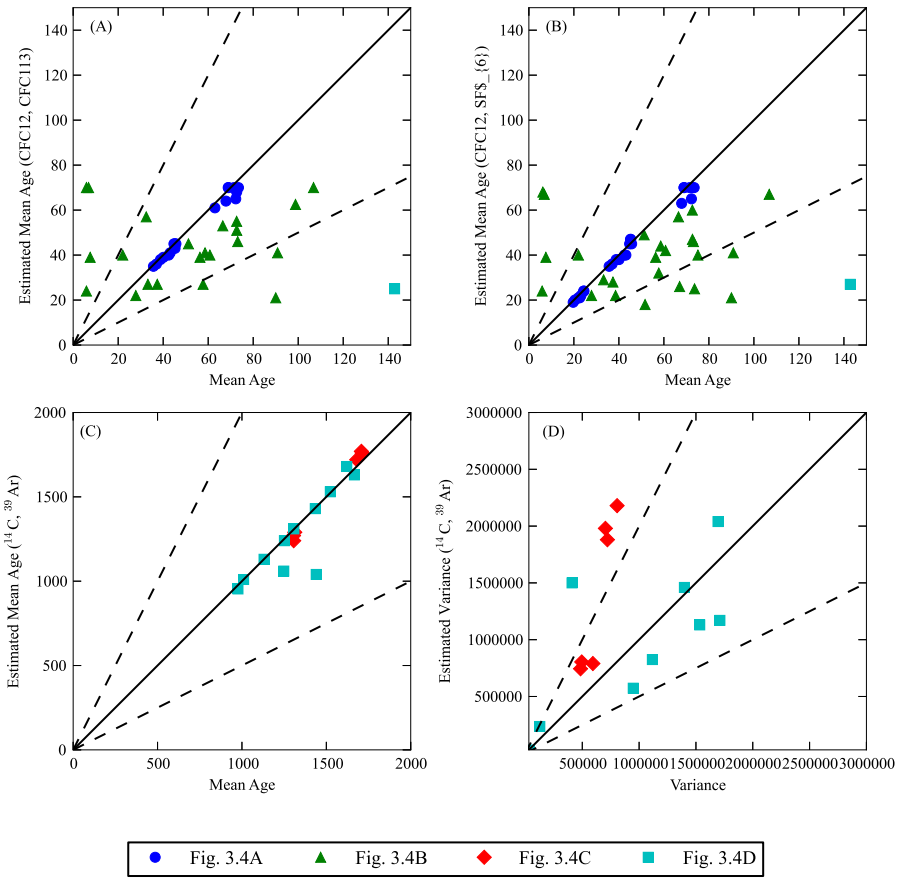


Figure 3.10: (A and B) Estimates of mean age using the differing bias of multiple tracers and estimates of mean age (C) and Variance of the age distribution (D) using the relationship presented by Varni and Carrera (1998) and ^{14}C and ^{39}Ar concentrations.

3.6 Discussion

In this paper, we have shown the concepts of groundwater ages and their relationships. We have demonstrated that the age of groundwater at a point in an aquifer is made up of a distribution of ages. Within this distribution, a mean age exists which represents the mean of the distribution. Outside of these two ages, others exist. These are advective ages and apparent ages. Advective ages will not relate to mean ages or age distributions in situations where streamlines with differing ages are in close proximity (Varni and Carrera, 1998). Apparent ages will represent mean ages only if distributions are narrow and tracers can represent ages over the entire distribution. For example, young tracers were unable to estimate mean ages when distributions became wide. However, in the scenarios investigated here, apparent 14C age provided good

approximations of mean ages even when uncorrected. This was due to the scale of the flow systems investigated. Our work has highlighted that in some cases mean ages may be determined by measuring multiple tracers and accounting for mixing biases. The errors however remain large outside of the homogenous case.

As stated above, the ability of apparent ages to estimate advective ages is dependent on the variations between the age of adjacent streamlines and the width of distributions. A number of papers have demonstrated cases where these comparisons have yielded accurate results. Solomon and Sudicky (1991) demonstrated the applicability of the tritium/ helium dating technique to the estimation of groundwater recharge when using two points close to the water table. These results have also been demonstrated in one-dimensional studies (Schlosser et al., 1989; Ekwurzel et al., 1994). The second group of authors also extended this to the use of CFCs and 85Kr. Here they demonstrated that despite errors of ages at individual depths, the use of a model for recharge estimation still resulted in only small errors in estimates of recharge. These studies were all undertaken in homogeneous conditions and are similar to those shown for the aquifer in Figure 3.3A where distributions were narrow and mean ages reflected advective ages.

In an assessment of environmental tracers in heterogeneous environments, Larocque et al. (2009) compared environmental tracer techniques with hydraulic techniques. The environmental tracer techniques used artificial tracers with ideal half-lives. The authors found that in general, an age differencing technique worked better with the best tracer being one with a half-life similar to the travel time. The authors used 2D multi-Gaussian fields, which behave as 1D equivalent media after sufficient distance (Gelhar, 1993). Additionally, the finding of the half-life association with accuracy is consistent with the study of Varni and Carrera (1998).

A large number of theoretical studies have also been associated with errors in comparisons between apparent ages and mean or advective ages. Sources of these errors in these relationships include transverse dispersion (Castro and Goblet, 2005), dual domain effects (LaBolle et al., 2006; Neumann et al., 2008), exchange with aquitards (Sudicky and Frind, 1981; Sanford, 1997; Bethke and Johnson, 2002b; Cornaton et al., 2011), matrix diffusion in fractured rock aquifers (Cook et al., 2005), mixing in coastal aquifers (Vandenbohede et al., 2011; Post et al., 2013) and heterogeneity (Weissmann et al., 2002; McCallum et al., In Press).

Although not the focus of this work, errors associated with the compounds used to date groundwater are also important. Cook and Solomon (1997) suggested that in the case of young tracers, analytical errors result in age errors less than 10%. Our work has demonstrated that in these cases, the assumption of the model used to determine apparent ages results in much greater errors. In the case of ^{14}C apparent ages, errors arise from calcite dissolution, dolomite dissolution and oxidation of old organics (Clark and Fritz, 1997). Additionally, although correction methods using multiple tracers appear adequate, they rely on the assumption that the two compounds move through the aquifer at the same rate. For example, CFC-113 has been demonstrated to have retarded transport (Bauer et al., 2001). In such cases these methods would not yield favourable results.

Although large errors are determined between apparent ages and mean or advective ages, their use in groundwater models may still provide favourable results. Janssen et al. (2008) found that the use of ad-

vective ages in groundwater models predicting solute breakthrough outperformed models calibrated with head and conductivity data, even with highly uncertain age data. Ginn et al. (2009) found that the inclusion of mean age data improved estimates of the K-field, even assuming the age observations were five time less accurate than head observations. Other studies have shown that these relationships may be highly non-unique in that apparent ages may represent many mean ages and parameter sets (Weissmann et al., 2002; Leray et al., 2012). Troldborg et al. (2007) suggested that environmental tracer data may be valuable if used to accept or reject conceptual models. Additionally, environmental tracer data may be included in models as concentrations rather than ages (Bauer et al., 2001). Starn et al. (2010) demonstrated that the use of tritium concentrations was able to constrain well catchment areas.

Our work highlights the importance of understanding exactly what measured ages represent. If the wrong comparison is made between what is measured and what is inferred by an interpretive model then estimates may be highly inaccurate. It appears however that correcting for mixing biases may be useful in estimating mean ages and these mean ages may be used for comparison when constraining groundwater flow problems. The accuracy of direct comparisons between advective or mean and apparent ages is dependent on conditions within the aquifer. These include the variability of ages between adjacent flow paths and the degree of mixing between flow paths. Generally, little is known about these conditions a priori hence, caution should be exercised.

3.7 Conclusions and recommendations

Knowledge of mean or advective ages allows for model calibration or estimates of water balance components. Unfortunately the methods we implement to infer these things do not necessarily represent what we want them to as the assumptions of models are often not met outside of ideal circumstances. These non-ideal circumstances arise due to heterogeneity, complex flow systems and the nature of tracer methods investigated. For this reason, an understanding of what an apparent age does and does not represent is important before application. Outside of ideal circumstances, corrections are required. Generally, corrections that relate apparent ages to advective ages require some information about the groundwater flow system. However, corrections for mean ages may be achieved in some cases by measuring multiple tracers with no underlying assumptions of aquifer structure or age distribution. Our work suggests that these corrected mean ages may provide better calibration targets than advective ages.

Acknowledgments

Funding for this work was provided by the Australian Research Council and the National Water Commission.

4 Manuscript 3: Impact of data density and geostatistical simulation technique on the estimation of residence times in a synthetic two-dimensional aquifer

This chapter is similar to an accepted manucript in Mathematical Geosciences, with the exception of revision comments.

Authors: James L. McCallum, Daan Herckenrath and Craig T. Simmons

Abstract

Connectivity patterns of heterogeneous porous media are important in the estimation of groundwater residence time distributions (RTDs). Understanding the connectivity patterns of a hydraulic conductivity (K) field often requires knowledge of the entire aquifer, which is not practical. As such, the method used to estimate unknown K values using known K values is important. Here we investigate how varying levels of conditioning data and four simulation techniques – one multi-Gaussian and three multi-point, are able to recreate key K field features and connectivity patterns of a synthetic two-dimensional bimodal distributed ln(K) field with highly connected high K features. We then assess these techniques in the context of RTD estimation. We found that the multi-Gaussian technique presented a bias towards earlier travel times with increased conditioning data. This was due to the inability of the method to recreate multiple scales of connecting features. Of the multi-point methods investigated, the facies method was unable to predict early arrival times. The use of a continuous variable training image produced good fits to the observed residence time distribution with a high number of conditioning points. The ability of the methods to predict the shape of residence time distributions appears to be related to their ability to reproduce the connection patterns of higher K features.

4.1 Introduction

Residence time distributions (RTDs) in groundwater flow systems are of interest as they infer the timescales of solute transport processes. As such, a large amount of interest has been given to their simulation and prediction (Varni and Carrera, 1998; Ginn, 1999; Cornaton and Perrochet, 2006a). This is due to the mixing of groundwater age being governed by the same principles that govern solute transport processes (Ginn, 1999). These distributions have been found to be sensitive to aquifer heterogeneity (Weissmann et al., 2002; Cornaton and Perrochet, 2006b; Engdahl et al., 2012), porosity (Cornaton and Perrochet, 2006b; Leray et al., 2012) and flow field dynamics (Cardenas, 2007; Cardenas and Jiang, 2010).

The relationship between the structure of hydraulic conductivity (K) fields and aquifer residence times and solute transport processes has received some attention. Desbarats (1990) found that RTDs in bimodal

sand and shale sequences were controlled by the connectivity of more permeable sandstone units. Gomez-Hernandez and Wen (1998) and Wen and Gomez-Hernandez (1998) found that residence time distributions in heterogeneous aquifers were dependent on the connection patterns of extreme hydraulic conductivity values. Zinn and Harvey (2003) demonstrated that K fields with the same variogram properties produced quite different solute transport properties due to differences in connectivity between low, medium and high K values. Schluter and Vogel (2011) showed that fields created with the use of connectivity metrics in conjunction with simulated annealing techniques were able to produce solute transport properties of reference fields, demonstrating the importance of these connectivity properties.

The difficulty is that a complete and accurate understanding of the connection patterns requires knowledge of the exhaustive K field when in reality only values of K at discrete points are known (Renard and Allard, 2013). As a result, simulation techniques are usually required to produce a complete K distribution with inherent connectivity features. These techniques can be the product of two-point models or multi-point models. Typical two-point models include kriging and multi-Gaussian models. The benefit of these methods is implementation only requires knowledge of the distribution and spatial correlation of K values. The assumptions of two point models have been showed to be inadequate when connection exists between extreme values (Gomez-Hernandez and Wen, 1998; Wen and Gomez-Hernandez, 1998; Zinn and Harvey, 2003; Kerrou et al., 2008). Multiple point methods offer an alternative to two-point methods as they maintain connectivity patterns of the K field (Strebelle, 2002). The drawback of these methods is that detailed information is required to construct an adequate training image (Mariethoz and Kelly, 2011).

In this study, we investigate the ability of four different simulation methods to predict connectivity patterns and subsequent residence time distributions in an aquifer with a bi-modal K distribution. The reference K field used to assess the predictions was generated based on a transformed image of a stream environment with highly connected channel features with varying widths. Three levels of conditioning data are used – consisting of 12, 130 and 1200 points. The four simulation techniques include a multi-Gaussian method, two methods that employ basic training images and a fourth method employing a full accurate training image. We then compare how well these methods recreate the connectivity patterns and residence time distributions of the reference field.

4.2 Methods

4.2.1 Conceptual model

The aim of this work was to assess a number of methods for recreating properties of a two dimensional K field with highly connected features and large contrasts in K. Given that the connection, flow and transport are all controlled by 3D processes, we are not attempting to replicate an exact field site but rather focus on the general abilities of the tested methods. For this reason, a base case was chosen such that the flow and transport processes were controlled by highly connected high K channels set in a relatively low K matrix.

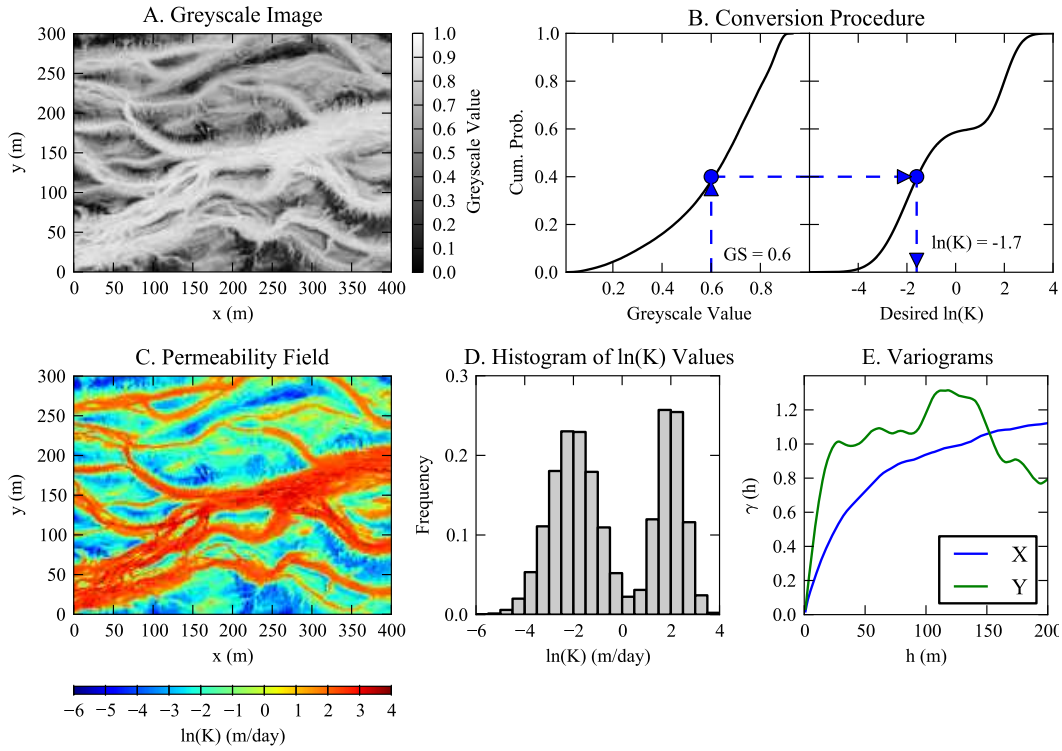


Figure 4.1: (A) Photo of braided stream environment. (B) Continuous permeability field generated by transforming the photo (A) to a permeability distribution. (C) Location of facies (red is channel and blue is floodplain) and (D) isotropic, x-directional and y-directional variograms (after conversion to a standard normal distribution).

For defining the base case, a greyscale image of the Diamantina River in QLD, Australia was used (Figure 4.1A). This image was chosen to create a hypothetical example that has complex curvilinear, highly connected forms at various scales. A connected river was used as example to amplify the effect of connecting features in our modelling study. Although connecting features in aquifers that stem from such river systems might undergo significant changes due to complex sedimentation and sediment reworking processes, similarities can be expected with the dynamic spatial characteristics of the river system. However, this case is not meant to represent an actual depositional environment, but rather a case to test methods for reproducing complex features, and how this impacts on the estimation of RTDs.

To map the image to a K field, greyscale values were converted to a desired distribution. This was achieved by creating a cumulative probability of the greyscale values, and assigning an ln(K) value based on the equivalent probability of the desired ln(K) field (Figure 4.1B). A bimodal distribution was chosen to accentuate differences between the connected high K channel features and the low K matrix. The resulting permeability field and the histogram of ln(K) values are presented in Figures 4.1C and 4.1D respectively. The selected K values cover a range from silts through to sands (Domenico and Schwartz, 1998).

