GRiMeDB data entry template- User's Guide

1. INTRODUCTION

This document provides information and advice about how to enter your data into the "GRiMeDB_data_template.xls" file and submit it for inclusion in the next update of the Global River Methane DataBase (GRiMeDB). A detailed description of this database (its content and construction) is available in Stanley et al. (2022a). Data can be provided in any form, but use of the excel template provided here at https://stanley.wisc.edu/grime, along with the User's Guide is intended to greatly increase the efficiency, extent, and accuracy of data assimilation into GRiMeDB. The figure below (Figure 1) illustrates the general structure of the published database (from Stanley et al. 2022b, available at: https://doi.org/10.6073/pasta/f48cdb77282598052349e969920356ef), illustrating the linkages between the four Tables (sheets) and the path of data entry (note that you are not being asked to provide/enter Source_IDs, Site_IDs, and observations names; we do this as part of the data checking/harmonization process). Details about the specific structure and data entry process begin on the following pages. If you have any questions during the data entry process, please don't hesitate to email us at grime@mailplus.wisc.edu

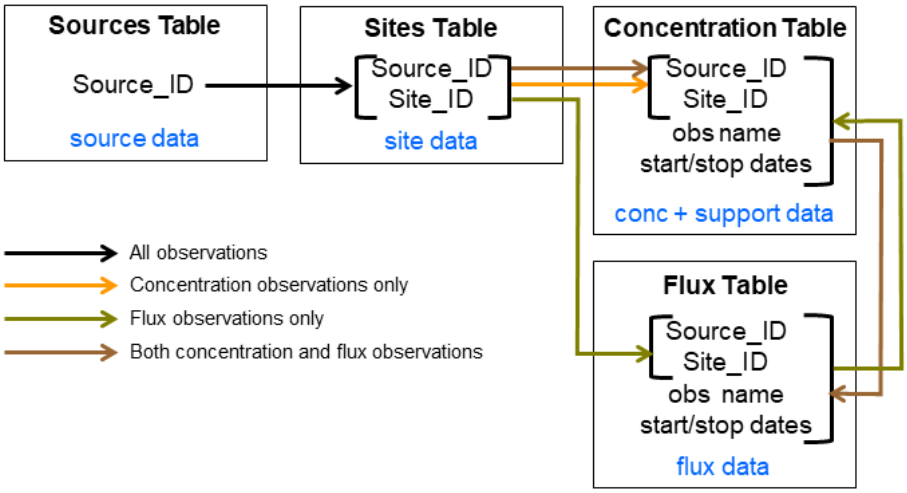

Figure 1: General structure of the GRiMe database and connections between its four tables. Information flow begins with entering information about each data source into the Sources Table. Site information for each site within a data source is then entered into the Sites Table. When we receive your data, each site will be given a unique Site_ID and linked to its data source by the Source_ID. Source_IDs and Site_IDs will also be carried over to all concentration and flux observations in their respective tables. CH_4 observations include site-date combinations with only concentration data (orange), only flux data (green), or both concentration and flux data (brown). Concentrations and available supporting data (described in Sect. 2.3) are entered into the Concentration Table, and each observation will be given a unique observation (obs) name. For site-date combinations that have both concentration and flux observations, the Source_ID, Site_ID, observation name, and date information are copied to the Flux Table for data entry. Site-date combinations with flux data only get entered into the Flux Table and given a unique observation name. If flux observations have associated supporting data, the Source_ID, Site_ID, observation name, and date information are copied to the Concentration Table for supporting data entry. However, if there is no supporting data, matching rows do not get added to the Concentration Table.

Before starting the data entry process, you need to be sure that your data meets basic database requirements. The primary goal of GRiMeDB is to provide consistent, organized greenhouse gas (GHG) data from sites in natural and human-made fluvial (lotic) waters. Thus, there are some data requirements and some exclusions that **must** be checked before data entry begins. And although it is not a requirement, we also prefer data that are site- and time-specific and not averages over more than one site and/or more than one time.

EXCLUSIONS:

Do **not** include sites and associated data from the following type of sites:

- Marine coastal zone systems with salinity > 1 ppt during sampling or for >50% of the year. In short, we do not want to include estuarine sites. If in doubt, take a conservative approach and do not include a site if you think it may be estuarine/marine influenced.
- Sites situated in lentic habitats, including ponds, lakes, reservoirs, and other impounded waters (e.g., upstream of small dams/weirs, beaver ponds)
- Reservoir spillways
- Lake outlets
- Streams lacking surface water
    - *Note*: sites in intermittent or ephemeral streams in the process of dewatering qualify for inclusion as long as surface water is present, even if there is no measurable streamflow.

