Reply to the reviewers' comments: Comparison of the oxygen isotope signatures in speleothem records and iHadCM3 model simulations for the last millennium (cp-2019-121)

Janica C. Bühler, Carla Roesch, Moritz Kirschner, Louise Sime, Max D. Holloway, Kira Rehfeld

November 6, 2020

Summary of changes

We thank the reviewer for her/his constructive comments and detailed reading. In response to the suggestions by the reviewer we plan to

- include the new plots on the influence of external forcing that the reviewer suggested in the discussion and add the figure to the supplement,
- additionally explore the impact of large scale climate modes on $\delta^{18}O$ variability and include the results in the discussion,
- update Fig. 8 and SFig. 7 to include more information on age-model uncertainty as suggested by the reviewer,
- revise the text throughout the manuscript to clarify statements,
- enhance the discussion especially regarding all analyses and interpretations that include age-model model ensembles,
- and fix formatting where necessary.

A detailed response to the helpful remarks of the referee is given below.

1 Reply to the first reviewer

(Original report cited in italics)

Bühler et al. explore the temporal and spatial variability of speleothem d18O for the past1000 time frame (850-1850) using a global compilation of speleothem data and a 1000-year run with an isotope enabled climate model. The authors briefly investigate the relation *of d18O to temperature and precipitation in the model, and compare the modelled temperature and precipitation to speleothem d18O. Next the authors explore the spatial relation between a number of variables such as latitude, annual mean temperature, precipitation, and the mean speleothem d18O. They go on to compare the spectrum of temporal variability in the model to the speleothem data. Finally they investigate the teleconnection patterns in speletothem data and in the model. The main conclusions are that i) high frequency variability is dampened in the speleothem data due to the hydrological residence time before reaching the cave ii) modelled centennial variability is underestimated iii) teleconnections are hard to find in the speleothem data, while more easily identified in the model iv) low signal-to-noise ratio for the speleothem data due to local processes makes it difficult to interpret.*

This study contains a lot of interesting and useful work to better understand speleothem and modelled d18O. However, I find some aspects missing that would motivate some of the things studied in the paper, and more analysis is needed to round off the study. Overall the paper is well-written with minor typographical issues. I hope that the authors will see my comments as a positive contribution, as I think the study has a lot of potential, but requires some more work. In summary, I recommend that the manuscript requires major revisions, but will then most certainly be a valued contribution to the topic.

We thank the reviewer for this positive assessment.

Major comments:

1) Why even look at teleconnections in the d18O data (Figure 8)? There is no mention of ITCZ variability, monsoon, NAO, or other mechanisms driving large-scale d18O variability. I can understand if the authors want to keep the analysis general, but through the whole paper there is no mention of any of the main climate patterns that could explain the teleconnections in Figure 8. See references listed in comment for L78-L80. This should at the very least be mentioned in the introduction, and included in the discussion. There is a lack in information of how HadCM3 performs when it comes to large-scale patterns, and what the imprint is on d18O. For example, add extra correlation maps in Figure 7 for the most important patterns, which quickly could be done. Correlating the monsoon index (e.g. Vuille et al., 2005) to d18O in precipitation should show a very clear pattern across the region around the Indian Ocean. This is not obvious when looking at grid point correlation of climate fields, because the main driving factor is not local precipitation amount, but down wind recycling of vapour in large-scale organized convection.

We thank the reviewer for this suggestion for additional analyses. We agree that the investigation of modes of variability and the modeled $\delta^{18}O$ response is very interesting. However, attributing the variability of speleothem $\delta^{18}O_{speleo}$ to specific modes is a challenge that we would not be able to address in a single manuscript. Also, a full-blown analysis on local drivers is not feasible and appropriate given the resolution of the model. The key idea is to show the general correlation patterns obtained from the data, and that of the model while taking into account the resolution difference. This does indeed not allow us to look into teleconnections or directly attribute to drivers of variability. Instead, it gives an overview over the general global spatial correlation structure, and whether there are similarities between the relationships obtained from proxy data, and from model simulations. We identify quite some gaps that require taking into account more proxy-related uncertainties and processes. Investigating to what extent large scale modes of variability impact $\delta^{18}O$ variability would of course be very interesting, and while an exhaustive investigation is beyond the scope here, we plan to explore the potential for it. We started to initially test the impact of some modes (NAO-index, ENSO-index, monsoon-index) and plan to extend the discussion to include some of these new insights.

2) The authors mention external forcing several times as a driver of variability, but never explains or does any analysis to show how this this is related to climate or d18O. This is of course a big topic (e.g. Swingedouw et al., 2017) and might be beyond the scope of the paper. Please either perform analysis of the impact of forcings or be more careful when making statements about what variability is forced and what is not forced.

