

“Groundwater mean transit times, mixing and recharge in faulted–hydraulic drop alluvium aquifers using chlorofluorocarbons (CFCs) and tritium isotope (^3H)” by Ma, B., Jin, M., Liang, X., Li, J., *Hydrol. Earth Syst. Sci.*, doi:10.5194/hess-2018-143.

We appreciate the many valuable suggestions and helpful comments of **Anonymous Referee #2**. We have seriously considered all of the suggestions and comments and have attempted to address each of the comments point-by-point. Detail explanations are as follows.

Author’s response – Line numbers referring to the old and revised version manuscripts are preceded by L and RL, respectively.

Anonymous Referee #2

1) As indicated by the title, this manuscript presents the results of a groundwater dating and mixing study conducted using two different atmospheric tracers (CFCs and tritium). The two aims of the study were to (i) relate “ages” to local and general hydrogeological conditions and (ii) explore the possibility to use mineralisation as proxy for environmental tracers. I agree with referee #1 concerning the style, which is a huge disservice to the manuscript by its approximate use of technical terms and the general turn of phrase. I disagree however with the novelty (I do not see any) and the “substantial conclusions” (very unsubstantial and too dependent on mean transit time calculations that at present look extremely weak). As far as comment 4 of referee #1 is concerned, I think it is simply a matter of opinion and taste to use “transit time” instead of “residence time” (I prefer transit time because my work is related to solute transport problems, and “transit time” conveys this very idea of transport). One can argue over that, but it is really a hair splitting exercise.

Response: We would like to thank you very much for taking the time to review our manuscript. We think that the many valuable suggestions and helpful comments will help to improve our manuscript quality greatly.

We agree with you about the aims of the study and the opinion on the terms “transit time” and “residence time”. We re-read the literatures and found that term “transit time” was numerously used to indicate the time for water to transit through the catchment and into the stream (Cartwright et al., 2018; Cartwright and Morgenstern, 2015, 2016; Hrachowitz et al., 2009, 2010; Morgenstern et al., 2010; Stewart and Morgenstern, 2016). Stewart et al. (2010) pointed out that “Residence time is the time spent in the catchment since arriving as rainfall. Transit time is the time taken to pass through the catchment and into the stream.” Leray et al. (2016) have adopted a general but robust definition for the residence time “the amount of time a moving element has spent in a hydrologic system”, and considered the terms residence time, transit time, travel time, age, and exposure time as equivalent in their discussions. Custodio et al. (2018) used both residence times and transit times for groundwater samples that collected from springs and deep wells. In our study, all of the groundwater samples were collected from the wells/artesian wells. Thus, we tend to use the term “residence time” instead of “transit time” in our manuscript.

In addition, we have reorganized the structure and tried our best to present a clear roadmap to readers. Some incorrect phrase or expressions, for example the “apparent CFC ages” (L277) and “The apparent ages (Table 2 and Fig. 5) estimated from the PFM ...” (L509), which mislead the study and thus

have been revised and rewritten. In the revised manuscript, we tried to discuss the mean residence times and hydrochemistry evolution clearly and reasonably.

2) Overall, the authors seem to have read sufficiently thoroughly the existing literature on the subject as well as the most recent developments (such as Kirchner's analysis of the effect of heterogeneity on mean transit time estimation using amplitude damping) and understood the different problems and pitfalls relevant for their study. However, the phrasing is sometimes very awkward and tends to obfuscate what the authors mean (see specific comments below).

Response: We have checked the manuscript carefully and have tried our best to modify some erroneous words and phrases that you and Ref #1 pointed out. Some ambiguous statements and sentences have also been rewritten. On the other hand, we will ask a proof reader to modify the language to help improve readability.

3) But above all, I am missing a strong reason for this study to be published at all. As case study, it does not go beyond the classical scheme of sampling a few boreholes, analyse the groundwater samples for one or more tracers, calculate some kind of "age" and correlate it to depth or water chemistry. Doing so however, the authors try to apply different methods (lumped-parameter modelling, binary mixing) without presenting a clear roadmap.

Response: Manas River with the longest channel length and the largest river flow is representative among the hundreds of inland rivers in the Junggar Basin in the arid northwest China (Fig. 1a). There are more than 1 million people in the Manas River Basin. Groundwater is the very important and even the only water resources for the local residents. However, the paucity of research on groundwater mean residence time, mixing and recharge features have hampered the rational development and utilization of water resources in the Junggar Basin. We think that the case study of our manuscript is very essential and can be a good proxy for other analogous areas in the Junggar Basin.

Yes, we tried to estimate the mean residence time using lumped-parameter models. In addition, we also have tried to recognize the modern and paleo-meteoric recharge features. However, it is a pity that there are many deficiencies for the initial manuscript. We have reorganized the structure and tried our best to present a clear roadmap to readers. The outline of the manuscript have been reorganized as follows:

Title: Application of environmental tracers for investigation of groundwater mean residence time and aquifer recharge in faulted-hydraulic drop alluvium aquifer

1. Introduction
2. Geological and hydrogeological setting
3. Materials and methods
 - 3.1 Water sampling
 - 3.2 Analytical techniques
 - 3.3 Groundwater dating
 - 3.3.1 CFCs indicating modern water recharge
 - 3.3.2 The apparent ^{14}C ages
 - 3.3.3 Groundwater mean residence time estimation
4. Results and discussion
 - 4.1 Stable isotope and major ion hydrochemistry

4.2 Modern and paleo–meteoric recharge features

4.2.1 Stable isotope indications

4.2.2 CFCs indications

4.2.3 ³H and ¹⁴C indications

4.3 Groundwater mean residence time

4.3.1 ³H and CFCs

4.3.2 Hydrochemistry evolution

5. Conclusions

4) Model choice in particular is strangely presented: first, “apparent age” is presented as “based on the hypothesis of pistonflow”. Then that very piston-flow model is used although mixing is supposed to be “most likely” either within the aquifer or at the sampling point. This is completely contradictory and there is no reason not to apply another model to the CFC data (and for that matter, to the 14C data as well. See Custodio et al., 2018).

Response: We have deleted “apparent CFC ages” and rewritten the content of the model choice for mean residence time estimation (Section 3.3.3 (RL219–272)):

“Groundwater mixing may occur both within the aquifer and in the long–screened wells (Cook et al., 2017; Custodio et al., 2018; Visser et al., 2013). The groundwater residence times (ages) often display a wide range due to the recharge under various climate conditions (Custodio et al., 2018). With the aid of gaseous tracers (e.g. ³H, CFCs, SF₆ and ⁸⁵Kr) one can describe the mixing distribution by a mixing model (Stewart et al., 2017; Zuber et al., 2005) and to obtain the groundwater mean residence times (MRTs). Lumped–parameter models (LPMs) is an alternative approach to interpret the MRTs for water flow through the subsurface systems to the output. For steady–state subsurface hydrologic system, on accounting of the ³H and CFCs tracers enter groundwater with precipitation are injected proportionally to the volumetric flow rates by nature itself, the output concentration in water at the time of sampling relating to the input ³H and CFCs can be described by the convolution integral (Małoszewski and Zuber, 1982):

$$C_{\text{out}}(t) = \int_0^{\infty} C_{\text{in}}(t-\tau) g(\tau) e^{-\lambda_{\text{3H}}\tau} d\tau \quad \text{for } ^3\text{H tracer} \quad (2a)$$

$$C_{\text{out}}(t) = \int_0^{\infty} C_{\text{in}}(t-\tau) g(\tau) d\tau \quad \text{for CFCs tracer,} \quad (2b)$$

where C_{out} is the tracer output concentration, C_{in} is the tracer input concentration, τ is the residence time, $t-\tau$ is the time when water entered the catchment, λ_{3H} is the ³H decay constant

($\lambda_{\text{3H}} = \ln 2/12.32$), and $g(\tau)$ is the system response function that describes the residence time distributions (RTDs) in the subsurface hydrologic system.

In this study the CFC concentrations from the time series trend of Northern Hemisphere atmospheric mixing ratio (Fig. 3) and ³H concentrations in precipitation in Urumqi (Fig. 4) are treated as proxies for CFC and ³H recharge concentrations (C_{in}), respectively. The historical precipitation ³H activity in Urumqi station (Fig. 4) was reconstructed from the available data in the International Atomic Energy Agency (IAEA) using a logarithmic interpolation method. Precipitation ³H activities between 1969 and 1983 at Hongkong and Irkutsk with different latitudes were used (data is available at <<https://www.iaea.org/>>). The time series ³H activities as the input data are still necessary that is based on the following two considerations. First, Manas River Basin is located in the Northern Hemi-

sphere, where the bomb–test ^3H activities were several orders of magnitude higher than in the Southern Hemisphere (Clark and Fritz, 1997; Tadros et al., 2014). ^3H activity in the atmosphere was superimposed over the China atmospheric nuclear tests from 1964 to 1974 in the arid northwest China, and thus the remnant ^3H activities are still affected by the tail–end of the bomb pulse. Second, the study area is more than 3500 km away from the western pacific, where ^3H activity in the atmosphere is evidently much higher than coastal sites due to the continental effect (Tadros et al., 2014). Furthermore, though ^3H activity in the atmosphere is known to vary between seasons (Cartwright and Morgenstern, 2016; Morgenstern et al., 2010; Tadros et al., 2014), the year–round mean values (Fig. 4) were adopted in this study.

