2011 Hurricane Tutorial
April 27, 2011
Boulder, CO

Introduction to Data Assimilation and Community Gridpoint Statistical Interpolation System (GSI)

Ming Hu

Developmental Testbed Center
NCAR-NOAA/GSD

DTC Developmental Testbed Center

What is Data Assimilation

- Numerical Weather Prediction is an initial-boundary value problem
  - Given an estimate of the present state of the atmosphere (initial conditions), and appropriate surface and lateral boundary conditions, the model simulates (forecasts) the atmospheric evolution
  - The more accurate the estimate of the initial conditions, the better the quality of the forecasts
- Data Assimilation: The process of combining observations and short-range forecasts to obtain an initial condition for NWP
- The purpose of data assimilation is to determine as accurately as possible the state of the atmospheric flow by using all available information

Three-dimensional Variational Data Analysis (3Dvar)
Concepts and Methods

Basic equation and concepts

$$J(x) = (x-x_b)^T \mathbf{B}^{-1}(x-x_b)+(y-H[x])^T \mathbf{R}^{-1}(y-H[x])$$
$$= J_b + J_o$$

- J is called the *cost function* of the analysis (*penalty function*)
- J_b is the *background term*; J_o is the *observation term*.
- The dimension of the model state is n and the dimension of the observation vector is p.
  - x_t true model state (dimension n)
  - $\mathbf{x}_b$ background model state (dimension n)
  - $\mathbf{x}_a$ analysis model state (dimension n)
  - $\mathbf{y}$ vector of observations (dimension p)
  - H observation operator (from dimension n to p)
  - **B** covariance matrix of the background errors ($\mathbf{x}_b - \mathbf{x}_t$) (dimension $n \times n$)
  - **R** covariance matrix of observation errors ($\mathbf{y} - H[\mathbf{x}_t]$) (dimension $p \times p$)

Hypotheses assumed

- **Linearized observation operator**: the variations of the observation operator in the vicinity of the background state are linear:
  - for any $\boldsymbol{x}$ close enough to $\boldsymbol{x}_b$:

    $H(\boldsymbol{x}) - H(\boldsymbol{x}_b) = \boldsymbol{H}(\boldsymbol{x} - \boldsymbol{x}_b)$, where $\boldsymbol{H}$ is a linear operator.
- **Non-trivial errors**: **B** and **R** are positive definite matrices.
- **Unbiased errors**: the expectation of the background and observation errors is zero i.e. $< x_b - x_t > = < y - \mathrm{H}(x_t) > = 0$
- **Uncorrelated errors**: observation and background errors are mutually uncorrelated i.e. $< (x_b - x_t)(y - \mathrm{H}[x_t])^T > = 0$
- **Linear analysis**: we look for an analysis defined by corrections to the background which depend linearly on background observation departures.
- **Optimal analysis**: we look for an analysis state which is as close as possible to the true state in an r.m.s. sense
  - i.e. it is a minimum variance estimate.
  - it is closest in an r.m.s. sense to the true state $\boldsymbol{x}_t$.
  - If the background and observation error pdfs are Gaussian, then $\boldsymbol{x}_a$ is also the *maximum likelihood estimator* of $\boldsymbol{x}_t$