A very interesting point, indeed. We will re-check our statements to ensure that forced and internal variability are appropriately distinguished. We have previously correlated the solar irradiance time series used for the forcing with $\delta^{18}O$ variability at the annual scale without seeing strong impacts. As visible in Fig. 1 below, this yields hardly any regions with correlation coefficients clearly distinct from zero. Volcanic forcing, on the other hand shows a clearer imprint on $\delta^{18}O$ variability. We will add a new figure to the supplement to underline these statements. These aspects are not explicitly discussed in the paper and we will amend the discussion to include them. Fig. 1 below illustrates the correlation map between a-c) volcanic forcing to ensemble mean temperature, precipitation and $\delta^{18}O$ changes, and in d-f) solar forcing to the climate variables. Especially for temperature we see a clear climatic influence by volcanic forcing, which is also visible in the timeseries of GMST. Precipitation and its isotopic composition however shows only a weak and non-uniform influence of volcanic forcing. Generally, the influence of solar forcing on all climate variables is very weak. The area-weighted mean correlation to solar forcing to the isotopic composition of precipitation is -0.01 (-0.04, 0.06 90% confidence interval) and -0.08 (-0.18, 0.00) for volcanic forcing. We conclude that the external forcing has little influence on the $\delta^{18}O$ signature in the simulation. We will include this figure in the supplement of the manuscript and incorporate these points in the discussion.

3) The authors have three simulations but appear to make very little use of the additional information to be gained from this. While three simulations is not a huge ensemble it still yields much more information on forced versus internal variability than a single

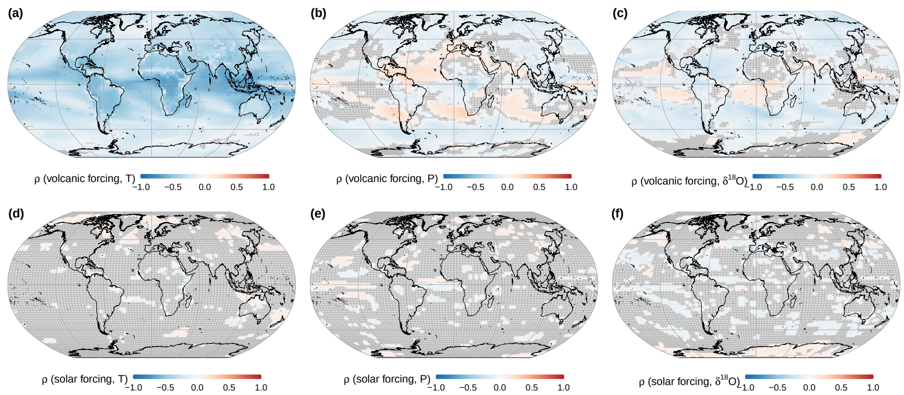

Figure 1: Correlation estimate fields of a) ensemble mean simulated temperature, b) precipitation, and c) $\delta^{18}O$ to volcanic forcing, and the same ensemble mean climate variables to solar forcing d-f). Empty tiles mask gridboxes with $p > 0.1$. The area-weighted average correlation estimates with volcanic forcing are ρ(T,volc) = −0.34 (−0.48, −0.11), ρ(P,volc) = −0.04 (−0.15, 0.05), $\rho(\delta^{18}O$,volc) = −0.08 (−0.18, 0.00). The area-weighted average correlation estimates with solar forcing are ρ(T,sol) = 0.01 (−0.004, 0.03), ρ(P,sol) = 0.003 (−0.009, 0.024), $\rho(\delta^{18}O$,sol) = −0.012 (−0.035, 0.056). The correlation estimates are calculated with 989 degrees of freedom.

simulation. When you perform correlation analysis between speleothem data and simulated d18O, this should be done using the ensemble mean. How similar are the ensemble runs in variability? How is the ensemble setup? There is very little information on this.

Thank you for this suggestion. Following them we will extend the description of the ensemble. The data, and its description, was uploaded to Pangaea prior to submission but apparently there is a backlog and they are not available yet. The three ensemble members were initialized from different years of the same spinup simulation. We will add this in the method section. We do not, however think that using the ensemble mean is necessarily appropriate in the assessment of variability changes, as this would amplify the forced response and dampen the dynamic part in the signal. The relative role of natural forcing is not the main focus of this paper. We will assess the degree of correlation between the $\delta^{18}O$ fields that is due to the common forcing. The insight from this analysis will also be added to the discussion.

4) The study uses a shot gun kind of approach to age-model uncertainties. As I understand the different age-models of individual speleothems are sampled independently when testing the range of possible age-models. But are all age-models really equally likely, for example for neighbouring speleothems that we expect to be correlated? Related to this.

