Reply to Michael (10-rejoinder …)

Michael's position reminds me of the never-ending discussion about the appropriateness of correlation between compositional parts. Sometimes correlation reflects real proportionality, and then people ask, why bother? The problem is you do not know when this is the case. Therefore, you have to do the analysis twice, one for correlation, and one for proportionality, something that makes little sense.

The same happens with the scatterplots using ilr or amalgamation representations. In some cases they will look similar, but you do not know when this is the case. Maybe it is possible to find criteria (a rule of thumb) to decide when this will be the case? Personally, I think it is not worthwhile, but maybe I am wrong.

Nevertheless, the essential thing behind the problem is interpretation. If you use amalgamation, you are interested in the absolute information carried by the data. If you use ilr coordinates (balances, clr-principal components), you are interested in the relative information carried by the data.

The Aitchison geometry of the sample space of compositional data is directed towards extracting relative, not absolute information, and ilr coordinates are coherent with this structure. The ilr coordinates satisfy the main principles of CoDa, i.e. scale invariance and subcompositional coherence, and this is something frequently fails using amalgamations.

Concerning the statement about interpretation of balances, saying that amalgamation is easy to interpret, while geometric means are not, I think the paper "The shock of the mean" by Simon Raper in the December 2017 issue of Significance, is a good reading in this respect.

But note that it is possible to look at the problem from a different perspective: amalgamation is a projection in the sample space of real random variables, while geometric means are a projection in the sample space of compositional parts. Amalgamation (sum) and perturbation (product) are the group operations in the respective sample spaces; they play mathematically the same role. However, I think this is much harder to explain to practitioners, although it should not be hard to understand for those with a solid mathematical background.

To see why this is so from an intuitive perspective, take a compositional data matrix X, which rows are samples and columns are parts, and transpose it. The data in the cells of X^T satisfy the same assumptions as before. If the sample space of the initial matrix X, with D columns (parts) and N rows (samples) is represented by the D-part simplex, the transposed matrix X^T, with N columns (samples) and D rows (parts) can be represented in the N-part simplex. The Aitchison geometry of this N-part simplex explains why it is a projection.

There is also a formal proof, but that would take too long to write down here. Hope this helps to clarify the claim for the use of the ilr: it is simply the safe approach.