Recent Results on Quality Control, Simulation and Assimilation of GPS RO Data

Xiaolei Zou
Florida State University
Email: zou@met.fsu.edu
Phone: 850-644-6025

Collaborators of this work: J. Hajj, H. Shao and Z. Zeng

FORMOSAT-3/COSMIC Science Summer Camp in Taiwan
30 May - 3 June 2005

Outline

- **An introduction to GPS RO data assimilation**
- **Quality control**
- **Comparison of numerical results between bending angle and local refractivity assimilation**
- **A non-local refractivity simulation&assimilation**
- **Summary**

GPS RO Measurements

1. **Excess phase:** caused by the bending of the radio signal at two frequencies*: 1227.6 MHz, 1575.4 MHz.*
2. **Excess Doppler frequency shift:** estimated by the time derivative of excess phase.
3. **Bending angle:** derived from Doppler frequency shift based on satellite geometry (impact parameter is assumed constant at GPS and LEO positions). Ionospheric effect is removed.
4. **Refractivity:** calculated from bending angle through the Abel inversion (the refractivity is assumed spherically symmetric).
5. **Temperature, water vapor and pressure:** retrieved from refractivity using the hydrostatic equation and large-scale analysis.

GPS Observable --- Phase Path

Phase path:

$$L_i = \int_{GPS}^{LEO} n(f_i)ds, \; i = 1,2$$

Index of refractivity:

$$n(f) = 10^{-6}\left(\overbrace{\underbrace{k_1\frac{p_d}{T} + k_2\frac{p_w}{T} + k_3\frac{p_w}{T^2}}_{\text{contribution from the neutral atmosphere}} - \underbrace{C\frac{N_e}{f^2}}_{\text{ionosphere contribution}}}^{N}\right) + 1$$

$$k_1 = 7760\times 10^{-5}, k_2 = 3739\times 10^{-5}1, k_1 = 7040\times 10^{-5}$$

$T \rightarrow$ the absolute temperature of the atmosphere (in Kelvin)

$p_d \rightarrow$ the partial pressure of the dry air – mass(in hPa)

$p_w \rightarrow$ the partial pressure of the water vapor content (in hPa)

GPS RO Data Processing Chain (1)

Phase path $L_i,\ i = 1,2$

two frequencies: *1227.6 MHz, 1575.4 MHz.*

Excess phase $\Delta L_i = L_i - R_{GL},\ i = 1,2$

Distance between GPS and LEO

Ionosphere free excess phase

$$\Delta L = \frac{f_1^{\,2}\Delta L_1 - f_2^{\,2}\Delta L_2}{f_1^{\,2} - f_2^{\,2}}$$

Excess Doppler shift

$$f_d = -fc^{-1}\frac{d\Delta L}{dt}$$

GPS RO Data Processing Chain (2)

Bending angle

$\alpha(a)$

$$f_d = f\left(\frac{c - n_{LEO}\left(v^r_{LEO}\cos\phi_{LEO} - v^t_{LEO}\sin\phi_{LEO}\right)}{c - n_{GPS}\left(v^r_{GPS}\cos\phi_{GPS} - v^t_{GPS}\sin\phi_{GPS}\right)} - 1\right)$$

$$r_{GPS}n_{GPS}\sin\phi_{GPS} = r_{LEO}n_{LEO}\sin\phi_{LEO} = a$$

$$\alpha = \phi_{GPS} + \phi_{LEO} + \arccos\left(\frac{\mathbf{r}_{GPS}\bullet\mathbf{r}_{LEO}}{r_{GPS}r_{LEO}}\right) - \pi$$

Index of refractivity

$$n(a) = EXP\left(\frac{1}{\pi}\int_a^\infty \frac{\alpha(x)}{\sqrt{x^2 - a^2}}dx\right)$$

GPS refractivity

$$N_{GPS} = 10^6(n - 1)$$

GPS RO Data Processing Chain (3)

Assuming $N \approx N_{GPS}$

$$N = c_1 \frac{p}{T} + c_2 \frac{pq}{T^2(0.622 + 0.378q)} \Leftrightarrow N = c_1 \frac{p}{T} + c_2 \frac{e}{T^2}$$

Dry regions

Ignoring water vapor q

Dry refractivity

$$N = c_1 \frac{p}{T}$$

$p = \rho R T$

Dry air density

$$\rho = N / c_1 R_d$$

Dry air pressure

$$p(z) = \int_z^{\infty} \rho g dz$$

Dry temperature

$$T = \frac{p}{R_d \rho}$$

Wet regions

Auxiliary temperature T

Air density

$$\rho = \frac{1}{c_1 R} N - c_2 \frac{e}{T^2}$$

Total air pressure

$$\frac{\partial p}{\partial z} = -\rho g$$

Partial water water pressure

$$e = \frac{T^2 (N - c_1 p / T)}{c_2}$$

Specific humidity

$$q = \frac{0.622e}{p - 0.378e}$$

Some Unique Features of GPS RO Observations

- Not a point measurement
- Indirect measurements of atmospheric thermodynamic state
- Very high vertical resolution
- A RO can occur at any geographical location, providing a vertica profile of measurements from the top of the atmosphere to certain height in the low troposphere
- No calibration error ⟶ Long-term stability
- Available in all weather conditions
- Global, all time coverage
- Independent to other remote-sensing systems, such as infrared and microwave sounding techniques

Part I: A Quality Control Procedure for GPS RO Data

The purpose of QC is to remove erroneous data.

Outliers {
- Erroneous data
- Data deviate greatly from background

Error sources

Systematic errors

- Local multi-path
- Position errors
- Velocity errors
- Retrieval errors

Retrieval errors

- Ionosphere calibration errors
- Upper altitude boundary errors
- Errors introduced by the spherical symmetry assumption
- Errors induced by atmospheric multi-path

Test Case: CHAMP RO Data Are Used for a QC test

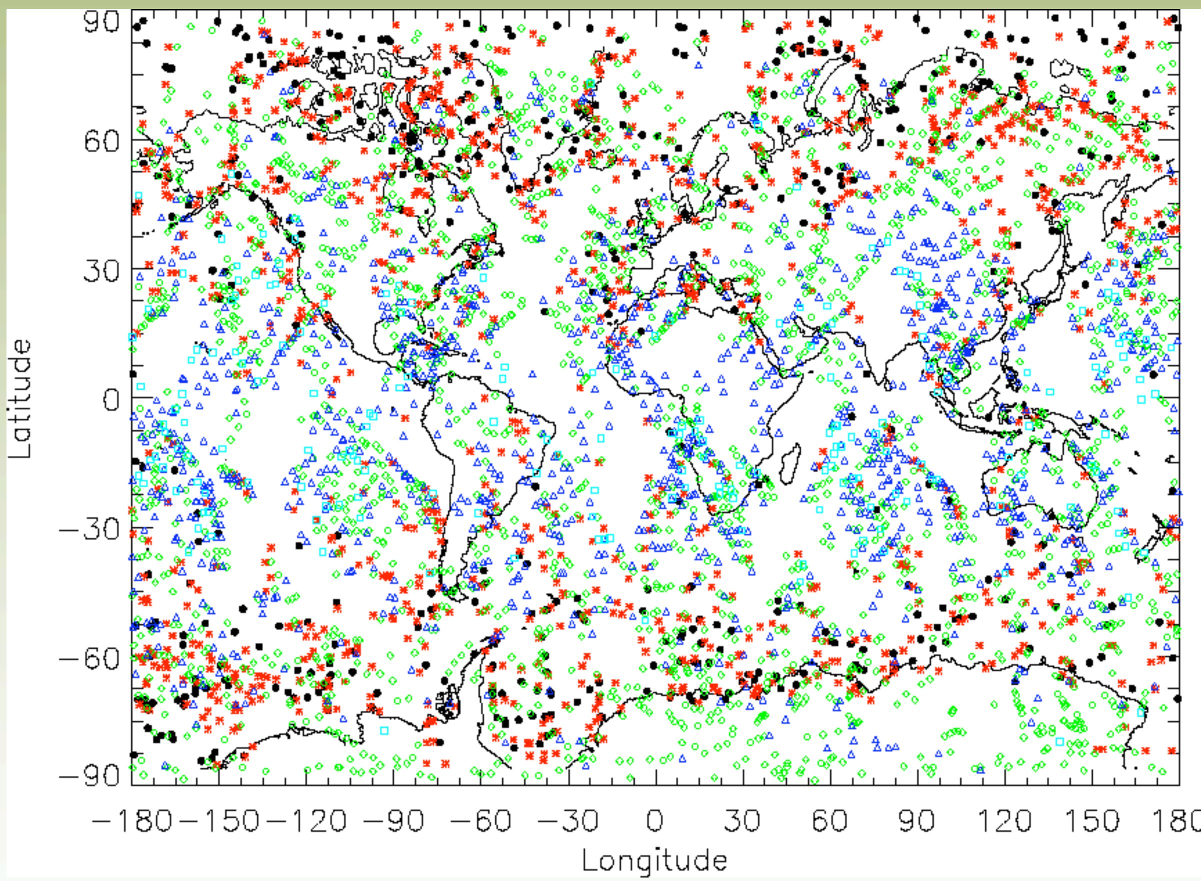
RO retrievals	Height range	Temporal range	Data source
α & N	0 ~ 20 km	One month (March 2004)	UCAR CDAAC

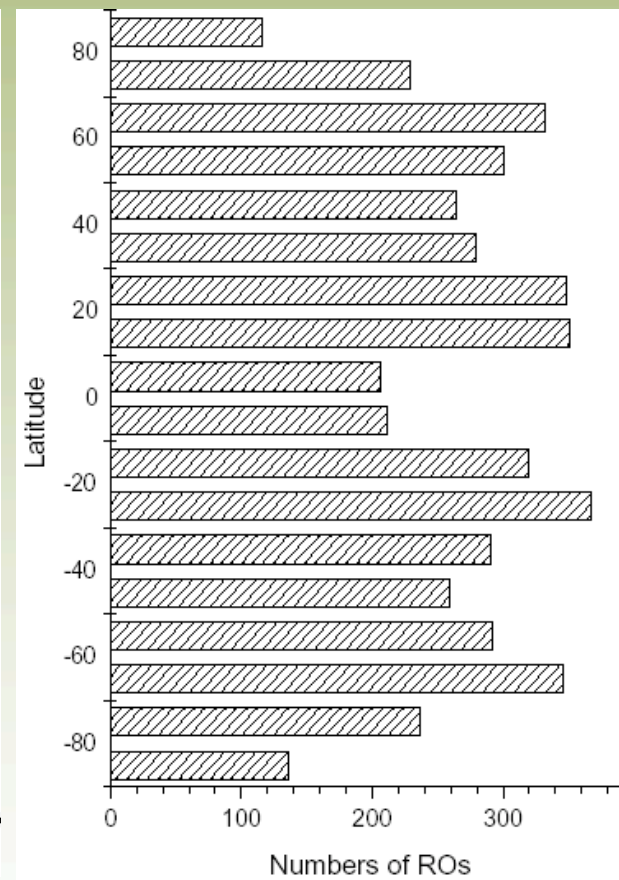
6169 ROs received totally

4884 α and N profiles retrieved

4514 profiles passed CDAAC QC

Geographical distribution of the 4884 ROs in March 2004





• < 1 km
* 1 ~ 2 km
◊ 2 ~ 4 km
Δ 4 ~ 6 km
> 6 km

Histogram of the number of RO events

QC1 --- range check:

Data points with negative values:

$\alpha < 0$ **or** $N < 0$

are removed.

