Natural Resources Canada | Ressources naturelles Canada

GEOLOGICAL SURVEY OF CANADA
OPEN FILE 9309e

Quality assurance and quality control of till compositional data using Python

E. Brouard and P.M. Godbout

2025

Canada

GEOLOGICAL SURVEY OF CANADA
OPEN FILE 9309e

Quality Assurance and Quality Control of Till Compositional Data using Python

E. Brouard and P.M. Godbout

2025

This publication is available for free download through the NRCan Open Science and Technology Repository (https://ostrnrcan-dostrncan.canada.ca/).

Recommended citation
Brouard, E. and Godbout, P.M. 2026. Quality Assurance and Quality Control of Till Compositional Data using Python; Geological Survey of Canada, Open File 9309, 1 zip file. https://doi.org/10.4095/ps1ht4x4vg

Publications in this series have not been edited; they are released as submitted by the author.

ISSN 2816-7155
ISBN 978-0-660-78789-3
Catalogue No. M183-2/9309E-PDF
https://doi.org/10.4095/ps1ht4x4vg

Quality Assurance and Quality Control of Till Compositional Data using Python

Etienne Brouard[1] and Pierre-Marc Godbout[2]

[1] Geological Survey of Canada – Quebec Division, Natural Resources Canada, 490 rue de la Couronne, Québec, QC G1K 9A9, Canada
[2] Geological Survey of Canada – Central Division, Natural Resources Canada, 601 Booth Street, Ottawa, ON K1A 0E8, Canada

Compositional data derived from tills play a pivotal role in geoscientific investigations, offering insights into geological processes, environmental conditions, and mineral exploration. However, ensuring the accuracy and reliability of these datasets is paramount for meaningful analyses and informed decision-making. To address this need, this report introduces a comprehensive suite of streamlined Quality Assurance and Quality Control (QA/QC) protocols tailored to till data. Designed to simplify QA/QC workflows—particularly for users with limited experience in coding or statistical analysis—these protocols combine simple procedures with user-friendly Python scripts. The protocols cover essential aspects of data quality assessment, including basic statistics, accuracy, precision, and variability, providing users with plots for visual interpretation and a systematic approach to evaluating the integrity and reliability of their datasets. The manuscript outlines key considerations for meticulous planning, proper data formatting, and the effective use of tools such as Microsoft Excel and Python. By following the outlined procedures and using the provided scripts, users can generate summary statistics, assess measurement accuracy, evaluate differences between routine samples and their analytical duplicates, and determine whether data variability is acceptable for mapping and statistical purposes.

Introduction

Compositional data from tills, or any other type of geological sample, obtained from various laboratories require meticulous evaluation for accuracy and precision before dissemination. A thoroughly planned Quality Assurance and Quality Control (QA/QC) program for sample collection, preparation, analysis, data management, and archiving ensures that no systematic bias is introduced to the data. At the Geological Survey of Canada (GSC), internal protocols for QA/QC have been developed and documented in various publications using software platforms, such as Microsoft Excel and the open-source R environment (Garrett and Chang, 2007; Garrett, 2013a; McCurdy & Garrett, 2016). Although robust QA/QC procedures exist for field sample collection (e.g., McClenaghan et al., 2020, 2023), a unified framework for evaluating the quality of resulting compositional data is still lacking. As a result, QA/QC practices vary across research groups, reflecting differences in methodology, software, and interpretation. This lack of standardization may stem from challenges related to complexity, usability, and software compatibility (McCurdy and Garrett, 2016). Overall, there is a clear need for simplified QA/QC procedures that are easy to understand, accessible to users with varying technical backgrounds, and designed to ensure reproducibility across studies. Here, we present a suite of five Python-based QA/QC protocols tailored to address specific objectives in data quality assessment, aiming to streamline the process of evaluating accuracy, precision, and variability in compositional datasets. By leveraging the capabilities of the Python programming language, the methods provide users, regardless of coding experience, with clear, efficient tools to assess the integrity and reliability of their data. To ensure ease of use, the protocols have been compiled into a user-friendly package (**Suppl. material 1**), containing input templates, executable scripts, and batch files that generate CSV outputs with summary statistics and interpretation guides. This report first discusses key planning considerations and software requirements, followed by a walkthrough of each protocol's purpose, implementation, and outputs.

Planning considerations for QA/QC

Proper planning for QA/QC procedures is essential throughout fieldwork and laboratory analyses to minimize errors, ensure data integrity, and enable meaningful interpretation. This planning includes the systematic collection of routine samples together with field duplicates, as well as the preparation of analytical duplicates:

- **Routine samples** are collected at specific locations to support scientific investigation.
- **Field duplicates** are additional samples collected at the same site as the routine sample to evaluate sampling variability and potential heterogeneity at the site level.
- **Analytical duplicates** are derived from routine samples or field duplicates to evaluate variance and ensure analytical consistency. Analytical duplicates should be separated from field duplicates to allow independent estimation of sampling and analytical variability (Garrett, 2013b).

Effective QA/QC protocol planning should also include **certified reference materials** (CRMs) within each sample batch submitted for analysis. CRMs—such as Till-1, Till-4, OREAS-46, and OREAS-47—should closely resemble the mineralogical matrix of routine samples to ensure compatibility (Horowitz, 1991). These materials are analyzed alongside routine samples to monitor analytical accuracy. The results are compared to expected or provisional values, either published online by laboratories (e.g., oreas.com/crm/oreas-46, oreas.com/crm/oreas-47) or reported in scientific publications (e.g., Lynch, 1996; Burnham and Schweyer, 2004), to detect potential issues like drift or systematic bias over time.

