Generic Carbon Budget Model (GCBM) Technical Guide: “gcbmwalltowall” Python Library

Natural Resources Canada

Canadian Forest Service

January 2025

Table of Contents

1. Introduction

The “gcbmwalltowall” tool for the Generic Carbon Budget Model (GCBM) helps users simplify preparing and running GCBM simulations by consolidating the manual workflow into a standardized configuration file format. Most GCBM projects can be run exclusively with this tool, with only advanced or complex simulations requiring the use of the Tiler library and custom scripting. The “gcbmwalltowall” tool performs the following steps for a GCBM project:

1. Tiles pre-rollback layers
2. Creates the input database
3. Runs the spatial inventory rollback (optional) including
    a. Tiling post-rollback layers
    b. Creating a post-rollback input database
4. Configures the GCBM to be run

2. Command-line Tools

GCBM users interact with the “gcbmwalltowall” library through a Python command-line application installed in the “Scripts” directory of the computer’s local Python installation. The application accepts a JavaScript Object Notation (JSON) configuration file containing the necessary information for tiling the spatial layers, creating the input database, performing a spatial inventory rollback, and configuring and running the GCBM. Command-line scripts are listed and described in Table 1.

Table 1. Descriptions of the command-line script included with the "gcbmwalltowall" library.

Command-Line Script	Description
walltowall build <builder config file path> [output config file path]	Allows the user to employ the builder configuration contained in the "config" file to fill in and configure the rest of the project; creates the fully specified configuration in a separate file in the same directory as the builder "config" file unless an output "config" file path is specified by the user
walltowall prepare <fully-specified config file path> [output root path]	Tells the script to use the fully specified project configuration in the "config" file, tile the spatial layers, generate the input database, run the spatial rollback, and configure the GCBM run; project files are generated in various subdirectories off the specified output root path or the directory containing the "config" file (the default)
walltowall merge <fully-specified config file path> <prepared project root 1> <prepared project root 2> [prepared project root n …] --output_path <output root path> [--include_index_layer]	Merges 2 or more projects together in descending priority order into a single project in the specified output path by using the fully specified project configuration from the first argument to get the GCBM "config" file template path and disturbance order; projects must be prepared by the "walltowall "scripts and have the same spatial extent and resolution; if "include_index_layer" is specified, a layer is generated showing which dataset was used for each pixel
walltowall run (local|cluster) <prepared project root> [--config_path <fully-specified config file path>] [--end_year <yyyy>] [--title <title> (cluster only)] [--compile_results_config <compile results config path> (cluster only)]	Runs the specified project either locally or on a "gcbm_celery cluster", using the optionally specified walltowall "config" file for the project title and overridden executable or distributed run client paths; if "end_year" is specified, the current simulation will run to that year without permanently altering the configured value; for cluster runs only, a title and a custom compile results configuration can also be specified

Abbreviation: GCBM, Generic Carbon Budget Model.

3. Configuration Format

The configuration format for a project using the "gcbmwalltowall" library is designed to be more streamlined and user-friendly than writing a Tiler script and running the Recliner2GCBM graphical user interface. It also allows known standardized collections of input data (e.g., Natural Resources Canada's Common Attribute Schema for Forest Resources [CASFRI]) to be configured automatically through Python code in the library instead of through explicit configuration by the user. There are two parts to the configuration, (1) the optional "builder"

configuration used with the "walltowall build" command that acts as a shorthand format for automatically configuring a project based on standardized sets of input data, and (2) the fully specified configuration read by the "walltowall prepare" command that contains the specific details needed to prepare and run the GCBM simulation.

3.1 Requirements

The configuration file (below) starts with the project name, followed by the optional builder configuration, and finally, the fully specified configuration, which can either be blank initially and generated by the builder using "walltowall build" or written manually.

```
{
   "project_name": <project name>,
   [optional builder configuration]
   (fully specified configuration)
}
```

3.2 Builder configuration

The builder configuration (below) is an optional shortcut for creating projects using input data that conforms to the requirements of a builder module, usually allowing a much shorter configuration to generate the more detailed fully specified configuration. It consists of the type of builder to use, followed by any configuration items recognized by the selected builder, and optionally, any fully specified configuration items that should take precedence over the configuration generated by the builder, such as:

```
   "builder": {
      "type": <builder type name>,
      (any configuration specific to the selected builder)
      [any main configuration items that take priority over the builder-generated configuration]
   }
```