Please do **not** include flux measurements that were generated with chambers placed on or in the sediment surface or attached to fixed collars on or in the sediment surface. We want to be sure we're capturing fluxes across the air-water interface.

REQUIREMENTS:

There are a small number of required fields in GRiMeDB that are highlighted by blue-colored cells in the excel sheet. At a minimum, we need to have the following information:

- The name and email of the data provider.
    - We need an email that will be in use for 2-3 years. If you are unsure about this, please *also include* the name and email for a secondary contact.
- The title of the data source. If the dataset is unpublished, please compose a name for it.
- The year of data source publication or if unpublished, when you are providing it to us.
- A way to cite or find the persistent citation for a published data source and data.
- The name of the sampling site
- The sampling site's latitude and longitude, resolved to a minimum of 2 decimal places- though resolution to 4 decimal places is strongly preferred.

At least one GHG (CH_4, CO_2, or N_2O) concentration or flux value for each site

- The date(s) of sample collection
- For flux data, how the flux was measured.
- Any supporting data (e.g., temp, DO, discharge) that is provided along with gas data should be *collected concurrently with the gas sample*. It should not be a general site or annual average.

OTHER IMPORTANT INFORMATION:

The template spreadsheet is composed of six different sheets:

- Introduction
- Sources
- Sites
- Concentrations
- Fluxes
- Drop-downs

Data gets entered into the first 4 or 5 sheets and details about the data entry process begin after this section.

***Before you begin entering data, please save your file with a new name by replacing 'template' in the file name with your last name. For example, change "GRiMeDB_data_template.xls" to GRiMeDB_data_Stanley.xls.*

Formatting issues:

Because future updates of GRiMeDB will continue to be hosted at the Environmental Data Initiative (https://edirepository.org/), we need to meet their formatting standards.

***Please do not use non-English symbols or letters when you enter data (e.g., letters with accents, etc.).*

Drop-down menus:

To facilitate consistent data entry, several columns have drop-down menus for the first 1000 rows. The only possible data to enter in these columns must be one of the options from these menus. All menus and their contents are listed in the Drop-down sheet. Note that the contents of this sheet are protected. You can see all the menus by visiting this sheet, or you can see a column-specific menu by clicking on down arrow on the right side of the cell that will appear when your cursor is in the cell (Figure 2).

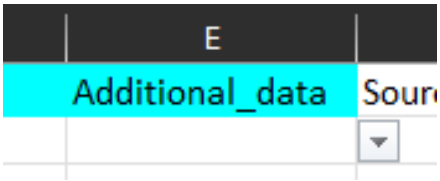

Figure 2: Example of a column with drop-down menus (the Additional_data column in the Sources sheet). The down arrow on the right side of the cell appears when your cursor is in the cell and clicking on this arrow opens the menu of data choices that can be entered into that cell. Note that the Additional_data column is blue, indicating that providing information for this column is required.

For columns with drop-down menus, data can be entered by (1) entering the value directly; (2) clicking on down arrow on the right side of the cell that will appear when your cursor is in the cell and choosing the appropriate value; or (3) copying the appropriate value from one cell into other cells in the column.

Note: If you have more than 1000 rows worth of data, please copy the menu-specified values into cells beyond row 1000.

Comments:

All sheets have a "Comments" column as the last column in the sheet. This is a place to add on any additional information about the data entered in the specific row, or to explain why "other"

was chosen from a drop-down menu. We hope that the "other" option is not needed, but if it is, then please use the comment field to report the appropriate units or category or datum.

2. GRIMEDB SHEETS AND DATA ENTRY

INTRODUCTION SHEET

This first sheet includes a request for your name and a contact email along with a list of the column names for the Sources, Sites, Concentrations, and Fluxes sheets with *brief* definitions for quick reference during data entry. More detailed explanations of the columns and how to enter data are provided below. And keep in mind: you are *required* to provide data for columns with headers highlighted in blue.

Primary contact name (enter to the right)- Provide your name in the black-outlined box.

Secondary contact name (enter to the right)- If you are unsure if you will be available to answer questions about the data you are about to enter or your email may change 2-3 years from now, please also provide the name of a secondary contact.

Primary contact email (enter to the right)- enter your email address so that we can contact you if we have any questions about your data.

