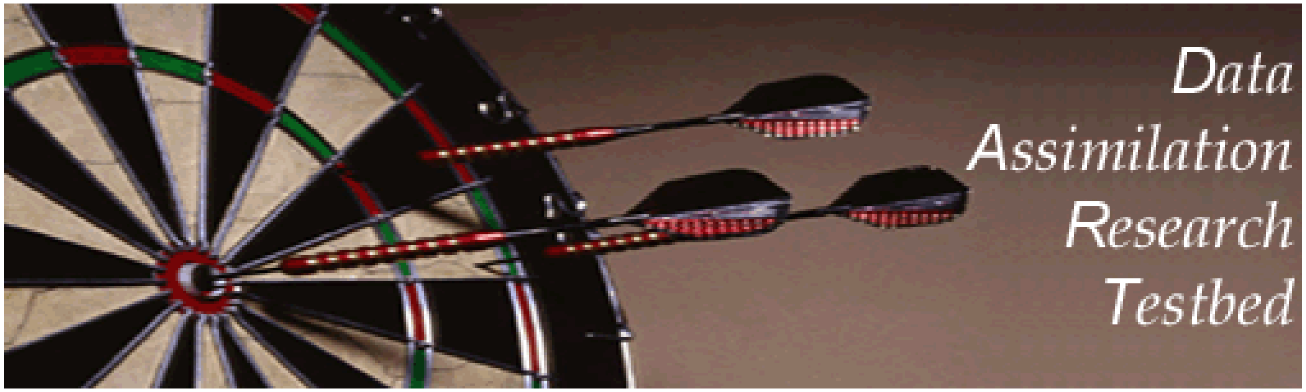
Data Assimilation Research Testbed Tutorial

