Evaluating model outputs using integrated global speleothem records of climate change since the last glacial

Laia Comas-Bru¹², Sandy P. Harrison¹, Martin Werner³, Kira Rehfeld⁴, Nick Scroxton⁵, Cristina Veiga-Pires⁶, and SISAL working group members*

¹School of Archaeology, Geography & Environmental Sciences, Reading University, Whiteknights, Reading, RG6 6AH, UK
²UCD School of Earth Sciences, University College Dublin, Belfield, Dublin 4, Ireland
³Division Climate Science – Paleoclimate Dynamics, Climate Science division, Helmholtz Centre for Polar and Marine Research, Alfred Wegener Institute, Bussestr. 24, 27570 Bremerhaven, Germany
⁴Institute of Environmental Physics, Ruprecht-Karls-Universität Heidelberg, Im Neuenheimer Feld 229, 69120 Heidelberg, Germany
⁵Department of Geosciences, University of Massachusetts Amherst, 611 North Pleasant Street, 01003-9297 Amherst, MA, USA
⁶Universidade do Algarve, Marine and Environmental Research Centre – CIMA, Campus de Gambelas, 8005-139 Faro, Portugal
*A full list of authors and their affiliations appears at the end of the paper.

Correspondence: Laia Comas-Bru (l.comasbru@reading.ac.uk)

Received: 19 February 2019 – Discussion started: 4 March 2019
Revised: 17 June 2019 – Accepted: 1 July 2019 – Published: 9 August 2019

Abstract. Although quantitative isotope data from speleothems has been used to evaluate isotope-enabled model simulations, currently no consensus exists regarding the most appropriate methodology through which to achieve this. A number of modelling groups will be running isotope-enabled palaeoclimate simulations in the framework of the Coupled Model Intercomparison Project Phase 6, so it is timely to evaluate different approaches to using the speleothem data for data–model comparisons. Here, we illustrate this using 456 globally distributed speleothem δ¹⁸O records from an updated version of the Speleothem Isotopes Synthesis and Analysis (SISAL) database and palaeoclimate simulations generated using the ECHAM5-wiso isotope-enabled atmospheric circulation model. We show that the SISAL records reproduce the first-order spatial patterns of isotopic variability in the modern day, strongly supporting the application of this dataset for evaluating model-derived isotopic trends. Our analyses provide a protocol for using speleothem isotope data for model evaluation, including screening the observations to take into account the impact of speleothem mineralogy on δ¹⁸O values, the optimum period for the modern observational baseline and the selection of an appropriate time window for creating means of the isotope data for palaeo-time-slices.

1 Introduction

Earth system models (ESMs) are routinely used to project the consequences of current and future anthropogenic forcing of climate, and the impacts of these projected changes on environmental services (e.g. Christensen et al., 2013;
ESMs are routinely evaluated using modern and historical climate data. However, the range of climate variability experienced during the period for which we have reliable historic climate observations is small, much smaller than the amplitude of changes projected for the 21st century. Radically different climate states in the geologic past provide an opportunity to test the performance of ESMs in response to very large changes in forcing, changes that in some cases are as large as the expected change in forcing at the end of the 21st century (Braconnot et al., 2012). The use of “out-of-sample” testing (Schmidt et al., 2014) is now part of the evaluation procedure of the Coupled Model Intercomparison Project (CMIP). Several palaeoclimate simulations are being run by the Palaeoclimate Modelling Intercomparison Project (PMIP) as part of the sixth phase of CMIP (CMIP6-PMIP4), including simulations of the Last Millennium (LM, 850–1850 CE, past1000), mid-Holocene (MH, ca. 6000 yr BP, midHolocene) Last Glacial Maximum (LMG, ca. 21 000 yr BP, lgm), the Last Interglacial (LIG, ca. 127 000 yr BP, lig127k) and the mid-Pleistocene Warm Period (mPWP, ca. 3.2 Myr BP, midPliocene-eoi400) (Kageyama et al., 2017).

Although these CMIP6-PMIP4 time periods were selected because they represent a range of different climate states, the choice also reflects the fact that global syntheses of palaeoenvironmental and palaeoclimate observations exist across them, thereby providing the opportunity for model benchmarking (Kageyama et al., 2017). However, both the geographic coverage and temporal coverage of the different types of data are uneven. Ice core records are confined to polar and high-altitude regions and provide regionally to globally integrated signals of forcings and climatic responses. Marine records provide a relatively comprehensive coverage of the ocean state for the LGM, but low rates of sedimentation mean they are less informative about the more recent past (Hessler et al., 2014). Lake records provide qualitative information of terrestrial hydroclimate, but the most comprehensive source of quantitative climate information over the continents is based on statistical calibration of pollen records (see for example Bartlein et al., 2011). However, pollen preservation requires the long-term accumulation of sediments under anoxic conditions and is consequently limited in semi-arid, arid and highly dynamic wet regions such as in the tropics.

Oxygen isotope records (δ18O) from speleothems, secondary carbonate deposits that form in caves from water that percolates through carbonate bedrock (Hendy and Wilson, 1968; Atkinson et al., 1978; Fairchild and Baker, 2012), provide an alternative source of information about past terrestrial climates. Although there are hydroclimatic limits on the growth of speleothems, their distribution is largely constrained by the existence of suitable geological formations and they are found growing under a wide range of climate conditions, from extremely cold climates in Siberia (Vaks et al., 2013) to arid regions of Australia (Treble et al., 2017). Therefore, speleothems have the potential to provide information about past terrestrial climates in regions for which we do not have (and are unlikely to have) information from pollen. As is the case with pollen, where quantitative climate reconstructions must be obtained through statistical or forward modelling approaches (Bartlein et al., 2011), the interpretation of speleothem isotope records in terms of climate variables is in some cases not straightforward (Lachniet, 2009; Fairchild and Baker, 2012). However, some ESMs now use water isotopes as tracers for the diagnosis of hydroclimate (Schmidt et al., 2007; Tindall et al., 2009; Werner et al., 2016), and this opens up the possibility of using speleothem isotope measurements directly for comparison with model outputs. At least six modelling groups are planning isotope-enabled palaeoclimate simulations as part of CMIP6-PMIP4.

As with other model evaluation studies, much of the diagnosis of isotope-enabled ESMs has focused on modern day conditions (e.g. Joussaume et al., 1984; Hoffmann et al., 1998, 2000; Jouzel et al., 2000; Noone and Simmons, 2002; Schmidt et al., 2007; Roche, 2013; Xi, 2014; Risi et al., 2016; Hu et al., 2018). However, isotope-enabled models have also been used in a palaeoclimate context (e.g. Schmidt et al., 2007; LeGrande and Schmidt, 2008, 2009; Langebroek et al., 2011; Caley and Roche, 2013; Caley et al., 2014; Jasechko et al., 2015; Werner et al., 2016; Zhu et al., 2017). The evaluation of these simulations has often focused on isotope records from polar ice cores and from marine environments. Where use has been made of speleothem records, the comparison has generally been based on a relatively small number of the available records. Furthermore, all of the comparisons make use of an empirically derived correction for the temperature-dependent calcite–water oxygen isotope fractionation at the time of speleothem formation that is based on synthetic carbonates (Kim and O’Neil, 1997). This fractionation is generally poorly constrained (McDermott, 2004; Fairchild and Baker, 2012), does not account for any kinetic fractionation at the time of deposition and is not suitable for aragonite samples. Thus, using a single standard correction and not screening records for mineralogy introduces uncertainty into the data–model comparisons.