The model domain was 400 m (x-direction) by 300 m (y-direction) and of unit depth, simulating two-dimensional, horizontal flow. Groundwater flow was driven by an upstream head of 1.8 m and a downstream head of 1.0 m to simulate a hydraulic gradient of 0.002. For solute transport simulations we used the

following properties: a porosity of 0.3, longitudinal and transverse dispersivities of 0.04 m and a tortuosity (defined as distance travelled over linear distance) of 1.7. A molecular diffusion coefficient of 0.0365 m^2/y was used to represent the free water diffusion coefficient. These dispersivities were chosen after Weissmann et al. (2002). Low values were chosen so that the dispersion of the RTDs reflected the heterogeneity. Additionally, it has been demonstrated that when simulating heterogeneous aquifers, the transport within an aquifer is controlled by diffusion rather than dispersion if the heterogeneity is adequately resolved (Labolle and Fogg, 2001).

4.2.2 Simulation techniques

Three levels of conditioning data were implemented with each simulation technique – 12, 130 and 1200 points (Figure 4.2A-C). These levels of conditioning points represent 0.01%, 0.1% and 1% of the K field respectively. The points were chosen at continuous distances. In reality, a large number of points at continuous distances are not available, but this represented an ideal setup for comparing the different simulation methods. In the context of this study, the term simulation refers to the generation of equally probable values between known conditioning points. These equally probable values are dependent on the selected model. Four simulation techniques were used to generate K values between the locations of conditioning data. The first method used the sequential Gaussian simulator (Deutsch and Journel, 1998). Based on variograms presented in Figure 4.1E, anisotropic correlation lengths were selected to better represent the K field. Correlation lengths of 80 m (x-directional) and 25 m (y-directional) were used in simulations. A sill of one was implemented in both the x and y directions. These variograms were based on data from the whole image and not from conditioning data only. This was done so the model was consistent between conditioning scenarios. In this case, a normal score transform was undertaken to convert the conditioning data to a standard normal distribution (zero mean and unit variance), the sequential Gaussian simulation was undertaken and the ensuing permeability field was back converted to reflect the initial distribution of the data. The two facies – the high K channel regions and the low K matrix regions were analysed with the same variogram. Although this may produce errors, the separate treatment of each would have required an additional step to simulate the location of the facies.

The second simulation technique used implemented a facies model technique. For this case, a $\ln(K)$ threshold value of 0 was selected, as this represented an intermediate value between the two facies (Figure 4.1D). Values above this were selected as high K channel and values below this were selected as low K matrix. These were then implemented with the direct sampling technique using elementary training images and transform invariant distances (Mariethoz and Kelly, 2011). In this method, transformations are applied to the search distances in a direct sampling method. An elementary training image is then searched at random until a match to the transformed pattern is found. The theory behind this method is that complex geological structures can be obtained using relatively simple training images by adjusting the statistics of the transformations. The transformations allow for a controlled increase in the variability that can be simulated with

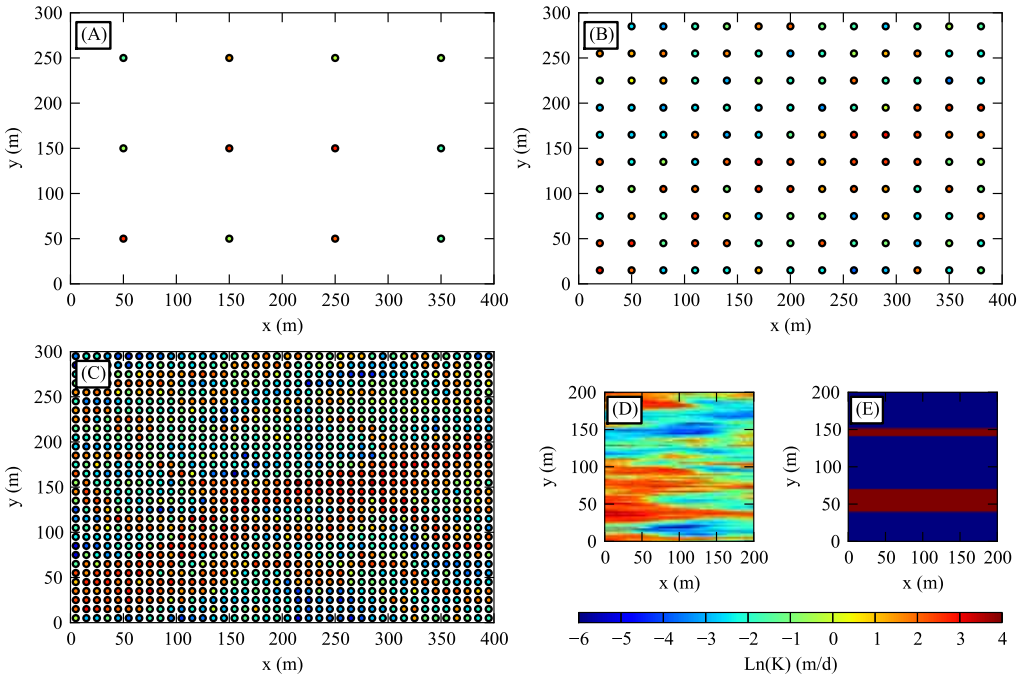


Figure 4.2: A, B and C represent the location of the 12, 130 and 1200 conditioning data points respectively. (D) represents the continuous variable elementary training image used in simulations. (E) Represents the facies training image used in simulations (red is channel and blue is floodplain).

a training image. The method is therefore dependent on the selected training image and the transformations allowed. For the facies model, the elementary training image was selected to represent two scales of channel features (Figure 4.2E). The transformations allowed were affinities of 3 (search pattern squeezed or expanded up to 3 times) and rotations of up to 30 degrees. The rotations were selected as they represented the general angles observed in the baseline images. In practice however, the constraints for these angles may not be known. The affinities were allowed to adjust the widths of the channel features that could be presented. The rotations were implemented to allow for the non-linear nature of the stream features. To obtain a continuous K field, it was assumed that all values in the low K matrix facies had an ln(K) value of -2.0 and all values in the high K channel facies had an ln(K) value of 2.0. These corresponded to the mean values assumed for each of the two facies (Figure 4.1D).

The third technique used was similar to the second but implemented a continuous variable training image. In this case, the training image was generated using the sequential Gaussian simulator (Deutsch and Journel, 1998). Correlation lengths of 1000 m (x-directional) and 25 m (y-directional) were selected to make the 200 m by 200 m elementary training image (Figure 4.2D). The image was initially generated using a standard normal distribution then converted to the distribution of the original K field. The selection of such a large correlation length in the x-direction was to ensure a high level of connection. The selection of the correlation length in the y-direction was based on the variogram of the base scenario. This image was selected as it could be made with limited knowledge of the connection structure of the sub-surface and to represent the continuity of high K channel features within the domain. To account for variations in width and direction of the high K channel features the same transformations applied to the facies model were allowed for the

continuous variable model.

The fourth technique represented an ideal scenario whereby the actual base K field was used as the training image. In reality, this method is impossible as the structure of the K field is unknown. However, this provided a best case scenario for comparison. In this case, the direct sampling method was used (Mariethoz et al., 2010). The training image used was the baseline K field presented in Figure 4.1C. For this case, a low error tolerance was used, representing a high confidence in the training image. Each method of simulation was implemented for the three levels of conditioning data. In all cases, 100 realisations of each K field were generated.

4.2.3 *K* field reproducibility

To assess how well the various methods reproduced the K fields, error maps and connectivity methods were implemented. Error maps were generated by finding the average difference between the base $\ln(K)$ field and the 100 realisations of each of the 12 scenarios. In such cases, negative values indicate the $\ln(K)$ value at a point is lower than the reference field and positive values indicate that the $\ln(K)$ value is higher than the reference field.

In addition to the $\ln(K)$ error maps, two measures of connectivity were implemented – the Euler number and the probability of connection (Renard and Allard, 2013). For both metrics, the domain is required to be partitioned into two phases – the permeable and impermeable phase defined by a threshold value. If the value of $\ln(K)$ at a point is greater than the threshold value then it is deemed to be in the permeable phase. If it is less than or equal to the threshold value then it is deemed to belong to the impermeable phase. The two metrics used measure how well these two phases are connected. The Euler number is a measure of topology. For a 2D grid, the Euler number represents the number of clusters minus the number of holes in the clusters. The Euler characteristic of the permeable or impermeable phase X in two dimensions is defined as (Renard and Allard, 2013):

$$\phi(X) = \sum_{i=0}^{2} (-1)^i \#e_i(X) \tag{4.1}$$

where $\#e_0$, $\#e_1$ and $\#e_2$ are the number of sites in X, the number of edges and the areas of the adjacent sites respectively. The specific method used was the algorithm outlined by Mantz et al. (2008). The algorithm was implemented for both the permeable phase (values above the threshold) and the impermeable phase (values equal to or below the threshold). The Euler characteristic was calculated for $\ln(K)$ threshold values between -6.0 and 4.0.

The second measure of connectivity is the probability of connection, which is defined as the proportion of connected pairs in the permeable or impermeable phase (Renard and Allard, 2013). This is given as:

$$\Gamma = \frac{1}{n_P^2} \sum_{i=1}^{N(X_P)} n_i^2 \tag{4.2}$$

where n_P is the number of permeable (impermeable) cells in X_P and N is the number of clusters in X_P and n_i is the number of cells in cluster i. If all cells in the permeable phase are connected, then the probability of connection for the specified threshold value will be one ($n_i = n_P$, $N = 1$). As with the Euler number, the probability of connection curve was calculated for numerous threshold values and for both the permeable and impermeable phases as described by Renard and Allard (2013).

4.2.4 Flow and transport simulation

The problem domain was discretised using a uniform 1 m by 1 m grid. A steady-state flow solution was obtained using the finite-element model HydroGeoSphere (Therrien et al., 2006). The head gradient was imposed using constant head values of 1.8 m at x = 0 m and 1.0 m at x = 400 m. Head boundaries were chosen over constant flux boundaries so that inflow and outflow fluxes reflected the heterogeneity of the field. The finite-element approximation of the saturated steady-state simulation forms a system of linear equations. These equations were solved iteratively until a residual value of 10^{-9} m had been reached.

Aquifer RTDs were simulated using a random walk particle tracking technique. The advantages of this technique in heterogeneous media include the elimination of numerical dispersion when using small dispersivities, and a reduction in computation times (Salamon et al., 2006). The method used was the Ito-Fokker-Plank Interpolation method (LaBolle et al., 1996). The accuracy of this method has been demonstrated in both multi-Gaussian and non-Gaussian K fields (Salamon et al., 2006). In this method, interpolation of the velocity field is used to prevent discontinuities (LaBolle et al., 1996). To determine the RTD of the whole domain, 10^5 particles were placed in a flux weighted manner along the upstream boundary (x = 0 m). Although neither a flux boundary or a head boundary are realistic, the use of a flux boundary and uniform particle injection would have led to a larger portion of slow flow paths. Additionally, this would have allowed injection conditions and flow rates between scenarios to be identical. As such the use of head boundaries can result in large differences in mean transit times (time to flush the pore water volume) and exaggerated differences in RTDs. Travel times of individual particles were determined to be the time taken to exit the domain at x = 400 m. Estimates of the RTD were made by constructing a cumulative distribution of the 10^5 particles (Weissmann et al., 2002).

4.3 Results

4.3.1 Baseline simulation

The baseline simulation is presented in Figure 4.3. Hydraulic heads and flow through the system are largely controlled by high permeability features (Figure 4.3A). In particular, large amounts of flow occur through the three channel features in the lower portion of the domain. This is demonstrated by the concentration of streamlines in these locations. A smaller amount of flow is observed in the upper portion of the domain. This is due to the occurrence of smaller channels that do not connect across the whole of the domain.

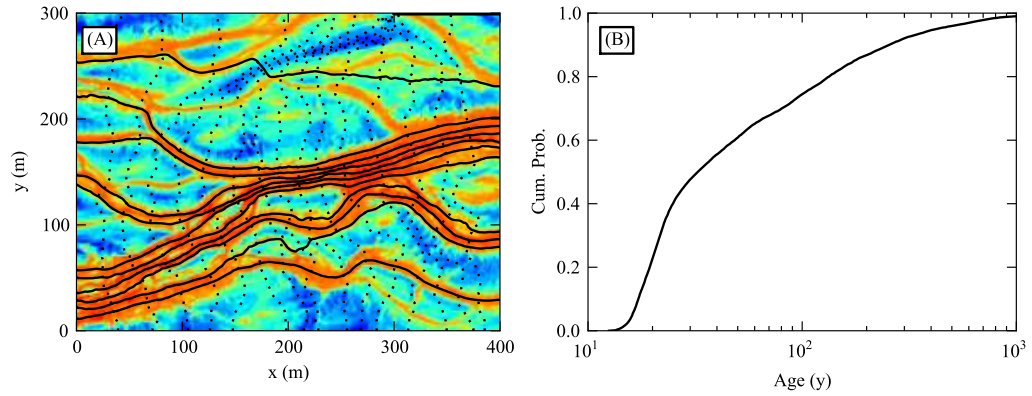


Figure 4.3: Results of baseline simulation head (dotted lines) and streamlines (solid lines) (A), and whole of domain residence time distributions (B).

Figure 4.3B presents the RTD for the whole domain. A large portion of travel times through the system are less than 100 years (approximately 70%). Half of the travel times are less than 30 years. This indicates the significant contribution of the large well-connected channel features. Transit times reach 1000 years in some cases. These later residence times are due to diffusion and slow advection in the less conductive regions of the domain.

4.3.2 Geostatistical simulations

Figure 4.4 represents a single realisation of each geostatistical simulation technique and level of conditioning data. When using the SGSIM, Facies and Elementary training image methods, and 12 and 130 conditioning data points, the methods are unable to recreate the main structures of the baseline field (Figure 4.1C). When using 1200 conditioning points, all methods reproduce larger channel features. The ability of the methods to reproduce the intermediate and smaller channel features appears to be quite variable between the methods. The SGSIM method seems to produce these as smaller isolated blobs. The multi-point methods (Facies and ETI) recreate some channel features but others are missing entirely. The use of the actual aquifer as a training image results in perfect re-creation of the K field with all levels of conditioning data. This was possible due to the use of a large search radius and low error tolerance.

4.3.3 Error maps

Initially, the ability of the simulation techniques to reproduce the actual $\ln(K)$ field was assessed. Figure 4.5 presents the error maps for the $\ln(K)$ field scenarios. In these maps, negative numbers infer that the actual $\ln(K)$ value was underestimated (positive numbers imply the opposite). In the case of multi-Gaussian fields when only 12 points are used, both high K channel and low K matrix regions are poorly recreated. This is consistent with observations from Figure 4.4 – that the method is unable to represent the structure of the K field. When using 130 conditioning points, the ability of the method to represent some portions of the high K channels is improved. However, narrow channel features and low K matrix features are poorly represented.

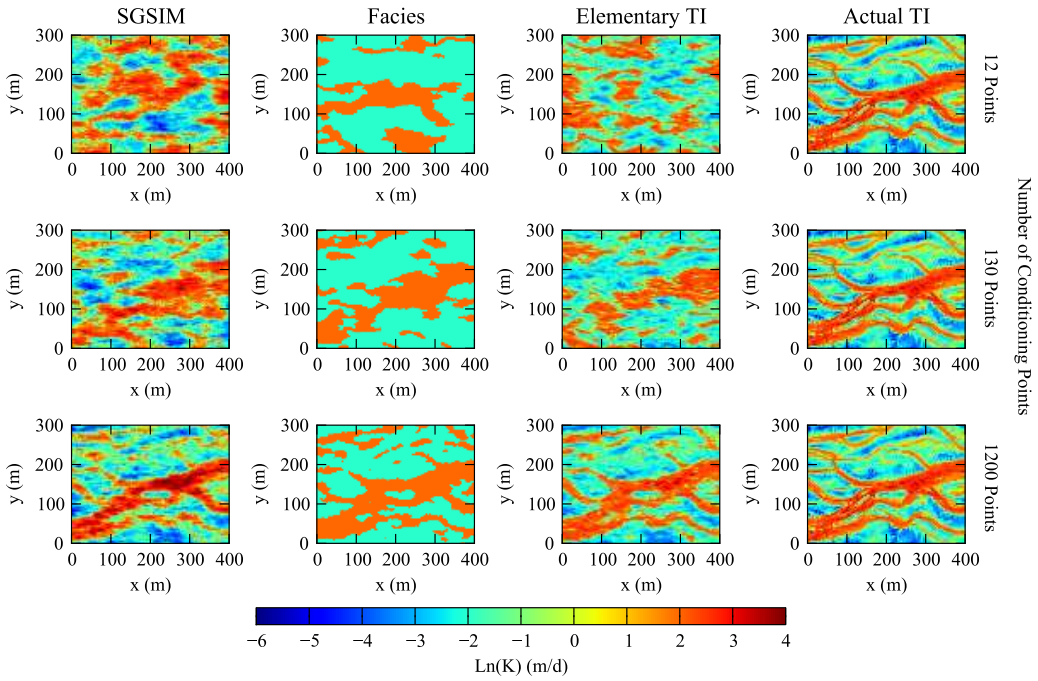


Figure 4.4: Example K fields for each of the simulation techniques and levels of conditioning data.

When using 1200 conditioning points, most channel features are represented well but the matrix features show a consistent bias towards higher values. This may be due to the use of the same variogram model for both facies.

