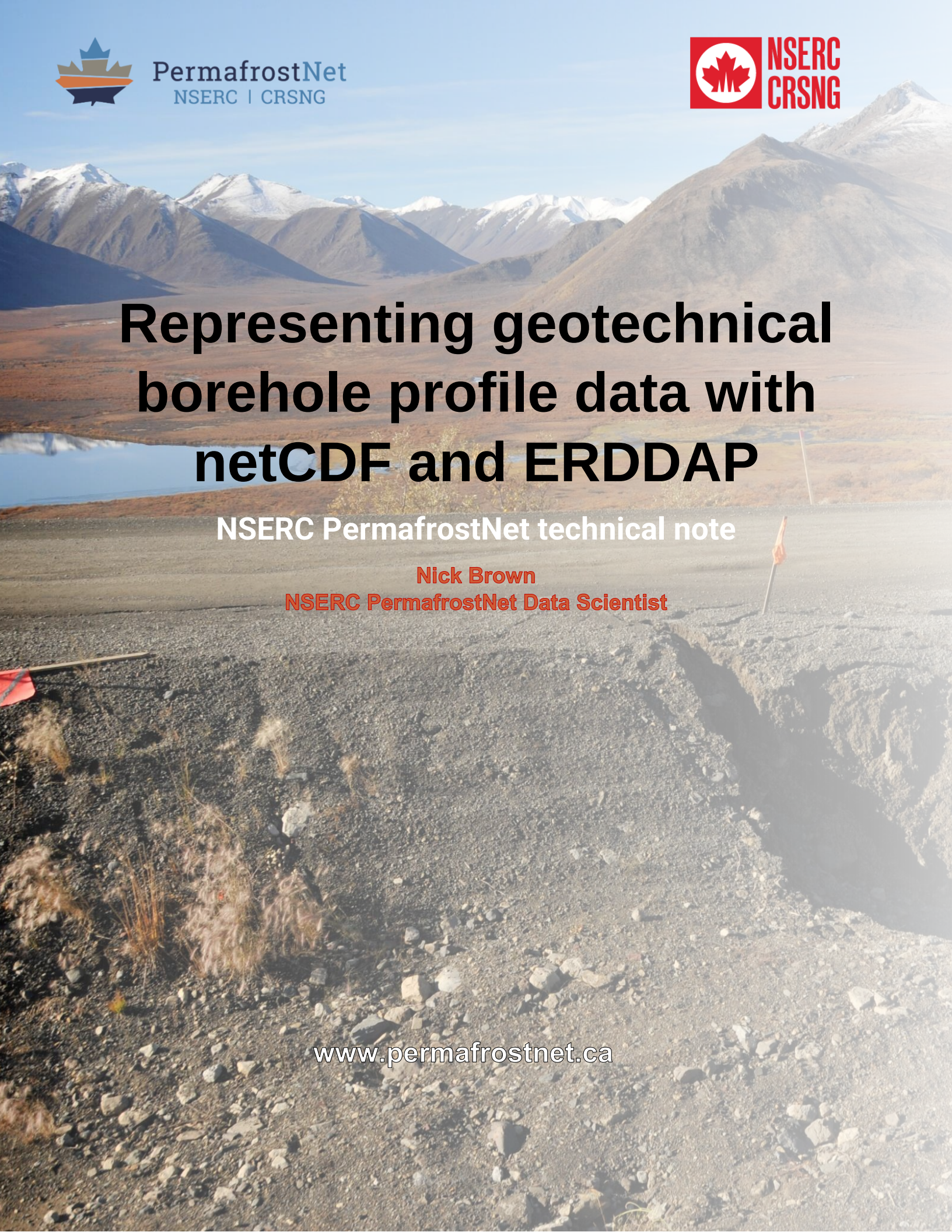

PermafrostNet
NSERC | CRSNG

NSERC CRSNG

Representing geotechnical borehole profile data with netCDF and ERDDAP

NSERC PermafrostNet technical note

Nick Brown
NSERC PermafrostNet Data Scientist

www.permafrostnet.ca

Context

NetCDF is a format for scientific data that is gaining popularity because it is self-documenting and uses a several well-established community-built standards to increase data interoperability. These include the CF conventions and the attribute convention for data discovery (ACDD). There are a number of data distribution platforms (ERDDAP, THREDDS, HYRAX) that can be used to easily distribute NetCDF files using a standardized API and web interface. NSERC PermafrostNet uses ERDDAP as a data distribution platform for these reasons and because it offers a graphical web interface for those who want to access data interactively. It also makes it possible to adopt the associated data standards as one way to increase the interoperability of permafrost data generated by and used within the network.

