Title: Peat profile database from peatlands in Canada

Authors:

Ilka E. Bauer[1]

Marissa A. Davies[2]

Kelly A. Bona[3]

Oleksandra Hararuk[4]

Cindy H. Shaw[4*]

Daniel K. Thompson[5]

Werner A. Kurz[6]

Kara L. Webster[5]

Michelle Garneau[7]

Jim W. McLaughlin[8]

Maara S. Packalen[8]

Emily Prystupa[8]

Nicole K. Sanderson[7]

Charles Tarnocai[10*]

Author Affiliations:

[1]Independent Contractor, Victoria, BC, Canada

[2]Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON, Canada

[3]Environment and Climate Change Canada, Gatineau, QC, Canada

[4]Canadian Forest Service, Northern Forestry Centre, Edmonton, AB, Canada

[5]Canadian Forest Service, Great Lakes Forestry Centre, Sault Ste Marie, ON, Canada

[6]Canadian Forest Service, Pacific Forestry Centre, Victoria, BC, Canada

[7]Département de Géographie, Université du Québec a Montréal, Montréal, QC, Canada

[8]Ontario Forest Research Institute, Ministry of Natural Resources and Forestry, Sault Ste. Marie, ON, Canada

[9]Hugh John Flemming Forestry Centre, New Brunswick Department of Natural Resources and Energy Development, Fredericton, NB, Canada

[10]Agriculture and Agri-Food Canada, Ottawa, ON, Canada

*retired

Corresponding Author:

Kara L. Webster: kara.webster@NRCan-RNCan.gc.ca

Open Research statement:

The complete data set is available as Supporting Information at [*to be completed at proof stage*]. Associated data will also be available on Open Government Canada (https://open.canada.ca/en) upon publication.

Class I. Data Set Descriptors

A. Data set identity:

Peat profile database from peatlands in Canada

B. Data set identification code:

N/A

C. Data set description:

1. Originators:

Ilka E. Bauer, Independent Contractor, PO Box 42086, Victoria RPO Oak Bay, British Columbia, Canada, V8R 6T4

Marissa A. Davies, Department of Geography and Environmental Management, 200 University Avenue W, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

Kelly A. Bona, Pollutant and Inventory Reporting Division, Environment and Climate Change Canada, 351 Saint-Joseph Boulevard, Gatineau, Quebec, K1A 0H3

Oleksandra Hararuk, Canadian Forest Service, Northern Forestry Centre, 5320 122 Street NW, Edmonton, Alberta, Canada, T6H 3S5

Cindy H. Shaw, Canadian Forest Service, Northern Forestry Centre, 5320 122 Street NW, Edmonton, Alberta, Canada, T6H 3S5 (retired)

Daniel K. Thompson, Canadian Forest Service, Great Lakes Forestry Centre, 1219 Queen St. East, Sault Ste. Marie, Ontario, Canada, P6A 2E5

Werner A. Kurz, Canadian Forest Service, Pacific Forestry Centre, 506 Burnside Road W, Victoria, British Columbia, Canada, V8Z 1M5

Kara L. Webster, Canadian Forest Service, Great Lakes Forestry Centre, 1219 Queen St. East, Sault Ste. Marie, Ontario, Canada, P6A 2E5, kara.webster@NRCan-RNCan.gc.ca

Michelle Garneau, Département de Géographie, Local A-4275, 1255, St-Denis, Université du Québec a Montréal, Montréal, QC, Canada, H2X 3R9

Jim W. McLaughlin, Ontario Forest Research Institute, Ministry of Northern Development, Mines, Natural Resources and Forestry, 1235 Queen Street E, Sault Ste. Marie, Ontario, Canada, P6A 2E5

Maara S. Packalen, Ontario Forest Research Institute, Ministry of Northern Development, Mines, Natural Resources and Forestry, 1235 Queen Street E, Sault Ste. Marie, Ontario, Canada, P6A 2E5

Emily Prystupa, Hugh John Flemming Forestry Centre, New Brunswick Department of Natural Resources and Energy Development, P.O Box 6000, 1350 Regent Street, Fredericton, New Brunswick, Canada, E3C 2G6

Nicole K. Sanderson, Département de Géographie, Local A-4275, 1255, St-Denis, Université du Québec a Montréal, Montréal, QC, Canada, H2X 3R9

Charles Tarnocai, Agriculture and Agri-Food Canada, 960 Carling Ave., Ottawa, ON, Canada, K1A 0C6 (retired)

2. Abstract:

Peatlands cover approximately 12% of the Canadian landscape and play an important role in the carbon cycle through their centennial to millennial-scale storage of carbon under waterlogged and anoxic conditions. In recognizing the potential of these ecosystems as natural climate solutions and therefore the need to include them in national greenhouse gas inventories, the Canadian Model for Peatlands module (CaMP v. 2.0) was developed by the Canadian Forest Service. Model parameterization included compiling peat profiles across Canada to calibrate peat decomposition rates from different peatland types, to define typical bulk density profiles, and to describe the hydrological (i.e., water table) response of peatlands to climatic changes. A total of 1,217 sites were included in the dataset from published and unpublished sources. The CORESITES table contains site location and summary data for each profile, as well as an estimate of total carbon mass per unit area (megagrams C ha^{-1}). Total carbon mass per unit area at each location was calculated using bulk density and carbon content through each profile. The PROFILES table contains data for depth (cm), bulk density (g cm^{-3}), ash and carbon content (%), and material descriptions for contiguous samples through each peat profile. Data gaps for bulk density and C content were filled using interpolation, regression trees, and assigned values based on material description and/or soil classification to allow for the estimation of total carbon mass per unit area. A subset of the sites (N = 374) also has pH and pore water trace-elemental geochemistry data and are found in the WATER table. The REFERENCES table contains the full citation of each source of the data and is linked to each core location through the SOURCEDATA table. The LOOKUP table defines codes in the database that required more space than what was sufficient in the metadata tables. The data can be accessed on Open Government Canada and will be useful for future work on carbon stock mapping and ecosystem modelling.

D. Key words/phrases:

ash content, biogeochemistry, bulk density, Canada, carbon content, peatlands

Class II. Research origin descriptors

A. Data compilation sources:

This database contains stratigraphic data (bulk density, ash, and carbon content) on 1,217 peat profiles located across Canada that were compiled from published and unpublished studies of 862 peatland sites [SOURCE_SITEID]. It also contains water chemistry data from a subset of peat profiles [374]. Locations of each profile can be found in the CORESITES table of the database. The profiles included in the database are summarized in Figures 1 and 2. Data was compiled from the following sources:

- Peat profile and water chemistry data from Zoltai *et al.* (2000)– "Zoltai data" [566 profiles].

- Peat profile data from 'Holocene perspectives on peatland biogeochemistry', an online peat core database originally hosted by Lehigh University. The cores in this database are now hosted on Pangaea https://www.pangaea.de. – "Lehigh data" [45 profiles].

- Peat profile data from soil survey reports (1970s to 1980s) from provincial and federal government agencies - "Soil Survey data" [83 profiles].

- Unpublished peat profile data provided by Charles Tarnocai that were included in the Soil Landscapes of Canada and Peatlands of Canada map products (Tarnocai *et al.*, 2011; Agriculture and Agri-food Canada, 2013) – "Tarnocai data" [154 profiles].

- Published and unpublished peat profile data for Ontario provided by Dr. Maara Packalen (Ontario Ministry of Natural Resources and Forestry) and from Kettles *et al.* (2000) – "Ontario data" [53 profiles].

- Published and unpublished peat profile data for Québec provided by Dr. Michelle Garneau (Université du Québec à Montreal) – "Québec data" [72 profiles].

- New Brunswick Peatland Database of Keys and Henderson (1988) and Thibault (1992) that was updated and revised for this database post-publication – "New Brunswick data" [244 profiles].

For full source information for individual profiles within these data compilation sources, see the REFERENCES table that is part of the database.

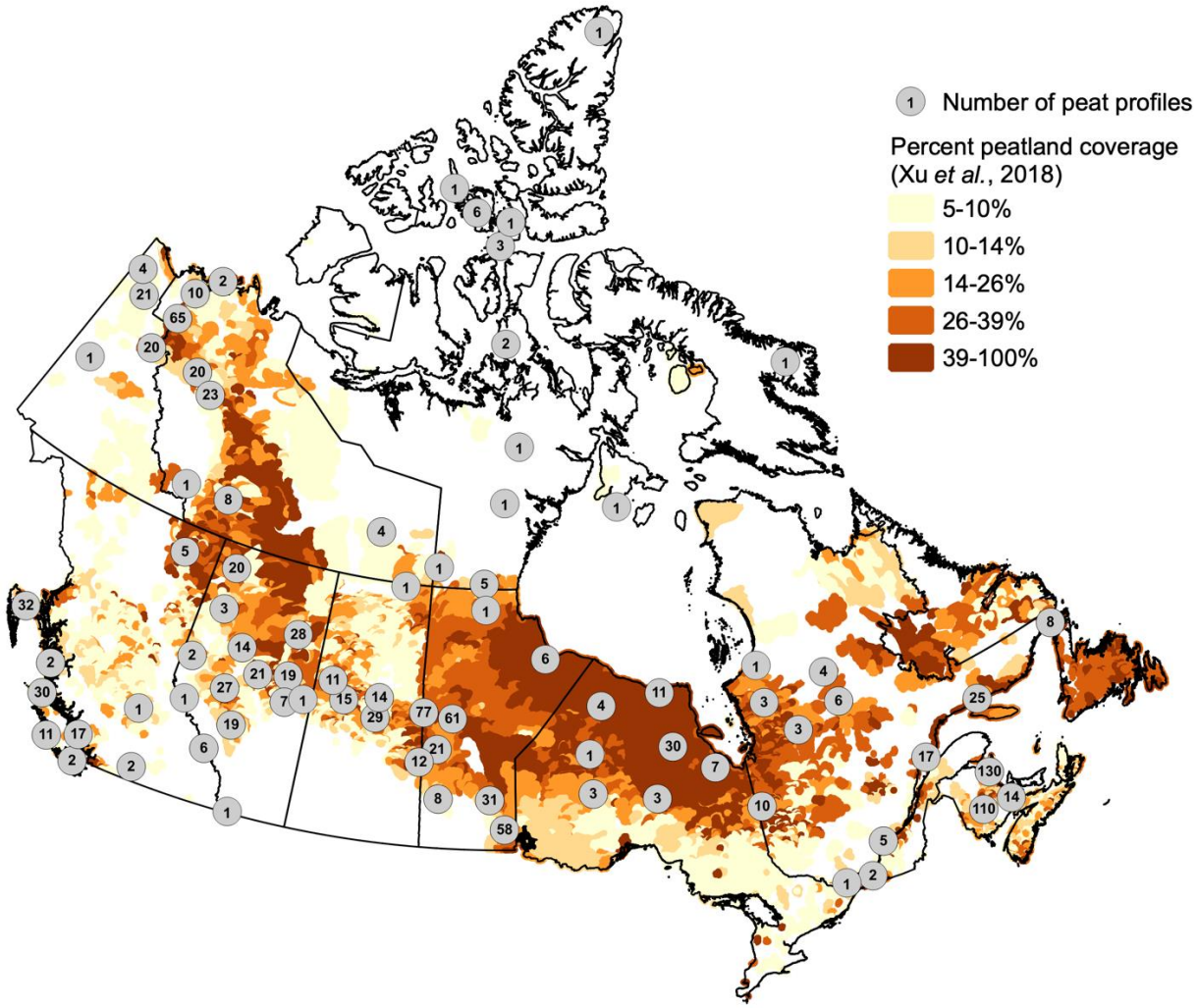

Figure 1: Map of Canada showing the location of the 1,217 peat profiles included in this database and the percent peatland coverage (Xu *et al*., 2018). Map was created by M. Davies using QGIS software (QGIS Development Team, 2020).

