Ressources naturelles Canada
Natural Resources Canada

COMMISSION GÉOLOGIQUE DU CANADA
DOSSIER PUBLIC 9309f

Contrôle de la qualité et validation des données compositionnelles des tills à l'aide de Python

E. Brouard et P.-M. Godbout

2026

Canada

COMMISSION GÉOLOGIQUE DU CANADA
DOSSIER PUBLIC 9309f

Contrôle de la qualité et validation des données compositionnelles des tills à l'aide de Python

E. Brouard et P.-M. Godbout

2026

On peut télécharger cette publication gratuitement à partir du Dépôt ouvert des sciences et technologie de RNCan (https://ostrnrcan-dostrncan.canada.ca/).

Notation bibliographique conseillée
Brouard, E. et Godbout, P.-M. 2026. Contrôle de la qualité et validation des données compositionnelles des tills à l'aide de Python; Commission géologique du Canada, Dossier public 9309f, 1 fichier zip. https://doi.org/10.4095/pzydve44br

Les publications de cette série ne sont pas révisées; elles sont publiées telles que soumises par l'auteur.

ISSN 2816-7163
ISBN 978-0-660-98068-3
Nº de catalogue M183-2/9309F-PDF
https://doi.org/10.4095/pzydve44br

Contrôle de la qualité et validation des données compositionnelles des tills à l'aide de Python

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

1 Commission géologique du Canada – Division de Québec, Ressources naturelles Canada, 490, rue de la Couronne, Québec (Québec) G1K 9A9, Canada

2 Commission géologique du Canada – Division centrale, Ressources naturelles Canada, 601, rue Booth, Ottawa (Ontario) K1A 0E8, Canada

Les données compositionnelles issues des tills sont couramment utilisées en géosciences pour la caractérisation de processus géologiques, de conditions environnementales et de contextes d'exploration minérale. La qualité de ces données est toutefois déterminante pour la fiabilité des analyses et des décisions qui en découlent. Cela nécessite des procédures rigoureuses et cohérentes de contrôle et de validation des données, appliquées de manière systématique tout au long du traitement analytique. Dans cette optique, le présent rapport présente un ensemble complet de protocoles simplifiés de contrôle de la qualité et de validation des données (ang., Quality Assurance/Quality Control; QAQC) adaptés aux données de till. Conçus pour simplifier les flux de travail de QAQC — en particulier pour les utilisateurs ayant une expérience limitée en programmation ou en analyse statistique — ces protocoles combinent des procédures simples à des scripts Python conviviaux. Ils offrent également des représentations graphiques facilitant l'interprétation des statistiques, ainsi qu'une approche systématique pour évaluer l'intégrité et la fiabilité des ensembles de données. Le manuscrit présente par ailleurs les principales considérations liées à une planification rigoureuse, au formatage approprié des données et à l'utilisation efficace d'outils tels que Microsoft Excel et Python. En suivant les procédures décrites et en utilisant les scripts fournis, les utilisateurs peuvent produire des statistiques sommaires, évaluer l'exactitude des mesures, analyser les différences entre les échantillons de routine et leurs duplicatas analytiques et déterminer si la variabilité des données est acceptable à des fins de cartographie et d'analyses statistiques.

Introduction

Les données compositionnelles provenant de tills, ou de tout autre type d'échantillon géologique, obtenues à partir de différents laboratoires doivent faire l'objet d'une évaluation rigoureuse de leur exactitude et de leur précision avant leur diffusion. Un programme de contrôle de la qualité et de validation des données (QAQC) soigneusement planifié pour la collecte, la préparation et l'analyse des échantillons, ainsi que pour la gestion et l'archivage des données, permet de s'assurer qu'aucun biais systématique n'est introduit dans les données. À la Commission géologique du Canada (CGC), des protocoles internes de QAQC ont été élaborés à l'aide de plateformes logicielles telles que Microsoft Excel et l'environnement libre R (Garrett et Chang, 2007 ; Garrett, 2013a ; McCurdy et Garrett, 2016). Ces travaux ont permis d'établir des approches solides pour certaines étapes du processus analytique. Toutefois, bien que des procédures robustes de QAQC existent pour la collecte d'échantillons sur le terrain (p. ex., McClenaghan et al., 2020, 2023), il manque encore un cadre unifié pour l'évaluation de la qualité des données compositionnelles qui en résultent. Par conséquent, les pratiques de QAQC varient d'un groupe de recherche à l'autre, reflétant des différences de méthodologie, de logiciels et d'interprétation. Ce manque de normalisation s'explique en partie par la complexité de certains logiciels, leur prise en main parfois difficile pour des utilisateurs sans formation en programmation, ainsi que par des enjeux de compatibilité entre plateformes (McCurdy et Garrett, 2016). Dans l'ensemble, il existe un besoin manifeste de procédures de QAQC simplifiées, faciles à comprendre, accessibles à des utilisateurs ayant des profils techniques variés et conçues pour assurer la reproductibilité des études. Nous présentons ici un ensemble de cinq protocoles de QAQC, adaptés à des objectifs précis d'évaluation de la qualité des données, et visant à simplifier le processus d'évaluation de l'exactitude, de la précision et de la variabilité des ensembles de données compositionnelles. En tirant parti des capacités du langage de programmation Python, ces méthodes offrent aux utilisateurs, quelle que soit leur expérience en programmation, des outils clairs et efficaces pour évaluer l'intégrité et la fiabilité de leurs données. Afin d'en faciliter l'utilisation, les protocoles ont été regroupés dans un ensemble convivial (voir **matériel supplémentaire**). Celui-ci comprend des gabarits de saisie et des scripts facilement exécutables, qui génèrent des tables de statistiques sommaires et des guides d'interprétation. Le présent rapport traite d'abord des principales considérations liées à la planification et aux exigences logicielles, puis présente le fonctionnement de chaque protocole, y compris ses objectifs, sa mise en œuvre et ses résultats. `