The new age-models provided with the SISAL chronology (Comas-Bru, Rehfeld, Roesch et al., ESSD 2020) are not ranked by likelihood. All of them are consistent with the radiometric chronological constraints. Therefore we indeed consider all of them in the correlation analyses. In all other analyses we use the corresponding original age models. We make this clearer in the method section.

When comparing the down-sampled modelled d18O to speleothem data in Figure 8, shouldn't the age-model uncertainties also be included for the model data to make the results truly comparable? For completeness there should be two more tests plotted in Figure 8: i) model data which is not down sampled (include SF7 a) and b) in Figure 8, I suppose?) ii) model data including age-model uncertainties. I think this issue with the comparison of model and speleothem data and differences in teleconnections depending on data treatment should be more emphasized.

We absolutely agree with the reviewer that adding this information makes Fig. 8 more informative. We will update Fig. 8 accordingly. It will include the record data, the simulated data and the downsampled-data, including the age-model uncertainty testing. We will also adjust the color scheme, as suggested in the detailed comments.

Detailed Comments

L5-L12 Briefly say that d18O is a climate proxy before discussing all the implications of sampling etc.

We agree that adding this general statement prior to the specific impacts improves readability. We adjusted the text as follows:

'The oxygen isotopic ratio $\delta^{18}O$ **, a proxy for many different climate variables,** is routinely measured in speleothem samples at decadal or higher resolution and single specimens can cover full Glacial-Interglacial cycles.'

L15 "We evaluate systematically. . . " change to "We systematically evaluate . . . "?

Done.

L16 ". . . and test for the main climate drivers for individual records or regions." change to ". . . and test for the main climate drivers recorded in d18O for individual records or regions."?

Done.

L17-L19 Maybe it is be better (worse) to explain in full sentences (fancy truncated syntax)?

We liked the short syntax. Nevertheless, we adjusted the text as follows: 'However, using robust filters and spectral analysis, we show that the observed proxy-based variability of $\delta^{18}O$ is lower than simulated by iHadCM3 on decadal**, and higher on centennial** timescales.'

L28 ". . . natural and human systems . . . " maybe change to ". . . human societies and the environment . . . "? We adjusted the text as follows:

The impacts of a changing climate have been observed over the last century (IPCC, 2013) and indicate a strong sensitivity of **human societies and natural systems** to changes in climate.

L36 The delta-notation comes in here before defining it or telling what the proxy is good for. Either move the definition up in the manuscript and description of the d18O proxy or call it "the relative abundance of 18-O" until you get to the definition, and then explain briefly that this is a climate proxy. You can't discuss the challenges of the interpretation before telling the basics. Thank you for pointing this out. We adjusted the section as follows:

'... Therefore, for model evaluation on longer than centennial time scales, we have to rely on evidence from paleoclimate archives, such as trees, ice cores, foraminifera from marine sediment cores, or speleothems. **The abundance of the heavy oxygen isotope ^{18}O, further denoted as $\delta^{18}O$, is a proxy for many climate variables and** can be measured on these, and quite a few other paleoclimate archives with high precision (Schmidt et al., 2014)...'

The formal definition then follows, starting in L63.

L41 You need to include the simulations with GISS ModelE2-R (Colose et al., 2016) and iCESM1 (Stevenson et al., 2019).

Thank you for pointing this out. We now include these studies in the literature review.

L42 Sjolte et al. (2018) compared the variability of the modelled ECHAM5/MPI-OM d18O to Greenland ice core d18O and used the model to assimilate the ice core data to produce gridded reconstructions. Never compare the proxy data to the model – it's the other way around!

L44 Again: Never compare the proxy data to the model! It's not the observations that are being evaluated. Thank you for the clarification. We will revise the manuscript for the expression and change the sentence stucture accordingly. The lines mentioned here were changed to:

'Few other transient model-data comparison studies focused on $\delta^{18}O$ (e.g., Wackerbarth, 2012; Dee et al., 2015; Parker et al., 2020). For example, Sjolte et al. (2018) **compared the variability of the simulated ECHAM5/MPI-OM $\delta^{18}O$ to Greenland ice cores** over the last millennium **to assimilate the ice core data to produce gridded reconstructions. They** were able to differentiate between solar and volcanic forcing effects from their reconstructions. On orbital timescales (150,000 yr), Caley et al. (2014)

compared **a transient isotope-enabled simulation with the model of intermediate complexity iLOVECLIM to speleothem records from South East Asia**. They found model-data similarity for the broad temporal trends, but differences at shorter timescales, highlighting the role of seasonality.'

