2013 Beijing GSI Tutorial May 28, 2013 Beijing, China

Introduction to 3DVAR, Gridpoint Statistical Interpolation System (GSI), and its Community Support

Ming Hu

Developmental Testbed Center NCAR-NOAA/GSD

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*)
 - x_b background model state (dimension n)
 - x_a analysis model state (dimension *n*)
 - *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 x close enough to x_b : $H(x) - H(x_b) = H(x - x_b)$, where 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. $\langle x_b x_t \rangle = \langle y H(x_t) \rangle = 0$
- **Uncorrelated errors**: observation and background errors are mutually uncorrelated i.e. $< (x_b x_t)(y H[x_t])^T >= 0$
- 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 x_t .
 - If the background and observation error pdfs are Gaussian, then x_a is also the maximum likelihood estimator of x_t

Background term

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

- Background (forecast field): x_b
- Analysis: x
- Analysis increment: *x*-*x*_b
- Background error covariance: B
 - Variance
 - Correlation
 - Horizontal and vertical
 - balance



Observation term

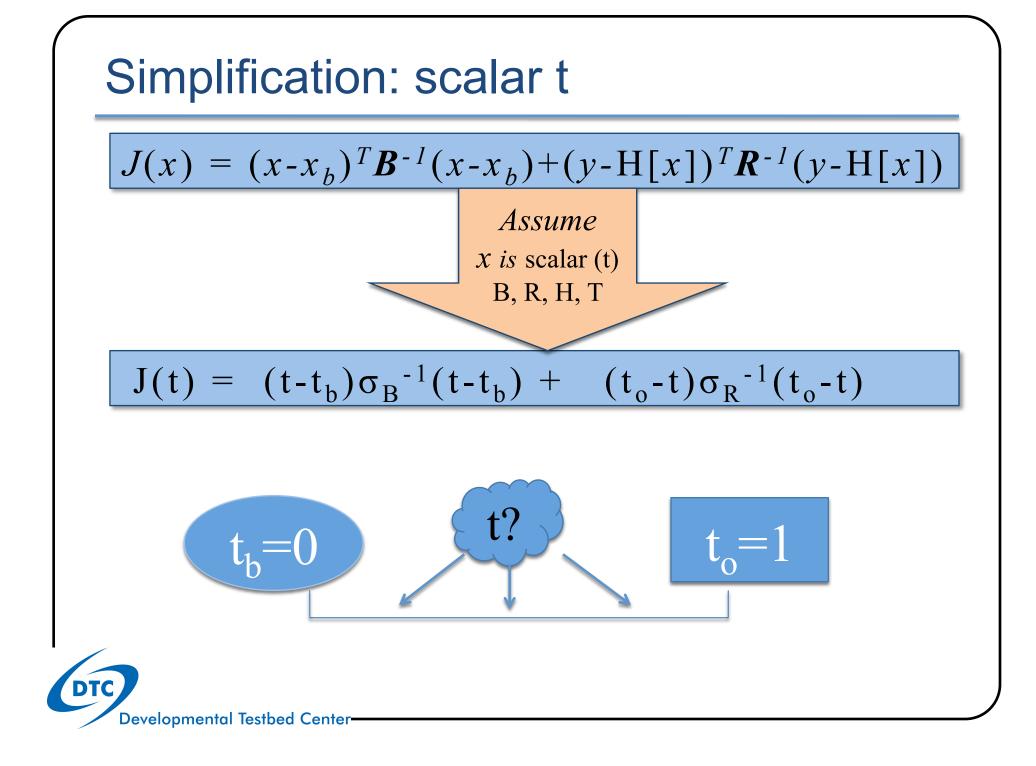
$$J(x) = (x - x_b)^T B^{-1} (x - x_b) + (y - H[x])^T R^{-1} (y - H[x])$$

