

Overview of GSI

John C. Derber
NOAA/NWS/NCEP/EMC



28 June 2010



DTC – Summer Tutorial



History

- The Spectral Statistical Interpolation (SSI) analysis system was developed at NCEP in the late 1980's and early 1990's.
- 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



28 June 2010



DTC – Summer Tutorial



History

- While the SSI system was a great improvement over the prior OI system – it still had some basic short-comings
 - 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



28 June 2010



DTC – Summer Tutorial



History

- The Global 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



28 June 2010



DTC – Summer Tutorial



History

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



28 June 2010



DTC – Summer Tutorial



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
 - Fast evolution creates difficulties for slower evolving research projects



28 June 2010



DTC – Summer Tutorial



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



28 June 2010



DTC – Summer Tutorial



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
 - Detailed options
- Options make code more complex – difficult balance between options and simplicity



28 June 2010



DTC – Summer Tutorial



Basic analysis problem

$$\mathbf{J} = \mathbf{J}_b + \mathbf{J}_o + \mathbf{J}_c$$

$$\mathbf{J} = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{H}(\mathbf{x}) - \mathbf{y}_0)^T (\mathbf{E} + \mathbf{F})^{-1} (\mathbf{H}(\mathbf{x}) - \mathbf{y}_0) + \mathbf{J}_C$$

\mathbf{J} = Fit to background + Fit to observations + constraints

\mathbf{x} = Analysis

\mathbf{x}_b = Background

\mathbf{B} = Background error covariance

\mathbf{H} = Forward model

\mathbf{y}_0 = Observations

$\mathbf{E} + \mathbf{F} = \mathbf{R}$ = Instrument error + Representativeness error

\mathbf{J}_C = Constraint terms



28 June 2010



DTC – Summer Tutorial



Jc term

- Currently Jc term includes 2 terms
 - Weak moisture constraint ($q > 0, q < q_{\text{sat}}$)
 - Can substantially slow convergence if coefficient made too large.
 - Conservation of global dry mass
 - not applicable to regional problem



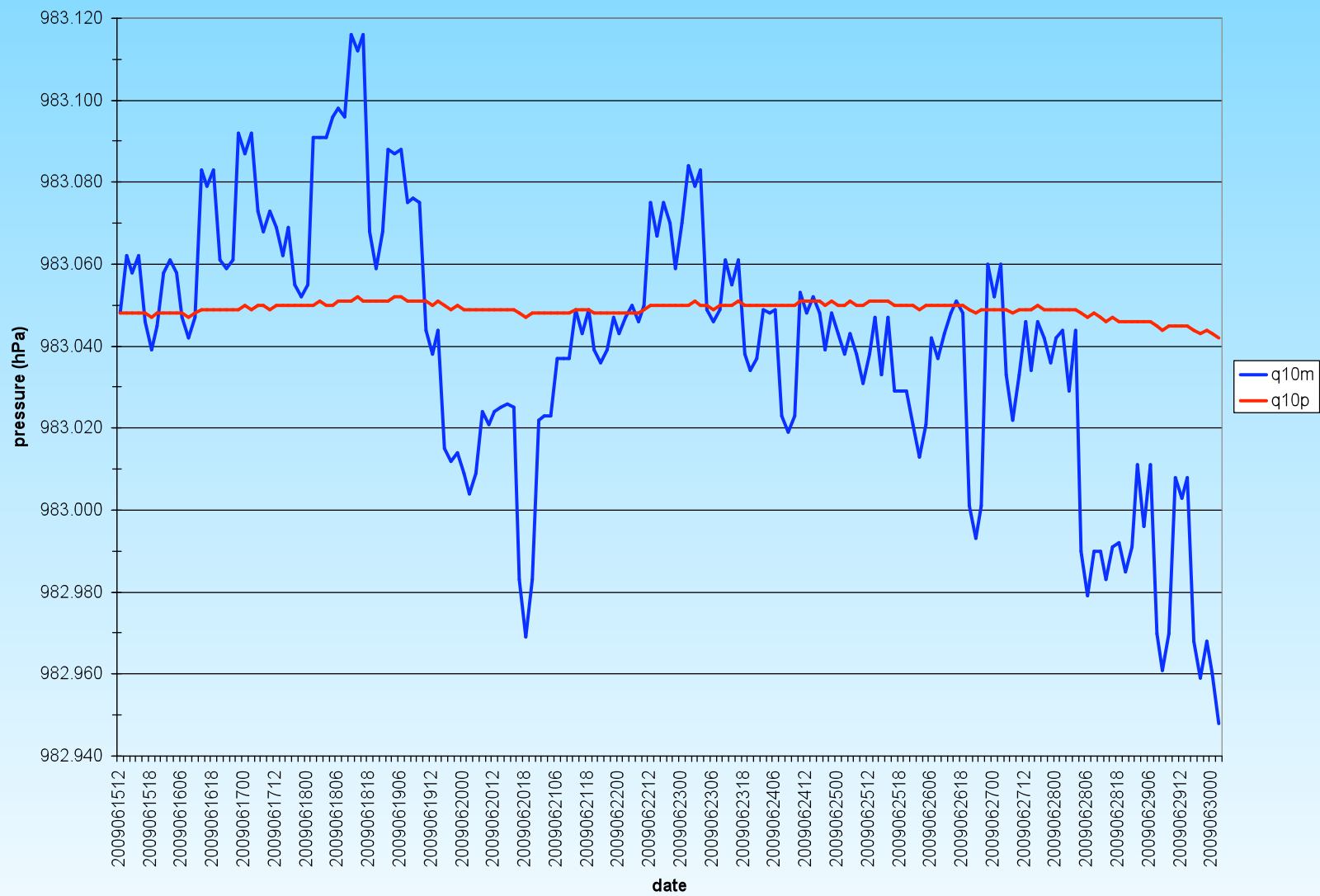
28 June 2010



DTC – Summer Tutorial



global mean_pdry



28 June 2010



DTC – Summer Tutorial



Solution

At minimum, $\text{Grad } J = 0$, Note this is a necessary condition – it is sufficient only for a quadratic J

$$\text{Grad } J = 2B^{-1}(\mathbf{x}-\mathbf{x}_b) + H^T(E+F)^{-1}(H(\mathbf{x})-y_0) + \text{Grad } J_C$$