When using a facies model the errors are two-fold. Firstly, errors occur due to the inability to match the structure of the field and secondly due to differences between the mean facies ln(K) value and the actual ln(K) value at a point. When using less than 1200 conditioning points, the errors with the facies method appear to be structurally related (i.e. – error patterns are determined by facies). This suggests that the method is unable to recreate the features of the K field . With only 12 points, the method appears unable to match the features of the reference image. As with the SGSIM method, the use of 130 points allows for some of the main channel features to be recreated. When using 1200 conditioning points, some small channel features are consistently under predicted. This is consistent with Figure 4.4, where some channel features are not represented. Interestingly, errors in the low K matrix do not present the same bias as the multi-Gaussian methods.

K fields generated with a continuous variable elementary training image and transform invariant distances fail to recreate channel features of the reference field with only 12 conditioning points. This is demonstrated by the patterns of the error map generally reflecting those of the reference field. Increasing the number of conditioning points to 130 improves the fit to data in the region of large channel features. Large errors are observed in the location of small channel features. These features can be under-predicted by a magnitude greater than 3.0 ln(K). When using 1200 conditioning points, the magnitude of ln(K) errors are generally less than 2.0 ln(K). Additionally, more of the small channel features are represented. Errors in the low K matrix region are similar to the facies model and do not present the bias of the multi-Gaussian model.

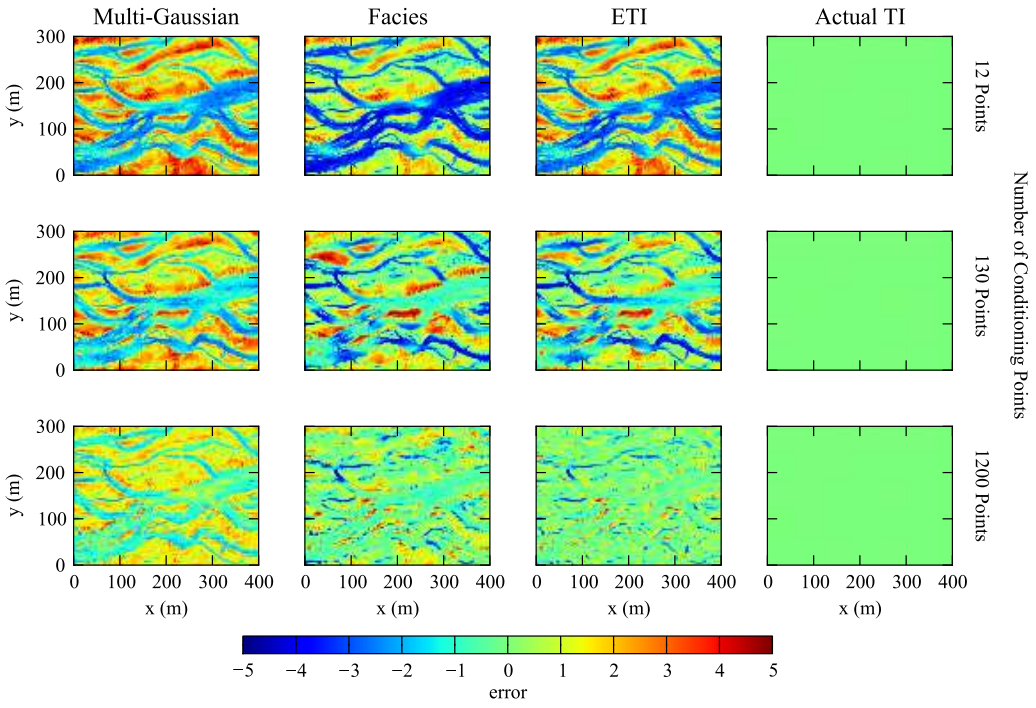


Figure 4.5: Errors between the ln(K) values of the baseline scenario and of simulated hydraulic conductivity fields.

For the ideal case where the actual structure of the K field is used as a training image, all features are predicted well for all levels of conditioning data. This is due to the exact training image being used.

4.3.4 Connectivity metrics

As stated above the Euler number represents the number of clusters in a phase minus the number of holes in those clusters. For the baseline scenario, high values of the Euler value are observed for the permeable phase at low thresholds (Figure 4.6). This suggests that a large number of isolated clusters are present in this phase. The impermeable phase shows the opposite behaviour. The low values indicate that a relatively low number of clusters contains a relatively large number of holes. This may suggest that connection occurs through narrow but tortuous channel features. For higher threshold values, high values of the Euler number for the impermeable phase suggest isolated clusters. Conversely, low values of the Euler number for the permeable phase over the same range may suggest that high K paths are connected by narrow tortuous paths.

Figure 4.6 also presents the Euler number with varying threshold value for the various K field simulations. The general patterns of the Euler number of K fields generated using the multi-Gaussian method are similar to those of the baseline scenario. This is particularly true for fields generated using 12 and 130 conditioning points. The largest differences appear to be for peak values. Larger differences are observed when using 1200 conditioning points particularly at higher threshold values. These differences likely relate to the method generally overestimating K values in the matrix facies (Figure 4.5), and the tendency for the method to predict the smaller high K channel features as isolated high K points (Figure 4.4).

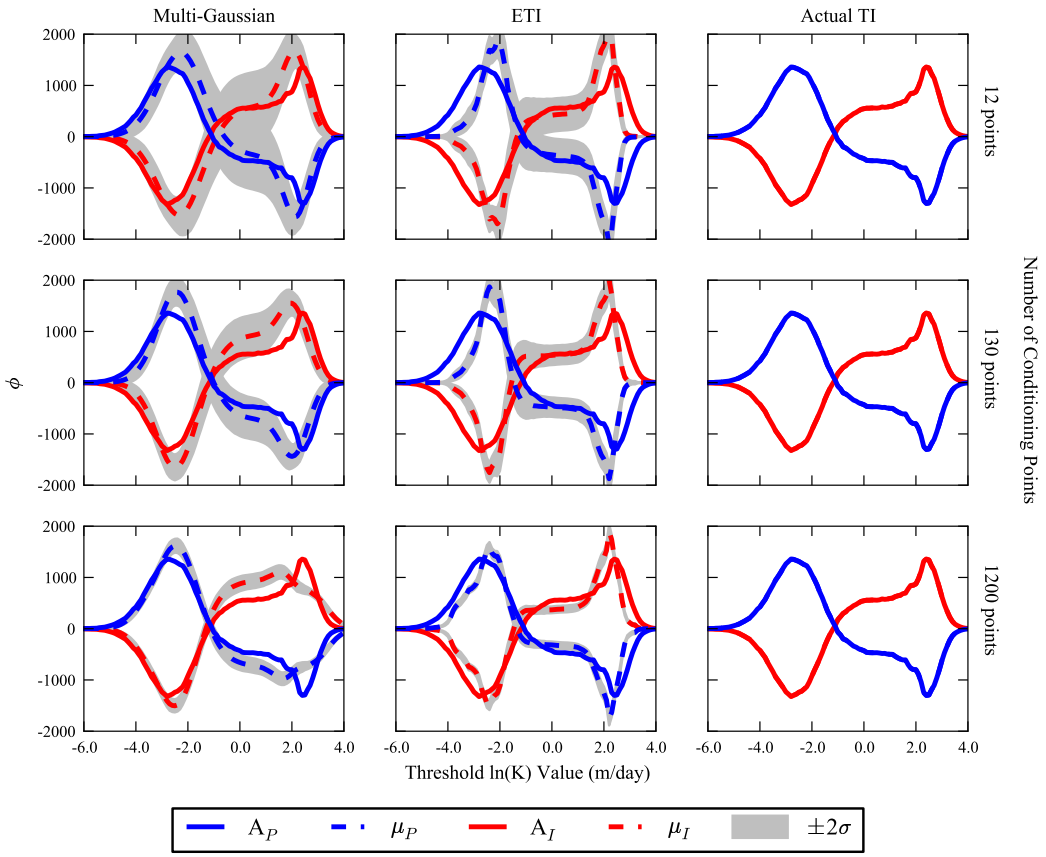


Figure 4.6: Euler number (ϕ) as a function of threshold conductivity for the simulated hydraulic conductivity fields and the baseline scenario. Results of simulated conductivity fields are presented as means and standard deviations.

For the continuous variable elementary training image, Euler number patterns appear averaged for intermediate ln(K) values between -2 and 2, as little change is observed between these values. At values close to -2 and 2, the extreme values are overestimated. For values less than -2 and greater than 2, Euler numbers appear to be underestimated. The patterns are similar for 12 and 130 points. Interestingly, when using 1200 points, threshold values less than -2.0 and greater than 2.0 appear closer to the baseline scenario. For the case where the actual K field was used as a training image, the Euler numbers are exact for all levels of conditioning data.

The probability of connection curve represents the probability that all locations in a single phase are connected. Figure 4.7 demonstrates that in the baseline scenario, the permeable phase has a high level of connection for threshold values less than -2. The probability of connection then reduced to roughly 0.75 for values between -2 and 2 before reducing to almost 0 for values greater than two. The impermeable phase shows a probability of connection generally lower than 0.25 for threshold values less than 2. For threshold values greater than 2, a high level of connection is observed for the impermeable phase.

For multi-Gaussian fields, the pattern of the probability of connection curve is different from the reference field for most of the range (Figure 4.7). For threshold values between -2 and 2, the probability of connection of the permeable phase is under predicted whilst the probability of connection of the impermeable phase is

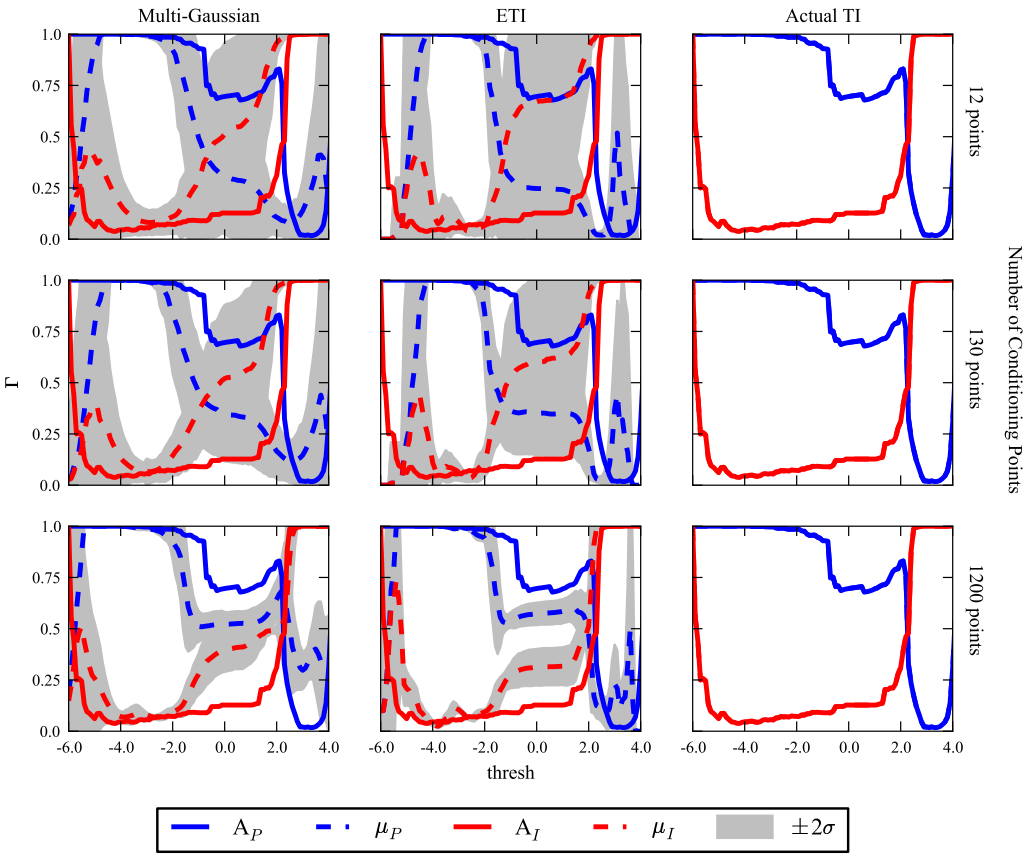


Figure 4.7: Probability of connection (Γ) as a function of threshold conductivity for the simulated hydraulic conductivity fields and the baseline scenario. Results of simulated conductivity fields are presented as means and standard deviations.

over predicted. This is apparent for all levels of conditioning data. When using 1200 conditioning points, the probability of connection of the permeable phase is consistently over-predicted for threshold values between 2 and 4. A similar case is observed for the continuous variable elementary training images whereby probability of connection patterns of the permeable phase are under-predicted for ln(K) threshold values between -2 and 2. Fields obtained using the actual K field as a training image reproduce the probability of connection at all threshold values.

4.3.5 RTDs

Figure 4.8 presents the RTDs for each interpolation scenario. When using multi-Gaussian techniques with 12 and 130 conditioning points, early arrival times are over predicted. Additionally, RTDs under predict the fraction of later arrival times. Interestingly, when 1200 conditioning points are used, predicted earliest arrival times are less than for the baseline scenario.

When using a facies model, early arrival times are not predicted well for all levels of conditioning data. An increase in the level of conditioning data does improve this. Even with 1200 points, earliest arrival times are predicted to be 10-20 years later than the baseline scenario. The facies model also has a lower proportion of later arrival times than the baseline scenario. Later arrival times become more agreeable with an increasing

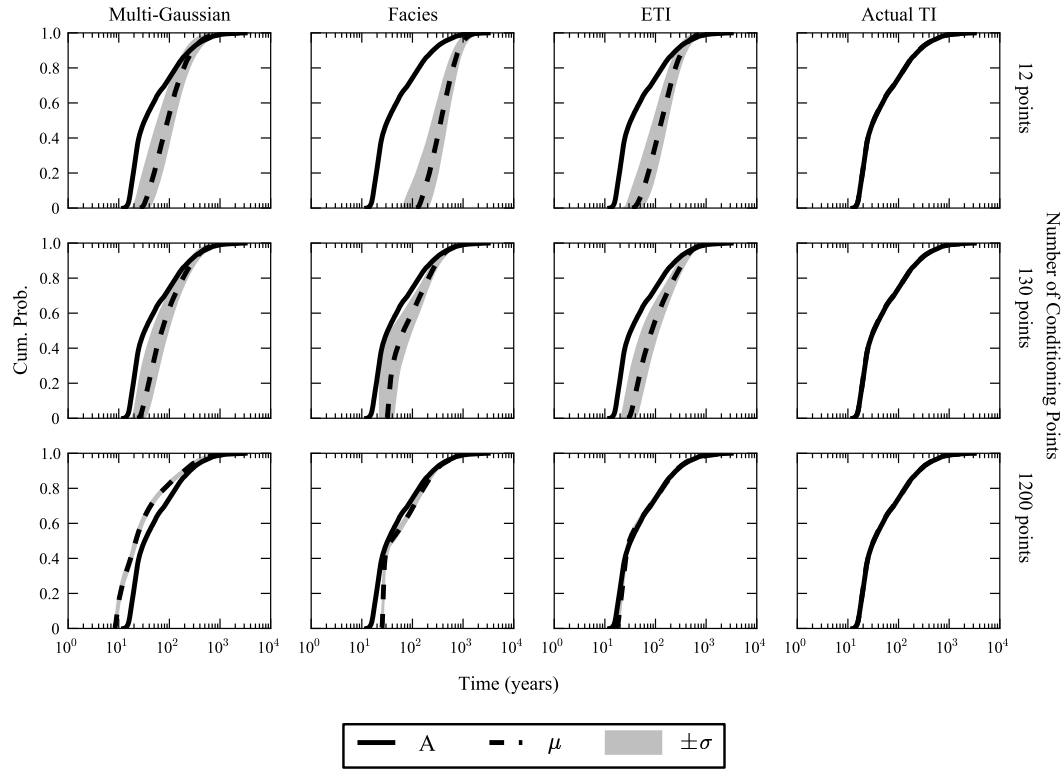


Figure 4.8: Whole of domain residence time distributions for 100 realisations of each interpolation method and level conditioning data (presented as means (μ) and standard deviations (σ)) and baseline scenario (A).

amount of conditioning data.

When using the continuous variable elementary training image with transform invariant distances with only 12 conditioning points, RTDs are poorly predicted. For the scenario where 130 points are conditioning data are used, estimates over-predict early arrival times and under-predict the fraction of later arrival times. When using 1200 points, predictions become very similar to the baseline scenario. Both early and late arrival times are predicted well. When using the actual K field as a training image, RTDs are reproduced accurately for all levels of conditioning data.

