WRFDA 4DVAR

Zhiquan (Jake) Liu

NCAR/MMM

Outline

- Incremental formulation of 3DVAR
- Incremental formulation of 4DVAR
- Introduction to 4DVAR practice

Incremental formulation of 3DVAR and outer loop

1.1 Non-linear 3DVAR Formulation

Non-linear 3DVAR cost function

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}[H(\mathbf{x}) - \mathbf{y}]^{\mathrm{T}}\mathbf{R}^{-1}[H(\mathbf{x}) - \mathbf{y}]$$

1.2 Incremental 3DVAR Formulation

Linearization, let $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^g$ and $\delta\mathbf{x}^g = \mathbf{x}^b - \mathbf{x}^g$, thus $\mathbf{x} = \delta\mathbf{x} + \mathbf{x}^g$, we have

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x} - \delta\mathbf{x}^g)^{\mathrm{T}}\mathbf{B}^{-1}(\delta\mathbf{x} - \delta\mathbf{x}^g) + \frac{1}{2}[H(\delta\mathbf{x} + \mathbf{x}^g) - \mathbf{y}]^{\mathrm{T}}\mathbf{R}^{-1}[H(\delta\mathbf{x} + \mathbf{x}^g)) - \mathbf{y}]$$

Do Taylor Expansion for observation term

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x} - \delta\mathbf{x}^g)^{\mathrm{T}}\mathbf{B}^{-1}(\delta\mathbf{x} - \delta\mathbf{x}^g) + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})$$

where $\mathbf{d} = \mathbf{y} - H(\mathbf{x}^g)$ and $\mathbf{H}$ is the linearized version of H in the vicinity of $\mathbf{x}^g$.

NOTE: X^g is the first guess, not to confuse with the background X^b even though they are the same for the first outer loop. From the 2nd outer loop, X^g is equal to the analysis X^a from previous outer loop.

1.3 Control Variable Transform (CVT)

To avoid the inverse calculation of large $\mathbf{B}$ matrix, do a change of variable $\delta\mathbf{x} = \mathbf{U}\mathbf{v}$ and $\delta\mathbf{x}^g = \mathbf{U}\mathbf{v}^g$ with $\mathbf{U}$ the square root of $\mathbf{B}$, namely $\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{\mathrm{T}/2} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$ or $\mathbf{U} = \mathbf{B}^{1/2}$. Also $\mathbf{B}^{-1} = \mathbf{U}^{-\mathrm{T}}\mathbf{U}^{-1}$. Then the cost function with respect to the control variable $\mathbf{v}$ becomes

$$J(\mathbf{v}) = \frac{1}{2}(\mathbf{v}-\mathbf{v}^g)^{\mathrm{T}}(\mathbf{v}-\mathbf{v}^g) + \frac{1}{2}(\mathbf{H}\mathbf{U}\mathbf{v}-\mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{U}\mathbf{v}-\mathbf{d}) \tag{4}$$

1.4 Solution of Incremental 3DVAR

The minimization of the cost function requires its gradient with respect to $\mathbf{v}$ to be zero, namely

$$\nabla_{\mathbf{v}} J(\mathbf{v}) = (\mathbf{v}-\mathbf{v}^g) + \mathbf{U}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{U}\mathbf{v}-\mathbf{d}) = 0 \tag{5}$$

After minimization, we get the analysis increment $\mathbf{v}^a$ in control variable space,

