

## *The stability and calibration of water vapor isotope ratio measurements during long-term deployments*

The Faculty of Oregon State University has made this article openly available.  
Please share how this access benefits you. Your story matters.

<b>Citation</b>	Bailey, A., Noone, D., Berkelhammer, M., Steen-Larsen, H. C., & Sato, P. (2015). The stability and calibration of water vapor isotope ratio measurements during long-term deployments. <i>Atmospheric Measurement Techniques</i> , 8(10), 4521-4538. doi:10.5194/amt-8-4521-2015
<b>DOI</b>	10.5194/amt-8-4521-2015
<b>Publisher</b>	Copernicus Publications
<b>Version</b>	Version of Record
<b>Terms of Use</b>	<a href="http://cdss.library.oregonstate.edu/sa-termsfuse">http://cdss.library.oregonstate.edu/sa-termsfuse</a>



# The stability and calibration of water vapor isotope ratio measurements during long-term deployments

A. Bailey<sup>1,2,a</sup>, D. Noone<sup>1,2,3</sup>, M. Berkelhammer<sup>4</sup>, H. C. Steen-Larsen<sup>5</sup>, and P. Sato<sup>6</sup>

<sup>1</sup>Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, Boulder, Colorado, USA

<sup>2</sup>Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado, USA

<sup>3</sup>College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, USA

<sup>4</sup>Department of Earth and Environmental Sciences, University of Illinois at Chicago, Chicago, Illinois, USA

<sup>5</sup>Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette, France

<sup>6</sup>Joint Institute for Marine and Atmospheric Research, NOAA, Hilo, Hawaii, USA

<sup>a</sup>now at: Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, Washington, USA

Correspondence to: A. Bailey (adriana.bailey@colorado.edu)

Received: 4 March 2015 – Published in Atmos. Meas. Tech. Discuss.: 28 May 2015

Revised: 5 September 2015 – Accepted: 2 October 2015 – Published: 27 October 2015

**Abstract.** With the recent advent of commercial laser absorption spectrometers, field studies measuring stable isotope ratios of hydrogen and oxygen in water vapor have proliferated. These pioneering analyses have provided invaluable feedback about best strategies for optimizing instrumental accuracy, yet questions still remain about instrument performance and calibration approaches for multi-year field deployments. With clear scientific potential for using these instruments to carry out monitoring of the hydrological cycle, this study examines the long-term stability of the isotopic biases associated with three cavity-enhanced laser absorption spectrometers – calibrated with different systems and approaches – at two remote field sites: Mauna Loa Observatory, Hawaii, USA, and Greenland Environmental Observatory, Summit, Greenland. The analysis pays particular attention to the stability of measurement dependencies on water vapor concentration and also evaluates whether these so-called concentration dependences are sensitive to statistical curve-fitting choices or measurement hysteresis. The results suggest evidence of monthly-to-seasonal concentration-dependence variability – which likely stems from low signal-to-noise at the humidity-range extremes – but no long-term directional drift. At Mauna Loa, where the isotopic analyzer is calibrated by injection of liquid water standards into a vaporizer, the largest source of inaccuracy in characterizing the concentration dependence stems from an insufficient density of calibration points at low water vapor volume mixing ra-

tios. In comparison, at Summit, the largest source of inaccuracy is measurement hysteresis associated with interactions between the reference vapor, generated by a custom dew point generator, and the sample tubing. Nevertheless, prediction errors associated with correcting the concentration dependence are small compared to total measurement uncertainty. At both sites, changes in measurement repeatability that are not predicted by long-term linear drift estimates are a larger source of error, highlighting the importance of measuring isotopic standards with minimal or well characterized drift at regular intervals. Challenges in monitoring isotopic drift are discussed in light of the different calibration systems evaluated.

## 1 Introduction

The isotope ratios of hydrogen and oxygen ( $^2\text{H}/^1\text{H}$ ,  $^{18}\text{O}/^{16}\text{O}$ ) are powerful tracers of water cycle processes. Due to their lower saturation vapor pressure, the heavier isotopes ( $^2\text{H}$  and  $^{18}\text{O}$ ) preferentially condense, while the lighter isotopes preferentially evaporate (Bigeleisen, 1961; Dansgaard, 1964). Paired with humidity information, isotope ratios thus provide clues about sources of moisture to the atmosphere and about the integrated condensation history of air masses (Gat, 1996).

With the recent advent of commercial vapor isotopic analyzers, measurements of isotope ratios in water vapor have become increasingly widespread. As a result, field experiments once limited to a small number of flask samples (e.g. Ehhalt, 1974; Galewsky et al., 2007) – whose vapor content must be captured through a cryogenic trap for later liquid analysis in the lab – have been replaced by field experiments in which in situ observations can be made at a temporal resolution better than 0.1 Hz. Researchers using these new commercial technologies are resolving water cycle processes on a range of local-to-regional scales, investigating, for example, water recycling within the forest canopy (Berkelhammer et al., 2013), evapotranspiration (Wang et al., 2010) and its contribution to atmospheric moisture (Noone et al., 2013; Aemisegger et al., 2014), mixing and convective processes in the atmosphere (Noone et al., 2011; Tremoy et al., 2012; Bailey et al., 2013, 2015), evaporation processes in the marine boundary layer (Steen-Larsen et al., 2014b, 2015), and large-scale condensation and advection dynamics (Galewsky et al., 2011; Hurley et al., 2012; Steen-Larsen et al., 2013).

Lessons learned from these early field programs are informing designs for longer-term observational campaigns, including the National Ecological Observatory Network's (NEON) plans to measure water vapor isotope ratios for 3 decades at sites across the United States (Luo et al., 2013). Yet questions still remain about the long-term stability of commercial isotopic analyzers and the best approaches for calibrating in the field to ensure high-quality long time series. The goal of this paper is to provide guidance for field deployments by evaluating the stability of the isotopic biases identified in three spectroscopic analyzers operated over periods of 3 years at two remote sites. Unlike laboratory tests that are more common in the literature, the calibration experiments described herein were conducted under variable and often adverse environmental conditions, which required that the instruments operate without maintenance for extended periods. It is precisely because of these challenges that the results of this work may be particularly relevant to future long-term measurement programs.

While previous studies suggest isotopic biases are specific to each individual commercial analyzer (e.g. Tremoy et al., 2011; Aemisegger et al., 2012; Wen et al., 2012), there are nevertheless shared characteristics upon which “best practices” for instrument operation and calibration can be based. A prominent and ubiquitous source of measurement bias stems from the tendency of the isotope ratio measured to change as a function of the water vapor volume mixing ratio, creating a so-called “concentration dependence,” which numerous studies describe (e.g. Lis et al., 2008; Schmidt et al., 2010; Sturm and Knohl, 2010; Johnson et al., 2011; Noone et al., 2011, 2013; Rambo et al., 2011; Tremoy et al., 2011; Aemisegger et al., 2012; Wen et al., 2012; Bailey et al., 2013; Steen-Larsen et al., 2013, 2014b; Bastrikov et al., 2014; Bonne et al., 2014; Samuels-Crow et al., 2014). While a few have found the dependence of isotope ratio on water va-

por concentration to be near linear (e.g. Lis et al., 2008; Wen et al., 2012), most have found it to be nonlinear and specific to both the instrument used and the isotope ratio measured (i.e.  $\delta D$  or  $\delta^{18}O$ , where  $\delta = (R_{\text{observed}}/R_{\text{standard}} - 1) \times 1000$  and  $R = {}^2H/{}^1H$  or  ${}^{18}O/{}^{16}O$ , respectively). Moreover, biases in the individual isotope ratios can be quite significant; Sturm and Knohl (2010) showed that failing to account for the concentration dependence of their analyzer resulted in a bias in the second-order deuterium excess parameter ( $d = \delta D - 8 \times \delta^{18}O$ ) of upwards of 25 %.

Instrumental drift creates another source of isotopic bias by influencing measurement repeatability on hourly timescales or longer. In this analysis, repeatability is defined as the ability of the analyzer to measure the same value for replicate samples at distinct times. It should not be confused with instrumental precision, which is defined here as the analyzer's ability to repeat measurements on very short timescales (e.g. seconds to minutes). Random errors associated with precision are not corrected through calibrations but are instead minimized by optimizing the measurement averaging time – a procedure typically aided by Allan variance plots (cf. Sturm and Knohl, 2010; Aemisegger et al., 2012).

Drift-induced changes in repeatability may be corrected by measuring the same isotopic standard at a constant humidity level at different times. Using such an approach, some studies have found no significant drift over multiple hours (Koehler and Wassenaar, 2011; van Geldern and Barth, 2012), while others claim significant variability in measurement repeatability on daily timescales (Gupta et al., 2009; Tremoy et al., 2011; Aemisegger et al., 2012). Steen-Larsen et al. (2013), for instance, reported large daily variability – as high as 4 ‰ in  $\delta^{18}O$  and 16 ‰ in  $\delta D$  – and observed seasonal drift in one of two isotope ratios and one of two instruments deployed. Sturm and Knohl (2010) similarly observed consistent enrichment in one isotope ratio over the course of 2 weeks but no change in the other. Steen-Larsen et al. (2014b), meanwhile, observed drift in both isotope ratios but of opposite sign during a 500-day deployment in Bermuda. Possible sources of such variability may be instrument sensitivities to fluctuations in environmental factors such as temperature (Sturm and Knohl, 2010; Rambo et al., 2011; Steen-Larsen et al., 2013) or uncertainties in the characterization of the concentration dependence with time (Sturm and Knohl, 2010).

Correcting concentration-dependent and drift-induced measurement biases requires a source of water vapor of known concentration and known isotope ratio; yet, unlike most trace gases, no gas-phase isotopic standard exists for water. Instead, a vapor stream, whose isotope ratio is sufficiently stable and whose volume mixing ratio can be adjusted to span the ambient humidity range, must be produced on site from a liquid standard – a task that is not trivial (e.g. Kurita et al., 2012). Previous studies have experimented with a variety of calibration systems for this purpose. Several experiments, for example, have used commercial (Wang et al., 2009) or custom (Wen et al., 2012; Ellehoj et al., 2013; Samuels-Crow

et al., 2014; Steen-Larsen et al., 2014b) dew point generators. These produce a range of water vapor volume mixing ratios but are sensitive to isotopic fractionation as the liquid progressively evaporates from the reservoir. An alternative strategy involves pumping, dripping, or nebulizing liquid water into a stream of dry air (Iannone et al., 2009; Gkinis et al., 2010; Sturm and Knohl, 2010; Rambo et al., 2011; Tremoy et al., 2011; Aemisegger et al., 2012; Kurita et al., 2012; Wen et al., 2012; Steen-Larsen et al., 2013, 2015; Bastrikov et al., 2014; Bonne et al., 2014). Some such systems require that the dry air be heated to ensure complete evaporation of the liquid while others – such as microdrop generators (cf. Iannone et al., 2009) – do not. These systems are advantageous because they produce a continuous stream of water vapor whose isotope ratio should equal that of the liquid standard. Kurita et al. (2012), however, found that their commercial nebulizer produced concentration-dependent biases in the measured isotope ratios. A third strategy involves flash-evaporating discrete liquid samples within a heated chamber (Lis et al., 2008; Schmidt et al., 2010; Penna et al., 2010; van Geldern and Barth, 2012; Noone et al., 2013). These systems require small amounts of liquid standard but do not reproduce the sampling conditions under which ambient measurements are made. Wen et al. (2012) provides a more in-depth discussion of each of these three types of calibration systems. In the experiments described here, only a custom dew point generator and a discrete liquid autosampler are discussed in detail.