Several RTDs have been described (Małozzewski and Zuber, 1982; Jurgens et al., 2012) and been widely used in studies of variable timescales and catchment areas (Cartwright and Morgenstern, 2015, 2016; Cartwright et al., 2018; Hrachowitz et al., 2009; Morgenstern et al., 2010, 2015; McGuire et al., 2005), of which the selection of each model depends on the hydrogeological situations in the hydrologic system to which it is applicable. The exponential–piston flow model (EPM) describes aquifer that containing a segment of exponential flow followed by a segment of piston flow. Piston flow assumes that water mixing from different flow lines is minimal and receiving little or no recharge in the confined aquifer, and the exponential flow assume a full mixing of water in the unconfined aquifer and receiving areally distributed recharge (Jurgens et al., 2012; Małozzewski and Zuber, 1982). Its RTD is

$$g(\tau) = 0 \quad \text{for } \tau < \tau_m(1-1/\eta) \quad (3a)$$

$$g(\tau) = \frac{\eta}{\tau_m} e^{(-\eta\tau/\tau_m + \eta - 1)} \quad \text{for } \tau \geq \tau_m(1-1/\eta) \quad (3b)$$

The dispersion model (DM) mainly measures the relative importance of dispersion to advection, and is applicable for confined or partially confined aquifers (Małozzewski, 2000). Its RTD is given by

$$g(\tau) = \frac{1}{\tau \sqrt{4\pi D_p \tau / \tau_m}} e^{-\left(\frac{(1-\tau/\tau_m)^2}{4\pi D_p \tau / \tau_m}\right)} \quad (4)$$

The exponential mixing Model (EMM) is given by

$$g(\tau) = \frac{1}{\tau_m} e^{(-\tau/\tau_m)}, \quad (5)$$

where τ_m is the mean residence time, η is the ratio defined as $\eta = (l_p + l_E)/l_E = l_p/l_E + 1$, where l_E (or l_p) is the length of area at the water table (or not) receiving recharge, D_p is the dispersion parameter, which is the reciprocal of the Peclet number (Pe) and defined as $D_p = D/(vx)$, where D is the dispersion coefficient ($\text{m}^2 \text{day}^{-1}$), v is velocity (m day^{-1}), and x is distance (m).

Each RTD has one or two parameters, which the MRT (e.g. τ_m) is determined by convoluting the input (the time series ^3H and CFCs input in rainfall) to each model in a way that matches the output (the measured ^3H and CFC concentrations in groundwater), and other parameters (e.g. η and D_p) are determined depended on the hydrogeological conditions. To interpret the ages of the Manas River Basin data set, the exponential–piston flow model (EPM, with $\eta=1.5$ and 2.2), the dispersion model (DM, with $D_p=0.03$ and 0.1), and the exponential mixing model (EMM) were used and then the MRTs with different RTDs were cross–referenced.”

We read the literature carefully and think that both the method and idea are very innovative by Cus-

todio et al. (2018), who related chloride (Cl) and ^{14}C activity changes in recharge for aquifers in the arid area to indicate the effect of variations in recharge rate during the previous wetter-than-present period. We think that it would be also good applications in other arid areas around the world because the climatic change is global. In our manuscript, the apparent ^{14}C ages estimation was adopted. The changes in groundwater reserves and ^{14}C content will be conducted in the future in the Manas River Basin.

5) I know it is customary to interpret CFC data assuming piston-flow, but it is nonetheless a priori wrong. Model choice must be substantiated from knowledge of the hydrogeological situation and the sampling scheme (Maloszewski and Zuber, 1982; Leray et al., 2016). Later on in the manuscript however different models are used in the binary mixing plots, and model choice is discussed briefly. Why use the “apparent age” concept at all, then ? This is confusing and reads like the two authors have written separately different parts of the manuscript and then pasted the two parts together.

Response: Sorry, we have realized that the apparent CFC ages are not appropriate. As we know that the CFCs are synthetic organic compounds and largely released to the air since 1930s, and thus they have been treated as good tracers for dating young water recharge time (post-1940 recharge).

We totally agree with you that “Model choice must be substantiated from knowledge of the hydrogeological situation and the sampling scheme”. Therefore, in the revised manuscript, we have rewritten the content and have discussed the model choice in detail in Section 3.3.3 (See Response 4).

The mean residence times estimated by the lumped-parameter models have been deleted in the binary mixing plots (Fig. 7). However, the binary plots of CFC vs. CFC and CFCs vs. ^3H are widely used in the literatures and can provide useful information on the co-existence of young water with old water (Cook et al., 2017; Han et al., 2007; Han et al., 2012; Koh et al., 2012; Qin et al., 2011), as well as can provide a powerful tool to recognize contamination, degradation and irrigation infiltration and so on (Cook et al., 2017; Han et al., 2015; Mahlknecht et al., 2017; Qin et al., 2012). However, these information mentioned above have not been clarified clearly before in the Manas River Basin, and even not in the Junggar Basin (Fig. 1a). Therefore, we think that the binary mixing plots are essential in the manuscript in order to tell the reader more information about groundwater mixing or recharge features.

6) I also have my doubts concerning the calculations of the mean transit times as they are presented.

Response: Take the CFC-12 and exponential-piston flow model (1.5) for example, the calculation process of the mean residence times is as follows: First, we chose the Eq. (2b) as the convolution integral, and chose the exponential-piston flow model (Eq. (3a, 3b)) as the system response function. Second, we used the time series CFC-12 trend of Northern Hemisphere atmospheric mixing ratio (1940–2014, <http://water.usgs.gov/lab/software/air/cure/>) as input concentrations. We treated the calendar year 2015 (groundwater sampling time) as age=0 yrs by convoluting the input (times series of CFC-12 input) to the EPM (1.5), and increased the mean residence time τ_m from 1 to 500 yrs with the time step of 1 yr within 100 yrs and with the time step of 5 yrs between 101 to 500 yrs. Then we can get a sequence of results of output CFC-12 concentrations and mean residence times (vary from 1 to 500 yrs). Third, we plotted the output CFC-12 concentrations vs. mean residence times and then compared the measured groundwater CFC-12 concentrations to get the groundwater mean residence

times. The computational procedures were conducted by using MATLAB (version R2014a). The output CFC–12 concentrations decreased from 526.4 pptv with τ_m of 1 yr to 3.0 pptv with τ_m of 155 yrs. As the detection limit for each CFC is about 0.01 pmolL^{-1} of water (be equal to 3.54 pptv with the laboratory temperature of 25°C), the output CFC–12 concentrations lower than 3.54 pptv can be neglected.

7) The method with which the tritium input has been reconstructed is not documented properly (which stations were used, and how long were the available time series?) and the estimated modern value (31 TU) seems extremely high compared to Western Europe for instance (about 6 TU). Is that because of the Chinese nuclear tests of the 60s and 70s that are being referred to in the introduction, or the result of some kind of regional effect?

Response: Agree and changes made. The input tritium (^3H) activity in Urumqi station was reconstructed as follows: First, as we know that there is a proportionality between the logarithmic precipitation ^3H activity and the latitude in the Northern Hemisphere, we can reconstruct the Urumqi precipitation ^3H activity according to the following equation:

$$\lg C_U = \lg C_H + \frac{\lg C_I - \lg C_H}{X_I - X_H} * (X_U - X_H) \quad (\text{S1})$$

where C_U is Urumqi precipitation ^3H activity, C_H is Hongkong precipitation ^3H activity, C_I is Irkutsk (in Russia) precipitation ^3H activity, X_U is Urumqi latitude, X_H is Hongkong latitude, and X_I is Irkutsk latitude. The precipitation ^3H activity are available in the International Atomic Energy Agency (IAEA). The results based on Eq (S1) are show in Table S1.

Table S1. The reconstructed Urumqi precipitation ^3H activity between 1969 and 1983.

Date	C_H (TU)	C_I (TU)	C_U (TU)	Ottawa precipitation (TU)
1969	46.93	454.79	248.63	253.66
1970	29.38	464.15	222.82	190.77
1971	24.38	516.61	229.38	206.1
1972	27.86	247.23	138.36	92.34
1973	13.22	173.34	87.44	90.41
1974	17.47	216.58	110.90	98.07
1975	12.42	190.93	92.33	75.86
1976	11.71	148.8	75.69	58.91
1977	12.24	150.72	77.31	73.93
1978	10.89	103.11	56.72	73.63
1979	8.96	92.06	49.55	49.62
1980	7.63	141.8	65.20	49.54
1981	11.53	138.3	71.44	55.09
1982	9.73	66.92	40.08	47.29
1983	21.14	127.26	78.96	40

Second, a good linear regression between the C_U and Ottawa precipitation ^3H activity can be obtained as follows:

$$C_{U_i} = 1.0102 C_{\text{Ottawa}} + 11.647 \quad (R^2 = 0.933, n = 15) \quad (\text{S2})$$

where C_{Ottawa} is Ottawa precipitation ^3H activity, C_{Ui} is the Urumqi annual precipitation ^3H activity.

Third, considering the precipitation amount effect, the reconstructed Urumqi precipitation ^3H activity can be obtained as follows:

$$C_{\text{RU}} = C_{\text{Ui}} * \frac{P_i}{P_m} \quad (\text{S3})$$

where C_{RU} is the reconstructed Urumqi precipitation ^3H activity shown in Fig. 6 (estimated ^3H), P_i is the annual precipitation amount (1953–2015), P_m is the annual mean precipitation amount in Urumqi between 1953 and 2015.

Yes, the estimated modern precipitation ^3H activity are indeed extremely high (mean value of 31.55 TU between 2000 and 2015). We speculate that the very high precipitation ^3H activity may be ascribed to both the Chinese atmospheric nuclear tests (around 350 km away to Manas River Basin) from 1964 to 1974 (Zhou, 1992) and the continental effect (Tadros et al., 2014), where the Manas River Basin is more than 3500 km far away from the western pacific. The high precipitation ^3H activity was also shown in Fig. S1 (Cauquoin et al., 2015).

