README

Prepared by Lamisa Malik, Juliane Mai, and Nandita B. Basu

March 3, 2025

This document is the README for the data used and produced in a publication currently under review with Nature Scientific Data. Proper citation and DOI forthcoming.

gTREND-P-Canada - A 250-m Gridded and Unified Phosphorus Mass Budget Dataset Across Canada (1961–2021)

Lamisa Malik [1], Juliane Mai [2], Danyka K. Byrnes [1], Kimberly J. Van Meter [3,4], Shuyu Chang [3], Meghan McLeod [5], Nandita B. Basu [1,2]

[1] Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada

[2] Department of Earth and Environmental Science, University of Waterloo, Waterloo, ON, Canada

[3] Department of Geography, The Pennsylvania State University, University Park, PA 16802, USA

[4] Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, PA 16802, USA

[5] Datastream, The Gordon Foundation. Toronto, ON, Canada.

The data is shared on Figshare. The proper citation and DOI for the dataset are:

Lamisa Malik, Juliane Mai, Danyka K. Byrnes, Kimberly J. Van Meter, Shuyu Chang, Meghan McLeod, Nandita B. Basu: A 250-m Unified Gridded Phosphorus Mass Budget Dataset Across Canada (1961–2021) - gTREND-P-Canada. DOI: https://doi.org/10.6084/m9.figshare.c.7701308

The dataset comprises of a county-scale data product (TREND-P-Canada) and a 250-m gridded data product (gTREND-P-Canada) of phosphorus inputs (fertilizer P, livestock manure P, domestic waste P), outputs (crop and pasture P removal), and surplus across ten provinces and territories of Canada, excluding the Northwest Territories, Nunavut, and Yukon.

This data product contains six zipped dataset folders:

A. County boundary shapefiles in ESRI shapefile format
   in folder "**County_shapefiles**"
B. County-scale data product (TREND-P-Canada) in CSV format
   in folder "**County_TREND-P-Canada**"
C. Gridded data product (gTREND-P-Canada) in georeferenced TIFF format
   in folder "**Gridded_gTREND-P-Canada**"
D. Data required for Figure 3:
   Reaggregated province-scale data product (pTREND-P-Canada) in CSV format
   in folder "**Figure_3**"
E. Data required for Figure 4:
   Independent other county-scale datasets in EXCEL format
   in folder "**Figure_4**"
F. Data required for Figure 5:
   Reaggregated gridded data product to county-scale in CSV format used for validation purposes only
   in folder "**Figure_5**"

The codes to create the figures are available on GitHub under https://github.com/julemai/gTREND-P-Canada_v1. Other codes are available upon request.

Correspondence should be directed to nandita.basu@uwaterloo.ca and/or juliane.mai@uwaterloo.ca.

County boundary shapefiles

The county boundaries may change between census years in Canada. The shapefiles are provided in ESRI shapefile format following the file naming pattern YYYY.shp/cpg/dbf/prj/shx where YYYY is the according census year the data are valid for. The county boundary shapefiles for the years 1961, 1966, 1971, 1976, 1981, 1986, 1991, 1996, 2001, 2006, 2011, 2016, and 2021 are provided. Please note that for the years 1966, 1971, and 1976 the boundary shapefiles were not available for and required manual processing which might lead to mismatches between the used and actual county boundaries during those years. Each shapefile attribute includes a column

labeled "ID," which represents the county ID. This ID is used to link the shapefile with the CSV files located in the folder named "**County_shapefiles**" after unzipping.

County-Scale TREND-P-Canada Dataset

The county-scale P dataset is provided in a folder called "**County_TREND-P-Canada**" after unzipping. It contains the county-scale P value of four phosphorus mass balance components and P surplus in kg-P ha^{-1} yr^{-1} (where ha is ha of land area). In that dataset the counties are referred to by their IDs. The IDs coincide with the counties and their associated IDs defined in the shapefile. Please be aware that some counties only exist for certain years and even if an ID exists in multiple years the spatial extent of that county might have changed. Details on the organization of the files provided in folder "County_TREND-P-Canada" are presented in **Table A**.