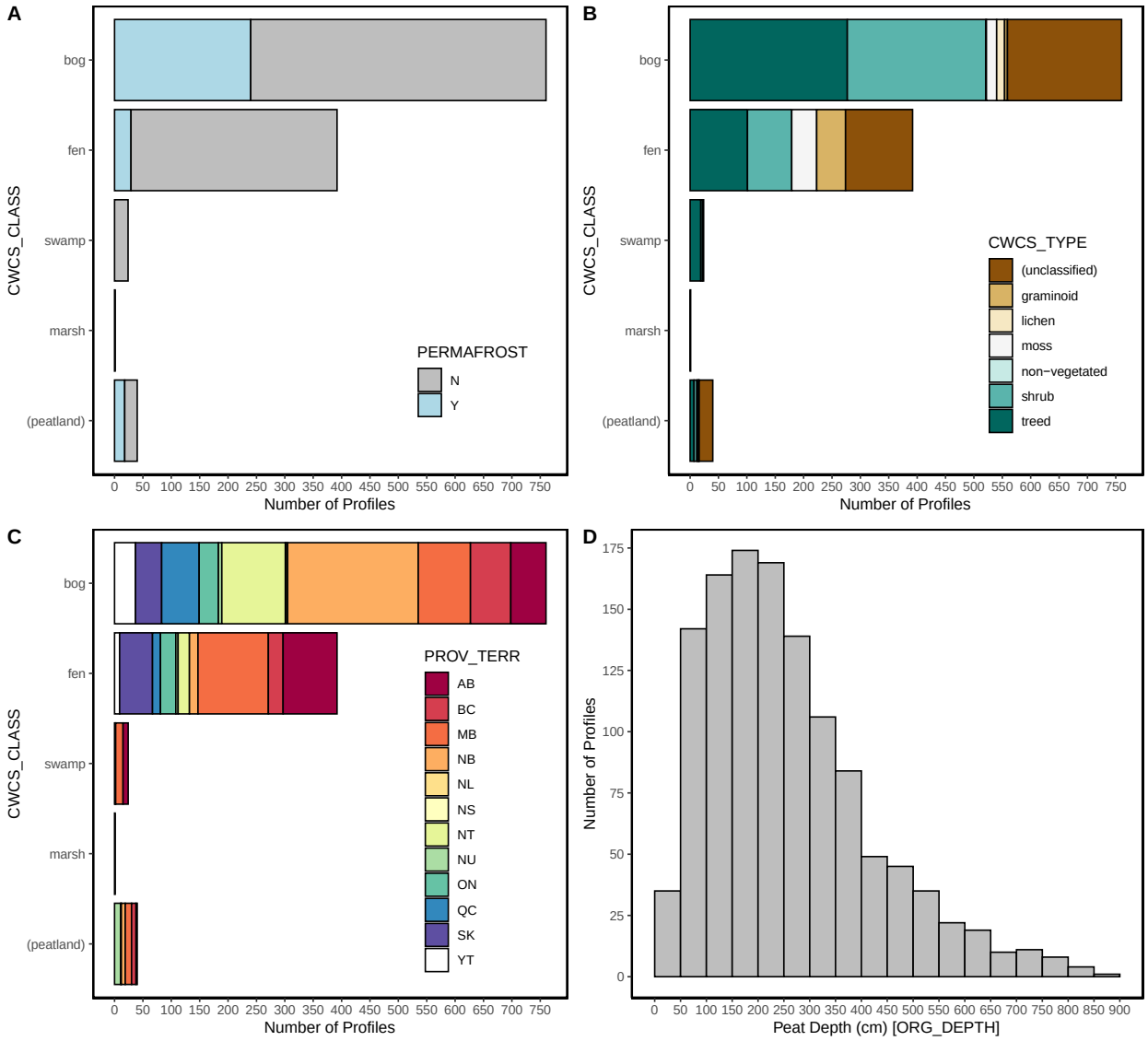

Figure 2: Summary of the peat profiles included in this database. **A.** Number of profiles with permafrost present (Y= yes, N = no), sorted by Canadian Wetland Classification System (CWCS) class. Profiles in the (peatland) category are unclassified in the database (value = NULL). **B.** Number of profiles in each CWCS type, sorted by CWCS class. Profiles in the (peatland) category are unclassified in the database (value = NULL). Profiles in the (unclassified) CWCS type are also unclassified in the database (value = NULL). **C.** Number of profiles in each province and territory (PROV_TERR) sorted by CWCS class. **D.** Number of profiles with a given peat depth (ORG_DEPTH). Bins are in 50 cm increments.

2. Statistical Methods:

To report total soil organic carbon mass per unit area of each soil profile in the database [ORC_C_MGHA variable in the PROFILES table], gap filling methods were used to predict missing bulk density [BULK_DENSITY] and carbon content values [C_TOT_PCT] in the PROFILES table of the database. Prediction of bulk density values was done using two different methods (as indicated in the BD_EST_TYPE variable in the PROFILES table):

1. Interpolation [LINEAR] and Extrapolation [EXT]

Extrapolation and interpolation were used when bulk density was measured in the surrounding samples and the gap between those samples was ≤ 3 cm for extrapolation and ≤ 10 cm for interpolation. It also was only used if there was no change in material between the two samples. A total of 19 samples were predicted using extrapolation and 262 using linear interpolation.

2. Regression Tree [BD_TREE_ORGANIC and BD_TREE_MINERAL]

A regression tree was used when the gap between samples with bulk density was greater than the threshold for extrapolation and interpolation or no bulk density data was available for the profile. Two regression trees (ANOVA-type) were constructed using the 'rpart' function within the 'rpart' package in R (Figure 3; Therneau *et al.*, 2022), one for mineral soils and another for organic soils. The training data was composed of all samples (without case weights) with measured bulk density available in the database, even when other data was missing, since regression trees are robust for partially missing data (N = 882 MINERAL and N = 33,386 ORGANIC and N = 133 N/A in SAMP_OM_CSSC. N/A samples included were MA (marl) samples in MATERIAL_1 and were split into organic or mineral based on their C_ORG_PCT being >17%, i.e., N=72 mineral and N = 61 organic). The final model was selected automatically by the algorithm (with no manual trimming) and included ash content [ASH], sample depth [MIDDEPTH = UPPER_SAMP_DEPTH + SAMP_THICK/2], wetland class [CWCS_CLASS], and material type [MATERIAL_1] as predictors. The attributes permafrost presence [PERMAFROST], Van Post scale of decomposition [VON_POST], and total carbon content [C_TOT_PCT] were also included as inputs into the regression tree, but not kept by the algorithm in the final model. Leave-one-out cross-validation was then performed at the profile level, which showed the regression trees had a mean absolute error of 0.038 g cm^{-3} and a median percent error of 29%. A total of 130 samples were predicted using BD_TREE_MIN and 1,854 using BD_TREE_ORG.

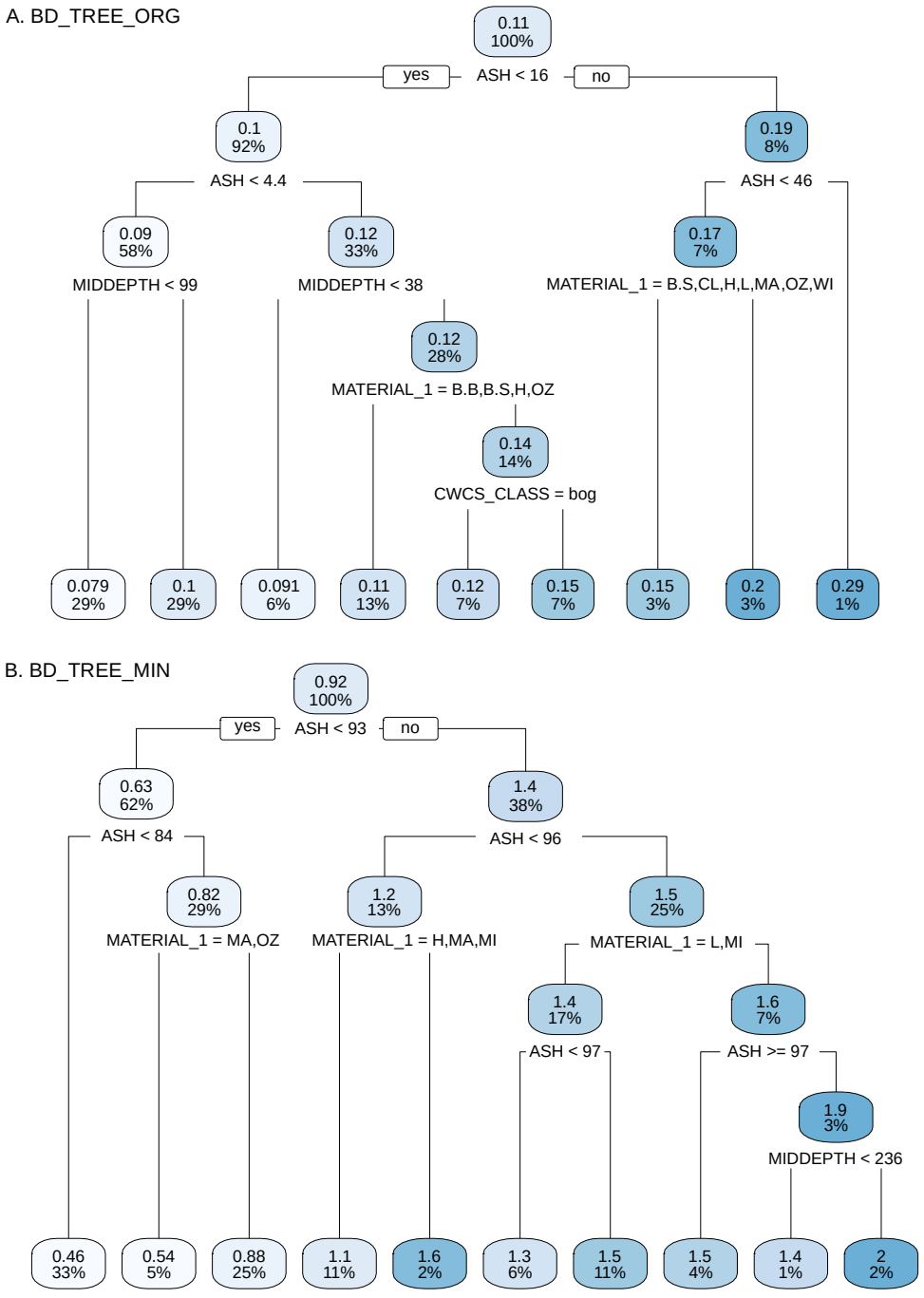

Figure 3: Regression-tree model used to gap-fill the BULK_DENSITY variable for (A) organic and (B) mineral samples in the PROFILES table of the database. For material codes, please refer to the LOOKUP table entries for MATERIAL_1. At each node, the left branch corresponds to the statement being true (yes) while the right branch is false (no). Within each node (blue shape), the top number represents the mean bulk density of that node (g cm^{-3}), while the bottom number is the % of the total data that falls into that node.

Prediction of total percent carbon [C_TOT_PCT] was based on five different methods which were selected based on the available data within each profile. The mean absolute error for value predictions of C_TOT_PCT based on the below methods was 3.6% for the N=2,741 samples for which measured values of C_TOT were available.

1. Interpolation [LINEAR] and Extrapolation [EXT]

Extrapolation and interpolation were used when total percent carbon was measured in the surrounding samples and the gap between those samples was ≤ 3 cm for extrapolation and ≤ 10 cm for interpolation. Interpolation was also used if the surrounding samples were estimated by ash content [C/OM; see below] and that gap was ≤ 5 cm. Both were also only used if there was no known change in material between the two samples. A total of 20 samples were estimated using extrapolation and 1,251 using linear interpolation.

2. Ash Content [C/OM]

If the ash content of the sample was available, total percent carbon was predicted using the following equation (Bauer *et al.*, 2006):

$$C\_TOT\_PCT = (100 - ASH) * 0.522 \quad (1)$$

where 0.522 is the median ratio of total carbon to organic matter content when both ASH and C_TOT_PCT are measured in a sample from the dataset (N = 2,089). This value was updated from Bauer *et al.* (2006) to include data from eastern Canadian sites. A total of 28,490 samples were estimated using this method.

3. Material [MATERIAL_1_MED]

The sample material code was used to assign C_TOT_PCT when ash content was not available for the sample. The value was based on the median value for that material type in the database where C_TOT_PCT was measured. Values are listed in Table 1. A total of 1,527 samples were estimated using this method.

