Intercomparison of tritium and noble gases analyses, 3H/3He ages and derived parameters excess air and recharge temperature

A. Visser, E. Fourré, F. Barbecot, L. Aquilina, T. Labasque, V. Vergnaud, B.K. Esser

Intercomparison of tritium and noble gases analyses, $^{3}H/^{3}He$ ages and derived parameters excess air and recharge temperature

Ate Visser[1*], Elise Fourré[2], Florent Barbecot[3], Luc Aquilina[4], Thierry Labasque[4], Virginie Vergnaud[4], Bradley K. Esser[1] and contributors from participating laboratories.

1- Chemical Sciences Division, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA, T: +1 925 423 0956, E: visser3@llnl.gov, esser1@llnl.gov

2- LSCE, CEA-Orme des Merisiers, F-91191 GIF-SUR-YVETTE CEDEX, France, T: +33 1.69.08.41.72, E: Elise.Fourre@lsce.ipsl.fr

3- GEOTOP, Université du Québec à Montréal, C.P. 8888 Succ. Centre-Ville, Montréal Qc, H3C 3P8, Canada, T: +1 514-987-3000 #7786, E: barbecot.florent@uqam.ca

4- Géosciences Rennes/OSUR – bat 15, Campus de Beaulieu, 263 av du général Leclerc, 35042 RENNES, France, T: +332.23.23.65.89, E: luc.aquilina@univ-rennes1.fr, thierry.labasque@univ-rennes1.fr, virginie.vergnaud@univ-rennes1.fr

* Corresponding author

Abstract

Groundwater age dating with the tritium-helium ($^3H/^3He$) method has become a powerful tool for hydrogeologists. The uncertainty of the apparent $^3H/^3He$ age depends on the analytical precision of the ^3H measurement and the uncertainty of the tritiogenic ^3He component. The goal of this study, as part of the groundwater age-dating interlaboratory comparison exercise, was to quantify the analytical uncertainty of the ^3H and noble gas measurements and to assess whether they meet the requirements for $^3H/^3He$ dating and noble gas paleotemperature reconstruction.

Samples for the groundwater dating intercomparison exercise were collected on 1 February, 2012, from three previously studied wells in the Paris Basin (France). Fourteen laboratories participated in the intercomparison for tritium analyses and ten laboratories participated in the noble gas intercomparison. Not all laboratories analyzed samples from every borehole.

The reproducibility of the tritium measurements was 13.5%. The reproducibility of the $^3He/^4He$ ratio and ^4He, Ne, Ar, Kr and Xe concentrations was 1.4%, 1.8%, 1.5%, 2.2%, 2.9%, and 2.4% respectively.

The uncertainty of the tritium and noble gas measurements results in a typical $^3H/^3He$ age precision of better than 2.5 years in this case. However, the measurement uncertainties for the noble gas concentrations are insufficient to distinguish the appropriate excess air model if the measured helium concentration is not included. While the analytical uncertainty introduces an unavoidable source of uncertainty in the $^3H/^3He$ apparent age estimate, other sources of uncertainty are often much greater and less well defined than the analytical uncertainty.

1 Introduction

Groundwater age dating with the tritium-helium ($^3H/^3He$) method has become a powerful tool for hydrogeologists. The principle of $^3H/^3He$ dating [*Tolstikhin and Kamenski*, 1969] is the decay of radioactive ^3H (half-life $\tau_{½}$= 12.32 a [*Lucas and Unterweger*, 2000]) to helium-3 (^3He) and the accumulation of the decay product ^3He in groundwater. Combined determination of the ^3H and tritiogenic ^3He ($^3He_{trit}$) concentrations in groundwater allows for the calculation of the apparent $^3H/^3He$ age (τ) given the decay constant of ^3H ($\lambda = \ln(2)/\tau_{½} = 0.05626\ a^{-1}$) (Eq. 1).

$$\tau = \ln(1 + [^3He_{trit}] / [^3H])\ \lambda^{-1} \qquad \text{[Eq. 1]}$$

The apparent $^3H/^3He$ age corresponds to the groundwater travel time under the assumption that ^3He is confined in groundwater below the water table and both ^3H and ^3He are transported at the same rate as the groundwater flow. First applications of $^3H/^3He$ dating were published in 1987 [*Takaoka and Mizutani*, 1987] and 1988 [*Poreda et al.*, 1988; *Schlosser et al.*, 1988].

Tritium is naturally produced in the atmosphere by cosmic radiation resulting in a tritium concentration in precipitation of less than 10 tritium units (1 TU corresponds to a $^3H/^1H$ ratio of 10^{-18}). Large quantities of ^3H were released into the stratosphere since 1953 by above ground testing of thermonuclear devices. ^3H enters the groundwater system by infiltration of tritiated water ($^3H^1HO$) in precipitation. The historical concentrations of tritium in precipitation have been monitored at several locations around the world by the Global Network of Isotopes in

Precipitation (GNIP) network and are available online from the International Atomic Energy Agency (IAEA).

Groundwater contains ^{3}He from five sources: equilibration with the atmosphere, excess air from bubble entrainment in the unsaturated zone, nuclear fission of ^{6}Li (nucleogenic: $^{6}Li(n,\alpha)$ $^{3}H \rightarrow {}^{3}He$) associated with the production of ^{4}He by U-Th decay (radiogenic ^{4}He) [*Schlosser et al.*, 1989], mantle helium and tritium decay (tritiogenic). $^{3}H/^{3}He$ dating requires calculating the tritiogenic component by subtracting the atmospheric components (equilibrium and excess air) from the measured ^{3}He concentration (if the nucleogenic and mantle helium components are negligible). If the recharge temperature is known, the excess air is assumed to be unfractionated with respect to the atmosphere and the $^{4}He/Ne$ ratio indicates that no terrigenic helium (radiogenic or mantle helium) is present, the atmospheric component can be derived from the concentration of ^{4}He. If terrigenic helium is present, the atmospheric helium component can be derived from the concentration of neon [*Schlosser et al.*, 1989] assuming a known recharge temperature and unfractionated or no excess air. A more advanced method is to derive the atmospheric ^{3}He component from inverse fitting the concentrations of the other noble gases to excess air fractionation models. The resulting "derived parameters" (noble gas recharge temperature, excess air amount and fractionation) are in itself useful proxies of past recharge conditions [*Aeschbach-Hertig et al.*, 2000; *Stute et al.*, 1995]. Radiogenic helium is also a tracer for groundwater age, typically in the range of 10^{3} to 10^{6} years [*Marine*, 1979].

The uncertainty of the apparent $^{3}H/^{3}He$ age (σ_{τ}) depends on the analytical precision of the ^{3}H measurement (σ_{3H}) and uncertainty of the tritiogenic ^{3}He component ($\sigma_{3Hetrit}$) derived from the propagation of the noble gas measurement uncertainty [*Solomon et al.*, 1993]. $\sigma_{3Hetrit}$ includes the

uncertainty of the helium isotope ratio of a terrigenic component, if present. A linear approximation of the age uncertainty is given by Eq. 2.

$$\sigma_\tau = \lambda^{-1} ([^3H] + [^3He_{trit}])^{-1} (\sigma_{3Hetrit}{}^2 + ([^3He_{trit}]/[^3H])^2 \sigma_{3H}{}^2)^{½} \quad \text{[Eq. 2]}$$

The uncertainty of the tritiogenic ^3He component includes the uncertainty in the determination of the recharge temperature and excess air fractionation [*Ballentine and Hall*, 1999]. The purpose of this study, as part of the groundwater age-dating interlaboratory comparison exercise [*Labasque et al.*, 2014] including also CFC and SF_6 dating techniques (Labasque, this issue), was to quantify the uncertainty related to field-sampling procedures as well as the analytical uncertainty of the ^3H and noble gas measurements, and the resulting uncertainty of the apparent $^3H/^3He$ ages and derived parameters.

For low-level tritium activity measurements in water, the International Atomic Energy Agency (IAEA) regularly organizes interlaboratory comparison exercises ["TRIC", *Gröning et al.*, 2009]. The last exercise, TRIC2008, included the results from 63 participating laboratories anonymously submitting the analysis results of five low level (<15 TU) tritium samples. The five samples are prepared by IAEA by gravimetric dilution of tritiated standard water with water of near-zero tritium concentration.

No such exercise exists for the analysis of noble gases in water samples. Interlaboratory comparison occurs occasionally when two laboratories with the same analytical capabilities participate in the same research project. Developments of new sampling or analytical techniques are often validated against accepted methods, often at the same laboratory [*Roether et al.*, 2013]. This is the first interlaboratory comparison exercise for noble gas analyses in water samples. The

goal of this study was to assess how the demonstrated sampling and analytical uncertainty propagate into the $^{3}H/^{3}He$ groundwater age and noble gas paleotemperature reconstruction.

2 Methods

A detailed presentation of the experiment design, site geological and hydrogeological context and participants is given in [*Labasque et al.*, 2014]. Samples for the intercomparison exercise were collected from existing and previously studied wells. The observed variability is due to both analytical procedures and sample collection. The stability of the well was confirmed by repeat samples and field measurements during the time it took to collect all samples for the intercomparison exercise [*Labasque et al.*, 2014]. We therefore assume the observed variability is representative of the uncertainty that can be expected in real world studies.

2.1 Sample collection and hydrogeological description

Samples for the groundwater dating intercomparison exercise were collected on 1 February, 2012, from three previously studied wells in the Paris Basin (France). One well (Albian) is located in the confined Albian Aquifer; the other two wells (SLP4 and SLP5) are located in the shallower unconfined Fontainebleau Sands Aquifer.

The Paris basin is a 600 km wide and less than 3 km deep sedimentary basin with small dips oriented toward the depocenter. The stratigraphic succession records a geological history from beginning of the Triassic (250 Ma BP) to the present. From a hydrogeologic standpoint, this basin is a multi-layered aquifer-aquitard system associated with a southeast to northwest topographically-driven flow with recharge zones at the highest outcrops and discharge zones in

the rivers or in the Channel. Some of these aquifers are exploited, e.g. the Albian sand formation, a deep protected aquifer, is used for water supply. The Albian sands of the Paris Basin represent one of the shallowest (600 m below ground level (mbgl)) confined aquifers in this thick sedimentary basin. Hydrogeologic investigations have identified an interesting stream line, between the recharge area (Gien-Auxerre, SE) and the middle of the Basin (Paris). This flow line is characterized by a strong depression of hydraulic head beneath Paris, induced by the massive pumping since last century. A hydrological and geochemical study [*Raoult et al.*, 1997] demonstrates that locally the Albian groundwater is variously mixed with water seeping up from the underlying Neocomian aquifer. The Albian borehole is screened between 556 to 592 mbgl and was pumped continuously for one month prior to the sampling.

