


RESEARCH LETTER

Open Access



# Comparison between total least squares and ordinary least squares in obtaining the linear relationship between stable water isotopes

Jeonghoon Lee<sup>1\*</sup> , Won Sang Lee<sup>2</sup>, Hyejung Jung<sup>1</sup> and Seung-Gu Lee<sup>3</sup>

## Abstract

The linear relationship between two stable water isotopes ( $\delta D$  and  $\delta^{18}O$ ) has been used to examine the physical processes and movements or changes of three water phases (water vapor, liquid water and ice), including deuterium excess. The ordinary least squares (OLS) method has been the most commonly used method to fit the linear relationship between two isotopic compositions of water. However, an alternative method, the total least squares (TLS) method, has been proposed because it considers the presence of errors in the explanatory variable (horizontal axis,  $\delta^{18}O$ ). However, not many studies have examined the differences of the relationship using two stable isotopes between the OLS and TLS for various types of water. In this work, these two methods were compared using isotopic compositions of three types of water (Antarctic snow, water vapor and summer and winter rainfall). Statistically, the slopes and intercepts obtained by the two linear regression methods were not significantly different except for summer rainfall, which has the smallest coefficient of variations ( $R^2$ ). The TLS method produced larger slopes than the OLS method and the degrees of difference between the two methods were greater when the coefficient of variation was lower. In addition, with a Monte Carlo method, we showed that the differences between the two methods increased as the uncertainty increased. Moreover, the results of Bayesian linear regression were consistent with the two linear regressions. Although the TLS method is theoretically more suited to the linear regression for the stable water isotopes than the OLS method is, the application of the widely used OLS method can be recommended in the case of small measurements uncertainties after testing whether the linear parameters, slopes and intercepts, derived from the two methods are statistically significant different.

**Keywords:** Ordinary least squares, Total least squares, Stable water isotopes, Monte Carlo

## Background

Stable isotopic measurements of water are helpful for quantifying global or local distributions of exchange processes between water vapor, liquid water and ice. One of the isotopic techniques widely used in isotope hydrology is to investigate the slope of the  $\delta D$  vs.  $\delta^{18}O$

regression line and deuterium excess (*d-excess*), defined as  $\delta D - 8 \times \delta^{18}O$  (Dansgaard 1964a, b). The slope and intercept of the linear regression may reveal evidence of water movement, moisture source regions, groundwater recharge processes, isotopic exchange among water vapor, liquid water and ice and so on (Lee et al. 1999, 2010). The *d-excess* is useful for identifying moisture source regions (Merlivat and Jouzel, 1979). Data from the Global Network of Isotopes in Precipitation (GNIP) database have been investigated to obtain a linear relationship of the Global Meteoric Water Line (GMWL)

\*Correspondence: jeonghoon.d.lee@gmail.com

<sup>1</sup> Department of Science Education, Ewha Womans University, Seoul 120-750, Korea

Full list of author information is available at the end of the article

or Local Meteoric Water Line (LMWL). The GMWL ( $\delta D = 8 \times \delta^{18}O + 10$ ), defined by Craig (1961) and LMWLs investigated by many works, have been derived using an ordinary least squares (OLS) method (Crawford et al. 2014). Typically, the evaporation of soil or lake water results in a linear slope less than  $\sim 8$ , the slope of the GMWL or LMWL.

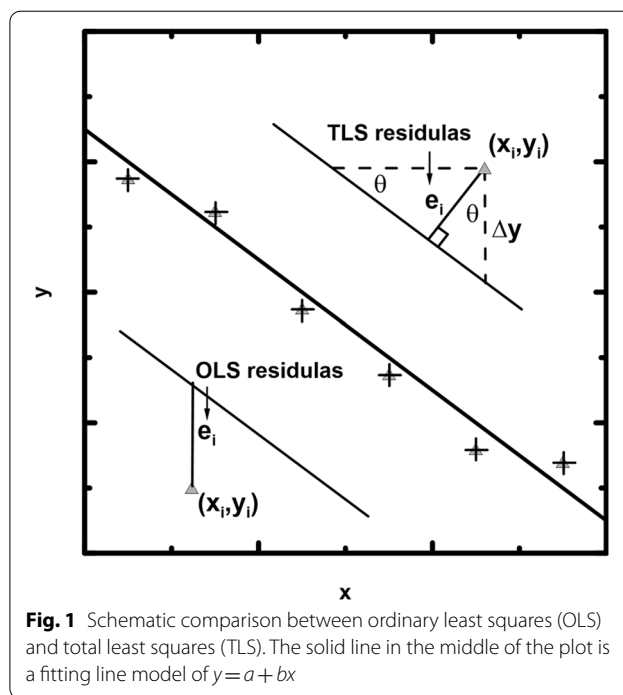
Craig (1961) presented a key finding concerning the distribution of isotopic ratios in precipitation. He noted that the global isotopic compositions of precipitation ( $\delta D$  and  $\delta^{18}O$ ) are highly correlated and plots along a regression slope of 8, which defines the GMWL. The slope value of the GMWL approximates the ratio of the equilibrium fractionation factors between liquid water and water vapor for D and  $^{18}O$ ,