Ionospheric residual in the inverse of upper troposphere.

Biweight Mean and Standard Deviation (1)

The biweight estimate is a weighted average such that weighting decreases away from the center of the distribution. All values beyond a certain critical distance from the center (controlled by a parameter "c") are given zero weight.

Sample: X = [1.01, 1.02, …, 1.09, 1000]

	Traditional	Biweight
Mean	100.95	1.05
Standard deviation	315.90	0.03
Maximum Z-score	2.85	34340.29

Biweight Mean and Standard Deviation (2)

(Hoaglin et al., 1983)

For n observations: $X_i, \quad (i = 1, 2, \ldots, n)$

1) Estimate the median (M) and median absolute deviation (MAD).
2) Calculate the weight (w_i) corresponding to each of the n observations (X_i).

$$w_i = \frac{X_i - M}{c \bullet MAD} \qquad if\ |w_i| \geq 1.0,\ set\ w_i = 1.0$$

3) Estimate the biweight mean and biweight standard deviation.

$$\overline{X}_{bi} = M + \frac{\sum_{i=1}^{n} (X_i - M)(1 - w_i^2)^2}{\sum_{i=1}^{n} (1 - w_i^2)^2}$$

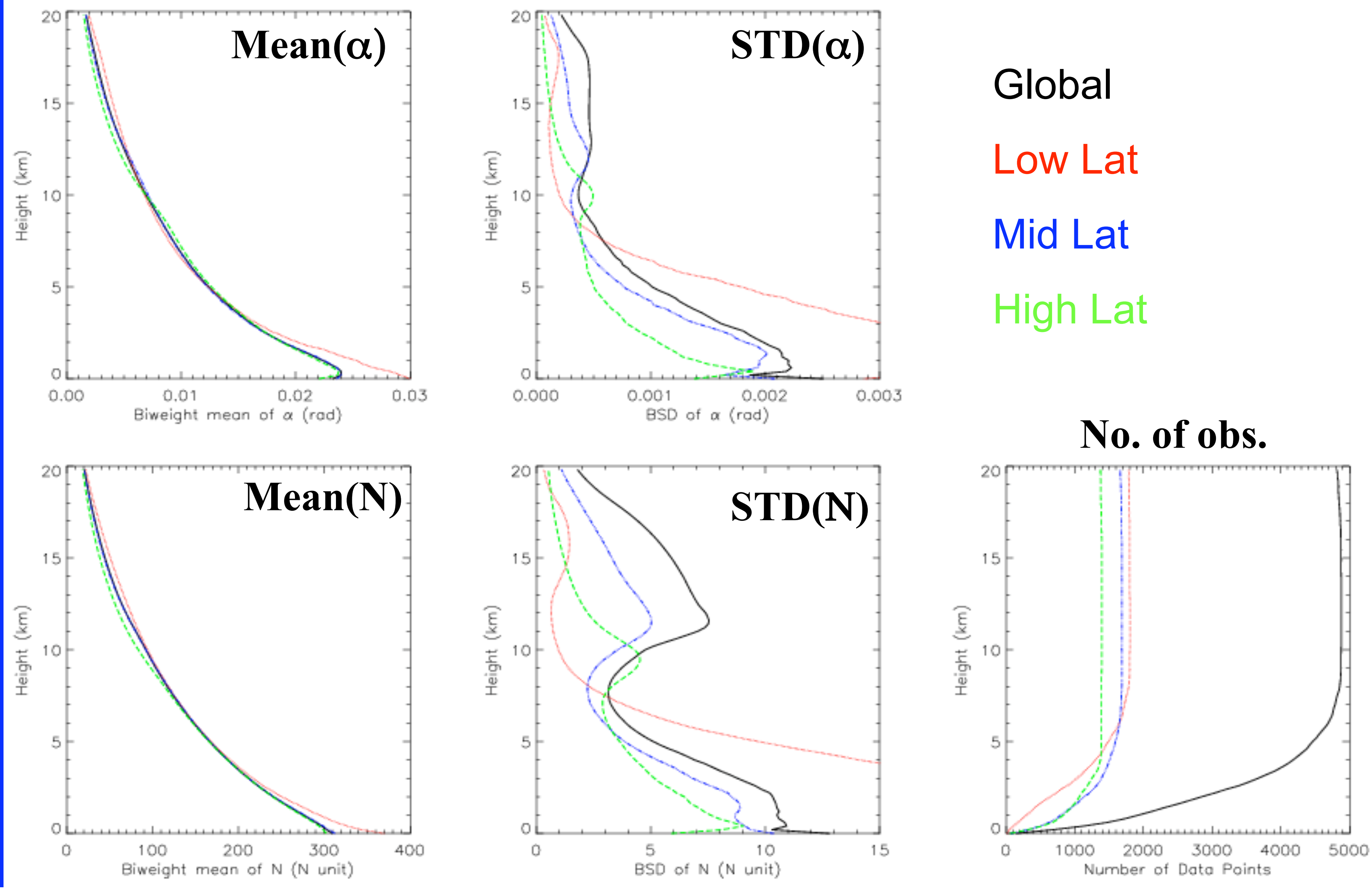
$$BSD = \frac{\left[n \sum_{i=1}^{n} (X_i - M)^2 (1 - w_i^2)^4 \right]^{0.5}}{\left| \sum_{i=1}^{n} (1 - w_i^2)(1 - 5w_i^2) \right|}$$

QC2 --- A Horizontal Spatial Consistency Check

Flag values outside of a confidence interval about the mean using a range of ±3, ±4, ±5 times the standard deviation:

$$Z_i = \left| \frac{X_i - mean(X)}{STD(X)} \right| \geq 4$$

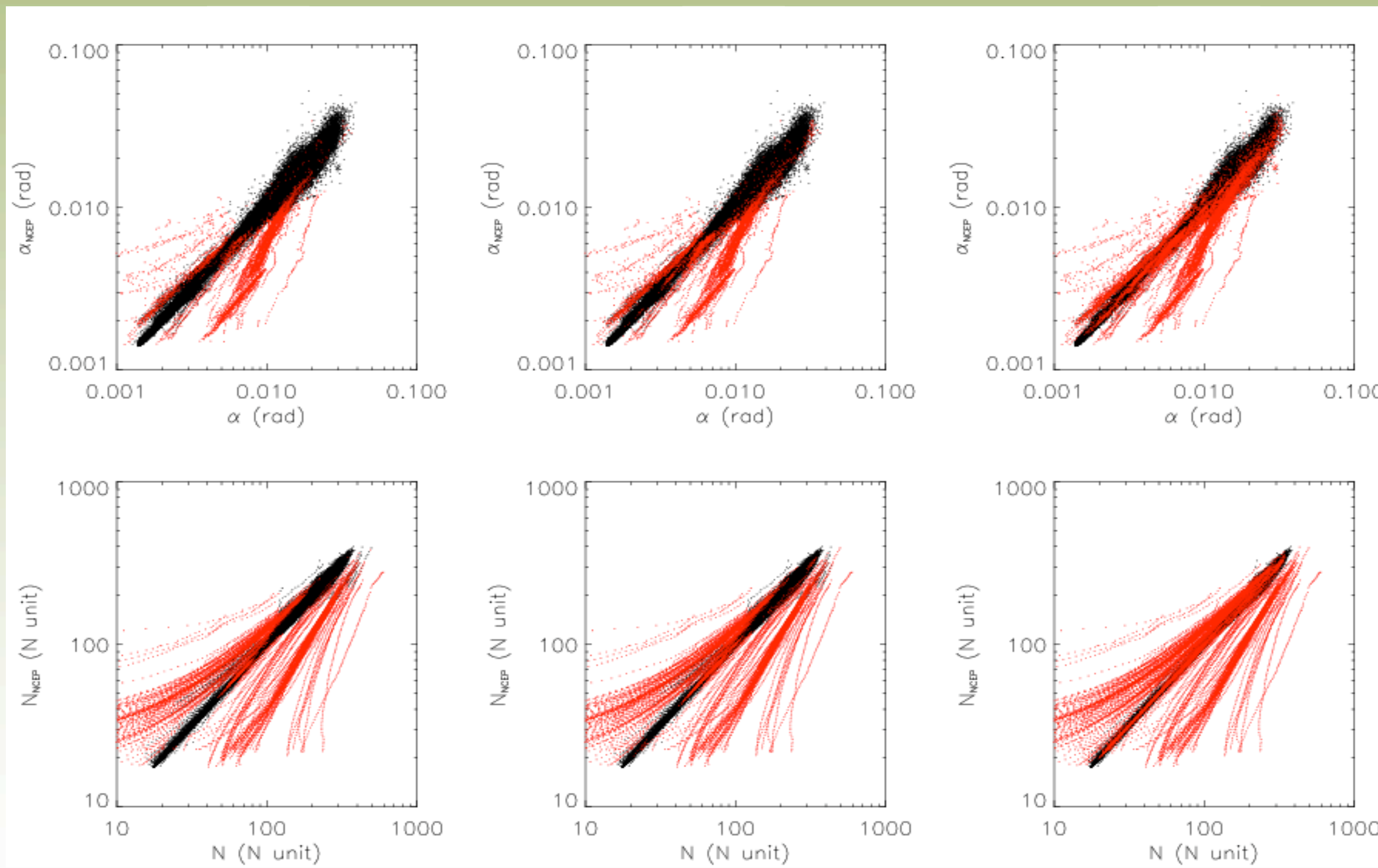
Biweight Mean and Standard Deviation of α and N



