SoilTemp: towards a global database of soil and near-surface temperatures

Responsible contact :

Jonas J. Lembrechts, Centre of Excellence Plants and Ecosystems (PLECO), University of Antwerp, 2610 Wilrijk, Belgium, jonas.lembrechts@uantwerpen.be

Gewijzigde veldcode

SoilTemp website: soiltemp.weebly.com

Gewijzigde veldcode

Short: *we are looking for in-situ measured microclimate time series from all over the world, optionally with additional biodiversity data. We focus on soil and near-surface temperature recordings from all heights and depths above and below the soil surface, yet also accept associated other microclimatic measurements (e.g., soil moisture, relative humidity…).*
For details on the projects' rationale, check our call for data in Global Change Biology: https://doi.org/10.1111/gcb.15123.

Data submission details

Ultimately, the database will be a collection of datasets with georeferenced **soil and near-surface microclimate time series** (temperature, soil moisture, relative humidity…), **possibly associated with species composition and/or trait data** from the same or a nearby location. The database consists of 4 types of data tables, as discussed below. Datasets must be submitted in the recommended format (also see the associated data submission example file) and be cleaned and checked (i.e. QA/QCed) in order to be included in the database.

1) **Raw microclimate data**

The raw microclimate data over time (hourly, every 30 mins, every 4 hours, etc). Time series need to be at least **one month** long, with a maximum **interval of 4 hours between measurements** (start and end date, as well as temporal resolution (time interval between two consecutive records) needs to be provided in the metadata file: see next section). The database focusses on near-surface (both above- and belowground) data (exact sensor depth/height to be provided in the metadata file: see next section). For **experimental sites**, data from control plots can be submitted, but data from manipulated systems (e.g. open-top chambers, pots) will not be included into SoilTemp.

The table contains the following information:

- **Raw data identifier** (e.g. NU14), a name that matches **EXACTLY** with the name in the metadata file (see next section). Preferably, the LOGGER CODE.

    - When several sensors measure in parallel at one location (e.g. for **TMS-loggers)**, data for each logger can then be submitted as 4 columns with the same LOGGER CODE, with columns labelled with the SENSOR CODE (e.g. T1, T2, T3 and Soil Moisture according to the TMS-definition).
    - If data for multiple sensors for the same logger are submitted in long format (underneath each other), the Raw Data Identifier should reflect this (e.g. Logger_code_Sensor_code, or 245629_T1)
- The (local or UTC) date (with a separate column for year, month and day) and time (in XX:XX, 24h format) of each temperature measurement. This time format is asked for as there are too many different time formats to automatically detect what you used. We thus put the burden on you to unify this!
- The microclimate measurements
- Please replace erroneous measurements by NAs or remove the rows.

Notes:

- This data can also be submitted in .csv or .txt, as long as the column structure is the same as proposed.
- The data can be submitted in two formats:
    - 'long format', with all sensors and loggers underneath each other (see example Excel sheet)
    - Separate files or sheets for each logger. In that case, the file or sheet name should match EXACTLY with the location or logger code in the metadata file.
- Name matching is HUGELY important for us to link your metadata with your raw data.

2) Temperature metadata

A table providing information about the sensor, logger device and measurement location with ONE ROW OF DATA FOR EACH UNIQUE SENSOR. Provide as much information as possible, but note that columns CAN BE LEFT EMPTY if data is not available.

- **Country_code**: the A2 (ISO) country codes for the country where the sensor is installed: https://www.worldatlas.com/aatlas/ctycodes.htm) (e.g. SE for Sweden)
- **Location_code**: a unique identifier for the physical location (e.g. the plot) where the sensor was installed.
- **Logger_code**: a unique identifier for the logger (preferably the ID number that is written on the device itself, but other unique codes work as well).
- **Sensor_code**: a unique identifier for the sensors on a logger (e.g. T1, T2, T3 and Soil Moisture for TMS-sensors)
- **Sensor_height:** The exact sensor depth/height (relative to the soil surface, excluding the vegetation. Negative values for sensors located below the soil surface, positive values for temperatures measured above). Critical for correct analysis of the data. For sensors measuring over a broader range (e.g. the TOMST TMS4 soil moisture sensor), provide here the average depth.
- **Sensor_length:** Total length of sensors measuring over a broader range (e.g. 15 cm for a TOMST TMS4 soil moisture sensor).

