How to Avoid Errors in Error Propagation: Monte Carlo and Addition in Quadrature can give Right or Wrong Answers when Properly or Improperly Applied

Ge (Jeff) Pu[1], John L. Campbell[2], John E. Drake[1], Mark B. Green[2], Craig Wayson[3], and Ruth D. Yanai[1]

[1]SUNY College of Environmental Science and Forestry, Syracuse, NY; [2]USDA Forest Service, Durham, NH, [3]USDA Forest Service, International Programs, Washington, DC, USA

Introduction

In ecosystem studies and in country-level carbon accounting, where using replication to establish confidence is not an option, other approaches are needed for quantifying uncertainty. Options for propagating uncertainty include **Addition in Quadrature** and **Monte Carlo Simulation**. However, it is easy to make mistakes, and incorrect implementation has been advised, for example to reduce reported **Uncertainty in Country-Level Carbon Accounting**, which affects the payments made to countries for reducing emissions from deforestation and forest degradation. It is easy for even well educated researchers to make mistakes that inflate or reduce errors. We share errors we have made involving **Independence of Errors in Monte Carlo Simulation**.

Uncertainty in Country-Level Carbon Accounting

Deforestation and forest degradation are important sources of net carbon emissions to the atmosphere. Countries seeking payments for emissions reductions must report uncertainty in their estimates, with high uncertainties resulting in up to 15% reductions in payments. A review of submissions of forest reference levels submitted to the United Nations Framework Convention on Climate Change (37 countries) and the Forest Carbon Partnership Facility (18 countries) reveals multiple mistakes in error propagation. The most egregious is the practice of treating Monte Carlo iterations as samples, and dividing by the number of iterations, sometimes resulting in combined uncertainties smaller than any of the input uncertainties (the lowest combined uncertainty reported by a country was 1.3%). Guidelines should be more specific as to how to obtain uncertainties from Monte Carlo simulations, and tools should be developed to support efficient computation and proper application of error terms. Investments in emission reductions should be made with known confidence, correctly estimated, even if uncertainties are high.

Addition in Quadrature

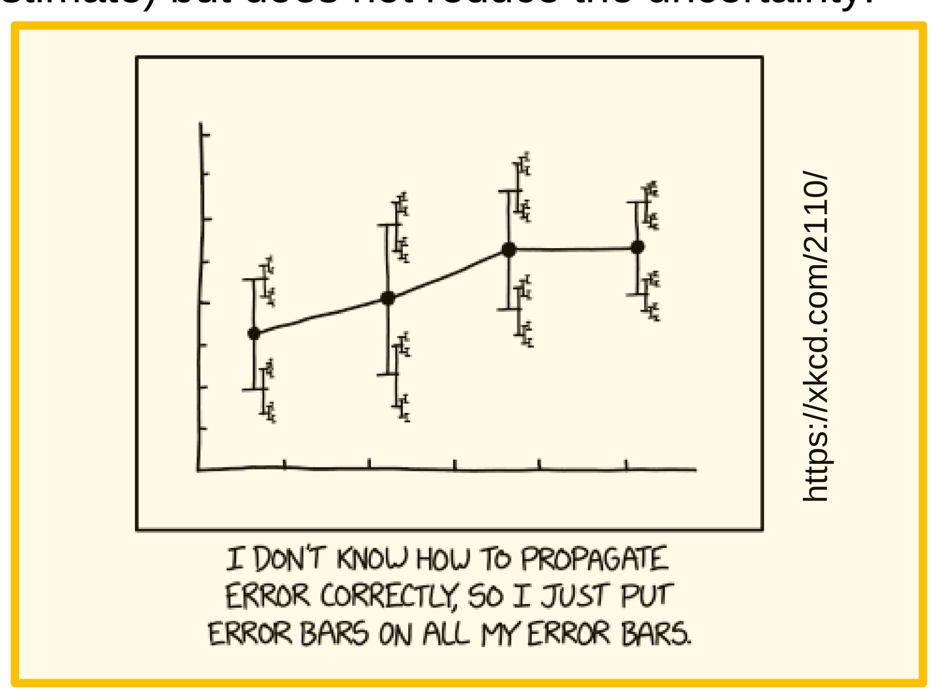
One approach to error propagation makes use of the mathematical rule that the variance of a sum is the sum of the variances: squaring the SDs, adding them together, and taking the square root gives the SD of the sum, if the errors are independent and normally distributed. This approach is referred to as Approach 1 by the IPCC guidelines (Eggleston, 2006).