Considérations pour la planification de la QAQC

Une planification adéquate des procédures QAQC est essentielle tout au long des travaux de terrain et des analyses en laboratoire afin de minimiser les erreurs, d'assurer l'intégrité des données et de permettre une interprétation pertinente. Cette planification comprend la collecte systématique d'échantillons de routine accompagnés de duplicatas de terrain, ainsi que la préparation de duplicatas analytiques :

- Les **échantillons de routine** (*ang., routine*) sont prélevés à des emplacements précis afin de soutenir l'investigation scientifique.
- Les **duplicatas de terrain** (*ang., field duplicate*) sont des échantillons supplémentaires prélevés au même site que l'échantillon de routine afin d'évaluer la variabilité d'échantillonnage et l'hétérogénéité potentielle à l'échelle du site.
- Les **duplicatas analytiques** (*ang., analytical duplicate*) sont dérivés des échantillons de routine ou des duplicatas de terrain afin d'évaluer la variance et d'assurer la cohérence analytique. Les duplicatas analytiques doivent être prélevés à partir des duplicatas de terrain afin de permettre une estimation indépendante de la variabilité liée à l'échantillonnage et à l'analyse (Garrett, 2013b).

Une planification efficace des protocoles de QAQC doit également inclure l'insertion de matériaux de référence certifiés (*ang., certified reference material*; CRM) dans chaque lot d'échantillons soumis à l'analyse. Il est suggéré que les CRMs — tels que Till-1, Till-4, OREAS-46 et OREAS-47 — présente une matrice minéralogique comparable à celle des échantillons de routine afin d'assurer leur compatibilité (Horowitz, 1991). Ces matériaux sont analysés conjointement avec les échantillons de routine afin de surveiller l'exactitude analytique. Les résultats d'analyse des CRMs sont ensuite comparés aux valeurs attendues ou provisoires. Celles-ci sont publiées soit en ligne par les laboratoires (p. ex., oreas.com/crm/oreas-46, oreas.com/crm/oreas-47), soit dans la littérature scientifique (p. ex., Lynch, 1996 ; Burnham et Schweyer, 2004). Cette comparaison permet de détecter d'éventuels problèmes, tels qu'une dérive analytique ou un biais systématique au fil du temps.

La séquence d'analyse des échantillons doit être soigneusement planifiée avant la soumission d'un lot à un laboratoire externe. La CGC recommande l'utilisation d'un « système de blocs de 20 échantillons », chaque bloc étant composé de 17 échantillons de routine, d'un CRM, d'un duplicata analytique et d'un duplicata de terrain (**tableau 1**). Dans ce cadre, des blancs doivent également être insérés près du début, du milieu et de la fin de chaque lot afin de surveiller une contamination potentielle provenant d'échantillons traités précédemment. Il est recommandé de répartir de façon aléatoire la position des échantillons de routine, des duplicatas et des CRM au sein de chaque bloc afin d'éviter tout biais lié à la position. L'échantillon de CRM peut être positionné à n'importe quel endroit dans le bloc, à l'exception de la première position, généralement réservée à l'échantillon blanc. Les éléments présentés ci-dessus résument les principes de base de la planification des procédures de QAQC. Pour des lignes directrices détaillées et une description complète des bonnes pratiques, les chercheurs sont invités à consulter les publications antérieures de la CGC (Garrett, 2013a, b ; McCurdy et Garrett, 2016 ; McClenaghan et al., 2020).

Tableau 1. Exemple d'un bloc de 20 échantillons comprenant un blanc (blank), des échantillons de routine (routine), un matériau de référence certifié (certified reference material), un duplicata analytique (analytical duplicate) et un duplicata de terrain (field duplicate). La numérotation des échantillons suit la convention utilisée dans l'application de terrain de la CGC (https://github.com/NRCan/GSC-Field-Application) et dans le Modèle de données en géologie de surface (Deblonde et al., 2024), où « 23 » représente l'année (2023), « BVB » correspond au code du géologue (Brouard), « 00XX » au numéro de station, « A » désigne le (ici premier) matériau terrestre (c.-à-d. le matériau à partir duquel l'échantillon a été prélevé à la station) et « 0X » le numéro de l'échantillon.

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

Logiciels requis

Les protocoles de QAQC présentés dans ce rapport nécessitent Microsoft Excel ou un logiciel équivalent (p. ex., Google Sheets ou LibreOffice Calc) pour organiser et gérer les tables de données. Nous recommandons Notepad++ pour l'édition flexible des scripts ; il est disponible gratuitement à l'adresse notepad-plus-plus.org. Python 3 doit être installé sur le système, et le programme d'installation peut être téléchargé à partir de python.org. Assurez-vous de sélectionner l'option 'Add Python to PATH' lors de l'installation afin de permettre l'exécution des scripts en ligne de commande. Plusieurs bibliothèques Python sont requises pour le bon fonctionnement des scripts. Celles-ci comprennent *Pandas* pour la manipulation des données, *NumPy* pour le calcul numérique, *SciPy* pour les opérations statistiques, *Matplotlib* pour la visualisation des données et *tqdm* pour le suivi de la progression du script. Ces bibliothèques peuvent être installées à partir d'une fênetre de commandes (appuyer sur Windows + R, saisir « cmd », puis appuyer sur Entrée) en entrant successivement les commandes suivantes : "pip install pandas", "pip install numpy", "pip install scipy", "pip install matplotlib" et "pip install tqdm".