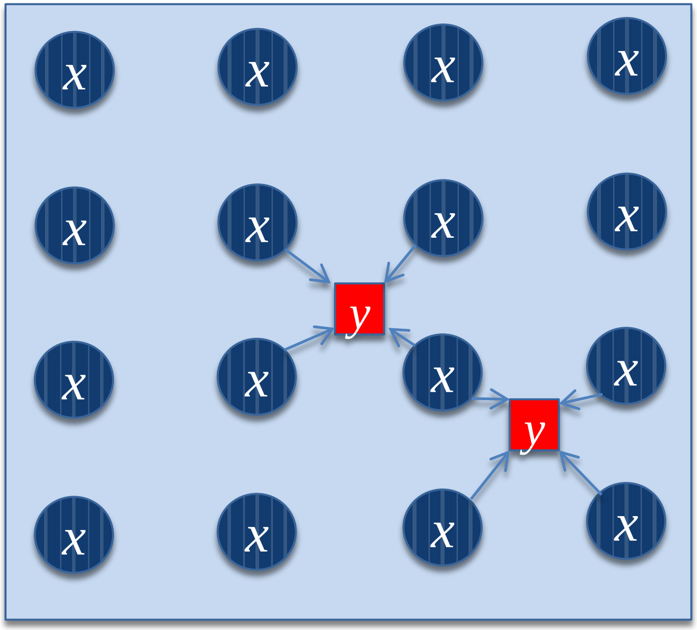
Background term

$$J(x) = (x-x_b)^T \boldsymbol{B}^{-1}(x-x_b)+(y-\mathrm{H}[x])^T \boldsymbol{R}^{-1}(y-\mathrm{H}[x])$$

- Background (forecast field): x_b
- Analysis: x
  - Start from $x = x_b$
- Analysis increment: x-x_b
- Background error covariance: B
  - Variance
  - Correlation
    - Horizontal and vertical
    - balance

Observation term

$$J(x) = (x - x_b)^T \mathbf{B}^{-1} (x - x_b) + (y - \mathrm{H}[x])^T \mathbf{R}^{-1} (y - \mathrm{H}[x])$$



- Observation: y
- Observation operator: H[x]
  - Most: 3D interpolation
  - Some: Complex function
    - Radiance (CRTM) = $f(t,q)$
    - Radar Reflectivity = $f(q_r, q_s, q_h)$
- Observation innovation: y-H[x]
- Observation error variance: R
  - No correlation between two observations

Simplification: scalar case

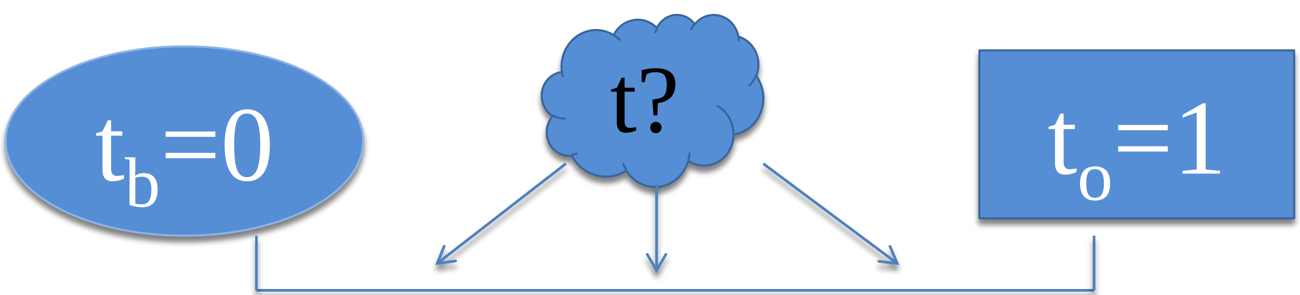
$$J(x) = (x-x_b)^T \boldsymbol{B}^{-1}(x-x_b)+(y-\mathrm{H}[x])^T \boldsymbol{R}^{-1}(y-\mathrm{H}[x])$$

Assume

x is scalar (t)

B, R, H, T

$$\mathrm{J(t)} = (t-t_b)\sigma_B^{-1}(t-t_b) + (t_o-t)\sigma_R^{-1}(t_o-t)$$



DTC Developmental Testbed Center

The scalar case: solution

$$J(t) = (t-t_b)\sigma_B^{-1}(t-t_b) + (t_o-t)\sigma_R^{-1}(t_o-t)$$

$$J(t) = \frac{(t-t_B)^2}{\sigma_B} + \frac{(t_o-t)^2}{\sigma_B}$$

When $\frac{\partial J(t)}{\partial t}=0$, J(t) is minimum and t is the best guess to truth