While each type of system has its advantages and disadvantages, a common complication with any isotopic calibration system is hysteresis, caused by water adhering to either the instrument cavities or inlet materials. Calibration tests using flash-evaporated liquid isotopic standards, for example, have demonstrated that “memory effects” frequently affect the first injections following a change in standard water (Lis et al., 2008; Gröning, 2011; Penna et al., 2012; van Geldern and Barth, 2012). Other studies have shown that tubing material connecting the calibration system to the analyzer can slow the analyzer’s response time, with Synflex particularly problematic for  $\delta D$  (Sturm and Knohl, 2010; Tremoy et al., 2011). Both Lee et al. (2005) and Sturm and Knohl (2010) speculated that failing to account for such measurement inaccuracies might result in a poor characterization of the concentration dependence and, ultimately, influence interpretation of scientific results.

Building on these previous analyses, this study extends our understanding of long-term stability in water vapor isotopic analyzers by evaluating concentration-dependent and drift-induced isotopic biases using distinct calibration systems at two remote measurement sites: the Mauna Loa Observatory on the Big Island of Hawaii, USA, and the Greenland Environmental Observatory at Summit, Greenland. First, the sensitivity of the concentration-dependence characterization to statistical treatments and measurement hysteresis is examined in the context of the distinct calibration systems used.

Possible variations in the concentration dependence with isotope ratio and time are then tested. Second, the influence of isotopic drift on measurement repeatability is analyzed for periods spanning 6 to 36 months. Finally, uncertainties in correcting concentration-dependent and drift-induced biases are compared with instrumental precision and ambient variability at each field site. Recommendations for calibration strategies are presented in the conclusions of the paper.

## 2 Methods

The data for this study were collected using three Picarro, Inc. water vapor isotopic analyzers, which were operated at two baseline observatories: the Mauna Loa Observatory (MLO, 3400 m) on the Big Island of Hawaii, USA, and the Greenland Environmental Observatory (3200 m) at Summit, Greenland. Frequent arid conditions at these locations provide an ideal test of the signal-to-noise strength of current commercial spectroscopy. Monitoring of the concentration dependence and isotopic drift at Mauna Loa was primarily achieved by flash-evaporating injections of discrete liquid samples from five standard waters. In comparison, a single isotopic standard and a custom dew point generator were used for this purpose at Summit. The instrumental setup and calibration systems are described in greater detail below and are summarized in Table 1.

### 2.1 Instruments

Picarro’s spectroscopic analyzers are one of several commercial water vapor isotopic analyzers based on cavity-enhanced near-infrared laser absorption spectroscopy. Los Gatos Research, Inc. also makes an analyzer used widely in the field (e.g. Noone et al., 2011; Rambo et al., 2011). These instruments exploit near-infrared light to measure the absorption-line features for three water isotopologues:  $^1H_2^{16}O$ ,  $^1H^2H^{16}O$  (i.e.  $^1HD^{16}O$ ), and  $^1H_2^{18}O$ . Cavity-enhanced techniques help create a longer effective absorption path length, which mitigates the very weak absorption of water vapor isotopologues in the near infrared.

### 2.2 Mauna Loa, Hawaii

Since October 2010, water vapor isotope ratios have been measured at MLO with a Picarro analyzer model L1115-i. Data from the period October 2010–September 2013 are examined in this analysis. The instrument, which is housed in the Charles Keeling building at the observatory, samples ambient air through 0.25-inch OD stainless steel tubing at a rate of approximately  $300 \text{ cc min}^{-1}$ . The stainless steel tubing protrudes through the roof of the building, through a plastic pipe, which has a rain cap to prevent precipitation from entering. The bulk of the stainless steel inlet line is housed inside the building and thus maintained at room temperature, which far exceeds the ambient dew point.

**Table 1.** Isotopic analyzers, calibration systems, and calibration approaches discussed in the analysis.

Location	Model	Calibration system	Calibration approach	Motivation	Secondary system	Secondary approach	Motivation
Mauna Loa Observatory Hawaii, USA	Picarro L1115-i	LEAP autosampler with five standard waters	Oct 2010–Feb 2012 Three injections 6-hourly, one standard, multiple $q$ Feb 2012–Sept 2013* 24 injections weekly, two standards, multiple $q$ (2–20 mmol mol <sup>-1</sup> )	Correct concentration dependence Monitor isotopic drift Normalize data to VSMOW-SLAP	Syringe pump with three standard waters	Feb 2012 Sampled each standard at 9–14 distinct water vapor volume mixing ratios (0.2–20 mmol mol <sup>-1</sup> )	Refine concentration-dependence correction
Greenland Environmental Observatory Summit, Greenland	Two Picarro L2120-i analyzers	Custom dew point generator with one standard water	Jul 2012–Jul 2013 One $q$ -level 6-hourly (~3.2 mmol mol <sup>-1</sup> ) Performed “extended” concentration calibrations in Jun, Jul, and Sep 2012 and in Apr and May 2013 (0.1–8 mmol mol <sup>-1</sup> ) Jul 2013–Dec 2014* Three $q$ -levels 6-hourly (0.05–4 mmol mol <sup>-1</sup> )	Correct concentration dependence Monitor isotopic drift	Syringe pump with three standard waters	Jul 2013 Sampled each standard at a single $q$ (~10 mmol mol <sup>-1</sup> )	Normalize data to VSMOW-SLAP

\* Calibrations continued past these dates, but the resulting data are not discussed in the analysis.

To verify the quality of the water vapor volume mixing ratio measurements ( $q$ ) used in this analysis, the data are compared with MLO's hourly-averaged dew point values, which are measured by hygrometer. A simple linear regression between the two data sets – after converting the MLO dew points to volume mixing ratios and averaging and interpolating the Picarro data – produces a slope of 1.00, and an offset of  $0.33 \text{ mmol mol}^{-1}$ . This suggests a small uniform low bias in the uncalibrated  $q$  measurements. However, since the accuracy of the MLO dew point measurements is not fully known, no adjustments are made to the Picarro volume mixing ratio data.

For most of the instrument's deployment, isotopic measurements at MLO have been calibrated using a LEAP Technologies PAL (Prep and Load) autosampler. Liquid samples from five secondary standards spanning approximately  $-45$  to  $0 \text{ ‰}$  in  $\delta^{18}\text{O}$  and  $-355$  to  $0 \text{ ‰}$  in  $\delta\text{D}$  are injected by syringe into a vaporizer, which flash-evaporates the liquid into commercial dry air (e.g. “zero-grade”) before delivery to the instrument. The volume of water injected controls the concentration of the sample. During the first 500 days of the analyzer's deployment, three injections of a single standard were made every 6 h. For the remainder of the 3-year period analyzed, weekly calibrations were performed in which one standard was injected 18 times at a variety of volume mixing ratios, typically spanning  $2\text{--}20 \text{ mmol mol}^{-1}$ , and a second standard was injected six times at a volume mixing ratio near  $10 \text{ mmol mol}^{-1}$  or greater. Monitoring of both the concentration dependence and deviation from the VSMOW-SLAP (Vienna Standard Mean Ocean Water – Standard Light Antarctic Precipitation) reference scale were accomplished in this manner.

The concentration dependence of the instrument was investigated in greater detail during a few days in February 2012 by performing calibrations over a large range of water vapor volume mixing ratios ( $0.2\text{--}20 \text{ mmol mol}^{-1}$ ). These calibrations were performed with a custom syringe pump, which steadily injects liquid standard into a stream of heated dry air. The system is similar to those described by Lee et al. (2005) and Gkinis et al. (2010). Unlike the PAL autosampler, the syringe-pump system provides a continuous flow of vapor to the instrument; and, by altering the rates of both the liquid injection and the dry airflow, much lower volume mixing ratios can be achieved. However, because the syringe pump was only used for a short time period, its performance is not discussed in the analysis.

### 2.3 Summit, Greenland

Two model L2120-i Picarro analyzers (named “Spiny” and “Gulper” after two types of dogfish shark) were deployed at Summit, Greenland in boreal summer 2011 through boreal summer 2014. The instruments were housed in an enclosed rack in an underground laboratory. The temperature of the laboratory was approximately  $10 \text{ °C}$  for the duration of the

experiment, and the temperature of the enclosed rack was maintained at  $15.0 \pm 0.2 \text{ °C}$ . The water vapor volume mixing ratio measurements were calibrated at the beginning of the deployment with an LI-610 portable dew point generator made by LI-COR.

Due to the need for isotopic calibrations in Greenland to run without maintenance for 11-month periods, a custom dew point generator (DPG) was developed to produce water vapor and calibrate both Summit instruments simultaneously approximately every 6 h. A system with similar design elements was used by Ellehoj et al. (2013). Commercially available calibration systems were found unsuitable for this purpose.

In the custom system, dry air from an industrial regenerative dryer (with a dew point temperature of  $-100 \text{ °C}$ ;  $q \leq 0.003 \text{ mmol mol}^{-1}$ ) was supplied to a 10-L Schott laboratory bottle containing water of known isotope ratio. By bubbling dry air through the liquid in this manner, vapor was produced whose isotope ratio ( $R_v = R_l/\alpha$ ) could be calculated as a function of the temperature-dependent fractionation factor  $\alpha$  and the isotope ratio of the liquid ( $R_l$ ). The temperature of the bottle was maintained near  $20 \text{ °C}$  by applying heat to a copper sleeve enveloping the glass. The water vapor volume mixing ratio of the air stream delivered to the instruments was altered through dry-air dilution; and a second-stage dilution immediately upstream of the analyzers was used to achieve the lowest volume mixing ratios.

Since the DPG was designed to maintain the vapor and liquid within the bottle in equilibrium, one would expect the removal of liquid water from the reservoir with time to have caused the isotopic composition of the water vapor produced to follow a predictable distillation described by Wang et al. (2009):

$$R_v = \frac{R_{l0}}{\alpha} \cdot \left(1 - \frac{t}{\tau}\right)^{\frac{1}{\alpha}-1} \quad (1)$$

Here,  $R_{l0}$  is the initial isotope ratio of the liquid water,  $\tau$  is the time necessary to evaporate all liquid from the bottle, and  $t$  is the time elapsed. Had it been possible to access the Summit site more frequently, isotopic depletion of the liquid standard could have been verified through periodic sampling; however, only water vapor measurements are available for the field campaign period. The analysis thus considers whether any or all long-term drift in the Summit water vapor data can be explained by the distillation described by Eq. (1).

Several distinct calibration approaches were used at Summit. From the summer of 2012 until the summer of 2013, drift calibrations were performed every 6 hours at a single isotope ratio and volume mixing ratio. The concentration dependences of Spiny and Gulper were evaluated in June, July, and September of 2012 and in April and May of 2013 by accessing the mass flow controller remotely and slowly altering dry air dilution of the DPG vapor stream to produce a large range of volume mixing ratios ( $\sim 0.1\text{--}8 \text{ mmol mol}^{-1}$ ) over

the course of several hours. These “extended” concentration–calibration periods were performed both by increasing and by decreasing the water vapor concentration progressively. Beginning in summer 2013, the protocol was modified to monitor the analyzers’ concentration dependences more frequently: 6-hourly calibrations were performed at three volume mixing ratios spanning 0.05–4 mmol mol<sup>-1</sup>, and no additional extended concentration calibrations were performed.

For all of the 6-hourly calibrations where  $q > 0.5$  mmol mol<sup>-1</sup>, the first 9 of 20 minutes spent sampling at a given volume mixing ratio are excluded from the analysis in order to eliminate possible memory effects. For calibrations performed at 0.5 mmol mol<sup>-1</sup>, the first 14 of 20 minutes spent sampling are excluded from the analysis. And at lower volume mixing ratios, where longer sampling was prescribed, the first 19 of 40 minutes spent sampling are excluded.

Since the DPG only produced a single isotopic standard, additional calibrations were performed in July 2013 to normalize the data to the VSMOW-SLAP scale. Three standard waters and the same syringe-pump system used at Mauna Loa (see Sect. 2.2) were employed for this purpose.