Exigences de formatage pour les données d'entrée QAQC

Les scripts de QAQC utilisent un seul fichier CSV commun, structuré de manière cohérente, nommé input.csv. Ce fichier doit contenir au moins quatre colonnes clés : un identifiant unique d'échantillon, un groupe d'échantillons, une utilité d'échantillon et une ou plusieurs valeurs de mesure. Un formatage approprié est essentiel pour assurer la compatibilité avec les scripts (voir **matériel suppl.** et **tableau 2**). Les colonnes suivantes doivent être incluses, avec des en-têtes en minuscules et correspondant exactement aux conventions de nommage requises (sample, group, purpose) telles que présentées dans le tableau 2 :

- **sample** : nom, étiquette ou identifiant unique pour chaque échantillon.
- **group** : identifiant reliant les échantillons de routine à leurs duplicatas de terrain et analytiques associés. Il relie également les blancs et les CRMs à leurs valeurs et incertitudes attendues. Pour les échantillons collectés à l'aide de l'application de terrain de la CGC, ce champ peut correspondre à l'identifiant "station".
- **purpose** : étiquette indiquant le rôle de l'échantillon dans le processus de QAQC. Une liste déroulante est fournie dans input.xlsx afin d'éviter des erreurs de frappes avant l'enregistrement du fichier sous format csv (input.csv). Les valeurs acceptées comprennent :
    - **expected** : valeur de référence certifiée, valeur connue d'un blanc ou limite de détection.
    - **uncertainty** : écart-type à 1 sigma ou niveau de tolérance/incertitude accepté.
    - **routine** : mesure originale.
    - **lab** : mesure issue d'un duplicata analytique.
    - **field** : mesure issue d'un duplicata de terrain.
    - **blank** : mesure provenant d'un échantillon blanc.
    - **crm** : mesure provenant d'un matériau de référence certifié.
- **measurements** : une ou plusieurs colonnes contenant les valeurs mesurées. Le noms des colonnes est flexible, mais l'utilisation d'une convention telle que Analyse_Methode_Unité est recommandée pour assurer la cohérence et la lisibilité. Par exemple, le cuivre mesuré par ICP-MS après une digestion Aqua Regia pourrait être libellé Cu_AR_ICPMS_ppm.

Les cas particuliers dans les mesures doivent être traités comme suit :

- Les valeurs inférieures à la limite de détection (p. ex., "<DL" ou "<0.01") doivent être remplacées par la moitié de la limite de détection. L'utilisation de la mise en forme conditionnelle dans Excel est recommandée pour identifier ces valeurs.
- Les valeurs supérieures à la limite de quantification doivent être remplacées par le double de la limite supérieure.
- Les valeurs manquantes doivent être remplacées par "NaN" (Not a Number).

Tableau 2. Exemple de tableau d'entrée formaté contenant tous les champs requis pour les protocoles de QAQC. Le tableau comprend des échantillons de routine, des duplicatas de terrain et analytiques, des blancs et des CRMs, ainsi que les valeurs attendues (expected) et l'incertitude (uncertainty) associées à chaque CRM. CRM1 et CRM2 désignent des matériaux de référence certifiés (p. ex., Till-4, OREAS-46, OREAS-47) utilisés pour le suivi de l'exactitude analytique. Les valeurs associées aux entrées CRM dans les champs expected et uncertainty sont utilisées pour évaluer l'exactitude et la précision analytiques. Plusieurs échantillons de CRM et de blancs (p. ex., 23BVB0005A01 et 23BVB0000A01) ont été intégrés à l'ensemble de données et étiquetés comme routine afin d'empêcher les laboratoires d'en reconnaître la nature et d'influencer les résultats. Les valeurs de limite de détection sont saisies comme valeurs attendues (expected), et les mesures inférieures à la limite de détection ont été remplacées par la moitié de cette limite afin de permettre le traitement statistique.

* Valeur inférieure à la limite de détection et remplacée par la moitié de la limite de détection.

** Cet échantillon est un blanc, étiqueté comme routine afin d'assurer l'anonymat lors de l'analyse.

*** Cet échantillon est un CRM (CRM1), étiqueté comme routine afin d'assurer l'anonymat lors de l'analyse.

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

Protocoles de QAQC

Évaluation des données compositionnelles des tills : éléments à examiner et à évaluer

L'évaluation des données compositionnelles repose généralement sur l'analyse de trois aspects fondamentaux : l'exactitude, la précision et la variabilité des résultats produits par le laboratoire. Les variations ou les erreurs potentielles observées dans un ensemble de données peuvent provenir de nombreuses sources, notamment des instruments (problèmes d'étalonnage, dérive des mesures ou dysfonctionnements durant l'analyse), de la contamination des échantillons (lors de la collecte, de la préparation ou de l'analyse), des enjeux liés aux limites de détection et aux données manquantes, de la variabilité analytique (variations entre différentes séries analytiques ou instruments), des erreurs de saisie des données (erreurs de transcription ou d'entrée), de la présence de valeurs aberrantes (données s'écartant fortement de l'intervalle attendu) et/ou des inversions d'échantillons (étiquetage ou manipulation incorrects).

Le protocole 1 (Statistiques descriptives de base) constitue la première étape de l'évaluation des données. Il permet de détecter les valeurs aberrantes et les problèmes potentiels de saisie, de vérifier la présence d'un nombre suffisant de mesures supérieures aux limites de détection et de produire des représentations graphiques facilitant l'interprétation.

Le protocole 2 (Analyse de l'exactitude analytique) vise à évaluer la concordance entre les matériaux de référence et les valeurs attendues. Il compare les mesures produites par le laboratoire aux valeurs de référence associées aux blancs, aux standards et aux CRMs, afin de mettre en évidence d'éventuels biais analytiques ou écarts systématiques.

