The manuscript "Novel Keeling plot based methods to estimate the isotopic composition of ambient water vapor" presents two methods to use existing Keeling plot data not only to calculate the isotopic composition of a source (here ET), but also that of ambient water vapor delta_a. Using these two methods might provide new insights into the variability of delta_a, but a rigorous evaluation and discussion of the limitations and biases of these methods would be needed.

I cannot recommend publication of the submitted manuscript in this form. The paper lacks detailed and clear descriptions of methods and evaluation steps in many points. Due to the small number of data points that fulfilled the quality criteria, it is not clear which significance the results have and if the strong conclusions of the manuscript are justified.

Response: We thank the reviewer for the insightful and critical comments. Addressing these comments will certainly improve the quality of our manuscript. We made thorough changes through expanding the field data set, adding more descriptions of the methods and evaluation procedures as well as providing more details of the theoretical derivations. More details are in the sections bellow.

In particular I am worried about the following points:

The sparsity of the data is a major problem of the submitted manuscript. Out of four months of data, only 4 days were used for data evaluation.

Response: We thank the reviewer for the constructive comment. Our goal is to provide two new methods to estimate a parameter that is rarely estimated or measured in the past. Our key contribution of this study is the theoretical derivation of the two new methods and the data evaluation component is less important. However, we agree more field data evaluation will strengthen the manuscript. As such, we expanded our database from 4 days to 49 days including all the possible field observations during May to September, 2017 to evaluate the two methods.

Further, many data points had to be removed because they produced contradictions with the assumption. (line 197 ff). If both methods produce so many data points that are obviously wrong, it is not clear to me why we should trust the other data points. At least it needs a detailed discussion why there are roughly 50% respectively 80% of obviously wrong values. Is the used data set inaccurate and/or are limitations of the methods producing these values? Are we sure that these problems do not occur for the remaining data points?

Response: Thanks for the comments. We think the expression of “XX% of δ_a values were acceptable” generates some confusion. With 30-min interval in 49 days, we should have gotten 2352 δ_a values. However, with a filter ($\delta_{ET}<\delta_v<\delta_a$ or $\delta_{ET}>\delta_v>\delta_a$), some of the field observations do not meet the criteria. We added a new **Table 1** here to provide more details. The total δ_a is 48 for each day as each 30-minute interval should have one δ_a value. However, we do not have 48 usable δ_a for most days. On May 19th, for instance, the number of $\delta_{a(IP)}$ and $\delta_{a(IVT)}$ values passing the filter ($\delta_{ET}<\delta_v<\delta_a$ or $\delta_{ET}>\delta_v>\delta_a$) are 27 and 8, respectively. After the expansion of field observations, there are 53.8% and 4.8% of δ_a values meeting the criteria using IP and IVT method, respectively. We think the unacceptable δ_a values does not mean “wrong values”. For one thing, if the filter is not applied, 100% of $\delta_{a(IP)}$ and 5.8% of $\delta_{a(IVT)}$ is acceptable. The low percentage of acceptable $\delta_{a(IVT)}$ is mainly because only 5.8% of the 30-min intervals obey the precondition of $k_1k_2<0$. For another, if the filter is not applied, the linear regression between $\delta_{a(IP)}$ and $\delta_{a(IVT)}$ is still significant ($\delta_{a(IP)} = 0.9975\delta_{a(IVT)} - 0.0425$, $R^2 = 0.9959$, $p<0.001$, $n=150$). In fact, the filter ($\delta_{ET}<\delta_v<\delta_a$ or $\delta_{ET}>\delta_v>\delta_a$) is necessary for further calculation (e.g., C_{ET} and f_{ET} in line 245-246) rather than an assumption of our methods. We made these clearer in the revision.

