Overview of GSI

John C. Derber
NOAA/NWS/NCEP/EMC

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

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

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

History

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

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
  - Fast evolution creates difficulties for slower evolving research projects

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

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

Basic analysis problem

$J = J_b + J_o + J_c$

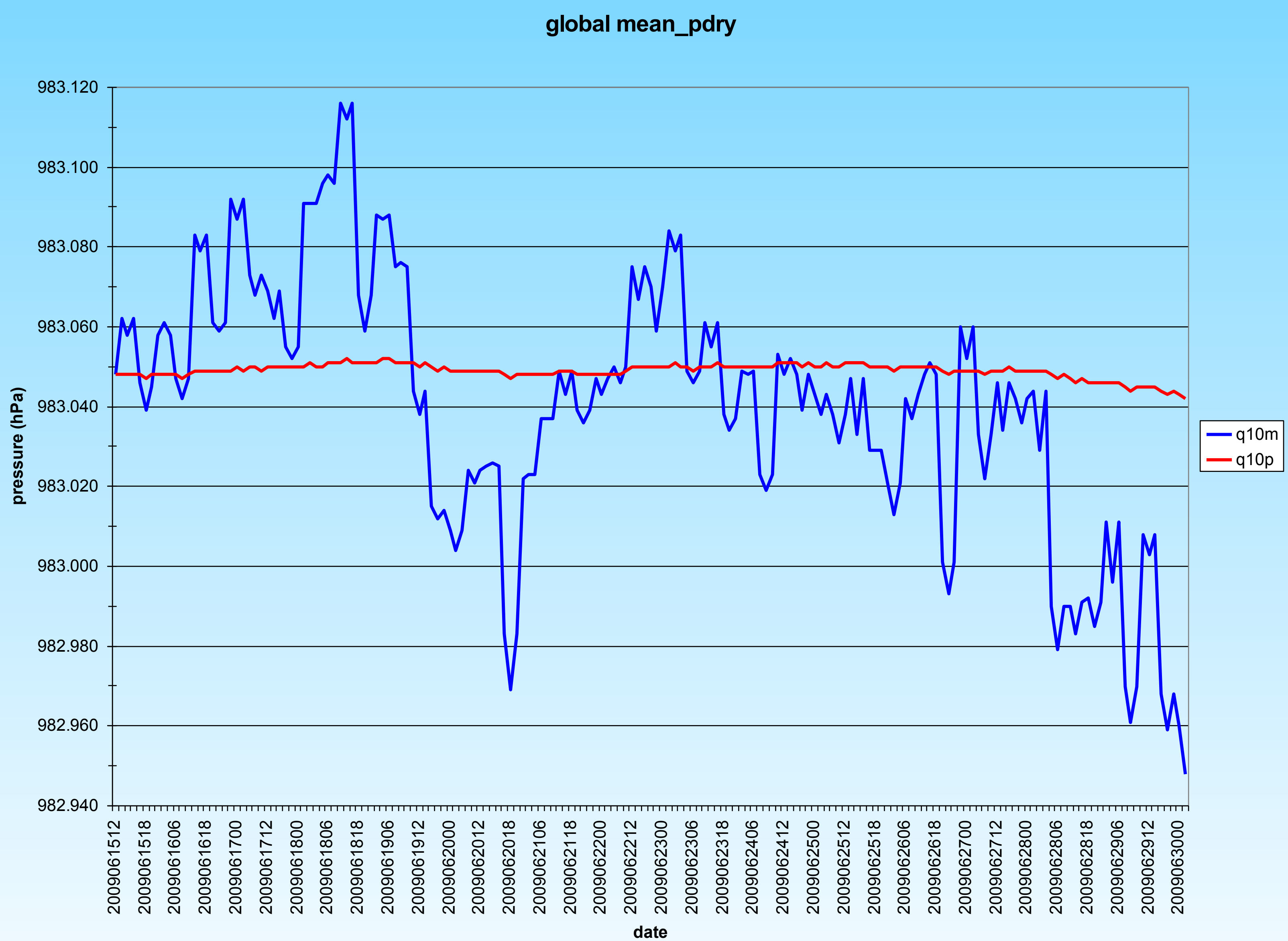
$J = (x\text{-}x_b)^T B^{-1}(x\text{-}x_b) + (H(x)\text{-}y_0)^T(E+F)^{-1}(H(x)\text{-}y_0) + J_C$