Le protocole 3 (Analyse de la précision analytique) permet d'évaluer la cohérence et la répétabilité des mesures en comparant les échantillons de routine à leurs duplicatas analytiques correspondants, ce qui permet de quantifier la variabilité analytique.

Le protocole 4 (Analyse de la variabilité liée à l'échantillonnage sur le terrain) est utilisé pour examiner les différences entre les échantillons de routine et leurs duplicatas de terrain, et pour évaluer la variabilité intra-groupe associée à l'échantillonnage.

Enfin, le protocole 5 (Analyse intégrée de la variabilité analytique et de l'échantillonnage) propose une évaluation globale de la variabilité associée aux différents types d'échantillons. Il vise à quantifier à la fois la variabilité analytique et celle liée à l'échantillonnage, et à déterminer si les données sont suffisamment robustes pour être interprétées ou utilisées à des fins de cartographie.

Protocole 1 : Statistiques descriptives de base

Le protocole 1 constitue une première étape essentielle de l'évaluation de la qualité des données. Il génère des statistiques sommaires, identifie les valeurs inférieures aux limites de détection et produit des graphiques destinés à l'inspection visuelle. Ces résultats permettent de repérer d'éventuels problèmes avant de poursuivre avec les autres protocoles de QAQC.

Pour exécuter le protocole, il suffit de double-cliquer sur le fichier run_protocol1.bat. Cette action lance le script Python (protocol1.py) et traite la table d'entrée (input.csv). Le script génère un fichier CSV de sortie (protocol1_output.csv) contenant les statistiques sommaires (voir **tableau 3**), ainsi qu'un dossier regroupant les visualisations associées à chaque colonne de mesure.

La sortie statistique comprend des indicateurs clés tels que les mesures de tendance centrale (moyenne, médiane et mode), de variabilité (écart-type, variance, coefficient de variation et écart absolu médian [MAD]) et de distribution (étendue, ordres de grandeur de l'étendue, asymétrie et kurtosis ; tableau 3). Des valeurs de percentiles, incluant le minimum, le maximum et certains percentiles sélectionnés (2e, 5e, 10e, 90e, 95e et 98e), sont également calculées. Ensemble, ces statistiques fournissent des informations sur la distribution des données, leur tendance centrale et leur variabilité. Par exemple, la compréhension de l'étendue et de la variabilité des concentrations élémentaires aide à identifier des valeurs aberrantes ou des anomalies, tandis que les mesures de tendance centrale, telles que la moyenne et la médiane, révèlent les valeurs typiques observées dans l'ensemble de données. Des paramètres comme l'asymétrie et la kurtosis renseignent sur la forme et la symétrie de la distribution, tandis que la médiane et le MAD constituent des alternatives robustes à la moyenne et à l'écart-type, en particulier pour les ensembles de données contenant des valeurs aberrantes.

Le protocole traite également des enjeux liés à l'hétéroscédasticité. Si l'étendue d'une mesure dépasse 1,5 ordre de grandeur, cela peut indiquer une hétéroscédasticité et suggérer que les données bénéficieraient d'une transformation logarithmique lors d'analyses subséquentes. Ce seuil peut être modifié en ajustant la ligne "high_magnitude = row['order_of_magnitude'] > 1.5". De plus, la sortie indique le nombre de valeurs inférieures à la limite de détection ainsi que la proportion correspondante dans l'ensemble de données, puisqu'un nombre élevé de valeurs censurées (sous la limite de détection) peut compromettre la fiabilité des analyses ultérieures. Si plus de 12 % des valeurs sont inférieures à la limite de détection, la distribution devient asymétrique et n'est plus normale (Fletcher, 1981). Les variables dépassant ce seuil sont signalées par défaut, bien que cette valeur puisse être personnalisée dans le script à l'aide de la ligne "below_limit = row['percentage below detection limit'] < 12". Bien que les variables signalées puissent toujours être utilisées dans des analyses ultérieures, la prudence est de mise, car une forte variabilité ou un fort taux de censure peut nuire à la robustesse des interprétations, en particulier dans le cadre d'analyses subéquentes telles qu'une analyse en composantes principales (PCA).

Des diagrammes de dispersion sont générés pour chaque colonne de mesure, les points de données étant ordonnés selon la séquence de la table d'entrée (**fig. 1**). Les points sont codés par couleur selon le champ d'utilisté (purpose). Une ligne rouge indique la limite de détection, tandis que des lignes en pointillés représentent la moyenne et ±2 écarts-types. Ces éléments visuels peuvent être personnalisés dans le script sous la fonction generate_plots. Chaque graphique est exporté à la fois aux formats Portable Network Graphics (.png) et Scalable Vector Graphics (.svg).

Dans l'ensemble, les statistiques et les visualisations générées par le protocole 1 constituent une base solide pour l'évaluation de la qualité des données et appuient la mise en œuvre de procédures de QAQC plus avancées ainsi que d'analyses multivariées.