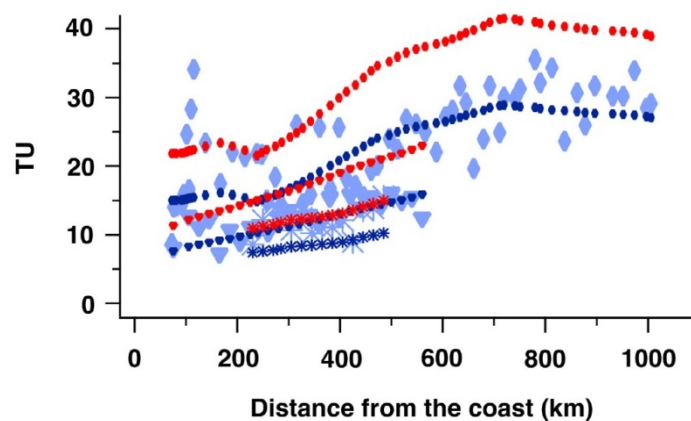


Figure S1. Comparison of modelled tritium in precipitation with surface snow samples from two ITASE Antarctic traverses (Proposito et al., 2002; Becagli et al., 2004) and some pre-bomb measurements. (The figure is from Cauquoin et al., 2015, Fig. 10d)

8) Furthermore, the authors do address the non-uniqueness problems that are bound to arise when calibrating an exponential piston-flow or a dispersion model (2 free parameters each) using a single tritium measurement in aquifers that still retain some of the “bomb tritium” (see Stewart et al., 2010 for details), but in a terribly confusing way and without first explaining the rationale and the approach taken.

Response: Sorry. We should have given more details about the two free parameters each for the exponential–piston flow and the dispersion model. These parameters are substantial based on the geological/hydrogeological conditions. We have added some more detailed explanations concerning the rationale and the approach taken in Section 3.3.3 (Response 4). The revised version can be seen in the “Response 4”.

9) I suppose figure 8 was meant to show the range of parameters that match the measured tracer concentrations. That's commendable, but badly explained.

Response: Yes, thanks. We have redrawn the figure 8 to show the most useful information to the readers (see the "Specific Comments Response 55").

10) In the final step relating mineralisation to transit time, the authors finally select the EPM calibrated with the CFC12 measurements, but this is once again presented in a unclear fashion.

Response: There are some reasons that we select the EPM (1.5) calibrated with the CFC-12 to relate hydrochemistry evolution: First, groundwater CFCs results show that the urban air with CFC compounds contaminations, which generally cause elevated CFC than the global background atmospheric CFC concentrations (Northern Hemisphere), are unlikely (Fig. 7) in the Manas River Basin. Second, CFC-12 has much less propensity for degradation and/or contamination than CFC-11 and CFC-113. Third, the exponential-piston flow model is one of the mostly commonly used system response functions. The choice of exponential-piston flow model is suitable for the aquifer that has two segments, in which a segment of exponential flow followed by a segment of piston flow. As shown in Fig. 1c and relative hydrogeological discussions, exponential-piston flow model is a good choice to be chose as the system response function. In addition, the mean residence times mathematic relation between EPM (1.5) and other system response functions are obtained (Fig. 11). We can relate mineralization to mean residence time based on these mathematic relations.

11) The discussion is too long, relies too much on untested and untestable hypotheses, and presents so many singular and unfocused results that it is difficult for the reader to grasp a clear picture of their meaning and significance. The paragraph on "apparent age" should be scraped altogether and the different estimates of "age" (i.e. mean transit time of the respective model) and mixing ratios organised in a clear and synthetic manner.

Response: To avoid misleading, we have deleted some irrelevant content, such as the incorrect statement of "apparent age" for the CFCs, and the incorrect hypothesis of piston flow as mixing is supposed to be "most likely" either within the aquifer or at the sampling point. We also tried to present a clear roadmap to readers, including modern and paleo-meteoric recharge features, mean residence time estimated by LPMs, and mixing features as well as irrigation infiltration indications.

12) All in all, the manuscript must be seriously reorganised and streamlined. The calculation of mean transit times of the different tracers must be redone, removing entirely the "apparent age" nonsense and explaining clearly the different steps taken by the authors to (I) select a model (II) explore model parameter range and (III) compare the different results obtained from tritium, the CFCs and carbon 14. Interpretation of the obtained "ages" in terms of hydrogeology and its correlation to hydrochemistry must then be presented in a clear and synthetic fashion. Only when this is done might the manuscript rise above an unoriginal and confusing rehash of previous studies, and could be considered for publication.

Response: We reorganized the structure and tried our best to present a clear roadmap to readers. We rewrote the Section 3.3.3 in which system response functions were selected by considering the applicable hydrogeological situations. Parameter and mean residence times were recalculated and then explained in a much clearer way. Figures 8 and 9 (They are Figs. 10 and 11 in the revised manuscript) were redrawn to show much useful information to readers. The controversial "apparent CFC age" was

deleted and then was shown in a more proper way to indicate the modern precipitation recharge.

Specific Comments

1) Please ask the help of a proof reader to help improve readability

Response: Agree. We will ask a proof reader to modify the language to help improve readability.

2) L11: Why is it crucial ? Please explain or leave that out.

Response: Agree. As we know that groundwater are the very important and even the only water resources for the local residents in the Manas River Basin of arid northwest China (It is also the same in many other arid areas around the world). Two recharge sources which, modern precipitation recharge and paleo–meteoric recharge, are known as the main sources for the groundwater in arid areas. Knowledge the recharge sources is thus can provide correct guidance for extraction of groundwater. In addition, groundwater renewability is also an important factor for guiding extracting groundwater. Groundwater residence time reveals information about water storage, mixing and transport in subsurface water systems, which has contributions to the rational development of groundwater (Cartwright et al., 2017; Custodio et al., 2018; Dreuzy and Ginn, 2016; Leray et al., 2016). Therefore, we tend to retain the sentence (L11) in consideration of the explanation mentioned above.

3) L15: “indicating the rainfall recharge...” You mean that the young water component is higher than in samples with lower tritium activity.

Response: The inexact expression generates misunderstanding. We mean the very high ^3H activities (41.1–60 TU) of groundwater indicate rainfall recharge during the nuclear bomb.

4) L29: The title of this section is not very telling, and this is not really what the study is about, is it?

Response: Agree and changes made. The title was changed to “Introduction”.

5) L33: “may be renewable”. What do you mean ? Something about short turnover time ?

Response: Changes made. We think that “renewable” may means groundwater being recharged by other water resources after extraction (or discharge). The ambiguous sentence was changed to “It is particularly crucial in the alluvium aquifer where the fresh groundwater renewability is generally strong (Huang et al., 2017) and thus as potable water resources in the arid areas, as well as more vulnerable to anthropogenic contaminants and ...”.

6) L37-39: Rewrite the entire sentence.

Response: Agree and changes made. The entire sentence was changed to (RL35–40): “Since the residence time distributions in the subsurface water systems cannot be empirically measured, a commonly used approach is parametric fitting of trial distributions to chemical concentrations (Leray et al., 2016; Suckow, 2014). The widely used lumped–parameter models (LPMs; Małoszewski and Zuber, 1982; Jurgens et al., 2012), which commonly assume that the hydrologic system is at a steady–state, have been applicable to subsurface water systems containing young water with modern tracers of variable input concentrations (e.g. seasonably variable stable isotope ^2H and ^{18}O , tritium, and ^{85}Kr ; Cartwright et al., 2018; McGuire et al., 2005; Morgenstern et al., 2015; Stewart et al., 2010).”

7) L38: “and may be inferred”. You mean “must be inferred”.

Response: Agree and changes made. Yes, the residence time distribution cannot be empirically measured. See “Specific Comments Response 6”.

8) L39: “at” steady-state, not “in” steady-state.

Response: Agree and changes made. “in” was replaced by “at”.

9) L40: “Three types of transit time”. You mean three time windows ?

Response: Changes made. The sentence was changed to (RL40–41): “The groundwater residence time tracers can be classified into three types possessing the different time scales”.

10) L46: It’s not variability, rather time span.

Response: Agree and changes made. “variabilities” was replaced by “time span”.

11) L48: “in a similar function with” should read “in a similar way to”.

Response: Agree and changes made. “in a similar function with” was replaced by “in a similar way to”.

12) L51: replace “over” with “than”.

Response: Agree and changes made. “over” was replaced by “than”.

13) L54: You mean that increasing transit time through the aquifer leads to increasing mineralisation.

Response: No change made. That “increasing transit time through the aquifer leads to increasing mineralisation” is reasonable under certain circumstance. However, groundwater mineralisation is a very complex process, which contains mineral dissolution, precipitation, hydrolysis, evaporation concentration, water–rock interaction, and so on (Ma et al., 2018). In addition, water corrosion, solubility of rocks, and groundwater runoff will also have strong impacts on mineralization (Shen et al., 1993). Thus, increasing transit time not necessarily leads to increasing mineralization. The sentence and cited literatures (L54) in the manuscript tend to show that there are some correlations between the transit time and mineralisation in some situations. Nevertheless, the correlation is still unclear in the Manas River Basin, and that is right the research objective in our manuscript.

14) L55: Please explain why tritium is “the only true age tracer”, namely because it is part of the water molecule.

Response: Agree and changes made. Yes, it is. The sentence was changed to (RL56): “Tritium (^3H), a component of the water molecule with a half-life of 12.32 yrs, is the only true age tracer for waters (Tadros et al., 2014)”.

15) L56 (entire paragraph): Why mention the southern hemisphere at all, since the study takes place in the northern hemisphere ? This is useless information.

Response: Agree and changes made. We have deleted the statements concerning the Southern Hemisphere ^3H activities and relevant references (L61–66).

16) L66: “may be used to estimate MTTs” should read “must be used to estimate MTTs”. And explain why (non-unicity problems...).

Response: Agree and changes made. Yes, we totally agree with your opinion. As the mean annual ^3H activities peaks were several hundred times natural levels in the Northern Hemisphere due to the atmospheric thermonuclear tests in the Northern Hemisphere between the 1950s and 1960s, the nowadays rainfall ^3H activities are still affected by the tail-end of the bomb-pulse in the Northern Hemisphere. Groundwater residence time cannot be assigned based on measurement of a single sample ^3H activity.

