University of Potsdam
Faculty of Science
Institute of Earth and Environmental Sciences

Stable Isotopes in Precipitation: Modelling Intra-Event Variations using Meteorological Parameters

Bachelor Thesis
for the B.Sc. Geoecology

Elena Macdonald

Supervisors:
Dr. Frauke Barthold
Dr. Till Francke

Potsdam, 14th September 2017

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-50661
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-506612

Zusammenfassung

Die kurzfristige Variabilität der Isotopenzusammensetzung von Niederschlägen in Golm, Deutschland wurde untersucht und modelliert. Dafür wurden Isotopendaten (D/H und $^{18}O/^{16}O$) mit einer hohen zeitlichen Auflösung sowie meteorologische Daten von einer Wetterstation und einem Mikroregenradar genutzt. Nach der Datenaufbereitung und dem Zusammenführen aller drei Datensätze wurde eine multivariate lineare Regressionsanalyse durchgeführt. Dies geschah für vier verschiedene, auf den Isotopendaten beruhende Response-Variablen und für den gesamten Datensatz sowie für die zwei Teildatensätze *Sommer* und *Winter*. Die verwendeten Response-Variablen sind die Differenzen der $\delta^{18}O$-Werte zu den ereignisbasierten Mittel- und Medianwerten und die Differenzen der Deuterium-Exzess-Werte zu den ereignisbasierten Mittel- und Medianwerten. Für die erhaltenen Modelle wurden die modellierten Werte mit den gemessenen Werten verglichen, wobei sich herausstellte, dass die Messwerte nicht zufriedenstellend wiedergegeben werden konnten. Daher werden am Ende mehrere Vorschläge gemacht, wie das Vorgehen und damit auch das Ergebnis der Modellierung möglicherweise verbessert werden kann.

Abstract

The short-term variability of the isotopic composition of precipitation in Golm, Germany was assessed and modelled. Isotopic data (D/H and $^{18}O/^{16}O$) on intra-event timescales as well as meteorological data from a weather station and a micro rain radar was used. After data preparation and the combination of all three data sets, a multivariate linear regression analysis was conducted. This was done for four different isotopic response variables and for the entire data set as well as for the two subsets *Summer* and *Winter*. The used response variables are the $\delta^{18}O$ values as the difference to the corresponding event-based mean and as the difference to the median, and the deuterium excess values as the difference to both the mean and the median. The models were evaluated by comparing the modelled values with the observed ones. This showed that the observations could not be reproduced in a satisfactory way. Therefore, several suggestions on how to possibly improve the methods and thus the modelling results are given in the end.

Contents

Zusammenfassung

Abstract

1. Introduction **1**

2. Methods **3**

2.1. Data Collection 3

2.1.1. Isotopic Data from Rainwater Samples 3

2.1.2. Meteorological Data from a Local Weather Station 5

2.1.3. Meteorological Data from a Micro Rain Radar 5

2.2. Data Analysis 6

2.2.1. Data Preparation 6

2.2.2. Regression Analysis and Model Application 8

3. Results **10**

3.1. The Isotopic Data 10

3.2. The Meteorological Data 14

3.3. Regression Analysis for the Entire Data Set 16

3.4. Regression Analysis for Data Subsets 18

3.4.1. Modelling δ^{18}O as the Difference to the Median 19

4. Discussion **22**

5. Conclusion **26**

References **27**

Acknowledgement **30**

Selbstständigkeitserklärung **30**

A. Appendix **31**

1. Introduction

The ratios of stable isotopes in precipitation are affected by many different processes and parameters. They depend on the source region of the water vapour and the evaporation processes taking place there, the trajectory and rainout history of the air mass as well as the isotopic exchange of it with surrounding vapour, and finally the conditions that are present when the rain is forming and the processes taking place during the fall of the raindrops (Miyake et al., 1968; Celle-Jeanton et al., 2004; Munksgaard et al., 2012). Four main factors that have an influence on the mean isotopic composition at a given location that were first mentioned by Dansgaard (1964) and have since been affirmed by various authors (e.g. Gat, 1996; Rozanski et al., 1993) are: an amount effect, an altitude effect, the distance from a coast and a latitude effect. The latter three are related to the moist adiabatic ascent and the cooling of the air mass. As the heavy isotopes condense more readily than the lighter, or "normal" ones, the air mass progressively depletes in heavy isotopes when travelling further inland or upwards (Gedzelman and Lawrence, 1990; Gat, 1996). These differences with regards to evaporation and condensation that are attributed to the differences in weight of the isotopes are the main prerequisites for all fractionation processes. Here, it is differentiated between equilibrium processes and non-equilibrium processes, also known as kinetic fractionation (Dansgaard, 1964; Hoefs, 2009). With respect to precipitation, especially the latter one is of interest as well as the so-called Rayleigh processes. These are processes where the formed condensate is removed immediately from the vapour, e.g. through rainout (Dansgaard, 1964).

Many of the factors mentioned above influence the isotopic composition at the moisture source or along the trajectory, but there are also many factors that have an impact on a more local scale at the precipitation site itself. The one that has probably been studied the longest and that Celle-Jeanton et al. (2004) call "the most important parameter" is temperature. It has a direct impact on atmospheric vapour condensation (Celle-Jeanton et al., 2004) resulting in a positive correlation with the enrichment of heavy isotopes in precipitation (Rozanski et al., 1992; Field, 2010). A second important factor is relative humidity which influences the isotopic composition via reevaporation (Berkelhammer et al., 2012). Here, a negative correlation was found, meaning that a higher relative humidity results in a lighter isotope signature (Muller et al., 2015). Two other factors that are probably linked to each other are the above-mentioned amount effect and the drop size. The amount effect, i.e. a higher degree of depletion of heavy isotopes with higher rain rates, has been found by various authors on different time- and spatial scales (e.g. Lee and Fung, 2008; Barras and Simmonds, 2009; Muller et al., 2015). The drop size seems to affect the rate of isotopic equilibration with the surrounding vapour and thus influences the ratio of light and heavy isotopes in the rain as well. It is probably linked to the amount effect as lighter rains are often associated with smaller rain drops (Lee and Fung, 2008).

A lot of work has already been done on analysing the inter-event variability of the isotopic composition of rain events and on understanding the underlying mechanisms and processes on monthly or annual timescales (e.g. Dansgaard, 1964; Gat, 1996). More recently though, several studies were conducted based on data with higher temporal resolutions (e.g. Celle-Jeanton et al., 2004; Coplen et al., 2008; Barras and Simmonds, 2009; Muller et al., 2015). These studies showed that variations of the isotopic composition can also occur within a single rain event as the physical conditions influencing the ratio of the isotopes also vary on smaller timescales (Munksgaard et al., 2012). So in addition to the large-scale trends that have thoroughly been studied before and that were mentioned above, there seems to be an intra-event variability that overlies these trends, suggesting that processes and effects on different timescales are of importance for the final isotopic composition (Barras and Simmonds, 2009).

The isotopic compositions reported in this study will be denoted as parts per thousand enrichments of the isotopic ratios relative to a standard:

$$\delta x = \left(\frac{R_{Sample}}{R_{Standard}} - 1 \right) \cdot 1000 \text{ ‰} , \tag{1}$$

with x being either one of the two heavy isotopes ^{18}O and D, and R being the corresponding isotopic ratio, that is ^{18}O/^{16}O or D/H respectively. As a standard the Vienna Standard Mean Ocean Water (VSMOW) is used (Craig, 1961; Coplen, 1994). In addition to the measured isotopic ratios, the secondary parameter deuterium excess (d) was calculated as follows:

$$d = \delta D \ - \ 8 \ \delta^{18} O \ . \tag{2}$$

This has first been defined by Dansgaard (1964) and has since been commonly used in other studies (e.g. Rindsberger et al., 1990; Gat, 1996; Barras and Simmonds, 2009).

Analysing the isotope composition in precipitation at higher temporal resolutions and assessing intra-event variability enables “new and powerful tracer applications in climatology, hydrology, ecophysiology and paleoclimatology” (Munksgaard et al., 2012). It has been noted that in measurements on sub-hourly timescales some of the typical relationships between factors and the isotopic composition that are commonly observed in monthly averaged data are absent (Munksgaard et al., 2012). That is why having a look at these shorter timescales can lead to new insights into mechanisms that have been concealed by the averaged data before (Muller et al., 2015). This can be of interest for various aspects such as cloud physics, atmospheric water vapour transport and the application of general circulation models (GCMs) for modelling microphysical and atmospheric processes (Muller et al., 2015). GCMs have increasingly been fitted with water isotope diagnostics and on annual timescales the simulations of isotope signals have proven to be fairly good already (Hoffmann et al., 2000; Muller et al., 2015). Nevertheless, the calibration of these models can be biased by local and short-term effects which could be improved when understanding the relevant processes better (Hoffmann et al., 2000).

The present study is based on a data set that is similar to the one Muller et al. (2015) used for a qualitative assessment of factors influencing the isotopic composition of precipitation on short timescales. In contrast to that, the goal here was to achieve a quantitative estimation of how different meteorological parameters influence the intra-event variability. From previous studies (e.g. Celle-Jeanton et al., 2004; Berkelhammer et al., 2012; Muller et al., 2015) it is known which parameters are likely to have an impact, but as the exact way and the interactive effects are not established yet the decision was made to use modelling as a method. More precisely, a multivariate regression analysis was conducted because of the multi-scale influences that are present at mid-latitude locations (Berkelhammer et al., 2012; Muller et al., 2015). A similar approach for the same location and partly the same data set as the study at hand was already taken by Breier (2015). In that study the focus of the quantitative analysis was less on the intra-event variability and more on inter-event variations, and δD instead of δ^{18}O and deuterium excess was used.

2. Methods

2.1. Data Collection

All data was collected in Potsdam on the campus Golm of the University of Potsdam (52°24’34”N, 12°58’36”E, 36 m a.s.l) between 2012 and 2016 in three different measurement phases. The first one was from August 2012 to April 2013, the second one took place in May 2014 and the last one was from June 2016 to October 2016.

2.1.1. Isotopic Data from Rainwater Samples

For the collection of the rainwater samples a different method was used in the last measurement phase than in the first two ones. The subsequent isotope analysis was the same for all samples though.