- 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

Developmental Testbed Center

DTC

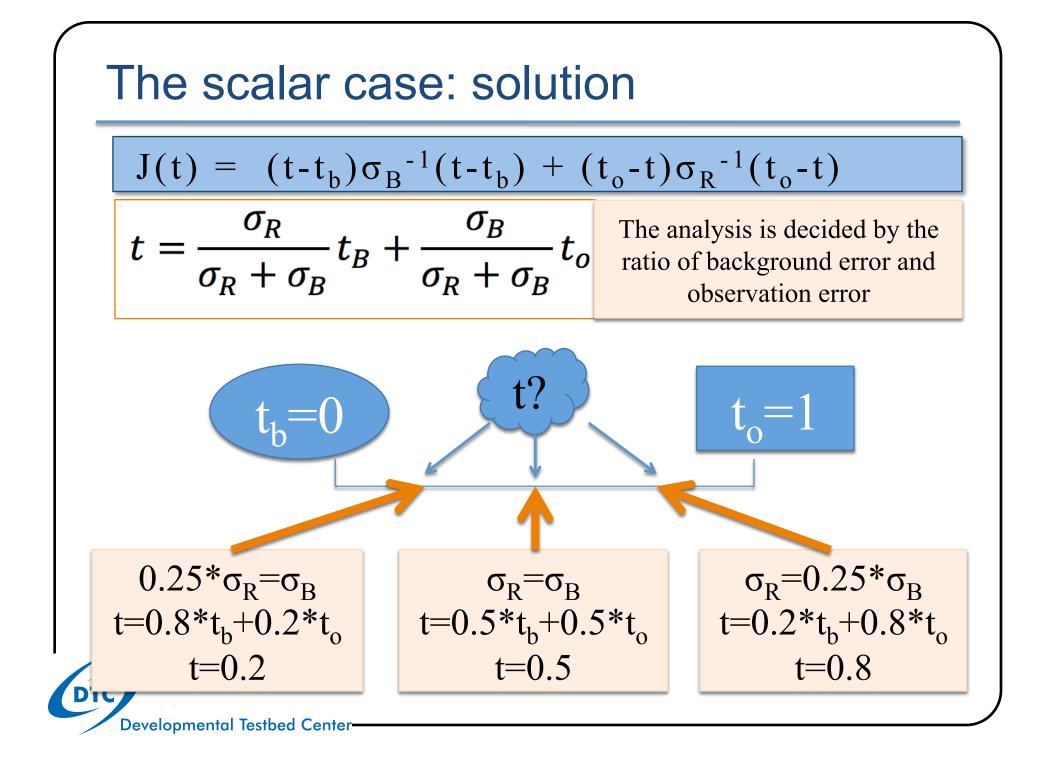


The scalar case: solve

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

$$J(t) = \frac{(t-t_B)^2}{\sigma_B} + \frac{(t_o-t)^2}{\sigma_B}$$
When $\frac{\partial J(t)}{\partial t} = 0$, J(t) is minimum and t is the best guess to truth

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



Background error covariance

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

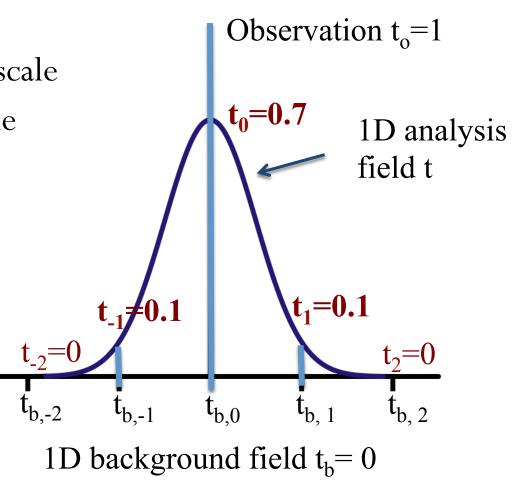
- Background error covariance: B
 - Variance
 - Ratio to the observation variance to decide how much analysis results fit to the observations
 - Correlation
 - Horizontal and vertical
 - Balance



Background error covariance: scale

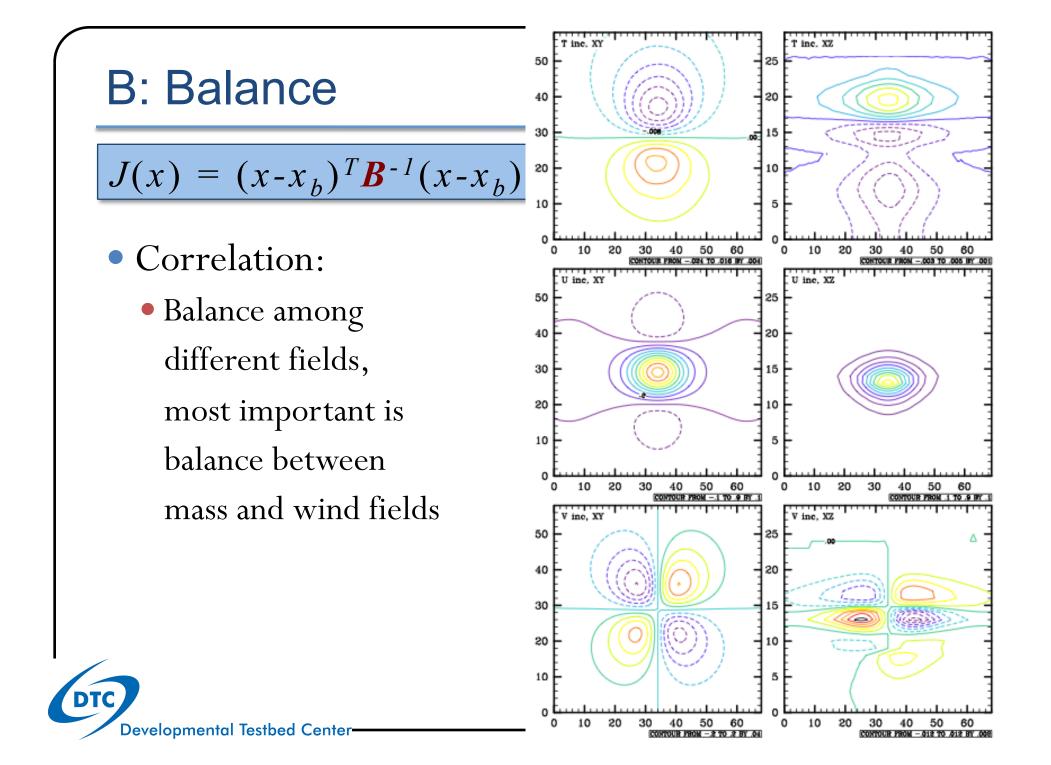
$$J(x) = (x - x_b)^T \mathbf{B}^{-1} (x - x_b) + (y - H[x])^T \mathbf{R}^{-1} (y - H[x])$$