To investigate why the multi-Gaussian method predicted earlier breakthrough, Figure 4.9 demonstrates the flow patterns for each of the simulations for one realisation with 1200 conditioning points (when using this level of conditioning data multiple realisations produced very similar results). Figure 4.3A presents the flow patterns in the baseline scenario. Here the streamlines show that the flow pattern is dictated by multiple channels of varying sizes. For example, flow is dictated by both the main channel and smaller channels in bottom portion of the domain. Some flow is also partitioned to the top part of the domain as evidenced by the single streamline. In Figure 4.9A, the streamlines are concentrated in the main channel feature as evidenced by the higher number of streamlines. This is due to the inability of the technique to simulate the connection of the smaller scale features. This results in larger portions of flow occurring in the main feature and hence an earlier arrival time. Conversely, the multipoint methods seem to better represent the different scales of features (Figure 4.9B-D). Whilst the flow patterns are not exact, the distribution of

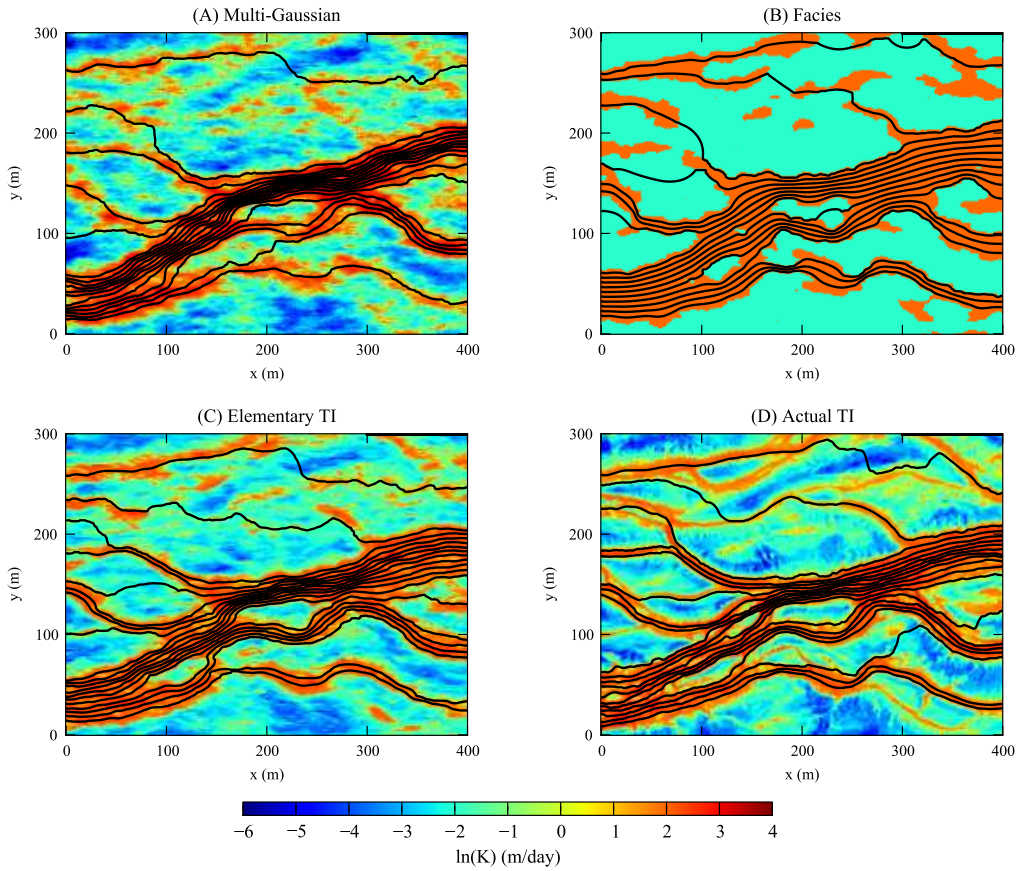


Figure 4.9: Hydraulic conductivity, and stream functions (solid lines) for a single realisation of each of the interpolation methods with 1200 conditioning points.

flow throughout the domain is in better agreement with the baseline scenario (Figure 4.3A).

Despite being able to match the general patterns of flow with 1200 conditioning points, the facies model was unable to match the early arrival times. To investigate this, the baseline K field was modelled as two distinct facies (Figure 4.1C). Figure 4.10 demonstrates that even though the pattern of the facies is exact, the use of a bulk conductivity value for each facies was unable to estimate the earliest arrival times. The magnitude of the lag is 20 – 30 years that is consistent with errors observed in Figure 4.8.

4.4 Discussion

Our results have highlighted some of the important differences and similarities in simulation techniques and hence the importance of choosing an appropriate simulation technique when estimating transit time distributions. We have demonstrated that the ability to reproduce RTDs is related to the partitioning of the flow within the domain. In the case presented here, flow patterns were determined by the location of high K channel features. Multiple scales of such features existed. Interestingly, although a majority of flow occurred through wider channels, the need to reproduce the smaller channels was also important. This was evidenced by the inability of the multi-Gaussian simulation method to reproduce smaller features despite predicting the location and connectivity of a large feature relatively well. This resulted in a higher propor-

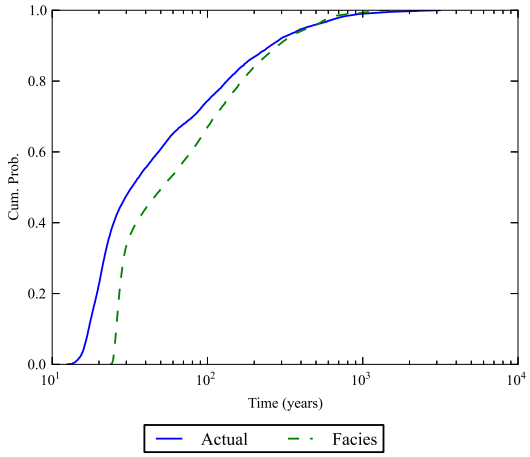


Figure 4.10: Comparison of transit time distributions for the baseline scenario (Figure 4.2B, Actual) and for the modelling of the baseline scenario as two facies (Depicted in Figure 4.1C) using mean K values (facies).

tion of flow occurring through the larger features leading to earlier arrival times. This is consistent with the findings of Knudby and Carrera (2005) that transit times were controlled by the presence of isolated, sometimes disconnected high K features (e.g. smaller disconnected features in the upper portion of the domain were important in the estimation of RTDs).

In addition to the choice of simulation technique, the number of conditioning points used is an extremely important variable. By far the most efficient conditioning was observed when using the actual training image. In this case, identical K fields could be produced with only 12 conditioning points. However, this requires the use of a large search radius and a low error tolerance. To execute this scenario in reality, detailed knowledge of the K field would be required a priori, which is often not possible. Interestingly, comparable results were achieved using continuous variable elementary training images with transform invariant distances and 1200 conditioning points. Here, the training image was designed only to promote connectivity in the horizontal direction and maintain the distribution of K values. However, the large amount of conditioning data required is unlikely to be available in field scenarios. Despite predictions becoming internally consistent, a bias was apparent when using multi-Gaussian techniques with even 1200 points. This is consistent with the findings of Kerrou et al. (2008). The facies model used here required larger amounts of conditioning data and was unable to reproduce early arrival times. Our work suggests that when using a continuous variable elementary training image, a large amount of conditioning data may be able to overcome limited knowledge of the sub-surface.

When using 1200 conditioning points the probability of connection curve was similar for threshold values between -2 and 2 for the multi-Gaussian and elementary training image methods. Both of these relate to an underestimation and overestimation of the proportion of connected pairs in the permeable and impermeable phases respectively when compared to the baseline scenario. However, the elementary training image method represents the equivalent RTD quite well. The largest difference between the connectivity patterns appears to be for threshold values greater than 2. Here the multi-Gaussian method over-predicts the propor-

tion of connected pairs in the permeable phase. This may relate to the inability of the method to recreate multiple scales of high K features. Consequently, the observed earliest residence time is earlier than the baseline scenario. The importance of the connection of extreme values is consistent with previous studies (Gomez-Hernandez and Wen, 1998; Wen and Gomez-Hernandez, 1998; Zinn and Harvey, 2003).

Our study has some limitations. Only two-dimensional fields are considered, which generally underestimate the connectivity of high K values as more connection occurs in three dimensions (Fleckenstein and Fogg, 2008). Furthermore, we used a simple flow field in that head boundaries induced flow from the left to the right of that domain. If an unconfined aquifer receiving distributed recharge were used, then the results may have been more sensitive to local conditions. Furthermore, this study assumed knowledge of the variogram pertaining to the base image (Figure 4.1). If only the known data points were used for generating variogram models, different results may be produced. Additionally, a constant porosity was used. (Cornaton and Perrochet, 2006b) showed that the relationship between K and porosity was important in the prediction of RTDs. Further, the range of K values was larger than would be expected in a braided stream environment. If a narrower range was used, the differences in RTD prediction between the simulation techniques might not have been as great.

Additional methods for the interpolation between points of known K values exist but have not been used in this study. These include sequential indicator simulation (SISIM; Goovaerts, 1997), transition probability geostatistical simulation (TPROGS; Carle and Fogg, 1996) and HOSIM (Mustapha et al., 2011). Klise et al. (2009) used SISIM to interpolate between conditioning points in a braided stream environment. The results were compared to baseline simulations obtained from lidar scans. The authors found that the two-point SISIM method was unable to recreate the connectivity patterns of the field obtained with the lidar scan. Increasing conditioning data was able to improve K field estimates but the method was still unable to accurately predict solute transport. dell'Arciprete et al. (2011) compared SISIM, TPROGS and MPS. The authors found that while MPS was able to better determine the geometries of the most represented hydrofacies and SISIM and TPROGS were better able to account for the distribution of less prominent hydrofacies. Mustapha et al. (2011) demonstrated that the HOSIM method was able to reproduce solute transport properties of both single phase and dual phase flow. The authors used a similar reference permeability pattern to that used here. These methods may also offer alternatives to methods presented in this study.

4.5 Conclusions

In this paper, we have demonstrated how the choice of a simulation method impacts on predicted RTDs for a theoretical aquifer with a multi-modal $\ln(K)$ distribution. We found that the RTDs in the aquifer were controlled by multiple scales of channel features. In general, with large amounts of conditioning data, multiple point methods were better able to recreate both large and small features. With 1200 conditioning points the multi-Gaussian method was able to reproduce larger features but not the narrower channel features, resulting in an overestimation of connectivity as larger amounts of flow occurred through the large-scale

channels. This was reflected by the shape of connectivity metrics for ln(K) values greater than 2.0. This flow distribution created a bias to earlier residence times.

When using multiple point methods, efficient conditioning was observed when using the baseline K field as a training image. Here with only 12 points the original image could be re-created. When using continuous variable elementary training images, the RTD was recreated using 1200 conditioning points despite differences in the K field. In reality, this level of conditioning data is unavailable. However, this suggests that the use of large amounts of conditioning data may be able to overcome the choice of the model. The facies model was unable to reproduce early arrival time RTD behaviour, due to the use of geometrically averaged hydraulic conductivity values for the separate facies.

This study has shown that choice of simulation method and number of conditioning points are critical, and highlighted in a systematic and quantitative manner some of the key controls on the RTD behaviour. Further work is required to validate these methods in three-dimensional real world cases.

Acknowledgments

Funding for this work was provided by the Australian Research Council and the National Water Commission. We thank Philipe Renard, Gregoire Mariethoz, Jean-Raynauld de Dreuzy and two annonymous reviewers for their reviews of our work.

5 Manuscript 4: Non-parametric estimation of residence time distributions: What can environmental tracers tell us about groundwater residence time?

Submitted to Water Resources Research

Authors: James L. McCallum, Nicholas B. Engdahl, Timothy R. Ginn and Peter G. Cook

Abstract

Residence time distributions (RTDs) have been used extensively for quantifying flow and transport in sub-surface hydrology. In geochemical approaches environmental tracer concentrations are used in conjunction with simple lumped parameter models (LPMs). Conversely numerical simulation techniques require large amounts of parameterisation and estimated RTDs are certainly limited by associated uncertainties. In this study we develop a non-parametric approach to estimate RTDs using environmental tracer concentrations. The model is based only on the assumption that flow is steady enough that the observed concentrations are well approximated by linear superposition of the input concentrations with the RTD; that is, the convolution integral holds. Even with large amounts of environmental tracer concentration data, the entire shape of an RTD remains highly non-unique. However, accurate estimates of mean ages and in some cases prediction of young portions of the RTD may be possible. The most useful type of data was found to be the use of a time series of tritium. This was due to the sharp variations in atmospheric concentrations and a short half-life. Conversely, the use of CFC compounds with smoothly varying atmospheric concentrations was more prone to non-uniqueness. This work highlights the benefits and limitations of using environmental tracer data to estimate whole RTDs with either LPMs or through numerical simulation. However, the ability of the non-parametric approach developed here to correct for mixing biases in mean ages appears promising.

5.1 Introduction

Determining the distribution of groundwater ages at wells and discharge features continues to be of interest. The reason for this is that these distributions directly relate to timescales of flow and transport in groundwater systems. This is because groundwater age distributions arise from the same processes that dictate solute transport processes Ginn (1999); Engdahl et al. (2012). Techniques for estimating these distributions can generally fall into two categories: geochemical techniques and numerical modelling techniques.

In geochemical techniques, environmental tracers are used in conjunction with lumped parameter models (LPMs) (Corcho Alvardo et al., 2007; Land and Huff, 2009; Knowles Jr. et al., 2010; Genereux et al., 2009; Solomon et al., 2010; Stolp et al., 2010). LPMs define the mixing in aquifer systems assuming simple geometries, recharge rates or advective-dispersive relationships (Maloszewski and Zuber, 1982). Multiple conceptual models may also be combined to define mixing in systems, which can greatly improve the

resulting LPMs, but the assumptions of models are rarely interrogated. When fitting these models with observations of multiple tracer concentrations, the parameters that define these models can be constrained by data from multiple tracers (Massoudieh and Ginn, 2011; Massoudieh et al., 2012) .

The numerical simulation of groundwater age distributions has also received much attention (Varni and Carrera, 1998; Ginn, 1999; Cornaton and Perrochet, 2006a). These methods have allowed for an understanding of how residence time distributions behave in advective dispersive groundwater systems. Applications of such methods have demonstrated RTDs in simple flow systems (Varni and Carrera, 1998; Etcheverry and Perrochet, 2000; Cornaton and Perrochet, 2006a; Woolfenden and Ginn, 2009), topographically driven flow systems (Cardenas, 2007; Cardenas and Jiang, 2010) and heterogeneous flow systems (Weissmann et al., 2002; Cornaton and Perrochet, 2006b; Engdahl et al., 2012). The use of environmental tracers in the context of such models generally results in a hypothesis testing scenario whereby models can be validated based on transit time distributions that explain one or multiple environmental tracer concentrations (Varni and Carrera, 1998; Weissmann et al., 2002; Leray et al., 2012). However, these concentrations can often be fit by multiple RTDs resulting in non-uniqueness and limited improvement through the use of environmental tracer data.

The problem is that the quantification of RTDs using LPMs relies on the use of simple models that may not represent the groundwater systems they are applied too, limiting the results obtained. Additionally, the extensive parameterisation of numerical models can be time consuming and the RTDs predicted by these will be limited by the associated parameter and structural uncertainties. In this paper we propose and investigate a method for determining RTDs using environmental tracer data. The method assumes only that the convolution relationship holds and hence the restrictions of LPMs do not apply. Additionally, far less information about the groundwater system is required than for the use of numerical groundwater models.

5.2 Theory

5.2.1 Deconvolution

The relationship between environmental tracer concentrations in groundwater, historical atmospheric concentrations, and steady-state aquifer residence time distributions can be given by the convolution integral (Maloszewski and Zuber, 1982, equation 6):

$$c_{out} = \int_0^t g(\tau) c_{in}(t-\tau) e^{-\lambda\tau} d\tau \tag{5.1}$$

where c_{out} is the observed concentration in the aquifer (corrected for solubility for gas tracers and heterogeneous reactions for reactive tracers), g is the age frequency distribution (i.e. RTD), c_{in} is the atmospheric concentration, λ is the radioactive decay constant and t is the time of sampling. If we have nc concentrations and ng discrete bins of the age frequency distribution, we can re-write this equation in discretized

form (Cirpka et al., 2007):

$$c = \mathbf{X}\mathbf{g} \tag{5.2}$$

where the rows of $\mathbf{X}$ are defined as:

$$\mathbf{X_{row}} = c_{in}(t-\tau)e^{-\lambda\tau}d\tau \tag{5.3}$$

where τ represents discrete values of age and $\mathbf{X}$ is a $nc \times ng$ matrix. Multiple rows for the matrix $\mathbf{X}$ can arise from either for a time series of a single tracer, multiple tracers at a single time or a mixture of both. (It is important to note that if tracers are obtained at different times then the flow and transport is assumed to be steady state.) It is important that the age frequency distribution integrates to one, and that no discrete value of $\mathbf{g}$ is less than zero. To enforce the first constraint, we use the equation:

$$\mathbf{u}^{\mathrm{T}}\mathbf{g}d\tau = 1 \tag{5.4}$$

where $\mathbf{u}$ is a $ng \times 1$ vector of unit entries. To implement the second constraint we use a method similar to Cirpka et al. (2007) where:

$$\mathbf{H}\mathbf{g} = 0 \tag{5.5}$$

where $\mathbf{H}$ is a $n\nu \times ng$ (where $n\nu$ is the number of active constraints) matrix defined as:

$$\mathbf{H_{ij}} = \begin{cases} 1, & \text{if } g_j \text{ is affected by the } ith \text{ constraint} \\ 0, & \text{otherwise} \end{cases} \tag{5.6}$$

By applying the maximum likelihood approach to Equation 5.2 and constraining the solution using Equations 5.4 and 5.5 through Lagrange multipliers we arrive at:

$$\begin{bmatrix} \mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{X} & \mathbf{u}d\tau & \mathbf{H}^{\mathrm{T}} \\ \mathbf{u}^{\mathrm{T}}d\tau & 0 & 0 \\ \mathbf{H} & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{g} \\ \nu_1 \\ \nu \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{c_o} \\ 1 \\ 0 \end{bmatrix} \tag{5.7}$$

where the matrix $\mathbf{Q}$ is a weighting matrix whereby the diagonal elements are $1/\sigma^2$ of the measured concentration (i.e. $C(\varepsilon) = \mathbf{Q}^{-1}$), ν_1 is the Lagrange multiplier associated with the constraint required for integration to one and ν is the vector of Lagrange multipliers implemented for non negative values. The addition and removal of constraints relating to non-negativity requires an iterative procedure. Constraints

are added or removed as follows:

$$\mathbf{g_j} < 0 \rightarrow \text{add constraint for element j}$$
$$v_i \leq 0 \rightarrow \text{keep constraint i}$$
$$v_i > 0 \rightarrow \text{remove constraint i}$$

The method differs from that of Cirpka et al. (2007) in that no variogram regularization is implemented, it is required that the age frequency distribution integrates to 1 and that a constraint of a zero frequency at the zero travel time is not enforced. The concept of zero age water is difficult as water is only of zero age instantaneously at the interface between recharge and groundwater. As such, any well screened below the water table should have no zero age water. Additionally, when applying this model at a discharge feature the assumption would require that water travel no distance since recharge. Although these assumptions are of considerable limitation, we have allowed for a zero age distribution so as to assess the sensitivity of entire age distributions to environmental tracer concentrations.

