Dear editor,

Many thanks for giving us the opportunity to comment on the reviewers' feedback.

Both reviewers had only minor comments. We followed all of their suggestions as detailed below.

Dear Daren,

We greatly appreciate your valuable feedback on our manuscript. Below, we detail our responses to your comments. We closely followed all of your suggestions.

Comments from D. Gooddy (Referee) dcg@bgs.ac.uk Received and published: 30 March 2017

This manuscript presents and extensive data set for H1301, building on that previously published by the lead author in HESS 19, 2775-2789, 2015 and WRR, 50, WR015818, 2014.

Main Comments: Despite the impressive data set I was slightly disappointed at the level of greater understanding that was gained from this. In particular, the unavailability of other tracers at some sites which would have hopefully given greater lucidity as to the retardation/removal processes taking place. I think this is a significant weakness in the paper although not one that the authors can rectify.

- ➔ Thanks for pointing this out. We agree with your comment toned down our message re further insight into the causes of reduced Halon-1301 concentrations in abstract and conclusion.

I do think however that more thought needs to go into the discussion as this is key to the main knowledge advancement that the paper could potentially provide.

- ➔ Thanks, we agree with your comment and included a more detailed discussion around degassing following your suggestions below.

Where low concentrations of H1301 are found, have the authors considered degassing of N2 (as a result of denitrification) or CH4 as possible mechanisms for removal. Without any NO3 data this is hard for the reviewer to assess. I would therefore refer the authors to Visser et al 2007 (WRR 43, 10 W10434) and Visser et al. 2009 (JoH 369, 4-4, 427-439) where the issue of tracer degassing is discussed in extensive detail.

- ➔ Thanks for pointing this out. Unfortunately we do not have NO3 data for over half of the sites to further assess degassing as possible cause of reduced Halon-1301 concentrations. However, following your suggestions we determined excess N as described below and discuss degassing as possible cause of reduced Halon-1301 concentrations.
- ➔ Further, to highlight that we had previously considered a more comprehensive range of possible causes of Halon-1301 'removal' (in our HESS 19, 2775-2789, 2015 paper), we included a more detailed summary of these in the introduction (page 3, line 14ff.).

Related to this, I am interested that the authors are using N2/Ar ratios to correct for excess air, rather than the more normally accepted Ne. Could they comment on the possible issues relating to this, especially if denitrification is taking place.

➔ While Ne/Ar is more robust, N2/Ar is much simpler to measure and still provides a useful excess air and recharge temperature correction in most cases. Significant denitrification (excess N) can be identified by anomalously high recharge temperatures. In such cases, excess N is corrected for by applying the mean annual air temperature. That method also allows for estimation of excess N. We think further assessment of degassing into excess N as possible cause of Halon-1301 'removal' would add significantly to the discussion on reduced Halon-1301 concentrations. We therefore addressed the above using the estimated presence of excess N. Please see page 5, line 5 ff. for a description of the method and page 13, line 20ff for results.

As a general observation there are far too many figures and figures within figures – As these are not really discussed in any detail, the true significance is not clear.

➔ Thanks for your comment. We assessed the significance of each figure and removed figure 6 and 7.

Minor Comments:

P1 Line 13. Could not rather than "couldn't" → Agreed and changed accordingly.

P1 Line 29. More description on the (speculation?) causes of H1301 reduction is needed here. → Yes, please see changes on page 3, line 14ff.

In the Introduction section you are only really referring to recent groundwater age indicators and you need to be explicit about that. As a general over view of the state of the art I would refer the authors to Aqulina et al 2014 (Applied Geochemistry 50, 115-117) and Darling et al 2012 (Applied Geochemistry 27, 9, 1688-1697). P2 line 13 would add Darling and Gooddy 2007 (Science of the Total Environment 387, 353-362) → Agreed. Please see changes in page 2, line 18 and 19.

P2 Line 15 after ambiguous age interpretations add Suckow 2014 (Applied Geochemistry 50, 222-230) ; P3 Line 12 replace 'They' with Bartyzel et al (2016); P4 Line 8 add reference to Oster 1996 → Thanks, changed accordingly.

P6 Question. Is the input function for S Hemisphere and N hemisphere the same for H1301? Some reference to the differences would be helpful for other/future practitioners. → We added a note to the caption of figure 1 to clarify the above.

P8 Line 19. Delete 'in fact'; P9 Line 18. Give reference for the 'issues' eluded to. → Agreed and changes made accordingly.

P11 Line 10. Need to justify assertion that T is one of the 'most reliable'. What do you mean by reliable? → Limitations of the tracers were described in the introduction. We included a link to the introduction to clarify the above.

P11 Line 14. The input of SF6 is exponential and not 'near linear'. → Clarified: near linear from 1985 onwards.

P13 line 15. Add in text relating to degassing potential. → Thanks, done accordingly.

Dear Axel,

Many thanks for your valuable feedback on our manuscript. Below, we detail our responses to your comments. We closely followed your suggestions.

Comments from A. Suckow (Referee) axel.suckow@csiro.au Received and published: 28 May 2017

General Comments: This manuscript massively extents the data set for Halon 1301 in New Zealand (>300 measurements) compared to the 2015 paper in HESS of the same authors, and compares its usability with tritium, SF6 and the CFCs (CFC-11, CFC-12, CFC-113). The analysis uses a lumped parameter (LPM) approach – the exponential piston flow model (EPM) with an evaluation of mean residence time (MRT) and volume ratio (E/PM) for this model – to assess agreement in "groundwater age" as inferred by the different tracers. It is very well presented and besides the obvious demonstration of the usefulness of H1301 the article also shows some valuable new approaches in demonstrating and quantifying agreement in lumped parameter model results. It is clearly worth publishing with minor revisions in HESS.

Specific Comments: Although using only one specific shape of an age distribution (the EPM), the paper does a good job in evaluating comparability of LPM results for different tracers for the same water sample as "agreement in inferred age". Of special value here is the 2d-plot of the E/PM parameter versus the MRT for different tracers if the model results are in the 1-sigma range of the measurement. This is a very useful way to display these results that I have not seen before. The developed metric, however, is worth discussing in more detail.

While I completely agree not to use MRT only, I have a problem with the attempted metric. No overlap of the two data clouds in Fig. 4 simply means that the two tracers give differing results which cannot be brought into a 1-sigma agreement with any parameter combination. So what is the meaning of the "Euclidian distance between two data clouds"? And what is the "% difference in MRT and mixing parameter inferred with two tracers"? Is the percentage (distance divided by what?) taken from the MRT and mixing parameter of the two nearest points or from the whole axis? For example, if the nearest two point in figure 4 left have 20 and 30 years MRT, is the percentage in MRT then (30-20)/(20)=50%? Or is it (30-20)/100=10%?

- ➔ Thanks for pointing this out. We agree that more discussion around the developed metric would be great. We included your comment re "No overlap of the inferred LPM parameter clouds implies that the two tracers give differing results which cannot be brought into a 1-sigma agreement with any parameter combination."(page 8, line 25ff). We also included an equation for the metric '% difference in MRT' to clarify its definition. → see Eqn. 5

Most of the following paper uses a 10% criterion on this distance as "agreement" (Fig. 6, 7, 9, 11, 12) which is misleading, because any percentage >0 means the two results actually disagree. Perhaps a better way for quantifying agreement or disagreement would be to use a 1-sigma and a 2-sigma evaluation. Overlap of the clouds generated with 1 sigma would be good agreement, overlap of clouds with 2 sigma still agreement with a certain smaller probability. No overlap of data clouds generated with 3 sigma would be clear disagreement. In case of disagreement (e.g. 6% of the sites with SF6 and 3H available) an evaluation of the uncertainty in recharge temperature, recharge altitude and excess air may be valuable – perhaps this would bring the results into agreement within the 1 sigma uncertainty of these parameters? Similar for H1301, since its dependency on temperature, altitude and excess air is different than for SF6.

- ➔ Thanks for your suggestions. We generally agree that it is a good idea to use multiple distance measure levels to assess agreement/disagreement. As this is one of the first attempts to use that type of approach, we decided to use 1 sigma uncertainty for the following reason. As we illustrate, most inferred LPM parameter clouds are already relatively large when using the 1 sigma criterion implying large uncertainties in inferred age parameters (this is not surprising as we only use 1 measurement to infer age). We are worried that when using a 2 sigma (or higher) distance criterion, huge inferred LPM parameter may suggest to the reader that the tracers are of no use. We added the 10% criterion to add another level of check assessing relative distance in addition to absolute distance of the cloud as per 1 sigma criterion.
- ➔ In general, a variety of objective functions (which our criterion in principle is), is increasingly used in the hydrological modelling community suggesting that there is no one criterion that should be applied everywhere. We included a note on the relative novelty of this approach and highlight that its general applicability needs to be assessed further for other datasets. – see page 9 ff.

Technical Corrections

Attempting to determine a MRT of 150 years with the given tracers (P. 3 line 22) is too ambitious and does not acknowledge the high quality of LPM presentation of the rest of the paper. None of the discussed tracers is sensitive to water recharged prior to 1950 (not even with the high sensitivity reached by Uwe for tritium). This is 67 years, not 150. Even using MRT instead of "age": an EPM with an E/PM of 0.1 (bottom row in plots of figure 15-20) and MRT of 75 years contains none of the tracers (all water in it is >67.5years old), and only <> 10%". This is a strange use of the symbols "<>". I think a better way to express what you mean is "disagreement

- ➔ Thanks for the above comment. To clarify, a water with MRT 75y and E/PM 0.1 contains 59% of water younger than 67.5 years. Water with MRT 150 years still contains 33% of water younger than 67.5 years and therefore contains significant amounts of tritium. For this model example we can detect up to 250 years MRT. So by saying 150 years we are not over-ambitious. However, we agree that for such old (nearly tritium-free) waters, the aggregation error is large – the old component of the water could be thousands of years old and therefore the true mean age be significantly older [see e.g. Stewart et al., 2016 - doi:10.5194/hess-2016-532]. We included a footnote on page 4 to explain the above more clearly.

Halon-1301 – further evidence of its performance as an age tracer in New Zealand groundwater

Monique Beyer[1]; Uwe Morgenstern[1]; Rob van der Raaij[1] and Heather Martindale[1]

[1] GNS Science, PO Box 30368, Lower Hutt, New Zealand

Correspondence to: Monique Beyer (Monique.Beyer@es.govt.nz)

Summary (500 characters)

The determination of groundwater age can aid characterization of aquifers, providing information on groundwater mixing, flow, volume and recharge rates. Here we assess a recently discovered groundwater age tracer, Halon-1301. Its performance as an age tracer is assessed against 6 other well established, widely used age tracers in 302 groundwater samples. We show Halon-1301 reliably inferred age, thus has potential to become a useful groundwater age tracer where other tracers are compromised, or complementary to other tracers.

Abstract. We recently discovered a new groundwater age tracer, Halon-1301, which can be used to date groundwater recharged after the 1970s. In a previous study, we showed that Halon-1301 reliably inferred groundwater age at the majority of studied groundwater sites. At those sites, ages inferred from Halon-1301 agreed with those inferred from SF_6 and tritium, two reliable widely applied groundwater age tracers. A few samples, however, showed reduced concentrations of Halon-1301, preventing meaningful age interpretation from its concentration. These reduced concentrations were likely a result of degradation or retardation of Halon-1301 in the aquifer. However, we could not provide full evidence for this due to the limited number of analysed groundwater samples (18 in total). In this study, we assess the potential of Halon-1301 as a groundwater age tracer for a larger dataset of groundwater samples under specific groundwater conditions, including highly anoxic young groundwater which can significantly degrade Halon-1301, to gain more information on the magnitude ~~of occurrence~~ and the causes of reduced Halon-1301 concentrations.

In this study, we analysed 302 groundwater samples for Halon-1301, SF_6, tritium and the CFCs CFC-11, CFC-12 and CFC-113. Comparison of age information inferred from the concentrations of these tracers allows assessment of the performance of Halon-1301 compared to other well established and widely used age tracers. The samples are taken from different groundwater environments in New Zealand and include anoxic and oxic waters with mean residence times ranging from < 2 years to over 150 years (tritium-free).

The majority of assessed samples have reduced or elevated concentrations of CFCs, which makes it impossible to infer a reliable age using the CFCs for these samples. Halon-1301, however, reliably infers ages for CFC-contaminated waters. Three other groundwater samples were found to have elevated SF_6 concentrations (contaminated). Again, at these SF_6 contaminated sites, ages inferred from Halon-1301 agree with ages inferred from tritium. A few samples (14 sites) exhibit reduced concentrations of Halon-1301, which result in elevated inferred Halon-1301 ages in comparison to those inferred from SF_6, tritium and/or CFC-113. Assessment of the groundwater environment at these sites gives further insight into the potential causes of Halon-1301 reduction in groundwater.

Overall, Halon-1301 gives age information that matches ages inferred from SF_6 and/or tritium for the majority (97 %) of the assessed groundwater sites. These findings suggest that Halon-1301 is a reasonably reliable groundwater age tracer, and is in particular significantly more reliable than the CFCs, which may have contamination and degradation problems. Halon-1301 thus has potential to become a useful groundwater age tracer where SF_6 and the CFCs are compromised, and where additional independent tracers are needed to constrain complex mixing models.

1 Introduction

Groundwater age or residence time is the time water has resided in the subsurface since recharge. The determination of groundwater age can aid understanding and characterization of groundwater resources, because it can provide information on groundwater mixing, flow and recharge rates, and volumes of groundwater (Maloszewski and Zuber, 1982; Morgenstern et al., 2010; Gusyev et al., 2014; Hrachowitz et al., 2016).