$$\text{Slope} = \frac{\ln(\alpha D_{w-v})}{\ln(\alpha^{18}O_{w-v})}, \tag{1}$$

where  $\alpha^{18}O_{w-v}$  and  $\alpha D_{w-v}$  are the equilibrium isotopic fractionation factor between water and water vapor for oxygen and hydrogen, respectively. The value of 8 is an approximation of the  $\frac{\ln(\alpha D_{w-v})}{\ln(\alpha^{18}O_{w-v})}$  ratio as observed in the

global meteoric waters. The +10‰ value of *d-excess* in the GMWL indicates that at a global scale, the kinetic evaporation of water vapor from oceanic water occurring. The non-equilibrium evaporation of ocean water under relative humidity conditions of less than 100% produces water vapor. This water vapor is depleted from the parent water yet displaced above the meteoric water line due to the enhanced diffusion of HDO (mass of 19) over  $H_2^{18}O$  (mass of 20) from the water-surface boundary layer ( $\frac{\ln(\alpha D_{w-v})}{\ln(\alpha^{18}O_{w-v})} < 8$ ; kinetic fractionation). Accordingly, variations in *d-excess* in precipitation reflect changes in the relative humidity of the air in the source area. Any process that changes two water isotopes along a slope with a value not equal to 8 would cause a change in the value of *d-excess* and thus a loss or misinterpretation of its sources information (Lee et al. 2009).

In many scientific studies, the OLS method has not been considered appropriate to define linear relationships as it is assumed that there are no measurement errors associated with the explanatory variable (horizontal axis or *x*-axis) (Keleş 2018). Conversely, alternative regression methods, such as the total least squares (TLS) method and linear regression model using a Bayesian approach, consider the presence of errors in the explanatory variable. In the TLS, the orthogonal (perpendicular) distances from the regression line to the data points are minimized (Fig. 1). The crucial difference between the OLS and TLS methods is that the former minimizes the error only for the vertical variable whereas the latter



minimizes the errors in both the horizontal and vertical directions (Markovsky and Van Huffel 2007). In principle, this makes the TLS approach more suitable for interpolating isotopic data as the two stable water isotopes are independent of each other. Recently, it has been shown that a linear regression model using a Bayesian approach can be used to determine the distribution of the regression parameters (Stow et al. 2006).

Obtaining a reliable slope and intercept for the linear relationships for various water types is particularly crucial for the studies, focusing on groundwater recharge, evaluating the effect of the evaporation processes of various water types, examining water sources using mixing calculations and differentiating isotopic exchange among water vapor, liquid water and ice (Lee et al. 2009, 2010; Earman et al. 2006). Thus, the objectives of this study are to: (1) evaluate the difference in the slopes and intercepts of the linear relationship between two water isotopes computed by the OLS and TLS methods for various water types (i.e., snow, ice and meltwater from Antarctica, water vapor and precipitation from volcanic island); (2) investigate whether the differences in the slopes and intercepts calculated by the OLS and a linear regression model using a Bayesian approach are significant; and (3) explore a cause of the difference in the slopes and intercepts between the OLS and TLS methods.

**Methodology**

**Ordinary least squares (OLS) vs. total least squares (TLS)**

A linear relationship between  $y$  and  $x$  can be expressed for each data pair as

$$\hat{y} = a + bx \tag{2}$$

where  $b$  is the slope and  $a$  is the intercept and the hat over the  $y$  in  $\hat{y}_i$  indicates that it is predicted value of  $y$ . Then, we assume that  $y$  is linked to  $x$  by

$$y_i = a + bx_i \tag{3}$$

where  $\Delta \varepsilon_i \sim N(0, \sigma^2)$  is referred to as an “error” or “residual”, which is the departure of an actual  $y_i$  from the value of  $y_i$  predicted using Eq. (2) and the sum of  $\Delta y_i$ ’s is zero. In OLS,  $x$  is not subject to error; i.e., all departures of the data from the straight line are caused by errors in  $y$ . Then the error is

$$\Delta \varepsilon_i = y_i - (a + bx_i) \tag{4}$$

Detailed explanations of OLS can be found in many statistical textbooks and only a brief overviews is presented here. Geometrically, the error in the OLS is the vertical distance between  $y_i$  (observed) and  $\hat{y} = a + bx_i$  (predicted) (see Fig. 1). In OLS, we can find the two variables,  $a$  and  $b$  that minimize the sum of the squares of the vertical distances between the line and the data; i.e., the quantity will be minimized, so the OLS solution is given by

$$a = \bar{y} - b\bar{x} \tag{5a}$$

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \tag{5b}$$

where  $\bar{x}$  and  $\bar{y}$  are the means of the  $x_i$  and  $y_i$  values, respectively.