Date	number of acceptable $\delta_{a\ (IP)}$ value in a whole day	number of acceptable $\delta_{a\ (IVT)}$ value in a whole day
5/19	27	8
5/27	13	3
5/28	30	3
5/31	25	5
6/4	38	5
6/5	28	0
6/7	29	6
6/9	32	5
6/10	26	2
6/11	21	4
6/12	22	4
6/15	32	0
6/16	33	0
6/17	24	1
6/18	26	0
6/21	26	3
6/22	22	0
6/26	22	0
6/27	29	3
7/4	23	0
7/5	23	1
7/7	30	0
7/8	29	0
7/14	28	4
7/16	28	0
7/18	25	1
7/19	28	6
7/20	27	6
7/21	29	0
7/22	19	0
8/3	18	1
8/4	22	3
8/5	25	3
8/6	28	1
8/12	13	8
8/18	19	3
8/19	30	0
8/28	23	0
8/29	22	1
8/30	27	1
8/31	27	0
9/20	25	0
9/21	24	1
9/22	31	1
9/23	28	1
9/27	28	2
9/28	25	1
9/29	30	5
9/30	25	1

Table 1. The number of acceptable estimated isotope composition of ambient vapor using the intersection point method ($\delta_{a(IP)}$) and the Intermediate Value Theorem method ($\delta_{a(IVT)}$) among all 49 days

The conclusion in line 39 (consistency between results and HYSPLIT modelling) and the statement in line 235 ff ("The calculated delta_a values on 11th June and 12th August ... were higher than on the other days") is not very well supported by the data. Firstly, there are only four data points. Secondly, the data is not that clear for the IP method: In line 204 it is written that the values were -12.95permil on 19th of May and -12.77permil on 12th of August. This is a difference of only approximately 0.2 permil. As there are so few data points for the comparison to modelling, the conclusion in line 39 is far to strong.

Response: We thank the reviewer for the constructive comment. Instead of using four representative days, we utilized all 49 days in this revision but made quality controls. After we expanded the database, we made a new time series of isotopic variation (**Fig. 2**) to replace the original **Fig. 2**. To ensure the representativeness of diurnal average $\delta_{a(IVT)}$, we removed 28 of 49 days (**Table 1**) because the number of acceptable $\delta_{a(IVT)}$ is no more than one in these 28 days. After this kind of quality control, we made two new figures (**Fig. 3a** and **Fig. 3b**). Fig. 3a shows external origins and **Fig. 3b** shows local origins based on HYSPLIT. Almost all of the $\delta_{a(IP)}$ and $\delta_{a(IVT)}$ in **Fig. 3a** is smaller than that of Fig. 3b.

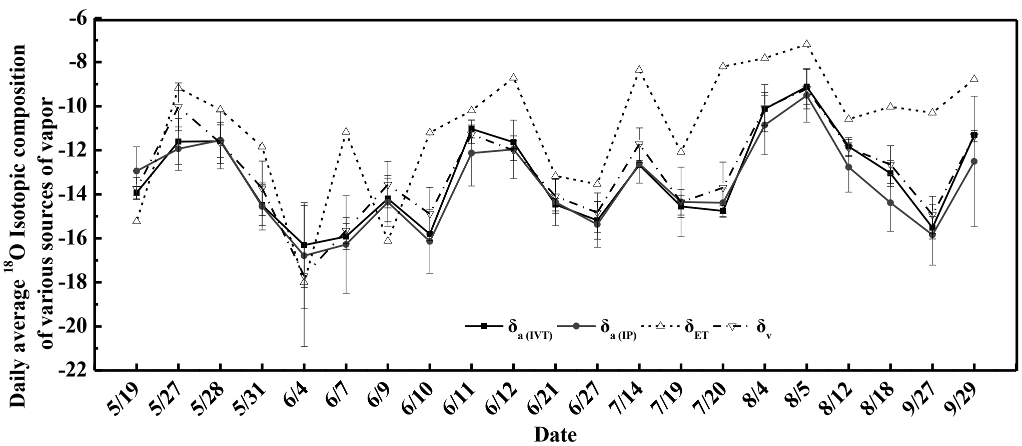


Fig. 2 The daily average values of the isotope composition of evapotranspiration vapor (δ_{ET}), the isotope composition of atmospheric vapor (δ_v), the estimated isotope composition of ambient vapor using the intersection point method ($\delta_{a(IP)}$) and the Intermediate Value Theorem method ($\delta_{a(IVT)}$) in all 49 days.