The unconfined Oligocene Fontainebleau Sands aquifer was selected for this investigation because its hydrogeology is well known. This aquifer has been the subject of previous tracer investigations [*Corcho Alvarado et al.*, 2009; *Corcho Alvarado et al.*, 2007; *Schneider*, 2005]. It is located in the shallower part of the Paris Basin. Constituted by very fine well-sorted silica grains, the Fontainebleau Sands formation has a thickness of 50-70 m, a hydraulic transmissivity of 1×10^{-3} to 5×10^{-3} m^2s^{-1} and a mean total porosity of about 25% [*Mégnien*, 1979; *Ménillet*, 1988; *Mercier*, 1981]. The hydrogeological situation is characterized by spatially extended recharge at rates varying between 100 and 150 mm/yr. Groundwater mean residence times in this aquifer vary between modern and a few hundred years [*Corcho Alvarado et al.*, 2007].

Wells in the Fontainebleau sands aquifer have generally long screened intervals. The boreholes SLP4 and SLP5 were selected for the study for two reasons: a) the use of these wells for water supply is continuous and the drawdown was stabilized weeks before sampling, and b) the age structure of groundwater is relatively well constrained. Flow paths intercepted by the well have

ages that expand from modern to a few hundreds of years, with a mean exponential age of about 100 years. The SLP4 and SLP5 wells have a long screened interval completely contained in the sands formation. The screened intervals are between 40 - 54 mbgl and 45 - 68 mbgl, respectively, and are constructed with a gravel pack and 0.6 m inner diameter stainless steel (CUAU inox) screened casing [*Corcho Alvarado et al.*, 2007].

2.2 Participation

Fourteen laboratories (Table 1) participated anonymously in the intercomparison for tritium analyses. Laboratories are identified with a letter, corresponding to the letters used for the CFC and SF_6 intercomparison exercise [*Labasque et al.*, 2014]. Not all laboratories analyzed a sample from every borehole (Table 2). Fifteen laboratories participated in the noble gas intercomparison. Only laboratories reporting a helium isotope ratio (10) were included in the analysis. All ten laboratories reported results for the SLP4 borehole, nine reported results for the SLP5 borehole and only five received and reported samples from the Albian borehole due to time limitations at the borehole. All ten laboratories reported the helium isotope ratio and the concentrations of helium and neon. Eight of the ten laboratories also reported the argon, krypton and xenon concentrations. A smaller number of laboratories reported stable isotope ratios of neon, argon, krypton and xenon.

Table 1: Contributors and participating laboratories

Table 2: Number of analyses reported by participating laboratories (in parentheses) for each of the boreholes.

2.3 General description of methods of sampling, analysis and calibration for tritium and noble gases

For the intercomparison exercise, the analytical procedures of the different laboratories are not described in detail to ensure anonymity. The following section describes sampling procedures used for the GDAT exercise, and common procedures for the analysis and calibration of tritium and noble gases.

Tritium samples are collected unfiltered without preservative in glass or plastic bottles. Tritium can be measured in two fundamentally different ways: counting the radioactive decay or measurement of the accumulated ^{3}He in a closed high vacuum container [*Clarke et al.*, 1976]. Decay counting can be either liquid scintillation counting or gas proportional counting. Liquid scintillation counting of samples at environmental tritium levels requires enrichment of the ^{3}H by electrolytic methods.

Helium accumulation and subsequent ^{3}He analysis requires complete degassing of a sub-sample to remove all ^{3}He before accumulation of new tritiogenic ^{3}He over a period of typically 20-80 days. Samples are degassed by evacuation of the headspace, after heating or ultrasonic vibrations to drive the dissolved gases including ^{3}He into the headspace. The accumulated ^{3}He is measured on a high resolution sector field mass spectrometer used for the determination of the helium isotope ratio in noble gas samples, possibly after introducing an ultrapure ^{4}He isotope spike [*Palcsu et al.*, 2010].

Standards of tritiated water are commercially available. For calibration of analytical systems measuring tritium at environmental levels, the commercial standards are diluted with tritium free water by gravimetric methods.

Samples for noble gas analysis are typically collected by flushing the water sample through annealed copper tubes, which are then pinched-off at either end to prevent atmospheric contamination. The volume of the water sample inside the copper tube varies between 9 and 45mL. Alternative sampling techniques are diffusion samplers [*Gardner and Solomon*, 2009], glass ampoules [*Roether et al.*, 2013] or a gas extraction system all having the advantage that the dissolved gases are extracted from the water sample in the field. For this comparison, only copper tube samples have been considered, to exclude the additional uncertainty due to different sampling procedures. For copper tubes, the dissolved gases are extracted from the sample in the laboratory using a vacuum manifold. The noble gases are purified from reactive gases by chemical gettering and separated by cryogenic adsorption to stainless steel [*Lott III*, 2001] or activated charcoal using one or more traps. The noble gas isotopes are analyzed by pressure measurement or by a quadrupole or sector field mass spectrometer [*Beyerle et al.*, 2000]. The helium isotope ratio is determined using a high resolution sector field mass spectrometer to separate ^{3}He from isobaric interference of the $^{1}H^{2}H$ and H_3 molecules, typically equipped with a Faraday cup for the ^{4}He ions and an electron multiplier for the ^{3}He ions [*Sültenfuß et al.*, 2009]. No commercial standards for dissolved noble gases exist. Analyses are calibrated against atmosphere containing fixed noble gas abundances and/or air equilibrated water standards. Special devices have been developed to speed the equilibration of noble gases between the water and gas phase. Special attention to contamination is required for creating air equilibrated water standards in laboratory buildings where helium or other noble gases are used for leak-checking or cryogenic purposes.

2.4 Data reduction

All laboratories reported noble gas concentrations as cm^3 STP per mass of sample (g or kg). Sample salinity was below 1 mg/L and not considered. Some laboratories reported concentrations of noble gas isotopes, rather than atomic concentrations of noble gases (e.g. ^{20}Ne and ^{22}Ne instead of Ne). The noble gas concentrations in these samples were calculated from the most abundant isotope, assuming atmospheric isotope ratios (e.g. Eq. 3 and Eq. 4).

$$Ne_{sample} = {}^{20}Ne_{sample} / 0.905 \quad \text{[Eq. 3]}$$

$$Ar_{sample} = {}^{40}Ar_{sample} / 0.996 \quad \text{[Eq. 4]}$$

If a measured noble gas isotope ratio was not reported directly but concentrations of multiple isotopes were reported, the isotope ratio was calculated by dividing the reported isotope concentrations (e.g. Eq. 5 and Eq. 6). If the measured helium isotope ratio was reported in terms of the atmospheric ratio (Ra), the value was multiplied by 1.384×10^{-6}.

$$R_{sample} = {}^{3}He_{sample} / {}^{4}He_{sample} \quad \text{[Eq. 5]}$$

$$^{20}/_{22}Ne_{sample} = {}^{20}Ne_{sample} / {}^{22}Ne_{sample} \quad \text{[Eq. 6]}$$

It is not certain that all labs used the same atmospheric compositions to relate their results to. This is especially critical in case of Ar, where different values for $^{40}Ar/^{36}Ar$ differ by 1%. The scatter between different labs may in part be due to different air normalizations.

Before the statistical analyses were performed, outliers were identified by a non-parametric statistical method and visualized using box-and-whisker plots. Measured values were identified as outliers if the distance from the top or bottom 25^{th} percentile was larger than 1.5 times the

interquartile range. The box in box-whisker plots represents the interquartile range; the median is represented by a bold horizontal line. The whiskers extend to 1.5 times the interquartile range from the 25^{th} and 75^{th} percentile or the extreme value, whichever is closest to the median. Outliers as identified by this method are plotted separately.

2.5 Statistical analyses

The purpose of the statistical analyses was to derive four metrics of the analytical performance of the tritium and noble gas laboratories individually and as a community. These metrics include some inevitable uncertainty due to sample collection. The first two metrics are of reproducibility across all participating laboratories in aggregate. The last two metrics consider the performance of individual participating laboratories. All metrics use a borehole mean (μ_b) as a reference. To avoid bias towards laboratories reporting more replicates, the mean of the replicates from the same borehole analyzed by the same laboratory ($\mu_{b,l}$) was first calculated. The borehole mean (μ_b) was then calculated as the mean of the laboratory replicate means ($\mu_{b,l}$). The borehole mean was not weighted, i.e. the uncertainties in the laboratory means were not taken into account.

The first metric is the laboratory mean reproducibility (σ_{lm}). This metric is a measure of the reproducibility of replicate sample analyses across all laboratories, i.e. the reproducibility across laboratories of the mean of a replicate sample set analyzed by any of the participating laboratories. It was calculated for each borehole as the standard deviation of all laboratory replicate means, divided by the borehole mean (μ_b).

The second metric is the single sample reproducibility (σ_{ss}). This metric is a measure of the reproducibility of an individual sample analysis across all laboratories, i.e. the reproducibility across laboratories of a single sample analysis by any of the participating laboratories. The single

sample reproducibility was calculated as the standard deviation of all reported results (excluding the outliers) from a single borehole, divided by the borehole mean (μ_b).

The third and fourth metric pertain to the performance of the participating laboratories. These are the laboratory bias (β_l) and the laboratory sample reproducibility (σ_{ls}), calculated as the mean and standard deviation of the laboratory residuals (ρ_l). The laboratory residuals were calculated as the differences between the measured values in each sample and the borehole mean values (μ_b), including the samples defined as outliers. These residuals were combined for SLP4 and SLP5. The Albian residuals were not included because the small number of samples reduces the reliability of the borehole mean.

Differences between measured concentrations in replicate samples and the borehole mean value are due to a combination of random measurement error and possibly a systematic laboratory bias (e.g. due to inaccurate calibration). If deviations are due to random measurement error, the laboratory bias (β_l) is expected to be less than the laboratory sample reproducibility (σ_{ls}). A laboratory bias significantly greater than the laboratory sample reproducibility indicates an inaccurate calibration.

2.6 Noble gas dissolution models

The measured noble gases were fitted to three commonly used excess air models (unfractionated air [UA: *Heaton and Vogel*, 1981], closed equilibrium [CE: *Aeschbach-Hertig et al.*, 1999], and partial re-equilibration [PR: *Stute et al.*, 1995]). The recharge elevation was assumed to be 100 meters above mean sea level, based on the reported surface elevation of 150 m and a water table depth of 50 m. The salinity was assumed to be zero. The best fit was obtained by minimizing the uncertainty-weighted squared deviations between modeled and measured concentrations

[*Ballentine and Hall*, 1999], denoted by χ^2 (Eq. 7), using a bound constrained quasi-Newton method [*Byrd et al.*, 1995].

$$\chi^2 = \sum_i \frac{(C_{i,o} - C_{i,m})^2}{\sigma_i^2} \quad \text{(i=Ne, Ar, Kr, Xe)} \qquad \text{[Eq. 7]}$$

$C_{i,m}$ are the modeled noble gas concentrations; $C_{i,o}$ and σ^2_i are the observed noble gas concentrations and their uncertainties. The probability of the obtained fit can be expressed if the degrees of freedom are greater than zero [*Johnson et al.*, 1995].

