Lembrechts, J.J., Aalto, J., Ashcroft, M.B., et.al. (2020), SoilTemp: a global database of near-surface temperature. *Global Change Biology*. Accepted Author Manuscript. doi:10.1111/gcb.15123

***Supplementary material***

# *Details on data and metadata for submission to the SoilTemp-database*

***Responsible contact:*** *Jonas J. Lembrechts, Centre of Excellence Plants and Ecosystems (PLECO), University of Antwerp, Belgium, jonas.lembrechts@uantwerpen.be*

***SoilTemp website: soiltemp.weebly.com***

*We are looking for in-situ measured microclimate time series from all over the world. We accept all soil and near-surface temperature recordings from all heights (up to 2 m) and depths above and below the soil surface. Temperatures need to have been measured for at least one consecutive month, 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 below). 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.*

Ultimately, the database will be a collection of datasets with georeferenced soil and near-surface data. For now, the database consists of 2 types of data tables, as discussed below, with optionally 2 extra tables for associated species composition or trait data from the same or a nearby location. The latter will however be only integrated in a second phase of the SoilTemp-project.

Datasets must be submitted in the suggested 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. We strongly encourage sensor calibration before submission, ideally before employing loggers in the field. The latter can be achieved by putting temperature sensors for a several hours or days in rooms with a constant low (e.g. freezer, fridge) and high (e.g. growth camber, oven) temperature to assess consistent deviations in measurements.

1) Temperature data

The raw temperature data over time (hourly, every 30 mins, every 4 hours, etc.), either separately submitted for each georeferenced location (with the unique plot code as name of the file), or compiled into one table (with the unique plot code in the first column). 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).

The table contains the following information:

- Unique plot code (e.g. SE_JL_NU14), consisting of the two first letters of the country (following the A2 (ISO) country codes: https://www.worldatlas.com/aatlas/ctycodes.htm) in which the data is measured (e.g. SE for Sweden), the first letters of the first and last name of the dataset manager (e.g. JL for Jonas Lembrechts), and then the unique plot code you used in your study (e.g. NU14), all separated by an underscore. Sensors at the same location (e.g. at a different

depth) should be given a unique plot code (e.g. NU14_0, NU14_5 for a sensor at the surface and 5 cm below the surface, respectively).
- For TMS-loggers, air, surface and soil temperature + soil moisture data can be submitted as 4 columns with the same plot code, with columns labelled T1, T2, T3 and Soil Moisture according to the TMS-definition.
- 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.
- The temperature values.
- If soil moisture has been measured continuously at the same time, this data can be added as an additional column.

Note:

- This data can also be submitted in .csv or .txt, as long as the column structure is the same as proposed.

2) Temperature metadata

A table providing information about the logger device used at a given location with one row of data for each unique georeferenced location. Provide as much information as possible, but note that not all columns should be filled out if data is not available.

- The same unique plot code as used in 1).
- The exact longitude and latitude (*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 temperature sensor used (logger_make = the company, logger_model = the specific type, accuracy, notes on wrapping) to allow amongst others inter-sensor comparison.
- 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.
- 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.
- Whether time was recorded in UTC or local time (UTC/Local)
- Whether species composition and/or species trait data is available (yes/no).
- 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 the end of this document for the full list.
- Any other information related to (local) vegetation structure, which is from fundamental importance to understand soil temperatures (e.g. forest canopy cover, total vegetation cover, depth of moss layer, Leaf Area Index, vegetation height). Also mark the plot size in which these variables are measured.

Lembrechts, J.J., Aalto, J., Ashcroft, M.B., et.al. (2020), SoilTemp: a global database of near-surface temperature. *Global Change Biology*. Accepted Author Manuscript. doi:10.1111/gcb.15123

- Any other metadata you have of that measurement 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 with other data types, or if measurements have been done in another unit, can be added at the end of the file.
- A statement on if the data and/or metadata can be published open access (see below under 'Terms of use and data ownership').

3) Species data (optional, will only be cross-linked with the database in a second phase)

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 unique plot code as in 1).
- A list of all the species observed within that plot.
- Possible additional columns with additional data on each species (e.g. cover percentage, density).
- Possible additional columns with trait data for each species (e.g. size, specific leaf area, phenological stage).

4) Species metadata (optional, will only be cross-linked with the database in a second phase)

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).
- 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.

## Terms of use and data ownership

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

Lembrechts, J.J., Aalto, J., Ashcroft, M.B., et.al. (2020), SoilTemp: a global database of near-surface temperature. *Global Change Biology*. Accepted Author Manuscript. doi:10.1111/gcb.15123

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.
- Until a certain dataset is published open access in the SoilTemp database, permission to use it needs to be asked to the data contributor every time a dataset is used for a new analysis. For submissions after November 2019, permission is automatically given for data use in ongoing analyses by: Jonas Lembrechts (University of Antwerp, all SoilTemp analyses), Johan van den Hoogen (ETH Zürich, global soil temperature map), Stef Haesen (KULeuven, microclimate map of Europe) and Joseph Bailey (University of York, geomorphology as driver of soil temperature). For more details on those projects, get in touch with Jonas Lembrechts.
- Dataset contributors can at any stage withdraw their dataset from the database, but then are no longer a SoilTemp member.
- Under no circumstances can non-open access data received through the SoilTemp network be circulated to others without permission.

2) Database publication

- One of the key goals of SoilTemp is to create a dynamic, open access database of soil temperature and associated species and metadata, to which anybody, even non-members, can add their datasets.
- Publication of this SoilTemp database will not happen before the first paper is published.
- Open access publication entails derived products (e.g. monthly averages and bioclimatic variables). Raw temperature data can only be downloaded with permission of the data owner.
- Any dataset contributor can decide upon publication of this database if his/her data can be included in the open access version of the database. SoilTemp will not publish any data without permission of the data contributor.
- Datasets can be put under a moratorium and be added to the open access SoilTemp database as soon as the dataset contributor deems 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). Again, dataset contributors can 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.

Field Code Changed

Field Code Changed

3) Publications and co-authorship

- Co-authorship is offered to all dataset contributors for every core publication of the SoilTemp network. These core publications include 1) the SoilTemp database and 2) the SoilTemp maps (and the papers related to development and refinement of those maps, see above). This rule also applies to all papers concerning application of the database that are initiated before publication of the database. Publications using the SoilTemp database or the SoilTemp maps after its publication are not required to include data contributors as co-authors, yet will have to cite respectively the database or the maps.
- Each dataset manager 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.

Lembrechts, J.J., Aalto, J., Ashcroft, M.B., et.al. (2020), SoilTemp: a global database of near-surface temperature. *Global Change Biology*. Accepted Author Manuscript. doi:10.1111/gcb.15123

- 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 the database, 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 project leaders. They will check for any conflict of interest with ongoing or planned proposals, contact dataset managers for dataset access and co-authorship, compile the approved data and send them to the proposer. After publication of the database, contacting the project leaders to verify conflicts of interest is not required, yet strongly encouraged.

Lembrechts, J.J., Aalto, J., Ashcroft, M.B., et.al. (2020), SoilTemp: a global database of near-surface temperature. *Global Change Biology*. Accepted Author Manuscript. doi:10.1111/gcb.15123

**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**