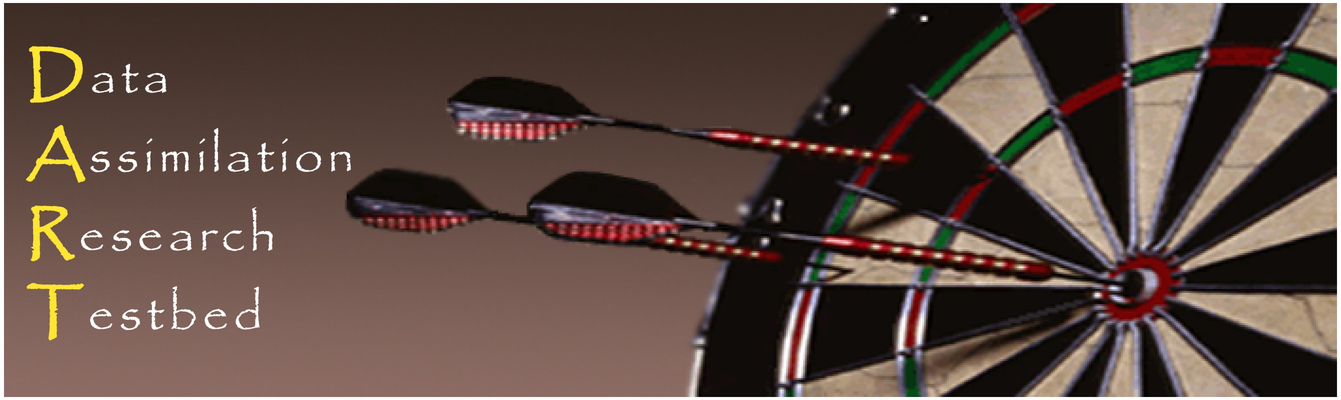
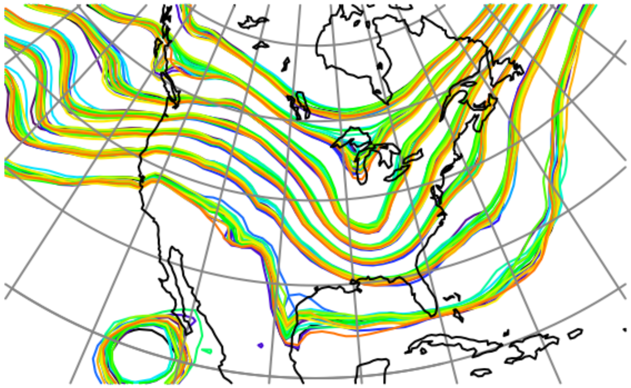


DART Tutorial Section 13:
Hierarchical Group Filters and Localization



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.

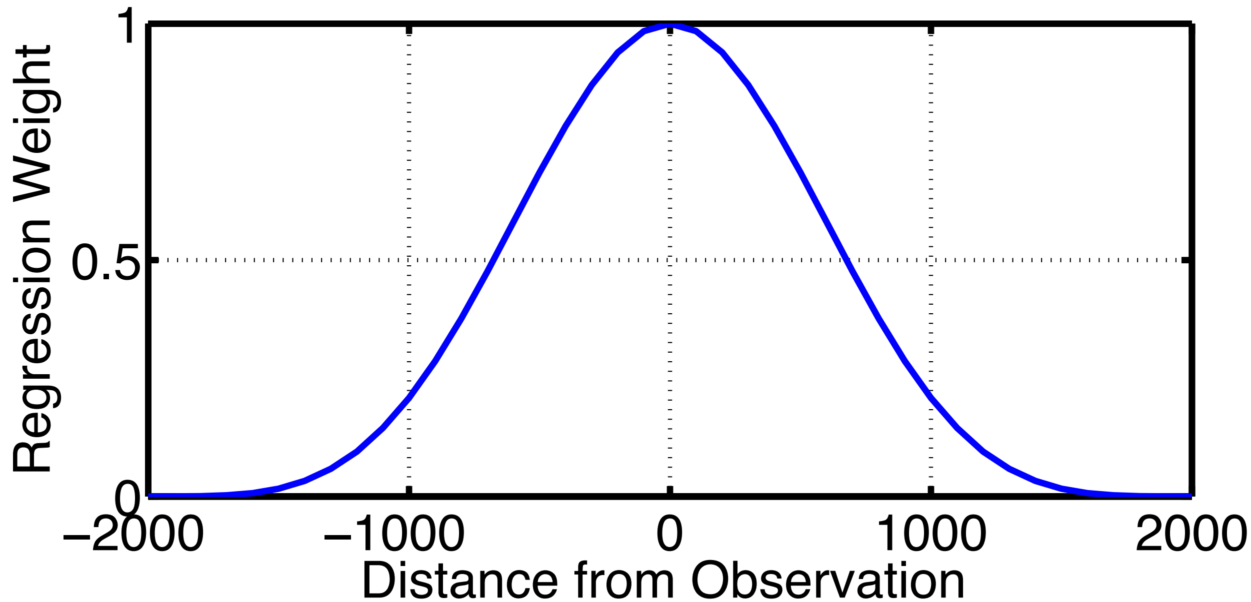
NCAR UCAR | National Center for Atmospheric Research

This tutorial section has not yet been updated for the Manhattan release of DART.

A new version is expected in April 2017.

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

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).

DART Tutorial Index to Sections