Supplementary information for "A rainfall-tracking travel time distribution model to quantify mixing and storage release preference in a large shallow lake by two-year stable isotopic data"

Rong Mao[1], Xin Luo[1], Jiu Jimmy Jiao[1], Xiaoyan Shi[1], and Wei Xiao[2]

[1]Department of Earth Sciences, The University of Hong Kong, P. R. China.
[2]Yale-NUIST Center on Atmospheric Environment, International Joint Laboratory on Climate and Environment Change (ILCEC), Nanjing University of Information Science and Technology, Nanjing, Jiangsu Province, China

Correspondence: Jimmy Jiao (jjiao@hku.hk)

Contents of this file

Text S1. The flow rate data of surface runoff, lake volume data and precipitation data excerpted from the *Monthly Hydrological Report*

The flow rate data of surface runoff, lake volume data and precipitation data excerpted from the *Monthly Hydrological Report* between 2012 and 2014 are presented in Figs. S1-S2 of this file.

Text S2. The daily temperature, relative air humidity and wind speed monitored in Lake Taihu

The monitored temperature, relative air humidity and wind speed data with daily frequency at the daily sampling point of Lake Taihu are presented in Fig. S3 of this file.

The calculated daily evaporation rate is shown in Fig. S4.

Base on the evaporation rate, Fig. S5 illustrated calibrated results of the mass balance model of Lake Taihu.

Text S3. The monitored isotope data in the lake and rivers connected to the lake

The monitored isotope data in the lake and in the rivers connected to the lake at the sampling points shown in the main text are presented in Figs. S6 and S7 of this file.

Text S4. Numerical routines of the implementation of the age master equation model and isotope mass balance model

The age master equation model and isotope mass balance model are implemented in Euler-forward method. The input data of the numerical routine include precipitation, river inflow rate, river outflow rate, calibrated evaporation rate, shape parameters in age-ranked SAS function, evaporation fractionation factor, deuterium isotope content in precipitation, and deuterium isotope contents in rivers. The output is age-ranked storage $S_T(t,\tau)$ and mean isotope content in the lake. The numerical routines of the Euler-forward scheme is as follows:

(1) Initialize age-ranked storage $S_T(t=0,\tau=0)$ to be the initial lake volume $S(t=0)$; initialize the mean isotope content in the lake $\delta^2\mathrm{H}_{mean}(t=0,\tau=0)=\delta^2\mathrm{H}_{in}(t_{in}=0)$; calculate age-ranked SAS functions Ω_Q and Ω_E at $t=0$.

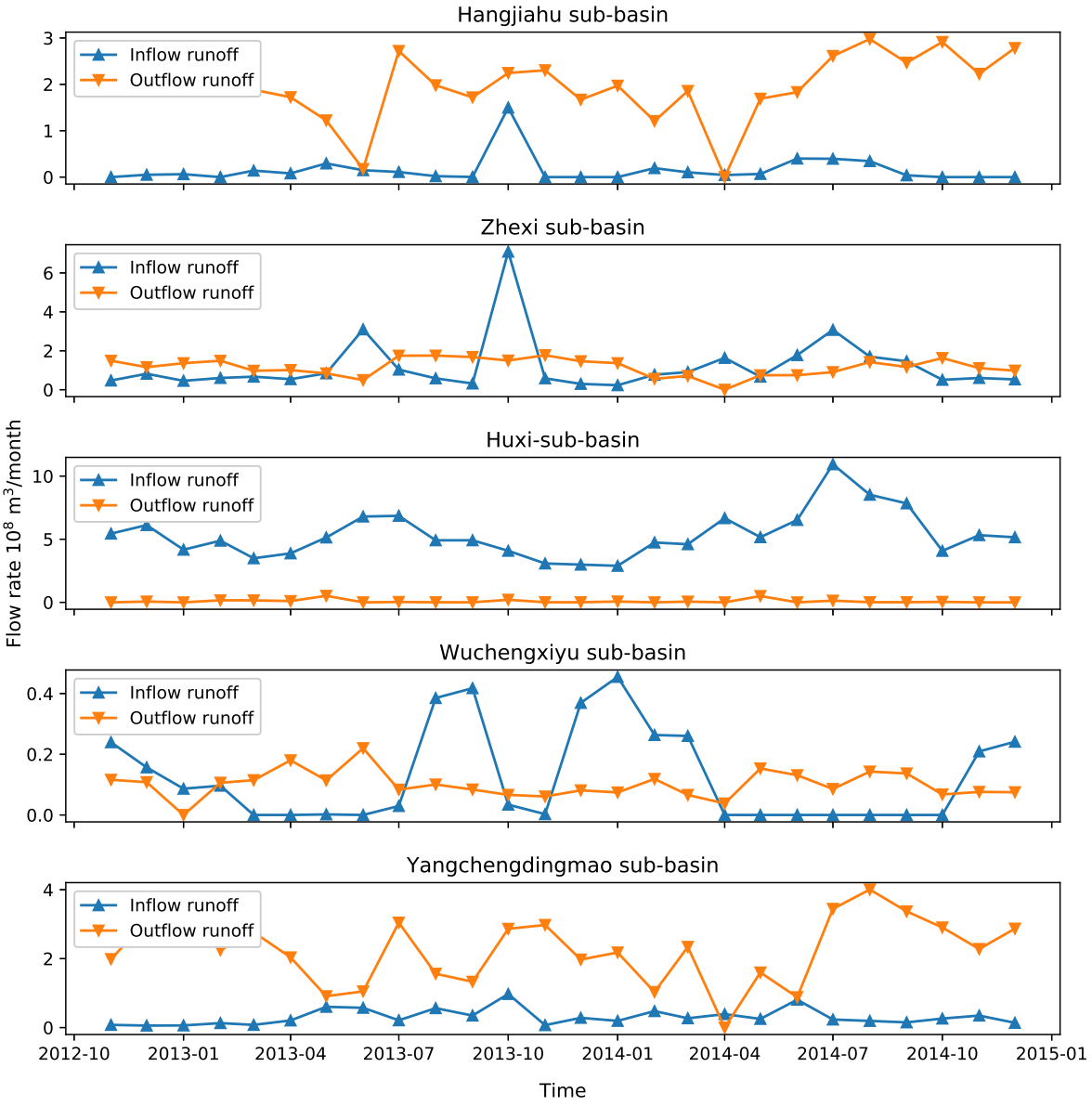

Figure S1. Flow rate of inflow runoff and outflow runoff in sub-basins connected to Lake Taihu.