Secondary contact email (enter to the right)- If you named a secondary contact name, please add this individual's email address in the box to the right.

SOURCES SHEET

This sheet asks for information about your data source (e.g., published papers, reports, published or unpublished datasets) and has 5 required fields to ensure that we are able to provide an accurate and persistent citation for the data source.

Title- The title of the data source. If this is a published outlet (e.g., journal article, report, dataset), please use the same title that was used for the journal article/report/published dataset. If the dataset is unpublished, please compose a name for it.

Author - The primary author of the data source OR name of the contact person providing the data.

- If the contact person is different from the primary author,
  - Please note this difference in the Comments field.
  - Please be sure that the primary author has authorized data sharing.

Source- of the journal, report type, manuscript archive, data repository, student product

- If this is a graduate thesis or dissertation, please specify M.S. or Ph.D., and add the name of the university and country; e.g.,
  - M.S. thesis, University of Wisconsin-Madison, USA
- If this is an unpublished dataset, enter "unpublished" into this field

Pub_year- year of publication or when data became available.

- For unpublished datasets, enter the present year (the year the data is submitted to us).

Additional_data- This is a Yes/No question.

- Yes- this answer means that there are additional relevant data (e.g., water chemistry) that was collected concurrently with gas data, AND that these data come from a different data source. For example, there may be discharge data derived from the U.S. Geological Survey to support the gas data.
    - This is a required field because we want to be sure to cite ALL data included in GRiMeDB accurately and completely.
- No- there is no additional data to report.

Source_DOI or cite- DOI of a data source or a full citation. Although this is not required, including a DOI is incredibly helpful. Again, we want to be sure that we can cite ALL data in GRiMeDB accurately and completely.

Data_DOI_primary- If your gas (CH_4, CO_2, N_2O) data are published in a public data repository (e.g., EDI, Zenodo, Pangaea, etc.), please provide the dataset DOI here.

- Provide this DOI for all published datasets even if there is a Source_DOI that you've entered. We need both pieces of information.

Data_DOI_secondary- If there are additional relevant datasets that are a source of data for your entry, please provide the DOI or other citation information here. Secondary data sources are those that contain supporting information (e.g., water chemistry, site information, discharge) and do NOT contain gas concentration or flux data.

- Some data sources currently in GRiMe have multiple secondary data sources. Frequently, these are collaborative projects in which gas data may appear in one (published) data set and hydrology, chemistry, etc. data appear in other datasets.

Comments- Provide any additional relevant information about the data source here…

SITES SHEET

As the name indicates, data entered into this sheet describes each unique sampling site in your dataset.

Site_Name- Please provide a *unique* name for each of your sampling sites. If possible, this should be the same or similar to the site name used in the data source document or data set.

Stream_Name- Name of the stream, river, ditch, spring, etc, that contains the site. If the system does not have a name, feel free to generate one.

Aggregated- This is a Yes/No question.

- Yes- indicates that the gas data you will enter in the Concentrations and/or Fluxes sheet are averages from more than one distinct site.
    - Please note: we prefer to have data entered for individual sites (i.e., non-aggregated data)
- No- indicates that your gas data to be entered are values for 1 specific site.

N_Sites_Aggregated- If you entered "Yes" for the "Aggregated" field, please enter the number of sites that went into generating your average gas values.

Basin_Region- The name of the larger drainage basin or region that contains the sampling site. This information is particularly useful for datasets that include sites from a large geographic area.

Latitude- Please enter sampling site latitude in decimal degrees (WGS84 ensemble: EPSG:4326 coordinate system). If your latitude/longitude units are in other units, there are many web pages that can convert to decimal format.

- We prefer latitude information to be resolved to 4 decimal places.

Longitude- Please enter sampling site latitude in decimal degrees (WGS84 ensemble: EPSG:4326 coordinate system).

- We prefer longitude information to be resolved to 4 decimal places.

Country- Sample site country. If the sampling site is on a stream/river that is a border between 2 countries, enter just 1 country name.

Elevation_m- elevation of sampling site in meters above sea level

Slope- channel slope for the sample site. Ideally, the reported slope is for the reach upstream of the sampling site. We prefer units of m elevation change per m channel length (m m^{-1}) if possible.

- It would be helpful to provide some details about the slope in the comments field; e.g., is it for a 100-m reach upstream of the sampling point?

Slope_units- Select appropriate units for channel slope from the drop-down menu. If your units do not appear in the menu, please select "other" and list units in the "Comments" column of this sheet.