In TLS, the error is the sum of Euclidean distances from the points (observed) to the regression line (predicted). The analogous estimator for the TLS put the cloud of measured  $(x_i, y_i)$  as close as possible to the regression line using a different measure of distance, in this case, the perpendicular distance,  $R_{\text{TLS}}$ , which can be written as

$$R_{\text{TLS}}(a, b) = \Delta \varepsilon_i \cos \theta = \frac{\Delta y_i}{\sqrt{1 + \tan^2 \theta}} = \frac{\Delta y_i}{\sqrt{1 + b^2}} \tag{6}$$

where we have identified the slope of the desired line as  $b = \tan \theta$  and  $\Delta y_i$  as the usual OLS vertical residual. With this notation, the TLS estimators  $(a, b)$  are found by

minimizing, which differs from the OLS only in the pre-multiplier. Therefore, the TLS solution is given as

$$a = \bar{y} - b\bar{x} \tag{7}$$

which is the same as the OLS solution. We can eliminate  $a$  by substituting the result from Eq. (11) into Eq. (10b). Rewriting the equation yields a quadratic equation in  $b$

$$0 = \beta_2 b^2 + \beta_1 b + \beta_0 \tag{8}$$

with coefficients

$$\beta_2 = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \tag{9a}$$

$$\beta_1 = \sum_{i=1}^n [(x_i - \bar{x})^2 - (y_i - \bar{y})^2] \tag{9b}$$

$$\beta_0 = - \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \tag{9c}$$

The value of  $b$  can be expressed as

$$b = \omega \pm (v - u) \sqrt{w^2 + 1} \tag{10a}$$

where

$$u = \sum_{i=1}^n (x_i - \bar{x})^2, \quad v = \sum_{i=1}^n (y_i - \bar{y})^2, \tag{10b}$$

$$w = \frac{v - u}{2r}, \quad r = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

The standard error of the slope and the intercept can be calculated via the following the equations.

$$SE_{\text{slope}} = \frac{\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y})^2}{n-2}}}{\sqrt{\sum_{i=1}^n (x_i - \hat{x})^2}} \tag{11a}$$

$$SE_{\text{intercept}} = SE_{\text{slope}} \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}} \tag{11b}$$

**Linear regression model using Bayesian approach**

Bayes’ theorem can be summarized by

$$\text{posterior} \propto \text{prior} \times \text{likelihood}. \tag{12}$$

Therefore, it is necessary to determine the likelihood and decide on the prior for the linear model. The linear relationship in Eq. (2) under the assumption that  $\Delta$  is distributed normally, the likelihood function is

$$p(y_i|x_i, a, b, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - (a + bx_i))^2}{2\sigma^2}\right) \tag{13}$$

It is assumed that the joint prior distribution of  $a$ ,  $b$  and  $\sigma^2$  is proportional to the inverse of  $\sigma^2$ , which can be expressed as following:

$$p(a, b, \sigma^2) \propto \frac{1}{\sigma^2} \tag{14}$$

Finally, the posterior distribution of  $b$  conditioning on  $\sigma^2$  is

$$b|\sigma^2, \text{ data} \sim \text{Normal}\left(\hat{\beta}, \frac{\sigma^2}{S_{xx}}\right), \tag{15}$$

where  $\hat{\beta}$  is the slope estimate using the OLS and  $S_{xx} = \sum_i (x_i - \bar{x})^2$ . The posterior distribution of a conditioning on  $\sigma^2$  is

$$a|\sigma^2, \text{ data} \sim \text{Normal}\left(\hat{\alpha}, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right), \tag{16}$$

In this work, we compare the linear relationship between two isotopes determined using OLS and the Bayesian approach, as computed with the Markov chain Monte Carlo (MCMC) package in Stan.