The NetCDF format and associated standards were originally developed for atmospheric, oceanographic and climatic data, so there are few examples of these files being used to represent observed soil profiles or geotechnical borehole data, and the CF standard_name table lacks many variables that are relevant to permafrost science. This document suggests a structure for observed borehole profile data.

Considerations

Geotechnical or borehole data are typically discretized into intervals with ‘from’ and ‘to’ coordinates that specify an interval along a vertical axis. This differs from ground temperature time series which are point measurements taken at a specific time, depth coordinate. Occasionally, only a depth measurement will be provided such as when small samples are taken (doi: 10.5281/zenodo.5873789)

The CF conventions define the *bounds* attribute to specify the extent of intervals or cells when data values are not representative of single points in time and/or space, but rather of intervals or multidimensional cells. This is an obvious solution, but unfortunately ERDDAP does not support this because it is not possible to include the bounds variables which have different dimensions as the rest of the data.

The chapter in the CF conventions on discrete sampling geometries is designed to solve the problem of representing features when available data does not fully cover the dimensions in a dataset, i.e. when data represent point observations rather than gridded data. In the case of borehole time series, such as ground temperature measurements, a “Time Series Profile” feature type is used. Other types include “Profile”, “Time Series” and “Point”.

Recommendations

Representing borehole data

The following CDL (Common Data Language: human-readable notation for netCDF objects and data) representation describes a recommended structure for observed borehole profile data. It uses the “Profile” feature type and makes use of the bounds attribute to indicate the vertical interval of the measurement. The depth_below_ground_surface dimension is given at the midpoint of each interval rather than at the bottom or the top of the interval. This follows the convention of most model outputs which use the midpoint of a discretized model cell to represent the constant value over that cell. It also produces reasonable data visualizations.

However, it should be noted that any of the depth values within a cell are acceptable choices in the CF conventions.

The ERDDAP data server uses an internal data representation that is not compatible with coordinate information being stored in an extra dimension. In order to enable distribution of data on ERDDAP servers, and to increase useability, two redundant variables (*top_of_interval* and *bottom_of_interval*) are included to describe the depth at the top and bottom of the described intervals. The EDDTableFromMultidimNcFiles dataset type can then be used by omitting the depth bounds variable.

In the simplest case, there is one borehole per file, and the netCDF file structure does not have a profile dimension, as below.

```
netcdf borehole_demo {
dimensions:
      depth_below_ground_surface = 8 ;
      nv = 2 ;
      name_strlen = 32 ;

variables:
    char borehole_name(name_strlen) ;
          borehole_name:cf_role = "profile_id";
          borehole_name:standard_name = "platform_id";
    float lon;
        lon:standard_name = "longitude";
        lon:long_name = "longitude" ;
        lon:units = "degrees_east" ;
        lon:axis = "X" ;
    float lat;
        lat:standard_name = "latitude";
        lat:long_name = "latitude" ;
        lat:units = "degrees_north" ;
        lat:axis = "Y" ;
    float depth_below_ground_surface(depth_below_ground_surface) ;
        z:standard_name = "depth";
        z:long_name = "Midpoint of described soil layer" ;
        z:units = "m" ;
        z:positive = "down" ;
        z:axis = "Z" ;
        z:bounds = "z_bnds" ;
    top_of_interval(depth_below_ground_surface) ;
        z:standard_name = "depth";
        z:long_name = "Top of described soil layer" ;
        z:units = "m" ;
        z:positive = "down" ;
    bottom_of_interval(depth_below_ground_surface) ;
        z:standard_name = "depth";
        z:long_name = "Bottom of described soil layer" ;
        z:units = "m" ;
        z:positive = "down" ;
    float z_bnds(depth_below_ground_surface, nv) ;
    float clay(depth_below_ground_surface) ;
        clay:standard_name = "volume_fraction_of_clay_in_soil" ;
        clay:long_name = "Volumetric fraction of clay in soil" ;
        clay:units = "1" ;
        clay:coordinates = "lon lat depth_below_ground_surface" ;

// global attributes:
        :featureType = "profile";
        :cdm_data_type = "Profile";
```

```
        :Conventions = "CF-1.7" ;

data:

 depth_to_bottom_of_interval = 10, 20, 30, 50, 60, 80, 100, 200 ;
 depth_to_top_of_interval = 0, 10, 20, 30, 50, 60, 80, 100 ;
 depth_below_ground_surface = 5, 15, 25, 45, 55, 70, 90, 150 ;

 z_bnds =
  0, 10,
  10, 20,
  20, 30,
  40, 50,
  50, 60,
  60, 80,
  80, 100,
  100, 200 ;

  lat = 60 ;

  lon = -112.14 ;

  clay = 0.21, 0.23, 0.15, 0.10, 0.20, 0.12, 0.05, 0.05 ;

  borehole_name = "borehole_01";
}
```