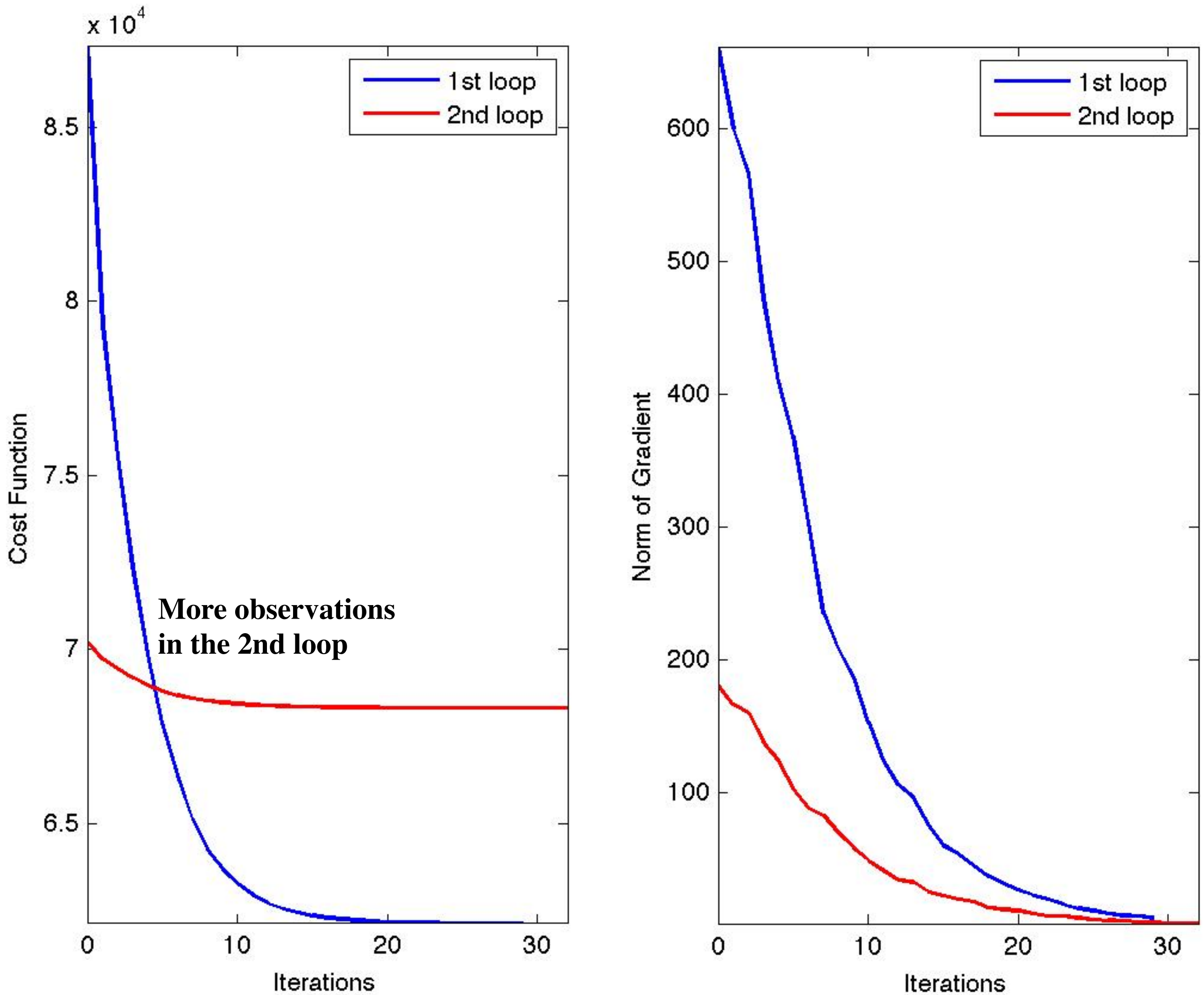
$$\mathbf{v}^a = (\mathrm{I} + \mathbf{U}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{U})^{-1}(\mathbf{v}^g + \mathbf{U}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{d})$$

The analysis increment and the analysis in model space are

$$\mathbf{x}^a = \mathbf{x}^g + \delta\mathbf{x}^a = \mathbf{x}^g + \mathbf{U}\mathbf{v}^a$$