The sample analysis sequence should be carefully planned before submitting a batch to an external laboratory. The GSC recommends the "20-sample block system, " in which each block of 20 consecutive sample numbers includes 17 routine samples, one CRM, one analytical duplicate, and one field duplicate (**Table 1**). In addition, blanks should be inserted near the beginning, middle, and end of each batch to monitor potential contamination from previously processed samples. The CRM sample can be positioned within the block at any location except the first, which is often reserved for the blank sample. Randomizing the placement of routine, duplicate, and CRM samples within each block is advised to avoid positional bias. For comprehensive guidelines concerning QA/QC procedures, researchers should consult previous GSC publications (Garrett, 2013a, b; McCurdy & Garrett, 2016; McClenaghan et al., 2020).

Table 1. Example of a 20-sample block including a blank, routine samples, a CRM, an analytical duplicate, and a field duplicate. Sample numbering follows the convention used in the GSC Field Application (https://github.com/NRCan/GSC-Field-Application) and the Surficial Data Model (Deblonde et al., 2024), where here "23" represents the year (2023), "BVB" is the geologist code (Brouard), "00XX" is the station number, "A" denotes the (first) earth material (i.e., the material from which the sample was collected at the station), and "0X" is the sample number.

Sequence No	Sample ID	Sample purpose
1	23BVB0000A01	blank
2	23BVB0001A01	routine
3	23BVB0002A01	routine
4	23BVB0000A02	certified reference material
5	23BVB0004A01	routine
6	23BVB0005A01	routine
7	23BVB0006A01	routine
8	23BVB0007A01	routine
9	23BVB0008A01	routine
10	23BVB0009A01	routine
11	23BVB0010A01	routine
12	23BVB0010A02	field duplicate
13	23BVB0011A01	routine
14	23BVB0012A01	routine
15	23BVB0013A01	routine
16	23BVB0014A01	routine
17	23BVB0010A03	analytical duplicate
18	23BVB0015A01	routine
19	23BVB0016A01	routine
20	23BVB0017A01	routine

Software requirements

The QA/QC protocols presented in this report require Microsoft Excel or equivalent spreadsheet software (e.g., Google Sheets or LibreOffice Calc) to organize and manage data tables. We recommend Notepad++ for flexible script editing; it is freely available at notepad-plus-plus.org. Python 3 must be installed on the system, and the installer can be downloaded from python.org. Be sure to select the option 'Add Python to PATH' during installation to enable command-line execution of the scripts. Several Python libraries are required for successful operation of the QA/QC scripts. These include Pandas for data manipulation, NumPy for numerical computing, SciPy for statistical operations, Matplotlib for data visualization, and tqdm for progress tracking. These packages can be installed from a command prompt (press Windows + R, type "cmd", and hit Enter) by entering the following commands in sequence: "pip install pandas", "pip install numpy", "pip install scipy", "pip install matplotlib", and "pip install tqdm".

Formatting Requirements for QA/QC Data Input

The QA/QC scripts expect a single, consistently structured CSV file named input.csv. This file must contain four key columns: a unique sample identifier, a sample group, a sample purpose, and one or more measurement values. Proper formatting is essential to ensure compatibility with the scripts (see **Suppl. material 1** and **Table 2**). The following columns must be included, with headers in lowercase and exactly matching the required naming conventions (sample, group, purpose) as outlined in Table 2:

- **sample**: A unique name, label, or identifier for each sample.
- **group**: An identifier that links routine samples with their associated field and analytical duplicates. It also links blanks and certified reference materials (CRMs) with their expected values and uncertainties. For samples collected using the GSC Field Application (https://github.com/NRCan/GSC-FieldApplication), this field should correspond to the "station" identifier.
- **purpose**: A label indicating the role of the sample in the QA/QC process. A drop-down list is provided in input.xlsx to help ensure consistency before saving the file as input.csv. Accepted values include:
  - **expected**: a certified reference value, known blank value, or detection limit.
  - **uncertainty**: the 1-sigma standard deviation or an accepted tolerance level.
  - **routine**: an original measurement.
  - **lab**: a measurement derived from an analytical duplicate.
  - **field**: a measurement derived from a field duplicate.
  - **blank**: a measurement from a blank sample.
  - **crm**: a measurement from a certified reference material.
- **measurements**: One or more columns containing the measured values. Column naming is flexible, but using a convention such as Analyte_Method_Unit is recommended for consistency and readability. For example, copper measured by ICP-MS following an Aqua Regia digestion could be labeled Cu_AR_ICPMS_ppm.

Special Special cases in measurement data should be handled as follows:

- Values below detection limit (e.g., "<DL" or "<0.01") should be replaced with half the detection limit. Use conditional formatting in Excel to identify these values.
- Values above quantification limit should be replaced with twice the upper limit.
- Missing values should be replaced with "NaN" (Not a Number).

Table 2. Example of a formatted input table containing all required fields for the QA/QC protocols. The table includes routine samples, field and analytical duplicates, blanks, and CRMs, as well as expected and uncertainty values for each CRM. CRM1 and CRM2 refer to certified reference materials (e.g., Till-4, OREAS-46, OREAS-47) used for accuracy monitoring. Values associated with CRM entries under expected and uncertainty are used to evaluate analytical accuracy and precision. Several CRM and blank samples (e.g., 23BVB0005A01 and 23BVB0000A01) have been integrated into the dataset and labeled as routine to prevent laboratories from recognizing their nature and biasing results. Detection limit values are entered as expected values, and measurements below detection have been replaced with half the detection limit to allow statistical processing.

* Value below detection limit and replaced with half the detection limit.