5.2.2 Truncated singular value decomposition (tSVD) and parameter informativeness

In a typical environmental system, the number of measured tracer concentrations is likely to be much smaller than the desired number of discrete values of the age frequency distributions. As such, the solution will be ill-posed and non-unique. For this reason, truncated Singular Value Decomposition (tSVD) was used as a solution technique. SVD is a useful method for analysing and solving singular or almost singular sets of equations (Press et al., 1986). This method also acts as a form of regularization. In this case the matrix on the left hand side of Equation 5.7 is represented as:

$$\mathbf{A} = \begin{bmatrix} \mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{X} & \mathbf{u}d\tau & \mathbf{H}^{\mathrm{T}} \\ \mathbf{u}^{\mathrm{T}}d\tau & 0 & 0 \\ \mathbf{H} & 0 & 0 \end{bmatrix} \tag{5.8}$$

The singular value decomposition of **A** results in it's approximation with three matrices:

$$\mathbf{A}{=}\mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{T}} \tag{5.9}$$

where **U** and **V** are the left and right orthogonal matrices and **S** is a diagonal matrix containing the singular values of **A**. In the truncation method the matrices **U**, **V** and **S** are approximated as:

$$\mathbf{A} = \begin{bmatrix} \mathbf{U_1} & \mathbf{U_2} \end{bmatrix} \begin{bmatrix} \mathbf{S_1} & 0 \\ 0 & \mathbf{S_2} \end{bmatrix} \begin{bmatrix} \mathbf{V_1^T} \\ \mathbf{V_2^T} \end{bmatrix} \tag{5.10}$$

Here, subscripts 1 and 2 denote elements in the two subspaces - solution and null respectively. The criteria

whereby this point is determined will be discussed below. The (non-unique) solution of the problem is achieved by calculating a pseudo inverse:

$$\mathbf{A^{-1}} = \mathbf{U^T} \begin{bmatrix} \mathbf{S_1} & 0 \\ 0 & 0 \end{bmatrix}^{-1} \mathbf{V} \tag{5.11}$$

whereby only the singular values in the solution space are used to estimate the best solution. The choice of the separation of $\mathbf{S}$ between $\mathbf{S_1}$ and $\mathbf{S_2}$ is important because it quantifies the non uniqueness of the solution in terms of the dimensionality of the nullsplace (the dimension of $\mathbf{S_2}$). One suggested location for the number of singular values used was the number that minimized the error of the prediction (Moore and Doherty, 2005; Doherty and Hunt, 2009). To determine the point we perform the singular value decomposition such that:

$$\mathbf{X^T Q X = U S V^T} \tag{5.12}$$

It is important to note that $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{S}$ and the left orthogonal matrix, the right orthogonal matrix and the singular values of $\mathbf{X^T Q X}$ respectively (not $\mathbf{A}$). The error of these estimates is defined as:

$$[\sigma^2]_i = \mathbf{i}^{\mathrm{T}} \mathbf{V_2} \mathbf{V_2}^{\mathrm{T}} C(p) \mathbf{V_2} \mathbf{V_2}^{\mathrm{T}} \mathbf{i} + \mathbf{i}^{\mathrm{T}} \mathbf{G} C(\varepsilon) \mathbf{G}^{\mathrm{T}} \mathbf{i} \tag{5.13}$$

where $C(p)$ is the covariance matrix of the parameters, $C(\varepsilon)$ is the covariance matrix of the observations, $\mathbf{i}$ is a vector where the entries are one for all locations relating to the \$ith\$ parameter and zero elsewhere and:

$$\mathbf{G} = (\mathbf{U^T S_1^{-1} V}) \mathbf{X^T Q} \tag{5.14}$$

The diagonal entries of $C(p)$ and $C(\varepsilon)$ contain the variance of the parameters and observations respectively (off-diagonal entries are zero). $\mathbf{G}$ represents the use of the pseudo inverse to map the observational error back to the parameters. Essentially, this equation relates the error of the estimates to the pre-calibration error and parameter error based on how much of the parameter is in the null space and how much of the parameter is in the solution space. The proportion of parameters undefined (i.e. portion in the null-space) are given the pre-calibration error of the parameter and the proportion that exists in the solution space is constrained by propagating the measurement error. This proportion, defined by Doherty and Hunt (2009) as identifiability is given as:

$$f_i = \mathbf{i}^{\mathrm{T}} (\mathbf{V_1} \mathbf{V_1}^{\mathrm{T}}) \mathbf{i} \tag{5.15}$$

In the case presented here identifiability is a measure of what proportion of each discrete value of the age frequency can be resolved with the measured environmental tracer concentrations. This reflects the sensitivity of measured environmental tracer concentrations to the discrete value of the age frequency distribution.

The choice of values for $C(p)$ and $C(\varepsilon)$ is also important. In general, if nothing is known about the distri-

bution, then a discrete value of g has the potential to have a value between 0 and $1/d\tau$. If we assume that three standard deviations cover this range then we can approximate a pre-calibration variance of $(1/d\tau/6)^2$. In determining the separation of $\mathbf{S}$ between $\mathbf{S_1}$ and $\mathbf{S_2}$, the error in Equation 5.13 is calculated for the use of one to ng singular values. The number of singular values that give the minimum error are then used to establish the truncation point for the solution of Equation 5.7. In application, the Lagrange multipliers are multiplied by the highest singular value obtained in Equation 5.12. This ensures that the constraints are maintained in the solution procedure.

5.2.3 Other metrics

In addition to the estimates of the entire distribution, estimates of the mean age (A_μ), new water fraction (f_{new}), a metric for early arrival time (A_{10}) and the second central moment of the age distribution (A_{σ^2}) were estimated. The mean age is obtained from the dot product of g and the vector of discrete time values multiplied by the time step. The fraction of new water was the proportion of the distribution younger than 50 years. The early arrival time metric was taken as the point where the cumulative age distribution reached 10%. The second central moment or variance is a measure of the width of the distribution. It has been demonstrated in other applications that environmental tracer concentrations are sensitive to the first two central moments of the age distribution (Varni and Carrera, 1998; Waugh et al., 2003). Variance was calculated as:

$$A_{\sigma^2} = \int_0^\infty \left(\tau - A_\mu\right)^2 g(\tau)\, d\tau \tag{5.16}$$

5.2.4 Choice of data to test

The robustness of the estimation of the non-parametric model was tested on three separate data sets, in order to evaluate the performance of the inversion in a real context. The first was a time series of tritium data. This type of data was based on data from the Fischa-Dagnitz system (Rank and Papesch, 2003; Stolp et al., 2010). The error applied to tritium estimates was 10%. This was selected due to analytical errors and also the spatial variability of atmospheric concentrations resulting in further errors. The second type of data was based on the sampling of CFCs, and SF_6 at multiple times. This was based on the data available from the Delmarva Peninsula (Dunkle et al., 1993; Ekwurzel et al., 1994; Busenberg and Plummer, 2000, 2008). The errors of CFC and SF_6 data were assumed to be 5%. This is for analytical errors and is lower than for tritium as the atmospheric mixing ratios of these elements are consistent and fairly well known. The third type of data was based on data collected from Costa Rica by Solomon et al. (2010). Here, we assumed that Bomb Peak ^{14}C, tritium, CFCs and SF_6 had been collected at a single time in 2010. Bomb Peak ^{14}C was assumed to have a 10% error.

The errors stated above were used to form the $C(\varepsilon)$ matrix in Equation 5.13. It was assumed that percentages, when multiplied by observation values, represented errors as standard deviations. As such, the

diagonal elements represent squared standard deviation values and off diagonal elements were set to a value of zero.

5.3 Synthetic RTDs and tracer concentrations

The performance of the proposed method will first be considered under known conditions using synthetically generated RTDs and environmental tracer concentrations. Random noise is added to the simulated concentrations to represent measurement error, and the resulting concentrations are used in the deconvolution to estimate the RTDs for two test cases: a closed form RTD (Section 5.3.1), and a synthetic aquifer model (Section 5.3.2).

5.3.1 Analytical RTDs

5.3.1.1 Generation of RTDs The theoretical model for the RTD in this section uses an ensemble of independent streamtubes. The governing equation for the age frequency distribution is:

$$\frac{\partial \theta g}{\partial t} + \nabla \cdot (v\theta g) + \nabla \cdot (D\theta \nabla g) + \frac{\partial \theta g}{\partial a} = 0 \tag{5.17}$$

where t is time, a is age, v is groundwater velocity, θ is porosity, D is the dispersion tensor and g is value of the age frequency distribution for a given age, at a given point in space and time (Ginn, 1999).

If we employ two terms, $x' = x/v$ and $Pe = vx/D$, the solution to the constant coefficient (i.e. completely homogeneous), steady state (in time) form of Equation 5.17 is(Ginn et al., 2009):

$$g(a) = \sqrt{\frac{x'Pe}{4\pi a^3}} \exp\left(\frac{-Pe\,(t-x')^2}{4x'a}\right) \tag{5.18}$$

The contribution from multiple streamtubes to mixing e.g. in a well or discharge feature would then give rise to a superposition of multiple instances of Equation 5.17, each weighted by the fraction of solute flux along each respective streamtube. The combined age frequency distribution of multiple steamtubes with distributions $g_1, g_2, \ldots g_n$ is:

$$g = f_1 g_1 + f_2 g_2 + \ldots f_n g_n \tag{5.19}$$

where f_i is the fraction of the total flux that occurs along the ith streamtube for the total sample (the sum of all values of f equals 1). This type of distribution may occur typically in layered or heterogeneous aquifers where multiple streamtubes contribute or in long well screens in aquifers with multiple recharge sources. A random approach was taken to simulating the RTDs. The number of streamtubes was arbitrarily chosen to be between one and three. To do this initially a random number (z) between zero and one was generated.

The number of streamtubes (ns) was then determined as:

$$ns = \begin{cases} 1, & z \leq 0.5 \\ 2, & 0.5 < z \leq 0.8 \\ 3, & z > 0.8 \end{cases} \tag{5.20}$$

Values for x' were randomly generated from a uniform distribution between zero and 100. Values for Pe were selected from a distribution based on data from (Gelhar et al., 1992). Only data for regional and ambient tests were used. When multiple streamtubes were implemented (values of $z > 0.5$), a constant value of Pe was used and only values of x' were altered. Values of f_i were generated randomly and normalised to sum to one.

The range of values of x' being limited to 100 years was implemented as only transient tracers were tested. These included compounds only produced since 1950 and compounds used based on sharp increases due to nuclear testing. As such, these tracers are generally only used for residence times less than 60 years.

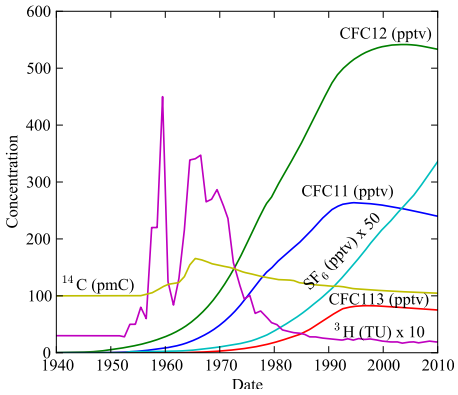


Figure 5.1: Atmospheric concentrations used for inversion (c_{in})

Environmental tracer concentrations associated with the the randomly generated RTDs were obtained using the convolution integral presented in Equation 5.1 and atmospheric concentrations (c_{in}) presented in Figure 5.1. Noise was added to the simulated concentrations to simulate some degree of measurement error. The final concentrations were selected by drawing a value from a normal distribution with the true concentration as the mean and a standard deviation that is some percentage of the mean (as outlined in Section 5.2.4).

5.3.1.2 Results The most reliable estimates of RTDs from the deconvolution of the environmental tracer data were the estimates made using tritium data (Figure 5.2). These estimates accurately represented the shape of the distributions for ages younger than 50 years (Figure 5.2A, 5.2D and 5.2G). Estimates made using CFCs and SF_6 at multiple times produced a more averaged behavior. This is especially evident in Figure 5.2G. Estimates made using concentrations of ^{14}C, ^{3}H, CFCs and SF_6 at a single time produced a similar averaging fit and over-predicted the frequency of younger residence times.

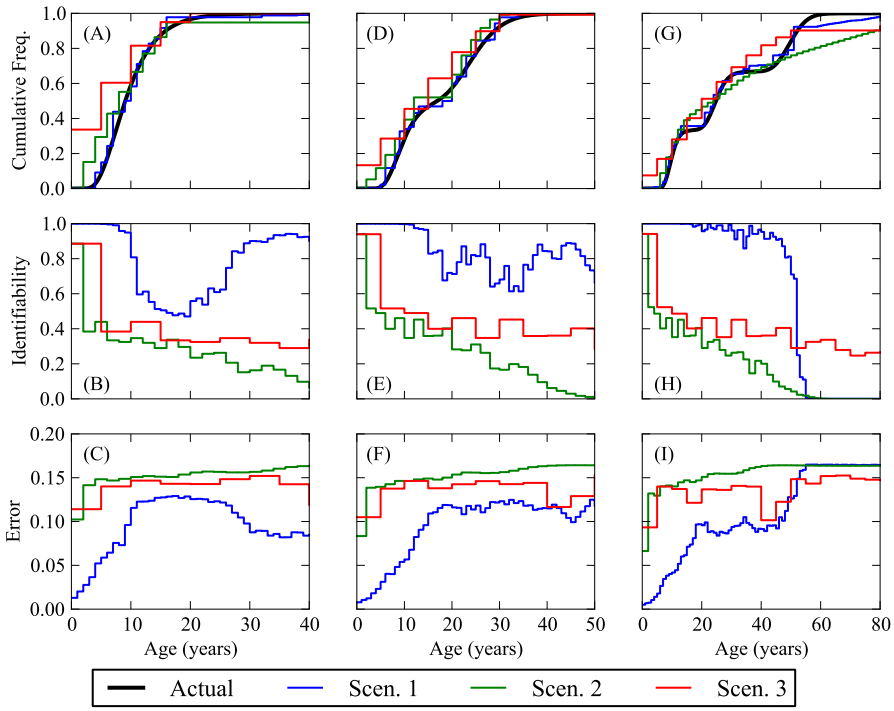


Figure 5.2: Example of the procedure for fitting RTD data for a time series of tritium (Scenario 1), CFCs and SF_6 for 5 year intervals between 1990 and 2010 (Scenario 2), and a single sampling of ^3H, ^{14}C, CFCs and SF_6 (Scenario 3). Figures A, D and G represent actual and estimated RTDs, Figures B, E and H represent the identifiability (calculated using Equation 5.15) of individual bins and Figures C, F and I represent the error of estimates (subject to assumptions outlined in methods).

The use of tritium time series is the most informative (i.e. gave the highest values of identifiability) of age frequency distributions particularly at young ages (Figures 5.2B, 5.2E and 5.2H). The pattern is different for each scenario. This range will be partly determined by the atmospheric concentrations but also the sampling concentrations as the matrix **Q** in Equation 5.7 is determined as a proportion of the measurement. The other two data types implemented show a high value for identifiability at early times before dropping significantly. In most cases less than half of the discrete interval can be informed by the environmental tracer concentrations. These reflect that the two data types are not sensitive to discrete values of the age frequency distribution. In the case of scenario two, this may be due to correlation between tracers. In the case of scenario three the sensitivity reflects the lack of data.

Errors increase with decreasing identifiability. Even where identifiability is high errors are not reduced by much (with the exception of early small values of age when using tritium time series). This suggests that the estimation is still quite sensitive to the errors in sample concentrations. Additionally, the relatively large errors suggest that estimated RTDs will remain highly non-unique, meaning that the constraint on mixtures provided by environmental tracers is somewhat limited.