You can verify this rule using Monte Carlo simulation. Randomly choose a number with a mean of 0 and a standard deviation of 3, and add it to a random number with a mean of 0 and a SD of 4, and record the sum. Do this a few times. The average answer should be close to 0. The SD of the estimates should be close to 5. The more estimates you have, the closer the answer will be to 5.

Summing in Quadrature verified by Monte Carlo Simulation: $3^2 + 4^2 = 5^2$

Number of Monte Carlo iterations summing 2 terms with SD of 3 & 4	SD of the sum, for multiple trials
100	4.7, 5.5, 4.9, 4.9...
10,000	5.02, 5.015, 5.006, ...
1,000,000	5.0002, 4.998, 4.996, ...

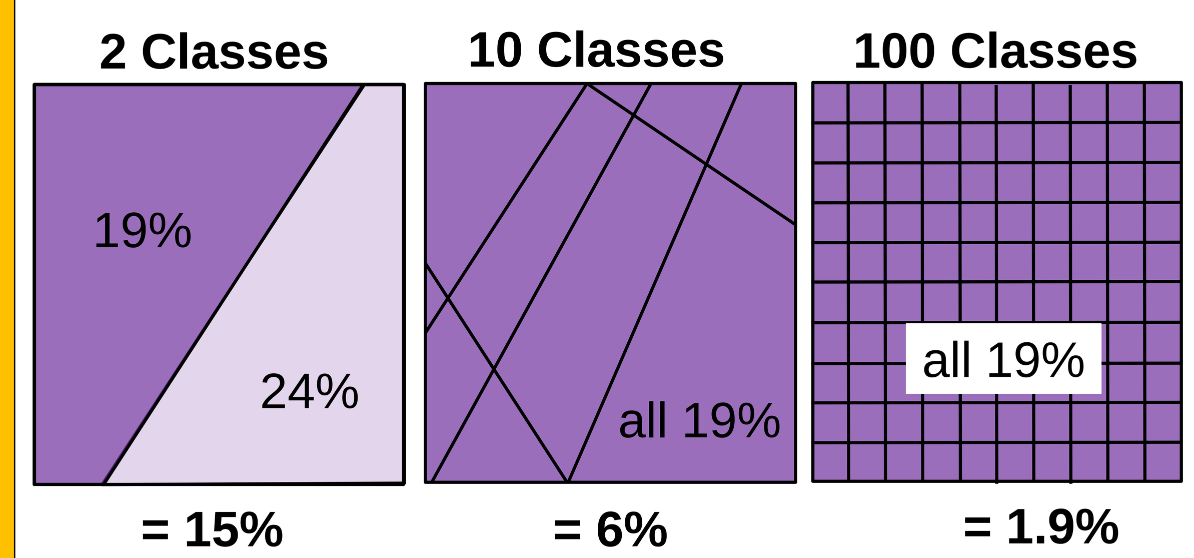
This exercise illustrates that more Monte Carlo samples gives a better estimate of the true uncertainty (lower uncertainty in the uncertainty estimate) but does not reduce the uncertainty.



It also shows how to select the number of Monte Carlo iterations needed to obtain uncertainty estimates of known confidence--you can repeat the exercise and see if you get a similar answer.

Dangers of Addition in Quadrature

The reference level of emissions for the Ivory Coast was based on two forest types. Rainforest emissions had an uncertainty of 19%, and mesic forest had an uncertainty of 24%. The combined uncertainty, obtained by addition in quadrature, came to 15%, although you would think a weighted average would have to fall between these two values. Here is the math: the square root of this sum $(19\% * 18\text{ M ha})^2 + (24\% * 9\text{ M ha})^2$, divided by the sum of the areas, is 15%. If we treated the rainforest in 10 subcategories, each with an uncertainty of 19%, the combined uncertainty using this propagation rule would be 6%. For a million subcategories, the uncertainty would be 0.02%, and with an infinite number of subcategories, there would be no uncertainty.