Tableau 3. Exemple de table de sortie générée par le protocole 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

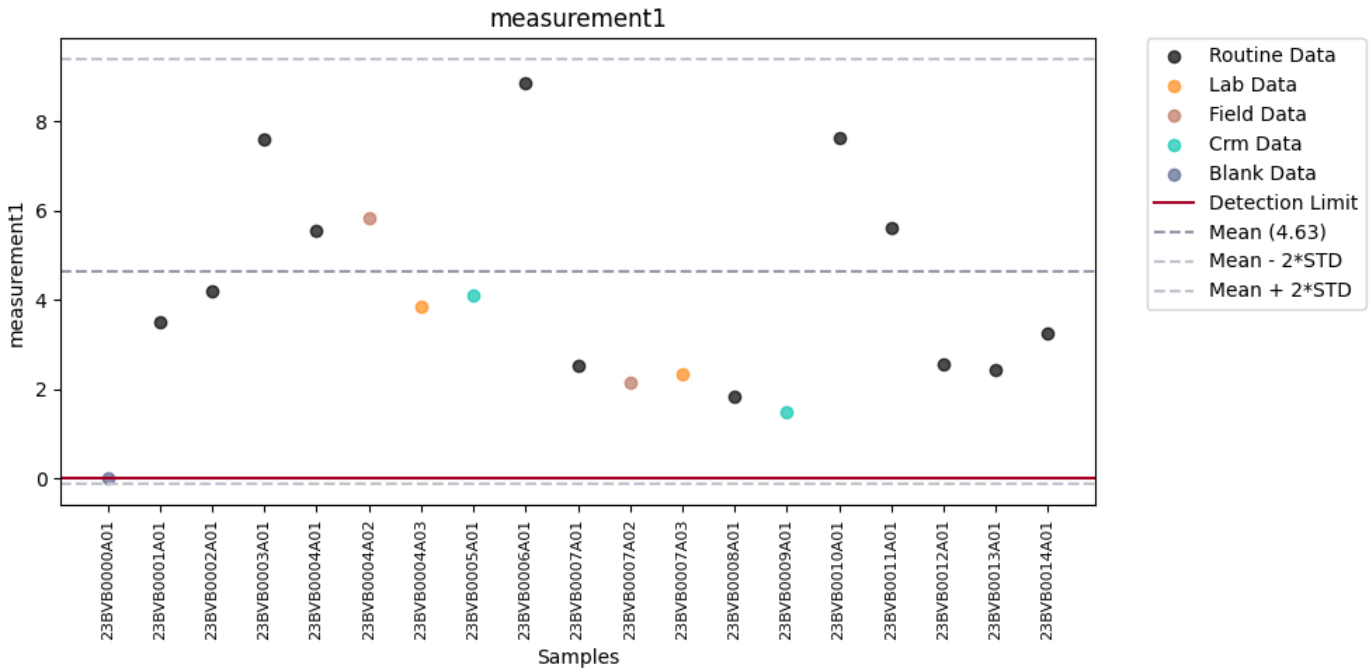

Figure 1. Exemple de diagramme de dispersion produit par le protocole 1 pour les données associées à measurement1 (tirées du tableau 2). Ce diagramme permet une inspection visuelle des tendances, des valeurs aberrantes et de la variabilité globale le long de la séquence d'échantillons.

Protocole 2 : Analyse de l'exactitude analytique

Le protocole 2 est conçu pour comparer les valeurs mesurées aux valeurs attendues pour les matériaux de référence (blancs et CRMs). Ce protocole permet d'évaluer l'exactitude analytique en déterminant si les valeurs mesurées se situent dans une plage acceptable. Afin d'atténuer les problèmes potentiels liés à l'hétéroscédasticité, il est fortement recommandé d'appliquer une transformation logarithmique à toute mesure présentant une étendue supérieure à 1,5 ordre de grandeur, telle qu'identifiée par le protocole 1, avant de procéder au protocole 2.

Le script est exécuté en double-cliquant sur le fichier run protocol2.bat, qui lance le script Python protocol2.py. Lors de l'exécution, une fenêtre de commandes s'ouvre et invite l'utilisateur à sélectionner le groupe à évaluer (p. ex., CRM1, CRM2, blank, Till-1, Till-4). L'utilisateur sélectionne le groupe en entrant le numéro correspondant. Le protocole peut évaluer un seul groupe ou l'ensemble des groupes présents dans l'ensemble de données. Par exemple, si l'ensemble de données comprend des échantillons CRM1, CRM2 et blank, le script peut les analyser simultanément. À noter que, dans l'ensemble de données d'exemple (tableau 2), seul CRM1 contient à la fois des valeurs expected et des valeurs mesurées. Par conséquent, lorsque tous les groupes sont sélectionnés, une sortie est générée uniquement pour CRM1.

Si le groupe sélectionné ne contient qu'une seule valeur mesurée, le script vérifie si chaque mesure se situe dans une plage de tolérance acceptable. Lorsque des incertitudes publiées sont disponibles, celles-ci sont utilisées pour définir cette plage. À défaut, une tolérance par défaut de ±10 % est appliquée. Ce seuil peut être modifié manuellement en éditant le script à l'aide de Notepad++, dans la section # Check if the single measurement is within 10% of the expected value (or uncertainty if available).

Si le groupe contient plusieurs mesures, comme les deux mesures associées à CRM1 dans le tableau 2, le script effectue un test t à un échantillon afin de comparer la moyenne du groupe à la valeur attendue. Pour les groupes comportant plus de 30 mesures, l'écart-type relatif (RSD%) est également calculé afin d'évaluer la cohérence interne. Le RSD% permet d'estimer la précision pour chaque matériau de référence et appuie l'évaluation de l'exactitude, puisqu'une faible précision peut masquer des biais analytiques (p. ex., Abzalov, 2011). Toutefois, l'exactitude doit être évaluée principalement à l'aide de la valeur p, qui fournit un seuil statistique pour l'interprétation. Par défaut, un seuil de valeur p de 0,05 (niveau de confiance de 95 %) est utilisé pour déterminer si l'exactitude est acceptable. Ce seuil peut être modifié dans le script, dans la section # Interpretation of p-value. En complément de l'évaluation basée sur la valeur p, le script vérifie si les mesures se situent à l'intérieur de ±2 écarts-types par rapport à la moyenne (McCurdy et Garrett, 2016). Si la valeur p dépasse le seuil défini ou si plusieurs mesures se situent en dehors de la plage acceptable, la mesure est signalée comme présentant une faible exactitude. Cela n'exclut pas automatiquement les données d'une utilisation ultérieure, mais indique qu'un examen plus approfondi est recommandé avant leur intégration dans des analyses subséquentes.