The excess air models were initially fitted to the noble gas concentrations excluding helium, because the presence of terrigenic helium cannot be reproduced by the excess air models. The excess air models were also fitted to all noble gas concentrations including helium, for boreholes where no terrigenic helium was detected, to derive a better constrained estimate for the fractionation of the excess air component. The fitting results and derived parameters are presented for each separate excess air model for the entire data set.

Terrigenic ^4He and tritiogenic ^3He concentrations were derived by subtracting the modeled atmospheric concentrations (He_m) from the sample concentrations (He_s) (Eq. 8-10).

$$^3He_s = R_s \times {}^4He_s \qquad \text{[Eq. 8]}$$

$$^4He_{rad} = {}^4He_s - {}^4He_m \qquad \text{[Eq. 9]}$$

$$^3He_{trit} = {}^3He_s - {}^3He_m \qquad \text{[Eq. 10]}$$

For nine samples, Ar, Kr and Xe measurements are not available. The helium and neon isotope measurements collected from the production wells do not provide sufficient information to estimate excess air fractionation and these samples were therefore not included in the excess air model fit.

The terrigenic and tritiogenic helium in all samples was also calculated based on the helium and neon measurements alone, assuming a recharge temperature equal to the mean annual air temperature (11.0±0.6 °C). For this model, the terrigenic helium concentration is calculated from the equilibrium neon and helium concentrations at the recharge temperature and elevation (Ne_{eq} = 1.97×10^{-7} cm^3STP/g and $^4He_{eq}$ = 4.57×10^{-8} cm^3STP/g) and the ratio of helium and neon in the atmosphere ($(He/Ne)_{atm}$ = 0.2882) (Eq. 11).

$$^4He_{rad} = He_s - (He/Ne)_{atm} \times (Ne_s - Ne_{eq}) - {}^4He_{eq} \quad \text{[Eq. 11]}$$

For this model, referred to as the helium-model, the tritiogenic helium concentration ($^3He_{trit,He}$) is calculated from the measured sample helium isotope ratio (R_s) and concentration (He_s), the atmospheric helium isotope ratio (R_a = 1.384×10^{-6}), the equilibrium helium concentration ($^4He_{eq}$) and the helium isotope dissolution fractionation factor (α = 0.983) (Eq. 12).

$$^3He_{trit,He} = (R_s - R_a) \times {}^4He_s + (1 - \alpha) \times R_a \times {}^4He_{eq} \quad \text{[Eq. 12]}$$

The uncertainty associated with the tritiogenic helium calculation by the helium-model is derived from the measurement uncertainties of the helium isotope ratio and helium concentration,

assuming these are uncorrelated and the uncertainty of the equilibrium helium concentration is negligible:

$$\sigma_{3Hetrit,He} = (((R_s - R_a) \times \sigma_{4Hes})^2 + (\sigma_{Rs} \times {}^4He_s)^2 + (\sigma_{4Hes} \times \sigma_{Rs})^2)^{½} \quad \text{[Eq. 13]}$$

The helium-model could be extended to include the radiogenic helium component, assuming a $^3He/^4He$ ratio for the radiogenic component.

Apparent tritium-helium ages were calculated from estimated tritiogenic helium concentrations calculated from the helium-model as well as more advanced excess air models. The mean tritium concentration in each borehole was used first to quantify the uncertainty in the age estimate derived from the noble gas and excess air model uncertainties. The tritium concentrations measured in each separate sample were used second to illustrate the uncertainty in the age estimate derived from the tritium and noble gas measurement uncertainties.

2.7 Numerical example of age uncertainty sources for typical Northern Hemisphere groundwater ages

The effect of the tritium and noble gas uncertainty on the $^3H/^3He$ age uncertainty depends non-linearly on the measured values of ^3H and $^3He_{trit}$. In addition, the proportion of tritiogenic helium to total helium depends on the concentration of atmospheric helium in the groundwater sample, and on the historical concentrations of tritium in precipitation which have varied several orders of magnitude. To illustrate these effects, the measurement uncertainties were propagated through the age calculations of the helium model, in a numerical exercise of hypothetical groundwater samples collected under ideal conditions. These conditions assume (1) pure piston flow (i.e. no

mixing of ages by dispersion of well pumping, (2) a known recharge temperature of 10 °C and an unfractionated excess air component of 0, 5 or 10 cm^3STP/kg, (3) tritium concentrations in recharging groundwater as measured by the IAEA in precipitation in Vienna, Austria, (4) no decay of tritium in the unsaturated zone and no loss of tritiogenic helium to the atmosphere, and (5) no terrigenic helium. The decay of the annual mean concentrations of ^3H to $^3He_{trit}$ from the time of recharge to 2012 was calculated, and the $^3He_{trit}$ was added to the atmospheric ^3He component to calculate the resulting helium isotope ratio. The uncertainties of the tritium, helium isotope ratio and ^4He concentration measurements were set to the observed laboratory mean reproducibility. The apparent $^3H/^3He$ ages (Eq. 1) and uncertainties (Eq. 2) were calculated from the tritiogenic helium (Eq. 12) and associated uncertainty (Eq. 13) following the helium-model. To quantify the contribution of the tritiogenic helium uncertainty ($C\sigma_{3Hetrit}$) to the linear approximation of the apparent $^3H/^3He$ age uncertainty (Eq. 2), it was calculated as:

$$C\sigma_{3Hetrit} = \sigma_{3Hetrit}^2 / (\sigma_{3Hetrit}^2 + ([^3He_{trit}]/[^3H])^2 \sigma_{3H}^2) \quad \text{[Eq. 14]}$$

3 Results

3.1 Measured tritium concentrations

Tritium was reported for 17 samples from the Albian borehole. Measured tritium concentrations are near zero in the Albian borehole (Figure 1g, Table 3). Negative reported values for the Albian are expressed in the Figure 1g boxplot. Two measured values were identified as outliers for the Albian samples. The tritium concentration was below the detection limit for nine of the

12 laboratories (14 of the 19 samples analyzed). Tritium was measured above the measurement uncertainty by two laboratories at concentrations between 0.049 and 0.09 TU, with measurement uncertainty between 0.019 and 0.03 TU. These measurements would support a small fraction of modern water in the Albian samples. The reproducibility of the Albian was not calculated because the mean value was close to zero.

Tritium was reported for 22 samples from borehole SLP4 and 20 from borehole SLP5. Measured tritium concentrations vary between 6.3 and 2.6 TU (Figure 1g). These values for boreholes SLP4 and SLP5 are lower than the tritium concentrations of 7.8±0.8 and 4.0±0.8 TU respectively in samples collected in 2001 reported by Corcho Alvarado et al. [2007].

The single sample reproducibility of the tritium measurements was 13.9% and 15.6% for the SLP4 and SLP5 samples. A linear regression of tritium single sample analysis standard deviation against concentration shows that the standard deviation of the tritium measurements can be approximated by 13% of the measured value plus 0.04 TU. Analysis of the tritium data reported in the 2008 IAEA tritium interlaboratory comparison (TRIC2008) yields a similar reproducibility, showing a trend with an offset of 0.3TU and a slope of 10% of the measured value (after excluding the outliers). Three of the 53 laboratories participating in TRIC2008 analyzed tritium by ^{3}He accumulation, three by gas proportional counting and 47 by liquid scintillation counting after enrichment. In contrast, ten of the 14 laboratories reporting tritium to this exercise also report a helium isotope ratio, and it is likely that these laboratories use ^{3}He accumulation for tritium determination.

3.2 Measured noble gas concentrations and isotope ratios

The helium isotope ratio (Figure 1a, Table 3) was measured in 8 samples of the Albian borehole and 23 and 21 samples of the boreholes SLP4 and SLP5. The helium isotope ratio of the Albian borehole ($0.122\pm0.035\times10^{-6}$) indicates the presence of terrigenic helium. The helium isotope ratios of both SLP4 and SLP5 are elevated above the atmospheric ratio indicating the presence of tritiogenic helium. Outliers are present in the data from all three boreholes. The reproducibility of the Albian helium isotope ratio measurement is 2.8%, and higher than that of the SLP4 and SLP5 boreholes (1.2% and 2.0%), possibly due to the low ratio or the smaller number of samples analyzed.

The terrigenic helium is confirmed in the measured concentrations of ^{4}He (Figure 1b, n=10). The mean helium concentrations in SLP4 (n=21) and SLP5 (n=21) are similar ($6.1\text{-}6.4\times10^{-8}$ cm^3STP/g). The helium concentrations measured in borehole SLP4 do not follow a normal distribution (Shapiro-Wilk normality test [*Shapiro and Wilk*, 1965], p-value = 0.024). Besides the two outliers identified in the boxplot, visual inspection showed two more data points from a single laboratory that were substantially lower than the rest of the distribution. These were not manually removed to avoid subjectively biasing the data. The reproducibility of the helium measurements varies from less than 1% for the Albian borehole, to 1.5% and 3% for the other boreholes.

The measured neon concentrations (Figure 1c) are similar ($2.5\text{-}2.8\times10^{-7}$ cm^3STP/g) for the three boreholes as well, and contain a few outliers. The reproducibility of the Albian measurements is 8%. The relatively poor reproducibility is possibly due to an outlier that was not identified as such, due to the small number of samples, or - depending on the instrumental details - because the high concentration of ^{4}He limited the amount of neon to be admitted to the mass

spectrometer. The reproducibility for the other boreholes is comparable to the reproducibility of the helium concentrations.

Argon, krypton and xenon concentrations (Figure 1d,e,f) were measured in 6, 19 and 18 samples from the Albian, SLP4 and SLP5 boreholes. All heavy noble gas concentrations are significantly higher in the Albian borehole than the SLP4 and SLP5 boreholes. The reproducibility varies from 1.2% for the argon measurements from the Albian borehole, to 4.7% for the xenon measurements in the borehole SLP5. No outliers are identified in these data.

Figure 1: Box-whisker plots for noble gas and tritium concentrations and noble gas isotope ratios measured in each sample. Bold horizontal line represents median, box represents 25%-75% quantiles, whiskers extend to 1.5 times the interquartile range or the maximum value (whichever is closer to the median), outlying values are plotted if beyond the whiskers.

A smaller number of laboratories have reported isotope ratios for neon, argon, krypton and xenon. Figure 1 (h-l) shows boxplots for the isotope ratios that were reported by more than one laboratory. Both high and low outliers are reported for the $^{20}Ne/^{22}Ne$ isotope ratio. The isotope ratios of argon (38 and 40 over 36) and xenon (129 and 136 over 132) show small deviations from the atmospheric isotope ratios.

The measured values for the noble gases are close to the values published previously for boreholes SLP4 and SLP5 [*Corcho Alvarado et al.*, 2007]. Surprisingly, the helium isotope ratios for SLP4 and SLP5 have changed between 2001 and 2012. In 2001, SLP4 showed a near atmospheric helium isotope ratio of $1.40\pm0.01 \times 10^{-6}$ and SLP5 showed an elevated ratio of $1.53\pm0.01 \times 10^{-6}$. In 2012, helium isotope ratios for SLP4 and SLP5 are $1.57\pm0.02 \times 10^{-6}$ and $1.41\pm0.03 \times 10^{-6}$.

