Response to 2nd Referee's Comments on
What controls the stable isotope composition of precipitation in the Asian monsoon region?
by Le Duy Nguyen et al

General Comments:

In this paper, the authors used their new weekly precipitation isotope dataset in Vietnam's Mekong river delta region for 1.5 years, and they tried to reveal the controls of the temporal variation of the precipitation isotope ratio. To do so, they conducted some statistical analyses, and they concluded that the isotope ratio is controlled by mainly regional scale phenomena (mainly by the previous rainfall activity along the trajectory of air mass) especially during the early rainy season, and the contribution of the control varies by season.

We thank the second anonymous referee for the comments, to which we reply below (in blue). Our answers will be included in a revised version of the manuscript.

In my opinion, even though they conducted multiple methods, nothing is quite new. The control of precipitation isotope had been discussed by many researchers as the authors mentioned, and the authors' findings were already pointed out by many, too. For example, the quantification of the controls was attempted by several model studies including Yoshimura et al., 2003; Risi et al., 2008; Kurita et al., 2011; Ishizaki et al., 2012; etc.

First of all, we would like to emphasize that the main objective of this study is to develop an approach to quantitatively estimate the relative contribution of regional and local factors controlling the isotopic variation of precipitation. The proposed approach is based on multiple linear regression (MLR) specifically considering the widespread issue of multicollinearity of the regression factors, in combination with a regression factor importance analysis.

We acknowledge that our methods (trajectory analysis, multiple regression and relative importance analysis) are simple and easy to apply and that each of it has already been used in previous studies. However, to our knowledge, the combination of these methods to investigate factors controlling isotopic composition in precipitation has never been applied before. In our opinion a study based on simple methodology is better than a study based on complex methods containing a larger number of uncertainty sources, if similar results are obtained (in line with the concept of parsimonious modelling). Any scientist can easily apply our method in order to investigate factors controlling isotopic composition in precipitation at any given study area around the world without the requirement of setting up and running a complex numerical circulation model. This is the novelty of this study. In the paper we already acknowledge that we don't come to new conclusion regarding the factors controlling the isotopic composition in rainfall in tropical areas, but we present a method that is universal and easy to apply (as mentioned above), and delivering a solid and reproducible quantitative analysis of the contribution of different factors.

The comment of the first anonymous referee: *"This study, like many previous studies, shows that local rainfall amount and temperature play a minor role in controlling the isotopic composition of the rainfall with upstream precipitation amount emerging as the dominant regional control again a result consistent with previous studies, but the author's conclusion is backed by solid quantitative analysis."* supports our point of view.

Moreover, this study quantitatively focuses on the interplay of various factors controlling isotopic composition in precipitation which has also never been studied before. The relative importance of these factors in controlling isotopic composition in precipitation has not been quantified as presented although first steps in this direction were taken by Ishizaki et al. (2012). However, Ishizaki et al. (2012) limited the analysis to two factors only (local precipitation amount and distillation of the moisture along its transport trajectories).