SISAL (Speleothem Isotopes Synthesis and Analysis), an international working group under the auspices of the Past Global Changes (PAGES) project (http://pastglobalchanges.org/sisal, last access: 2 August 2019), is an initiative to provide a reliable, well-documented and comprehensive synthesis of isotope records from speleothems worldwide (Comas-Bru and Harrison, 2019). The first version of the SISAL database (SISALv1: Atsawawaranunt et al., 2018a, b) included 381 speleothem-based isotope records and metadata to facilitate quality control and record selection. A major motivation for the SISAL database was to provide a tool for the benchmarking of palaeoclimate simulations using isotope-enabled models.
In this paper, we examine a number of issues that need to be addressed in order to use speleothem data, specifically data from the SISAL database, for model evaluation in the palaeoclimate context and make recommendations about robust approaches that should be used for model evaluation in CMIP6-PMIP4. We focus particularly on interpretation issues that could be overlooked in using speleothem records and we show the strengths and limitations of different comparison techniques. We use the MH and LGM time periods, partly because the midHolocene and lgm experiments are the “entry cards” for the CMIP6-PMIP4 simulations and partly because these are the PMIP time periods with the best coverage of speleothem records. We use an updated version of the SISAL database (SISAL v1b: Atsawawaranunt et al., 2019) and simulations made with the ECHAM5-wiso isotope-enabled atmospheric circulation model (Werner et al., 2011) to explore the various issues in making data-model comparisons. Our goal is not to evaluate the ECHAM5-wiso simulations but rather to use them to illustrate generic issues in data-model comparison with speleothem isotope data.

Section 2 introduces the data and the methods used in this study. Section 2.1 introduces the isotope-enabled model simulations for the modern (1958–2013), the midHolocene and the lgm experiments, explains the methods used to calculate weighted simulated $\delta^{18}O$ values, and provides information about the construction of time slices. Section 2.2 presents the modern observed $\delta^{18}O$ in precipitation ($\delta^{18}O_p$) used. Section 2.3 introduces the speleothem isotope data from the SISAL database and explains the rationale for screening records. Section 3 describes the results of the analyses, specifically the spatio-temporal coverage of the SISAL records (Sect. 3.1), the representation of modern conditions (Sect. 3.2), anomaly-mode time-slice comparisons (Sect. 3.3) and the comparison of $\delta^{18}O$ gradients in absolute values along spatial transects to test whether the model accurately records latitudinal variations in $\delta^{18}O$ across time periods (Sect. 3.4). Section 4 provides a protocol for using speleothem isotope records for data-model comparisons and Sect. 5 summarizes our main conclusions.

2 Methods

2.1 Model simulations

ECHAM5-wiso (Werner et al., 2011; Werner, 2019) is the isotope-enabled version of the ECHAM5 Atmosphere Global Circulation Model (Roeckner et al., 2003, 2006; Hagemann et al., 2006). The water cycle in ECHAM5 contains formulations for evapotranspiration of terrestrial water, evaporation of ocean water, and the formation of large-scale and convective clouds. Vapour, liquid and frozen water are transported independently within the atmospheric advection scheme. The stable water isotope module in ECHAM5 computes the isotopic signal of different water masses through the entire water cycle, including in precipitation and soil water.

ECHAM5-wiso was run for 1958–2013 using an implicit nudging technique to constrain simulated fields of surface pressure, temperature, divergence and vorticity to the corresponding ERA-40 and ERA-Interim reanalysis fields (Butzin et al., 2014). The midHolocene simulation (Wackerbarth et al., 2012) was forced by orbital parameters and greenhouse gas concentrations appropriate to 6 ka following the PMIP3 protocol (https://pmip3.lsce.ipsl.fr, last access: 2 August 2019). The control simulation has modern values for the orbital parameters and greenhouse gas (GHG) concentrations (Wackerbarth et al., 2012). The change in sea surface temperatures (SSTs) and sea ice cover between 6 ka and the pre-industrial period were calculated from 50-year averages from each interval extracted from a transient Holocene simulation performed with the fully coupled ocean–atmosphere Community Climate System Model (CCSM3; Collins et al., 2006). The anomalies were then added to the observed modern SST and sea ice cover data to force the midHolocene simulation (Wackerbarth et al., 2012). For the lgm experiment (Werner et al., 2018), orbital parameters, GHG concentrations, land-sea distribution, and ice sheet height and extent followed the PMIP3 guidelines. Climatological monthly sea ice coverage and SST changes were prescribed from the GLAMAP dataset (Paul and Schäfer-Neth, 2003). A uniform glacial enrichment of sea surface water and sea ice of $+1\%\delta^{18}O$ and $+8\%\deltaD$ on top of the present-day isotopic composition of surface seawater was applied. For the ocean surface state of the corresponding control simulation, monthly climatological SST and sea ice cover for the period 1979–1999 were prescribed. All the ECHAM5-wiso simulations were run at T106 horizontal grid resolution (approx. $1.1^\circ \times 1.1^\circ$) with 31 vertical levels. The midHolocene and lgm experiments were run for 12 and 22 years, respectively, and the last 10 (midHolocene) and 20 (lgm) years were used to construct the anomalies. Model anomalies for the MH and the LGM were calculated as the differences between the averaged midHolocene or lgm simulation and its respective control simulation. We also calculated the anomaly between lgm and midHolocene, taking account of the difference between their control simulations in the following way: $(\text{lgm} - \text{lgm}_{\text{PI}}) - (\text{midHolocene} - \text{midHolocene}_{\text{PI}})$.

At best, the speleothem isotopic signal will be an average of the precipitation $\delta^{18}O$ ($\delta^{18}O_p$) signals weighted towards those months when precipitation is greatest (Yonge et al., 1985). However, the signal is transmitted via the karst system and is therefore modulated by storage in the soil, recharge rates, mixing in the subsurface, and varying residence times ranging from hours to years (e.g. Breitenbach et al., 2015; Riechelmann et al., 2017). These factors could all exacerbate differences between observations and simulations. We investigated whether weighting the simulated $\delta^{18}O$ signals by soil moisture or recharge amount provided a better global comparison than weighting by precipitation amount by calculating three indices: (i) $\delta^{18}O_p$ weighted according to monthly precipitation amount ($w\delta^{18}O_p$), (ii) $\delta^{18}O_p$ weighted accord-
ing to the potential recharge amount calculated as precipitation minus evaporation \((P - E)\) for months where \(P - E > 0\) (\(w\delta^{18}O_{\text{recharge}}\)), and (iii) soil water \(\delta^{18}O\) weighted according to soil moisture amount (\(w\delta^{18}O_{\text{sw}}\)). To investigate the impact of transit time on the comparisons, we smoothed the simulated \(w\delta^{18}O\) using a range of smoothing from 1 to 16 years. Finally, we investigated whether differences in elevation between the model grid and speleothem records had an influence on the quality of the data–model comparisons by applying an elevational correction of \(-2.5\% \text{ km}^{-1}\) (Lachniet, 2009) to the simulated \(w\delta^{18}O\).

2.2 Modern observations

We use two sources of modern isotope data for assessment purposes: (i) \(\delta^{18}O_p\) measurements from the Global Network of Isotopes in Precipitation (GNIP) database (IAEA/WMO, 2018) and (ii) a gridded dataset of global water isotopes from the Online Isotopes in Precipitation Calculator (OIPC: Bowen and Revenaugh, 2003; Bowen, 2018).

The GNIP database (IAEA/WMO, 2018) provides raw monthly \(\delta^{18}O_p\) values for some part of the interval March 1960 to August 2017 for 977 stations. Individual stations have data for different periods of time and there are gaps in most individual records; only two stations have continuous data for over 50 years and both are in Europe (Valentina Observatory, Ireland, and Hohe-Warte, Vienna, Austria). Most GNIP stations are more than 0.5° away from the SISAL cave sites, precluding a direct global comparison between GNIP and SISAL records. However, the GNIP data can be used to examine simulated interannual variability. Annual \(w\delta^{18}O_p\) averages were calculated from GNIP stations with enough months of data to account for more than 80% of the annual precipitation and 5 or more years of data. Annual \(w\delta^{18}O_p\) data were extracted from the ECHAM5-wiso simulations at the location of the GNIP stations for the years for which GNIP data are available at each station. We exclude GNIP stations from coastal locations that are not land in the ECHAM5-wiso simulation. This dual screening results in only 450 of the 977 GNIP stations being used for comparisons. Boxplots are calculated with the standard deviation of annual \(w\delta^{18}O_p\) data.