## Results

### Data sets used in this work

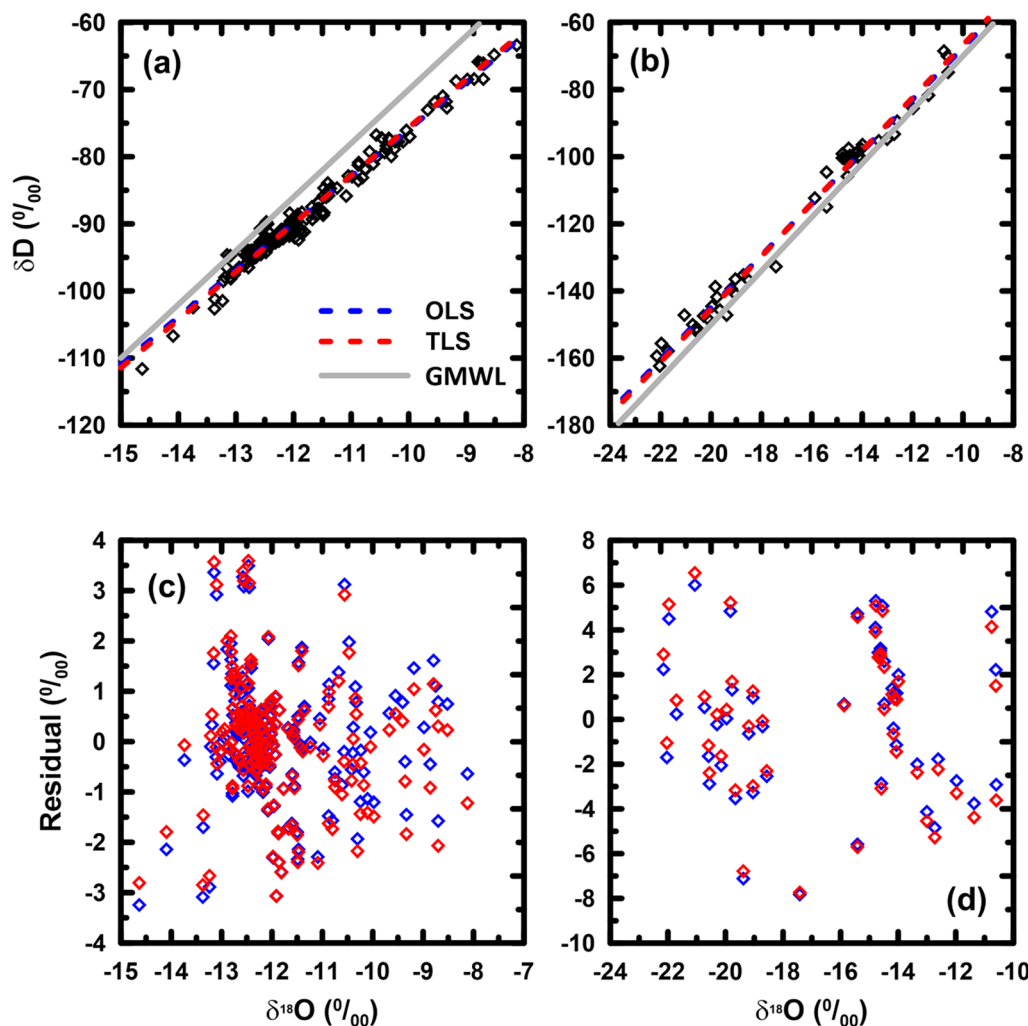
The isotopic ratios are expressed in the  $\delta$  notation as differences in parts per thousand relative to Vienna Standard Mean Ocean Water (VSMOW),  $\delta = [(R_x/R_s) - 1] \times 1000$ . Here  $R_x$  is the isotopic ratio  $^{18}\text{O}/^{16}\text{O}$  or D/H of the water sample and  $R_s$  is the isotopic ratio of the VSMOW. For this study, the following three data sets were used to perform comparisons between the OLS and TLS methods:

1. Isotopic compositions of snow and meltwater from the Barton peninsula, Antarctica were selected (Lee et al. 2020). Snow ( $n=62$ ) and meltwater ( $n=116$ ) samples were collected for isotopic analysis from 9 to 30 January, 2014. Compared to the Global Meteoric

Water Line (GMWL,  $\delta\text{D} = 8 \times \delta^{18}\text{O} + 10$ ), the linear slope between the oxygen and hydrogen isotopes of snow and meltwater was found to be 7.0 using OLS, which is significantly less than the GMWL (Fig. 2a).

2. Isotopic compositions of water vapor observed by Lee et al. (2013) were chosen. Continuous monitoring of water vapor isotopes before and after Typhoon Bolaven from 27 to 29 August, 2012 was conducted and a largely depleted isotopic ratios in surface water vapor were observed in association with the passage of the Typhoon Bolaven. The linear slope between oxygen and hydrogen isotopes of water vapor was 7.8 as determined by OLS, which is similar to the GMWL (Fig. 2b).
3. Precipitation isotopes from Jeju volcanic island, located about 100 km off the southwestern tip of the Korean peninsula, were selected for the comparison of linear relationship between OLS and TLS (Lee et al. *in prep*). The samples were collected from fifteen sites distributed all over Jeju island every month between September 2000 and December 2003 (Fig. 3). The isotopic compositions of samples were determined using a stable isotope ratio mass spectrometer (Isoprime model, GV Instruments) at the Korea Basic Science Institute. The analytical precisions for the oxygen and hydrogen isotopes were less than  $\pm 0.1\%$  and less than  $\pm 1.0\%$ , respectively. Figure 3 shows the linear relationships between two water isotopes in both Case 1 (Antarctic snow and snowmelt) and Case 2 (water vapor isotopes) using the OLS and TLS methods. In Case 1, the  $\text{slope}_{\text{OLS}}$  and  $\text{intercept}_{\text{OLS}}$  differed from those of the GMWL (Table 1). Lee et al. (2020) concluded that the linear isotopic relationship for the whole samples (snow and snowmelt) indicates that the original snow experienced isotopic fractionation through significant melting ( $\text{slope}_{\text{OLS}} \sim 7.0$ ). The  $\text{slope}_{\text{OLS}}$  and  $\text{intercept}_{\text{OLS}}$  of Case 2 are 7.8 and 10.1, respectively, which is close to the LMWL (Lee et al. 1999,  $\delta\text{D} = 7.9 \times \delta^{18}\text{O} + 8.8$ ).