Gewijzigde veldcode

- **Raw_data_identifier:** the exact same code as used to identify loggers and sensors in the raw data file (see above). Often the same as the Logger_code
- **Microclimate_measurement:** the variable you measured with this specific sensor, e.g. temperature, soil moisture, relative humidity…
- **Unit:** unit in which this variable is expressed (°C for temperature, % for volumetric soil moisture and relative humidity).
- The exact **latitude** and **longitude** (*either* WGS84, in decimal degrees, with a minus for southern/western hemisphere *or* with your local coordinate system, specifying the **EPSG-code** (4326 for WGS84)) of the measurement location, as well as **GPS-accuracy** (important for microclimate-related questions).
- The **brand** (company) and **type** of the **temperature logger** used (e.g. TOMST, TMS4), as well as the sensor accuracy.
- **Logger_shielding:** If a radiation shield is in place for that specific sensor (e.g. for the air temperature sensor of a TOMST TMS4): yes/no. Home-made shielding for air temperature sensors is also a 'yes'; soil temperature sensors usually has no shielding.
- **Logger_shielding_type:** If the radiation shield is provided by the logger company (e.g. TOMST radiation shields) or not (either bought elsewhere, or homemade).
- **Logger_shielding_notes:** Extra details on the type of shielding if not provided by the company, or on sensor wrapping etc.
- **Habitat_type**: we use the global habitat identification scheme of the IUCN (https://www.iucnredlist.org/resources/habitat-classification-scheme) to synchronize habitat types across studies. See Supplementary Material at the bottom of this file for the full list. Very important, as satellite data cannot always accurately estimate local, in-situ conditions at the moment of sampling.
- The exact **start and end date of the measurements** (i.e. the first and last full day of in-situ temperature measurements), with a separate column for year, month and day.
- The **temporal resolution** or time interval between two consecutive records, in minutes.
- **Time_zone:** either 'UTC' or 'local'
- **Time_difference:** difference in time with UTC. Either 0 if provided in UTC, or a number relative to UTC if measured in local time (e.g. +2).
- Whether **species composition and/or species trait data** is available (yes/no). Also mark 'yes' if available upon request.
- **LAI:** Leaf Area Index estimates from hemispherical pictures. This is the standardized way to provide canopy cover estimates that can be readily compared with other studies.
- **Forest_canopy_cover:** Forest canopy cover estimates using another method than LAI
- The **name** and **email address** of the dataset manager, as well as the **institute** under which flag the data was collected. A comma-separated list of **all other data contributors**.
- **DOI:** Digital Object Identifier of a reference paper or published dataset for the sensor in question, allows data users to cite original work.
- A statement on if the data and/or metadata can be published **open access** (see below under 'Terms of use and data ownership').
- **Optional columns**: Any other information related to (local) vegetation structure, which is from fundamental importance to understand soil and near-surface temperatures (e.g. forest canopy cover, total vegetation cover, depth of moss layer, vegetation height). Also mark the plot size in which these variables are measured. Any other metadata you have of that measurement

Gewijzigde veldcode

location (for example variables related to soil conditions (e.g. soil type, soil moisture), topography (e.g. elevation, slope and aspect measured in situ) and land use practices (e.g. disturbance types, disturbance estimates)). Again, mark the plot size in which these variables are measured. New columns can be added at the end of the file.

3) Raw species data (optional)

A table containing species composition and/or species trait data for each of the measurement locations (or in the vicinity (distance < 1 km)) for which this is available. Data can – as in 1) – be submitted either separately for each georeferenced location, or compiled into one table (plots pasted below each other). The table contains the following information:

- The same LOCATION CODE as in 1).
- A list of all the species observed within that plot.
- Optional additional columns with additional data on each species (e.g. cover percentage, density). Especially cover percentage is greatly appreciated.
- Possible trait data for each species (e.g. size, specific leaf area, phenological stage) can be provided as a separate file (see the sheet 'Species_traits_example').