#### 2.4 Statistical methods for characterizing and correcting isotopic biases

To evaluate the stability of the concentration dependences at both sites, isotopic data are first normalized to a reference humidity during set time intervals. For each standard used at Mauna Loa, isotope ratio measurements are normalized to the 9–11 mmol mol<sup>-1</sup> range. This is done once for the syringe-pump data and every 3 months for the autosampler data. At Summit, where a single isotopic standard was used to monitor both concentration dependence and drift, 1-minute averages of calibration data are normalized to the 2.5–3.5 mmol mol<sup>-1</sup> range weekly. Shrinking this range or shifting it to higher volume mixing ratios does not change the qualitative features of the results presented.

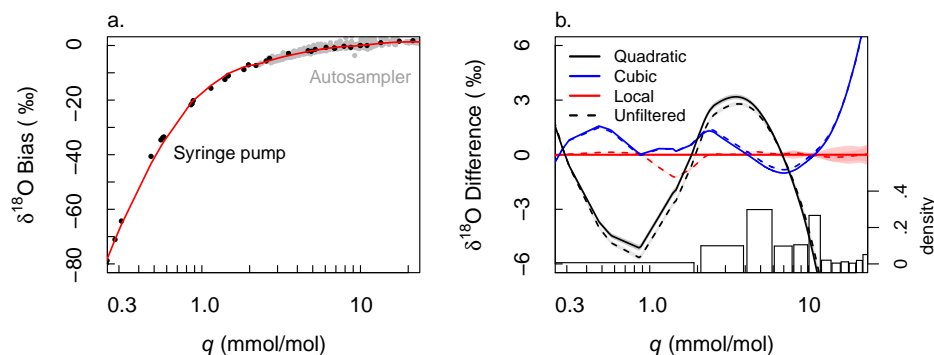
At both sites, concentration-dependent biases are characterized as a function of the natural logarithm of the water vapor volume mixing ratio, which ensures that fractional changes in  $q$  at both high and low volume mixing ratios are equally represented in the calibration fit. Best-fit quadratic polynomial (cf. Rambo et al., 2011), cubic polynomial (cf. Aemisegger et al., 2012; Noone et al., 2013), and nonparametric functions (cf. Bailey et al., 2013) are evaluated by least squares estimation. Nonparametric characterizations of the concentration dependence are derived by fitting a locally weighted polynomial regression with R’s “locfit” package (Loader, 1999). The local regression fits polynomial functions to subsets of the data in order to predict the best value for a given calibration point. An important advantage of this approach is that it ensures that variations in the isotope ratio across small subsets of the larger ambient  $q$  range are well calibrated. A bisquare kernel weights neighboring ob-

servations within the fitting window by their proximity to the prediction location. The degree of the polynomial and the smoothing parameter (i.e. the fraction of nearest neighbors included in the fitting window) are selected by minimizing the generalized cross validation score. One- and two-degree polynomials and smoothing parameters ranging from 0.50 to 1.00, every 0.05, are evaluated. All predictor values are scaled before fitting.

To account for the smaller number of low- $q$  calibration points at the Hawaii site, all functions associated with the Mauna Loa analyzer weight the predictor values by  $1/q^2$ . Although this weighting is strictly arbitrary, it is motivated by the fact that the isotope ratio is approximately proportional to the heavy isotopologue concentration divided by the water vapor concentration. It is therefore equivalent to weighting by  $1/q$  to predict the conserved quantity of the heavy water vapor concentration. Weighting the predictors by  $1/q$  or  $1/\ln(q)$ , in comparison, results in a poor fit at the lowest volume mixing ratios. No weighting is performed for Spiny or Gulper since the Summit calibration data are more evenly distributed across the humidity range of interest for Greenland ambient conditions. Standard errors associated with the fitted values (here referred to as “prediction errors”) are used to evaluate uncertainties in the curve fitting and to identify statistically significant variations in the concentration-dependence characterizations with isotope ratio and time.

Before evaluating measurement repeatability, the calibration data are first corrected for concentration dependence. Distinct analytical procedures are used due to differences in the data available at each site. At Mauna Loa, a locally weighted polynomial regression in two dimensions (i.e. a surface) is fit to the total isotopic bias, using both the natural logarithm of the water vapor volume mixing ratio and the isotope ratio measured as predictors. In this manner, normalization to VSMOW-SLAP is performed simultaneously with the concentration correction. As previously stated, the nonparametric regression is weighted by  $1/q^2$  to give larger consideration to the infrequent lower humidity measurements. The local regression predictions are then subtracted from the autosampler isotope ratios and the residual biases examined for long-term drift.

The simultaneous correction is advantageous for Hawaii since it maximizes all of the calibration information without requiring that the isotopic data be normalized to a reference humidity level – a challenge at this site since the autosampler does not produce consistent volume mixing ratios. (Note that it would also be advantageous were the concentration dependence sensitive to isotope ratio, since prediction errors associated with correcting the bias and normalizing the data to VSMOW-SLAP would be estimated jointly and double-counting of correlated systematic error avoided.) The approach does not, however, guarantee that the normalization to VSMOW-SLAP is perfectly linear (cf. Hut, 1987; IAEA, 2009), although it is nearly so for this data set. Moreover,



**Figure 1.** (a) The concentration dependence of the Mauna Loa analyzer shown with both the autosampler (gray) and syringe-pump (black) calibration points normalized to a reference humidity (see Sect. 2.4). The red line is the locally weighted polynomial regression used as the reference curve in panel (b). (b) Curve-fitting and hysteresis effects on the Mauna Loa concentration-dependence characterization. Lines represent the calibration differences that would result if fitting a quadratic (black) or cubic (blue) polynomial instead of a locally weighted polynomial regression (red). All are shown as a function of  $q$  – the water vapor volume mixing ratio – on a log scale. Curves fit to all data (unfiltered for hysteresis) are represented by dashed lines. Curves fit to filtered data are represented by solid lines. Prediction errors are represented by shaded envelopes around each curve. The calibration data density is depicted by bars (rightmost ordinate).

future deployment programs will need to consider the trade-off between performing calibrations across a wide  $q$  range using multiple isotopic standards and maximizing ambient measurement time.

In comparison, the concentration-dependent biases of the Summit analyzers are corrected using a one-dimensional locally weighted polynomial regression fit to the extended concentration–calibration data and the 6-hourly data measured at three volume mixing ratios. Only the natural logarithm of the water vapor volume mixing ratio is used as a predictor, since the DPG calibrations consist of a single isotopic standard. Measurement repeatability is evaluated by assessing linear drift in the concentration–corrected 6-hourly calibrations made at volume mixing ratios close to  $3 \text{ mmol mol}^{-1}$ .

### 3 Results and discussion

In this section, the long-term stability of the water vapor isotopic analyzers is examined in two parts. First, the concentration dependence of each analyzer is characterized, and uncertainties associated with curve-fitting procedures and measurement hysteresis are evaluated. Variations in the concentration dependence with isotope ratio and with time are analyzed. Second, measurement repeatability is evaluated by examining linear drift in the concentration–corrected calibration data. Results are discussed with respect to the distinct calibration systems used at each site.

#### 3.1 Concentration dependence

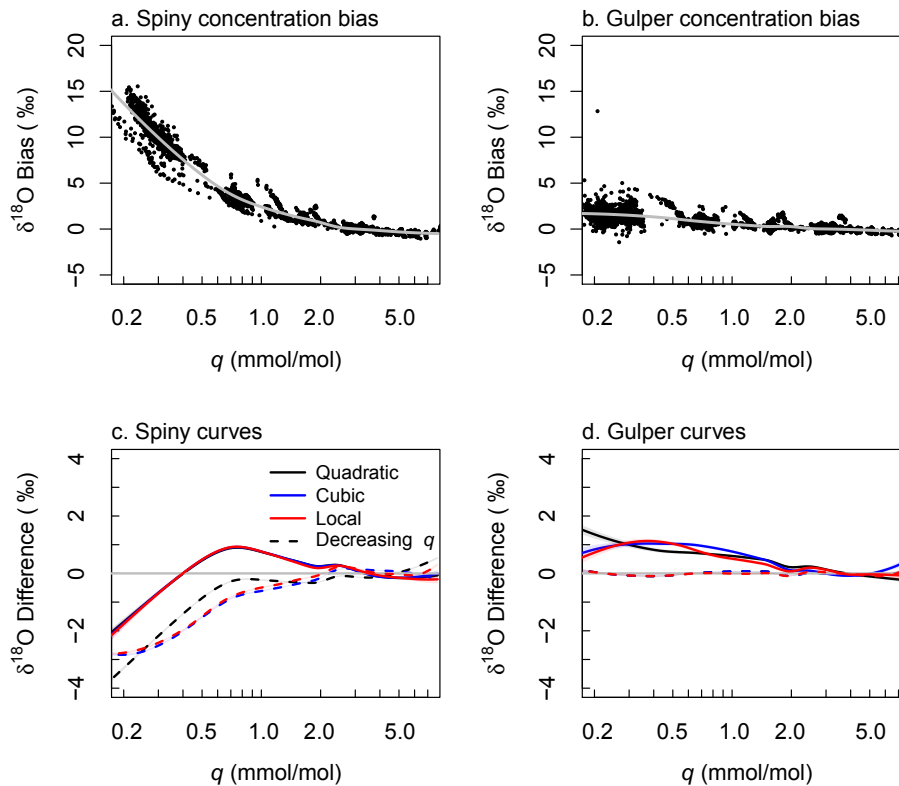
Characterizing the concentration dependence is a key step in correcting the isotopic measurements made by commercial laser analyzer, particularly for older instruments, like the one

in use at Mauna Loa, for which concentration dependence is the dominant isotopic bias. This subsection considers the importance of statistical curve-fitting procedures and sampling hysteresis in modifying the accuracy of the concentration-dependence characterization. Assumptions about the stability of the concentration dependence with isotope ratio and with time are also tested.

##### 3.1.1 Curve fitting and hysteresis

To evaluate uncertainties in the concentration-dependence characterization introduced by curve fitting and sampling hysteresis, two subsets of the Mauna Loa calibration data are considered. The first includes *filtered* autosampler injection points, where the first two injections of each standard are eliminated during every calibration period in order to reduce memory effects (cf. Penna et al., 2010). The second or *unfiltered* subset includes all autosampler injections. Both subsets include the syringe-pump data. Although it is unlikely that this filtering procedure eliminates all memory effects associated with the autosampler, due to the fact that the injections are asymptotic, it should be nontrivial since the first injections following a change in liquid standard are typically farthest in value from the final measured isotope ratio (cf. Lis et al., 2008; Gröning, 2011; Penna et al., 2012; van Geldern and Barth, 2012). While discarding more injections can help reduce memory effects further, it reduces the calibration sample size, which can also affect the accuracy of the bias characterization. Figure 1 shows the difference in the isotope ratio adjustment that would result if fitting the filtered data (solid lines) or the unfiltered data (dashed lines) with a quadratic polynomial (black), a cubic polynomial (blue), or a locally weighted polynomial regression (red). Clearly, the choice of characterization function is much more important in determining the isotopic correction for the large concentration





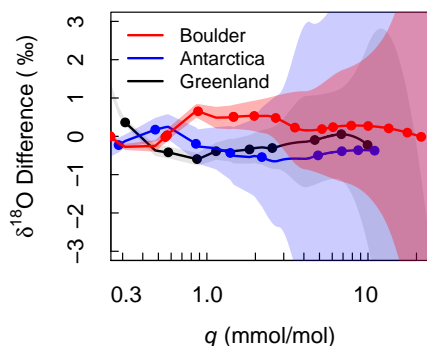
**Figure 2.** (a–b) The concentration dependences of the Summit analyzers shown normalized to a reference humidity (see Sect. 2.4). The gray lines are the locally weighted polynomial regressions used as reference curves in panels (c–d). (c–d) Curve-fitting and hysteresis effects on the Summit concentration-dependence characterizations. Lines represent the isotopic differences that would result if fitting data from the extended concentration-dependence calibrations in which the water vapor volume mixing ratio progressively increased (solid lines) or decreased (dashed lines), compared with the concentration dependence obtained from all Summit concentration–calibration data (the zero line is shown in gray for clarity). Characterizations are shown as a function of  $q$  – the water vapor volume mixing ratio – on a log scale for the various fitting procedures described in the legend and text. Prediction errors are represented by shaded envelopes around each curve.

dependence at Mauna Loa than filtering of the autosampler data. While the typical difference between the filtered and unfiltered data is less than a few tenths ‰, the average difference between the cubic polynomial and the local regression is greater than 1 ‰. Root mean square errors (RMSEs) suggest the parametric functions are insufficient for properly capturing the true curvature of the concentration dependence at Mauna Loa (RMSE: quadratic, 7.75 ‰; cubic, 2.36 ‰; local regression, 0.35 ‰).