In Case 3 (precipitation from Jeju volcanic island), two different LMWLs can be drawn using the OLS to describe the isotopic data for different seasons, for instance, summer (June, July and August) and winter precipitation (December, January and February) on the island (Fig. 3). The two LMWLs were  $\delta\text{D} = 8.1 \times \delta^{18}\text{O} + 7.9$  and  $\delta\text{D} = 7.9 \times \delta^{18}\text{O} + 21.0$  for summer and winter, respectively. The values of *d-excess* for summer precipitation



**Fig. 2** **a** The linear relationship between oxygen and hydrogen isotopes of snow and snowmelt sampled from the King Sejong Station obtained using OLS and TLS. **b** The linear relationship between oxygen and hydrogen isotopes of water vapor obtained using OLS and TLS (Lee et al. 2013). **c** A comparison of the residuals of oxygen isotopes obtained using OLS and TLS. **d** A comparison of the residuals oxygen isotopes obtained using OLS and TLS

(~ 8‰) were clearly distinct from those of winter precipitation because of different origin of air masses between two seasons (> 20‰).

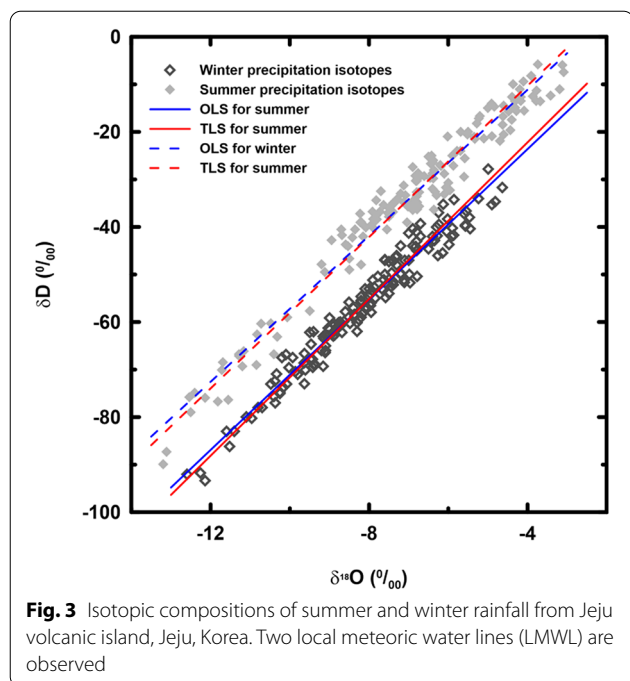
**Discussion**

**Linear relationship between two water isotopes**

Evaporated or convected water vapor over the ocean can be transported to the atmosphere and condensed and precipitated in the form of rain and snow, which results in the depletion of isotopic compositions ( $\delta D$  and  $\delta^{18}O$ ) compared to the original value of the water vapor. The isotopic linear relationship between oxygen and hydrogen of water originating from the ocean has a slope of 8 and an intercept of 10, which is the Global

Meteoric Water Line (GMWL, Dansgaard 1964). Water vapor convected from nearby ocean and transported to polar regions will be precipitated in the form of snow instead of rain. The isotopic compositions of snow also have a slope of around 8 and an intercept of 8~12, depending on relative humidity. In Antarctica, the LMWL determined by Masson-Delmotte et al. (2008) is  $7.75 (\pm 0.02) - 4.93 (R^2 = 0.998, n = 789)$ .

Typically, the evaporation of soil or lake water results in a slope less than ~8, which is the slope of the local meteoric water line (LMWL). However, processes other than evaporation, such as an isotopic exchange between liquid water and ice (melting) or between water vapor and ice (sublimation), may also affect the



slope of  $\delta D$  vs.  $\delta^{18}O$  relationship (Earman et al. 2006). As the isotopic fractionation between liquid water and ice is 3.1‰ for oxygen and 19.5‰ for hydrogen, respectively, Lee et al. (2009) and Lee et al. (2010) predicted and demonstrated that the slope between the two water isotopes of snowmelt would be close to  $19.5/3.1 \approx 6.3$ . Near the surface, snow undergoes isotopic exchanges

with atmospheric water vapor (isotopic fractionation between water vapor and ice, 88.2‰ for hydrogen and 11.4‰ for oxygen, respectively), which can yield an ice–vapor relationship of  $88.2/11.4 \approx 7.7$  at equilibrium ( $0^\circ C$ ).