Outliers Identified by QC2 under different Z Scores

α

N

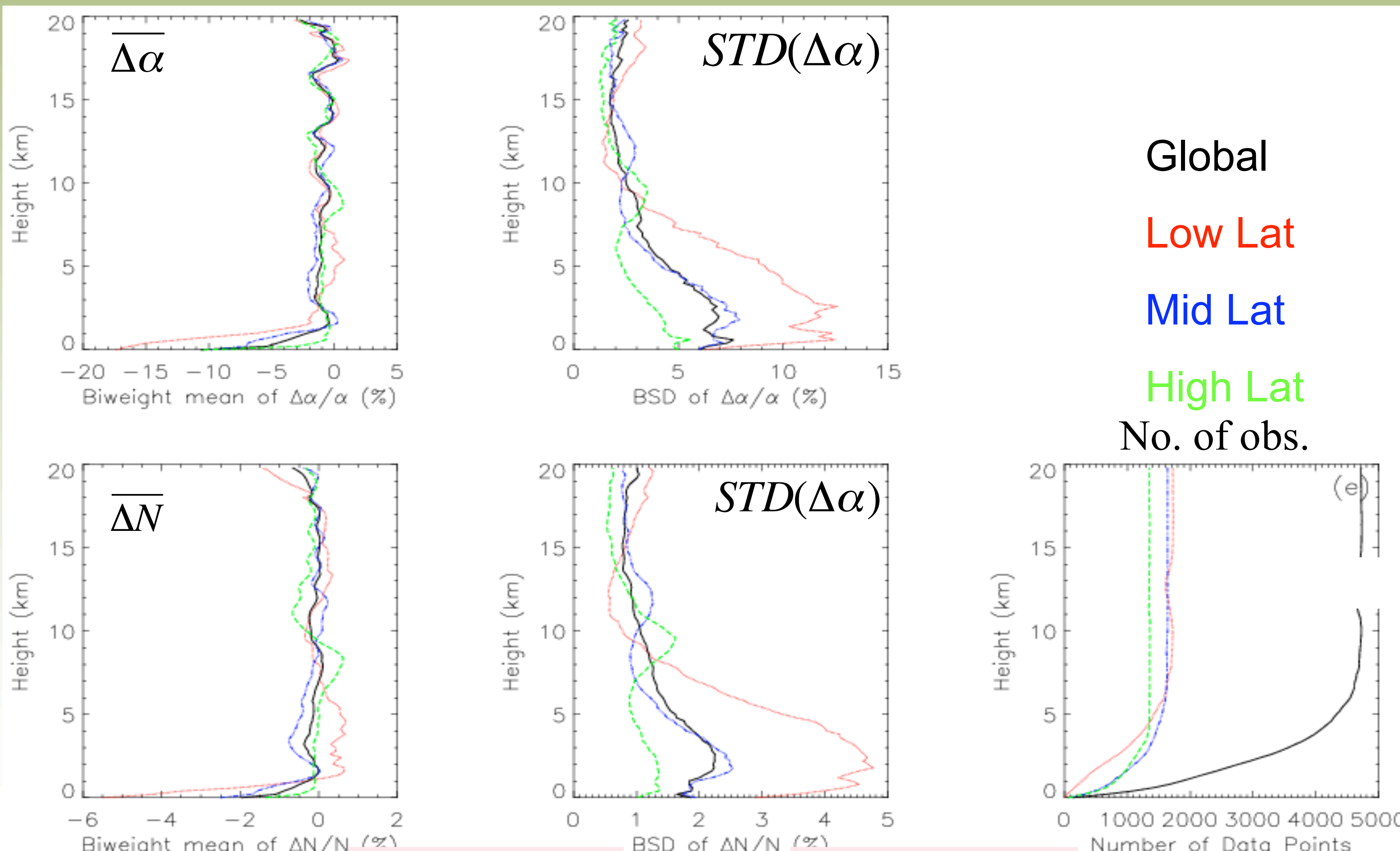


Z>5 Z>4 Z>3

Percentages of Erroneous Data Points Suspected by QC2

Outliers (%)	Z > 5	Z > 4	Z > 3
α	1.1%	1.4%	2.2%
N	2.5%	2.9%	3.7%
$\alpha \cup N$	2.6%	3.1%	4.4%

QC3 --- A Consistent Check with NCEP Analysis

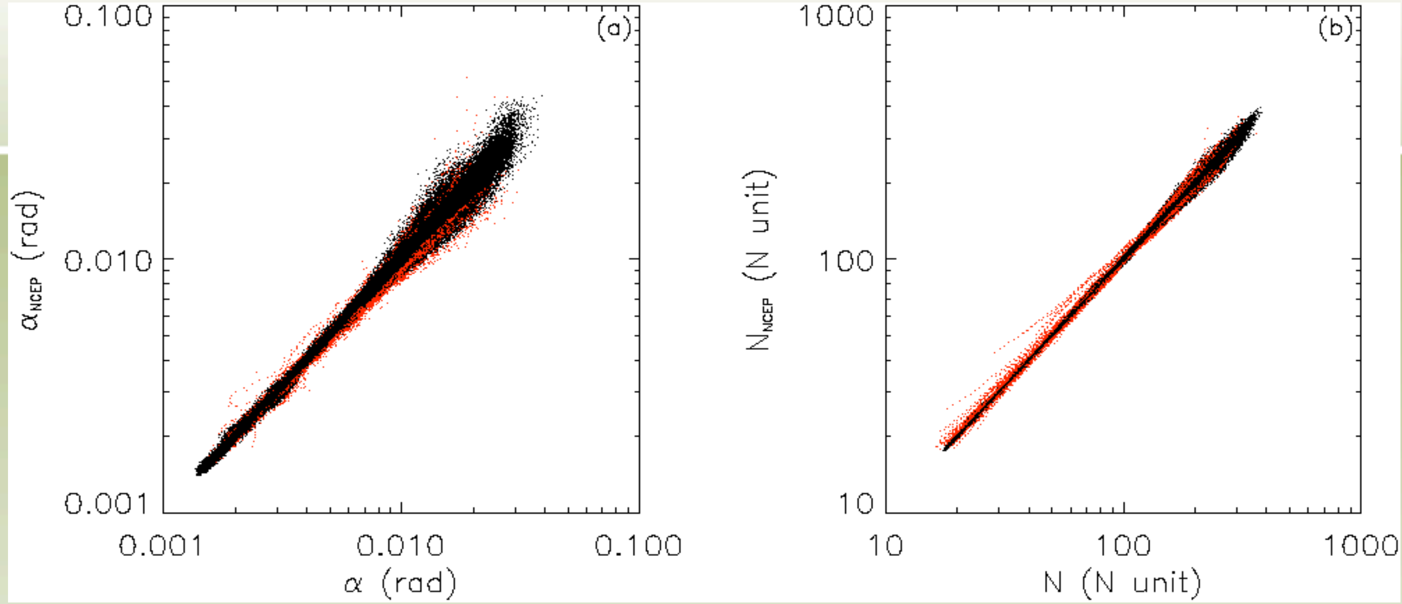


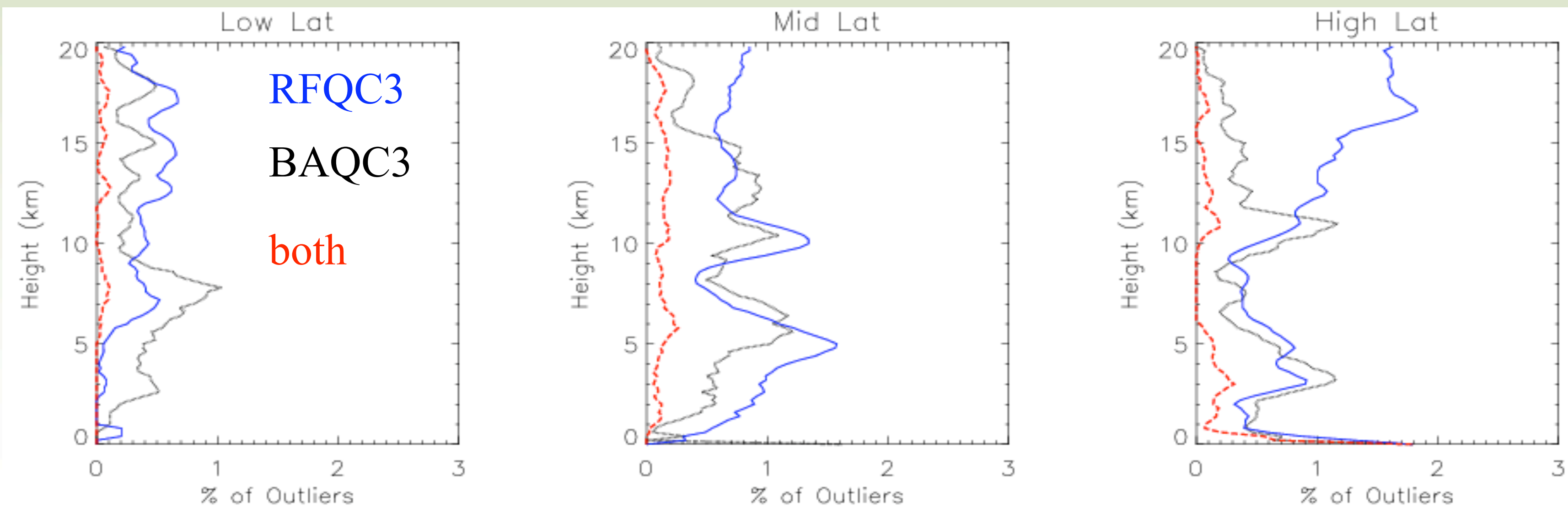
$$\Delta\alpha = \frac{\alpha^{obs} - \alpha^{model}}{\alpha^{model}}$$

$$\Delta N = \frac{N^{obs} - N^{model}}{N^{model}}$$

QC3

Z > 4


(a)
(b)
0.100
0.010
0.001
α_{NCEP} (rad)
α (rad)
1000
100
10
N_{NCEP} (N unit)
N (N unit)

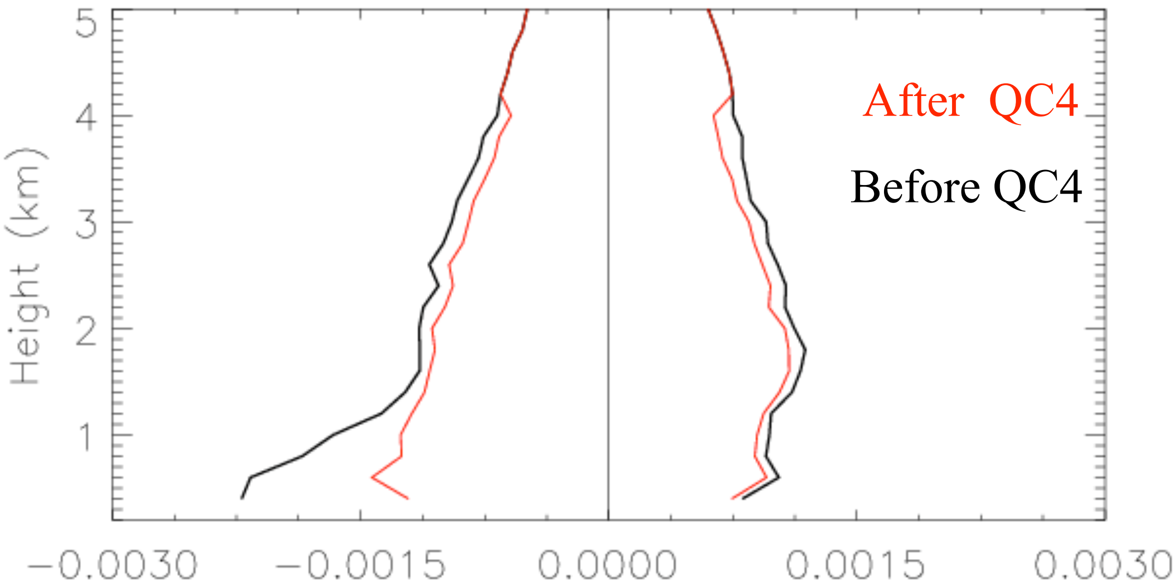

Low Lat
Mid Lat
High Lat
RFQC3
BAQC3
both
Height (km)
% of Outliers
20
15
10
5
0
1
2
3