L56-L61 These are a very important points and is written in almost bullet point-style. Please add more details to make it more comprehensible to non-experts. For example, Laepple and Huybers (2014a) are talking about decadal and longer time scales. Laepple and Huybers (2014b) say that the models are too diffusive which is not the same as saying "too high diffusivity", depending on context. Here, they mean that the energy dissipates too quickly across the spectra of temporal variability, which is not clear in your text. My advice is to spend a bit more space on this part of the introduction and don't mix topics, such as too diffusive models and missing processes and feedbacks in the same sentence, unless linking these things directly together.
We agree that these are very important points. Therefore, thank you for pointing out that the discussion should be extended. We will grant the section more space in the revised version.

L64 Add white space after "climate system".
Done

L78-L80 There is quite some evidence that d18O is not primarily a proxy of neither local temperature nor precipitation, but strongly related to circulation modes, large-scale climate patterns and downwind fractionation. For example, in the North Atlantic region the North Atlantic Oscillation (NAO) is important for d18O variability (Vinther et al., 2010; Sjolte et al., 2011; Deininger et al., 2016), while downwind fractionation connected with the Indian summer monsoon impacts the cave d18O in the region around the Northern Indian Ocean, China and South-East Asia (Vuille et al., 2005; Fleitmann et al., 2007; Pausata er al., 2011; Kurita et al., 2013; Lekshmy et al., 2014; Liu et al., 2014; Sjolte et al., 2014; Zhang and Jin, 2015). I think these factors should be highlighted in the introduction.
Thank you for pointing this out. We adjusted the section as follows:

'... $\delta^{18}O$ can be regarded as a proxy **for example** for surface temperature variations in higher latitudes, and precipitation amount in the tropics (Dansgaard, 1964), overlayed with distinct observable signatures of source water evaporation**,** transportation over longer distances (Bradley, 1999; Dansgaard, 1964)**, and large scale-climate patterns of circulation such as e.g. the North Atlantic Oscillation (NAO) (e.g. Vinther et al., 2010) or the El-Niño Southern Oscillation (ENSO) (Tindall et al., 2009)**. These signatures $\delta^{18}O$ **in precipitation may** also visible in speleothem records, including additionally a fractionation process involved in the calcite formation, which is primarily temperature-dependent (Urey, 1948; McCrea, 1950)...'

L91-L99 Here you mainly list the contents of the paper. Can you make the science questions that you are pursuing more clear? Maybe you are testing the climate controls on the variability in simulated d18O using an isotope enabled climate model and compare this to speleothem d18O in a global dataset? Formulate more like hypothesis testing rather than say what kind of analysis you are doing.

We plan to restructure the paragraph such that it reads:

Here, we present three new last millennium isotope-enabled simulations from the iGCM version 3 of the Hadley Model (iHadCM3) and test how similar the $\delta^{18}O$ variations in iHadCM3 and speleothem records are (Sec. 4.1). A characterization of the datasets and relevant forcing can be found in Fig. 1. The robustness of the findings and methods are evaluated over the last millennium, for which a large number of high-resolution proxy datasets from the SISAL v.2. database (Comas-Bru et al., 2020) are available. **Our key question are: i) how similar are the modeled $\delta^{18}O$ signatures to the speleothem records especially regarding variability, ii) can we distinguish main drivers for these signatures, and iii) how representative are the speleothem records for their region. To address these questions,** we explore these similarities on both spatial and temporal scales, to distinguish patterns of the mean state (Sec. 4.1), the variability (Sec. 4.2 and Sec. 4.3), and the spatial representativity of speleothem climate records (Sec. 4.4 and Sec. 4.5). We examine the simulation's capability to simulate and the records' capability to capture variability on different time scales to improve our understanding of processes and uncertainties of both.

L108 Add white space ". . . 30min . . . "

Done.

L115 What is meant by "ice sheet" here? I suppose the model doesn't have an ice sheet model?

We apologize for this mistake. This statement refers to the sea ice model as documented in Valdes et al. (2017).

'Compared to instrumental observations, the model represents sea surface temperature (SST), **sea ice**, and ocean heat content well (Gordon et al., 2000).'

L120 ". . . features like latitude effect, amount effect, or the continental effect . . . " this is partly repetition from L118. Why not lump these things together?

Thank you for pointing this out, this is truly a repetition. We rearrange the sentence as follows:

'...The model simulates the major isotopic fractionation effects defined by Dansgaard

(1964) (e.g. the latitude effect, the amount effect, and the continental effect) appropriately compared to GNIP data (Zhang et al., 2012). Additionally, a broad agreement in isotopic output with GNIP data in the general spatial distribution can be observed and the **above mentioned** general oxygen isotopic ratio features are **represented** well (Tindall et al., 2009). As such, iHadCM3 captures large scale features of climate and oxygen isotope ratios while remaining computationally efficient for the simulation of timescales such as the last millennium...'