A conjugate gradient minimization algorithm is used to solve for $\text{Grad } J = 0$



28 June 2010



DTC – Summer Tutorial



Solution Strategy

- Solve series of simpler problems with some nonlinear components eliminated
- Outer iteration, inner iteration structure
 - $x = x_{\text{outer iteration}} + x_{\text{inner iteration}} + x_b$
- Outer iteration
 - QC
 - More complete forward model
- Inner iteration
 - Preconditioned conjugate gradient
 - Estimate search direction
 - Estimate optimal stepsize in search direction
 - Often simpler forward model
 - Variational QC
 - Solution used to start next outer iteration



28 June 2010



DTC – Summer Tutorial



Inner iteration – algorithm 1

- $J = \mathbf{x}^T \mathbf{B}^{-1} \mathbf{x} + (\mathbf{H}\mathbf{x}-\mathbf{o})^T \mathbf{O}^{-1} (\mathbf{H}\mathbf{x}-\mathbf{o})$ (assume linear)
- define $\mathbf{y} = \mathbf{B}^{-1} \mathbf{x}$
- $J = \mathbf{x}^T \mathbf{y} + (\mathbf{H}\mathbf{x}-\mathbf{o})^T \mathbf{O}^{-1} (\mathbf{H}\mathbf{x}-\mathbf{o})$
- $\text{Grad } J_{\mathbf{x}} = \mathbf{B}^{-1} \mathbf{x} + \mathbf{H}^T \mathbf{O}^{-1} (\mathbf{H}\mathbf{x}-\mathbf{o}) = \mathbf{y} + \mathbf{H}^T \mathbf{O}^{-1} (\mathbf{H}\mathbf{x}-\mathbf{o})$
- $\text{Grad } J_{\mathbf{y}} = \mathbf{x} + \mathbf{B} \mathbf{H}^T \mathbf{O}^{-1} (\mathbf{H}\mathbf{x}-\mathbf{o}) = \mathbf{B} \text{Grad } J_{\mathbf{x}}$
- Solve for both \mathbf{x} and \mathbf{y} using preconditioned conjugate gradient (where the \mathbf{x} solution is preconditioned by \mathbf{B} and the solution for \mathbf{y} is preconditioned by \mathbf{B}^{-1})



28 June 2010



DTC – Summer Tutorial



Inner iteration - algorithm

Specific algorithm

$$x^0=y^0=0$$

Iterate over n

$$\text{Grad } x^n = y^{n-1} + H^T O^{-1}(Hx^{n-1} - o)$$

$$\text{Grad } y^n = B \text{ Grad } x^n$$

$$\text{Dir } x^n = \text{Grad } y^n + \beta \text{ Dir } x^{n-1}$$

$$\text{Dir } y^n = \text{Grad } x^n + \beta \text{ Dir } y^{n-1}$$

$$x^n = x^{n-1} + \alpha \text{ Dir } x^n \quad (\text{Update } x_{\text{save}} \text{ (outer iter. } x) \text{ - as well)}$$

$$y^n = y^{n-1} + \alpha \text{ Dir } y^n \quad (\text{Update } y_{\text{save}} \text{ (outer iter. } y) \text{ - as well)}$$

Until max iteration or gradient sufficiently minimized



28 June 2010



DTC – Summer Tutorial



Inner iteration – algorithm 2

- $J = \mathbf{x}^T \mathbf{B}^{-1} \mathbf{x} + (\mathbf{H}\mathbf{x} - \mathbf{o})^T \mathbf{O}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{o})$ (assume linear)
- define $\mathbf{y} = \mathbf{B}^{-1/2} \mathbf{x}$
- $J = \mathbf{y}^T \mathbf{y} + (\mathbf{H}\mathbf{B}^{1/2} \mathbf{y} - \mathbf{o})^T \mathbf{O}^{-1} (\mathbf{H}\mathbf{B}^{1/2} \mathbf{y} - \mathbf{o})$
- $\text{Grad } J_{\mathbf{y}} = \mathbf{y} + \mathbf{B}^{1/2} \mathbf{H}^T \mathbf{O}^{-1} (\mathbf{H}\mathbf{B}^{1/2} \mathbf{y} - \mathbf{o})$
- Solve for \mathbf{y} using preconditioned conjugate gradient
- For our definition of the background error matrix, $\mathbf{B}^{1/2}$ is not square and thus \mathbf{y} is (3x) larger than \mathbf{x} .