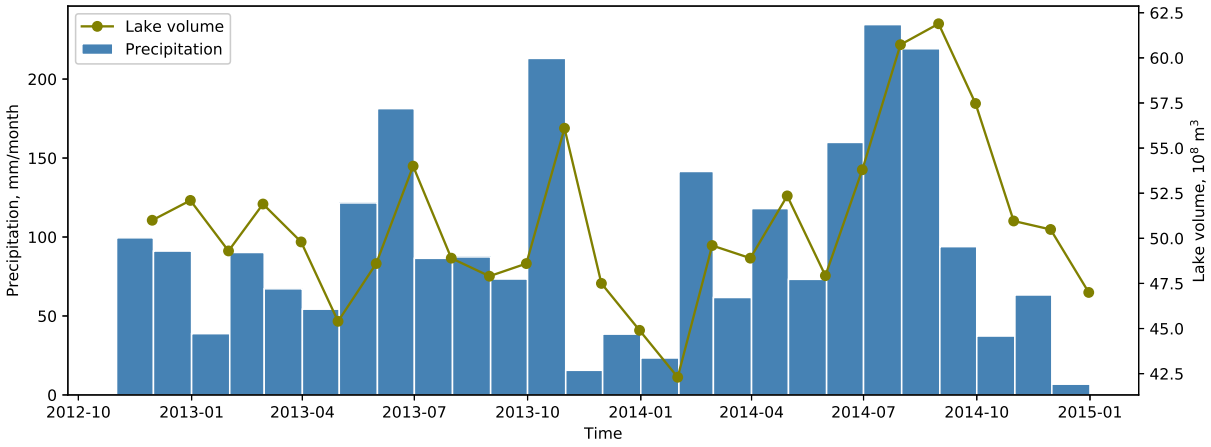

Figure S2. Monthly precipitation and volume of Lake Taihu.

(2) Increase time step from i to $i+1$, calculate the volume of input water based on precipitation and river inflow rate at i^{th} time step. The volume of event water emerges in output water should be subtracted from the input water; the volume of event water is evaluated by the age-ranked SAS function Ω_Q obtained from i^{th} time step;

(3) Update age-ranked storage $S_T(t=i+1,\tau)$ at $i+1$ time step by adding the new input water, and calculate age-ranked SAS functions Ω_Q and Ω_E;

(4) Update age-ranked storage $S_T(t=i+1,\tau)$ at $i+1$ time step again by subtracting age-ranked volume of water in river outflow and vapour according to the age-ranked SAS functions Ω_Q and Ω_E evaluated at step (3);

(5) Calculate mean isotope content in the lake based on the isotope mass balance model and the age-ranked storage obtained at step (4);

(6) Return to step (2) until $i+1$ is equal to the end of modeling time. The modeling time is set to be 12 years by repeating the two-year observed data six times to remove the influence of unrealistic initial condition on the modeling results.

Text S5. The determination of shape parameters of beta distribution on storage 'selection' preference

The relation between the shape parameters (α and β) in beta distribution and the storage 'selection' preference is presented in Fig. S8.

Text S6. Discrete Fourier transform to rebuild monthly isotope content in lake water

The calculation of mean isotope content in Lake Taihu is based on the daily sampled point in Lake Taihu and 29 seasonally sampled points in Lake Taihu. Discrete Fourier transform is applied to rebuild the monthly isotope content at the 29 seasonally sampled points in Lake Taihu. The discrete Fourier transform decomposes the time-series data into a series of signals with

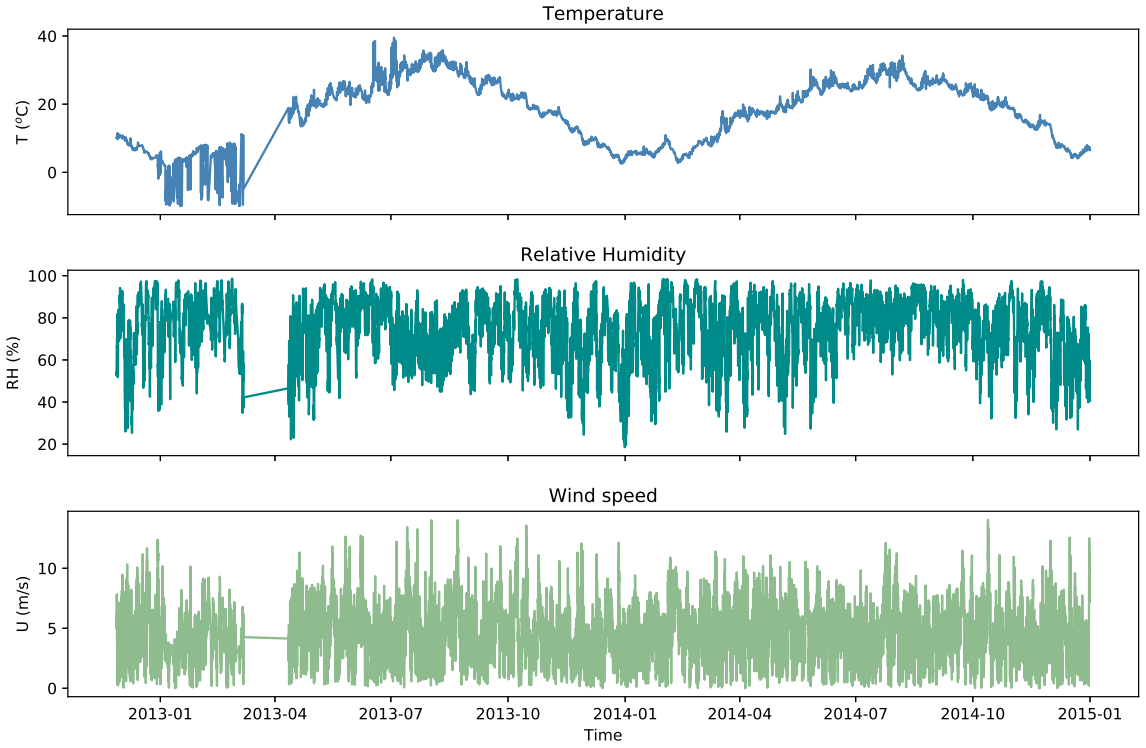

Figure S3. Daily monitored temperature of lake water, relative humidity in air and the wind speed at 10 meters above the lake.