Table 3: Mean measured values, excluding the identified outliers, relative laboratory mean reproducibility (σ_{lm}) and single sample reproducibility (σ_{ss}) and number of outliers (see section 2.5 for definitions).

3.3 Sources of measurement uncertainty

The standard deviations of all measured noble gas concentrations are systematically higher for SLP5 than for SLP4. Therefore we assume that the reproducibility observed in borehole SLP4 represents the analytical uncertainty, while some of the variation observed in borehole SLP5 may be caused by variations during sampling. Particular attention was paid to pumping a homogeneous fluid during the whole sampling experiment [*Labasque et al.*, 2014]. No significant changes in field parameters were noticed during the sampling on SLP5.

Differences between the single sample reproducibility and the laboratory mean reproducibility are insightful of the origins of the variation. If measurement uncertainty was the result of unbiased white noise, the laboratory mean reproducibility would be smaller than the single sample reproducibility. This is true for Kr and Xe, and to a lesser extent for Ne and Ar. The single sample reproducibility and laboratory mean reproducibility of the helium concentration and isotope ratio measurements are equally large. This implies that measurement uncertainty is the result of the variation in laboratory means, rather than random variations between samples, indicating that improvements would be possible by cross-calibration against a common standard.

To investigate the laboratory bias and reproducibility, Figures 2 and 3 summarize for each of the laboratories the residuals of samples from SLP4 and SLP5 from the respective borehole means.

The samples from the Albian borehole were excluded from the analysis, because the mean values calculated for this borehole are more uncertain due to the small number of analyses. Because not all laboratories provided sufficient measurements to reliably calculate the laboratory sample reproducibility, no quantitative analysis was performed (e.g. z-scores or χ^2 analysis). The horizontal bar represents the laboratory bias, the vertical bar the laboratory reproducibility, both calculated from the limited number of reported analyses. The outliers as defined by the box-whisker plot analyses are included here.

Figure 2: Differences between measured sample tritium concentrations and the borehole mean concentration.

The differences between all the measured sample tritium concentrations and the borehole mean concentrations may illustrate whether the deviations are due to random error or systematic bias. For example, laboratories Q and U appear to provide very precise measurements, close to the borehole means. Laboratory F provides less precise tritium measurements, but the mean of the differences is close to zero. Laboratories W and X provide precise measurements, with a bias from the borehole means.

Figure 3: Differences between measured sample concentrations and the borehole mean concentration, for the helium isotope ratio and noble gas concentrations.

For noble gases (Fig. 3), the differences between the measured concentrations and the mean concentrations of boreholes SLP4 and SLP5 show that some laboratories provide very precise (reproducible) measurements that are significantly biased (e.g. krypton measured by laboratory

F), while other laboratories provide bias free measurements with a poor reproducibility (e.g. neon measured by laboratory J).

3.4 Excess air model fit

Each of the three excess air models was fitted to the measured concentrations of Ne, Ar, Kr and Xe in all samples (Table 4). All samples but one fit to all three excess air models. Two samples lack a Kr measurement and could not be fitted to the CE and PR model. The goodness of a fit to each of the three models was assessed from the sum of error-weighted squared deviations (χ^2) for all samples from all boreholes, given the total number of degrees of freedom, defined by the number of samples (43) and the number of model parameter degrees of freedom (2 for UA, 1 for CE, and 1 for PR). The total sum of χ^2 was higher for the UA model (51.4) than for the CE (37.0) or PR model (38.0). Nevertheless, the unfractionated air model was the most probable model because of its larger number of degrees of freedom. For the UA model, the combined probability of the observed χ^2 is 99.9%. Given the measurement uncertainties, the UA model appears to be capable of accommodating the variations in measured concentrations by varying the parameters of the model.

Table 4: Results from fitting the three excess air models to the noble gas concentrations.

3.5 Noble gas recharge temperature, excess air amount and fractionation

The noble gas recharge temperatures estimated by the three different models are consistent with one another. Groundwater sampled from the Albian borehole has a model recharge temperature of 4.2 to 4.3 ± 0.5 °C, while the groundwater in boreholes SLP4 and SLP5 have a recharge temperature of 9.9 to 10.7 ±1.6 °C, slightly below present day mean annual air temperature of

11.0±0.6 °C [*Corcho Alvarado et al.*, 2007]. These results are consistent with a previously published and more elaborate discussion of the noble gas recharge temperatures of these wells [*Corcho Alvarado et al.*, 2009].

Figure 4: Derived parameters recharge temperature (a), modeled excess air amount expressed as ΔNe (b) and excess air fractionation parameter (c) derived from the excess air models fitted to four noble gases (Ne, Ar, Kr, Xe: horizontal axis label 4) and five noble gases (He, Ne, Ar, Kr, Xe: horizontal axis label 5).

3.6 Differences between excess air model fits to 4 and 5 noble gases

The UA excess air model fitted to all five noble gases consistently estimates a lower recharge temperature than when excluding helium (paired t-test: $\Delta T = -0.3$ °C, p-value < 0.01). The CE model shows the opposite pattern, estimating a higher recharge temperature when including helium into the fit (paired t-test: $\Delta T = 0.4$ °C, $P < 0.01$). No significant difference is found for the PR model. The estimated amount of excess air does not change when fitting the CE and PR model to five noble gases, but is lower in the UA model for 5-gas fits. The excess air fractionation parameters are higher for the CE model (expressing more fractionation), nearly the same for the PR model at borehole SLP4, and higher for the PR model at borehole SLP5. The reproducibility of the excess air fractionation parameters is improved for both models at both boreholes, showing a better constraint of the model parameters by the additional helium measurements. Including helium in the model fit may not always be possible due to non-atmospheric helium. If non-atmospheric helium is not detected fitting to four noble gases (Ne-Xe), including helium is recommended.

3.7 Terrigenic and tritiogenic helium

The terrigenic helium concentrations were calculated as the difference between the measured helium and modeled atmospheric helium concentrations. The Albian borehole contains 2.5×10^{-6} cm^3STP/g terrigenic helium. The terrigenic helium isotope ratio in these samples is $1.0\pm0.14\times10^{-7}$.

The calculated terrigenic helium concentrations for SLP4 and SLP5 show consistently negative values for the UA and CE models fitted to four noble gases. This suggests that the excess air in these boreholes is more fractionated than the CE model can derive from the noble gas concentrations excluding helium. Each of the three excess air models was fitted again to the measured concentrations, this time including helium. The helium residuals of the models including helium in the fit (Figure 5b) show that the PR model is capable of precisely reproducing the measured helium concentrations. Both the UA and CE model show a consistent negative residual, although the CE model residual falls within the single sample measurement uncertainty observed for SLP4 (dashed horizontal lines).

Figure 5: Terrigenic helium observed in boreholes Albian, SLP4 and SLP5, derived from excess air models fitted to Ne, Ar, Kr and Xe (a). Helium residual of the excess air models fitted to He, Ne, Ar, Kr and Xe shows that the PR model is better capable of reproducing the measured helium concentrations (b). Dashed lines show the measurement uncertainty of 0.095×10^{-8} cm^3STP/g for helium.

3.8 Apparent $^3H/^3He$ age and propagated uncertainty

Tritiogenic helium concentrations in SLP4 and SLP5 were calculated from the noble gas fit including helium. The median concentration of tritiogenic helium in SLP4 varies between 1.20

and 1.45×10^{-14} cm^3STP/g, equivalent to 4.9 TU (UA) and 5.8 TU (PR). The median concentration of tritiogenic helium in SLP5 varies between 1.4×10^{-15} and 4.4×10^{-15} cm^3STP/g, equivalent to 0.6 TU (UA) and 1.8 TU (PR). Physically impossible negative values for tritiogenic helium were set to zero. The better fit to the ^4He concentration by the PR model is reflected in the smaller ^4He residuals, and results in systematically higher estimates of the tritiogenic ^3He concentrations.

The helium-model (Eq. 12) results in an estimate of the tritiogenic helium concentration close to that of the CE model (Fig. 6a). The CE model is based on fractionation between noble gases and isotopes due to solubility differences. Fractionation of the helium isotopes is therefore negligible. The PR model is based on fractionation between noble gases and isotopes based on their diffusivities. Because of the difference in ^3He and ^4He diffusivity, corresponding to the square root of the inverse of their masses, the derived fractionation between the noble gases is exaggerated by the diffusive fractionation of the helium isotopes. The result is a lower estimate of the atmospheric ^3He concentration, resulting in a higher concentration of tritiogenic^3He.

The helium model behaves similarly to the CE model fitted to all 5 noble gases, in the sense that it closely captures the concentration of atmospheric ^4He and ^3He, but it considers no additional fractionation between the helium isotopes.

Figure 6: a) Tritiogenic ^3He, derived from each excess air model (UA, CE or PR) and from the measured helium concentration and isotope ratio alone (helium-model, He). b) Apparent $^3H/^3He$ age, calculated from the mean tritium concentration in the borehole. c) Apparent $^3H/^3He$ age, calculated from the measured tritium concentrations and mean tritiogenic helium in each borehole, according to the three excess air models.

3.9 $^{3}H/^{3}He$ ages

The variability between models of the $^{3}H/^{3}He$ groundwater age will depend on the variability of the tritium concentration, model estimates of the excess air and tritiogenic helium concentration, and the relative magnitude of the ^{3}H and $^{3}He_{trit}$ estimates. To evaluate the contributions of tritium and tritiogenic helium, ages were calculated first from well mean tritium concentration and sample specific estimated tritiogenic helium estimates (Fig. 6b) expressing the tritiogenic helium uncertainty, and second from well mean tritiogenic helium concentrations and sample specific measured tritium concentrations (Fig. 6b) expressing the tritium uncertainty.

In borehole SLP4, assuming a mean measured tritium concentration of 6.2 TU, the median apparent $^{3}H/^{3}He$ ages vary between 10.3 years (UA) and 11.7 years (PR), reflecting the differences in tritiogenic helium estimates resulting from corrections for non-tritiogenic ^{3}He between models. For borehole SLP5, assuming the mean measured tritium concentration of 2.4 TU, the variation between the median estimated $^{3}H/^{3}He$ ages is much larger, between 3.8 and 9.8 years. For all models, the variability of the estimated ages is also larger for SLP5 than for SLP4 because the concentrations of ^{3}H and $^{3}He_{trit}$ are lower while the analytical uncertainty is similar.

The uncertainty in the groundwater age estimation that results from uncertainty in tritium concentration can be assessed by calculating $^{3}H/^{3}He$ ages using the borehole mean for tritiogenic helium concentrations and tritium concentrations measurements from all laboratories. For SLP4, the variation in ages for each of the models is similar to the variation between the excess air models. In this case, the contribution of the tritium uncertainty is equivalent to the subjective uncertainty of the excess air model choice.