Table A: Structure, file names and description of the county-scale TREND-P-Canada Dataset in the data repository.

Folder Name after unzipping	Files in Main Folder	Description
County_TREND-P-Canada	P_FERT.csv	County scale P from fertilizer use in kg-P ha-county-area^{-1} yr $^{-1}$
	P_LVSK.csv	County scale P generated form livestock manure in kg-P ha-county-area^{-1} yr $^{-1}$
	P_WASTE.csv	County scale P generated from domestic waste use in kg-P ha-county-area^{-1} yr $^{-1}$
	P_CROP.csv	County scale P removal by crop and pastureland in kg-P ha-county-area^{-1} yr $^{-1}$
	P_Surplus.csv	County scale P surplus in kg-P ha-county-area^{-1} yr $^{-1}$

Guide to Interpreting the Data

In each Comma-Separated Values (CSV) file within the "County_TREND-P-Canada" folder, the first column indicates the available years, and the first row lists county ID numbers. In the

county shapefile, the "ID" column in the attribute table corresponds to the county ID for that particular year. The county ID numbers in the CSV files represent the county boundaries as shown in the corresponding year's shapefile, matching the "ID" column in the attribute table. For example, to determine the P fertilizer contribution for the year 1996 in county ID 3559, the magnitude of P in kg $ha^{-1}yr^{-1}$ can be located in the "P_FERT.csv" file, in the cell corresponding to county ID 3559 and the year 1996. To locate this county spatially, refer to the census boundary map and find ID 3559 in the "ID" column of the shapefile attribute table. Note that "NAN" values in the CSV files indicate the absence of a corresponding county ID for that year. For example, the cell value for county ID 3559 in the year 1961 is "NAN," indicating that this county ID is absent in the county boundary shapefile for that year due to changes in county boundaries and/or county ID assignments.

Gridded (250-meter) Dataset gTREND-P-Canada

Gridded data is provided in the folder titled "**Gridded_gTREND-P-Canada**". The folder contains TIF files with rasters of 250 meter resolution for phosphorus estimates of four P mass balance components and P surplus for the year 1961, 1966, 1971, 1976, 1981, 1986, 1991, 1996, 2001, 2006, 2011, 2016, and 2021 (see **Table B**). The P mass balance component raster is provided for each province individually. The unit of each variable in TIF files is kg-P ha-of-grid-$area^{-1}$ yr^{-1}. Each province is encoded using the 2-digit standard geographical classification (SGC) code form Statistics Canada [1] .

Table B: ***Structure, file names and description of the gridded gTREND-P-Canada Dataset in the data repository.*** The data is housed within the "Gridded_gTREND-P-Canada" zipped folder that has five subfolders for the five components. Each sub-folder contains 130 TIFF files corresponding to the 13 census years designated as YYYY in the file name (1961, 1966, 1971, 1976, 1981, 1986, 1991, 1996, 2001, 2006, 2011, 2016, and 2021) and ten provinces (designated using the province-specific geographical code (SGC)). The SGC for the provinces are ten for Newfoundland and Labrador, 11 for Prince Edward Island, 12 for Nova Scotia, 13 for New

Brunswick, 24 for Quebec, 35 for Ontario, 46 forManitoba, 47 for Saskatchewan, 48 for Alberta, and 59 for British Columbia.

Sub-Folder Name	File Naming Convention within the subfolder	Description	Number of Files
Fertilizer_P	FERT_SGC_YYYY.tif	P from fertilizer use in kg ha^{-1} yr^{-1} at 250 m grid resolution.	130
Livestock_Manure_P	LVSK_SGC_YYYY.tif	P generated from livestock manure in kg ha^{-1} yr^{-1} at 250 m grid resolution.	130
Domestic_Waste_P	WSTE_SGC_YYYY.tif	P generated from domestic waste in kg ha^{-1} yr^{-1} at 250 m grid resolution.	130
Crop_and_Pasture_P_ Removal	CROP_SGC_YYYY.tif	P removal by crop and pastureland in kg ha^{-1} yr^{-1} at 250 m grid resolution.	130
P_Surplus	TSur_SGC_YYYY.tif	P surplus in kg ha^{-1} yr^{-1} at 250 m grid resolution.	130

Guide to Interpreting the Data

Each grid value in the Tagged Image File (TIF) represents the amount of phosphorus in kg per hectare of land-area within that county per year. To estimate the total phosphorus within a specific county or watershed boundary, calculate the mean of all grid cells within the desired boundary. This mean value will provide the phosphorus magnitude in kg ha of county $area^{-1}$ yr^{-1} for that particular county or watershed.