28 June 2010



DTC – Summer Tutorial



Inner iteration - algorithm

- intall routine calculate $H^T O^{-1}(Hx-o)$
- bkerror routines multiplies by B
- dprod x calculates β and magnitude of gradient
- stpcalc calculates stepsize



28 June 2010



DTC – Summer Tutorial



Inner iteration – algorithm

Estimation of α (the stepsize)

- The stepsize is estimated through estimating the ratio of contributions for each term

$$\alpha = \sum a / \sum b$$

- The a's and b's can be estimated exactly for the linear terms.
- For nonlinear terms, the a's and b's are estimated by fitting a quadratic using 3 points around an estimate of the stepsize
- The estimate for the nonlinear terms is re-estimated iteratively using the stepsize for the previous estimate (up to 5 iterations)



28 June 2010



DTC – Summer Tutorial



Analysis variables

- Background errors must be defined in terms of analysis variable
 - Streamfunction (Ψ)
 - Unbalanced Velocity Potential ($\chi_{\text{unbalanced}}$)
 - Unbalanced Temperature ($T_{\text{unbalanced}}$)
 - Unbalanced Surface Pressure ($P_{\text{S}_{\text{unbalanced}}}$)
 - Ozone – Clouds – etc.
 - Satellite bias correction coefficients



28 June 2010



DTC – Summer Tutorial



Analysis variables

- $\chi = \chi_{\text{unbalanced}} + A \Psi$
- $T = T_{\text{unbalanced}} + B \Psi$
- $P_s = P_{s_{\text{unbalanced}}} + C \Psi$
- 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.



28 June 2010



DTC – Summer Tutorial



Analysis variables

- A, B and C matrices can involve 2 components
 - A pre-specified statistical balance relationship – part of the background error statistics file
 - Optionally a incremental normal model balance
 - Not working well for regional problem
 - See references for details