This unreasonable result is due to the assumption (unwarranted, in this case) that the sources of error are independent. For example, if you had all your savings invested in ExxonMobil, the uncertainty in the value of your holdings would be higher than if you had a diverse portfolio of stocks, bonds, currencies, commodities, and hedge fund strategies, because these have somewhat independent risks. However, if you treated your 100 shares in ExxonMobil as independent, compared to a single holding of 100 shares, it would be absurd to claim that you reduced your uncertainty 10-fold. When forest parcels share uncertainties--for example, in tree allometry and wood density estimates--these need to be applied at the proper (larger) scale. See **Independence of Errors**.

References

Eggleston, Simon, et al., 2006. *2006 IPCC guidelines for national greenhouse gas inventories*. Vol. 5, Chapter 3. Hayama, Japan: Institute for Global Environmental Strategies

McMurray, A, T. Pearson, and F. Casarim. 2017. Guidance on applying the Monte Carlo approach to uncertainty analyses in forestry and greenhouse gas accounting. Winrock International Report.

Campbell, J.L., R.D. Yanai, M.B. Green, G.E. Likens, C.R. See, A.S. Bailey, D.C. Buso, and D. Yang. 2016. Uncertainty in the net hydrologic flux of calcium in a paired-watershed harvesting study. *Ecosphere* 7(6).

Acknowledgements

The QUEST RCN is funded by the National Science Foundation, DEB 1257906.

Monte Carlo Simulation

In Monte Carlo error propagation, the contributing uncertainties are randomly sampled in each of many iterations of a calculation, and the distribution of these multiple estimates indicates the uncertainty of the estimate. The most egregious error in error reporting is to report the standard error of the Monte Carlo estimates.

Number of Monte Carlo iterations summing 2 terms with SD of 3 & 4	SE, incorrectly treating iterations as samples
100	0.5
10,000	0.05
1,000,000	0.005

While the standard error is appropriate for reporting confidence in an estimate based on random samples of a population, it is incorrect to treat each of the estimates from iterations of a Monte Carlo simulation as if it were an observation and calculate standard error from the set of iterations. However, this approach is recommended for reducing error estimates! (McMurray, 2017)

Instead, uncertainty should be obtained from the dispersion of the Monte Carlo iterations, for example as a 95% confidence interval. This gives the correct answer (5, not 0.05 or 0.005, in this example), with precision increasing with the number of iterations (see Table 1).

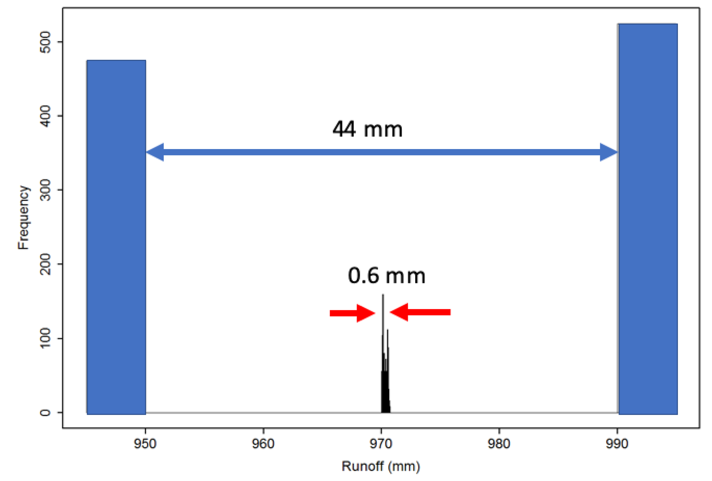
Join QUEST -- or QUERCA !

QUEST (Quantifying Uncertainty in Ecosystem Studies) is a Research Coordination Network devoted to promoting and improving the practice of error propagation. We hope to continue as QUERCA (Quantifying Uncertainty Estimates Required for Carbon Accounting) to help countries involved in reducing emissions from deforestation and forest degradation (REDD+) for climate mitigation. Please join us!