QC4 --- A Asymmetry Check

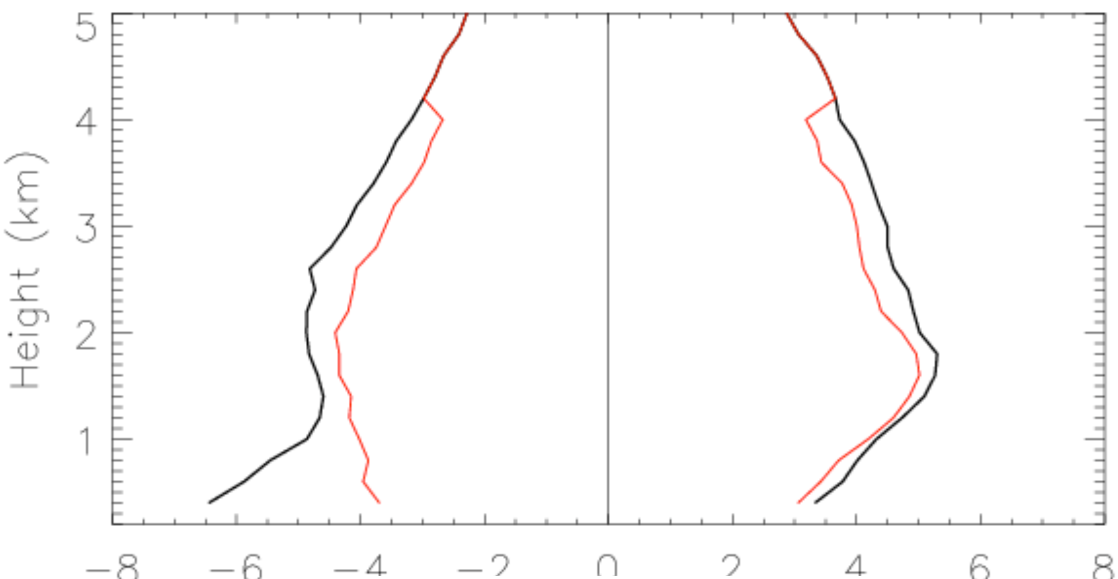
Remove the negative bias of occultation observation in the lower troposphere

Asymmetry check procedure (Lanzante, 1996):

1) estimated the median (M) of the sample;
2) divide the total data sample into two groups according to the sign of data difference from the median: XL denotes all data points less than M and XG is all data points greater than the M;
3) Transform XL data into a new group Y by reflecting them across the median: $Y = M + (M-XL)$;
4) XL and Y represent a new symmetry sample; XG and its reflecting part form another symmetry sample in the same way

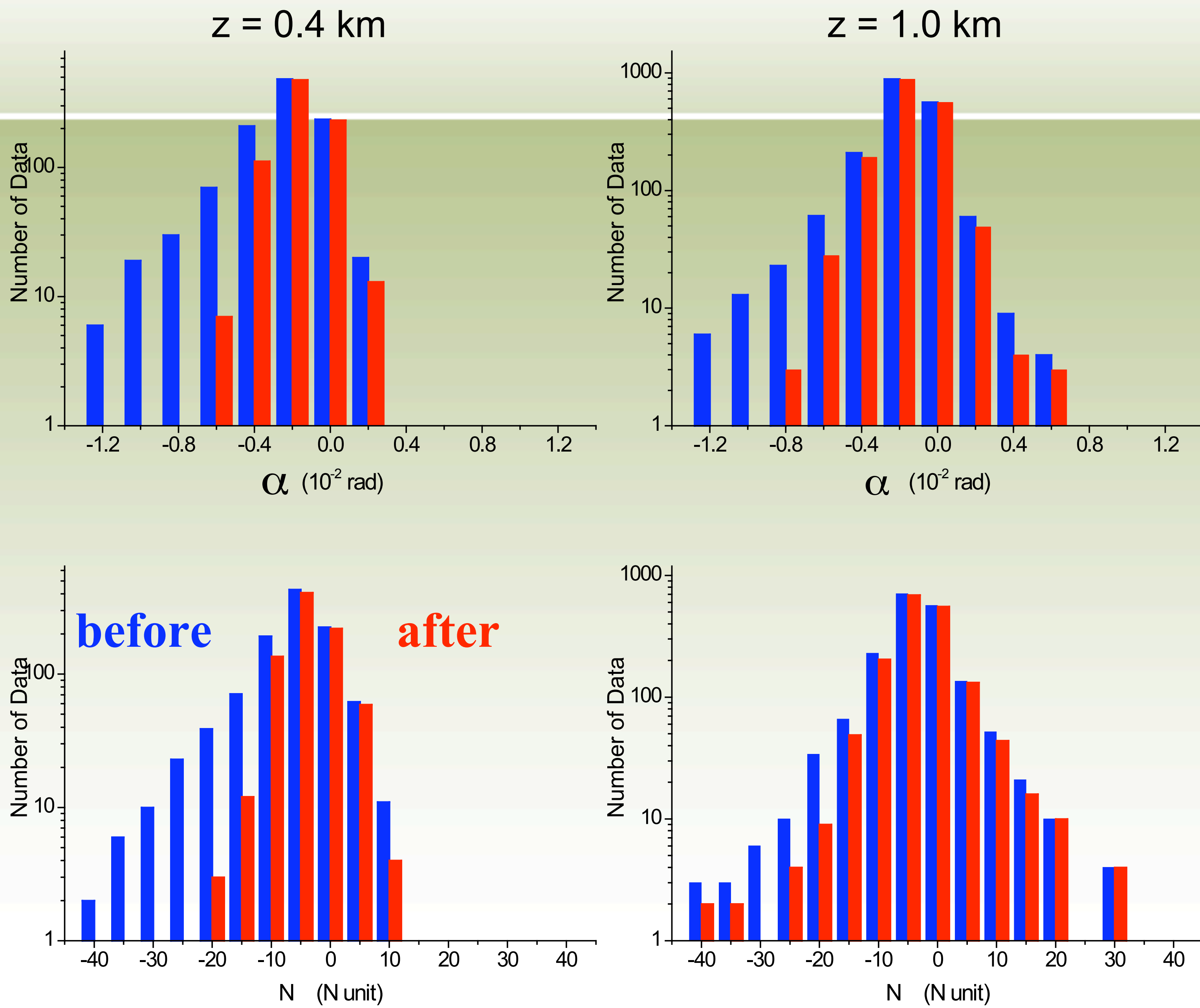


$\overline{\alpha}|_{\alpha>0}$(right) and $\overline{\alpha}|_{\alpha<0}$(left)



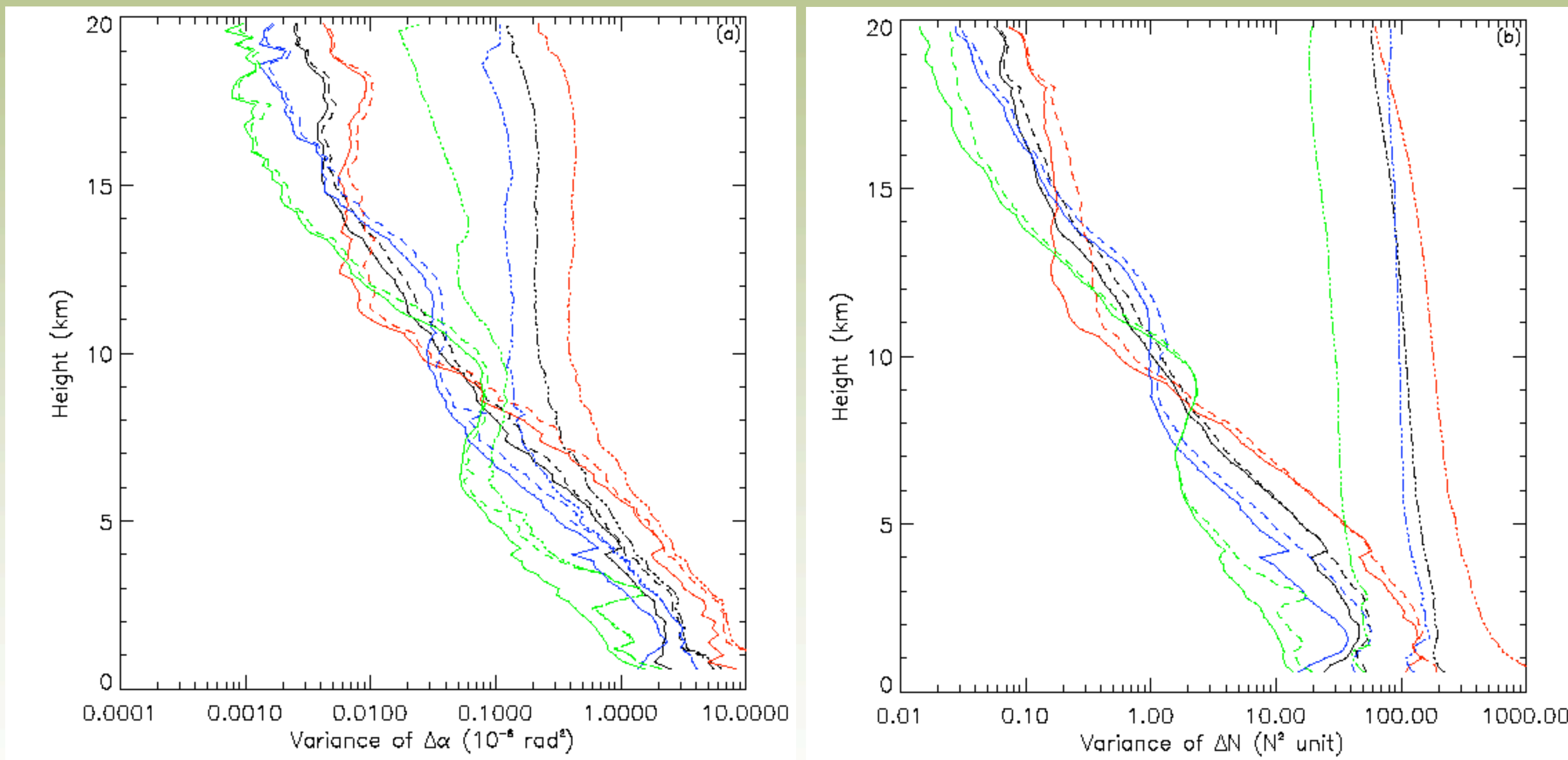
$\overline{N}|_{N>0}$(right) and $\overline{N}\|_{<<0}$(left)

QC4



z = 0.4 km
z = 1.0 km
Number of Data
1
10
100
1000
-1.2
-0.8
-0.4
0.0
0.4
0.8
1.2
α (10^-2 rad)
before
after
-40
-30
-20
-10
0
10
20
30
40
N (N unit)