**Differences between the OLS and TLS**

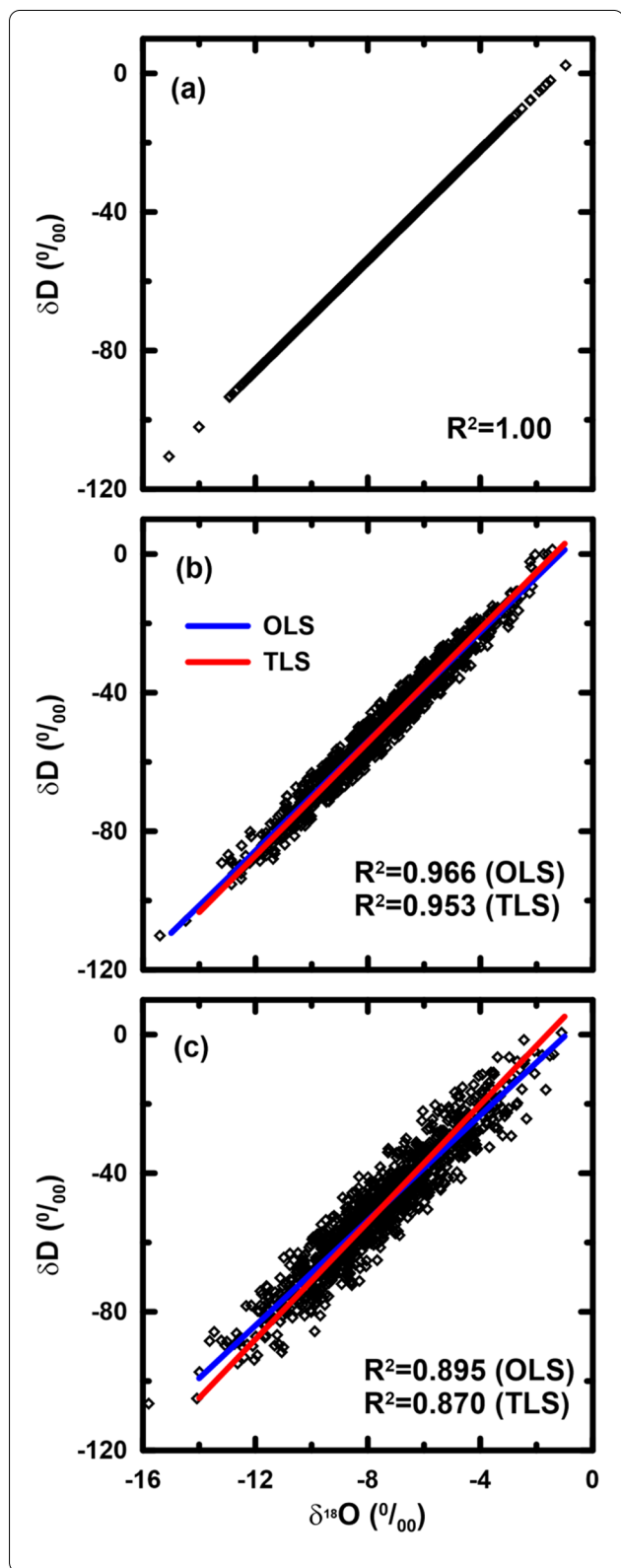
For all three cases with the two regression methods, the meteoric waters (snow, water vapor and rainfall) have statistically similar slopes and intercepts for OLS and TLS ( $H_0 = TLS > OLS$ , one tailed) except for those of summer for Case 3 (Table 1). A similar distribution of residuals ( $y_i - \hat{y}_i$ ) was observed in Cases 1 and 2 (Fig. 2c, d). The probability ( $p$ ) of the differences in slope for the three cases obtained using the two methods are 0.08 (Case 1), 0.27 (Case 2) and 0.04 (Case 3, summer rainfall) and the differences in intercept are 0.08 (Case 1), 0.28 (Case 2) and 0.03 (Case 3 summer rainfall) and 0.10 (Case 3 winter rainfall), respectively. Again, the slopes and intercepts obtained using the two linear regressions methods are not statistically different except for Case 3 summer rainfall.

Compared to other cases, Case 3 summer rainfall has the smallest  $R^2$  values, that is, coefficient of variation or correlation (0.956, Table 1). Our results indicate that the magnitude of the differences in the slope and intercepts obtained by the two methods are primarily due to the coefficient of variation ( $R^2 = 1 - \frac{\sum \epsilon_i}{\sum (y_i - \bar{y})^2}$ ) computed

**Table 1** Comparisons between parameters obtained using ordinary least squares (OLS) and total least squares method (TLS)

Data set	Parameters	OLS	TLS
Antarctic snow (Case 1, $n = 178$ )	Slope $\pm$ standard error	$7.00 \pm 0.08$	$7.16 \pm 0.08$
	Intercept	$- 5.89 \pm 0.94$	$- 4.05 \pm 0.95$
	$R^2$	0.978	0.977
Water vapor (Case 2, $n = 48$ )	Slope	$7.75 \pm 0.14$	$7.87 \pm 0.14$
	Intercept	$10.06 \pm 2.40$	$12.01 \pm 2.42$
	$R^2$	0.985	0.984
Precipitation of Jeju island ( $S^*$ , $n = 156$ ; $W^*$ , $n = 147$ , Case 3)	Slope	$8.09 \pm 0.14$ (S)	$8.45 \pm 0.15$ (S)
		$7.86 \pm 0.11$ (W)	$8.09 \pm 0.13$ (W)
	Intercept	$9.56 \pm 1.16$ (S)	$12.54 \pm 1.16$ (S)
		$21.03 \pm 0.82$ (W)	$22.60 \pm 0.94$ (W)
	$R^2$	0.956 (S)	0.954 (S)
		0.972 (W)	0.962 (W)
Random samples (1, 2, $n = 1000$ )	Slope	$7.9 \pm 0.05$ (1)	$8.18 \pm 0.05$ (1)
		$7.59 \pm 0.08$ (2)	$8.47 \pm 0.09$ (2)
	Intercept	$9.14 \pm 0.37$ (1)	$11.17 \pm 0.37$ (1)
		$7.08 \pm 0.65$ (2)	$13.67 \pm 0.68$ (2)
$R^2$	0.966 (1)	0.964 (1)	
	0.895 (2)	0.870 (2)	

