# Statement of problem

Antonio Paez

2023-06-02

## Background

Compositional variables are [define and use quick example].

It has long been known that calculating correlations between compositional variables is prone to error. In brief, spurious correlations are caused by the negative covariances that necessarily emerge when the variables sum to a constant (Pearson 1896; Chayes 1971).

Vera Pawlosky (Pawlowsky 1984) discussed a related problem that emerges when compositional variables are regionalized, in other words, when they are collected in space and for whom a georeference was recorded. Succintly, this is the situation that Pawlosky dealt with:

$$z(x) = [z_1(x), \cdots, z_D(x)]$$

where:

$$\sum_d^D z_d(x) = k$$

with $k$ a given constant. What this means is that a variable that is compositional is georeferenced. For example, components $P$, $Q$, $R$ whose sum is 100% of the composition *at* $x$ are measured, and measurements are made in multiple locations $x_s$, $s = 1, \cdots, S$. The sum of the components over locations is not assumed to be a constant in a meaningful way. Pawlowsky calls this a *regionalized composition*; we might call this a *locally compositional process* to distinguish it from the case that we discuss next. Pawlowsky shows how calculating the autocovariance (spatial covariance) of this locally compositional process $z(x)$ is prone to error and bound to produce spurious autocovariances (Pawlowsky 1984, 109).

A different problem of interest can be stated as follows:

$$z = [z(x_1), \cdots, z(x_D)]$$

In other words, only one variable is observed at each location, but the variable is compositional *over space*, that is:

$$\sum_d^D z(x_d) = 1$$

This is a process that is globally compositional over the region of interest. [Give examples].

## Statement of problem

We don't know whether the autocovariance of a globally compositional process is also affected.

## References

Chayes, Felix. 1971. *Ratio Correlation: A Manual for Students of Petrology and Geochemistry.* Book. University of Chicago Press.

Pawlowsky, V. 1984. "On Spurious Spatial Covariance Between Variables of Constant Sum." Journal Article. *Sciences de La Terre. Informatique Géologique*, no. 21: 107–13.

Pearson, K. 1896. "On a Form of Spurious Correlation Which May Arise When Indices Are Used, Etc." Journal Article. *Proc. Roy. Soc* 9: 489–98.