The OIPC dataset provides a gridded long-term (1960–2017) global record of modern \(w\delta^{18}O_p\), based on combining data from 348 GNIP stations covering part or all the period 1960–2014 (IAEA/WMO, 2017) and other \(w\delta^{18}O_p\) records from the Water Isotopes Database (Waterisotopes Database, 2017). The OIPC data can be used to evaluate modern spatial patterns in both the SISAL records and the simulations.

2.3 Speleothem isotope data

We use an updated SISAL database (SISALv1b: Atsawawarunut et al., 2019), which provides revised versions of 45 records from SISALv1 and includes 60 new records (Table 1). SISALv1b has isotope records from 455 speleothems from 211 cave sites distributed worldwide. Because the isotopic fractionation between water and CaCO\(_3\) differs between calcite and aragonite, we use calcite speleothems, aragonite speleothems where the correction to calcite values was made by the original authors and speleothems with uncorrected aragonite mineralogy. We exclude speleothems where some samples are calcite and some aragonite (mixed mineralogy speleothems) and speleothems with unknown mineralogy. As a result of this screening, we use 407 speleothem records from 193 cave sites for comparisons. However, the number of speleothem records covering specific periods (i.e. modern, MH, LGM) is considerably lower.

Recent data suggest that many calcite speleothems are precipitated out of isotopic equilibrium with waters (Daëron et al., 2019). Therefore, we have converted speleothem calcite data to their drip-water equivalent using an empirical speleothem-based fractionation factor that accounts for any kinetic fractionation that may arise in the precipitation of calcite speleothems in caves (Tremaine et al., 2011):

\[
\delta^{18}O_{\text{dripw-SMOW}} = \delta^{18}O_{\text{calcite-SMOW}} - \left(\frac{16.1 \cdot 1000}{T} - 24.6\right) \quad (T \text{ in } K).
\]

We use the fractionation factor from Grossman and Ku (1986) as formulated in Lachniet (2015) to convert aragonite speleothems to their drip-water equivalent:

\[
\delta^{18}O_{\text{dripw-SMOW}} = \delta^{18}O_{\text{aragonite-SMOW}} - \left(\frac{18.34 \cdot 1000}{T} - 31.954\right) \quad (T \text{ in } K).
\]

We use the V-PDB to V-SMOW conversion from Coplen et al. (1983) as in Sharp (2007):

\[
\delta^{18}O_{\text{SMOW}} = 1.03092 \cdot \delta^{18}O_{\text{PDB}} + 30.92.
\]

We have used mean annual surface air temperature from CRU-TS4.01 (Harris et al., 2014) for the OIPC comparison and ECHAM5-wiso simulated mean annual temperature for the SISAL-model comparison as a surrogate for modern and past cave air temperature (Moore and Sullivan, 1997). There are uncertainties in this conversion because several factors are unknown, e.g. cave temperature and \(p\)CO\(_2\) of soil.

We compare the modern temporal variability in the SISAL records with ECHAM5-wiso by extracting simulated \(w\delta^{18}O_p\) at the cave site location for all the years for which there are speleothem isotope samples within the period 1958–2013. The speleothem isotope ages were rounded to exact calendar years for this comparison.

Data–model comparisons are generally made by comparing (1) anomalies between a palaeoclimate simulation and a control period with (2) data anomalies with respect to a modern baseline. There is no agreed standard defining the interval
Table 1. List of speleothem records that have been added to SISALv1 (Atsawawaranunt et al., 2018a, b) to produce SISALv1b (Atsawawaranunt et al., 2019) sorted alphabetically by site name. Elevation is in metres above sea level (m a.s.l.), latitude in degrees north and longitude in degrees east.