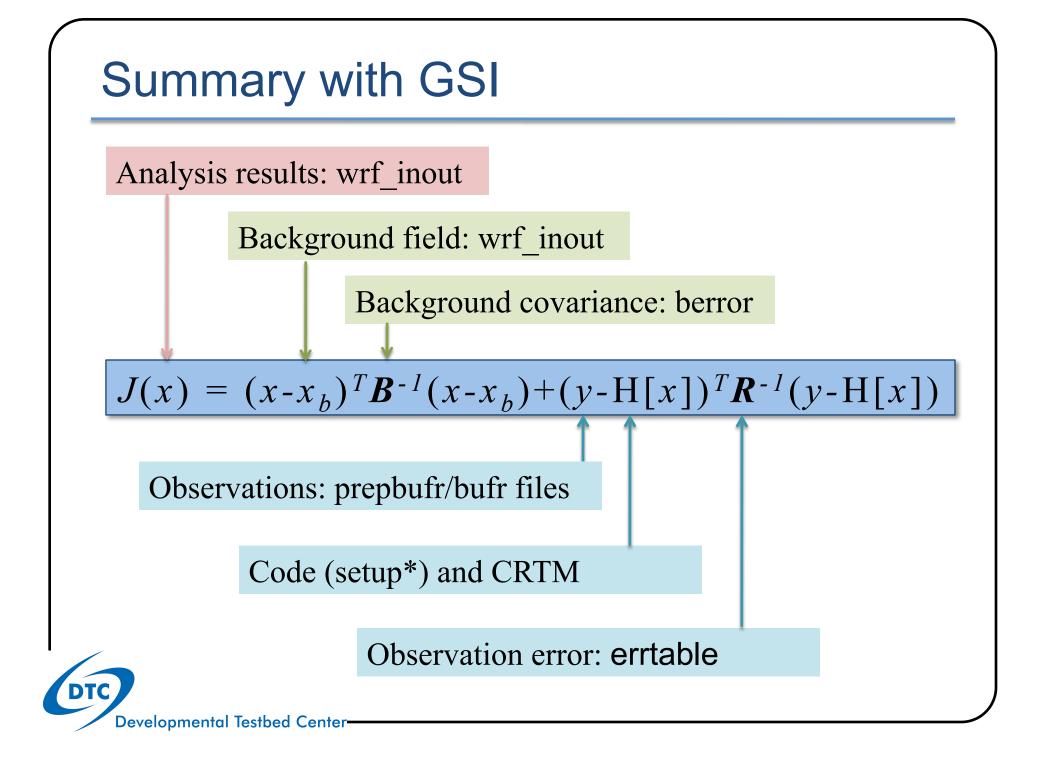
- Correlation:
 - Horizontal influence scale
 - Vertical influence scale
 - In Gaussian Shape





DTC





GSI History and Current Status

Based on John Derber's talk in 2012 summer GSI tutorial

- The Spectral Statistical Interpolation (SSI) analysis system was developed at NCEP in the late 1980's and early 1990's.
 - Originally called Spectral Optimal Interpolation(SOI) references still in code
- Main advantages of this system over OI systems were:
 - All observations are used at once (much of the noise generated in OI analyses was generated by data selection)
 - Ability to use forward models to transform from analysis variable to observations
 - Analysis variables can be defined to simplify covariance matrix and are not tied to model variables (except need to be able to transform to model variable)
- The SSI system was the first operational
 - variational analysis system
 - system to directly use radiances



- While the SSI system was a great improvement over the prior OI system – it still had some basic shortcomings
 - Since background error was defined in spectral space not simple to use for regional systems
 - Diagonal spectral background error did not allow much spatial variation in the background error
 - Not particularly well written since developed as a prototype code and then implemented operationally



- 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
 - 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



- Used in NCEP operations for
 - Regional
 - Global
 - Hurricane
 - Real-Time Mesoscale Analysis
 - Rapid Refresh (ESRL/GSD)
- GMAO collaboration
- Preparation for AFWA implementation
- Modification to fit into WRF and NCEP infrastructure
- Evolution to NEMS



General Comments

- GSI analysis code is an evolving system.
 - Scientific advances
 - situation dependent background errors -- hybrid
 - new satellite data
 - new analysis variables
 - Improved coding
 - Bug fixes
 - Removal of unnecessary computations, arrays, etc.
 - More efficient algorithms (MPI, OpenMP)
 - Bundle structure
 - Generalizations of code
 - Different compute platforms
 - Different analysis variables
 - Different models
 - Improved documentation
 - Removal of legacy options
 - Fast evolution creates difficulties for slower evolving research projects



DTC – Summer Tutorial

14 August 2012

Developmental Testbed Center

General Comments

- Code is intended to be used Operationally
 - Must satisfy coding requirements
 - Must fit into operational infrastructure
 - Must be kept as simple as possible
 - Must run fast enough and not use too many computer resources.
- External usage intended to:
 - Improve external testing
 - Transition research science into operations
 - Reduce transition time/effort to operations
 - Reduce duplication of effort



