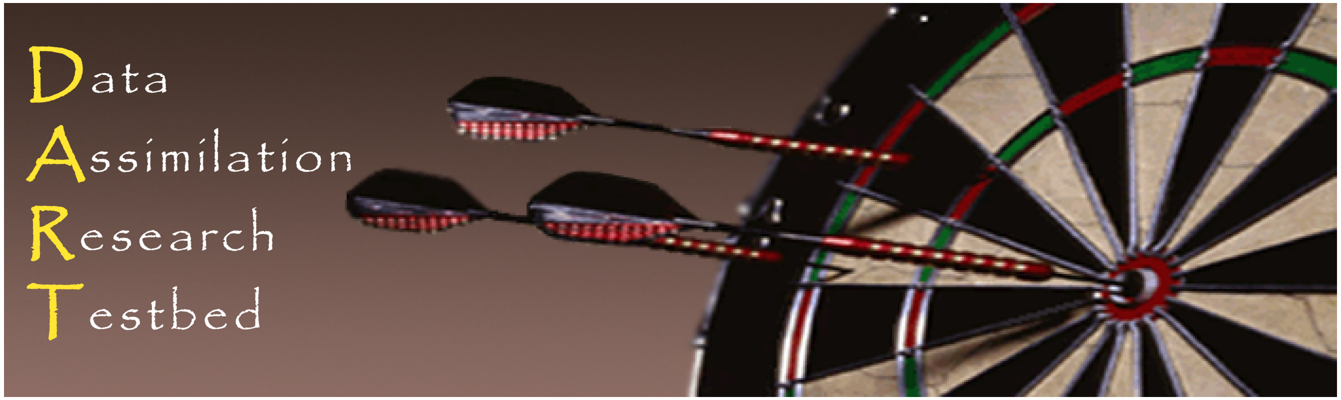
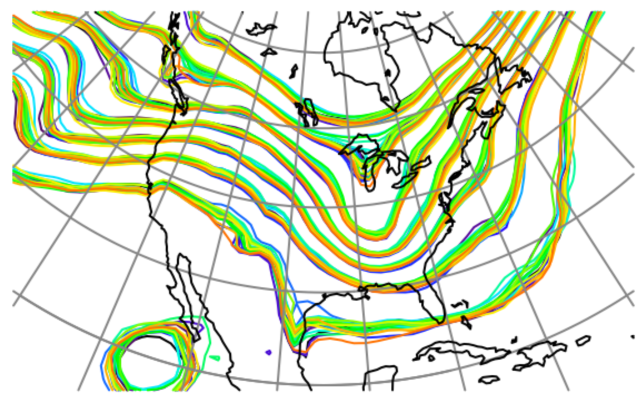


DART Tutorial Section 13: Hierarchical Group Filters and Localization



The National Center for Atmospheric Research is sponsored by the National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

NCAR UCAR | National Center for Atmospheric Research

Ways to deal with regression sampling error

1. Ignore it: if number of unrelated observations is small and there is some way of maintaining variance in priors.
   We did this in the 3 and 9 variable models.

2. Use larger ensembles to limit sampling error (test in lorenz_96).
   This can get expensive for big problems.
   Try modifying *ens_size* in *filter_nml* (try 40, 80, 160).