<table>
<thead>
<tr>
<th>Site name</th>
<th>Elev.</th>
<th>Lat.</th>
<th>Long.</th>
<th>Entity name</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arch cave</td>
<td>660</td>
<td>50.55</td>
<td>−127.07</td>
<td>DM05-01</td>
<td>Marshall et al. (2009)</td>
</tr>
<tr>
<td>Beatus cave</td>
<td>875</td>
<td>46.38</td>
<td>7.49</td>
<td>EXC3, EXC4</td>
<td>Boch et al. (2011)</td>
</tr>
<tr>
<td>Bribin cave</td>
<td>500</td>
<td>−8.05</td>
<td>110.633</td>
<td>JB2</td>
<td>Hartmann et al. (2013)</td>
</tr>
<tr>
<td>Cesare Battisti cave</td>
<td>1880</td>
<td>46.08</td>
<td>11.02</td>
<td>CB25, CB39, CB47</td>
<td>Johnston et al. (2018)</td>
</tr>
<tr>
<td>Chan Hol cave</td>
<td>−8.5</td>
<td>20.16</td>
<td>−89.0925</td>
<td>CH-7</td>
<td>Stinnesbeck et al. (2017)</td>
</tr>
<tr>
<td>Chen Ha cave</td>
<td>550</td>
<td>16.67</td>
<td>−89.863</td>
<td>CH04-02</td>
<td>Pollock et al. (2016)</td>
</tr>
<tr>
<td>Cold Water cave</td>
<td>356</td>
<td>43.4768</td>
<td>−91.9755</td>
<td>CW-1s, CW-2s, CW-3l</td>
<td>Denniston et al. (1999)</td>
</tr>
<tr>
<td>Devil’s Icebox cave</td>
<td>250</td>
<td>38.15</td>
<td>−92.05</td>
<td>DIB-1, DIB-2</td>
<td>Denniston et al. (2007b)</td>
</tr>
<tr>
<td>Dos Anas cave</td>
<td>120</td>
<td>22.38</td>
<td>−83.975</td>
<td>CG</td>
<td>Fensterer et al. (2010, 2012)</td>
</tr>
<tr>
<td>El Condor cave</td>
<td>860</td>
<td>−5.93</td>
<td>77.3</td>
<td>ELC_composite</td>
<td>Cheng et al. (2013)</td>
</tr>
<tr>
<td>Frasassi cave system – Grotta Grande del Vento</td>
<td>257</td>
<td>43.4008</td>
<td>12.9619</td>
<td>FR16</td>
<td>Vanghi et al. (2018)</td>
</tr>
<tr>
<td>Goshute cave</td>
<td>2000</td>
<td>40.0333</td>
<td>−114.7833</td>
<td>GC_2, GC_3</td>
<td>Denniston et al. (2007a)</td>
</tr>
<tr>
<td>Harrison’s cave</td>
<td>300</td>
<td>13.2</td>
<td>59.6</td>
<td>HC-1</td>
<td>Mangini et al. (2007), Mickler et al. (2006, 2004)</td>
</tr>
<tr>
<td>Jaraguá cave</td>
<td>570</td>
<td>−21.0833</td>
<td>−56.853</td>
<td>JAR4, JAR1, JAR_composite</td>
<td>Novello et al. (2017, 2018)</td>
</tr>
<tr>
<td>Karaca cave</td>
<td>1536</td>
<td>40.5443</td>
<td>39.4029</td>
<td>K1</td>
<td>Rowe et al. (2012)</td>
</tr>
<tr>
<td>Klaus Cramer cave</td>
<td>1964</td>
<td>47.26</td>
<td>9.52</td>
<td>KC1</td>
<td>Boch et al. (2011)</td>
</tr>
<tr>
<td>KNI-51 cave</td>
<td>100</td>
<td>−15.18</td>
<td>128.37</td>
<td>KNI-51-A1, KNI-51-P</td>
<td>Denniston et al. (2013)</td>
</tr>
<tr>
<td>Korallengrottan cave</td>
<td>570</td>
<td>64.88</td>
<td>14.15</td>
<td>K1</td>
<td>Sundqvist et al. (2007)</td>
</tr>
<tr>
<td>Lianhua</td>
<td>455</td>
<td>29.48</td>
<td>109.53</td>
<td>A1</td>
<td>Cosford et al. (2008a)</td>
</tr>
<tr>
<td>Lynds cave</td>
<td>300</td>
<td>−41.58</td>
<td>146.25</td>
<td>Lynds_BCD</td>
<td>Xia et al. (2001)</td>
</tr>
<tr>
<td>Mawmluh cave</td>
<td>1160</td>
<td>25.2622</td>
<td>91.8817</td>
<td>MAW-0201</td>
<td>Myers et al. (2015)</td>
</tr>
<tr>
<td>McLean’s cave</td>
<td>300</td>
<td>38.07</td>
<td>−120.42</td>
<td>ME2</td>
<td>Oster et al. (2014)</td>
</tr>
<tr>
<td>Minnetonka cave</td>
<td>2347</td>
<td>56.5833</td>
<td>119.655</td>
<td>MC08-1</td>
<td>Landeen et al. (2013)</td>
</tr>
<tr>
<td>Moondyne cave</td>
<td>100</td>
<td>−34.27</td>
<td>146.12</td>
<td>MND-S1</td>
<td>Treble et al. (2003, 2005), Fischer and Treble (2008), Nagra et al. (2017)</td>
</tr>
<tr>
<td>Paraiso cave</td>
<td>60</td>
<td>−4.0667</td>
<td>−55.455</td>
<td>Paraiso composite</td>
<td>Wang et al. (2017)</td>
</tr>
<tr>
<td>Pequin cave</td>
<td>650</td>
<td>32.58</td>
<td>35.19</td>
<td>PEK_composite, PEK 6, PEK 9, PEK 10</td>
<td>Bar-Matthews et al. (2003)</td>
</tr>
<tr>
<td>Piani Eterni karst system</td>
<td>1893</td>
<td>46.16</td>
<td>11.99</td>
<td>MN1, GG1, IS1</td>
<td>Columbus et al. (2018)</td>
</tr>
<tr>
<td>Poloeva cave</td>
<td>390</td>
<td>44.7144</td>
<td>21.7469</td>
<td>PP10</td>
<td>Constantin et al. (2007)</td>
</tr>
<tr>
<td>São Bernardo cave</td>
<td>631</td>
<td>−13.81</td>
<td>−46.35</td>
<td>SBE3</td>
<td>Novello et al. (2018)</td>
</tr>
<tr>
<td>São Matheus cave</td>
<td>631</td>
<td>−13.81</td>
<td>−46.35</td>
<td>SM5</td>
<td>Novello et al. (2018)</td>
</tr>
<tr>
<td>Shatuca cave</td>
<td>1960</td>
<td>−5.7</td>
<td>−77.9</td>
<td>Sha-2, Sha-3, Sha-composite</td>
<td>Bustamante et al. (2016)</td>
</tr>
<tr>
<td>Sofular cave</td>
<td>440</td>
<td>41.42</td>
<td>31.93</td>
<td>So-17A, So-2</td>
<td>Badertscher et al. (2011), Fleitmann et al. (2009), Göktürk et al. (2011)</td>
</tr>
<tr>
<td>Soylegrotta cave</td>
<td>280</td>
<td>66</td>
<td>14</td>
<td>SG93</td>
<td>Lauritzen and Lundberg (1999)</td>
</tr>
<tr>
<td>Tangga cave</td>
<td>600</td>
<td>−0.36</td>
<td>100.76</td>
<td>TA12-2</td>
<td>Wurtzel et al. (2018)</td>
</tr>
<tr>
<td>Uluu-Too cave</td>
<td>1490</td>
<td>40.4</td>
<td>72.35</td>
<td>Uluu2</td>
<td>Wolff et al. (2017)</td>
</tr>
<tr>
<td>White moon cave</td>
<td>170</td>
<td>37</td>
<td>−122.183</td>
<td>WMCI</td>
<td>Oster et al. (2017)</td>
</tr>
<tr>
<td>Xiangshui cave</td>
<td>380</td>
<td>25.25</td>
<td>110.92</td>
<td>X3</td>
<td>Cosford et al. (2008b)</td>
</tr>
<tr>
<td>Xibalba cave</td>
<td>350</td>
<td>16.5</td>
<td>−89</td>
<td>GU-Xi-1</td>
<td>Winter et al. (2015)</td>
</tr>
<tr>
<td>Yaoba Don cave</td>
<td>420</td>
<td>21.8</td>
<td>109.83</td>
<td>YB</td>
<td>Cosford et al. (2008b)</td>
</tr>
</tbody>
</table>

used as a modern baseline for palaeoclimate reconstructions. Some studies have used modern observational datasets which cover a specific and limited period of time and some use the late 20th century as a reference. We investigate the appropriate choice of modern baseline for the speleothem records by comparing the interval centred on 1850 CE with alternative intervals covering the late 20th century, specifically 1961–1990 and 1850–1990 CE, and we assess the impact of these choices on both mean $\delta^{18}O$ values and the number of records available for comparison. The MH time slice was defined as 6000 ± 500 yr BP (where present is 1950 CE) and the LGM time slice as 21 000 ± 1000 yr BP, following the conventional definitions of these intervals used in the construction of other benchmark palaeoclimate datasets (e.g. MARGO Project Members, 2009; Bartlein et al., 2011). However, we also examined the impact of using shorter intervals for each time slice.

We use the published age–depth models for each speleothem record. There is no information about the temporal uncertainties on individual isotope samples for most of the...
records in SISALv1b. This precludes a general assessment of the impact of temporal uncertainties on data–model comparisons. Nevertheless, we assess these impacts for the LGM for two records (entity BT-2 from Botuverá cave: Cruz et al., 2005; and entity SSC01 from Gunung-buda cave: Partin et al., 2007) for which new age–depth models have been prepared using COPRA (Breitenbach et al., 2012). We created 1000-member ensembles of the age–depth relationship using the original author’s choice of radiometric dates and pchip (piecewise cubic hermite interpolating polynomial) interpolation. Isotope ratio means were calculated using time windows of increasing width (±100 to ±1500 years) around 21 kyr BP for the original age–depth model, the COPRA median age model and all ensemble members. All COPRA-based uncertainties have been projected to the chronological axes.

To explore the use of absolute isotope data for model evaluation, we extracted absolute data for two transects illustrating key features of the geographic isotope patterns during the modern, MH and LGM periods. Each transect follows the great circle line between two locations. The span of each regional transect varies to maximize the number of SISAL records included. We extracted model outputs for the same transects at 1.12° steps to match the model grid size and using the model land–sea mask to remove ocean grid cells. Comparisons are made between the SISAL mean δ18O value and the simulated wδ18O values averaged within the latitudinal or longitudinal range defined for each transect.