Simplification to operational 3-D for presentation

- For today's introduction, I will be talking about using 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
 - NSST analysis
 - Detailed options
- Options make code more complex difficult balance between options and simplicity



Analysis variables

- Background errors must be defined in terms of analysis variable
 - Streamfunction (Ψ)
 - Unbalanced Velocity Potential ($\chi_{unbalanced}$)
 - Unbalanced Temperature (T_{unbalanced})
 - Unbalanced Surface Pressure (Ps_{unbalanced})
 - Ozone Clouds etc.
 - Satellite bias correction coefficients
- Size of problem
 - NX x ÑY x NZ x NVAR
 - Global = ~130 million component control vector
 - Requires multi-tasking to fit on computers



Analysis variables

- $\chi = \chi_{\text{unbalanced}} + A \Psi$
- $T = T_{unbalanced} + B \Psi$
- $Ps = Ps_{unbalanced} + C \Psi$
- Streamfunction is a key variable defining a large percentage T and P_s (especially away from equator). Contribution to χ is small except near the surface and tropopause.



u,v

- Analysis variables are streamfunction and unbalanced velocity potential
- u,v needed for many routines (int,stp,balmod, etc.) on different domains
- u,v updated along with other variables by calculating derivatives of streamfunction and velocity potential components of search direction x and creating a dir x (u,v)



Background fields

- Current works for following systems
 - NCEP GFS NEMS, GFSIO and spectral coefficients
 - NCEP NMM binary and netcdf
 - NCEP RTMA
 - NCEP Hurricane
 - GMAO global
 - ARW binary and netcdf
- FGAT (First Guess at Appropriate Time) enabled up to 100 time levels



Background Errors

- Three paths
 - Isotropic/homogeneous
 - Most common usage.
 - Function of latitude/height
 - Vertical and horizontal scales separable
 - Variances can be location dependent
 - See talk by Syed Rizvi
 - Anisotropic/inhomogeneous
 - Function of location / state
 - Can be full 3-D covariances
 - Still relatively immature
 - Hybrid
 - Dual resolution
 - Operational in global
 - See talk by J. Whitaker



Observations

- Observational data is expected to be in BUFR format (this is the international standard)
- See presentation by Ruifang Li
- Each observation type (e.g., u,v,radiance from NOAA-15 AMSU-A) is read in on a particular processor or group of processors (parallel read)
- Data thinning can occur in the reading step.
- Checks to see if data is in specified data time window and within analysis domain



Data processing

- Data used in GSI controlled 2 ways
 - Presence or lack of input file
 - Control files input (info files) into analysis
 - Allows data to be monitored rather than used
 - Each ob type different
 - Specify different time windows for each ob type
 - Intelligent thinning distance specification



Input data – Satellite currently used

Regional **GOES-13** Sounder Channels 1-15 Individual fields of view 4 Detectors treated separately Over ocean only Thinned to 120km AMSU-A NOAA-15 Channels 1-10, 12-13, 15 NOAA-18 Channels 1-8, 10-13, 15 Channels 1-7, 9-13, 15 NOAA-19 **METOP** Channels1-6, 8-13, 15 Thinned to 60km AMSU-B/MHS Channels 1-5 NOAA-18 Channels 1-5 **METOP** Thinned to 60km HIRS NOAA-19 Channels 2-15 **METOP** Channels 2-15 Thinned to 120km AIRS 148 Channels AOUA Thinned to 120km IASI **METOP** 165 Channels DTC – Summer Tutorial

Global all thinned to 145km

Geo Sounders/imagers GOES-13 and 15 Sounders Channels 1-15 Individual fields of view 4 Detectors treated separately Over ocean only **SEVIRI Clear Sky Radiances** Channels 2-3 AMSU-A NOAA-15 Channels 1-10, 12-13, 15 NOAA-18 Channels 1-8, 10-13, 15 Channels 1-7, 9-13, 15 NOAA-19 Channels 1-6, 8-13, 15 METOP Channels 6, 8-13 AOUA ATMS NPP Channels 1-14, 16-22 AMSU-B/MHS Channels 1-5 NOAA-19 Channels 1-5 NOAA-18 METOP Channels 1-5 HIRS NOAA-19 Channels 2-15 METOP Channels 2-15 AIRS 148 Channels AQUA IASI **METOP** 165 Channels

14 August 2012

Developmental Testbed Center

Input data – Conventional currently used

- Radiosondes
- Pibal winds
- Synthetic tropical cyclone winds
- wind profilers
- conventional aircraft reports
- ASDAR aircraft reports
- MDCARS aircraft reports
- dropsondes
- MODIS IR and water vapor winds
- GMS, JMA, METEOSAT and GOES cloud drift IR and visible winds
- GOES water vapor cloud top winds

- Surface land observations
- Surface ship and buoy observation
- SSM/I wind speeds
- QuikScat and ASCATwind speed and direction
- SSM/I and TRMM TMI precipitation estimates
- Doppler radial velocities
- VAD (NEXRAD) winds
- GPS precipitable water estimates
- GPS Radio occultation refractivity and bending angle profiles
- SBUV ozone profiles and OMI total ozone