Table 1: Total carbon content values [C_TOT_PCT] for each MATERIAL_1 code in the database. The value for B (moss) is the median for B, B.B. and B.S. codes.

MATERIAL_1	Definition	N	C_TOT_PCT (%)
B	Moss	1291	47.94
B.B	moss (non-*Sphagnum*)	78	48.19
B.S	moss (*Sphagnum*)	1198	48.00
H	Herbaceous	517	49.97
HU	Humus / highly humified	110	45.20
L	Wood	176	46.22
MA	Marl	29	22.23
OZ	Ooze / lacustrine	92	33.16
Mineral (all types)	-	122	3.77

4. Canadian System of Soil Classification (CSSC) Horizon Designation [CSSC_MED]:

The Canadian System of Soil Classification (CSSC) horizon designation was used when ash content and material type were not available for the sample. The value was based on the median value for that CSSC horizon designation in the database. Values are listed in Table 2. A total of 48 samples were estimated using this method.

Table 2: Total carbon content values for each CSSC_HORIZON code in the database.

CSSC_Horizon	Definition	N	C_TOT_PCT (%)
Of or F	Fibric organic horizon	48	41.65
Oh	Humic organic horizon	84	47.90
Om	Mesic organic horizon	137	43.60
A or B or C	Mineral horizons	48	4.10

5. Average Carbon Content [OVERALL_MED]

If there was no sample taken (but peat is assumed to be present) and there was no ash content or material information available for the sample, the median value for all organic samples in the database was used (48.45%; N = 4,709). A total of 428 samples were estimated using this method.

Class III. Data set status and accessibility

A. Status

1. Latest update:

   January 26, 2024

2. Latest archive date:

   January 26, 2024

3. Metadata status:

   Last updated January 26, 2024

4. Data verification:

   All data has been checked through QAQC procedures that are described in Class IV-B (Variable Information).

B. Accessibility

1. Storage location and medium:

   Dataset will be available on Open Government Canada upon publication and within the Supplementary Files.

2. Contact persons:

   Kelly A. Bona, Pollutant and Inventory Reporting Division, Environment and Climate Change Canada, 351 Saint-Joseph Boulevard, Gatineau, Quebec, K1A 0H3, Kelly.Bona@ec.gc.ca, 1-514-909-7123

   Daniel K. Thompson, Canadian Forest Service, Great Lakes Forestry Centre, 1219 Queen St. East, Sault Ste. Marie, Ontario, Canada, P6A 2E5, Daniel.thompson@NRCan-RNCan.gc.ca, 1-825-510-1226

   Kara L. Webster, Canadian Forest Service, Great Lakes Forestry Centre, 1219 Queen St. East, Sault Ste. Marie, Ontario, Canada, P6A 2E5, kara.webster@NRCan-RNCan.gc.ca, 1-705-987-0596

3. Copyright restrictions:

4. Citation:

Bauer, I.E., Davies, M.A., Bona, K.A., Hararuk, O., Shaw, C.H., Thompson, D.K., Kurz, W.A., Webster, K.L., Garneau, M., McLaughlin, J.W., Packalen, M.S., Prystupa, E., Sanderson, N.K., and Tarnocai, C., 2024, Peat profile database from peatlands in Canada.

5. Database updates:

This database will be updated periodically on Open Government Canada. Data can be submitted to be included in database updates through contacting the listed contact persons in Class III B (Accessibility).

Class IV. Data structural descriptors

A. Data set file

1. Identity

The dataset is distributed over six tables:

1. CORESITES

The CORESITES table contains basic information about the location of each of the 1,217 peat profiles. This includes geographic coordinates, wetland type, water table depth, and the presence or absence of permafrost. It also includes information about the how the profile was sampled, its carbon storage, vegetation present at the site as noted by the original authors, and whether it represents the total depth of organic deposits at the core location.

2. PROFILES

The PROFILES table contains information on 37,072 peat samples. Data include stratigraphic depth, bulk density, ash content, material type, and carbon content. Where measured data were unavailable, bulk density and carbon were modelled and gap-filled to calculate carbon storage for each organic sample (see Class II-B (Statistical Methods) for details).

3. WATER

The WATER table contains information on 514 water samples, each associated with a profile location. Data in this table includes pH, conductivity, and trace-elemental concentrations. Most of the information and all trace-elemental data are from a single source (i.e., “Zoltai data”; Zoltai *et al.*, 2000).

4. SOURCEDATA

The SOURCEDATA table links the CORESITES and the REFERENCE tables. It lists known sources for each core location and associated profile, as well as the kind of information (if any) that was taken from each source in compiling the database.

5. REFERENCES

The REFERENCES table provides the full citation for each source. It also contains links to online versions of sources for which a free pdf could be located.

6. LOOKUP

   The LOOKUP table defines codes used in the database that were too complex to be included in the metadata tables.

2. Size:

   Compressed .zip folder 765 KB

3. Format and storage mode:

   Comma-separated values

B. Variable information

VARIABLE INFORMATION	
Table CORESITES [1,217 records]	
Variable identity	CORE_ID
Variable definition	Unique identifier for each core
Units of measure	Not applicable
Data type	
1. Storage type	Integer
2. Variable codes	Not applicable
3. Range for numeric values	1-1217
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	SOURCE_SITEID
Variable definition	Site name or other descriptor used in the source publication or database Because sources differ in their sampling strategies and naming conventions, information in fields SOURCE_SITEID and SOURCE_SUB differs between studies and may be repetitive rather than complementary Indexing in the New Brunswick Database [CORE_ID 974-1217] uses both a site + subsite model [called PEATLAND_NUM and SAMPLE_LOCATION_NUM in the NB tables] and a code unique to each profile [called SITES_ID in the NB tables]. Entries in SOURCE_SITEID are composites of the NB SITES_ID and the NB PEATLAND_NUM For the Zoltai data [CORE_ID 1-566], Z- was added in front of the SOURCE_SITEID and SOURCE_SUB to prevent conversion to date format when opening the file in Microsoft Excel.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	SOURCE_SUB
Variable definition	Profile name or other descriptor used in the source publication or database

	Because sources differ in their sampling strategies and naming conventions, information in fields SOURCE_SITEID and SOURCE_SUB differs between studies and may be repetitive rather than complementary Indexing in the New Brunswick Database [CORE_ID 974-1217] uses both a site + subsite model [called PEATLAND_NUM and SAMPLE_LOCATION_NUM in the NB tables] and a code unique to each profile [called SITES_ID in the NB tables]. Entries in SOURCE_SITEID are composites of the NB SITES_ID and the NB SAMPLE_LOCATION_NUM For the Zoltai data [CORE_ID 1-566], Z- was added in front of the SOURCE_SITEID and SOURCE_SUB to prevent conversion to date format when opening the file in Microsoft Excel.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	no missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	LATITUDE
Variable definition	Core location latitude
Units of measure	Decimal degrees
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	45.008-82.1
4. Missing value codes	No missing values
5. Precision	0.1-0.0001 Also see LOC_SOURCE_FORMAT and LOC_ACCURACY_NOTES
Data format	
1. Fixed, variable length	Variable
QA/QC	Ensured location is within PROV_TERR boundaries
Variable identity	LONGITUDE
Variable definition	Core location longitude
Units of measure	Decimal degrees
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	(-140.2)-(-55.927)
4. Missing value codes	No missing values
5. Precision	0.1-0.0001

	Also see LOC_SOURCE_FORMAT and LOC_ACCURACY_NOTES
Data format	
1. Fixed, variable length	Variable
QA/QC	Ensured location is within PROV_TERR boundaries
Variable identity	LOC_SOURCE_FORMAT
Variable definition	The format of location data in the source dataset This field is meant to be indicative of the potential maximum precision of co-ordinates given in the LATITUDE and LONGITUDE fields. It assumes that researchers are aware of the limitations of their own location data and report co-ordinates accordingly. If researchers reported data at a higher precision than is warranted given the equipment or methodology used, information in this field will not provide a reliable precision estimate. In some cases, coordinates are supplemented by actual precision estimates (see LOC_ACCURACY_NOTES)
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Codes defined in LOOKUP table
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
QA/QC	Codes LOOKUP table are the only permissible values in CORESITES table for LOC_SOURCE_FORMAT
Variable identity	LOC_ACCURACY_NOTES
Variable definition	Additional notes on location accuracy
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 274)
Data format	
1. Fixed, variable length	Variable
Variable identity	PROV_TERR
Variable definition	Code for the Canadian province or territory the profile is in
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Codes defined in LOOKUP table
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Fixed

QA/QC	Codes in LOOKUP table are the only permissible values in SITES table for PROV_TERR Ensured location is within PROV_TERR boundaries
Variable identity	SAMPLING_YR
Variable definition	The year in which the profile was sampled in the field
Units of measure	Years CE
Data type	
1. Storage type	Integer
2. Variable codes	Not applicable
3. Range for numeric values	1967-2017
4. Missing value codes	Null (n = 106)
Data format	
1. Fixed, variable length	Fixed
QA/QC	Ensured no date is in the future Double-checked all dates older than 1970 against the source.
Variable identity	SAMPLING_MONTH
Variable definition	The month in which the profile was sampled in the field
Units of measure	Not applicable
Data type	
1. Storage type	Integer
2. Variable codes	6: June 7: July 8: August
3. Range for numeric values	6-8
4. Missing value codes	Null (n = 857)
Data format	
1. Fixed, variable length	Fixed
Variable identity	SAMPLING_DATE
Variable definition	The day in which the profile was sampled in the field
Units of measure	Not applicable
Data type	
1. Storage type	Integer
2. Variable codes	Not applicable
3. Range for numeric values	1-31
4. Missing value codes	Null (n = 857)
Data format	
1. Fixed, variable length	Fixed
Variable identity	SAMPLING_ST
Variable definition	Indicates the general strategy for how the profile was sampled. Sampling strategy is either taken from the source, or, if the source is silent, inferred from how the data are reported. If the uppermost sample for which information was reported by the source starts at 10 cm depth or lower, a number appended to the code indicates the depth at which reporting started. While missing sections were gap-filled using estimated values, bulk density tends to change rapidly at the top of the peat profile, so having to estimate large sections at

	the top of otherwise well-sampled profiles reduces the confidence in calculated ORG_C_MGHA. In the Lehigh dataset (CORE_ID 567-599 and 601-611), sample depths were reported as point depths rather than depth intervals. For these cores, SAMPLING_ST is coded as point-based, reflecting the information that was available for entry. This may be different from the sampling strategy used by the original researchers. For how upper and lower sample depths were assigned to such samples, see UPPER_SAMP_DEPTH and SAMP_THICK in the PROFILES table.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	see LOOKUP table
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable length
QA/QC	Codes in LOOKUP table are the only permissible values in the CORESITES table for SAMPLING_ST
Variable identity	CSSC_CODE
Variable definition	Soil great group and subgroup abbreviation according to the Canadian System of Soil Classification. This field is only filled if soil taxonomy information is provided by the source. Taxonomy is based on the following reference: Soil Classification Working Group, 1998. The Canadian System of Soil Classification. (3rd Ed.) Publication 1646. Ottawa, ON, Agriculture and Agri-Food Canada. 187 pp. Manual available online at http://sis.agr.gc.ca/cansis/publications/manuals/1998-cssc-ed3/cssc3_manual.pdf.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n=983)
Data format	
1. Fixed, variable length	Variable
QA/QC	CSSC_CODEs used are only those found in the Canadian System of Soil Classification. All CORE_IDs with a CSSC_CODE have values listed in the CSSC_HORIZON column in the PROFILES table.
Variable identity	CSSC_ORDER

Variable definition	Soil order according to the Canadian System of Soil Classification. This field is only filled if soil taxonomy information is provided by the source. Taxonomy is based on the following reference: Soil Classification Working Group, 1998. The Canadian System of Soil Classification. (3rd Ed.) Publication 1646. Ottawa, ON, Agriculture and Agri-Food Canada. 187 pp. Manual available online at http://sis.agr.gc.ca/cansis/publications/manuals/1998-cssc-ed3/cssc3_manual.pdf.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	null (n=983)
Data format	
1. Fixed, variable length	Variable length
QA/QC	CSSC_ORDER names used were only those found in the Canadian System of Soil Classification. Order name matches assigned CSSC_CODE. All CORE_IDs with a CSSC_ORDER value have values listed in the CSSC_HORIZON column in the PROFILES table. Ensured that all sites classified as 'Cryosolic' had PERMAFROST = Y and ACTIVE_DPT was < 100 cm as defined by the Canadian System of Soil Classification. Ensured that all the sites classified as 'Organic' had an ORG_DEPTH > 40 cm as defined by the Canadian System of Soil Classification.
Variable identity	CSSC_GREAT_GROUP
Variable definition	Soil great group according to the Canadian System of Soil Classification. This field is only filled if soil taxonomy information is provided by the source. Taxonomy is based on the following reference: Soil Classification Working Group, 1998. The Canadian System of Soil Classification. (3rd Ed.) Publication 1646. Ottawa, ON, Agriculture and Agri-Food Canada. 187 pp. Manual available online at http://sis.agr.gc.ca/cansis/publications/manuals/1998-cssc-ed3/cssc3_manual.pdf.

Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	null (n=983)
Data format	
1. Fixed, variable length	Variable
QA/QC	Ensured that great group names used are only those found in the Canadian System of Soil Classification. Ensured that great group name matches assigned CSSC_CODE. Ensured that all CORE_IDs with a CSSC_GREAT_GROUP have values listed in the CSSC_HORIZON column in the PROFILES table. Ensured that sites classified as a 'Fibrisol', 'Humisol', 'Folisol', and 'Mesisol' within the Organic CSSC_GREAT GROUP had soils with >17% organic carbon [C_ORG_PCT] to the ORG_DEPTH within the PROFILES table as defined by the Canadian System of Soil Classification. Ensured that sites classified as a 'Organic Cryosol' within the Cryosolic CSSC_GREAT GROUP had soils with >17% organic carbon [C_ORG_PCT] to the ORG_DEPTH within the PROFILES table as defined by the Canadian System of Soil Classification.
Variable identity	CSSC_GGSG
Variable definition	Soil great group subgroup according to the Canadian System of Soil Classification. This field is only filled if soil taxonomy information is provided by the source. Taxonomy is based on the following reference: Soil Classification Working Group, 1998. The Canadian System of Soil Classification. (3rd Ed.) Publication 1646. Ottawa, ON, Agriculture and Agri-Food Canada. 187 pp. Manual available online at http://sis.agr.gc.ca/cansis/publications/manuals/1998-cssc-ed3/cssc3_manual.pdf.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	null (n=983)
Data format	

1. Fixed, variable length	Variable length
QA/QC	Ensured that great group subgroup names used are only those found in the Canadian System of Soil Classification. Ensured that great group subgroup name matches assigned CSSC_CODE. Ensured that all CORE_IDs with a CSSC_GGSG have values listed in the CSSC_HORIZON column in the PROFILES table. Ensured that all sites classified as 'Terric' had an ORG_DEPTH < 120 cm for "Organic' soils and < 100 cm for 'Cryosolic' soils in the CSSC_ORDER variable column as defined by the Canadian System of Soil Classification. If BASE_Y_N = N and BASE_RSR = N, then it was assumed that the mineral contact was greater than these thresholds for classification.
Variable identity	ORG_DEPTH
Variable definition	Depth of the organic deposit at the sampling location ORG_DEPTH is the base of the lowest sample with known SAMP_THICK value that is classified as ORG-IN in field SAMP_OM_CSUM in the PROFILES TABLE. With very few exceptions, it is the lowest sample with known SAMP_THICK value that is classified as ORGANIC in field SAMP_OM_CSSC. ORG_DEPTH depth is determined within the database. It may differ from peat- or organic-matter depth value(s) given in source publications. Whether the ORG_DEPTH represents the entire peat profile at a site is indicated in the variable BASE_Y_N.
Units of measure	cm
Data type	
1. Storage type	double
2. Variable codes	Not applicable
3. Range for numeric values	40-851
4. Missing value codes	No missing values
5. Precision	0-0.1
Data format	
1. Fixed, variable length	Variable
QA/QC	Made sure information is consistent with UPPER_SAMP_DEPTH, SAMP_THICK, and SAMP_OM_CSSC in PROFILES table Made sure all profiles meet threshold for inclusion of at least 40 cm ORG_DEPTH.
Variable identity	BASE_Y_N

Variable definition	Indicates if the profile reaches the base of the organic deposit as defined within the database. The base of organic deposits is considered reached if one of the following conditions is met: The lowest sample coded ORGANIC in SAMP_OM_CSSC in the PROFILES table is underlain by a sample coded MINERAL. The lowest sample coded ORGANIC in SAMP_OM_CSSC in the PROFILES table is underlain by marl or water/ice, which, in turn, is underlain by a sample coded MINERAL. Because of the way SAMP_OM_CSSC is filled, identifying mineral material requires one of the following: Measured C_TOT_PCT or C_ORG_PCT Measured ASH Information on MATERIAL_1 Information on CSSC_HORIZON If none of this information is available, it is assumed that the basal samples did not reach the base of the organic deposit.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	"Y" if the core reached the base of the organic deposit "N" if the core did not reach the base of the organic deposit, or if is unclear if it did by the criteria used in this database.
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Fixed
QA/QC	Make sure information is consistent with UPPER_SAMP_DEPTH, SAMP_THICK, and SAMP_OM_CSSC in PROFILES table
Variable identity	BASE_RSR
Variable definition	Indicates if the profile reaches the base of the organic deposit as defined by the original researcher. Not all researchers clearly state their intent or criteria for separating organic and mineral material, so entries by necessity involve some interpretation. BASE_RSR is only populated if BASE_Y_N is "N",
Units of measure	Not applicable
Data type	
1. Storage type	String

2. Variable codes	"Y" if the researcher considered the core to represent the entire organic deposit "N" if the researcher indicated that the profile does not reach mineral material. "U" if it is unknown if the researcher considered the core to represent the entire organic deposit. "DNT" if reaching the base of the peat was never an objective of the study. Cores in this category often contain surficial peat only and should not be used as a measure of total organic matter depth.
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 769) where BASE_Y_N = Y
Data format	
1. Fixed, variable length	Variable
QA/QC	Ensured that field is null if BASE_Y_N is "Y" Ensured field is filled if BASE_Y_N is "N"
Variable identity	PERC_BD_MEAS
Variable definition	The proportion of profile depth included in calculating ORG_C_MGHA that has measured bulk density values. Only samples used in calculating ORG_C_MGHA are considered in PERC_BD_MEAS.
Units of measure	Percent
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0.0-100.0
4. Missing value codes	No missing values
5. Precision	0.1
Data format	
1. Fixed, variable length	Variable
Variable identity	PERC_C_MEAS
Variable definition	The proportion of profile depth included in calculating ORG_C_MGHA that has measured C_TOT_PCT and/or C_ORG_PCT values. Only samples used in calculating ORG_C_MGHA are considered in PERC_C_MEAS.
Units of measure	Percent
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0.0-100.0
4. Missing value codes	No missing values
5. Precision	0.1
Data format	
1. Fixed, variable length	Variable

Variable identity	PERMAFROST
Variable definition	Indicates if the profile location contained permafrost at the time of sampling.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	"Y" if permafrost was present "N" if no permafrost was present "U" if it is unknown whether permafrost was present
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Fixed
Variable identity	ACTIVE_DPT
Variable definition	For profiles where permafrost was present, depth of the active layer at the time of sampling. Since seasonal frost and permafrost are often hard to distinguish in the field, the value in this field technically represents depth to frost at the time of sampling in profiles where PERMAFROST = "Y".
Units of measure	cm
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0-180
4. Missing value codes	Null (n = 940) where permafrost absent or depth to frost unknown
5. Precision	0-0.1
Data format	
1. Fixed, variable length	Variable
QA/QC	Made sure all records with a value for ACTIVE_DPT have "Y" in PERMAFROST
Variable identity	WT_DEPTH
Variable definition	Depth of the water table at the time of coring Negative values indicate that the water level was above the surface.
Units of measure	cm
1. Data type	
2. Storage type	Integer
3. Variable codes	Not applicable
4. Range for numeric values	(-40)-105
5. Missing value codes	Null (n = 827) where no liquid water present or value unknown
Data format	
1. Fixed, variable length	Variable
Variable identity	ORG_C_MGHA

Variable definition	Total soil organic carbon mass per unit area of the soil profile, from the peat surface (UPPER_SAMP_DEPTH = 0) to ORG_DEPTH. Calculated as the sum of SAMP_CARB_MGHA for horizons identified as ORG-IN or MIN-IN in field SAMP_OM_CSUM of the PROFILES table.
Units of measure	Mg ha^{-1}
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	80.5-4889.0
4. Missing value codes	No missing values
5. Precision	0.1
Data format	
1. Fixed, variable length	Variable
QA/QC	Ensured no negative values Examined regression plots of ORG_C_MGHA on ORG_DEPTH and checked data for sites that were apparent outliers. QA/QC of data in the PROFILES table used to calculate SAMP_OM_CSUM
Variable identity	CWCS_CLASS
Variable definition	Wetland Class under the Canadian Wetland Classification System (CWCS). National Wetlands Working Group, 1997. The Canadian Wetland Classification System (2nd ed). Wetlands Research Centre, University of Waterloo, Waterloo, Ontario. Available online at https://nawcc.wetlandnetwork.ca/Wetland%20Classification%201997.pdf. Assignment to a wetland class generally follows the source if the original study was a wetland study. For soil science or non-specialist sources, the entire site and profile information, along with any designation made by the author(s) was considered in assigning a coring site to a CWCS_CLASS. Information associated with soil survey data (CORE_ID 612-694) was often general in nature, describing vegetation and environments associated with an entire soil series or type. To assign profiles from such sources to a CWCS_CLASS, descriptions were confirmed and/or supplemented with profile-level data, especially pH and peat material types. In cases where available information was ambiguous or conflicting, no CWCS_CLASS assignment was made.
Units of measure	Not applicable
Data type	
1. Storage type	String

2. Variable codes	Allowable entries are bog, fen, swamp, and marsh (also see LOOKUP table)
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 40)
Data format	
1. Fixed, variable length	Variable
QA/QC	Ensured entry is one of the allowable wetland classes
Variable identity	FEN_TYPE
Variable definition	The nutrient status of the CWCS_CLASS "fen" if reported by the original source. Nutrient status is generally defined by pH and electrical conductivity (EC) (see Alberta Environment and Sustainable Resource Development, 2015). Typical values are: Poor: EC < 100 μS cm^{-1} and pH <5.5 Moderately rich: 100-250 μS cm^{-1} and pH 5.5-7.0 Extremely rich: 250-2000 μS cm^{-1} and pH >7 Note that if original source did not distinguish between moderately or extremely rich, then it was classified as "rich" for fen type.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Allowable entries are poor, moderately rich, extremely rich, and rich (also see LOOKUP table)
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 958)
Data format	
1. Fixed, variable length	Variable
QA/QC	Ensured that only rows that only sites classified as "fen" in CWCS_CLASS have this value filled, if the information was available for the site.
Variable identity	CWCS_FORM
Variable definition	Wetland form and (if known) subform, under the Canadian Wetland Classification System (CWCS). National Wetlands Working Group, 1997. The Canadian Wetland Classification System (2nd ed). Wetlands Research Centre, University of Waterloo, Waterloo, Ontario. Available online at https://nawcc.wetlandnetwork.ca/Wetland%20Classification%201997.pdf. Wetland descriptions or classifications in source publications that do not conform to the CWCS were converted to CWCS forms if possible.
Units of measure	Not applicable
Data type	
1. Storage type	String