This example uses the percentage of clay in the soil as the single variable described. Additional variables can be added so long as the interval bounds remain the same between intervals. (For instance, if visible ice were measured at 10 cm intervals and soil bulk density was measured at 15 cm intervals, both sets of data could not fit in this simple structure without at least some minor changes). To accommodate variables with non-overlapping intervals this, a more complex structure could be defied, but the more convenient and recommended approach is either to split the data into multiple files or to redefine the sampling intervals to include interval breaks from all variables – analogous to a *Union* operation. In the example above with 10 and 15 cm sample spacing, the combined sampling grid would be as follows:

Visible ice	0, 10, 20, 30, 40, 50, 60,…
Bulk density	0, 15, 30, 45, 60, …
Combined	0, 10, 15, 20, 30, 40, 45, 50, 60,…

The simple file structure presented here assumes that the length of the profile dimension (“profile”) is equal to one, corresponding to a single borehole location. As before, a more complicated file structure would permit the aggregation of data from multiple locations with uneven sampling intervals into one file. More specifically, an additional profile dimension must be added. Data variables then use both the profile and z dimensions. If measurements are not taken at the same depth for all profiles, the ‘depth_below_ground_surface’ variable must also depend on the profile dimension in addition to the z dimension.

```
netcdf multi_borehole_demo {
dimensions:
```

Last revision: 2022-06-14

```
      z = 8 ;
      profile = 2 ;
      nv = 2 ;
      name_strlen = 32 ;

variables:
    char borehole_name(profile, name_strlen) ;
          borehole_name:cf_role = "profile_id";
          borehole_name:standard_name = "platform_id";
    float lon(profile);
        lon:standard_name = "longitude";
        lon:long_name = "longitude" ;
        lon:units = "degrees_east" ;
        lon:axis = "X" ;
    float lat(profile);
        lat:standard_name = "latitude";
        lat:long_name = "latitude" ;
        lat:units = "degrees_north" ;
        lat:axis = "Y" ;
    float depth_below_ground_surface(profile, z) ;
        z:standard_name = "depth";
        z:long_name = "Midpoint of described soil layer" ;
        z:units = "m" ;
        z:positive = "down" ;
        z:axis = "Z" ;
        z:bounds = "z_bnds" ;
    depth_to_top_of_interval(profile, z) ;
        z:standard_name = "depth";
        z:long_name = "Top of described soil layer" ;
        z:units = "m" ;
        z:positive = "down" ;
    depth_to_bottom_of_interval(profile, z) ;
        z:standard_name = "depth";
        z:long_name = "Bottom of described soil layer" ;
        z:units = "m" ;
        z:positive = "down" ;
    float z_bnds(profile, z, nv) ;
    float clay(profile, z) ;
        clay:standard_name = "volume_fraction_of_clay_in_soil" ;
        clay:long_name = "Volumetric fraction of clay in soil" ;
        clay:units = "1" ;
        clay:coordinates = "lon lat depth_below_ground_surface " ;

// global attributes:
        :featureType = "profile";
        :cdm_data_type = "Profile";
        :Conventions = "CF-1.7" ;

depth_to_bottom_of_interval = 10, 20, 30, 50, 60, 80, 100, 200 ;
                              10, 20, 30, 50, 60,;

 depth_to_top_of_interval = 0, 10, 20, 30, 50, 60, 80, 100 ;
                            0, 10, 20, 30, 50;

 depth_below_ground_surface = 5, 15, 25, 45, 55, 70, 90, 150 ;
                              5, 15, 25, 45, 55;


 z_bnds =
  0, 10,
  10, 20,
  20, 30,
  40, 50,
```

```
  50, 60,
  60, 80,
  80, 100,
  100, 200 ;

  0, 10,
  10, 20,
  20, 30,
  40, 50,
  50, 60,

  lat = 60, 61;

  lon = -112.14, -113.1  ;

  clay = 0.21, 0.23, 0.15, 0.10, 0.20, 0.12, 0.05, 0.05 ;
         0.11, 0.43, 0.11, 0.21, 0.10;

  borehole_name = "borehole_01", "borehole_02”;


}
```