J = Fit to background + Fit to observations + constraints

x = Analysis
x_b = Background
B = Background error covariance
H = Forward model
y_0 = Observations
$E+F$ = R = Instrument error + Representativeness error
J_C = Constraint terms

Jc term

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



global mean_pdry
pressure (hPa)
983.120
983.100
983.080
983.060
983.040
983.020
983.000
982.980
982.960
982.940
date
2009061512
2009061518
2009061606
2009061618
2009061700
2009061712
2009061800
2009061806
2009061818
2009061906
2009061912
2009062000
2009062012
2009062018
2009062106
2009062118
2009062200
2009062212
2009062300
2009062306
2009062318
2009062406
2009062412
2009062500
2009062512
2009062518
2009062606
2009062618
2009062700
2009062712
2009062800
2009062806
2009062818
2009062906
2009062912
2009063000
q10m
q10p

Solution

At minimum, Grad J = 0, Note this is a necessary condition – it is sufficient only for a quadratic J

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

A conjugate gradient minimization algorithm is used to solve for Grad J = 0

Solution Strategy

- Solve series of simpler problems with some nonlinear components eliminated
- Outer iteration, inner iteration structure
  - $x = x_{outer\ iteration} + x_{inner\ iteration} + x_b$
- Outer iteration
  - QC
  - More complete forward model
- Inner iteration
  - Several different minimization options – preconditioned Conjugate Gradient (algorithm 1)
    - Estimate search direction
    - Estimate optimal stepsize in search direction
  - Often simpler forward model
  - Variational QC
  - Solution used to start next outer iteration

Inner iteration – algorithm 1

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

Inner iteration – algorithm 1

Specific algorithm

$x^0 = y^0 = 0$

Iterate over n

- Grad $x^n = y^{n-1} + H^T O^{-1}(Hx^{n-1} - o)$
- Grad y^n = B Grad x^n
- Dir x^n = Grad $y^n + \beta$ Dir x^{n-1}
- Dir y^n = Grad $x^n + \beta$ Dir y^{n-1}
- $x^n = x^{n-1} + \alpha$ Dir x^n (Update xhatsave (outer iter. x) - as well)
- $y^n = y^{n-1} + \alpha$ Dir y^n (Update yhatsave (outer iter. y) - as well)

Until max iteration or gradient sufficiently minimized

Inner iteration – algorithm 2

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

Inner iteration – algorithm 1

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

Inner iteration – algorithm 1 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)