System_size- Select an appropriate size for the channel containing your sampling site from the drop-down menu. This is a qualitative category, but categories and some guidance to help with this filed are:

- Small- these are wadeable headwater channels typically with a Strahler stream order of 0, 1, 2, or 3.
- Medium- mid-sized systems in the range of Strahler orders 4-6
- Large- large 7^{th}-8^{th} order non-wadeable systems
- Major- very large, regionally significant rivers (e.g., mainstems of major world drainages)

Strahler_order- Strahler stream order

Depth_m- typical/average water depth of the channel at the sampling point. This field is used to characterize the channel size.

Width_m- typical/average channel width at the sampling point. This field is used to characterize the channel size.

Basin_size- area of the drainage basin *at the point of sample collection*

Basin_size_units- Select the appropriate units from the pull-down menu for the catchment size information entered previously.

- m^2- square meters
- ha- hectares
- km^2- square kilometers
- Other- units other than m^2, ha, km^2; if other units are used, please specify your units in the comments field of this sheet.

Channel_type- This field is intended to capture information for site situated in marginal or human-modified locations. If the channel area does not fit any of channel types described below, then select "NORM".

NOTE: more than 1 category code may be appropriate, so there is no drop-down menu. If 2 codes apply, separate the codes with a comma (e.g., DD, CH for a site that is channelized and downstream from a dam). Channel type codes are:

- DD- A site between 0 and ~7 km downstream of a large dam and 0-2 km of a small dam. Do not include sites located in dam spillways.
    - If you know the site's distance downstream from the dam, please enter this in the Comments field.
- IMP- Site is in a channel with many (typically small) weirs, lock-and-dam structures, or other sorts of impoundments in close proximity that likely modify water velocity and sediment transport/deposition. This condition is common in European rivers and rivers used for navigation.
- PS- Site is immediately downstream of a point source (within approximately 100-200 m). Some judgement is required for this field. If you feel that the site is influenced by the point source effluent, please choose the PS option.
    - Additional information about the expected effluent influence may be useful. If so, please add comments to the Comments field.
- TH- Site is suspected of receiving CH_4 with a thermogenic origin. The thermogenic source may be natural or the result of mining, fracking, or oil extraction activities in the area.
- AR- Site is located in a channel with armored banks (e.g., with large rocks, rip-rap, materials such as concrete mats or blocks or plastic liners) or have structures such as wing dams or groynes along the channel edges intended to prevent bank erosion or channel migration.
- CH- A site in what is likely to have been a natural channel at some time in the past but now appears to be channelized; that is, it has long straight sections of uniform width, and changes in channel direction are typically distinct angular features rather than curves.
- DIT- Ditch, typically for agricultural drainage, without channel hardening
- CAN- Canal or other artificial channel with hardened channel boundaries
    - We recognize that ditches, canals and channelized systems may be difficult to identify or distinguish from DIT, CH, and AR categories.
- DC- Site in a river delta channel
- WS- a stream site within or at the outlet of a wetland that it drains
- SP- Spring channel
    - This does not include sites characterized as seeps (features with low flow volume adjacent to channels)
- FP- Site in a floodplain water body connected to the main channel that appears lentic or is described as a floodplain lake or backwater.
    - This category does not include side-channels of a braided channel situated within a floodplain or tributary channels transecting a floodplain.
- NORM- Site is situated in a channel that is *not* conspicuously modified or marginal.

Comments- Provide any additional relevant information about the site here.

CONCENTRATIONS SHEET

In addition to gas concentration data, supporting information on physico-chemical conditions (temperature, nutrients, etc.) at the time of sampling also gets reported on this sheet. In the case of datasets that have gas flux and supporting physicochemical data but not gas concentration, rows should be created in this sheet to report the supporting physicochemical data. In this case, it is easier to enter the flux data first, then copy the contents for the first 7

columns (Site_Name → Season) from the Fluxes Sheet and paste it into the Concentrations Sheet before entering the supporting physicochemical data.

Site_Name- Please copy the appropriate unique name for your sampling site from the Sites sheet.

Date_start- First sampling date; please use "yyyy-mm-dd" format.

Date_end- Last sampling date; please use "yyyy-mm-dd" format. This date will be the same as Date_start if you are reporting an observation from a single day.

Aggregated_Space- This is a Yes/No question. "Yes" means that CH_4 data entered are averages from >1 site.