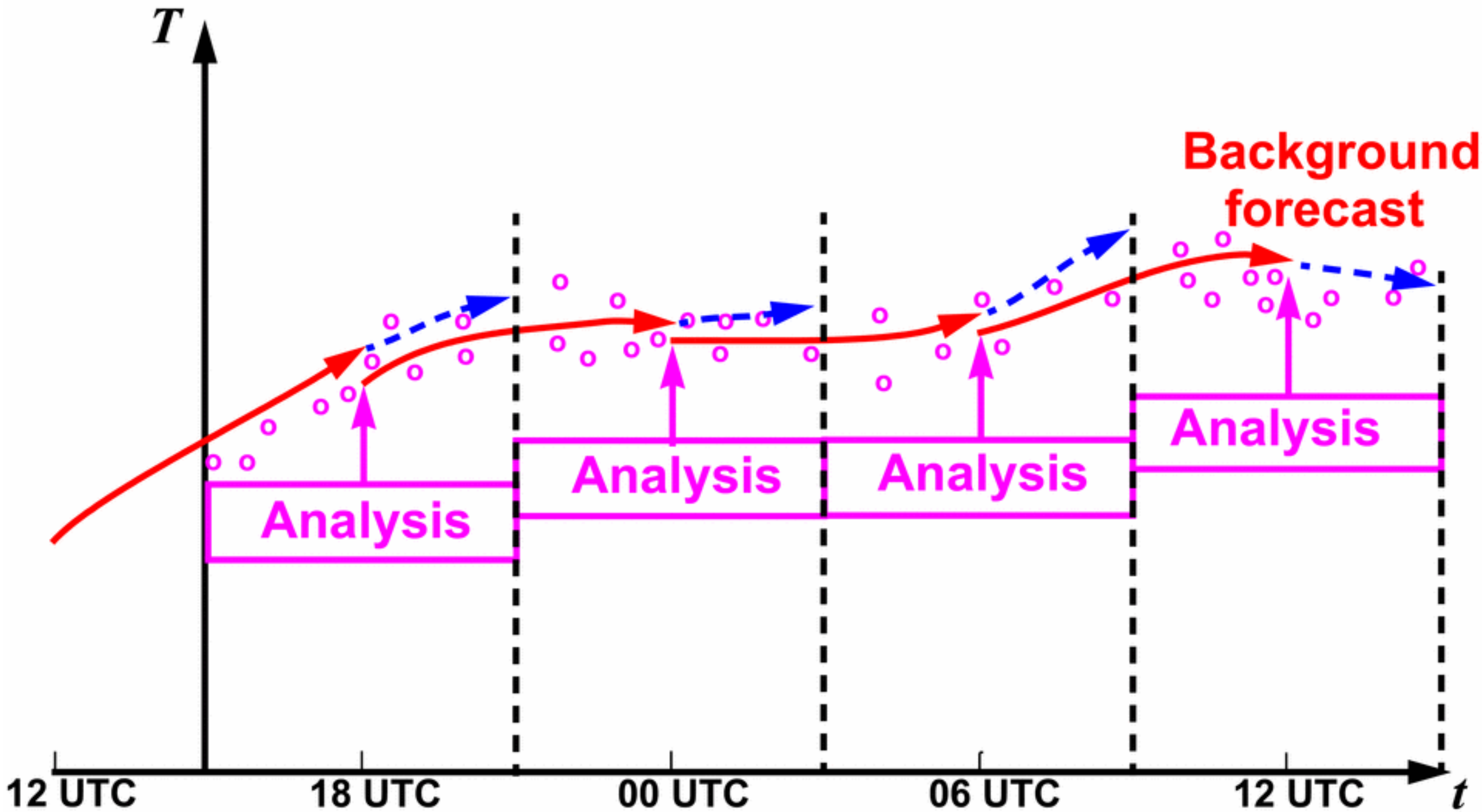
NOTE: (1)outer loop-1: $X^g = X^b$; V^g=0; loop-2: $X^g = X^a$, V^g=V^a from previous loop.
(2) For each outer loop, *H* needs to be re-linearized around new X^g ;
(3) d=y-*H*(X^g) is also re-calculated and re-do QC (OMB check).