The sentence was changed to (RL62–65): “Thus, measurement of a single sample ^3H activity does not allow the groundwater residence time to be assigned (Cook et al., 2017), and time series ^3H measurements must be used to estimate MRTs in the Northern Hemisphere by LPMs (Han et al., 2007; Han et al., 2015).”

17) L69: You are confusing residence time and degradation half-life. The residence time of the CFCs in the atmosphere is no different from that of tritium or any other tracer. The difference lies in their half-lives (degradation for CFCs, decay for tritium), which are very long for the CFCs.

Response: Agree and changes made. We agree with your opinion on the “residence time” and “degradation half-lives” for the atmospheric environmental tracers (e.g. CFCs and ^3H). CFCs degrade slowly in the atmosphere and have relatively long degradation half-lives, while the decay half-life for ^3H is much shorter with 12.32 yrs. The input function for CFCs is not area-specific as is the case with ^3H . Therefore, the atmospheric concentrations for CFCs are uniform over large areas (Cartwright et al., 2017; Cook et al., 2017), while that for ^3H often varies across latitudes and between seasons (Tadros et al., 2014).

The sentence was changed to (RL66–67): “Contrasting to ^3H , CFCs degrade slowly in the atmosphere and have relatively long degradation half-lives, which permits their uniform atmospheric distributions over large areas, ...”.

18) L78-82: Please rephrase the entire sentence.

Response: Agree and changes made. The sentence was rewritten as follows (RL76–81): “Furthermore, inaccurate groundwater MRTs might be obtained when the CFCs entrapped excess air or contaminated in urban environments due to the CFCs degradation and/or contamination. For example, MRTs are much less than actual values if CFCs inputs are entrapped excess air in the unsaturated zone during recharge (Cook et al., 2006; Darling et al., 2012) or contaminated in urban and industrial environments (Carlson et al., 2011; Han et al., 2007; Mahlknecht et al., 2017; Qin et al., 2007), and are much higher if CFCs inputs are degraded in anaerobic groundwater (most notably CFC-11 and CFC-113; Cook and Solomon, 1995; Horneman et al., 2008; Plummer et al., 2006b).”

19) L89: “Mixing [...] is particularly true...”. You don’t know that, it’s a probable hypothesis!

Response: Agree and changes made. Though the previous study (Ma et al., 2018) have confirmed that the groundwater lateral-flow mixing is common in the alluvium aquifers of the Manas River Basin, more work and studies should be conducted to confirm the mixtures within the aquifers and long-screened wells. On the other hand, groundwater mixing can occur within the aquifer itself (Cook et al., 2017) as well as occur during pumping from long screened wells. However, for the case study of the Manas River Basin, mixing ascribed to pumping from long-screened wells has not been demonstrated before.

The sentence was changed to (RL88–89): “The hypothesis mixing within the aquifers and pumping from the long–screened wells is particularly common in the faulted–hydraulic drop alluvium aquifers of the Manas River Basin in the arid northwest China (Fig. 1).”

20) L93-95: “The MTTs that impacted....”. This sentence makes no sense. Rewrite.

Response: Agree and changes made. The sentence was changed to (RL91–93): “The MRTs that result from a deep unsaturated zone (water table depth is 180 m) and contrasting geological settings (a level difference of 130 m hydraulic drop caused by the thrust fault) are still insufficiently recognized in the alluvium aquifer (Fig. 1c).”

21) L106: “with totally length” should read “with total length”.

Response: Agree and changes made. “totally” was replaced by “total”.

22) L107: “was intermittent activity” should read “was intermittently active”.

Response: Agree and changes made. “intermittent activity” was replaced by “intermittently active”.

23) L110: So the different aquifers are all fractured rock aquifers.

Response: Yes, the substantial fractures and fissures of rock aquifers in the mountain area allow groundwater migration to be possible. The literatures by Cui et al. (2007) and Zhou (1992) were cited in the relevant discussion.

24) L114: “is macroscopically similar”. What do you mean with “macroscopically” ?

Response: Changes made. The sentence was changed to (RL113): “The groundwater flow direction is consistent with the Manas River flow direction.”

25) L120: “is as large” should read “is as deep”.

Response: Agree and changes made. “large” was replaced by “deep”.

26) L124: How many samples were taken altogether ? And are there any information concerning screening depth and size (fully penetrating wells or not) ? This is important information to guide model choice.

Response: Agree and changes made. The important information concerning sample numbers as well as screening depth and size were added in the revised manuscript.

The sentences were changed to (RL123–127):

“In total, 29 groundwater (pumped from fully penetrating well, of which 3 are from spring and 3 are from the artesian well) were collected along the Manas River motion during June to August, 2015 (from G1 to G29 in Table 1 and Fig. 2). Groundwater were separated into three clusters including the upstream groundwater (UG, south of the Wuyi Road), midstream groundwater (MG, area between the Wuyi Road and the West main canal–Yisiqi), and downstream groundwater (north of the West main canal–Yisiqi) based on the hydrochemistry and stable isotope data.”

27) L126: What was the rationale for separating the samples into three groups ? For instance, why is G13 MG while G26 is DG ? DG seems like the downgradient boundary. Did you use the piper diagrams to separate the samples ?

Response: Agree and changes made. Yes, you are exactly right that the three groundwater groups are based on the result of hydrochemistry and stable isotope. Yisiqi (Fig. 2) is the dividing line where there is an obvious change for the hydrochemistry and stable isotope along the Manas River motion (Fig. S2 from Ma et al., 2018). We also use piper diagrams to separate the samples (Fig. 4).

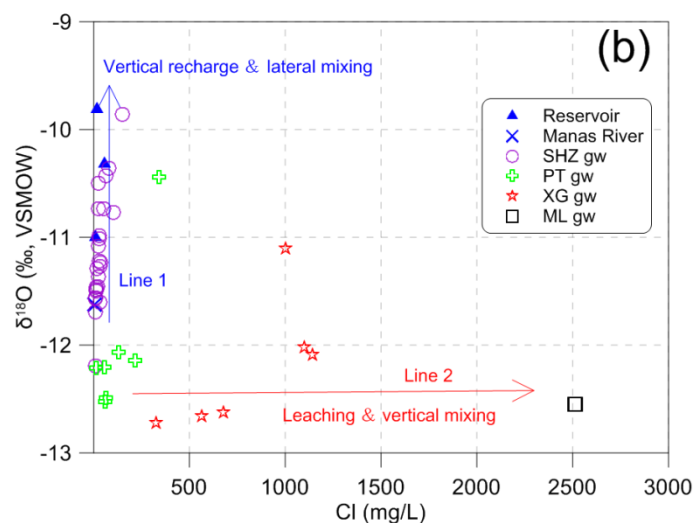


Figure S2. Relationship between $\delta^{18}\text{O}$ and Cl concentrations for waters as a means to differentiate hydrogeological processes in the Manas River Basin. The purple circles represent the upstream and midstream groundwater, and green, red, and black legends represent downstream groundwater (from Ma et al., 2018).

28) L152: “were followed” is used multiple times but should read “after” or “following”.

Response: Agree and changes made. We have checked the manuscript carefully and seriously have corrected the erroneous words. “were followed” was changed to “after” and we insisted on the “after” throughout the manuscript.

29) L173: “refers” is not the proper verb. Use “depends” for instance.

Response: Agree and changes made. The sentence was changed to (RL174–175): “The difference between the local and global background atmospheric CFC concentrations (Northern Hemisphere), which we intitule as CFC excess, varies largely based on the industrial development.”

30) L179: What are low latitude countries ?

Response: Changes made. The vague expression was deleted. As the atmospheric CFC concentrations are affected by the industrial development, the time series of Northern Hemisphere atmospheric CFC concentrations were widely treated as the background values for the underdeveloped areas (it is also applied to the Manas River Basin in our manuscript).

31) L189: The entire procedure is correctly explained, and also the fact that “apparent age” implies piston-flow transit time distribution, but why use apparent age in the first place ? Piston-flow is one model among many, as the authors explain later in the manuscript. Furthermore, the entire concept of “age” is problematic and should be replaced by mean transit time or mean residence time (for an in-depth discussion, see Suckow, The age of groundwater-Definitions, models and why we do not need this term, Applied Geochemistry 50, 2014, 222-230).

Response: Agree and changes made. The erroneous “apparent CFC age” has been deleted. The concept of “age” is replaced by mean residence times and estimated by lumped-parameter models with the exponential-piston flow model, dispersion model, and exponential model, which are discussed in Section 4.3.1. In addition, the groundwater recharge sources (e.g. to distinguish the modern and paleo-meteoric recharge features) identified by the CFCs are also discussed in Section 3.3.1 and Section 4.2.2.

32) L194: What do you mean by “closed system” ? Physically bounded ?

Response: Changes made. As we know that the measured tracer output concentration in the groundwater is compared to its historical input using the convolution integral, in which the system response function is selected based on the different hydrogeological subsurface flow systems (Jurgens, et al., 2012; Małoszewski and Zuber, 1982; the details of calculation is shown in “Response 6”). The tracer output concentration can be measured from the spring, stream, well, and so on. There is no absolutely physical boundary for the subsurface flow system. Therefore, the ambiguous statements “closed system” is not rigorous and thus is deleted. See Section 3.3.3 (Response 4).

33) L208: “were given below as transit time distribution function” should read “were selected and are given below”.

Response: Agree and changes made. This sentence was deleted and more detailed interpretations were added (See Section 3.3.3 in the Response 4).

34) L219: You should also explain here how you planned to choose between these competing models.

Response: Agree and changes made. The system response function is selected based on the hydrogeological settings and the sampling position of tracer output concentration measured. We explain the basis on the choice of the models. See Section 3.3.3 (Response 4).