Simulation of observations

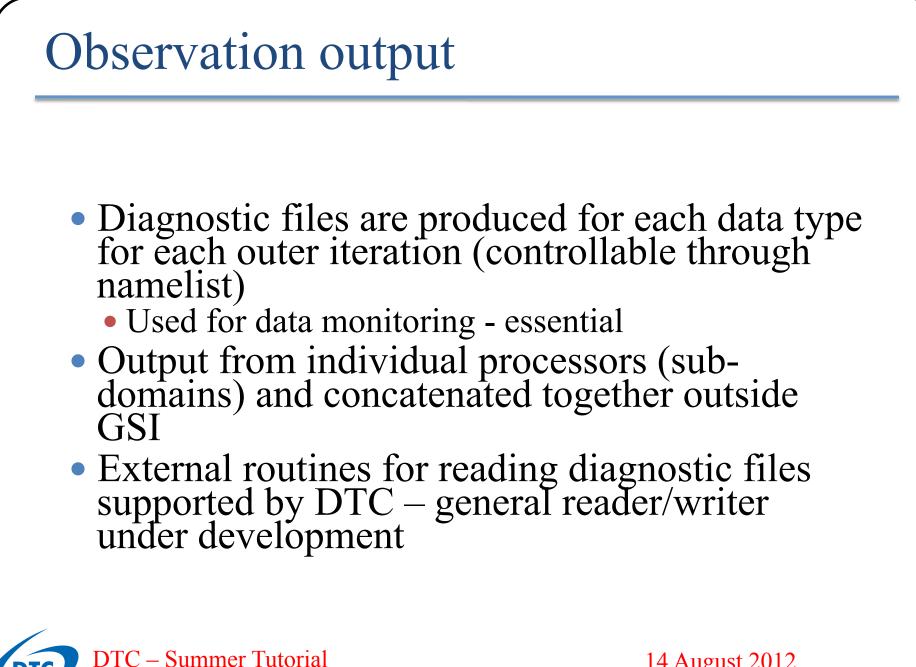
- To use observation, must be able to simulate observation
 - Can be simple interpolation to ob location/time
 - Can be more complex (e.g., interpolation plus radiative transfer)
- For radiances we use CRTM
 - Vertical resolution and model top important



Quality control

- External platform specific QC
- Some gross checking in PREPBUFR file creation
- Analysis QC
 - Gross checks specified in input data files
 - Variational quality control
 - Data usage specification (info files)
 - Outer iteration structure allows data rejected (or downweighted) initially to come back in
 - Ob error can be modified due to external QC marks
 - Radiance QC much more complicated. Andrew Collard!





14 August 2012

Developmental Testbed Center

Community GSI

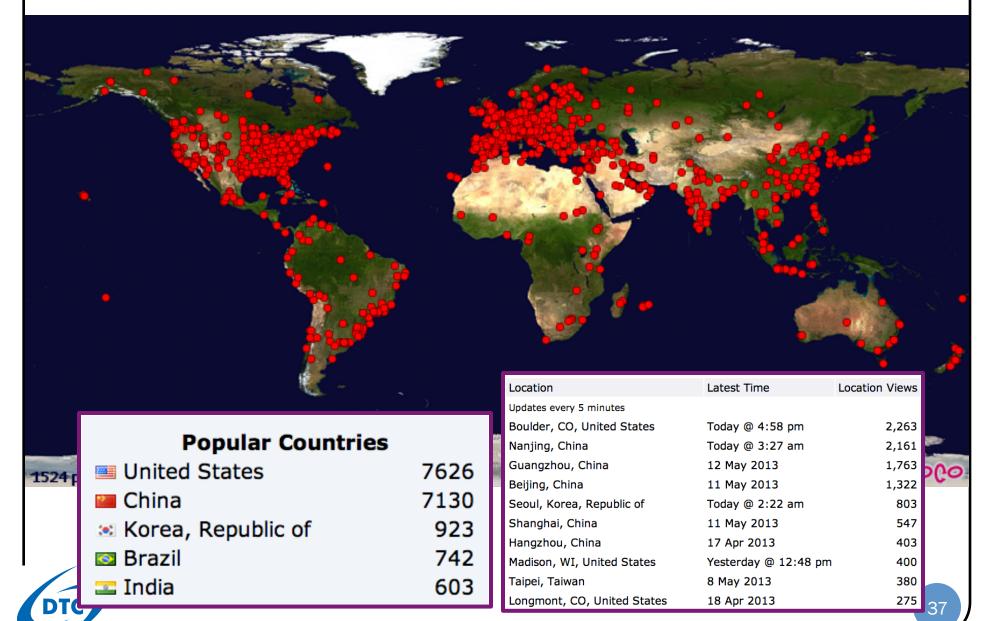
Community GSI

- 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 (over 700 registered users)