For example,

```
{
   "builder": {
      "type": "casfri",
      "casfri_data": "../00_preprocessing/casfri_data/processed/YT03",
      "other_data": "../00_preprocessing",
      "yield_table": "../00_preprocessing/aspatial/yields/afforestation_national/YT_yield_table.csv",
      "dm_xls": "../00_preprocessing/aspatial/archive_index/casfri_dms.xlsx",
      "aidb": "../00_preprocessing/aspatial/archive_index/casfri_archive_index.mdb",
      "age_distribution": "../../00_preprocessing/aspatial/BGI_age_class_dist.xlsx"
      "rollback": {
         "age_distribution": "rollback/age_distribution.json",
         "inventory_year": 2022
      }
   }
}
```

3.3 Fully specified configuration

The fully specified configuration contains the details required to prepare and run a GCBM simulation. Because the objectives of this tool are to increase user-friendliness and efficiency, many of the configuration items are optional, and the tool will attempt to detect or use defaults for omitted configuration. Configuration items include layer definition, required configuration items, classifiers, optional configuration items, a bounding box, layers, disturbances, and rollback.

Layer definition

There are several places in the configuration that accept either a layer path for simple layers that comply with all the default settings or a layer definition where the details of the layer can be specified more explicitly. Layer definitions can take these settings:

"layer": <path to the layer>

[optional] "attribute": <attribute to read; default: search by layer name, or use first attribute found>
[optional] "attribute": {
  <attribute to read>: <attribute value to filter for; only matching polygons are rasterized>
}

[optional] "lookup_table": <path to .csv file>

The lookup table replaces the original pixel or attribute values with new ones. For rasters, the first column is the pixel value, and subsequent columns are the attributes. For vectors, the columns are paired up; the first column in each pair is named after the attribute to replace and contains the original values, and the second column's name does not matter and contains the values to substitute. If a substitution does not appear in the lookup table, the original value is used. If no lookup table is specified, the gcbmwalltowall tool searches first for <layer filename>.csv in the same directory as this configuration file, and then in the layer directory.

If "strict_lookup_table" is enabled in the script below and a vector layer is used, polygons with attribute values that do not have a substitution in the lookup table are excluded from rasterization.

[optional] "strict_lookup_table": <true/false>

Required configuration items

Items required in the configuration include:

"aidb": <path to the Archive Index Database (AIDB)>,
"yield_table": <path to the yield table>,
"yield_interval": <yield table age/volume interval in years>,

Classifiers

Classifiers link the yield table to the spatial landscape. They are configured at a minimum with a layer definition for the spatial component. By default, the “gcbmwalltowall” tool will attempt to match the spatial layer to the yield table or the corresponding yield column. Even a separate file containing all possible classifier values can be configured explicitly. At least one classifier must be included in the "classifiers" section, with the following configuration format:

```
"classifiers": {
  <classifier name>: {
    <layer definition items for spatial layer linked to classifier>,
    [optional] "values_path": <file with classifier values; default: use the yield table>,
    [optional] "values_col": <column name/# in values_path; default: use only column in single-column
            file, or classifier name or spatial attribute, or column where values intersect with
            spatial values>,
    [optional] "yield_col": <column name/# in the yield table; default: use values_col if values_path is
the same as yield table path, otherwise search by classifier name, or search for column where values are
a subset of values from values_path>
  }
}
```

Optional configuration items

Optional configuration items include the following:

```
"resolution": <pixel resolution in degrees latitude/longitude>,
"gcbm_config_templates": <path to directory containing the GCBM simulation config file
      templates>,
"disturbance_order": <path to disturbance_order text file>,
```

Bounding box

The bounding box configuration (below) defines the simulation area, and NoData pixels in the bounding box are propagated to all other spatial layers. If the bounding box is not configured, the "initial_age" layer is used.

"bounding_box": <path to the layer or full layer definition>

Layers

The layer configuration (below) contains all the "miscellaneous" spatial layers other than the bounding box, classifiers, or disturbances.

```
"layers": {
   <layer name>: <path to the layer or full layer definition>
}
```

Disturbances

The disturbance configuration (below) contains any disturbance layers to be included with the simulation. The keys are the file names or file patterns to search for, followed by the disturbance details. If the disturbance layer is a shapefile, the “gcbmwalltowall” tool will automatically split it into tiled layers by year.

```
"disturbances": {
   <layer path or glob file pattern>: {
      [optional] "year": <specific year, or name of attribute containing year, or "filename">
```

If a specific year is given or the special keyword "filename" is used, that exact year should be used or the user should try to parse it from the file name. If the year is not specified, the “gcbmwalltowall” tool will first search for an attribute named "year", then search the first column where all the values are 4-digit integers, and finally it will assess if the disturbance year can be parsed from the file name.

