New Control Variables in Var DA

Wan-Shu Wu

Holm et al.(2002) ECMWF Tech Memo
Liu et al.(2017) WMO DAS 7
S. Levine (NCEP)

Examples of 'new' control variables

Beside stream function (ψ); unbalanced part of velocity potential (χ); unbalanced part of temperature (T); unbalanced part of surface pressure (P); and normalized relative humidity (nrh), examples of other analysis control variables are

- Ozone
- Visibility
- Radar reflectivity
- Cloud water
- Cloud ice
- Cloud condensate
- Wind gust
- Cloud ceiling height
- …

Use of nrh as example on Gaussian CV

Issues related to implementing new control variables will be discussed

GSI Hybrid Variational-Ensemble

- Incorporate ensemble perturbations directly into variational cost function through extended control variable

$$J(\mathbf{x}_f', \boldsymbol{\alpha}) = \beta_f \frac{1}{2}(\mathbf{x}_f')^T \mathbf{B}^{-1}(\mathbf{x}_f') + \beta_e \frac{1}{2}\sum_{n=1}^{N} (\boldsymbol{\alpha}^n)^T \mathbf{L}^{-1}(\boldsymbol{\alpha}^n) + \frac{1}{2}\sum_{t=1}^{T} (\mathbf{y}_t' - \mathbf{H}\mathbf{x}_t')^T \mathbf{R}^{-1}(\mathbf{y}_t' - \mathbf{H}\mathbf{x}_t')$$

$$\mathbf{x}_t' = \mathbf{x}_f' + \sum_{n=1}^{N} \boldsymbol{\alpha}^n \circ \mathbf{x}_{e\,t}'^{\,n} \qquad \frac{1}{\beta_f} + \frac{1}{\beta_e} = 1$$

$\mathbf{x}_f'$ is the increment associated with the static covariance

β_f & β_e: weighting coefficients for fixed and ensemble covariance respectively

$\mathbf{X}_t'$: (total increment) sum of $\mathbf{x}_f'$ increment from fixed/static **B** ($\mathbf{x}_f$) and ensemble **B**

α_k: extended control variable; χ_e: ensemble perturbation

L: correlation matrix (localization on ensemble perturbations)

B: background error covariance matrix

R: observational and representativeness error covariance matrix

Background Error Variance

-finding Gaussian distribution

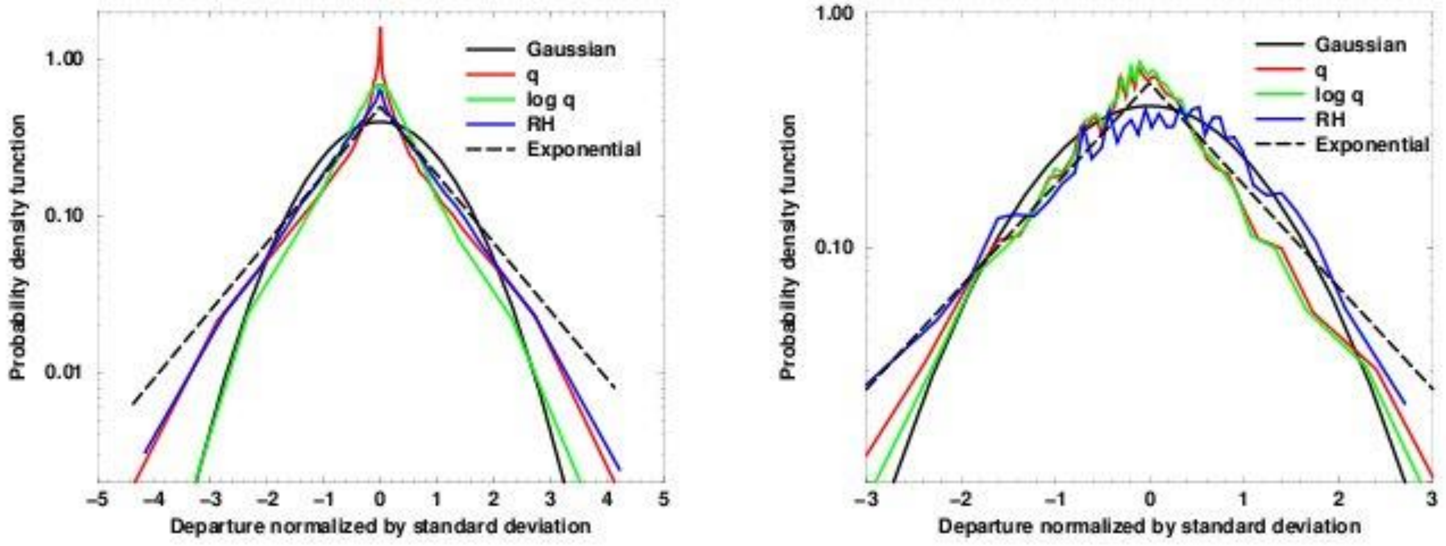


Figure 21: The pdf for a single forecast difference at approximately 850 hPa for q, $\log q$ and RH (see legend). The left panel shows all differences, and the right panel shows differences for similar values of the background field (a 2.5% interval centered around the median of the background values of each variable). For comparison, a Gaussian (black line) and an exponential pdf (dashed black) are also shown. The right panel is noisy due to the limited sample in each interval, but the result remains similar when more fields are added to the statistics.

Holm et al.(2002) ECMWF Tech Memo

Humidity background error
-finding Gaussian distribution

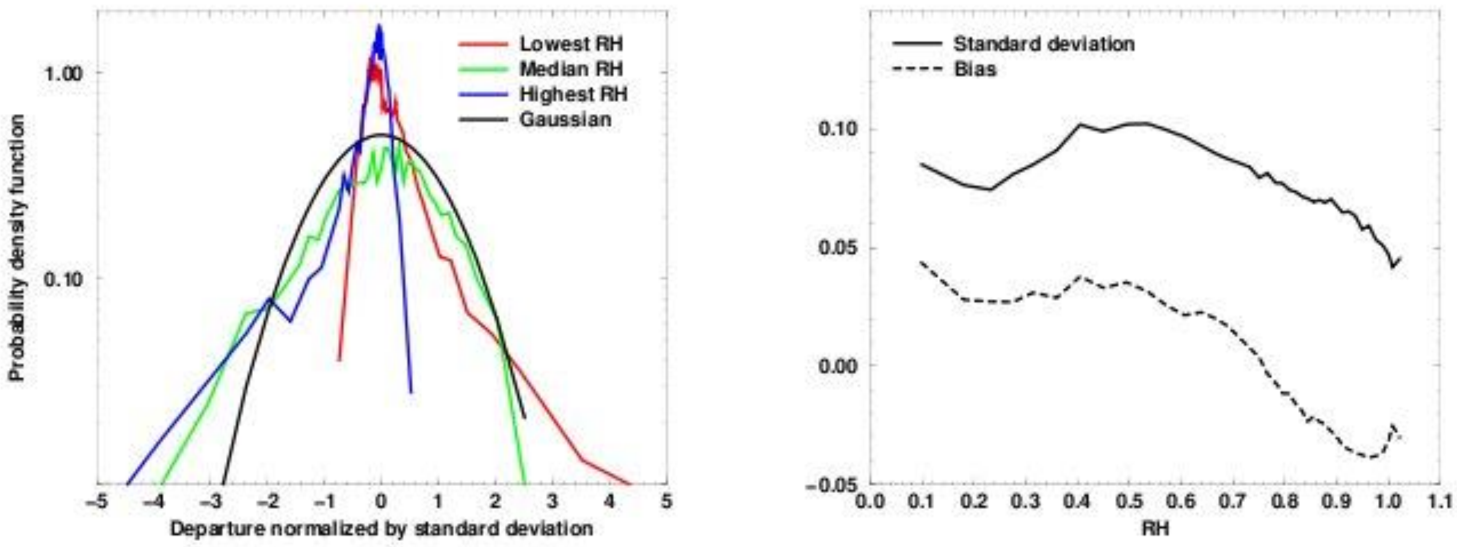


Figure 22: Forecast differences for the 'linear' $\delta RH(RH^b)$ at approximately 850 hPa. The left panel shows the pdf's for the lowest, median and highest 2.5% values of RH^b, and the right panel shows the standard deviation (full line) and bias (dashed) as a function of RH^b.

Holm et al.(2002) ECMWF Tech Memo

Humidity background error

-finding Gaussian distribution

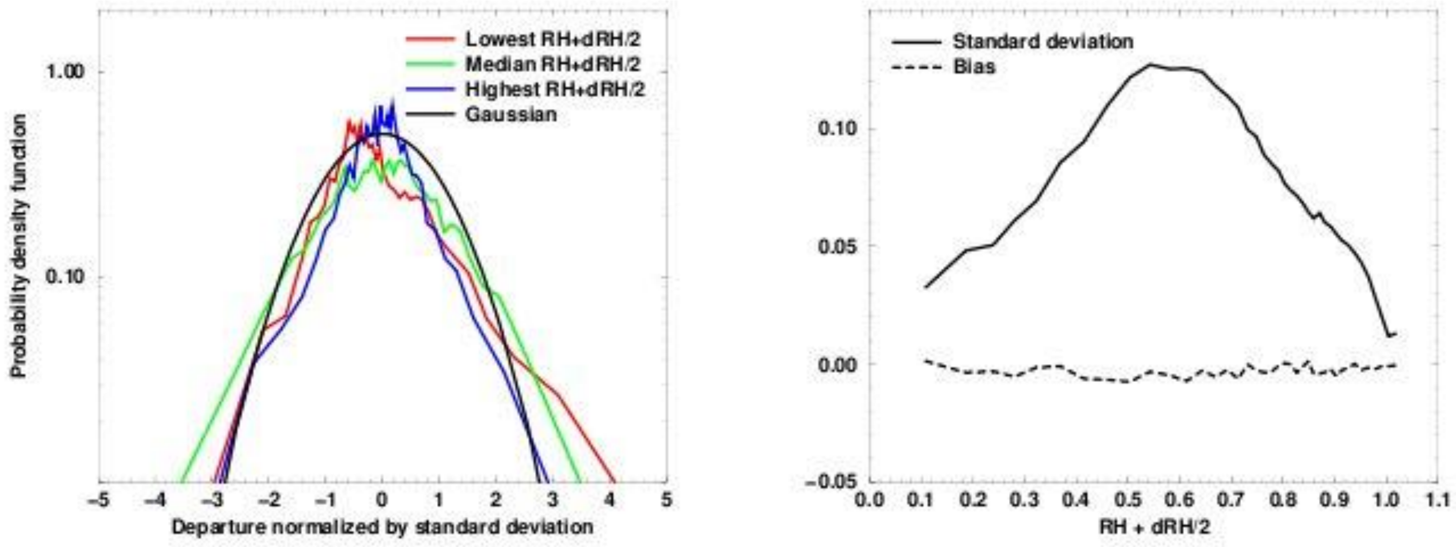


Figure 23: Forecast differences for the 'symmetric' $\delta RH(RH^b + \frac{1}{2}\delta RH)$ at 850 hPa. The left panel shows the pdf's for lowest, median and highest 2.5% values of $RH^b + \frac{1}{2}\delta RH$, and the right panel shows the standard deviation and bias as a function of $RH^b + \frac{1}{2}\delta RH$. All three curves compare reasonably well with the Gaussian (black line). Note that bins for the extreme values of RH^b are particularly affected by model and analysis effects of super-saturation clipping and resetting of humidity to positive values.