** This sample is a blank, labeled as routine to ensure anonymity during analysis.

*** This sample is a CRM (CRM1), labeled as routine to ensure anonymity during analysis.

sample	group	purpose	measurement1	measurement2	measurement3	measurement4
detection_limit	detection_limit	expected	0.01*	0.01*	10.00*	0.50*
CRM1	CRM1	expected	4.05	16.50	12.20	3.50
CRM1	CRM1	uncertainty	0.60	0.10	NaN	0.80
CRM2	CRM2	expected	6.70	25.20	50.00	4.20
CRM2	CRM2	uncertainty	0.70	1.55	5.50	0.20
23BVB0000A01**	silica_blank	blank	0.01*	0.01*	10.00*	0.50*
23BVB0001A01	23BVB0001	routine	3.50	2.20	22.10	0.50
23BVB0001A01	23BVB0001	lab	NaN	2.15	23.50	6.75
23BVB0002A01	23BVB0002	routine	4.20	4.40	75.45	5.50
23BVB0003A01	23BVB0003	routine	7.60	NaN	NaN	4.35
23BVB0004A01	23BVB0004	routine	5.55	5.25	48.70	0.50
23BVB0004A02	23BVB0004	field	5.85	5.55	45.65	3.85
23BVB0004A03	23BVB0004	lab	3.84	5.75	46.95	3.45
23BVB0005A01***	CRM1	crm	4.10	2.31	11.80	3.20
23BVB0006A01	23BVB0006	routine	8.85	7.52	26.25	3.32
23BVB0007A01	23BVB0007	routine	2.52	6.63	85.63	6.36
23BVB0007A02	23BVB0007	field	2.15	7.10	85.50	NaN
23BVB0007A03	23BVB0007	lab	2.35	6.10	84.89	6.55
23BVB0008A01	23BVB0008	routine	1.85	5.25	55.55	7.63
23BVB0009A01***	CRM1	crm	1.51	2.32	12.10	NaN
23BVB0010A01	23BVB0010	routine	7.62	14.12	99.85	11.10
23BVB0011A01	23BVB0011	routine	5.63	5.55	20.35	9.80
23BVB0012A01	23BVB0012	routine	2.56	9.56	33.25	5.46
23BVB0013A01	23BVB0013	routine	2.45	8.26	35.65	5.96
23BVB0014A01	23BVB0014	routine	3.26	7.62	38.57	9.23

QA/QC protocols

Evaluation of till compositional data, what to look for and evaluate

The evaluation of compositional data generally involves assessing three core aspects: accuracy, precision, and variability in the results produced by the analytical laboratory. Variations in the dataset or errors could come from a wide range of sources including instrumental errors (calibration issues, drift in instrument readings, or malfunctions during analysis), sample contamination (during collection, preparation, or analysis), detection limit issues and missing data, analytical variability (variability in measurements between different analytical runs or instruments), data entry errors (mistakes in data entry or transcription), outliers (data points that deviate significantly from the expected range), and/or sample mix-ups (incorrect labelling or handling of samples).

To detect outliers and potential data entry issues, **Protocol 1: Basic Statistics** should be used. This protocol generates fundamental statistics, verifies whether there are sufficient measurements above detection limits, and produces visual plots for interpretation. To assess whether reference materials align with expected values, **Protocol 2: Accuracy Analysis** compares laboratory measurements to expected values for blanks, standards, and CRMs, helping to identify any analytical bias or systematic discrepancies. To evaluate measurement consistency and repeatability, **Protocol 3: Precision Analysis** compares routine samples with their corresponding analytical duplicates to quantify analytical variability. To examine differences between routine samples and their field duplicates and to assess within-group variability, **Protocol 4: Field Variance Analysis** should be applied. Finally, **Protocol 5: Routine, Analytical, and Field Variance Analysis** provides a combined assessment of variability across all three sample purposes. This protocol is designed to quantify both analytical and sampling-related variability and to evaluate whether the data are sufficiently robust for interpretation or mapping.

Protocol 1: Generate Basic Statistics

Protocol 1 serves as a critical first step in assessing data quality. It generates summary statistics, identifies values below detection limits, and produces plots for visual inspection. These outputs help identify potential issues before proceeding to further QA/QC protocols.

To run the protocol, double-click the run_protocol1.bat file. This executes the Python script (protocol1.py) and processes the input table (input.csv). The script generates a CSV output (protocol1_output.csv) with summary statistics (see **Table 3**) and a folder of visualizations for each measurement column.

The statistical output includes key metrics such as measures of central tendency (mean, median, and mode), variability (standard deviation, variance, coefficient of variation, and median absolute deviation [MAD]), and distribution (range, orders of magnitude of the range, skewness, and kurtosis; **Table 3**). Percentile values, including minimum, maximum, and selected percentiles (2nd, 5th, 10th, 90th, 95th, and 98th), are also calculated. Together, these statistics provide insights into the data's distribution, central tendency, and variability. For example, understanding the range and variability of element concentrations aids in identifying outliers or anomalies, while measures of central tendency such as the mean and median reveal the typical values observed in the dataset. Metrics such as skewness and kurtosis shed light on the shape and symmetry of the distribution, while the median and MAD offer robust alternatives to the mean and standard deviation, particularly for datasets with outliers.

The protocol also addresses issues related to heteroscedasticity. If the range of a measurement exceeds 1.5 orders of magnitude, this may indicate heteroscedasticity, suggesting that the data would benefit from a log transformation in subsequent analyses. This threshold can be changed by modifying the line "*high_magnitude = row['order_of_magnitude'] > 1.5*". In addition, the output reports how many values fall below the detection limit and what proportion of the dataset they represent, as a high number of censored values can compromise the reliability of further analyses. If more than 12% of values are below detection limit, the distribution is skewed and is no longer normal (Fletcher, 1981). Variables exceeding this threshold are flagged by default, though this value can be customized in the script: "*below_limit = row['percentage below detection limit'] < 12*". While flagged variables can still be used in downstream analyses, caution is advised, as high variability or censoring may compromise the robustness of interpretations, particularly in multivariate analyses such as principal component analysis (PCA).