If disturbance type is not specified in the script below, the “gcbmwalltowall” tool will search for the first attribute where all the values appear in the “tblDisturbanceTypeDefault” table in the Archive Index Database (AIDB).

[optional] "disturbance_type": <specific disturbance type, or name of attribute containing disturbance type>

If “age_after” is not specified in the script below, the “gcbmwalltowall” tool will search for an attribute named "age_after" or default to no transition rule if it is not found.

[optional] "age_after": <specific age after, or name of attribute containing reset age>

If “age_after” or “regen_delay” is present, this will cause a transition rule to be attached directly to the disturbance layer. The “regen_delay” can be omitted even if “age_after” is present.

[optional] "regen_delay": <specific regeneration delay, or name of attribute containing regeneration delay;
default: 0>

If “regen_delay” is not present, the “gcbmwalltowall” tool will search for an attribute named "regen_delay" or default to no transition rule if it is not found, and “age_after” is not present. If “age_after” or “regen_delay” are present, a transition rule will to be attached directly to the disturbance layer. The “age_after” must be configured if “regen_delay” is present because there is no default.

[optional] <other Layer definition items>
}
}

Rollback

If present, the rollback configuration (below) will initiate a spatial inventory rollback:

```
  "rollback": {
    "age_distribution": <path to the age distribution JSON or Excel file>,
    "inventory_year": <path to the inventory vintage layer or full layer definition, or global
      inventory vintage year>,
    [optional] "rollback_year": <year to roll back to; default: 1990>,
    [optional] "prioritize_disturbances": <true/false, default: false>,
    [optional] "single_draw": <true/false, default: false>,
    [optional] "establishment_disturbance_type": <default establishment disturbance type
      name>,
    [optional] "disturbance_order": <path to the disturbance_order text file>
  }
```

4. Examples of applications

Two “gcbmwalltowall” tool configuration examples are provided below, one for Natural Resources Canada’s Common Attribute Schema for Forest Resources (CASFRI), and one for the Standalone Template.

CASFRI – Yukon Territory example

This configuration uses the CASFRI builder in the “gcbmwalltowall” tool to set up a simulation for a project in the Yukon. In this example, it is assumed that the project is stored in a file called yt_casfri.json:

```
{
  "project_name": "casfri_yt",
  "builder": {
    "type": "casfri",
    "casfri_data": "../00_preprocessing/casfri_data/processed/YT03",
    "other_data": "../00_preprocessing/other_data",
    "yield_table": "../00_preprocessing/yields/afforestation_national/YT_yield_table.csv",
    "dm_xls": "../00_preprocessing/archive_index/casfri_dms.xlsx",
    "aidb": "../00_preprocessing/archive_index/casfri_archive_index.mdb",
    "rollback": {
      "age_distribution": "rollback/age_distribution.json",
      "inventory_year": 2022
    }
  }
}
```

After running the “yt_casfri.json” gcbmwalltowall build, the builder fills in the remaining project details based on assumptions about the way CASFRI projects are structured:


```
{
    "project_name": "casfri_yt",
    "builder": {
        "type": "casfri",
        "casfri_data": "../00_preprocessing/casfri_data/processed/YT03",
        "other_data": "../00_preprocessing/other_data",
        "yield_table": "../00_preprocessing/yields/afforestation_national/YT_yield_table.csv",
        "dm_xls": "../00_preprocessing/archive_index/casfri_dms.xlsx",
        "aidb": "../00_preprocessing/archive_index/casfri_archive_index.mdb",
        "rollback": {
            "age_distribution": "rollback/age_distribution.json",
            "inventory_year": 2022
        }
    },
    "resolution": 0.001,
    "aidb": "../00_preprocessing/archive_index/casfri_archive_index.mdb",
    "yield_table": "../00_preprocessing/yields/afforestation_national/YT_yield_table.csv",
    "yield_interval": 10,
    "classifiers": {
        "RU": {
            "layer": "../00_preprocessing/pspu/pspus_2016.shp",
            "attribute": "Reconcilia"
        },
        "LeadingSpecies": {
            "layer": "../00_preprocessing/casfri_data/processed/YT03/layer_1/leading_species.tiff",
            "values_col": "casfri_species_name"
        }
    },
    "layers": {
        "initial_age": "../00_preprocessing/casfri_data/processed/YT03/layer_1/age_2022.tiff",
        "mean_annual_temperature": "../00_preprocessing/other_data/NAmerica_MAT_1971_2000.tif",
        "admin_boundary": {
            "layer": "../00_preprocessing/other_data/pspu/pspus_2016.shp",
            "attribute": "ProvinceNa"
        },
        "eco_boundary": {
            "layer": "../00_preprocessing/other_data/pspu/pspus_2016.shp",
            "attribute": "EcoBound_1"
        }
    },
    "disturbances": {
        "../00_preprocessing/casfri_data/processed/YT03/layer_1/disturbances_*.tiff": {}
    },
    "rollback": {
        "age_distribution": "rollback/age_distribution.json",
        "inventory_year": 2022
    }
}
```

