Calculating Uncertainties

Stanford Hooker
NASA/GSFC/CVO
Halethorpe, Maryland

The Premise and Utility of Round-Robin Intercomparisons

The premise of a round robin is *all participants use a validated method, which are equally capable of estimating a true result for each "sample," and each sample is analyzed no differently than any other normally analyzed by the method*.

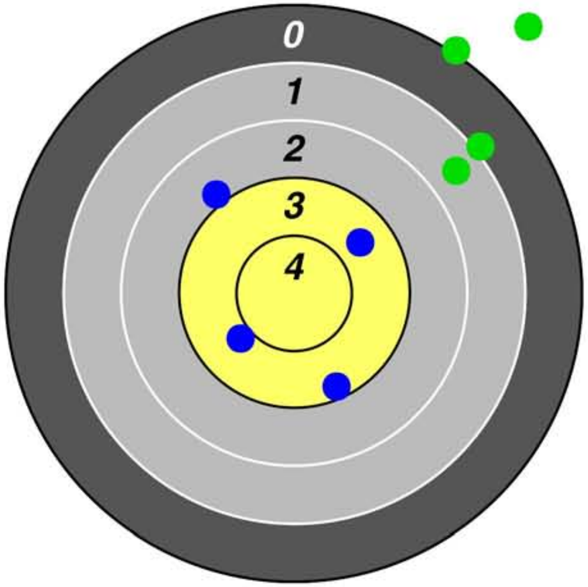
The result from each method is expected to be close to the truth (which is usually unknown for field samples), and the dispersion of the results is equally expressed above and below the true value. *A validated method has no inherent biases*, because if one existed it would have been removed by the validation process. *The accuracy (or uncertainty) for each method is estimated by computing the difference of each result from the truth (usually the average of all data) for each product*.

Accuracy estimates how close the result is to the true value while *precision* is an estimate of how exactly the result is determined independently of any true value.

Accuracy is telling a story truthfully, and precision is how similarly the story is repeated over and over again.

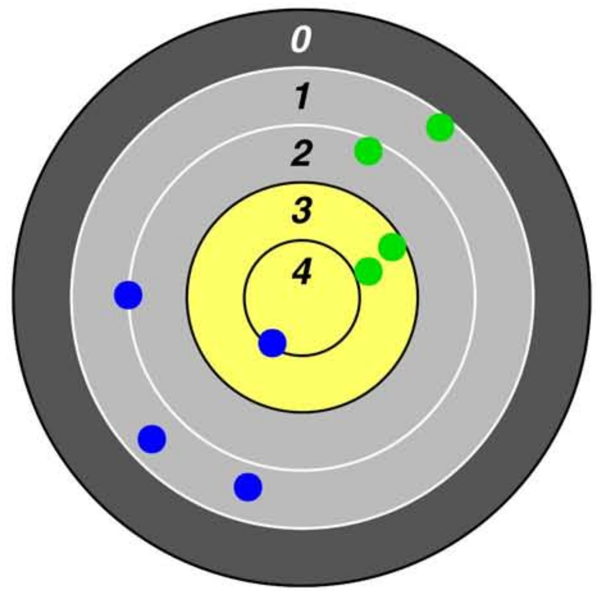
Examples of round-robin inquiries for ocean color include the SeaWiFS Intercalibration Round-Robin Experiment (SIRREX), which investigated optical calibrations, and the SeaWiFS Data Analysis Round Robin (DARR), which looked at data products from measurements of the apparent optical properties (AOPs) of seawater.

Computing Uncertainties in a Round-Robin Intercomparison Involving Unknowns



Consider two groups, composed of four archers each, all competing for placement on an archery team. With respect to the known bull's-eye (left graphic), the green archers are the least accurate — *in fact two of them missed the target completely* — although, as a group, their precision is better than the blue archers. If a minimum score of 1 is needed to make the archery team, all of the blue archers, but only one green archer, will qualify.

Imagine that the archers are not evaluated with respect to the known bull's-eye (or the average of the blue archers, which might have been a pre-established qualified set of archers), but are instead evaluated with respect to the average of all the shots (right graphic). In this case, *all of the archers qualify for the team*, and the green archers are slightly more accurate as a group than the blue archers. The latter is a recurring consequence of data clusters having bad accuracy, but good precision.



Estimating the Uncertainty for a Field Sample

The estimation of uncertainties begins with computing the average concentration for each of the sets of replicates (left), and the average of each sample for the quality-assured (QA) subset of laboratories provides the proxy for truth (right):

$$\bar{C}_{P_i}^{L_j}(S_k) = \frac{1}{N_R} \sum_{l=1}^{N_R} C_{P_i}^{L_j}(S_{k,l}) \qquad \bar{C}_{P_i}^{A}(S_k) = \frac{1}{N_L} \sum_{j=1}^{N_L} \bar{C}_{P_i}^{L_j}(S_k)$$

The uncertainty for each sample is based on the difference between the laboratory sample and the average from the QA subset (left), the absolute value of which is used to ensure the variance of the differences is not artificially reduced (right):

$$\psi_{P_i}^{L_j}(S_k) = 100 \, \frac{C_{P_i}^{L_j}(S_k) - \bar{C}_{P_i}^{A}(S_k)}{\bar{C}_{P_i}^{A}(S_k)} \qquad |\bar{\psi}|_{P_i}^{L_j} = \frac{1}{N_S} \sum_{k=1}^{N_S} \left|\psi_{P_i}^{L_j}(S_k)\right|$$

Averages of the number of laboratories in each of the identified subsets establish overall performance:

$$|\bar{\psi}|_{P_i}^{A} = \frac{1}{N_L} \sum_{j=1}^{N_L} |\bar{\psi}|_{P_i}^{L_j}$$