2. Variable codes	see LOOKUP table
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 453)
Data format	
1. Fixed, variable length	Variable
QA/QC	Ensured that form and subform names in CWCS_LOOKUP are only those found in the Canadian Wetland Classification System. Ensured codes in CWCS_LOOKUP are the only permissible values CWCS_FORM Ensured CWCS_FORM is an allowable wetland form under CWCS_CLASS of the same record. (If the CWCS_FORM identified occurs under more than one CWCS_CLASS, CWCS_CLASS may be null)
Variable identity	CWCS_TYPE
Variable definition	Wetland type under the Canadian Wetland Classification System (CWCS). National Wetlands Working Group, 1997. The Canadian Wetland Classification System (2nd ed). Wetlands Research Centre, University of Waterloo, Waterloo, Ontario. Available online at https://nawcc.wetlandnetwork.ca/Wetland%20Classification%201997.pdf. CWCS_TYPE contains information about the dominant vegetation at the site. Sites were assigned their original classification for CWCS_TYPE if the original data collectors provided this information. IF CWCS_TYPE was not assigned and vegetation percent cover data was available for the site, then it was assigned at CWCS_TYPE based on the following criteria (after National Wetlands Working Group, 1997and Alberta Environment and Sustainable Resource Development, 2015): If trees > 5m tall were greater than or equal 25% coverage, then the site was classified as treed. Sites were then classified into a subtype if the dominant tree species was known (coniferous or hardwood). If trees > 5m tall were less than 25 % and shrub cover was greater than or equal to 25% then the site was classified as shrub. Sites were then classified into a subtype if the dominant shrub species were low (<0.5 m tall) or tall (>0.5 m tall). If >25% of low and tall species were present, then the site was considered mixed. If tress < 5m tall and shrubs were less and 25% then the site was assigned to graminoid, lichen, or moss, based on which of these categories has the highest percent ground cover. Then subtypes were assigned based on which vegetation group

	within that type dominated if in the moss or graminoid category. Those were the following: moss: Sphagnum or brown graminoid: sedge, tall rush, or low rush If the site has < 5% vegetation cover, it was assigned to the non-vegetated category.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	see LOOKUP table
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 345)
Data format	
1. Fixed, variable length	Variable
QA/QC	Ensured that wetland types are only those found in the Canadian Wetland Classification System. Ensured codes in CWCS_LOOKUP are the only permissible values CWCS_TYPE
Variable identity	TREED
Variable definition	Site is considered treed under the Canadian Wetland Classification System (CWCS). National Wetlands Working Group, 1997. The Canadian Wetland Classification System (2nd ed). Wetlands Research Centre, University of Waterloo, Waterloo, Ontario. Available online at https://nawcc.wetlandnetwork.ca/Wetland%20Classification%201997.pdf. A site is typically considered treed when there is 25% cover of trees > 5 m in height (National Wetlands Working Group, 1997; Alberta Environment and Sustainable Resource Development, 2015).
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	"Y" if site is considered treed "N" if site is not considered treed "U" if it is unknown whether site would be considered treed
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Fixed
QA/QC	Ensured that all sites that have CWCS_TYPE "treed", "treed : coniferous", or "treed : hardwood" have the value "Y: in the TREED column and all other CWCS_TYPE values have "N".
Variable identity	CORE_POSITION

Variable definition	A description of where coring took place in relation to the entire peatland
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	"center" "edge, margin, or rim"
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n=1040)
Data format	
1. Fixed, variable length	Variable
Variable identity	DISTURBANCE
Variable definition	Description of disturbance(s) that occurred at the site. Note that additional details on the disturbance may be listed in the OTHER_DESC column. A null cell means that there was no noted disturbance at the site by the original source.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 1174)
Data format	
1. Fixed, variable length	Variable
Variable identity	MICROFORM
Variable definition	The specific peatland microform that the core was taken from when listed by the original source
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	"flark" "hummock" "lawn" "pool" "string"
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 1114)
Data format	
1. Fixed, variable length	Variable
Variable definition	OTHER_DESC
Units of measure	Additional descriptors used by the original author(s) to describe wetland type, disturbances, and other site characteristics that are not captured in the other columns within the CORESITES file .
Data type	
1. Storage type	String
2. Variable codes	Not applicable

3. Range for numeric values	Not applicable
4. Missing value codes	Null (n=1143)
Data format	
1. Fixed, variable length	Variable
Variable identity	VEGETATION
Variable definition	Vegetation noted as present at and/or surrounding the coring location from the original source. Note that some sites do not list all species found but only major or noteable species or broad vegetation groups.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 406)
Data format	
1. Fixed, variable length	Variable
QAQC	Ensured that vegetation names are consistent in each cell.
Table PROFILES [37,072 records]	
Variable identity	CORE_ID
Variable definition	Unique identifier for each core in the database. Used in conjunction with SAMPLE_NO to produce unique identifiers for the PROFILES table.
Units of measure	Not applicable
Data type	
1. Range for numeric values	Integer
2. Missing value codes	Not applicable
3. Data format	1-1217
4. Fixed, variable length	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	SAMPLE_NO
Variable definition	Rank position of the sample within its profile, starting at 1 for uppermost sample and increasing with depth. Used in conjunction with CORE_ID to produce unique identifiers for samples in the PROFILES table
Units of measure	Not applicable
Data type	
1. Storage type	Integer
2. Variable codes	Not applicable
3. Range for numeric values	1-605
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	UPPER_SAMP_DEPTH

Variable definition	The upper depth limit of the sample. At the surface of the peat profile UPPER_SAMP_DEPTH = 0. Values increase with depth in the profile. For CORE_IDs 567-599 and 601-611, the source data tables cited a point depth rather a depth range for each sample. For these profiles, boundaries between samples were assumed to fall half-way between the stated point depths of successive samples. See relevant entries in LOOKUP table for field SAMPLING_ST (CORESITES table) for further details. Some sample depths were reported in inches in the original data source. These values were converted and rounded to the nominal precision.
Units of measure	cm
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0-848
4. Missing value codes	No missing values
5. Precision	0-0.1
Data format	
1. Fixed, variable length	Variable
QA/QC	Ensured UPPER_SAMP_DEPTH for peat surface = 0 Ensured all values for UPPER_SAMP_DEPTH are positive Ensured that UPPER_SAMP_DEPTH + SAMP_THICK of a sample equal the UPPER_SAMP_DEPTH for the next (lower) sample.
Variable identity	SAMP_THICK
Variable definition	Vertical thickness of the sample A key requirement of QA/QC was that the SAMP_THICK values of a samples within a profile had to add up to UPPER_SAMP_DEPTH + THICK (if known) of the lowest sample. In other words, there could be no overlap or unexplained gaps between samples. Not all profiles met this requirement, so adjustments had to be made in a number of cases. For CORE_IDs 567-599 and 601-611 from the Lehigh data the source data tables provided a point depth rather a depth range for each sample. For these profiles, boundaries between samples were assumed to fall half-way between the stated point depths of successive samples, and the lower boundary (UPPER_SAMP_DEPTH + THICK) of the lowest sample in the core was assumed to equal either the total core length stated in the data tables, or the total core

	length given for the profile in another published source. If neither was available, the point depth given for the lowest sample was assumed to represent the middle of the sample. For CORE_IDs 1-566 from Zoltai *et al.*, 2000 (Zoltai SC, Siltanen RM and Johnson JD. 2000. A wetland data base for the western boreal, subarctic, and arctic regions of Canada. Natural Resources Canada, Canadian Forest Service, Northern Forest Centre, Information Report NOR-X-368. Edmonton, AB.), discrepancies, where present, were resolved by keeping the published UPPER_SAMP_DEPTH of affected samples and re-calculating SAMP_THICK. For 'missing' core sections, unless there was evidence to the contrary, it was assumed that peat had been present but not been sampled or analyzed, so 'dummy' samples were added and coded as gaps in MATERIAL_1. Some sample thicknesses were reported in inches in the original data source. These values were converted and rounded to the nominal precision.
Units of measure	cm
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0.5-338
4. Missing value codes	Null (n = 349). Allowed only for the bottom sample in a profile, and only if relevant data (e.g., CSSC_HORIZON or peat material type) is available for the sample.
5. Precision	0-0.1
Data format	
1. Fixed, variable length	Variable length
QA/QC	Ensured all SAMP_THICK values are positive Ensured that UPPER_SAMP_DEPTH + SAMP_THICK of a sample equal the UPPER_SAMP_DEPTH for the next (lower) sample.
Variable identity	BULK_DENSITY
Variable definition	The bulk density of the sample
Units of measure	g cm^{-3}
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0.001-2.484
4. Missing value codes	Null (n = 406) Used only for samples where BULK_DENSITY would have to be estimated, and SAMP_THICK is null and/or MATERIAL_1 is WI (water/ice)
5. Precision	0.1-0.001

Data format	
1. Fixed, variable length	Variable
QA/QC	Examined BULK_DENSITY by MATERIAL_1 and checked values more than 2 standard deviations from the mean against the source. Examined relationship between BULK_DENSITY and ASH and checked outliers against the source.
Variable identity	BD_MEAS_EST
Variable definition	Indicates if bulk density was measured or estimated (see BD_EST_TYPE)
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	"MEAS" if BULK_DENSITY was measured "EST" if BULK_DENSITY was estimated within the database. Estimated values from source data (if identified as such) were not used.
3. Range for numeric values	Not applicable
4. Missing value codes	null (n = 406) where BULK_DENSITY is null
Data format	
1. Fixed, variable length	Variable
QA/QC	Ensured that all samples coded EST have a corresponding entry in BD_EST_TYPE Ensured that samples coded MEAS or null in BD_MEAS_EST are null for BD_EST_TYPE
Variable identity	BD_EST_TYPE
Variable definition	The way BULK_DENSITY was estimated for samples where BD_MEAS_EST is "EST"
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Codes defined in LOOKUP table.
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 34,807) where BD_MEAS_EST is MEAS or null
Data format	
1. Fixed, variable length	Variable
QA/QC	Ensured that estimated BULK_DENSITY values are from the correct estimation source Made sure BD_EST_TYPE is null when BD_MEAS_EST is MEAS or Null
Variable identity	ASH
Variable definition	The ash content of the sample. For studies that report Loss on ignition (LOI) at 550°C or organic matter (OM), ASH was calculated as ASH (%) = 100 – LOI (%), or ASH (%) = 100 - OM (%)

	Measured ASH contents <0 were entered as zero
Units of measure	Percent
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0-99.7
4. Missing value codes	Null (n = 5,215)
5. Precision	0-0.1
Data format	
1. Fixed, variable length	Variable
QA/QC	Examined ASH by MATERIAL_1 and checked values more than 2 standard deviations from the mean against the source. Examined relationship between BULK_DENSITY and ASH and checked outliers against the source. Ensured no ash value is <0.
Variable identity	VON_POST
Variable definition	The humification of the sample as measured on the von Post scale. (Von Post, 1922. Sveriges geologiska undersökings torvinventering och några a dess hittills vunna resultat. Svenska Mosskulturföreningens Tidskrift 37: 1-27). For an explanation of the system in English, see e.g. Riley and Michaud, 1994. Ontario Peatland Inventory: Field-Work Methods. Ontario Geological Survey Miscellaneous Paper 55.
Units of measure	Von Post
Data type	
1. Storage type	Integer
2. Variable codes	Not applicable
3. Range for numeric values	1–10
4. Missing value codes	Null (n = 27,845)
Data format	
5. Fixed, variable length	Variable
QA/QC	Ensured all values are integers ≥ 1 and ≤ 10
Variable identity	MATERIAL_1
Variable definition	The primary (botanical or mineral) constituent of the sample. Different sources use different ways to describe peat material, so data had to be standardized to fit the material descriptors used in the database (see LOOKUP). Documenting details for each profile is impossible, but the three groups of cores for which material information was most widely available are as follows: CORE_IDs 1-566 from Zoltai *et al.*, 2000 (Zoltai SC, Siltanen RM and Johnson JD. 2000. A wetland data base for the western boreal, subarctic, and arctic regions of Canada. Natural Resources Canada, Canadian Forest Service, Northern Forest Centre, Information Report NOR-X-368. Edmonton, AB.) used a limited and consistent set of material descriptors that could be converted using simple decision rules.