For the SLP5 borehole, the uncertainty in the age estimate as the result of the tritium measurement uncertainty is smaller than the differences between the three excess air models. Here the choice of excess air model dominates the uncertainty of the age estimate.

3.10 Numerical example of age uncertainty sources for typical Northern Hemisphere groundwater ages

For groundwater with an age of 1 year and an initial ^{3}H concentration of 11.3 TU, the measurement uncertainties for tritium (based on the linear regression model σ_{3H} = 0.04 TU + 13% × [^{3}H], section 3.1) and noble gases in groundwater (1.4% and 1.8% for the helium isotope ratio and concentration respectively, Figure 7) result in uncertainties of 1.4 TU for ^{3}H and the equivalent of 0.4 TU for $^3He_{trit}$. In this case, the resulting uncertainty of a ^{3}H/^{3}He groundwater age is 0.5 years under the ideal circumstance of no excess air. Groundwater age uncertainty in this scenario is dominated by the uncertainty of tritiogenic ^{3}He, which accounts for 94% of the total uncertainty, because the contribution of the ^{3}H uncertainty is weighted by the ^{3}He/^{3}H ratio (Eq. 14). The age uncertainty increases to 1.1 years if unfractionated excess air equivalent to a ΔNe of 90% is present. In older groundwater (up to 40 years) the age uncertainty increases to 2.1 years. The contribution of the tritiogenic ^{3}He uncertainty to the age uncertainty decreases to 3% and the contribution of excess air uncertainty becomes correspondingly less significant.

Figure 7: Uncertainty of groundwater ^{3}H/^{3}He age estimates resulting from the uncertainty of the tritium and noble gas measurements (solid lines) and the contribution of the uncertainty of the tritiogenic helium estimate as a percentage of the total uncertainty (dashed lines) in groundwater samples with excess air amounts equivalent to a ΔNe of 0%, 45% and 90%.

3.11 Effect of bias on derived parameters recharge temperature and excess air

Noble gas concentrations were calculated for an imaginary sample with a recharge temperature of 10°C and 1.0×10^{-3} cm^3STP/g of unfractionated excess air (equivalent to ΔNe = +9%). The mean differences between the laboratory concentrations for SLP4 and SLP5 and the borehole mean concentrations were added to the synthetic noble gas concentrations. Then these disturbed noble gas concentrations were fitted to the unfractionated excess air model. The differences between the noble gas recharge temperature and excess air amount estimated by each laboratory and the predetermined conditions (10°C, ΔNe = 9%) are presented in Table 5. These differences can be compared to estimated model uncertainties for recharge temperature and excess air that are propagated using the error-weighted, nonlinear inverse technique of Ballentine and Hall [1999]. The propagated uncertainty of the recharge temperature estimate (0.7°C) is smaller but similar to the differences obtained from the model fit (-2.5°C to 1.3°C, 12 of 15 laboratories producing estimates within ±0.7°C of the predetermined value). The propagated uncertainty of the excess air estimate (1.5%) can be compared to excess air amount differences ranging from -8.3% to 3.8% (9 of 15 laboratories producing estimates within ±1.5% of the predetermined value).

Table 5: Differences between estimated noble gas temperature and excess air amount in synthetic sample with added residuals with a recharge temperature of 10°C and ΔNe = 9%.

4 Discussion

4.1 Measurement uncertainties

The uncertainty of a tritium measurement is less than 15% for the two samples containing 2.6 and 6.4 TU of tritium. Some laboratories provide measurements that appear to be biased from the borehole means of SLP4 and SLP5, pointing towards opportunities for improvement by recalibrating. Validating the results against a new tritium standard and participating in the IAEA TRIC exercises is recommended.

The single sample reproducibility was 1.3% for the helium isotope ratio, 1.5% for the helium and neon concentrations, 2.5% for argon, 3.1% for krypton and 3.5% for xenon measurements.

As with the tritium measurements, some laboratories appear to provide precise but biased measurements, based on the few replicates available. For these laboratories it is recommended to check whether these differences with the borehole mean are significant, with respect to their reported uncertainty. (Laboratory uncertainty was not systematically reported for the first GDAT exercise.) Calibration against air equilibrated water standards and cross-validation against other laboratories is recommended. Improving the reproducibility of replicates is more challenging.

4.2 Uncertainty of apparent $^3H/^3He$ calculations, excess air fractionation and noble gas recharge temperature

The propagated uncertainty of the tritium and noble gas measurements meets the desired precision (1.5 years at 1 year to 2 years at 40 years) for typical $^3H/^3He$ dating applications. However, the measurement uncertainties for the noble gas concentrations were insufficient to distinguish the appropriate excess air model, if the measured helium concentration is not included. As a result of the larger measurement uncertainty, the UA model appears to fit the noble gas concentration data better than the other models due to fewer parameters. In this study, these two models produced significant helium concentration residuals and negative tritiogenic

^{3}He concentrations, indicating that even if a model fits the Ne/Ar/Kr/Xe data well, it may not be well suited for determination of He excess, even though the results are not obviously unreliable as it is the case here.

If no terrigenic helium is detected when the excess air models are fit to the other noble gases, including helium in the excess air fit will improve the estimate of the tritiogenic ^{3}He component. The helium-model captures noble gas fractionation due to solubility differences as described by the closed equilibrium model but not the diffusive fractionation as described by the partial re-equilibration model.

The difference in noble gas recharge temperatures between the three models (0.6 °C) fitted to one set of noble gases (excluding helium) is less than the reproducibility in noble gas recharge temperatures (± 1.6 °C) by each of the models separately. The difference we observed between the noble gas temperatures derived from the CE and PR models (0.1 °C) is smaller than that observed in the noble gas data set from Brazil (~1 °C) [*Aeschbach-Hertig et al.*, 2000; *Stute et al.*, 1995]. Whether this is sufficient for paleoclimate reconstruction depends on the climate signal under study. Given the measurement uncertainty of the noble gases demonstrated here, the choice of an excess air model is not critical to derive noble gas recharge temperatures, but it is crucial for calculating $^3He_{trit}$.

5 Conclusions

This first intercomparison exercise for ^{3}H/^{3}He groundwater ages investigated the measurement uncertainty for the ^{3}H/^{3}He age dating community as a whole thanks to the participation of a large

number of laboratories. Performing the exercise has not only provided an independent assessment of laboratory performance and a demonstration that such assessments are needed and are useful, but also with invaluable experience in organization and preparation for the joint sampling event.

A future $^{3}H/^{3}He$ groundwater dating intercomparison exercise can be improved by demanding a fixed number of analyses from each laboratory and requesting the reported analytical uncertainty. These will allow for a more reliable statistical evaluation of the data and an evaluation of the reported laboratory precision.

It should be noted that the investigated measurement uncertainty includes both sampling and analytical uncertainty, and it is therefore assumed to be representative of the variation expected in real world studies. The purely analytical uncertainty, as investigated in the IAEA TRIC exercises for example, does not include the sampling artifacts. Sampling artifacts are expected to be minimal for tritium, but may be significant for noble gas samples. Besides natural samples with unknown true values, analyses of synthetic noble gas samples with known concentrations would allow the independent assessment of laboratory bias.

The analytical uncertainty introduces an unavoidable source of uncertainty in the $^{3}H/^{3}He$ apparent age estimate. It is not the limiting factor in groundwater age dating. Other sources of uncertainty are often much greater and at the same time less well defined than the analytical uncertainty. For example, the interpretation of apparent $^{3}H/^{3}He$ groundwater ages in a hydrogeological context (beyond piston-flow) is complicated by a lack of knowledge of mixing and dispersion in the aquifer and during sampling. Shifting the focus of future intercomparison exercises from the analytical capabilities to the hydrogeological interpretation of multiple age

tracers will benefit the age dating community in assessing the uncertainty of the groundwater age distribution derived from age tracers in a specific context.

Acknowledgements

We gratefully acknowledge the participating laboratories for providing the GDAT team with tritium and noble gas isotope analyses. We thank "Source du Val Saint-Lambert" for permission to sample the boreholes. We thank Peter Cook, Daren Gooddy, Ed Busenberg, Niel Plummer, Axel Suckow, Kip Solomon for discussions on the design of the GDAT intercomparison exercise and László Palcsu for his comments on the manuscript. We thank three anonymous reviewers for their valuable comments.

Part of this work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-JRNL-640319.

References

Aeschbach-Hertig, W., F. Peeters, U. Beyerle, and R. Kipfer (1999), Interpretation of dissolved atmospheric noble gases in natural waters, *Water Resour Res*, *35*(9), 2779-2792.

Aeschbach-Hertig, W., F. Peeters, U. Beyerle, and R. Kipfer (2000), Palaeotemperature reconstruction from noble gases in ground water taking into account equilibration with entrapped air, *Nature*, *405*(6790), 1040.

Ballentine, C. J., and C. M. Hall (1999), Determining paleotemperature and other variables by using an error-weighted, nonlinear inversion of noble gas concentrations in water, *Geochim Cosmochim Ac*, *63*(16), 2315-2336.

Beyerle, U., W. Aeschbach-Hertig, D. M. Imboden, H. Baur, T. Graf, and R. Kipfer (2000), A mass spectrometric system for the analysis of noble gases and tritium from water samples, *Environmental Science and Technology*, *34*(10), 2042-2050.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu (1995), A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Comput.*, *16*(5), 1190-1208, doi:10.1137/0916069.

Clarke, W. B., W. J. Jenkins, and Z. Top (1976), Determination of tritium by mass spectrometric measurement of ^{3}He, *The International Journal of Applied Radiation and Isotopes*, *27*(9), 515-522, doi:http://dx.doi.org/10.1016/0020-708X(76)90082-X.

Corcho Alvarado, J. A., F. Barbecot, R. Purtschert, M. Gillon, W. Aeschbach-Hertig, and R. Kipfer (2009), European climate variations over the past half-millennium reconstructed from groundwater, *Geophysical Research Letters*, *36*(15), L15703, doi:10.1029/2009gl038826.

Corcho Alvarado, J. A., R. Purtschert, F. Barbecot, C. Chabault, J. Rueedi, V. Schneider, W. Aeschbach-Hertig, R. Kipfer, and H. H. Loosli (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*(3), W03427.

Gardner, P., and D. K. Solomon (2009), An advanced passive diffusion sampler for the determination of dissolved gas concentrations, *Water Resour Res*, *45*(6), W06423.

Gröning, M., H. Tatzber, A. Trinkl, P. Klaus, and M. v. Duren (2009), Eigth IAEA Interlaboratory Comparison on the Determination of Low-Level Tritium Activities in Water (TRIC2008)*Rep.*, International Atomic Energy Agency, Vienna.

Heaton, T. H. E., and J. C. Vogel (1981), "Excess air" in groundwater, *Journal of Hydrology*, *50*, 201-216.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995), *Continuous Univariate Distributions, chapters 18 (volume 1) and 29 (volume 2)*, Wiley, New York.