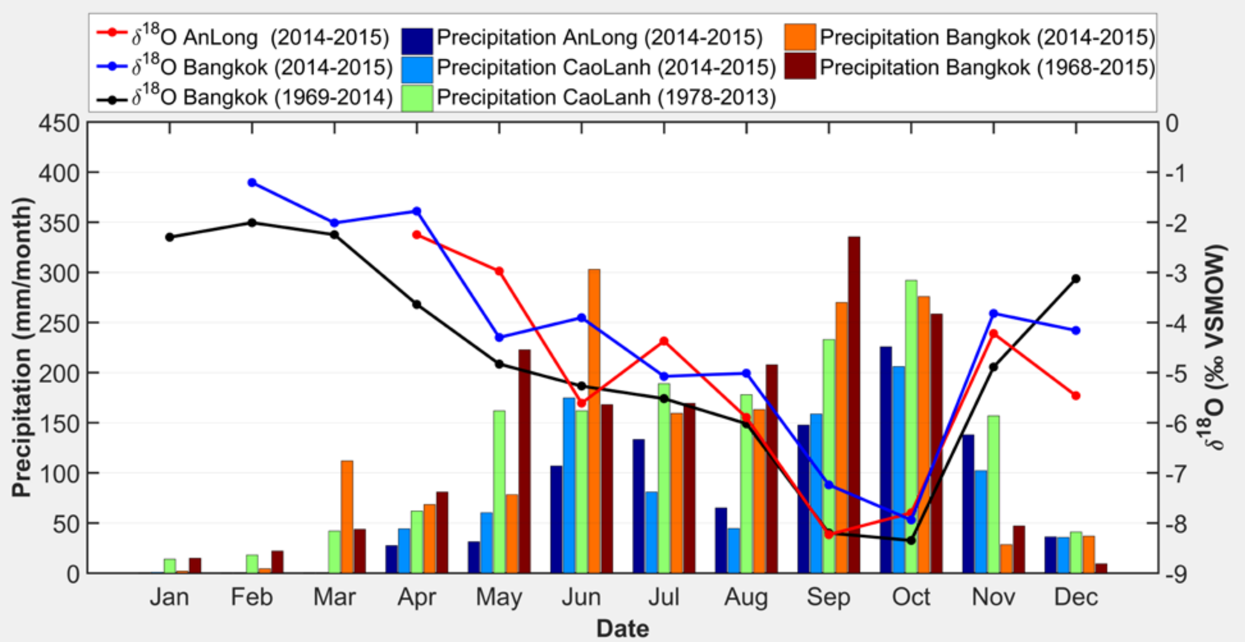
Some of these studies do not necessarily focus only Asian monsoon regions, but basically, they tried to reveal more general controls. In these studies, they used GCM or equivalent models to reveal the controls, whereas the present paper used statistical models.
Indeed we have taken climate reanalysis data derived from circulation models, extracted water transport trajectories by Lagrangian backtracing (HYSPLIT), and analyzed them with statistical models and relative importance analysis. So the presented study is not just a simple statistical data analysis. In fact, in can be seen as a substantial extension of the approach of Ishizaki et al. (2012).

Furthermore, by the recent efforts, researchers already began to realize that it is indeed not appropriate to make a simple relationship between precipitation isotopes and climate parameters. The present paper's conclusion of necessity of consideration of multiple climate impacts and temporal (and spatial) dependency on the controls have been explicitly or implicitly stated many times. Therefore, nowadays, more advanced techniques of utilization of isotopic information have been utilized. One of them is data assimilation.
As mentioned above, we acknowledge the fact that the results are not new, and that the focus of the paper is the development and testing of the combined method instead. We assume that you refer to the assimilation of data in atmospheric circulation models, which explicitly simulate the separation of water isotopes in the hydrological cycle. Of course, this would be one way to use the data and to derive information about the dominating factors for isotopic composition of rainfall. Indeed, these models could provide much more detailed information about the fractionation processes along the transport pathways of water in the atmosphere. However, the complexity of this approach is much higher compared to the one we propose, and requires in-depth knowledge about atmospheric modelling and data assimilation. In any case it would take much more effort to establish such a system if it is not already present. That means that such an approach is rather for the specialists in modelling of isotope-enabled general circulation models. For an application in a study in another field, as e.g. the mentioned paleo-climate studies, we believe that our proposed approach would be much more suitable. Besides this, even if circulation models are directly used, it is not straight forward to extract the impact of the different factors from the complex models and weight their relative importance. Some statistical procedure surely needs to be applied to come to similar conclusion as provided by our method. There are a lot of studies using isotope-enabled global climate models (GCMs) combined with some statistics to investigate the physical links between climate and water isotopes, e.g. (Vuille et al., 2005;LeGrande and Schmidt, 2009;Tindall et al., 2009;Ishizaki et al., 2012;Conroy et al., 2013). Some studies applied statistics such as principal component analysis (PCA) (Vuille et al., 2003;Curio and Scherer, 2016); machine learning technique random forests (Sánchez-Murillo et al., 2016); sensitivity experiments (Ishizaki et al., 2012) to investigate dominant factors of isotopic composition in precipitation. However, the relative importance of these parameters has not been quantitatively investigated yet (Ishizaki et al., 2012). Moreover, to our best knowledge, there is no study considering the interplay of both local and regional factors in controlling isotopic composition in precipitation, which is carefully taken into account in our study by the relative importance analysis dealing with the multicollinearity of controlling factors.