The presence or absence of speleothems in the temperate zone has long been interpreted as a direct indication of an interstadial or stadial climate state (Gordon et al., 1989; Kashiwaya et al., 1991; Baker et al., 1993), while in dry regions speleothem growth indicates a pluvial climate (Vaks et al., 2006) and in episodically cold regions responds to the absence of permafrost (Atkinson et al., 1978; Vaks et al., 2013). Speleothem distribution through time approximates an exponential curve in many regions around the world (e.g. Ayliffe et al., 1998; Jo et al., 2014; Scroxton et al., 2016). This relationship suggests that the natural attrition of stalagmites is independent of the age of the specimens and approximately constant through time, despite potential complications from erosion, climatic changes and sampling bias. The underlying exponential curve can, therefore, be thought of as a prediction of the number of expected stalagmites given the existing population. Intervals when climate conditions were more or less favourable to speleothem growth can then be identified from changes in the population size by subtracting this underlying exponential curve (Scroxton et al., 2016). We apply this approach at a global level to the unscreened SISAL data by counting the number of individual caves with stalagmite growth during every 1000-year period from 500 kyr BP to the present. Growth was indicated by a stable isotope sample at any point in each 1000-year bin, giving 3866 data points distributed in 500 bins. We use cave numbers, rather than the number of individual speleothems, to minimize the risk of over-sampled caves influencing the results. Random resampling (100 000) of the 3866 data points was used to derive 95 % and 5 % confidence intervals. The number of speleothems cannot be reliably predicted by a continuous distribution when numbers are low, so we do not consider intervals prior to 266 kyr BP – the most recent interval with less than four records.

3 Results

3.1 Spatio-temporal coverage of speleothem records

There are many regions of the world where the absence of carbonate lithologies means that there will never be speleothem records (Fig. 1a). Nevertheless, SISALv1b represents a substantial improvement in spatial coverage compared to SISALv1, particularly for Australasia and Central and North America (Fig. 1a, Table 1), and the sampling for regions such as Europe and China is quite dense. Thus, SISALv1b provides a sufficient coverage to allow the data to be used for model evaluation. The temporal distribution of records is uneven, with only ca. 40 at 21 kyr increasing to > 100 records at 6 kyr and > 110 for the last 1000 yr (Fig. 1b). A pronounced regional bias exists towards Europe during the Holocene. Regional coverage is relatively even during the LGM, except for Africa, which is under-represented throughout (< 4 % of total). Nevertheless, there is enough coverage to facilitate data–model comparisons for the MH and LGM for most regions of the world.

The global occurrence of speleothems through time approximates an exponential distribution (Fig. 2a). Anomalously high numbers of speleothems are found in the last 12 kyr, between 128 and 112 kyr BP and during interglacials MIS 1 and 5e (and the early glacial MIS 5d). There are fewer speleothems than expected between 73 and 63 kyr BP and during MIS 2 (Fig. 2b). These deviations could arise from sampling biases, but it is unlikely that such biases would lead to differences between the tropics and temperate regions. Differences between curves constructed for both tropical and temperate regions (Fig. 2c) suggest that, at least for the last 130 ka, deviations from expected stalagmite growth in the extra-tropics correspond to variability on glacial and interglacial scales. Thus, the speleothem data indicate similar climatic sensitivity, even at a global level, to that demonstrated for sub-continental and regional scales by earlier authors, despite their use of much smaller numbers and far less precise age data than in the SISAL dataset.

3.2 How well do the speleothem records represent modern δ18O in precipitation?

The first-order spatial patterns shown by the SISAL speleothem records during the modern period (1960–2017; n = 87) are in overall agreement with the OIPC dataset of interpolated wδ18Op (R2 = 0.76), with more negative values
Figure 1. Spatio-temporal distribution of SISALv1b database. (a) Spatial distribution of speleothem records. Filled circles are sites used in this study (SISALv1 in blue; SISALv1b in green). Triangles are SISAL sites that do not pass the screening described in Sect. 2.3 and/or do not cover the time periods used here (modern, MH and LGM). The background carbonate lithology is that of the World Karst Aquifer Mapping (WOKAM) project (Chen et al., 2017). (b) Temporal distribution of speleothem records according to regions. The non-overlapping bins span 1000 years and start in 1950 CE. Regions have been defined as follows: Oceania (−60°<lat<0°; 90°<long<180°); Asia (0°<lat<60°; 60°<long<130°); Middle East (7.6°<lat<50°; 26°<long<59°); Africa (−45°<lat<36.1°; −30°<long<60°; with records in the Middle East region removed); Europe (36.7°<lat<75°; −30°<long<30°; plus Gibraltar and Siberian sites); South America (−60°<lat<8°; −150°<long<−30°); North and Central America (8.1°<lat<60°; −150°<long<−50°).

Figure 2. Distribution of the number of single caves with speleothem growth through time. (a) Number of single caves with growth over the last 500 000 yr BP (where present is 1950 CE) in 1000-year bins (solid line), bootstrapped estimate of uncertainty (shading between 5% and 95% percentiles) and fitted exponential distribution (darker solid line). Horizontal bars denote previous interglacials. (b, c) Same as (a) but with the fitted exponential distribution subtracted to highlight anomalies from the expected number of caves over the last 300 kyr BP. Horizontal bars indicate periods with significantly greater (dark grey) or fewer (light grey) numbers of caves with speleothem growth than expected. Green indicates the full global dataset; blue and red indicate temperate and tropical subdivisions respectively.
at higher latitudes and in more continental climates (Fig. 3a). The fact that the speleothem records reflect the $\delta^{18}\text{O}_\text{p}$ patterns in modern precipitation confirms at a global scale the findings of McDermott et al. (2011) for the continental scale in Europe. There are no systematic biases between OIPC and SISAL data at different latitudes (Fig. 3b). However, low latitude sites tend to show more positive $\delta^{18}\text{O}$ values than simulated $w\delta^{18}\text{O}_\text{p}$, whereas sites from middle to high latitudes tend to be more negative (Fig. 3c, d). The discrepancies between the SISAL data and the observations or simulations may be due to cave-specific factors (such as a preferred seasonality of recharge, e.g. Bar-Matthews et al., 1996, or non-equilibrium fractionation processes during speleothem deposition, e.g. Ersek et al., 2018), by complex soil–atmosphere interactions affecting evapotranspiration (e.g. Denniston et al., 1999) and thus the isotopic signal of the effective recharge (Baker et al., 2019), or uncertainties in the isotope fractionation factors with respect to temperature (Fig. S1 in the Supplement) amongst others (e.g. Hartmann and Baker, 2017). However, the overall level of agreement suggests that the SISAL data provide a good representation of the impacts of modern hydroclimatic processes.

Comparison of the SISAL records with $\delta^{18}\text{O}_\text{p}$ weighted according to the potential recharge amount or with $\delta^{18}\text{O}_{\text{sw}}$ weighted to the moisture amount does not significantly improve the data–model comparison (Fig. S2). The best relationship is obtained with soil water $\delta^{18}\text{O}$ weighted according to soil moisture amount ($w\delta^{18}\text{O}_{\text{sw}}$; $R^2 = 0.76$). However, smoothing the simulated $w\delta^{18}\text{O}$ records on a sample-to-sample basis to account for multi-year transit times in the karst environment produces a slightly better geographic agreement with the SISAL records (Fig. S3). Accounting for differences between the model grid cell and cave elevations does not yield any overall improvement in the global correlations.

Simulated inter-annual variability is less than that shown in the GNIP data (Fig. 4). Although there are missing values for the GNIP station data, we have also removed these intervals from the simulations, so incomplete backward comparison is unlikely to explain the difference between the observed and simulated inter-annual variability. Our results are consistent with the general tendency of climate models to underestimate the sensitivity of extreme precipitation to temperature variability or trends (Flato et al., 2014). ECHAM5 is known to underestimate inter-annual variability in regions where precipitation is dominantly convective (i.e. the tropics), as well as in summer in extra-tropical regions (e.g. in southern Europe) because convective precipitation operates on small spatial scales and has a large random component, even for a given large-scale atmospheric state (Eden et al., 2012). The inter-annual variability of the modern speleothem records is lower than both the simulated and the GNIP data, reflecting the impact of karst and in-cave processes that effectively act as a low-pass filter on the signal recorded during speleothem growth (Baker et al., 2013). Thus, smoothing the simulated $\delta^{18}\text{O}_\text{p}$ signal produces a better match to the SISAL records: application of a smoothing window of $>6\text{yr}$ to simulated $w\delta^{18}\text{O}_\text{p}$ produces a good match (95 % confidence) with the inter-annual variability shown by the speleothems (Fig. 4). This result indicates that global data–model comparisons using speleothem records should focus on quasi-decadal or longer timescales. However, the temporal smoothing caused by karst processes varies from site to site; where transmission from the surface to the cave can be shown to be rapid, individual speleothems may preserve annual or even sub-annual signals.