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

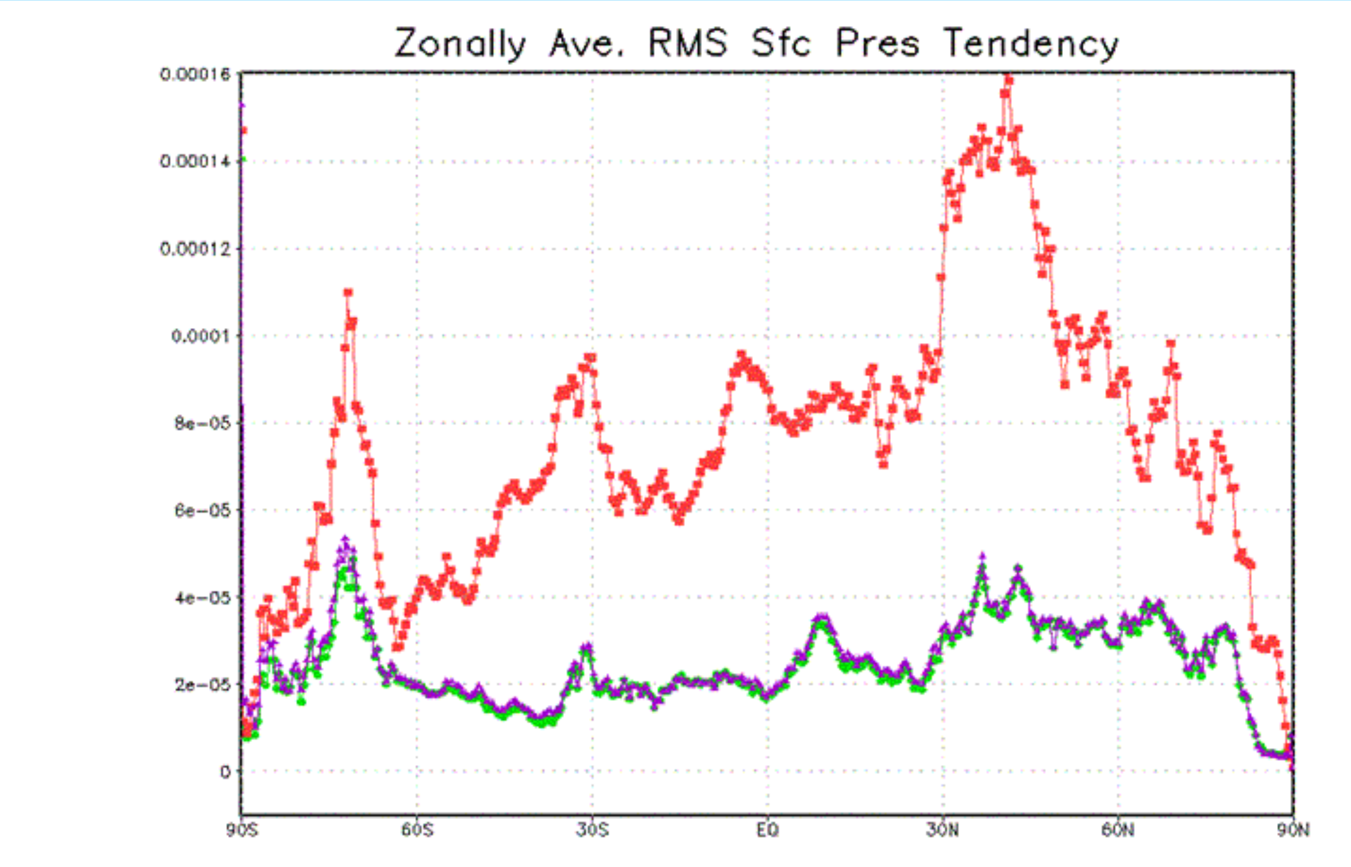
Analysis variables

- $\chi = \chi_{unbalanced} + A\,\Psi$
- $T = T_{unbalanced} + B\,\Psi$
- $Ps = Ps_{unbalanced} + C\,\Psi$
- 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.