Labasque, T., L. Aquilina, V. Vergnaud, R. Hochreutener, F. Barbecot, and G. Casile (2014), Inter-comparison exercises on dissolved gases for groundwater dating – (1) Goals of the exercise and site choice, validation of the sampling strategy, *Applied Geochemistry*, *40*(0), 119-125, doi:http://dx.doi.org/10.1016/j.apgeochem.2013.11.007.

Lott III, D. E. (2001), Improvements in noble gas separation methodology: A nude cryogenic trap, *Geochem. Geophys. Geosyst.*, *2*(12), doi:10.1029/2001gc000202.

Lucas, L. L., and M. P. Unterweger (2000), Comprehensive review and critical evaluation of the half-life of Tritium, *J. Res. Natl. Inst. Stand. Technol.*, *105*(4), 541-549.

Marine, I. W. (1979), The use of naturally occurring helium to estimate groundwater velocities for studies of geologic storage of radioactive waste, *Water Resour Res*, *15*(5), 1130-1136, doi:10.1029/WR015i005p01130.

Mégnien, C. (1979), Hydrogéologie du centre du Bassin de Paris, Mem. BRGM, 98, 144–149.*Rep.*

Ménillet, F. (1988), Meulières, argiles a` meulières et meuliérisation - Historique évolution des termes et hypothèses génétiques, *Bull. Inf. Geol. Bassin Paris*, *25*(4), 71– 79.

Mercier, R. (1981), Inventaire des ressources aquifères et vulnérabilité des nappes du département des Yvelines, Rapp. 81SGN348IDF, Bur. de

Rech. Géol. et Min., Paris.*Rep.*

Palcsu, L., Z. Major, Z. Köllő, and L. Papp (2010), Using an ultrapure ^{4}He spike in tritium measurements of environmental water samples by the ^{3}He-ingrowth method, *Rapid Communications in Mass Spectrometry*, *24*(5), 698-704, doi:10.1002/rcm.4431.

Poreda, R. J., T. E. Cerling, and D. K. Salomon (1988), Tritium and Helium-Isotopes as Hydrologic Tracers in a Shallow Unconfined Aquifer, *Journal of Hydrology*, *103*(1-2), 1-9.

Raoult, Y., J. Boulègue, J. Lauverjat, and P. Olive (1997), Contribution de la géochimie à la compréhension de l'hydrodynamisme de l'aquifère de l'Albien dans le Bassin de Paris., *Comptes Rendus de l'Academie des Sciences - Series IIA: Earth and Planetary Science*, *325*(6), 419-425.

Roether, W., M. Vogt, S. Vogel, and J. Sültenfuß (2013), Combined sample collection and gas extraction for the measurement of helium isotopes and neon in natural waters, *Deep Sea Research Part I: Oceanographic Research Papers*, *76*(0), 27-34, doi:http://dx.doi.org/10.1016/j.dsr.2013.02.006.

Schlosser, P., M. Stute, H. Dorr, C. Sonntag, and K. O. Munnich (1988), Tritium/^{3}He Dating of Shallow Groundwater, *Earth and Planetary Science Letters*, *89*(3-4), 353-362.

Schlosser, P., M. Stute, C. Sonntag, and K. O. Munnich (1989), Tritiogenic ^{3}He in Shallow Groundwater, *Earth and Planetary Science Letters*, *94*(3-4), 245-256.

Schneider, V. (2005), Apports de l'hydrodynamique et de la géochimie à la caractérisation des nappes de l ` Oligocène et de l`Eocene, et à la reconnaissance de leurs relations actuelles et passées: Origine de la dégradation de la nappe de l`Oligocène (sud-ouest du Bassin de Pris), Ph.D. thesis, Univ. Paris-Sud, Orsay, France.

Shapiro, S. S., and M. B. Wilk (1965), An analysis of variance test for normality (complete samples), *Biometrika*, *52*, 591-611.

Solomon, D. K., S. L. Schiff, R. J. Poreda, and W. B. Clarke (1993), A Validation of the ^{3}H/^{3}He Method for Determining Groundwater Recharge, *Water Resour Res*, *29*(9), 2951-2962.

Stute, M., M. Forster, H. Frischkorn, A. Serejo, J. F. Clark, P. Schlosser, W. S. Broecker, and G. Bonani (1995), Cooling of Tropical Brazil (5ºC) during the Last Glacial Maximum, *Science*, *269*(5222), 379.

Sültenfuß, J., W. Roether, and M. Rhein (2009), The Bremen mass spectrometric facility for the measurement of helium isotopes, neon, and tritium in water†, *Isotopes in Environmental and Health Studies*, *45*(2), 83-95, doi:10.1080/10256010902871929.

Takaoka, N., and Y. Mizutani (1987), Tritiogenic ^{3}He in groundwater in Takaoka, *Earth and Planetary Science Letters*, *85*(1-3), 74-78.

Tolstikhin, I. N., and I. L. Kamenski (1969), Determination of groundwater age by the T-^{3}He method, *Geochemistry International*, *6*, 810-811.

Table 1: Contributors and participating laboratories

Contributor	Laboratory	^{3}H	Noble gases
Suckow A., Leaney F.	CSIRO, Autralia		x
Matsumoto T., Han L.,	IAEA, Austria	x	x
Aeschbach-Hertig W.	IUP,Heidelberg, Germany	x	x
Yoon Y. Y.	KIGAM, Korea	x	
Sliwka I., Bielewski J.	INP, Poland		x
Solomon K., Rigby A.	Utah University, USA	x	x
Barbecot F., Lefebvre K.	IDES, France	x	
Travi Y, Babic M.	Lab Hydrol. Avignon, France	x	
Pauwels H., Fléhoc C.	BRGM, France	x	x
Fourré E. Jean-Baptiste P., Dapoigny A.	LSCE, France	x	x
Palcsu L.	Atomki, Hungary	x	x
Niedermann S.	GFZ Potsdam, Germany		x
Sültenfuß J.	Univ. Bremen, Germany	x	x
Otha T.	H.K.A.T. Japan	x	x
Rosanski K., Bartyzel J.	AGH Poland	x	
Gumm L., Hiscock K., Dennis P.	University of East Anglia, GB		x
Hunt A.	USGS, NG lab, USA	x	x
Visser A.	LLNL, USA	x	x
Lavielle B., Thomas B.	CENBG, France		x

Table 2: Number of analyses reported by participating laboratories (in parentheses) for each of the boreholes.

		Borehole	
Parameter	Albian	SLP4	SLP5
^{3}H	19 (12)	22 (14)	21 (13)
$^{3}He/^{4}He$	10 (5)	25 (10)	23 (9)
^{4}He	10 (5)	25 (10)	23 (9)
Ne	10 (5)	25 (10)	22 (9)
$^{20}Ne/^{22}Ne$	5 (2)	12 (5)	11 (4)
$^{21}Ne/^{22}Ne$	0 (0)	1 (1)	0 (0)
Ar	6 (3)	19 (8)	18 (7)
$^{38}Ar/^{36}Ar$	3 (1)	6 (2)	5 (1)
$^{40}Ar/^{36}Ar$	5 (2)	12 (5)	11 (4)
Kr	6 (3)	18 (8)	17 (7)
$^{86}Kr/^{84}Kr$	0 (0)	2 (1)	2 (1)
Xe	6 (3)	19 (8)	18 (7)
$^{128}Xe/^{132}Xe$	0 (0)	1 (1)	0 (0)
$^{129}Xe/^{132}Xe$	3 (1)	6 (2)	5 (1)
$^{130}Xe/^{132}Xe$	0 (0)	3 (2)	2 (1)
$^{131}Xe/^{132}Xe$	0 (0)	1 (1)	0 (0)
$^{134}Xe/^{132}Xe$	0 (0)	1 (1)	0 (0)
$^{136}Xe/^{132}Xe$	3 (1)	6 (2)	5 (1)

Table 3: Mean measured values, excluding the identified outliers, relative laboratory mean reproducibility (σ_{lm}) and single sample reproducibility (σ_{ss}) and number of outliers.

		$^3He/^4He$	^4He	Ne	Ar	Kr	Xe
		-			cm^3STP/g		
		10^{-6}	10^{-8}	10^{-7}	10^{-4}	10^{-8}	10^{-8}
Mean							
	Albian	0.122	267	2.42	4.64	11.02	1.62
	SLP4	1.57	6.15	2.63	4.13	9.29	1.32
	SLP5	1.41	6.44	2.77	4.26	9.50	1.34
Laboratory Mean Reproducibility							
	Albian	3.0%	11.2%	11.0%	0.2%	1.3%	2.9%
	SLP4	1.4%	1.8%	1.5%	2.2%	2.9%	2.4%
	SLP5	2.4%	3.1%	2.1%	3.1%	2.7%	2.5%
Single Sample Reproducibility							
	Albian	2.8%	9.1%	8.5%	1.2%	2.2%	2.6%
	SLP4	1.3%	1.5%	1.5%	2.5%	3.1%	3.5%
	SLP5	2.0%	3.1%	2.7%	3.7%	3.6%	4.7%
Outliers							
	Albian	2	0	0	0	0	0
	SLP4	2	4	7	0	0	0
	SLP5	2	2	2	0	0	0
		^3H	$^{20}Ne/^{22}Ne$	$^{38}Ar/^{36}Ar$	$^{40}Ar/^{36}Ar$	$^{129}Xe/^{132}Xe$	$^{136}Xe/^{132}Xe$
		TU			-		
Mean							
	Albian	0.02	9.78	0.193	295	0.980	0.332
	SLP4	6.34	9.79	0.189	295	0.986	0.333
	SLP5	2.58	9.79	0.188	294	0.983	0.334
Laboratory Mean Reproducibility							
	Albian	-	0.13%	-	1.3%	-	-
	SLP4	13.5%	0.05%	0.7%	0.4%	0.6%	1.4%
	SLP5	16.2%	0.05%	-	0.7%	-	-
Single Sample Reproducibility							
	Albian	-	0.11%	2.5%	1.2%	0.2%	0.9%
	SLP4	13.9%	0.06%	0.5%	0.5%	0.5%	1.0%
	SLP5	15.6%	0.06%	0.8%	0.8%	0.2%	0.2%
Outliers							
	Albian	2	2	0	0	0	0
	SLP4	0	4	2	0	0	0
	SLP5	1	4	1	0	1	1

Table 4a: Results from fitting the three excess air models to the Ne, Ar, Kr and Xe concentrations.

Parameter	Borehole	Excess air model		
		UA	CE	PR
χ^2	All	51.4	37.0	38.0
χ^2 probability	All	99.9%	73%	69%
Noble gas recharge temperature				
	Albian	4.2±0.4	4.2±0.5	4.3±0.5
	SLP4	10.2±0.9	10.6±1.0	10.7±1.1
	SLP5	9.9±1.4	10.4±1.5	10.5±1.6

Table 4b: Results from fitting the three excess air models to all noble gas concentrations including helium.