4) Species metadata (optional)

A table with one row of data for each unique georeferenced location, containing the following information:

- The same unique plot code as in 1).
- The size of the plot (in m^2) in which the species data was obtained.
- The exact longitude and latitude (WGS84, decimal degrees, with a minus for southern/western hemisphere) of the measurement location.
- The date on which the observation was done, with a separate column for year, month and day.
- The species groups that were monitored (e.g. vascular plants, mosses, ground beetles, arbuscular mycorrhizae). If multiple species groups are monitored, submit the raw species data (under 3)) separately for each group.
- Any other metadata relevant for the species or traits not yet provided in 2).

All this data can be sent as a Zip-file via WeTransfer to jonas.lembrechts@uantwerpen.be.

Gewijzigde veldcode

Anybody with suitable datasets is encouraged to contribute data to SoilTemp. Approval of the datasets is done by the project leaders. Upon approval, the dataset contributor(s) will automatically become a SoilTemp-member.

Network rules

Member types

All members are asked to submit and update their personal information and specific interests through the following FORM: (https://forms.gle/emVNknWMB7qd4zNk8). Any changes (e.g. to email address) should be communicated through this form.

The database is maintained by SoilTemp-members, organized at four levels regarding data contribution and decision making:

1) SoilTemp Steering Committee

The SoilTemp SC consists of:

- Caroline Greiser (University of Stockholm, Sweden, ECR)
- David Klinges (University of Florida, USA, ECR)
- Maroof Hamid (University of Kashmir, India, representative Asia)
- Ilya Maclean (University of Exeter, UK)
- Ivan Nijs (University of Antwerp, Belgium)
- Jonas Lembrechts (University of Antwerp, Belgium, representative Europe)
- Jonathan Lenoir (Université de Picardie Jules Verne, France)
- Juha Aalto (University of Helsinki; Finnish Meterological Institute, Finland)
- Julia Kemppinen (University of Oulu, Finland, ECR)
- Liesbeth van den Brink (University of Tübingen, Germany, representative South America)
- Martin Kopecký (Institute of Botany of the Czech Academy of Sciences, Czech Republic)
- Mick Ashcroft (University of Wollongong, Australia, representative Australia)
- Miska Luoto (University of Helsinki, Finland)
- Pieter De Frenne (Ghent University, Belgium)
- Rachel Penczycowski (Washington University in St. Louis, USA, representative North America)

From 2024 onwards, election of the twelve SoilTemp steering committee is done by the members. Elections take place for 4-year renewable terms in January by means of an electronic ballot extending over one month. The candidate with the highest number of votes for each continent is automatically elected (6 continental representatives). From the remaining candidates, the six candidates with the highest number of votes are elected; in case of a tie for the sixth position, all persons with the same number of votes are elected. The SC determines one of its members as chairperson.

In addition, there are rotating (every 2 years) and invited (by the SC members) seats for 3 early career scientists (PhD students, postdocs < 3 year after PhD).

The steering committee is responsible for:

- Management and updating of the database
- Judging project proposals using the database and keeping track of ongoing projects
- Meeting every four months to discuss the general status of the network, its management and scientific outputs.

2) Dataset contributor

(Co-)owner of a dataset submitted to SoilTemp. A *dataset* is a unit of data as submitted from a certain project, with a fixed set of dataset contributors. Datasets are defined by a common methodology and contributor team. They can be small (e.g. one month of data from one sensor) or large (e.g. hundreds of sensors measuring over multiple years) and can have or have not associated biodiversity data. There is no limit to the amount of contributors to one dataset; this is left to the discretion of the contributors yet should clearly be communicated to the SC.

3) Dataset custodian

Each dataset as defined above can have multiple contributors, yet has one dataset custodian. Dataset managers serve as contact and mediators between dataset contributors and the SC and the principal investigators (PI) of a given project submitted to SoilTemp and approved by the SC using this dataset (see below). They are the first contact regarding all questions (practical, scientific, data use, etc.) for the dataset and are responsible to pass on information to their co-contributors when necessary.