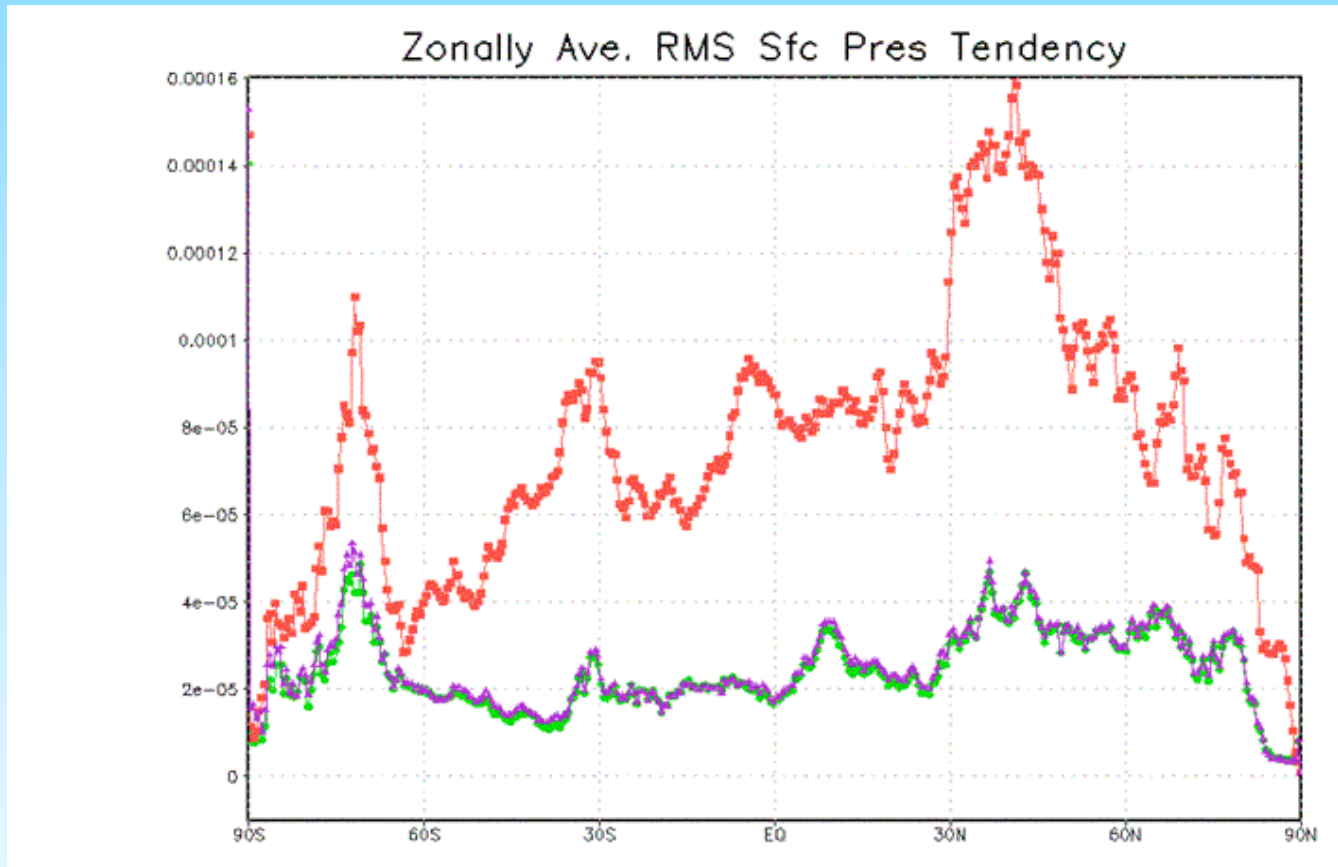
28 June 2010



DTC – Summer Tutorial



Impact of TLNM constraint



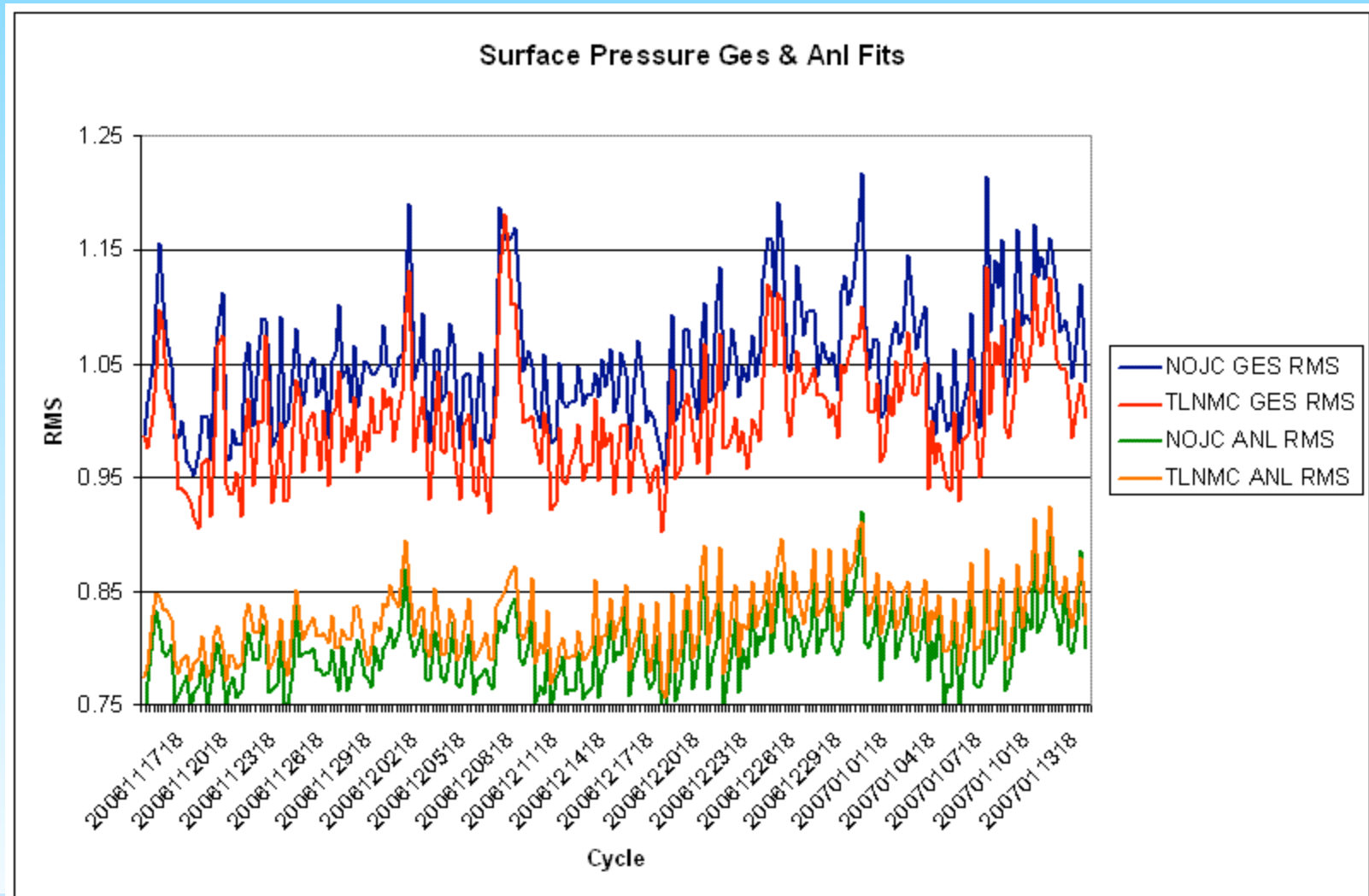
Zonal-average surface pressure tendency for background (green), unconstrained GSI analysis (red), and GSI analysis with TLNMC (purple).

28 June 2010

DTC – Summer Tutorial



Fits of Surface Pressure Data in Cycled Experiment with and without TLNM constraint



28 June 2010



DTC – Summer Tutorial



Analysis variables

- Size of problem
 - $NX \times NY \times NZ \times NVAR$
 - Global = 25.7 million component control vector
 - Requires multi-tasking to fit on computers



28 June 2010



DTC – Summer Tutorial



Grid Sub-domains

- The analysis and background fields are divided across the processors in two different ways
 - Sub-Domains – an x-y region of the analysis domain with full vertical extent – observations defined on sub-domains
 - Horizontal slabs – a single or multiple levels of full x-y fields
- Since the analysis problem is a full 3-D problem – we must transform between these decompositions repeatedly



28 June 2010



DTC – Summer Tutorial



u,v

- Analysis variables are streamfunction and velocity potential
- u,v needed for many routines (int,stp,balmod, etc.) routines
- 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)



28 June 2010



DTC – Summer Tutorial



Background fields

- Current works for following systems
 - NCEP GFS
 - NCEP NMM – binary and netcdf
 - NCEP RTMA
 - NCEP Hurricane (not using subversion version yet)
 - GMAO global
 - ARW – binary and netcdf – (not operationally used – so not fully tested by NCEP)
- FGAT (First Guess at Appropriate Time) enabled up to 100 time levels



28 June 2010



DTC – Summer Tutorial



Background Errors

- Two paths – more in talk by S. Rizvi
 - Isotropic/homogeneous
 - Most common usage.
 - Function of latitude/height
 - Vertical and horizontal scales separable
 - Variances can be location dependent
 - Anisotropic/inhomogeneous
 - Function of location /state
 - Can be full 3-D covariances
 - Still relatively immature



28 June 2010



DTC – Summer Tutorial



Observations

- Observational data is expected to be in BUFR format (this is the international standard)
- See presentation by Stacy Bender
- 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



28 June 2010



DTC – Summer Tutorial



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



28 June 2010



DTC – Summer Tutorial



Input data – Satellite currently used

- Regional

- GOES-11 and 12 Sounders
 - 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
 - METOP Channels 1-6, 8-13, 15
 - Thinned to 60km
- AMSU-B/MHS
 - NOAA-15 Channels 1-3, 5
 - NOAA-18 Channels 1-5
 - METOP Channels 1-5
 - Thinned to 60km
- HIRS
 - NOAA-17 Channels 2-15
 - METOP Channels 2-15
 - Thinned to 120km
- AIRS
 - AQUA 148 Channels
 - Thinned to 120km