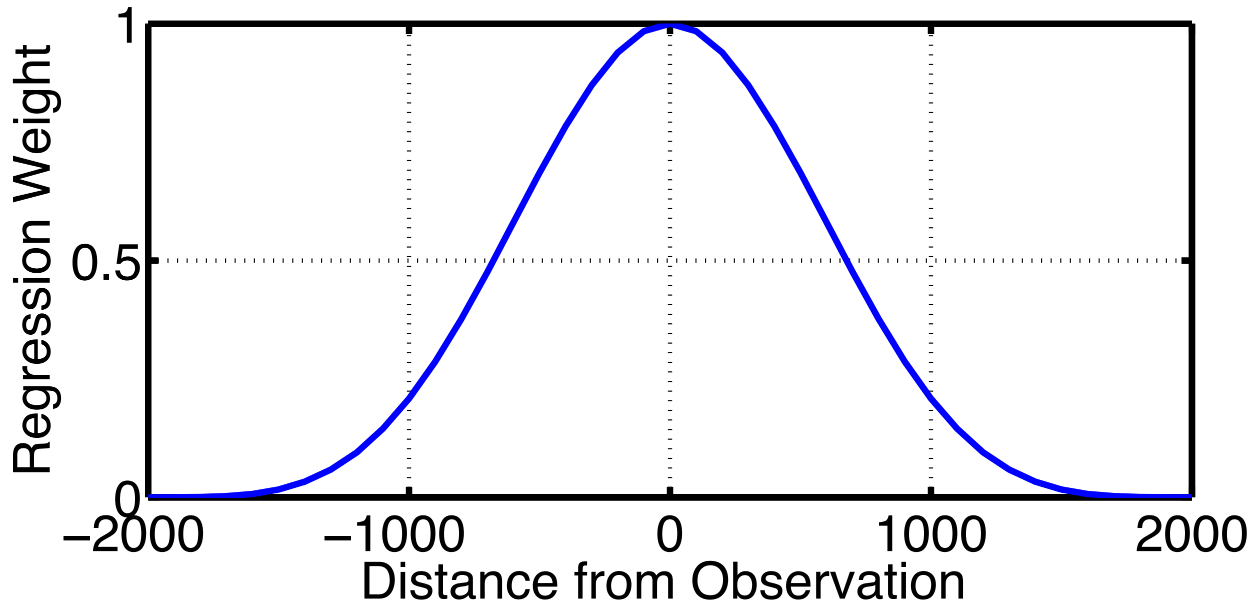
Reminder from Section 8.

3. Use additional a priori information about relation between observations and state variables.
   Don't let an observation impact state if they are known to be unrelated.

4. Try to determine the amount of sampling error and correct for it.
   There are many ways to do this; some simple, some complex.

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

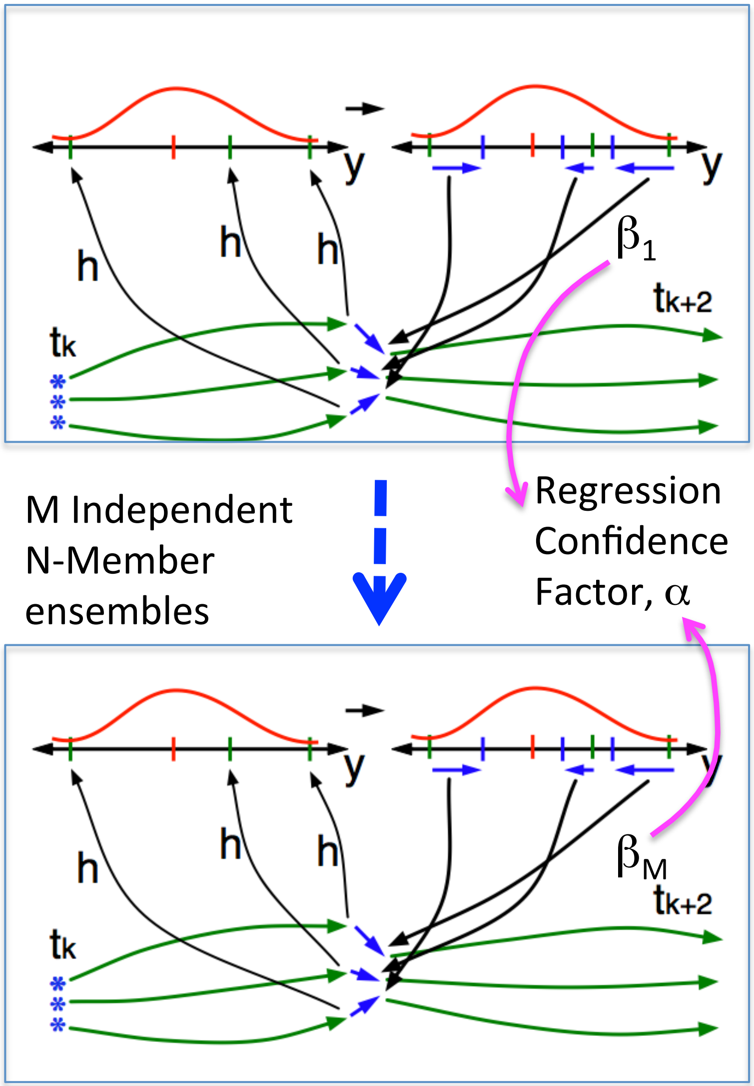
4. Try to determine the amount of sampling error and correct for it:

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.