- it is a minimum variance estimate.
- it is closest in an r.m.s. sense to the true state $\boldsymbol{x}_t$.
- If the background and observation error pdfs are Gaussian, then $\boldsymbol{x}_a$ is also the *maximum likelihood estimator* of $\boldsymbol{x}_t$

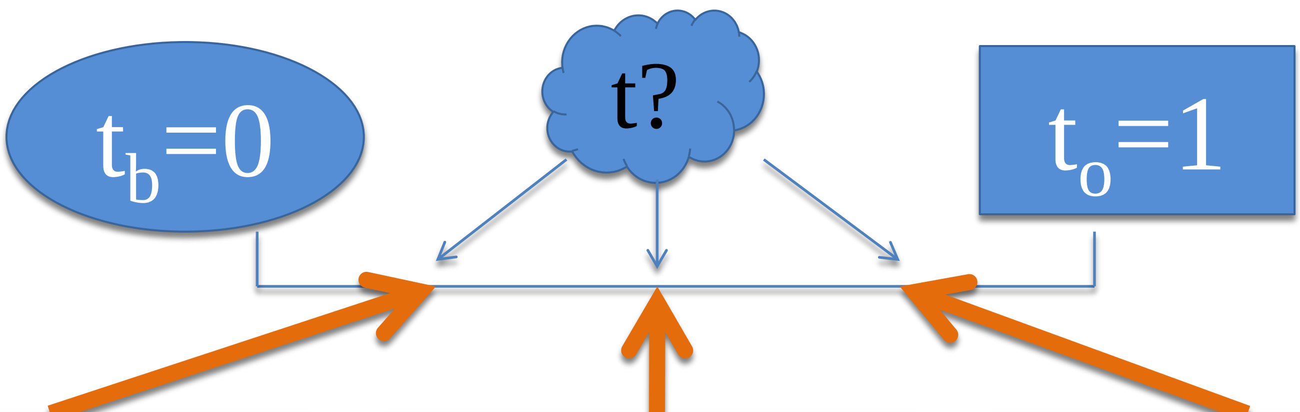
$$t = \frac{\sigma_R}{\sigma_R+\sigma_B} t_B + \frac{\sigma_B}{\sigma_R+\sigma_B} t_o$$

The scalar case: results

$$J(t) = (t-t_b)\sigma_B^{-1}(t-t_b) + (t_o-t)\sigma_R^{-1}(t_o-t)$$

$$t = \frac{\sigma_R}{\sigma_R + \sigma_B} t_B + \frac{\sigma_B}{\sigma_R + \sigma_B} t_o$$

The analysis is decided by the ratio of background error and observation error



$0.25*\sigma_R=\sigma_B$	$\sigma_R=\sigma_B$	$\sigma_R=0.25*\sigma_B$
$t=0.8*t_b+0.2*t_o$	$t=0.5*t_b+0.5*t_o$	$t=0.2*t_b+0.8*t_o$
$t=0.2$	$t=0.5$	$t=0.8$