After the fully specified project configuration is generated, the project can be prepared by running:

```
walltowall prepare yt_casfri.json
```

This tiles the spatial layers, generates the input database, runs the spatial inventory rollback, generates the rollback input database, and configures the GCBM simulation to be run. Finally, the simulation can be run either locally or on a cluster (with the "gcbmwalltowall" tool correctly configured and tunnels already connected):

```
walltowall run local . --config_path yt_casfri.json
walltowall run cluster . --config_path yt_casfri.json
```

Standalone Template example

This Standalone Template example assumes the same directory structure as the GCBM Standalone Template training project, with the gcbmwalltowall configuration file located in an extra "config" directory (walltowall_config.json), and the input files in the same locations. There is no builder shortcut for this type of custom project, so the builder step is skipped, and the fully specified configuration is entered, relying on defaults to keep everything concise as follows:

```
{
    "project_name": "standalone_template",
    "start_year": 2010,
    "end_year": 2020,
    "resolution": 0.00025,
    "aidb": "../input_database/ArchiveIndex_Beta_Install.mdb",
    "yield_table": "../input_database/yield.csv",
    "yield_interval": 10,
    "classifiers": {
        "Classifier1": {
            "layer": "../layers/raw/inventory/inventory.shp",
            "attribute": "Classifer1"
        },
        "Classifier2": {
            "layer": "../layers/raw/inventory/inventory.shp",
            "attribute": "Classifer2"
        }
    },
    "layers": {
        "initial_age": {
            "layer": "../layers/raw/inventory/inventory.shp",
            "attribute": "AGE_2010"
        },
        "mean_annual_temperature": {
            "layer": "../layers/raw/inventory/inventory.shp",
            "attribute": "AnnualTemp"
        }
    },
    "disturbances": {
        "../layers/raw/disturbances/disturbances.shp": {
            "age_after": 0,
            "regen_delay": 0
        }
    }
}
```


After the fully specified project configuration is written, the project can be prepared by running the following command from the project root directory (i.e., 1_Standalone_Template\):

walltowall prepare walltowall_config.json

This tiles the spatial layers, generates the input database, runs the spatial inventory rollback, generates the rollback input database, and configures the GCBM simulation to be run. Finally, the simulation can be run either locally or on the cluster (with the "gcbmwalltowall" tool correctly configured and tunnels already connected):

```
walltowall run local . --config_path config\walltowall_config.json
walltowall run cluster . --config_path config\walltowall_config.json
```

Appendix 1

Configuration file components

Appendix 1. Summary of GCBM configuration file components, elements, requirements, value types, functions, and validation (if any)

Element	Required (Yes/No)	Value Type	Function
Section: <root>			
project_name	Yes	string	Project name
builder	No	dict	Optional builder configuration
resolution	No	decimal	Resolution for the bounding box when tiling spatial layers
		null/empty	Use 0.001 as the default value
aidb	Yes	string	Path to the Archive index database (AIDB)
gcbm_config_templates	No	null/empty	Use the default GCBM config file templates included with "gcbmwalltowall"
		string	Path to the directory containing the GCBM config file templates
disturbance_order	No	string	Path to the disturbance order file, a text file with a list of disturbance type names in descending priority order; the list is used to order the sequence of disturbances if there are multiple disturbances in the same year in the same pixel
		null-empty	A default disturbance order is generated based on a disturbance code (default disturbance type ID), in which the lowest code/ID has the highest priority
yield_table	Yes	string	Path to the yield table
yield_interval	Yes	int	The yield increment interval in years
classifiers	Yes	dict	Top-level keys are classifier names with the details in the values
rollback	No	dict	Rollback configuration (enables spatial rollback if present)
		null/empty	The rollback section is optional; if not present, no rollback will be performed
bounding_box	No	string	Path to the bounding box layer
layers	No	dict	List of miscellaneous spatial layers to tile, where "key" equates to the layer name (for the tiler), and the values are either simple layer paths or full layer definitions
		null/empty	The layers section is technically optional but will usually contain the "initial_age" layer at a minimum
disturbances	No	dict	The list of disturbance layers to tile; keys are paths to layers or glob file patterns, and values are disturbance layer definitions
		null/empty	The disturbances section is optional
Section: <builder>			