different wavelengths, and each signal has its own wavelength and amplitude. Then, the monthly mean isotope contents at these seasonally sampling points are calculated in the following steps:

(1) Perform discrete Fourier transform for isotope content data at the daily sampling point from November 2012 to November 2014:

$$F(n) = \sum_{t=0}^{N\cdot\Delta t} f(t) e^{-i\frac{2\pi\cdot n}{N\cdot\Delta t}t} \quad \begin{cases} 0 \le n \le \frac{N}{2} - 1, N \text{ is even} \\ 0 \le n \le \frac{N-1}{2}, N \text{ is odd} \end{cases} \tag{1}$$

where $f(t)$ is the time series data and represents the daily isotope content in this step; N is the number of sampling points in $f(t)$; Δt is the sampling spacing. i is the imaginary unit. The wavelength of the n-th signal is $\frac{N\cdot\Delta t}{n}$ for $n \neq 0$. The amplitude of the n-th signal is $2\frac{|F(n)|}{N}$.

(2) Fit the seasonal isotope data at each sampling point by the equation below:

$$\delta^2\mathrm{H}(t) = a_0 + \sum_{n=1}^{4}(a_n cos(n\pi t) + b_n sin(n\pi t)) \tag{2}$$

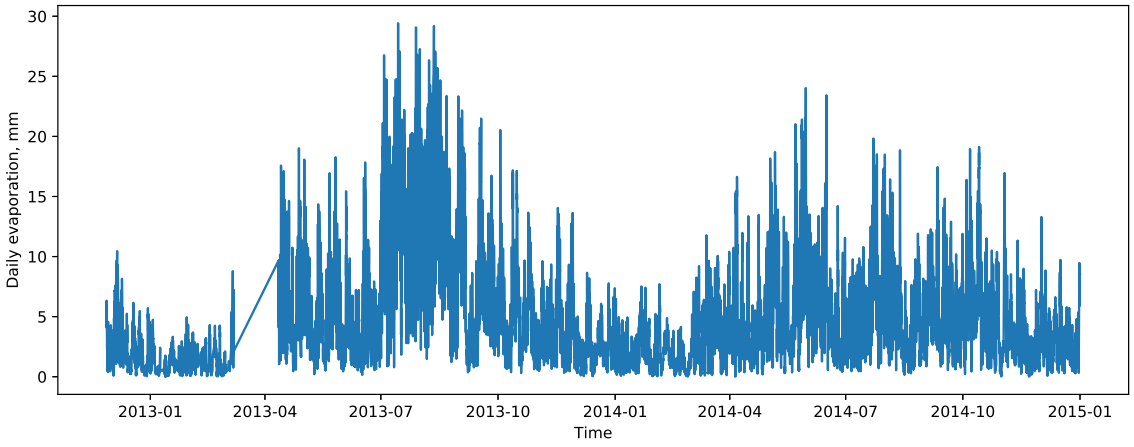

Figure S4. Evaluated daily evaporation rate. There is a gap between March 8, 2013 and April 10, 2013 due to the missing data. This gap is filled by the evaporation rate at the same period of the year 2014 for further calculation.

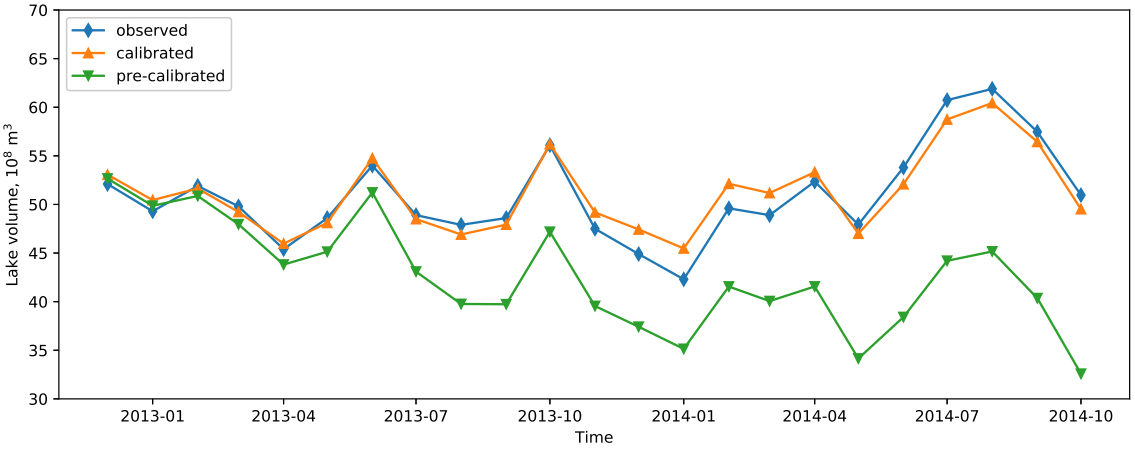

Figure S5. Calibration of mass balance model for lake water. The blue line is the lake volume; the pre-calibrated line lake volume is calculated using mass balance equation and the calculated evaporation rate by Penman equation; the calibrated lake volume is calculated using the mass balance equation and the calibrated evaporation rate.