From the above aspect, I have to tell that this paper's methods (multiple regression and trajectory analysis) is no longer insufficient to fulfill the objectives of this study. What I mean is, there is no guarantee that this study's number of 70% regional control can be applied to any other year's temporal variation of precipitation isotopes. In this regard, 1.5-yr long data is not sufficient, too.
Of course, due to the limited length of the time series we cannot be 100% sure that the identified contribution of local and regional factors will be the same in other years. However, as shown in figure 7, the long term monthly isotopic values in Bangkok and the values of our two rainy seasons in the Mekong delta are quite similar. Considering also the climatic similarities between the two locations, this indicates that the recorded isotopic variation is likely to be representative

for a longer period and a wider area. This suggests in turn that the identified contribution of the factors could also be the same in other years. Also, the fact that our findings agree with the ones of Ishizaki et al. (2012) supports this assumption. Ideally one would perform a similar analysis for Bangkok for longer time series, but this is not possible due to the low resolution (monthly) of the publicly available isotope and rainfall data.

Figure 7 in the manuscript will be edited as follows to include also the short-term mean monthly isotopic signature of precipitation of Bangkok:



Major issue:

1. Drop unnecessary and unrelated analyses. Especially the parts with local meteoric line is not directly related to the conclusion of the study. It is too simple analysis. Even global meteoric line is just conceptual idea (slope of 8 and intercept of 10 is not certain). There maybe some physical reason to have smaller slope, especially by kinetic effect, but in this study, it is out discussed enough. It's better to drop the part.

You are right that the derivation of a local meteoric water line is a very simple analysis. We still think it provides valuable information for the following reasons:

- From our point of view the analysis of isotopic data in terms of meteoric water lines is a standard for such kind of data and should always be conducted, just as descriptive statistics of other data.
- Up to now, there is no LMWL for Vietnamese Mekong Delta (VMD) and Indochinese Peninsula, which can be used as a baseline for other studies using isotopic data to investigate hydrological processes in this area.
- The close fit of all considered regressions is one evidence indicating that secondary fractionation processes, e.g. sub-cloud evaporation, are insignificant in study area. This provides support for the discussion of sub-cloud evaporation in Sec. 4.3.1 in the manuscript.

2. One point data cannot represent Asian monsoon. Perhaps Mekong river delta data had some similarity with Bangkok, but with only 1.5-yr long data, the authors cannot reject possibility of "by chance". Furthermore, such similarity is nothing related to that Mekong data represent all Asian monsoon region. The title is quite misleading.

We acknowledge that the title is too generic. We will change it to "What controls the stable isotope composition of precipitation in the Mekong Delta?" and discuss the transferability to the greater region, i.e. SE-Asia. Actually, isotopic data of rainfall has never been collected for the

Mekong delta, and therefore the fact that the isotopic variation of the Mekong data is similar to that of Asian monsoon region has never been confirmed before.
We also went at length to illustrate that the variability of the isotopic data is similar to the long term data from Bangkok in order to provide evidence that the derived results might be representative for SE-Asia. This is already discussed in section 4.2, but we will add some critical discussion of the issue of representability in the discussion and conclusion of the revised manuscript, stating that there are indications that the obtained results could be representative for the southern part of SE-Asia.

3. Organize the previous literature with focused temporal and spatial scales. The authors listed many previous studies, which partly investigated on precipitation isotope controls, and (implicitly) stated that there is still huge discussion on the controls. However, it is misleading and not true. What is confusing is the controls can be different dependent on temporal and spatial scales. For example, daily variation of precipitation isotopes in some parts of the world is quite likely determined by synoptic-scale moisture circulation, in which previous rainfall activity along the trajectory matters a lot, rather than local precipitation or temperature, and nowadays there is consensus on this in the research community. However, even in the same place, the controls of monthly or interannual time series can be different. It is simply because those smaller scale impact can be offset each other in those scales, so that local signal only remains.
We completely agree that scales matter. This is fundamental to hydrology. What we present is the result for daily variation (or bi-weekly, to be exact) in rainfall, in a monsoonal climate region with a strong seasonal variation. We will stress this more in the discussion and conclusion, and sort the cited literature according to the scales considered.