In the first two measurement phases a sequential precipitation collector was used, which can be seen in figure 1. It collects rainwater in a catchment tray with a receiving surface of 0.4 m^2. Via a funnel the water is led into a tipping counter and from there it is filled into collecting flasks. The flasks have a capacity of 200 ml each and are filled sequentially. A data logger, that is attached to the tipping counter, registers the time of each tipping so that for every flask the period of time in which the rainwater was collected is known. The flasks were then sampled with a filter syringe to avoid suspended particles in the water. Afterwards the samples were filled into bottles of 1.5 ml, sealed airtight and stored in a cool place until further analysis (Breier, 2015).

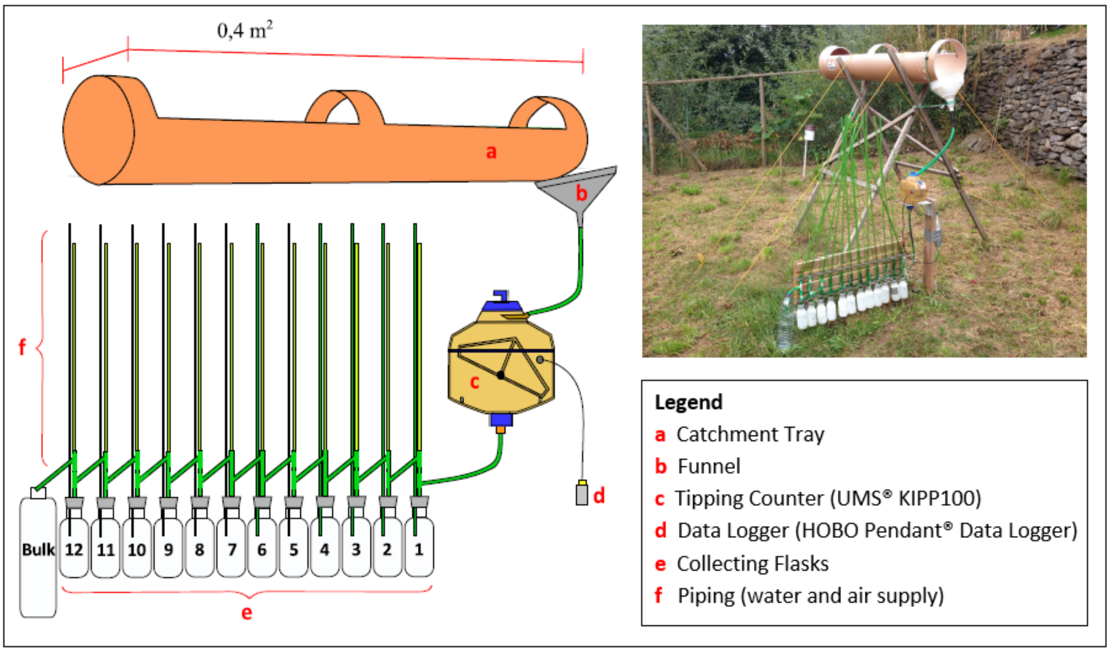


Figure 1: Schematic representation and photo of the sequential precipitation collector. Adapted from Breier (2015).

For the last measurement phase a different way of sampling had to be used as the sequential precipitation collector was not available. Instead, the rainwater was collected using two buckets alternately. When there was enough water in one bucket, the buckets got changed and the water in the first one got sampled. Again, a filter syringe was used to fill the samples into small bottles which got sealed airtight and kept cool. In this procedure, the time period in which each sample was collected was registered by hand.

Following the collection of the samples, the rainwater was analysed for hydrogen and oxygen isotopes. The analysis was carried out with the laser spectrometer Liquid Water Isotope Analyzer from Los Gatos Research according to the IAEA standards. Each sample was measured six times, but only the last three measurements were taken into account when calculating the mean to avoid the influence of a possible contamination from the previous sample. The values are expressed in the δ notation, relative to the standard VSMOW.

2.1.2. Meteorological Data from a Local Weather Station

Part of the meteorological parameters that were used for modelling were obtained from the local weather station of the University of Potsdam (see figure 2). It consists of an ombrometer, two hygro-thermo transmitters, two soil temperature transmitters, a wind transmitter, two pyranometers and a data logger that records the data every minute. All equipment is from ThiesClima (Adolf Thies GmbH & Co. KG). The parameters that were used in this study are air temperature [°C] and relative humidity [%], both measured 2 m above the ground, wind speed [m/s] and precipitation height [mm].

Figure 2: Weather station of the University of Potsdam in Golm. From Breier (2015).

Figure 3: Micro Rain Radar of the University of Potsdam in Golm. From Breier (2015).

2.1.3. Meteorological Data from a Micro Rain Radar

The rest of the considered meteorological parameters were acquired using a vertically pointing micro rain radar (MRR) from METEK (Meteorologische Messtechnik GmbH), which can be seen in figure 3. The MRR is a Frequency Modulated Continuous Wave (FM-

CW) radar which operates at 24 GHz. Using the Doppler frequency, the measurement of height profiles of the drop size distribution and, derived from that, the rain rate, the liquid water content and the falling velocity can be achieved (METEK GmbH, 2009). The profiles are resolved in 31 range gates, with each range comprising of 35 m (until 2013) or 100 m (since 2014) and therefore allowing measurements up to a height of 1085 m or 3100 m, respectively. Again, data was recorded every minute.

2.2. Data Analysis

The aim was to see if the intra-event variability of the isotopic data could be modelled using the parameters from the weather station and the MRR. In order to do so, the data from the three different sources had to be prepared to allow further analysis and comparison. Once all needed data was merged, a regression analysis could be performed and the resulting models could be assessed. All steps of the data analysis were carried out using the programme *RStudio* which is based on the programming language *R* (R Core Team, 2016).

2.2.1. Data Preparation

To be able to perform a regression analysis, all data had to be compiled for the same timestamps and periods. Furthermore, postprocessing was needed for some parameters and secondary parameters were calculated.

As mentioned before, the data from the weather station as well as from the MRR were available continuously for every minute whereas the periods of the isotopic data lay between one minute and several hours, depending on the rain intensity. Thus, the periods of the meteorological data had to be adapted to the periods of the isotopic data. As the sampling of the rainwater over these longer time spans results in mean values of the isotopic composition, the same was done for the other two data sets: The arithmetic mean of the minute-by-minute values was calculated over the same periods from which the rainwater samples derived. Another aspect that had to be taken into account to be able to merge the associated data was the time zone. All three data sets were recorded using different time zones which made conversions necessary.

Using the isotopic data, a secondary parameter - the deuterium excess - was calculated. As mentioned in the Introduction, deuterium excess is defined as $\delta D - 8\ \delta^{18}O$. Furthermore, the differences of each single value to both the mean and median of all values belonging to the same rain event were calculated as measures of the intra-event variability.

The data from the weather station was also used to calculate secondary parameters, namely gradients of the measured variables. Apart from very few exceptions, several minute-by-minute values of the meteorological parameters were available for each time span derived from the isotopic data. These values were used to fit a linear model and the regression coefficient was then taken as the gradient of the respective parameter for that time span.

The preparation of the MRR data was a bit more extensive as there were not only several values for each reference time span but also for several heights. To import the relevant data for each precipitation event the functions *parse_mrr* and *read_mrrdata* (see appendix; Francke and Brenner, 2011) were used. For the calculation of the arithmetic mean for each period as described above one height had to be chosen. Preferably, the values of a range gate close to the ground should be used to assure that the relevant fractionation processes, which were mentioned in the Introduction, acted already and that the precipitation measured by the MRR is as similar to the collected rain as possible. Because of preliminary assessments it was known that the range gates below 300 m tend to be defective which led to the decision to use the values of a height of 315 m (until 2013) or 300 m (since 2014), respectively. In order to take account of the fact that the precipitation collected at the ground was at a height of roughly 300 m some seconds before it was collected the time spans got shifted a little bit. To do so, the mean falling velocity of all events was calculated and using that the time the rain needs to fall to the ground from the chosen height could be computed.

For the drop sizes the MRR measures drop size distributions, providing a range of different size classes per time and height as well as the corresponding spectral drop densities (SDD). The weighted mean of the sizes with the SDD as a weighting was calculated to attain a single value – the average drop size - for each timestamp and height. Furthermore, the gradients of the MRR parameters were computed but in a different way than it was done for the parameters from the weather station. Here, the advantage of the MRR measuring in different heights was used. For the data that was collected until 2013 the highest range gate is at 1085 m and thus this height was used as a maximum for all data. For each range gate between 300 m and 1085 m the mean was calculated for each time span, and in doing so the time span was shifted backwards from range gate to range gate according to the mean falling velocity. Afterwards a linear model was fitted to the calculated values and the regression coefficient was taken as the gradient. Therefore, the gradients reflect how the precipitation changed on its way to the ground.

In the end, all parameters were merged into one data frame for every event. All procedures as described above were executed with *R* using the functions *AssembleData_35* or *AssembleData_100* (see appendix). Two different functions were necessary because of the different range gates of the MRR and the different methods of collecting the isotopic data. For the rain samples that were collected using the sequential precipitation collector only the end time of each period was known, whereas for the samples taken manually both, the starting and the end time got registered, allowing a slightly different way of data preparation.

2.2.2. Regression Analysis and Model Application

For the analysis of the data the 21 event-based data sets were combined into one table. Using the isotope data of all events the relation between the δD and the δ^{18}O values was analysed. As expected the correlation was very high, suggesting the modelling of just one of the two variables and a measure of deviation from this linear relationship. In line with some other studies (e.g. Rindsberger et al., 1990; Barras and Simmonds, 2009; Berkelhammer et al., 2012) the δ^{18}O values and deuterium excess were chosen.

In a next step, the predictors were made fit for the regression analysis. The distributions of all possible predictors were analysed using histograms. Where the distribution was skewed a transformation of the data was considered to achieve a more symmetrical distribution. A logit-transformation was used for the data of the relative humidity and a log-transformation for the rain rate, the liquid water content and the drop size. The latter transformations and the histograms of these variables revealed that the MRR data set contained some values which seemed to be unrealistic. First, there were a few timestamps where the rain rate and the liquid water content were recorded as being zero, even though the data was from a time when it was raining according to the meteorological data. As a consequence, the decision was made to exclude all data from these timestamps from the regression analysis. Secondly, some of the drop size values were below zero, which is not reasonable. This is probably a result of the postprocessing of the MRR data (METEK GmbH, 2017). As it could not be detected whether the rest of the MRR data for the respective timestamps was defective as well, all data from these few timestamps was again excluded from the regression analysis.