Analysis variables

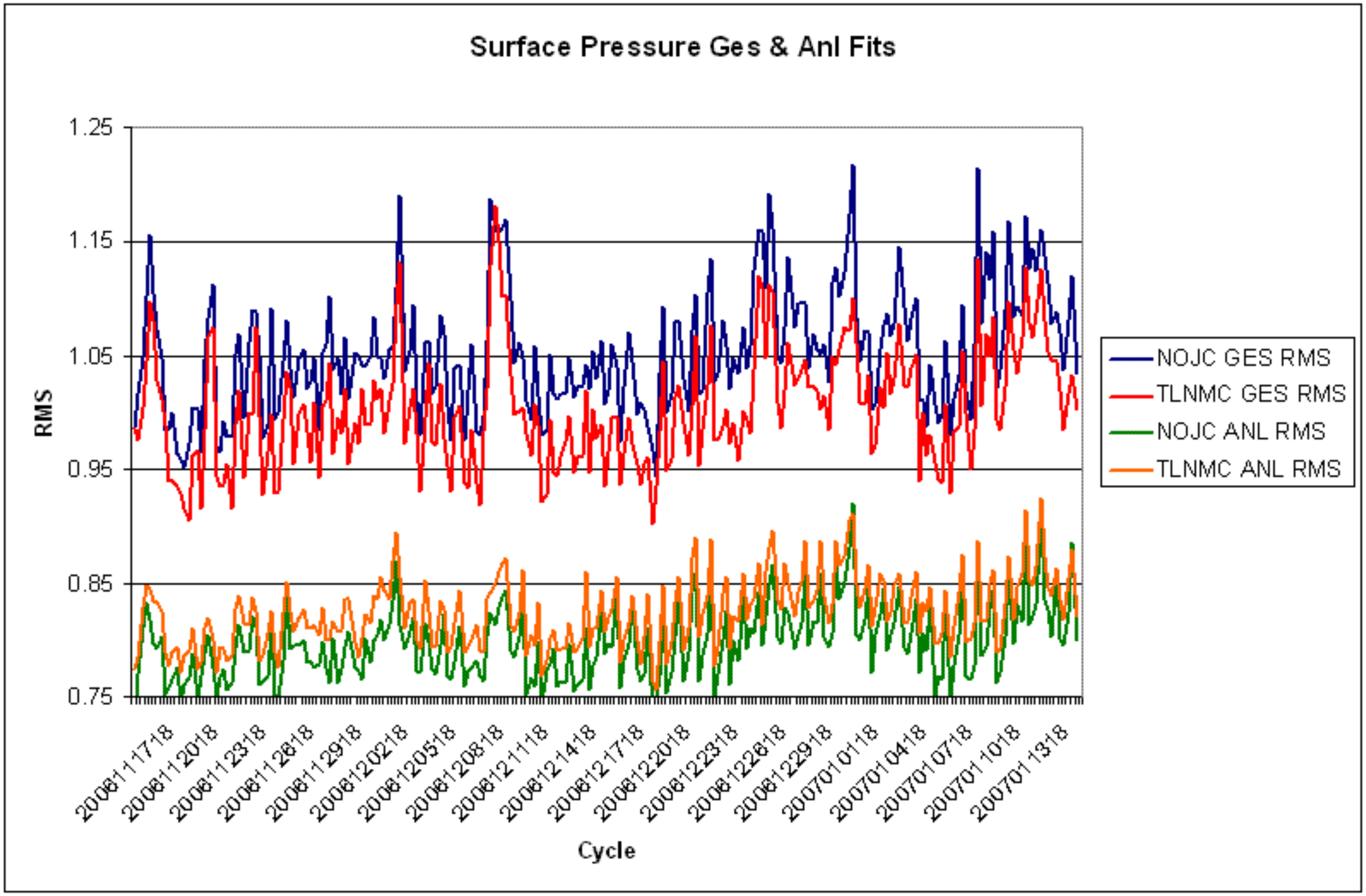
- A, B and C matrices can involve 2 components
  - A pre-specified statistical balance relationship – part of the background error statistics file
  - Optionally a incremental normal model balance
    - Not working well for regional problem
    - See references for details

Impact of TLNM constraint



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

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



Analysis variables

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

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

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)

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 yet RR - GSD)
- FGAT (First Guess at Appropriate Time) enabled up to 100 time levels

Background Errors

- Three paths – more in talk by D. Kleist
  - 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
  - Hybrid

Observations

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

AMSU-A		
	NOAA-15	Channels 1-10, 12-13, 15
	NOAA-18	Channels 1-8, 10-13, 15
	METOP-A	Channels1-6, 8-13, 15
	AQUA	Channels 5, 8-13
	Thinned to 60km	
AMSU-B/MHS		
	NOAA-15	Channels 1-3, 5
	NOAA-18	Channels 1-5
	METOP-A	Channels 1-5
	Thinned to 60km	
HIRS		
	NOAA-17	Channels 2-15
	METOP-A	Channels 2-15
	Thinned to 120km	
AIRS		
	AQUA	148 Channels
	Thinned to 120km	
IASI		
	METOP-A	165 Channels

- Global

all thinned to 145km

GOES-15 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-A	Channels 1-6, 8-13, 15
	AQUA	Channels 6, 8-13
ATMS		
	NPP	Channels 1-14,16-22
MHS		
	NOAA-18	Channels 1-5
	NOAA-19	Channels 1-5
	METOP-A	Channels 1-5
HIRS		
	METOP-A	Channels 2-15
AIRS		
	AQUA	148 Channels
IASI		
	METOP-A	165 Channels