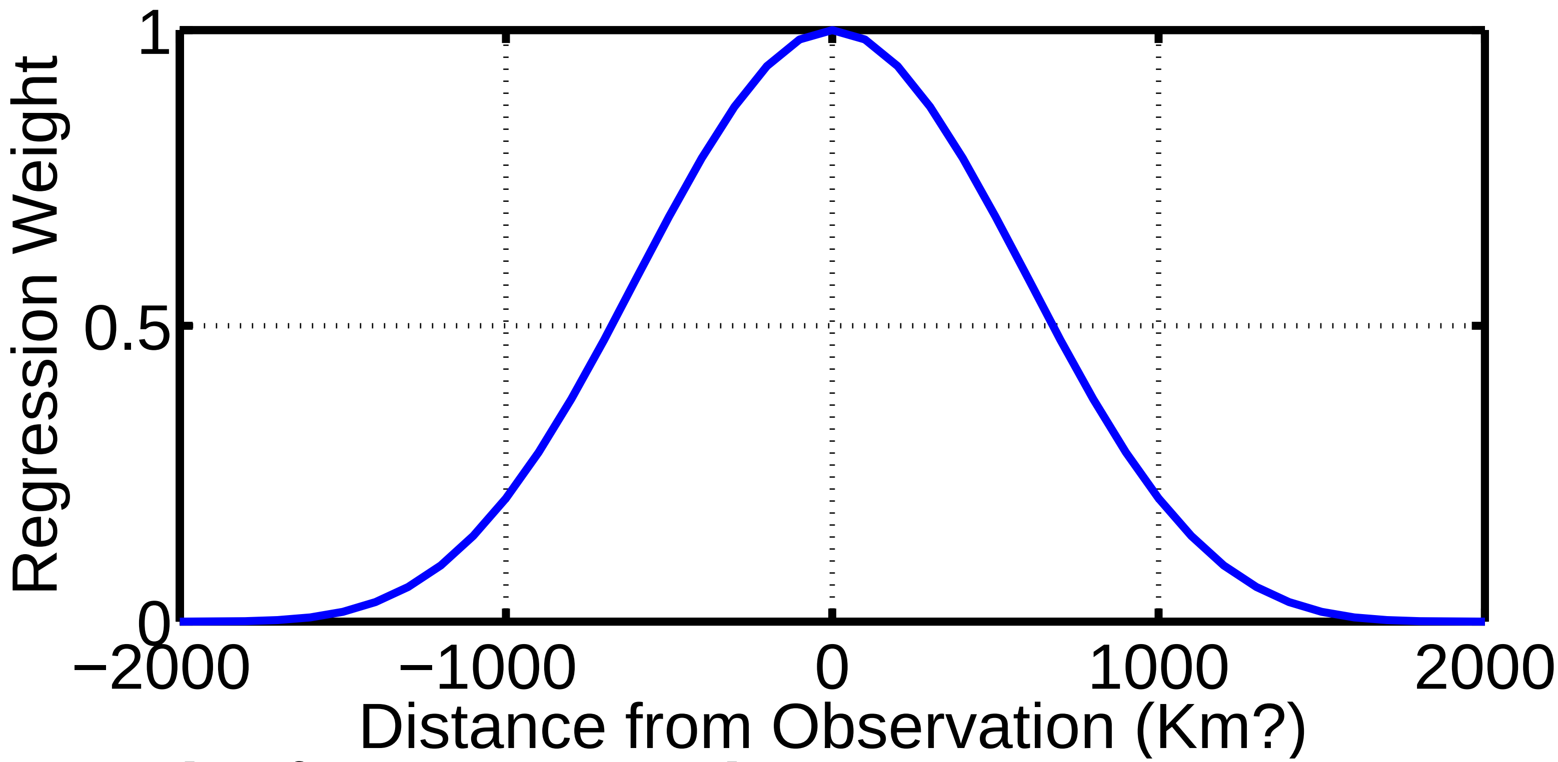


Section 13: Hierarchical Group Filters and Localization

Version 1.0: June, 2005

Ways to deal with regression sampling error:

3. Use additional a priori information about relation between observations and state variables.



Can use other functions to weight regression.

Unclear what *distance* means for some obs./state variable pairs.

Referred to as LOCALIZATION.

Localization is function of expected correlation between obs and state.

Often, don’t know much about this.

Horizontal distance between same type of variable may be okay.

What is expected correlation for co-located temperature and pressure?

What about vertical localization? Looks pretty complex.

What about complicated forward operators:

Expected correlation of satellite radiance and wind component?

Note: DART does allow vertical localization for more complex models.

Ways to deal with regression sampling error:

4. Try to determine the amount of sampling error and correct for it:

A. Could weight regressions based on sample correlation.

Limited success in tests.

For small true correlations, can still get large sample correl.

B. Do bootstrap with sample correlation to measure sampling error.

Limited success.

Repeatedly compute sample correlation with a sample removed.