To achieve a stable model and to avoid the negative effects that multicollinearity can have on a model, the collinearity of the predictors was tested (Cambridge dictionary: Multicollinearity, 2010). To do so, Spearman's rank correlation coefficient was used (Spearman, 1904). All predictor pairs with a coefficient greater than 0.5 were classified as *highly correlated* and the ratios of these pairs were calculated and added as new predictors. After doing so, one of the predictors of each highly correlated pair was deleted from the list. This reduces the problems associated with multicollinearity while keeping the information of all initial predictors.

The remaining 14 predictors were joined with each of the four response variables: the δ^{18}O values as the difference to the corresponding event-based mean and as the difference to the median as well as the deuterium excess values as the difference to both the mean and the median. The resulting table contained some rows where either data from the weather station or from the MRR was missing which is not suitable for the regression analysis. Thus, all incomplete rows were removed from the table, leaving 114 data lines for the analysis.

The regression analysis itself was carried out using the function *colin_fun* (see appendix; Priest, 2015). Here, the best combination of predictors for modelling the response variable is determined based on the Akaike Information Criterion (AIC; Akaike, 1981). The AIC is defined as

$$AIC = -2L_m + 2m \ , \tag{3}$$

with L_m being the maximized log-likelihood and m being the number of parameters in the model (Cambridge dictionary: AIC, 2010). Hence, the AIC takes both the quality of the models and their complexity into account which makes it useful for selecting the best set of variables. In addition, a maximum number of predictors was defined, namely five, to avoid overfitting. A multivariate linear model was then fitted to each of the four sets of selected predictors and the corresponding response variable. This was done using the least squares method.

To assess how good each of the fitted models is the coefficient of determination (R^2) was calculated as follows:

$$R^2 = 1 - \frac{\sum (y_o - y_m)^2}{\sum (y_o - \bar{y_o})^2} \ , \tag{4}$$

with y_o being the observed values, y_m being the modelled values and $\bar{y_o}$ being the mean of the observed values. This coefficient indicates the proportion of variance of the response variable that can be explained by the predictor variables that were used (Cambridge dictionary: Coefficient of determination, 2010). It is therefore a measure for the goodness of fit of a model. The closer the R^2-value is to 1 the better the model, while an R^2-value close to 0 means that the chosen parameters are not suitable for modelling the response variable. Furthermore, diagnostic plots based on the residuals of the models were used as another way of evaluating the results of the regression analysis.

The same steps, from testing collinearity over predictor selection and fitting a model to assessing the quality of the models, were also carried out for two subsets of the data. The first subset comprised of the data from all events recorded in the summer half-year, that is between April and September, and the second subset contained all data from winter, that is from October to March.

3. Results

Data from a total of 21 rain events was included in the analysis. Table 1 gives an overview over all of them by summarising the basic data as well as isotopic and meteorological parameters for each individual event. It can be seen that there is a lot of variability between the events, regarding the basic data such as the number of samples or the duration of the event as well as the measured parameters. The δ^{18}O values for example range from -18.37 ‰ to -2.30 ‰ and the deuterium excess values from -5.76 ‰ to 16.00 ‰. Similar ranges for the inter-event variability of δ^{18}O values were also encountered by Celle-Jeanton et al. (2004) and Berkelhammer et al. (2012). The focus of this study though is on the intra-event variability. Here, the event with the least variations was MA07 with a change of as little as 0.48 ‰ δ^{18}O over the course of almost two hours. The greatest variability was measured for the event RE07, where the δ^{18}O values decreased by 9.61 ‰. However, this was also by far the event with longest duration, lasting more than 44 hours. Another event with a high variability but a much shorter duration was BO04.2, where δ^{18}O values changed by as much as 5.14 ‰ in less than two hours. Figures of the temporal courses of the isotopic composition of all events can be found in the appendix.

3.1. The Isotopic Data

As mentioned above, one of the first steps after the data preparation was analysing the relation of δD and δ^{18}O for all events. In figure 4 the two parameters are plotted against each other, showing the distinct linear relationship that is typical for these variables. The blue line in the figure represents the global meteoric water line (GMWL), which is defined as

$$\delta D = 8\ \delta^{18}O + 10\ ‰\ . \tag{5}$$

It was first established by Craig (1961) who analysed this relationship based on hundreds of precipitation samples from all over the world. The black line is the linear model fitted to the data shown in the graph. It is a local meteoric water line (LMWL) as it represents the relationship of the isotopic parameters for one location. The formula for the LMWL is:

$$\delta D = 7.54\ \delta^{18}O + 3.93\ ‰\ . \tag{6}$$

Deviations of LMWLs from the GMWL as were observed here can for example occur when there is a seasonal variation in the moisture source of the precipitation or when secondary evaporation processes take place during the rainfall (Araguás-Araguás et al., 2000). Because of the high correlation of δD and δ^{18}O and in concordance with other studies (e.g. Rindsberger et al., 1990; Barras and Simmonds, 2009; Berkelhammer et al., 2012) the decision was made to focus the regression analysis only on δ^{18}O and the secondary parameter deuterium excess which is a measure of the deviation of a data point from the GMWL (Araguás-Araguás et al., 2000). So by choosing these two variables for the analysis the expectation was to be able to model the general trend of the isotopic data in dependence on the meteorological data as well as deviations from this trend.

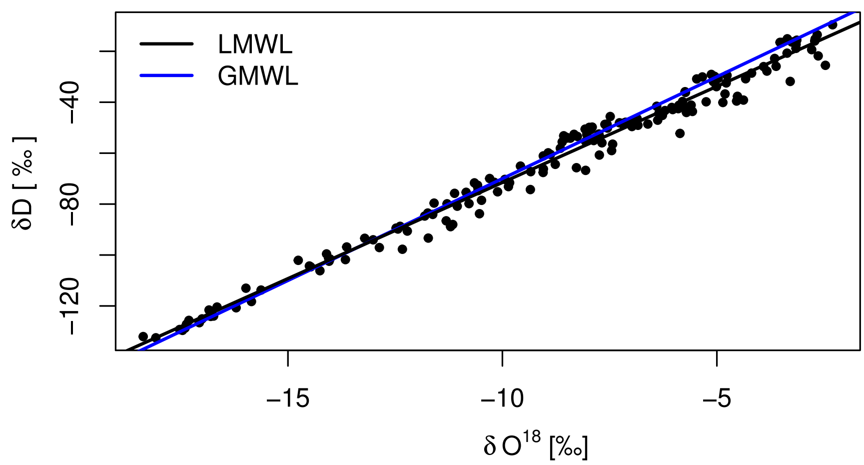


Figure 4: δD versus δ^{18}O for rain water samples collected in Golm, Germany. The blue line represents the global meteoric water line (GMWL, $\delta D = 8\ \delta^{18}O + 10$ ‰); the black line is the least mean square average of the data points (LMWL, $\delta D = 7.54\ \delta^{18}O + 3.93$ ‰).

Both parameters - δ^{18}O and deuterium excess - were not used in absolute numbers for the regression as the goal was to model intra-event variability. Instead, differences to the event-based mean and median were calculated. This is visualised exemplarily for the median and event MA06 in figure 5: the black dots are the measured values and the blue dashed line is the median for this event. The coloured areas symbolise the fluctuation of the values around this median. At first, mean and median were calculated by simply using the given data points for each event, making them volume-based measures. The regression analysis was then carried out for the resulting response variables, but the outcome was rather poor. Hence, a new approach was taken: by interpolating between the data points, time-based mean and median values were calculated. For some events the volume- and time-based measures were almost identical, for others they differed drastically. Again, MA06 was chosen as an example for the visualisation. Figure 6 displays both the volume- and time-based median for this event and shows the moderate increase of the measure. Using the time-based measures as a reference level for the response variables improved the outcome of the regression analysis which is why in the following only the results of the second approach will be presented and discussed.

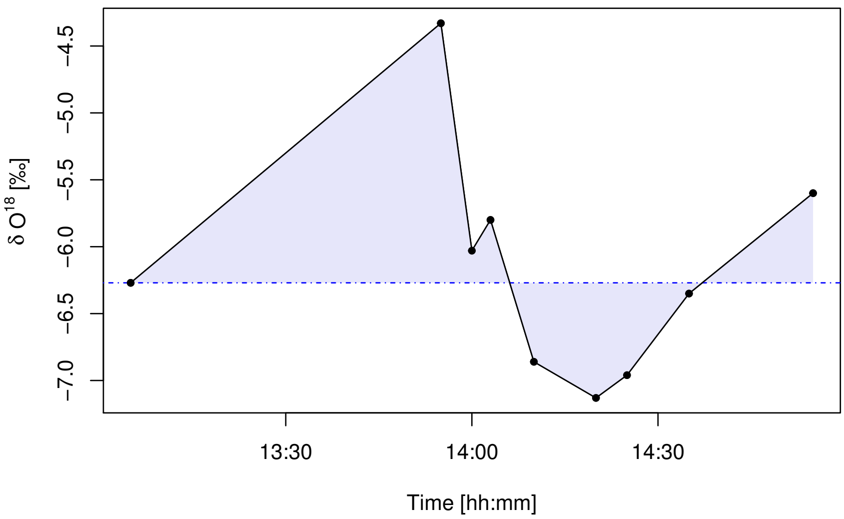


Figure 5: The intra-event variability of δ^{18}O for the event MA06. The black dots represent the measured values, the blue dashed line is the median for this event.