Figure 5.3 presents the results of the estimation of metrics discussed in Section 5.2.3 using the deconvolution method and differing types of environmental tracer data. When using a long time series of tritium, mean ages less than 50 years are predicted well by the method (Figure 5.3A). For values greater than this, the prediction is less accurate and biases to younger age. This reflects the inability of the concentrations to

inform older components of the distribution. Estimates of the fraction of water younger than 50 years (Figure 5.3B) and the age that 10% of water is younger than (Figure 5.3C) are also well predicted with this data type. The 2nd moment of the age distribution is predicted poorly by the method (Figure 5.3D). This likely reflects the inability of the method to inform greater ages which contribute to this estimate

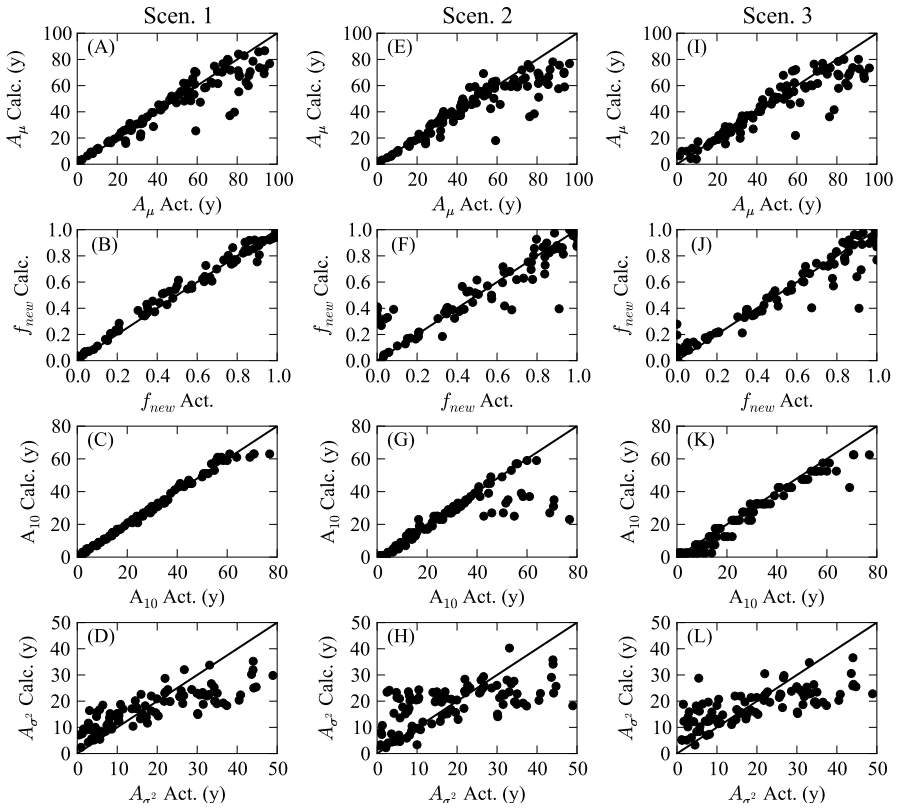


Figure 5.3: Estimates of mean age (A_μ), new water fraction (f_{new}), 10% arrival time (A_{10}) and second moment of the age distribution (A_{σ^2}) for 100 synthetic RTDs using a time series of tritium (Scenario 1), CFCs and SF_6 for 5 year intervals between 1990 and 2010 (Scenario 2), and a single sampling of ^3H, ^{14}C, CFCs and SF_6 (Scenario 3).

When using CFCs and SF_6 at multiple times, mean ages less than 50 years are predicted well by the method (Figure 5.3E). The fraction of water less than 50 years is not predicted as well as for the use of tritium particularly for waters that have almost no water less than 50 years or almost all water greater than 50 years of age (Figure 5.3F). The age for which 10% of all water is younger is predicted well with this type of data for values less than 40 years (Figure 5.3G). As with the use of tritium data, the second moment of the RTD is predicted poorly (Figure 5.3H).

When using ^{14}C, ^3H, CFCs and SF_6 at a single time, predictions of mean ages show the most scatter however, they are generally of the correct order for values less than 50 years (Figure 5.3I). Estimates of the fraction of water less than 50 years show similar biases as Scenario 2, whereby predictions are more incorrect near extreme values of zero and one (Figure 5.3J). The predictions of ages for which 10% of waters are younger appear strongly biased towards smaller values (Figure 5.3K). This is consistent with what was observed in Figure 5.2 – that this type of data over predicted the youngest discrete values of the age frequency distribution. As with the other two data types, accurate estimates of the second moment of

the RTD were not obtained.

With regard to the ability of each of the data types to represent RTDs, variations in atmospheric concentrations as well as short half lives result in sequential tritium measurements being uncorrelated, providing greater information about RTDs. Conversely, CFC compounds have smoothly varying atmospheric concentrations and individual CFC11, CFC12 and CFC113 have similar variations in atmospheric concentrations. This limits the information that can be obtained about the RTD when using multiple CFCs. Hence tritium time series data is better able to provide information about portions of the RTD for ages less than 50 years.

5.3.2 Synthetic heterogeneous aquifer

5.3.2.1 Generation of RTDs The method was also assessed in a synthetic heterogeneous aquifer. Here, RTDs and environmental tracer concentrations were simulated using numerical models. Further details are provided in McCallum et al. (In Press), however the method is briefly described here. The Synthetic aquifer was simulated using the direct sampling method (Mariethoz et al., 2010) with the use of an elementary training image and transform invariant distances (Mariethoz and Kelly, 2011). Although flow and transport processes are strictly three dimensional, a two dimensional example was used. The K-field was simulated to represent a fluvial system. For numerical modeling the problem domain was set up to extend 500 m (x-directional) by 200 m (y-directional) by 1 m (z-directional) (Figure 5.4A). Numerical flow modeling was undertaken using Hydrogeosphere(Therrien et al., 2006). For flow modeling a flux of 0.365m/y was applied to the boundary at x = 0 m and a constant head of 1.5 m was imposed to simulate two-dimensional saturated flow from the left to the right of the domain. A uniform porosity of 0.3 was implemented and RTDs and environmental tracer concentrations were simulated using the Ito-Fokker-Plank interpolation method (LaBolle et al., 1996).

The simulated concentrations were given measurement errors based on the percentages outlined above then implemented with the deconvolution method to estimate RTDs at each of the 34 wells presented in Figure 5.4A. Additionally the metrics for mean age, fraction of new water and early arrival time were estimated.

5.3.2.2 Results Figure 5.4B presents simulated RTDs estimated from concentrations obtained at the white point in Figure 5.4A. As with the analytical RTDs, the use of a time series of tritium data best reproduces the actual RTD. The use of both CFCs and SF_6 at multiple times and ^{14}C, ^{3}H, CFCs and SF_6 at a single time results in averaged behavior. Additionally the metrics predicted showed consistent patterns with results presented in Figure 5.3. Interestingly however, these results suggest that even in highly heterogeneous aquifers, it may be possible to estimate some aspects of the RTD such as mean ages, or early portions of the RTD.

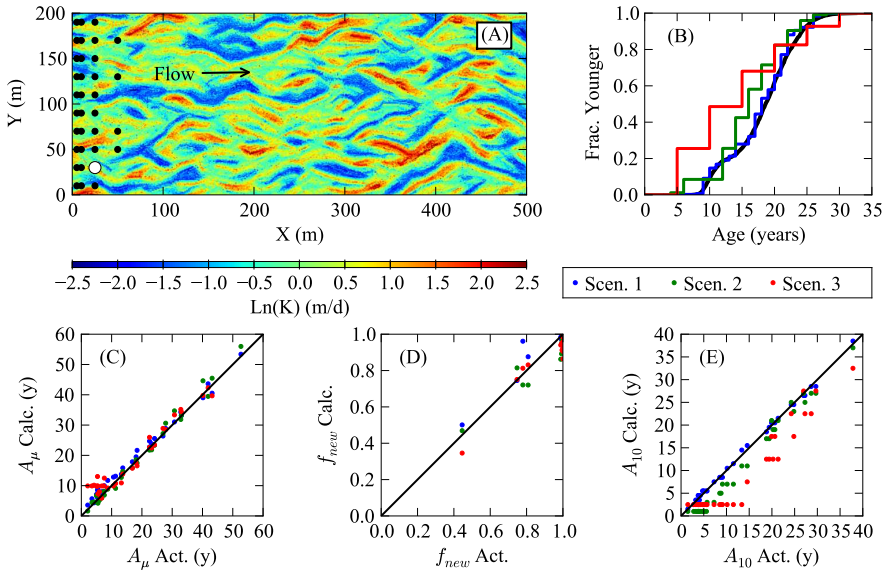


Figure 5.4: (A) Heterogeneous field and locations of 100 wells. (B) Estimation of RTD at Well022 (identified in Figure A). Figures C, D and E represent estimates of mean age (A_μ), new water fraction (f_{new}) and earliest arrival time (t_0) for 34 wells presented in Figure A using a time series of tritium (Scenario 1), CFCs and SF_6 for 5 year intervals between 1990 and 2010 (Scenario 2), and a single sampling of ^3H, ^{14}C, CFCs and SF_6 (Scenario 3).

5.4 Application to literature data

In this section we apply the non-parametric estimation method to field data published from previous studies. In these cases the age distributions are not known therefore making comparison impossible. However this allows us to test the methods applicability to real world data sets.

5.4.1 Fischa-Dagintz

5.4.1.1 Data This section performs an analysis on data presented by Stolp et al. (2010). The data is collected from a spring in the Fischa-Dagintz system in the Vienna basin in Austria. Tritium data has been obtained regularly from the site since the 1960s. The authors interpreted the data using a hybrid exponential model (which assumes mixing via vertical stratification) and the dispersion model (which represents Fickian mixing). The authors also used a single measurement of ^3He however we have not used this point here as it is unlikely that these two tracers move through the subsurface at the same rate (LaBolle et al., 2006).

The Fischa-Dagintz system is composed of three springs and the stream they create. Recharge to the springs is thought to occur from a river 20 km away. Additional recharge is thought to have been contributed by irrigation and precipitation (Stolp et al., 2010).

5.4.1.2 Results The inversion procedure was able to recreate observed concentrations very well (Figure 5.5A). The estimated RTD had a small fraction of water in the youngest bin (Figure 5.5B). Some detail is observed for ages less than 40 years however the data is unable to inform the RTD after this point.

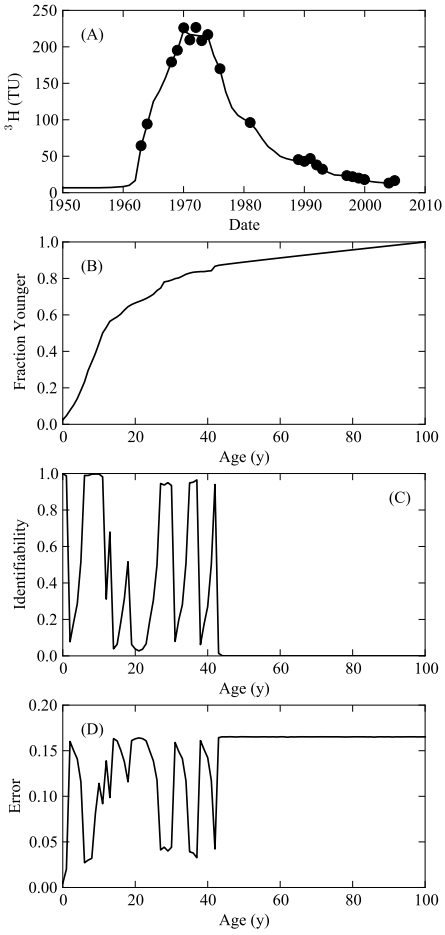


Figure 5.5: Measured (dots) and estimated (lines) tritium concentrations (A) and Estimated RTD (B), Identifiably (C) and error of estimate (D) using tritium time series data from the Fischa-Dagnitz system in Austria.

The identifiability also shows interesting patterns (Figure 5.5C). It appears that bins are either entirely identifiable, or not identifiable. This is most likely due to the spacing of the sampling times in addition to the atmospheric concentrations. The error follows a similar pattern as it is determined by the amount of identifiability. In bins where identifiability is high, errors are due to propagation of measurement errors, rather than pre-calibration errors. The estimated mean age of the distribution was approximately 21 years compared to an estimate from just fitting the tritium data of 14 years by Stolp et al. (2010). It was estimated that the fraction of water younger then 50 years was 0.89.

The fraction of water in the youngest bin is consistent with the model of recharge through precipitation and irrigation near the stream. The general roughness of the curve may represent various paths taken by water to the discharge points.

5.4.2 Delmarva Peninsula

5.4.1.1 Data This section preforms analysis on data from a single well located in Locust Grove on the Delmarva peninsula in Maryland, USA. The data used was presented in Busenberg and Plummer (2000, 2008); Dunkle et al. (1993); Ekwurzel et al. (1994). At this site, CFC-11 and CFC-12 were originally

obtained in the early 1990's. Subsequent data collection has occurred resulting in the addition of CFC-113 and SF_6.

Bore KEBE052 forms part of a greater monitoring network on the Delmarva Peninsula. The aquifer is formed of gravels sands and clays deposited in transgressive and regressive sequences. The surficial aquifer is characterised by shallow water tables and low gradients (Ekwurzel et al., 1994).

5.4.1.2 Results Although some trends are re-created, the deconvolution procedure does not fit all data exactly (Figure 5.6A-D). This may be due to tracers moving at different rates, the system not being in steady state as assumed by the method or simply that the concentrations could not be well represented by the number of singular values that minimized the error. Estimated concentrations show little sensitivity to the RTD as demonstrated by the identifiability in Figure 5.6F. This suggests that this level of data may not be informative of the RTDs. As such, errors remain large and estimates of the RTD are likely to be highly non-unique.

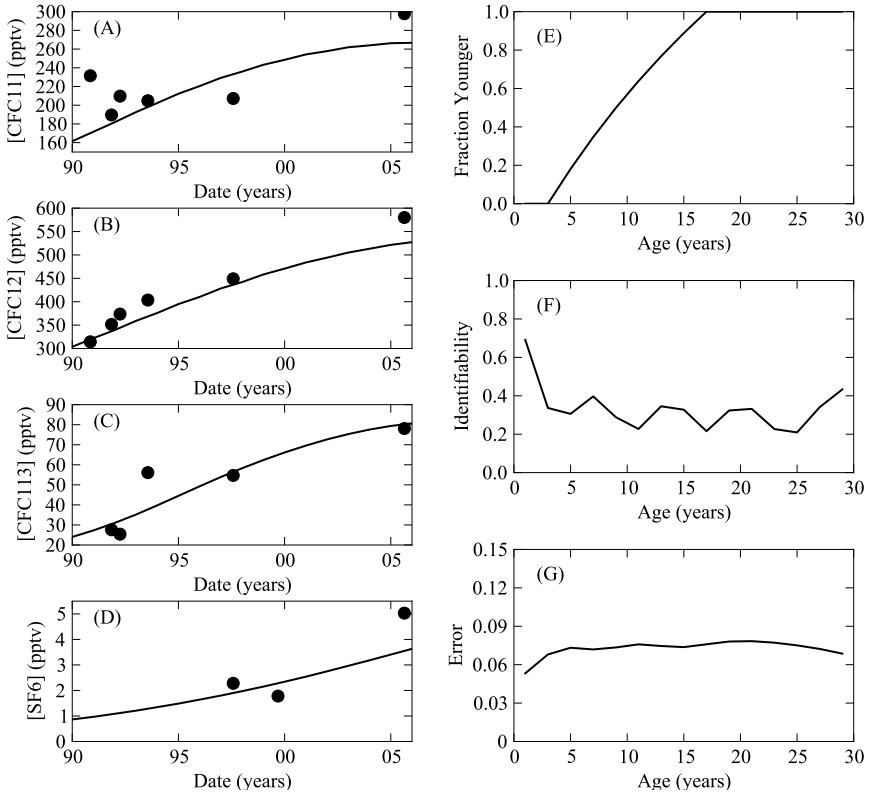


Figure 5.6: Measured (dots) and estimated (lines) concentrations of CFC-11 (A) , CFC-12 (B), CFC-113 (C) and SF_6 (D) and Estimated RTD (E), Identifiably (F) and error of estimated (G) data from monitoring well KEBE52 on the Delmarva peninsula in Maryland.

The method predicted a mean age of 5 years. This compares to apparent ages ranging between 7.1 years and 9.8 years for the same bore Busenberg and Plummer (2008) and an advective age of 8.1 years from groundwater modeling Reilly et al. (1994). In general however, little information can be accepted from the deconvolution undertaken here, as the method is clearly unable to capture the important processes and details that lead to groundwater concentrations at the site. However, the measure of identifiability gives

insight into what portions of the RTD can be informed by the type of data collected here.