- Global

all thinned to 145km

- GOES-11 and 12 Sounders
 - Channels 1-15
 - Individual fields of view
 - 4 Detectors treated separately
 - Over ocean only
- AMSU-A
 - NOAA-15 Channels 1-10, 12-13, 15
 - NOAA-18 Channels 1-8, 10-13, 15
 - NOAA-19 Channels 1-7, 9-13, 15
 - METOP Channels 1-6, 8-13, 15
 - AQUA Channels 6, 8-13
- AMSU-B/MHS
 - NOAA-15 Channels 1-3, 5
 - NOAA-18 Channels 1-5
 - METOP Channels 1-5
- HIRS
 - NOAA-17 Channels 2-15
 - NOAA-19 Channels 2-15
 - METOP Channels 2-15
- AIRS
 - AQUA 148 Channels
- IASI
 - METOP 165 Channels



28 June 2010



DTC – Summer Tutorial



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, 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 wind speed and direction
- SSM/I precipitable water
- SSM/I and TRMM TMI precipitation estimates
- Doppler radial velocities
- VAD (NEXRAD) winds
- GPS precipitable water estimates
- GPS Radio occultation refractivity profiles
- SBUV ozone profiles (other ozone data under test)



28 June 2010



DTC – Summer Tutorial



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., radiative transfer)
- For radiances we use CRTM
 - Vertical resolution and model top important



28 June 2010



DTC – Summer Tutorial



Atmospheric analysis problem (Practical)

Outer (K) and Inner (L) iteration operators

Variable	K operator	L operator
Temperature – surface obs. at 2m	3-D sigma interpolation adjustment to different orography	3-D sigma interpolation Below bottom sigma assumed at bottom sigma
Wind – surface obs. at 10m over land, 20m over ocean, except scatt.	3-D sigma interpolation reduction below bottom level using model factor	3-D sigma interpolation reduction below bottom level using model factor
Ozone – used as layers	Integrated layers from forecast model	Integrated layers from forecast model
Surface pressure	2-D interpolation plus orography correction	2-D interpolation
Precipitation	Full model physics	Linearized model physics
Radiances	Full radiative transfer	Linearized radiative transfer

Data Sub-domains

- Observations are distributed to processors they are used on. Comparison to obs are done on sub-domains.
 - If an observation is on boundary of multiple sub-domains will be put into all relevant sub-domains for communication free adjoint calculations.
 - However, it is necessary to assign the observation only to one sub-domain for the objective function calculation
 - Interpolation of sub-domain boundary observations requires the use of halo rows around each sub-domain