Background error covariance

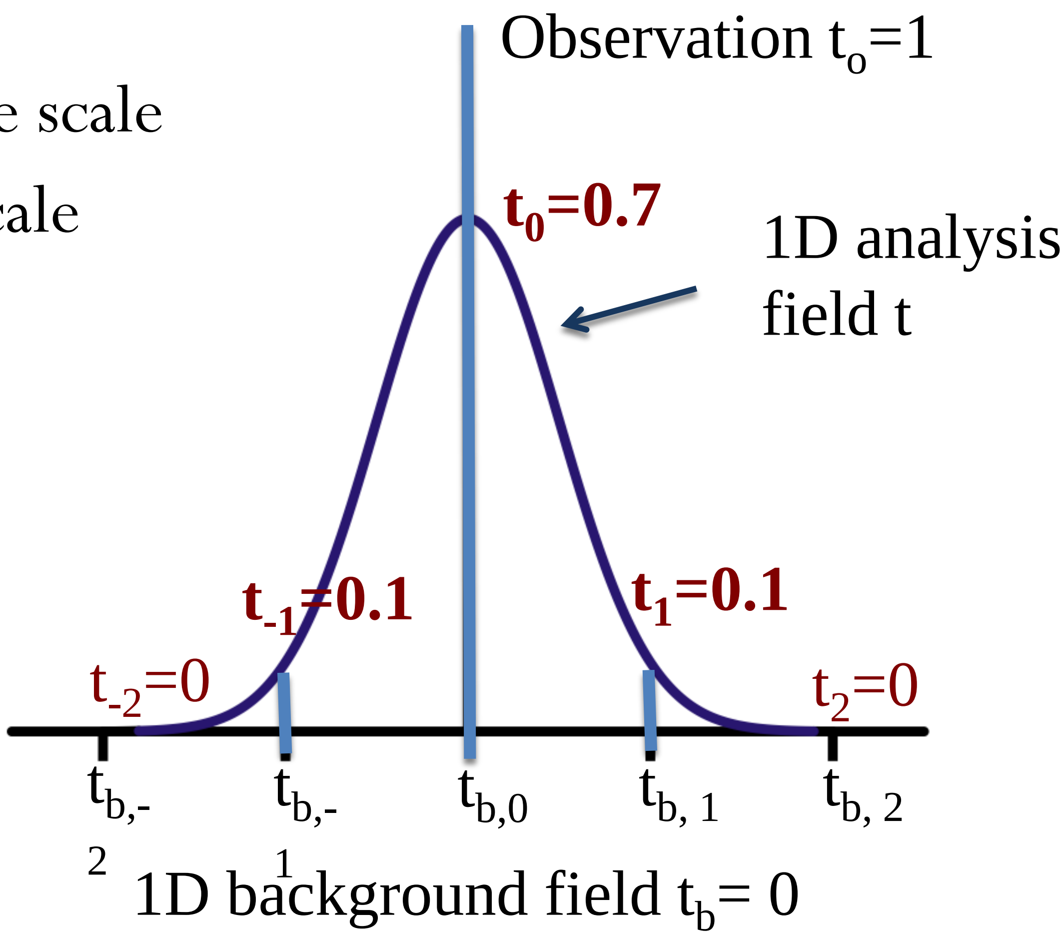
$$J(x) = (x-x_b)^T \mathbf{B}^{-1}(x-x_b)+(y-H[x])^T \mathbf{R}^{-1}(y-H[x])$$

- Background error covariance: B
  - Variance
    - Diagonal of the B matrix
    - Ratio to the observation variance to decide how much analysis results fit to the observations
  - Correlation
    - Horizontal and vertical
    - Balance

Background error covariance: scale

$$J(x) = (x-x_b)^T \mathbf{B}^{-1}(x-x_b)+(y-H[x])^T \mathbf{R}^{-1}(y-H[x])$$

- Correlation:
  - Horizontal influence scale
  - Vertical influence scale
  - In Gaussian Shape
    - background and observation error pdfs are Gaussian