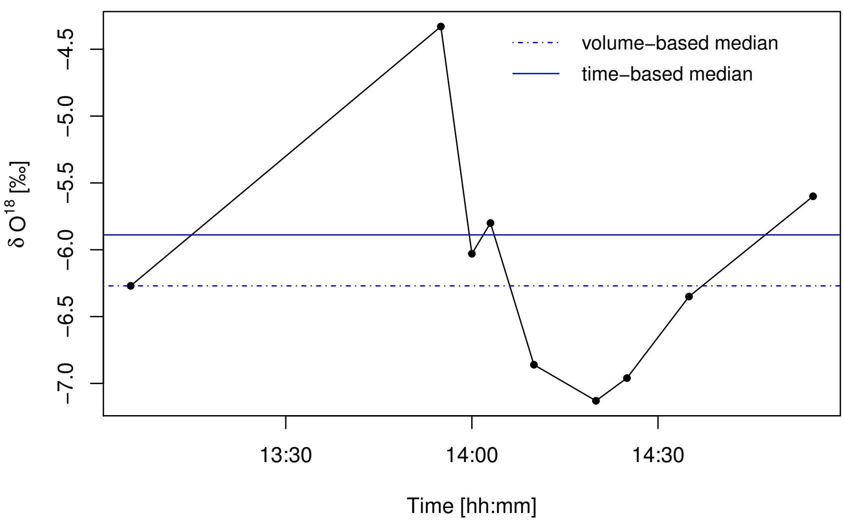


Figure 6: The difference between the volume- and the time-based median of the δ^{18}O-values of event MA06.

Table 1: Summary of the isotopic and meteorological parameters of all events. Abbreviations: Deut. - deuterium excess, WV - wind velocity, RH - relative humidity, T - temperature, RR - rain rate, LWC - liquid water content, FV - falling velocity, DS - drop size.

					Isotopic Parameters [‰]						Meteorological Parameters (Mean values)						
Event	Date	No. of samples	Start time	Duration [h:m]	Mean δO^{18}	Min δO^{18}	Max δO^{18}	Mean Deut.	Min Deut.	Max Deut.	WV [m/s]	RH [%]	T [°C]	RR [mm/h]	LWC [g/m³]	FV [m/s]	DS [mm]
RE04	31.08.2012	5	01:46	3:10	-4,90	-5,47	-3,83	8,97	2,88	12,90	0,75	96,92	16,85	0,41	0,068	2,54	0,28
RE05	18.09.2012	6	23:37	2:35	-3,05	-3,53	-2,30	9,74	7,77	11,72	0,79	93,94	13,73	11,44	2,666	1,76	0,30
RE06	04.10.2012	4	14:54	0:51	-4,12	-6,05	-2,64	3,41	-0,72	6,35	2,06	92,61	15,01	0,42	0,127	7,07	0,34
RE07	28.11.2012	30	04:19	44:44	-14,45	-18,37	-8,77	10,42	5,72	14,97	1,86	99,20	5,21	0,65	0,047	4,45	0,76
BO01	18.12.2012	7	12:34	6:26	-11,85	-13,01	-10,19	10,13	8,94	11,40	0,66	99,66	4,20	0,29	0,024	3,64	0,45
BO02	03.01.2013	3	04:30	3:49	-5,53	-5,88	-4,97	8,23	5,87	9,72	2,68	89,55	5,73	0,28	0,025	3,69	0,38
BO03	04.01.2013	5	13:16	1:14	-4,82	-5,74	-3,36	10,24	8,66	11,94	1,55	95,34	9,40	0,82	0,056	5,08	0,50
BO04.1	06.01.2013	8	02:29	24:30	-3,24	-3,91	-2,71	6,76	5,26	9,67	1,15	99,47	7,27	0,08	0,013	1,75	0,33
BO04.2	08.01.2013	7	04:18	1:25	-10,60	-12,87	-7,73	6,52	4,82	8,41	1,03	100,00	5,90	0,87	0,052	5,45	1,06
BO06	11.04.2013	3	09:59	2:20	-7,73	-8,06	-7,45	1,28	-2,28	5,50	1,91	90,40	9,52	0,13	0,009	3,70	0,35
BO07	12.04.2013	8	03:39	4:13	-10,74	-12,33	-8,27	1,10	0,47	3,91	0,45	97,22	9,49	0,30	0,018	4,95	0,55
BR01.1	27.05.2014	12	21:09	1:16	-7,68	-8,42	-6,40	11,12	7,93	14,28	1,49	95,43	12,85	1,07	0,059	6,04	0,21
BR01.2	28.05.2014	19	15:42	7:33	-8,69	-10,65	-7,27	12,25	9,23	15,38	1,35	99,27	10,46	0,32	0,029	4,14	0,39
MA02	14.06.2016	3	12:00	0:35	-7,65	-9,34	-5,86	1,08	-5,38	7,45	1,32	87,45	17,40	0,35	0,016	5,86	0,34
MA03	05.07.2016	3	14:45	0:10	-3,42	-5,01	-2,47	1,13	-5,76	6,19	1,02	68,74	23,09	0,43	0,018	6,42	0,58
MA04	13.07.2016	3	10:05	0:45	-4,07	-4,54	-3,29	-4,29	-5,52	-3,23	0,51	83,19	17,76	0,22	0,012	5,12	0,46
MA05.2	27.07.2016	14	15:05	1:50	-5,91	-7,43	-4,52	2,96	-1,53	6,40	0,91	94,18	19,53	1,79	0,070	8,07	0,29
MA06	02.08.2016	9	13:05	1:50	-6,15	-7,13	-4,33	6,33	3,66	9,17	0,71	86,76	17,90	0,70	0,041	5,19	0,05
MA07	11.10.2016	4	14:05	1:50	-14,26	-14,50	-14,02	11,13	10,62	11,69	0,64	94,61	10,35	0,25	0,025	3,08	0,63
MA08	12.10.2016	3	08:45	3:35	-13,01	-14,10	-11,29	11,96	10,43	13,24	0,50	98,75	7,94	0,04	0,003	3,62	0,29
MA09	19.10.2016	5	07:50	2:50	-13,48	-14,76	-11,74	11,20	7,48	16,00	0,59	96,28	9,64	0,41	0,019	5,89	0,24

3.2. The Meteorological Data

For all meteorological parameters that were considered to be potential predictors the distribution of the data was analysed. The respective histograms can be seen in figure 7. The distribution of the relative humidity is skewed left and as the values of it can only range from 0 to 1 (i.e. 0 % – 100 %) a logit-transformation was used to achieve a more symmetrical distribution. The formula is as follows:

$$RH_{logit} = \log\left(\frac{RH/100}{1 - RH/100}\right) , \tag{7}$$

where RH is the relative humidity expressed as a percentage. Predictors that were distinctively right skewed are rain rate (RR), liquid water content (LWC) and drop size (DS). Here, log-transformations were performed to shift the distributions closer to being normally distributed. Histograms of the transformed variables are also displayed in figure 7. The distributions of some of the gradients were not very symmetrically either, nevertheless the decision was made not to transform these as it would have been only a minor improvement.

Table 2 gives an overview of the results from testing the collinearity of the predictors. The pairs that were classified as being highly correlated for the entire data set as well as for the two subsets *Summer* and *Winter* are listed. When taking all data into account six parameter pairs were found to be highly correlated and for the two subsets even eight pairs are listed. For a total of 13 parameters these numbers are fairly high, making it clear that multicollinearity is definitely an issue that needs to be considered for this data set. When looking at table 2 it can also be seen that especially data from the same source tends to be highly correlated. For all but one pair both parameters are either from the weather station or the MRR but not mixed. The only exception is that for the entire data set the falling velocity (MRR data) is correlated to the temperature (weather station data). Another point that is striking is that for the winter data set none of the data of the weather station was classified as being highly correlated. All eight pairs only comprise of parameters measured by the MRR. The table also includes information on which parameters were kept as predictors and which ones were deleted. After calculating the ratios of the predictor pairs one of the two was excluded from the list as mentioned above. As some parameters were highly correlated to more than one other parameter there are some pairs where both original variables were excluded and only the ratio was used as a potential predictor.

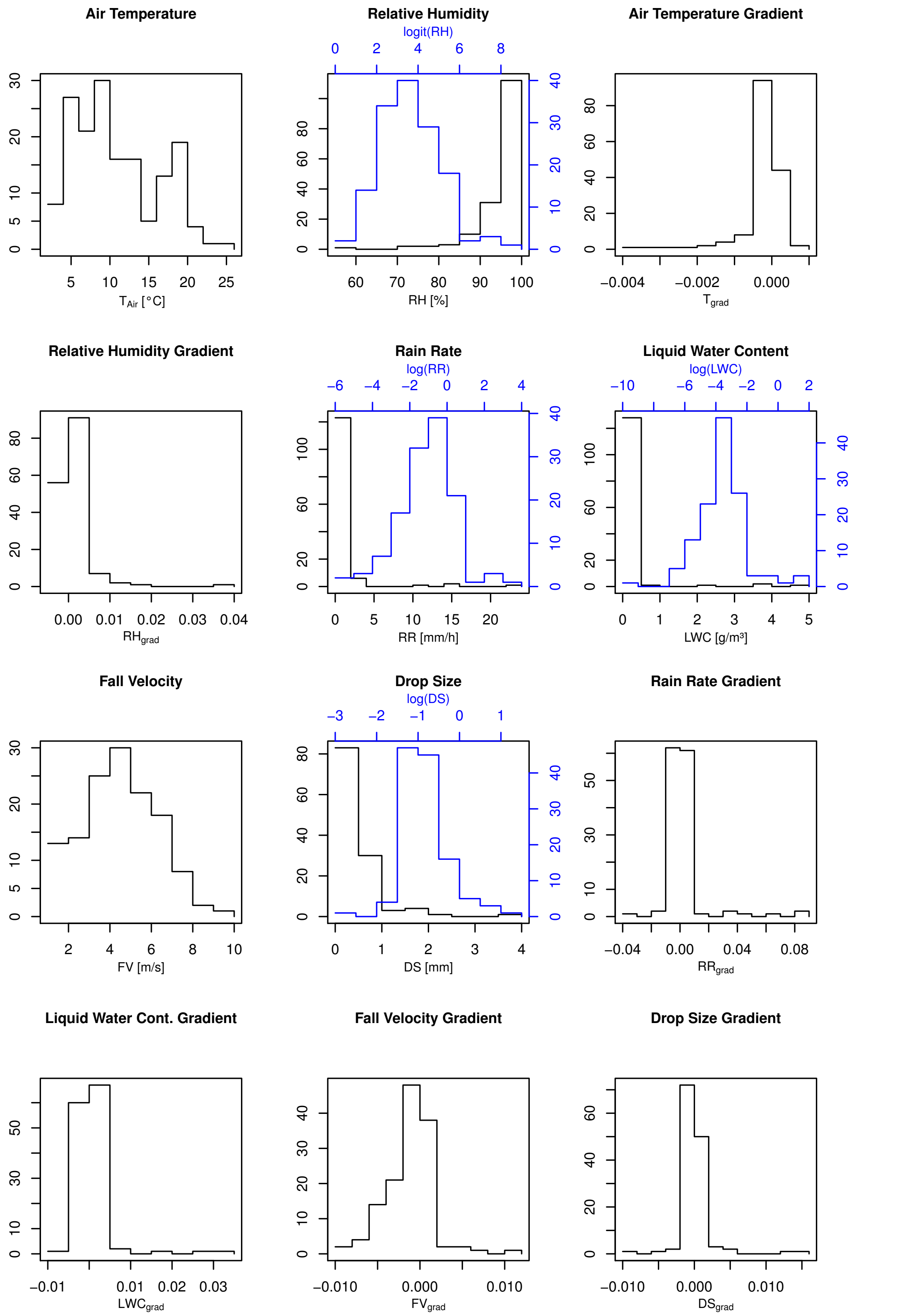


Figure 7: Histograms of the distributions of meteorological parameters and their gradients. The data was obtained from a weather station and a micro rain radar during the considered precipitation events. Black lines are for initial values, blue lines are for transformed values.