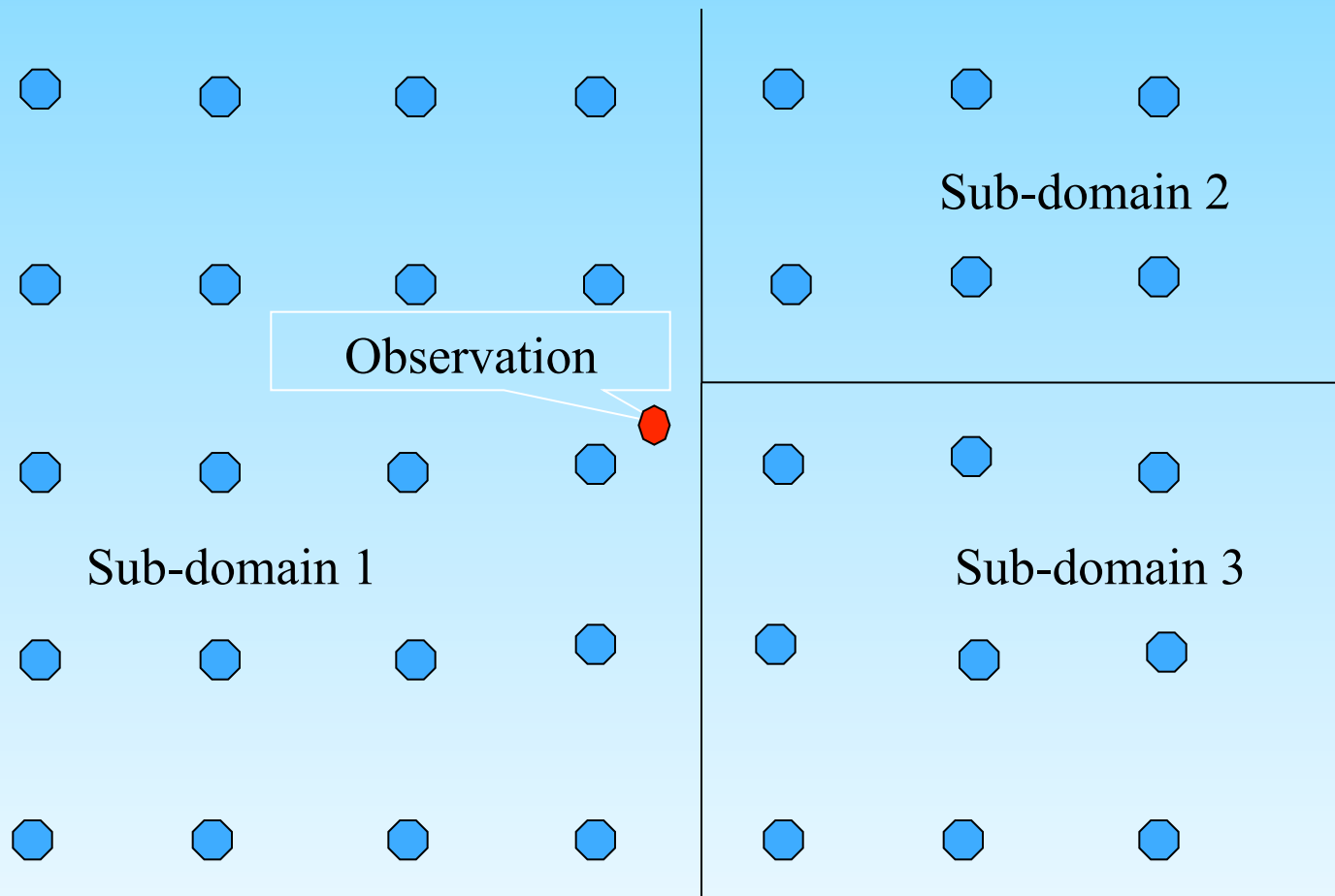
28 June 2010



DTC – Summer Tutorial



Observation/Sub-domain layout



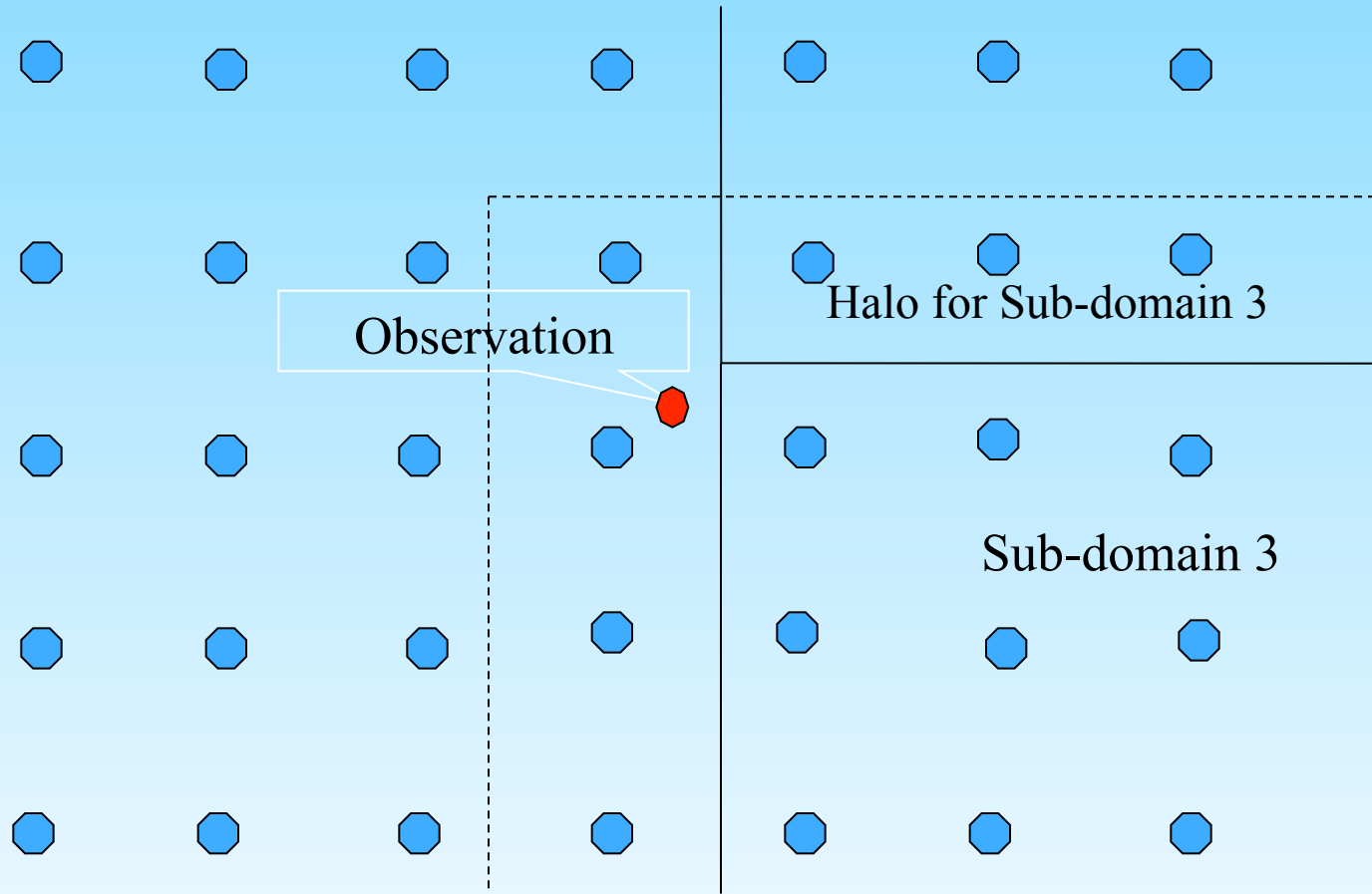
28 June 2010