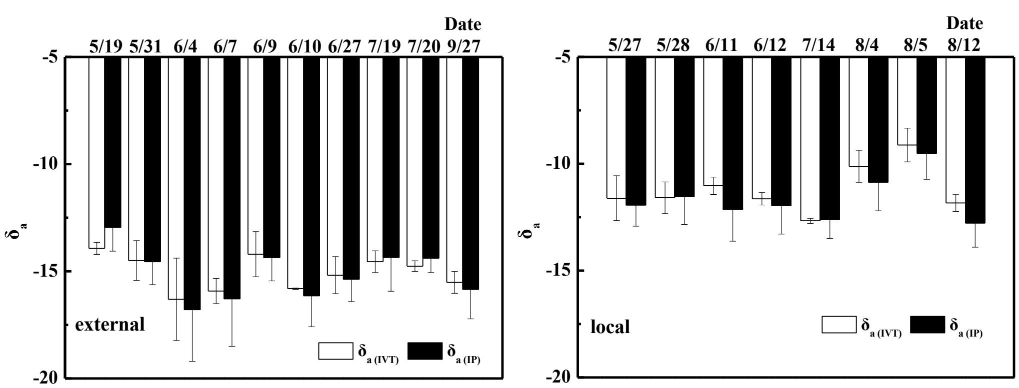


Fig 3 The daily average values of the estimated isotope composition of ambient vapor using the intersection point method ($\delta_{a(IP)}$) and the Intermediate Value Theorem method ($\delta_{a(IVT)}$) after filter. Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) backward trajectory showed external origin (a) and local origin (b), respectively.

The diurnal averages of the methods (lines 202ff and line209ff) are quite different between the two models (up to 1 permil but in both directions). The difference between day and night values is app. -1.6 permil for the IP method and 0.02 permil for the IVT method (lines 205 and 211 resp.) Thus, on a daily scale, the method comparison (Fig. 4) is much worse than on a point to point scale. Without providing a time series, it is hard to understand what is the problem here and to see e.g. in how far the diurnal means are uncertain and contain more or less data points. Thus, the conclusion in line 239 is not very well supported by the data.

Response: We thank the reviewer for the constructive comment. We added new **Fig. 4** in the revision to address these comments. After we expanded the database, the method comparison at daily scale (**Fig. 4a**) is still not as good as the point to point scale (**Fig. 4b**). This is mainly because the number of acceptable $\delta_{a(IP)}$ is far more than that of acceptable $\delta_{a(IVT)}$ for the point to point scale due to the precondition of IVT method ($k_1k_2<0$). As such, the daily results for $\delta_{a(IP)}$ and $\delta_{a(IVT)}$ are based on different numbers of data points. We revised the conclusion in line 239 as follow.

"The reliability of two methods on point to point scale were also supported by the close relationship of δ_a using these two independent methods. Daily time scale result is less reliable than point to point scale."

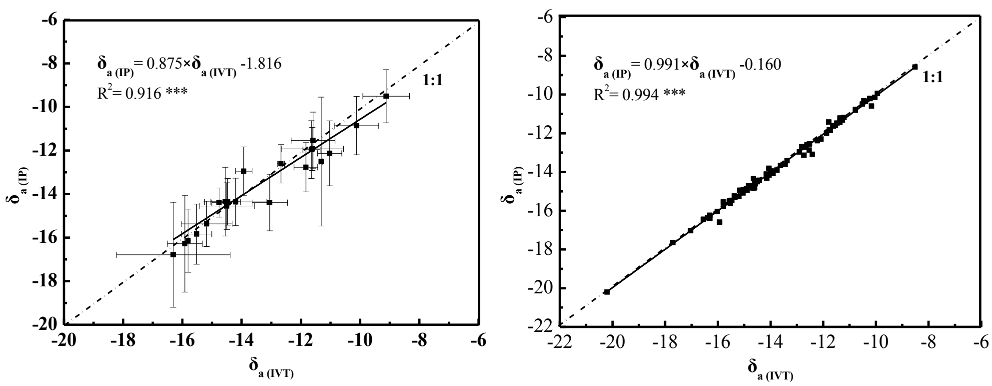


Fig 4 Linear regression between the estimated isotope composition of ambient vapor using the intersection point method ($\delta_{a(IP)}$) and the Intermediate Value Theorem method ($\delta_{a(IVT)}$) on daily scale (a) and point to point scale (b), respectively