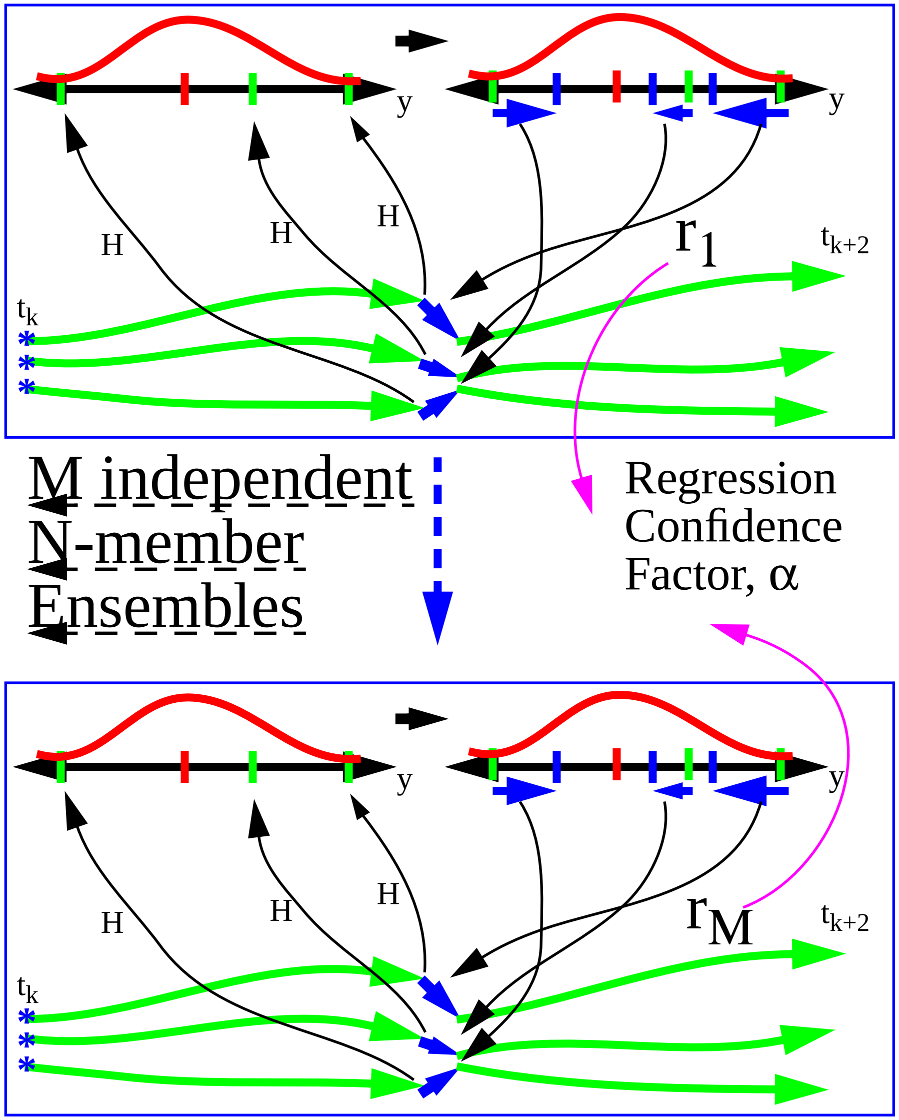
C. Use hierarchical Monte Carlo.

Have a ‘sample’ of samples.

Compute expected error in regression coefficients and weight.

Ways to deal with regression sampling error:

4C. Use hierarchical Monte Carlo: ensemble of ensembles.



M groups of N-member ensembles.

Compute obs. increments for each group.

For given obs. / state pair:

1. Have M samples of regression coefficient, β.
2. Uncertainty in β implies state variable increments should be reduced.
3. Compute regression confidence factor, α.

4C. Use hierarchical Monte Carlo: ensemble of ensembles.

Split ensemble into M independent groups.

For instance, 80 ensemble members becomes 4 groups of 20.