Cost Function/Gradient with 2 outer loops



3DVAR

Assume observations valid at the center of time window



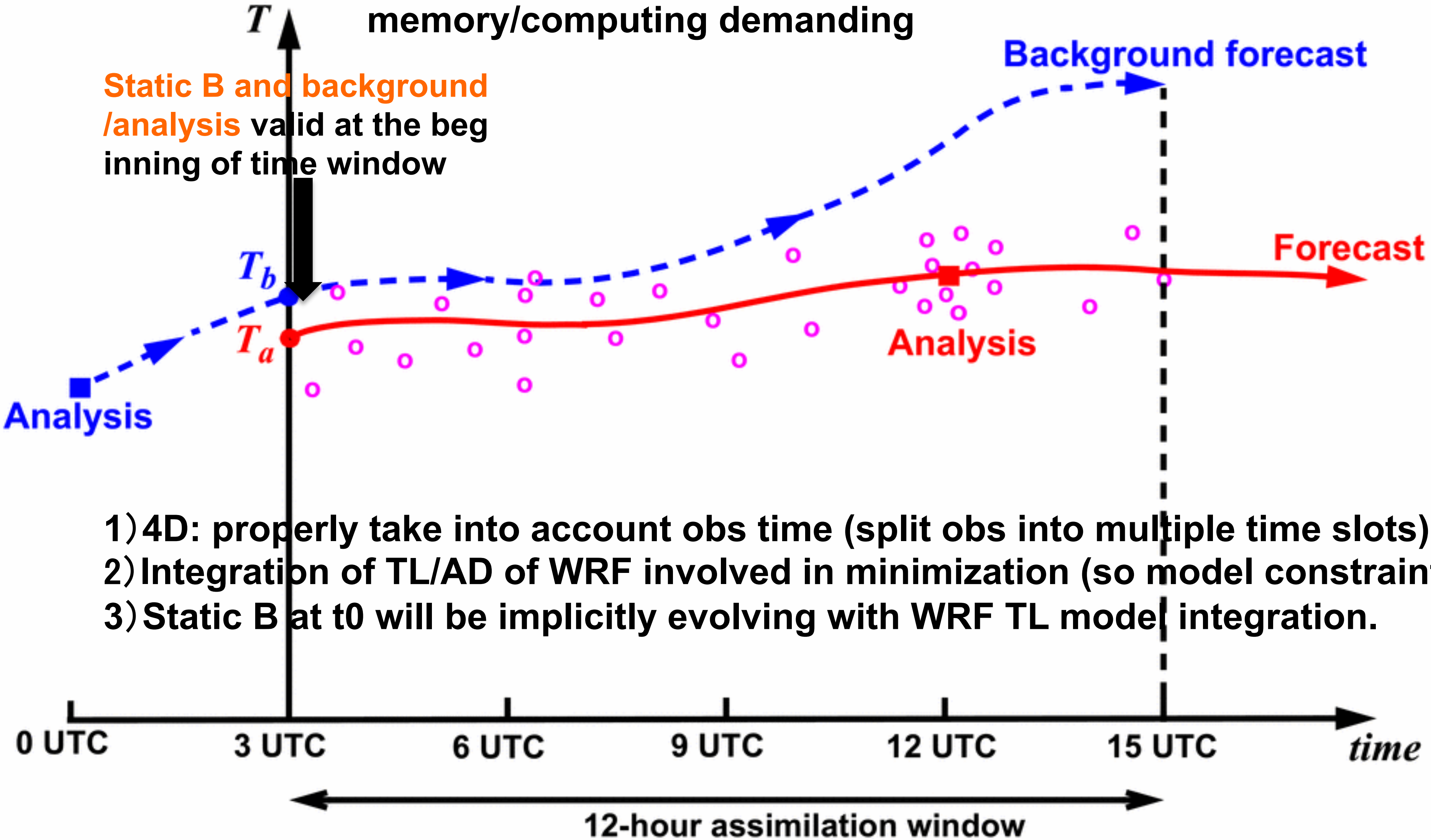
Outline

- Incremental formulation of 3DVAR
- Incremental formulation of 4DVAR
- Introduction to 4DVAR practice

4DVAR

Need develop/maintain TL/AD version of a NWP model.
memory/computing demanding

Static B and background /analysis valid at the beg inning of time window



1) **4D: properly take into account obs time (split obs into multiple time slots).**
2) **Integration of TL/AD of WRF involved in minimization (so model constraint)**
3) **Static B at t0 will be implicitly evolving with WRF TL model integration.**

Incremental formulation of 4DVAR

2.1 Non-linear 4DVAR Formulation

Non-linear 4DVAR cost function

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{i=1}^{N}[H_i(\mathbf{x}_i) - \mathbf{y}_i]^{\mathrm{T}}\mathbf{R}_i^{-1}[H_i(\mathbf{x}_i) - \mathbf{y}_i] \tag{15}$$

where the subscript "0" indicates the beginning of the 4DVAR time window. Substitute the NWP model into the cost function, we obtain

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{i=1}^{N}[H_i(M_i(\mathbf{x}_0)) - \mathbf{y}_i]^{\mathrm{T}}\mathbf{R}_i^{-1}[H_i(M_i(\mathbf{x}_0)) - \mathbf{y}_i] \tag{16}$$

2.2 Incremental 4DVAR Formulation

Linearization, let $\delta\mathbf{x}_0 = \mathbf{x}_0 - \mathbf{x}_0^g$ and $\delta\mathbf{x}_0^g = \mathbf{x}_0^b - \mathbf{x}_0^g$, thus $\mathbf{x}_0 = \delta\mathbf{x}_0 + \mathbf{x}_0^g$, we have

$$J(\delta\mathbf{x}_0) = \frac{1}{2}(\delta\mathbf{x}_0 - \delta\mathbf{x}_0^g)^{\mathrm{T}}\mathbf{B}^{-1}(\delta\mathbf{x}_0 - \delta\mathbf{x}_0^g) + \frac{1}{2}\sum_{i=1}^{N}[H_i(M_i(\delta\mathbf{x}_0 + \mathbf{x}_0^g) - \mathbf{y}_i]^{\mathrm{T}}\mathbf{R}_i^{-1}[H_i(M_i(\delta\mathbf{x}_0 + \mathbf{x}_0^g)) - \mathbf{y}_i]$$

Do Taylor Expansion for observation term

$$J(\delta\mathbf{x}_0) = \frac{1}{2}(\delta\mathbf{x}_0 - \delta\mathbf{x}_0^g)^{\mathrm{T}}\mathbf{B}^{-1}(\delta\mathbf{x}_0 - \delta\mathbf{x}_0^g) + \frac{1}{2}\sum_{i=1}^{N}(\mathbf{H}_i\mathbf{M}_i\delta\mathbf{x}_0 - \mathbf{d}_i)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{H}_i\mathbf{M}_i\delta\mathbf{x}_0 - \mathbf{d}_i)$$

where $\mathbf{d}_i = \mathbf{y}_i - H_i[M_i(\mathbf{x}_0^g)]$.