~~Groundwater a~~Age of recently recharged groundwater (<100 years ago) can be inferred from environmental tracers, such as SF_6 and tritium. The currently used age tracers have limited application ranges and reliability (Darling et al., 2012). For example, SF_6 has natural sources (e.g. Bunsenberg and Plummer, 2000, 2008; Stewart and Morgenstern, 2001; Koh et al., 2007), which can interfere with the interpretation of age from its concentration. The commonly used CFCs (CFC-11, -12 and -113) have stagnant input functions over the last 25 years (Bullister, 2011), have anthropogenic point sources (e.g. in industrial and horticultural areas) (e.g. Oster et al., 1996; Stewart and Morgenstern, 2001; Bunsenberg and Plummer, 2008, 2010; Cook et al., 2006) and are known to be degradable in anoxic environments (e.g. Lesage et al., 1990; Bullister and Lee, 1995; Oster et al., 1996; Shapiro et al., 1997). Ambiguous age interpretations can occur from tritium measurements due to similar rates of radioactive decay and decrease in atmospheric concentration, which leads to similar concentrations of tritium in groundwater recharged at different times (Suckow, 2014). This is particularly true for the northern hemisphere, where concentrations in young groundwater are still elevated due to atmospheric H-bomb testing in the 1960s (Taylor et al., 1992; Morgenstern and Taylor, 2009; Morgenstern et al., 2010). To reduce these limitations of ambiguity and input uncertainty, multiple tracers should be applied complementarily. New groundwater age tracers and/or new groundwater dating techniques are needed to supplement the existing ones. New complementary age tracer techniques are also necessary to resolve the multi-parameter age distributions for more complex mixing models (Stewart et al., 2016).

We only recently discovered a new groundwater age tracer, namely Halon-1301, which can be used to date groundwater recharged after the 1970s (depending on the limit of detection and mixing model assumptions) (Beyer et al., 2014). Halon-1301 can be measured simultaneously with SF_6, which adds only a little to the cost of analysis. Measurement of Halon-1301 and SF_6 in the same water sample allows identification of contact with air during sampling (which can lead to contamination of the water sample with the higher modern atmospheric concentration of both SF_6 and Halon-1301), contamination from other sources, or degradation (elevated/reduced concentration of either Halon-1301 or SF_6). Comparison of Halon-1301 and SF_6 inferred ages to tritium inferred ages also allows assessment of processes in the unsaturated zone and confirmation of degradation/contamination of one or both Halon-1301 and SF_6.

We previously assessed Halon-1301's performance as an age tracer against two relatively reliable established tracers, SF_6 and tritium, at 18 sites (Beyer et al., 2015), and found that in the majority of assessed groundwater samples ages inferred from Halon-1301 agreed with those inferred from SF_6 and tritium. None of the samples showed significantly elevated concentration of Halon-1301, despite the presence of CFC contamination from industrial sources. This suggests that Halon-1301 from anthropogenic or geologic sources that could interfere with age interpretations are insignificant in aquifers. In the remaining (anoxic) water samples, reduced concentrations of Halon-1301 were found. We could exclude degassing into headspace created by de-nitrification, production of methane or loss of Halon-1301 when groundwater is brought to the ground surface, since this would have affected all determined gas tracers, to the highest extend the least water soluble SF_6, which we did not find in any of our samples. Based on these findings, we concluded that reduced Halon-1301 concentrations were likely caused by degradation of Halon-1301 under anoxic conditions or sorption of Halon-1301 to aquifer material. Both contamination and degradation can result in misleading age estimates.

A recent study by Bartyzel et al. (2016) compared tritium inferred mean residence times (MRTs), using lumped parameter models (LPMs) with piston flow ages inferred from SF_6, CFC-12 and Halon-1301 at six sites in Poland. They found Halon-1301 ages agreed well with SF_6 ages. None of their samples indicated reduced Halon-1301 concentrations, despite assessing relatively old and anoxic waters. However, we observed reduced Halon-1301 concentrations in old anoxic waters in our previous study. Bartyzel et al (2016) ~~They~~ may not have observed reduced concentrations of Halon-1301 because they assessed only six sites and we previously found reduced concentrations in 29% of the assessed sites. Another explanation is that different groundwater environments were encountered in both their and our previous study.

This study aimed to further assess the performance of Halon-1301 as a groundwater age tracer on a larger dataset and covering a wider range of groundwater conditions than previously assessed. Of particular interest was to confirm the absence of local geologic or anthropogenic sources of Halon-1301 that could cause elevated concentrations in groundwater samples, and to confirm the causes of reduced Halon-1301 concentrations found in anoxic groundwater. We analysed 302 groundwater samples for Halon-1301 and SF_6 simultaneously and subsequently inferred ages from their concentrations. The samples were taken from different groundwater environments in New Zealand. The samples were also analysed and dated with tritium and the CFCs (CFC-11, CFC-12, and CFC-113), with mean residence times (MRTs) ranging from < 2 years to over 150 years (tritium-

free water)[1]. A large proportion (ca. 30 %) of these samples showed reduced or elevated concentrations of CFCs (in particular CFC-11 and CFC-12), which made it impossible to infer a reliable age using the CFCs in these samples. CFC-113 performed considerable better than CFC-11 and -12, with matching ages to that of tritium for 83 % of the sites. Comparison of groundwater ages inferred from Halon-1301 to those inferred from SF_6, tritium and CFC-113, the three most widely applied and reliable age tracers, allowed further assessment of the performance of Halon-130 as an age tracer. In particular, the reliability of Halon-1301 for groundwater dating of CFC-contaminated, CFC-degraded, or SF_6-contaminated waters was examined, to gain further insight into the contamination and degradation potential of Halon-1301.

2 Method

2.1 Sampling and analysis

In this study, we sampled 302 groundwater samples across New Zealand from over 20 aquifers, ranging from highly anoxic to oxic conditions (Figure 1~~Figure 1~~). Not all age tracers were determined at all 302 sites, as summarized in Table 1~~Table 1~~. To prevent sampling of stagnant water, the well was flushed with at least three times its volume or until DO and EC stabilized. Tritium was analysed in 1 L water samples, using electrolytic enrichment and liquid scintillation counting (LSC) detailed in Morgenstern and Taylor (2009). For analysis of the gaseous tracers Halon-1301, SF_6 and the CFCs, groundwater was sampled under rigorous exclusion of air to avoid contamination of the samples with modern air (Oster et al., 1996). For determination of CFC-11, CFC-12 and CFC-113, 125 ml glass bottles with aluminium foil cap liners were used. For determination of Halon-1301 and SF_6, 1 L brown borosilicate bottles were used. The sampling methods are detailed in van der Raaij and Beyer (under review). The gas samples were subsequently purged and analysed on a gas chromatograph with attached electron capture detector (GC/ECD). Simultaneous analysis of Halon-1301 and SF_6 is detailed in Beyer et al. (2014, 2015). The analytical setup for determination of Halon-1301 and SF_6 also allowed the simultaneous determination of CFC-12 (Busenberg and Plummer, 2008; Beyer et al., 2014; Bartyzel et al., 2016). However, an appropriately concentrated standard gas is needed to establish its calibration curve. CFC-12 concentrations were therefore not determined simultaneously with Halon-1301and SF_6 in this study. CFC-12 was analyzed separately, together with CFC-11, CFC-113, Ar and N_2, as described in van der Raaij and Beyer (under review).

The amount of gaseous tracers in all groundwater samples was determined by establishing a calibration curve (least square fit, forced through 0/0) with certified air standard at various pressures. We analysed blank samples (only containing N_2), which indicated 0 signal for SF_6 and Halon-1301. In addition, the statistical difference between the intercept of the calibration curves

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) +Body (Times New Roman)

[1] Water with MRT 150 years still contains 33% of water younger than 67.5 years and therefore contains significant amounts of tritium, which are easily detectable with our method (Morgenstern and Taylor, 2009). However, for such old (nearly tritium-free) waters, the aggregation error is large – the old component of the water could be thousands of years old and therefore the true mean age be significantly older (Stewart et al., 2016).

for SF6 and Halon-1301 (when not forced through 0/0) were not significant (at 99 % confidence). The intercept of the calibration curve was therefore considered insignificantly different from 0, hence the calibration curve was forced through 0/0 to simplify the calibration procedure and to ensure 0 signal is interpreted as a concentration of 0 (fmol/L, e.g.). This procedure follows the suggestion of Helsel and Hirsch (2002) and Caulcutt and Boddy (1983).

The compositions of the certified air standards for analysis of Halon-1301 and SF_6 (supplied by the NOAA in 2014) and for analysis of the CFCs (supplied by the Scripps Institution of Oceanography in 2011) are summarized in Table 2~~Table 2~~. A calibration curve was established every day before measurement commenced, since the performance of the GC/ECD can change from day to day, due to fluctuations in the environment (e.g. temperature) or aging of the material (e.g. column fill). If applicable, the amount of gaseous tracer in the water sample was corrected for headspace and/or excess air (by dissolved Ar and N_2 determination described in Heaton and Vogel (1981). We note that that method is sensitive to excess N_2 produced e.g. by denitrification. We determined the presence of excess N_2 (as a product of denitrification) on the basis of anomalously high inferred recharge temperatures determined from N_2/Ar ratios. Specifically, we identified samples for which the inferred recharge temperature were significantly higher than the mean annual air temperature at that location (using climate data from NIWA (2016)). Assessment of the performance of Halon-1301 as a groundwater age tracer at these sites allows for assessment of degassing into excess N_2 as possible cause of reduced Halon-1301 concentrations.

Formatted: Font: (Default) Times New Roman

TABLE 1 here

TABLE 2 here

FIGURE 1 here

The equivalent atmospheric molar ratio at time of equilibrium (for groundwater samples at recharge) was determined using the solubility relationship or Clarke-Glew-Weiss fit (Warner and Weiss, 1985) given in Eq. (1). The solubility fit parameters for Halon-1301, SF_6, CFC-11, CFC-113 and CFC-12 are summarized in Table 3~~Table 3~~. In contrast to the solubility of the CFCs and SF_6, which have been well studied and directly measured (Bullister et al., 2002; Wilhelm et al., 1977), the solubility parameters of Halon-1301 have only been estimated by Deeds (2008), using the solubility estimation methods of Meylan and Howard (1991) and Meylan et al. (1996). Actual solubility measurements of Halon-1301 are not available in the literature (according to our searches and further backed up by personal communication with Daniel Deeds, 06/03/2015). In our previous study (Beyer et al., 2015), we used modern (equilibrated tap and river) water to estimate solubility and to validate the solubility estimates. In this study, we confirmed our previous estimate by using solubility estimated from four additional modern (river) water samples.

$$\ln K_x = A + B\frac{100}{T} + C\ln\frac{T}{100} \quad (1)$$

with K_x as solubility; estimated as one of Henry's (K_H), Bunsen (K_B) or Ostwald (K_O) coefficient, T as recharge temperature (in K) and A, B, C as solubility fit parameters, given in Table 2. A salinity term can be added to Eq. (1), but this is negligible for most groundwater applications and so is ignored here.

TABLE 3 here

To determine analytical uncertainty, the EURACHEM/CITAC Guide CG4 (Ellison and Williams, 2012) was followed. Analytical uncertainty included the following uncertainties related to:

- The least square regression (calibration curve),
- The standard gas concentration and recharge temperature,
- Repeatability error from relative standard deviation of replicates, and
- Correction for headspace and excess air.

Uncertainty related to solubility is unknown or has never been reported, so it was not considered in this study. Uncertainty of the solubility of Halon-1301 is relatively high, approx. 10 % (Beyer et al., 2015), and therefore would add 10 % to the total analytical uncertainty for the determination of Halon-1301. We believe that Halon-1301's solubility will be determined with sufficient precision and become available in the near future. To enable comparison of Halon-1301's performance as an age tracer compared to other tracers after availability of a sufficiently accurate solubility value, we did not include the currently high uncertainty in its solubility in the following analysis. For the interested reader, the effect of the uncertainty on the age estimate when adding 10 % uncertainty for solubility is shown in Beyer (2015).

2.2 Inferring groundwater ages

To infer the recharge year or residence time of the groundwater, the equivalent concentration of tritium, Halon-1301, SF_6 and the CFCs in the atmosphere at time of recharge (determined as described above) was compared to their historic atmospheric records (illustrated in Figure 2). For tritium, radioactive decay also is applied, with its half-life of 12.32 years. Southern hemisphere atmospheric SF_6, CFC-12, CFC-113 and CFC-11 records are available at the GMD/NOAA (http://www.esrl.noaa.gov/gmd/; Thompson et al., 2004) and CDIAC websites (Miller et al., 2008); data from 1973–1995 have been reconstructed by Maiss and Brenninkmeijer (1998). Southern hemisphere (Cape Grim) atmospheric Halon-1301 concentrations have been summarized and smoothed by Newland et al. (2013). Data from 1969 to 1977 have been reconstructed by Butler et al. (1999). Tritium records for New Zealand are available at Kaitoke, New Zealand. Since seasonal variability of groundwater recharge can affect tritium recharge to groundwater, the tritium recharge is often estimated using recharge weighting techniques (Allison and Hughes, 1978; Stewart and Taylor, 1981; Engesgaard et al., 1996; Knott and Olipio, 2001). Morgenstern et al. (2010) showed that this is less of a problem in New Zealand, because infiltration is relatively constant through the seasons and the summer gap in infiltration occurs at average tritium concentration in rain, so there is little bias. We therefore did not weight the tritium input in this study. However, the tritium input function was scaled according to elevation and altitude (Morgenstern et al., 2010; Stewart and Morgenstern, 2016).

To account for mixing of waters of different age in the aquifer or during sampling, lumped parameter models (LPMs) were used (Maloszewski and Zuber, 1982). The use of LPMs allows inference of an age distribution rather than the mean or apparent age of a groundwater sample. The age distribution is increasingly used as an indicator for quality and contamination risks (e.g. the New Zealand drinking water standard (Ministry of Health, 2008) and the European Water Framework Directive (EU Legislature, 2000)). Since we did not have reliable estimates of the best-fitting LPM, we initially used a range of LPMs to tests the tracers' performance to infer age. Specifically, the exponential piston flow model (EPM), the dispersion model (DM) and the partial exponential model (PEM) were used (Eq. (2) to (4)). However, since the performance of the age tracers was very similar for the different LPMs employed, we only present results with regard to the EPM.