Data for Figure 3: Reaggregated Provincial Data Aggregated from County-Scale Dataset

Provincial data is provided in the sub-folder titled "**Figure_3**" after unzipping. The folder contains ten Comma-Separated Values (CSV) files; one for each province.

Guide to Interpreting the Data

Each CSV file is named after the abbreviation of a province following the pattern "<province-abbreviation>_Kton.csv". The abbreviations used for the file titles are described as follows: AB for Alberta, BC for British Columbia, MB for Manitoba, NB for New Brunswick, NL for Newfoundland and Labrador, NS for Nova Scotia, ON for Ontario, PE for Prince Edward Island, QC forQuebec, and SK for Saskatchewan.

Each CSV file contains six columns:

1. **years**: The years for which we have data (census years starting 1961 ending 2021).
2. **fert**: Fertilizer P magnitude.
3. **man**: P generated from livestock manure.
4. **waste**: P generated from domestic waste.
5. **crop**: P removal by crops and pastureland.
6. **surplus**: Total surplus is the difference between the sum of "fert", "man", and "waste," and the "crop" values.

All values in the table are reported in metric kilotons per year.

Data for Figure 4: Various Provincial-scale Datasets

The folder named "**Figure_4**" after unzipping contains two validation datasets that were not produced in this publication but serve as independent validation datasets. They are shared here in case the data format or availability is changing in future. The two files are in Excel format, i.e., "IPNI_data.xlsx" and "Wang_Paper_data.xlsx".

The first file contains data of "Estimates of Nutrient Uptake and Removal" published by IPNI. The data were previously accessible under http://www.ipni.net/article/IPNI-3296 but at the time of publication the entire IPNI website was down and the authors were not able to verify this link[2].

The second file is the data published by Wang et al. (2022) [3]. The data are publicly accessible under https://doi.org/10.5281/zenodo.6877750.

Guide to Interpreting the Data

The reader may refer to Wang et al. (2022) [3] and IPNI [2] for details on their data structure and units.

Data for Figure 5: Reaggregated Gridded Data Product to County-Scale for validation

For validation purposes of the gridded data gTREND-P-Canada, they were reaggregated to county-scale and compared with the originally derived county-scale dataset TREND-P-Canada. We provide these data for the convenience of the data users in a folder named "**Figure_5**" indicating that it is the county-scale aggregates of the gridded dataset. The data are organized exactly like the county-scale dataset TREND-P-Canada (see section "County-Scale TREND-P-Canada Dataset" above).

Guide to Interpreting the Data

The reaggregated gridded data to county-scale are provided only for validation purposes. If a county-scale dataset is needed, the original county-scale dataset in folder "County_TREND-P-Canada" is recommended to be used.

Find more information on how to interpret the files found in folder "Figure_5" by referring to the section on "County-Scale TREND-P-Canada Dataset" above. The same filenames and units are used (see **Table A**).

Please refer to the journal publication for details on the methodology used to prepare these data (DOI and proper citation forthcoming).

References

1. Statistics Canada. Dictionary, Census of Population, 2021 - Table 1.8 Abbreviations and codes for provinces and territories, 2021 Census [Accessed- Nov, 2024]. https://www12.statcan.gc.ca/census-recensement/2021/ref/dict/tab/index-eng.cfm?ID=t1_8 (2021).
2. Crop Nutrient Balance - Canada - Phosphorus [Accessed- Feb, 2023]. http://phosphorus.ipni.net/article/NANE-3104.
3. Wang, J., Qi, Z. & Bennett, E. M. Changes in Canada's phosphorus cycle 1961–2018: Surpluses and deficits. *Global Biogeochem. Cycles* **36**, (2022).