Incremental 4DVAR with control variable transform

Again, control variable transform $\delta \mathbf{x}_0 = \mathbf{U}\mathbf{v}$ and $\delta \mathbf{x}_0^g = \mathbf{U}\mathbf{v}^g$. $\delta \mathbf{x}_0$ indicates that analysis increment is valid at the beginning of the 4DVAR time window. Then the cost function with respect to the control variable $\mathbf{v}$ becomes

$$J(\mathbf{v}) = \frac{1}{2}(\mathbf{v} - \mathbf{v}^g)^{\mathrm{T}}(\mathbf{v} - \mathbf{v}^g) + \frac{1}{2}\sum_{i=1}^{N}(\mathbf{H}_i\mathbf{M}_i\mathbf{U}\mathbf{v} - \mathbf{d}_i)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{H}_i\mathbf{M}_i\mathbf{U}\mathbf{v} - \mathbf{d}_i) \tag{19}$$

NOTE:

(1) For each outer loop, need to store forecast trajectory (each time step) and V^g in the memory.

(2) For each loop, H and M needs to be re-linearized around new forecast trajectory; $d_i = y_i - H_i(X^g_i)$ is also re-calculated and re-do QC (OMB check).

(3) 4DVAR outer loops could run at different (typically lower) resolutions, common practice at operational NWP centers (capability under development with WRFDA)

Multi-Resolution Incremental 4DVAR

- 4DVAR minimization runs at lower resolutions than WRF model's to allow substantial speed-up
- Now works for cv_options = 3
  - Need more development to make it work properly with cv_options = 5/6/7

TABLE 2. Computational performance comparison of the full-resolution WRF 4D-Var and multi-incremental WRF 4D-Va on NCAR Yellowstone; Each test has three outer loops with 20 iterations inner loops for each. Unit: Minutes

Cores	8	16	32	64	128	256	512	1024
Full-Res.	–	–	4191	2169	1230	728	392	257
Multi-Inc.	455	217	135	83	53	37	–	

12-h vs. ~ 40min!

15km/15km/15km versus 135km/45km/45km

Xin Zhang et al., 2014: Development of an Efficient Regional Four-Dimensional Variational Data Assimilation System for WRF. J. Atmos. Oceanic Technol., 31, 2 777–2794.

Advantages of 4DVAR

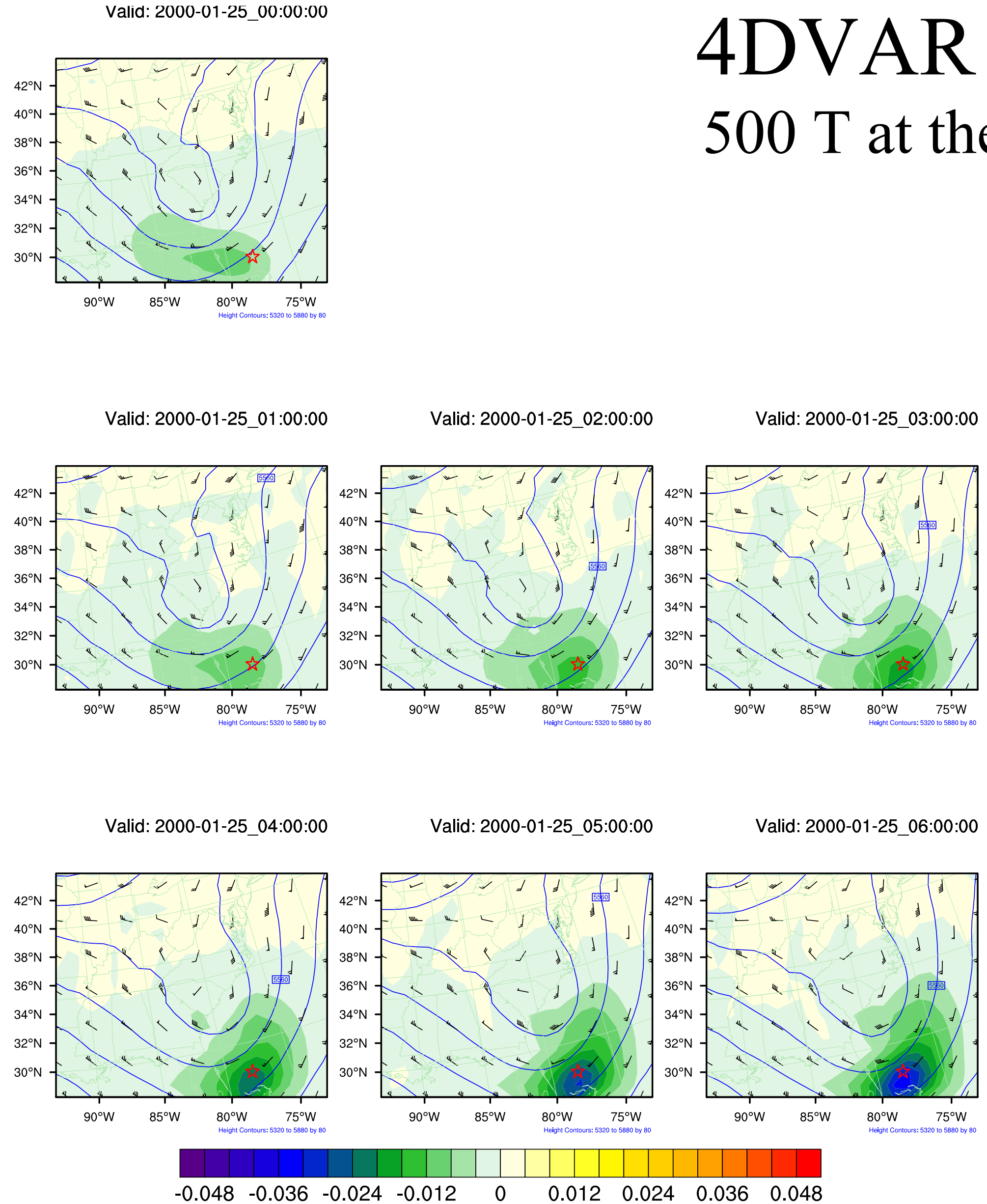
- Data can be assimilated at appropriate time, so can use frequently reported observations
- Can use "future" observations to constrain the analysis at earlier time
- NWP model as part of constraints, so propagating observation information via model dynamics and physics

