

Supplementary material of Comas-Bru et al. (manuscript esd-2020-39)

Data Quality Control

1. Pre-upload checks:

Quality Control checks for SISALworkbook_v12 (Comas-Bru and Harrison, 2019):

- Python script for automatic checks and R script for manual checks are available at <https://10.5281/zenodo.3631403>

2. Post-upload checks:

Quality Control of database:

- Julia script for automatic checks of U/Th dating metadata is available at https://github.com/jensfohlmeister/QC_SISALv2_dating_metadata

We also prepared an Excel file with a calculator for converting atomic ratios and their associated uncertainties to activity ratios using a range of decay constants. This is available at <https://10.5281/zenodo.3631443>

Note: The file used to format and submit the speleothem records for the SISAL database (SISAL_workbook_v12.xlsx) is available as supplementary material of Comas-Bru and Harrison, 2019.

AUTOMATIC CHECKS:

Colour code:

Blue (*): print informative message

Purple (**): print warning

Red (***): print warning and terminate

Site table:

- There can only be one site (i.e. one row) (***)
- All drop-down menus must be filled in with the pre-defined options (**)
- Latitude and longitude must be within a valid range (i.e. -90 to 90 for latitude and -180 to 180 for longitude (**))
- "site_name" cannot start/end with a space (**)
- Print coordinates for visual inspection (*)
- "elevation" should not be missing (*)

Entity metadata table:

- This table must be filled in (***)
- Entity names in this table must be unique (***)
- "entity_name" cannot start/end with a space (**)
- All drop-down menus must be filled in with the pre-defined options (**)
- "contact_name" cannot be numeric, it must contain two words (i.e. name and surname) and cannot start/end with a space. (**)

- If "one_and_only" = "yes", "entity_status_info" must be "not applicable" and "entity_status_notes" must be empty. If "one_and_only" = "no", "entity_status_info" cannot be "not applicable" and "entity_status_notes" cannot be empty, "NA", "unknown", "not known" (and their respective variants). (**)
- The format of the fields without dropdown menus should be correct (**). For example, the NOAA/PANGEA URL or DOI of the data cannot be "unknown", "N/A", "not known", etc. or have spaces before/after the text. The URL or DOI must be empty or start with either "10.", "ftp", or "http". (**)
- If "entity_type" = "composite", "drip_type" must be "not applicable" (**)
- If "entity_type" = "composite", request that the person entering the data lists the names of the entities and their references of the records used to construct the composite in the Notes tab. (*)

Sample table:

- This table must be filled in. An empty sample table is only acceptable for entities that are part of a composite. (***)
- If interp_age is missing, the entity is likely to be missing an age model. This is not allowed except for some very special cases. The person entering the data is encouraged to add dummy interp_age values so that all other checks can be performed. (***)
- All drop-down menus must be filled in with the pre-defined options (**)
- If "mineralogy" is not "aragonite" or "mixed", "arag_corr" must be "not applicable". (**)
- If "mineralogy" = "aragonite" or "mixed", "arag_corr" must be something different than "not applicable" (**)
- If "mineralogy" = "unknown", "arag_corr" must be "unknown" (**)
- If "mineralogy" = "calcite", "secondary calcite" or "vaterite", "arag_corr" must be "not applicable" (**)
- If the entity has no age model, "interp_age_uncert_pos", "interp_age_uncert_neg", "age_model_type", "modern_reference", "ann_lam_check" and "dep_rate_check" should be empty (**)
- Check that the uncertainties coexist and that "corr_age" exist if they exist (**)
- "interp_age" must be within the correct age range depending on the chosen "modern_reference" (i.e. it cannot be -70 if "modern_reference" = "BP (1950)", but this would be correct if "modern_reference" = "BCE/CE") (**)
- The depth of the hiatus must be within the range of samples' depths.
- If "interp_age" exists, "interp_age_uncert_pos" and "interp_age_uncert_neg" should not be missing (*).
- If "arag_corr" = "yes" for samples with "mineralogy" = "mixed", an informative warning will be printed asking for details to be added in the Notes tab.
- Check overall length of the entity based on the "sample_depth". Print informative warning if the total length is less than 100mm to ensure that depth units have been entered correctly (i.e. in mm instead of cm).

Dating metadata table:

- If "entity_type" is not "composite", this table should not be empty (***)
- If "entity_type" = "composite", this should have been left empty (**)
- All drop-down menus must be filled in with the pre-defined options taking into account that:
 - o "calib_used" must be filled in when "date_type" = "C14". (**)
 - o "calib_used" must be empty when "date_type" is not "C14". (**)
 - o "decay_constant" must be filled in when "date_type" is of U/Th type (**)

- “decay_constant” must not be filled in when “date_type” is not of U/Th type (**)
- “decay_constant” must be empty when “date_type” is not of U/Th type (**).
- Check that the entity has dating info other than hiatuses and/or not used dates. (**)
- If date_used = "yes" or "unknown", “corr_age”, “corr_age_uncert_pos”, 'corr_age_uncert_neg" and “modern_reference” should be provided (**)
- “min_weight” must be smaller than “max_weight” (**)
- Hiatuses have to be located at depths different than the dates (**)
- “corr_age_uncert_pos” and “corr_age_uncert_neg” must coexist and “corr_age” must exist if they exist.
- “corr_age” must be within the correct age range depending on the chosen “modern_reference” (i.e. it cannot be -70 if “modern_reference” = "BP (1950)", but this would be correct if “modern_reference” = "BCE/CE") (**)
- If there is a “date_type” = "Event; end of laminations" there should also be a “date_type” = "Event; start of laminations" (**)

