

Asunto: RE: FOREGS Stream water data
De: "Ron S. Kenett" <ron@kpa-group.com>
Fecha: 22/06/2017 21:16
Para:

Vera

First of all, let me congratulate you on addressing the discussion in such a detailed and coherent way.

Please let me reiterate two points I made on various occasions at CoDaWork17:

1. Amalgamation: I looked into this aspect in the CoDa context. It does indeed seem different from data fusion and other data integration aspects I was involved with. The practical issue I have looked at, was driven from the need to combine sensor or survey data to provide a unified decision support model. This topic is dealt with in different contexts such as small area estimation (SAE), where external calibration is used to improve small area estimates, or improving temporal relevance (more on this is presented in my book with Galit Shmueli on information quality). The challenges relate to the combination of various data structures that are generated with different timing cycles and have different forms such as semantic data and physical measurements. My approach is to look at this from a practical perspective in order to derive practical solutions. Mathematical coherence is what you seem focused on and is definitely a complementary aspect of the discussion.
2. "Practice before theory": This quote is from Terry Speed who is one of the leading omics statisticians. Terry is focused on developing methods that help derive biological insights. His approach starts with the data and the application context and he is using mathematics for support. For more on his work see <https://www.stat.berkeley.edu/~terry/>. I had some interesting discussions with him about CoDa when he visited Israel a month ago. Part of his response was: "Of course I know Aitchison's work on compositional data. I haven't used it in my genomics work so far, but the issues are not far beneath the surface. The work by my former student Mark Robinson addresses the problem in RNAseq that when you have a lot of one or a few things, you necessarily have less of everything else, if the total is more or less constant. But it is rarely as simple as this, and so more needs to be done to normalize (and remove batch effects from) RNAseq data. In other words, recognizing it as compositional data doesn't take you too far." It seems that CoDa methods did not really impact his work.

In my biased view, the CoDa community needs to better integrate its contributions with aspects such as those mentioned above. The potential is endless, and this is what got me interested in CoDa. As you mentioned, the range of applications presented in CoDaWork17 was impressive. I hope this potential eventually materializes. This requires addressing data driven problems and providing added value, i.e. insights in the terms of the domain experts.

Thank you again for sharing your thoughts.

Bst

ron

.....
Prof Ron Kenett
ron@kpa-group.com
www.kpa-group.com
www.amazon.com/author/rkenett
Mob+972522434491
.....