35) L235: Why present an equation you will not be using for lack of appropriate data ?

Response: Agree and changes made. We have deleted this equation and the accompanying description.

36) L240: This is true for the piston-flow model only ! See Custodio et al. for details.

37) L250: So the entire paragraph boils down to using literature values for the initial ^{14}C activity. Make it shorter and to the point.

Response: Agree and changes made. Custodio et al. (2018) indicated that the calculated groundwater ^{14}C age is really an apparent age due to the mixture of waters with different transit times, which was also indicated by other researchers (e.g. Cook et al., 2017; Suckow, 2014). Therefore, we deleted some unnecessary description and unimportant information in the text. Some sentences were reduced and revised. The revised content can be seen in Section 3.3.2 (RL193–218):

“Carbon-14 (^{14}C , with a half-life of 5730 yrs) activity in groundwater is often used to estimate groundwater ages over time periods of approximately 200 and 30 000 yrs, and to determine the recharge from mixing waters in various climate conditions (Cook, et al., 2017; Custodio et al., 2018; Huang et al., 2017). The calculation of ^{14}C ages may be complicated if groundwater dissolved inor-

ganic carbon (DIC) is derived from a mixture of sources or/and the ^{14}C originating from the atmosphere or soil zone is often significantly diluted by the dissolution of ^{14}C -free carbonate minerals in the aquifer matrix and biochemical reactions along the groundwater flow paths (Clark and Fritz, 1997). While only minor carbonate dissolution is likely, determination of groundwater residence times requires ^{14}C correction to be taken into account (Atkinson et al., 2014). When dissolution of carbonate during recharge or along the groundwater flow path may dilute the initial soil CO_2 , $\delta^{13}\text{C}$ can be used to trace the process (Clark and Fritz, 1997). An equation for the reaction between carbon-dioxide-containing water with a carbonate mineral is commonly written as (modified after Pearson and Hanshaw, 1970):



where $\delta^{13}\text{C}_{\text{carb}}$ is the dissolved carbonate $\delta^{13}\text{C}$ value (approximately 0; Clark and Fritz, 1997), and $\delta^{13}\text{C}_{\text{DIC}}$ is the measured $\delta^{13}\text{C}$ value in groundwater.

Depending on knowing the measured ^{14}C activity after adjustment for the geochemical and physical dilution processes in the aquifer (without radioactive decay), then the groundwater apparent ^{14}C ages (t) can be calculated from the following decay equation:

$$t = -\frac{1}{\lambda_{^{14}\text{C}}} \times \ln \frac{a^{14}\text{C}}{a_0^{14}\text{C}}, \quad (1)$$

where $\lambda_{^{14}\text{C}}$ is the ^{14}C decay constant ($\lambda_{^{14}\text{C}} = \ln 2/5730$), and $a^{14}\text{C}$ is the measured ^{14}C activity of the DIC in groundwater. As mentioned above, the estimated ages are really apparent ages due to the mixture of waters with wide range of ages (Custodio et al., 2018; Suckow, 2014).

Previous studies in the arid northwest China (Edmunds et al., 2006; Huang et al., 2017) have concluded that a volumetric value of 20 % “dead” carbon derived from the aquifer matrix was recognized, which is consistent with the value (10–25 %) obtained by Vogel (1970). Therefore, the initial ^{14}C activity ($a_0^{14}\text{C}$) of 80 pMC is used to correct groundwater ^{14}C ages (results are shown in Table 1), despite this simple correction makes no attempt to correct the age of individual samples that may have experienced different water–rock interaction histories.”

38) L264: Check the discussion paper by Benettin et al. in review in HESS for the latest developments on the “evaporation slope”.

Response: Agree and changes made. We agree with the opinion on the evaporation slope that a steeper slope would be obtained due to the large source variability under different meteorological conditions in different seasons indicated by Benettin et al. (2018). As we know that water fractionation is affected by various factors, like the water surface temperature, air humidity, wind speed, and so on, but we can always obtain a linear trend from a source water with low salinity (Ma et al., 2015). However, in practice, the evaporation trend line is often obtained from various sources water, which is often not true evaporation line. In our manuscript, the evaporation line is estimated according to the surface water (ditch water, reservoir water, and Manas River water), which are all collected in summer of 2015. The ditch, reservoir and Manas River are always connected to one another (Fig. 1b), and all are recharged from the mountain areas in the same season. Though there are minor differences of the water sources for the surface water, the linear trend obtained based on these surface water may have implications for the surface water evaporation. Therefore, we tend to use the evaporation slope and to add some statements (as follows) for the rationality in the revised manuscript (RL286–290):

“A recent study (Benettin et al., 2018) has indicated that the evaporation line that obtained from various sources water is often not the true evaporation line. Our surface water were all collected in summer of

2015 and were recharged from the mountain areas in the same season, though they were collected from the different point (ditch water, reservoir water and Manas River water), the linear trend obtained based on these surface water may have implications for surface water evaporation.”

39) L274: The entire paragraph is too short and should explain clearly the approach adopted to calculate “ages” from the tracer data (model and model parameter choices !). I strongly advise against using binary mixing diagrams, and encourage the authors to use a multi-tracer modelling approach trying to find a single optimum or optimal parameter regions for the different tracers.

Response: Changes made. The entire paragraph was deleted. We reorganized the structure of the Section “4. Results and discussion”, rewrote some contents and deleted some incorrect expressions, and added statements concerning the model choice in Section “3. Materials and methods”. However, we tend to retain the binary mixing diagrams (the mean residence times with different models have been deleted in the diagrams) to explain the young and old groundwater mixing features and to identify the impact of human activity on the groundwater.

40) L277: The paragraph on “apparent age” makes no sense for the reasons given above. I disagree with the proposition that “they [apparent ages] provide a good first approximation for groundwater age”. There is no reason to prefer the piston-flow model which is implied by the “apparent age” concept over other models. This argument has been for years a lazy way to skip responsibility in choosing one model based on knowledge of the hydrogeological situation and sampling.

Response: Agree and changes made. The “apparent CFC ages” has been deleted. In addition, we tend to discuss the modern precipitation recharge features and to distinguish the modern and paleo-meteoric recharge using CFCs in Section 4.2.2 (RL344–414):

“It is seen from Table 1 that groundwater with well depths between 13 and 150 m contain detectable CFC concentrations ($0.17\text{--}3.77\text{ pmol L}^{-1}$ for CFC–11, $0.19\text{--}2.18\text{ pmol L}^{-1}$ for CFC–12, and $0.02\text{--}0.38\text{ pmol L}^{-1}$ for CFC–113) both in the upstream and midstream areas, indicating at least a small fraction of young groundwater components (post–1940). The highest concentration was observed in the UG (G3), south of the fault, median and the lowest were respectively observed in the west and east bank of the ‘East main canal’ in the MG, north of the fault. In the midstream area (Fig. 2), CFC concentrations generally decrease with well depth at the south of reservoirs (G25, G8, and G9), while increase with well depth at the north of reservoirs (G15 and G16), which might indicate the different groundwater flow paths (e.g., downward or upward flow directions).

Groundwater aerobic environment (Table 1, DO values vary from 0.7 to 9.8 mg L⁻¹) make CFC degradation under anoxic conditions unlikely. Nevertheless, CFC–11 has shown a greater propensity for degradation and/or contamination than CFC–12 (Plummer et al., 2006b). Therefore, we use the CFC–12 to interpret the modern groundwater recharge in the following discussions. The estimated CFC atmospheric partial pressures and possible recharge year are shown in Table 2 and Fig. 3. The UG (G3) CFC–113 and CFC–12 both indicate the 1990 precipitation recharge (Table 2), probably indicating piston flow recharge in the upstream area. The MG CFC–11–based modern precipitation recharge agreed within 2–8 yrs with that based on CFC–12 concentrations, while that the CFC–113–based recharge were much 4–11 yrs later than that based on CFC–11 and CFC–12 concentrations, indicating mixtures of young and old groundwater components recharge in the midstream area. The latest groundwater recharge is in the upstream area (G3 with 1990 recharge), which is most likely due to the shortest flow paths from recharge sources compared to the piedmont groundwater samples in

the midstream area.

Groundwater G5 and G7, which are located in the East 'East main canal' in the midstream area with much shorter distance than G15 and G16 in the reservoir north, show that the modern recharge are much earlier than G15 and G16 (Table 2). This could be explained by the lower groundwater velocities in the East 'East main canal', where the hydraulic gradient (Fig. 2) is much smaller than the West. Furthermore, it can be seen from Table 2 and Fig. 2 that groundwater recharge were much earlier with well depth increasing from 48 to 100 m at the reservoir south (G25, G8 and G9), while that in the reservoir north were much later with well depth increasing from 23 to 56 m (G15 and G16). The different trends for the relationship between groundwater recharge year and well depth might be ascribed to the different flow paths among the two sites (e.g., reservoir south and north).

Comparing CFC concentrations has provided a powerful tool to recognize samples containing co-existence of young (post-1940) water with old (CFC-free) water (Han et al., 2007; Han et al., 2012; Koh et al., 2012) or exhibiting contamination or degradation (Plummer et al., 2006b). The cross-plot of the concentrations for CFC-113 and CFC-12 (Fig. 7a) demonstrates that all of the groundwater can be characterized as binary mixtures between young and older components, though there is still room for some ambiguity around the crossover in the late 1980s (Darling et al., 2012). As shown in Fig. 7a, all of the MG samples are located in the shaded region, representing no post-1989 waters recharge. The UG (G3) sample is clearly quite 'modern' and seems to be recharged in 1990 through piston flow or mixed by the old water and post-1989 water. Using the method described by Plummer et al. (2006b) with the binary mixing model (BMM), the fractions of young water vary from 12 to 91 % (Table 2) for the MG samples with the relatively low young fractions of 12 and 18 % in the east bank of the 'East main canal' of MG samples (G5 and G7). These two well water table depths are more than 40 m, probably indicating a relatively slow and deep circulated groundwater flow. This hypothesis is also suggested by lower DO ($3.7\text{--}4.6\text{ mg L}^{-1}$; Table 1) and nitrate concentrations ($8.6\text{--}9.5\text{ mg L}^{-1}$ from Ma et al., 2018) and relatively much smaller hydraulic gradient (Fig. 2). Furthermore, as high as 100 % fraction of young water for G3 sample is obtained with the recharge water from 1990, or 87 % fraction is obtained by the binary mixture between post-1989 water and old water (Table 2). The quite 'modern' recharge for G3 sample is likewise explained by its highest DO (9.8 mg L^{-1} ; Table 1) and relatively low nitrate concentration (7.9 mg L^{-1} from Ma et al., 2018), which represent the contribution of high-altitude recharge rather than the old age water.