	For CORE_IDs 612-848, material descriptors, where available, were usually part of soil-science-based profile descriptions that varied in format and completeness and often used descriptors such as “peat”, “rootlets”, or “amorphous” that were hard to reconcile with material types used in the database. In the result, MATERIAL_1, MATERIAL_2, and MATERIAL_3 differ in completeness, and in how well they describe the total material composition of the sample. CORE_IDs 974-1217 from the New Brunswick Peatland Database (Keys and Henderson, 1988. An investigation of the peat resources of New Brunswick. New Brunswick Department of Natural Resources. Open File Report 83-10. [Digital dataset supplied by New Brunswick DNRED, 26 January 2021]) used a limited and consistent set of material descriptors that could be converted using simple decision rules.
Units of measure	Not Applicable
Data type	
1. Storage type	String
2. Variable codes	See LOOKUP Table
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 22,684)
Data format	
1. Fixed, variable length	Variable
QA/QC	Codes in LOOKUP table are the only permissible values for MATERIAL_1 Flagged and checked against original data if MATERIAL_1 seemed inconsistent with SAMP_OM_CSSC. (E.g., MATERIAL_1 = H, while SAMP_OM_CSSC = MINERAL) Checked against source if MATERIAL_1-based QAQC for BULK_DENSITY or ASH flagged the sample.
Variable identity	MATERIAL_2
Variable definition	A second (botanical or mineral) constituent of the sample. See MATERIAL_1 for further information
Units of measure	Not Applicable
Data type	
1. Storage type	String
2. Variable codes	See LOOKUP Table
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 29,404)
Data format	
1. Fixed, variable length	Variable
QA/QC	Codes in LOOKUP table are the only permissible values for MATERIAL_2
Variable identity	MATERIAL_3
Variable definition	A second (botanical or mineral) constituent of the sample.

	See MATERIAL_1 for further information
Units of measure	Not Applicable
Data type	
1. Storage type	String
2. Variable codes	See LOOKUP Table
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 34,452)
Data format	
1. Fixed, variable length	Variable
QA/QC	Codes in LOOKUP table are the only permissible values for MATERIAL_3
Variable identity	CSSC_HORIZON
Variable definition	Soil horizon designation according to the Canadian System of Soil Classification. This field was only filled if soil taxonomy information is provided by the source. Taxonomy is based on the following reference: Soil Classification Working Group, 1998. The Canadian System of Soil Classification. (3rd Ed.) Publication 1646. Ottawa, ON, Agriculture and Agri-Food Canada. 187 pp. Manual available online at http://sis.agr.gc.ca/cansis/publications/manuals/1998-cssc-ed3/cssc3_manual.pdf.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 35,959)
Data format	
1. Fixed, variable length	Variable
QA/QC	CSSC_Horizon names used are only those found in the Canadian System of Soil Classification. CSSC_Horizon value(s) for a CORE_ID match the assigned CSSC_CODE in the CORESITES table. All CORE_IDs with a CSSC_CODE value in the CORESITES table have values listed in the CSSC_HORIZON column. Ensured that O horizons had an organic carbon content >17% defined by the Canadian System of Soil Classification. When a horizon was designated by the original source as O but had a C_ORG_PCT value that was < 17%, the sample was checked to ensure that it was designated as MINERAL in the SAMP_OM_CSSC and the original horizon classification from the source was retained.

Variable identity	C_TOT_PCT
Variable definition	Total carbon content of the sample, either measured by the source or estimated within the database No estimates of C_TOT_PCT were made for samples coded WI (profile water or ice), TEPH (Tephra) or RO (rock) in MATERIAL_1. Estimates of C_TOT_PCT for mineral samples based on MATERIAL_1 or CSSC_HORIZON are medians derived by pooling all samples with mineral MATERIAL_1 or CSSC_HORIZON types, respectively. Thus, they are very broad estimates. C_TOT_PCT estimates for samples containing marl (MATERIAL_1 = MA), are considered unreliable unless C_ORG_EST_TYPE = MATERIAL_1 MED or CSSC_MED. Other estimation methods (C/OM and OVERALL_MED) were optimized for organic samples and training data expressly excluded samples known to contain marl.
Units of measure	Percent
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0.02-64.25
4. Missing value codes	Null (n = 390) Used for samples where C_TOT_PCT would have to be estimated, and SAMP_THICK is null, and/or MATERIAL_1 is WI (profile water or ice), TEPH (tephra), or RO (rock), unless an ASH measurement is available. Also used for samples without ASH or MATERIAL_1 that could not be classified as either organic or mineral based on CSSC_HORIZON.
5. Precision	0-0.01
Data format	
1. Fixed, variable length	Variable
QA/QC	Checked measured values >55% against the source to verify. If both ASH and C_TOT_PCT had been measured, calculated the CT/organic matter ratio, and checked organic samples with extreme ratios against the source. If both C_TOT_PCT and C_ORG_PCT had been measured, checked values where C_ORG_PCT > C_TOT_PCT against the source.
Variable identity	C_TOT_MEAS_EST
Treed treedVariable definition	Indicates whether C_TOT_PCT was measured or estimated within the database.

	See notes under C_TOT_PCT and additional documentation for information on estimation methods and limitations of estimated values
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	"MEAS" if C_TOT_PCT was measured "EST" if C_TOT_PCT was estimated using the procedure in C_TOT_EST_TYPE
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 390) where C_TOT_PCT is null
Data format	
1. Fixed, variable length	Variable length
QA/QC	If "EST", ensured entries have valid entry in C_TOT_EST_TYPE that is explained in LOOKUP table
Variable identity	C_TOT_EST_TYPE
Variable definition	Indicates how C_TOT_PCT was estimated if C_TOT_MEAS_EST is "EST" See notes under C_TOT_PCT and additional documentation for information on estimation methods and limitations of estimated values.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Codes defined in LOOKUP table.
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 5,308) where C_TOT_MEAS_EST is "MEAS" or null
Data format	
1. Fixed, variable length	Variable
QA/QC	Ensured that values for C_TOT_PCT are from the correct C_TOT_EST_TYPE Ensured C_TOT_EST_TYPE is null when C_TOT_MEAS_EST is MEAS or null.
Variable identity	C_ORG_PCT
Variable definition	Total organic carbon content of the sample, either measured by the source or estimated within the database Measured C_ORG_PCT values that exceed ~60%, should be viewed with caution. While such values are reported in (especially older) literature and retained in the database, they are questionable given analyses such as those by Loisel *et al.* (2014) that examine peat properties using large datasets. (Loisel et al., 2014. A database and synthesis of northern peatland soil properties and Holocene carbon and nitrogen accumulation. The Holocene 24: 1028-1042). The database estimates C_ORG_PCT (if not directly measured) as C_ORG_PCT = C_TOT_PCT, i.e., it assumes

	that all carbon present in the sample is organic. This assumption is commonly made for peat. (See Chambers *et al.*, 2010. Methods for determining peat humification and for quantifying peat bulk density, organic matter and carbon content for palaeostudies of climate and peatland carbon dynamics. Mires and Peat 7(07): 1-10: http://www.mires-and-peat.net/pages/volumes/map07/map0707.php.) This assumption, however, is questionable for mineral samples where a significant portion of the carbon present may be inorganic. It is, by definition, inappropriate for marl, which contains high amounts of calcium carbonate. Estimated C_ORG_PCT values for these material types should thus be viewed with caution. Irrespective of whether a sample is organic or mineral, estimates of C_ORG_PCT in samples with pH > ~7 should be viewed with some skepticism. Most C_TOT_PCT values used to estimate C_ORG_PCT are themselves estimates. See notes under C_TOT_PCT and fields C_TOT_MEAS_EST and C_TOT_EST_TYPE for more information.
Units of measure	Percent
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
2. Range for numeric values	0.00-74.79
3. Missing value codes	Null (n = 407) Used for samples where C_ORG would have to be estimated, and SAMP_THICK is null, and/or MATERIAL_1 is WI (profile water or ice), TEPH (tephra), or RO (rock), unless an ASH measurement is available.
4. Precision	0-0.01
Data format	
1. Fixed, variable length	Variable
	Check measured values >55% against the source to verify. If both ASH and C_TOT_PCT had been measured, calculated the C/organic matter ratio, and checked samples with extreme ratios against the source. If both C_TOT_PCT and C_ORG_PCT had been measured, checked values where C_ORG_PCT > C_TOT_PCT against the source.
Variable identity	C_ORG_MEAS_EST
Variable definition	Indicates whether C_ORG_PCT was measured or estimated within the database

	See notes under C_ORG_PCT and additional documentation for information on estimation methods and limitations of estimated values.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	"MEAS" if C_ORG_PCT was measured "EST" if C_ORG_PCT was estimated using the procedure in C_ORG_EST_TYPE
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 407) where C_ORG_PCT is null
Data format	
1. Fixed, variable length	Variable length
QA/QC	If EST, ensured entries have valid entry in C_ORG_EST_TYPE that is explained in LOOKUP table
Variable identity	C_ORG_EST_TYPE
Variable definition	Indicates how C_ORG_PCT was estimated if C_ORG_MEAS_EST is "EST" See notes under C_ORG_PCT and additional documentation for information on estimation methods and limitations of estimated values.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Codes defined in LOOKUP table.
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 1143) where C_ORG_MEAS_EST is "MEAS" or null
Data format	
1. Fixed, variable length	Variable
QA/QC	Ensured that values for C_ORG_PCT are from the correct C_ORG_EST_TYPE Ensured C_ORG_EST_TYPE is null when C_ORG_MEAS_EST is MEAS or null
Variable identity	SAMP_CARB_MGHA
Variable definition	Organic carbon mass per unit area for the sample calculated as SAMP_CARB_MGHA = C_ORG_PCT/100 * BULK_DENSITY * SAMP_THICK * 100 See notes under C_TOT_PCT and C_ORG_PCT for limitations of estimated C values used to calculate this number. SAMP_CARB_MGHA values for samples with MATERIAL_1 = MA (marl) especially should be viewed with extreme skepticism unless C_ORG_PCT was measured directly (i.e., C_ORG_MEAS_EST = EST).
Units of measure	Mg ha^{-1}
Data type	
1. Storage type	Double
2. Variable codes	Not applicable

3. Range for numeric values	0.1-2392.8
4. Missing value codes	Null (n = 407) where SAMP_THICK is null, and/or MATERIAL_1 is WI (profile water or ice), TEPH (tephra), or RO (rock), unless an ASH measurement is available.
5. Precision	0.1
Data format	
1. Fixed, variable length	Variable
QA/QC	QA/QC of variables in the PROFILES table used to calculate SAMP_CARB_MGHA. Ensured SAMP_CARB_MGHA is only filled for samples where THICK is known
Variable identity	SAMP_OM_CSSC
Variable definition	Indicates whether the sample is organic or mineral according to definitions of the Canadian System of Soil Classification. (Soil Classification Working Group, 1998. The Canadian system of soil classification. (3rd Ed.) Publication 1646. Ottawa, ON, Agriculture and Agri-Food Canada. 187 pp. Manual available online at http://sis.agr.gc.ca/cansis/publications/manuals/1998-cssc-ed3/cssc3_manual.pdf) The CSSC uses a cutoff of 17% organic C to differentiate organic and mineral horizons, so a sample was classified as ORGANIC if it was not marl or water/ice (MATERIAL_1 = MA or WI) and had C_ORG_PCT (measured or estimated) > 17%. Because of the way C_ORG_PCT is filled in the database, this means a sample is classified as ORGANIC if one of the following is true for the sample: it had a measured C_ORG_PCT >17% if ORG_C_PCT was not measured, it had a measured C_TOT_PCT >17% if neither C_ORG_PCT nor C_TOT_PCT was measured, it had ASH < 67.5%. if measured C_ORG_PCT, measured C_TOT_PCT, and ASH were all unavailable, MATERIAL_1 was not a mineral material type if measured C_ORG_PCT, measured C_TOT_PCT, ASH and MATERIAL_1 were all unavailable, CSSC_HORIZON was not a mineral horizon type.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	“ORGANIC” if the sample is organic “MINERAL” if the sample is mineral “N/A” if the sample is water/ice or marl* *An exception was made for two samples that have MATERIAL_1 = WI but measured BULK_DENSITY and