Le fichier de sortie fournit un résumé de plusieurs statistiques pour chaque mesure analysée (**tableau 4**). Celles-ci comprennent la valeur attendue (expected; utilisée comme valeur de référence), l'incertitude(uncertaintyl; si disponible), le nombre de mesures incluses dans l'évaluation, la moyenne et l'écart-type, l'écart-type relatif et la valeur p. Une ligne d'interprétation indique si la mesure présente une exactitude acceptable ou faible selon le seuil défini. De plus, la sortie comprend un message indiquant le nombre de mesures situées hors de la plage de tolérance attendue (définie soit par l'incertitude établie, soit par ±10 %), ainsi que le décompte de ces valeurs.

Le protocole génère également des diagrammes de contrôle de Shewhart (**fig. 2**), qui offrent une représentation visuelle de l'exactitude des mesures. Ces graphiques permettent d'identifier des problèmes analytiques tels qu'une dérive ou un biais et fournissent un aperçu rapide de la performance analytique associée aux matériaux de référence (p. ex., Piercey, 2014 ; Garrett et al., 2017).

Tableau 4. Exemple de sortie générée par le protocole 2 montrant les résultats de l'évaluation de l'exactitude pour quatre mesures. Le tableau comprend des statistiques telles que la valeur attendue (*ang., expected*), la moyenne des échantillons (*ang., mean*), l'écart-type (*std_dev*), l'écart-type relatif (RSD), la valeur p et l'interprétation.

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

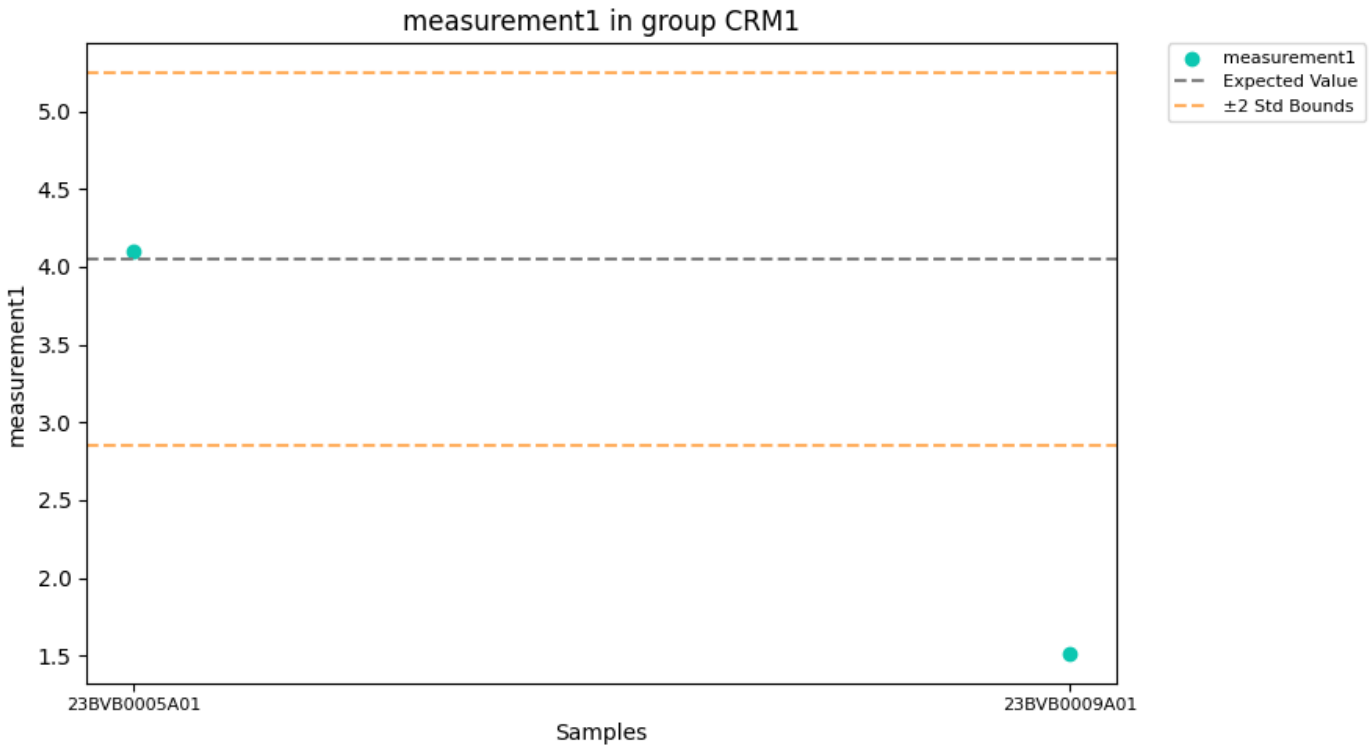

Figure 2. Exemple de diagramme de dispersion produit par le protocole 2 pour CRM1 et measurement1 (tirés du tableau 2). Ce graphique montre visuellement que l'échantillon 23BVB0009A01 se situe nettement en dehors de la plage attendue (±2 écarts-types), ce qui indique la nécessité de procéder à des investigations supplémentaires en matière de contrôle de la qualité.