4) Other members

Membership of SoilTemp within a given project is possible without contributing datasets. 'Other members' can be co-author on any of the resulting papers if they contribute sufficiently to idea development, data analysis and/or paper writing (decisions on this at the discretion of the project leaders).

Projects

Projects using the full database follow the terms of use as specified below, inviting all contributors as co-author. Projects using SoilTempOpen (once it is published) are exempt of this rule.

Each project using the database should submit a one-page summary of the project idea (including a preliminary title, brief outline, core group of people working with the data and list of datasets that will be used) to the SC, which can then approve or reject the idea based on potential overlap with other projects, rationale and feasibility.

Each project has a Principal Investigator (PI). In addition, each project gets assigned a responsible contact within the Steering Committee, who is in charge of keeping track of the project, evaluating its progress and reporting back to the SC.

PIs are asked to provide (ir)regular updates on their progress to the interested SoilTemp members through the SoilTemp Slack environment. The PI can agree upon a 'supporting committee' of interested SoilTemp-members that are closely involved in the progress and can provide feedback through semi-regular meetings throughout its execution. Call for membership of these supporting committees go out for each project separately, to all members of SoilTemp.

Terms of use and data ownership

Participation in SoilTemp and use of the SoilTemp database are subject to the following conditions:

1) Dataset ownership and use

- The datasets within the SoilTemp database remain the property of the respective dataset contributors. Power of decision regarding a certain dataset remains entirely with them.
- Permission to use a dataset needs to be asked to the data contributor every time a dataset is used for a new analysis (unless stated differently in the FORM mentioned above). New submissions automatically accept data being used in ongoing projects, a list of which is kept in the Slack Environment.
- Dataset contributors can at any stage withdraw their dataset from the database, but then are no longer a SoilTemp member.
- Data contributors mark at submission if their data can be part of SoilTempOpen. If so, their data becomes free for the whole scientific community. SoilTemp will not publish any data without permission of the data contributor(s) through the data custodian.
- Under no circumstances can non-open access data received through the SoilTemp network be circulated to others without permission. Datasets can be put under a moratorium and be added to the open access SoilTemp database as soon as the dataset contributor(s) deem this possible.
- We developed a SoilTemp website (https://soiltemp.weebly.com) linking to a global map (https://microclimate.shinyapps.io/loggerapp/) on which some metadata on each dataset will be published (i.e. coordinates, contact information, number of plots, temporal extent and whether species data is available). Dataset contributors can separately decide to be excluded from the website or the map, impose a moratorium, or ask to blur coordinates. It is the responsibility of the dataset contributor to communicate this to the project leaders (e.g. through the submission form).

2) Publications and co-authorship

- Co-authorship is offered to all dataset contributors for every core publication of the SoilTemp network (unless stated differently in the FORM mentioned above). These core publications include 1) the publication of the call for data (Lembrechts et al. 2020), 2) the SoilTemp maps and 3) the open access SoilTemp database itself (SoilTempOpen). This rule also applies to all papers concerning application of the database that are initiated before publication of the database. Publications using SoilTempOpen or the SoilTemp maps after its publication are not required to include data contributors as co-authors, yet will have to cite respectively the SoilTemp database or the maps.
- IMPORTANT: analyses are now well on their way, so new data submissions might be too late for some of the ongoing paper projects. Submissions before March 31st 2022 will however be guaranteed inclusion in at least publication of SoilTempOpen itself. Submissions after this date will still be included in ongoing analyses as much as possible and will be included in publication of an update of the database at a later stage.
- Each dataset custodian has the responsibility to check with possible other contributors involved in a particular dataset whether or not they should be offered co-authorship as well.
- All persons that will act as co-author are expected to, at least, review and explicitly approve the publication. If a dataset contributor has initially agreed that his/her data can be used, yet does not respond to emails approving the final publication, he will be removed from the author list (yet the dataset remains included).