Aggregated_Time- This is a Yes/No question. "Yes" means that CH_4 data entered are averages from >1 date, as would be indicated by different Date_start and Date_end entries.

Time_of_day- The time of day of sample collection is especially important if multiple samples were taken within one day, but will still be useful if only 1 sample/day was collected. Please use your local time, 24-hour time scale (not a.m. or p.m.), minutes, and ideally, your time zone. For example, enter "13:17 CDT" if a sample was collected at 1:17 p.m. Central Daylight Time.

Season- Please select from drop-down menu choices:

- Winter
- Spring
- Summer
- Fall
- Wet
- Dry
- Other
- Multiple

These options may not be good fits for your site. If this is the case, then select "Other" and provide some explanation in the Comments field. If your data are temporally aggregated (Aggregated_Time field = "Yes") and spanned more than one season, please select "Multiple".

FluxYesNo- This is a Yes/No question. "Yes" means there is a corresponding flux measurement that was taken at the same time as the concentration measurement for this record (row).

CH4min- Minimum measured CH_4 concentration- Only provide this information if:

- data are aggregated spatially or temporally, or
- there are multiple within-day measurements (e.g., a diel study)

CH4max- Maximum measured CH_4 concentration- Only provide this information if:

- data are aggregated spatially or temporally, or
- there are multiple within-day measurements (e.g., a diel study)

CH4mean- Mean or sole reported CH_4 concentration

CH4_SD- standard deviation of the mean CH_4 concentration. Please enter this value for aggregated concentrations or if replicate gas samples were collected

CH4median- median CH_4 concentration

CH4_unit- units of CH_4 concentration. Select gas units from the drop-down menu:

- mgC/L

- mgCH4/L
- ugC/L
- ugCH4/L
- ngC/L
- ngCH4/L
- molCH4/L
- mmCH4/L
- umolCH4/L
- nmolCH4/L
- ppm_CH4
- ppb_CH4
- uatm_CH4
- %sat_CH4
- other

N_CH4- Number of samples or observations corresponding to the mean or median concentration CH_4 concentration.

CO2min- Minimum measured CO_2 concentration- Only provide this information if:
- data are aggregated spatially or temporally, or
- there are multiple within-day measurements (e.g., a diel study)

CO2max- Maximum measured CO_2 concentration- Only provide this information if:
- data are aggregated spatially or temporally, or
- there are multiple within-day measurements (e.g., a diel study)

CO2mean- Mean or sole reported CO_2 concentration

CO2 _SD- standard deviation of the mean CO_2 concentration. Please enter this value for aggregated concentrations or if replicate gas samples were collected.

CO2median- median CO_2 concentration

CO2 _unit- units of CO_2 concentration. Select gas units from the drop-down menu:
- gC/L
- gCO2/L
- mgC/L
- mgCO2/L
- ugC/L
- ugCO2/L
- ngC/L
- ngCO2/L
- molCO2/L
- mmCO2/L
- umolCO2/L
- nmolCO2/L
- ppm_CO2
- ppb_CO2
- uatm_CO2
- %sat_CO2
- other

N_CO2- Number of samples or observations corresponding to the mean or median concentration CO_2 concentration.

N2Omin- - Minimum measured N_2O concentration- Only provide this information if:

- data are aggregated spatially or temporally, or
- there are multiple within-day measurements (e.g., a diel study)

N2Omax- Maximum measured N_2O concentration- Only provide this information if:

- data are aggregated spatially or temporally, or
- there are multiple within-day measurements (e.g., a diel study)

N2Omean- Mean or sole reported N_2O concentration

N2O_SD- standard deviation of the mean N_2O concentration. Please enter this value for aggregated concentrations or if replicate gas samples were collected.

N2Ounit- units of N_2O concentration; select gas units from the drop-down menu:

- mgN/L
- mgN2O/L
- ugN/L
- ugN2O/L
- ngN/L
- ngN2O/L
- mmN2O/L
- umolN2O/L
- nmolN2O/L
- ppm_N2O
- ppb_N2O
- uatm_N2O
- %sat_N2O
- other

N_N2O- Number of samples or observations corresponding to the mean or median N_2O concentration.

WaterTemp_degC- Water temperature in degrees C measured at the sampling site concurrently with gas sampling.