Scatterplots are generated for each measurement column, with data points ordered according to the sequence in the input table (**Fig. 1**). Points are colour-coded by purpose. A red line indicates the detection limit, while dashed lines show the mean and ±2 standard deviations. These visual elements can be customized in the script under the *generate_plots* function. Each plot is exported in both Portable Network Graphics (.png) and Scalable Vector Graphics (.svg) formats.

Together, the statistics and visualizations generated by Protocol 1 provide a solid foundation for evaluating data quality and support more advanced QA/QC procedures and multivariate analyses.

Table 3. Example output from Protocol 1.

measurement	measurement1	measurement2	measurement3	measurement4
count	12	11	11	12
mean	4.63	6.94	49.21	5.81
std	2.37	3.12	26.96	3.36
min	1.85	2.20	20.35	0.50
2%	1.98	2.64	20.70	0.50
5%	2.18	3.3	21.23	0.50
10%	2.46	4.4	22.10	0.78
90%	7.62	9.56	85.63	9.74
95%	8.17	11.84	92.74	10.39
98%	8.58	13.21	97.01	10.81
max	8.85	14.12	99.85	11.10
mode	1.85	5.25	20.35	0.50
variance	5.63	9.76	726.66	11.29
skewness	0.61	1.02	0.83	-0.17
kurtosis	-1.05	2.13	-0.54	-0.55
coefficient of variation	0.51	0.45	0.55	0.58
median (50%)	3.85	6.63	38.57	5.73
MAD	1.55	1.38	16.47	2.16
range	7.00	11.92	79.50	10.60
order_of_magnitude	0.85	1.08	1.90	1.03
entries below detection limit	0	0	0	2
percentage below detection limit	0	0	0	16.67
interpretation	Number of values below detection limit is acceptable	Number of values below detection limit is acceptable	**Caution:** Number of values below detection limit is acceptable, but the order of magnitude of the range **necessitates log transformation** in subsequent protocols	**Caution:** Number of values below detection limit is too high

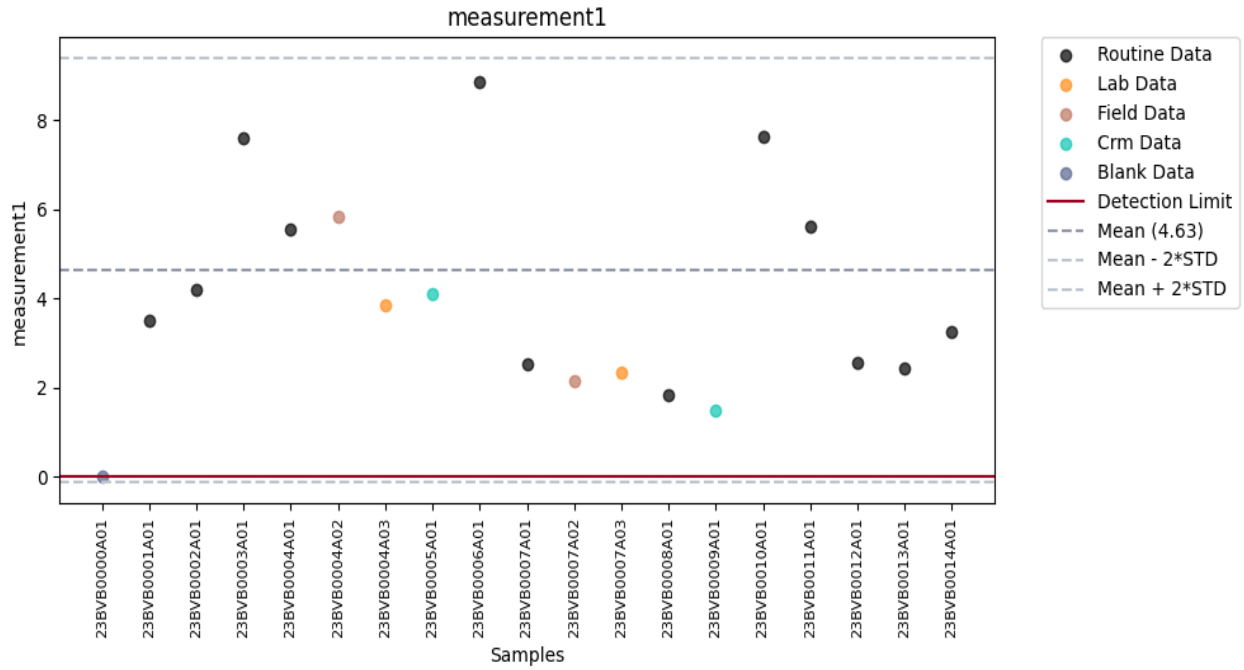

Figure 1. Example scatterplot produced by Protocol 1 for the data associated with measurement1 (from table 2). This scatterplot allows for visual inspection of trends, outliers, and overall variability across the sample sequence.

Protocol 2: Accuracy Analysis

Protocol 2 is designed to compare measured values against expected values for reference materials, including blanks, standards, and CRMs. This protocol supports the evaluation of analytical accuracy by determining whether measured values fall within an acceptable range. To mitigate potential issues related to heteroscedasticity, we strongly recommend log-transforming any measurements exhibiting a range greater than 1.5 orders of magnitude, as identified in Protocol 1, before proceeding with Protocol 2.

The script is run by double-clicking the file run_protocol2.bat, which executes the Python script protocol2.py. Upon execution, a command prompt will appear, prompting the user to select the group to evaluate (e.g., CRM1, CRM2, blank, Till-1, Till-4). The user selects the group by entering its corresponding number. The protocol can evaluate either a single group or all groups present in the dataset. For example, if the dataset includes CRM1, CRM2, and blank samples, the script can analyze all of them simultaneously. Note that, in the example dataset (**Table 2**), only CRM1 contains both expected and measured values. Therefore, when all groups are selected, output is generated only for CRM1.