4. Limitation of statistical approach with such short-term data. The conclusion of the study is based on the statistical regression using all samples. The authors should validate their statistical model(s) with different independent samples. In this regard, the observation data is perhaps too short.

As described in section 3.6, we use PRESS for selecting the best model. Within PRESS the model is fitted to all data except one, and the missing value is predicted with the fitted model, i.e. not all data is used for fitting the models at once. This procedure is repeated for every data point. Thus PRESS is equivalent to a so called leave-one-out cross validation (LOOCV), as described in section 3.6. LOOCV is the cross validation procedure appropriate for a limited data set, when a standard split sample validation cannot be applied. There are numerous papers available employing this method in different fields of environmental sciences. LOOCV is actually a split sample validation of the regression, where the data is split as often as data points are available. This means that our results are in fact validated.

5. Most importantly, what is new in this study? As I wrote above, it is well known that precipitation isotope is not controlled by a single factor and the relationship can be different in time and space. The finding in this paper is nothing more than these.
As we have stated previously, we acknowledge the fact that our methods (trajectory analysis, multiple regression and relative importance analysis) are relatively simple and easy to apply, but we would like to stress again that the combination of these methods to investigate factors controlling isotopic composition in precipitation has never been applied before.
Moreover, our study focuses on the quantification of the impact of the various factors controlling isotopic composition in precipitation. This has not been performed in such an exhaustive way as presented here (as reviewer 1 actually points out particularly). Of course, the qualitative outcome of the study is not novel in itself, but the way we achieved these results constitutes a novel approach. Furthermore, this approach is easily reproducible and contains a rigorous analysis and quantification of the interplay of the different factors. Thus we argue that the manuscript indeed

goes beyond just stating that regional factors are more important than local factors in the daily rainfall isotopic composition for the study region. It rather supports this finding by a thorough and reproducible method that combines circulation modelling and statistical analysis.

Minor issues:

P2L17: what is “circulation effect”? Describe.

The term “circulation effect” (Tan, 2009;Tan, 2014) is used to describe the changes in isotopic composition in precipitation which originate from the changes in Indian/Pacific Ocean atmospheric circulation. We will add this explanation to the manuscript.

P2L23: what is difference between “distillation during vapor transport” and “upstream rainout”. Aren’t they essentially the same?

Yes, thank you. We will use only the term “distillation during vapor transport” in the manuscript.

P2L22-P3L3: Different temporal scales are mixed.

As mentioned above, we will sort the references according to scale.

P3L21: Before the authors’ conclusion, there are many studies which state necessity of consideration of multiple parameters.

Yes, the paragraph is misleading. We will replace the whole paragraph with:

“Since it has been frequently stated and agreed to that local factors (e.g. local rainfall effect or temperature effect) and regional factors (e.g. circulation effect) should be considered simultaneously to explain the isotopic variation in rainfall (e.g. Johnson and Ingram, 2004), it can be hypothesized that using multiple factors in a single linear model is able to explain a larger share of the observed variance in isotopic composition.”

P3L27: For quantification of the controls, usually researchers try to develop a physical simulator. Any statistical model principally cannot explain the real control.

Physical models are one way to address this problem. But statistical models are an alternative way to do this, and have in fact be applied many times in all sorts of environmental studies. Both approaches have their advantages and disadvantages, and they coexist, respectively supplement each other. And while statistical models are not able to represent the actual process causing a phenomenon, they are able to detect results of a process. And this is what we actually are aiming at. The statement on P3L27 expresses just this. Therefore we are arguing that the proposed approach is a) valid, and b) accepted by the majority of researchers, as long as the limitations are clearly taken into consideration. We will underline this point in more detail in the revised manuscript.

P4L20: There are many other definition of dry/wet season. What is the impact?