AirPressure- Air pressure measured concurrently with gas sampling

AirPressure_unit- Select air pressure units from the drop-down menu:

- atm
- Pa
- MPa
- KPa
- bar
- mbar
- mmHg- mm mercury
- other

If “other” is selected, please provide information about your units in the Comments field

Cond_uScm- Specific conductance in µS cm^{-1}

pH- pH (unitless) measured at the sampling site concurrently with gas collection

DO- dissolved oxygen measured at the sampling site concurrently with gas collection

DO_unit- Please select dissolved oxygen units from the drop-down menu choices:

- mol/L
- mmol/L
- mg/L
- other

If “other” is selected, please provide information about your units in the Comments field

DO_percentsat- dissolved oxygen expressed as percent saturation

NO3- NO_3 or NO_2+NO_3 concentration measured at the sampling site concurrently with gas collection

NO3_unit- Select the appropriate NO3 unit from the drop-down menu:

- mgNO2+NO3-N/L- nitrite and nitrate as mg N/L
- mgNO3-N/L- nitrate as mg N/L
- mgNO3/L nitrate as mg NO3/L
- ugNO2+NO3-N/L- - nitrite and nitrate as ug N/L
- ugNO3-N/L- nitrate as µg N/L
- ugNO3/L- nitrate as µg NO3/L
- mmNO2+NO3/L- nitrite and nitrate as mmol/L
- mmolNO3/L- nitrate as mmol/L
- umNO2+NO3/L- nitrite and nitrate as µmol/L
- umNO3/L- nitrate as µmol/L
- other

If “other” is selected, please provide information about your units in the Comments field

NH4- NH4 concentration measured at the sampling site concurrently with gas collection

NH4_unit- Please select the appropriate NH4 unit from the drop-down menu. Unit definitions are similar to those for NO3_unit.

- mgNH4-N/L
- mgNH4/L
- ugNH4-N/L
- ugNH4/L
- mmolNH4/L
- umolNH4/L
- other

If “other” is selected, please provide information about your units in the Comments field

DIN- Dissolved inorganic nitrogen concentration measured at the sampling site concurrently with gas collection. If you collected NO3 and NH4 data, you can leave this field blank; it will be calculated automatically during the unit conversion and checking process.

DIN_unit- Select the appropriate DIN unit from the drop-down menu:

- mgDIN/L
- ugDIN/L
- mmolDIN/L
- umolDIN/L

- other

If “other” is selected, please provide information about your units in the Comments field.

TN_or_TDN- Total Nitrogen or Total Dissolved (filtered) nitrogen concentration measured at the sampling site concurrently with gas collection.

TN_unit- Select the appropriate TN or TDN unit from the drop-down menu:

- mgTN/L
- ugTN/L
- mmolTN/L
- umolTN/L
- mgTDN/L
- ugTDN/L
- mmolTDN/L
- umolTDN/L
- other

If “other” is selected, please provide information about your units in the Comments field.

SRP- Phosphate (PO_4) or soluble- (dissolved-, molybdate-) reactive phosphorus concentration measured at the site concurrently with gas sampling.

SRP_unit- Select the appropriate PO4 unit from the drop-down menu:

- mgSRP/L
- mgPO4-P/L
- mgPO4-L
- ugSRP/L
- ugPO4-P/L
- ugPO4-L
- mmolSRP/L
- mmolPO4/L
- umolSRP/L
- umolPO4/L
- other

If “other” is selected, please provide information about your units in the Comments field.

TP_or_TDP- Total Phosphorus or Total Dissolved (filtered) Phosphorus concentration measured at the sampling site concurrently with gas collection.

TP_unit- Select the appropriate TP or TDP unit from the drop-down menu:

- mgTP/L
- ugTP/L
- mmolTP/L
- umolTP/L
- mgTDP/L
- ugTDP/L
- mmolTDP/L
- umolTDP/L
- other

If “other” is selected, please provide information about your units in the Comments field

DOC_or_TOC- Dissolved (filtered) organic carbon or Total (unfiltered) organic carbon concentration measured at the sampling site concurrently with gas collection

DOC_unit- Select the appropriate DOC or TOC unit from the drop-down menu. Dissolved OC units are denoted by "C" only and total OC have units of TC.

- mgC/L
- ugC/L
- mmmol/L
- umol/L
- mgTC/L
- ugTC/L
- mmolTC/L
- umolTC/L
- other

If "other" is selected, please provide information about your units in the Comments field.

Q- channel discharge at the sampling site measured at the time of gas sample collection.