3.3 Anomaly-mode time-slice comparisons

The selection of a modern or pre-industrial base period is a first step in reconstructing speleothem $\delta^{18}\text{O}$ anomalies for comparisons with simulated changes in specific model experiments. There are 76 speleothem records from 62 sites that cover the pre-industrial interval (PI) 1850$\pm15\text{CE}$, commonly used as a reference in model experiments. However, using this short interval as the base period for comparisons with MH or LGM simulations would result in the reconstruction of anomalies for only 21 records for the MH and only 7 records for the LGM – which are the number of speleothem records with isotope samples in both the base period and either the MH or LGM (Table 2). There is no significant difference in the mean $\delta^{18}\text{O}$ values for this pre-industrial period and the modern $\delta^{18}\text{O}$ values ($R^2 = 0.96$; Fig. S4). Using an extended modern baseline (1850–1990 CE) increases the data uncertainties by only $\pm0.5\%$ but raises the number of MH records for which MH–modern anomalies can be calculated to 36 entities from 32 sites around the world. There is also an improvement in the number of LGM sites for which it is possible to calculate anomalies, from 7 to 13 entities at 12 sites. Although longer base periods have been used for data–model comparisons, for example the last 1000 years (e.g. Werner et al., 2016), this would increase the uncertainties in the observations without substantially increasing the number of records for which it would be possible to calculate anomalies, particularly for the LGM (Table 2). We, therefore, recommend the use of the interval 1850–1990 CE as the baseline for calculation of $\delta^{18}\text{O}$ anomalies from the speleothem records.

A relatively good agreement exists between the sign of the simulated and observed $\delta^{18}\text{O}$ changes at the MH and LGM: 94 % of the MH entities and 84 % of the LGM entities show changes in the same direction after allowing for an uncertainty of $\pm0.5\%$ (Fig. 5a, b). However, the magnitude of the changes is larger in the SISAL records than the simulations. The MH–modern speleothem anomalies range from $-3.60\%e$ to $1.29\%e$ (mean $\pm SD$: $-0.50 \pm 1.01\%e$), but the simulated anomalies only range from $-0.49\%e$ to $0.28\%e$ (mean $\pm SD$: $0.00 \pm 0.32\%e$). Observed anomalies are 4–20 times larger than simulated anomalies in the Asian monsoon region, and in individual sites in North and South
Figure 3. Comparison of SISAL data with observational and simulated $\delta^{18}O_p$ for the modern period. (a) Comparison between SISAL $\delta^{18}O$ averages ($\%e$; V-SMOW) for the period 1960–2017 CE with OIPC data ($\%e$; V-SMOW). (b) Scatterplot of SISAL modern $\delta^{18}O$ averages as in (a) versus $\delta^{18}O_p$ extracted from OIPC (i.e. background map in a) at the location of each cave site. (c) Same as (a) with simulated $\delta^{18}O_p$ data for the period 1958–2013 in the background. (d) Scatterplot of SISAL modern $\delta^{18}O$ as in (e) versus the simulated $\delta^{18}O_p$ for the period 1958–2013 CE. Dashed lines in (b) and (d) represent the 1:1 line. All SISAL isotope data have been converted to their drip-water equivalent, following the approach described in Sect. 3.2. Mean annual air surface temperature from CRU-TS4.01 (Harris et al., 2014) and mean annual simulated ECHAM5-wiso air surface temperature were used as surrogates for cave temperatures in the OIPC and ECHAM5-wiso comparison, respectively. See Sect. 2.3 for details on data extraction and conversion.

Table 2. Number of SISALv1b speleothem records available for key time periods. Mid-Holocene (MH): 6 ± 0.5 kyr BP; Last Glacial Maximum (LGM): 21 ± 1 kyr BP. The term “kyr BP” refers to thousand years before present, where present is 1950 CE.

<table>
<thead>
<tr>
<th>Time period</th>
<th>Number of speleothems (entities) and cave sites in both periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modern (1961–1990 CE)</td>
<td>73 entities (59 sites)</td>
</tr>
<tr>
<td>PI (1835–1865 CE)</td>
<td>76 entities (62 sites)</td>
</tr>
<tr>
<td>Extended PI (1850–1990 CE)</td>
<td>100 entities (81 sites)</td>
</tr>
<tr>
<td>MH and PI</td>
<td>21 entities (20 sites)</td>
</tr>
<tr>
<td>MH and extended PI</td>
<td>36 entities (32 sites)</td>
</tr>
<tr>
<td>MH and Last Millennium (LM, 850–1850 CE)</td>
<td>51 entities (41 sites)</td>
</tr>
<tr>
<td>LGM and PI</td>
<td>7 entities (7 sites)</td>
</tr>
<tr>
<td>LGM and extended PI</td>
<td>13 entities (12 sites)</td>
</tr>
<tr>
<td>LGM and Last Millennium (LM, 850–1850 CE)</td>
<td>14 entities (12 sites)</td>
</tr>
<tr>
<td>LGM and MH</td>
<td>22 entities (18 sites)</td>
</tr>
</tbody>
</table>

America and Uzbekistan (Fig. 5a). The data–model mismatch is smallest in Europe, with a mean data–model offset of $-0.13 \pm 0.42 \%e$ ($n = 9$ entities from 7 sites). A two-tailed Student $t$ test shows that most of the simulated MH values are not significantly different from those of the present (at 95% confidence). This may reflect the fact that the midHolocene simulation was only run for 10 years but is also consistent with previous studies which show that climate models substantially underestimate the magnitude of MH changes (Harrison et al., 2014), particularly in monsoon regions (e.g. Perez-Sanz et al., 2014).
The simulated changes in δ18O at the LGM are much larger than those simulated for the MH and are significant (at 95% confidence) over much of the globe. There is no regionally coherent pattern in the observed LGM anomalies because of the limited number of speleothems that grew continuously from the LGM to the present. However, the sign of the observed changes is coherent with the simulated change in δ18O for 11 of the 13 records (Fig. 5b). The magnitude of the LGM anomalies differs by less than 1‰ between model and data in two-thirds of the locations. A strong offset is found in the two records from Sofular Cave, which are ca. 5.5‰ more negative than the simulated δ18O. This offset may be related to the glacial changes in the Black Sea region, which are not well represented in the lgm simulation. Thus, although overall the comparison with the speleothem records suggests that the simulated changes in hydroclimate are reasonable, the simulated changes in the Middle East differ from observations.

An alternative approach to examine the realism of simulated changes is to compare the LGM and MH periods directly, which improves the number of records for which anomalies can be calculated (Fig. 5c; n = 22). However, the pattern of change is similar to the LGM–modern anomalies. The simulated and observed direction of change is coherent...
at 86% of the locations with an offset smaller than 1‰ occurring in 12 sites, and again the largest discrepancy is Sofular Cave. Thus, in this particular example, a direct comparison of the LGM–MH anomalies does not provide additional insight to the comparison of LGM–modern anomalies. Nevertheless, such an approach might be useful for other time periods (e.g. comparison of early versus mid-Holocene) when there are likely to be many more speleothem records available.