Element	Required (Yes/No)	Value Type	Function
type	Yes; only when the associated optional configuration section is present	string	The builder name
Section: <any layer definition>			
layer	Yes	string	The path to the spatial layer
attribute	No	string	Use as an attribute to read from
		dict	A single-item dictionary where the key is the attribute to read from, and the value is the attribute value to filter for; only polygons matching the filter value are rasterized
		null/empty	Use the attribute matching layer name if it exists; otherwise, use the first attribute in the layer if it has an attribute or lookup table or use the pixel value directly if it is a raster layer
lookup_table	No	string	The path to the lookup table to replace the original pixel or attribute values with new ones; the lookup table should be a .csv file, in which, for rasters, the first column is a pixel value and the subsequent columns are the attributes to attach to it; for vectors, columns are paired so that the first column matches the original attribute name and values, and the second column provides substitutions
		null/empty	Search for <layer filename>.csv in the same directory as the config file and then search in the same directory as the spatial layer; it is also valid to have no lookup table
strict_lookup_table	No	boolean	Set “strict mode” for lookup tables associated with vector layers (default = false); “strict mode” means that only values that appear in the lookup table are included in the final rasterized layer; otherwise, the polygons are skipped and become “NoData”; “non-strict mode” uses the original attribute value when no substitution is provided in the lookup table
(any mojadata library Layer arguments)	No	any	Any additional keyword arguments to pass directly to the underlying “mojadata” library layer constructor
Section: <classifier definition>			
<layer definition>	Yes	(layer def keys)	The spatial layer corresponding to the classifier

Element	Required (Yes/No)	Value Type	Function
values_path	No	string	The path to file containing all possible values for the classifier
		null/empty	Use the yield table
values_col	No	string	The column name for reading from the classifier values file
		null/empty	If “values_path” only has a single column, use it; otherwise, search for the column name in “values_path”, matching the classifier name or spatial attribute; if not found, search for a column name in “values_path” where the column values intersect with this classifier's values from its spatial layer
yield_col	No	string	The column name or index linked to the classifier in the yield table
		int	
		null/empty	Use “values_col” if “values_path” is the same as the yield table path or search the yield table for the column matching the classifier name, or search for the column where the values are a subset of the values from “values_path”
Section: <bounding box>			
(layer def keys)	No	(layer def keys)	The layer definition for the bounding box
Section: <disturbance layer definition>			
year	No	string	The name of the attribute containing the year of disturbance; if a special "filename" keyword is used, parse the year from “filename”
		int	The specific year of disturbance for the entire layer
		null/empty	Search for an attribute named "year"; if not found, search for the first attribute where all the unique values parse into years; otherwise, parse the year from the “filename”
disturbance_type	No	string	The specific GCBM disturbance type for the entire layer
		null/empty	Search for first attribute where all unique values appear in the “tblDisturbanceTypeDefault” table in the AIDB
age_after	No	string	The name of the attribute containing the reset age
		int	The specific reset age for the entire layer
		null/empty	Search for an attribute named "age_after"; if not found and “regen_delay” is not present, do not create a transition rule; if “regen_delay” is present but “age_after” is not, throw an exception
regen_delay	No	string	The name of the attribute containing the regeneration delay
		int	The specific regeneration delay for the entire layer
		null/empty	Search for an attribute named "regen_delay"; if not found and “age_after” is not present, do not create a transition rule; otherwise, the default is set to 0

Element	Required (Yes/No)	Value Type	Function
Section: <rollback>			
age_distribution	Yes	string	The path to the age distribution JSON or Excel file for the rollback tool
inventory_year	Yes	int	The global inventory vintage year for the rollback tool
		string	The path to the inventory vintage layer for the rollback tool; this layer should be tiled with the other layers, with the resulting layer passed to the rollback
		<layer definition>	The full layer definition for the inventory vintage layer
rollback_year	No	int	The year to roll back to
		null/empty	Use 1990 as default value
prioritize_disturbances	No	boolean	Set “prioritize disturbances” mode in the rollback tool (the default is “false”)
		null/empty	Use “false” as the default value (prioritize inventory)
single_draw	No	boolean	Set “single_draw” mode in the rollback tool (the default is “false”); draws a single age per work unit requiring a draw from the age distribution instead of a draw per pixel
		null/empty	Use “false” as the default value (per-pixel age draw)
establishment_disturbance_type	No	string	The default establishment disturbance type for pixels that require one to be created due to a lack of historic disturbance information (the default is “Wildfire”)
		null/empty	Use “Wildfire” as the default value

Abbreviations: AIDB, Archive Index Database; GCBM, Generic Carbon Budget Model; N/A, not applicable.