Holm et al.(2002) ECMWF Tech Memo

Normalized relative humidity as analysis control variable

$$\widetilde{\delta RH} = \frac{\delta RH}{\sigma(RH^b + \frac{1}{2}\delta RH)}$$

$\sigma(RH^b)$: standard deviation of background error as function of RH^b

Holm et al.(2002) ECMWF Tech Memo

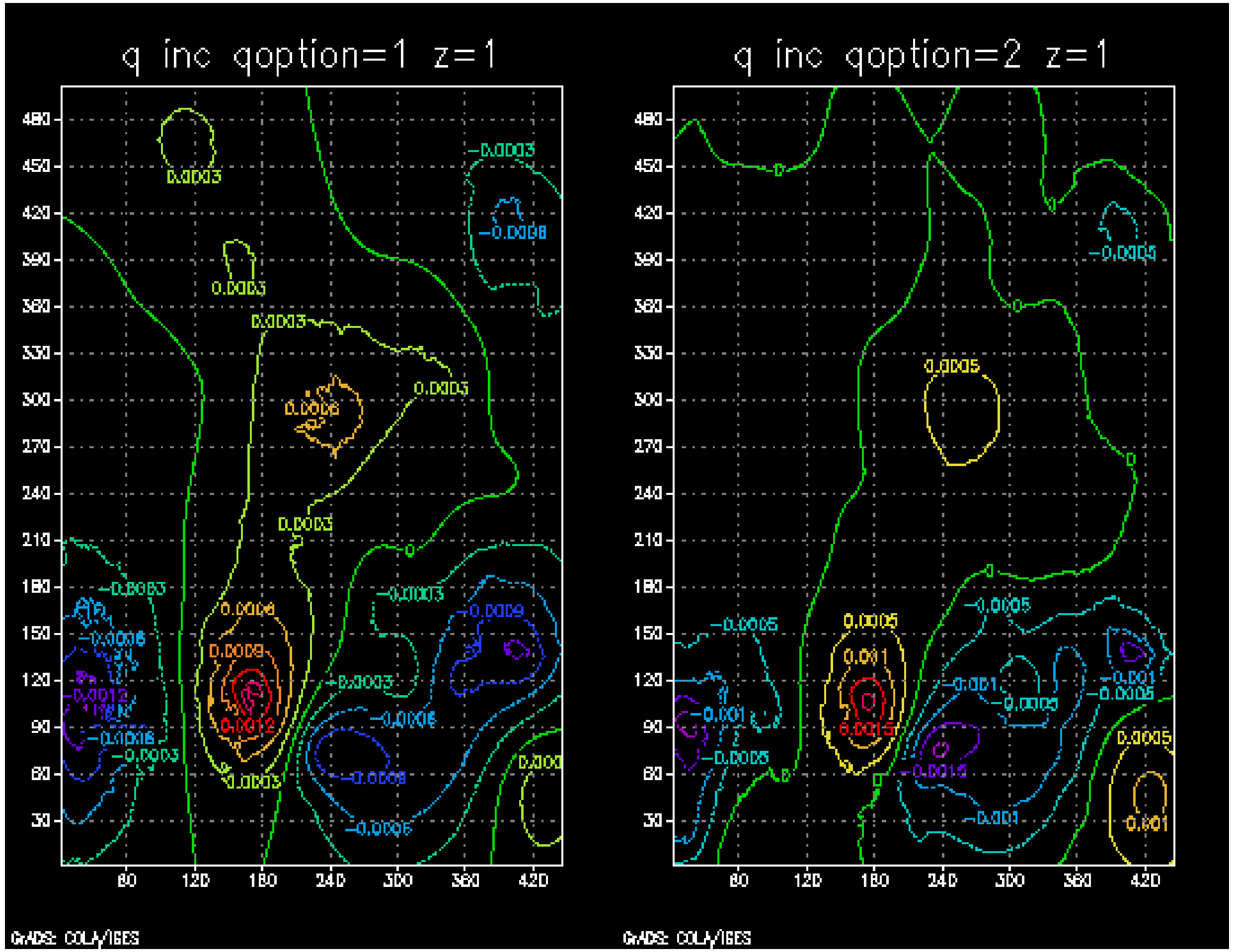
Normalized relative humidity

$$\delta RH \approx RH^b \left(\frac{\delta p}{p^b} - \frac{\delta q}{q^b} - \frac{\delta T}{\alpha_T^b} \right)$$

$$\begin{pmatrix} \delta T \\ \delta p \\ \widetilde{\delta RH} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{RH^b}{\sigma(RH^b)}\frac{1}{\alpha_T^b} & \frac{RH^b}{\sigma(RH^b)}\frac{1}{p^b} & \frac{RH^b}{\sigma(RH^b)}\frac{1}{q^b} \end{pmatrix} \begin{pmatrix} \delta T \\ \delta p \\ \delta q \end{pmatrix}$$

Holm et al.(2002) ECMWF Tech Memo

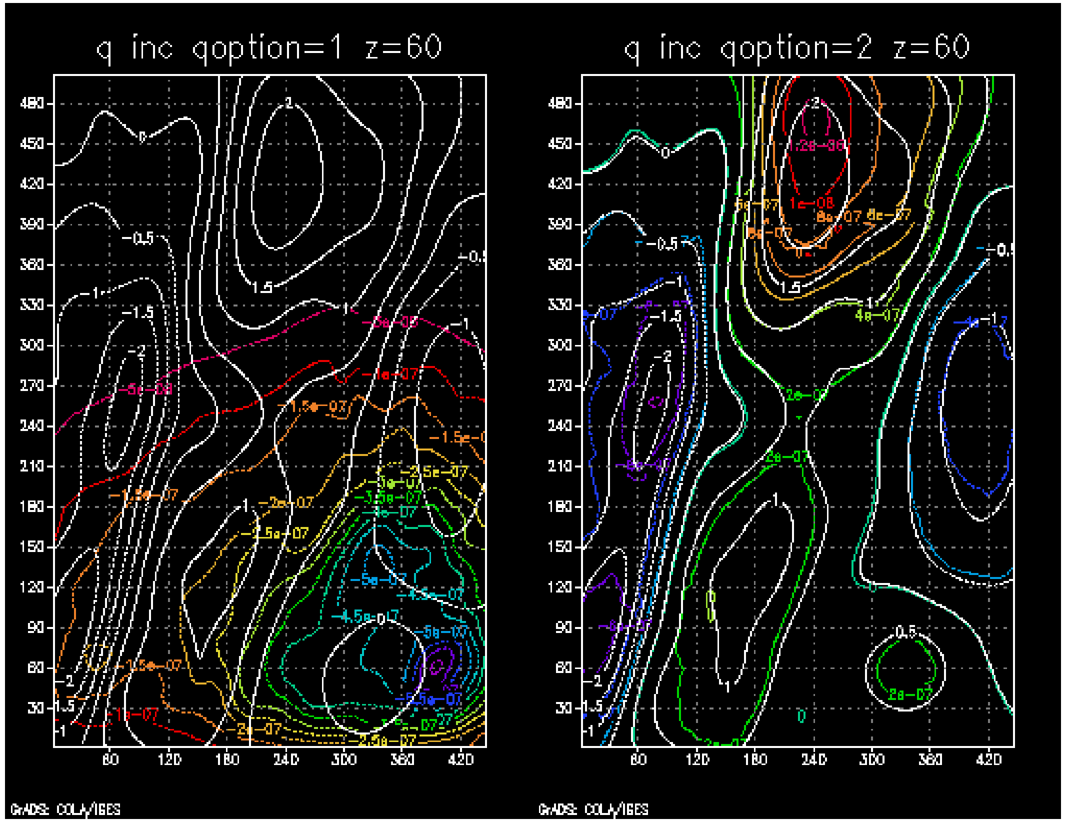
CV_q and CV_nrh produce similar analysis increments
NMM central 8km domain



At top of the model domain

CV-q: extend the RH forcing from below or from satellite radiance

CV_nrh: keep zero RH forcing with T and q multivariate relation.



With positive forecast impact CV_nrh is used operationally in both global and regional systems at NCEP

Importance of background error variances

Humidity background error variances from the EDA

- Pre-July 2017: Humidity background error variances were climatological average for given background relative humidity value and model level through a climatological statistically determined fit.
- Now: Use relative humidity background errors σ_{rh} from EDA like for other variables.
- Humidity sensitive data used better, in particular MW/IR where the radiance signal is more accurately apportioned between humidity and temperature.
- In the tropics in particular, where absolute humidity is highest, this leads to more accurate wind adjustments through the 4D-Var tracing effect.
- Results show improved O-B fits for wind and humidity sensitive observations and improved scores of wind in particular.

Holm (2017) NCEP seminar

Adding Ozone as analysis variable

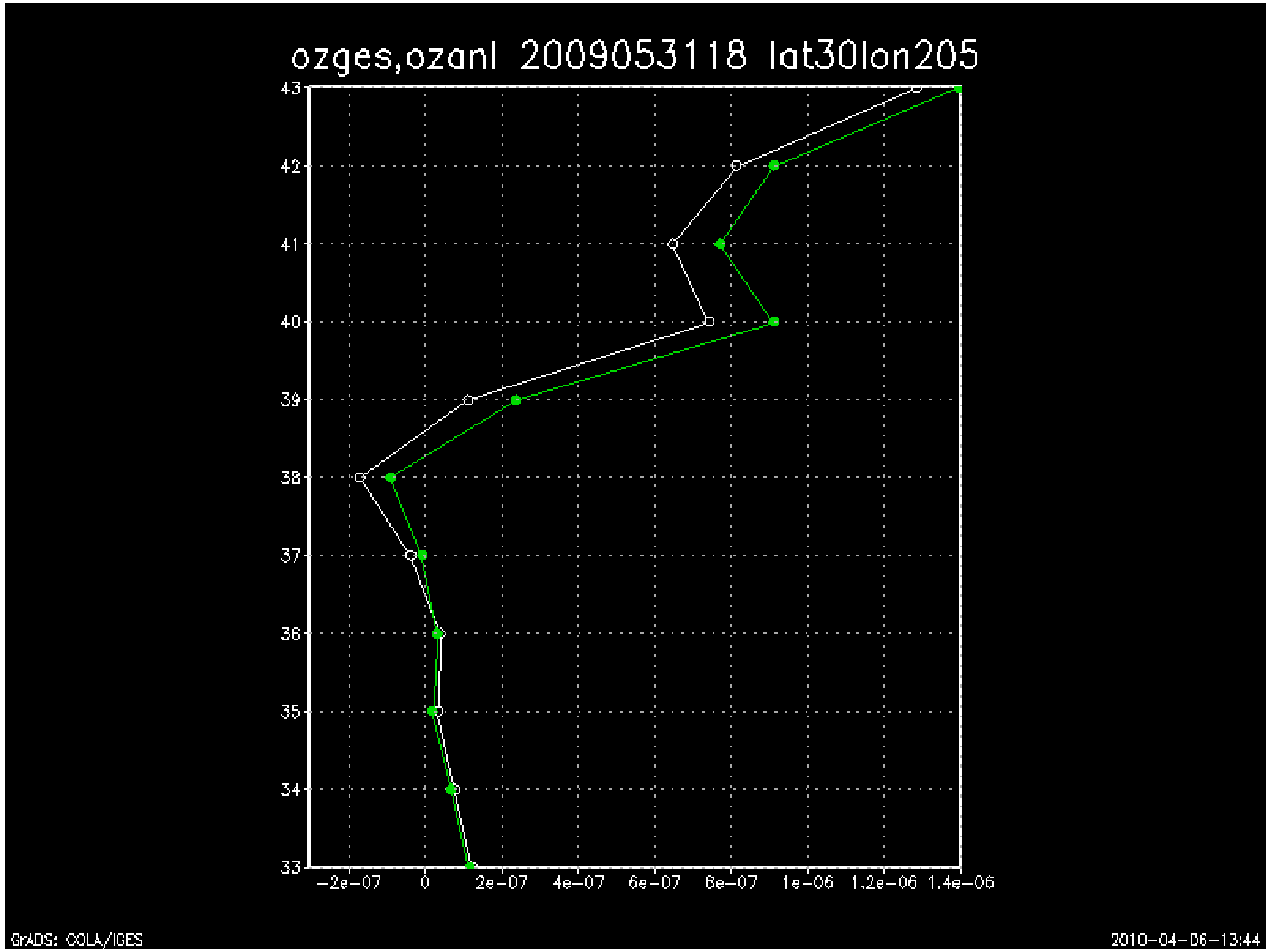
- Defined layer-ozone as control variable (instead of ozone mixing ratio)
- Layer ozone observations from SBUV used
- Sharp changes of ozone mixing ratio in vertical direction