- Any member of SoilTemp can propose a publication using the SoilTemp database or SoilTemp maps. Until publication of SoilTempOpen, publication proposals (including a preliminary title, brief outline, core group of people working with the data and list of datasets that will be used) should be sent to the SC. They will check for any conflict of interest with ongoing or planned proposals, contact dataset custodians for dataset access and co-authorship, compile the approved data and send them to the proposer. After publication of SoilTempOpen, contacting the SC to verify conflicts of interest is not required, yet strongly encouraged. When contacting the SC, the full database might be provided.
- Any SoilTemp member can express interest in a particular SoilTemp-analysis by 'liking' it in the Slack-environment, and will then be kept in the loop about progress through Slack.

Supplementary material

S1: Habitat classification types

- **1. Forest**
  - 1.1. Forest – Boreal
  - 1.2. Forest - Subarctic
  - 1.3. Forest – Subantarctic
  - 1.4. Forest – Temperate
  - 1.5. Forest – Subtropical/tropical dry
  - 1.6. Forest – Subtropical/tropical moist lowland
  - 1.7. Forest – Subtropical/tropical mangrove vegetation above high tide level
  - 1.8. Forest – Subtropical/tropical swamp
  - 1.9. Forest – Subtropical/tropical moist montane

- **2. Savanna**
  - 2.1. Savanna - Dry
  - 2.2. Savanna - Moist

- **3. Shrubland**
  - 3.1. Shrubland – Subarctic
  - 3.2. Shrubland – Subantarctic
  - 3.3. Shrubland – Boreal
  - 3.4. Shrubland –Temperate
  - 3.5. Shrubland – Subtropical/tropical dry
  - 3.6. Shrubland – Subtropical/tropical moist
  - 3.7. Shrubland – Subtropical/tropical high altitude
  - 3.8. Shrubland – Mediterranean-type shrubby vegetation
  - *3.9. Shrubland - Heathlands*

- **4. Grassland**
  - 4.1. Grassland – Tundra *(alpine/Arctic)*
  - 4.2. Grassland – Subarctic
  - 4.3. Grassland – Subantarctic
  - 4.4. Grassland – Temperate
  - 4.5. Grassland – Subtropical/tropical dry
  - 4.6. Grassland – Subtropical/tropical seasonally wet/flooded
  - 4.7. Grassland – Subtropical/tropical high altitude

- **5. Wetlands (inland)**
  - 5.1. Wetlands (inland) – Permanent rivers/streams/creeks (includes waterfalls)
  - 5.2. Wetlands (inland) – Seasonal/intermittent/irregular rivers/streams/creeks
  - 5.3. Wetlands (inland) – Shrub dominated wetlands
  - 5.4. Wetlands (inland) – Bogs, marshes, swamps, fens, peatlands

- 5.5. Wetlands (inland) – Permanent freshwater lakes (over 8 ha)
- 5.6. Wetlands (inland) – Seasonal/intermittent freshwater lakes (over 8 ha)
- 5.7. Wetlands (inland) – Permanent freshwater marshes/pools (under 8 ha)
- 5.8. Wetlands (inland) – Seasonal/intermittent freshwater marshes/pools (under 8 ha)
- 5.9. Wetlands (inland) – Freshwater springs and oases
- 5.10. Wetlands (inland) – Tundra wetlands (inc. pools and temporary waters from snowmelt)
- 5.11. Wetlands (inland) – Alpine wetlands (inc. temporary waters from snowmelt)
- 5.12. Wetlands (inland) – Geothermal wetlands
- 5.13. Wetlands (inland) – Permanent inland deltas
- 5.14. Wetlands (inland) – Permanent saline, brackish or alkaline lakes
- 5.15. Wetlands (inland) – Seasonal/intermittent saline, brackish or alkaline lakes and flats
- 5.16. Wetlands (inland) – Permanent saline, brackish or alkaline marshes/pools
- 5.17. Wetlands (inland) – Seasonal/intermittent saline, brackish or alkaline marshes/pools
- 5.18. Wetlands (inland) – Karst and other subterranean hydrological systems (inland)