If the selected group contains only one row of measured values, the script checks whether each measurement falls within an acceptable tolerance range. If published uncertainties are available, they are used to define this range. Otherwise, a default tolerance of ±10 percent is applied. This threshold can be adjusted manually by editing the script using Notepad++ under the section *# Check if the single measurement is within 10% of the expected value (or uncertainty if available)*.

If the group contains multiple measurements, such as the two measurements for CRM1 in Table 2, the script performs a one-sample t-test to compare the mean of the group to the expected value. For groups with more than 30 measurements, the relative standard deviation (RSD%) is also calculated to assess internal consistency. The RSD% estimates the precision for each reference material and supports accuracy assessment, as poor precision can mask analytical biases (e.g., Abzalov, 2011). However, accuracy should be primarily assessed using the p-value, which provides a statistical threshold for evaluation. By default, a p-value threshold of 0.05 (95% confidence) is used to determine whether the accuracy is acceptable. This threshold can be modified in the code under the section # Interpretation of p-value. In addition to the p-value assessment, the script evaluates whether the measurements fall within ±2 standard deviations of the mean (McCurdy and Garrett, 2016). If the p-value exceeds the threshold or if multiple measurements fall outside the acceptable range, the measurement is flagged as having low accuracy. This does not automatically disqualify the data from further use but indicates that further scrutiny is recommended before inclusion in downstream analyses.

The output file provides a summary of several statistics for each measurement analyzed (**Table 4**). These include the expected value (used as the reference value), the uncertainty (if available), the number of measurements included in the evaluation, the mean and standard deviation, the relative standard deviation, and the p-value. An interpretation row indicates whether the measurement has acceptable or low accuracy based on the defined threshold. In addition, the output includes a message stating how many measurements fall outside the expected tolerance range (either defined by published uncertainty or ±10 percent), along with a count of those values.

The protocol also generates Shewhart control charts (**Fig. 2**), which provide a visual representation of measurement accuracy. These plots help identify analytical problems such as drift or bias and offer a quick overview of measurement performance across reference samples (e.g., Piercey, 2014; Garrett et al., 2017).

Table 4. Example output from Protocol 2 showing accuracy evaluation results for four measurements. The table includes statistics such as the expected value, sample mean, standard deviation, relative standard deviation (RSD), p-value, and interpretation.

measurement	measurement1	measurement2	measurement3	measurement4
expected_value	4.05	16.5	12.2	3.5
uncertainty	0.60	0.10	1.22	0.80
sample_size	2	2	2	1
mean	2.81	2.32	11.95	3.20
std_dev	1.83	0.01	0.21	
rsd	65.29	0.31	1.78	
p_value	0.51	0.00	0.34	
interpretation	**Caution:** accuracy is low as sample mean falls outside the expected range (± 2 * uncertainty) and not significantly different from the expected value	**Caution:** accuracy is low as sample mean falls outside the expected range (± 2 * uncertainty) and significantly different from the expected value	Accuracy is acceptable as sample mean falls within the expected range (± 2 * uncertainty) and not significantly different from the expected value	Accuracy is acceptable as measurement is within tolerance (± 10% of expected value).
out_of_range_message	**Caution:** 1 measurement(s) fall outside the expected range (±2 * uncertainty)	**Caution:** 2 measurement(s) fall outside the expected range (±2 * uncertainty)	No values outside the expected range (± 2 * uncertainty)	No values outside the expected range (± 2 * uncertainty)
out_of_range_count	1	2	0	0