    -Vertical correlation with layer ozone / ozone mixing ratio

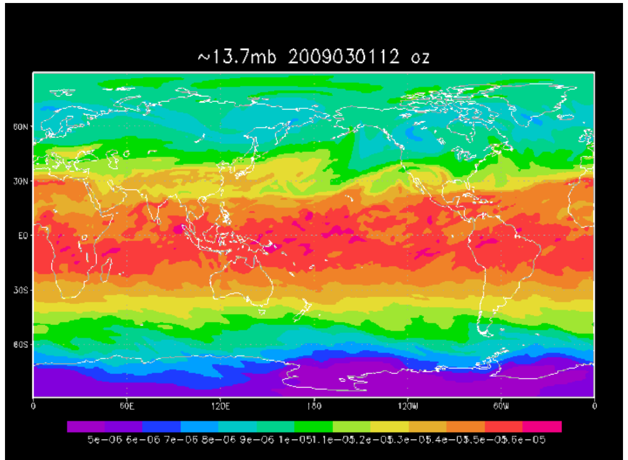
- Fraction of zonal mean of first guess ozone defined as B variance to account for seasonal change of ozone variance

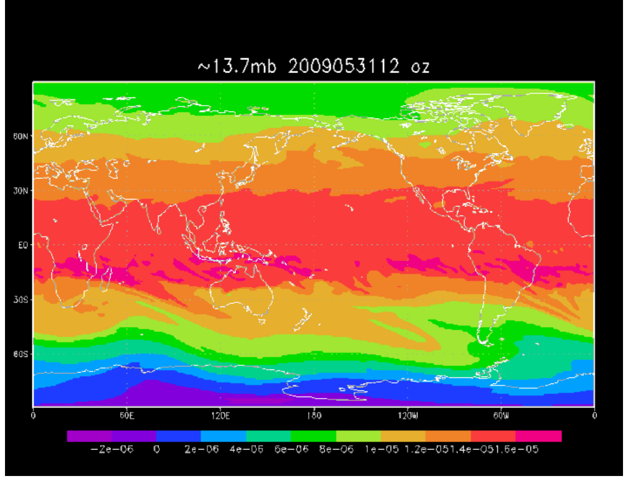
Problem with ozone analysis

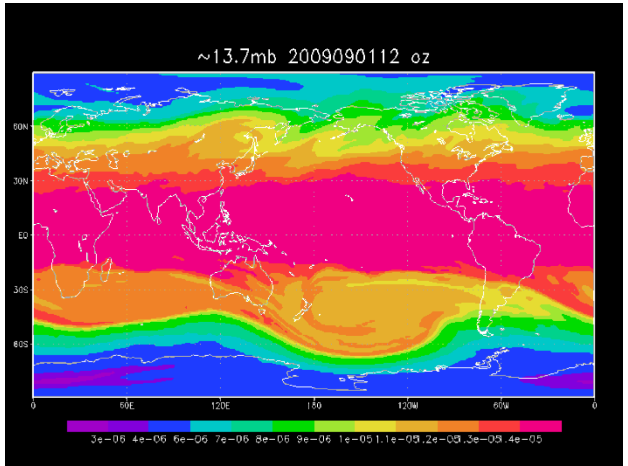
-negative o3 at z=38 & 37

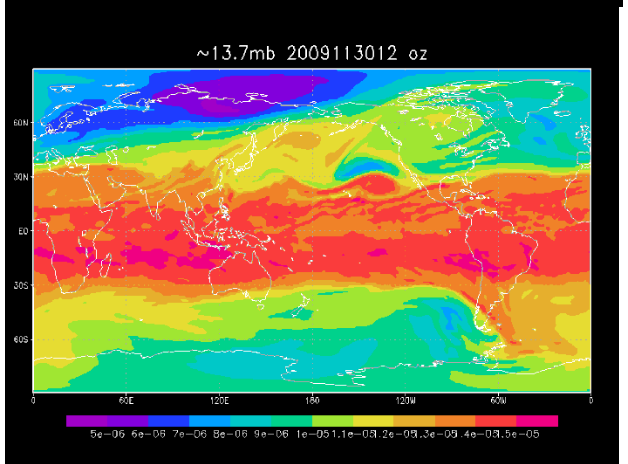


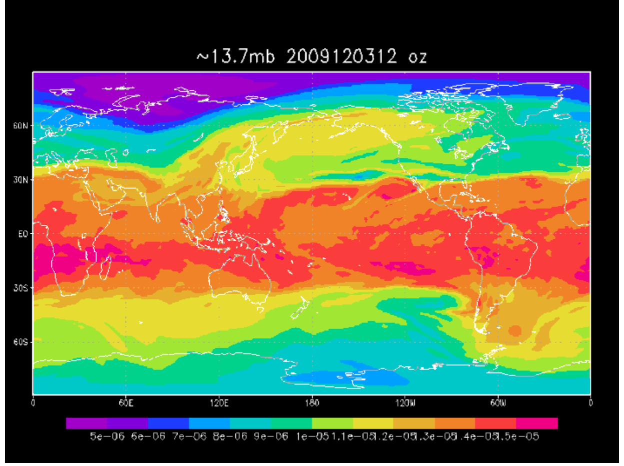
Seasonal change of ozone







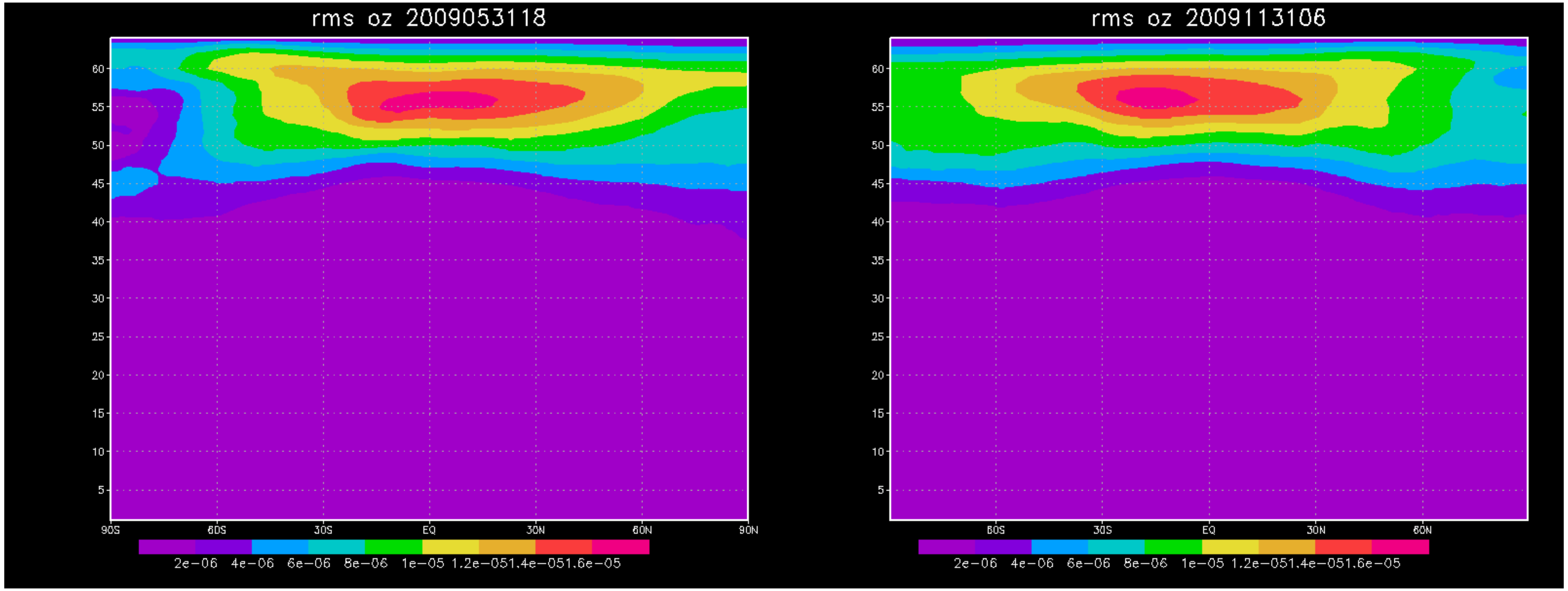




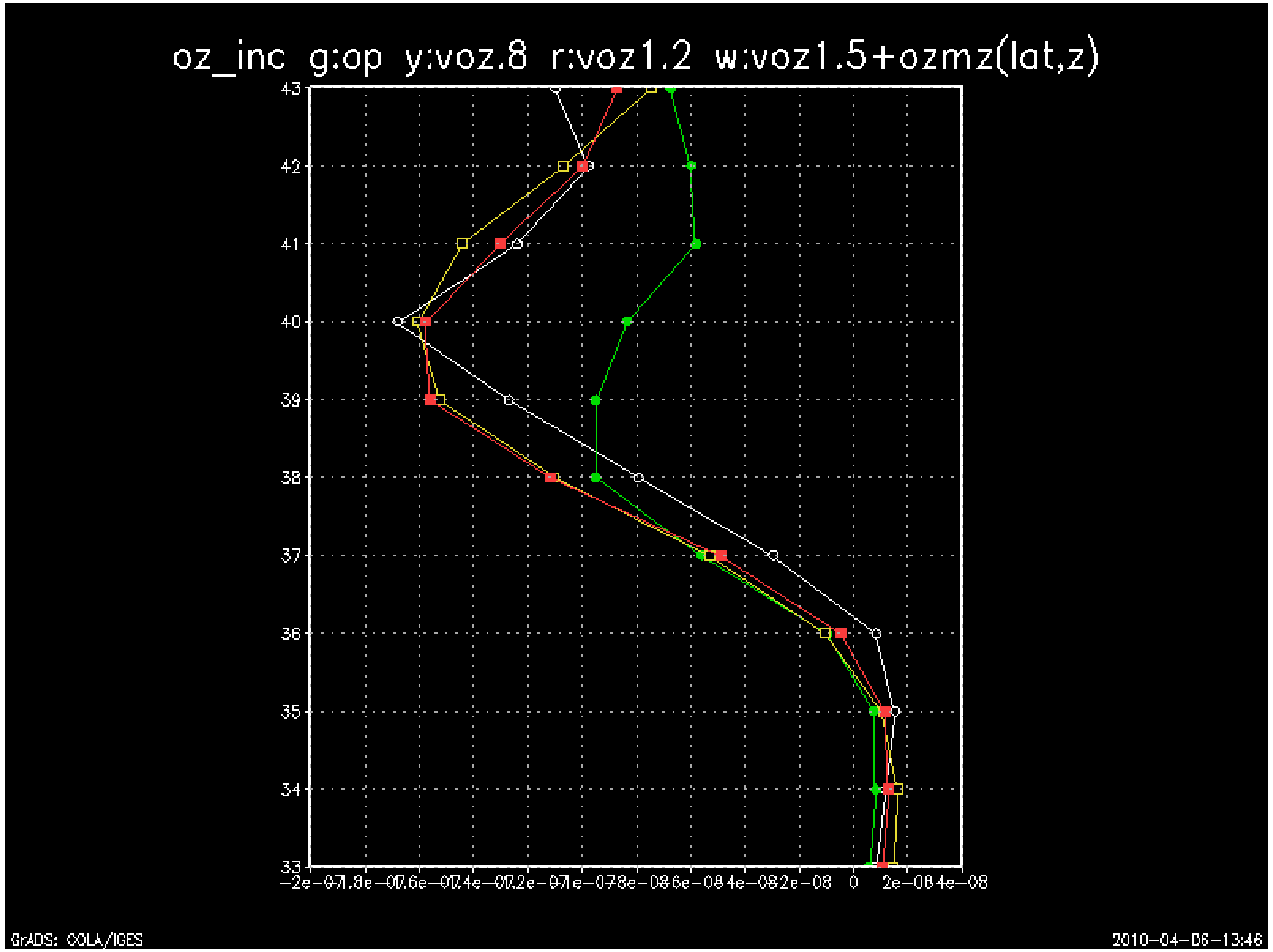
B variances defined as function of latitude and height

Oz background error variances

-derived from 20% of ozone first guess zonal mean

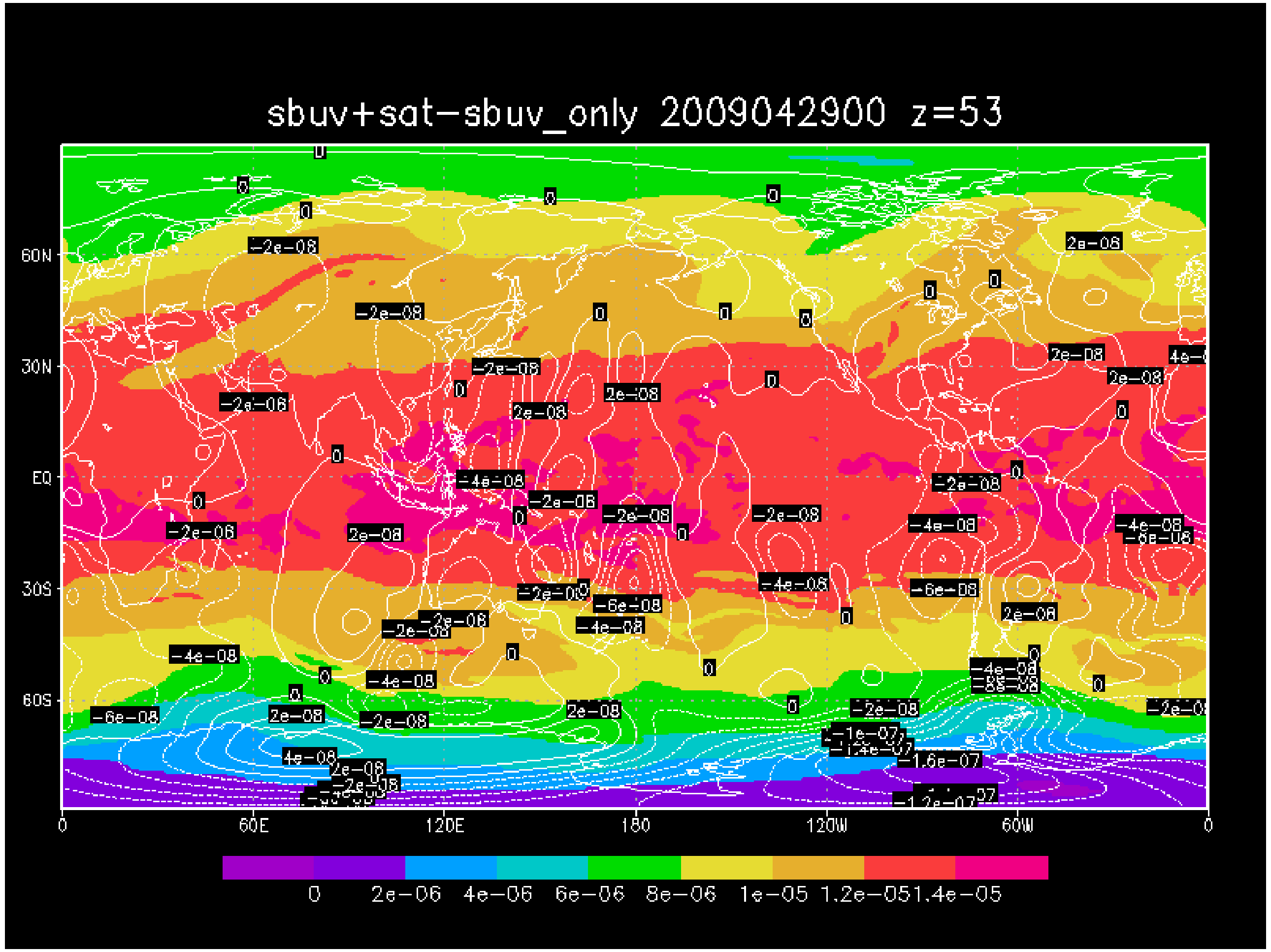


impact of variance & vertical covariance



Problem with ozone analysis

-Importance of CV variances



Note the negative ozone in the ozone hole near south pole

- through use of satellite data (t,q,oz,p)

Real-Time Mesoscale Analysis (RTMA)

- RTMA: 2D Variational surface analysis run hourly
  - Use NCEP's Gridpoint Statistical Interpolation (GSI) system
  - Temperature, moisture, pressure, winds (speed/direction/gust), visibility
  - 2.5 km grid over CONUS, 3 km grid over Alaska, Hawaii and Puerto Rico
  - Assimilated surface obs include METARs, buoys, and mesonets
  - Background field: 1-hour forecast from HRRR over CONUS
    Local NAM nests for Alaska, Hawaii, Puerto Rico
- Purpose of RTMA: Used for situational awareness, verification and tuning of blended model output
- Unique challenge: Analysis should match observations as closely as possible; well-balanced initial conditions for model initialization is not the priority
- URMA: runs 6 hours after analysis time for verification/tuning

