

Supplement of

SISAL_monv1: a global database of cave monitoring observations

Treble et al.

- 5 The copyright of individual parts of the supplement might differ from the CC BY 4.0 License.

S1 Data Quality Control for Cave Monitoring database

S1.1 Pre-upload checks:

An R script for automatic QC checks of the monitoring workbook. It loads, validates, and visualizes data from each worksheet. The script is available at https://github.com/istvan60/SISAL_monv1/blob/main/QC%20script_v5.3.R

- 10 **Note:** The file used to format and submit the speleothem records for the SISAL monitoring database (SISAL_monitoringdb.xlsx) is available alongside the database in the repository at doi: 10.5158/ARP/29W5J3.

S1.2 Automatic checks

Site table

- Only one row allowed.
- 15 • site_name must not start or end with a space.
- latitude $\in [-90, 90]$; longitude $\in [-180, 180]$.
- elevation must be present.

Reference table

- All non-empty rows share the same site_name.
- 20 • Each citation maps to a single publication_doi.
- Duplicate DOIs link to identical citations (unless “unpublished”).
- No repeated citations or DOIs.
- No spaces around publication_doi; DOI must start with "10", "http://" or be "unpublished".
- No null site_name or citation; table must have at least one row.

25 **Precipitation site metadata table**

- Unique precip_entity_name.
- precip_site_name has no extra spaces.
- Valid precipitation lat/lon ranges.
- precip_elevation and precip_distance_cave_entrance must not be missing.

30 **Precipitation entity metadata table**

- precip_method from allowed list ("IAEA/GNIP", "other (see notes)", "unknown").
- precip_entity_contact is non-numeric, multi-word, without extra spaces.

Precipitation sample table

- precip_accumulation_unit valid (days, hours, minutes, unknown).
- 35 • Start/end date and time fields follow 4-digit year, 1–2 digit month/day, 4-digit hhmm format within ranges.
- Numeric checks for accumulation time, amount, $\delta^{18}\text{O}$ & $\delta^2\text{H}$ measurements and precisions.

Cave entity metadata table

- Unique cave_entity_name.
- Flags (cave_temperature, cave_relative_humidity, cave_pCO₂), frequency, and instrument fields from allowed options (yes, no, unknown, plus frequency/instrument lists).
- cave_entity_contact non-numeric, multi-word, no extra spaces.

Drip entity metadata table

- Unique drip_entity_name.
- entity_id integer 1–2000.
- drip_entity_contact non-numeric, multi-word.
- Geology and rock-age fields from predefined lists.
- Flags (drip_iso, drip_rate, mod_carb) ∈ {yes,no,unknown}.

Drip isotope sample

- drip_iso_accumulation_unit ∈ {days,hours,minutes,seconds,unknown}.
- Date/time checks for start/end components.
- Numeric checks for accumulation time and isotope measurements/precision.

Drip rate sample

- drip_rate_accumulation_unit ∈ {days,hours,minutes,seconds,unknown}.
- Date/time checks for start/end.
- Numeric checks for accumulation time, measurement, precision.

Modulated carbonate sample

- mod_carb_accumulation_unit ∈ {years,months,days,unknown}.
- mod_carb_surface and mod_carb_mineralogy from allowed lists.
- Date/time checks for start/end.
- Numeric checks for accumulation time, $\delta^{18}\text{O}$ / $\delta^2\text{H}$ measurements and precisions.

S1.3 Table integrity checks

- **Cross-table consistency (T15):**
 - site_name present across site, reference, notes, precipitation-site tables.
 - precip_entity_name alignment across site-metadata, entity-metadata, sample tables.
 - cave_entity_name alignment across entity-metadata and sample sheets.
 - drip_entity_name alignment across entity-metadata and sample sheets.
 - At least one reference row exists.

S2 Output reports

S2.1 Text QC summary

- write_vars_to_pdf(): outputs PASS/FAIL summary of each check to text_output_<site>.pdf, coloring failures red.

S2.2 Graphics outputs

- **World map** with site and precipitation locations labeled.
- **Zoomed-in map** $\pm 5^\circ$ around the points.
- **Scatter of precip_amount vs. precip_accumulation_time**, colored by unit.
- **Time series of precip_amount** by entity over date.

80

- **Scatter & regression** of d^2H vs. $d^{18}O$ for precipitation and drip-water, with Global Meteoric Water Line reference, annotated with regression equations and R^2 .
- **Time-series plots** for each sample type (precipitation $\delta^{18}O$, δ^2H ; cave temperature, humidity, pCO_2 ; drip-ISO $\delta^{18}O$, δ^2H ; drip rate; mod_carb $\delta^{18}O$, $\delta^{13}C$), each only if data are present; otherwise blank pages with a “No data available” message.

S3 Manual checks (Analogous to plotting-based review):

- Evaluate text and graphics output for anomalies not caught automatically (incorrect units, missing values, outliers).

85