Variance Values Before and After QC



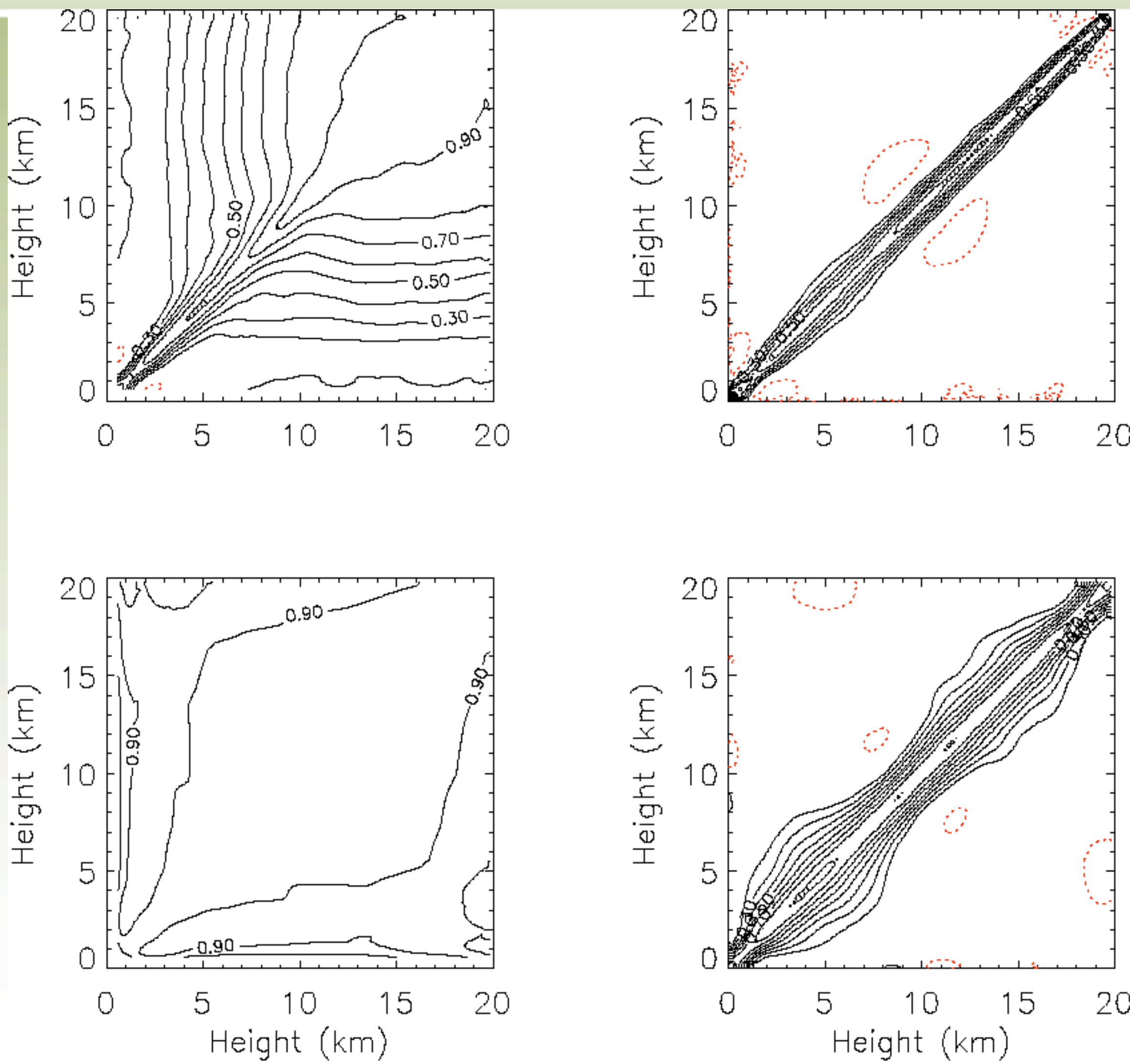
Global — Low Lat — Mid Lat — High Lat

After QC — CDAAC QC — Before QC

Vertical Correlations

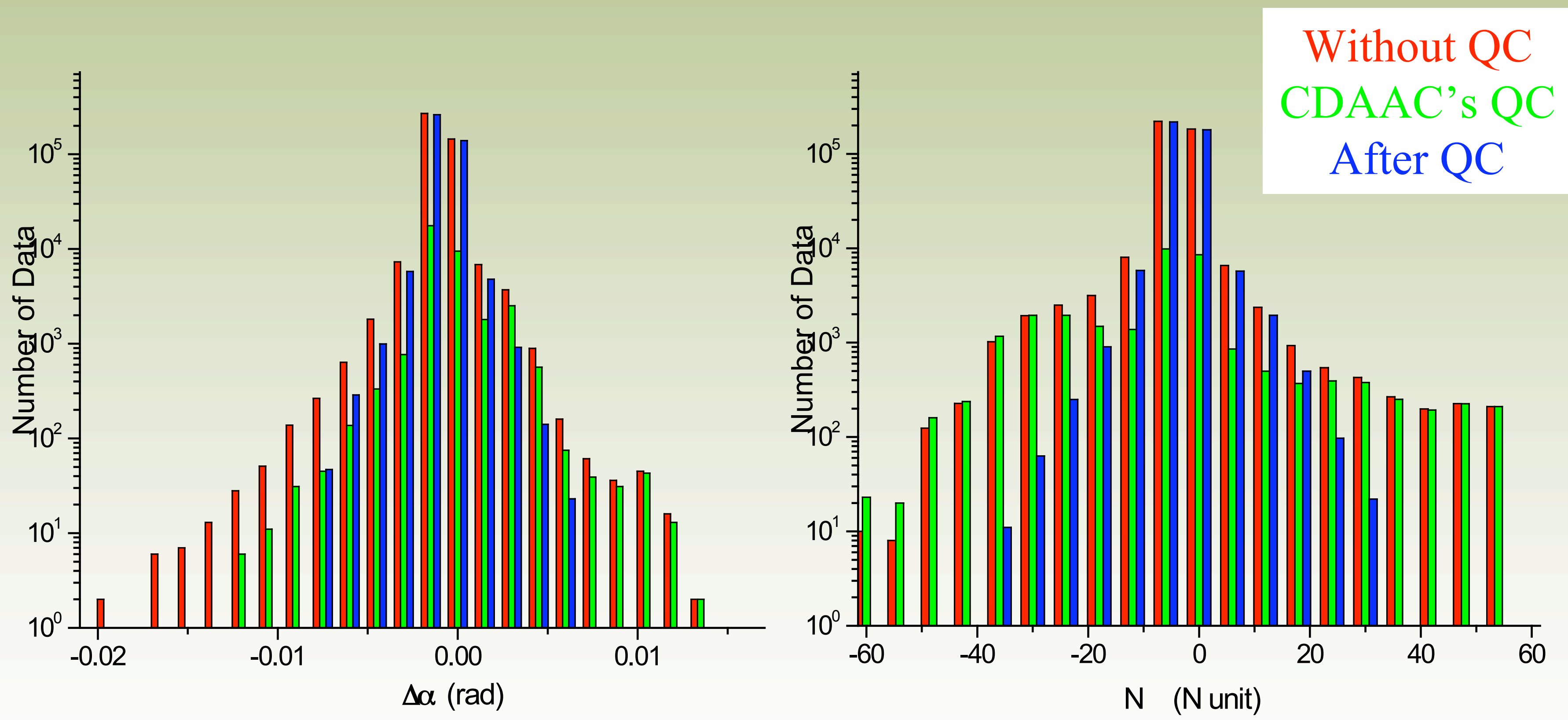
$\Delta\alpha$

ΔN



Before QC **After QC**

Histograms of the Differences between CHAMP Retrievals and NCEP Analysis



Part II: Comparing GPS RO Bending Angle and Local Refractivity Assimilation

An effort to assess the need to incorporate a non-local refractivity assimilation scheme

Three sets of data assimilation experiments were carried out, incorporating CHAMP RO data that occurred during 21-31 May, 2002:

1) NOGPS: All observations without GPS RO data

2) BA: All observations & CHAMP bending angle data

2) REF: All observations & CHAMP refractivity data

Cost function for assimilation of α or N

REF:

$$J = (\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b)$$
$$+ (H_N(\mathbf{x}_0) - N^{obs}(z))^T \mathbf{R}_N^{-1} (H_N(\mathbf{x}_0) - N^{obs}(z))$$
$$+ (\text{other observations})$$

BA:

$$J = (\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b)$$
$$+ (H_{\boldsymbol{\alpha}}(\mathbf{x}_0) - \boldsymbol{\alpha}^{obs}(a))^T \mathbf{R}_{\boldsymbol{\alpha}}^{-1} (H_{\boldsymbol{\alpha}}(\mathbf{x}_0) - \boldsymbol{\alpha}^{obs}(a))$$
$$+ (\text{other observations})$$

Gradients for Assimilation of α and N_{GPS}

$$\nabla J = \mathbf{B}^{-1}\left(\mathbf{x}_0 - \mathbf{x}^b\right) + \mathbf{H}_N^T \mathbf{R}_N^{-1}\left(H_N(\mathbf{x}_0) - \mathbf{N}^{obs}\right) + \cdots$$

$$\nabla J = \mathbf{B}^{-1}\left(\mathbf{x}_0 - \mathbf{x}^b\right) + \mathbf{H}_\alpha^T \mathbf{R}_\alpha^{-1}\left(H_\alpha(\mathbf{x}_0) - \boldsymbol{\alpha}^{obs}\right) + \cdots$$

Required input for GPS RO assimilation:

$\mathbf{N}^{obs}$, $\mathbf{R}_N^{-1}$, $H_N(\mathbf{x}_0)$, $\mathbf{H}_N^T$ for refractivity assimilation

$\boldsymbol{\alpha}^{obs}$, $\mathbf{R}_\alpha^{-1}$, $H_\alpha(\mathbf{x}_0)$, $\mathbf{H}_\alpha^T$ for bending angle assimilation

H_α --- A Ray-Tracing Model

The GPS bending angle is derived through an integration of the following ray equation

$$\frac{d^2\vec{x}}{d\tau^2} = n\nabla n$$

where $d\tau = \frac{ds}{n}$, *s* is the length of the ray, *n* is the refractive index, and $\vec{x}$ is the Cartesian coordinate vector.

H_N --- A Forward Model for Simulating N_{loc}

$$N = k_1 \frac{p_d}{T} + k_2 \frac{p_w}{T} + k_3 \frac{p_w}{T^2}$$

$$= 77.6 \frac{p}{T} + 3.73 \times 10^5 \frac{pq}{T^2 (0.622 + 0.378q)}$$

$$= N(T, q, p)$$

(Bean and Dutton, 1968)

$$N(T,q,p)\Big|_{\text{ó level}} \xrightarrow{convert} N(T,q,p)\Big|_{\Phi(\text{geopotential height})} \xrightarrow{convert} N(T,q,p)\Big|_{z(\text{geometric height})}$$

$$N(T,q,p)\big|_{\sigma \text{ level}} \xrightarrow{\textit{convert}} N(T,q,p)\big|_{\Phi(\text{geopotential height})}$$

Calculation of Φ is based on the hydrostatic equation (Sela, 1980):

$$\Phi_{i-1,j,k} - \Phi_{i,j,k} = \frac{c_p}{2}\left[T_{v,i-1,j,k}\left(\frac{p_{i,j,k}^{R_d/c_p}}{p_{i-1,j,k}^{R_d/c_p}} - 1\right) + T_{v,i,j,k}\left(1 - \frac{p_{i-1,j,k}^{R_d/c_p}}{p_{i,j,k}^{R_d/c_p}}\right)\right]$$

where

$p_{i,j,k} = \sigma_i p_{s,j,k}$ **--- pressure at the *i*th vertical level**

$p_{s,j,k}$ **--- surface pressure for the *j*th latitude and *k*th longitude**

$T_v = T\left[1 + \left(\frac{R_v}{R_d} - 1\right)q\right]$ **--- virtual temperature**

$$N(T,q,p)\big|_{\Phi(\text{geopotential height})} \xrightarrow{\text{convert}} N(T,q,p)\big|_{z(\text{geometric height})}$$

Calculation of the geometric height z is based on the definition of the geopotential (List, 1989):