Q_unit- Select the appropriate discharge unit from the drop-down menu:

- L/s
- m3/s
- ft3/s
- m3/min
- m3/d
- other

If "other" is selected, please provide information about your units in the Comments field

Comments- Report any relevant additional information about the data entered in the row

FLUXES SHEET

Flux data is entered here. For site-dates that also have concentrations collected along with flux measurements, it is easiest to copy contents of the first 5 columns (Site_Name → Season). Or conversely, for a site-date entry that has flux data and supporting physicochemical data, but not gas concentration data, information entered for these 7 columns here should be copied into the Concentrations sheet when entering this supporting data.

Note- there are columns for entering diffusive, ebullitive, and total flux measurements for CH_4, but only single columns for entering CO_2 and N_2O flux data, which we assume represents diffusive flux. Even though these gases may be emitted via ebullition, this template does not currently have columns for this pathway. You can create these columns after the comments field if you want to report these fluxes. Similarly, there are no columns for any gas for plant-associated fluxes.

Site_Name- Please copy the appropriate unique name for your sampling site from the Sites sheet or the Concentrations sheet.

Date_start- First sampling date; please use "yyyy-mm-dd" format

Date_end- Last sampling date; please use "yyyy-mm-dd" format. This date will be the same as Date_start if you are reporting an observation from a single day.

Aggregated_Space- This is a Yes/No question. "Yes" means that CH_4 data entered are averages from >1 site.

Aggregated_Time- This is a Yes/No question. “Yes” means that CH_4 data entered are averages from >1 date, as would be indicated by different Date_start and Date_end entries.

Time_of_day- The time of day of sample collection is especially important if multiple samples were taken within one day, but will still be useful if only 1 sample/day was collected. Please use your local time, 24-hour time scale (not a.m. or p.m.), minutes, and ideally, your time zone. For example, enter “13:17 CDT” if a sample was collected at 1:17 p.m. Central Daylight Time.

Season- Please select from drop-down menu choices:

- Winter
- Spring
- Summer
- Fall
- Wet
- Dry
- Other
- Multiple

These options may not be good fits for your site. If this is the case, then select “Other” and provide some explanation in the Comments field. If your data are temporally aggregated (Aggregated_Time field = “Yes”) and spanned more than one season, please select “Multiple”.

Diffusive_CH4_Flux_Min- Minimum measured CH_4 diffusive flux. Only provide this information if:

- data are aggregated spatially or temporally, or
- there are multiple (>10) within-day measurements.

Diffusive_CH4_Flux_Max- Maximum measured CH_4 diffusive flux. Only provide this information if:

- data are aggregated spatially or temporally, or
- there are multiple (>10) within-day measurements.

Diffusive_CH4_Flux_Mean- Mean or sole reported CH_4 diffusive flux.

Diffusive_CH4_Flux_SD- standard deviation of the mean CH_4 diffusive flux. Enter this value if flux is aggregated or if replicate measurements were made.

Diffusive_CH4_Flux_Median- Median CH_4 diffusive flux

Diffusive_Flux_unit- units of CH_4 diffusive flux. Select gas units from the drop-down menu. For brevity, these units are not listed here, as there are >35 options. Unit formatting follows prior examples above.

N_Diff_CH4- sample size for reported diffusive CH_4 flux

Diff_CH4_Method- Methodological category used to measure diffusive gas flux. Select the appropriate method from the drop-down menu. ‘Unspecified response’ indicates that it’s not clear if gas concentrations increased in a linear or non-linear fashion during flux measurements; floating chambers are those that are able to move freely on the water surface and are not held in place. Conc+k refers to flux calculated from dissolved gas concentration and some gas exchange (k) value.

- suspended/tethered chamber- unspecified response
- suspended/tethered chamber- linear response
- floating chamber- unspecified response
- floating chamber- linear response
- conc+k

Eb_CH4_Flux_Min- Minimum measured CH_4 ebullitive flux. Only provide this information if:

- data are aggregated spatially or temporally, or
- there are multiple (>10) within-day measurements

Eb_CH4_Flux_Max- Maximum measured CH_4 ebullitive flux. Only provide this information if:

- data are aggregated spatially or temporally, or
- there are multiple (>10) within-day measurements

Eb_CH4_Flux_Mean- Mean or sole reported CH_4 ebullitive flux

Eb_CH4_Flux_SD- standard deviation of the mean CH_4 ebullitive flux. Enter this value if flux is aggregated or if replicate measurements were made.

Eb_CH4_Flux_Median- Median CH_4 ebullitive flux

Eb_CH4_Flux_unit- units of CH_4 ebullitive flux. Select gas units from the drop-down menu. For brevity, these units are not listed here. Unit formatting follows prior examples above.