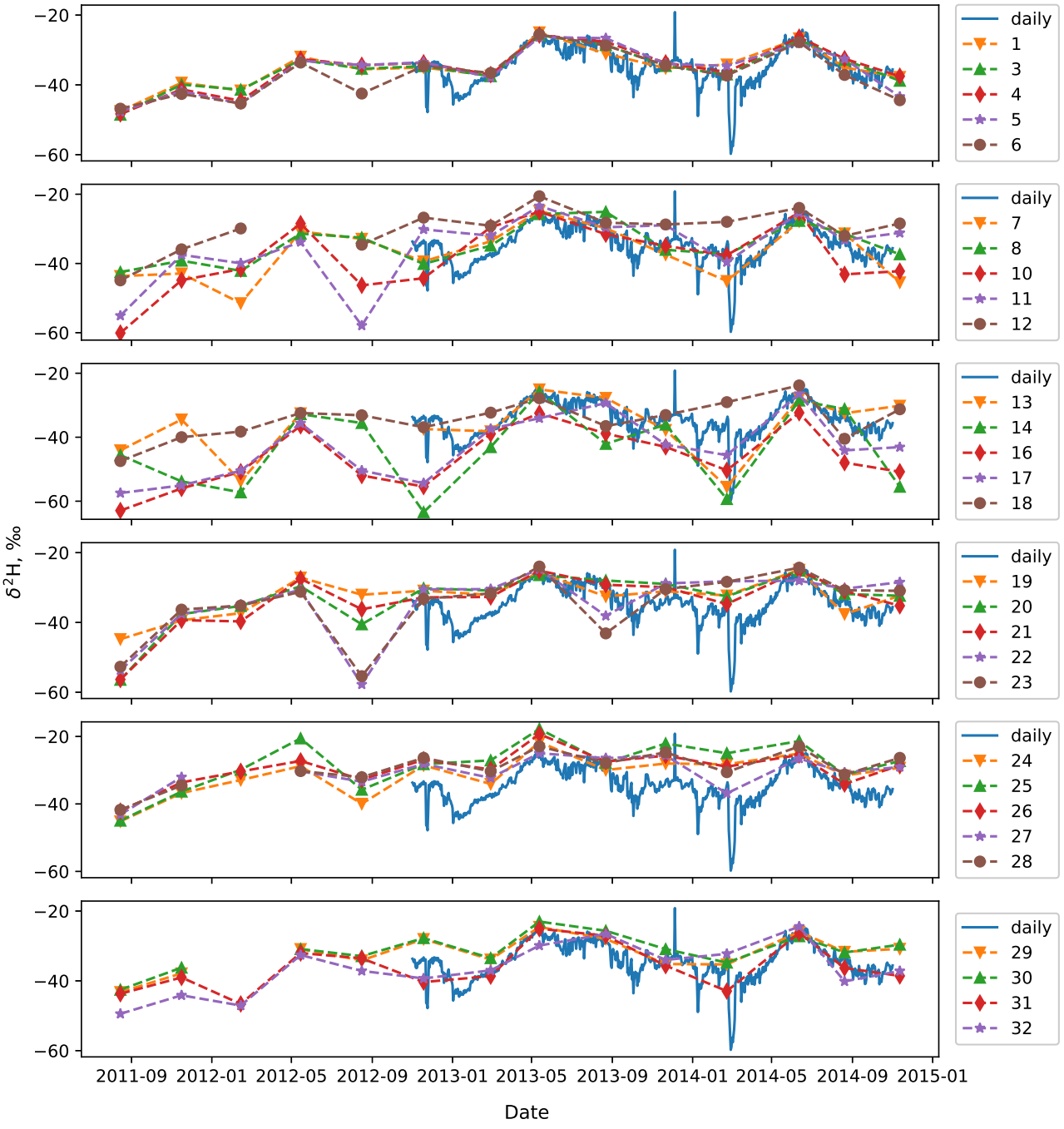

Figure S6. Observed isotope content in Lake Taihu, the dashed lines in each sub-figure represent the isotope content at the seasonal sampling points in the lake, the solid line represent the isotope content at the daily sampling point in the lake. There is 1 daily sampling point and 29 seasonally sampling points in total.The solid line in each sub-figure is for comparing the spatial variation of isotope content in the lake.

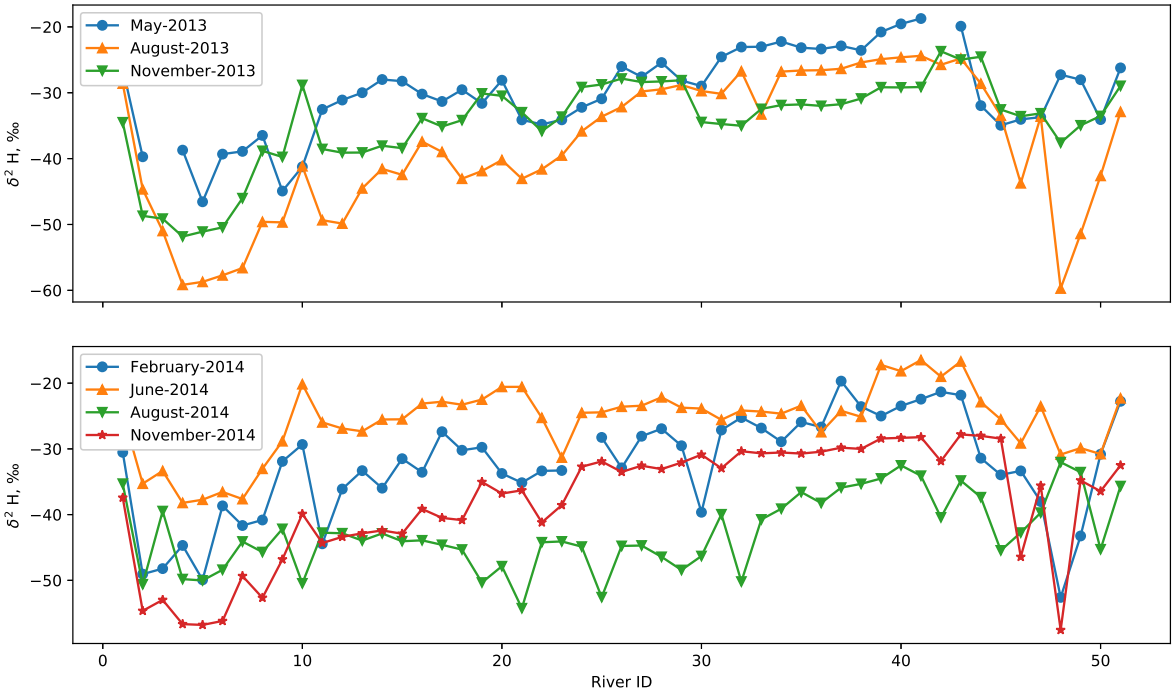

Figure S7. Spatial distribution of isotope content in rivers connected to Lake Taihu during 2013 and 2014. The label of x-axis is the river ID which is shown in figure **??**.