Protocole 3 : Analyse de la précision analytique

Le protocole 3 est conçu pour évaluer la précision des données de mesure en comparant les échantillons de routine à leurs duplicatas analytiques, qu'il s'agisse de sous-échantillons issus d'un même matériau ou de fractions envoyées à différents laboratoires analytiques. Il permet d'identifier et de signaler des problèmes de précision susceptibles de compromettre la fiabilité des ensembles de données compositionnelles. Avant d'exécuter le protocole, il convient d'évaluer si les données présentent une hétéroscédasticité, en particulier pour les mesures couvrant plus de 1,5 ordre de grandeur (tel qu'identifié par le protocole 1). Dans de tels cas, une transformation logarithmique est fortement recommandée afin de stabiliser la variance et d'assurer la validité de l'analyse de la précision.

Le script est exécuté en double-cliquant sur le fichier run_protocol3.bat, qui lance le script Python protocol3.py. Le protocole commence par lire les données à partir du fichier CSV d'entrée et par filtrer l'ensemble de données afin de ne conserver que les entrées valides contenant à la fois des échantillons de routine et des duplicatas analytiques. Toute ligne présentant des valeurs manquantes pour la mesure évaluée est exclue afin de garantir que l'ensemble de données est complet pour l'analyse statistique.

Pour chaque mesure, une analyse de la variance (ANOVA) est effectuée afin de comparer les moyennes des échantillons de routine à celles de leurs duplicatas analytiques correspondants, ou entre différents laboratoires. L'ANOVA produit une statistique F et une valeur p permettant d'évaluer si les moyennes diffèrent de manière significative. Une valeur p inférieure à 0,05 indique une forte évidence contre l'hypothèse nulle, suggérant une différence significative entre les mesures de routine et les duplicatas. Une valeur p supérieure à 0,05 indique qu'aucune différence significative ne peut être conclue.

En complément de l'ANOVA, le protocole calcule des statistiques descriptives pour chaque mesure, notamment la moyenne, la médiane, l'écart-type et la variance, calculées séparément pour les échantillons de routine et les duplicatas. Le coefficient de variation (CV), défini comme le rapport entre l'écart-type et la moyenne, est utilisé comme indicateur clé de la précision des mesures. Un CV faible indique une forte précision, tandis qu'un CV élevé traduit une variabilité accrue.

Un seuil de 20 % d'écart-type relatif (RSD = 0,2) est utilisé pour évaluer la précision. Si le RSD dépasse ce seuil, la mesure est signalée comme présentant une faible précision. Ce seuil repose sur l'hypothèse que, en moyenne, 95 valeurs sur 100 devraient se situer à l'intérieur de deux écarts-types autour de la moyenne (Fletcher, 1981). Des valeurs élevées de RSD, en particulier lorsqu'elles sont associées à une plage de valeurs restreinte, peuvent masquer des tendances ou des anomalies significatives (Howarth et Martin, 1979). Dans de tels cas, le protocole alerte l'utilisateur de potentiels enjeux liés à la qualité des données.

Les résultats sont enregistrés dans un fichier CSV de sortie et comprennent : le nombre de groupes uniques conservés et retirés après le filtrage, les valeurs moyennes et les écarts-types pour les échantillons de routine et les duplicatas, le RSD calculé et une interprétation de la précision. La sortie rapporte également la variabilité intra- et inter-groupes, ainsi que les résultats de l'ANOVA pour les comparaisons selon la finalité des échantillons et selon les groupes (**tableau 5**). Les utilisateurs peuvent personnaliser les seuils de RSD et de valeur p en modifiant les sections correspondantes du code (c.-à-d. # Add interpretation based on RSD et # Check if there are at least two groups present).

Tableau 5. Exemple de sortie générée par le protocole 3 présentant les résultats de l'analyse de la précision pour les duplicatas analytiques.

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

Protocole 4 : Analyse de la variabilité liée à l'échantillonnage sur le terrain

Le protocole 4 évalue la variabilité liée à l'échantillonnage sur le terrain en comparant les échantillons de routine à leurs duplicatas de terrain au moyen d'une analyse de la variance (ANOVA). L'objectif principal est de décomposer la variabilité observée en composantes inter-groupes (entre les sites) et intra-groupe (au site). Un objectif clé consiste à déterminer si la variabilité intra-groupe (ou au site) — qui reflète les différences entre des échantillons prélevés au même site — est significativement inférieure à la variabilité inter-groupes. Dans un tel cas, cela indique que les signatures compositionnelles propres aux groupes dominent les erreurs aléatoires d'échantillonnage ou le bruit analytique.

Afin d'assurer la validité de l'ANOVA, les mesures couvrant plus de 1,5 ordre de grandeur doivent être transformées logarithmiquement avant l'analyse, tel que décrit dans le protocole 1. Cette transformation permet de réduire l'impact de l'hétéroscédasticité et de satisfaire l'hypothèse d'homogénéité des variances requise pour l'ANOVA.

Le script est exécuté en double-cliquant sur le fichier run_protocol4.bat, qui lance le script Python protocol4.py. La sortie du protocole comprend plusieurs statistiques clés (**tableau 6**), notamment le nombre de groupes uniques (après exclusion des échantillons incomplets ou inférieurs à la limite de détection), les moyennes et les écarts-types des échantillons de routine et des duplicatas, la variance intra-groupe, l'écart-type relatif, la variabilité intra- et inter-groupes, la statistique F et la valeur p associée. Ces paramètres permettent de déterminer si la variabilité observée reflète des patrons géologiques réels ou une erreur aléatoire. Un seuil couramment utilisé pour l'interprétation des résultats est un rapport F (variance inter-groupes divisée par la variance intra-groupe) supérieur à 4,0 à un niveau de confiance de 95 %, indiquant une différence statistiquement significative. Le dépassement de ce seuil suggère que les erreurs analytiques et d'échantillonnage sont faibles par rapport à la variabilité régionale, ce qui appuie l'utilisation des données à des fins de cartographie et d'analyses statistiques (McCurdy et Garrett, 2016). Même lorsque le rapport F est inférieur au seuil de 4:1, les résultats peuvent néanmoins mettre en évidence des valeurs aberrantes ou des groupes anormaux potentiellement liés à une minéralisation ou à d'autres caractéristiques géologiques. Dans ces cas, les valeurs p constituent un outil supplémentaire pour signaler des mesures nécessitant une investigation plus approfondie. Bien que le seuil de 4:1 ait été initialement établi pour des levés d'échantillons d'eau de lacs comportant de grands ensembles de données, il peut être ajusté pour des ensembles de données plus restreints en modifiant le code dans Notepad++, dans la section intitulée # Check if variability between is at least 4 times larger than the variability within. Cette flexibilité permet d'adapter le protocole à différents types de levés et à diverses densités d'échantillonnage.