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

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

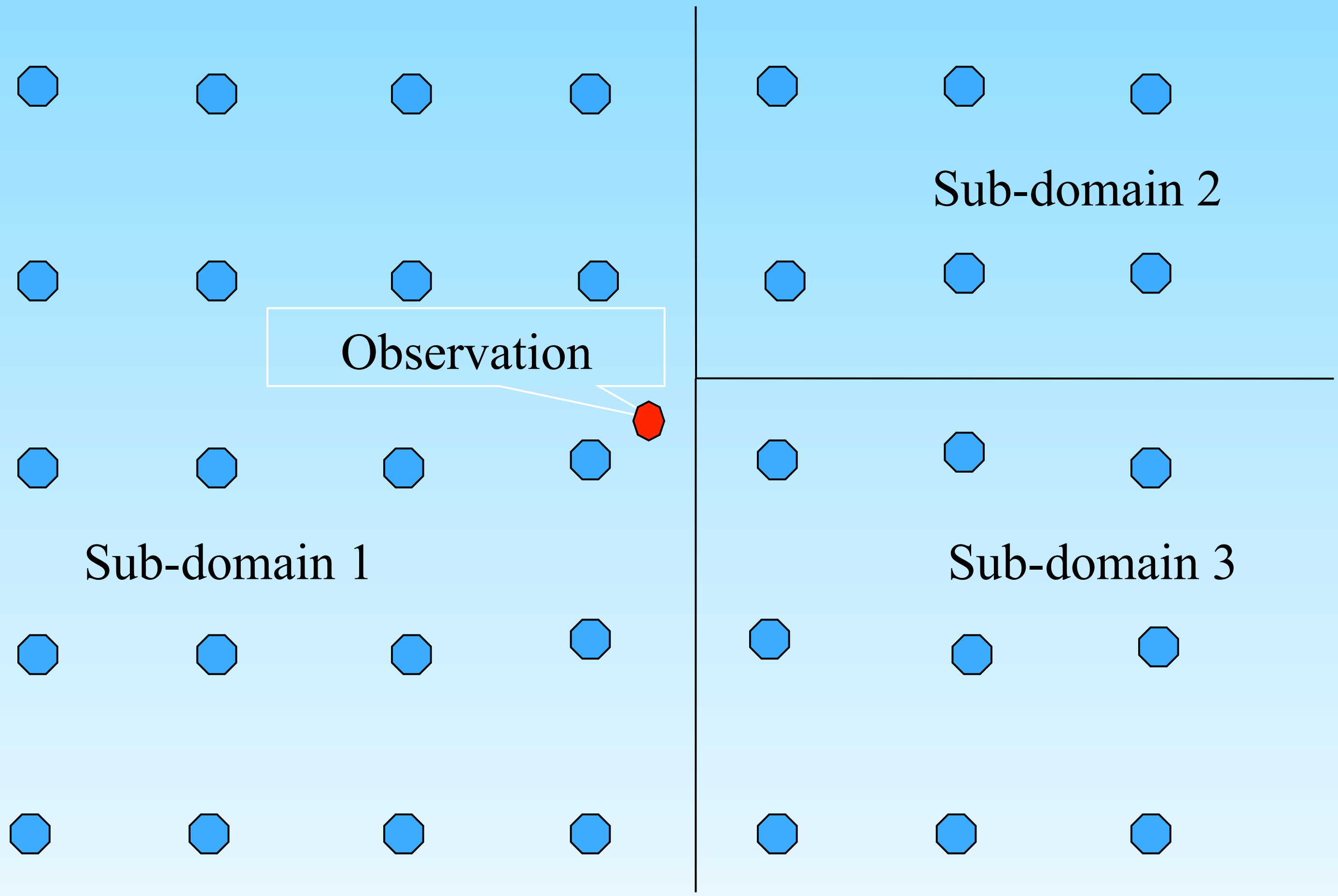
Simulation of observations