	ASH. Based on ASH, these are organic samples and treated as such.
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable length
QA/QC	Made sure field is coded N/A for samples that are water or marl.
Variable identity	SAMP_OM_CSUM
Variable definition	Indicates how the sample is treated for the purpose of calculating field ORG_C_MGHA in the CORESITES table. Samples with MATERIAL_1 = MA (marl) are always excluded from ORG_C_MGHA, as are samples with MATERIAL_1 = WI (profile water or ice). Otherwise, organic samples (SAMP_OM_CSSC="ORGANIC") are usually included in calculating ORG_C_MGHA while mineral samples (SAMP_OM_CSSC= "MINERAL") are usually excluded. Exceptions from this rule were made on a sample-by-sample basis under a limited number of circumstances and while examining the entire profile for the purpose of setting ORG_DEPTH and BASE_Y_N in the CORESITES table. Mineral samples (SAMP_OM_CSSC="MINERAL") were included in calculating ORG_C_MGHA for the profile under two main scenarios: The sample was underlain by at least one organic sample (i.e., above ORG_DEPTH for the profile) and based on the weight of all evidence, it represented peat mixed with mineral or lacustrine deposits rather than a pure mineral layer interbedded with peat. The sample was part of a permafrost profile that had high mineral content throughout and was classified by the source as organic. For CORE_IDs 700-706 (lowland polygons), soil horizon designations by C. Tarnocai took precedence over ASH or C measurements in deciding whether to include or exclude a sample. Organic samples (SAMP_OM_CSSC="ORGANIC") were not included in calculating ORG_C_MGHA under two main scenarios: Isolated organic samples buried under massive (~>1m) mineral deposits at the core base. These samples were usually close to organic/mineral threshold, and in profiles where they occurred (CORE_IDs 121, 227, 448, 455), ORG_DEPTH was set to the top of the mineral section, i.e., ignoring the isolated organic layer

	Organic samples at the very base of the profile for which SAMP_THICK was unknown (n=10).
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	ORG-IN: SAMP_OM_CSSC is ”ORGANIC” and sample is included in calculating ORG_C_MGHA. ORG-EX: SAMP_OM_CSSC is ”ORGANIC”, but sample is not included in setting the core base or in calculating ORG_C_MGHA. MIN-IN: SAMP_OM_CSSC is ”MINERAL”, but sample is included in calculating ORG_C_MGHA. MIN-EX: SAMP_OM_CSSC is ”MINERAL”, and sample is excluded from calculating ORG_C_MGHA. N/A: Sample is water/ice or marl
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
QA/QC	Made sure information for lowest sample where SAMP_OM_CSUM = ORG_IN is consistent with ORG_DEPTH and BASE_Y_N in CORESITES table Made sure field is coded “N/A” for samples where MATERIAL_1=”MA” (marl) or “WI” (water/ice).
Variable identity	SAMP_PH
Variable definition	The pH value of the sample For the method by which pH was measured, see SAMP_PH_METHOD
Units of measure	pH
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	2.5-8.1
4. Missing value codes	Null (n = 30,018)
5. Precision	0.1
Data format	
1. Fixed, variable length	Variable
QA/QC	Check source data if SAMP_PH differs by a pH unit or more from BOTH the over- and the underlying sample.
Variable identity	SAMP_PH_METHOD
Variable definition	Indicates the method used to measure pH
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	See LOOKUP table

3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 30,018)
Data format	
1. Fixed, variable length	Variable
QA/QC	Only codes listed in LOOKUP table are permissible in SAMP_PH_METHOD Ensured all samples with a measured pH value have an entry in SAMP_PH_METHOD
Variable identity	SAMP_CA
Variable definition	Calcium concentration of the sample All SAMP_CA measurements are from either Zoltai *et al.*, 2000 (Zoltai SC, Siltanen RM and Johnson JD. 2000. A wetland data base for the western boreal, subarctic, and arctic regions of Canada. Natural Resources Canada, Canadian Forest Service, Northern Forest Centre, Information Report NOR-X-368. Edmonton, AB.) or the New Brunswick peatland database (Keys and Henderson, 1988. An investigation of the peat resources of New Brunswick. New Brunswick Department of Natural Resources. Open File Report 83-10. [Digital dataset supplied by New Brunswick DNRED, 26 January 2021])
Units of measure	mg kg^{-1}
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0.8087-402238.7
4. Missing value codes	Null (n = 28,590)
5. Precision	0-0.0001
Data format	
1. Fixed, variable length	Variable
Variable identity	SAMP_MG
Variable definition	Magnesium concentration of the sample All SAMP_MG measurements are from either Zoltai *et al.*, 2000 (Zoltai SC, Siltanen RM and Johnson JD. 2000. A wetland data base for the western boreal, subarctic, and arctic regions of Canada. Natural Resources Canada, Canadian Forest Service, Northern Forest Centre, Information Report NOR-X-368. Edmonton, AB.) or the New Brunswick peatland database (Keys and Henderson, 1988. An investigation of the peat resources of New Brunswick. New Brunswick Department of Natural Resources. Open File Report 83-10. [Digital dataset supplied by New Brunswick DNRED, 26 January 2021])
Units of measure	mg kg^{-1}
Data type	
1. Storage type	double
2. Variable codes	Not applicable
3. Range for numeric values	0.1741-115882.5

4. Missing value codes	Null (n = 28,588)
5. Precision	0-0.0001
Data format	
1. Fixed, variable length	Variable
Variable identity	SAMP_NA
Variable definition	Sodium concentration of the sample All SAMP_NA measurements are from either Zoltai *et al.*, 2000 (Zoltai SC, Siltanen RM and Johnson JD. 2000. A wetland data base for the western boreal, subarctic, and arctic regions of Canada. Natural Resources Canada, Canadian Forest Service, Northern Forest Centre, Information Report NOR-X-368. Edmonton, AB.) or the New Brunswick peatland database (Keys and Henderson, 1988. An investigation of the peat resources of New Brunswick. New Brunswick Department of Natural Resources. Open File Report 83-10. [Digital dataset supplied by New Brunswick DNRED, 26 January 2021]) A value of zero indicates that the sample was analyzed and returned a result of zero or trace (below the detection limit of the equipment).
Units of measure	mg kg^{-1}
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0-7241.5
4. Missing value codes	Null (n = 27,796)
5. Precision	0-0.0001
Data format	
1. Fixed, variable length	Variable
Variable identity	SAMP_K
Variable definition	Potassium concentration of the sample All SAMP_K measurements are from either Zoltai *et al.*, 2000 (Zoltai SC, Siltanen RM and Johnson JD. 2000. A wetland data base for the western boreal, subarctic, and arctic regions of Canada. Natural Resources Canada, Canadian Forest Service, Northern Forest Centre, Information Report NOR-X-368. Edmonton, AB.) or the New Brunswick peatland database (Keys and Henderson, 1988. An investigation of the peat resources of New Brunswick. New Brunswick Department of Natural Resources. Open File Report 83-10. [Digital dataset supplied by New Brunswick DNRED, 26 January 2021]) A value of zero indicates that the sample was analyzed and returned a result of zero or trace (below the detection limit of the equipment).
Units of measure	mg kg^{-1}

Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0-20993.5
4. Missing value codes	Null (n = 27,795)
5. Precision	0-0.0001
Data format	
1. Fixed, variable length	Variable
Variable identity	SAMP_TRACE_METHOD
Variable definition	Method used to determine SAMP_CA, SAMP_MG, SAMP_NA and SAMP_K
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Codes defined in LOOKUP Table
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 27,795) where none of SAMP_CA, SAMP_MG, SAMP_NA and SAMP_K was measured
5. Precision	Not applicable
Data format	
1. Fixed, variable length	Variable
QA/QC	Only codes listed in LOOKUP table are permissible in SAMP_TRACE_MEDHOD Ensured all samples with measured SAMP_CA, SAMP_MG, SAMP_NA or SAMP_K have an entry in SAMP_TRACE_METHOD
Variable identity	SAMP_N
Variable definition	Total nitrogen concentration of the sample
Units of measure	Percent
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0-5.13
4. Missing value codes	Null (n = 32,949)
5. Precision	0.1-0.001
Data format	
1. Fixed, variable length	Variable
Variable identity	SAMP_N_METHOD
Variable definition	Method used to determine total N
Units of measure	Not Applicable
Data type	
1. Storage type	String
2. Variable codes	Codes defined in LOOKUP table
3. Range for numeric values	Not Applicable
4. Missing value codes	Null (n = 32,949) where SAMP_N unknown
Data format	
1. Fixed, variable length	Variable
QA/QC	Only codes listed in LOOKUP table are permissible in SAMP_N_METHOD

	Ensured all samples with a measured total N value have an entry in SAMP_N_METHOD
Table WATER [514 Records]	
Variable identity	CORE_ID
Variable definition	Unique identifier for each core in the database. Used in conjunction with WATER_NO to produce unique identifiers for the WATER table
Units of measure	Not applicable
Data type	
1. Storage type	Integer
2. Variable codes	Not applicable
3. Range for numeric values	184-824
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	WATER_NO
Variable definition	A counter (1-n) for the water samples associated with a given CORE_ID Used in conjunction with CORE_ID to produce unique identifiers for the WATER table
Units of measure	Not applicable
Data type	
1. Storage type	Integer
2. Variable codes	Not applicable
3. Range for numeric values	1-3
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	WATER_METHOD
Variable definition	How/where the water sample was taken (surface water, pit, etc.) Data for CORE_ID 184-543 (where available) are from Zoltai *et al.*, 2000 and represent specific samples. For other profiles, methodological descriptions are often vague, and values reported may represent averages from multiple samples.
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	See LOOKUP table
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	WATER_PH
Variable definition	pH of water

Units of measure	pH
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	3.9-7.9
4. Missing value codes	Null (n = 4)
5. Precision	0.1
Data format	
1. Fixed, variable length	Variable
Variable identity	WATER_PH_METHOD
Variable definition	Method used to determine water pH
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Codes defined in LOOKUP table
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 4) where pH not measured
Data format	
1. Fixed, variable length	Variable
Variable identity	WATER_COND
Variable definition	Conductivity of the water sample
Units of measure	millisiemens cm^{-1}
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0.009-0.970
4. Missing value codes	Null (n = 41) where conductivity not measured
5. Precision	0.001
Data format	
1. Fixed, variable length	Variable
Variable identity	COND_CORR_Y_N
Variable definition	Indicates if WATER_COND was corrected for Hydrogen ions (e.g., Riley, 1989. Laboratory Methods for Testing Peat – Ontario Peatland Inventory Project. Ontario Geological Survey Miscellaneous Paper 145, p 10, http://www.geologyontario.mndmf.gov.on.ca/mndmfiles/pub/data/imaging/MP145/MP145.pdf)
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	"Y" If the conductivity reported was corrected for H+ "N" If conductivity was not corrected for H+ "U" If it is unclear if conductivity was corrected
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 41)
Data format	
1. Fixed, variable length	Fixed
Variable identity	WATER_CA
Variable definition	Water total calcium concentration

	All WATER_CA measurements in the database are from Zoltai *et al.* (2000) (Zoltai SC, Siltanen RM and Johnson JD. 2000. A wetland data base for the western boreal, subarctic, and arctic regions of Canada. Natural Resources Canada, Canadian Forest Service, Northern Forest Centre, Information Report NOR-X-368. Edmonton, AB.).
Units of measure	mg kg^{-1}
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0.2309-146.61
4. Missing value codes	Null (n = 36)
5. Precision	0.01-0.0001
Data format	
1. Fixed, variable length	Variable
Variable identity	WATER_MG
Variable definition	Water total magnesium concentration All WATER_MG measurements in the database are from Zoltai *et al.* (2000) (Zoltai SC, Siltanen RM and Johnson JD. 2000. A wetland data base for the western boreal, subarctic, and arctic regions of Canada. Natural Resources Canada, Canadian Forest Service, Northern Forest Centre, Information Report NOR-X-368. Edmonton, AB.). A value of zero indicates that the sample was analyzed and returned a result of zero or trace (below the detection limit of the equipment).
Units of measure	mg kg^{-1}
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0-47.74
4. Missing value codes	Null (n = 36)
5. Precision	0.001-0.0001
Data format	
1. Fixed, variable length	Variable
Variable identity	WATER_NA
Variable definition	Water total sodium concentration All WATER_NA measurements in the database are from Zoltai *et al.* (2000) (Zoltai SC, Siltanen RM and Johnson JD. 2000. A wetland data base for the western boreal, subarctic, and arctic regions of Canada. Natural Resources Canada, Canadian Forest Service, Northern Forest Centre, Information Report NOR-X-368. Edmonton, AB.).
Units of measure	mg kg^{-1}
Data type	
1. Storage type	Double
2. Variable codes	Not applicable