- Background error covariance (BEC) implicitly evolving within time window through linearized model, though B (BEC at the beginning of time window) typically the same for each analysis cycle. BEC at time t_i,

$$\mathbf{B_i = M_i B M_i^T}$$

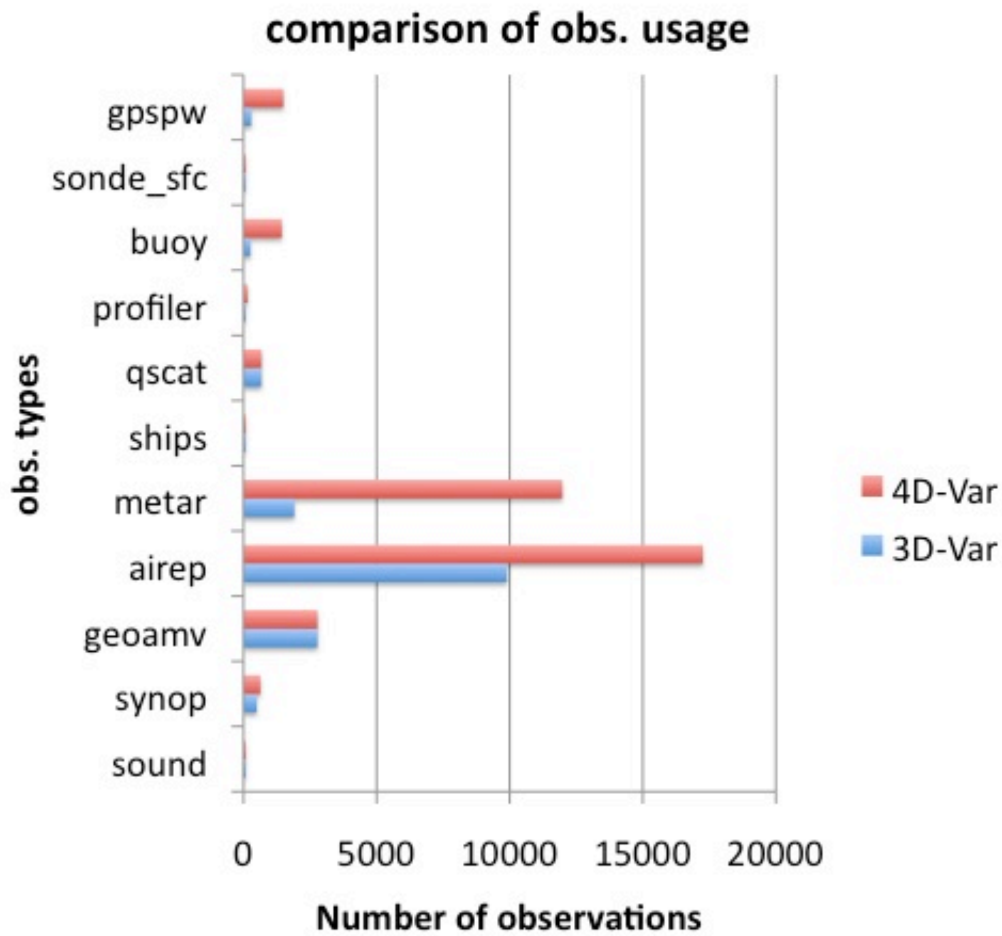
4DVAR Single Obs Test

500 T at the end of time window



From Xin Zhang

Number of obs assimilated: 3DVAR vs. 4DVAR



From Xin Zhang

Some word about WRFDA-3DVAR/4DVAR for WRF/Chem

- Under development for aerosol/chemistry data assimilation

- Including WRFPlus-Chem for GOCART
  - J. J. Guerrette and D. K. Henze, 2015

- Will be very useful for air-quality forecast and source emission inversion.

See Lecture by J. J. Guerrette

Outline

- Incremental formulation of 3DVAR
- Incremental formulation of 4DVAR
- Introduction to 4DVAR practice

Compile WRFDA in 4DVAR mode

- Download WRFPlus code
  - Include non-linear and TL/AD code of WRF
- Download WRFDA code
- Install WRFPLUS V3.8
  - ./configure (-d) wrfplus
    ./compile wrf (only compile wrf.exe)
  - wrf.exe should be generated under the WRFPLUSV3/main directory.
- for csh, tcsh : setenv WRFPLUS_DIR path of wrfplusv3
  for bash, ksh : export WRFPLUS_DIR=path of wrfplusv3
- Install WRFDA V3.8
  - ./configure (-d) 4dvar
    ./compile all_wrfvar
    da_wrfvar.exe should be generated in the var/build directory.

Notes about WRFPlus

- WRFPLUS only works with regional ARW core, not for NMM core or global WRF.

- WRFPLUS only works with single domain, not for nested domains.

- WRFPLUS can not work with Adaptive Time Stepping options.
- WRFPLUS TL/AD code only has 3 simplified physics processes:
  - surface drag (bl_pbl_physics=98);
  - large scale condensate or Kessler (mp_physics=98 or 99)
  - a simplified cumulus scheme (cu_physics=98)