Developmental Testbed Center

Community GSI Website Visitors



Developmental Testbed Center

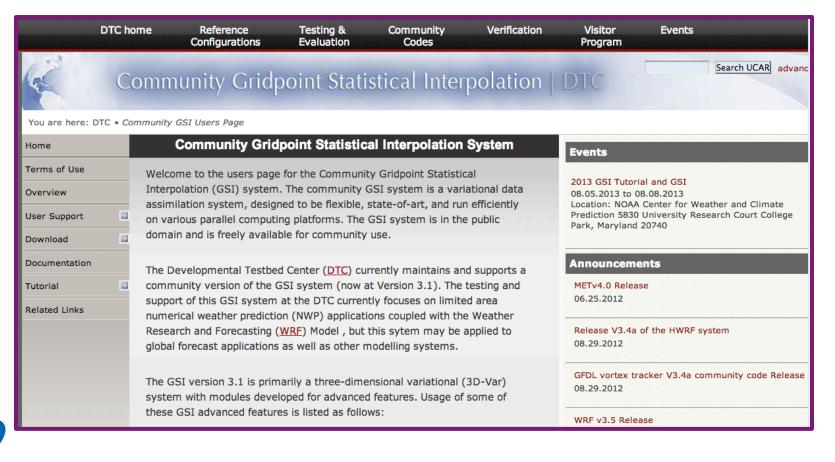
"History"

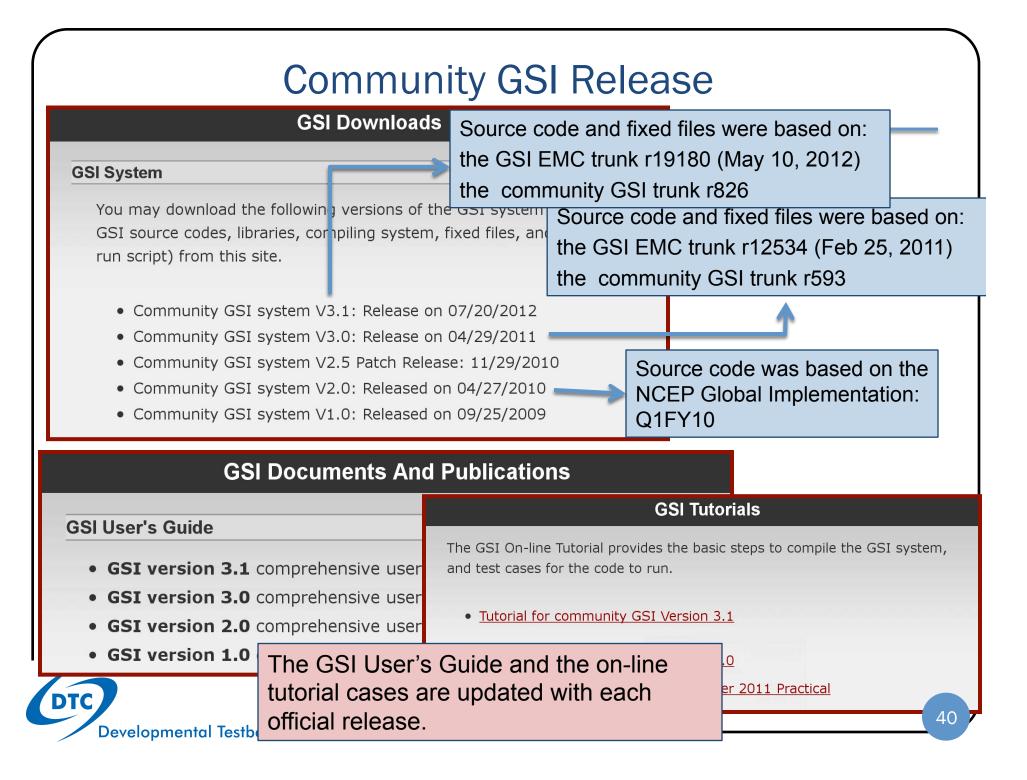
- June, 2006:North American Mesoscale (NAM) System, NCEP
- May, 2007: Global Forecast System (GFS), NCEP
- 2007: DTC started to document GSI.
- 2009:
 - Created GSI code repository over NCEP and DTC
 - First GSI release V1.0
 - Started user support
- 2010:
 - First Community GSI tutorial
 - Created GSI Review Committee
- 2011
 - First Community GSI workshop



Community GSI – User's Page

• Mainly support through User's Page and help desk: <u>http://www.dtcenter.org/com-GSI/users/index.php</u>





Community GSI - Documents

Hon

Terr Ove

Dow

Doc

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

1	GSI Documents And Publications
of Use	GSI User's Guide
ew	GSI Üser's Guide
	 GSI version 3.1 comprehensive user documentation [pdf]
upport	 GSI version 3.0 comprehensive user documentation [pdf]; [ps]
ad	 GSI version 2.0 comprehensive user documentation [pdf]
entation	 GSI version 1.0 comprehensive user documentation [pdf]
al i	DTC Community GSI Workshop Presentations
d Links	Summer 2011 DTC Community GSI Workshop
	(June 28: NCAR, Boulder CO) [presentations], [photos]
	DTC Community GSI Tutorial Presentations
	 2013 Community GSI Tutorial Beijing
	(May 28 to 29: Beijing China) [presentations]
	Summer 2012 DTC Community GSI Tutorial
	(August 21 to 23: NCAR, Boulder CO) [presentations], [photos]
	 Summer 2011 DTC Community GSI Tutorial
	(June 29 to July 1: NCAR, Boulder CO) [presentations], [photos]
	 Summer 2010 DTC Community GSI Tutorial
	(NCAR, Boulder CO) [presentations], [photos]
	GSI Technique doumentation
	 Technical Report on the Development of the GSI-based WRF
	4DVAR by Xin Zhang & Xiang-Yu Huang [pdf]
	 Generation of WRF-AW Background Errors (BE) for GSI: by Syed
	RH Rizvi, Zhiquan Liu & Xiang-Yu Huang [pdf]