- 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

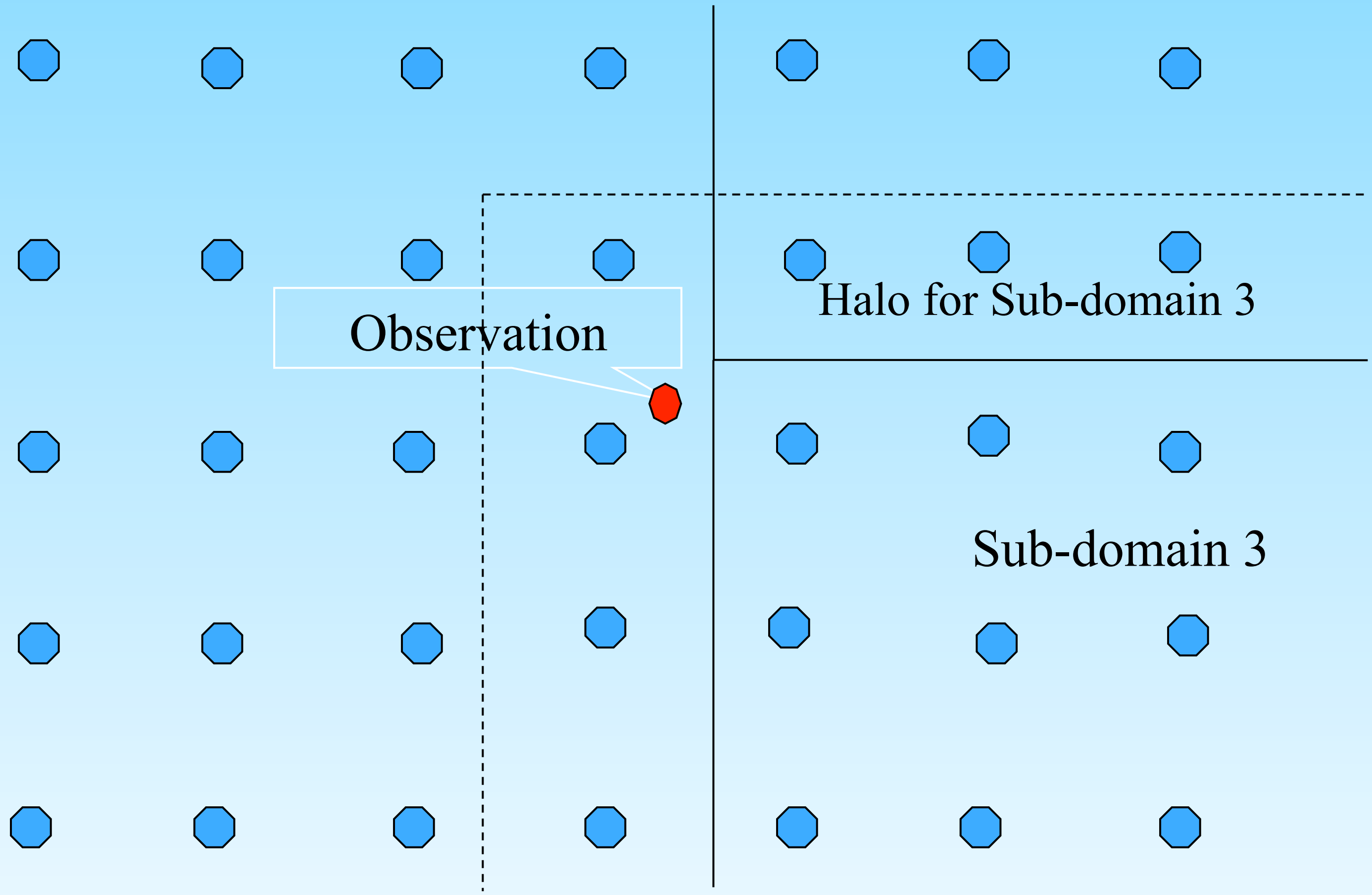
Atmospheric analysis problem (Practical) Outer (K) and Inner (L) iteration operators