\* S and W denote summer and winter, respectively



**Fig. 4** Results of a Monte Carlo experiments. **a** Global Meteoric Water Line (GMWL) of oxygen and hydrogen isotopes with mean values and one standard deviation; the values were  $-7.5 \pm 2.21$  and  $-50.1 \pm 17.65$ , respectively. **b** New data set generated from the distribution from the previous one (means values with one standard deviation for oxygen and hydrogen,  $-7.5 \pm 2.23$  and  $-50.2 \pm 17.90$ ). **c** Another data set, means values with one standard deviation for oxygen and hydrogen,  $-7.5 \pm 2.25$  and  $-50.2 \pm 18.05$

between two stable water isotopes. When considerable vertical scatter was present in the relationship between the two water isotopes, larger differences in the slopes were observed between the two regression methods. The TLS method is more sensitive to outliers and as a result, it appears that the TLS produced large slopes when the coefficient of variation was lower. The mean values of the slopes obtained using TLS are statistically significantly larger than those obtained using OLS (*t* test, one-tailed,  $p < 0.05$ ).

A Monte Carlo method was used to test how the coefficient of variation affects the difference in the slopes and intercepts between the two regressions (Anderson 1976). A data set following the GMWL ( $\delta\text{D} = 8 \times \delta^{18}\text{O} + 10$ , Fig. 4a) was created and then another two data sets were generated using the means of this data set and added uncertainties. As the uncertainty increased, the coefficient of variation decreased as shown in Fig. 4. As the coefficient of variation decreased, the slopes and intercepts were deviated more from the GMWL and the standard errors of the slopes and intercepts were increased using the two regressions. The slopes obtained using the TLS method are larger than those obtained the OLS method in the two calculations (Random samples 1 and 2 in Table 1). The degrees of the deviations from the GMWL for the TLS method are greater than those for the OLS method. With this experiment, the significant difference in slope obtained for Case 3 summer rainfall can be explained by the fact that the coefficient of variation controls the differences in the linear slopes and intercepts.

#### Comparison with the Bayesian linear regression

In the Bayesian perspective, the linear regression can be formulated using probability distributions rather than point estimates. The aim of Bayesian linear regression is not to find the single “best” value of the regression parameter, but rather to determine the posterior distribution for the model parameters (Bolstad and Curran 2016). By comparing the mean values for the slope and

**Table 2** Bayesian results for Antarctic snow (Case 1)

	Mean	Standard deviation	Quantiles (2.5%)	Quantiles (25%)	Quantiles (50%)	Quantiles (75%)	Quantiles (97.5%)
Intercept	- 5.899	0.949	- 7.748	- 6.536	- 5.892	- 5.262	- 4.038
Slope	7.003	0.080	6.847	6.950	7.004	7.057	7.160

intercept obtained using Bayesian linear regression to those obtained using OLS, it was found that they are not different. While we can use the mean as a single point estimate, we also have a range of possible values for the regression parameters from the Bayesian perspective (Permai and Tanty 2018).

In this work, the Bayesian linear regression was applied to the Antarctic snow/snowmelt (Case 1) to investigate that different regression method results in changing the physical process of water. The slope of the two water isotopes (Case 1) obtained using OLS was 7.0, which indicates that the original snow experienced isotopic fractionation through significant melting. The Bayesian posterior distribution results of slope (*b*) and intercept (*a*) show that the posterior credible intervals are numerically equivalent to the confidence intervals obtained using the OLS method. Table 2 provides 95% confidence intervals, which coincides with the confidence intervals obtained using the OLS method. However, the primary difference is in the interpretation of the results. Based on the data, there is 95% chance that the slope will fall in the range between 6.85 and 7.16, which does not change the conclusion obtained using the OLS method.

## Summary

In this work, we quantified the differences in the slopes and intercepts of two stable water isotopes computed by the OLS and TLS methods and investigated whether the magnitude of the differences was affected by the coefficient of variation ( $R^2$ ). As expected, based on the intrinsic mathematical characteristics of the two methods, we found that the TLS method always produced the larger slopes and intercepts than the OLS method for three water types, Antarctic snow/snowmelt, water vapor and summer and winter rainfall. The slopes and intercepts obtained using the two linear regression methods are not statistically different except for the summer rainfall, which has the smallest coefficient of variations ( $R^2$ ). With the Monte Carlo method, we showed that the differences between the two methods increased as the uncertainty increased. Furthermore, the results of the Bayesian linear regression were consistent with those of the two linear regression methods.

Based on our findings, regarding isotope hydrology, we suggest that researchers should consider the

measurement uncertainties for both  $\delta D$  and  $\delta^{18}O$  and to test whether the slopes and intercepts calculated by the OLS and TLS are statistically significantly different. Although the TLS method is theoretically more suited to linear regression for stable water isotopes than OLS, the application of the widely used OLS method can be still legitimate in the case of small measurement uncertainties.