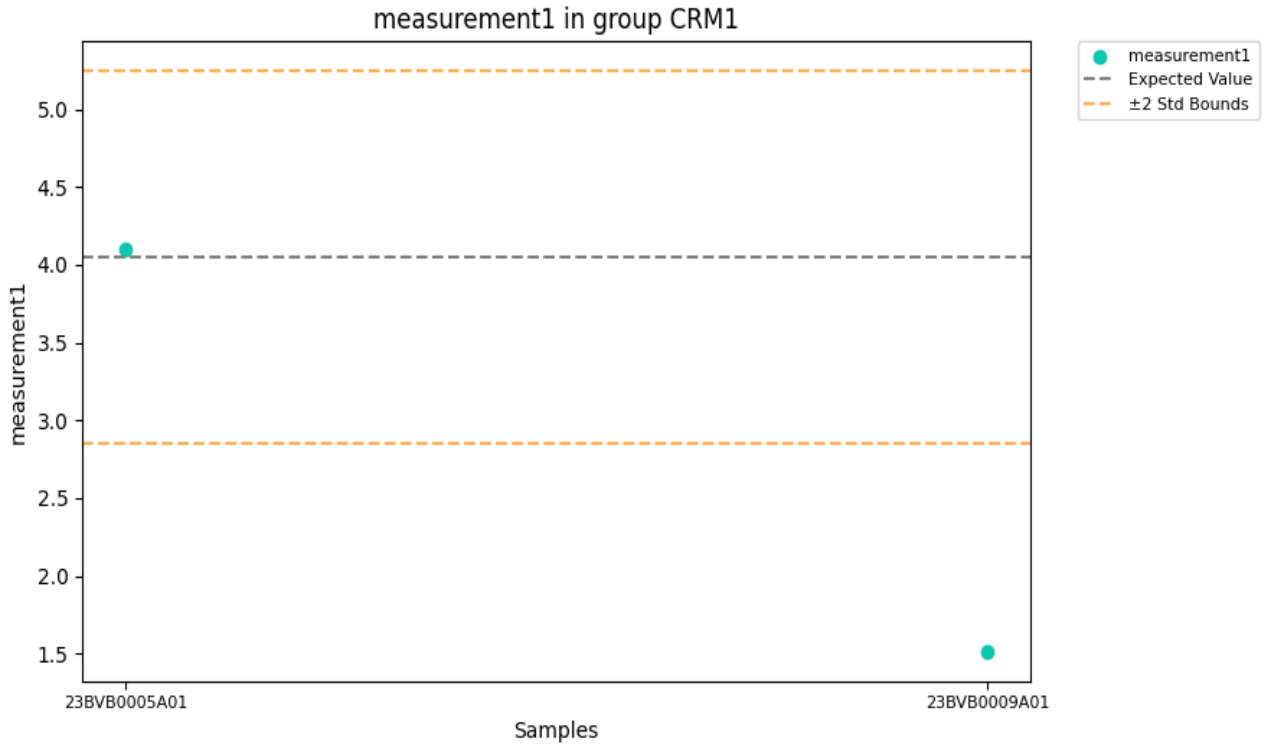

Figure 2. Example scatter plot produced by Protocol 2 for CRM1 and measurement1 (from Table 2). This plot visually demonstrates that sample 23BVB0009A01 falls well outside the expected range (±2 standard deviations), indicating the need for further quality control investigations.

Protocol 3: Precision Analysis

Protocol 3 is designed to evaluate the precision of measurement data by comparing routine samples with their analytical duplicates, either subsamples from the same material or splits sent to different analytical laboratories. It identifies and flags precision issues that may compromise the reliability of compositional datasets.

Before running the protocol, assess whether the data exhibit heteroscedasticity, particularly for measurements that span more than 1.5 orders of magnitude (as identified in Protocol 1). In such cases, log-transformation is strongly recommended to stabilize variance and ensure the validity of the precision analysis.

The script is run by double-clicking the file run_protocol3.bat, which executes the Python script protocol3.py. The protocol begins by reading the data from the input CSV file and filtering the dataset to include only valid entries containing both routine and analytical duplicate samples. Any rows with missing values for the measurement being evaluated are excluded to ensure the dataset is complete for statistical analysis.

For each measurement, an analysis of variance (ANOVA) is performed to compare the means of routine samples against their corresponding analytical duplicates or between different laboratories. The ANOVA produces an F-statistic and a p-value to assess whether the means differ significantly. A p-value below 0.05 indicates strong evidence against the null hypothesis, suggesting a significant difference between routine and duplicate measurements. A p-value above 0.05 implies no significant difference can be concluded.

In addition to ANOVA, the protocol computes descriptive statistics for each measurement, including the mean, median, standard deviation, and variance, calculated separately for routine and duplicate samples. The coefficient of variation (CV), defined as the ratio of the standard deviation to the mean, is used as a key indicator of measurement precision. A low CV indicates high precision, while a high CV signals greater variability.

A threshold of 20% relative standard deviation (RSD = 0.2) is used to evaluate precision. If the RSD exceeds this threshold, the measurement is flagged as having low precision. This threshold is based on the assumption that, on average, 95 out of 100 values should fall within two standard deviations of the mean (Fletcher, 1981). Elevated RSD values, especially when combined with a narrow range of values, can mask meaningful patterns or anomalies (Howarth and Martin, 1979). In such cases, the protocol alerts the user to possible data quality concerns.

The results are saved to a CSV output file and include: the number of unique groups retained and removed after filtering, mean values and standard deviations for routine and duplicate samples, the calculated RSD, and an interpretation of the precision. The output also reports variability within and between groups, along with the ANOVA results for both sample purpose and group comparisons (**Table 5**). Users may customize the thresholds for both RSD and p-value by modifying the corresponding sections of the code (i.e., *# Add interpretation based on RSD* and *# Check if there are at least two groups present*).

Table 5. Example output from Protocol 3 showing precision analysis results for analytical duplicates.

measurement	measurement1	measurement2	measurement3	measurement4
unique group after filtering	2	3	3	3
unique group dropped	1	0	0	0
mean	3.57	4.68	51.96	4.02
Mean of routine	4.04	4.69	52.14	2.45
Mean of lab	3.10	4.67	51.78	5.58
median	3.18	5.50	47.83	4.91
standard deviation	1.48	1.99	28.13	2.98
Std of routine	2.14	2.27	31.90	3.38
Std of lab	1.05	2.19	30.98	1.85
variance within group	0.74	0.09	0.93	7.97
standard deviation within group	0.86	0.30	0.96	2.82
relative standard deviation	24.10	6.37	1.85	70.24
precision interpretation	**Caution:** Precision is lower than acceptable	Precision is acceptable	Precision is acceptable	**Caution:** Precision is lower than acceptable
variability between	0.70	0.41	0.00	0.30
variability within	0.30	0.59	1.00	0.70
f-statistic (purpose)	0.31	0.00	0.00	1.98
p-value (purpose)	0.63	0.99	0.99	0.23
interpretation (puspose)	The mean from the laboratory duplicates is not statistically different from the mean of the routine groups	The mean from the laboratory duplicates is not statistically different from the mean of the routine groups	The mean from the laboratory duplicates is not statistically different from the mean of the routine groups	The mean from the laboratory duplicates is not statistically different from the mean of the routine groups

Protocol 4: Field Variance Analysis

Protocol 4 evaluates field sampling variability by comparing routine samples with their field duplicates using an ANOVA. The primary goal is to partition the observed variability into between-group (between sites) and within-group (at site) components. A key objective is to assess whether within-group (or at site) variability – which reflects differences between samples collected at the same site – is significantly smaller than between-group variability. If so, this suggests that group-specific compositional signals dominate over random sampling error or analytical noise.

To ensure the validity of the ANOVA, measurements that span more than 1.5 orders of magnitude should be log-transformed prior to analysis, as outlined in Protocol 1. This transformation helps reduce the impact of heteroscedasticity and helps satisfy the ANOVA assumption of homogeneity of variance.