EPM: for $' > MRT(1-\frac{1}{n})$, $f_{EPM} = \frac{n}{MRT} * exp(-n * \frac{t'}{MRT} + n - 1)$; else $f_{EPM} = 0$. (2)

with MRT = the mean residence time; n = the reciprocal of the ratio of exponential in total flow, which we refer to as E/PM = 1/n, the ratio of exponential to total flow in the following (n has been defined as ratio of total to exponential flow after Maloszewski and Zuber, 1982). At E/PM = 0 pure piston flow is obtained, and at E/PM = 1 pure exponential flow is obtained. The EPM matches well tritium time series data and therefore is the most commonly used LPM in New Zealand (Morgenstern and Daughney, 2012).

DM: $f_{DM} = \frac{1}{\mathrm{MRT}} \times \frac{1}{\sqrt{4\pi \mathrm{DP} \frac{t'}{\mathrm{MRT}}}} \times \mathrm{e}^{-\frac{\left(1-\frac{t'}{\mathrm{MRT}}\right)^2}{4\mathrm{DP}\frac{t'}{\mathrm{MRT}}}}$ (3)

The DM conceptualizes one-dimensional advection-dispersion, with DP as the dispersion parameter, which is defined as $DP = \frac{\mathrm{D}}{\mathrm{Vx}}$ with D as the dispersion coefficient, V as velocity and x as outlet position. When DP = 0, piston flow behaviour is obtained.

PEM: for $t' > MRT_{aq} * \ln(m)$, $f_{PEM} = \frac{m}{MRT_{aq}} * exp(\frac{-t'}{MRT_{aq}})$; else $f_{EPM} = 0$. (4)

with $MRT_{aq} = \frac{MRT_s}{\ln(m)+1}$, MRT_s is the MRT of the sample and m is the reciprocal of the ratio of sampled to total volume (P/EM). This version of the PEM conceptualises mixing of water in an aquifer that can be described by the exponential model (EM) with only part of the well being screened/sampled. At n = 1 (for wells screened across the entire aquifer) the EM is obtained.

FIGURE 2 here

To quantify uncertainty in the inferred LPM parameters as a result of uncertainties in the determination of the tracers in groundwater, age modelling was placed into a probabilistic framework illustrated in Figure 3. The framework included the generation of tracer concentrations by random sampling of the model inputs from within their uncertainty. LPMs which generated tracer concentrations within +/- 1 SD of observations were considered as behavioural, i.e. adequately fitting and representative. The remaining LPMs were considered as non-behavioural and were disregarded. Consequently, and in contrast to the commonly inferred single LPM parameter point estimate, age information in this study is determined as behavioural

LPM parameter populations (i.e. clouds of MRTs and mixing parameter pairs that produce tracer concentrations within +/- 1 SD of the observation) illustrated in Figure 4.

For this study, we considered only uncertainty in the determination of the tracers (i.e. analytical uncertainty, determined as described above). This commonly used approach may underestimate the uncertainty in the age interpretation, but gives a first insight into the performance of Halon-1301 as an age tracer compared to other, better established age tracers. For a more comprehensive analysis, all model uncertainties, such as the uncertainty in the tracer's recharge estimate, as well as assessment of the appropriateness model components, need to be included in the uncertainty modelling approach, as demonstrated in Beyer et al. (under review) and Beyer (2015); Green et al. (2014), Massoudieh et al. (2014, 2012) and Timbe et al. (2013).

FIGURE 3 here

2.3 Comparison of tracer performance

Figure 4 illustrates examples of behavioural age interpretations (i.e. population or cloud of LPM parameters that produce tracer concentrations within +/- 1 SD of the observation) for three sites determined with two tracers. To assess whether Halon-1301 gives comparable age estimates to the ones inferred from SF_6, CFC-12, CFC-11 and tritium, we determined if the inferred LPM parameters populations overlapped (i.e. agreed). No overlap of the inferred LPM parameter clouds implies that the two tracers give differing results which cannot be brought into a 1-sigma agreement with any parameter combination. If they were non-overlapping ~~(i.e. different)~~, we determined the shortest distance of inferred LPM parameter populations as a measure of difference. As a measure of the shortest distance, we determined the nearest neighbour and minimum Euclidian distance between two data clouds in Matlab software (Muja and Lowe, 2009). From that, the % difference in MRT and mixing parameter inferred with two tracers (e.g. SF_6 and Halon-1301) was determined as:~~.~~

$$\Delta\varepsilon\%(MRT) = \frac{d_{\min tracer1,tracer2}}{\mathrm{mean}(\mathrm{MRT}_{\mathrm{tracer1}},\mathrm{MRT}_{\mathrm{tracer2}})} * 100 \quad (5).$$

We note that a variety of objective or fitting functions is increasingly used in the hydrological modelling community suggesting that there is no one criterion that should be applied everywhere (e.g. Beven and Binley, 2014). The general applicability of the approach suggested here (using 1 sigma and 10% distance criterion) needs to be assessed further on other datasets.

FIGURE 4 here

We decided not use the widely applied one-dimensional comparison of MRTs inferred from different tracers (i.e. MRT(tracer1) versus MRT(tracer 2) plots), since this type of comparison may result in misleading interpretations of the agreement/disagreement between age information inferred from the different tracers. For example, for site C illustrated in Figure 4, one may conclude that both tracers' inferred MRTs agree. However, when assessing the behavioural MRT and mixing

parameter population in Fig. 4, it is evident that both tracers' age interpretations do ~~in fact~~ not agree (i.e. the LPM parameter clouds do not overlap), although the tracers give similar MRT estimates.

3 Results

3.1 Solubility

The estimated solubility of Halon-1301 using modern equilibrated water samples in this study was comparable to the solubility estimated previously (Beyer et al., 2015) (Figure 5). We therefore considered the use of the previously estimated solubility coefficients as reasonable for estimating equivalent atmospheric mixing ratios from concentrations of Halon-1301 in water (procedure described in method section). To more accurately determine the solubility of Halon-1301 and reduce uncertainty in its determination, further study is needed (also pointed out in Beyer et al. (2015)). Accurate measurement of the solubility of Halon-1301 is beyond the scope of this study, as due to its extremely low solubility, specialised equipment is required.

FIGURE 5 here

3.2 Age interpretation

In the following, we discuss age interpretations when employing the EPM only. We note that although the EPM is the most commonly employed LPM in New Zealand and other places around the world, we could not with confidence exclude that groundwater mixing at the studied sites is better represented by the DM, PEM or other more complex models, because time series age tracer data is lacking for the vast majority of assessed sites. However, very similar conclusions in terms of the tracer's age interpretation and agreement/disagreement of the inferred age information can be drawn when using the DM and PEM.

Halon-1301 data were limited to one measurement at each site (Halon-1301 has only recently been discovered). Time series SF_6, tritium and CFC data, although available for a few sites, were not employed in this study. We decided to use only one observation at each site for each tracer to infer age, allowing an unbiased comparison of the tracers' performance as age tracers. As a result, relatively large uncertainties in inferred age information were obtained, as shown subsequently. To constrain the uncertainty in inferred age information further and assess the value of time series Halon-1301 data, we are aiming to collect and analyse time series Halon-1301 data in New Zealand groundwater for a direct comparison of time series Halon-1301 and other tracer data.

We emphasize that one should not conclude that the assessed age tracers are not useful because the uncertainties in the age interpretations presented in this study appear large. Instead, this study reiterates what is increasingly recognized in the literature – that there may be issues related to uncertainty in the age estimate and that one needs to apply multiple tracers or time series tracer data to better constrain age information (Cook and Herczeg, 2000; Gooddy et al., 2006). Further, this study does not attempt to discuss which tracer has the lowest uncertainty in its age interpretation, as this is a complex matter. The uncertainty

in the tracer's age estimate is dependent on multiple factors; these include the conditions the tracers' input function (this is dependent on the location of the site), groundwater age and mixing (i.e. the age distribution) at the particular site, in addition to sampling conditions and uncertainty in the determination of the tracers in groundwater. This study only compares Halon-1301 ages with other tracer ages.

First we compared inferred CFC ages with those inferred from tritium to identify the CFC that gives the most reliable age estimates. Misleading CFC age estimates are a common problem (Shapiro et al., 1997; Bartyzel et al., 2016; Stewart and Morgenstern, 2001), because CFCs are prone to degradation and contamination. Thereafter, we assessed the performance of Halon-1301 as an age tracer relative to tritium, SF_6 (assuming that SF_6 and tritium give the most reliable age estimates) and with the CFC that has been found 'most reliable' in this study (CFC-113). We use the term inferred 'age' as a synonym for inferred 'age interpretation', referring to the cloud of behavioural EPM parameters (Figure 4).

~~Figure 6 illustrates the performance of the CFCs as age tracers relative to tritium (refer to figures in Appendix A for details).~~ Figure 6~~Figure 7~~ illustrates typical LPM parameter populations that were inferred based on the CFCs and tritium (refer to figures in Appendix A for details). For the majority of assessed groundwater samples, the CFCs gave similar age estimates to tritium. However, 29 % and 38 % of the sites were contaminated or degraded in CFC-12 or CFC-11, respectively, which made it impossible to reliably infer age from the CFCs at these sites ~~(Figure 6)~~. CFC-113 performed considerably better than the other CFCs ~~(Figure 6)~~ with only ~~some (~~5 %~~)~~ of the sites being contaminated with CFC-113 ~~or (~~and 12 %~~)~~ being degraded in CFC-113. ~~and is~~CFC-113 is therefore considered the most reliable age tracer of the three CFCs in this study.

FIGURE 6 here

~~**FIGURE 7 here**~~

Figure ~~8~~ 7 confirms that SF_6 is more reliable than the CFCs. At 94 % of the sites, where both tritium and SF_6 data were available, SF_6 and tritium MRTs matched. Only six sites were contaminated with SF_6. SF_6 concentrations of these samples were at least 15 % higher, but in some cases several hundred % above current-day atmospheric concentrations. For these samples, comparison of SF_6 and Halon-1301 inferred MRTs was not possible. At all except one of these sites, matching Halon-1301 and tritium inferred MRTs were found. At that one SF_6 contaminated site where the Halon-1301 and tritium inferred MRTs did not match, evaluation of tritium data was inconclusive, as it gave an ambiguous age interpretation (suggesting the water could be either very young (<2 years) or older (>50 years)). That site was Halon-1301 free (and also free of CFC-11 and CFC-12), suggesting the water is older than ca. 75 years. Further, CFC-113 concentrations at this site were very low, suggesting that the water is older than 50 years, which does align with CFC-12, -11 and Halon-1301 data suggesting this site may not be degraded in Halon-1301, although we cannot exclude that the CFCs may be degraded too (this is an anoxic site).

FIGURE ~~8~~ 7 here

In the following, we further discuss the performance of Halon-1301 as an age tracer in comparison to SF_6, tritium, and CFC-113. Of particular interest was the magnitude of occurrence of reduced Halon-1301 concentrations and their possible cause. In our previous study, we concluded that the most likely reasons for reduced Halon-1301 concentrations are degradation and sorption of Halon-1301 to aquifer material (Beyer et al., 2015). To further assess the reasons for Halon-1301 reduction in groundwater in this study, we studied the groundwater environment (i.e. redox state and excess N_2) and compared the performance of Halon-1301 with that of the CFCs more closely in samples that indicated presence of reduced concentrations of Halon-1301 compared to tritium and SF_6.

Figure 8~~Figure 9~~ illustrates typical LPM parameter populations that were inferred from SF_6, Halon-1301, CFC-113 and tritium data in this study (figures for all sites are presented in Appendix A). As mentioned previously, we found that the inferred LPM parameter populations were relatively large for most sites, and most tracers, particularly the mixing parameter (E/PM), was difficult to constrain. This is mostly due to the use of only one tracer observation to infer age at each site, which cannot sufficiently constrain the uncertainties in the LPM parameters. To reduce the uncertainty in the inferred LPM parameters further, time series tracer data are needed and/or multiple tracers need to be applied complementarily.

FIGURE ~~9~~ 8 here

Overall, none of the samples indicated significantly elevated Halon-1301 concentrations (i.e. > 10 % reduced Halon-1301 inferred ages, or concentrations above 10 % of modern-day air). This finding is in line with our previous ~~findings and one~~assessment on the limited data set, suggesting~~s~~ the absence of local sources of Halon-1301 that could lead to contamination of Halon-1301 in groundwater. Considering that these sites cover a large fraction of New Zealand's groundwater systems, Halon appears to not be impacted by geologic or anthropogenic local sources in general.

Figure 9~~Figure 10~~ illustrates that 99 % of the sites where both tritium and Halon-1301 have been determined showed matching tritium and Halon-1301 inferred ages (+/-10 %). Since tritium is seen as one of the most reliable age tracers in New Zealand (mentioned previously), this finding is really positive, suggesting Halon-1301 is equally as reliable as tritium at 230 sites, although at 90 sites significant difference in E/PM (>10 %) were found (Figure 9~~Figure 10~~), which can be related to both tracers' input functions. Specifically, as tritium input is a pulse function, it is easier to constrain the mixing model (mixing parameter). Halon-1301, however, has an S-shaped input, making it more difficult to constrain the mixing model (mixing parameter). We note that with SF_6, with its near linearly increasing input from 1985 onwards~~,~~ we expect that constraining of the mixing parameter is even poorer than with Halon-1301, which is assessed subsequently.

FIGURE ~~10~~ 9 here

At 79 % of the sites where comparison between Halon-1301 and SF_6 inferred ages was possible (i.e. where both SF_6 and Halon-1301 were determined and SF_6 was not contaminated), inferred Halon-1301 MRTs agreed within +/-10 % of the SF_6 inferred MRTs (Figure 10~~Figure 11~~). E/PMs inferred from Halon-1301 also agreed with the ones inferred from SF_6 at the majority of sites. Further, Figure 8~~Figure 9~~ (and Appendix A) suggest that in most cases Halon-1301 can constrain the mixing parameter and the MRT better than SF_6, as indicated by the generally 'narrower' LPM parameter ranges inferred with Halon-1301, although the difference was not statistically significant. We note that in any case, time series tracer data are necessary to better constrain the mixing parameter and allow a more conclusive comparison of the tracers' performance as age tracers.