Courtesy of Steve Levine , NCEP

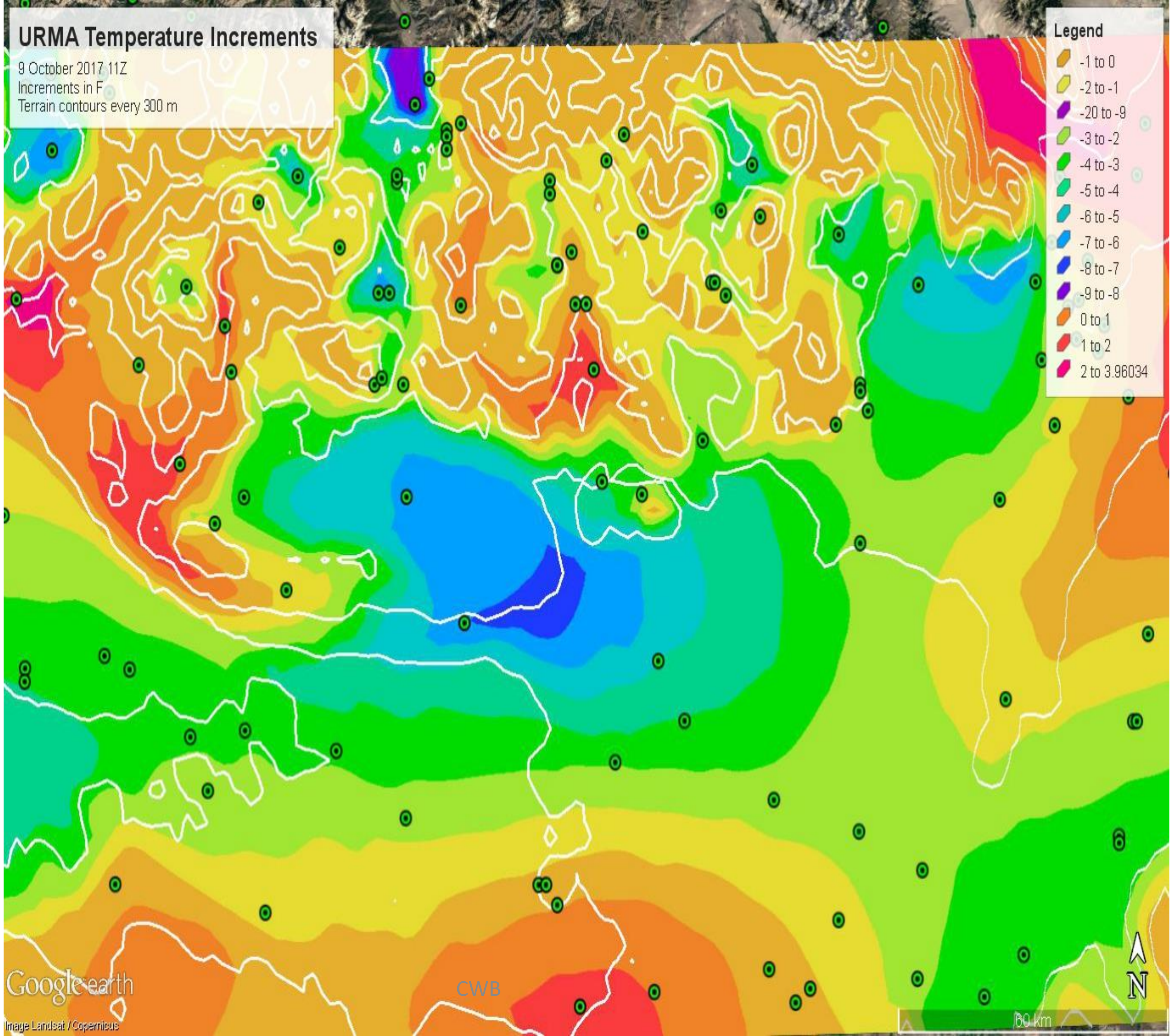
Anisotropic covariances

Increments follow terrain

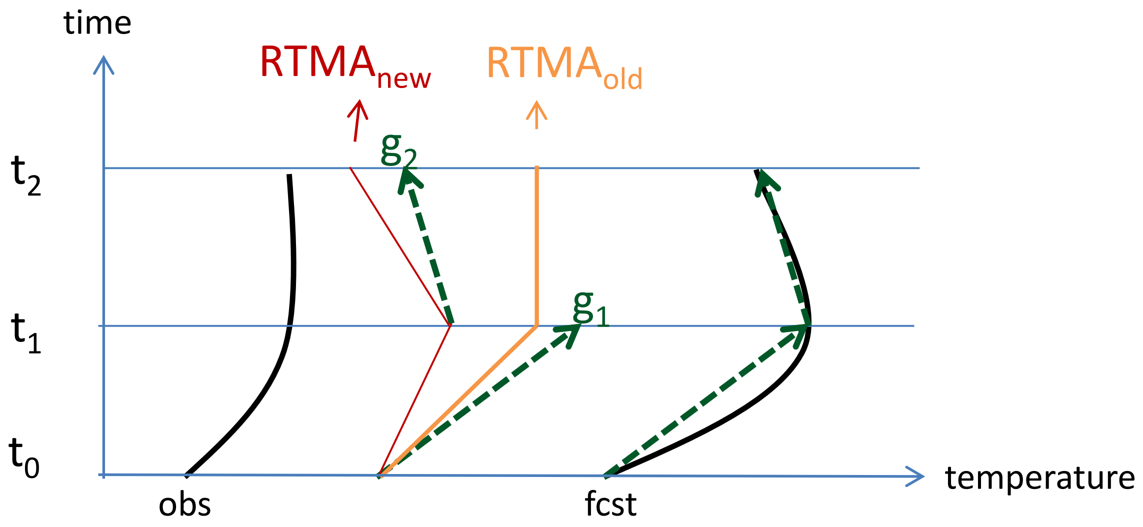
Green dots are obs used

October 9, 2017 Southern Idaho

-Steve Levine, NCEP



Proposed new first guess for RTMA



Last RTMA analysis + forecast tendency as first guess (see g_1 & g_2)
Converge to the observation (desired results?)

changing first guess for RTMA

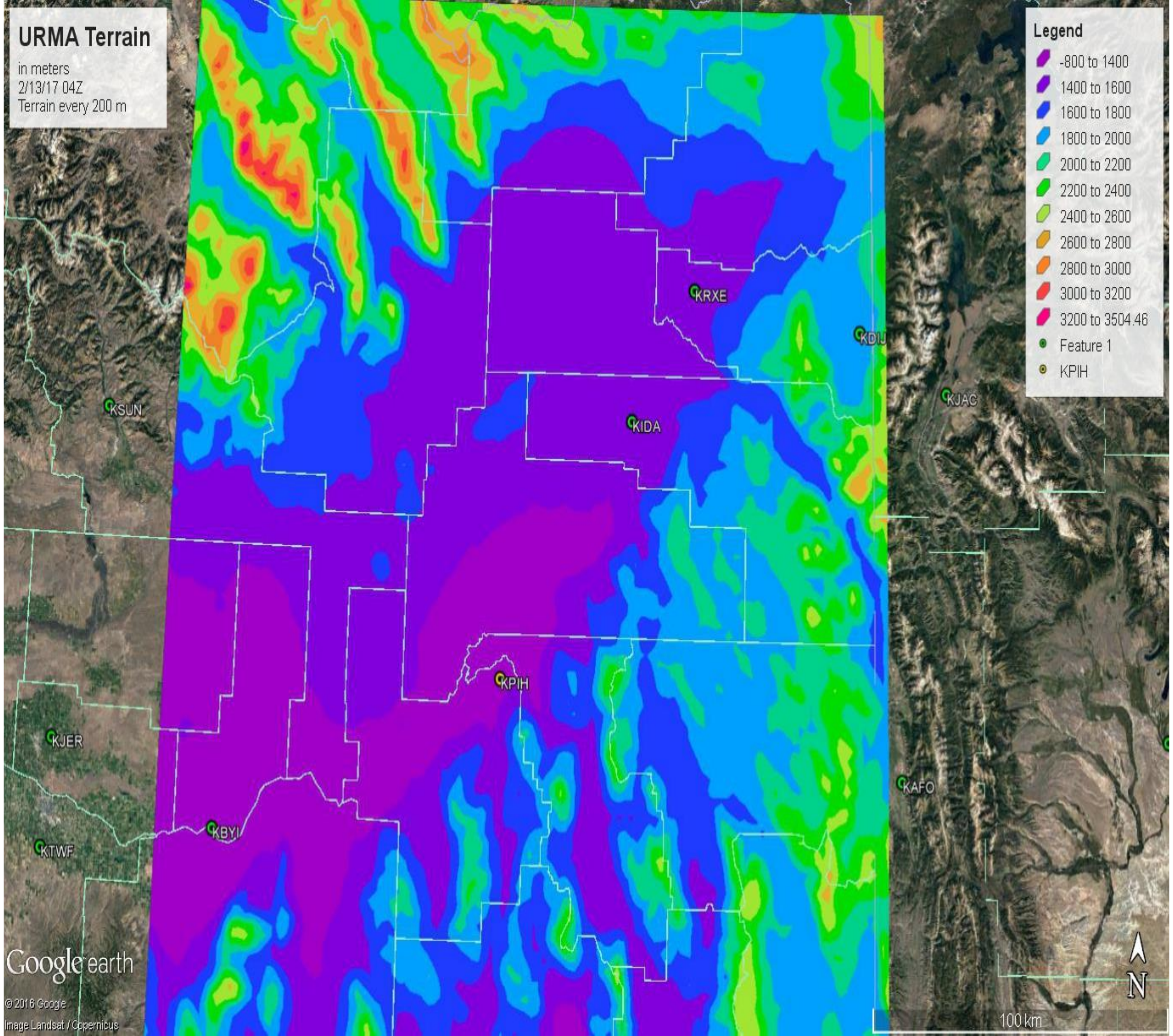
- Use previous RTMS **analysis** + time tendency from **NWP model** as first guess
- Benefit over using **climatology**: seasonal and diurnal change, large scale weather pattern change, factor in influence of terrain, radiation, moisture, wind...
- Benefit over using **NWP forecasts**: easier to correct model bias with observation, avoid some of representativeness error.

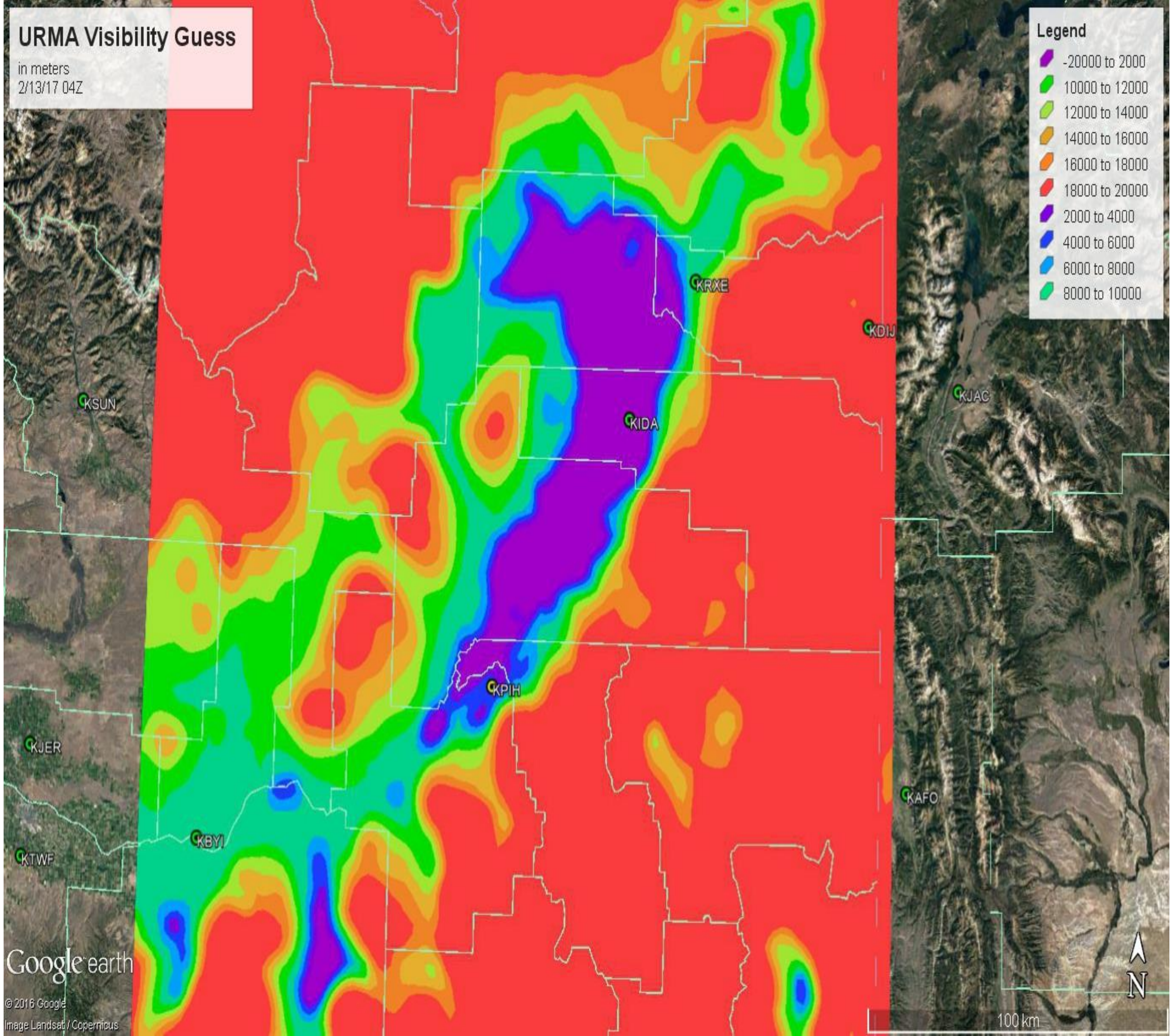
- Benefit: improve QC
- Problem: converge to bad data if they are used
- Day two: Locally choose time tendencies from various models with the guidance of observations in the area; anisotropically fill in the gap between observations
  - Tool: hybrid ensemble-variational method (provides anisotropy)