Age uncertainties inherent to the speleothem samples representing the LGM could partially explain the LGM data–model mismatches. A global assessment of the impact of time-window width on the MH and LGM anomalies shows that reducing the window width from ±500 to ±200 years in the MH has little impact on the average values (Fig. S5) but reduces the inter-sample variability and produces a better match to the simulated anomalies. A similar analysis for the LGM (Fig. S6) suggests that a window width of ±500 years (rather than ±1000 years) would be the most appropriate choice for comparisons of this interval. The number of SISAL sites available for such comparisons is not affected. However, analyses of the relative error of the isotope anomalies calculated at individual sites for different LGM window widths (Fig. 6) show a clear increase in all relative error components as window size decreases for BT-2 (Botuverá cave; Fig. 6a; Cruz et al., 2005) but no clear changes in the relative error terms for SSC01 (Gunung-buda cave; Fig. 6b; Partin et al., 2007). These results suggest that, with an LGM window width of ±1000 years, the relative contribution of age uncertainty to the anomaly uncertainty is small (Fig. 6). Thus, although it is clear that it would be useful to propagate age uncertainties for individual sites, changing the conventional definitions of the MH and LGM time slices in deriving speleothem anomalies does not seem warranted at this stage.

3.4 Analysis of spatial gradients

The number of sites available in SISALv1b means that quantitative data–model comparisons using the traditional anomaly approach are limited in scope. Approaches based on comparing trends in absolute δ¹⁸O values could provide a way of increasing the number of observations and an alternative way to evaluate the simulations. Comparison of trends places less weight on anomalous sites and allows large-scale systematic similarities and dissimilarities between model and observations to be revealed. We illustrate this approach using spatial gradients across Asia and across Europe and showing how they differ between the modern, MH and LGM periods, although such an approach could also be used for temporal trends.

The first-order spatial gradient in observed δ¹⁸O during the modern period is broadly captured by the model in both examples (Figs. 7, 8), with the largest offsets found mainly for high-altitude sites. There is a fundamental change in the latitudinal gradient across Asia during the MH (Fig. 7). In this period, the gradient observed in the data is clearly not reproduced by the model, which systematically simulates higher δ¹⁸Op values between 20 and 35° N, suggesting that the model underestimates the insolation-driven intensification of the hydrological cycle in monsoon regions during this period. The limited number of speleothem records available between

Figure 6. LGM period definitions and their impact on SISAL δ¹⁸O mean estimate uncertainty. The impact of the window definition and age uncertainty is explored for two entities: (a) entity BT-2 from Botuverá cave (Cruz et al., 2005) and (b) entity SSC01 from Gunung-buda cave (Partin et al., 2007). The relative error is defined as 2 standard deviation for the original age model and the COPRA median; and the upper minus lower 95% quantiles for the COPRA median uncertainty as well as the COPRA ensemble spread of standard deviations. Black solid lines give the relative error of the mean isotope estimate for the LGM for the original age model and grey solid lines give the estimate based on the COPRA median age model. The pink dotted line shows the uncertainty of the COPRA median estimate, and the green dashed line the average relative error estimate across the 1000-member COPRA ensemble. For both speleothems, relatively stable error estimates are found for window sizes larger than ±750 years, whereas the relative error increases towards smaller window sizes.
25 and 35° N for the LGM agree with the simulated $\delta^{18}O$ gradient. The longitudinal gradient across Europe (Fig. 8) does not change substantially in the MH compared to modern times. However, the model simulates $w\delta^{18}O$ values $\sim 2$‰ lower than observed in low-altitude sites in south central Europe between 0 and 15° E during the MH. This suggests that the model may be underestimating the role of atmospheric circulation (i.e. weaker westerlies) during this period, an aspect of the climate system that models have difficulty simulating (Mauri et al., 2014). The large latitudinal variability of simulated values eastwards of $\sim 5°$ E during the LGM is consistent with a larger spread in the observations, despite the limited number of data available. These examples show the potential to use trends in absolute values for model evaluation and diagnosis.

4 Protocol for data–model comparison using speleothem data

Our analyses illustrate a number of possible approaches for using speleothem isotope data for model evaluation. The discontinuous nature of most speleothem records means that the number of sites available for conventional anomaly-mode comparisons is potentially limited. To some extent this is mitigated by the fact that differences between the modern and pre-industrial isotope values are small, permitting the calculation of anomalies using a longer baseline interval (1850–1990 CE). The use of smaller intervals of time in calculating MH or LGM anomalies (Figs. S5 and 6) does not have a significant impact either on the mean values or the number of records provided the interval is $> \pm 300$ yr for the MH and $> \pm 500$ yr for the LGM. Although the use of shorter intervals is possible, we recommend using the conventional definitions of each time slice to facilitate comparison with other benchmark datasets. Although patterns in the isotope anomalies can provide a qualitative assessment of model performance, site-specific factors could lead to large differences from the simulations at individual locations. Improved spatial coverage would allow such sites to be identified and screened out before making quantitative comparisons of observed and simulated anomalies. Although there are only a limited number of records that cover both the modern baseline period and the MH (or the modern baseline period and the LGM), there are many more records that provide information about one or other of these periods. The examination of spatial gradients in absolute $\delta^{18}O$ provides one way of exploiting this larger data coverage. Even when an offset between the observed and simulated $\delta^{18}O$ exists, comparing the trends along such gradients is possible. Thus, both absolute values and anomalies of the isotope data for data–model comparison are useful.

Screening of published speleothem isotope data is essential to produce meaningful data–model comparisons. The SISAL database facilitates screening for mineralogy, which has a substantial effect on isotope values because of differences in water-carbonate fractionation factors for aragonite or calcite that are more pronounced at lower temperatures (Fig. S1).

Based on the limited number of records available at the LGM, speleothem age uncertainties have only a limited impact on mean isotope values, and propagation of such uncertainties as well as any model uncertainties would nevertheless substantially improve the robustness of data–model comparisons.

Based on our analyses, we therefore recommend that model evaluation using speleothem records should do the following:

1. filter speleothem records with respect to their mineralogy and use the appropriate equilibrium fractionation factor: Tremaine et al. (2011) for converting isotope data from either calcite or aragonite-corrected-to-calcite samples to their drip water equivalent, and Grossman and Ku (1986) as reformulated by Lachniet (2015) for converting isotope data from aragonite samples;
2. use the interval between 1850 and 1990 as the reference period for speleothem isotope records;
3. use speleothem isotope data averaged for the intervals 6000±500 yr (21 000±1000 yr) for comparability with other MH (LGM) palaeoclimate benchmark datasets;
4. use speleothem isotope data averaged for the interval 6000±200 yr or 21 000±500 yr for best approximation of midHolocene and lgm experiments;
5. use absolute values only to assess data–model first order spatial patterns;
6. focus on multi-decadal to millennial timescales if using transient simulations for data–model comparisons.

5 Conclusions

Speleothem records show the same first-order spatial patterns as are available in the Global Network of Isotopes in Precipitation (GNIP) data and are therefore a good reflection of the $\delta^{18}O$ patterns in modern precipitation. This observation suggests that stalagmites are a rich source of information for model evaluation. However, the inter-annual variability in the modern speleothem records is considerably reduced compared to the simulations, which in turn show less inter-annual variability than the GNIP observations. The low variability shown by the SISAL records – most likely from the low-pass filter effectively applied to the speleothem record by the karst system – precludes the use of this database for global studies focused on timescales shorter than quasi-decadal.

Using the traditional anomaly approach to data–model comparisons, there is consistency between the sign of observed and simulated changes in both the MH and the LGM. However, the ECHAM5-wiso model underestimates
Figure 7. Latitudinal isotopic transect for Asia during the (a) modern (1958–2013 CE), (b) Mid-Holocene (MH; 6 ± 0.5 kyr BP) and (c) Last Glacial Maximum (LGM; 20 ± 1 kyr BP) periods. Background maps at the top of each panel show the simulated $w_\delta^{18}O_{p}$ from ECHAM5-wiso. Bottom plots in each panel show the simulated $w_\delta^{18}O_{p}$ data extracted for each transect: black circles and grey whiskers are mean ±2 standard deviation of the data extracted along longitudinal sections in between the two great circle lines shown in solid black lines in the top maps. The red line is the median of the extracted data. All data were extracted at steps of 1.12° to coincide with the average model grid size. These bottom panels also show SISAL $\delta^{18}O$: circles for low-elevation sites, < 1000 m a.s.l.; triangles for high-elevation sites, > 1000 m a.s.l.