The manuscript generally lacks a careful discussion of the (propagated) uncertainties and limitations of both methods as well as of the used data (e.g. Fig. 4 is without errorbars)

Response: Thanks for the comment. We added more discussion of the uncertainty and limitations in the revised manuscript. In the revision, we also added error bars on daily scale method comparison (**Fig. 4a**) and time series of δ_{ET}, δ_v, $\delta_{a(IP)}$ and $\delta_{a(IVT)}$ from May to September (**Fig. 2**).

Throughout the manuscript, there are many unclearities/missing details that makes it hard for the reader to understand what has been done and makes it hard to assess the results. For many of these points it might indeed help to use more references.

Response: Thanks for the comment. We thoroughly revised the materials & method part and results part. More details are shown in the following responses.

In the methods section, there is barely no detail about the calculation of the Keeling plots such as the following: How many data points were used for one Keeling plot calculation? Which data points were used (spatial and temporal) in a single Keeling plot)? Was the calibration procedure a) a standard procedure that has been used elsewhere (if so, please provide a reference) or b) carefully evaluated?

Response: Thanks for the comment. As was shown in the materials and methods section, one Keeling plot calculation contained eight heights of δ_v and C_v (line152). The eight heights data was collected in 30 mins (line165-line166). Therefore, a single Keeling plot considered eight different heights in one time point (line 101). The calibration procedure was a standard procedure that had been used. Our calibration procedure mainly referred to the study by Steen-Larsen et al. (2013). In their study, six steps of calibration protocols were provided in 2.4 section. The calibration protocols were followed in our study. Moreover, we had some different measures to fit our study, which were carefully evaluated. For example, compared with their 15-min-interval switch of different heights, our study shortened this interval into 225s to ensure a relative stable value of δ_a, C_a and C_{ET}. Data from No. 195 to No. 253 was used. The absolute value of coefficient of variations (|CV|) of δ_v and C_v were no more than 0.016 and 0.002, respectively, which was far below the critical value of 15% (Lovie, 2005). C_v gradients calibration was the third calibration step in Steen-Larsen et al. (2013).

Lovie, P.: Coefficient of variation, **Encyclopedia of statistics in behavioral science**, doi: 10.1002/0470013192.bsa107, 2005.

Steen-Larsen, H. C., Johnsen, S. J., Masson-Delmotte, V., Stenni, B., Risi, C., Sodemann, H., Balslev-Clausen, D., Blunier, T., Dahl-Jensen, D., and Ellehøj, M. D.: Continuous monitoring of summer surface water vapor isotopic composition above the Greenland Ice Sheet, **Atmospheric Chemistry and Physics**, 13, 4815-4828, doi: 10.3929/ethz-b-000067919, 2013.

For the IP method, there are more details needed such as: What is the time step between the two Keeling plots that are used? Which of them are used as delta_v and c_v? If all of them are used, you get 8 different delta_a from one single Keeling plot. What did you do with them? Are they treated as individual measurements or are they averaged or did you pick one of them?

Response: We are grateful for the constructive comments from the reviewer. We apologize that we mistakenly thought that the original Eq. 4 and 5 were able to represent the process that δ_a was estimated through the y (or δ_v) value of the point at which two Keeling-plot lines intersect. In fact, the result of IP method was exactly based on intersection point adjacent moments of two Keeling Plots. That was the reason why we call it "intersection point" method. Original Eq. 4 and 5 were not used in our study. We revised the IP method component. The following is our newly added description of IP method. Since the actual calculations in the original manuscript followed the revised procedure, there are no changes in our results.