The initial gauss of the parameters a_0, a_n and b_n is calculated from $F(n)$:

$$a_0 = \frac{|F(0)|}{N};\ a_n = 2\frac{Re(F(n))}{N};\ b_n = 2\frac{Im(F(n))}{N} \tag{3}$$

where $Re(F(n))$ is the real part of the complex number $F(n)$, and $Im(F(n))$ is the imaginary part of $F(n)$.

(3) Infer the daily isotope content at the seasonal sampling point by inverse discrete Fourier transform:

$$f'(t) = \sum_{n=0}^{N}\left[\frac{1}{N}\cdot F'(n)\cdot e^{i\frac{2\pi\cdot n}{N\cdot\Delta t}t}\right] \tag{4}$$

where $F'(n)$ is calculated as follows:

$$F'(n) = \begin{cases} F'(0) = a_0 N & n = 0 \\ a_n\frac{N}{2} + b_n\frac{N}{2}\cdot i & n = 1\ldots4 \\ F'(n) = F(n) & n > 4 \end{cases} \tag{5}$$

$f'(t)$ is the inferred daily isotope data at the seasonal sampling points. As is shown in Eq. (5), the assumption of this step is that the fluctuations of isotope content whose wavelengths are smaller than half a year are the same everywhere in the

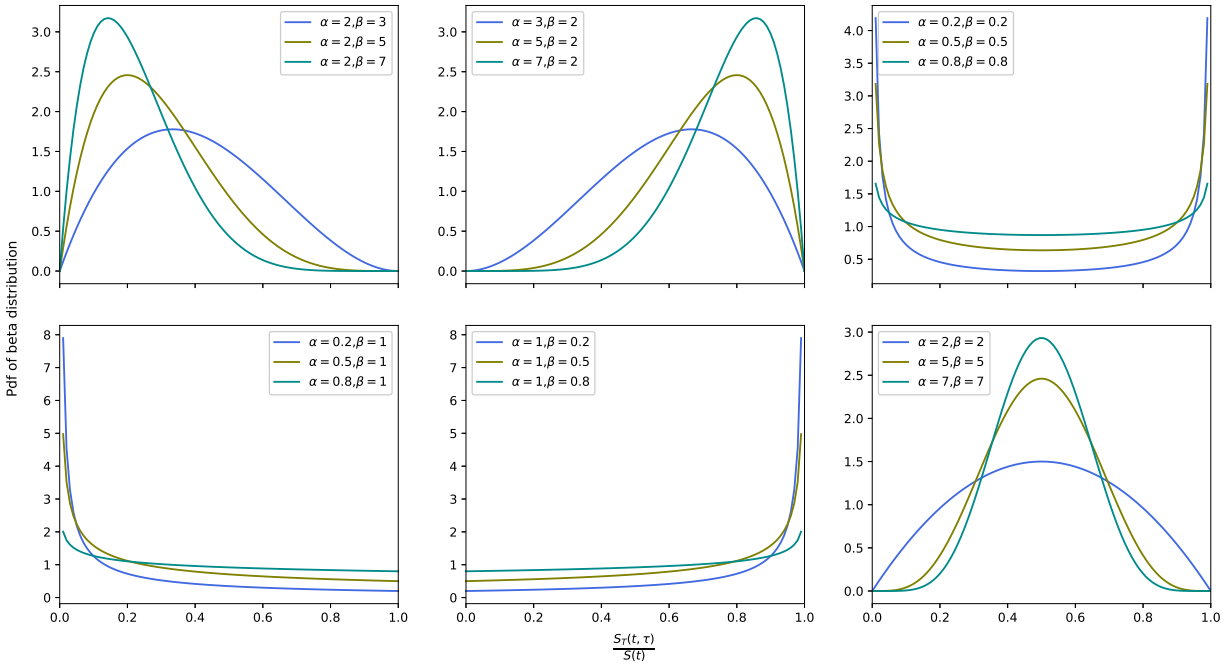

Figure S8. Relation between shape parameters (α and β) in Beta distribution and the storage selection preference: (1) $\beta > \alpha > 1$ or $\alpha < 1, \beta = 1$: the storage prefer releasing young water; (2) $\alpha > \beta > 1$ or $\alpha = 1, \beta < 1$: the storage prefer releasing old water; (3) $\alpha < 1, \beta < 1$: the storage prefer releasing both young and old water.

lake including the daily sampling point and the seasonal sampling point. Meanwhile, the fluctuations with wavelength larger than half a year are determined by the seasonally observed isotope content data. The influence of this assumption on the final evaluated time-variant TTD can be reduced by chosing a proper resolution in time.

(4) The inferred daily isotope content $f'(t)$ at the seasonal sampling points can be used to calculate the mean isotope content at the other time.

This method not only considers the similarity of the fluctuation between the isotope content at the daily sampling point and the isotope content at the seasonal sampling points, but also maintains the difference of the fluctuation among the isotope content at these sampling points.

The result of discrete Fourier transform for daily lake isotope data is presented in figure Fig. S9. It can be seen that the signal with the wavelength of 1 year has the maximum amplitude, and this indicates that the fluctuation of isotope content with the wavelength of 1 year is the most significant. When wavelength becomes shorter, the amplitude decreases rapidly. Meanwhile, Fig. S9 compares the spectrum of daily isotope data with smoothed isotope data. The amplitudes of signals with short wavelengths show a large decrease caused by smooth method. Therefore, the smooth method can reduce the contribution of fluctuations with short wavelength to the total fluctuation of isotope content. This study set the resolution of time-variant

TTD to be one month, so the monthly average isotope content is calculated as one of the input data for the evaluation of time-variant TTD. The precipitation, river inflow and river outflow data should also be averaged monthly. The lake volume data is not the monthly average but the volume data is recorded at the interval of one month. In this way, the change of lake volume and the change of isotope content in one month can be quantified with these data.