N_Eb_CH4- sample size for reported ebullitive CH_4 flux

Eb_method- Methodological category used to measure ebullitive gas flux. Select the appropriate method from the drop-down menu.

Total_CH4_Flux_Min- Minimum measured CH_4 total (diffusion + ebullition) flux. Only provide this information if:

- data are aggregated spatially or temporally, or
- there are multiple (>10) within-day measurements

Total_CH4_Flux_Max- Maximum measured CH_4 total (diffusion + ebullition) flux. Only provide this information if:

- data are aggregated spatially or temporally, or
- there are multiple (>10) within-day measurements

Total_CH4_Flux_Mean- Mean or sole reported CH_4 total (diffusion + ebullition) flux

Total_CH4_Flux_SD- standard deviation of the mean CH_4 total (diffusion + ebullition) flux. Please enter this value for aggregated concentrations or if replicate gas samples were collected

Total_CH4_Flux_Median- Median CH_4 total (diffusion + ebullition) flux

Total_CH4_Flux_unit- units of CH_4 total flux. Select gas units from the drop-down menu. For brevity, these units are not listed here. Unit formatting follows prior examples above.

N_Total_CH4- sample size for reported total CH_4 flux

Total_method- - Methodological category used to measure total gas flux. Select the appropriate method from the drop-down menu.

CO2_Flux_Min- Minimum measured CH_4 diffusive flux. Only provide this information if:

- data are aggregated spatially or temporally, or
- there are multiple (>10) within-day measurements

CO2_Flux_Max- Maximum measured CH_4 diffusive flux. Only provide this information if:

- data are aggregated spatially or temporally, or
- there are multiple (>10) within-day measurements

CO2_Flux_Mean- Mean or sole reported CH_4 diffusive flux

CO2_Flux_SD- standard deviation of the mean CH_4 diffusive flux. Enter this value if flux is aggregated or if replicate measurements were made.

CO2_Flux_Median- Median CH_4 diffusive flux

CO2_Flux_unit- units of CH_4 diffusive flux. Select gas units from the drop-down menu. For brevity, these units are not listed here, as there are >35 options. Unit formatting follows prior examples above.
N_CO2_flux- sample size for reported total CH_4 flux

CO2_flux_method- Methodological category used to measure ebullitive gas flux. Select the appropriate method from the drop-down menu.

N2O_Flux_Min- Minimum measured CH_4 diffusive flux. Only provide this information if:

- data are aggregated spatially or temporally, or
- there are multiple (>10) within-day measurements

N2O_Flux_Max- Maximum measured CH_4 diffusive flux. Only provide this information if:

- data are aggregated spatially or temporally, or
- there are multiple (>10) within-day measurements

N2O_Flux_Mean- Mean or sole reported CH_4 diffusive flux

N2O_Flux_SD- standard deviation of the mean CH_4 diffusive flux. Enter this value if flux is aggregated or if replicate measurements were made.

N2O_Flux_Median- Median CH_4 diffusive flux

N2O_Flux_unit- units of CH_4 diffusive flux. Select gas units from the drop-down menu. For brevity, these units are not listed here, as there are >35 options. Unit formatting follows prior examples above.
N_N2O_flux- sample size for reported total CH_4 flux

N2O_flux_method- Methodological category used to measure ebullitive gas flux. Select the appropriate method from the drop-down menu.

k_method- Methodological category used for estimating the gas exchange coefficient, *k*. Select appropriate category from the drop-down menu

k_ref- article or report citation for the k method used in flux calculations

Comments- Report any relevant additional information about the data entered in the row

Citations

Stanley, E.H., L.C. Loken, N.J. Casson, S.K. Oliver, R.A. Sponseller, M.B. Wallin, L. Zhang, and G Rocher-Ros. 2022a. GRiMeDB: The global river database of methane concentrations and fluxes. Earth System Science Data https://doi.org/10.5194/essd-2022-346.

Stanley, E.H., L.C. Loken, N.J. Casson, S.K. Oliver, R.A. Sponseller, M.B. Wallin, L. Zhang, and G. Rocher-Ros. 2023. GRiMeDB: a comprehensive global database of methane concentrations and fluxes in fluvial ecosystems with supporting physical and chemical information ver 2. Environmental

Data Initiative. https://doi.org/10.6073/pasta/f48cdb77282598052349e969920356ef (Accessed 2023-04-11).