Complaint on visibility in RTMA

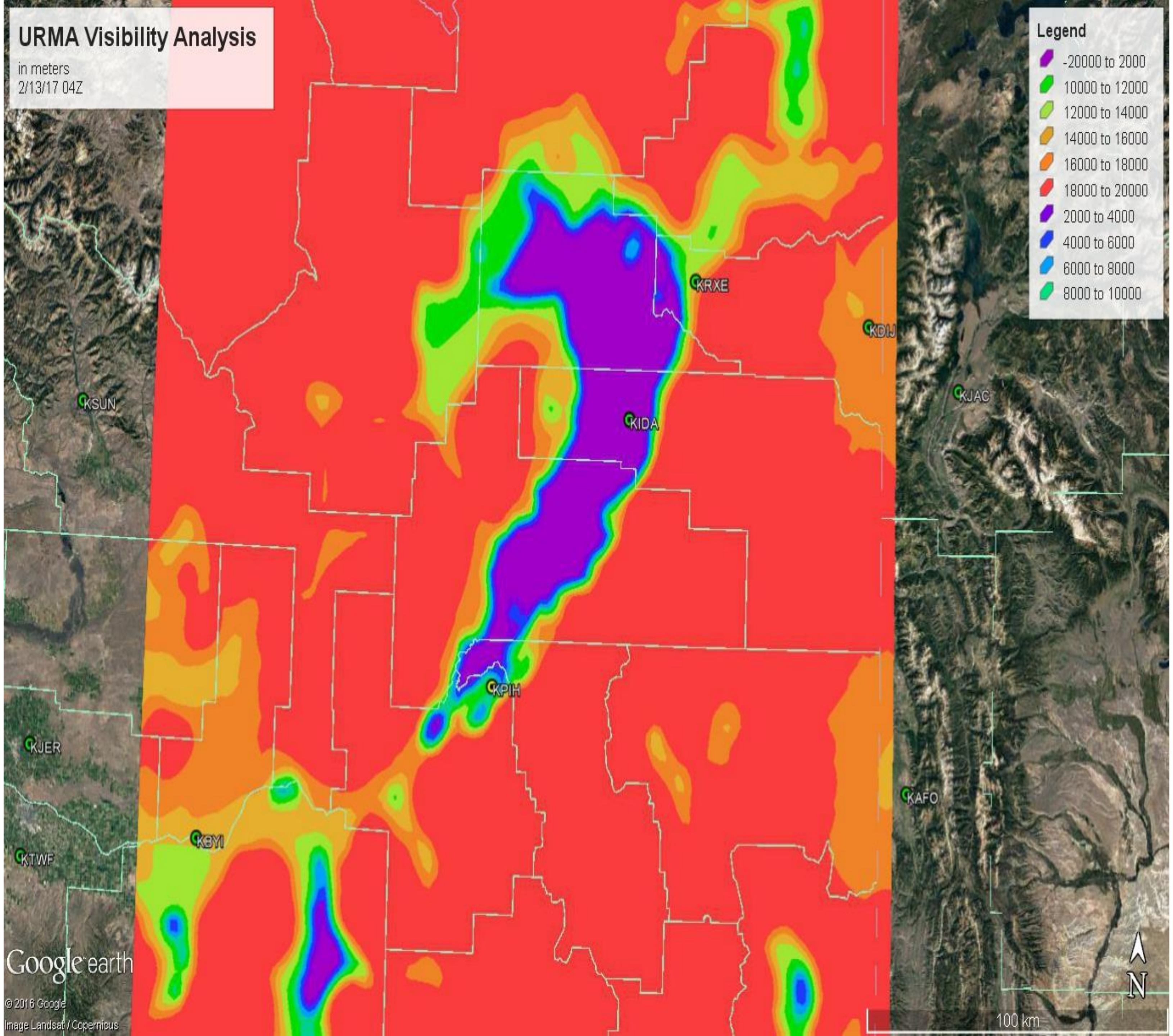
Low visibilities eastern ID 04Z, 2/13/2017

…The overnight Visibility over eastern ID were analyzed as very low on the RTMA...and URMA. I was unable to find such low visibilities reported, but did note that the low Vis was forecast on the HRRR. Looks like it made it to the RTMA….


URMA Terrain
in meters
2/13/17 04Z
Terrain every 200 m
Legend
-800 to 1400
1400 to 1600
1600 to 1800
1800 to 2000
2000 to 2200
2200 to 2400
2400 to 2600
2600 to 2800
2800 to 3000
3000 to 3200
3200 to 3504.46
Feature 1
KPIH
KRXE
KDIJ
KJAC
KSUN
KIDA
KPIH
KJER
KAFO
KBYI
KTWF
Google earth
© 2016 Google
Image Landsat / Copernicus
100 km
N



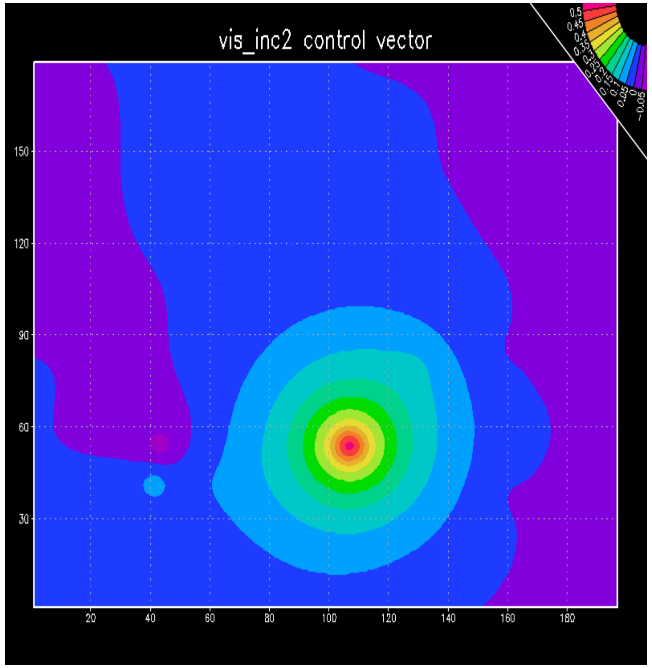
URMA Visibility Guess
in meters
2/13/17 04Z
Legend
-20000 to 2000
10000 to 12000
12000 to 14000
14000 to 16000
16000 to 18000
18000 to 20000
2000 to 4000
4000 to 6000
6000 to 8000
8000 to 10000
KRXE
KDIJ
KJAC
KSUN
KIDA
KPIH
KJER
KAFO
KBYI
KTWF
Google earth
© 2016 Google
Image Landsat / Copernicus
100 km
N

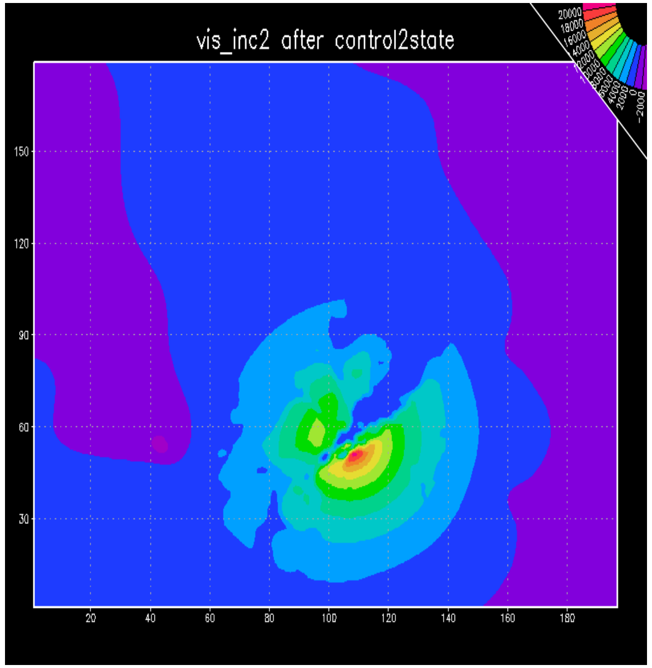

URMA Visibility Analysis
in meters
2/13/17 04Z
Legend
-20000 to 2000
10000 to 12000
12000 to 14000
14000 to 16000
16000 to 18000
18000 to 20000
2000 to 4000
4000 to 6000
6000 to 8000
8000 to 10000
KRXE
KDIJ
KSUN
KJAC
KIDA
KPIH
KJER
KAFO
KBYI
KTWF
Google earth
© 2016 Google
Image Landsat / Copernicus
N
100 km

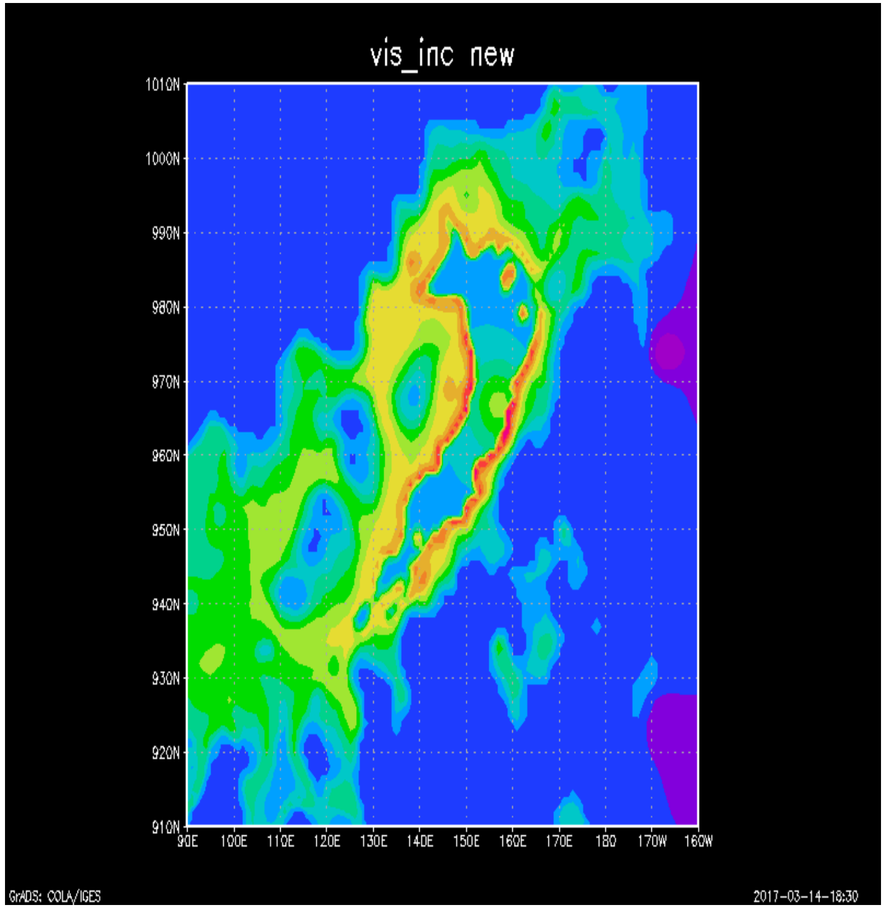
CV: log-vis

- tangent linear conversion to visibility

$$J(\mathbf{x}_f', \boldsymbol{\alpha}) = \frac{1}{2}(\mathbf{x}_f')^T \mathbf{B}^{-1}(\mathbf{x}_f') + \frac{1}{2}(\mathbf{y}_t' - \mathbf{H}\mathbf{x}_t')^T \mathbf{R}^{-1}(\mathbf{y}_t' - \mathbf{H}\mathbf{x}_t')$$