For the remaining 21 % of the samples (where comparison of both SF_6 and Halon-1301 was possible), Halon-1301-inferred ages were elevated compared to the ages inferred from SF_6. In the following, we assess whether Halon-1301 is likely to be reduced at these sites (summary shown in Table 4).

At 39 of these 62 sites where Halon-1301 inferred ~~ages~~ MRTs were elevated compared to SF_6, tritium inferred ages agreed with the Halon-1301 inferred ages (+/-10 %). This suggests that concentrations of Halon-1301 in these samples were not reduced, despite elevated Halon-1301 inferred MRTs compared to those inferred using SF_6. For only one of the remaining sites was the Halon-1301 inferred MRT also elevated compared to the MRT inferred from tritium, suggesting reduced concentrations of Halon-1301 were present in this sample.

FIGURE ~~11~~ 10 here

The remaining 22 sites at which Halon-1301 inferred MRTs were higher than those inferred using SF_6 did not have tritium information. CFC-113 and CFC-12 ~~concentrations~~ inferred MRTs at only two of these sites agreed with Halon-1301 inferred MRTs, suggesting that at these sites Halon-1301 reliably infers groundwater age. At ~~seven~~ 10 of the remaining 21 sites, CFC-113 data were unavailable. At five of these ~~seven~~10 sites, Halon-1301 inferred MRTs agreed with those inferred with CFC-12 (and at one site also with CFC-11). Although ~~all~~ two of these five samples were anoxic and CFCs are also known to degrade under anoxic conditions, the fact that age interpretations inferred from Halon-1301 or CFC-12 data match at these sites suggests that these sites are not likely degraded in Halon-1301 (or CFC-12), but further data is needed to confirm this exclusively.

At the remaining ~~two~~ five of the seven sites, both CFC-12 and CFC-11 data were unavailable (in addition to unavailable tritium and CFC-113 data). As we do not have further data, we cannot exclude that Halon-1301 concentrations are reduced in these samples. Anoxic conditions in two of these samples could suggest degradation of Halon-1301 occurred at these sites, but further data is needed to confirm this supposition.

In summary, this leaves ~~14~~ 10 sites with likely reduced concentrations of Halon-1301 (as per comparison to tritium and CFC-113), of which ~~10~~ seven are anoxic, two have an unknown redox state, and one is oxic. At the one oxic site, it is unlikely that reduced concentrations of Halon-1301 were caused by degradation, suggesting that degradation may not (at least solely) be causing a reduction in Halon-1301 concentrations. Sorption in the aquifer remains a possible reason, also suggested by the relatively high MRT (12 years for Halon-1301 and 25 years for SF_6). However, mixing of anoxic and oxic water at this site

may have occurred, or the redox state may have been otherwise wrongly evaluated. Another possible cause of reduced concentrations at this one oxic site is uncertainties in the solubility of Halon-1301 and uncertainties in the input of Halon-1301 (due to its relatively flat atmospheric concentration over the last 5 years). At the remaining anoxic sites (10 sites or 3 % of the assessed sites), the lack of oxygen strongly suggests degradation of Halon-1301 that can only occur under anoxic conditions.

Discrepancy of recharge temperature to mean ~~recharge~~annual air temperature of the region suggesting excess N_2 was found in 29 sites. Of these 25 showed matching SF6, Halon-1301, CFC-12, CFC-113 and tritium inferred ages suggesting that degassing into headspace created by denitrification did not affect any of the gaseous tracers at these sites. At the remaining four sites, Halon-1301 inferred MRTs were significantly different to CFC-113 and SF_6 inferred MRTs. However, at these four sites Halon-1301 inferred MRTs matched those inferred from CFC-12 (tritium was unavailable at 2 of these sites, at 1 site Halon-1301 inferred MRT and was different to that inferred with tritium, and at the remaining site Halon-1301 inferred MRT matched that inferred from tritium). This suggests that degassing did effect the gaseous tracers, the most the least soluble ones (SF_6 and CFC-113). Halon-1301 (and similarly CFC-12) appear to be less effected by degassing into headspace created by denitrification, production of methane or when groundwater is brought to the ground surface, making it a more reliable age tracer than SF_6 and CFC-113 in these environments.

TABLE 4 here

In summary, our findings suggest that Halon-1301 performed well as an age tracer at the majority of groundwater sites. Although reduced Halon-1301 concentrations were found in a few samples, resulting in misleading Halon-1301 inferred age estimates, overall Halon-1301 performed significantly better than the CFCs, which are prone to degradation and contamination (shown in this and our previous study). Figure ~~12~~ 11 summarizes the performance of the tracers used in this study, highlighting that Halon-1301 performs almost as well as SF_6 and tritium, and with a much higher success rate than the CFCs in this study. In particular, Halon-1301 is significantly more reliable than the CFCs (which were either degraded or contaminated at as many as 30 % of the sites) and in some cases SF_6 (for six contaminated SF_6 samples, Halon-1301 still matched age estimates from tritium).

FIGURE ~~12~~ 11 here

4 Conclusion

In summary, this study presented an extensive assessment of the performance of Halon-1301 in 302 groundwater samples across New Zealand. We showed that Halon-1301 had a high reliability as an age tracer, similar to that of SF_6 and tritium. It performed much better than the CFCs, CFC-11 and -12, which are prone to degradation and contamination. Both degradation

and contamination lead to non-conforming age estimates. For example, despite some groundwater samples showing evidence of contamination from industrial or agricultural sources (inferred by elevated CFC concentrations), no sample showed significantly elevated concentration of Halon-1301, which suggests there were no local anthropogenic or geologic sources of Halon-1301 contamination.

Like any other tracer, the use of Halon-1301 as a groundwater age tracer has its limitations. In this and our previous study, reduced concentrations of Halon-1301 were found. Causes for these are likely degradation, ~~and/or~~ sorption and/or degassing into headspace or excess N_2. Halon-1301 appears to be less affected by degassing ~~into excess N_2~~ than SF_6 and CFC-113 due to ~~their differing~~its higher solubility. Although we provided further evidence for degradation being the main reason for reduced Halon-1301 concentrations, we could not fully determine the reasons for the reduced concentrations. We hope that future studies will explore this matter further. Knowing the cause of reduced Halon-1301 concentrations is important as it can help predict its reliability as an age tracer in different groundwater environments.

Further study is also needed on time series Halon-1301 data to better understand how uncertainty in inferred age information can be constrained with multiple Halon-1301 data compared to other tracers, e.g. tritium and SF_6. In addition, the solubility of Halon-1301 needs to be better estimated to reduce uncertainty in the determination of Halon-1301 in groundwater and inferred age.

Overall, we highly recommend the use of Halon-1301 as an age tracer, in particular its use in combination with SF_6. The simultaneous determination of Halon-1301 with SF_6 (and CFC-12) at no additional cost to sole SF_6 analysis, can reduce both tracer's limitations to ultimately obtain a more reliable inferred age than through the use of a single age tracer.

Acknowledgments

The New Zealand Ministry of Business and Innovation is thanked for funding in line with the Smart Aquifer Characterization (SAC) project.

References

Allison, G. B. and Hughes, M. W.: The use of environmental chloride and tritium to estimate total recharge to an unconfined aquifer, Australian Journal of Soil Research, 16, 181–195, 1978.

Beyer, M., van der Raaij, R., Morgenstern, U. and Jackson, B.: Potential groundwater age tracer found: Halon-1301 (CF_3Br), as previously identified as CFC-13 (CF_3Cl), Water Resources Research, 50, doi:10.1002/2014WR015818, 2014.

Bartyzel J, Rozanski K. Dating of young groundwater using four anthropogenic trace gases (SF6, SF5CF3, CFC-12 and Halon-1301): methodology and first results. Isotopes in environmental and health studies, 52(4-5):393-404, 2016.

Beven, K. and A. Binley (2014) GLUE: 20 years on, Hydrol. Process. 28; 5897–5918, doi: 10.1002/hyp.10082.

Beyer, M., van der Raaij, R, Morgenstern, U. and Jackson, B.: Assessment of Halon-1301 as a groundwater age tracer, Hydrol. Earth Syst. Sci., 19, 2775–2789, doi:10.5194/hess-19-2775-2015, 2015.

Beyer, M., Jackson, B., Daughney, C. and Morgenstern, U.: Quantifying uncertainty in age estimation from environmental tracers, J. Hydrol. (under review).

Beyer, M.: Characterization of Groundwater with Complementary Age Tracers: Ph.D. Thesis, the Victoria University of Wellington, School of Geography, Environment Earth Sciences; available online: http://researcharchive.vuw.ac.nz/handle/10063/5202, 2015.

Bu, X. and Warner M. J.: Solubility of chlorofluorocarbon 113 in water and seawater, Deep-Sea Res. 42(7), 1151–1161, 1995.

Bullister, J. L. and Lee, B. S.: Chlorofluorocarbon-11 removal in anoxic marine waters. Geophysical Research Letters, 22(14), 1893–1896, 1995.

Bullister, J. L., Wisegarver, D. P. and Menzia, F. A.: The solubility of sulfur hexafluroide in water and seawater. Deep-Sea Res. I, 49, 175–188, 2002.

Bullister, J. L. (NOAA/PMEL): Atmospheric CFC-11, CFC-12, CFC-113, CCl_4 and SF_6 Histories (1910-2011), Ocean CO_2, Carbon Dioxide Information Analysis Centre, online resource http://cdiac.ornl.gov/oceans/new_atmCFC.html, accessed 1/10/2012, 2011.

Busenberg, E. and Plummer, L. N.: Dating young groundwater with sulphur hexafluoride: natural and anthropogenic sources of sulfur hexafluoride, Water Resources Research, 36, 3011–3030, 2000.

Busenberg, E. and Plummer, L. N.: Dating groundwater with trifluoromethyl sulfurpentafluoride (SF_5CF_3), sulfurhexafluoride (SF_6), CF_3Cl (CFC-13) & CF_2CL_2 (CFC-12), Water Resources Research, 44, W02431, doi:10.1029/2007WR0061502008, 2008.

Busenberg, E. and Plummer, L. N.: A rapid method for the measurement of sulfur hexafluoride (SF_6), trifluoromethyl sulfur pentafluoride (SF_5CF_3), and Halon 1211 (CF_2ClBr) in hydrologic tracer studies, Geochem.Geophysics Geosystems, 11(11), 2010.

Butler, J.H., Battle, M., Bender, M. L., Montzka, S. A., Clarke, A.D., Saltzman, E. S., Sucher, C. M., Severinghausand, J. P. and Elkins, J. W.: A record of atmospheric halocarbons during the twentieth century from polar firn air, Nature, 399, no. 6738, 749–755, 1999.

Caulcutt, Roland, and Richard Boddy. *Statistics for analytical chemists*. Vol. 80. London: Chapman and Hall, 1983.

Cook, P. and A. Herczeg (2000) Environmental Tracers in Subsurface Hydrology, Kluwer Academic Press, Boston, USA, 479–510.

Cook, P.G. and Solomon, D. K.: Transport of atmospheric trace gases to the water table: Implications for groundwater with chlorofluorocarbons and dating krypton 85. Water Resources Research 31: doi: 10.1029/94WR02232, 1995.

Cook, P.G., Plummer, N.L., Solomon, D.K., Busenberg, E., Han, L.F.: Effects and processes that can modify apparent CFC age. In: Use of chlorofluorocarbons in hydrology. IAEA, Vienna, 31-58, 2006.

Darling W.G., Gooddy D.C., MacDonald A.M., Morris B.L. 2012. The practicalities of using CFCs and SF6 for groundwater dating and tracing. Applied Geochemistry, 27, 1688–1697.

Deeds, D.A.: The Natural Geochemistry of Tetrafluoromethane and Sulfur Hexafluoride: Studies of Ancient Mojave Desert Groundwaters, North Pacific Seawaters and the Summit Emissions of Kilauea Volcano, Ph.D. Dissertation, University of California, San Diego, 2008; available as Scripps Institution of Oceanography Technical Report, UC, San Diego; Permalink: http://www.escholarship.org/uc/item/1hp1f3bd, 2008.

Ellison, S.L.R. and A. Williams (eds.) (2012) EURACHEM/CITAC Guide CG4. Quantifying Uncertainty in Analytical Measurement, 3rd Edition.

Engesgaard, P., Jensen, K. H., Molson, J., Frind, E. O. and Olsen, H. (1996) Large-Scale Dispersion in a Sandy Aquifer: Simulation of Subsurface Transport of Environmental Tritium, Water Resour. Res., 32(11), 3253–3266, DOI: 10.1029/96WR02398, 1996.

EU Legislature: Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for the Community action in the field of water policy, 2000.

Gooddy, D.; Darling, G.; Abesser, C.; Lapworth, D.J. (2006) Using chlorofluorocarbons (CFCs) and sulphur hexafluoride (SF6) to characterise groundwater movement and residence time in a lowland Chalk catchment. J. Hydrol., 330 (1-2). 44-52. 10.1016/j.jhydrol.2006.04.011

Green, C. T.; Zhang, Y.; Jurgens, B. C.; Starn, J. J. and M. K. Landon (2014) Accuracy of travel time distribution (TTD) models as affected by TTD complexity, observation errors, and model and tracer selection, Water Resour. Res. 50, 6191–6213, doi:10.1002/2014WR015625.

Gusyev, M.A., Abrams, D., Toews, M.W., Morgenstern, U., and Stewart, M.K.: A comparison of particle-tracking and solute transport methods for simulation of tritium concentrations and groundwater transit times in river water. Hydrol. Earth Syst. Sci., 18, 3109–3119, doi:10.5194/hess-18-3109-2014, 2014

Heaton, T.H.E. and Vogel, J.C. (1981) “Excess air” in groundwater. Journal of Hydrology, 50, 201–216, 1981.