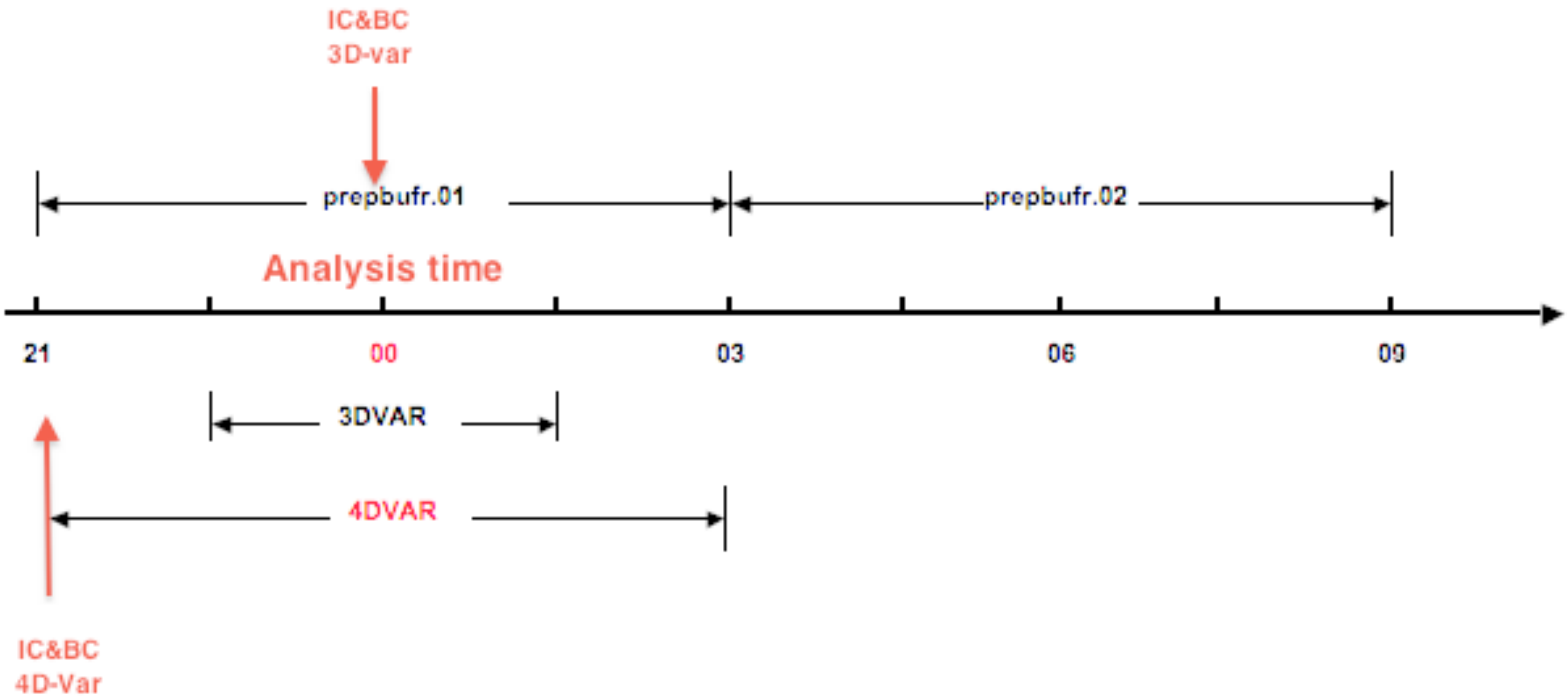
Prepare obs for 4DVAR

- Conventional observations
  - LITTLE_R format
  - NCEP PREPBUFR format

- Satellite radiance BUFR data

- ASCII format precipitation and radar data

4DVAR time window



Run a 4DVAR test case

- enter WRFDA/var/test/4dvar (or working directory of your choice)
- get the test dataset from:
  - http://www2.mmm.ucar.edu/wrf/users/wrfda/download/testdata.html
- ln -fs wrfinput_d01 fg
- ln –fs wrfbdy_d01 .
- ln -fs ../../build/da_wrfvar.exe .
- ln -fs ../../run/be.dat.cv3 be.dat
- ./da_wrfvar.exe
- Typically you should run in parallel with MPI (mpirun -np # da_wrfvar.exe) or your system's custom run command (on Yellowstone: bsub))

Run a 4DVAR test case

- WRFPlus/WRFDA compiled in double precision
- So link double-precision version of following files for 4DVAR run
  - ln -sf ${WRF_DIR}/run/RRTM_DATA_DBL RRTM_DATA
  - ln -sf ${WRF_DIR}/run/RRTMG_LW_DATA_DBL RRTMG_LW_DATA
  - ln -sf ${WRF_DIR}/run/RRTMG_SW_DATA_DBL RRTMG_SW_DATA
- And other WRF related files
  - ln -sf ${WRF_DIR}/run/SOILPARM.TBL .
  - ln -sf ${WRF_DIR}/run/VEGPARM.TBL .
  - ln -sf ${WRF_DIR}/run/GENPARM.TBL .
  - ln -sf ${WRF_DIR}/run/LANDUSE.TBL .

Important namelist variables

- &wrfvar1
  - **var4d**: logical, set to .true. to use 4D-Var
  - **var4d_lbc**: logical, set to .true. to include lateral boundary condition control in 4D-Var
  - **var4d_bin**: integer, seconds, length of sub-window to group observations in 4D-Var
- &wrfvar18,21,22
  - **analysis_date** : the start time of the assimilation window