In contrast, at Summit, where the Spiny and Gulper analyzers are calibrated with the DPG, memory effects are much more critical in influencing the concentration-dependence correction. Figure 2, like Fig. 1, shows differences in the bias characterization that result from applying three functions: a global quadratic polynomial, a global cubic polynomial, and a locally weighted polynomial regression. Hysteresis in the Summit system is evaluated by considering the different characterizations that result when isolating the extended concentration–calibration periods (see Sect. 2.3) in which  $q$  progressively increased (solid lines, Fig. 2) or decreased (dashed lines, Fig. 2) from the rest of the concentration

calibrations. Unlike data from the extended concentration–calibration periods, data from the 6-hourly calibrations are already filtered for memory effects, since the first few minutes spent sampling at each volume mixing ratio are excluded from the analysis (see Sect. 2.3).

Despite being very different in shape and magnitude, the concentration dependences of Spiny and Gulper are both sensitive to whether  $q$  is progressively increased or decreased. Differences in the concentration-dependence characterization resulting from curve fitting are, in contrast, negligible. This may be due to the fact that each curve in Fig. 2 is comprised of at least 68 1-minute averages, and these data points are well distributed across the humidity range of interest. Differences between the solid and dashed lines, in comparison, suggest memory effects can be quite substantial for the DPG system. Indeed, the tubing connecting the DPG to the instrument amplifies the surface area that the reference vapor contacts. As a result, equilibration may be slow and longer sampling at each water vapor concentration required. Differences in calibration system and approach clearly affect the relative contributions of curve fitting and hysteresis to uncertainties

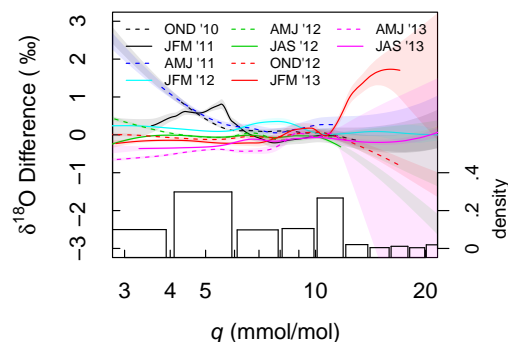


**Figure 3.** Mauna Loa concentration-dependence characterizations for three different isotopic standards named in the legend. Curves – fit with a local regression – are shown as differences from the concentration dependence obtained when data from all three standards are considered and are depicted as a function of  $q$  – the water vapor volume mixing ratio – on a log scale. Solid circles represent the  $q$  values at which calibration points were measured. Prediction errors are represented by shading.

associated with the concentration-dependence characterization.

### 3.1.2 Sensitivity to isotope ratio

To test whether the concentration dependence changes with isotope ratio, data from the Mauna Loa syringe-pump calibrations – normalized to the 9–11 mmol mol<sup>-1</sup> volume mixing ratio range – are considered. Figure 3 shows the difference in characterization that would be obtained from a single isotopic standard versus the characterization obtained from all the syringe-pump data. All curves are fit using a locally weighted polynomial regression. While there are clear differences between curves, there is no evidence of a monotonic shift in concentration dependence with isotope ratio. This is statistically supported by the fact that for any given water vapor volume mixing ratio, at least two curves overlap within the prediction error envelopes shown. Bastrikov et al. (2014) found similar results when using commercial dry air to perform concentration calibrations. Moreover, between 0.3 and 8 mmol mol<sup>-1</sup>, the largest absolute difference between any individual curve and the curve obtained when using all the syringe-pump data is approximately 0.6‰. In comparison, the average difference between the local and cubic fits shown in Fig. 1 is larger. This finding again highlights the fact that great care is needed when fitting the concentration dependence in situations where calibration data at the low end of the humidity range are infrequent. In such cases, variations in the concentration-dependence characterization are more likely to be caused by limitations in statistical fitting than by isotopic differences in the standard.

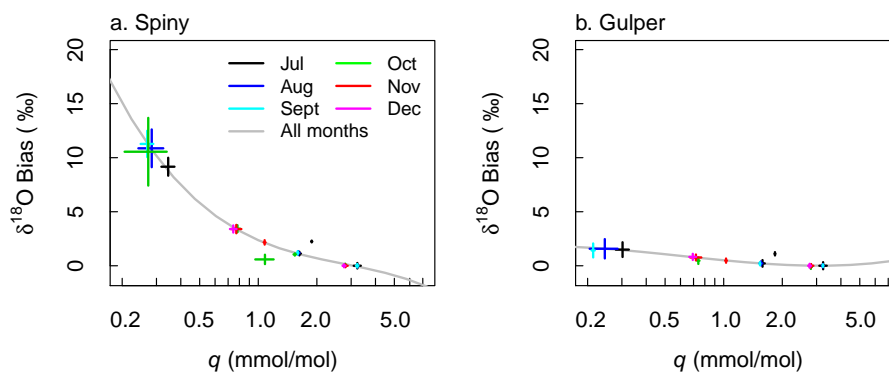


**Figure 4.** Mauna Loa concentration-dependence characterizations for 3-month periods described in the legend. Only periods where  $n \geq 30$  are shown. Curves – fit with a local regression – are depicted as a difference from the concentration dependence obtained when using all 3 years of data and are plotted as a function of  $q$  – the water vapor volume mixing ratio – on a log scale. Prediction errors are represented by shading. The calibration data density is depicted by bars (rightmost ordinate).

### 3.1.3 Variations with time

To identify possible drift in the concentration dependence with time, the concentration dependence at Mauna Loa is characterized every 3 months between October 2010 and September 2013. Unfiltered autosampler calibrations are used in order to increase the sample size within each period; the data are normalized to a reference humidity, as described in Sect. 2.4. All curves are fit with a locally weighted polynomial regression and are displayed as a difference from the characterization obtained when all 3 years of data are considered (Fig. 4). Although differences between characterizations are apparent, there is no evidence of a long-term directional trend. Instead, there is variability on the order of  $\pm 0.5$ ‰ for  $q$  values less than 10 mmol mol<sup>-1</sup> – similar to reports by Steen-Larsen et al. (2014b) of variability in the concentration dependence on the order of several tenths ‰ between 10 and 30 mmol mol<sup>-1</sup>. The prediction error envelopes suggest this variability is only statistically significant between approximately 4 and 6 mmol mol<sup>-1</sup>. Meanwhile, sparse data coverage accounts for periods in which the absolute  $\delta^{18}\text{O}$  difference exceeds 1.0‰ and for large prediction errors near the  $q$ -range extremes. When all 3 years of calibrations are considered, increasing the sample size, the prediction error drops below 0.2‰ for the entire  $q$  range.

The 6-hourly calibration data at Summit, which span much lower volume mixing ratios, also show no trend in the concentration dependence over a 6-month period. Because of the design of the Summit calibration protocol (e.g. a large number of observations at a small number of volume mixing ratio levels), the data in a given month do not span a sufficient  $q$  range to characterize the concentration dependence meaningfully. To address this challenge, Fig. 5 considers instead how well a single characterization – derived from all 6 months of



**Figure 5.** Monthly mean concentration-dependent biases in  $\delta^{18}\text{O}$  (‰) for the Summit instruments shown as crosses whose width and height represent  $1\sigma$  (standard deviations) in  $\delta^{18}\text{O}$  and  $q$ . The concentration-dependence characterizations obtained when fitting a locally weighted polynomial regression to all 6 months of data are represented by gray lines.

1-minute-averaged data (where  $q > 0.15 \text{ mmol mol}^{-1}$ ) – fits monthly mean calibration values. This “all data” characterization – fit with a locally weighted polynomial regression – is shown as a gray line for both Spiny and Gulper. The colored crosses depict calibration data that are clustered by  $q$  and averaged in 30-day segments, approximately corresponding to the months shown in the legend. The width and height of the crosses show the  $1\sigma$  (standard deviations) of the water vapor volume mixing ratios and isotope ratios, respectively. While there is clear variability within monthly means, the “all data” curves pass within  $1\sigma$  of nearly every cluster. Importantly, there are no instances in which more than one cluster per month lies farther than  $1\sigma$  away from the gray line. Thus, despite the fact that the Summit instruments exhibit substantially different biases, there is no statistical evidence for long-term drift in the concentration dependences of either Spiny or Gulper. These results suggest that concentration calibrations need not be performed frequently at Mauna Loa and Summit to resolve long-term directional drift; however, they may nevertheless be useful for characterizing the concentration dependence precisely, particularly at low water vapor concentrations where the signal-to-noise ratio of the analyzers is very small.

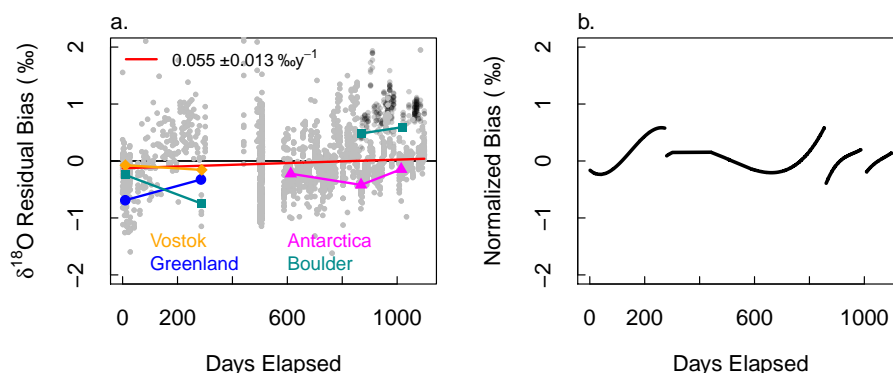
### 3.2 Long-term measurement repeatability

While the previous section suggests the concentration dependence may be considered stable with time for measurement calibration purposes, this section considers whether instrumental drift causes long-term directional changes in measurement repeatability and influences isotopic deviations from the VSMOW-SLAP scale. The analysis pays particular attention to the differences in calibration systems at Mauna Loa and Summit. Measurement repeatability is evaluated at Mauna Loa by examining trends in the residual bias of the autosampler data once corrected for concentration dependence and normalized to VSMOW-SLAP, as described in Sect. 2.4. While at first glance Fig. 6 suggests the residual

isotopic bias does increase with time, closer inspection suggests this apparent multi-year trend is the result of changes to the standards used to calibrate the instrument. The colored symbols in Fig. 6a represent the mean residual biases, associated with each standard, calculated for 20-day periods following dates on which the standards were replaced with new water. (Note that the standards have been used for different periods throughout the measurement campaign.) While variability from one standard-replacement date to the next is discernible, there is no clear evidence of directional drift. The jump in the bias measured with Boulder water between days 279 and 860 coincides with the period in which the standard was replaced with new water that was more than 1.5 ‰ more depleted than the original. (For this reason, the symbols are not connected in the figure.)