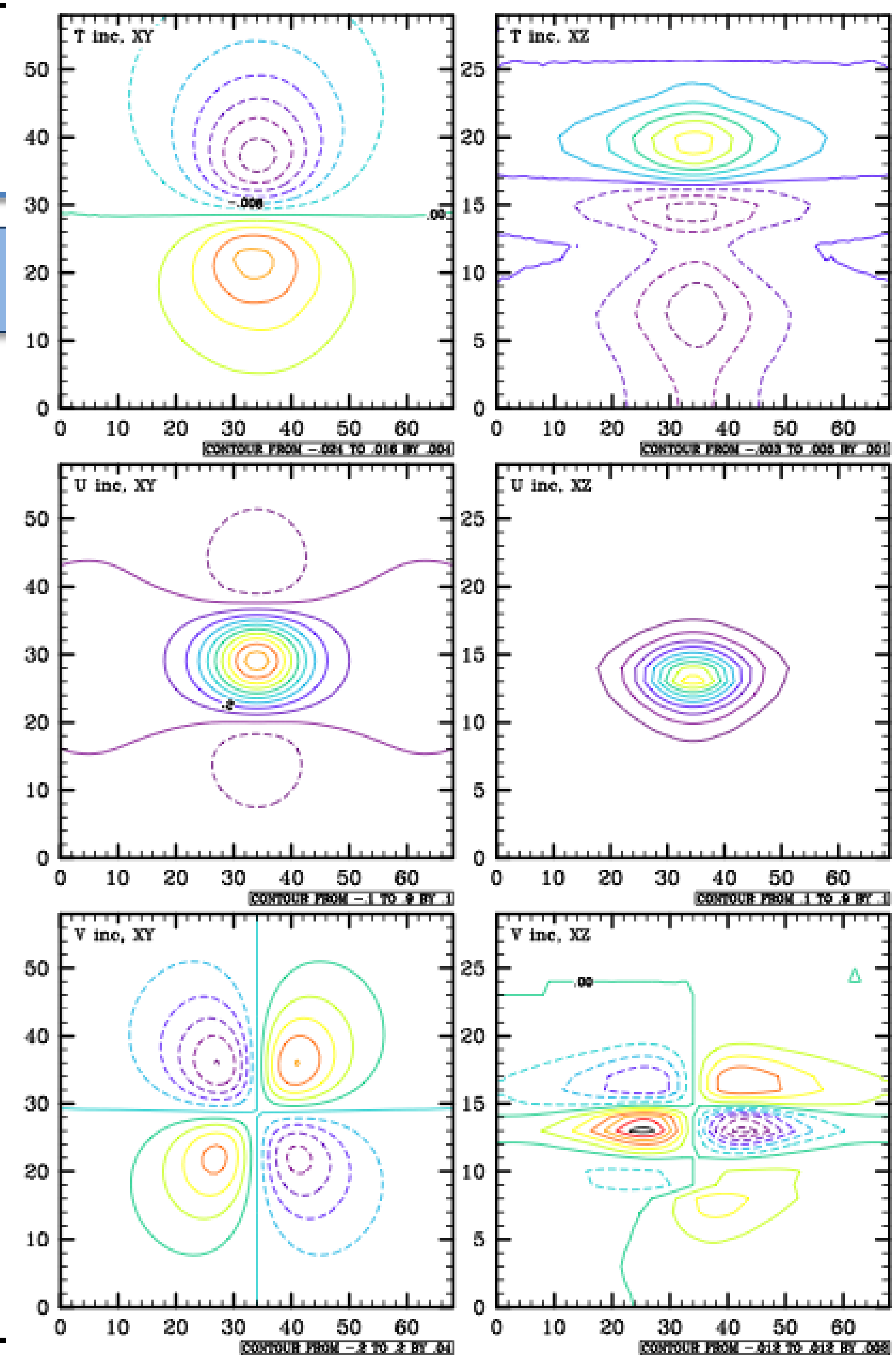


Developmental Testbed Center

B: Balance

$$J(x) = (x - x_b)^T \mathbf{B}^{-1} (x - x_b)$$

- Correlation:
  - Balance among different fields, most important is balance between mass and wind fields



Summary with GSI

Analysis results: wrf_inout

Background field: wrf_inout

Background covariance: berror

$$J(x) = (x-x_b)^T \boldsymbol{B}^{-1}(x-x_b)+(y-\mathrm{H}[x])^T \boldsymbol{R}^{-1}(y-\mathrm{H}[x])$$

Observations: prepbufr/bufr files

Code (setup*) and CRTM

Observation error: errtable

GSI
History, Current, and Future

Based on John Derber's talk

in 2010 summer GSI tutorial

History

- OI systems, first statistic data analysis system
- The Spectral Statistical Interpolation (SSI) analysis system was developed at NCEP in the late 1980's and early 1990's.
  - first operational variational analysis system
  - system to directly use radiances
- The Gridpoint Statistical Interpolation (GSI) analysis system was developed as the next generation global/regional analysis system
  - Wan-Shu Wu, R. James Purser, David Parrish, 2002: *Three-Dimensional Variational Analysis with spatially Inhomogeneous Covariances. Mon.Wea. Rev., 130, 2905-2916.*
  - Based on SSI analysis system
  - Replace spectral definition for background errors with grid point version based on recursive filters