L123 So, what are the differences between the three model runs? Initial state of the ocean?
See response to major comment #3 above.

Figure 2, caption. Add white space "600yr".
L153 "600y" add white space, and I believe Clim Past uses "yr" shorthand for year.
Thank you for your detailed reading. We revised the document for white spaces and modified the abbreviation.

L161-169 As I understand you allow any type of age model to be used out of the many options, and you pick the best fit independently for each site/speleothem? What are the criteria for accepting an age model, be sides that it is the best fit? Are there cases where the "best" age model is outside of the uncertainty range of the U/Th dating?
In general we use the authors original chronology for the analysis. In the cross-correlation analysis (esp. Fig 8 and 9) we test, how much correlation estimates change depending on the choice of age models. Here we use all age-model realizations provided by the SISALv2 database. The SISALv2 database defines one median best-fit estimate for each age modeling method, following their selection criteria (see Comas-Bru et al., 2020). As a median age model, it cannot lie outside the range of the ensemble members. Nevertheless, age controls only exist at the radiometric dates, and depending on the dating density and whether reversals were found some age models may of course lie outside individual dates if the dating evidence is contradictory. These age models are method-dependent, consistent with the evidence and free of reversals, as described in Comas-Bru et al. (2020). In the correlation assessment we use all available age model realizations. We will adjust the section. See also our response to major comment # 4.

L176 How do you decide on the nine clusters? Is this what you describe L236-L239. Please clarify.
We decided on eight distance-based clusters and manually added a ninth cluster, to separate a cluster containing all East Asian speleothems above 20°N from those below. We made the link to the latter section, where it is explained in more detail, clearer as follows:

L176: 'For the investigation of spatial correlation patterns by network analysis the set of speleothems is divided into nine regional clusters (Fig. 2)**, as explained in detail in Sec.3.3**.'

L236: 'We split the network into **eight** sub-networks by hierarchical distance-based clustering of the node locations. The cluster that includes all East Asian caves is manually split into two clusters, one for East Asia (all caves above 20°N) and a cluster of South East Asia (all caves below 20°N). **With this, we end up with nine clusters as depicted in Fig. 2.**...'

L180 '. . . 10 or more d18O sampled.' should it be '. . . 10 or more d18O samples.'? Otherwise please rephrase.
Thank you for noticing this typo. We corrected it.

L181 "We exclude six speleothems of mixed mineralogy." Why? This is an excellent question. We require information on the mineralogy of the samples for the conversion to drip water $\delta^{18}O_{dw.eq}$. For samples of mixed mineralogy it is unclear to what extent the correction is appropriate. Therefore, and following Comas-Bru et al. (2019), we excluded those speleothems with mixed mineralogy. We will add this clarification to L181 in the manuscript.

L228 If you chose the highest correlation out of a large ensemble of possible solutions, how do you account for this when determining the significance of the correlation?
We only choose the highest correlation estimate from significant cross-correlation estimates. If the cross-correlation between two speleothems using a specific pair of age-models is non-significant, it is not chosen as a 'best fit'. We will clarify this in the manuscript at the respective section.

L256-L257 If you calculate the regional lapse rate of 18O in the model you can estimate the contribution of the model orography to the d18O biases. L265 Did you try doing multivarite regression? To know the influence on one parameter you need to isolate it from the other parameters.
Thank you for these two interesting suggestions which we will add to the discussion section. We will consider them for a further planned study, where we will look more closely at the biases and the influence of different parameters. In this manuscript, however, we only want to give a first glance at the potential of the analysis.

L273 Add white space "both in the annual mean andfor ...".
Thank you for noticing. We added the space.

L277-L278 "To analyze . . ." please rewrite this sentence more concisely and remember, again, that you are comparing the model to the data.
We adjusted the sentence in accordance with your comment to L42 and clarified the state-

ment as follows:

'..To analyze how similar the variability of the isotopic signal is **in the iHadCM3 climate model and in the speleothems**, we compare the total variance of **the simulation to the speleothem records** over the last millennium. The global distribution of variance ratios between $\delta^{18}O_{dw.eq}$ and $\delta^{18}O$ (Fig. 2a) shows overall higher variability in the speleothem records than in the simulation, with local exceptions...'

L311 Add white space "3yr". I see space missing many places before "yr". Please check in general.
Sorry for the trouble. We now checked the manuscript thoroughly for this particular mistake.

Figure 7, caption. Here, "insignificant" should be "non-significant". I assume you use the term "significant" in a statistical sense?
Adjusted.

Section 4.4 Did you look at the relation of d18O to climate modes? See comment above to L78-L80. For example, the monsoon index (Vuille et al., 2005) might have a stronger imprint on d18O in the tropical Indian Ocean than local precipitation amount.
See response to major comment #1.