The definition used here is appropriate for a) the climatic condition, b) the problem to be solved and c) the data available, and is thus reasonable from our point of view. This is also supported by the reasonable results obtained for the two seasons. The impact of other definitions has not been studied in detail, but if anything would change, then some data points would be assigned to the other season. In general we do not expect any significant changes in the results, as long as the definition of the seasons is reasonable, i.e. that samples surely belonging to the dry or wet season are not assigned to the other season. Or expressed in other words: the definition of the seasons will most likely affect the samples from the transition period from one season to the other, i.e. samples that have the least explanatory value for the actual dry and wet seasons.

P5L5: “three methods” are not really regarded as different “method”.

Thank you for this point. “three methods” will be changed to “three regression methods”

P6L4-L20: drop

As discussed in the 1st comment under 'major issues', we consider this part relevant and important for the manuscript.

P7L18: what is TRATIO?

We will modify the sentence from P7L17-L19 as follow:

"Secondly, we use the shortest possible integration time step (i.e. 1 h) and a small value for the parameter TRATIO (0.25), which is the fraction of a grid cell that a trajectory is permitted to transit in one advection time step. Smaller values of TRATIO help to minimize the trajectory computation error using the HYSPLIT model".

P7L20: The uncertainty of trajectory analysis is not quantified. Perhaps it is minimized in the suggested framework, but how large is the "minimized" uncertainty and what is its potential consequence?

This paragraph will be added to the manuscript to discuss about the uncertainty of trajectory analysis.

"While errors in trajectory calculation computed from analyzed wind fields seem to be typical on the order of 20% of the distance travelled (Stohl, 1998), the statistical analysis of a large number of trajectories arriving at a study site would increase the accuracy of the trajectory analysis (Cabello et al., 2008).

Harris et al. (2005) studied trajectory model sensitivity to the input meteorological data (focusing on ERA-40 and NCEP/NCAR reanalysis data) and vertical transport method. They pointed out five causes of trajectory uncertainty, expressed as percentage of deviation of the average travel distance: 1) minor differences in the computational methodology: 3–4%; 2) time interpolation: 9–25%; 3) vertical transport method: 18–34%; 4) meteorological input data: 30–40%; and 5) combined two-way differences in the vertical transport method and meteorological input data: 39–47%. However, it would be difficult to prove that in all situations a single meteorological data set or a single method of trajectory modelling was superior to another one (Gebhart et al., 2005;Harris et al., 2005). More details about the uncertainties in trajectory modelling were provided by Stohl (1998), later by Fleming et al. (2012) and references therein."

P8L4: PRESS is essentially the same as root mean square error (RMSE), which is more popular in the community.

We challenge this view. RMSE is calculated from the residuals of the model fitted to all data, while PRESS is based on the residuals resulting from the model fitted to all data except one, for which the residual is calculated. Repeating this for all data points and summing the calculated residuals results in PRESS. PRESS is therefore a cross validation method. See also our comment above, and for example the definition in WIKIPEDIA as reference (https://en.wikipedia.org/wiki/PRESS_statistic).

P8L5: what is "leave-one-out cross validation"? and what does it mean by "equivalent to" it?

See our reply to major comment 4 and the previous comment.

P8L16: what is physical meaning of using "mean values of their combinations"? Combination of 800hPa and 850hPa represent 825hPa level (somehow the precipitation was formed at that level at that time)? In this regard, what is meaning of 800/850/900hPa combination?

In P7L4-L7 we discuss that three levels at 1000, 1500, and 2000 m above ground are corresponding to barometric surfaces of approximately 900, 850, and 800 hPa. These barometric surfaces were chosen because the 850 hPa vorticity is highly indicative of the strength of the boundary layer moisture convergence and of rainfall in regions away from the equator (Wang et al., 2001). Hence rainfall is expected to mostly originate from these altitudes.

Consequently, the combination of 800 hPa and 850 hPa barometric surfaces accounts for the fact that rainfall is expected to mostly originate between 1500 and 2000 m above ground level. Similarly, the combination of the barometric surfaces of 800, 850 and 900 hPa represents that rainfall is expected to mostly originate between 1000 and 2000 m above ground level.