## Acknowledgements

This work was sponsored by a research grant from the Korean Ministry of Oceans and Fisheries (KIMST20190361) and partially supported by the principal Research Fund of the Korea Institute of Geoscience and Mineral Resources (GP2017-018).

## Authors' contributions

JL: Analysis, interpretation and writing the manuscript. WSL: Initiation of this study. HJ: Implementation of the reference collection. SGL: Constructive comments and funding. All authors read and approved the final manuscript.

## Funding

This work was sponsored by a research grant from the Korean Ministry of Oceans and Fisheries (KIMST20190361) and partially supported by the principal Research Fund of the Korea Institute of Geoscience and Mineral Resources (GP2017-018).

## Availability of data and materials

The data used in this work may be obtained from the paper indicated in the manuscript.

## Declarations

### Competing interests

The authors declare that they have no competing interest.

### Author details

<sup>1</sup>Department of Science Education, Ewha Womans University, Seoul 120-750, Korea. <sup>2</sup>Korea Polar Research Institute, Incheon 21990, Korea. <sup>3</sup>Korea Institute of Geoscience and Mineral Resources, Daejeon 34132, Korea.

Received: 15 September 2020 Accepted: 26 January 2022

Published online: 22 February 2022

## References

- Anderson GM (1976) Error propagation by the Monte Carlo method in geochemical calculation. *Geochim Cosmochim Acta* 40:1533–1538
- Bolstad WM, Curran JM (2016) Introduction to Bayesian statistics. Wiley, New Jersey
- Craig H (1961) Isotopic variations in meteoric waters. *Science* 133(3465):1702–1703
- Crawford J, Hughes CE, Lykoudis S (2014) Alternative least squares methods for determining the meteoric water line, demonstrated using GNIP data. *J Hydrol* 519:2331–2340
- Dansgaard W (1964) Stable isotopes in precipitation. *Tellus* 16:436–468



- Earman S, Campbell AR, Phillips FM, Newman BD (2006) Isotopic exchange between snow and atmospheric water vapor: estimation of the snow-melt component of groundwater recharge in the southwestern United States. *J Geophys Res* 111:D09302. <https://doi.org/10.1029/2005JD006470>
- Gautam MK, Lee K-S, Bong Y-S, Song B-Y, Ryu J-S (2017) Oxygen and hydrogen isotopic characterization of rainfall and throughfall in four South Korean cool temperate forests. *Hydrolog Sci J* 62(12):2025–2034
- Hollins SE, Hughes CE, Crawford J, Cendon DI, Meredith KT (2018) Rainfall isotope variations over the Australian continent—implications for hydrology and isoscape applications. *Sci Total Environ* 645:630–645
- Keleş T (2018) Comparison of classical least squares and orthogonal regression in measurement error models. *Int Online J Educ Sci* 10(3):200–214
- Lee K-S, Wenner DB, Lee I (1999) Using H- and O-isotopic data for estimating the relative contributions of rainy and dry season precipitation to groundwater: example from Cheju Island, Korea. *J Hydrol* 222:65–74
- Lee J, Feng X, Posmentier E, Faiia A, Taylor S (2009) Stable isotopic exchange rate constant between snow and liquid water. *Chem Geol* 260:57–62
- Lee J, Feng X, Faiia A, Posmentier E, Kirchner J, Osterhuber R, Taylor S (2010) Isotopic evolution of a seasonal snowcover and its melt by isotopic exchange between liquid water and ice. *Chem Geol* 270:126–134
- Lee J, Choi H, Oh J, Na US, Kwak H, Hur SD (2013) Moisture transport observed by water vapor isotopes in the vicinity of coastal area, Incheon, Korea. *Econ Environ Geol* 46:339–344
- Lee J, Hur SD, Lim HS, Jung H (2020) Isotopic characteristics of snow and its meltwater over the Barton Peninsula, Antarctica. *Cold Reg Sci Technol* 173:102997
- Markovsky I, Van Huffel S (2007a) Overview of total least square methods. *Signal Process* 87(10):2283–2302
- Masson-Delmotte V, Hou S, Ekaykin A et al (2008) A review on Antarctic surface snow isotopic compositions: observations, atmospheric circulation and isotopic modeling. *J Clim* 21(13):3359–3387
- Merlivat L, Jouzel J (1979) Global climatic interpretation of the deuterium-oxygen 18 relationship for precipitation. *J Geophys Res* 84:5029–5033
- Permai SD, Tanty H (2018) Linear regression model using Bayesian approach for energy performance of residential building. *Procedia Comput Sci* 135:671–677
- Pospiech S, Tolosana-Delgado R, van den Boogaart KG (2020) Discriminant analysis for compositional data incorporating cell-wise uncertainties. *Math Geosci*. <https://doi.org/10.1007/s11004-020-09878-x>
- Stow CA, Reckhow KH, Qian SS (2006) A Bayesian approach to retransformation bias in transformed regression. *Ecology* 87:1472–1477

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---