Helsel, D.R. and R.M. Hirsch (2002) Statistical Methods in Water Resources Techniques of Water Resources Investigations, Book 4, chapter A3. U.S. Geological Survey. 522 p.

Hrachowitz, M., Benettin, P., van Breukelen, B.M., Fovet, O., Howden, N.J.K., Ruiz, L., van der Velde, Y., and Wade, A.J.: Transit times—the link between hydrology and water quality at the catchment scale. WIREs Water, 3, 629–657. doi: 10.1002/wat2.1155, 2016.

Jurgens, B.C., Böhlke, J.K., and Eberts, S.M. (2012) TracerLPM (Version 1): An Excel® workbook for interpreting groundwater age distributions from environmental tracer data: U.S. Geological Survey Techniques and Methods Report 4-F3, 60 p, 2012.

Knott, J. F. and Olimpio, J. C.: Estimation of Recharge Rates to the Sand and Gravel Aquifer Using Environmental Tritium, Nantucket Island, Massachusetts, U.S. Geological Survey, Water-Supply Paper 2297, 2001

Koh, D. C., Plummer, N. L., Busenberg, E., and Kim, Y.: Evidence for terrigenic SF_6 in groundwater from basaltic aquifers, Jeju Island, Korea: Implications for groundwater dating. Journal of Hydrology 339(1-2), 93–104, 2007.

Lesage, S., Jackson, R. E., Priddle, M. W., and Riemann, P. G.: Occurrence and fate of organic solvent residues in anoxic groundwater at the Gloucester landfill, Canada. Environmental Science and Technology24, 559–566, 1990.

Maiss, M. and Brenninkmeijer, C.A.M.: Atmospheric SF_6: Trends, sources and prospects. Environmental Science & Technology, 32, 3077–3086, 1998.

Małoszewski, P. and Zuber, A.: Determining the turnover time of groundwater systems with the aid of environmental tracers: Models and their applicability, J. Hydrol., 57, 207–231, doi:10.1016/0022-1694(82)90147-0, 1982.

Massoudieh, A., Leray, S. and de Dreuzy, J.-R.: Assessment of the value of groundwater age time-series for characterizing complex steady-state flow systems using a Bayesian approach, Applied Geochemistry 50, 240–251, doi:10.1016/j.apgeochem.2013.10.006, 2014.

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

Meylan, W. M. and Howard, P. H.: Bond contribution method for estimating Henry's law constants. Environmental Toxicology and Chemistry 10, 1283–1293, doi: 10.1002/etc.5620101007, 1991.

Meylan, W. M., Howard, P. H. and Boethling, R. S.: Improved method for estimating water solubility from octanol/water partition coefficient. Environmental Toxicology and Chemistry 15, 100–106. doi: 10.1002/etc.5620150205, 1996.

Miller, B. R., R. F. Weiss, P. K. Salameh, T. Tanhua, B. R. Greally, J. Muhle, and P. G. Simmonds (2008) Medusa: A Sample Preconcentration and GC/MS Detector System for in Situ Measurements of Atmospheric Trace Halocarbons, Hydrocarbons, and Sulfur Compounds, Anal. Chem. 80(5), 1536-1545.

Ministry of Health: Drinking-water Standards for New Zealand 2005 (Revised 2008). Ministry of Health, Wellington, New Zealand, 163 p., Available online: http://www.health.govt.nz/system/files/documents/publications/drinking-water-standards-2008_0.pdf (accessed July 2013), 2008.

Morgenstern, U. and Taylor, C.B.: Ultra low-level tritium measurement using electrolytic enrichment and LSC. Isotopes in environmental and health studies 45(2), 96–117, 2009.

Morgenstern, U., Stewart, M.K., and Stenger, R.: Dating of streamwater using tritium in a post nuclear bomb pulse world: continuous variation of mean transit time with streamflow. Hydrology and Earth System Sciences, 14, 2289–2301, 2010.

Morgenstern, U. and Daughney, C.J.: Groundwater age for identification of baseline groundwater quality and impacts of landuse intensification – The National Groundwater Monitoring Programme of New Zealand. J. Hydrol., Vol. 456–457: 79–93, 2012.

Muja, M. and Lowe, D. G.: Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. In VISAPP International Conference on Computer Vision Theory and Applications. 2009. pp. 331–340. Specific information online at: http://au.mathworks.com/ help/vision/ref/pointcloud.findnearestneighbors.html (last accessed 7/1/16), 2009.

Newland, M. J., Reeves, C. E., Oram, D. E., Laube, J. C., Sturges, W. T., Hogan, C., Begley, P., and Fraser, P. J.: Southern hemispheric halon trends and global halon emissions, 1978–2011, Atmos. Chem. Phys., 13, 5551–5565, doi:10.5194/acp-13-5551-2013, 2013.

NIWA. (2016). National Institute of Water and Air. National Climate Database. NIWA Science, New Zealand. Available online at http://cliflo.niwa.co.nz/

Oster, H., Sonntag, C., and Munnich, K.O.: Groundwater age dating with chlorofluorocarbons. Water Resources Research, 32(10), 2989–3001, 1996.

Shapiro, S. D., Schlosser, P., Smethie, W. M., and Stute, M.: The use of H-3 and tritiogenic He-3 to determine CFC degradation and vertical mixing rates in Framvaren Fjord, Norway. Marine Chemistry 59 (1–2), 141–157, 1997.

Stewart, M. K. and Morgenstern, U.: Age and Source of Groundwaters from Isotopic Tracers. In: Groundwaters of New Zealand. Rosen, M.R. and White P.A. (eds.), 161–183, 2001.

Stewart, M.K., and Morgenstern, U.: Importance of tritium-based transit times in hydrological systems. *WIREs Water 3*, 145–154. doi: 10.1002/wat2.1134, 2016.

Stewart, M. K. and Taylor, C. B.: Environmental isotopes in New Zealand hydrology; 1. Introduction. The role of oxygen-18, deuterium, and tritium in hydrology. New Zealand Journal of Science, 24(3/4), 295–311, 1981.

Stewart, M. K., Morgenstern, U., Gusyev, M. A., and Maloszewski, P.: Aggregation effects on tritium-based mean transit times and young water fractions in spatially heterogeneous catchments and groundwater systems, and implications for past and future applications of tritium, Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-532, in review, 2016.

Suckow, A. (2014) The age of groundwater – Definitions, models and why we do not need this term, Applied Geochemistry, Volume 50, 2014, Pages 222-230, ISSN 0883-2927, http://dx.doi.org/10.1016/j.apgeochem.2014.04.016.

Taylor, C.B., Brown, L. J., Cunliffe, J. J., and Davidson, P. W.: Environmental tritium and ^{18}O applied in a hydrological study of the Wairau Plain and its contributing mountain catchments, Marlborough, New Zealand, J. Hydrol., 138, 269–319, 1992.

Thompson, T.M.; Butler, J.H.; Daube, B.C.; Dutton, G.S.; Elkins, J.W.; Hall, B.D.; Hurst, D.F.; King, D.B; Kline, S.; Lafleur, B.G.; Lind, J.; Lovitz, S.; Ondeel, D. J.; Montzka, S.A.; Moore, F.L.; Nance, J.D.; Neu, J.L.; Romashkin, P.A.; Scheffer, A. and W.J . Snible (2004) Halocarbons and other atmospheric trace species, in Climate Monitoring and Diagnostics Laboratory Summary Report 2002–2003, Rep. 27, chap. 5, 174 pp., Natl. Oceanic and Atmos. Admin., Boulder, Colorado.

Timbe, E., Windhorst, D. Crespo, P., Frede, H.-G., Feyen, J. and L. Breuer (2013) Understanding mean transit times in Andean tropical montane cloud forest catchments: combining tracer data, lumped parameter models and uncertainty analysis. Hydrol. Earth Syst. Sci. Discuss. 10, 15871-15914.

van der Raaij, R. and M. Beyer (under review) Use of CFCs and SF_6 as groundwater age tracers in New Zealand, Journal of Hydrology New Zealand.

Warner, M.J. and Weiss, R.F.: Solubilities of chlorofluorocarbons 11 and 12 in water and seawater. Deep-Sea Res. 32, 1485–1497, 1985.

Wilhelm, E., Battino, R. and Wilcock, R. J.: Low pressure solubility of gases in liquid water. Chem. Rev. 77, 219–262, 1977.

TABLES

Formatted: Heading 1

Table 1. Summary of number of sites at which each age tracer has been determined in this study.

Age tracer	# of sites with available data
Halon-1301	302
SF_6	302
tritium	229
CFC-11	297
CFC-12	297
CFC-113	288

Table 2. Concentrations of compounds in calibrated air standard and custom made standard gas in parts per trillion by volume (ppt) SIO-2005 scale. * estimated from atmospheric concentration at time of filling.

Compound	Calibrated air (Scripps) 2011	Custom made standard gas (NOAA)
SF_6	7.53 (±0.81) ppt	10.97 (±0.04) ppt (analysed)
CFC-13	approx. 5.3 ppt (not reported*)	None
Halon 1301	3.27 (±1.55) ppt	29.3 (±0.2) ppt (gravimetric blend)
SF_5CF_3	approx. 0.16ppt (not reported*)	18.6 (±0.2) ppt (gravimetric blend)
CFC-12	530.8 (± 0.06) ppt	511.4 (±2.0) ppt (analysed)
CFC-11	238.43 (±0.06) ppt	None
CFC-113	74.88 (±0.11) ppt	None
others	Other CFCs and Halon gases usually contained in air	Halon-1201 (6.31 ± 0.03)

Table 3. Reported solubility parameters for Halon-1301 and SF_6 and * solubility parameters for Halon-1301 estimated in Beyer et al. (2015) with an uncertainty of 10 %

Compound	Reference	Parameters for Henry solubility coefficient [mol/L/atm]		
		A	B	C
SF_6	Bullister et al., 2002	-96.5975	139.883	37.8193

CFC-11	Warner & Weiss, 1985	-134.1536	203.2156	56.2320
CFC-12		-122.3246	182.5306	50.5898
CFC-113	Bu and Warner, 1995	-134.243	203.898	54.9583
Halon-1301	Deeds, 2008	-92.9683	140.1702	36.3776
	Beyer et al., 2015*	-91.878	139.001	35.478

Table 4. Breakdown of 62 sites that indicated elevated MRTs inferred with Halon-1301 compared to those inferred with SF_6, and assessment of possibility that the differences in SF6 and Halon-1301 inferred MRTs have been caused by degradation of Halon-1301 in groundwater

Agreement of Halon-1301 inferred MRT with those inferred with CFCs and tritium	**Redox state**	**excess N_2**	**# of sites affected**	**Halon-1301 likely degraded?**
Matching tritium and Halon-1301 inferred MRTs	Various		39	No
Tritium and Halon-1301 MRTs do not match	Anoxic	Yes	1	**Yes, also possibly degassing into excess N_2**
Matching CFC-12, CFC-113 and Halon-1301 inferred MRTs, unavailable tritium data	Anoxic/unknown	No	2	No
Matching CFC-12 and Halon-1301 inferred MRTs, unavailable CFC-113 and tritium data	Various	Yes, 1 anoxic site	5	No
CFC-113 and Halon-1301 MRTs do not match, unavailable tritium and CFC-12 data	Oxic	No	4	No, Halon-1301 possibly retarded
	Anoxic	No	1	**Yes**
unavailable CFC and tritium data	Anoxic	Yes, 1	7	**Cannot be excluded**
	Unknown	No	2	**Cannot be excluded**
	Oxic	No	1	**No**

FIGURES

Formatted: Heading 1

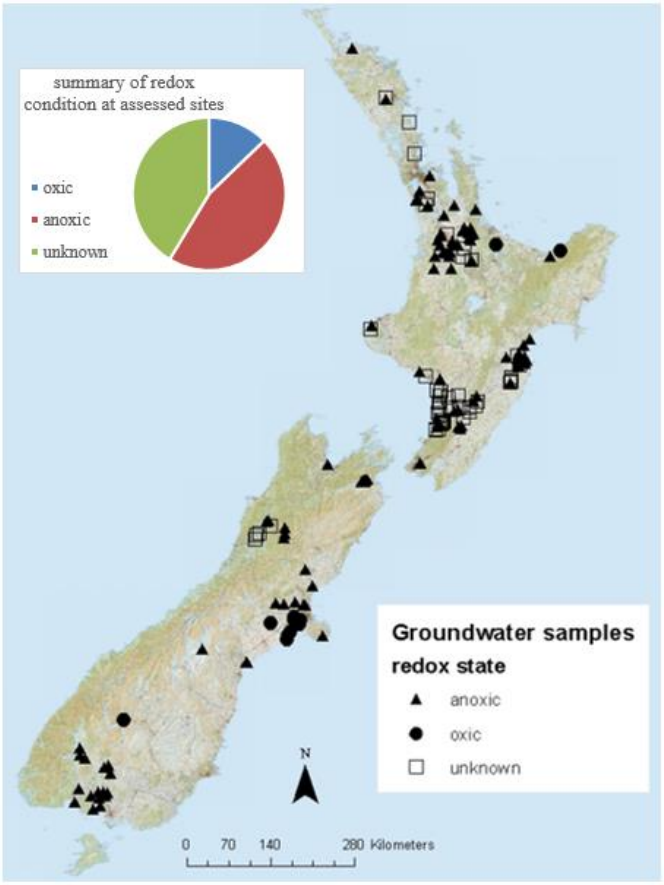


Figure 1: Location of groundwater samples analysed for Halon-1301 in New Zealand. Groundwater was considered as oxic if the concentration of dissolved oxygen exceeded 0.5 mg/L and/or the concentration of dissolved iron and/or manganese was below 0.05 mg/L and methane was not present (and vice versa for anoxic water).