L320 "and the climate variable is shown." Change to "and the climate variable is also shown."?
Done.

L325 What about the correlation of LM2 and LM3 to the proxy data? Using the model ensemble could give a clue if the variability is related to forcing.
We agree with the reviewer. See response to major comment # 3. We will adjust this paragraph according to our findings there.

L332 $p < 0.1$ *is not a strong significance criterion. How many samples are there? And in the first place can we expect much correlation between a single model run and observed climate? Changing the initial conditions of the model run would likely affect these correlations, since this is just one realisation, no?*
Indeed, $p < 0.1$ is not generally strong criterion for significance of correlation estimates. However, we aim to choose criteria that are appropriate for both palaeoclimate archive and model data time series. Therefore, we need to balance this strictness and the expected level of false positives against that of data demands and the available number of samples N. In Fig. 7, in particular, we show both model-model ($N = 1000$) and model-proxy (N varying) correlations. For irregular time series the effective degrees of freedom differ

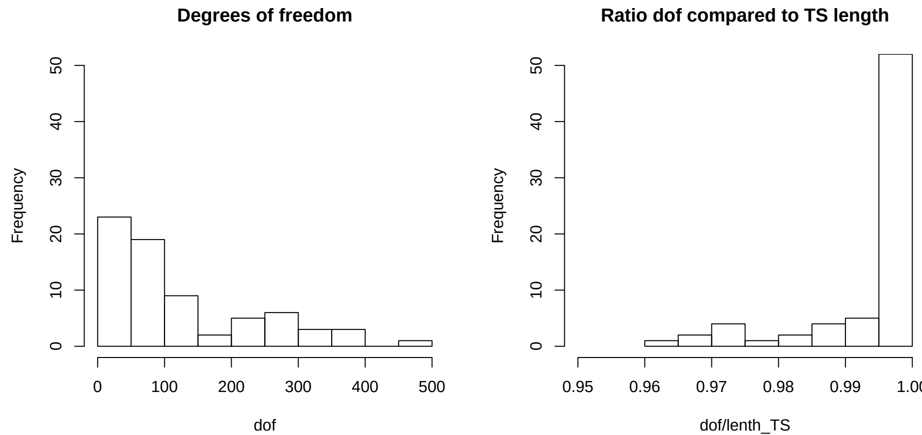

Figure 2: Left: Visualization of the degrees of freedom corresponding to the analysis of the correlation estimates between simulated temperature and speleothem $\delta^{18}O$ as in Fig.7 of the manuscript. Right: The degrees of freedom in relation to the length of the timeseries.

from the nominal value of N. The p-values for irregular series are estimated based on a t-distribution, with the degrees of freedom estimated from the temporal coverages $R_{x,y}$ and the persistence time $\tau_{x,y}$ as $N_{\text{eff}} = \min(\max(R_x/\tau_x, R_y/\tau_y, \text{na.rm=TRUE}), \max(N_x, N_y))$. This is implemented in the R package `nest` (`https://github.com/krehfeld/nest`, Rehfeld et al., 2011; Rehfeld and Kurths, 2014). For the regular time series p-values are calculated via Pearson's product moment correlation (via the function `cor.test`). We will add the degrees of freedom in the manuscript where different correlation estimators are used. In the case of the records, the estimated effective degrees of freedom range from $N_{eff} = 20$ to $N_{eff} = 470$, and they are generally similar to the length of the records (see histogram in Fig. 2). This indicates that the estimated persistence time is often of the order of magnitude of the sampling resolution.

Figure 8: I found the choice of colours confusing in Figure 8d. The smoothed lines are red and blue in the same shade as the markers for the correlation, which made me think at first that the smoothed lines were for the data marked of similar colours, which doesn't make sense. It's quite a busy plot. Consider making it easier to read by choosing different colours or making an extra subplot.
We will adjust Figure 8 according to your suggestions in major comments #4 and also account for the color-confusion. Thank you, for pointing this out.

L396-L398 "In general, . . . " I don't follow this sentence. Seems like a leap in topic. How can you say anything about forced variability without analysing it? Also, I

believe Jungclaus et al. (2010) are discussing the hemispheric mean temperature, while the speleothem d18O data is temperature, precipitation, evaporation and circulation dependent.
We agree with the reviewer, the sentence structure is unfortunate. We removed the statement to forcings, and will include it in the analysis to major comment #2. For the analysis here we only want to adress the total variability over the last millennium and adjusted the sentence as follows:

'... In general, the total variance **of the simulated $\delta^{18}O$ and of the speleothem isotopic signatures** over the last millennium are consistent. Differences in variance can, to some extent, be attributed to the sample resolution of the records, whereas down-sampling of simulated $\delta^{18}O$ decreases the variability on decadal time scales... '