DTC – Summer Tutorial



Sub-domain 3 calculation w/halo



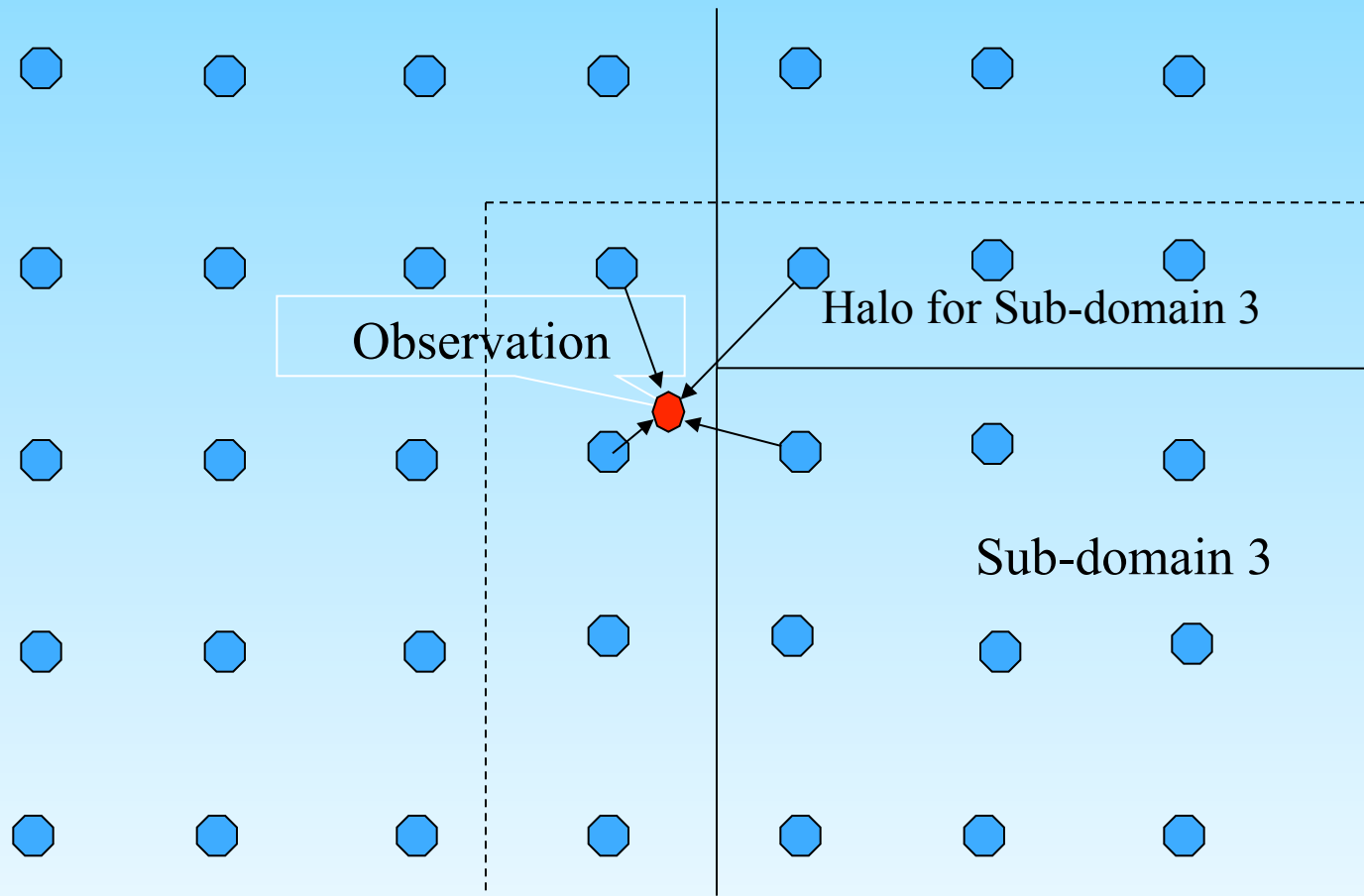
28 June 2010



DTC – Summer Tutorial



Forward interpolation to ob.