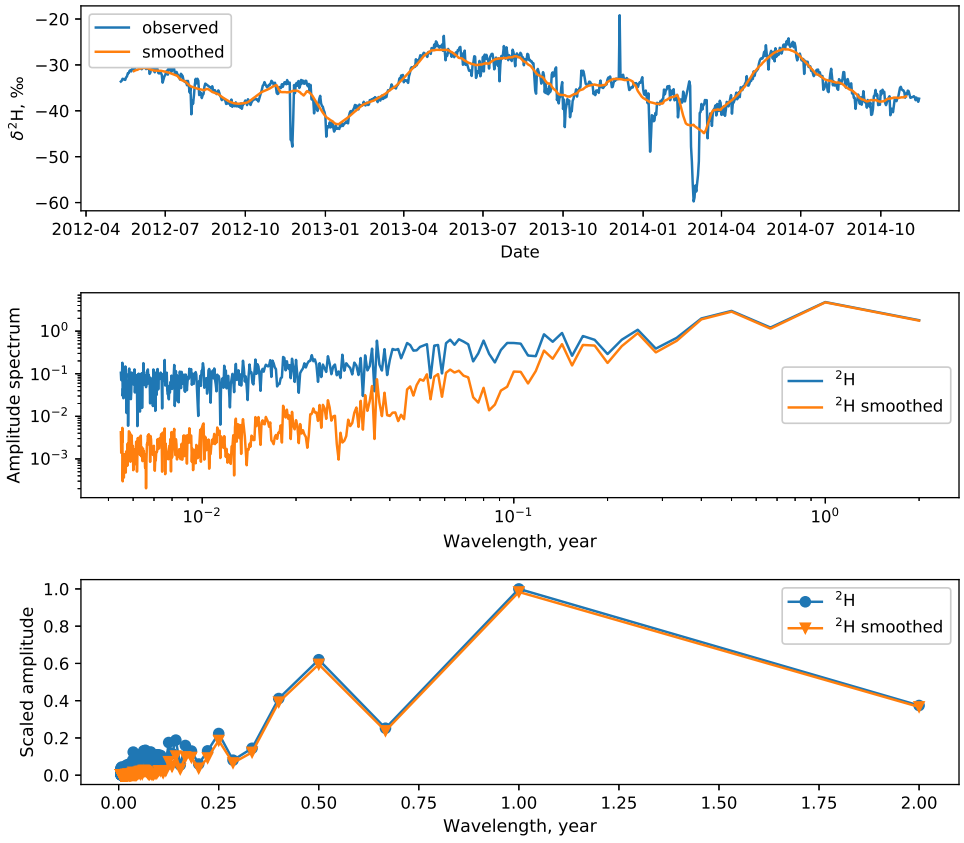

Figure S9. Discrete Fourier transform for original daily lake isotope data and smoothed lake isotope data. The first sub-figure shows the original lake isotope data and smoothed isotope data with smoothed windows which is equal to 30 days; the second sub-figure shows the results of discrete Fourier transform for original data and smoothed data; the third sub-figure scaled the results shown in the second sub-figure. The amplitudes in third sub-figure are scaled by the amplitude of the signal with 1 year wavelength.

Text S7. Implementation of sequential Monte Carlo method

The sequential Monte Carlo method is used to solve the inverse problem described in the main text. The implementation starts with the prior uniform distribution of shape parameters and evaporative fractionation factor. The posterior distribution of model parameters is the expected results of this inverse problem, and the relation between the posterior distribution and prior

distribution is expressed as:

$$p(\Theta|Y) = \frac{p(\Theta)p(Y|\Theta)}{p(Y)} \tag{6}$$

where $\Theta = [a, b, k, \epsilon]$ represents the model parameters; Y represents the calculated mean isotope content in the lake from observed data. The term $p(\Theta|Y)$ is the posterior probability distribution of Θ given Y; $p(\Theta)$ is the prior uniform distribution of Θ; $p(Y|\Theta)$ is the likelihood which means the probability of Y given the model parameters Θ. $p(Y)$ is a constant which represents the probability of Y. Eq. (6) is known as the Bayes' theorem which updates the hypothesis, i.e., the prior uniform distribution of model parameters given the evidence, i.e., the observed isotope content in the lake. The likelihood $p(Y|\Theta)$ links the hypothesis and evidence, and describes how the model parameters determines the mean isotope content in the lake. The age master equation model and isotope mass balance model are wrapped in the likelihood. A multi-variable Gaussian distribution is assigned to the likelihood, and this likelihood means that the probability distribution of the errors between the calculated mean isotope content from the model and the calculated mean isotope content from the observed isotope content in the lake are a multi-variable Gaussian distribution.

As the posterior distribution $p(\Theta|Y)$ of model parameters by Eq. (6) is difficult to be calculated analytically, a numerical sampling method, i.e., the sequential Monte Carlo (SMC) method is adopted to obtain $p(\Theta|Y)$. Several numerical methods are combined in SMC including tempering method, importance sampling and Markov Chain Monte Carlo (MCMC) kernel. The tempering method introduces an auxiliary temperature parameter λ to control the sampling process:

$$p(\Theta|Y)_\lambda = \frac{p(\Theta)p(Y|\Theta)^\lambda}{p'(Y)} \tag{7}$$

where $p'(Y)$ can be regarded as a normalized constant and will vanish in the sampling step. Specifically, for $\lambda = 0$, $p(\Theta|Y)_{\lambda=0} = p(\Theta)$; and for $\lambda = 1$, $p(\Theta|Y)_{\lambda=1} = p(\Theta|Y)$. During the sampling process, the auxiliary temperature parameter λ increases from 0 to 1 to control the transition from prior distribution which is easy to sample to posterior distribution which is hard to sample. The parameter λ is like a knob that turns on the likelihood gradually. This tempering method is capable of sampling from distribution with multiple peaks and exploring the whole state space of model parameters efficiently.