Table 2: Overview of highly correlated predictor pairs (Spearman's rank correlation coefficient > 0.5). The ones in bold were kept in addition to the calculated ratios, while the others were excluded from the list. Abbreviations: RH - relative humidity, T - temperature, RR - rain rate, LWC - liquid water content, FV - falling velocity, DS - drop size, log/logit - parameter is log-/logit-transformed, grad - gradient.

All Data		Summer		Winter	
T	$\boldsymbol{RH_{logit}}$	T	$\boldsymbol{RH_{logit}}$	LWC_{log}	$\boldsymbol{RR_{log}}$
RH_{grad}	$\boldsymbol{RH_{logit}}$	RH_{grad}	$\boldsymbol{RH_{logit}}$	FV	$\boldsymbol{RR_{log}}$
LWC_{log}	$\boldsymbol{RR_{log}}$	RH_{grad}	T	FV	LWC_{log}
LWC_{grad}	$\boldsymbol{RR_{grad}}$	T_{grad}	T	LWC_{grad}	$\boldsymbol{RR_{grad}}$
FV_{grad}	T	LWC_{log}	$\boldsymbol{RR_{log}}$	FV_{grad}	FV
FV_{grad}	LWC_{grad}	RR_{grad}	$\boldsymbol{RR_{log}}$	FV_{grad}	$\boldsymbol{DS_{log}}$
		RR_{grad}	LWC_{log}	FV_{grad}	LWC_{grad}
		LWC_{grad}	RR_{grad}	DS_{grad}	$\boldsymbol{DS_{log}}$

3.3. Regression Analysis for the Entire Data Set

As mentioned above, for each response variable the best set of predictors was selected based on the AIC. They are listed in table 3. For both response variables based on the δ^{18}O values three of the five predictors are the same, namely the logit-transformed relative humidity RH_{logit}, the log-transformed rain rate RR_{log}, and the ratio of the falling velocity and the temperature FV_{grad}/T. This means that these three parameters seem to be important for modelling the intra-event variability of δ^{18}O, independent of whether the mean or the median is chosen as a reference level. Remarkable is that for both response variables based on the deuterium excess only one predictor was selected, namely the ratio of the gradient and the logit-transformed relative humidity RH_{grad}/RH_{logit}. If this was to result in a good model it would mean that the intra-event variability of the deuterium excess is only dependent on one measure based on the relative humidity.

Table 3 also contains the coefficients of determination for the multivariate linear models that were fitted to the data based on the selected predictors. Here, it becomes clear that deuterium excess is in fact not only dependent on that one variable, as the coefficients of 0.086 and 0.085 respectively indicate that the goodness of fit of the models is very poor. That is to say that even with the "best" set of predictors the given data is not suitable for modelling the intra-event variability of deuterium excess. Almost the same goes for δ^{18}O. Even though the coefficients of determination are much higher here, they are still too low to be classed as good, suitable models. With the difference to the mean as the response variable the coefficient is 0.351 and for the median it is 0.415.

To further evaluate the models and to examine whether some – possibly defective – data lines may be the cause for the poor fit some diagnostic plots were used. First, the distribution of the residuals was analysed. One of the assumptions when fitting a model is

Table 3: The best set of predictors for each response variable as determined by using the AIC and the coefficients of determination (R^2) for the corresponding models. Abbreviations: WV - wind velocity, RH - relative humidity, T - temperature, RR - rain rate, LWC - liquid water content, FV - falling velocity, DS - drop size, log/logit - parameter is log-/logit-transformed, grad - gradient.

(a) based on the entire data set

Response Variable	δO^{18} diff. to mean	δO^{18} diff. to median	Deut. excess diff. to mean	Deut. excess diff. to median
Predictors	RH_{logit} RR_{log} T/RH_{logit} RH_{grad}/RH_{logit} FV_{grad}/T	RH_{logit} RR_{log} WV RR_{grad} FV_{grad}/T	RH_{grad}/RH_{logit}	RH_{grad}/RH_{logit}
R^2	0.351	0.415	0.086	0.085

(b) based on the data subset *Summer*

Response Variable	δO^{18} diff. to mean	δO^{18} diff. to median	Deuterium excess diff. to mean	Deuterium excess diff. to median
Predictors	RH_{logit} RR_{log} RH_{grad}/T FV WV	RH_{logit} RR_{log} RH_{grad}/T FV	RH_{logit} RH_{grad}/RH_{logit} T/RH_{logit} T_{grad}/T	RH_{logit} RH_{grad}/RH_{logit} T/RH_{logit} T_{grad}/T FV
R^2	0.371	0.406	0.262	0.303

(c) based on the data subset *Winter*

Response Variable	δO^{18} diff. to mean	δO^{18} diff. to median	Deut. excess diff. to mean	Deut. excess diff. to median
Predictors	RH_{logit} T RH_{grad} LWC_{grad}/RR_{grad} FV/LWC_{log}	WV RR_{grad} RH_{grad} LWC_{grad}/RR_{grad} FV_{grad}/FV	RH_{grad}	WV T RR_{grad} FV_{grad}/FV
R^2	0.491	0.543	0.064	0.238

that the residuals should be normally distributed (Cambridge dictionary: Residual plots, 2010). This was indeed more or less the case for all four models. Secondly, the impact of single observations on the estimation of all regression coefficients was assessed using the Cook's distance. A value greater than one indicates that the corresponding observation has a disproportionate influence on the regression coefficients (Cambridge dictionary: Cook's distance, 2010). That means that the accuracy of that observation needs to be checked and if there is any evidence of it being defective it should be excluded from the regression. For both regression analyses based on deuterium excess there were two observations for which the values of the Cook's distance were close to one. Nevertheless, excluding them from the analysis did not improve the results and there was also no meaningful reason why they should be excluded.

As the calculated models were not capable of reproducing the data in a satisfactory way the idea was to split the data set in order to achieve better results for the subsets. Several authors, including Friedman et al. (1992), Feng et al. (2009) and Berkelhammer et al. (2012), have mentioned before that there is a seasonality to δ^{18}O values. During warmer months the values tend to be enriched, while in cooler months more depleted values are recorded. Thus, the data set was divided into a summer and a winter subset to see if an improvement for modelling the intra-event variability could be achieved.

3.4. Regression Analysis for Data Subsets

As for the preceding analysis the best sets of predictors were selected based on the AIC and are displayed in table 3. For the summer subset it is again the case that the sets of predictors are very similar for the response variables based on the mean and the ones based on the median. For both δ^{18}O and deuterium excess there are four common predictors each which are independent of the reference level. For the winter subset this looks a bit different: there are only two common predictors for the two response variables based on δ^{18}O and none for the ones based on deuterium excess. Here, the same remarkable issue as mentioned for the entire data set is present again, with RH_{grad} being the only selected predictor for the intra-event variability of deuterium excess as a difference to the mean. Another point that is striking when looking at all three tables is that the relative humidity is present in almost all predictor sets, either as logit-transformed values or as a gradient or both. This indicates that there is a close link between the relative humidity and the isotopic values.

The selected predictor sets were again used to fit multivariate linear models to the data which were then evaluated. The coefficients of determination can be found in table 3. Overall it seems as if a slight improvement to the models based on all data could be achieved. Apart from "δ^{18}O difference to median" from the summer subset and "deuterium excess difference to mean" from the winter subset an increase in the R^2-values of all models compared to their respective models based on all data can be seen. Nevertheless, all coefficients of determination for models regarding the intra-event variability of deuterium excess are still fairly low, with the highest one being only 0.303. Thus, even though these models might be better than the previous ones they are still far from being a good fit to the data. The highest R^2-value for either data subset was achieved for the δ^{18}O values as the difference to the median. For the summer subset the coefficient is 0.406 and for the winter subset it is 0.543. These values are still not really high, especially when taking into account that the models were only tested on the same data that was used when fitting them. If a new data set was used for the validation, the resulting coefficients would most likely be even lower again. However, as these two are the best models that could be achieved with the given data they will be looked at in more detail below.

But first, the same diagnostic plots as mentioned above were used again for further evaluation. This time, the distributions of the residuals as well as the Cook's distances were unremarkable for all models, meaning that no further inspection of single observations was necessary.