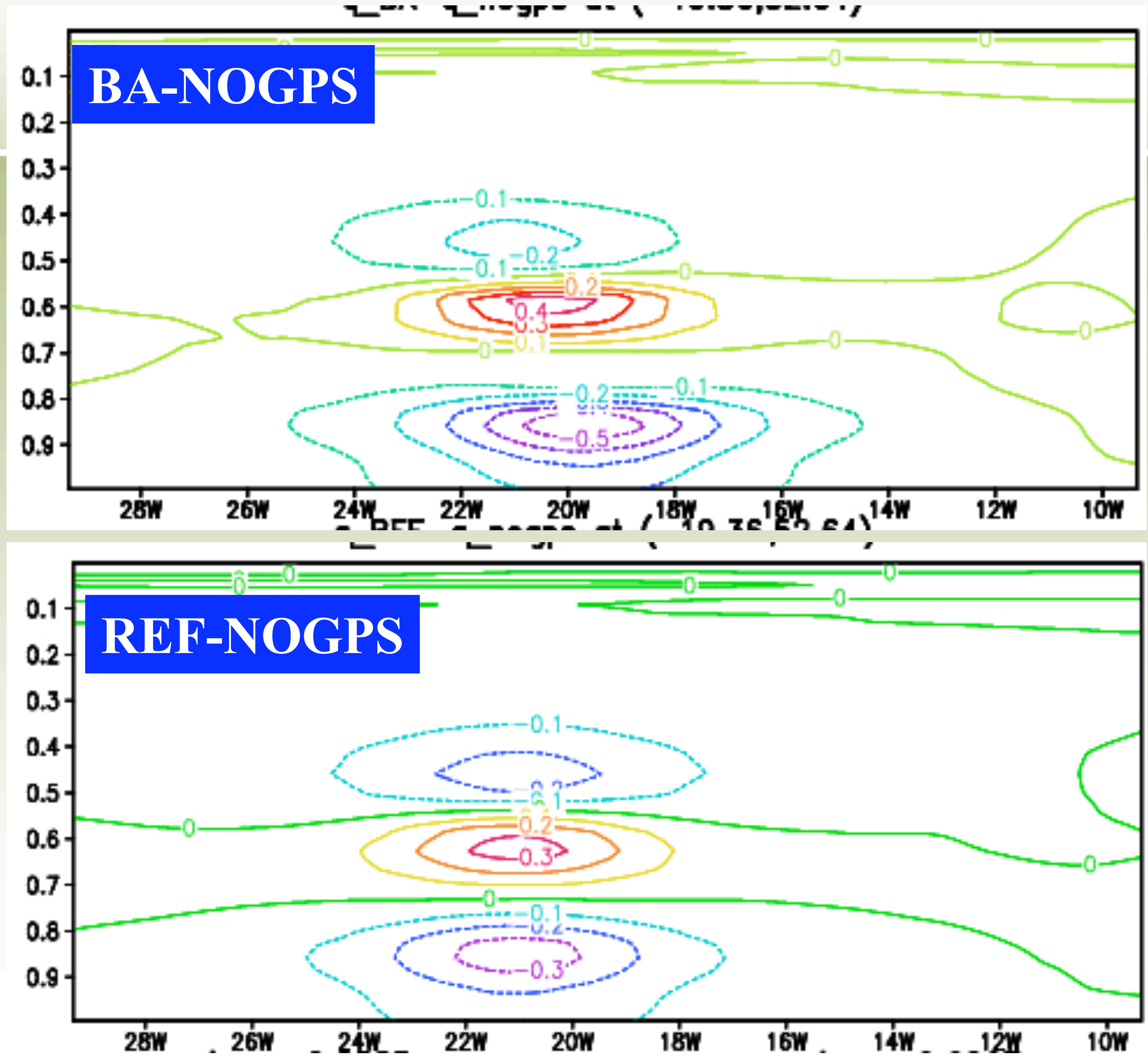
$$\Phi_{i,j,k} = \int_0^{z_{i,j,k}} g(\varsigma, \varphi_j, \lambda_k)\, d\varsigma$$

where *g* is gravity acceleration at a given altitude ζ and latitude ϕ:

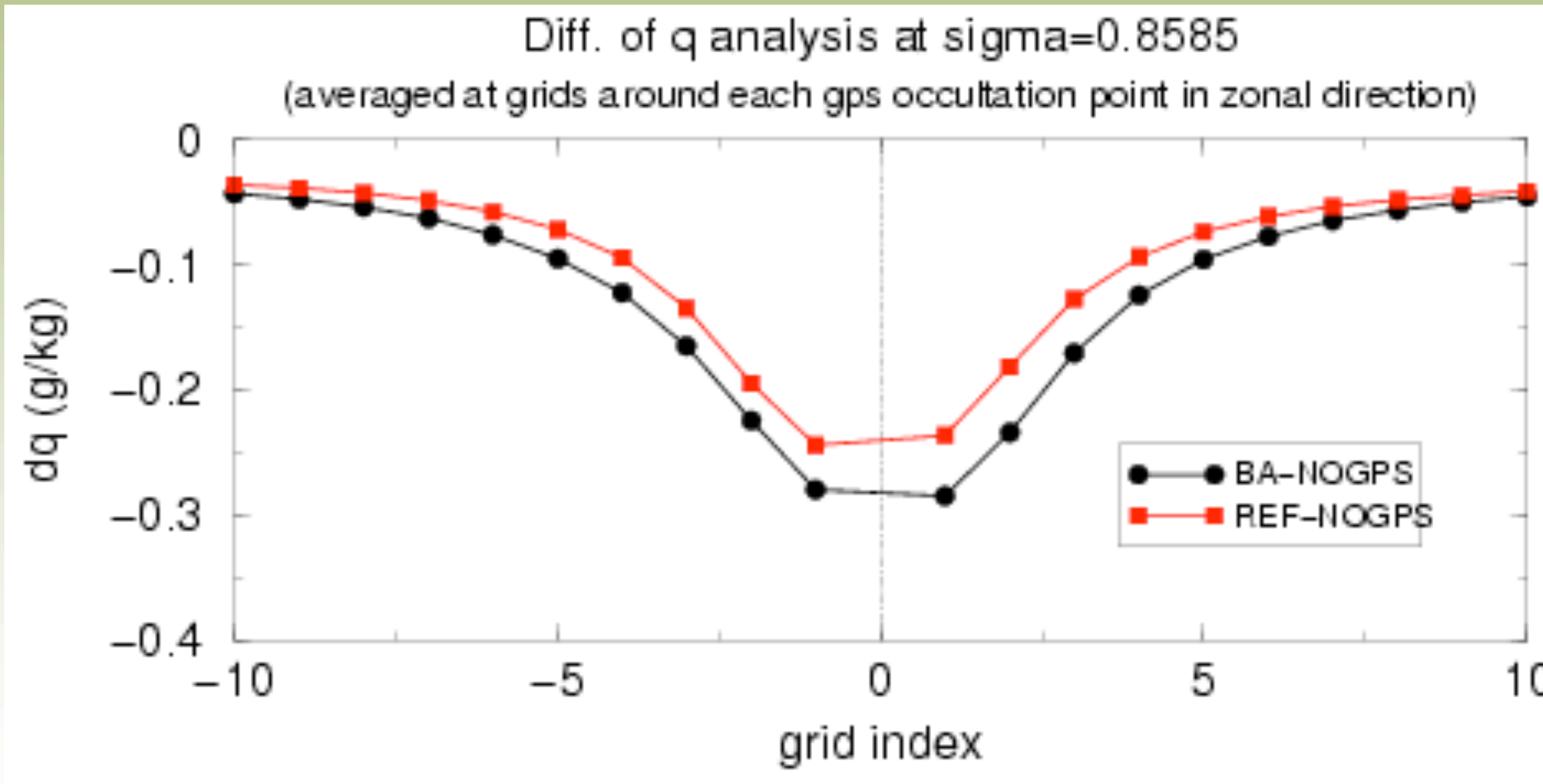
$$g(\varsigma, \varphi, \lambda) = \frac{g_{re} R^2}{(R + \varsigma)^2}$$

$$\Rightarrow \quad \Phi = \frac{g_{re} R z}{R + z} \quad \Rightarrow \quad z = \frac{R\Phi}{g_{re} R - \Phi}$$

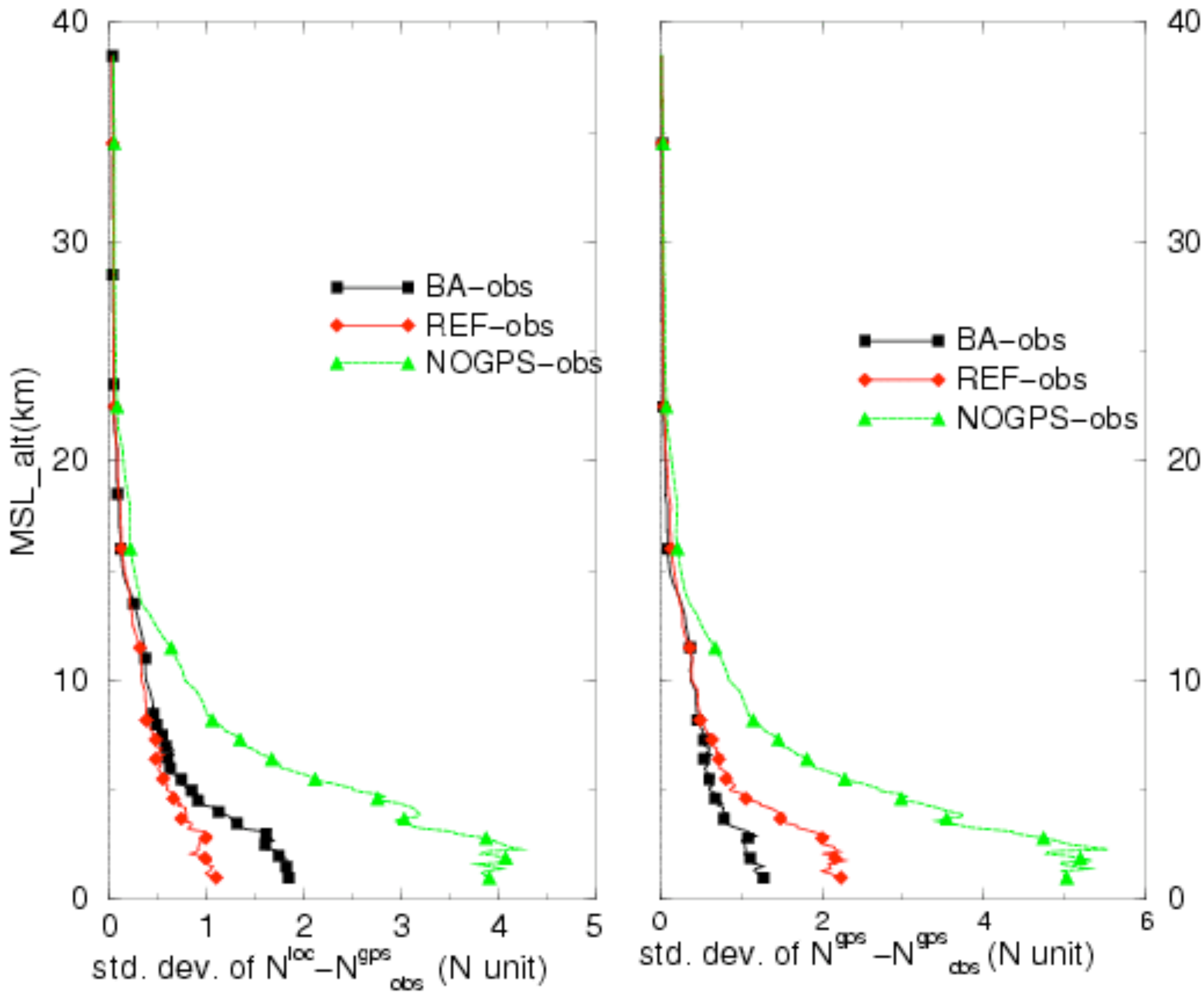
Δq


BA-NOGPS
REF-NOGPS
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
28W
26W
24W
22W
20W
18W
16W
14W
12W
10W