CFC contamination and sorption in unsaturated zone during recharge have great influence on interpretation of groundwater recharge. Points lying off the curves in the cross-plot CFC concentrations may indicate that contaminations from the urban air with CFC compounds during sampling (Carlson et al., 2011; Cook et al., 2006; Mahlknecht et al., 2017) or degradation/sorption of CFC-11 or CFC-113 (Plummer et al., 2006b). Figure 7 demonstrates that the urban air with CFC compounds contaminations, which generally cause elevated CFC concentrations than the global background atmospheric CFC concentrations (Northern Hemisphere), are unlikely. Elevated CFC concentrations have been reported in the air of urban environments such as Las Vegas, Tucson, Vienna and Beijing (Barletta et al., 2006; Carlson et al., 2011; Han et al., 2007; Qin et al., 2007), contrary to that in the arid northwest China (Barletta et al., 2006). Hence, the anomalous CFC-11/CFC-12 (Fig. 7b) ratios plotting off the model lines might be ascribed to the sorption in the unsaturated zone during recharge rather than the degradation of CFC-11 (Cook et al., 2006; Plummer et al., 2006b) under anoxic conditions (Table 1, DO values vary from $0.7\text{ to }9.8\text{ mg L}^{-1}$). Nevertheless, the small deviations (Fig. 7b) indicate that the hypothesized sorption rate was low. Higher CFC sorption rate with high clay fraction and high organic

matter in soils have been proved (Russell and Thompson, 1983), and vice versa (Carlson et al., 2011). Therefore, the hypothesis of a low sorption rate due to the low clay fraction and low organic matter content in the intermountain depression and the piedmont plain (Fig. 1c) seems reasonable.

The time lag for CFCs transport through the thick unsaturated zone (Cook and Solomon, 1995), as well as degradation especially for CFC-11 is being common in the anaerobic groundwater (Horneman et al., 2008; Plummer et al., 2006b), which both are important consideration when interpretation of groundwater recharge using CFC concentrations. The time lag for CFCs movement both in dissolved and gas phases through deep unsaturated zone. The time lag for the diffusive transport of CFCs through deep unsaturated zone in simple porous aquifers, a function of the tracer solubility in water, tracer diffusion coefficients and soil water content (Cook and Solomon, 1995), have been widely proved (Darling et al., 2012; Qin et al., 2011). The small differences in CFC-11 and CFC-12 recharge years (Table 2) demonstrates that the time lag would be short in the faulted-hydraulic drop alluvium aquifers with deep unsaturated zone (Fig. 1c). Previous studies in the Manas River Basin (Ma et al., 2018; Wang, 2007; Zhou, 1992) showed that groundwater were mainly recharged by the river fast leakage in the upstream area and piedmont plain, where the soil texture is consisted of pebbles and sandy gravel (Fig. 1c), which makes us to assume that the unsaturated zone air CFC closely follows that of the atmosphere and thus the recharge time lag through the unsaturated zone is not consideration.”

41) L297: “which confirms”. A performative statement confirms nothing. You are supposing this is the case !

Response: Agree and changes made. We have revised this sentence: “..., which makes one to assume that ...”

42) L317: Shortly explain the method used to estimate the tritium input (linear regression ? And how long were the time series used ?). The reference to Han et al. is not very useful as the authors of that paper themselves refer to an IAEA publication without further explanations.

Response: Agree and changes made. The explanation is seen in “Response 7”. We added short statements in the revised manuscript as follows (RL236–239):

“The historical precipitation ^3H activity in Urumqi station (Fig. 4) was reconstructed from the available data in the International Atomic Energy Agency (IAEA) using a logarithmic interpolation method. Precipitation ^3H activities between 1969 and 1983 at Hongkong and Irkutsk with different latitudes were used (data is available at <<https://www.iaea.org/>>).”

We think that the reconstructed tritium (^3H) activity in Urumqi station can be added as the supplement if necessary.

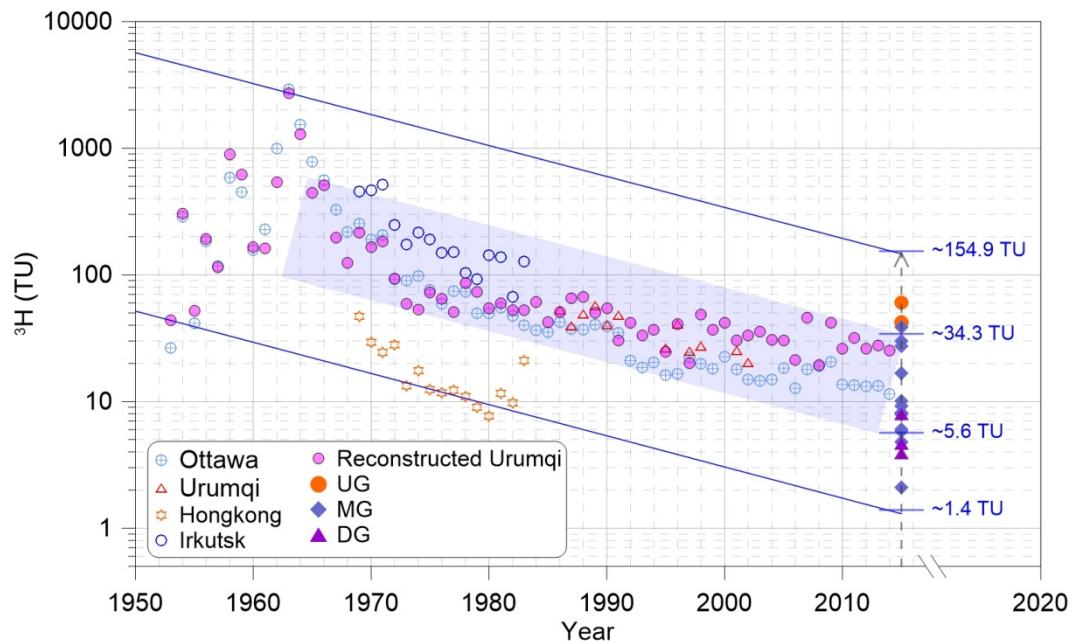


Figure 6. Tritium concentration (TU) of groundwater water samples of upstream groundwater (UG), midstream groundwater (MG), and downstream groundwater (DG). Time series of tritium concentration in precipitation at Ottawa, Urumqi, Hongkong, and Irkutsk were obtained by GNIP in IAEA (<https://www.iaea.org/>). The blue dashed lines and shaded field were drawn using the half-life (12.32 yrs) of tritium decayed to 2014. (It is Fig. 4 in the revised manuscript)

43) L318: A background of 31 TU is very high compared to Western Europe (about 6 TU). How come ?

Response: As shown in Fig. 4 and reconstructed tritium ^3H activity in Urumqi, the estimated modern precipitation ^3H activity are indeed extremely high (mean value of 31.55 TU between 2000 and 2015). We speculate that the very high precipitation ^3H activity may be ascribed to both the Chinese atmospheric nuclear tests (around 350 km away to Manas River Basin) from 1964 to 1974 (Zhou, 1992) and the continental effect (Tadros et al., 2014), where the Manas River Basin is more than 3500 km far away from the western pacific.

44) L413: What do you mean by “serious” ?

45) L414: “tend toward more discrete with their increase”. I do not understand this part of the sentence.

Response: Changes made. The spelling error “serious” was changed to “series”. The sentence was changed to (RL479–480): “Figure 11 shows that different LPMs yield different MRTs for the same time series ^3H activities and CFC concentrations. The MRTs obtained from different LPMs tend to more discrete among each model with the increase of MRTs.”

46) L448: The paragraph on hydrochemistry is not bad, but underdeveloped and bad organized. State again what you’re looking for first. A good correlation between hydrochemistry and “ages” calculated using some of the TTD models might be a way to constrain or guide model choice, but the authors do not really state that explicitly, although that would be interesting and relatively new.

Response: Yes, we agree with your opinion on the relationship between hydrochemistry and “ages”

calculated using the TTD models. Thanks very much, the suggestion of using hydrochemistry to guide model choice is definitely interesting and relatively new. We will bear in mind in the future studies in the Manas River Basin. But we think that we tend to seek the relationship between groundwater hydrochemistry and mean residence times firstly in this manuscript.

47) L491: The entire chapter 4.5.1 makes no sense. You must first decide which model is the most appropriate, and then calculate metrics such as mean transit time, young water fraction, etc... You cannot both calculate water fractions using a binary mixing strategy (assuming piston-flow) AND later use an EPM. The same remark applied to chapter 4.5.2.

Response: We agree with your opinions that “first decide which model is the most appropriate, and then calculate metrics such as mean transit time, young water fraction, etc...”. We think that the application of mixing models (even combining two lumped-parameter models) is a good method to quantitatively analysis groundwater mixing ratios, and this method is also used more and more widespread. However, CFCs are also good tracers to distinguish groundwater mixing from the different recharge sources, like to recognize modern and paleo-meteoric recharge features, to distinguish the fraction of young groundwater. We have totally revised the entire chapter 4.5.1 and reorganized the structure of the manuscript (See “Response 3”). The erroneous phrases and expressions including apparent CFC age and groundwater ages estimated from the piston flow model have been deleted.