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

Protocole 5: Analyse intégrée de la variabilité analytique et de l'échantillonnage

Le protocole 5 vise à quantifier les contributions relatives des différentes sources de variabilité en comparant les échantillons de routine à leurs duplicatas de terrain et analytiques. En analysant ces groupes séparément, le protocole permet de distinguer les sources de variabilité liées à l'échantillonnage (terrain) de celles liées à l'analyse en laboratoire (Garrett, 2013b). Lorsque les duplicatas analytiques sont systématiquement prélevés à partir des duplicatas de terrain, ou lorsque les échantillons de routine sont associés à leurs duplicatas de terrain respectifs, il devient possible de distinguer clairement ces différentes sources de variabilité. Afin d'éviter les problèmes potentiels liés à l'hétéroscédasticité — en particulier pour les colonnes de mesures dont l'étendue dépasse 1,5 ordre de grandeur, tel qu'indiqué par le protocole 1 — il est recommandé d'appliquer une transformation logarithmique aux données avant de procéder à ce protocole.

Le script est exécuté en double-cliquant sur le fichier run_protocol5.bat, qui lance le script Python protocol5.py. Le protocole génère un fichier de sortie contenant plusieurs statistiques (**Tableau 7**), notamment le nombre de groupes uniques après filtrage, le nombre de groupes exclus en raison de données manquantes, ainsi que des statistiques descriptives pour l'ensemble des données et pour les échantillons de routine, les duplicatas de terrain et les duplicatas analytiques (moyenne, médiane et écart-type). La sortie inclut également la variance et l'écart-type intra-groupe, l'écart-type relatif, ainsi que les contributions respectives de la variabilité inter-groupes et intra-groupe à la variabilité totale. La statistique F, la valeur p et une interprétation de la valeur p sont également fournies.

Le protocole réalise une analyse de la variance (ANOVA) afin de comparer les moyennes des échantillons de routine et des duplicatas (de terrain et analytiques). Ce test permet de vérifier si les moyennes des groupes diffèrent de manière significative, selon un seuil standard de $p < 0,05$. Si un seul groupe est disponible (c.-à-d., un seul ensemble d'échantillons de routine et de duplicatas), le script compare les moyennes et les écarts-types des échantillons de routine et des duplicatas afin de déterminer si les valeurs se situent dans une plage attendue. Si toutes les valeurs sont comprises dans cette plage, la sortie indiquera que les moyennes ne sont pas différentes statistiquement ("The means are not statistically different"). Si une ou plusieurs valeurs se situent en dehors de la plage attendue, la ligne d'interprétation indiquera une avertissement qu'une des moyennes est différente ("Caution: one of the means is statistically different"). Si les données sont insuffisantes, la ligne d'interprétation indiquera : "Not enough data for interpretation".

En complément de la comparaison des moyennes, un rapport de variabilité est calculé afin de déterminer la proportion de la variabilité totale attribuable à la variabilité intra-groupe par rapport à la variabilité inter-groupes. Ce rapport compare la variabilité entre les groupes à la variabilité à l'intérieur des groupes et, s'il dépasse 4,0 à un niveau de confiance de 95 %, cela indique que les erreurs liées à l'échantillonnage et à l'analyse sont nettement inférieures à la variabilité régionale. Le dépassement de ce seuil suggère que les données présentent une forte cohérence intra-groupe et qu'elles sont adaptées à des fins de cartographie et d'analyses statistiques. Le rapport de variabilité peut être ajusté au besoin en modifiant le code dans Notepad++, plus précisément la valeur définie dans la section "# Variability interpretation".

Tableau 7. Exemple de résultats générés par le protocole 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

La mise en œuvre de ces protocoles simplifiés de contrôle de la qualité et de validation des données constitue une avancée significative pour assurer l'exactitude, la fiabilité et l'utilité des données compositionnelles dérivées des tills et d'autres matériaux géologiques. Conçus pour compléter les méthodologies de QAQC existantes développées en R à la CGC (Garrett et Chen, 2007 ; Garrett, 2013a, b ; McCurdy et Garrett, 2016), ces protocoles offrent des avantages supplémentaires, notamment en termes de simplicité, d'accessibilité et de clarté des lignes directrices pour l'interprétation.

Les améliorations futures devraient viser l'élargissement du soutien aux utilisateurs, par exemple par l'intégration de représentations graphiques supplémentaires et la mise en œuvre d'outils statistiques plus avancés. Le perfectionnement continu de ces protocoles contribuera à mieux répondre aux besoins des chercheurs et à renforcer la robustesse de l'analyse des données compositionnelles dans un large éventail d'applications géoscientifiques.

Remerciements

Ce rapport constitue une contribution au programme GEM-GeoNord. Nous remercions tout particulièrement Jessey M. Rice pour avoir testé les premières versions des codes. Nous exprimons également notre reconnaissance à Marty McCurdy et Bob Garrett pour leur révision approfondie du manuscrit et des codes. ChatGPT a été utilisé pour appuyer le développement du code.

Références

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.