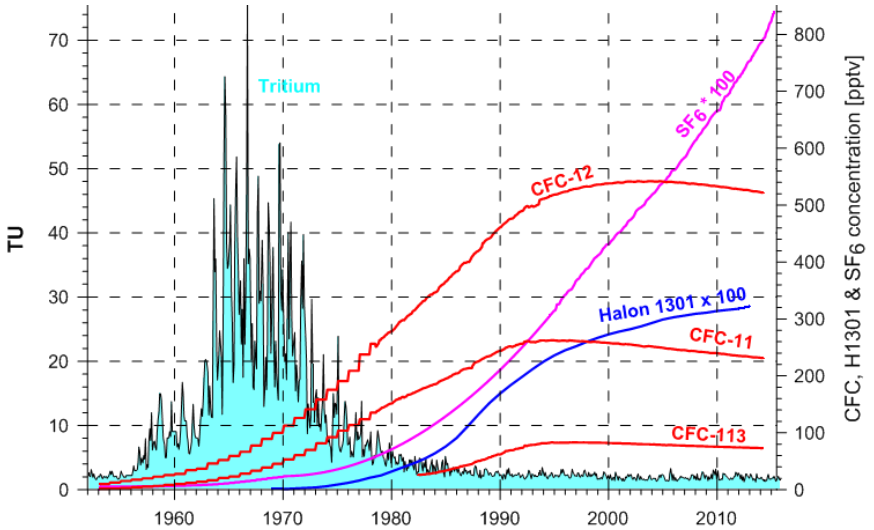


Figure 2: Southern hemisphere atmospheric concentrations of CFC-12, CFC-11, CFC-113, SF_6, Halon-1301 and tritium, using data from NOAA (available at ftp://ftp.cmdl.noaa.gov/hats) for the CFCs and SF_6; Morgenstern and Taylor (2009) for tritium. Concentrations of Halon-1301 in northern hemisphere are very similar to those in the southern hemisphere (see Figure 1 in Beyer et al., 2014 and references therein) suggesting Halon-1301 is well mixed in the atmosphere and can be applied as an age tracer in both hemispheres.

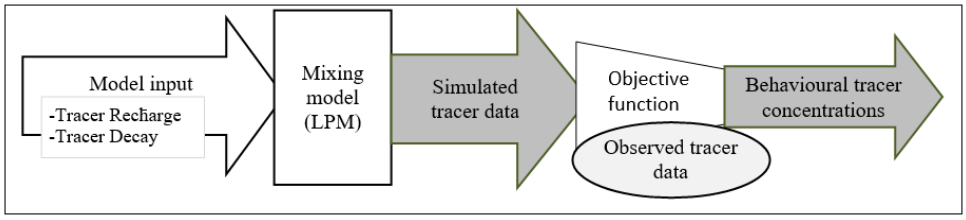


Figure 3: Schema of the modelling approach with framework components: model input, mixing model and objective function

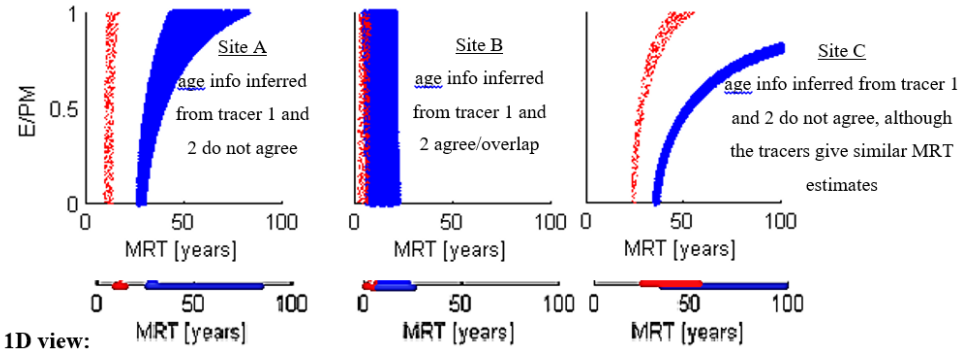


Figure 4: Example of behavioural lumped parameter model (LPM) parameter clouds inferred from two different tracers (tracer 1 in red and tracer 2 in blue). For site B (MIDDLE) the inferred LPM parameter clouds overlap, indicating that the age interpretation inferred from each tracer agree. For site A and C (left and right) the inferred LPM parameter clouds do not overlap, indicating that the two tracers give different age interpretations. For site C the tracers give similar MRT estimates giving the impression that the tracers age info agree when only looking at one dimension (bottom of each figure). To quantify difference/disagreement, the distance between the data clouds can be determined. [it might be useful to have one more example where the age info 1 and 2 overlap only at a certain E/PM ratio – this is a very realistic scenario using two tracers to constrain the E/PM ratio]

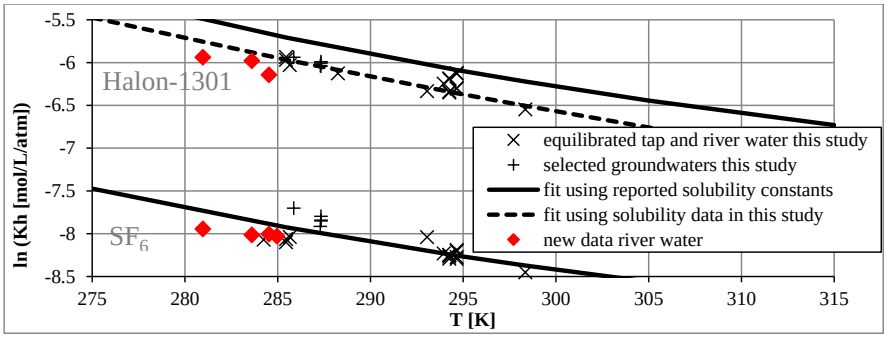
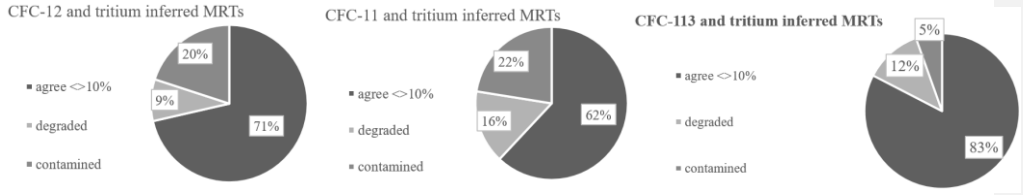


Figure 5: estimated solubility of Halon-1301 and SF_6 in equilibrated tap water, river water, and oxic young groundwater in comparison to reported solubility data, * data from Deeds (2008) for Halon-1301 and Bullister et al. (2011) for SF_6



~~**Figure 6: Agreement of CFC-12 and tritium (right), CFC-11 and tritium (middle), and CFC-113 and tritium inferred MRTs. The data suggest that for most sites CFC-11/ CFC-12 inferred MRTs match tritium inferred MRTs. However, over 30 % of the sites are contaminated or degraded in CFC-11/CFC-12, common causes for erroneous age information with CFC-11/CFC-12. Some (5 %) of the sites are contaminated with CFC-113 or (12 %) degraded in CFC-113. Overall CFC-113 performs considerably better than CFC-11 or CFC-12.**~~

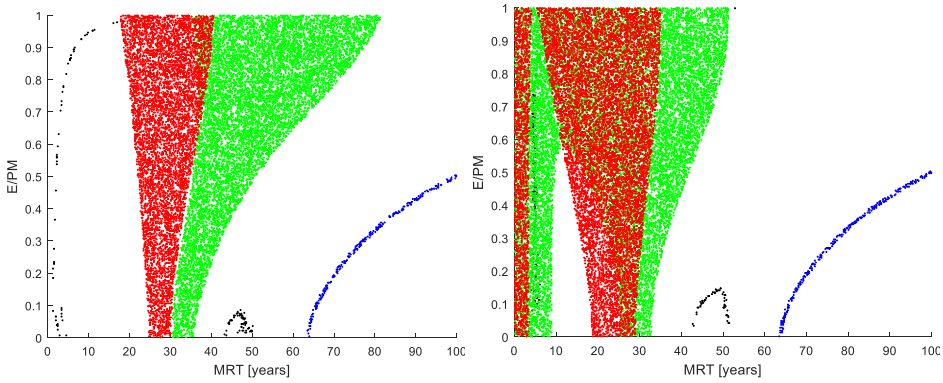


Figure 67: Typical EPM parameter could be inferred with tritium (black), CFC-11 (green), CFC-12 (red) and CFC-113 (blue). In both figures tritium and CFC-11 and CFC-12 inferred LPM parameter clouds overlap, i.e. age interpretations agree; in both figures, tritium gives ambiguous age estimates, in right figure CFC-11 and CFC-12 also give ambiguous age estimates due to recently falling atmospheric concentrations.

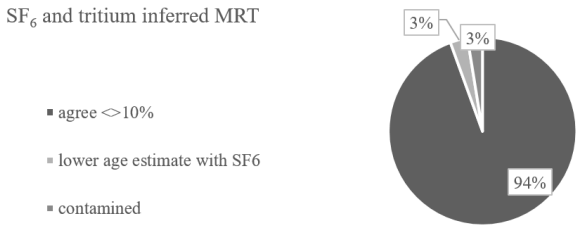


Figure 78: Agreement of SF_6 and tritium inferred MRTs. The data suggest that for most sites SF_6 inferred MRTs match tritium inferred MRTs. Only few sites (3 %) were contaminated with SF_6 and another 3 % showed lower age estimates with SF_6 than with tritium, which could result from a thick unsaturated zone and associated travel time to the aquifer.

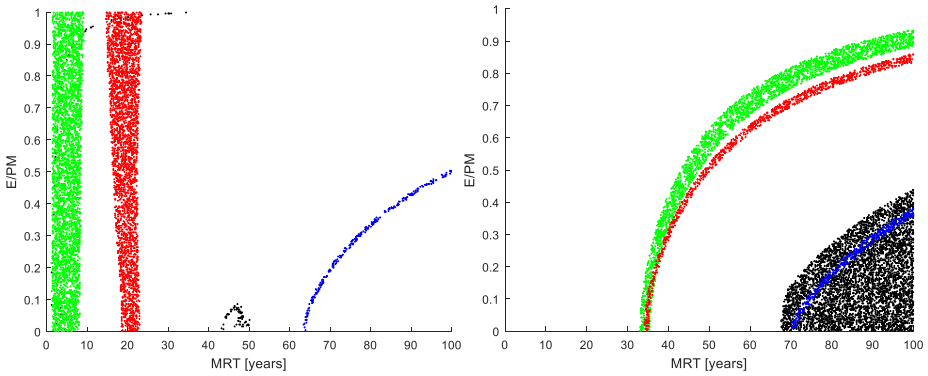


Figure 89: Typical EPM parameter could be inferred with tritium (black), SF_6 (green), Halon-1301 (red) and CFC-113 (blue). In both figures some agreement of age interpretations is evident through overlapping LPM parameter clouds. In the left figure tritium age interpretation is ambiguous; the younger part agrees with the SF_6 and Halon-1301 inferred age. In the right figure the tritium and CFC-113 inferred age agrees, but differs from Halon-1301 and SF_6 inferred ages (which agree with each other).

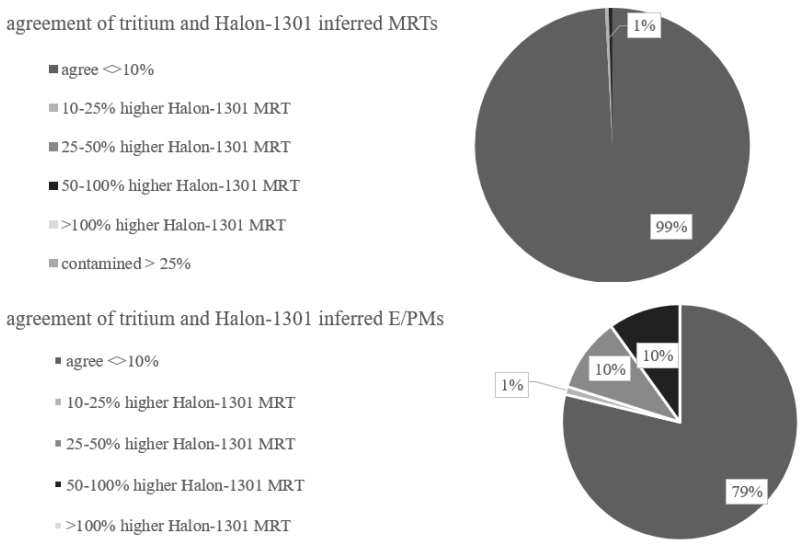


Figure 9~~10~~: Summary of performance of Halon-1301 as an age tracer compared to tritium in predicting the MRT (upper) and the mixing parameter E/PM (lower). Overall, at 99 % of the sites MRTs agree within +/- 10 %.

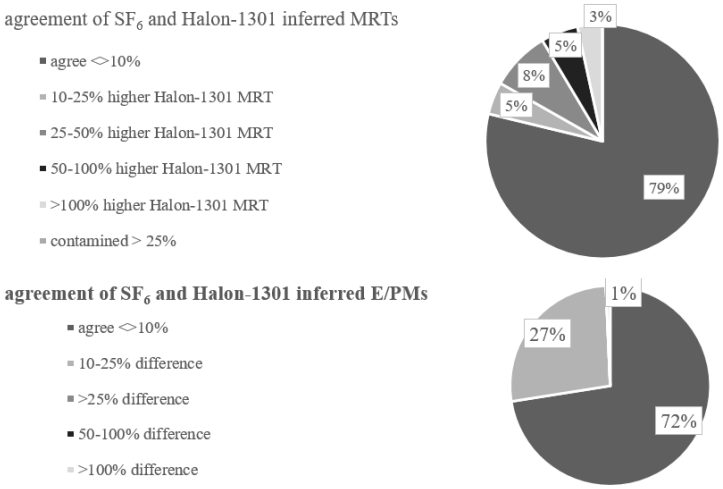


Figure 10~~11~~: Summary of performance of Halon-1301 as an age tracer compared to SF6 in predicting the MRT (upper) and the mixing parameter E/PM (lower). Overall, at 79 % of the sites MRTs agree within +/- 10 %.