To evaluate the effect of changes in the Boulder standard on the appearance of long-term directional drift at Mauna Loa, a simple linear regression is fit to all of the residuals and another to all residuals except those associated with the new Boulder calibration points (darker shading, Fig. 6a). The first regression suggests a linear drift of  $0.162 \text{ ‰ year}^{-1}$  (not shown); however, the second suggests a drift of only  $0.055 \text{ ‰ year}^{-1}$  (Fig. 6a, red line). The latter would have caused a total increase in the isotopic bias remaining after calibration of  $0.166 \text{ ‰}$  over the course of the 3-year campaign – a change smaller than the mean uncertainty associated with the locally weighted polynomial regression used to calibrate the data. While drift of this magnitude would be significant over many years, it is negligible relative to the atmospheric variability (i.e. the signal) for the 3-year field deployment at Mauna Loa.

Figure 6b provides additional evidence that changes in the calibration standards create the appearance of long-term drift at Mauna Loa. The figure depicts cubic polynomials that have been fit to the residuals, after the residuals are normalized by standard for the four periods in which different bottles of standard water were used. During every period, there is a consistent increase in the residual bias with time, sug-



**Figure 6.** (a) The residual isotopic bias at Mauna Loa – after correcting the calibration data for concentration dependence and normalizing to the VSMOW-SLAP scale – plotted as a function of time (gray points). Long-term drift is estimated by the simple linear regression (red line), which is fit to all data except those associated with the second set of Boulder standard water (points with darker shading). Mean residual biases for 20-day periods following the replacement of each standard with new water are represented by the larger symbols and are colored by standard (see legend). (b) Cubic polynomials fit to gray points in panel (a) – once normalized by isotopic standard for the four periods in which different standard bottles were used – indicate that the standards themselves may be drifting at Mauna Loa.

gesting that the standards themselves have drifted in the field. This may be due to the fact that each week the Mauna Loa instrument is calibrated by pipetting standard water from the glass bottle in which it is stored into small vials, which are loaded onto the autosampler. Over time, weekly opening and closing of the standard bottles allows isotopic exchange between the water reservoir and the ambient vapor. Though this unfortunately impedes characterization of shorter-term measurement repeatability, it importantly showcases the susceptibility of standards to drift in the field if not stored in airtight containers. Long-term campaigns wishing to characterize weekly-to-seasonal variability in measurement repeatability will thus need to replace standards sent to the field every few weeks or use an on-site, airtight storage system for standards such as that described by Tanweer et al. (2009).

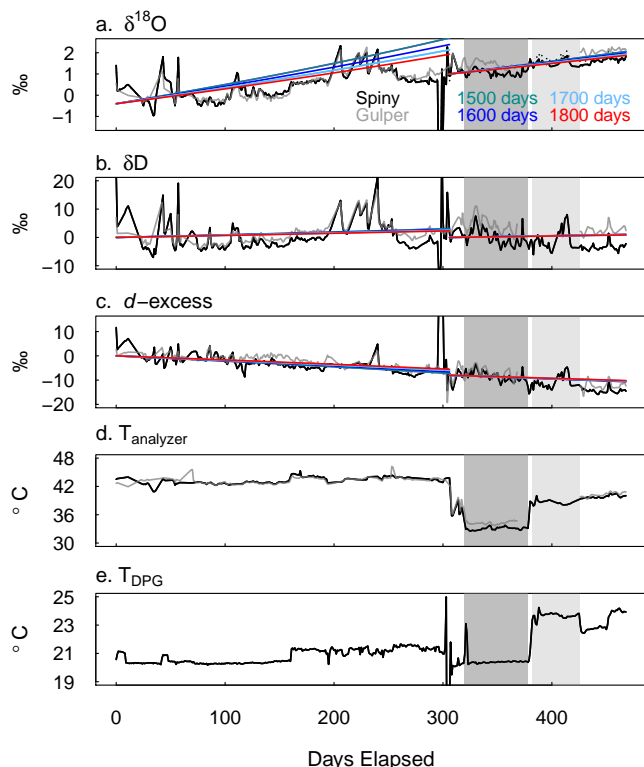
At Summit, in comparison, the 6-hourly DPG calibrations – once corrected for concentration and normalized to the July 2013 isotopic mean – show an expected enrichment in  $\delta^{18}\text{O}$  of  $1.4 \text{ ‰ year}^{-1}$  (Fig. 7a). As water is continuously vaporized and removed from the DPG bottle, the isotopic composition of the remaining liquid in the reservoir should become more enriched, following a theoretical Rayleigh distillation (Dansgaard, 1964). Moreover, as a result of fractionation differences, the  $\delta^{18}\text{O}$  of the remaining liquid should increase faster than  $\delta\text{D}$  (Fig. 7b), resulting in a decrease in the deuterium excess ( $d = \delta\text{D} - 8 \times \delta^{18}\text{O}$ ) (Fig. 7c) (Craig, 1961; Sharp, 2007).

The relevance of the Rayleigh model for describing changes in measurement repeatability at Summit is evaluated by comparing the predicted drift in  $\delta^{18}\text{O}$  and  $d$  from Eq. (1) with the Spiny observations. However, since  $\tau$ , the time necessary to evaporate all liquid, is not precisely known for the DPG, a range of possible values is considered for the period following July 2013 – when the DPG bottle was last refilled – (i.e. the last 150 days of the experiment). These estimates are

then compared with the slopes of isotopic change observed, which – for such a short period – are well described by linear models.  $\tau$ 's spanning 1500–1800 days (or approximately 4–5 years) produce the smallest differences in slope between model and observations (e.g.  $< 0.04 \text{ ‰ year}^{-1}$  in  $\delta^{18}\text{O}$ ). This range of  $\tau$  values is then used to extrapolate the distillation that would have occurred between 421 and 727 days of continuous operation of the DPG, which corresponds to the first year of calibration points plotted. Normalizing and plotting these extrapolated curves suggests the 1800-day distillation model best reproduces the observations and explains the directional drift in both isotope ratios and in deuterium excess (Fig. 7a–c).

Assuming the linear enrichment in  $\delta^{18}\text{O}$  in Spiny and Gulper is entirely explained by distillation of the calibration system, uncertainty due to shorter-term variability (i.e. measurement repeatability errors) can be estimated once the linear trend is removed. For the period following July 2013 the RMSE associated with characterizing the long-term linear drift is  $0.21 \text{ ‰}$  for Spiny and  $0.18 \text{ ‰}$  for Gulper. For the period before July 2013 the RMSE is  $0.71 \text{ ‰}$  for Spiny and  $0.43 \text{ ‰}$  for Gulper. Importantly, the fact that both analyzers record much of the same variability suggests a source either in the surrounding environment or inherent to the DPG system.

One possible cause of variability in measurement repeatability suggested by previous studies is ambient temperature fluctuations (cf. Sturm and Knohl, 2010; Rambo et al., 2011); however, the data presented here do not support this claim. During the summer of 2013 (represented by dark gray shading in Fig. 7), the temperatures of the Summit analyzers plummeted following a change in the temperature of the underground chamber in which they were housed (Fig. 7d). No concurrent change in either isotope ratio occurred.



**Figure 7.** The (a)  $\delta^{18}\text{O}$ , (b)  $\delta\text{D}$ , and (c) deuterium excess of the Summit (Spiny in black, Gulper in gray) concentration-corrected calibration points, normalized to the July 2013 mean and plotted as a function of time. In panel (a) Spiny values are shown corrected (solid) and uncorrected (dotted) for DPG temperature fluctuations. Colored lines represent distillation models where  $\tau$ , the total time required to remove all liquid from the DPG bottle, equals 1500 (green), 1600 (blue), 1700 (light blue), or 1800 (red) days. Variations in the temperatures of the (d) Summit analyzers and (e) DPG bottle are also shown as a function of time. The dark (light) gray shaded regions indicate periods in which the analyzer (DPG) temperature changed significantly.

In comparison, changes in the temperature of the DPG do appear to influence the isotopic calibrations. During the fall of 2013 (the period represented by light gray shading in Fig. 7), the DPG bottle temperature increased from 20 to 24 °C (Fig. 7e). A change of this magnitude should have altered the fractionation in the DPG, resulting in an increase in  $\delta^{18}\text{O}$  of the water vapor produced of about 0.25 ‰. Such enrichment is evident in the temperature-uncorrected data shown in Fig. 7a (dotted line). Yet this factor alone cannot explain all of the isotopic variations around the linear trend, once again pointing to the possibility that the vapor stream of the DPG is not entirely non-biasing (cf. Kurita et al., 2012). Clearly, long-term deployments choosing a dew point generator for a calibration system should monitor both the system's temperature and distillation over the course of the campaign. And, the calibration approach needs to be flexible enough to allow for inaccuracies in continuous vapor gener-

ation that will influence the total error associated with calibrating ambient measurements.

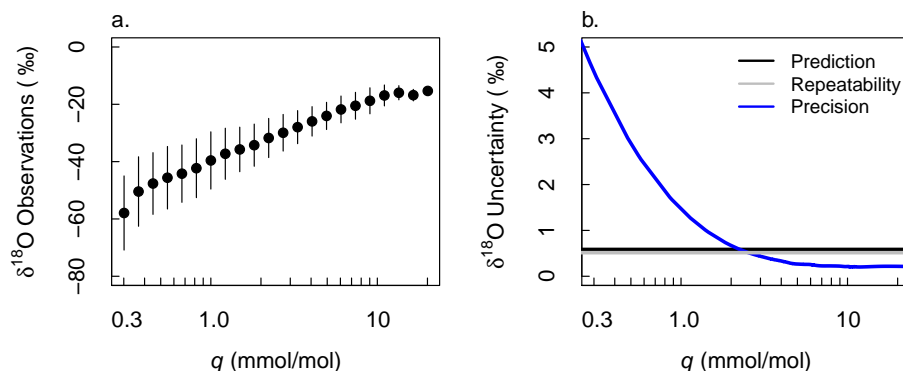
## 4 Uncertainty estimation

Having examined two key biases in isotopic measurements made with spectroscopic analyzers, we now consider how the uncertainties associated with characterizing and correcting these biases influence the signal-to-noise ratio at each site and discuss how the total error may be effectively reduced given the calibration strategies described. Measurement uncertainty at both Mauna Loa and Summit primarily stems from three sources: (1) prediction errors associated with characterizing the isotopic biases of the field data, (2) variability in measurement repeatability that is not captured by linear drift estimates, and (3) instrumental precision. At Mauna Loa, instrumental precision is estimated using the standard deviations associated with each autosampler injection, which includes approximately 2 minutes of water vapor flow through the instrument's optical cavity. At Summit, instrumental precision is estimated using the standard deviations associated with the 1-minute-averaged calibration points. Keep in mind that the actual error contribution by precision to a measurement depends on the measurement averaging time (e.g. the sample size).

### 4.1 Mauna Loa, Hawaii

Ambient isotopic measurements are calibrated at Mauna Loa using the local surface regression described in Sect. 2.4, fit to the syringe-pump and filtered (see Sect. 3.1.1) autosampler calibration data. Measurements are sufficiently stable so that no long-term drift correction is made. Mass-weighted averages of the calibrated field data are shown as a function of  $q$  for the period October 2010–September 2013 in Fig. 8a. Even for  $q$  values exceeding 2 mmol mol<sup>-1</sup>, the mean isotope ratios range from -31.7 to -15.3 ‰, a 16 ‰ span, which is not atypical during the diurnal growth and collapse of the convective boundary layer at MLO (Bailey et al., 2015). Standard deviations are greater than 6.0 ‰ at the lower end of this range and near 2.0 ‰ at the high end.