The changes in $\delta^{18}O$ between time periods compared to the speleothem records (i.e. the amplitude of modelled $\delta^{18}O$ changes is lower). Thus, these kinds of comparisons should only focus on the large-scale spatial patterns. Based on the available SISAL data, the use of smaller time windows than the conventional definitions for each time slice does not have a strong impact on the mean values and could be used to reduce the uncertainties associated with the palaeodata. However, this would preclude comparisons with existing benchmark datasets that use the conventional windows for the MH and LGM time slices.

Only a limited number of speleothem records are continuous over long periods of time and the need to convert these to anomalies with respect to modern times is a drawback. The limited number of records covering the LGM make the comparisons for this period particularly challenging. Nevertheless, continued expansion of the SISAL database will increase its usefulness for model evaluation in future. Furthermore, we have shown that alternative approaches using absolute values could help examine spatial trends and diagnose systematic offsets.

Mismatches between simulations and observations can reflect the issues with the experimental design, problems with the model or uncertainties in the observations (Harrison et al., 2015). The failure to include changes in atmospheric dust loading, for example, has been put forward as an explanation of data–model mismatches in both the MH and the LGM (e.g. Hopcroft et al., 2015; Messori et al., 2019). Missing processes and feedbacks, such as climate-induced vegetation or land-surface changes, could also contribute to mismatches (e.g. Yoshimori et al., 2009; Swann et al., 2014). Uncertainties caused by the specific structure of the model or assigned model parameter values could also contribute to data–model mismatches (Qian et al., 2016). Ultimately, there needs to be
an assessment of the contribution of all these factors to data–
model mismatches, but here we have only focused on poten-
tial uncertainties associated with the speleothem data. Our
initial analyses suggest age uncertainty contributes little to
the uncertainties in the estimates of LGM speleothem isotope
values. However, it is still important to propagate dating un-
certainties for data–model comparison. Site-specific controls
may have a much larger effect on the $\delta^{18}$O record recorded
by individual speleothems and thus may contribute signifi-
cantly to uncertainties in local or regional signals. We have
not screened for regionally anomalous records that could be
influencing the results of our analyses, but this should cer-
tainly be done. Despite these challenges, SISAL appears to
be an extremely useful tool for describing past patterns of
variability, highlighting its potential for evaluating CMIP6-
PMIP4 experiments.

Comparisons with speleothem data can be seen as a com-
plement to model evaluation using other types of palaeoen-
vironmental data and palaeoclimatic reconstructions (see
for example MARGO Project Members, 2009; Harrison et
al., 2014). They are particularly useful because they pro-
vide insights into how well state-of-the-art models reproduce
the hydrological cycle and atmospheric circulation patterns.
The ability to reproduce past observations provides addi-
tional confidence in the ability of climate models to simu-
late large climate changes, such as those expected by the end
of the 21st century (Braconnot et al., 2012; Schmidt et al.,
2014). However, mismatches between model simulations and
palaeo-observations are also useful because they can help

Figure 8. Longitudinal isotopic transect for Europe during the (a) modern (1958–2013 CE), (b) Mid-Holocene (MH; 6 ± 0.5 kyr BP) and (c) Last Glacial Maximum (LGM; 20 ± 1 kyr BP) periods. Details as in caption of Fig. 7.
to pinpoint issues that may need to be addressed in developing improved models or in better experimental protocols (Kageyama et al., 2018), providing that these mismatches do not arise because of misunderstanding or misinterpretation of the observations themselves. By providing a protocol for using speleothem data for data–model comparisons that accounts for uncertainties in the observations, we anticipate that at least such causes of data–model mismatches will be minimized.

Data availability. The SISAL (Speleothem Isotopes Synthesis and Analy...
drea Borsato (School of Environmental and Life Science, University of Newcastle, 2308 NSW, Australia), Meighan Boyd (Department of Earth Sciences, Royal Holloway University of London, Egham, Surrey TW20 0EX, UK), Chris Brierley (Department of Geography, University College London, WC1E 6BT, UK), Yanjun Cai (State Key Lab of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi’an 710061, China), Stacy Carolin (Institute of Geology, University of Innsbruck, Innrain 52, 6020 Innsbruck, Austria), Hai Cheng (Institute of Global Environmental Change, Xi’an Jiaotong University, China), Silviu Constantin (Emil Racovita Institute of Speleology, str. Frumoasa 31, Bucharest, Romania), Isabelle Coughoud (EDYTEM, UMR 5204 CNRS, Université Savoie Mont Blanc, Université Grenoble Alpes, 73370 Le Bourget du Lac, France), Francisco Cruz (Instituto de Geociências, Universidade de São Paulo, São Paulo, Brazil), Rhawn Denniston (Department of Geography, Cornell College, Mount Vernon, IA, 52314, USA), Virgil Drăgușin (Emil Racovita Institute of Speleology, Str. Frumoasa 31, Bucharest, Romania), Wuhui Duan (Key Laboratory of Cenozoic Geology and Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, 100029, China), Vasile Ersek (Department of Geography and Environmental Sciences, Northumbria University, Newcastle upon Tyne, UK), Martin Finné (Department of Archaeology and Ancient History, Uppsala University, Sweden), Dominik Fleitmann (Department of Archaeology, School of Archaeology, Geography and Environmental Science, Whiteknights, University of Reading RG6 6AB, UK), Jens Fohlmeister (Institute for Earth and Environmental Sciences, University of Potsdam, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Germany), Amy Frappier (Department of Geosciences, Skidmore College, Saratoga Springs, NY 12866, USA), Dominique Genty (Laboratoire des Sciences du Climat et de l’Environnement, CNRS, L’Orme des Merisiers, 91191 Gif-sur-Yvette Cedex, France), Steffen Holzkämper (Department of Physics and Environmental Analysis) is a working group of the Past Global Changes (PAGES) programme. We thank PAGES for their support for this activity. We thank the World Karst Aquifer Mapping project (WOKAM) team for providing us with the karst map presented in Fig. 1a. The authors would like to thank the following data contributors: Petra Bajo, Dominique Blamart, Russell Drysdale, Frank McDermott and Jean Rioult. Laia Comas-Bru and Sandy P. Harri-
Financial support. Financial support for SISAL activities that have lead to this research has been provided by the Past Global Changes (PAGES) programme; the European Geosciences Union (grant no. W2017/413); the Irish Centre for Research in Applied Geosciences (iCRAG); the European Association of Geochemistry (Early Career Ambassadors program 2017); the Quaternary Research Association UK; the Navarino Environmental Observatory, Stockholm University; University College Dublin (grant no. SF1428), Savilllex (UK); John Cantle; Ibn Zohr University, Morocco; the University of Reading; the European Research Council (NERC). Laia Comas-Bru also acknowledges funding from the Portuguese Science Foundation (FCT) through the CIMA research centre project UID/MAR/00350/2013; the Geological Survey Ireland (Short Call 2017; grant number 2017-SC-056) and the Royal Irish Academy’s Charlemont Scholar award 2018. Cristina Veiga-Pires acknowledges support from the Geological Survey Ireland (FCT) through the CIMA research centre project UID/MAR/00350/2013; and the Deutsche Forschungsgemeinschaft (DFG) grant no. RE3994/2-1.

Review statement. This paper was edited by Zhengtang Guo and reviewed by three anonymous referees.

References


Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B., and Zhao, Y.: Evaluation of climate models us-


Schmidt, G. A., LeGrande, A. N., and Hoffmann, G.: Water isotope expressions of intrinsic and forced variability in a cou-