Developmental Testbed Center

DTC

Community GSI - Practice

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

Home	GSI Tutorials				
Terms of Use	The GSI On-line Tutorial provides the basic steps to compile the GSI system,				
Overview	and test cases for the code to run.				
User Support 🛛 🗳	Tutorial for community GSI Version 3.1				
Download 🛛	Further exercises from the Summer 2012 Practical				
Documentation Tutorial	 <u>Tutorial for community GSI Version 3.0</u> <u>Further exercises from the Summer 2011 Practical</u> 				
	 <u>Tutorial for community GSI Version 2.0</u> <u>Further exercises from the Summer 2010 Practical</u> 				
	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.				

Basic Practical Cases

0). BUFR/PrepBUFR tools [practice]

To run any of the following practical cases, start with the basic **run_gsi.ksh**, found in the **comGSI_v3.1/run** directory, and m modifications to it. These changes are used by all the basic exe specific to using the COMET classroom. If you are running the elsewhere, additional modifications will be necessary.

1). Single Observation Tests:

- [a]. ARW background with global BE
- [b]. ARW background with NAM BE
- [c]. NMM background with global BE
- [d]. <u>NMM background with NAM BE</u>

2). Test with conventional data (prepbufr):

- [a]. <u>ARW background</u>
- [b]. <u>NMM background</u>

3). Test with conventional and satellite radiance data:

- [a]. ARW background
- [b]. <u>NMM background</u>

4). Test with conventional and gpsro data:

- [a]. ARW background
- [b]. <u>NMM background</u>

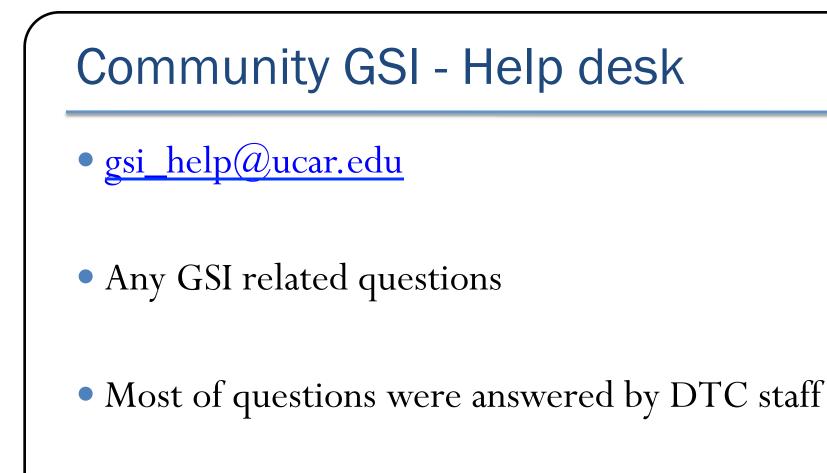
5). Test with conventional data and radar radial wind:

- [a]. <u>ARW background</u>
- [b]. <u>NMM background</u>

Community GSI - tutorial

- 2010 summer tutorial (28-30 June, 2010)
 - 14 lectures and 8-h practice session
- 2011 summer tutorial (28-30 June, 2012)
 - Lectures + basic practice
 - Optional advanced practice (Full day on 30 June)
- The 1st GSI Workshop on 27 June, 2011
- 2012 summer tutorial (21-23 August, 2012)
- 2013 summer tutorial (5-7 August, 2013 at NCEP)
- The 2nd GSI workshop on 8 August, 2013 at NCEP
- 2013 Beijing tutorial: 1st oversea

Developmental Testbed Center



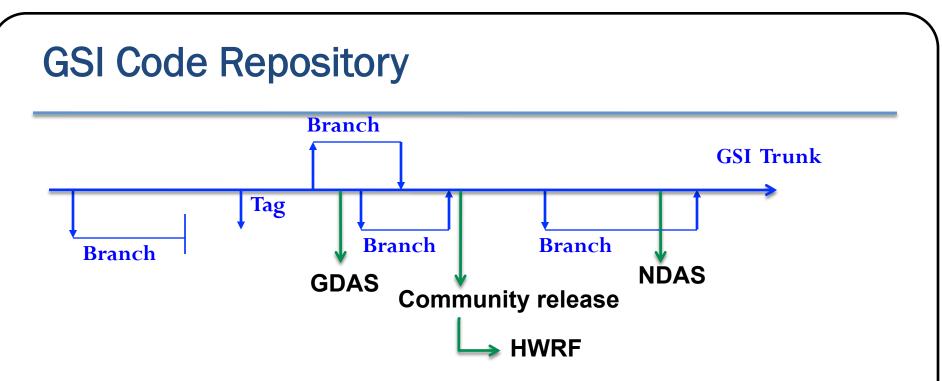
• Forward complex questions to NCEP colleague



GSI Developers

- DTC supports GSI developers through:
 - Maintaining community GSI repository
 - Organizing GSI review committee
 - Testing and reviewing the new code
 - Organizing monthly community GSI developer's meeting
 - Help desk
- R2O: Developers need to:
 - Identify the work that will contribute to operation
 - Send development proposal to the committee early
 - Develop the new code based on GSI trunk