Known measurement uncertainties at Mauna Loa are shown beside the ambient data in Fig. 8b. Because prediction errors associated with the nonparametric calibration approach are not a monotonic function of  $q$ , the 99th percentile of prediction error (0.58 ‰) is used as a conservative estimate of uncertainty associated with the isotopic bias characterization (black line, Fig. 8b). Meanwhile, uncertainty related to measurement repeatability (gray line, Fig. 8b) is estimated from the RMSE associated with the linear regression shown in red in Fig. 6a: 0.51 ‰. In effect, this estimate captures the variability in the calibration data that is unaccounted for when applying a linear drift correction to the time series. It would increase by only 0.04 ‰ were the second



**Figure 8.** (a) The calibrated Mauna Loa  $\delta^{18}\text{O}$  ambient data averaged as a function of the water vapor volume mixing ratio ( $q$ ), with  $1\sigma$  deviations given by vertical lines. (b) Sources of uncertainty in the  $\delta^{18}\text{O}$  ambient data include the prediction errors from correcting the total isotopic bias with a locally weighted polynomial regression (black), measurement repeatability errors (gray), which are estimated from the root mean square error associated with fitting a linear regression to the corrected calibration data, and instrument precision (blue), estimated from the standard deviations of the autosampler injections.

set of Boulder calibration points (darker shading, Fig. 6a) also included in the long-term drift analysis. While these two sources of uncertainty dominate the total error of measurements made at water vapor volume mixing ratios greater than  $2.2\text{ mmol mol}^{-1}$ , the precision of the autosampler and isotopic analyzer assembly could potentially contribute large errors at lower  $q$  (blue line, Fig. 8b). The standard deviations of the autosampler injections shown in Fig. 8 are similar to values reported by Galewsky et al. (2011) for the same model Picarro analyzer. Note, however, that the actual error associated with instrumental precision will depend on the measurement averaging time.

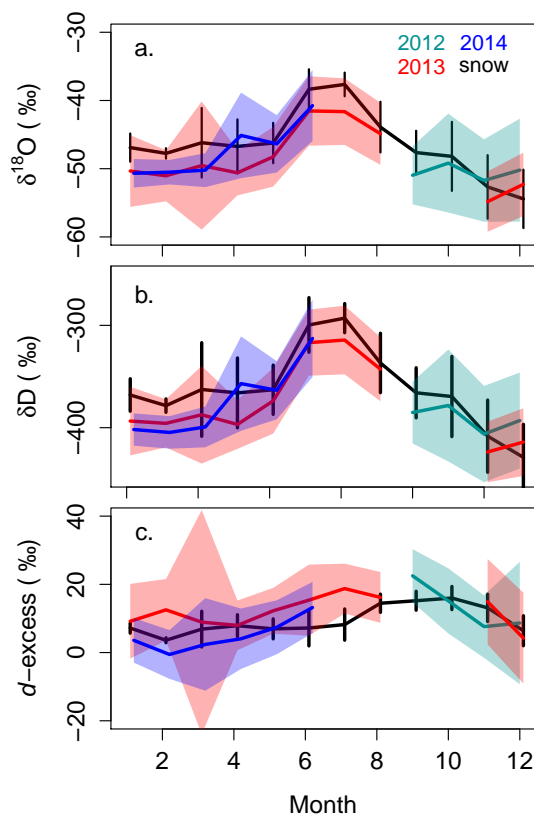
Summing the prediction and repeatability errors in quadrature, the error (on individual measurements) associated with characterizing the isotopic biases of the Mauna Loa analyzer is  $0.8\%$ . This combined error is much smaller than the range of  $\delta^{18}\text{O}$  observed at Mauna Loa Observatory. And, for  $q < 20\text{ mmol mol}^{-1}$  (where 99.99% of the ambient data lie), it is less than half the standard deviation of the ambient isotope ratio at a given water vapor concentration (Fig. 8a). Since humidity at Mauna Loa is largely a function of whether the observatory is exposed to boundary layer or free tropospheric air, these findings suggest diurnal variations in isotope ratio would be clearly discernible, as would variations associated with synoptic variability, which influences day-to-day changes in boundary layer and free tropospheric air. While differences of about twice the total error (e.g. the sum of prediction, repeatability, and precision errors in quadrature) might be desirable to distinguish any paired field observations statistically, smaller isotopic changes could be detectable if a long-term linear trend is identified whose slope differs significantly from the linear trend in the calibration data ( $0.055\text{‰ year}^{-1}$ ). Measurement sensitivity at Mauna Loa would be greatly improved by increasing the number of calibration points at low  $q$  (to reduce prediction errors) and by monitoring changes in measurement repeatability more frequently.

Optimizing the measurement averaging time could also bolster the signal-to-noise ratio by improving the precision at low  $q$  (Aemisegger et al., 2012).

#### 4.2 Summit, Greenland

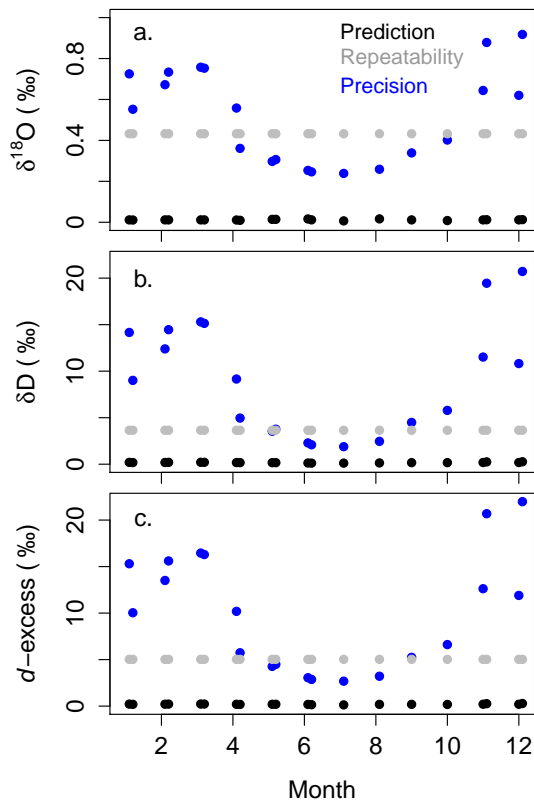
Ambient isotopic data at Summit are calibrated sequentially, first by correcting the concentration dependence, using a locally weighted polynomial regression, fit to the extended and 6-hourly concentration–calibration points, and second by normalizing the concentration-corrected field data to VSMOW-SLAP by applying a scale and offset derived from fitting a linear model to the syringe-pump calibrations described in Sect. 2.3 (cf. Hut, 1987; IAEA, 2009). No long-term drift correction is made. The calibrated Gulper field data from the 40 m inlet on the instrument tower at Summit are shown in Fig. 9 as monthly averages (for months with at least 10 days of data) spanning mid-2012 to mid-2014. While there is some variability from year to year, particularly near boreal spring and fall, there is a clear annual cycle in both  $\delta^{18}\text{O}$  and  $\delta\text{D}$ , with ranges of about 14 and 111‰, respectively. The standard deviation shading gives an indication of the synoptic variability within each month and ranges from 1.8 to 9.4‰ in  $\delta^{18}\text{O}$  and 15.8 to 54.7‰ in  $\delta\text{D}$ , with means of 4.8 and 33.2‰, respectively. Isotope ratios from surface snow samples collected in 2013, which were analyzed at the University of Colorado (A. Kaushik, personal communication, 2014) and are plotted in black for comparison, suggest a similar annual signal and similar monthly variability.

Uncertainties associated with the Gulper measurements are shown in Fig. 10, with  $d$  uncertainties estimated by summing  $\delta\text{D}$  and  $8 \times \delta^{18}\text{O}$  in quadrature. Prediction errors associated with characterizing the concentration dependence (black dots, Fig. 10) contribute negligibly to the total error, since the characterization is strongly constrained by the large number of calibration points at low  $q$ . Uncertainties in mea-



**Figure 9.** Gulper (Summit) mass-weighted monthly means of (a)  $\delta^{18}\text{O}$ , (b)  $\delta\text{D}$ , and (c) deuterium excess in water vapor for 2012 (green), 2013 (red), and 2014 (blue). All monthly means represent  $>20$  days of sampling except for the October 2012 and June 2013 means, which represent 16 and 11 days of sampling, respectively. Data are from the 40 m inlet on the instrument tower. Shading represents  $1\sigma$  deviations. Monthly mean Summit surface snow values from 2013 (black, with vertical lines representing standard deviations) are plotted for comparison and offset by (a)  $-10\%$  and (b)  $-80\%$ , as an approximation to equilibrium fractionation.

surement repeatability (gray dots, Fig. 10) are estimated from the RMSEs associated with fitting linear distillation trends to the concentration-corrected 6-hourly calibrations. (Recall that the DPG bottle was refilled in summer 2013; therefore the distillation curve is estimated independently for periods before and after). Only the larger RMSE for each isotope ratio is shown:  $0.43\%$  in  $\delta^{18}\text{O}$  and  $3.6\%$  in  $\delta\text{D}$ . These suggest a repeatability error of  $5.0\%$  for  $d$ . Annual and synoptic variability in both isotope ratios and synoptic variability in  $d$  are clearly large enough to be distinguished despite the size of these errors. Nevertheless, repeatability is the dominant source of error during half the year. While uncertainty associated with instrumental precision (blue dots, Fig. 10) is larger in winter, when ambient humidity is lowest, the actual error contribution could be lower than the repeatability error, depending on the measurement averaging time (e.g. the sample size). To detect isotopic changes associated



**Figure 10.** Contributions to measurement uncertainty at Summit include prediction errors from correcting the concentration dependence with a locally weighted polynomial regression (black), measurement repeatability errors (gray), which are estimated from the root mean square error associated with fitting a linear regression to the concentration-corrected calibration data, and instrument precision (blue), estimated from the standard deviations of the 1-minute calibration measurements made at volume mixing ratios matching the mean  $q$  for each month. These uncertainties are shown for (a)  $\delta^{18}\text{O}$ , (b)  $\delta\text{D}$ , and (c) deuterium excess for the Gulper analyzer.

with a  $\sim 1^\circ\text{C}$  change in air temperature – assuming temporal  $\delta^{18}\text{O}$ –temperature relationships in Greenland water vapor are approximately  $0.8\%$   $^\circ\text{C}^{-1}$  (Steen-Larsen et al., 2014a) – increasing the instrumental precision at low  $q$  and correcting changes in measurement repeatability on daily-to-weekly time scales would both be desirable. Better characterization of DPG variability is therefore important for improving measurement sensitivity at this site.

## 5 Implications for long-term deployments

There are a number of different factors that need to be considered when calibrating isotope ratio data, and many of these factors have been documented previously (e.g. Sturm and Knohl, 2010; Johnson et al., 2011; Tremoy et al., 2011; Aemisegger et al., 2012; Gröning, 2011; Noone et al., 2013; Steen-Larsen et al., 2013). However, only recently have data

sets been long enough to evaluate the long-term stability of laser isotopic analyzers in the field. This analysis has evaluated the stability of the isotopic biases in three field-operational water vapor isotopic analyzers. The instruments, which are deployed at the Mauna Loa Observatory in Hawaii and at Summit, Greenland, include both the early Picarro model L1115-i and the later L2120-i models, which have been used in previous studies to study atmospheric hydrological phenomena (e.g. Galewsky et al., 2011; Noone et al., 2011; Hurley et al., 2012; Bailey et al., 2013) and water exchange between the land surface and atmosphere (e.g. Berkelhammer et al., 2013; Noone et al., 2013; Aemisegger et al., 2014, Steen-Larsen et al., 2014a). Factory pre-calibration has helped reduce concentration-dependent biases in the later 2000 series (cf. Aemisegger et al., 2012), yet concentration calibrations are still necessary.