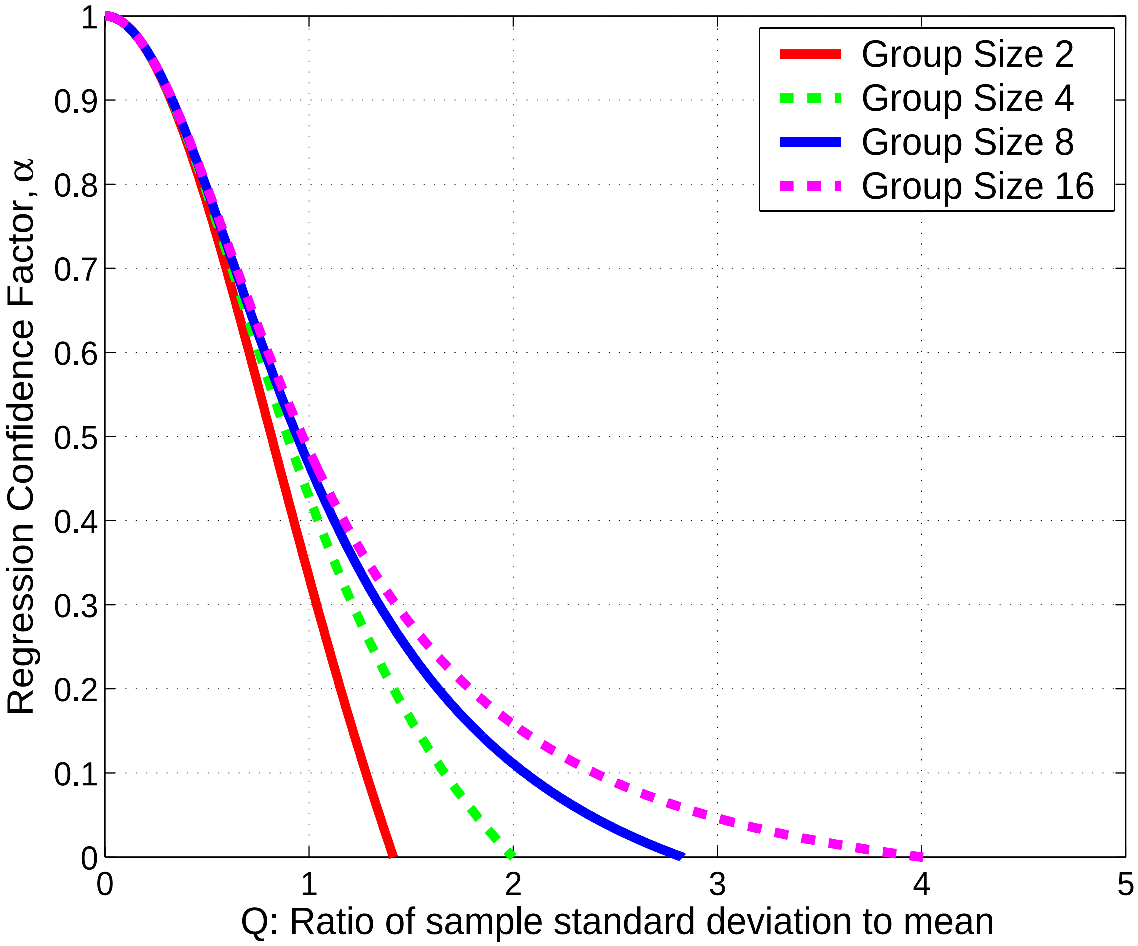
With M groups get M estimates of regression coefficient, β_i.

Find regression confidence factor α (weight) that minimizes:

$$\sqrt{\sum_{j=1}^{M} \sum_{i=1, i \neq j}^{M} [\alpha\beta_i - \beta_j]^2}$$

Minimizes RMS error in the regression (and state increments).

4C. Use hierarchical Monte Carlo: ensemble of ensembles.



Weight regression by α.

If one has repeated observations, can generate sample mean or median statistics for α.

Mean α can be used in subsequent assimilations as a localization.

α is function of M and $Q = \Sigma_\beta / \bar{\beta}$ (sample SD / sample mean regression)

4C. Use hierarchical Monte Carlo: ensemble of ensembles.

Hierarchical filter controlled by setting number of groups, M.
num_groups in *filter_nml*.

If we don't know how to localize to start with, can use groups to help.

Try splitting 80 ensemble members into 4 groups for Lorenz-96.
(4 groups of 20 each).
Use adaptive inflation (0.05 lower bounds) to make things nice.

Can look at the time mean and median value of α.

Essentially an estimate of a 'good' localization for a given observation.

4C. Use hierarchical Monte Carlo: ensemble of ensembles.

After running the 80 by 4 ‘group’ filter, look at plots of α.

Use *plot_reg_factor* in matlab.

Select default input file name.

Only observations 1, 2, 3, and 4 are available:
Located at: 0.39, 0.17, 0.64, 0.86

Think about value of time median vs. time mean.

Could use time mean or median as prior localization functions.

Play around with model error again. What happens to localization?