The script is run by double-clicking the file run_protocol4.bat, which executes the Python script protocol4.py. The output of the protocol includes several key statistics (**Table 6**), such as the number of unique groups (after filtering out incomplete or below-detection-limit samples), means and standard deviations for both routine and duplicate samples, within-group variance, relative standard deviation, variability within and between groups, the F-statistic, and the associated p-value. These metrics help determine whether observed variability reflects geological patterns or random error. A common threshold for interpreting results is an F-ratio (between-group variance divided by within-group variance) greater than 4.0 at a 95% confidence level, indicating a statistically significant difference. Exceeding this threshold suggests that analytical and sampling errors are small relative to regional variability, supporting the suitability of the data for mapping and statistical analysis (McCurdy and Garrett, 2016). Even when the F-ratio falls below the 4:1 threshold, results may still highlight outliers or anomalous groups potentially related to mineralization or other geological features. In such cases, p-values provide an additional tool to flag measurements for further investigation. Although the 4:1 threshold was originally established for lake sediment surveys with large datasets, it can be adjusted for smaller datasets by modifying the code in Notepad++ under the section titled *# Check if variability between is at least 4 times larger than the variability within*. This flexibility allows the protocol to be adapted for various survey types and sample densities.

Table 6. Example output from Protocol 4 showing results of field variance analysis.

measurement	measurement1	measurement2	measurement3	measurement4
unique groups	2	2	2	1
groups dropped	0	0	0	1
mean	4.02	6.13	66.37	2.18
median	4.04	6.09	67.10	2.18
standard deviation	1.95	0.88	22.20	2.37
Mean of routine	4.04	5.94	67.17	0.50
Mean of field	4	6.33	65.58	3.85
Std of routine	2.14	0.98	26.11	
Std of field	2.62	1.10	28.18	
variance within group	0.06	0.08	2.33	5.61
standard deviation within group	0.24	0.28	1.53	2.37
variability between	0.99	0.93	1	
variability within	0.01	0.07	0	
variability interpretation	Within-group variability is significantly smaller	Within-group variability is significantly smaller	Within-group variability is significantly smaller	**Caution**: Within-group variability is not significantly smaller
f-statistic	0.00	0.14	0.00	
p-value	0.99	0.75	0.96	
interpretation p-value	No significant difference	No significant difference	No significant difference	Insufficient data

Protocol 5: Routine, Analytical, and Field Variance Analysis

Protocol 5 is designed to assess the relative contributions of between-group and within-group variability by comparing routine samples with their associated field and analytical duplicates. By analyzing these groups separately, the protocol distinguishes between sampling (field) and analytical sources of variability (Garrett, 2013b). When analytical duplicates are systematically "sampled" from field duplicates, or when routine samples are associated with their respective field duplicates, it is possible to clearly distinguish these sources of variability.

To avoid potential issues with heteroscedasticity (particularly for measurements with a range exceeding 1.5 orders of magnitude, as noted in Protocol 1), we recommend log-transforming the data before proceeding with this protocol.

The script is run by double-clicking the file run_protocol5.bat, which executes the Python script protocol5.py. The protocol produces an output file containing several statistics (**Table 7**), including the number of unique groups after filtering, the number of groups dropped due to missing data, and descriptive statistics for the overall dataset, as well as for routine samples, field duplicates, and analytical duplicates (mean, median, and standard deviation). It also includes variance and standard deviation within-group, relative standard deviation, and the contributions of both between-group and within-group variability to the total variability. The F-statistic, p-value, and an interpretation of the p-value are provided as well.

The protocol performs an ANOVA to compare the means of routine samples and duplicates (both field and analytical). This test checks whether the means of the groups differ significantly, with a standard threshold of $p < 0.05$. If only one group (i.e., one set of routine and duplicate samples) is available, the script compares their means and standard deviations of the routine and duplicate samples to determine whether the values fall within an expected range. If all values are within range, the output will state: "The means are not statistically different". If any value falls outside the range, the interpretation row will state: "Caution: one of the means is statistically different". If insufficient data are available, the interpretation row will state: "Not enough data for interpretation".

In addition to the mean comparison, the variability ratio is calculated to determine the portion of total variability attributed to within-group versus between-group variability. This ratio compares the variability between groups to the variability within groups, and if the ratio exceeds 4.0 at a 95% confidence level, it suggests that the sampling and analytical errors are significantly smaller than the regional variability. If the ratio exceeds this threshold, it indicates that the data are highly consistent within groups, making it suitable for mapping and statistical analysis. The variability ratio can be adjusted if needed by modifying the code in Notepad++, specifically the value under the section "# Variability interpretation".

Table 7. Example of results from protocol 5.

measurement	measurement1	measurement2	measurement3	measurement4
unique groups analyzed	2	2	2	1
groups dropped	0	0	0	1
mean	3.71	6.06	66.22	2.60
median	3.18	5.93	66.80	3.45
standard deviation	1.65	0.70	20.97	1.83
Mean of lab	3.10	5.93	65.92	3.45
Mean of routine	4.04	5.94	67.17	0.50
Mean of field	4.00	6.33	65.58	3.85
Std of lab	1.05	0.25	26.83	
Std of routine	2.14	0.98	26.11	
Std of field	2.62	1.10	28.18	
variance within group	0.61	0.16	1.25	
standard deviation within group	0.78	0.40	1.12	
variability between	0.86	0.79	1.0	
variability within	0.14	0.21	0.00	
variability interpretation	Within-site variability is significantly smaller than regional variability	**Caution**: Within-site variability is not significantly smaller than regional variability	Within-site variability is significantly smaller than regional variability	Insufficient data
f-statistic	0.14	0.14	0.00	
p-value	0.88	0.88	1.00	
interpretation p-value	No significant difference between group means	No significant difference between group means	No significant difference between group means	Insufficient data

Conclusions

The implementation of these streamlined QA/QC protocols represents a significant advancement in ensuring the accuracy, reliability, and usability of compositional data derived from tills and other geological materials. Designed to complement existing QA/QC methodologies developed in R at the GSC (Garrett and Chen, 2007; Garrett, 2013a,b; McCurdy and Garrett, 2016), these protocols offer added benefits, particularly in terms of simplicity, accessibility, and clearer guidance for interpretation.