3. Range for numeric values	0.0304-73.159
4. Missing value codes	Null (n = 36)
5. Precision	0.0001-0.001
Data format	
1. Fixed, variable length	Variable
Variable identity	WATER_K
Variable definition	Water total elemental potassium concentration All WATER_K measurements in the database are from Zoltai *et al.* (2000) (Zoltai SC, Siltanen RM and Johnson JD. 2000. A wetland data base for the western boreal, subarctic, and arctic regions of Canada. Natural Resources Canada, Canadian Forest Service, Northern Forest Centre, Information Report NOR-X-368. Edmonton, AB.). A value of zero indicates that the sample was analyzed and returned a result of zero or trace (below the detection limit of the equipment).
Units of measure	mg kg^{-1}
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0-20.643
4. Missing value codes	Null (n = 36)
5. Precision	0.0001-0.001
Data format	
1. Fixed, variable length	Variable
Variable identity	WATER_TRACE_METHOD
Variable definition	Method used to determine WATER_CA, WATER_MG, WATER_NA and WATER_K
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Codes defined in LOOKUP Table
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 36) where trace metals not measured
Data format	
1. Fixed, variable length	Variable
Variable identity	WATER_P
Variable definition	Water total phosphorous concentration All WATER_P measurements in the database are from Zoltai *et al.* (2000) (Zoltai SC, Siltanen RM and Johnson JD. 2000. A wetland data base for the western boreal, subarctic, and arctic regions of Canada. Natural Resources Canada, Canadian Forest Service, Northern Forest Centre, Information Report NOR-X-368. Edmonton, AB.).

	A value of zero indicates that the sample was analyzed and returned a result of zero or trace (below the detection limit of the equipment).
Units of measure	mg kg^{-1}
Data type	
1. Storage type	Double
2. Variable codes	Not applicable
3. Range for numeric values	0-3.1407
4. Missing value codes	Null (n = 36)
5. Precision	0.0001
Data format	
1. Fixed, variable length	Variable
Variable identity	WATER_P_METHOD
Variable definition	Method used to determine total P
Units of measure	Not Applicable
Data type	
1. Storage type	String
2. Variable codes	Codes defined in LOOKUP table
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 36) where WATER_P not measured
Data format	
1. Fixed, variable length	Variable
Table: LOOKUP [175 Records]	
Variable identity	TABLE
Variable definition	Name of table in the database in which field codes are used
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	FIELD
Variable definition	Name of fields in which codes are used
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	CODE
Variable definition	The code used in FIELD
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable

3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	DEFINITION
Variable definition	The definition of CODE
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Table: SOURCEDATA [2,096 Records]	
Variable identity	CORE_ID
Variable definition	Unique identifier for each core location in the database.
Units of measure	Not applicable
Data type	
1. Storage type	Integer
2. Variable codes	Not applicable
3. Range for numeric values	1-1217
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	REF_NO
Variable definition	A counter (1-n) for sources associated with a given CORE_ID Sources listed are the ones that were examined in compiling the database and may represent only a subset of data (published or unpublished) available for the profile. A source that is listed may or may not have been used directly in data entry (see REF_DATA).
Units of measure	Not applicable
Data type	
1. Storage type	Integer
2. Variable codes	not applicable
3. Range for numeric values	1-5
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	REF_ID
Variable definition	Unique identifier for each reference
Units of measure	Not applicable
Data type	
1. Storage type	Integer
2. Variable codes	Not applicable
3. Range for numeric values	1-132

4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	CITE_SHORT
Variable definition	Short citation as used when citing the reference in a publication
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	REF_DATA
Variable definition	The data taken from REF_ID in compiling data for CORE_ID
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable length
Variable identity	REF_NOTES
Variable definition	Additional notes regarding use of REF_ID for CORE_ID
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	null (n = 1589)
Data format	
1. Fixed, variable length	Variable
Table: REFERENCES [132 Records]	
Variable identity	REF_ID
Variable definition	Unique identifier for each reference
Units of measure	Not applicable
Data type	
1. Storage type	Integer
2. Variable codes	Not applicable
3. Range for numeric values	1-132
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	CITE_SHORT
Variable definition	Short citation as used when citing the reference in a publication

Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	CITE_LONG
Variable definition	Long citation as used in a reference list
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	AUTHORS
Variable definition	List of authors of the reference
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	YEAR
Variable definition	Year the reference was published
Units of measure	Not applicable
Data type	
1. Storage type	Integer
2. Variable codes	Not applicable
3. Range for numeric values	1964-2021
4. Missing value codes	Null (n = 2) where source not published
Data format	
1. Fixed, variable length	Fixed
Variable identity	TITLE
Variable definition	Title of the reference
Units of measure	Not applicable
Data type	
1. Storage type	String
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	No missing values
Data format	
1. Fixed, variable length	Variable
Variable identity	LINK_TO_PDF

Variable definition	Active link to the reference online Links are only provided for sources that are not behind a paywall
Units of measure	Not applicable
Data type	
1. Storage type	Hyperlink
2. Variable codes	Not applicable
3. Range for numeric values	Not applicable
4. Missing value codes	Null (n = 20) where no free pdf could be located
Data format	
1. Fixed, variable length	Variable
QAQC	Ensured each link was active and used digital object identifier (DOI) links when possible.

C. Data anomalies:

1. High carbon values

Some C_TOT_PCT (n=39) and C_ORG_PCT (n=55) values exceed 60% in the PROFILES table, which is higher than the range reported in a published synthesis of peat properties for northern peatlands (i.e., 30-60%; Loisel *et al.*, 2014). Since these values represent < 0.2% of the total values and they come from several data sources, they were retained and used to calculate SAMP_CARB_MGHA in the PROFILES table. However, since these values represent the extreme end of the distribution of carbon content in peat samples, they should be used with caution.

2. Precision

Values in the database for fields with double type storage are generally formatted to display the nominal precision listed under Class IV Section B Variable Information. In the instances where precision from a data source was less than the nominal precision, the number of decimal points shown was reduced for that CORE_ID. The range of precision is reported in brackets under the relevant field information.

For the SAMP_CA, SAMP_MG, SAMP_NA, SAMP_K, and SAMP_N fields in the PROFILES table and WATER_CA, WATER_MG, WATER_NA, WATER_K, and WATER_P in the WATER table, the precision of the value depended on the analysis method and the order of magnitude of the value measured. Therefore, the precision reported in the original source was retained in these fields. The range of precision is reported in brackets under the relevant field information.

Class V. Supplemental descriptors

A. Data acquisition

1. Data entry verification procedures:

   All data were entered manually and double checked against the original source and were also checked by several researchers in the QA/QC procedures listed with each relevant field.

B. Quality assurance/quality control procedures:

Specific QA/QC procedures are outlined in the Class IV Section B Variable Information for each of the relevant fields. Random rows in each table were also double checked against their source to ensure that the data values matched.

C. Publications:

Agriculture and Agri-food Canada, 2013, Soil Landscapes of Canada Version 3.2, https://open.canada.ca/data/en/dataset/5ad5e20c-f2bb-497d-a2a2-440eec6e10cd.

Bauer, I. E., Bhatti, J. S., Cash, K. J., Tarnocai, C., and Robinson, S. D., 2006, Developing statistical models to estimate the carbon density of organic soils: Canadian Journal of Soil Science, v. 86, no. Special Issue, p. 295-304, doi: 10.4141/S05-087.

Bona, K. A., Shaw, C., Thompson, D. K., Hararuk, O., Webster, K., Zhang, G., *et al.*, 2020, The Canadian model for peatlands (CaMP): A peatland carbon model for national greenhouse gas reporting: Ecological Modelling, v. 431, p. 109164, doi: 10.1016/j.ecolmodel.2020.109164.

Tarnocai, C., Kettles, I. M., and Lacelle, B., 2011, Peatlands of Canada, Open File 6561: Natural Resources Canada, doi: 10.10.4095/288786.

Literature Citations

Agriculture and Agri-food Canada, 2013, Soil Landscapes of Canada Version 3.2, https://open.canada.ca/data/en/dataset/5ad5e20c-f2bb-497d-a2a2-440eec6e10cd.

Alberta Environment and Sustainable Resource Development, 2015, Alberta Wetland Classification System, Water Policy Branch, Policy and Planning Division: Edmonton, AB, 54 pages, https://open.alberta.ca/dataset/92fbfbf5-62e1-49c7-aa13-8970a099f97d/resource/1e4372ca-b99c-4990-b4f5-dbac23424e3a/download/2015-alberta-wetland-classification-system-june-01-2015.pdf.

Bauer, I. E., Bhatti, J. S., Cash, K. J., Tarnocai, C., and Robinson, S. D., 2006, Developing statistical models to estimate the carbon density of organic soils: Canadian Journal of Soil Science, v. 86, no. Special Issue, p. 295-304, doi: 10.4141/S05-087.

Chambers, F. M., Beilman, D. W., and Yu, Z., 2010, Methods for determining peat humification and for quantifying peat bulk density, organic matter and carbon content for palaeostudies of climate and peatland carbon dynamics: Mires and Peat, v. 7, p. Article 7, 1-10.

Kettles, I. M., Garneau, M., and Jetté, H., 2000, Macrofossil, Pollen, and Geochemical Records of Peatlands in the Kinosheo Lake and Detour Lake Areas, Northern Ontario, Ottawa, ON, Canada, Natural Resources Canada, Geological Survey of Canada Bulletin 545, 16 pages.

Keys, D., and Henderson, R. E., 1988, An investigation of peat resources of New Brunswick. Open File Report 83-10, Department of Natural Resources.

Loisel, J., Yu, Z. C., Beilman, D. W., Camill, P., Alm, J., Amesbury, M. J., *et al.*, 2014, A database and synthesis of northern peatland soil properties and Holocene carbon and nitrogen accumulation: Holocene, v. 24, no. 9, p. 1028-1042, doi: 10.1177/0959683614538073.

National Wetlands Working Group, 1997, The Canadian Wetland Classification System (2nd Ed.), Waterloo, ON, Canada, Wetlands Research Centre, University of Waterloo, 68 pages.

QGIS Development Team, 2020, QGIS Geographic Information System, Open Source Geospatial Foundation Project.

Riley, J. L., 1989, Laboratory Methods for Testing Peat - Ontario Peatland Inventory Project. Ontario Geological Survey Miscellaneous Paper 145, Ontario Geological Survey: Toronto, ON, Canada, 52 pages.

Riley, J. L., and Michaud, L., 1994, Ontario Peatland Inventory: Field-Work Methods. Miscellaneous Paper 55, Ontario Geological Survey: Toronto, ON, Canada, 62 pages.

Soil Classification Working Group, 1998, The Canadian System of Soil Classification (3rd Ed.) Publication 1646: Ottawa, ON, Canada, Agriculture and Agri-Food Canada, 187 pages.

Tarnocai, C., Kettles, I. M., and Lacelle, B., 2011, Peatlands of Canada, Open File 6561: Natural Resources Canada.

Therneau, T., Atkinson, B., and Ripley, B., 2022, Recursive partitioning and regression trees Version 4.1.19, https://github.com/bethatkinson/rpart, https://cran.r-project.org/package=rpart.

Von Post, L., 1922, Sveriges geologiska undersökings torvinventering och några a dess hittills vunna resultat: Svenska mosskulturföreningens tidskrift, v. 37, p. 1-27.

Zoltai, S. C., Siltanen, R. M., and Johnson, J. D., 2000, A wetland data base for the western boreal, subarctic, and arctic regions of Canada, Information Report NOR-X-368, Canadian Forest Service: Edmonton, AB, Canada, Natural Resources Canada, 30 pages.