The results of this work suggest that while these types of laser analyzers exhibit some variability in the concentration dependence with time, there is no long-term directional drift. Moreover, it is very likely that smaller sample sizes at low  $q$  are partly responsible for the larger variability in the concentration dependence at Mauna Loa. Repeated concentration calibrations are thus more likely to improve statistical fitting of the bias rather than reveal long-term changes in its stability. Nevertheless, there is always the possibility that instrumental biases may drift as the optical components age (e.g. as laser power diminishes and/or the mirrors and lenses become dirty, particularly at sites exposed to heavy concentrations of pollutants) or if degradation of materials used in the inlet system (e.g. plastics) occurs.

In addition, there is no clear evidence of sensitivity in the concentration dependence to the isotope ratio of the standard used. Therefore, for field campaigns in which maximizing ambient sampling time is desirable, concentration-dependence characterizations should prioritize measuring a single standard at volume mixing ratios spanning the expected range of ambient humidity rather than using multiple standards. Maximizing the spread of the calibration points is particularly important for reducing uncertainties associated with statistically fitting the concentration-dependence curve.

How important statistical fitting is for the overall accuracy of the isotopic measurements depends, at least in part, on the calibration strategy. The Mauna Loa experiment showed that differences in the bias correction derived from fitting distinct functions to the autosampler data were much larger than any differences produced by filtering for memory effects. However, at Summit, where a custom dew point generator (DPG) was used and very low  $q$  values produced, the opposite was true. This result is likely influenced by the following two factors: (1) a reduced sensitivity to statistical fitting associated with the larger number of calibration points at Summit and (2) greater hysteresis associated with the DPG setup.

Both the Mauna Loa and Summit experiments suggest that, once the concentration dependence is measured, a single global characterization may be used to correct the ambient

data. Remaining isotopic biases represent deviations from the VSMOW-SLAP scale, and these may be corrected either simultaneously with the concentration dependence or sequentially by fitting a simple linear regression between the known values of three or more standards and the (concentration-corrected) isotope ratios observed (cf. Hut 1987; IAEA, 2009). Unfortunately, neither experiment presented here could determine conclusively whether the VSMOW-SLAP offset (i.e. measurement repeatability) changes directionally with time as a result of instrumental drift. At Mauna Loa, while calibration data appeared to change in between dates when the standard waters were replaced – suggesting the standards themselves drifted – there was no evidence of a linear trend from one standard-replacement date to the next. In comparison, at Summit, though the possibility of directional drift could not be ruled out, linear changes in the individual isotope ratios and deuterium excess of the calibration points were consistent with distillation of water in the DPG with time.

For field experiments using continual vapor generation to calibrate their analyzers, eliminating or accurately modeling drift in the standards used to calibrate the instruments is paramount. Modeling distillation in a DPG, for example, requires knowing the exact time to remove all water from the liquid reservoir. This, in turn, depends on the initial volume (or mass) of water and the flow rate of dry air through the liquid reservoir. In contrast, for calibrations with an autosampler, large amounts of secondary standards should either be stored on-site in an airtight container, such as described by Tanweer et al. (2009), or shipped to the site every 3–4 weeks to avoid excessive fractionation associated with opening standard bottles in the field. Such precautions will make it possible to evaluate more accurately whether changes in measurement repeatability are significant with time.

Advancements in commercial technology over the last few years suggest that the accuracy and precision of laser isotopic analyzers will continue to improve. Nevertheless, the biases examined in this paper remain critical to evaluate, particularly since they are unique to each instrument. To that end, the following recommendations are offered for long-term deployments.

The ideal calibration system should:

1. Enable very accurate evaluation of the concentration dependence by producing volume mixing ratios that span the ambient humidity range. For arid field sites, this requires producing very low and stable volume mixing ratios of water vapor for extended periods of time. This study found that a custom dew point generator could achieve volume mixing ratios  $<0.3 \text{ mmol mol}^{-1}$ , but that a vapor stream of at least 30 minutes was necessary to obtain a stable 10-minute average and minimize hysteresis. The autosampler, in contrast, was unable to produce low volume mixing ratios reliably.



2. Exhibit minimal or carefully controlled drift. While the former can be achieved by storing standards in specially designed airtight containers (Tanweer et al., 2009), the latter requires both that the liquid standard be temperature controlled and that distillation be carefully modeled and spot-checked.
3. Allow multiple standards to be delivered to the instrument. At least three standards are recommended for sequential calibration approaches, in which normalization to the VSMOW-SLAP reference scale is performed after the concentration dependence is corrected. More standards will be necessary to reduce prediction errors associated with correcting the biases simultaneously, such as with a local surface regression. While an autosampler can easily accommodate any number of standards through simple swapping of vials, a DPG would require different water reservoirs for each standard or for quantitative blending.

The ideal calibration approach should:

1. Characterize the concentration dependence by sampling a single isotopic standard at a large number of water vapor concentrations. Sampling at low volume mixing ratios (e.g.  $q < 1 \text{ mmol mol}^{-1}$ ) should be conducted for sufficient periods of time in order to bolster the signal-to-noise ratio. While frequently repeated characterizations of the concentration dependence are not necessary for verifying the stability of this bias, field programs may nevertheless wish to recheck their characterizations every few months, particularly if operating in polluted environments.
2. Choose a statistical-fitting procedure that minimizes the magnitude of fitting biases introduced into the data.
3. Consider whether simultaneously correcting the concentration dependence and normalizing to VSMOW-SLAP improves data accuracy.

*Acknowledgements.* Isotopic water vapor measurements at Mauna Loa Observatory would not have been possible without the dedication of NOAA-ESRL's J. Barnes and R. Schnell and instrument support from the NASA Tropospheric Emission Spectrometer team. The authors are indebted to C. Cox, M. O'Neill, D. Schneider, and the Summit technicians for assistance in maintaining water isotopic measurements in the challenging environment of the Greenland Environmental Observatory. Thanks are also due to A. Kaushik for analyzing the isotopic composition of the 2013 Summit snow samples at the University of Colorado with support from NSF CAREER Award AGS-0955841 and 1539234. Funding for the Greenland measurements was provided by NSF Award PLR-1023574. In addition, A. Bailey has been supported by a Ford Foundation Dissertation Fellowship and a NASA Earth and Space Science Fellowship. Comments by anonymous reviewers greatly

improved the quality and clarity of the manuscript.

Edited by: T. von Clarmann

## References

- Aemisegger, F., Sturm, P., Graf, P., Sodemann, H., Pfahl, S., Knohl, A., and Wernli, H.: Measuring variations of  $\delta^{18}\text{O}$  and  $\delta^2\text{H}$  in atmospheric water vapour using two commercial laser-based spectrometers: an instrument characterisation study, *Atmos. Meas. Tech.*, 5, 1491–1511, doi:10.5194/amt-5-1491-2012, 2012.
- Aemisegger, F., Pfahl, S., Sodemann, H., Lehner, I., Seneviratne, S. I., and Wernli, H.: Deuterium excess as a proxy for continental moisture recycling and plant transpiration, *Atmos. Chem. Phys.*, 14, 4029–4054, doi:10.5194/acp-14-4029-2014, 2014.
- Bailey, A., Toohey, D., and Noone, D.: Characterizing moisture exchange between the Hawaiian convective boundary layer and free troposphere using stable isotopes in water, *J. Geophys. Res.*, 118, 8208–8221, doi:10.1002/jgrd.50639, 2013.
- Bailey, A., Nusbaumer, J., and Noone, D.: Precipitation efficiency derived from isotope ratios in water vapor distinguishes dynamical and microphysical influences on subtropical atmospheric constituents, *J. Geophys. Res.*, 120, doi:10.1002/2015JD023403, online first, 2015.
- Bastrikov, V., Steen-Larsen, H. C., Masson-Delmotte, V., Gribanov, K., Cattani, O., Jouzel, J., and Zakharov, V.: Continuous measurements of atmospheric water vapour isotopes in western Siberia (Kourovka), *Atmos. Meas. Tech.*, 7, 1763–1776, doi:10.5194/amt-7-1763-2014, 2014.
- Berkelhammer, M., Hu, J., Bailey, A., Noone, D., Still, C., Barnard, H., Gochis, D., Hsiao, G., Rahn, T., and Turnipseed, A.: The nocturnal water cycle in an open-canopy forest, *J. Geophys. Res.*, 118, 10225–10242, doi:10.1002/jgrd.50701, 2013.
- Bigeleisen, J.: Statistical mechanics of isotope effects on the thermodynamic properties of condensed systems, *J. Chem. Phys.*, 34, 1485–1493, 1961.
- Bonne, J.-L., Masson-Delmotte, V., Cattani, O., Delmotte, M., Risi, C., Sodemann, H., and Steen-Larsen, H. C.: The isotopic composition of water vapour and precipitation in Ivittuut, southern Greenland, *Atmos. Chem. Phys.*, 14, 4419–4439, doi:10.5194/acp-14-4419-2014, 2014.
- Craig, H.: Isotopic variations in meteoric waters, *Science*, 133, 1702–1703, 1961.
- Dansgaard, W.: Stable isotopes in precipitation, *Tellus*, 16, 436–468, 1964.
- Ehhalt, D. H.: Vertical profiles of HTO, HDO, and H<sub>2</sub>O in the troposphere, NCAR Tech. Note NCAR–TN/STR–100, Natl. Cent. for Atmos. Res., Boulder, CO, USA, 131 pp., 1974.
- Ellehoj, M. D., Steen-Larsen, H. C., Johnsen, S. J., and Madsen, M. B.: Ice-vapor equilibrium fractionation factor of hydrogen and oxygen isotopes: experimental investigations and implications for stable water isotope studies, *Rapid Commun. Mass Sp.*, 27, 2149–2158, doi:10.1002/rcm.6668, 2013.
- Galewsky, J., Strong, M., and Sharp, Z. D.: Measurements of water vapor D/H ratios from Mauna Kea, Hawaii, and implications for