"**Intersection point method**. Note that for two nearby time points t_1 and t_2, we could use local constant approximation to estimate δ_a within this time interval since it remains relatively constant over a short period of time. By assuming local constant for C_a and δ_a within this time interval, we have

$$k_1 = C_a(\delta_a - \delta_{ET_1}) \quad , \qquad (4)$$

$$k_2 = C_a(\delta_a - \delta_{ET_2}) \quad , \qquad (5)$$

where k_i and δ_{ET_i} represent the value at t_i for i=1, 2. From (4) and (5), we can solve δ_a as:

$$\delta_a = \frac{k_1\delta_{ET_2} - k_2\delta_{ET_1}}{k_1 - k_2} \quad . \qquad (6)$$

The local constant approximation idea was first described in Yamanaka and Shimizu (2007) as an assumption to quantify the contribution of local ET to total atmospheric vapor. ”

It seemed that the new Eq. 4 and 5 had nothing to do with C_v and δ_v. However, δ_{ET_1} and δ_{ET_2} were estimated by clastic Keeling plots, which replied on C_v and δ_v from all eight heights.

Yamanaka, T., and Shimizu, R.: Spatial distribution of deuterium in atmospheric water vapor: Diagnosing sources and the mixing of atmospheric moisture, **Geochimica et cosmochimica acta**, 71, 3162-3169, doi: 10.1016/j.gca.2007.04.014, 2007.

In the conclusion, it is written "The results show an evidence that delta_a was constant ... among different heights". I would like to see the data on which this conclusion is based.

Response: We thank the reviewer for the constructive comment. We did not show any evidence in the text to support the constant δ_a assumption. We just used this assumption in our study. We will delete the related description and apologize for the oversight.

Please provide a time series of all results and indicate the 14 points used for the comparison. The boxplots in Figure 1 can hide interesting features of delta_a. A time series would help to discuss potential problems of the methods e.g. to test the assumption that delta_a is constant at a sufficient timescale.

Response: We thank the reviewer for the constructive comment. After we expanded the database, a time series of all results are shown in **Fig. 2**. Standard deviation (Std) values were selected here to evaluate the constancy among isotopic parameters at daily scale. Std(δ_{ET}), Std(δ_v), Std($\delta_{a(IP)}$) and Std($\delta_{a(IVT)}$) were 6.08, 0.91, 1.38 and 0.59, respectively. As a result, the constancy of δ_a was similar to the constancy of δ_v at daily scale.

The results are presented as showing "four typical days" without any indication how the term "typical" is used here and in particular no data driven evidence for the claim that there four days are "typical".

Response: Thanks for the comment. We have expanded the database as suggested.

In Fig 4, there is no statistics given on the deviation between these models - such as sqrt(mean(delta_IP - delta_IVT)2)). This would give helpful additional information. Additionally, slope and Offset of the regression line in Fig. 4 could be discussed separately. Thus, there seems to be an offset of 0.748 between the two methods. Please discuss this offset.

Response: Thanks for the comment. The $\text{sqrt}(\text{mean}(\delta_{a(IP)} - \delta_{a(IVT)})^2))$ between these two methods on daily scale and point to point scale were 0.618‰ and 0.167‰, respectively. This indicated that two methods were well matched on point to point scale. The slope and offset of point to point scale regression were closer to one than that of daily scale. As IVT method rely on an approximate valuation of $\delta_a \in [min(\delta_{v_1}, \delta_{v_2}), max(\delta_{v_1}, \delta_{v_2})]$, it is reasonable to have some error.

The derivation of the IVT method lacks some clearity. As it is not a direct implementation of the intermediate value theorem, it would be good to add references here and/or explain it more direct. E.g. the six cases in Figure 1 are not clearly written somewhere. It would be helpful to put headlines above the graphs mentioning the order. So for example it is not clear to me, why the Figure did not contain a case that is delta_a1<deltav1<deltav2<delta_a2, because this would also fulfill k1*k2<0.

Response: Thanks for the comment. Appendix of IVT method has been added as follows.