5

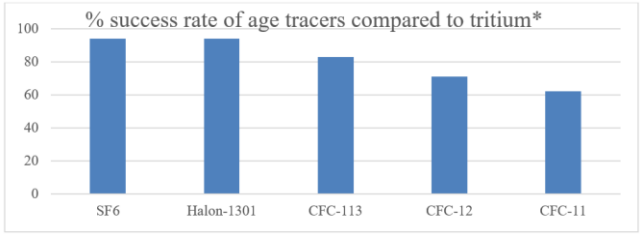


Figure 11~~12~~: Success rate of Halon-1301, SF_6 and the CFC -11 and -12 compared to tritium in this study; *assuming tritium has a success rate of 100 %

Appendix A

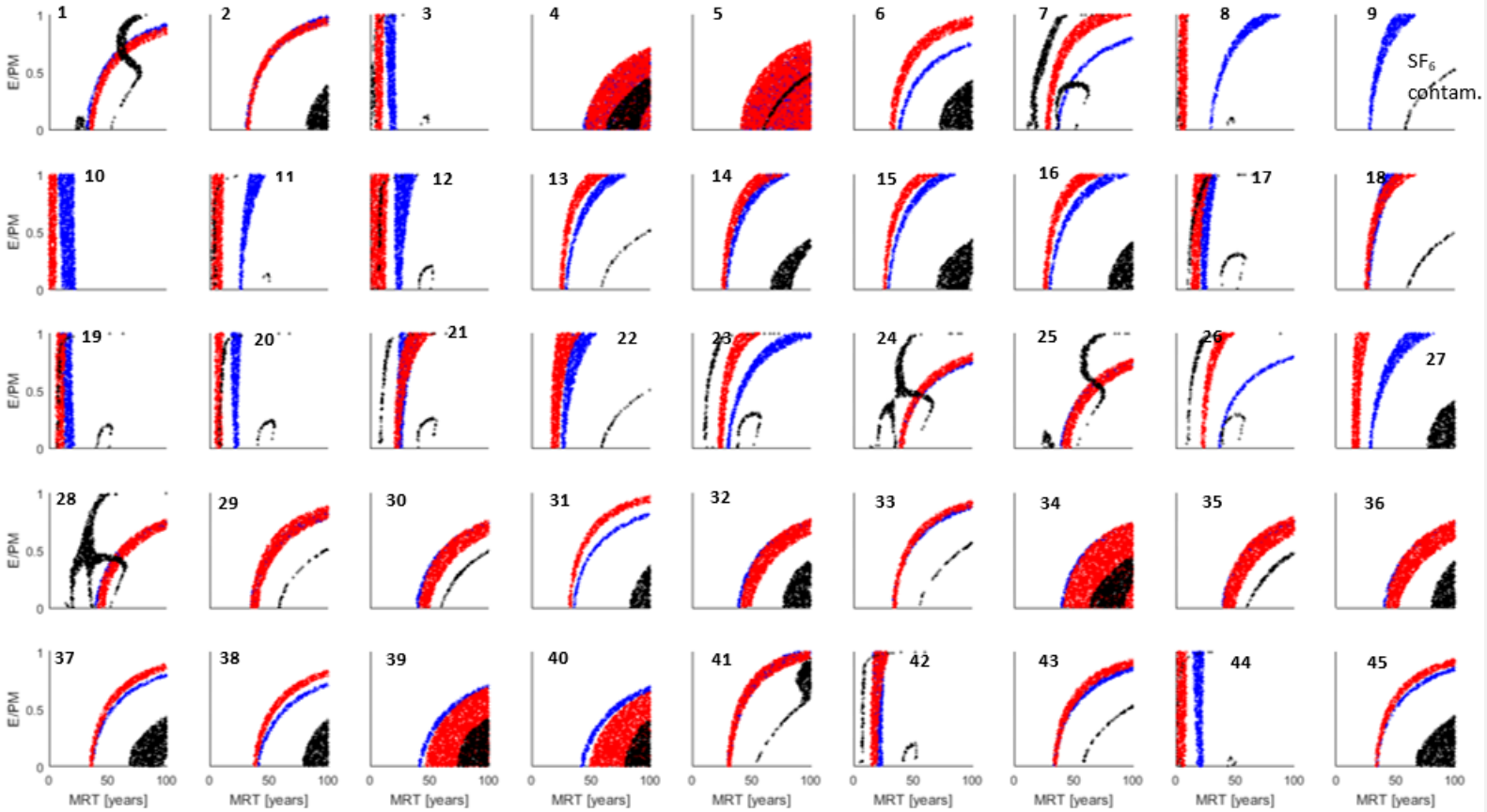


Figure 1213: Behavioural age information (MRT and E/PM) inferred from SF_6 (red), tritium (black) and Halon-1301 (blue) for sites 1 to 45

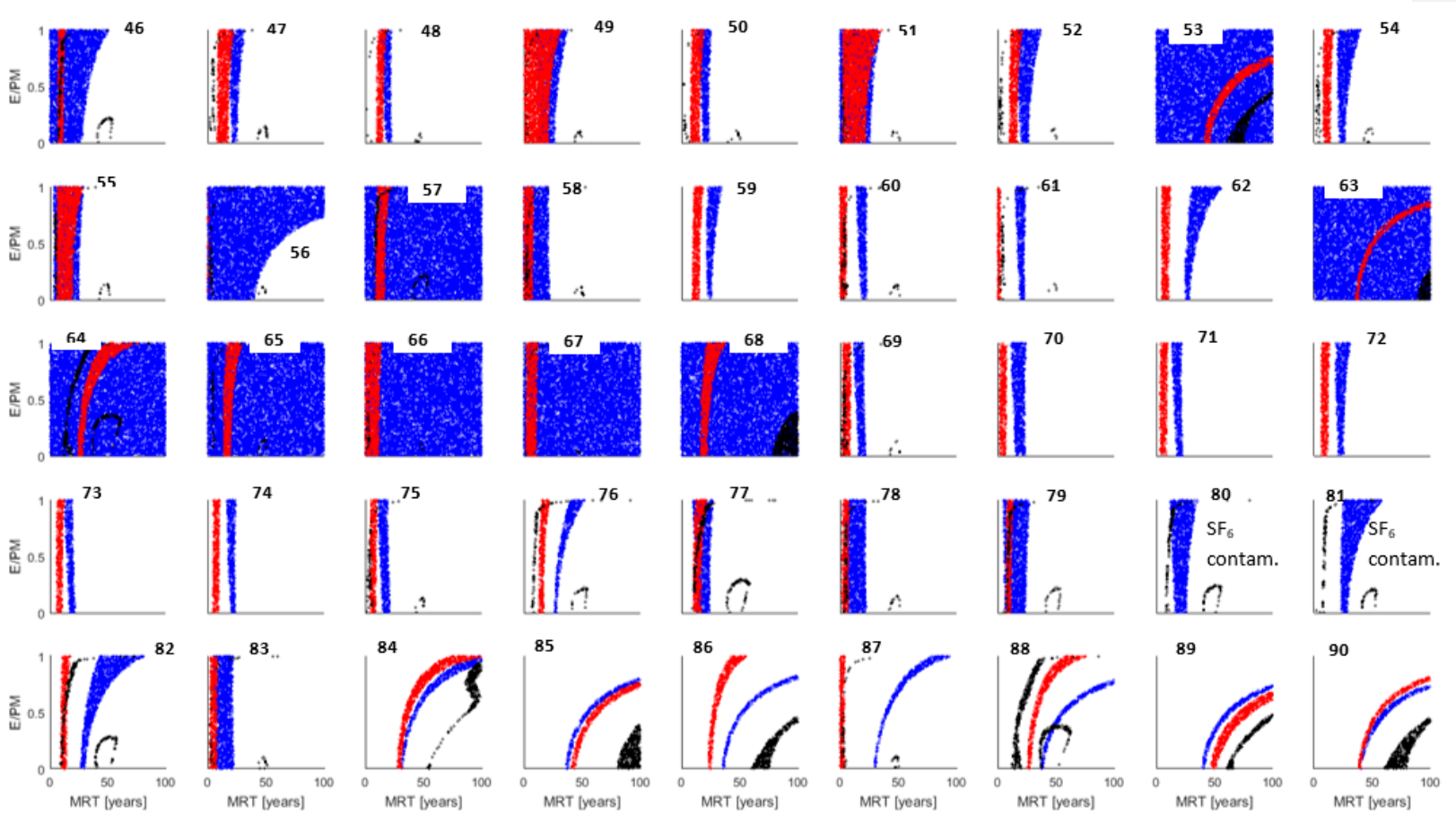


Figure 13~~14~~: Behavioural age information (MRT and E/PM) inferred from SF_6 (red), tritium (black) and Halon-1301 (blue) for sites 46 to 90

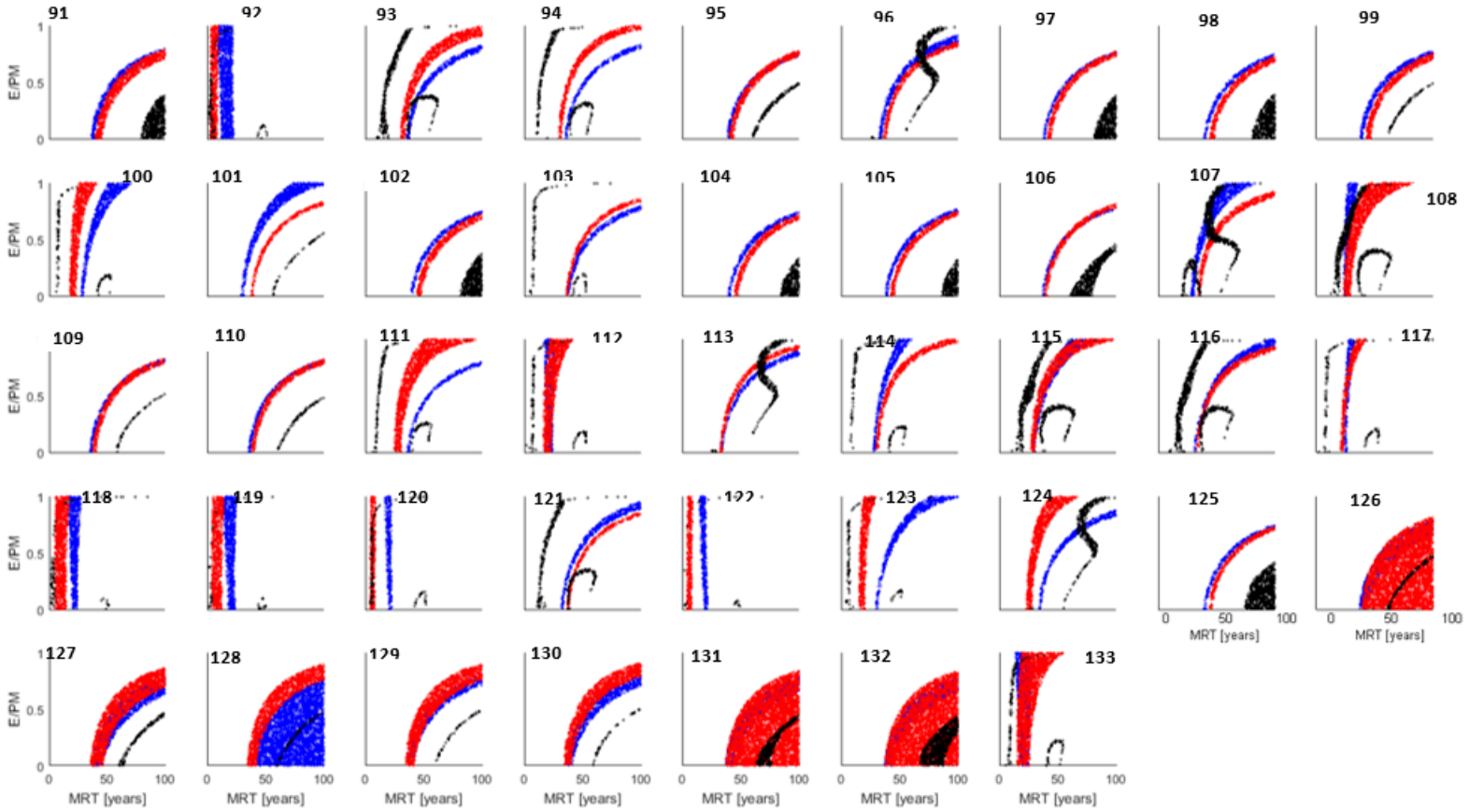


Figure 1415: Behavioural age information (MRT and E/PM) inferred from SF_6 (red), tritium (black) and Halon-1301 (blue) for sites 91 to 133