L410 "However, we find little regional consistency . . . " couldn't this be due to time scale uncertainties? You find no structure in correlation for the speleothem data in Figure 8, but there could be a correlation/regional climate signal, just as well as there could be no correlation.
We agree with the reviewer that the lack of correlation could be due to time scale uncertainties. Therefore we included these uncertainties in our cross-correlation analysis where we account for age-model sensitivity. There, as the reviewer points out, we find no structure in the correlation for the speleothem data. Under the assumption, that the true age time series is covered with the age-model ensembles, we account for all age uncertainties. However, this assumption may not be true. We will clarify this potential reasons for underestimated correlations in the paragraph as follows:

'...However, we find little regional consistency and high heterogeneity in the variance estimates from the speleothem records. These findings point to the strong influence of cave internal processes or the impact of seasonally filtered data captured by speleothems, which is in agreement with McDermott et al. (2001). **Age uncertainties, that are not covered by the age-model ensembles, could also be responsible for the low similarity between isotopic signals of neighboring speleothem entities.**

L428 "longer than50yr" Spaces!
Done.

L428 "by 4% (3, 4)" Upper confidence bounds same as median? Or is this due to the number of significant digits?
Thanks for noticing. This is a rounding error. We adjusted for one extra digit, so we arrive at 4.0% $(3.3, 4.4)$.

L434 "However, no systematic pattern and few significant correlations were found for the speleothem records (Fig. 7)." Again, I'm really not surprised that there is no correlation between a free running simulation and the proxy data. There might be forced common variability between model run and proxies (volcanic, solar), but then you need to check the model and proxy response to forcings.
We will relate to this according to our analysis to major comment #2 and include the results and findings in this paragraph in the discussion.

L464 "We use a three member initial-condition ensemble from a single iGCM in this study." Please describe the model ensemble initiation in the methods section.
Sorry for missing out here. See comment L123

L470 " . . . as suggested by Dalaiden et al. (2020)." There are lots of examples of offline data assimilation. Maybe provide a few more? E.g., see references in introduction of Sjolte et al. (2020). Ice core data is synchronized using volcanic markers. Any particular age-model related uncertainties to take into account that might complicate the assimilation of speleothem data?
Data assimilation would indeed be a very interesting application for our dataset. As speleothems are dated radiometrically, the uncertainty of the age depends primarily on the concentration of the relevant isotopes used and the age limits of the method, e.g. the secular equilibrium for U/Th-dating (Scholz and Hoffmann, 2008). An additional source of uncertainty stems from growth irregularities and outliers. However, in contrast to ice cores, the uncertainty does not per se increase with depth, and no synchronization with volcanic markers are needed. Similar to this study, age uncertainties can be accounted for by the provided age-model ensembles (Comas-Bru et al., 2020). We will add this explanation and also more examples as suggested to the section corresponding to L470.

L483 ". . . such as d13C cannot (yet) be implemented in GCMs . . . " It's not that far away (Scholze et al., 2008; Camino-Serrano et al., 2019).
We will include the references in the conclusion.

L503 ". . . low signal-to-noise ratios . . . " For the speleothem data?
Clarified:
'...We found low signal-to-noise ratios **for the isotopic signatures in the speleothem records**, which imply a low spatial representativity of individual entities...'

References

Bradley, R. S.: Paleoclimate: reconstructing climates of the Quaternary, Elsevier, 1999.

Caley, T., Roche, D. M., and Renssen, H.: Orbital Asian summer monsoon dynamics revealed using an isotope-enabled global climate model, Nature Communications, 5, 6–11, https://doi.org/10.1038/ncomms6371, 2014.

Comas-Bru, L., Harrison, S. P., Werner, M., Rehfeld, K., Scroxton, N., and Veiga-Pires, C.: Evaluating model outputs using integrated global speleothem records of climate change since the last glacial, Climate of the Past Discussions, pp. 1–53, https://doi.org/10.5194/cp-2019-25, 2019.

Comas-Bru, L., Rehfeld, K., Roesch, C., Amirnezhad-Mozhdehi, S., Harrison, S. P., Atsawawaranunt, K., Masood Ahmad, S., Ait Brahim, Y., Baker, A., Bosomworth, M., Breitenbach, S. F., Burstyn, Y., Columbu, A., and Deininger, M.: SISALv2: A comprehensive speleothem isotope database with multiple age-depth, Earth System Science Data Discussions, 225, 1–47, https://doi.org/10.5194/essd-2020-39, 2020.

Dansgaard, W.: Stable isotopes in precipitation, Tellus, 16, 436–468, https://doi.org/10.3402/tellusa.v16i4.8993, 1964.