"**Proposition.** Command $\delta_a = f(t)$, which is a continuous function of time. Then for two definite moments t_1 and t_2 ($t_1 < t_2$), the slope of corresponding keeling plot curve was $k_1 = C_{a_1}(\delta_{a_1} - \delta_{ET_1})$ and $k_2 = C_{a_2}(\delta_{a_2} - \delta_{ET_2})$, respectively. When $k_1 k_2 < 0$, there exists$[t_1{}', t_2{}'] \subset [t_1, t_2]$, such that $[min(f(t_1{}'), f(t_2{}')), max(f(t_1{}'), f(t_2{}'))] \subset [min(\delta_{v_1}, \delta_{v_2}), max(\delta_{v_1}, \delta_{v_2})]$ (hereafter $A \in B$).

"To make a proof of the proposition, classical Intermediate Value Theorem (IVT for short) was used. It states that if f is a continuous function from the interval $I = [a, b]$ to real number (R). Then *Version I*. if u is a number between$f(a)$ and$f(b)$, there is *c* in (a, b) such that $f(c) = u$. *Version II*. the image set $f(I)$ is also an interval, and it contains $[min(f(a), f(b)), max(f(a), f(b))]$. While in this study, IVT was able to be explained as follows: if f is a continuous function from the interval $I = [t_1, t_2]$ to R with $min[f(t_1), f(t_2)] < \delta_v$ and $max[f(t_1), f(t_2)] > \delta_v$, then *Version I* Inference. there is $t' \in (t_1, t_2)$ such that $f(t') = \delta_v$. *Version II* Inference. the image set $f(I)$ is also an interval, and it contains $[min(f(t_1), f(t_2)), max(f(t_1), f(t_2))]$.

"**Proof.** (1) When $[min(\delta_{v_1}, \delta_{v_2}), max(\delta_{v_1}, \delta_{v_2})] \subset [min(\delta_{a_1}, \delta_{a_2}), max(\delta_{a_1}, \delta_{a_2})]$ (Fig. 1 a),

"According to *Version I* Inference there exists $t_1^{'} \in [t_1, t_2]$, such that$f(t_1{}') = \delta_{v_1}$; similarly, there exists $t_2^{'} \in [t_1, t_2]$, such that$f(t_2{}') = \delta_{v_2}$. Based on Version II Inference, there exists$\left[t_1^{'}, t_2^{'}\right] \subset [t_1, t_2]$, such that

$$A = B \qquad (10)$$

"(2) When $[min(\delta_{a_1}, \delta_{a_2}), max(\delta_{a_1}, \delta_{a_2})] \subset [min(\delta_{v_1}, \delta_{v_2}), max(\delta_{v_1}, \delta_{v_2})]$ (Fig. 1 b),

"According to Version I Inference there exists $t_1^{'} \in [t_1, t_2]$, such that$f(t_1{}') = \delta_{a_1}$; similarly, there exists $t_2^{'} \in [t_1, t_2]$, such that$f(t_2{}') = \delta_{a_2}$. Based on Version II Inference, there exists $\left[t_1^{'}, t_2^{'}\right] \subset [t_1, t_2]$, such that

$$A = [min(\delta_{a_1}, \delta_{a_2}), max(\delta_{a_1}, \delta_{a_2})] \subset B \qquad (11)$$

"(3) When $\delta_{v_2} < \delta_{a_1} < \delta_{v_1} < \delta_{a_2}$, or $\delta_{a_2} < \delta_{v_1} < \delta_{a_1} < \delta_{v_2}$ (Fig. 1 c and Fig. 1 d),

"According to Version I Inference there exists $t_2^{'} \in [t_1, t_2]$, such that $f(t_2{}') = \delta_{v_1}$. Based on condition (2), When $[\min(\delta_{a_1}, \delta_{v_1}), \max(\delta_{a_1}, \delta_{v_1})] \subset [min(\delta_{v_1}, \delta_{v_2}), max(\delta_{v_1}, \delta_{v_2})]$, there exists $\left[t_1^{'}, t_2^{'}\right] \subset \left[t_1, t_2^{'}\right] \subset [t_1, t_2]$, such that

$$A \subset [min(\delta_{a_1}, \delta_{v_1}), max(\delta_{a_1}, \delta_{v_1})] \subset B \qquad (12)$$