Parameter	Borehole	Excess air model		
		UA	CE	PR
χ^2	SLP4+SLP5	114.4	63.6	56.7
χ^2 probability	SLP4+SLP5	82%	97%	99%
Noble gas recharge temperature				
	SLP4	9.6±0.8	10.8±1.0	10.5±1.2
	SLP5	9.3±1.3	10.5±1.3	10.4±2.1

Table 5: Differences between estimated noble gas temperature and excess air amount in synthetic sample with added residuals with a recharge temperature of 10°C and ΔNe = 9%.

lab	NGT [°C]	ΔNe [%]
F	1.3	-0.4
G	0.1	1.8
H	-0.3	3.2
I	0.1	3.8
J	-0.1	1.1
M	0.6	-1.6
N	0.6	0
O	-0.1	2.1
P	-1.7	0.6
Q	-0.7	1.1
R	-2.5	-0.4
U	-0.3	-1.5
V	-0.1	0.2
W	0	-8.3
X	-0.4	-1.9
Propagated Uncertainty	0.7	1.5

Figure captions

Figure 1: Box-whisker plots for noble gas and tritium concentrations and noble gas isotope ratios measured in each sample. Bold horizontal line represents median, box represents 25%-75% quantiles, whiskers extend to 1.5 times the interquartile range or the maximum value (whichever is closer to the median), outlying values are plotted if beyond the whiskers. Note different axes for $^{3}He/^{4}He$, ^{4}He and ^{3}H for Albian (left) and SLP4 and SLP5 (right).

Figure 2: Differences between measured sample tritium concentrations and the borehole mean concentration.

Figure 3: Differences between measured sample concentrations and the borehole mean concentration, for the helium isotope ratio and noble gas concentrations.

Figure 4: Derived parameters recharge temperature (a), modeled excess air amount expressed as ΔNe (b) and excess air fractionation parameter (c) derived from the excess air models fitted to four noble gases (Ne,Ar, Kr, Xe: horizontal axis label 4) and five noble gases (He, Ne, Ar, Kr, Xe: horizontal axis label 5).

Figure 5: Terrigenic helium (a) observed in boreholes Albian, SLP4 and SLP5, derived from excess air models fitted to Ne, Ar, Kr and Xe. Helium residual (a) of the excess air models fitted to He, Ne, Ar, Kr and Xe shows that the PR model is better capable of reproducing the measured helium concentrations. Dashed lines show the measurement uncertainty of 0.095×10^{-8} $cm^{3}STP/g$ for helium.

Figure 6: a) Tritiogenic ^{3}He, derived from each excess air model (UA, CE or PR) and from the measured helium concentration and isotope ratio alone (helium-model, He). b) Apparent $^{3}H/^{3}He$ age, calculated from the mean tritium concentration in the borehole. c) Apparent $^{3}H/^{3}He$ age, calculated from the measured tritium concentrations and mean tritiogenic helium in each borehole, according to the three excess air models.

Figure 7: Uncertainty of groundwater $^{3}H/^{3}He$ age estimates resulting from the uncertainty of the tritium and noble gas measurements (solid lines) and the contribution of the uncertainty of the tritiogenic helium estimate as a percentage of the total uncertainty (dashed lines) in groundwater samples with excess air amounts equivalent to a ΔNe of 0%, 45% and 90%.

Highlights

Key results of the groundwater age-dating inter-laboratory comparison exercise

The reproducibility of the tritium measurements was 13.5%.

The noble gas reproducibility was <2% (R, He, Ne) and <3% (Ar, Kr, Xe).

The measurement uncertainty meets the requirements for $^{3}H/^{3}He$ dating.

Other sources of uncertainty are less well defined than the analytical uncertainty