48) L498: “no post-1988.5”. Please round this off...

Response: Agree and changes made. “no post-1988.5” was changed to “no post-1989”. “1988.5” was changed to “1990”.

49) L509: Why do you treat “apparent age” as some kind of different measure of transit time than MTTs “estimated from the EPM” ? This is doing the analysis the wrong way around. First find a way to select a model, then discuss the obtained “ages” instead of hypothesizing on tons and tons of different “ages” that are meaningless because they were obtained disregarding the actual situation. This leads nowhere.

Response: Agree and changes made. The erroneous sentence and term “apparent CFC ages” been deleted. We reorganized the structure and revised the relative content to tried to present a clear roadmap to readers.

50) L541: Before engaging in complicated mixing scenarios, you should first try to find one model and one parameterisation that fits both the CFCs, tritium and carbon 14. Only if that search does not succeed should additional mixing be introduced. Please note that the binary mixture approach by Plummer et al. is only one way of doing so, and a particularly weak one at that because it assumes per default a piston-flow distribution of transit time of each component (other models can be integrated, but it becomes quickly very cumbersome). Another way to include the mixing of different reservoirs is to combine models (say two exponential models, each representing one distinct source) following Piotr Maloszewski and coworkers or Mike Stewart and Uwe Morgenstern. Binary plots such as those of figure 11 suffer from the limitation that you have to recalculate the mixing line for each parameterisation of each model, and they cannot really replace a multi-tracer lumped-parameter modelling approach, where the objective function reduces simultaneously the prediction error of all tracers.

Response: Thanks very much for the precious suggestions you offered. We believe that the modelling

approaches are commendable to interpret complicated mixing scenarios quantitatively. For example, the binary parallel lumped-parameter model (Morgenstern et al., 2015; Stewart et al., 2017), the binary mixing model that followed by two response function models (Jurgens et al., 2012), and the mathematical solutions that indicating the changes in water reserve in the relatively large systems with wide range of residence times (Custodio et al., 2018; Custodio and Custodio-Ayala, 2014). In addition, tracer-tracer binary plots were also the widely used method to determine the groundwater recharge mechanisms and to interpret the groundwater mixing (Cook et al., 2017; Darling et al., 2012; Han et al., 2015; Kagabu et al., 2017). We think that combining ^3H and CFCs is a good tool to distinguish the modern precipitation recharge and to indicate the groundwater mixing properties (for example mixing old with young water, mixing irrigation infiltration water and young water) in the Manas River Basin, which has not been reported before. Recharge features (e.g. modern and paleo-meteoric recharge features, and the fractions of modern recharge water) are also the essential contents in the manuscript, which will not be all on account of the mean residence times estimated by ^3H and CFCs in the revised manuscript. Further investigation work will be carried out in the Manas River Basin in the following several years (for example, time serious groundwater CFCs from some given wells will be measured). Deeper analysis concerning groundwater mixing obtained by the combination of two models using lumped-parameter models (binary mixing model) and to explore the changes in water reserve will be good choices in the next studies in the Manas River Basin.

51) L562: Solutions are obtained, explanations are devised.

Response: No change made.

52) L572: What are mixing rates ? You mean mixing ratios ?

Response: Agree and changes made. "rates" was changed to "ratios"

53) L575: "The thrust faults were found to play a paramount role on groundwater flow path". There are not conclusions, but hypotheses very weakly suggested by the analysis of the environmental tracers, which is itself very shaky. I hardly call that evidence. Please refrain from drawing conclusions if the data necessary to test hypotheses is not available (as is the case here).

Response: Agree and changes made. This sentence has been deleted. To make the conclusions to be more clear and well-founded, we have revised the conclusions and delete some incorrect statements. Yes, this conclusion is important but not supported by strong supporting evidences in the paper. Indeed, there are some results that show large differences on both sides of the thrust fault. For examples, there is a level difference of 130 m hydraulic drop (Fig. 1c) in the south margin in Shihezi (SHZ), ^3H activities of groundwater decrease rapidly along the Manas River motion in the north of the fault but show relatively the highest values in the south of the fault (Fig. 8). These results still can not support the conclusion explicitly "The thrust fault were found to play a paramount role on groundwater flow paths ...".

The revised conclusions are as follows (RL554–570):

"In this study, the environmental tracers and hydrochemistry have enabled us to identify the modern and paleo-meteoric recharge sources, to constrain the different end-members mixing rates, and to study the mixed groundwater mean residence times in faulted-hydraulic drop alluvium aquifer systems. The paleo-meteoric recharge in a cooler climate rather than the lateral flow from the higher elevation

precipitation in the Manas River downstream area was distinguished. The quite 'modern' groundwater with young (post-1940) water fractions of 87–100 % was obtained, indicating small mixing degree in the south of the fault. The short mean residence times (19 yrs) along with the higher NO_3^- concentration (7.86 mg L^{-1}) than natural groundwater (5 mg L^{-1}) in the south of the fault (headwater area) imply invasion of modern contaminants, which should arouse people's attention. Large amplitudes of mixing rate varying from 12 to 91 % were widespread in the north of the fault due to the varying depth of long-screened boreholes or within the aquifer itself. Furthermore, the large water table fluctuations during groundwater pumping, vertical recharge through the thick unsaturated zone, and young water mixtures in different decades highlight the mixing diversity. The obtained strong correlations between groundwater mean residence times and hydrochemistry concentrations allow the first-order proxy at different times to be made. In addition, our study has also highlighted that mean residence times estimated by CFCs rather than ^3H were more appropriate in the arid Manas River Basin with thick unsaturated zone."

54) L585: "due to the highly complex groundwater system...". This is no explanation at all ! Indeed, devising a conceptual model that could explain why CFC derived "ages" correlate well with mineralisation while tritium derived "ages" do not could be a useful task (but you should first redo the calculation of the "ages" as suggested above). On the one hand, the correlation between CFC12 and hydrochemistry might be an artifact, given that the area sampled is so large and hydrogeologically diverse. On the other hand, there might be some kind of systematic shift between tritium and CFC ages if differences are due to the unsaturated zone. Maybe a diffusion model using the unsaturated zone thickness might be useful. Still much work to do...

Response: The sentence has been deleted. Thanks very much for the useful suggestions. Previous studies (Ji, 2016; Wang, 2007; Zhou, 1992) have shown that the groundwater flow paths were very complicated from the mountain to the plain areas. Precipitation recharge in the ground in the mountain areas, one part groundwater discharge into spring in the south of the intermountain depression, one part groundwater discharge into stream in the mountain areas and then recharge groundwater in the intermountain depression, and one part groundwater flow in the ground through the intermountain depression. The complicated groundwater flow systems make devising a conceptual model very difficult to implement. We think that more and much detailed work should be conducted from the mountain to intermountain depression areas to find out more evidences that interpretation the conceptual flow model. In addition, the thick unsaturated zone mainly distribute in the intermountain depression and piedmont plain areas (Fig. 1c). The interpretation of CFCs in Section 4.2.2 make us to assume that the unsaturated zone air CFC closely follows that of the atmosphere and thus the recharge time lag through the unsaturated zone is not consideration. In some cases the unsaturated zone can be ignored to obtain workable solutions (Custodio et al., 2018) though the unsaturated zone is not so thick with the Manas River Basin. We think that a diffusion model using the unsaturated zone thickness is a good guide for the further studies in the arid Manas River Basin as well as in other arid basins around the world.

55) Figure 7, 8 and 9: The figures are incredibly cluttered and very difficult to read, especially figure 9 (not to mention the legend).

Response: Agree and changes made. Figures 7, 8 and 9 were redrawn as follows:

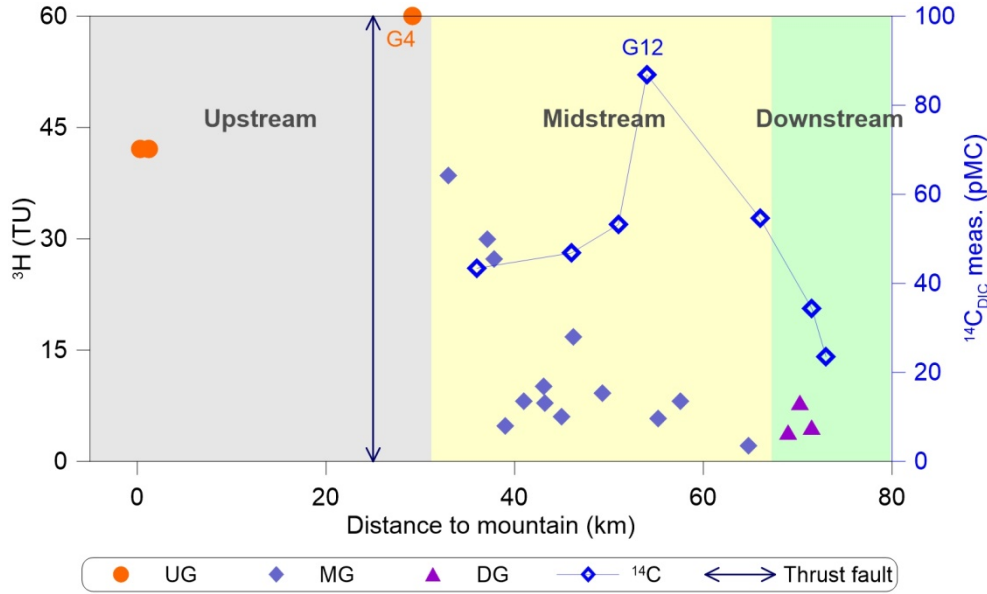


Figure 7. Distributions of ^3H and ^{14}C activities with distance to mountain. The shaded regions indicate the upstream, midstream and downstream of Manas River. (It is Fig. 8 in the revised manuscript)