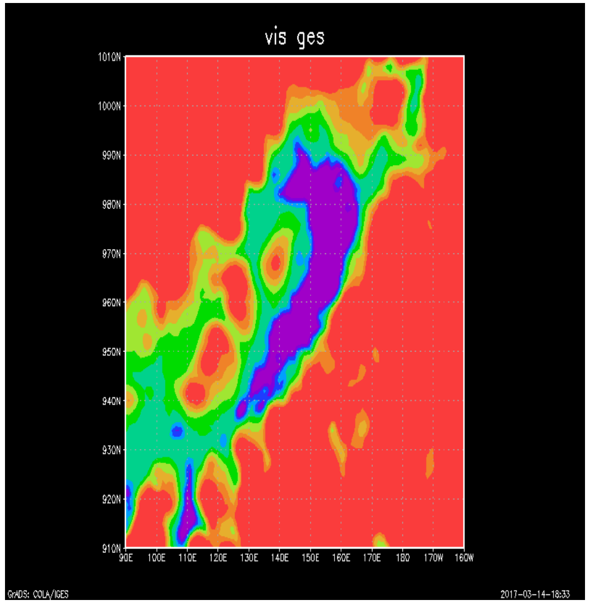




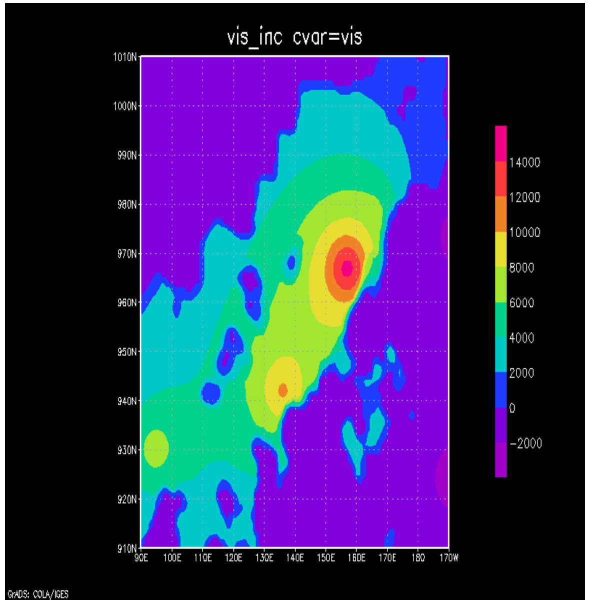
Adjust B to fit the data better, but inc with variable conversion signal from tangent linear assumption

fit the data better & without variable conversion signal

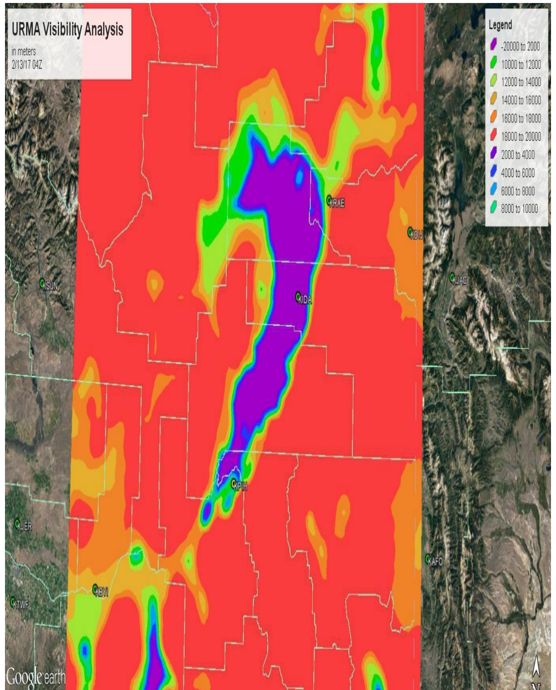
First guess



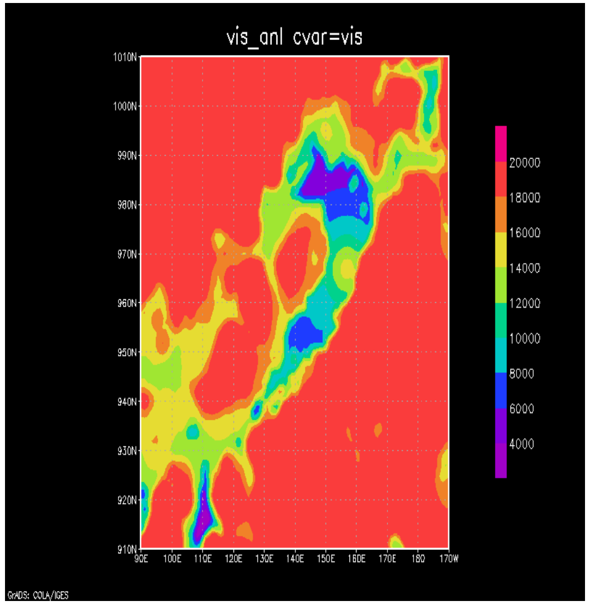
Analysis inc



Old analysis



New analysis



Options for CV_visibility

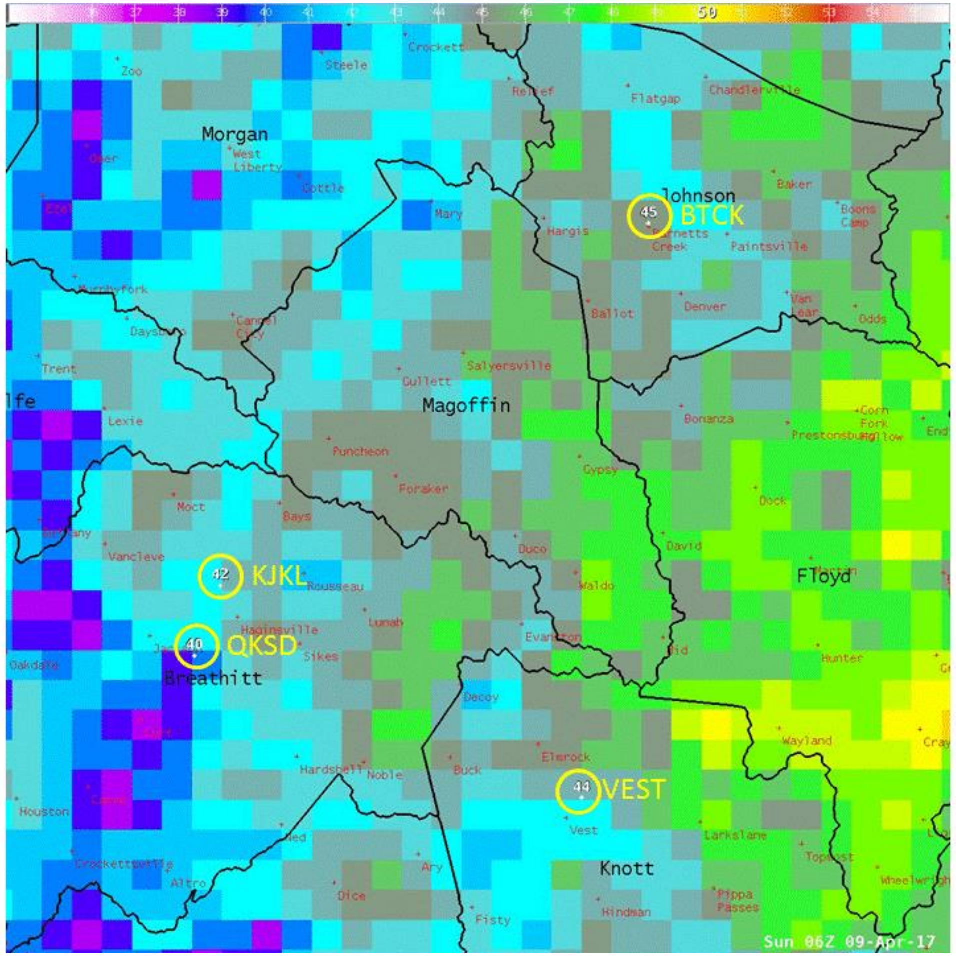
- Analysis with log-vis fits data better
- Avoid tangent linear assumption by using log-vis in inner iterations (both CV and observations)

(More discussion on logarithmic variables later)

Ridge/Valley Temp in E. Kentucky
06Z, 4/17/2017

Ridge/Valley Temp in E. Kentucky

URMA
06Z
Sunday
April 9th,
2017



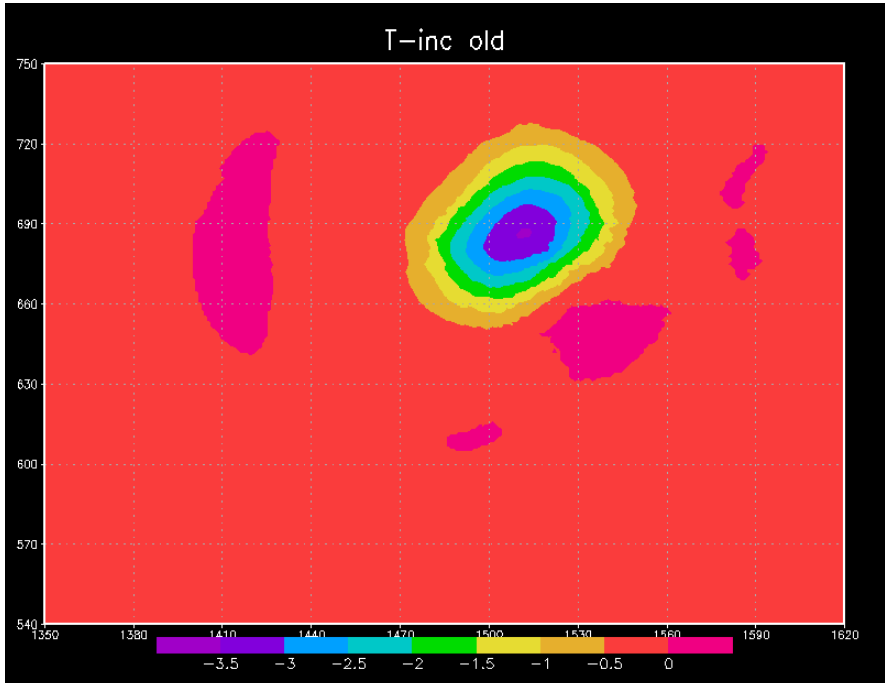
Ridge/Valley Temperature Differences Not Being Captured

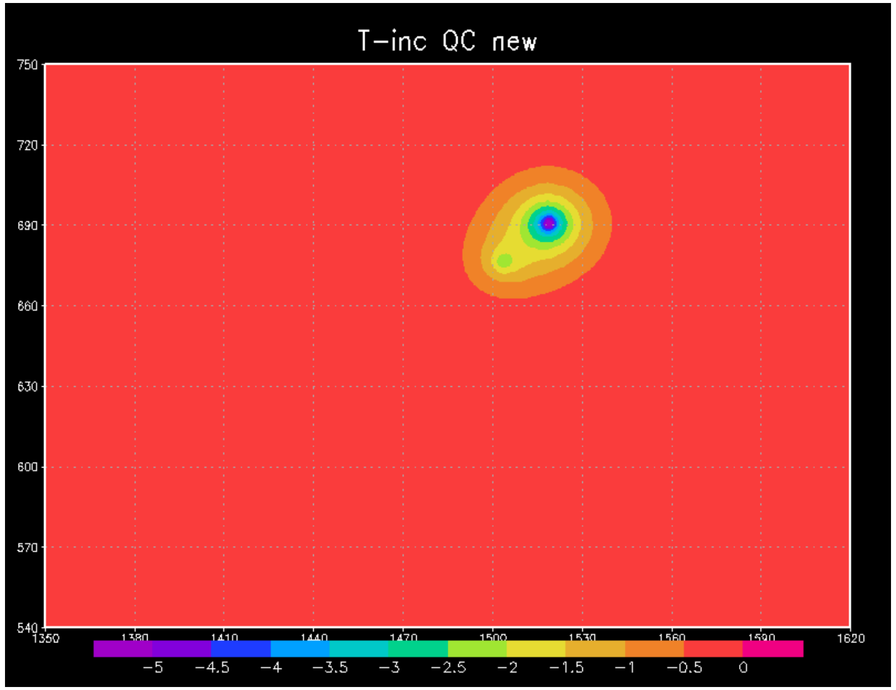
06Z Sunday, April 9th, 2017

Site	ID	Lat	Lon	Elevation	Temperature (Acutal)	Temperature (URMA)
Jackson 3SE (KY Mesonet)	QKSD	37.54	-83.34	688 feet (Valley)	37	40
Paintsville 4W (KY Mesonet)	BTCK	37.83	-82.88	755 feet (Valley)	34	45
Jackson, Carroll Airport	KJKL	37.59	-83.31	1381 feet (Ridge)	53	42
Hindman 5N	VEST	37.41	-82.99	1556 feet (Ridge)	46	44