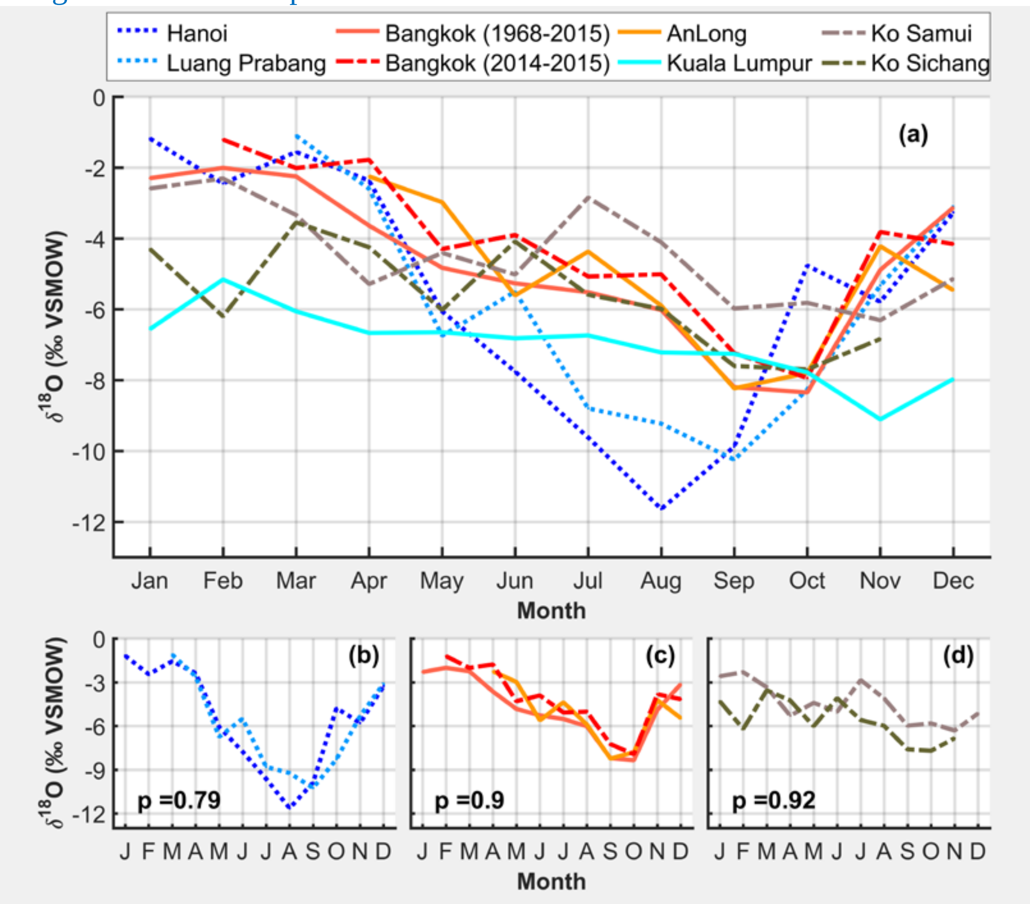
P10L4-L27: drop

As discussed in the 1st comment under 'major issues', we argue that this part is relevant for the manuscript.

P11L23-L24: I don't agree with this statement. More evidence is needed.

The Levene test (Levene, 1960) for equality of variances was used to compare the data of the different stations across the Indochinese Peninsula. We think that the similarity of the isotopic values and their seasonal variances between An Long and the long term time series of Bangkok (Fig. 8c) (of which the visible similarity is also confirmed with high significance by the statistical Levene test) provides sufficient evidence for our statement. In order to substantiate this finding we added the time series of Bangkok covering the same time span as our data collected in the Mekong Delta to the analysis (new figure 8c, shown below). This time series is even more similar to the one of An Long, resulting in a highly significant Levene test statistic of 0.98. This means that the isotopic variation of the An Long time series is almost identical to the one from Bangkok, and that the variation of the short term time series of Bangkok and An Long is also very similar to the long term time series. In turn, one can infer from this that the data collected in An Long are likely to be representative for the area (i.e. the southern part of SE-Asia).
However, we will modify the statement acknowledging the remaining uncertainty to: "In summary, the analyzed GNIP data suggests that the data and results from this study are likely to be representative for the Southern continental part of the Indochinese Peninsula."