  - **time_window_min** : the start time of the assimilation window
  - **time_window_max** : the end time of the assimilation window
- &perturbation
  - **jcdfi_use**: logical, if turn on the digital filter as a weak constraint.
  - **jcdfi_diag**: integer, 0/1, Jc term diagnostics
  - **jcdfi_penalty**: real, weight to jcdfi term

Important namelist variables

- &physics
  - all physics options must be consistent with those used in wrfinput
    - Non-linear WRF run can use different physics options from TL/AD
  - **mp_physics_ad** =
    98: large-scale condensation microphysics (default)
    99: modified Kessler scheme (new in V3.7)
  - **bl pbl physics** = any : but only surface drag available for TL/AD
  - **cu physics** = any : but only simplified cumulus scheme for TL/AD

- &time control
  - **run_xxxx** : be consistent with the length of the time window
  - **start_xxxx** : be consistent with the start time of the time window
  - **end_xxxx** : be consistent with the end time of the time window

WRFDA adjoint check before 4DVAR run

- &wrfvar10
  - test_transforms=true,
- run da wrfvar.exe

Check results

```
...
 wrf: back from adjoint integrate
 d01 2008-02-05_21:00:00 read nonlinear xtraj time stamp:2008-02-05_21:00:00
 Single Domain < y, y     > =   2.15435506772433E+06
 Single Domain < x, x_adj > =   2.15435506772431E+06

 Whole Domain < y, y     > =   2.15435506772433E+06
 Whole Domain < x, x_adj > =   2.15435506772431E+06

da_check_xtoy_adjoint: Test Finished:

 *** WRF-Var check completed successfully ***
```