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



M groups of N-member ensembles.

Compute observation increments for each group.

For given observation/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.

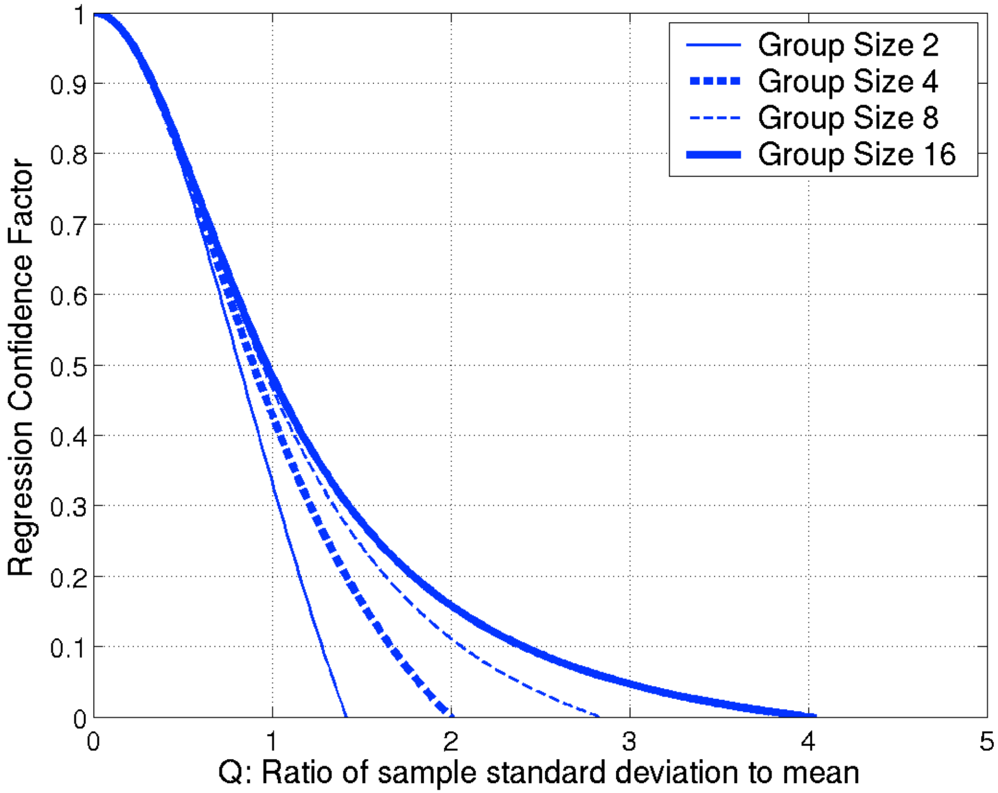
With M groups get M estimates of regression coefficient, β_i .

Find regression confidence factor a (weight) that minimizes:

$$\sqrt{\sum_{j=1}^{M}\sum_{i=1,i\neq j}^{M}\left(\alpha\beta_i-\beta_j\right)^2}$$

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.

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

4C. Use hierarchical Monte Carlo: ensemble of ensembles

models/lorenz_96/work/

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

Try splitting 80 ensemble members into 4 groups of 20 members for Lorenz 96.

```
&assim_tools_nml
   cutoff          = 1000000.0
   …
&filter_nml
   ens_size        = 80
   num_groups      = 4
   inf_flavor      = 0,        0
   …
```

No localization

No inflation

4C. Use hierarchical Monte Carlo: ensemble of ensembles

Turn on regression factor diagnostics.

```
&reg_factor_nml
   select_regression    = 1
   input_reg_file       = "time_mean_reg"
   save_reg_diagnostics = .true.
   reg_diagnostics_file = "reg_diagnostics"
  /
```

After running the 80 by 4 'group' filter, look at plots of α.
Essentially an estimate of a 'good' localization for a given observation.

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?

More Detailed Look at Hierarchical Filters

A more detailed look at some features of group filters is available in the tutorial directory in the file OLD_section_13.pdf.

Pages 10-54 complement the materials in this section.

WARNING: The material on pages 1-9 of OLD_section_13.pdf is outdated.

DART Tutorial Index to Sections