Dee, S., Emile-Geay, J., Evans, M. N., Allam, A., Steig, E. J., and Thompson, D. M.: PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems, Journal of Advances in Modeling Earth Systems, 7, 1220–1247, https://doi.org/10.1002/2015MS000447, 2015.

Gordon, C., Cooper, C., Senior, C. A., Banks, H. T., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., and Wood, R. A.: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, Climate Dynamics, 16, 147–168, https://doi.org/https://doi.org/10.1007/s003820050010, 2000.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, https://doi.org/10.1017/CBO9781107415324, 2013.

McCrea, J. M.: On the isotopic chemistry of carbonates and a paleotemperature scale, The Journal of Chemical Physics, 18, 849–857, https://doi.org/10.1063/1.1747785, 1950.

McDermott, F., Mattey, D. P., and Hawkesworth, C.: Centennial-scale holocene climate variability revealed by a high-resolution speleothem δ18O record from SW Ireland, Science, 294, 1328–1331, https://doi.org/10.1126/science.1063678, 2001.

Parker, S. E., Harrison, S. P., Comas-Bru, L., Kaushal, N., LeGrande, A. N., and Werner, M.: A data-model approach to interpreting speleothem oxygen isotope records from monsoon regions on orbital timescales, Climate of the Past Discussions, 2020, 1–30, https://doi.org/10.5194/cp-2020-78, 2020.

Rehfeld, K. and Kurths, J.: Similarity estimators for irregular and age-uncertain time series, Climate of the Past, pp. 107–122, https://doi.org/10.5194/cp-10-107-2014, 2014.

Rehfeld, K., Marwan, N., Heitzig, J., and Kurths, J.: Comparison of correlation analysis techniques for irregularly sampled time series, Nonlinear Processes in Geophysics, pp. 389–404, https://doi.org/10.5194/npg-18-389-2011, 2011.

Schmidt, G. A., Annan, J. D., Bartlein, P. J., Cook, B. I., Guilyardi, E., Hargreaves, J. C., Harrison, S. P., Kageyama, M., Legrande, A. N., Konecky, B., Lovejoy, S., Mann, M. E., Masson-Delmotte, V., Risi, C., Thompson, D., Timmermann, A., and Yiou, P.: Using palaeo-climate comparisons to constrain future projections in CMIP5, Climate of the Past, 10, 221–250, https://doi.org/10.5194/cp-10-221-2014, 2014.

Scholz, D. and Hoffmann, D.: 230Th/U-dating of fossil corals and speleothems, Quat. Sci. J, 57, 129, 2008.

Sjolte, J., Sturm, C., Adolphi, F., Vinther, B. M., Werner, M., Lohmann, G., and Muscheler, R.: Solar and volcanic forcing of North Atlantic climate inferred from a process-based reconstruction, Climate of the Past, 14, 1179–1194, https://doi.org/10.5194/cp-14-1179-2018, 2018.

Tindall, J. C., Valdes, P. J., and Sime, L. C.: Stable water isotopes in HadCM3: Isotopic signature of El Niño-Southern Oscillation and the tropical amount effect, Journal of Geophysical Research Atmospheres, 114, 1–12, https://doi.org/10.1029/2008JD010825, 2009.

Urey, H. C.: Oxygen Isotopes in Nature and in the Laboratory, Science, 108, 489–496, 1948.

Valdes, P. J., Armstrong, E., Badger, M. P. S., Bradshaw, C. D., Bragg, F., Crucifix, M., Davies-Barnard, T., Day, J., Farnsworth, A., Gordon, C., Hopcroft, P. O., Kennedy, A. T., Lord, N. S., Lunt, D. J., Marzocchi, A., Parry, L. M., Pope, V., Roberts, W. H. G., Stone, E. J., Tourte, G. J. L., and Williams, J. H. T.: The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0, Geoscientific Model Development, 10, 3715–3743, https://doi.org/10.5194/gmd-10-3715-2017, 2017.

Vinther, B., Jones, P., Briffa, K., Clausen, H., Andersen, K., Dahl-Jensen, D., and Johnsen, S.: Climatic signals in multiple highly resolved stable isotope records from Greenland, Quaternary Science Reviews, 29, 522 – 538, https://doi.org/https://doi.org/10.1016/j.quascirev.2009.11.002, URL `http://www.sciencedirect.com/science/article/pii/S0277379109003655`, 2010.

Wackerbarth, A.: Supplementary Material, pp. 90–92, 2012.

Zhang, X., Sun, Z., Guan, H., Zhang, X., Wu, H., and Huang, Y.: GCM simulations of stable isotopes in the water cycle in comparison with GNIP observations over east asia, Acta Meteorologica Sinica, 26, 420–437, https://doi.org/10.1007/s13351-012-0403-x, 2012.