AN EVALUATION OF STATISTICAL TECHNIQUES FOR THE INTERCOMPARISON OF TWO INSTRUMENTAL METHODS FOR A NATIONAL AIR QUALITY MONITORING NETWORK

Colleen Marciel F. Rosales[a,b], Nicholas E. Brunk[c], Nicholas J. Spada[a], Warren White[a], Kyoungmi Kim[d], Nicole P. Hyslop[a]

(a) Air Quality Research Center, University of California Davis, (b) OpenAQ, (c) American Regent, Inc.

(d) Department of Public Health Sciences and Environmental Health Science Center, University of California Davis

INTRODUCTION

- **Regression** (*expressing one variable in terms of another, and can be used for prediction*) is often confused with **correlation** (*degree to which two variables vary together*).
- Most comparisons of two instrumental techniques use ordinary least squares (OLS) regression—simple, but often ***incorrect***.
  - In OLS, one variable needs to be "independent" and the other "dependent".
- **Concordance analysis** and **measures of association (correlation)** are more robust metrics of agreement for comparing two instrumental techniques.

METHODOLOGY

A. ELEMENTAL ANALYSIS

Complete details available at: DOI: 10.1080/10962247.2023.2247376

***N* = 594 samples** from US EPA's Chemical Speciation Network (CSN) → **UCD XRF** X-Ray Fluorescence Spectroscopy at UC Davis → **ICPMS** Inductively Coupled Plasma-Mass Spectrometry at RTI International

29 elements analyzed in common; **13** out of **29** elements explored.
Na, Al, K, Ti, Fe, Cu, Zn, Pb, Cr, Ni, Mg, Mn, Sr -only select elements are shown here

B. STATISTICAL ANALYSIS- per element

Performed using Wolfram Mathematica 13.0

Concordance Correlation Coefficient (CCC)

Evaluates the degree to which measurement pairs fall on the 45° line.

$$\frac{2\,\text{Covariance}[x,\,y]}{(\text{Mean}[x]-\text{Mean}[y])^2+\text{StandardDeviation}[x]^2+\text{StandardDeviation}[y]^2}$$

-1: negative agreement
0: no agreement
1: positive agreement; >0.7 is strong

Kendall's Tau (τ)

A nonparametric measure of association based on the relative order of consecutive elements in two lists.

$$\frac{n_c-n_d}{\sqrt{(n_c+n_d+n_x)\,(n_c+n_d+n_y)}}$$

n_c # concordant pairs
n_d # discordant pairs
n_x or n_y # ties involving only x or y

-1: complete disagreement (discordant)
0: number of concordant & discordant pairs are equal
+1: complete agreement (concordant); >0.35 is strong.

Run *Cluster Analysis;* return N number of clusters

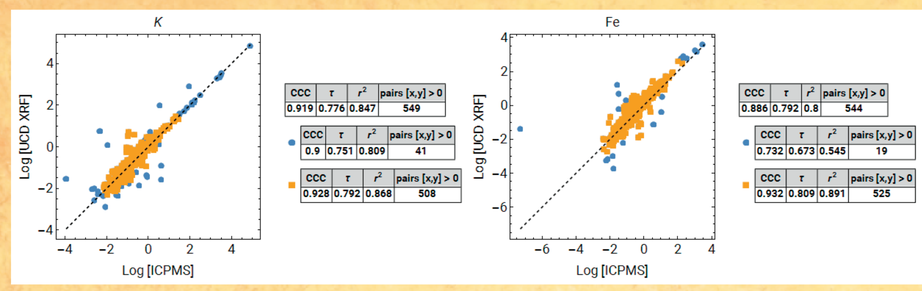
↓

For each cluster, calculate statistical metrics: CCC, τ, r^2

DISCUSSION

- ***Cluster analysis*** may provide insight on the natural trends or clusters that occur and can help differentiate noise from real measurements.
- The current dataset has underlying variance due to sampling differences (e.g., geographical, sampling mechanism)--thus, cluster analysis ***cannot*** be the only test for separating noise from real measurements.
  - Looking at collocated sites may remove underlying factors that influence clustering.
- Investigate metrics **(CCC, τ, r^2)** for whole set, and for individual clusters
  - → Do all metrics show strong agreement?
  - → What if only 2/3 metrics show strong agreement?
  - → With respect to instrumental conditions and other experimental factors, do the clusters make sense?

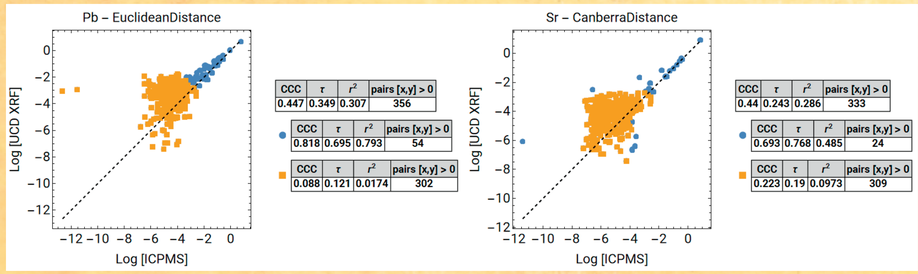
CASE 1 Strong agreement



For K, all metrics agree

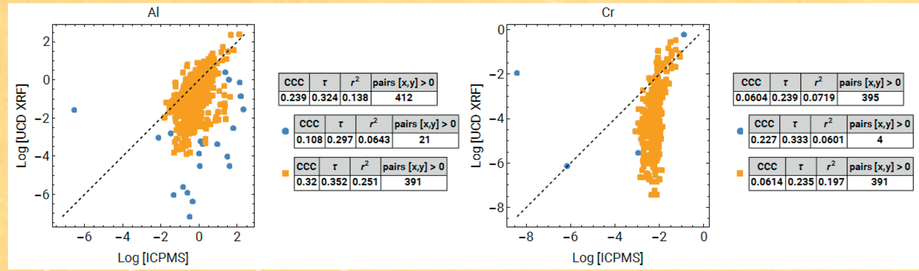
For Fe, 2/3 metrics agree (per cluster)

CASE 2 Less agreement due to Instrumental noise



For both Pb and Sr, there is obviously a "noise" cluster

CASE 3 Less agreement due to other experimental factors, e.g. extraction



For Al, one cluster passed 1/3 metrics

For Cr, all metrics failed

CONCLUSIONS

- Concordance analysis and non-parametric measures of association can be included with regression analysis when comparing agreements of two methods.
- Use multiple statistical metrics for association and do not just rely on one.
- For real world PM data, log scale is more insightful because of the complexity of environmental data and uncertainties at lower concentrations.
- It is paramount to understand the measurement process and interpret the data according to the real-world sampling and experimental process.

REFERENCES

- Sokal, R., Rohlf, F.J. (1995) *Chapter 14. Linear Regression* and *Chapter 15: Correlation.* Biometry, W.H. Freeman and Company, New York.
- Lin (1989). *A Concordance Correlation Coefficient to Evaluate Reproducibility.* Biometrics.

cfrosales@ucdavis.edu colleen.phd