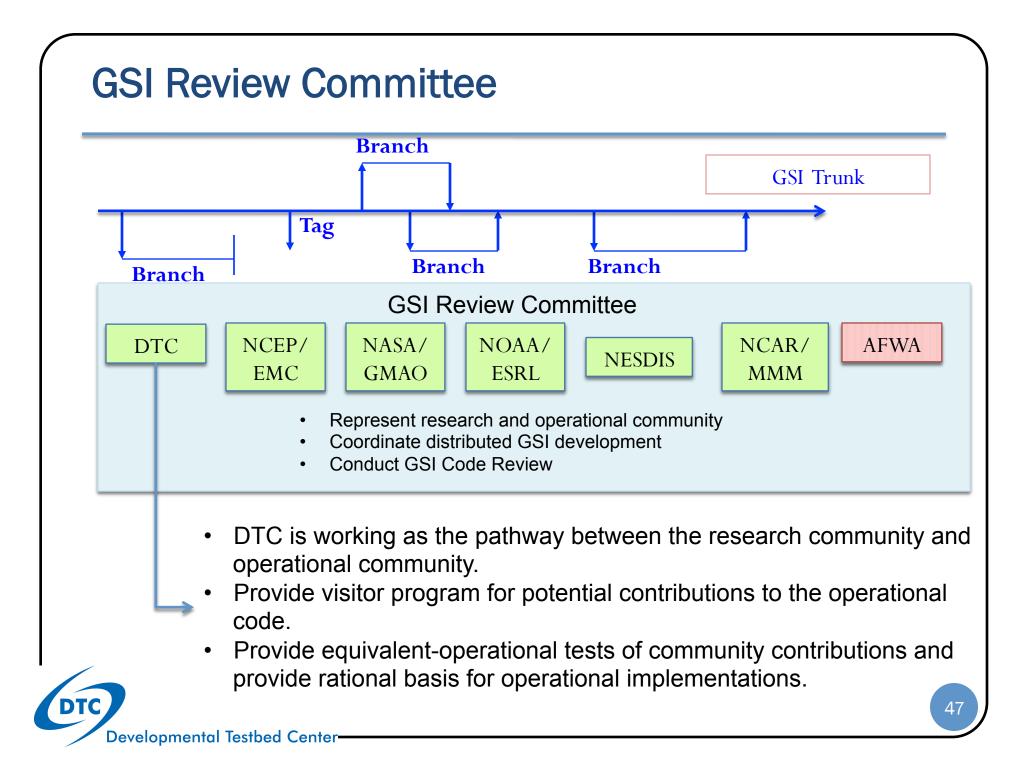




- Use tags or branches for: Release, new development, bug fix ...
- Applications may use different revisions in the trunk ("snapshot").
- ✓ Which GSI should I use ?

There is no "DTC GSI", "EMC GSI" or "global GSI". There is only one GSI! For a researcher, community release should be sufficient to use. If you are interested in getting new development back to the GSI trunk, contact GSI helpdesk (gsi_help@ucar.edu) get access to the developmental version of GSI.





Tutorial Agenda: Lectures and Practices

O \cdots i								
Overview and introductions (<i>Tue</i> Welcome, Background and Participants' Int Fundamentals of Data Assimilation	•	Group Picture: Tuesday Morning Break, 10:30 am, August 21, 2012						
Introduction to Practice Service	Community GSI fundamentals and tools (Tuesday-Wednesday)							
Practice Session : COMET room (right behind the front desk at the main entrance)	GSI Fundament GSI Fundament GSI Fundament GSI Fundament Community Too	ntals (1): Setup and Compilation: <i>Tuesday</i> ntals (2): Run and Namelist: <i>Tuesday</i> ntals (3): Diagnostics: <i>Wednesday</i> ntals (4): Applications: <i>Wednesday</i> ools (1): PrepBUFR/BUFR tools: <i>Wednesday</i> ools (2): GEN_BE: <i>Wednesday</i>						
Advanced topics (Wednesday-Thursday)								
GPSRO Data Assimilation: Wednesday Satellite Radiance Assimilation: Wednesday GSI Infrastructures: Thursday GSI Hybrid Data Assimilation: Thursday		on Practices						
Cloudy Radiance Data Assimilation: Thursd	Practice	Practice session: <i>Tuesday and Wednesday afternoon</i> Optional Practice Session: <i>Thursday</i> 48						

Tutorial Lecturers and Practical Session Instructors

NCEP/EMC	NASA/GMAO	NOAA/ESRL	NCAR/MMM	NCAR/RAL
John Derber Andrew Collard Lidia Cucurull** ** UCAR	Ricardo Todling	Jeff Whitaker Ming Hu*	Tom Augline Syed Rizvi Ruifang Li	Hui Shao* Don Stark* Kathryn Newman* Chunhua Zhou*

* Developmental Testbed Center (DTC)

Ming Hu, Hui Shao, Don Stark, Kathryn Newman, Chunhua Zhou Pamela Johnson



Summary

- DTC works with NCEP/EMC to provide current operation GSI capability to research community (O2R)
- DTC also works with researchers to bring research community contributions back to the GSI operation repository (O2R)

• Send your questions to gsi_help@ucar.edu

Developmental Testbed Center