Lamina vs age table:

- If “Entity_type” = composite, this table should be empty (**)
- If this table is filled in, ensure that all drop-down menus are filled in.
- “lam_age_uncert_pos” and “lam_age_uncert_neg” must coexist and “lam_age” exist if they exist
- “lam_age” must be within the correct age range depending on the chosen “modern_reference” (i.e. it cannot be -70 if “modern_reference” = "BP (1950)", but this would be correct if “modern_reference” = "BCE/CE") (**)

Reference table:

- If a citation appears more than once, its DOI must be the same (**)
- If two DOIs are the same, their linked citation must be the same unless the DOI is “unpublished” (**)
- There should not be repeated citations for one entity (**)
- There should be no spaces before/after the DOI/URL (**)
- “DOI/URL” must either be "unpublished" (e.g. PhDs and unpublished records) or start with "http" or "10".
- A citation cannot be empty, "unknown", "N/A", "not known", etc or have spaces before/after the text.

Checks across tables (table integrity):

- “entity_name” in dating information and sample data must match that on the entity metadata tab (***)
- If the oldest sample is at the lowest “sample_depth”, “depth_ref” must be “from base” (***)
- If the youngest sample is at the lowest “sample_depth”, “depth_ref” must be “from top” (***)
- If the Lamina age vs depth table is not empty, the dating information table must contain a “date_type” = "Event; start of laminations" and a “date_type” = "Event; end of laminations". (**)
- If the dating information table contains an 'Event; start of laminations', the Lamina age vs depth table should exist (**)
- The hiatuses depth in the sample tab must match the corresponding depth in the dating information tab (**)
- There must be at least one reference for each entity listed in the Entity metadata tab (**)

- If the entity is laminated (i.e. lamina age vs depth information and/or date_type = "Event; start of laminations" or "date_type" = "Event; end of laminations" exist) "ann_lam_check" cannot not be "not applicable". This is only allowed for non-laminated speleothems. (**)
- The depth of "Event; start of laminations" must not be 0 if "depth_ref" = "from top" (**)
- The depth of "Event; end of laminations" must not be 0 if "depth_ref" = "from base" (**)
- There should never be consecutive "Event; start of laminations" or "Event; end of laminations" (i.e., they must always alternate) (**)
- If "depth_ref" = "from top", then "date_type" = "Event; start of laminations" cannot exist at "depth_dating" = 0 (**)
- If "depth_ref" = "from base", then "date_type" = "Event; end of laminations" cannot exist at "depth_dating" = 0 (**)
- The youngest date attached to a lamina must be linked to an "Event; end of laminations" in the dating tab. (**)
- The oldest date attached to a lamina must be linked to an "Event; start of laminations" in the dating tab. (**)
- "data_DOI_URL" and "publication_DOI" cannot be identical (**)
- If "modern_reference" in the sample table is "year of chemistry", "year_done" must be filled in for that entity in the Dating information table. (**)
- A "gap" column in the sample data tab can only be filled if "entity_type" = "composite" (**)
- If youngest "corr_age" or "Event; end of laminations" is < -40 BP (1950), print an informative warning saying that it is likely that an "Event; actively forming" is missing (*)

MANUAL CHECKS:

These are based on four plots:

Plot 1: Age-depth model plot with dates, isotopes and hiatuses.

- If dates and isotope ages do not align, it can be indicative of wrong dates/samples depth units, incorrect "modern_reference" or "depth_ref".
- Non-systematic offsets between "interp_age" and "corr_age" may require extra documentation of the isotope chronologies in the Notes tab
- Visual evaluation that the depths of the hiatuses are correct. These should plot at depths with no isotope samples

Plot 2: Difference in interpolated ages between consecutive samples vs sample depth mid-points

- Peaks in this plot that are not marked as hiatuses may be indicative of hiatuses that have not been identified in the workbook.

Plot 3: Time series of d18O data

- Wrongly entered oxygen isotope values may be identified here

Plot 4: Time series of d13C data

- Wrongly entered carbon isotope values may be identified here

ADDITIONAL QUALITY CONTROL OF THE DATING METADATA TABLE APPLIED TO THE DATABASE:

1. Query database and filter all speleothems dated by U/Th.

2. Check if the U/Th metadata available is enough. If not, contact corresponding author or data provider.
 - a) Prepare selected data in a way that (uncorrected) age calculations are possible; flag `sample_ids`, for which no ages can be calculated otherwise calculate ages.

The following table provides an overview for which set of data it is possible to calculate uncorrected ages from (minimum requirements):

#	238U	232Th	230Th	230Th/232Th	230Th/238U	234U/238U
1					x	x
2	x		x			x
3	x	x		x		x

- b) as in a), but this time for (detritus) corrected ages

The following table provides an overview for which set of data it is possible to calculate detritus corrected ages from (minimum requirements):

#	238U	232Th	230Th	230Th/232Th	230Th/238U	234U/238U	lni230Th232Th
1	x	x			x	x	x
2	x	x	x			x	x
3				x	x	x	x
4	x	x		x		x	x
5		x	x		x	x	x

3. Evaluate individual data for correctness (as far as possible, e.g., check if d234U provided instead of activity 234U/238U), including the conversions between activity and ratio
4. Compare calculated corrected and uncorrected ages with those in SISAL database. QC test fails if they do not agree within 1 % and 5 %, respectively.