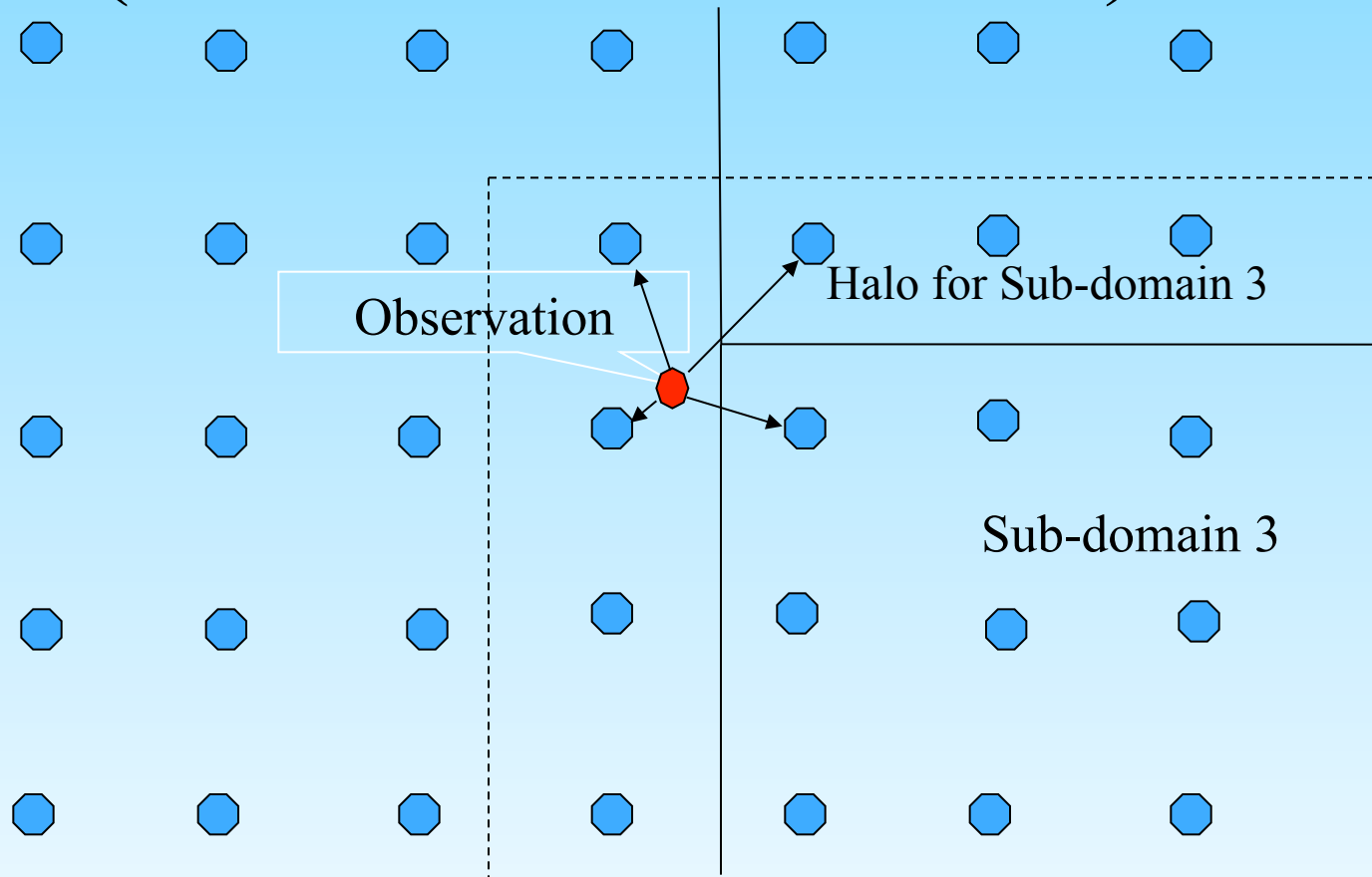
28 June 2010



DTC – Summer Tutorial



Adjoint of interpolation to grid (values in halo not used)



28 June 2010



DTC – Summer Tutorial



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. Tomorrow!



28 June 2010



DTC – Summer Tutorial



Observation output

- Diagnostic files are produced for each data type for each outer iteration (controllable through namelist)
- Output from individual processors (sub-domains) and concatenated together outside GSI
- External routines for reading diagnostic files should be supported by DTC



28 June 2010



DTC – Summer Tutorial



GSI layout (major routines) (generic names, 3dvar path)

- gsimain (main code)
 - gsimain_initialize (read in namelists and initialize variables)
 - gsimain_run
 - gsisub
 - deter_subdomain (creates sub-domains)
 - *read_info (reads info files to determine data usage)
 - glbsoi
 - » observer_init (read background field)
 - » observer_set (read observations and distribute)
 - » prewgt (initializes background error)
 - » setuprhsall (calculates outer loop obs. increments)
 - » pcgsoi or sqrtmin (solves inner iteration)
 - gsimain_finalize (clean up arrays and finalize mpi)



28 June 2010



DTC – Summer Tutorial



GSI layout (major routines)

- pcgsoi or sqrtmin
 - control2state (convert control vector to state vector)
 - intall (compare to observations and adjoint)
 - state2control (convert state vector to control vector)
 - bkerror (multiply by background error)
 - stpcalc (estimate stepsize and update solution)
 - update_guess (updates outer iteration solution)
 - write_all (write solution)



28 June 2010



DTC – Summer Tutorial



Challenges

- Negative Moisture and other tracers
- Diabatic analysis
- Hurricane initialization
- Advanced assimilation
 - Situation dependent background errors
 - Hybrid assimilation
 - 4d-var
- Use of satellite radiances in regional mode
- Use of satellite data over land/ice/snow
- AQ and constituent assimilation
- Improved bias correction
- New instruments – SSM/IS, NPP/JPSS, research satellites



28 June 2010



DTC – Summer Tutorial



Useful References

- Wan-Shu Wu, R. James Purser and David F. Parrish, 2002: **Three-Dimensional Variational Analysis with Spatially Inhomogeneous Covariances**. *Monthly Weather Review*, Vol. 130, No. 12, pp. 2905–2916.
- R. James Purser, Wan-Shu Wu, David F. Parrish and Nigel M. Roberts, 2003: **Numerical Aspects of the Application of Recursive Filters to Variational Statistical Analysis. Part I: Spatially Homogeneous and Isotropic Gaussian Covariances**. *Monthly Weather Review*, Vol. 131, No. 8, pp. 1524–1535.
- R. James Purser, Wan-Shu Wu, David F. Parrish and Nigel M. Roberts, 2003: **Numerical Aspects of the Application of Recursive Filters to Variational Statistical Analysis. Part II: Spatially Inhomogeneous and Anisotropic General Covariances**. *Monthly Weather Review*, Vol. 131, No. 8, pp. 1536–1548.
- Parrish, D. F. and J. C. Derber, 1992: **The National Meteorological Center's spectral statistical interpolation analysis system**. *Mon. Wea. Rev.*, 120, 1747 - 1763.
- Kleist, Daryl T; Parrish, David F; Derber, John C; Treadon, Russ; Wu, Wan-Shu; Lord, Stephen , **Introduction of the GSI into the NCEP Global Data Assimilation System**, *Weather and Forecasting*. Vol. 24, no. 6, pp. 1691-1705. Dec 2009
- Kleist, Daryl T; Parrish, David F; Derber, John C; Treadon, Russ; Errico, Ronald M; Yang, Runhua, **Improving Incremental Balance in the GSI 3DVAR Analysis System**, *Monthly Weather Review [Mon. Weather Rev.]*. Vol. 137, no. 3, pp. 1046-1060. Mar 2009..
- Zhu, Y; Gelaro, R, **Observation Sensitivity Calculations Using the Adjoint of the Gridpoint Statistical Interpolation (GSI) Analysis System**, *Monthly Weather Review*. Vol. 136, no. 1, pp. 335-351. Jan 2008.
- [DTC GSI documentation \(http://www.dtcenter.org/com-GSI/users/index.php\)](http://www.dtcenter.org/com-GSI/users/index.php)



28 June 2010



DTC – Summer Tutorial

