Statistics for Atmospheric and Oceanic Sciences
Class Exercises: Propagation of random error, boot strap, and Monte Carlo
24 July 2006

Part 1: Propagation of Random Error

In this exercise, we will be calculating ΔC_{gasex}, a tracer that is derived from observations of dissolved inorganic carbon (DIC), nutrients, and other hydrographic data (Gruber and Sarmiento, 2002) and stoichiometric ratios associated with biological processes in the interior ocean. This is a quasi-conserved tracer in the interior ocean and reflect the component of the observed DIC that is due to natural air-sea gas exchange that already existed in preindustrial times. The equation for ΔC_{gasex} is:

$$\Delta C_{gasex} = sDIC - r_{C:P}[PO_4] - 0.5(sAlk - r_{N:P}[PO_4]) - s\Delta C_{ant} - C$$

In this equation, the second term is used to remove the effects of the soft tissue pump, the third term is used to remove the effects of the alkalinity pump, ΔC_{ant} is an estimate of the oceanic uptake due to the atmospheric perturbation due to anthropogenic activities, and C is a constant that is selected so that the mean value of ΔC_{gasex} at the surface is equal to zero.

1. Use getnc to read in the following datasets from DCgasex_inputs.nc in Matlab:
   tco2_new = DIC (regridded from GLODAP)
   tco2_err_new = DIC error (regridded from GLODAP)
   salinity = salinity extrapolated from point observations in the crudest way imaginable
   PO4 = salinity normalized phosphate extrapolated point observations in the crudest way imaginable
   talk_new = total alkalinity regridded from GLODAP
   talk_new_err = total alkalinity error regridded from GLODAP
   antco2_new = Anthropogenic carbon regridded from GLODAP
   antco2_err_new = Anthropogenic carbon error regridded from GLODAP

   All units are umol/kg except salinity (psu).
   When you read in these files, set NaN's to some distinctive value such as 1e-34.

2. PO_4 and salinity have been crudely extrapolated and therefore have values over land. Create a mask to mask out land regions using "find". For example:
   missing=find(tco2_new== NaN value you selected);
   PO4(missing)==NaN value you selected

3. Compute ΔC_{gasex} using the above datasets and the following values:
$r_{C:P}$ = 117 +/- 14 (Anderson and Sarmiento, 1996)
$r_{N:P}$ = 15 +/- 1 (Anderson and Sarmiento, 1996)
C = 777.3

Don't forget that PO_4 is already salinity normalized!

3. Compute the uncertainty for each grid cell. Use the uncertainty fields from Dcgasex_inputs.nc for alkalinity, anthropogenic carbon, and DIC. The uncertainty for the other components is:
   PO4 = 0.05
   $r_{C:P}$ = 14 (Anderson and Sarmiento, 1994)
   $r_{N:P}$ = 1 (Anderson and Sarmiento, 1994)

4. Adapt the Matlab routine provided on the web site and write the results of #2 and #3 to a netCDF file, and read it into Ferret. (If you are short on time, just use the inferior graphics in Matlab.)

Save this workspace. You will be using it again.

5. Pick a transect that you like and plot ΔC_{gasex} and it's errors.
   For example, shade var[x=155W] plots a transect along 155 West.
   You can use shade/lev=(min value, max value, interval) var to plot both on the same scale.

6. Part 2: Estimating Uncertainty Due to Biases in $r_{c:p}$

We have just calculated the uncertainty with the assumption that all of the errors are random and uncorrelated. This is probably a pretty good assumption for the observations of DIC, Alk, PO_4. Many of them were made on different ships at different dates, but Bob Key has done extensive work to ensure that any sampling biases are accounted for. (There is substantial spatial coherence to errors incurred by extrapolating point data to a global grid, but let's put that aside for the moment.)

Unfortunately, there may well be spatially coherent errors in $r_{c:p}$ and $r_{n:p}$ which would lead to substantial biases in this tracer. We are now going to use Monte Carlo methods to estimate the uncertainty due to the possibility of spatially uniform errors in these ratios.

7. Go back to your Matlab work space.

8. Pick some nice, round number of iterations that you would like to use, niter. Create a new variable with dimensions x,y,z,niter (e.g. gasex_mc=zeros(im,jm,km,niter);).

9. Re-calculate ΔC_{gasex} niter times, adding a random perturbation scaled by the uncertainty estimates (above) to $r_{c:p}$ and $r_{n:p}$. This perturbation should be the same for all x,y,z grid points but different for each iteration. Use randn(niter,1); to get n randomly generated numbers with a Gaussian distribution and a standard deviation of +/-1 (which can then be multiplied by the actual uncertainty).

10. The standard deviation of the niter Monte Carlo simulations is the 1-sigma (68.5%) confidence limit, and twice that value is the 95% confidence interval. Write the 95% confidence level to a netCDF files

11. Plot this on the same scale as the random error estimate. Is the potential uncertainty due to this bias generally larger or smaller? Are there regions that are particularly sensitive to this?

References:

Anderson, L. A. and J. L. Sarmiento, Medfield ratios of remineralization determined by nutrient data analysis, *Global Biogeochem. Cycles*, 8, 65-80, 1994.

Gruber, N. and J. L. Sarmiento, Biogeochemical/Physical Interactions in Elemental Cycles, in *The Sea: Biological-Physical Interactions in the Oceans*, A. R. Robinson and J. J. McCarthy and B. J. Rothschild, eds., John Wiley and Sons, New York, 337-399, 2002.