- subtropical humidity dynamics, *Geophys. Res. Lett.*, 34, L22808, doi:10.1029/2007GL031330, 2007.
- Galewsky, J., Rella, C., Sharp, Z., Samuels, K., and Ward, D.: Surface measurements of upper tropospheric water vapor isotopic composition on the Chajnantor Plateau, Chile, *Geophys. Res. Lett.*, 38, L17803, doi:10.1029/2011GL048557, 2011.
- Gat, J. R.: Oxygen and hydrogen isotopes in the hydrologic cycle, *Annu. Rev. Earth Pl. Sc.*, 24, 225–262, 1996.
- Gkinis, V., Popp, T. J., Johnsen, S. J., and Blunier, T.: A continuous stream flash evaporator for the calibration of an IR cavity ring-down spectrometer for the isotopic analysis of water, *Isotop. Environ. Health S.*, 46, 463–475, doi:10.1080/10256016.2010.538052, 2010.
- Gröning, M.: Improved water  $\delta^2\text{H}$  and  $\delta^{18}\text{O}$  calibration and calculation of measurement uncertainty using a simple software tool, *Rapid Commun. Mass Spectrom.*, 25, 2711–2720, doi:10.1002/rcm.5074, 2011.
- Gupta, P., Noone, D., Galewsky, J., Sweeney, C., and Vaughn, B. H.: Demonstration of high-precision continuous measurements of water vapor isotopologues in laboratory and remote field deployments using wavelength-scanned cavity ring-down spectroscopy (WS-CRDS) technology, *Rapid Commun. Mass Sp.*, 23, 2534–2542, 2009.
- Hurley, J. V., Galewsky, J., Worden, J., and Noone, D.: A test of the advection-condensation model for subtropical water vapor using stable isotopologue observations from Mauna Loa Observatory, Hawaii, *J. Geophys. Res.*, 117, D19118, doi:10.1029/2012JD018029, 2012.
- Hut, G.: Consultants' group meeting on stable isotope reference samples for geochemical and hydrological investigations, Consultant's Group Meeting on stable isotope reference sample for geochemical and hydrological investigations, International Atomic Energy Agency (IAEA), Vienna, Austria, 16–18 September 1985, 42 pp., available at: [http://www.iaea.org/inis/collection/NCLCollectionStore/\\_Public/18/075/18075746.pdf](http://www.iaea.org/inis/collection/NCLCollectionStore/_Public/18/075/18075746.pdf), 1987.
- IAEA: Reference sheet for VSMOW2 and SLAP2 international measurement standards, International Atomic Energy Agency, Vienna, Austria, 5 pp., available at: [http://curem.iaea.org/catalogue/SI/pdf/VSMOW2\\_SLAP2.pdf](http://curem.iaea.org/catalogue/SI/pdf/VSMOW2_SLAP2.pdf), 2009.
- Iannone, R. Q., Romanini, D., Kassi, S., Meijer, H. A. J., and Kerstel, E. R. Th.: A microdrop generator for the calibration of a water vapor isotope ratio spectrometer, *J. Atmos. Oceanic Technol.*, 26, 1275–1288, doi:10.1175/2008JTECHA1218.1, 2009.
- Johnson, L. R., Sharp, Z. D., Galewsky, J., Strong, M., Van Pelt, A. D., Dong, F., and Noone, D.: Hydrogen isotope correction for laser instrument measurement bias at low water vapor concentration using conventional isotope analyses: application to measurements from Mauna Loa Observatory, Hawaii, *Rapid Commun. Mass Sp.*, 25, 608–616, 2011.
- Koehler, G. and Wassenaar, L. I.: Realtime stable isotope monitoring of natural waters by parallel-flow laser spectroscopy, *Anal. Chem.*, 83, 913–919, doi:10.1021/ac102584q, 2011.
- Kurita, N., Newman, B. D., Araguas-Araguas, L. J., and Aggarwal, P.: Evaluation of continuous water vapor  $\delta\text{D}$  and  $\delta^{18}\text{O}$  measurements by off-axis integrated cavity output spectroscopy, *Atmos. Meas. Tech.*, 5, 2069–2080, doi:10.5194/amt-5-2069-2012, 2012.
- Lee, X., Sargent, S., Smith, R., and Tanner, B.: In situ measurement of the water vapor  $^{18}\text{O}/^{16}\text{O}$  isotope ratio for atmospheric and ecological applications, *J. Atmos. Ocean. Tech.*, 22, 555–565, doi:10.1175/JTECH1719.1, 2005.
- Lis, G., Wassenaar, L. I., and Hendry, M. J.: High-precision laser spectroscopy D/H and  $^{18}\text{O}/^{16}\text{O}$  measurements of microliter natural water samples, *Anal. Chem.*, 80, 287–293, doi:10.1021/ac701716q, 2008.
- Loader, C.: *Local Regression and Likelihood*, Springer Verlag, New York, USA, 290 p., 1999.
- Luo, H., Goodman, K. J., Hinckley, E. S., West, J. B., Williams, D. G., and Bowen, G. J.: Continental-scale stable isotope measurements at NEON to address ecological processes across systems, Fall Meeting of the American Geophysical Union, San Francisco, California, 9–13 December 2013, H51K-1349, 2013.
- Noone, D., Galewsky, J., Sharp, Z. D., Worden, J., Barnes, J., Baer, D., Bailey, A., Brown, D. P., Christensen, L., Crosson, E., Dong, F., Hurley, J. V., Johnson, L. R., Strong, M., Toohey, D., Van Pelt, A., and Wright, J. S.: Properties of air mass mixing and humidity in the subtropics from measurements of the D/H isotope ratio of water vapor at the Mauna Loa Observatory, *J. Geophys. Res.*, 116, D22113, doi:10.1029/2011JD015773, 2011.
- Noone, D., Risi, C., Bailey, A., Berkelhammer, M., Brown, D. P., Buening, N., Gregory, S., Nusbaumer, J., Schneider, D., Sykes, J., Vanderwende, B., Wong, J., Meillier, Y., and Wolfe, D.: Determining water sources in the boundary layer from tall tower profiles of water vapor and surface water isotope ratios after a snowstorm in Colorado, *Atmos. Chem. Phys.*, 13, 1607–1623, doi:10.5194/acp-13-1607-2013, 2013.
- Penna, D., Stenni, B., Šanda, M., Wrede, S., Bogaard, T. A., Gobbi, A., Borga, M., Fischer, B. M. C., Bonazza, M., and Chárová, Z.: On the reproducibility and repeatability of laser absorption spectroscopy measurements for  $\delta^2\text{H}$  and  $\delta^{18}\text{O}$  isotopic analysis, *Hydrol. Earth Syst. Sci.*, 14, 1551–1566, doi:10.5194/hess-14-1551-2010, 2010.
- Penna, D., Stenni, B., Šanda, M., Wrede, S., Bogaard, T. A., Micheli, M., Fischer, B. M. C., Gobbi, A., Mantese, N., Zuecco, G., Borga, M., Bonazza, M., Sobotková, M., Cejková, B., and Wassenaar, L. I.: Technical Note: Evaluation of between-sample memory effects in the analysis of  $\delta^2\text{H}$  and  $\delta^{18}\text{O}$  of water samples measured by laser spectrometers, *Hydrol. Earth Syst. Sci.*, 16, 3925–3933, doi:10.5194/hess-16-3925-2012, 2012.
- Rambo, J., Lai, C. T., Farlin, J., Schroeder, M., and Bible, K.: On-site calibration for high precision measurements of water vapor isotope ratios using off axis cavity-enhanced absorption spectroscopy, *J. Atmos. Ocean. Tech.*, 28, 1448–1457, doi:10.1175/JTECH-D-11-00053.1, 2011.
- Samuels-Crow, K. E., Galewsky, J., Sharp, Z. D., and Dennis, K. J.: Deuterium excess in subtropical free troposphere water vapor: continuous measurements from the Chajnantor Plateau, northern Chile, *Geophys. Res. Lett.*, 41, 8652–8659, doi:10.1002/2014GL062302, 2014.
- Schmidt, M., Maseyk, K., Lett, C., Biron, P., Richard, P., Barriac, T., and Seibt, U.: Concentration effects on laser-based  $\delta^{18}\text{O}$  and  $\delta^2\text{H}$  measurements and implications for the calibration of vapour measurements with liquid standards, *Rapid Commun. Mass Sp.*, 24, 3553–3561, 2010.

- Sharp, Z.: Principles of Stable Isotope Geochemistry, Pearson Prentice Hall, Upper Saddle River, New Jersey, USA, 344 p., 2007.
- Steen-Larsen, H. C., Johnsen, S. J., Masson-Delmotte, V., Stenni, B., Risi, C., Sodemann, H., Balslev-Clausen, D., Blunier, T., Dahl-Jensen, D., Ellehøj, M. D., Falourd, S., Grindsted, A., Gkinis, V., Jouzel, J., Popp, T., Sheldon, S., Simonsen, S. B., Sjolte, J., Steffensen, J. P., Sperlich, P., Sveinbjörnsdóttir, A. E., Vinther, B. M., and White, J. W. C.: Continuous monitoring of summer surface water vapor isotopic composition above the Greenland Ice Sheet, *Atmos. Chem. Phys.*, 13, 4815–4828, doi:10.5194/acp-13-4815-2013, 2013.
- Steen-Larsen, H. C., Masson-Delmotte, V., Hirabayashi, M., Winkler, R., Satow, K., Prié, F., Bayou, N., Brun, E., Cuffey, K. M., Dahl-Jensen, D., Dumont, M., Guillevic, M., Kipfstuhl, S., Landais, A., Popp, T., Risi, C., Steffen, K., Stenni, B., and Sveinbjörnsdóttir, A. E.: What controls the isotopic composition of Greenland surface snow?, *Clim. Past*, 10, 377–392, doi:10.5194/cp-10-377-2014, 2014a.
- Steen-Larsen, H. C., Sveinbjörnsdóttir, A. E., Peters, A. J., Masson-Delmotte, V., Guishard, M. P., Hsiao, G., Jouzel, J., Noone, D., Warren, J. K., and White, J. W. C.: Climatic controls on water vapor deuterium excess in the marine boundary layer of the North Atlantic based on 500 days of in situ, continuous measurements, *Atmos. Chem. Phys.*, 14, 7741–7756, doi:10.5194/acp-14-7741-2014, 2014b.
- Steen-Larsen, H. C., Sveinbjörnsdóttir, A. E., Jonsson, Th., Ritter, F., Bonne, J.-L., Masson-Delmotte, V., Sodemann, H., Blunier, T., Dahl-Jensen, D., and Vinther, B. M.: Moisture sources and synoptic to seasonal variability of North Atlantic water vapor isotopic composition, *J. Geophys. Res. Atmos.*, 120, 5757–5774, doi:10.1002/2015JD023234, 2015.
- Sturm, P. and Knohl, A.: Water vapor  $\delta^2\text{H}$  and  $\delta^{18}\text{O}$  measurements using off axis integrated cavity output spectroscopy, *Atmos. Meas. Tech.*, 3, 67–77, doi:10.5194/amt-3-67-2010, 2010.
- Tanweer, A., Gröning, M., van Duren, M., Jaklitsch, M., and Pöhlstein, L.: Stable isotope internal laboratory water standards: preparation, calibration and storage, Technical Procedure Note No. 43, Isotope Hydrology Laboratory, International Atomic Energy Agency, Vienna, Austria, 2009.
- Tremoy, G., Vimeux, F., Cattani, O., Mayaki, S., Souley, I., and Favreau, G.: Measurements of water vapor isotope ratios with wavelength-scanned cavity ring-down spectroscopy technology: new insights and important caveats for deuterium excess measurements in tropical areas in comparison with isotope-ratio mass spectrometry, *Rapid Commun. Mass Sp.*, 25, 3469–3480, 2011.
- Tremoy, G., Vimeux, F., Mayaki, S., Souley, I., Cattani, O., Risi, C., Favreau, G., and Oi, M.: A 1 year long  $\delta^{18}\text{O}$  record of water vapor in Niamey (Niger) reveals insightful atmospheric processes at different timescales, *Geophys. Res. Lett.*, 39, L08805, doi:10.1029/2012GL051298, 2012.
- van Geldern, R. and Barth, J. A. C.: Optimization of instrument setup and post-run corrections for oxygen and hydrogen stable isotope measurements of water by isotope ratio infrared spectroscopy (IRIS), *Limnol. Oceanogr.-Meth.*, 10, 1024–1036, doi:10.4319/lom.2012.10.1024, 2012.
- Wang, L., Caylor, K. K., and Dragoni, D.: On the calibration of continuous, high-precision  $\delta^{18}\text{O}$  and  $\delta^2\text{H}$  measurements using an off axis integrated cavity output spectrometer, *Rapid Commun. Mass Sp.*, 23, 530–536, doi:10.1002/rcm.3905, 2009.
- Wang, L., Caylor, K. K., Camilo Villegas, J., Barron-Garfford, G. A., Breshears, D. D., and Huxman, T. E.: Partitioning evapotranspiration across gradients of woody plant cover: assessment of a stable isotope technique, *Geophys. Res. Lett.*, 37, L09401, doi:10.1029/2010GL043228, 2010.
- Wen, X. F., Lee, X., Sun, X. M., Wang, J. L., Tang, Y. K., Li, S. G., and Yu, G. R.: Intercomparison of four commercial analyzers for water vapor isotope measurement, *J. Atmos. Ocean. Tech.*, 29, 235–247, 2012.