Use only obs from the 4 stations

QKSD,BTCK,KJKL,VEST





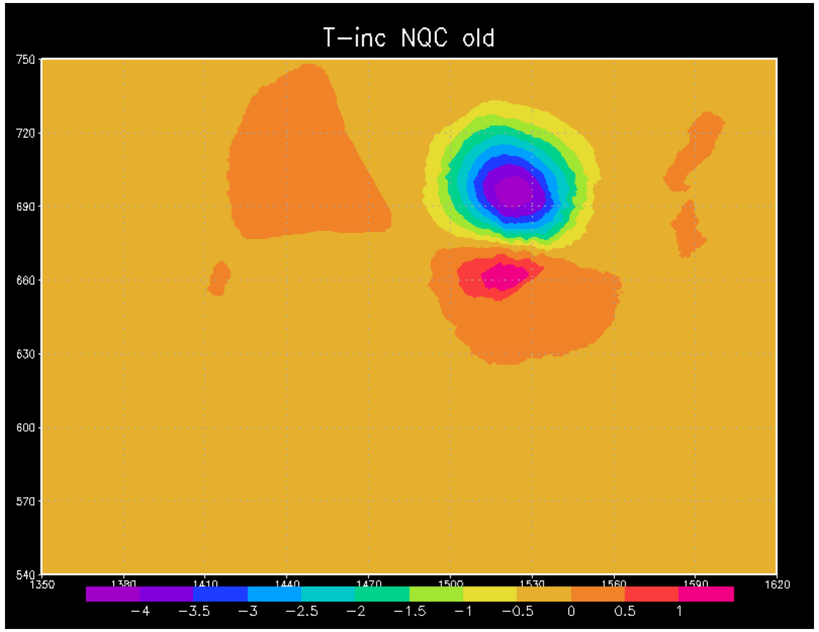
New: adjust horizontal scale

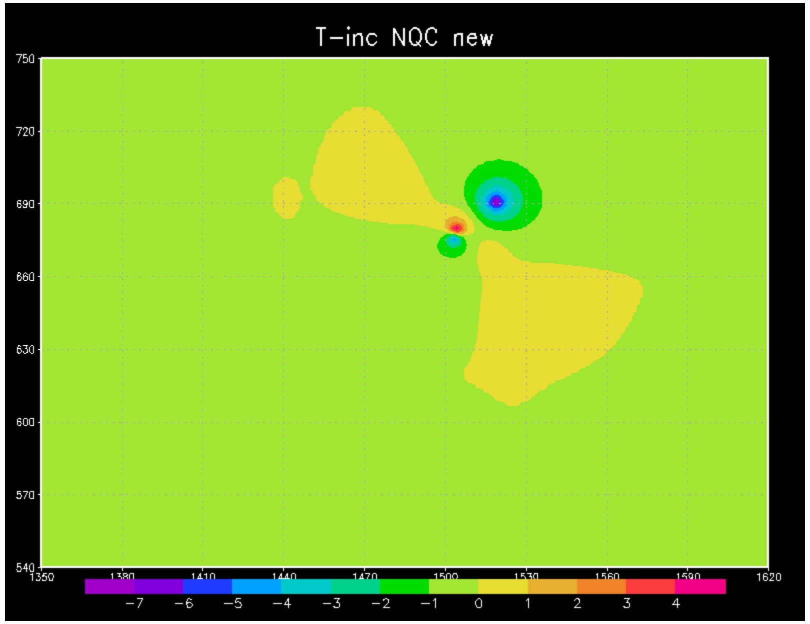
Type 187: KJKL gross checked (not used with 4.57 deg K innovation)

Type 195: 26 used, 13 Hilbert_curved

T-inc using all obs from the 4 stations

by relaxing gross check and turning off VarQC & Hilbert curve in both old and new





Adjust H scale & All 40 obs from 4 stations used

Anl rms fit: old: 3.04 new: 0.81

Max O-A : old:5.29/-3.89 new: 2.28/-1.42(degree K)

Assimilation of Radar Reflectivity

Liu et al.(2017)

- To directly assimilate reflectivity (Z) and radial velocity (Vr) data within a variational (either pure variational or hybrid) framework, some issues associated with the **nonlinearity of the reflectivity operator** arise and special treatments are needed.
- We consider two choices of **moisture control variable**, namely, mixing ratio (CV_q) and logarithmic mixing ratio (CV_logq)

Gradient of cost function and tangent linear operators for hydrometeors
(reflectivity from rain as example)

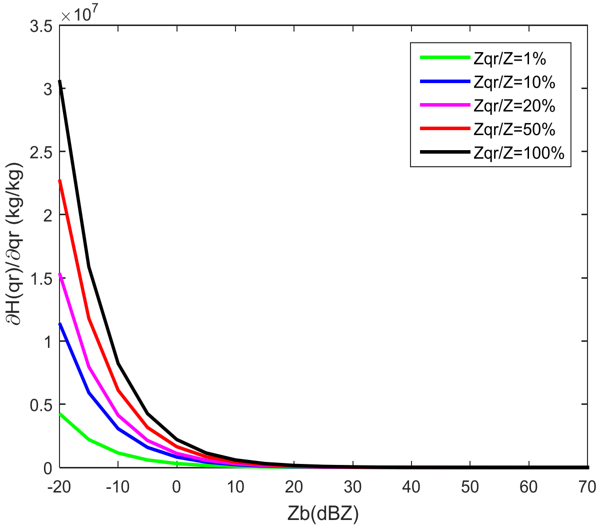
CV_q: $$Z_{er} = 3.63 \times 10^9 \times \left(\rho q_r\right)^{1.75}$$

CV_logq: $$Z_{er} = 3.63 \times 10^9 \times \left(\rho 10^{q_r}\right)^{1.75}$$

Jacobian of reflectivity

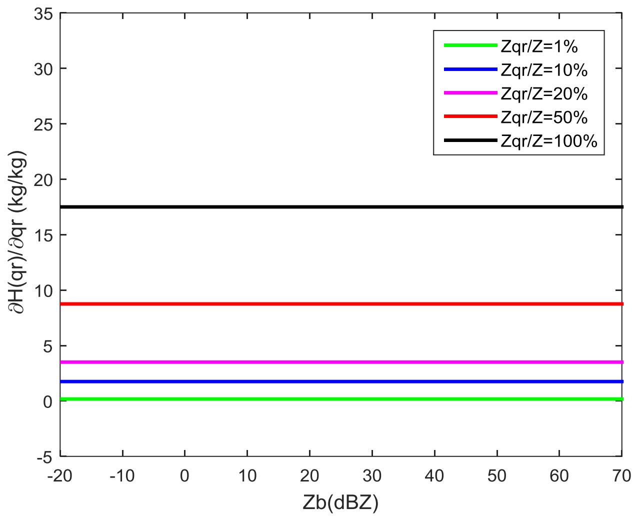
Liu et al (2017)

Issues with Using CV_q



- When background Z is below ~20 dbz, the gradient is very large.
- The gradient of the cost function becomes dominantly large when the background Z values are small, making the assimilation of Z where background Z has high values and of all Vr data ineffective (having little impact).

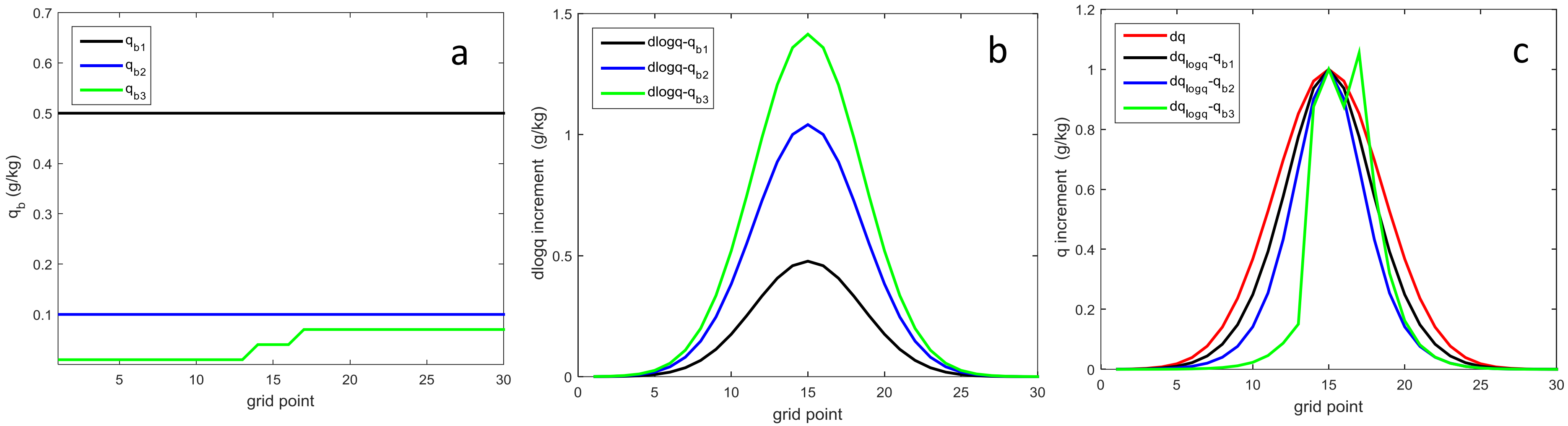
Benefits of Using CV_logq



- The gradient range is from 0 to 17. 5
- The ineffectiveness of assimilating Vr and Z where background Z with high values can be avoided.

Issue with CV_logq: background error structure

Liu et al (2017)



Single observation test for CV_logq and CV_q with different background. (a) background q_b; (b) logq analysis increments using Gaussian background error correlation with the same scale of CV_q; (c) q increments derived from CV_logq analysis increments (black, blue and green) and CV_q analysis increment (red).

Problem: Δ Logq -> Δ q is function of background q

fixing the problems on assimilating reflectivity

Liu et al (2017)

solutions:

1) setting a lower limit to mixing ratio in Z observation operator (qlim);
2) avoiding very weak reflectivity observations where background reflectivity is also very week (limit on data);
3) assimilating Z data in a separate analysis pass from Vr observations in the high background Z region when using CV_q (Zlim_Pass)
4) setting a lower limit to X_b (analysis increments) when converting analyzed logarithmic mixing ratio to regular mixing ratio ($X_b$lim).

Summary

Reasons for poor analysis fit to obs

- Variances, Gaussian CV, flow/seasonal dependent B
- Scales of B: vertical & horizontal structures
- Data quality control
- Source of penalty (clear area in reflectivity)
- Tangent linear assumption

Summary

background error variances

- O3_CV defined as layer ozone / ozone mixing ratio: convergence, vertical correlation, accurate variance
- Realistic variances of each variables for nonconventional obs

  i.e. satellite variances (t,q,oz,p), radar reflectivity(q,qr,..)
- Realistic variances of each variables for CV_nrh
- Linear visibility & log-visibility: Gaussian B, better fit to data
- Climatological averaged humidity variances & ensemble estimated humidity (ozone) variances: more accurate wind adjustments through the 4D-Var tracing effect