5.5 Discussion and conclusions

The results of this paper show that even with a large amount of data, it is still difficult to use sparse measurements to provide robust estimates of the groundwater age distribution. The exceptions to this were the examples that used a long time series of tritium data but tritium cannot be used in systems with a large fraction of old water. The terms "long" and "old" are also relative terms and the age distributions we considered had maximum ages less than 10^2, which is a much narrower distribution of ages than many modeling studies have shown. At longer time scales, the tracers used here would have to be replaced by other naturally occurring tracers and there may be large uncertainties in the source concentrations and locations of such tracers. However, even in the absence of a high-resolution, long-term record of tracer concentrations, it is still possible to use the proposed method to provide a plausible estimate of a portion of the age distribution. All of the tracers we tested were able to provide accurate estimates of the mean ages and this can be used in simplified piston-flow models. These models obviously cannot account for dispersion of groundwater age, but they may be sufficient for guiding future work.

The ability of the procedure to fit environmental tracer concentrations, despite highly non-unique RTDs also raises important questions. The use of error minimisation between observed and modelled concentrations as a process by which to validate a choice of LPM seems somewhat flawed. More work is required to validate the choice of a specific LPM in a particular environment as data fit is not a reliable metric. Examples of this are provided by Massoudieh et al. (2012) and Eberts et al. (2012).

Additional issues include the choice of tracers to use in the future. Given the decay of bomb-peak tritium, the information gained from a time-series of tritium will be limited in the future. Tritium proved useful as atmospheric concentrations were not smoothly variable and the short half-life meant that sequential measurements were less correlated. The use of time series of CFCs and SF_6 are limited by the smooth variation of atmospheric concentrations. Krypton-85 has a similar short half-life to tritium however atmospheric concentrations vary smoothly which may reduce the usefulness of this compound.

One of our fundamental assumptions is that the flow system has reached a steady-state but this is not necessarily a valid assumption in many real-world scenarios. Transient conditions from long-term pumping Cornaton (2012), seasonal changes in the velocity field Gomez and Wilson (2013), or even higher frequency diurnal fluctuations can have profound effects on the age distribution. These effects are reflected (or will eventually be reflected) in the observed concentrations but it is the assumption of a steady-state RTD that is difficult to overcome without explicit knowledge of the transient conditions. Another restrictive assumption was that all of the tracers behave as conservative tracers with identical transport properties. In reality, the CFC compounds are susceptible to sorption and they can also be affected by degradation reactions (Cook and Solomon, 1997). If these processes are known to occur within the domain, it may be possible to account

for them but any uncertainty in the assumed process will propagate through the entire estimate (Massoudieh et al., 2012).

Most of the estimated and simulated age distributions in this work were generally unimodal; however, the simulated age distribution in Figure 5.2G is an example of a more complex age distribution that is composed of three dominant flow paths with very different mean ages. The dispersion in that example system is insufficient to further mix the flow paths but this is partially an artifact from the streamline model we used which does not allow the flow paths to interact, forcing scale dependent dispersion (Engdahl et al., 2013). In natural porous media, lateral mixing should smooth out some of the contrast in these distributions but the sharply contrasting multi-modal distribution is likely in fractured or layered aquifers, and complex heterogeneous domains often have multi-modal distributions also. Even though the estimates from our method are imperfect (only reliable estimates of the mean age were found), the approach appears to work well regardless of the heterogeneity structure of the system. This is promising since accounting for the different effects of heterogeneities is often one of the most difficult tasks when attempting to estimate age distributions (Engdahl et al., 2013).

The main strength of the method attempted here is that estimates of the RTD do not rely on the prior assumption of one or multiple models. Instead the estimated RTDs will only be a reflection of the relationship between historic atmospheric concentrations and those measured in an aquifer or at a discharge feature. Ultimately our work has shown that in some cases these relationships are not very sensitive. This has significant implications for the choice of LPMs and the use of environmental tracers to constrain groundwater models. Clearly the approach will require refinement before it can provide more accurate estimates of the entire age distribution, or even the higher moments of the age distribution. We speculate that this might be accomplished by combining elements of this approach with mechanistic models of age and/or LPMs to create a hybrid approach that uses the advantages of one approach to overcome the limitations of the others. Ultimately, it seems that such combinations of various methods will be required in order to determine age distributions, but the wide array of potential applications of age data make further pursuit of these approaches a worthwhile endeavor.

Acknowledgments

The authors wish to thank L. Niel Plummer for providing data for bore KEBE52. Additionally, Craig Simmons provided editorial comments. JM and PC wish to thank the National Water Commission and Australian Research Council for funding this research. NBE is supported by Award number P42ES004699 from the institute of Environmental Health Sciences and National Science Foundation Grant No. WSC-1204787. TRG is supported by the U.S National Science Foundation under Grant Nos. 1114257, 1234367, and 1215756.

6 Conclusions

6.1 Summary of the findings

The findings of the four specific studies resulted in the following findings:

i. An increase in the level of heterogeneity in aquifers results in an increase in the width of age distributions at a single point. These wide distributions in turn result in errors when attempting to estimate apparent ages from these mixtures. This is due to the non-linear temporal variations of these compounds. Given that individual environmental tracers present different temporal variations, the patterns of bias manifesting as differences in apparent ages may give insight into the mean age of the distributions.

ii. Errors in apparent ages, when compared to mean or advective ages can occur, due to transverse dispersion, heterogeneities or diffusion with aquitards. Often, these uncorrected ages are compared directly to models that assume they represent mean or advective ages. This may lead to considerable errors or biases if left uncorrected. Correction schemes have been proposed to overcome such errors. In the cases we tested, we found that methods that correct for advective ages do not result in significantly reduced errors, as the assumptions of the model are not met explicitly (i.e. wells fully screened in mobile zones). However, correction schemes that account for mixing biases in correcting for mean ages may provide better results. These however are limited in environments where the nature of the assumed mixing patter deviates substantially from the actual distribution.

iii. The estimation of residence time distributions is highly sensitive to heterogeneity. In the case we presented, multiple scales of high-K channels were controlling flow and transport in our synthetic two-dimensional aquifer. We found that the best geostatistical method to use in such environments was one that was able to recreate multiple scales of features. We demonstrated that an elementary training image with a large amount of conditioning points was able to do so with a minimal assumption of the underlying aquifer structure.

iv. The first two studies in this thesis demonstrated that assuming the shape of a distribution was a limitation in determining groundwater ages. The third study also demonstrated the limitations of using highly parametrised groundwater models to estimate groundwater residence time distribution. In this study we proposed a new method for estimating the shape of groundwater residence time distributions. We found that the relationships between environmental tracer concentrations were highly non-unique, meaning that many distributions may result in similar concentrations. However, the method was able to give favorable estimates of mean ages and early arrival times. The findings of this study have implications for the use of lumped parameter model, in that the choice of a distribution can not be validated by the fit of the data.

6.2 Future work

An important extension of this work would be the application of the concepts presented to field scenarios. Although in such cases results are unknown for comparisons, some of this work forms a foundation for understanding the limitations of these methods in such contexts.

An extension of the second manuscript would be to apply the errors found in the paper to estimates made with environmental tracers (i.e. - calibrating groundwater models or estimating recharge and flow rates). This could be as an assessment of the method of Vogel (1967) in a quantitative manner. In the context of groundwater models, it would be useful to investigate the various ways to incorporate environmental tracer data into numerical models. Independent assessments of the errors associated with advective ages, mean ages and straight concentrations have been made by Janssen et al. (2008), Ginn et al. (2009) and Starn et al. (2010). If the methods were compared in the same framework, this would give some advice on the best ways to implement environmental tracers into groundwater models and the errors and biases associated with the choice of these methods.

Additional studies may incorporate the work of the third and fourth manuscripts. The incorporation of both environmental tracer data and hydraulic conductivity data in a compatible framework may improve the assessment of residence time distributions.

References

Bauer, S., C. Fulda, and W. Schafer, 2001: A multi-tracer study in a shallow aquifer using age dating tracers ^{3}H, ^{85}Kr, CFC-113 and SF_6 - indication for retarded transport of CFC-113. *J. Hydrol*, **248**, 14 – 34, doi:10.1016/S0022-1694(01)00381-X.

Bethke, C. M. and T. M. Johnson, 2002a: Ground water age. *Ground Water*, **40 (4)**, 337 – 339, doi: 10.1111/j.1745-6584.2002.tb02510.x.

Bethke, C. M. and T. M. Johnson, 2002b: Paradox of groundwater age. *Geology*, **30 (2)**, 107 – 110, doi: 10.1130/0091-7613(2002)030<0107:POGA>2.0.CO;2.

Bethke, C. M. and T. M. Johnson, 2008: Groundwater age and groundwater age dating. *Annual Review of Earth and Planetary Science 2008*, **36**, 121 – 152, doi:10.1146/annurev.earth.36.031207.124210.

Bohlke, J. K. and J. M. Denver, 1995: Combined use of groundwater dating, chemical, and isotopic analyses to resolve the history and fate of nitrate contamination in two agricultural watersheds, Atlantic coastal plain, Maryland. *Water Resour. Res.*, **31 (9)**, 2319 – 2339, doi:10.1029/95WR01584.

Busenberg, E. and L. N. Plummer, 1992: Use of chlorofluorocarbons (CCl_3F and CCl_2F_2) as hydrologic tracers and age-dating tools: The alluvium and terrace system of central Oklahoma. *Water Resour. Res.*, **28 (9)**, 2257–2283, doi:10.1029/92WR01263.

Busenberg, E. and L. N. Plummer, 2000: Dating young water with sulfur hexafluoride: Natural and anthropogenic sources of sulfur hexafluoride. *Water Resour. Res.*, **36 (10)**, 3011–3030, doi: 10.1029/2000WR900151.

Busenberg, E. and L. N. Plummer, 2008: Dating groundwater with trifluoromethyl sulfurpentafluoride (SF_5CF_3,sufur hexafluoride SF_6, CF_3Cl (CFC-13) and CF_2Cl_2 (CFC-12). *Water Resour. Res.*, **44**, W02431, doi:10.1029/2007WR006150.

Cardenas, M. B., 2007: Potential contribution of topography-driven regional groundwater flow to fractal stream chemistry: Residence time distribution analysis of Toth flow. *Geophys. Res. Lett.*, **34**, L05403, doi:10.1029/2006GL029126.

Cardenas, M. B. and X. W. Jiang, 2010: Groundwater flow, transport, and residence times through topography-driven basins with exponentially decreasing permeability and porosity. *Water Resour. Res.*, **46**, W11538, doi:10.1029/2010WR009370.

Carle, S. F. and G. E. Fogg, 1996: Transition probability-based indicator geostatistics. *Math. Geol.*, **28**, 453 – 477, doi:10.1007/BF02083656.

Castro, M. C. and P. Goblet, 2005: Calculation of ground water ages - a comparitive analysis. *Ground Water*, **43 (3)**, 368–380, doi:10.1111/j.1745-6584.2005.0046.x.

Cirpka, O. A., M. N. Fienen, M. Hofer, E. Hoehn, A. Tessarini, R. Kipfer, and P. K. Kitanidis, 2007: Analyzing bank filtration by deconvoluting time series of electric conductivity. *Ground Water*, **45 (3)**, 318–328, doi:10.1111/j.1745-6584.2006.00293.x.

Clark, I. D. and P. Fritz, 1997: *Environmental Isotopes in Hydrogeology*. CRC Press, USA.

Cook, P. G. and J. K. Bohlke, 2000: Determining time frames for groundwater flow and solute transport. *Environmental tracers in subsurface hydrology*, P. G. Cook and A. L. Herczeg, Eds., Kluwer, Massachusetts, 1–30.

Cook, P. G. and A. L. Herczeg, 2000: *Environmental Tracers in Subsurface Hydrology*. Kluwer, Boston.

Cook, P. G., A. J. Love, N. I. Robinson, and C. T. Simmons, 2005: Groundwater ages in fractured rock aquifers. *J. Hydrol.*, **308**, 284 – 301, doi:10.1016/j.jhydrol.2004.11.005.

Cook, P. G. and D. K. Solomon, 1995: Transport of atmospheric trace gasses to the water table: Implications for groundwater dating with chlorofluorocarbons and krypton 85. *Water Resour. Res.*, **31 (2)**, 263 – 270, doi:doi:10.1029/94WR02232.

Cook, P. G. and D. K. Solomon, 1997: Recent advences in dating young groundwater: chlorofluorocarbons, ^{3}H/^{3}He and ^{85}Kr. *J. Hydrol.*, **191**, 245 – 265, doi:10.1016/S0022-1694(96)03051-X.

Corcho Alvardo, J. A., et al., 2007: Constraining the age distribution of highly mixed groundwater using ^{39}Ar: A multiple environmental tracer (^{3}H/^{3}He, ^{85}Kr, ^{39}Ar and ^{14}C) study in the semiconfined Fontainebleau Sands Aquifer (France). *Water Resour. Res.*, **43**, W03427, doi:10.1029/2006WR005096.

Cornaton, F., 2012: Transient water age distributions in environmental flow systems: The time-marching Laplace transform solution technique. *Water Resour. Res.*, **48**, W03524, doi:10.1029/2011WR010606.

Cornaton, F., Y.-J. Park, and E. Deleersnijder, 2011: On the biases affecting water ages inferred from isotopic data. *J. Hydrol.*, **410**, 217–225, doi:10.1016/j.jhydrol.2011.09.024.

Cornaton, F. and P. Perrochet, 2006a: Groundwater age, life expectancy and transit time distributions in advective-dispersive systems: 1. Generalised reservoir theory. *Adv. Water Res.*, **29**, 1267 – 1291, doi:10.1016/j.advwatres.2005.10.009.

Cornaton, F. and P. Perrochet, 2006b: Groundwater age, life expectancy and transit time distributions in advective-dispersive systems: 2. Reservoir theory for sub-drainage basins. *Adv. Water Res.*, **29**, 1292 – 1305, doi:10.1016/j.advwatres.2005.10.010.

Danckwerts, P. V., 1952: Continuous flow systems: Distribution of residence times. *Chem. Engng Sci.*, **2**, 1–13, doi:10.1016/0009-2509(53)80001-1.

de Marsily, G., F. Delay, J. Goncalves, P. Renard, V. Teles, and S. Violette, 2005: Dealing with spatial heterogeneity. *Hydrogeol. J.*, **13**, 161–132, doi:10.1007/s10040-004-0432-3.

dell'Arciprete, D., R. Bersezio, F. Felleti, M. Giudici, A. Comunian, and P. Renard, 2011: Comparison of three geostatistical methods for hydrofacies simulation: a test on alluvial sediments. *Hydrogeol. J.*, **20 (2)**.

Desbarats, A. J., 1990: Macrodispersion in sand-shale sequences. *Water Resour. Res.*, **26 (1)**, 153–163, doi:10.1029/WR026i001p00153.

Deutsch, C. V. and A. G. Journel, 1998: *GSLIB Geostatistical Software Library and User's Guide, 2nd ed.* Oxford University Press, New York.

Doherty, J. and R. J. Hunt, 2009: Two statistics for evaluating paramater identifiability and error reduction. *J. Hydrol*, **366**, 119–127, doi:10.1016/j.jhydrol.2008.12.018.

Domenico, P. A. and F. W. Schwartz, 1998: *Physical and Chemical Hydrogeology, 2nd ed.* John Wiley & Sons, Brisbane, Australia.

Dunkle, S. A., L. N. Plummer, E. Busenberg, P. J. Phillips, J. M. Denver, P. A. Hamilton, R. L. Michel, and T. B. Coplen, 1993: Chlorofluorocarbons (CCl_3F and CCl_2F_2) as dating tools and hydrologic tracers in shallow groundwater of the Delmarva Penninsula, Atlantic Coastal Plain, United States. *Water Resour. Res.*, **29 (12)**, 3837–3860, doi:10.1029/93WR02073.

Eberts, E., J. Bohlke, L. Kauffman, and B. Jurgens, 2012: Comparison of particle-tracking and lumped-parameter age-distribution models for evaluating vulnerability of production wells to contamination. *Hydrogeol. J.*, **20**, 263–282, doi:10.1007/s10040-011-0810-6.

Ekwurzel, B., P. Schlosser, W. M. Smethie Jr., L. N. Plummer, E. Busenberg, R. L. Michel, R. Weppernig, and M. Stute, 1994: Dating of shallow groundwater: Compsarison of the transient tracers $^3H/^3He$, chlorofluorocarbons, and ^{85}Kr. *Water Resour. Res.*, **30 (6)**, 1693 – 1708, doi:10.1029/94WR00156.

Engdahl, N. B., T. R. Ginn, and G. E. Fogg, 2012: Non-Fickian dispersion of groundwater age. *Water Resour. Res.*, **48**, W07508, doi:10.1029/2012WR012251.

Engdahl, N. B., T. R. Ginn, and G. E. Fogg, 2013: Using groundwater age distributions to estimate the effective parameters of Fickian and non-Fickian models of solute transport. *Adv. Water Res.*, **54**, 11–21, doi:10.1016/j.advwatres.2012.12.008.

Etcheverry, D. and P. Perrochet, 2000: Direct simulations of groundwater transit-time distributions using the reservoir theory. *Hydrogeol. J.*, **8**, 200–208, doi:10.1007/s100400050006.

Fleckenstein, J. H. and G. E. Fogg, 2008: Efficient upscaling of hydraulic conductivity in heterogeneous alluvial aquifers. *Hydrogeol. J.*, **16**, 1239–1250, doi:10.1007/s10040-008-0312-3.

Freeze, R. and J. Cherry, 1979: *Groundwater*. Prentice-Hall, Sydney.