"(4) When $\delta_{v_1} < \delta_{a_2} < \delta_{v_2} < \delta_{a_1}$, or $\delta_{a_1} < \delta_{v_2} < \delta_{a_2} < \delta_{v_1}$ (Fig. 1 e and Fig.1 f),

"According to Version I Inference there exists$t_1^{'} \in [t_1, t_2]$, such that$f(t_1{}') = \delta_{v_2}$, based on condition (2), When $[\min(\delta_{a_2}, \delta_{v_2}), \max(\delta_{a_2}, \delta_{v_2})] \subset [min(\delta_{v_1}, \delta_{v_2}), max(\delta_{v_1}, \delta_{v_2})]$, there exists $\left[t_1^{'}, t_2^{'}\right] \subset \left[t_1^{'}, t_2\right] \subset [t_1, t_2]$, such that

$$A \subset [min(\delta_{a_2}, \delta_{v_2}), max(\delta_{a_2}, \delta_{v_2})] \subset B \tag{13}$$

"Because $k_1 k_2 < 0$, so $\delta_{a_1} < \delta_{v_1}$ and $\delta_{a_2} > \delta_{v_2}$, or $\delta_{a_1} > \delta_{v_1}$ and $\delta_{a_2} < \delta_{v_2}$. As a result, there is no more extra scenarios including $\delta_{a_1} < \delta_{v_1} < \delta_{a_2} < \delta_{v_2}$, $\delta_{v_1} < \delta_{a_1} < \delta_{v_2} < \delta_{a_2}$, $\delta_{v_2} < \delta_{a_2} < \delta_{v_1} < \delta_{a_1}$ and $\delta_{a_2} < \delta_{v_2} < \delta_{a_1} < \delta_{v_1}$.

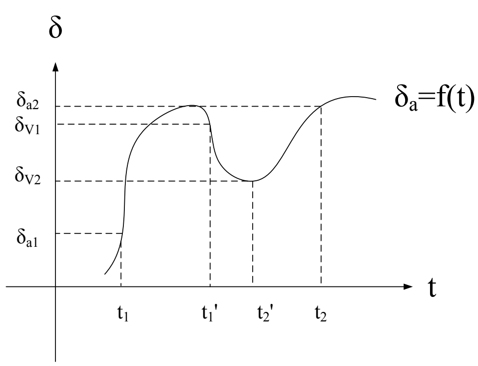
"The proposition is true for all four possible scenarios, which make the estimation of δ_a theoretically feasibly: when $k_1 k_2 < 0$ and when δ_{v_1} and δ_{v_2} adequately close, actual δ_a between t_1 and t_2 can be ensured as the interval below:

$$\text{“}\delta_a \in [min(\delta_{v_1}, \delta_{v_2}), max(\delta_{v_1}, \delta_{v_2})] \tag{14}$$

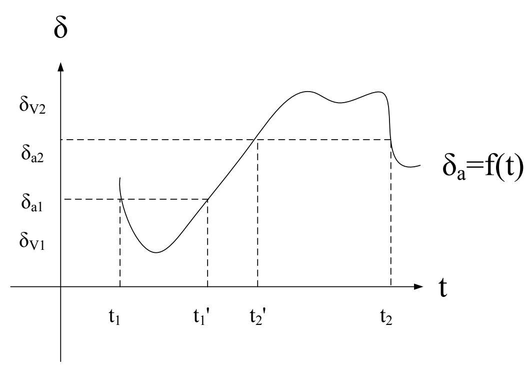
"To simplify the result, actual δ_a between t_1 and t_2 can be approximately regarded as:

$$\text{“}\delta_a \approx \frac{\delta_{v_1} + \delta_{v_2}}{2} \tag{15”}$$

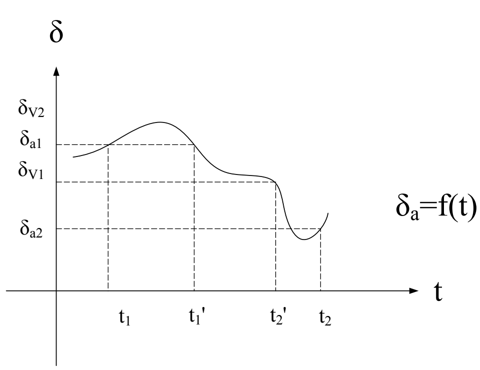
The case "$\delta_{a_1} < \delta_{v_1} < \delta_{v_2} < \delta_{a_2}$" was included in the first scenario.