Figure 1

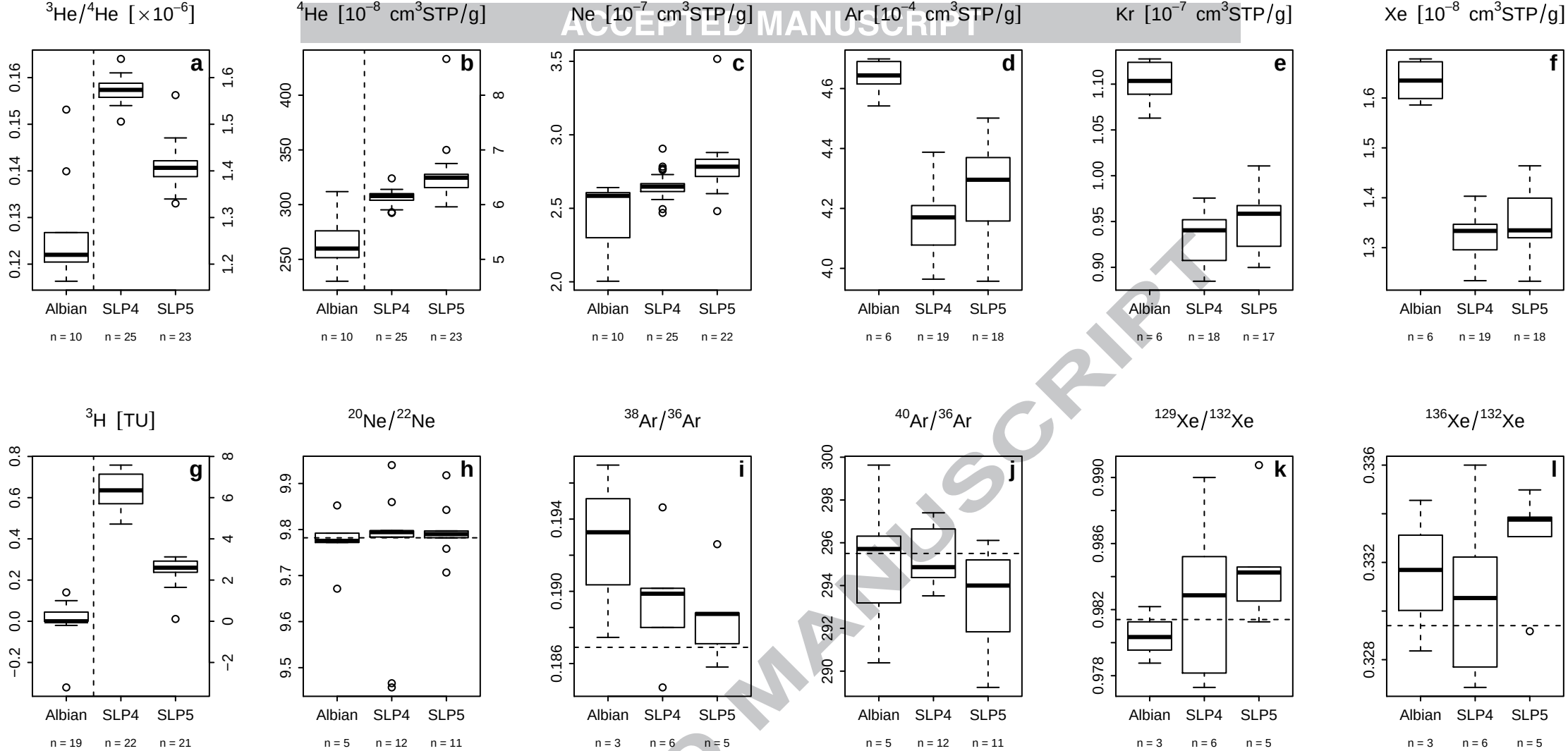

Figure 2

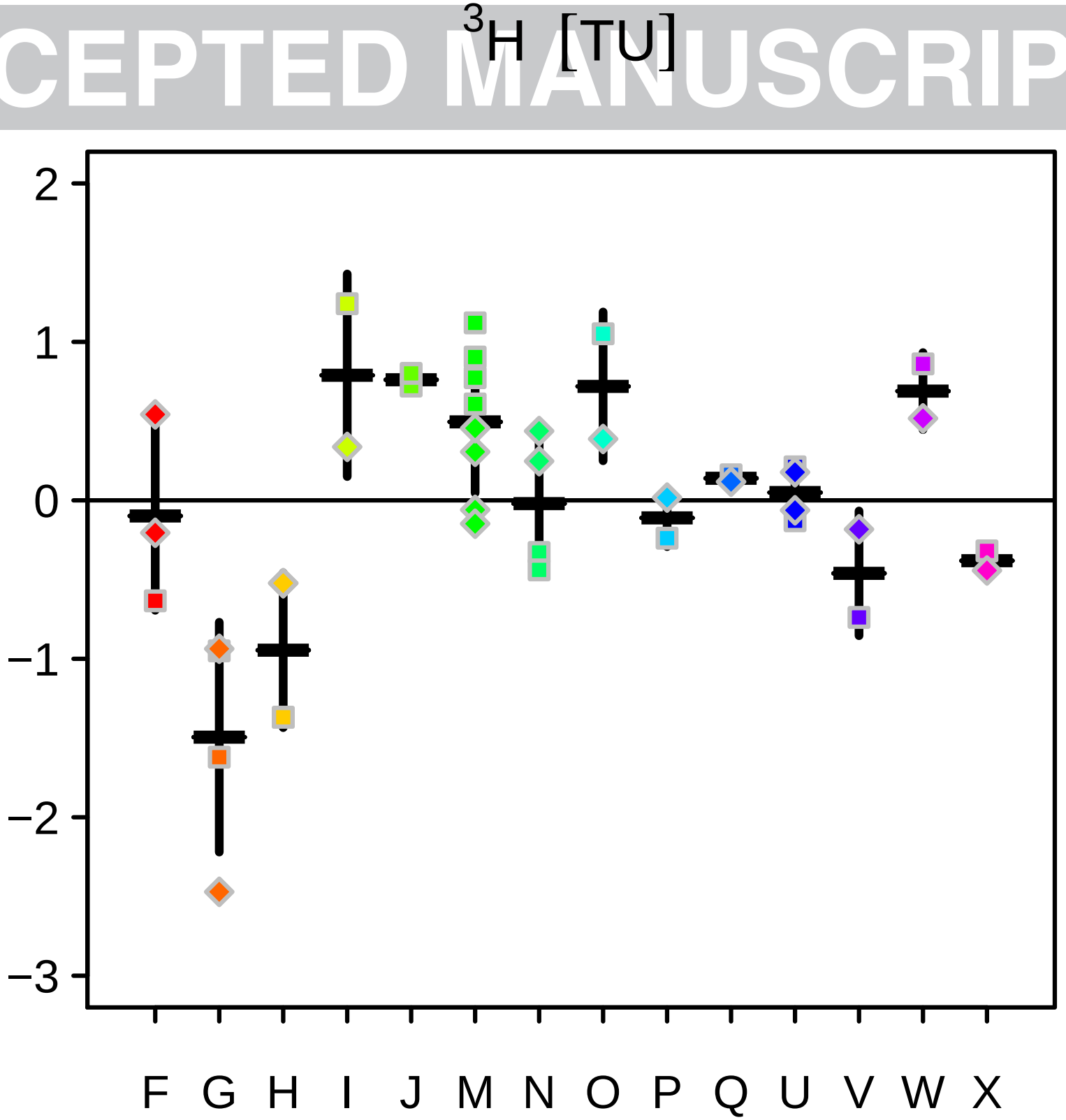

Figure 3

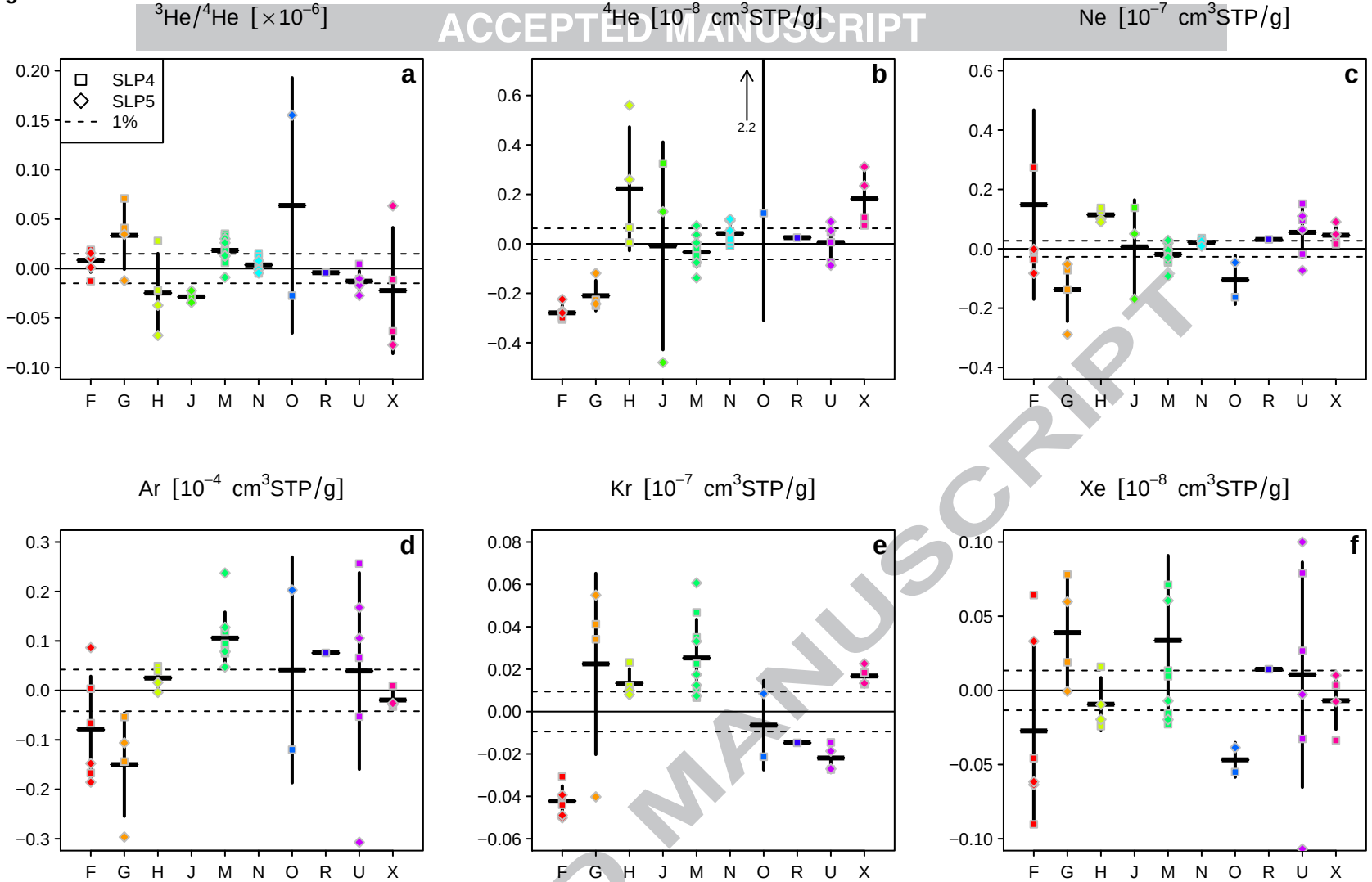

Figure 4

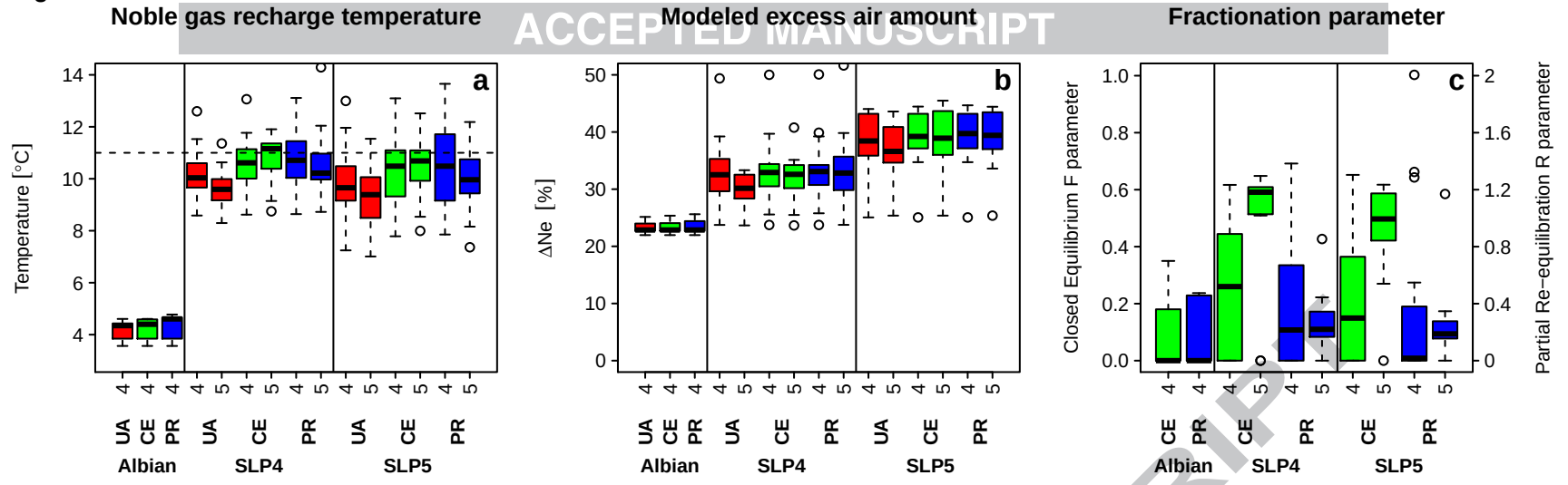

Figure 5

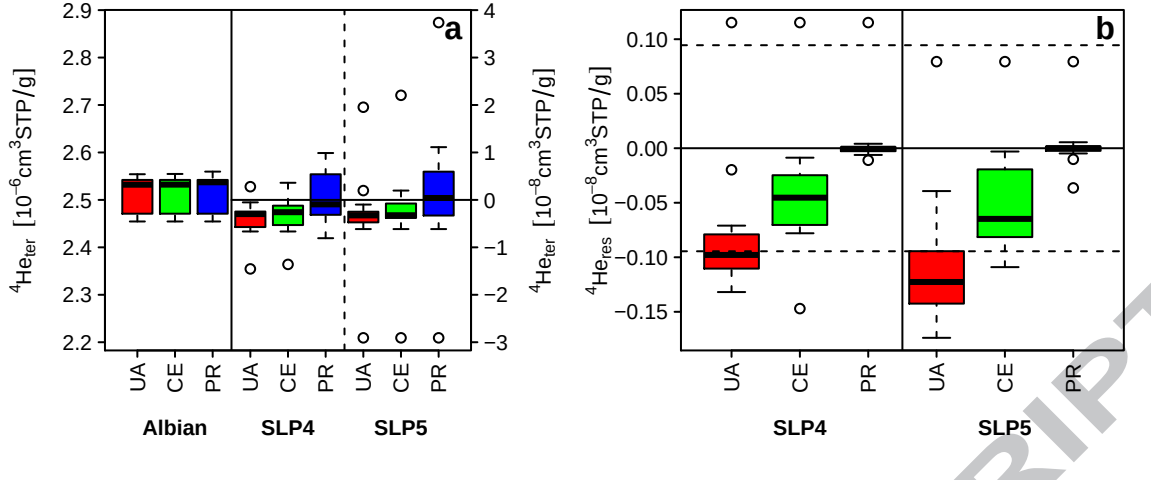

Figure 6

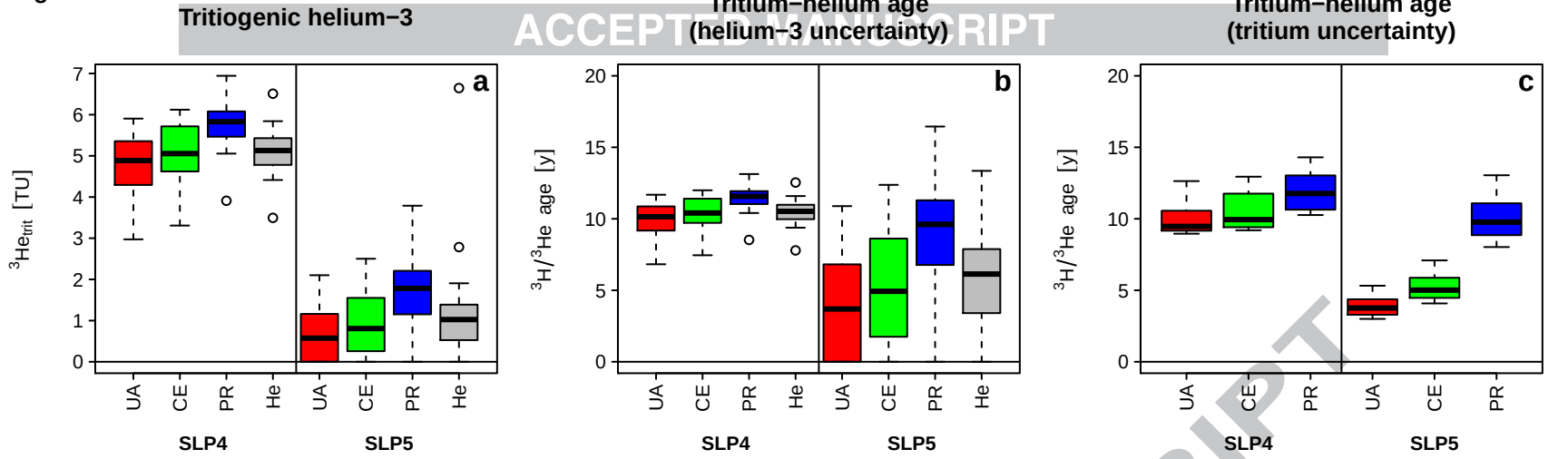

Figure 7

Groundwater Age Uncertainty