Gelhar, L., C. Welty, and K. Refheldt, 1992: A critical review of data on field-scale dispersion in aquifers. *Water Resour. Res.*, **29 (7)**, 1955–1974, doi:10.1029/92WR00607.

Gelhar, L. W., 1993: *Stochastic Subsurface Hydrology*. Prentice-Hall, Sydney.

Genereux, D. P., M. Webb, and D. K. Solomon, 2009: Chemical and isotopic signiture of old groundwater and magmatic solutes in a costa rican rain forest: Evidence from carbon, helium and chlorine. *Water Resour. Res.*, **45**, W08413, doi:10.1029/2008WR007630.

Ginn, T. R., 1999: On the distribution of multicomponent mixtures over generalized exposure time in subsurface flow and reactive transport: Foundations, and formulations for groundwater age, chemical heterogeneity, and biodegradation. *Water Resour. Res.*, **35 (5)**, 1395–1407, doi:10.1029/1999WR900013.

Ginn, T. R., H. Haeri, A. Massoudieh, and L. Foglia, 2009: Notes on groundwater age in forward and inverse modeling. *Transp. Porous Med.*, **79**, 117–134, doi:10.1007/s11242-009-9406-1.

Gomez, J. D. and J. L. Wilson, 2013: Age distributions and dynamically changing hydrologic systems: Exploring topographically-driven flow. *Water Resour. Res.*, **49**, 1503–1522, doi:10.1002/wrcr.20127.

Gomez-Hernandez, J. J. and X. Wen, 1998: To be or not to be multi-Gaussian? a reflection on stoichastic hydrogeology. *Adv. Water Res.*, **21 (1)**, 47–61, doi:10.1016/S0309-1708(96)00031-0.

Goode, D. J., 1996: Direct simulation of groundwater age. *Water Resour. Res.*, **32 (2)**, 289–296, doi:10.1029/95WR03401.

Goovaerts, 1997: *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.

Hess, K. M., S. H. Wolf, and M. A. Celia., 1992: Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts: 3. Hydraulic conductivity variability and calculated macrodispersivities. *Water Resour. Res.*, **28 (8)**, 2011–2027, doi:10.1029/92WR00668.

Janssen, G. M. C. M., J. R. Valstar, and S. E. A. T. M. van der Zee, 2008: Measurement network design including traveltime determinations to minimize model prediction uncertainty. *Water Resour. Res.*, **44**, W02405, doi:10.1029/2006WR005462.

Jones, E., T. Oliphant, P. Peterson, et al., 2001–: *SciPy: Open source scientific tools for Python*. URL `http://www.scipy.org/`.

Kerrou, J., P. Renard, H.-J. Hendricks Franssen, and I. Lunati, 2008: Issues characterising heterogeneity and connectivity in non-Gaussian media. *Adv. Water Res.*, **31**, 147 – 159, doi:10.1016/j.advwatres.2007.07.002.

Klise, K. A., G. S. Weissmann, S. McKenna, E. M. Nichols, J. D. Frechette, T. F. Wawryniec, and V. C. Tidwell, 2009: Exploring solute transport and streamline connectivity using lidar based outcrop images and geostatistical representations of heterogeneity. *Water Resour. Res.*, **45**, W05413, doi:10.1029/2008WR007500.

Knowles Jr., L., B. G. Katz, and D. J. Toth, 2010: Using multiple chemical indicators to characterize and determine the age of groundwater from selected vents of the Silver Springs Group, central Florida, USA. *Hydrogeol. J.*, **18 (1)**, 1825–1838, doi:10.1007/s10040-010-0669-y.

Knudby, C. and J. Carrera, 2005: On the relationship between indicators of geostatistical, flow and transport connectivity. *Adv. Water Res.*, **28**, 405–421, doi:10.1016/j.advwatres.2004.09.001.

Labolle, E. M. and G. E. Fogg, 2001: Role of molecular diffusion in contaminant migration and recovery in an alluvial aquifer. *Transp. Porous Med.*, **42**, 155–179, doi:10.1023/A:1006772716244.

LaBolle, E. M., G. E. Fogg, and J. B. Eweis, 2006: Diffusive fractionation of ^{3}H and ^{3}He in groundwater and it's impact on groundwater age estimates. *Water Resour. Res.*, **42**, W07202, doi:10.1029/2005WR004756.

LaBolle, E. M., G. E. Fogg, and A. F. B. Tompson, 1996: Random-walk simulation of transport in heterogeneous media: Local mass-conservation problem and implementation methods. *Water Resour. Res.*, **32**, 583–593, doi:10.1029/95WR03528.

Land, L. and G. F. Huff, 2009: Multi-tracer investigation of groundwater residence time in a karstic aquifer: Bitter Lakes National Wildlife Refuge, New Mexico, USA. *Hydrogeol. J.*, **18**, 455–472, doi: 10.1007/s10040-009-0522-3.

Larocque, M., P. G. Cook, K. Haaken, and C. T. Simmons, 2009: Estimating flow using tracers and hydraulics in synthetic heterogeneous aquifers. *Ground Water*, **47 (6)**, 786–796, doi:10.1111/j.1745-6584.2009.00595.x.

Leray, S., J.-R. de Dreuzy, O. Bour, T. Labasque, and L. Aquilina, 2012: Contribution of age data to the characterization of complex aquifers. *J. Hydrol.*, **464-465**, 54–68, doi:10.1016/j.jhydrol.2012.06.052.

Loosli, H. H., B. E. Lehman, and W. M. Smethie Jr., 2000: Noble gas radioisotopers ^{37}Ar, ^{85}Kr, ^{39}Ar and ^{81}Kr. *Environmental tracers in subsurface hydrology*, P. G. Cook and A. L. Herczeg, Eds., Kluwer, Massachusetts, 379–396.

Maharajh, D. M. and J. Walkley, 1973: The temperature dependance of the diffusion coefficients of Ar, CO_2, CH_4, CH_3Br and $CHCl_2F$ in water. *Can. J. Chem.*, **51**, 944 – 952.

Maloszewski, P. and A. Zuber, 1982: Determining the turnover time of groundwater systems with the aid of environmental tracers: 1. Models and their applicability. *J. Hydrol*, **57**.

Mantz, H., K. Jacobs, and K. Mecke, 2008: Utilizing minkowski functionals for image analysis: a marching square algorithm. *Journal of Statistical Mechanics:Theory and Experiment*, **12**, P12015, doi: 10.1016/0022-1694(82)90147-0.

Mariethoz, G. and B. F. J. Kelly, 2011: Modelling complex geological structures with elementary training images and transform-invariant distances. *Water Resour. Res.*, **47**, W07527, doi: 10.1029/2011WR010412.

Mariethoz, G., P. Renard, and J. Straubhaar, 2010: The Direct Sampling method to perform multiple-point geostatistical simulations. *Water Resour. Res.*, **46**, W11526, doi:10.1029/2008WR007621.

Massoudieh, A. and T. R. Ginn, 2011: The theoretical relationship between unstable solutes and groundwater age. *Water Resour. Res.*, **47**, W10523, doi:10.1029/2010WR010039.

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

McCallum, J. L., P. G. Cook, C. T. Simmons, and A. D. Werner, In Press: Bias of environmental tracer ages in heterogeneous environments. *Ground Water*, doi:10.1111/gwat.12052.

McMahon, P., L. Plummer, J. Bohlke, S. Shapiro, and S. Hinkle, 2011: A comparison of recharge rates in aquifers of the United States based on groundwater-age data. *Hydrogeol. J.*, **19**, 779–800, doi:10.1007/s10040-011-0722-5.

Michael, H. and C. Voss, 2009: Controls on groundwater flow in the Bengal Basin of India and Bangladesh: Regional modelling analysis. *Hydrogeol. J.*, **17**, 1561–1577, doi:10.1007/s10040-008-0429-4.

Molson, J. W. and E. O. Frind, 2012: On the use of mean groundwater age, life expectancy and capture probability for defining aquifer vulnerability and time-of-travel zones for source water protection. *J. of Contam. Hydrol.*, **35 (3)**, 523 – 537.

Moore, C. and J. Doherty, 2005: Role of the calibration process in reducing model predictive error. *Water Resour. Res.*, **41**, W05020, doi:10.1029/2001WR001146.

Mustapha, H., R. Dimitrakopoulos, and S. Chatterjee, 2011: Geologic heterogeneity representation using high order spatial cumulants fo subsurface flow and transport simulations. *Water Resour. Res.*, **47**, W08536, doi:10.1029/2010WR009515.

Neumann, R., E. Labolle, and C. Harvey, 2008: The effects of dual-domain mass transfer on the tritium-helium-3 dating method. *Environ. Sci. Technol.*, **42 (13)**, 4837–43, doi:10.1021/es7025246.

Park, J., C. M. Bethke, T. Torgerson, and T. M. Johnson, 2002: Transport modeling applied to the interpretation of groundwater ^{36}Cl age. *Water Resour. Res.*, **38 (5)**, 1043, doi:10.1029/2001WR000399.

Pint, C. D., R. J. Hunt, and M. P. Anderson, 2003: Flowpath deliniation and groundwater age, Allequash Basin, Wisconsin. *Ground Water*, **41 (7)**, 895 – 902, doi:10.1111/j.1745-6584.2003.tb02432.x.

Pohll, G. M., J. J. Warwick, and D. Benson, 2000: On the errors associated with two-dimensional stochastic solute transport models. *Transp. Porous Med.*, **40**, 281–293., doi:10.1023/A:1006786714947.

Portniaguine, O. and D. K. Solomon, 1998: Parameter estimation using groundwater age and head data, Cape Cod, Massachusetts. *Water Resour. Res.*, **34 (4)**, 637 – 645, doi:10.1029/97WR03361.

Post, V., A. Vandenbohede, A. Werner, Maimun, and M. Teubner, 2013: Groundwater ages in coastal aquifers. *Adv. Water Res.*, **57**, 1–11, doi:10.1016/j.advwatres.2013.03.011.

Press, W., B. Flannery, S. Teukolsky, and W. Vetterling, 1986: *Numreical recipes, the art of scientific computing*. Cambridge University Press, Melbourne, Australia.

Rank, D. and W. Papesch, 2003: Determination of groundwater flow velocity in the Southern Vienna Basin from long-term enviromnetal isotope records. Abstract volume of the First Conference on Applied Environmental Geology in Central and Eastern Europe: BE-228, 206–207.

Rehfeldt, K. R., J. M. Boggs, and L. W. Gelhar, 1992: Field study of dispersion in a heterogeneous aquifer 3. Geostatistical analysis of hydraulic conductivity. *Water Resour. Res.*, **28 (5)**, 3309–3324, doi: 10.1029/92WR01758.

Reilly, T. E., L. N. Plummer, P. J. Phillips, and E. Busenberg, 1994: The use of simulation and multiple environmental tracers to quantify groundwater flow in a shallow aquifer. *Water Resour. Res.*, **30 (2)**, 421 – 433, doi:10.1029/93WR02655.

Renard, P. and D. Allard, 2013: Connectivity metrics for subsurface flow and transport. *Adv. Water Res.*, **51**, 168 – 196, doi:10.1016/j.advwatres.2011.12.001.

Salamon, P., D. Fernandez-Garcia, and J. J. Gomez-Hernandez, 2006: A review and numerical assessment of the random walk particle tracking method. *J. of Contam. Hydrol.*, **35 (4)**, 943–952, doi: 10.1016/j.jconhyd.2006.05.005.

Sanford, W. E., 1997: Correcting for diffusion in carbon-14 dating of ground water. *Ground Water*, **35 (2)**, 357–361, doi:10.1111/j.1745-6584.1997.tb00093.x.

Sanford, W. E., 2011: Calibrating models using groundwater age. *Hydrogeol. J.*, **19**, 13–16, doi: 10.1007/s10040-010-0637-6.

Sanford, W. E. and S. Buapeng, 1996: Assessment of a groundwater flow model of the Bangkok Basin, Thailand using carbon-14 based ages and paleohydrology. *Hydrogeol. J.*, **4 (4)**, 26–40, doi: 10.1007/s100400050083.

Sanford, W. E., L. N. Plummer, D. McAda, L. Bexfield, and S. Anderholm, 2004: Hydrochemical tracers in the middle Rio Grande Basin, USA: 2. Calibration of a groundwater flow model. *Hydrogeol. J.*, **12**, 389–407, doi:110.1007/s10040-004-0326-4.

Schlosser, P., M. Stute, C. Sonntag, and K. O. Munnich, 1989: Tritiogenic ^{3}He in shallow groundwater. *Earth Planet. Sci. Lett.*, **94**, 245 – 256, doi:10.1016/0012-821X(89)90144-1.

Schluter, S. and H. Vogel, 2011: On the reconstruction and functional properties in random heterogeneous media. *Adv. Water Res.*, **34**, 314 – 325, doi:10.1016/j.advwatres.2010.12.004.

Solomon, D. K., D. P. Genereux, L. N. Plummer, and E. Busenberg, 2010: Testing mixing models of old and young groundwater in a tropical lowland rain forest with environmental tracers. *Water Resour. Res.*, **46**, W04518, doi:10.1029/2009WR008341.

Solomon, D. K., S. L. Schiff, R. J. Poreda, and W. B. Clarke, 1993: A validation of the $^3H/^3He$ method for determining groundwater recharge. *Water Resour. Res.*, **29 (9)**, 2951 – 2962, doi:10.1029/93WR00968.

Solomon, D. K. and E. A. Sudicky, 1991: Tritium and helium 3 isotope ratios for direct estimation of spatial variations in groundwater recharge. *Water Resour. Res.*, **27 (9)**, 2309 – 2319, doi:10.1029/91WR01446.

Starn, J., A. Bagtzoglou, and G. Robbins, 2010: Using atmospheric tracers to reduce uncertainty in groundwater recharge areas. *Ground Water*, **48 (6)**, 858–868, doi:10.1111/j.1745-6584.2010.00674.x.

Stolp, B. J., D. K. Solomon, A. Suckow, T. Vitvar, D. Rank, P. K. Aggarwal, and L. F. Han, 2010: Age dating base flow at springs and gaining streams using helium-3 and tritium: Fishca-Dagnitz system, southern Vienna Basin, Austria. *Water Resour. Res.*, **46**, W07503, doi:10.1029/2009WR008006.

Strebelle, S., 2002: Conditional simulation of complex geological structures using multiple-point statistics. *Math. Geol.*, **34 (1)**, 1–21.

Sudicky, E. A. and E. O. Frind, 1981: Carbon 14 dating of groundwater in confined aquifers: Implications of aquitard diffusion. *Water Resour. Res.*, **17 (4)**, 1060–1064, doi:10.1029/WR017i004p01060.

Sultenfuss, J., R. Purtschert, and J. F. Fuhrboter, 2011: Age structure and recharge conditions of a coastal aquifer (northern Germany) investigated with ^{39}Ar, ^{14}C, ^3H, He isotopes and Ne. *Hydrogeol. J.*, **19**, 221–236, doi:10.1007/s10040-010-0663-4.

Therrien, R., R. G. Mclaren, E. A. Sudicky, and S. M. Panday, 2006: *Hydrogeosphere*. Groundwater Simul. Group, Univ. of Waterloo, Waterloo, Ontario, Canada.

Troldborg, L., J. Refsgaard, K. Jensen, and P. Engesgaard, 2007: The importance of alternative conceptual models for simulation of concentrations in a multi-aquifer system. *Hydrogeol. J.*, **15**, 843–860, doi: 10.1007/s10040-007-0192-y.

Vandenbohede, A., K. Hinsby, C. Courtens, and L. Lebbe, 2011: Flow and transport model of a polder area in the Belgian coastal plain: example of data integration. *Hydrogeol. J.*, **19**, 1599–1615, doi: 10.1007/s10040-011-0781-7.

Varni, M. and J. Carrera, 1998: Simulations of groundwater age distributions. *Water Resour. Res.*, **34(12)**, 3271–3281, doi:10.1029/98WR02536.

Vogel, J. C., 1967: Investigation of groundwater flow with radiocarbon. *Isotopes in Hydrology*, IAEA, Vienna, 355–369.

Waugh, D. W., T. M. Hall, and T. W. N. Haine, 2003: Relationships amoung tracer ages. *J. Geophys. Res.*, **108**, doi:10.1029/2002JC001325.

Weissmann, G. S., Y. Zhang, E. M. Labolle, and G. E. Fogg, 2002: Dispersion of groundwater age in an alluvial aquifer system. *Water Resour. Res.*, **38 (10)**, 1198, doi:10.1029/2001WR000907.

Wen, X.-H. and J. Gomez-Hernandez, 1998: Numerical modelling of macrodispersion in heterogeneous media: a comparison of multi-Gaussian and non-multi-Gaussian models. *J. Contam. Hydrol.*, **30**, 129–156, doi:10.1016.S0169-7722(97)0035-1.

Woolfenden, L. R. and T. R. Ginn, 2009: Modeled ground water age distributions. *Ground Water*, **47 (4)**, 547–557, doi:10.1111/j.1745-6584.2008.0550.x.

Zinn, B. and C. F. Harvey, 2003: When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields. *Water Resour. Res.*, **39 (3)**, 1051, doi:10.1029/2001WR001146.