Current

- GSI Used in NCEP operations for
  - Regional
  - Global
  - Hurricane
  - Real-Time Mesoscale Analysis
  - Future Rapid Refresh (ESRL/GSD)
- GMAO collaboration (NASA 4DVAR)
- Can work with both WRF and NCEP model
- Evolution to ESMF

General Comments

- GSI analysis code is an evolving system.
  - Scientific advances
    - situation dependent background errors
    - new satellite data
    - new analysis variables
  - Improved coding
    - Bug fixes
    - Removal of unnecessary computations, arrays, etc.
    - More efficient algorithms (MPI, OpenMP)
    - Generalizations of code
      - Different compute platforms
      - Different analysis variables
      - Different models
    - Improved documentation

General Comments

- Code is intended to be used Operationally
  - Must satisfy coding requirements
  - Must fit into infrastructure
  - Must be kept as simple as possible
- External usage intended to:
  - Improve external testing
  - Reduce transition to operations work/time
  - Reduce duplication of effort

Future: options under development

- HWRF uses the GSI for standard operational 3-D var. analysis. Many other options available or under development
  - 4d-var
  - Hybrid assimilation
  - Observation sensitivity
  - FOTO
  - Additional observation types
  - SST retrieval
  - Detailed options
- Options make code more complex – difficult balance between options and simplicity

Community GSI

Community GSI

- Community GSI Goals:
  - Provide current operation GSI capability to research community (O2R) and a pathway for research community to contribute operation GSI (R2O)
  - Provide a framework to enhance the collaboration from distributed GSI developers
- GSI Community code includes:
  - Community GSI repository
  - User's webpage
  - Annual code release with user's guide
  - Annual residential tutorial
  - Help desk (271 registered users)

Community GSI – User's Page

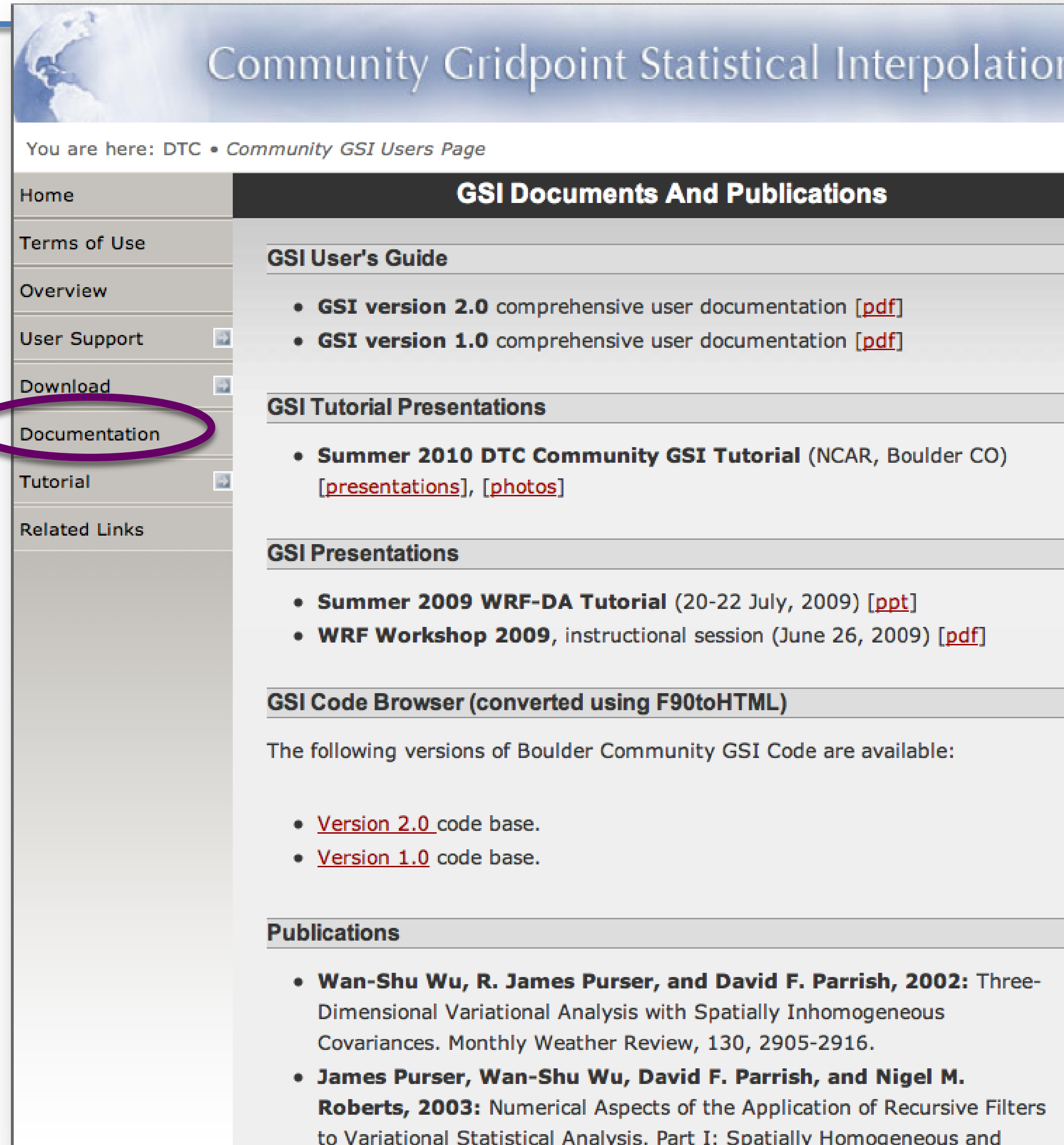
- Mainly support through User's Page and help desk:

  http://www.dtcenter.org/com-GSI/users/index.php

- Release versions
  - Tested with IBM, Linux PGI and ifort, Mac PGI
  - Latest version is Release V2.5
    - Used in this tutorial
    - in 2011 operational HWRF
- release annually: version 3 in April, 2011

Community GSI - Documents

- User's Guide
  - Match each official release
- Tutorial lectures
- Code browser
  - Calling tree
- Key publications



Community GSI - Practice

- On-line tutorial for each release
- Residential tutorial practice cases

Community Gridpoint Statistical Interpolation

You are here: DTC • *Community GSI Users Page*

- Home
- Terms of Use
- Overview
- User Support
- Download
- Documentation
- Tutorial
- Related Links

GSI Tutorials

The GSI On-line Tutorial provides the basic steps to compile the GSI system, and test cases for the code to run.

- Tutorial for community GSI Version 2.0
  - Further exercises from the Summer 2010 Practical
- Tutorial for community GSI Version 1.0

GSI Residental Tutorial classes for new users are generally offered once a year. These classes consist of lectures on the components of the GSI system, as well as practical sessions providing hands-on experience on using the software.

Practical cases

1). Single Observation Tests:

- [a]. ARW background w/ global BE
- [b]. ARW background w/ NAM BE
- [c]. NMM background w/ global BE
- [d]. NMM background w/ NAM BE

2). Test with conventional data (prepbufr):

- [a]. ARW background
- [b]. NMM background

3). Test with conventional and satellite radiance data:

- [a]. ARW background
- [b]. NMM background

4). Test with conventional and gpsro data:

- [a]. ARW background
- [b]. NMM background

5). Test with conventional data and radar radial wind:

- Baseline Test with conventional data
  - [a]. ARW background
  - [b]. NMM background
- Test with radar data
  - [c]. ARW background
  - [d]. NMM background

6). Prepbufr converter examples [practice]

DTC Developmental Testbed Center

Community GSI - Help desk

- wrfhelp@ucar.edu
- HWRF GSI related questions
- Most of questions were answered by DTC staff
- Forward complex questions to NCEP colleague