3.4.1. Modelling δ^{18}O as the Difference to the Median

For the summer subset of data, the formula of the fitted linear model is as follows:

$$\hat{y} = 1.89 - 0.356\ RH_{logit} - 0.144\ RR_{log} - 0.149\ FV - 1900\ \frac{RH_{grad}}{T}\ , \tag{8}$$

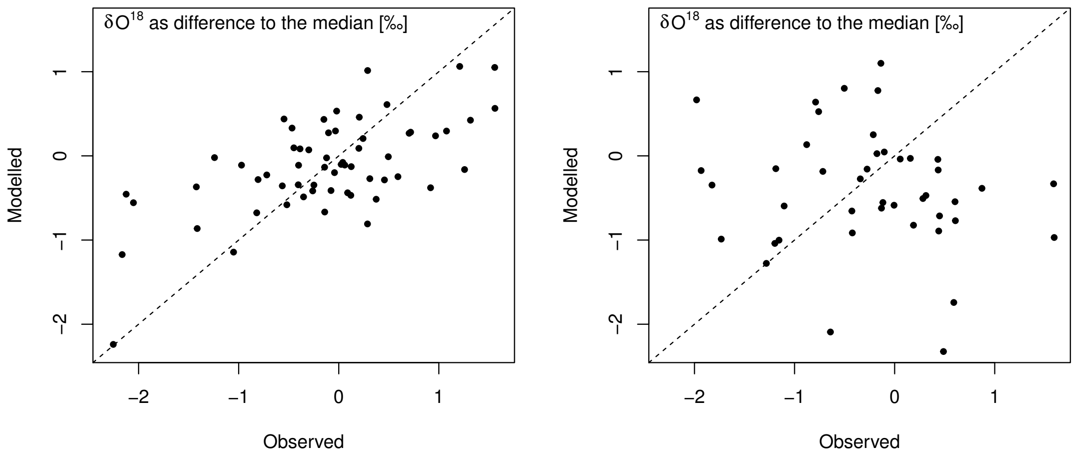
where $\hat{y}$ is the estimated value of the response variable, RH_{logit} is the logit-transformed relative humidity, RR_{log} is the log-transformed rain rate, FV is the falling velocity, and RH_{grad}/T is the ratio of the gradient of the relative humidity and the temperature. This means that the difference to the median of δ^{18}O is negatively correlated to all four predictor variables, i.e. an increase of one variable with all others remaining the same leads to a decrease of the difference and vice versa. As mentioned above the coefficient of determination for this regression equation is 0.406. Hence, only about 40 % of the variance of the response variable can be explained by the predictors, leaving the bigger part unexplained. The differences in the order of magnitude of the regression coefficients are probably due to the differences in the order of magnitude of the parameters, with the gradients being much smaller than the other values.

The regression equation based on the winter subset is as follows:

$$\hat{y} = 1.10 - 1.67\ WV + 1350\ RH_{grad} + 32.7\ RR_{grad} + 3.57\ \frac{LWC_{grad}}{RR_{grad}} + 1470\ \frac{FV_{grad}}{FV}\ , \tag{9}$$

with WV being the wind velocity, RH_{grad} being the gradient of the relative humidity, RR_{grad} being the gradient of the rain rate, LWC_{grad}/RR_{grad} being the ratio of the gradients of the liquid water content and the rain rate, and FV_{grad}/FV being the ratio of the gradient and the absolute value of the falling velocity. Here, the difference to the median of δ^{18}O is negatively correlated to the wind velocity and positively correlated to the four other predictors. The coefficient of determination for this equation is 0.543, meaning that a bit more than half of the variance of the response variable can be explained by the predictors.

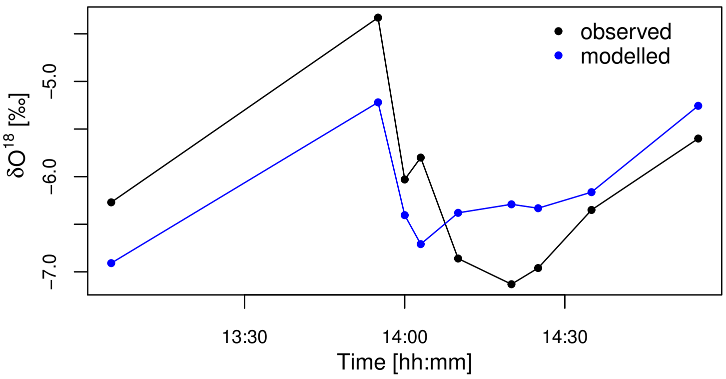
For both regression equations the modelled values were plotted against the observed values, as can be seen in figure 8. The dashed lines in these graphs are the bisecting lines and for a perfect model all points should lie on these lines. This is clearly not the case for these two models. Instead, the points are scattered around the lines and for the model based on the summer data there seems to be the tendency that small values are overestimated while high values are underestimated. The same applies to the model based on the winter data, but here the under- and overestimation of the highest and the lowest values is even more distinct. What is also quite remarkable for the winter subset is that all observed values that are greater than zero are modelled as negative values whereas for negative observations both positive and negative values were modelled.



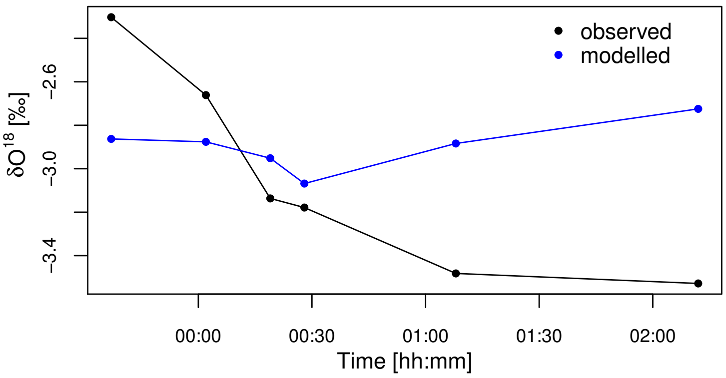
(a) based on the data subset *Summer* (b) based on the data subset *Winter*

Figure 8: Modelled values versus observed values for δ^{18}O as a difference to the event-based median for two data subsets. The dashed line is the bisecting line. For the summer subset the coefficient of determination of the model is 0.406 and for the winter subset it is 0.543.

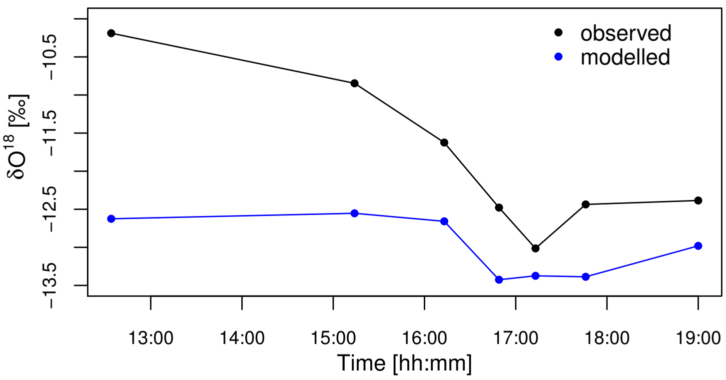
Even though the overall performance of the two models is not very good as was shown with figure 8 and the coefficients of determination, they were also tested graphically for modelling single events. The results of this were very mixed. For both, the summer and the winter model there are events of which the main trends can be reproduced fairly well. Other events can only partly be modelled well while for other parts of the same event the modelled values vary drastically from the observations. And then there are events for which the modelled values do not meet the observed ones at all and where modelled trends are the exact opposite to the observed trends. Four examples can be seen in figure 9. MA06 and RE05 are events belonging to the summer subset, whereas BO01 and BO04.1 belong to the winter subset. MA06 is one of the events where the main trends of the data, namely an increase followed by a decrease again followed by an increase, were reproduced well by the model. RE05 on the contrary is an example for opposite trends: at the end of the event the δ^{18}O values were decreasing but the modelled values show an increase instead. The same but even more distinct goes for BO04.1. Even small trends within the event were modelled as the opposite, making the observed and the modelled values almost look like mirror images. BO01 is one of the winter events where



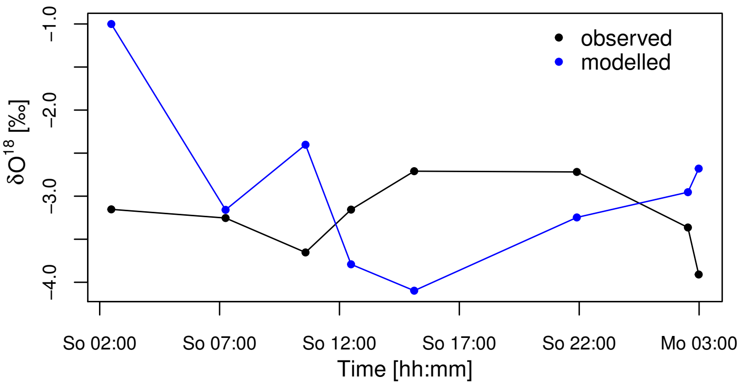
(a) Event MA06 - 02.08.2016



(b) Event RE05 - 18.09.2012



(c) Event BO01 - 18.12.2012



(d) Event BO04.1 - 06.01.2013

Figure 9: Comparing observed and modelled values for single events. The used response variable in all four cases is the difference of δ^{18}O to the event-based median. For (a) and (b) the model based on the summer subset was used as these two events belong to that subset, whereas (c) and (d) are from the winter subset and thus the corresponding model was tested here. (a) and (c) are examples for events where the main trends were modelled fairly well but for most events, including the ones in (b) and (d), the outcome was rather poor.

the main trends were modelled well, but still there is a huge deviation between modelled and observed values at the beginning of the event. So as assumed after the previous model evaluation the two models do not seem suitable for modelling single events, apart from a few exceptions.

4. Discussion

The resulting models do not reproduce the observed values very well for neither of the four response variables. Apart from one exception, all of them can only explain less than fifty percent of the variations of the respective response variable, making them not really suitable for any sort of prediction. What is surprising with regards to the δ^{18}O values is the discrepancy of the selected predictors between the summer and the winter subset. It was expected that the relative importance of the predictors might change or that there might even be a slight difference in the composition of the predictor sets, but they turned out to be very different. Some of the predictors are parameters that were mentioned in the Introduction and that were found to be related to the isotope signal by other studies before (e.g. Celle-Jeanton et al., 2004; Berkelhammer et al., 2012; Muller et al., 2015). Examples are the temperature (T), the relative humidity (RH) and the rain rate (RR) which is related to the rain intensity and thus the amount effect. For others, like the liquid water content (LWC) and the falling velocity (FV), the link to the isotopic composition has not been studied a lot before. But as the performance of the models is rather poor the presumably found relationships are not very reliable anyway. Nevertheless, it should be noted that for the regression equation for δ^{18}O as a difference to the median based on the summer data it is not only that T, RH and RR were found as predictors, but the correlations for these three parameters are also in accordance with the other studies: a positive correlation of T and the enrichment of heavy isotopes, and a negative correlation for RH as well as RR were found.