Importance sampling and MCMC kernel are used to draw samples of model parameters from posterior distribution. The posterior distribution of model parameters is obtained from the samples generated by sequential Monte Carlo method. The procedures to draw samples from posterior distribution $p(\Theta|Y)$ are summarized as follow:

(1) Initialize λ to be zero and draw samples from prior uniform distribution. Let S_λ denote this set of samples. The size of S_λ is N;

(2) Increase λ gradually, and compute the importance weight for each sample in S_λ: $W = \frac{p(Y|\Theta)_{\lambda_{i+1}}}{p(Y|\Theta)_{\lambda_i}}$;

(3) Re-sample S_λ according to the weight W. Let S_w denote the new set of samples. The size of sample is still N;

(4) Run N chains with MCMC kernel. The start of each chain is the sample in S_w;

(5) Return to step (2) until $\lambda \geq 1$;

(6) The final posterior distribution $p(\Theta|Y)$ is a collection of N samples from the ends of final N MCMC chains.

During the implementation of SMC, the increase rate of auxiliary temperature parameter λ is determined automatically according to the convergence rate, i.e., the effective sample size of S_W. Generally, the lower the convergence rate is, the smaller the increase rate of λ will be. Meanwhile, the number of steps that each MCMC chain takes is also determined automatically according to acceptance rate of previous N MCMC chains. The model parameters are updated at each step in all MCMC chains, so the age master equation model and isotope mass balance model are running repeatedly for each new set of model parameters during the sampling process. The number of MCMC chains running in this study is set to be 3000, that is, $N = 3000$.

At a certain sampling step in a certain MCMC chain, the age master equation model and isotope mass balance model are employed to calculate mean isotope content in the lake for the evaluation of the likelihood $p(Y|\Theta)$ of the new sampled model parameters. The age master equation models are solved numerically in Euler-Forward scheme which is presented in the section 1.

Text S8. Explanation of the biased estimation of mean deuterium concentration of lake water $c_L(t,\tau)$

Referring to equation (19) in the maintext, the deuterium concentrations in rain water and river water of age τ are:

$$c_{L\_rain}(t,\tau) = c_{in\_rain}(t-\tau)\cdot e^{(1-\epsilon)\int_0^{t-t_{in}} \frac{E_{rain}(t)\overleftarrow{p_{E\_rain}}(t,\tau)}{S_{rain}(t)p_{s\_rain}(t,\tau)}\mathrm{d}\tau} \tag{8}$$

$$c_{L\_river}(t,\tau) = c_{in\_river}(t-\tau)\cdot e^{(1-\epsilon)\int_0^{t-t_{in}} \frac{E_{river}(t)\overleftarrow{p_{E\_river}}(t,\tau)}{S_{river}(t)p_{s\_river}(t,\tau)}\mathrm{d}\tau} \tag{9}$$

Then, the mean deuterium concentrations $c_{L\_1}(t,\tau)$ in the mixed lake water of age τ is:

$$c_{L\_1}(t,\tau) = \frac{s_{rain}(t,\tau)c_{L\_rain}(t,\tau) + s_{river}(t,\tau)c_{L\_river}(t,\tau)}{s_{rain}(t,\tau) + s_{river}(t,\tau)} \tag{10}$$

Equation (15) is used in this study to calculate the mean deuterium concentration of lake water. It requires tracking the transport and mixing of rain water and river water in the lake, i.e., requires calculating age distributions of rain water $s_{rain}(t,\tau)$ and river water $s_{river}(t,\tau)$.

However, if the mean deuterium concentration is calculated like previous TTD study, i.e., the transport of rain water and river water are not differentiated. The only way to obtain the mean deuterium concentrations $c_{L\_2}(t,\tau)$ in the mixed lake water of age τ is:

$$c_{L\_2}(t,\tau) = c_{in\_mean}(t,\tau)\cdot e^{(1-\epsilon)\int_0^{t-t_{in}} \frac{E(t)\overleftarrow{p_E}(t,\tau)}{S(t)p_s(t,\tau)}\mathrm{d}\tau} \tag{11}$$

where $c_{in\_mean}(t,\tau)$ is the mean deuterium concentration in all input water and is calculated as follow:

$$c_{in\_mean}(t,\tau) = \frac{J(t-\tau)c_{in\_rain}(t-\tau) + F_{in}(t-\tau)c_{in\_river}(t-\tau)}{J(t-\tau) + F_{in}(t-\tau)} \tag{12}$$

Comparing equation (15) with equations (16) and (17), $c_{L\_1}(t,\tau)$ and $c_{L\_1}(t,\tau)$ can not be equal, unless in 2 very specific situations: (1) the deuterium concentrations in both the input rain water c_{in_rain} and input river water c_{in_river} are the same; and the SAS functions of evaporation of rain water $\frac{E_{river}(t)\overleftarrow{p_{E\_river}}(t,\tau)}{S_{river}(t)p_{s\_river}(t,\tau)}$ and river water $\frac{E_{river}(t)\overleftarrow{p_{E\_river}}(t,\tau)}{S_{river}(t)p_{s\_river}(t,\tau)}$ are the same; (2)

the rain water and river water are well mixed in the lake, i.e., the proportions of rain water and river water in input water are the same as the proportions in the lake.

In Lake Taihu, the SAS functions of evaporation of rain water and river water are the same, but the deuterium concentrations in the input rain water and input river water are not the same. Besides, the rain water and river water are only well mixed in the vertical direction, so the proportions of rain water and river water in input water are the same as the proportions in the lake. Therefore, the calculation of mean deuterium concentrations $c_L(t,\tau)$ in the mixed lake water using Equation (16) and (17) like previous TTD studies will be biased from $c_{L\_1}(t,\tau)$ obtained by equation (15).

Additionally, most lakes or other systems do not meet the two specific situations, so the calculation of $c_L(t,\tau)$ requires tracking the transport of river and rain water in the lake.