- Find us at **www.quantifyinguncertainty.org**
- Read papers, share sample code, stay updated with QUEST News
- Email us at quantifyinguncertainty@gmail.com
- Follow us on Twitter: @QUEST_RCN

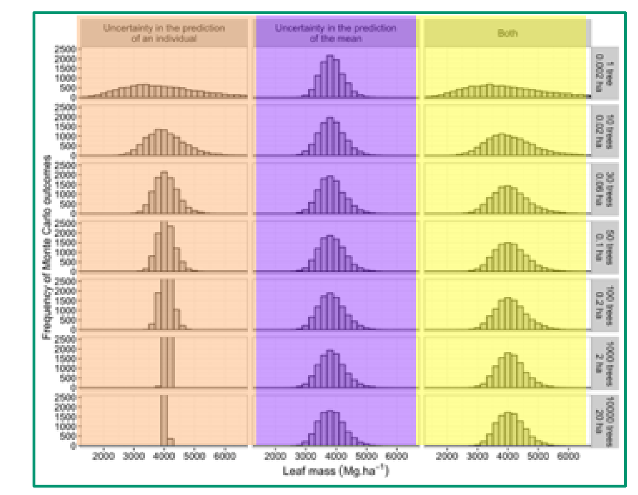
More Dangers Involv
Independence of Err

A common mistake in Monte Carlo simulatio
uncertainty sources as independent when th
For example, the area of a watershed contri
uncertainty to estimates of runoff, as it is ne
stream discharge to units of mm. Randomly
uncertainty each day is incorrect (Campbell
One watershed area needs to be used throu
iteration of the Monte Carlo.



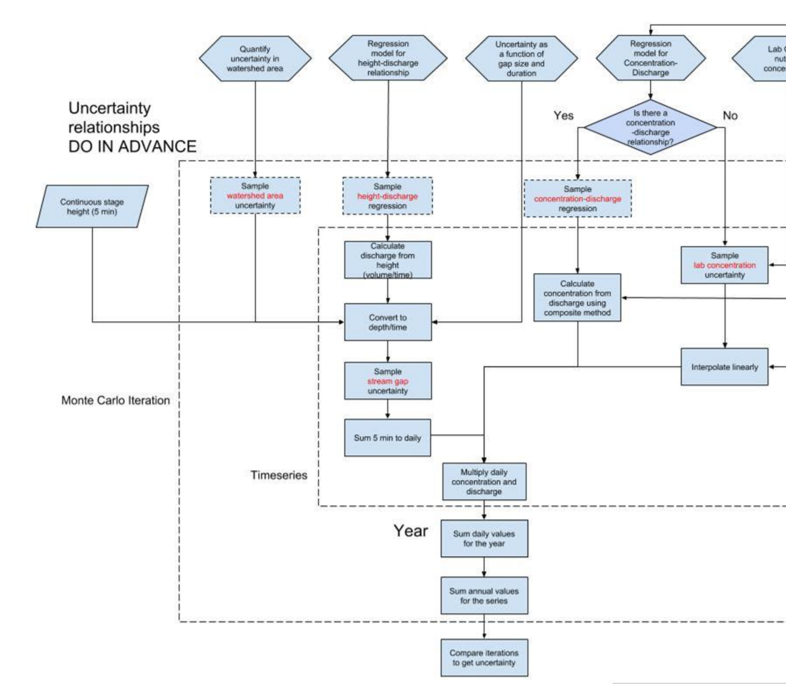
Uncerta
Watershed 3
Experimenta
uncertainty i
used two est
area, one su
from a LiDAR
Incorre
watershed a
observation
in a 95% CI
Correct
one watersh
Carlo iteratio
44 mm.

A similar mistake commonly occurs when pr
error in the allometric equations describing t
between tree diameter and mass. If this unc
applied independently for each tree, estimat
across Monte Carlo simulations when the nu
is large. The uncertainty in the model shoul
only once for each iteration; if the model unc
the average tree biomass, it should be unde
all the trees at once.



Uncertainty
maple leaves, b
propagation of u
allometric regre
tree diameter to
trees at the Hub
Experimental Fo
Uncertainty
individuals is no
numbers of tree
prediction of the
each iteration to
important. Inclu
error is importan
of trees.

Other sources of error, such as the precisio
analysis for stream chemistry or measureme
diameters, introduce errors that are indepen
observation.



It is important to conduct random sampling a
appropriate points in the simulation. To appl
errors independently to each observation re
underestimate of the true uncertainty.