Measuring parameters like LWC, FV or RR using an MRR for isotopic studies is not very common yet. In general, MRRs are still considered as a “relatively new method” for the measurement of local precipitation parameters (Muller et al., 2015). According to Muller et al. (2015), their study was the first one using an MRR at a mid-latitude location to examine microphysical characteristics in a high resolution and to link them to the isotopic composition of precipitation. They found qualitative relationships between the isotopic values and the MRR variables, and in the results of this present study these relationships seem to be existent as well. As mentioned above the evaluation of the models suggests not to put too much trust in the found relations though. Nevertheless, the fact that parameters from the MRR show up in the final regression equations means that including them could improve the quality of the models. Thus, they seem to be useful in explaining the variations of the isotopic data. Especially when analysing the isotopic composition of rain water on small temporal and spatial scales the usage of an MRR can lead to new insights by providing data of meteorological variations and precipitation characteristics on these scales (Muller et al., 2015). Integrating MRR data into studies like the one at hand is therefore still considered to be a promising approach.

The same goes for integrating gradients of the measured parameters in the regression analysis. Here, no other studies which have attempted that before were found. But as four out of the five predictors of the regression equation based on the winter subset consist of or include gradients they appear to have an explanatory content for the variation of the isotopic composition. When looking at table 3 it can be seen that in every single one of the 12 best predictor sets that were determined for different response variables and based on different data (sub-)sets at least one predictor including a gradient is present. Hence, including not only the absolute values of meteorological parameters but also how they changed during the course of the rain water sample collection seems to provide useful information, especially when analysing variations on intra-event timescales.

For the response variables based on deuterium excess the goodness of fit of the models was even worse than for δ^{18}O. The decision to take deuterium excess (d) into account was based on several other studies doing so and some of them stating that there is a potential relationship between d and local factors or meteorological parameters. Barras and Simmonds (2009) for example argued that the d-value of the residual water of a raindrop can be reduced by rapid reevaporation, with the extent of the reduction being dependent on drop size, falling distance and the ambient vapour content. A relationship of d and the relative humidity was also found by Berkelhammer et al. (2012). In addition, the same study suggested a weakly negative correlation with temperature. But there are also other studies stating that the d-value is primarily related to the climatic characteristics of the source region of the moisture and that it is mainly affected by the conditions under which evaporation takes place there (e.g. Rindsberger et al., 1990; Gat, 1996; Araguás-Araguás et al., 2000). According to Rindsberger et al. (1990) the rainout of the airmass after the moisture formation has no impact on d. This could explain the very poor results from the regression analysis. If the d-value is really primarily dependent on the conditions at the moisture source it will be hard to model its variation using only local parameters from the precipitation site. On the other hand, the studies stating that d can be influenced during the falling of the droplets are more recent. Thus, it is also possible that there are relations of meteorological parameters and d, and that the reasons for the fact that these relations could not be modelled lie somewhere else.

With regards to the methods, one of the aspects that might have had a negative influence on the outcome of the analysis could be the inconsistency in the measuring techniques. For the MRR data the vertical spatial resolution got changed during the considered time period: until 2013 data was recorded in 35 m intervals and from 2014 onwards it was only recorded every 100 m. It was attempted to take these differences into account in the data preparation in order to maximise the comparability of the two data sets, but it cannot be ruled out that the inconsistency still had an effect on the final results. Furthermore, two different methods for taking rain water samples were used. Until 2014 samples were taken with a sequential precipitation collector and in 2016 they were obtained manually. Again, the data preparation was conducted in a way to minimise the possible impact of this discrepancy, but some aspects could not be subsequently straightened out. For example, the sequential precipitation collector started collecting rain water as soon as it began to rain whereas with the manual method the very beginning of some rain events might have been missed. In addition, the manual samples were taken in specific time intervals and the corresponding rain volume is not known, and for the sequential collector the samples were based on a specific volume but the corresponding time interval got also recorded. This last aspect should not have had an influence on the results though, as the decision was made to use the time- instead of the volume-based mean and median values. Nevertheless, more conformity in the sampling of precipitation might improve the overall result of the regression analysis.

Another aspect with regards to the manual sampling method that might have been problematic is the higher risk of obtaining biased values. When analysing the isotopic composition of precipitation it is important to minimise the possibility of reevaporation from the samples as this would alter the initial composition and thus lead to a bias (Berkelhammer et al., 2012). Even for the sequential precipitation collector it cannot be said with certainty that none of the samples encountered reevaporation but the risk is much lower here. The water is led from the catchment tray into the collecting flasks almost immediately, therefore the time in which evaporation of the collected rain water could occur is shortened. For the manual method, precipitation was collected in buckets, leaving it exposed to the ambient air for the entire duration of each sample collection. These lasted from 3 up to 300 minutes and therefore the risk of reevaporation and biased values was much greater with this sampling method.

For the meteorological data it is also possible that errors in measurement occurred. As described above, there were data lines for which the LWC and the RR were recorded as being zero while the weather station registered rain fall. These data lines with obvious discrepancies were excluded from the analysis, but it might be that they were not the only defective ones. There might have been more biased values but since they were not in disagreement with other recorded values they were included in the final data set. One reason for potential errors regarding the MRR data would be that assumptions which were made for the measuring and the calculation of the parameters were not met. To name one example, "stagnant air" is assumed for the relation of terminal FV and DS, and if downwind is present during the measurement this will lead to an overestimation of DS and an underestimation of LWC and RR (METEK GmbH, 2009). For the instruments of the weather station it cannot be guaranteed either that all recorded data is correct even though there were no indications of defective values.

Modelling the intra-event variability of the isotopic composition was attempted using the differences to the event-based mean and median respectively. Maybe a different way of expressing this variability within the events can be found that leads to an improvement of the performance of the resulting models.

One of the main issues leading to a poor result is probably connected with the set of possible predictors. The number of parameter pairs that were classified as being highly correlated was very high in comparison to the total number of parameters. As mentioned above, multicollinearity can cause problems for a regression analysis. Correlating predictors will for example lead to a high variance of the estimated regression coefficients (Cambridge dictionary: Multicollinearity, 2010). That is why ratios were calculated and one of the parameters from each pair was excluded from the analysis. Which parameter was excluded and which one was kept was determined randomly but has a huge impact on the final regression equation. Two ways leading to two different parameter sets were attempted: excluding the variables in the first column of the pairs and excluding the ones in the second column. Even though the coefficients of determination varied only marginally for the different sets, meaning that the goodness of fit was about the same, the parameters showing up in the final equations were completely different. Thus, this one random decision has a major impact on the interpretation of the final result, namely which parameters have an influence on the variability of the isotopic composition.

In general, the way in which the issue of collinear predictors was handled could be improved. Excluding so many of the initial parameters is not very desirable. As mentioned above, it was striking that especially variables from the same source were correlated among each other. In particular for the MRR data this is not surprising as many of the parameters are calculated from the same initially measured value (METEK GmbH, 2009). Thus, a good alternative to avoid the impacts of collinearity while keeping the information of all initial variables could be to perform a principal component analysis (PCA). The Cambridge Dictionary of Statistics (Everitt and Skrondal, 2010) defines PCA as a "procedure for analysing multivariate data which transforms the original variables into new ones that are uncorrelated and account for decreasing proportions of the variance in the data". The aim of a PCA is to recognize patterns in a data set and achieve a reduction in the dimensionality of it (Backhaus et al., 2016). The new variables, called principal components, are based on groups of variables that are highly correlated to each other but less correlated to other groups (Backhaus et al., 2016). They are expressed as linear functions of the initial parameters (Cambridge dictionary: PCA, 2010). Here, two principal components could be based on the data sets coming from either of the two data sources, that is the MRR and the weather station. Whether this approach can lead to an improvement in the modelling of the intra-event variability needs to be tested.

One last aspect to be discussed with regards to the methods is the split of the data into a summer and a winter subset. As mentioned above, this was based on studies stating that there is a seasonality to δ^{18}O values. Most of these studies though were based on data with a lower temporal resolution than the data used here. The distinct differences between the final regression equations for the summer and the winter data show that there is indeed a deviation between the two subsets, nevertheless other ways of splitting the data could be considered as well. Two examples would be the moisture source region and the precipitation type. The first might be harder to realise but the second criterion was studied by several authors already. It was for instance used in a study by Barras and Simmonds (2009) where the observations were classified into the three circulation types "mixed frontal", "convective" and "stratiform precipitation events". It revealed that the influence of some factors differed in their strength depending on the precipitation type. Miyake et al. (1968) even noted that the relationship of the ^{18}O content and the precipitation intensity for frontal precipitation was inverse to the relationship observed during convective conditions. And Berkelhammer et al. (2012) stated that there is a "positive relationship between percentage of convective precipitation and the isotopic composition of precipitation". Thus, splitting the data according to commonly occurring precipitation types in Potsdam might be a promising approach for improving the results.

When comparing the results of this study with the results from the study by Breier (2015) some aspects are worth mentioning. First of all, a better model performance was achieved in the latter one, with an R^2 of 0.685. This was for a model with the absolute δD values as a response variable in contrast to the measures of intra-event variability used in the present study. It suggests that the used data is more suitable for modelling absolute isotopic values rather than the differences to event-based mean or median values. Nevertheless, some findings were quite similar in both studies. Breier (2015) also applied the obtained model to single events and came to the same conclusion that some events can be modelled well while others cannot. Even the attribution of the events to these two categories was similar. In this context, he also stated that some of the events that could not be reproduced well were convective events and that that might be the reason for the poor results. This supports the idea of splitting the data set according to precipitation types as was mentioned above.

5. Conclusion

The variations of δ^{18}O and deuterium excess on intra-event timescales were assessed. Using meteorological data from a weather station and a vertically pointing MRR the goal was to model these variations. For this a multivariate linear regression model was used. Several approaches were taken based on different subsets of the data as well as different response variables, but none of the resulting models turned out to be very good. The model performances did not exceed an R^2 of 0.543. Even the two best models were not capable of reproducing the observed data in a satisfactory way, as was shown for the data subsets as well as for single events. Here, some exceptions were found where trends of the isotopic composition within single events were modelled fairly well. Overall though, the models do not seem to be suitable for predicting the intra-event variations of the isotopic composition of precipitation.