The figure 8 in the manuscript will be replaced by the following figure, where the time series of Bangkok for the same period as our observation is added:



P14L9: Why was 124th model chosen as best?

Because the PRESS value of 124th model is smallest. The sentence provides this information. We also state this information in the methodology (P8L13).

P15L2: It is good idea. Why don't you do this trial?

We actually did this. The result are shown in Figure 12 and discussed in section 4.4 (from P15L6 to P16L8).

References

Cabello, M., Orza, J., Galiano, V., and Ruiz, G.: Influence of meteorological input data on backtrajectory cluster analysis? a seven-year study for southeastern Spain, Advances in Science and Research, 2, 65-70, 2008.

Conroy, J. L., Cobb, K. M., and Noone, D.: Comparison of precipitation isotope variability across the tropical Pacific in observations and SWING2 model simulations, Journal of Geophysical Research: Atmospheres, 118, 5867-5892, 2013.

Curio, J., and Scherer, D.: Seasonality and spatial variability of dynamic precipitation controls on the Tibetan Plateau, Earth System Dynamics, 7, 767, 2016.

Fleming, Z. L., Monks, P. S., and Manning, A. J.: Untangling the influence of air-mass history in interpreting observed atmospheric composition, Atmospheric Research, 104, 1-39, 2012.

Gebhart, K. A., Schichtel, B. A., and Barna, M. G.: Directional biases in back trajectories caused by model and input data, Journal of the Air & Waste Management Association, 55, 1649-1662, 2005.

Harris, J. M., Draxler, R. R., and Oltmans, S. J.: Trajectory model sensitivity to differences in input data and vertical transport method, Journal of Geophysical Research: Atmospheres, 110, 2005.

Ishizaki, Y., Yoshimura, K., Kanae, S., Kimoto, M., Kurita, N., and Oki, T.: Interannual variability of H218O in precipitation over the Asian monsoon region, Journal of Geophysical Research: Atmospheres, 117, 2012.

LeGrande, A. N., and Schmidt, G. A.: Sources of Holocene variability of oxygen isotopes in paleoclimate archives, Clim. Past, 5, 441-455, 10.5194/cp-5-441-2009, 2009.

Levene, H.: Robust tests for equality of variances, Contributions to probability and statistics, 1, 278-292, 1960.

Sánchez-Murillo, R., Birkel, C., Welsh, K., Esquivel-Hernández, G., Corrales-Salazar, J., Boll, J., Brooks, E., Roupsard, O., Sáenz-Rosales, O., and Katchan, I.: Key drivers controlling stable isotope variations in daily precipitation of Costa Rica: Caribbean Sea versus Eastern Pacific Ocean moisture sources, Quaternary Science Reviews, 131, 250-261, 2016.

Stohl, A.: Computation, accuracy and applications of trajectories—a review and bibliography, Atmospheric Environment, 32, 947-966, 1998.

Tan, M.: Circulation effect: climatic significance of the short term variability of the oxygen isotopes in stalagmites from monsoonal China—dialogue between paleoclimate records and modern climate research, Quaternary Sciences, 29, 851-862, 2009.

Tan, M.: Circulation effect: response of precipitation δ18O to the ENSO cycle in monsoon regions of China, Climate Dynamics, 42, 1067-1077, 2014.

Tindall, J., Valdes, P., and Sime, L. C.: Stable water isotopes in HadCM3: Isotopic signature of El Niño–Southern Oscillation and the tropical amount effect, Journal of Geophysical Research: Atmospheres, 114, 2009.

Vuille, M., Bradley, R., Werner, M., Healy, R., and Keimig, F.: Modeling δ18O in precipitation over the tropical Americas: 1. Interannual variability and climatic controls, Journal of Geophysical Research: Atmospheres, 108, 2003.

Vuille, M., Werner, M., Bradley, R., and Keimig, F.: Stable isotopes in precipitation in the Asian monsoon region, Journal of Geophysical Research: Atmospheres, 110, 2005.

Wang, B., Wu, R., and Lau, K.: Interannual variability of the Asian summer monsoon: contrasts between the Indian and the Western North Pacific-East Asian Monsoons*, Journal of climate, 14, 4073-4090, 2001.