Text S8. Explanation of the biased estimation of mean deuterium concentration of lake water $c_L(t,\tau)$

Referring to equation (19) in the maintext, the deuterium concentrations in rain water and river water of age τ are:

$$c_{L\_rain}(t,\tau) = c_{in\_rain}(t-\tau)\cdot e^{(1-\epsilon)\int_0^{t-t_{in}} \frac{E_{rain}(t)\overleftarrow{p_{E\_rain}}(t,\tau)}{S_{rain}(t)p_{s\_rain}(t,\tau)}\mathrm{d}\tau} \quad (13)$$

$$c_{L\_river}(t,\tau) = c_{in\_river}(t-\tau)\cdot e^{(1-\epsilon)\int_0^{t-t_{in}} \frac{E_{river}(t)\overleftarrow{p_{E\_river}}(t,\tau)}{S_{river}(t)p_{s\_river}(t,\tau)}\mathrm{d}\tau} \quad (14)$$

Then, the mean deuterium concentrations $c_{L\_1}(t,\tau)$ in the mixed lake water of age τ is:

$$c_{L\_1}(t,\tau) = \frac{s_{rain}(t,\tau)c_{L\_rain}(t,\tau) + s_{river}(t,\tau)c_{L\_river}(t,\tau)}{s_{rain}(t,\tau) + s_{river}(t,\tau)} \quad (15)$$

Equation (15) is used in this study to calculate the mean deuterium concentration of lake water. It requires tracking the transport and mixing of rain water and river water in the lake, i.e., requires calculating age distributions of rain water $s_{rain}(t,\tau)$ and river water $s_{river}(t,\tau)$.

However, if the mean deuterium concentration is calculated like previous TTD study, i.e., the transport of rain water and river water are not differentiated. The only way to obtain the mean deuterium concentrations $c_{L\_2}(t,\tau)$ in the mixed lake water of age τ is:

$$c_{L\_2}(t,\tau) = c_{in\_mean}(t,\tau)\cdot e^{(1-\epsilon)\int_0^{t-t_{in}} \frac{E(t)\overleftarrow{p_E}(t,\tau)}{S(t)p_s(t,\tau)}\mathrm{d}\tau} \quad (16)$$

where $c_{in\_mean}(t,\tau)$ is the mean deuterium concentration in all input water and is calculated as follow:

$$c_{in\_mean}(t,\tau) = \frac{J(t-\tau)c_{in\_rain}(t-\tau) + F_{in}(t-\tau)c_{in\_river}(t-\tau)}{J(t-\tau) + F_{in}(t-\tau)} \quad (17)$$

Comparing equation (15) with equations (16) and (17), $c_{L\_1}(t,\tau)$ and $c_{L\_1}(t,\tau)$ can not be equal, unless in 2 very specific situations: (1) the deuterium concentrations in both the input rain water c_{in_rain} and input river water c_{in_river} are the same; and the SAS functions of evaporation of rain water $\frac{E_{river}(t)\overleftarrow{p_{E\_river}}(t,\tau)}{S_{river}(t)p_{s\_river}(t,\tau)}$ and river water $\frac{E_{river}(t)\overleftarrow{p_{E\_river}}(t,\tau)}{S_{river}(t)p_{s\_river}(t,\tau)}$ are the same; (2) the rain water and river water are well mixed in the lake, i.e., the proportions of rain water and river water in input water are the same as the proportions in the lake.

In Lake Taihu, the SAS functions of evaporation of rain water and river water are the same, but the deuterium concentrations in the input rain water and input river water are not the same. Besides, the rain water and river water are only well mixed in the vertical direction, so the proportions of rain water and river water in input water are the same as the proportions in the lake. Therefore, the calculation of mean deuterium concentrations $c_L(t,\tau)$ in the mixed lake water using Equation (16) and (17) like previous TTD studies will be biased from $c_{L\_1}(t,\tau)$ obtained by equation (15).

Additionally, most lakes or other systems do not meet the two specific situations, so the calculation of $c_L(t,\tau)$ requires tracking the transport of river and rain water in the lake.

Text S9. Derivation of the weighted averaged absolute SAS function of all water

The absolute SAS function of all water for river outflows is defined:

$$aSAS_{all} = \frac{TTD_{all}}{RTD_{all}} \tag{18}$$

where TTD_{all} is the backward TTD of all water in outflows, and RTD_{all} is the RTD of water. The calculations of absolute SAS functions of river water and rain water are similar to Equation (18).

The TTD of all water is the weighted average TTDs of rain water and river water:

$$TTD_{all} = \frac{F_{out\_rain}}{F_{out}} TTD_{rain} + \frac{F_{out\_river}}{F_{out}} TTD_{rain} TTD_{river} \tag{19}$$

where $\frac{F_{out\_rain}}{F_{out}}$ and $\frac{F_{out\_river}}{F_{out}} TTD_{rain}$ are the fractions of rain water and river water in outflows.

Substitute Equation (19) into Equation (18):

$$aSAS_{all} = \frac{F_{out\_rain}}{F_{out}} \cdot \frac{TTD_{rain}}{RTD_{all}} + \frac{F_{out\_river}}{F_{out}} TTD_{rain} \cdot \frac{TTD_{river}}{RTD_{all}} \tag{20}$$

According to the definition of aSAS function, the TTD_{rain} and TTD_{river} can be calculated as:

$$TTD_{rain} = RTD_{rain} \cdot aSAS_{rain} \tag{21}$$

$$TTD_{river} = RTD_{river} \cdot aSAS_{river} \tag{22}$$

Substitute Equations (21) and (22) into Equation (20):

$$aSAS_{all} = aSAS_{rain} \cdot \frac{RTD_{rain}}{RTD_{all}} \cdot \frac{F_{out\_rain}}{F_{out}} + aSAS_{river} \cdot \frac{RTD_{river}}{RTD_{all}} \cdot \frac{F_{out\_river}}{F_{out}} \tag{23}$$

Therefore, the weights of absolute SAS functions of river water and rain water to calculate absolute SAS function of all water are $\frac{RTD_{river}}{RTD_{all}} \cdot \frac{F_{out\_river}}{F_{out}}$ and $\frac{RTD_{rain}}{RTD_{all}} \cdot \frac{F_{out\_rain}}{F_{out}}$ respectively.