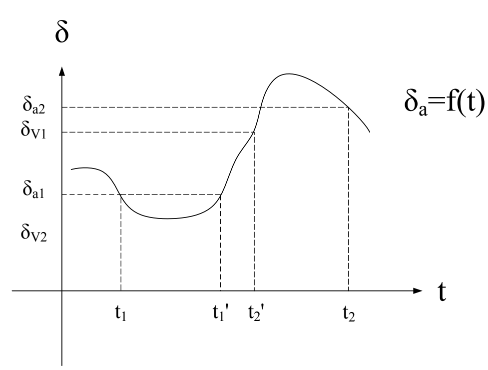
(a)



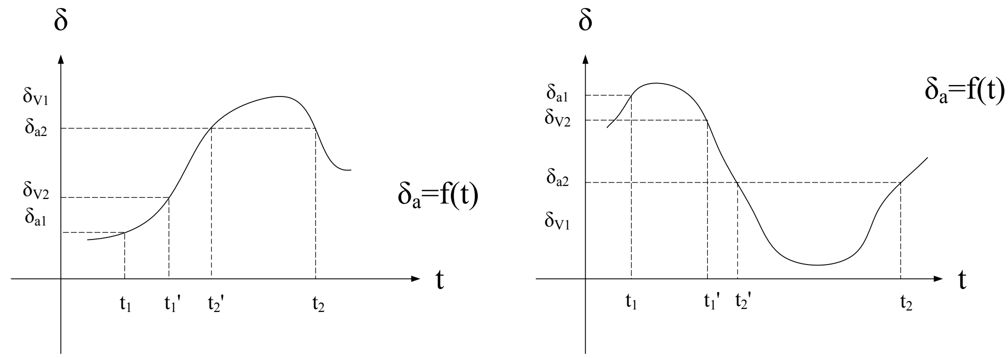
(b)



(c)



(d)



(e) (f)

Fig. 1 Theoretical diagrams of all possible combinations of the relationships between isotope composition of ambient vapor (δ_a) and observed isotope composition of atmospheric vapor (δ_v) of two continuous moments t_1 and t_2, ($t_1 < t_2$). δ_{a1} and δ_{a2} represent δ_a value in t_1 and t_2, respectively. δ_{v1} and δ_{v2} represent δ_v value in t_1 and t_2, respectively. t_1' and t_2' represent the time of two specific moments between t_1 and t_2 with $t_1 < t_1' < t_2' < t_2$. For all of the six situations, there exists some sub-intervals $[t_1', t_2'] \subset [t_1, t_2]$ such that the whole range of $\{\delta_a(t): t \in [t_1', t_2']\}$ is within $[min(\delta_{v_1}, \delta_{v_2}), max(\delta_{v_1}, \delta_{v_2})]$.

I would recommend a detailed language check and in general a more careful usage of definitions, because there are some language-related unclearities that might be avoided by a more precise description.

Response: Changed as suggested.

Some minor comments:

It is not clearly written how Eq. 6 is used. I guess delta_ET is taken from two adjacent Keeling plots, but which c_v and delta_v are taken. One more sentence would help here.

Response: Apologize. Eq. 6 is a mistake. Revise have been made.

The calibration procedure is not explained. E.g. it is not clear to me, what is meant in line 173. If this refers to a standard procedure, a reference would help.

Response: Explained above.

Line 270: I am not sure if the IVT method really gives an explanation for the figure as stated here, or if it is rather the other way around, that the figure can be used to understand the IVT method and in particular the change of slope.

Response: Explained above.

I think the reference to equation 1 in line 197 is wrong.

Response: Sorry we do not find any equations in line 197. Equation 1 is in line 93. It is a commonly used water balance equation.