Even though it was not possible to model the variability with the given data and the used methods that does not mean that there is no relation between the studied parameters and the isotopic composition at all. Many authors have found relationships in this field on different scales before as was described above. This suggests that altering the methods can lead to an improvement in the performance of the resulting models. These changes could affect the data collection as well as the definition of the response variables, the handling of collinear predictors or the subdivision of the data set. Using a principal component analysis for all potential predictors and differentiating between events according to the present precipitation type is considered to be most promising. If these alterations can indeed lead to a better modelling of the intra-event variability of the isotopic composition of precipitation this would provide new insights into the local parameters and mechanisms that are important on these short timescales.

References

Akaike's information criterion (AIC). In B. S. Everitt and A. Skrondal, editors, *The Cambridge dictionary of statistics*, page 10. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-76699-9.

Multicollinearity. In B. S. Everitt and A. Skrondal, editors, *The Cambridge dictionary of statistics*, page 287. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-76699-9.

Cook's distance. In B. S. Everitt and A. Skrondal, editors, *The Cambridge dictionary of statistics*, pages 104–105. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-76699-9.

Principal components analysis. In B. S. Everitt and A. Skrondal, editors, *The Cambridge dictionary of statistics*, page 337. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-76699-9.

Coefficient of determination. In B. S. Everitt and A. Skrondal, editors, *The Cambridge dictionary of statistics*, page 89. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-76699-9.

Residual plots. In B. S. Everitt and A. Skrondal, editors, *The Cambridge dictionary of statistics*, pages 367–368. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-76699-9.

H. Akaike. Likelihood of a model and information criteria. *Journal of Econometrics*, 16 (1):3–14, 1981. ISSN 03044076. doi: 10.1016/0304-4076(81)90071-3.

L. Araguás-Araguás, K. Froehlich, and K. Rozanski. Deuterium and oxygen-18 isotope composition of precipitation and atmospheric moisture. *Hydrological Processes*, 14(8): 1341–1355, 2000. ISSN 0885-6087.

K. Backhaus, B. Erichson, W. Plinke, and R. Weiber. *Multivariate Analysemethoden*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016. ISBN 978-3-642-14986-3. doi: 10.1007/978-3-662-46076-4.

V. Barras and I. Simmonds. Observation and modeling of stable water isotopes as diagnostics of rainfall dynamics over southeastern Australia. *Journal of Geophysical Research*, 114(D23):674, 2009. ISSN 0148-0227. doi: 10.1029/2009JD012132.

M. Berkelhammer, L. Stott, K. Yoshimura, K. Johnson, and A. Sinha. Synoptic and mesoscale controls on the isotopic composition of precipitation in the western United States. *Climate Dynamics*, 38(3-4):433–454, 2012. ISSN 0930-7575. doi: 10.1007/s00382-011-1262-3.

J. Breier. *Analyse und Modellierung der Niederschlagsisotopie anhand von meteorologischen Parametern am Standort Golm.* Bachelorarbeit, Universität Potsdam, Potsdam, 2015.

H. Celle-Jeanton, R. Gonfiantini, Y. Travi, and B. Sol. Oxygen-18 variations of rainwater during precipitation: Application of the Rayleigh model to selected rainfalls in Southern France. *Journal of Hydrology*, 289(1-4):165–177, 2004. ISSN 00221694. doi: 10.1016/j.jhydrol.2003.11.017.

T. B. Coplen. Reporting of stable hydrogen, carbon, and oxygen isotopic abundances (technical report). *Pure and Applied Chemistry*, 66(2), 1994. ISSN 1365-3075. doi: 10.1351/pac199466020273.

T. B. Coplen, P. J. Neiman, A. B. White, J. M. Landwehr, F. M. Ralph, and M. D. Dettinger. Extreme changes in stable hydrogen isotopes and precipitation characteristics in a landfalling pacific storm. *Geophysical Research Letters*, 35(21):1063, 2008. ISSN 0094-8276. doi: 10.1029/2008GL035481.

H. Craig. Isotopic variations in meteoric waters. *Science (New York, N.Y.)*, 133(3465): 1702–1703, 1961. ISSN 0036-8075. doi: 10.1126/science.133.3465.1702.

W. Dansgaard. Stable isotopes in precipitation. *Tellus*, 16(4):436–468, 1964. ISSN 00402826. doi: 10.1111/j.2153-3490.1964.tb00181.x.

B. S. Everitt and A. Skrondal, editors. *The Cambridge dictionary of statistics*. Cambridge University Press, Cambridge, 4. ed. edition, 2010. ISBN 978-0-521-76699-9.

X. Feng, A. M. Faiia, and E. S. Posmentier. Seasonality of isotopes in precipitation: A global perspective. *Journal of Geophysical Research*, 114(D8):28721, 2009. ISSN 0148-0227. doi: 10.1029/2008JD011279.

R. D. Field. Observed and modeled controls on precipitation $\delta 18$ O over Europe: From local temperature to the Northern Annular Mode. *Journal of Geophysical Research*, 115(D12):1341, 2010. ISSN 0148-0227. doi: 10.1029/2009JD013370.

T. Francke and J. Brenner. R functions: parse_mrr and read_mrrdata, unpublished, 2011.

I. Friedman, G. I. Smith, J. D. Gleason, A. Warden, and J. M. Harris. Stable isotope composition of waters in southeastern California 1. Modern precipitation. *Journal of Geophysical Research*, 97(D5):5795, 1992. ISSN 0148-0227. doi: 10.1029/92JD00184.

J. R. Gat. Oxygen and hydrogen isotopes in the hydrologic cycle. *Annual Review of Earth and Planetary Sciences*, 24(1):225–262, 1996. ISSN 0084-6597. doi: 10.1146/annurev.earth.24.1.225.

S. D. Gedzelman and J. R. Lawrence. The isotopic composition of precipitation from two extratropical cyclones. *Monthly Weather Review*, 118(2):495–509, 1990. ISSN 0027-0644. doi: 10.1175/1520-0493(1990)118<0495:TICOPF>2.0.CO;2.

J. Hoefs. *Stable Isotope Geochemistry*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 978-3-540-70703-5. doi: 10.1007/978-3-540-70708-0.

G. Hoffmann, J. Jouzel, and V. Masson. Stable water isotopes in atmospheric general circulation models. *Hydrological Processes*, 14(8):1385–1406, 2000. ISSN 0885-6087. doi: 10.1002/1099-1085(20000615)14:8<1385::AID-HYP989>3.3.CO;2-T.

J.-E. Lee and I. Fung. "Amount effect" of water isotopes and quantitative analysis of post-condensation processes. *Hydrological Processes*, 22(1):1–8, 2008. ISSN 0885-6087. doi: 10.1002/hyp.6637.

METEK GmbH. MRR Physical Basics: Version 5.2.0.1, 2009.

METEK GmbH. MRR FAQs: Quick answers to technical questions, 2017. URL `http://metek.de/faqs/`. Last accessed 07-09-2017.

Y. Miyake, O. Matsubaya, and C. Nishihara. An isotopic study on meteoric precipitation. *Papers in Meteorology and Geophysics*, 19(2):243–266, 1968. doi: 10.2467/mripapers1950.19.2_243.

C. L. Muller, A. Baker, I. J. Fairchild, C. Kidd, and I. Boomer. Intra-Event Trends in Stable Isotopes: Exploring Midlatitude Precipitation Using a Vertically Pointing Micro Rain Radar. *Journal of Hydrometeorology*, 16(1):194–213, 2015. doi: 10.1175/JHM-D-14-0038.1.

N. C. Munksgaard, C. M. Wurster, A. Bass, and M. I. Bird. Extreme short-term stable isotope variability revealed by continuous rainwater analysis. *Hydrological Processes*, 26(23):3630–3634, 2012. ISSN 0885-6087. doi: 10.1002/hyp.9505.

C. Priest. R function: colin_fun, 2015. URL `https://colinpriest.com/tag/glms/`. Last accessed 07-09-2017.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL `https://www.R-project.org/`. Last accessed 07-09-2017.

M. Rindsberger, S. Jaffe, S. Rahamim, and J. R. Gat. Patterns of the isotopic composition of precipitation in time and space: Data from the Israeli storm water collection program. *Tellus B: Chemical and Physical Meteorology*, 42(3):263–271, 1990. ISSN 1600-0889. doi: 10.3402/tellusb.v42i3.15218.

K. Rozanski, L. Araguás-Araguás, and R. Gonfiantini. Relation between long-term trends of oxygen-18 isotope composition of precipitation and climate. *Science (New York, N.Y.)*, 258(5084):981–985, 1992. ISSN 0036-8075. doi: 10.1126/science.258.5084.981.

K. Rozanski, L. Araguás-Araguás, and R. Gonfiantini. Isotopic patterns in modern global precipitation. *Climate Change in Continental Isotopic Records*, 1993. doi: 10.1029/GM078p0001.

C. Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72, 1904. ISSN 00029556. doi: 10.2307/1412159.

Acknowledgement

I would like to thank my two supervisors, Till Francke and Frauke Barthold, for their support and for taking the time to help me and answer my questions. Thanks also to Heide Kraudelt for taking the rain water samples during the summer 2016 and for showing me how to do it.

Furthermore I would like to thank Jannes for meeting up with me and helping me to get started (and for giving me advice and helping me with my decision about the master's degree).

And last but not least a big thanks to Juliana for the good writing company, and to Alicja, Christina and Tadhg for their feedback.

Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbstständig verfasst habe. Andere als die angegebenen Quellen und Hilfsmittel wurden nicht verwendet. Die den verwendeten Quellen wörtlich oder inhaltlich entnommenen Abschnitte sind als solche kenntlich gemacht. Des weiteren versichere ich, dass die vorliegende Arbeit noch nicht im Rahmen eines anderen Prüfungsverfahrens eingereicht wurde.

Potsdam,

Elena Macdonald

A. Appendix

Contents

A.1. Data and R scripts 31
A.2. Isotopic composition as a function of time for all events 32

A.1. Data and R scripts

The data as well as the R scripts and functions that were used for the analysis and visualization of the data can be found in the digital appendix. The following list provides an overview.

1. Data
    - Isotopic data from rainwater samples
    - Meteorological data from a weather station
    - Meteorological data from a micro rain radar
    - Assembled data from all sources for each rain event
2. R scripts and functions
    - *AssembleData_35* and *AssembleData_100*
    - *parse_mrr* and *read_mrrdata*
    - *allData_Regression*
    - *colin_fun*
    - *allData_graphs*

A.2. Isotopic composition as a function of time for all events