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

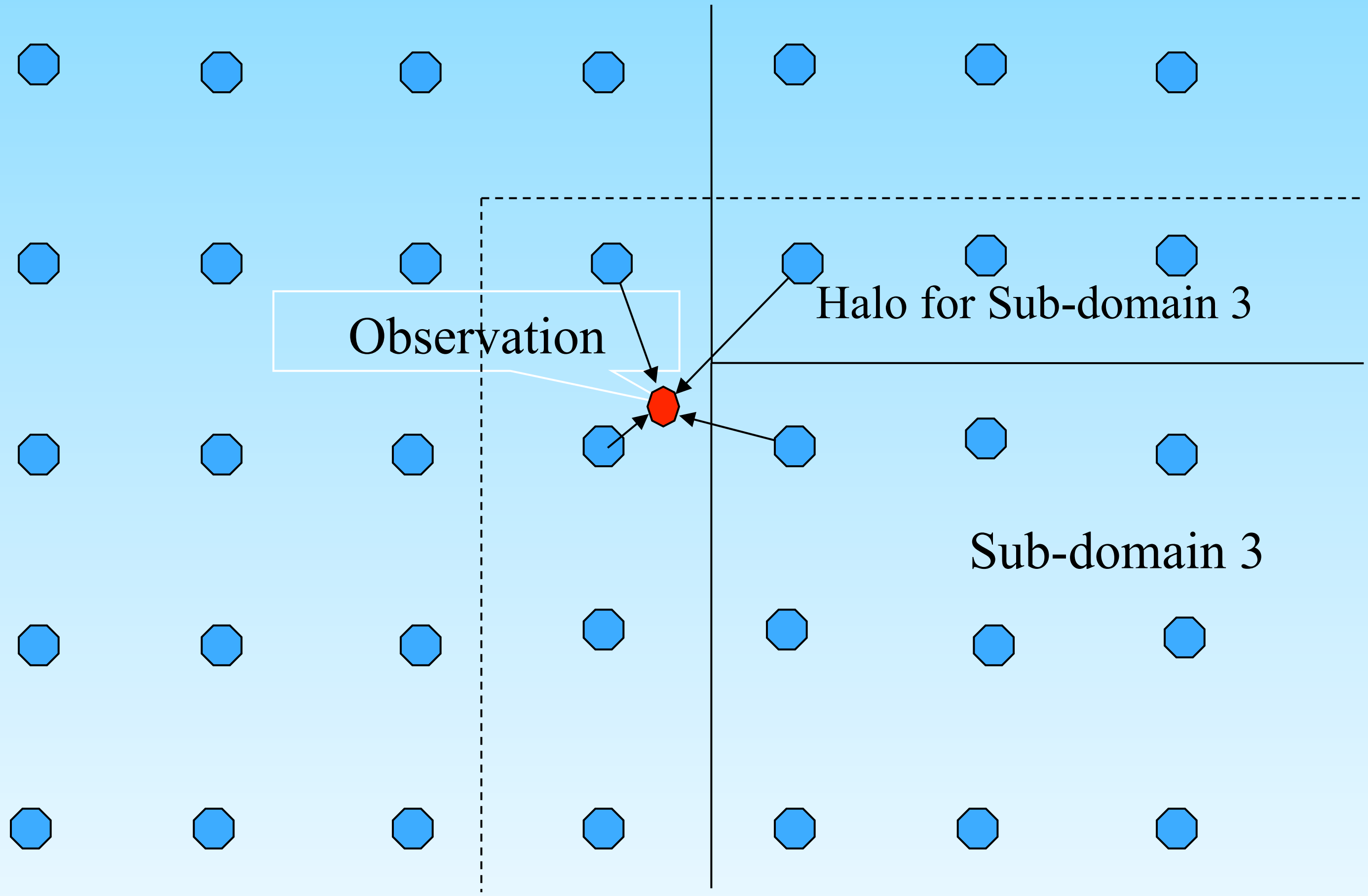
Observation/Sub-domain layout



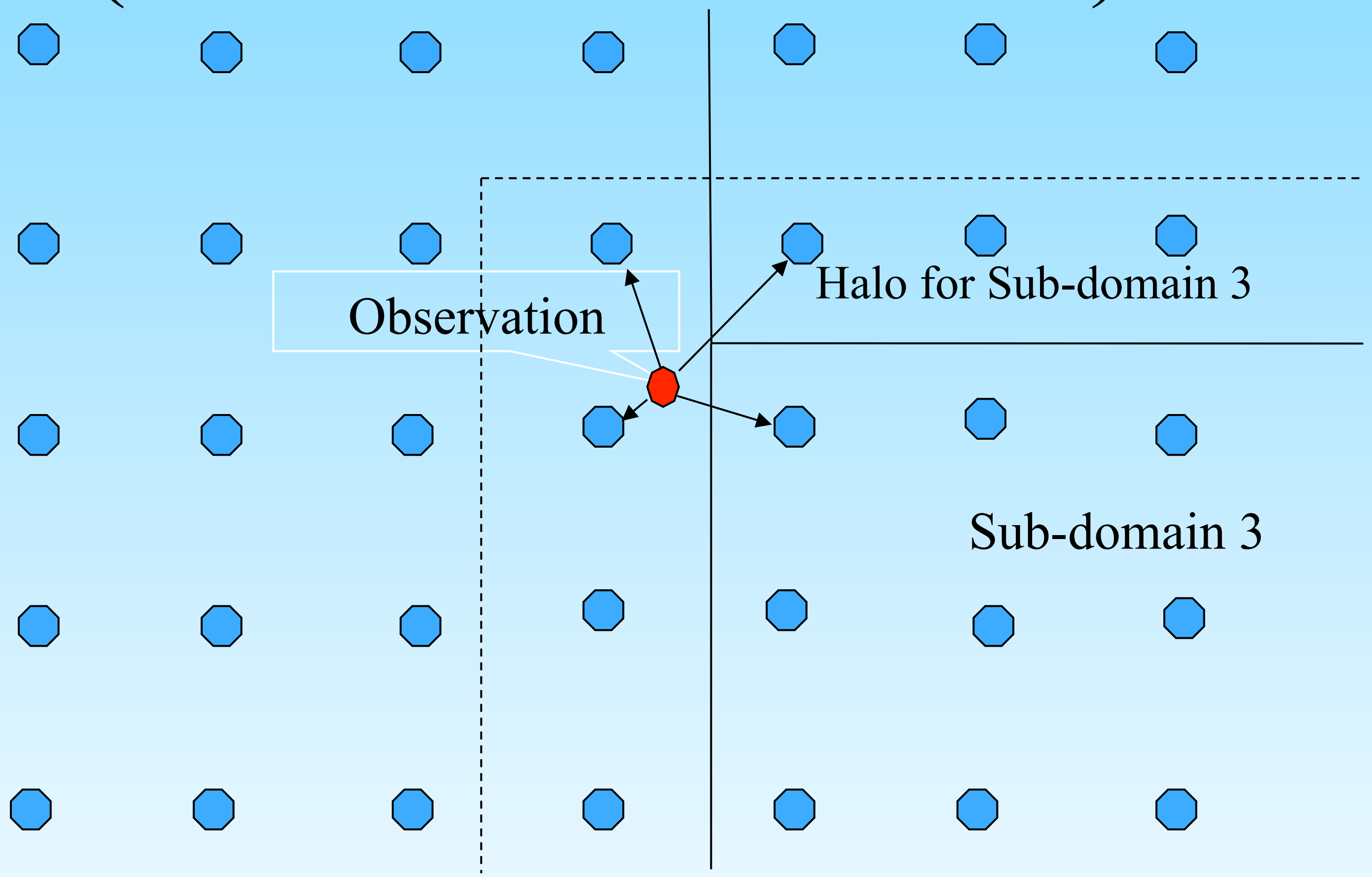
Sub-domain 3 calculation w/halo



Forward interpolation to ob.



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



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!

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

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, sqrtmin or other minimization (solves inner iteration)
  - gsimain_finalize (clean up arrays and finalize mpi)

GSI layout (major routines)

- pcgsoi (other minimizations similar)
  - 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)

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.
- McNally, A.P., J.C. Derber, W.-S. Wu and B.B. Katz, 2000: **The use of TOVS level-1B radiances in the NCEP SSI analysis system.** Q.J.R.M.S., 126, 689-724.
- Parrish, D. F. and J. C. Derber, 1992: **The National Meteorological Center's spectral statistical interpolation analysis system**. Mon. Wea. Rev., 120, 1747 - 1763.
- Derber, J. C. and W.-S. Wu, 1998: **The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system**. Mon. Wea. Rev., 126, 2287 - 2299.
- 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.
- Kazumori, M; Liu, Q; Treadon, R; Derber, JC, **Impact Study of AMSR-E Radiances in the NCEP Global Data Assimilation System** Monthly Weather Review,Vol. 136, no. 2, pp. 541-559. Feb 2008.
- 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)

Challenges

- Negative Moisture and other tracers
- Diabatic analysis
- Cloud and precipitation assimilation
- Trace gas and aerosol assimilation
- Cross-variable covariances – “balance”
- Situation dependent background and observation errors
- Use of satellite radiances in regional mode
- Use of satellite data over land/ice/snow