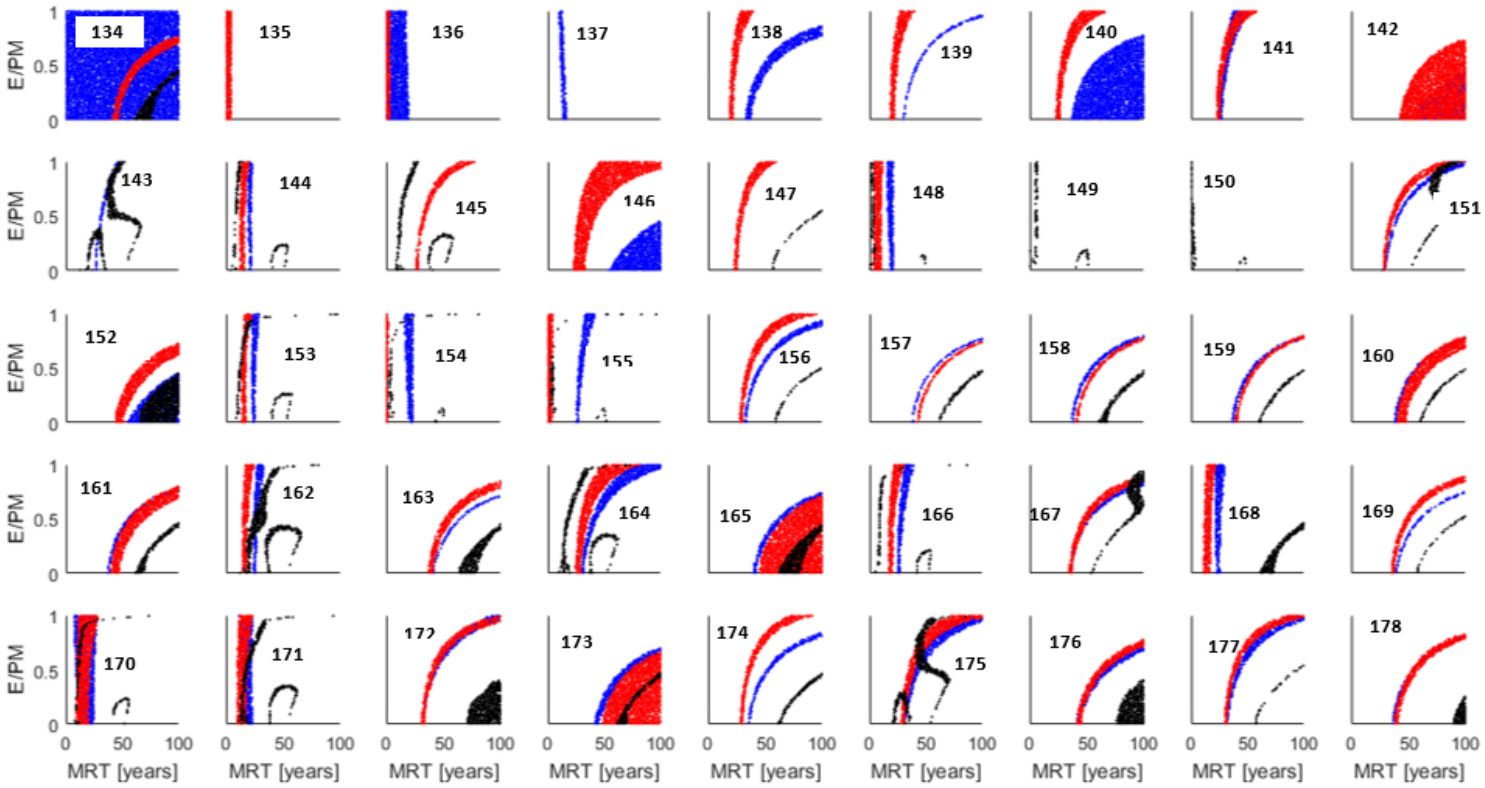


Figure 15~~16~~: Behavioural age information (MRT and E/PM) inferred from SF_6 (red), tritium (black) and Halon-1301 (blue) for sites 134 to 178

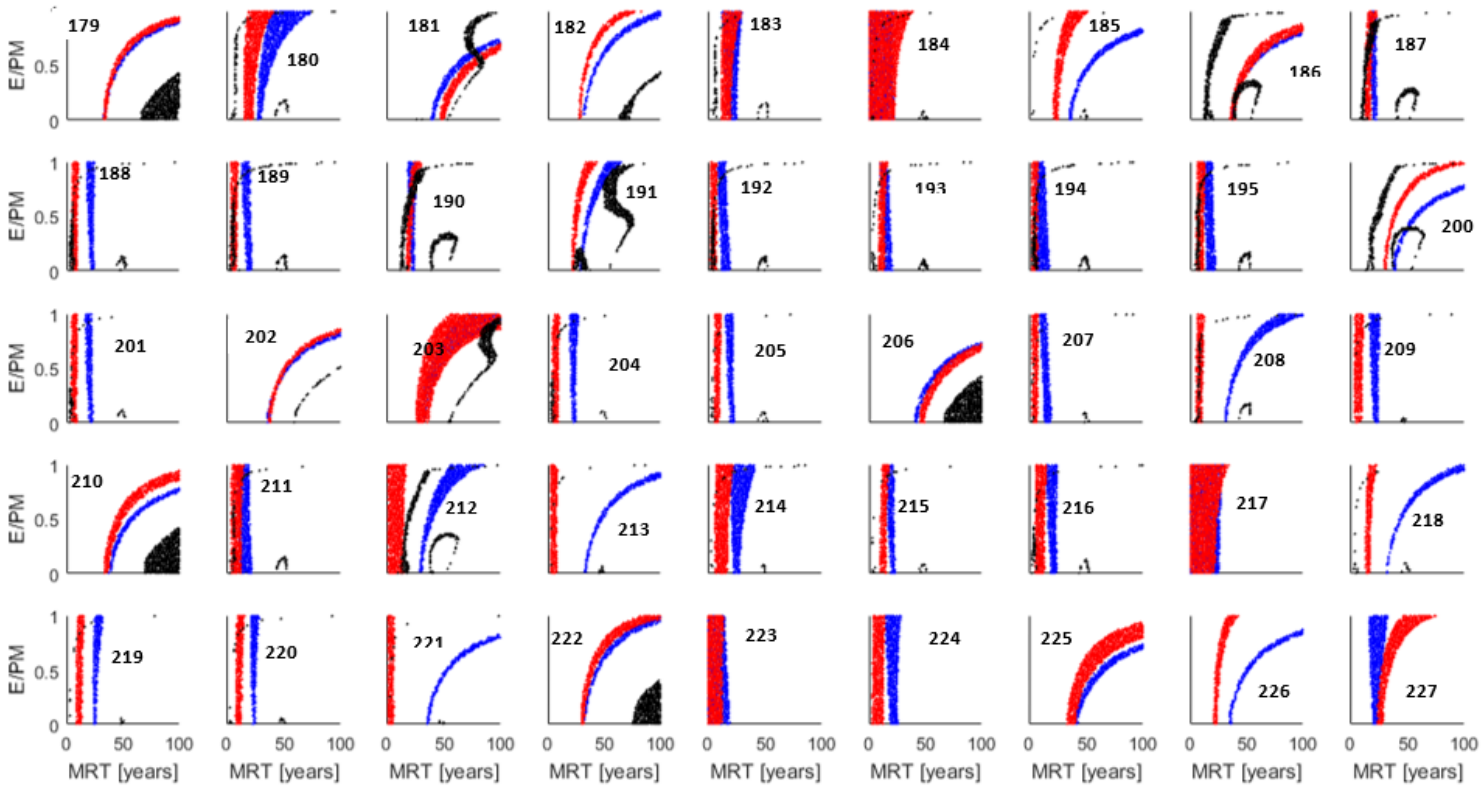


Figure 16~~17~~: Behavioural age information (MRT and E/PM) inferred from SF_6 (red), tritium (black) and Halon-1301 (blue) for sites 179 to 227

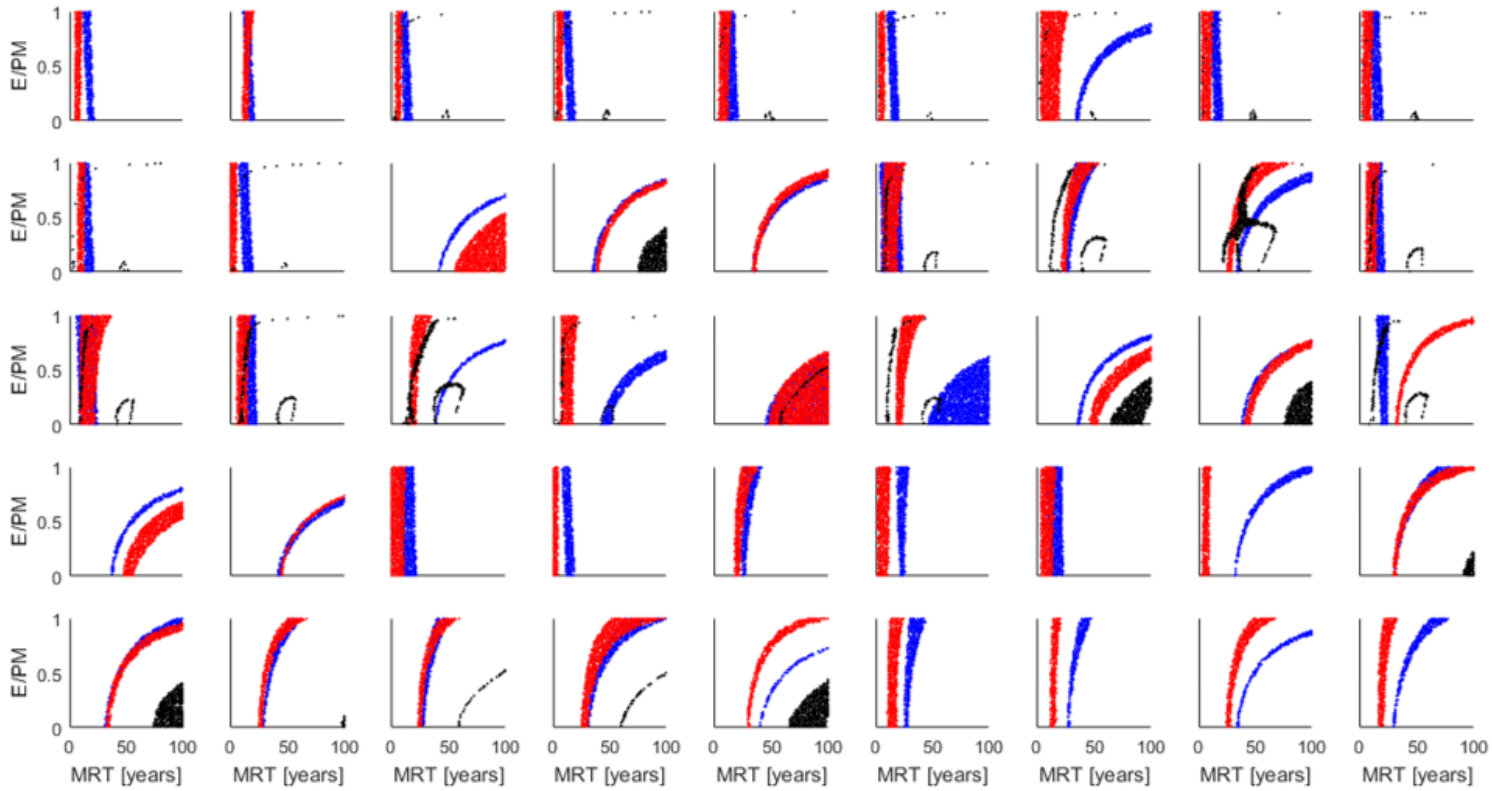


Figure 17~~18~~: Behavioural age information (MRT and E/PM) inferred from SF_6 (red), tritium (black) and Halon-1301 (blue) for sites 228 to 270

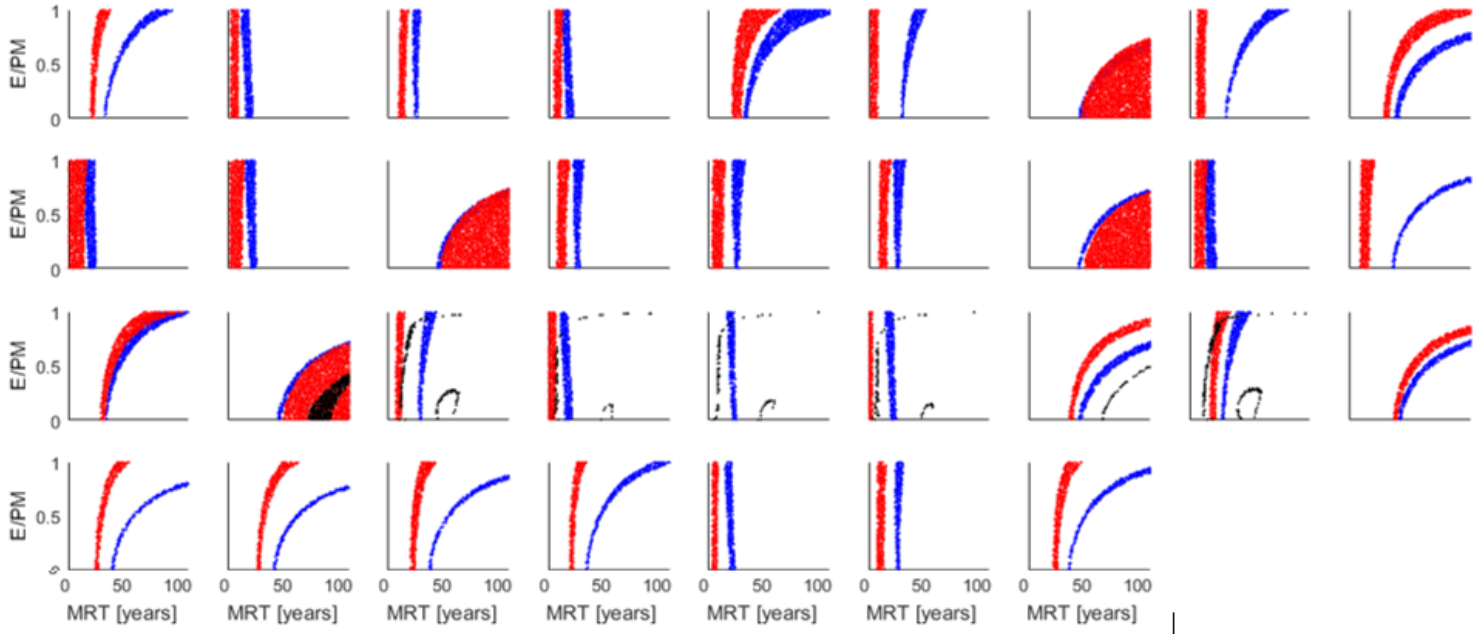


Figure 1819: Behavioural age information (MRT and E/PM) inferred from SF_6 (red), tritium (black) and Halon-1301 (blue) for sites 268 to 302