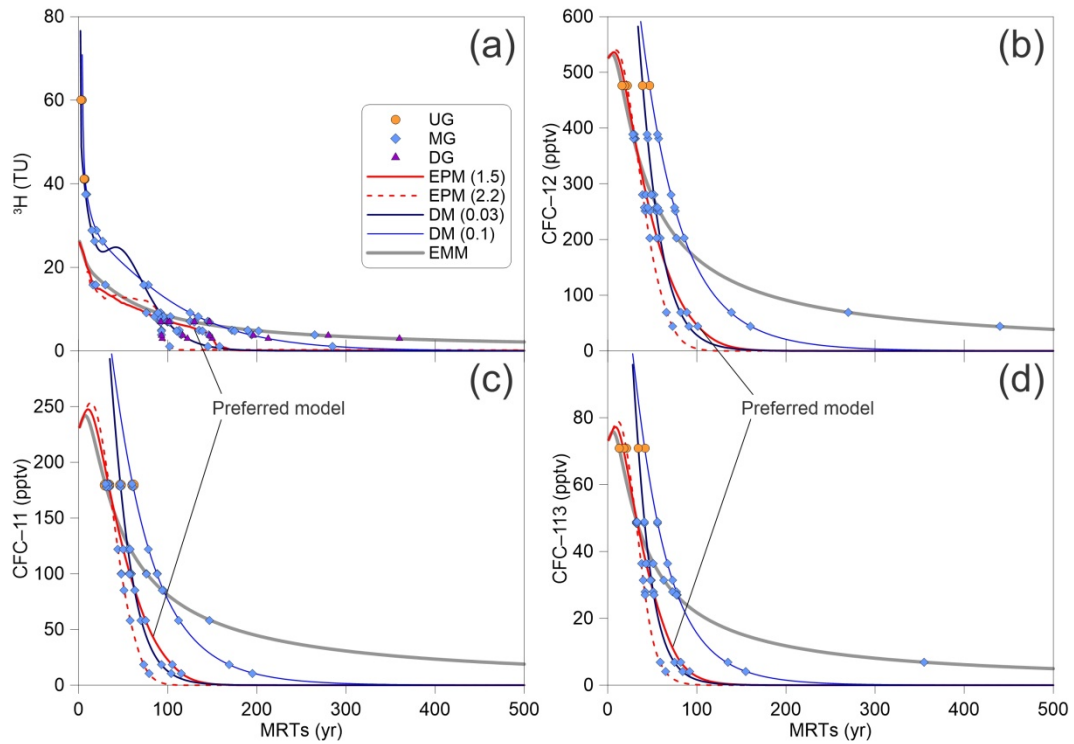


Figure 8. Tritium and CFCs (CFC-11, CFC-12 and CFC-113) output vs. mean residence times for different lumped-parameter models estimated using Eqs. (2) to (5). The input ^3H activity and CFCs concentration are using the estimated ^3H activities in precipitation in Urumqi station (Fig. 4) and Northern Hemisphere atmospheric mixing ratio (Fig. 3), respectively. (It is Fig. 10 in the revised manuscript)

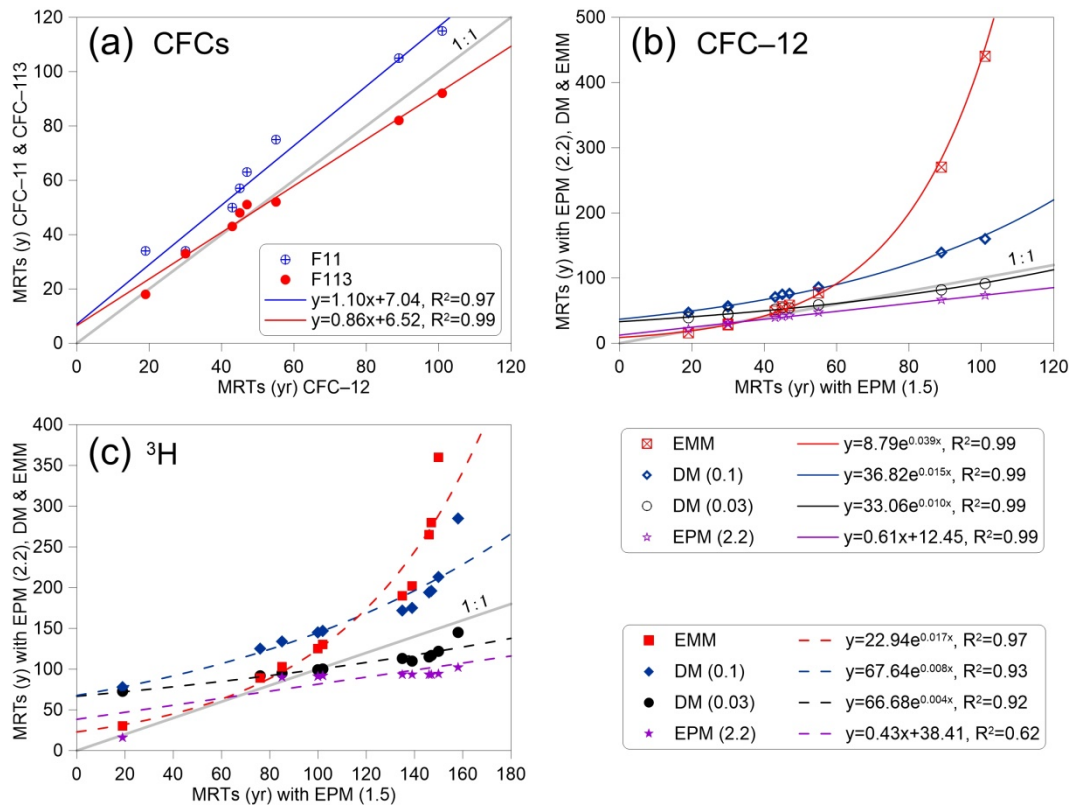


Figure 9. (a) MRTs with EPM (1.5) of CFC-12 vs. CFC-11 & CFC-113, (b) CFC-12 MRTs with EPM (1.5) vs. EPM (2.2), DM & EMM, and (c) 3H MRTs with EPM (1.5) vs. EPM (2.2), DM & EMM. (It is Fig. 11 in the revised manuscript)

56) Figure 10: Why are there so few points on each graph, since you sampled at 29 locations according to table 1 ?

Response: Figure 10 shows that the x-axis represents the CFC-12 mean residence times with EPM (1.5), and Table 1 shows that there are 9 groundwater CFC-12 samples. For the sample G15, the hydrochemistry show abnormal data with much higher concentrations of SO_4^{2-} , HCO_3^- , and electrical conductivity than that of perimeter zone. Moreover, in the plot of CFC-12 mean residence times vs. hydrochemistry, G15 sample distributes abnormal. Therefore, only 8 groundwater samples are potted in Fig. 10 (It is Fig. 12 in the revised manuscript).

57) Figure 11: As I wrote above, binary mixing diagrams rapidly tend to show their limits. After two or three mixing lines for different models are drawn, reading becomes nigh impossible. Importantly, error bars are missing for the CFCs and for tritium. I suspect that with error bars, selecting a model visually will become impossible (the lack of sensitivity is another limitation of binary mixing diagrams,).

Response: Agree and changes made. We deleted the mean residence times with different models but retain the measured CFC concentrations and 3H activities of groundwater in the diagrams.

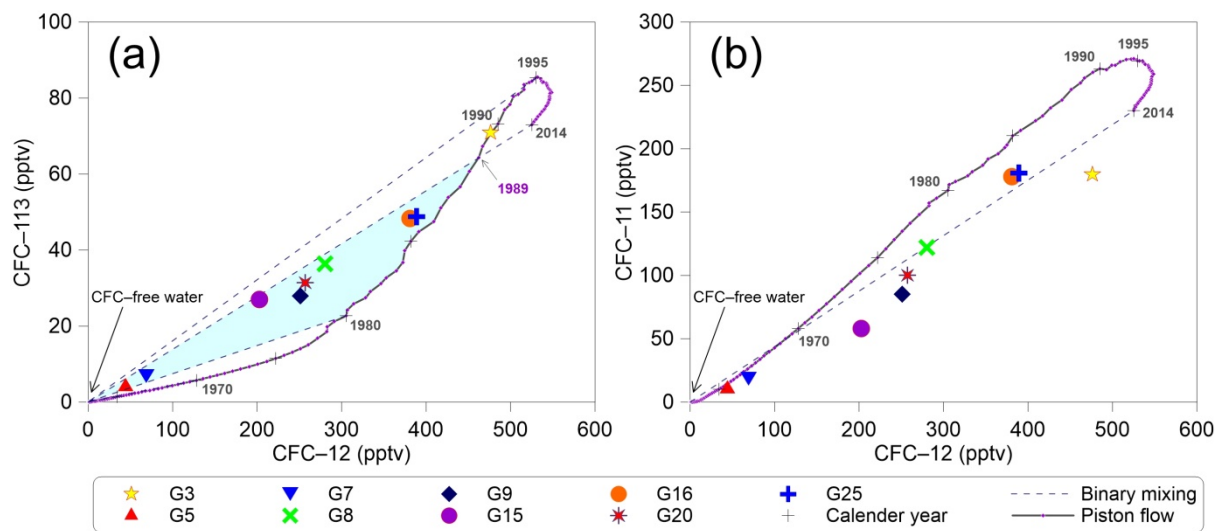


Figure 11. Plots showing relationships of (a) CFC-113 vs. CFC-12 and (b) CFC-11 vs. CFC-12 in pptv for Northern Hemisphere air. The '+' denotes selected calendar years. The solid lines correspond to the piston flow (PF) and the short-dashed lines show the binary mixing (BM). The shaded regions in (a) indicate no post-1989 waters mixing. (It is Fig. 7 in the revised manuscript)