- **6. Rocky Areas (e.g., inland cliffs, mountain peaks)**

- **7. Caves & Subterranean Habitats (non-aquatic)**
  - 7.1. Caves and Subterranean Habitats (non-aquatic) – Caves
  - 7.2. Caves and Subterranean Habitats (non-aquatic) – Other subterranean habitats

- **8. Desert**
  - 8.1. Desert – Hot
  - 8.2. Desert – Temperate
  - 8.3. Desert – Cold

- **9. Marine Neritic**
  - 9.1. Marine Neritic – Pelagic
  - 9.2. Marine Neritic – Subtidal rock and rocky reefs
  - 9.3. Marine Neritic – Subtidal loose rock/pebble/gravel
  - 9.4. Marine Neritic – Subtidal sandy
  - 9.5. Marine Neritic – Subtidal sandy-mud
  - 9.6. Marine Neritic – Subtidal muddy
  - 9.7. Marine Neritic – Macroalgal/kelp
  - 9.8. Marine Neritic – Coral Reef
    - 9.8.1. Outer reef channel
    - 9.8.2. Back slope
    - 9.8.3. Foreslope (outer reef slope)
    - 9.8.4. Lagoon
    - 9.8.5. Inter-reef soft substrate
    - 9.8.6. Inter-reef rubble substrate

- 9.9 Seagrass (Submerged)
- 9.10 Estuaries

- **10 Marine Oceanic**
  - 10.1 Epipelagic (0–200 m)
  - 10.2 Mesopelagic (200–1,000 m)
  - 10.3 Bathypelagic (1,000–4,000 m)
  - 10.4 Abyssopelagic (4,000–6,000 m)

- **11 Marine Deep Ocean Floor (Benthic and Demersal)**
  - 11.1 Continental Slope/Bathyl Zone (200–4,000 m)
    - 11.1.1 Hard Substrate
    - 11.1.2 Soft Substrate
  - 11.2 Abyssal Plain (4,000–6,000 m)
  - 11.3 Abyssal Mountain/Hills (4,000–6,000 m)
  - 11.4 Hadal/Deep Sea Trench (>6,000 m)
  - 11.5 Seamount
  - 11.6 Deep Sea Vents (Rifts/Seeps)

- **12 Marine Intertidal**
  - 12.1 Rocky Shoreline
  - 12.2 Sandy Shoreline and/or Beaches, Sand Bars, Spits, etc.
  - 12.3 Shingle and/or Pebble Shoreline and/or Beaches
  - 12.4 Mud Shoreline and Intertidal Mud Flats
  - 12.5 Salt Marshes (Emergent Grasses)
  - 12.6 Tidepools
  - 12.7 Mangrove Submerged Roots

- **13 Marine Coastal/Supratidal**
  - 13.1 Sea Cliffs and Rocky Offshore Islands
  - 13.2 Coastal Caves/Karst
  - 13.3 Coastal Sand Dunes
  - 13.4 Coastal Brackish/Saline Lagoons/Marine Lakes
  - 13.5 Coastal Freshwater Lakes

- **14 Artificial - Terrestrial**
  - 14.1 Arable Land
  - 14.2 Pastureland
  - 14.3 Plantations
  - 14.4 Rural Gardens
  - 14.5 Urban Areas
  - 14.6 Subtropical/Tropical Heavily Degraded Former Forest
  - *14.7 Roadsides*

- **15 Artificial - Aquatic**
  - 15.1 Water Storage Areas [over 8 ha]
  - 15.2 Ponds [below 8 ha]
  - 15.3 Aquaculture Ponds
  - 15.4 Salt Exploitation Sites
  - 15.5 Excavations (open)
  - 15.6 Wastewater Treatment Areas
  - 15.7 Irrigated Land [includes irrigation channels]
  - 15.8 Seasonally Flooded Agricultural Land
  - 15.9 Canals and Drainage Channels, Ditches
  - 15.10 Karst and Other Subterranean Hydrological Systems [human-made]
  - 15.11 Marine Anthropogenic Structures
  - 15.12 Mariculture Cages
  - 15.13 Mari/Brackish-culture Ponds

- **16 Introduced Vegetation**

- **17 Other**

- **18 Unknown**