Future enhancements should focus on expanding user support, such as integrating additional visualizations and implementing more advanced statistical tools. Continued refinement of these protocols will help better meet the needs of researchers and strengthen the robustness of compositional data analysis across a broad range of geoscientific applications.

Acknowledgements

This report is a GEM-GeoNorth contribution. Special thanks to Jessey M. Rice for diligently testing the early version of the codes. We also extend our appreciation to Marty McCurdy and Bob Garrett for their thorough review of the manuscript and the codes. ChatGPT was used to assist in the development of the code.

8. References

Abzalov, M. 2011. Sampling Errors and Control of Assay Data Quality in Exploration and Mining Geology, in: Ivanov, O. (Ed.), Applications and Experiences of Quality Control. IntechOpen, Rijeka, Croatia, 611-644, https://doi.org/10.5772/14965.

Burnham, O.M., Schweyer, J. 2004. Trace element analysis of geological samples by inductively coupled plasma mass spectrometry at the Geoscience Laboratories: Revised capabilities due to improvements to instrumentation. In: Baker, C.L., Debicki, E.J., Parker, J.R., Kelly, R.I., Ayer, J.A., Easton, R.M. (Eds.), Summary of field work and other activities. Open File Report 6145, Ontario Geological Survey, 54-51-54-20, https://prd-0420-geoontario-0000-blob-cge0eud7azhvfsf7.z01.azurefd.net/lrc-geology-documents/publication/OFR6145/OFR6145.pdf.

Deblonde, C., Campbell, J.E., Chow, W., Cocking, R.B., Huntley, D.H., Parent, M.P., Rice, J.M., Robertson, L., Smith, I.R., Weatherston, A.J., Zawadzka, K. 2024. Surficial Data Model: the science language of the integrated Geological Survey of Canada data model for surficial geology maps. version 2.5.1. Geological Survey of Canada, Open File Report 8236, 1-9, https://doi.org/10.4095/332530.

Fletcher, W.K. 1981. Analytical Methods in Geochemical Prospecting. Handbook of Exploration Geochemistry 1. Elsevier Science B.V., https://doi.org/https://doi.org/10.1016/B978-0-444-41930-9.50001-5.

Garrett, R. G. and Chen, Y. 2007. rgr: The GSC (Geological Survey of Canada) Applied Geochemistry EDA Package - R tools for determining background ranges and thresholds. Open file report 5583, Geological Survey of Canada, https://doi.org/10.4095/224575.

Garrett, R.G. 2013a. The "rgr" package for the R Open Source statistical computing and graphics environment - a tool to support geochemical data interpretation. Geochemistry: Exploration, Environment, Analysis 13(4), 355-378, https://doi.org/10.1144/geochem2011-106.

Garrett, R.G. 2013b. Assessment of local spatial and analytical variability in regional geochemical surveys with a simple sampling scheme. Geochemistry: Exploration, Environment, Analysis 13(4), 349-354, https://doi.org/10.1144/geochem2011-085.

Garrett R.G., Reimann, C., Hron, K., Kynčlová, P. and Filzmoser, P. 2017. Finally, a correlation coefficient that tells the geochemical truth. Explore Newsletter,176, 1-10, https://doi.org/10.70499/YKCX6512.

Horowitz, A.J. 1991. A Primer on Sediment-Trace Element Chemistry, 2nd Edition. Open File Report 91-76, United States Geological Survey, 1-136, https://pubs.usgs.gov/of/1991/0076/report.pdf.

Howarth, R.J., Martin, L. 1979. Computer-based techniques in the compilation, mapping and interpretation of exploration geochemical data. In: Hood, P.J. (Ed.), Geophysics and Geochemistry in the Search for Metallic Ores. Economic Geology Report 31, Geological Survey of Canada, Ottawa, Ontario, Canada, 545-574, https://doi.org/10.4095/106065.

Lynch, J. 1996. Provisional Elemental Values for four new Geochemical Soil and Till Reference Materials, TILL-1, TILL-2, TILL-3 and TILL-4. Geostandards Newsletter 20(2), 277-287, https://doi.org/10.1111/j.1751-908X.1996.tb00189.x.

McClenaghan, M.B., Spirito, W.A., Plouffe, A., McMartin, I., Campbell, J.E., Paulen, R.C., Garrett, R.G., Hall, G.E.M., Pelchat, P., Gauthier, M.S. 2020. Geological Survey of Canada till-sampling and analytical protocols: from field to archive, 2020 update. Geological Survey of Canada, Open File 8591, Natural Resources Canada, 1-73, https://doi.org/10.4095/326162.

McClenaghan, M.B., Paulen, R.C., Smith, I.R., Rice, J.M., Plouffe, A., McMartin, I., Campbell, J.E., Lehtonen, M., Parsasadr, M., Beckett-Brown, C.E. 2023. Review of till geochemistry and indicator mineral methods for mineral exploration in glaciated terrain. Geochemistry: Exploration, Environment, Analysis 0(ja), geochem2023-2013, https://doi.org/10.1144/geochem2023-013.

McCurdy, M.W., Garrett, R.G. 2016. Geochemical data quality control for soil, till and lake and stream sediment samples. Open File 7944, National Resources Canada, 1-35, https://doi.org/10.4095/297562.

Piercey, S. J. 2014. Modern Analytical Facilities 2. A Review of Quality Assurance and Quality Control (QA/QC) Procedures for Lithogeochemical Data. Geoscience Canada, 41 (1), 75-88, https://doi.org/10.12789/geocanj.2014.41.035.