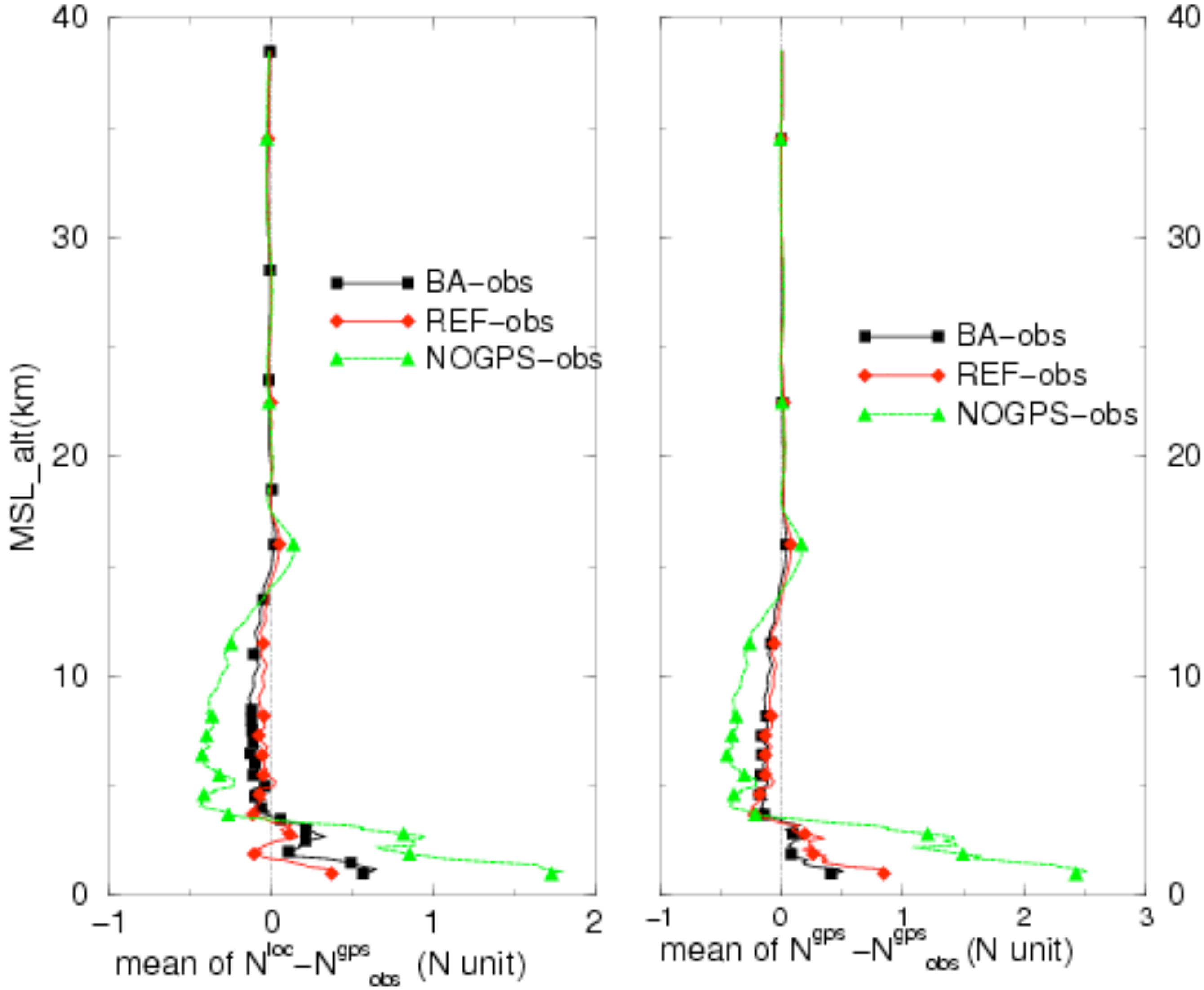
$\overline{\Delta q}\Big|_{\sigma=0.858}$ averaged over all occultations



Standard deviation of $N^{loc}_{BA,REF,NOGPS} - N^{obs}_{gps}$



Mean of $N^{loc}_{BA,REF,NOGPS} - N^{obs}_{gps}$



Part III: A Non-Local Refractivity Simulation&Assimilation

Along a straight line (m) tangent to the observed ray, the following relationship holds true:

$$\int_m N^{GPS} ds = \int_m N^{loc} ds \equiv L$$

$$\Downarrow$$

$$\mathbf{B}\vec{N}^{GPS} = \mathbf{A}\vec{N}^{loc}.$$

$$\Downarrow$$

$$\vec{N}^{GPS} = \mathbf{K}\vec{N}^{loc}, \qquad \mathbf{K} = \mathbf{B}^{-1}\mathbf{A}.$$

Kernel function

A Non-local Refractivity Simulation&Assimilation

Along a straight line (m) tangent to the observed ray, the following relationship holds true:

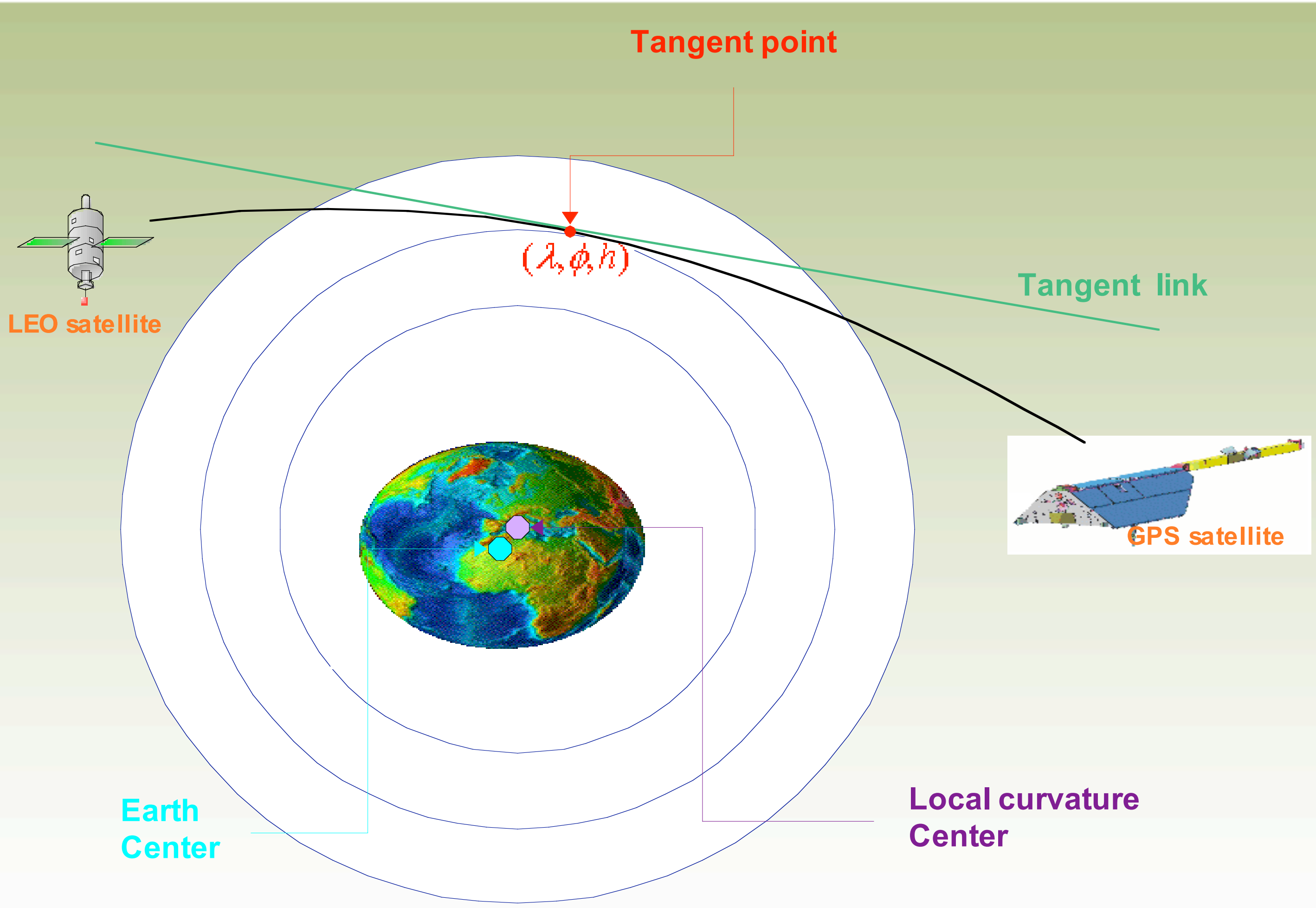
$$\int_m N^{GPS} ds = \int_m N^{loc} ds \equiv L \quad \Longrightarrow \quad \mathbf{B}\vec{N}^{GPS} = \mathbf{A}\vec{N}^{loc}.$$

Excess phase:

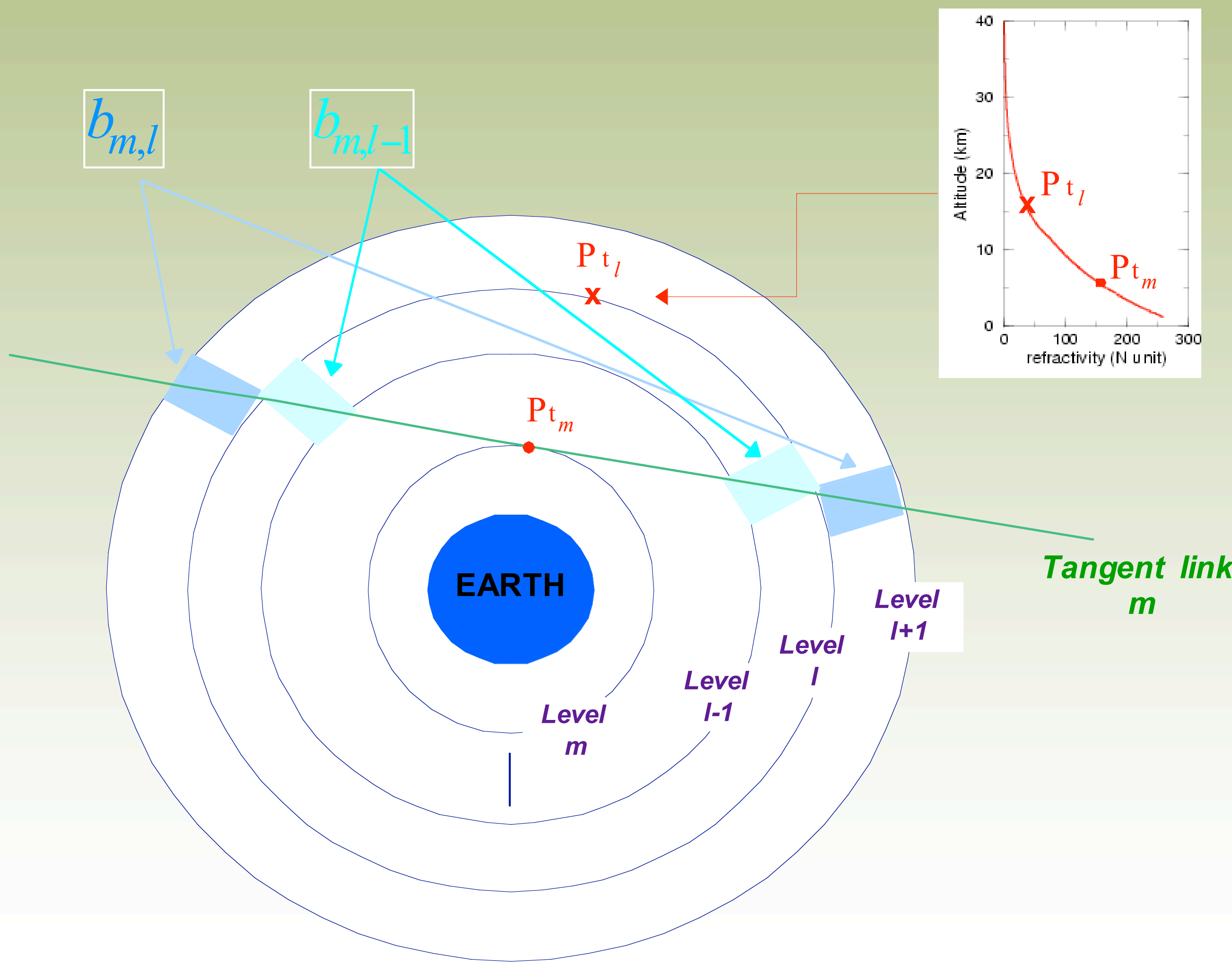
$$L_{non} = \mathbf{A}\vec{N}^{loc}$$

$$L_{obs} = \mathbf{A}\vec{N}_{obs}$$

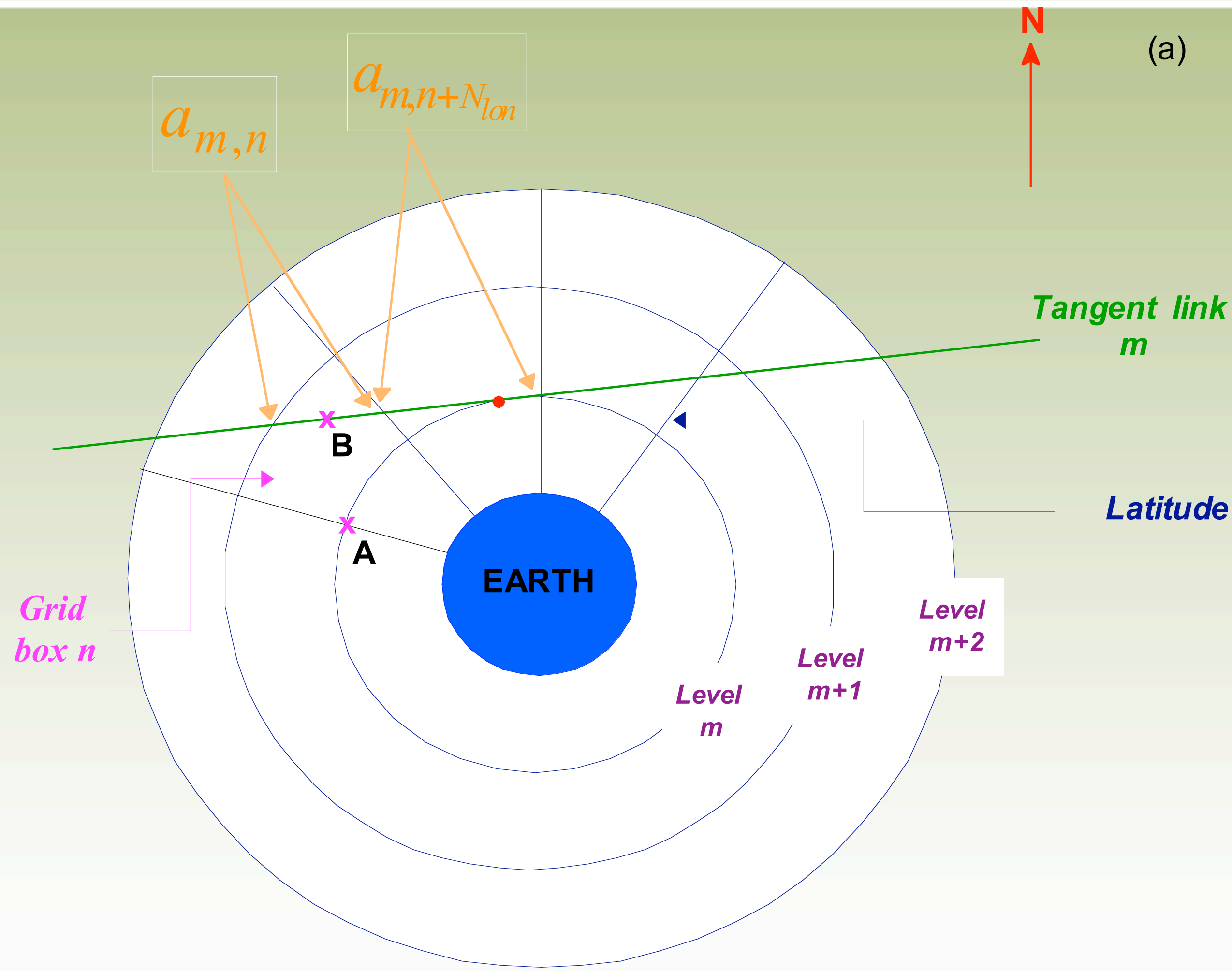
$$L_{loc} = \mathbf{A}\vec{N}^{loc}_{\text{perigee point}}$$

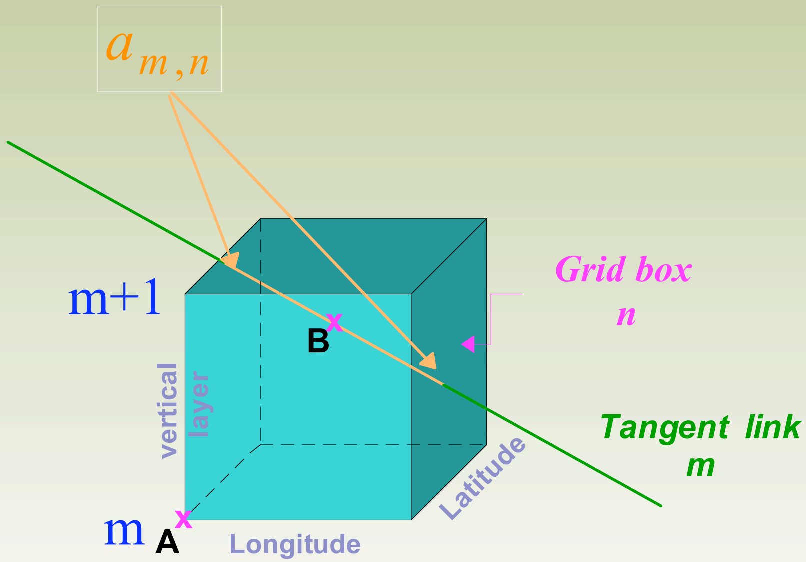

Tangent point
(λ,ϕ,h)
Tangent link
LEO satellite
GPS satellite
Earth
Center
Local curvature
Center

$$\mathbf{B}\vec{N}^{GPS} = \mathbf{A}\vec{N}^{loc}.$$



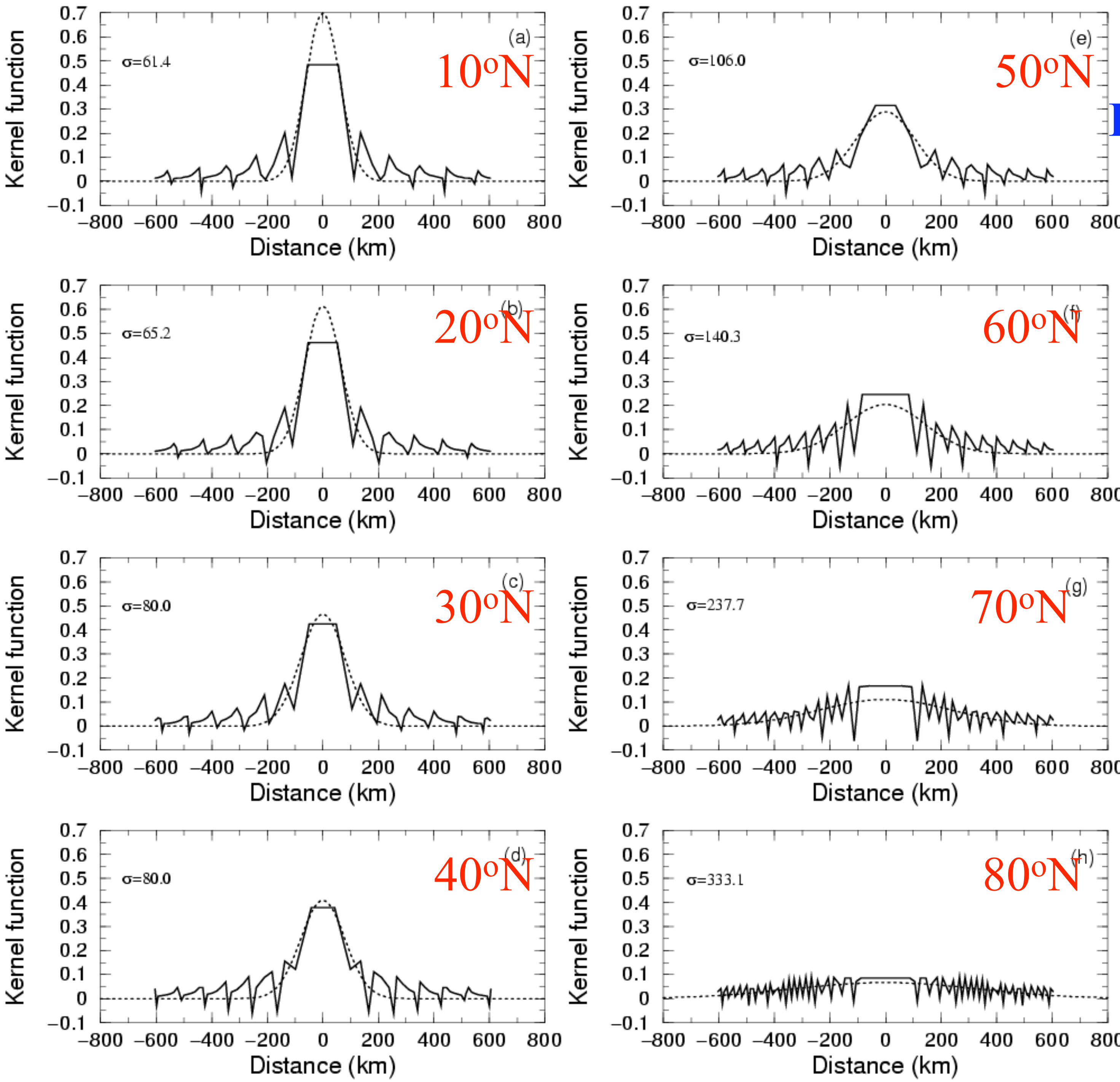
$$\mathbf{B}\vec{N}^{GPS} = \mathbf{A}\vec{N}^{loc}.$$





$a_{m,n}$
m+1
B
Grid box
n
vertical layer
Tangent link
m
Latitude
m
A
Longitude

Kernal function



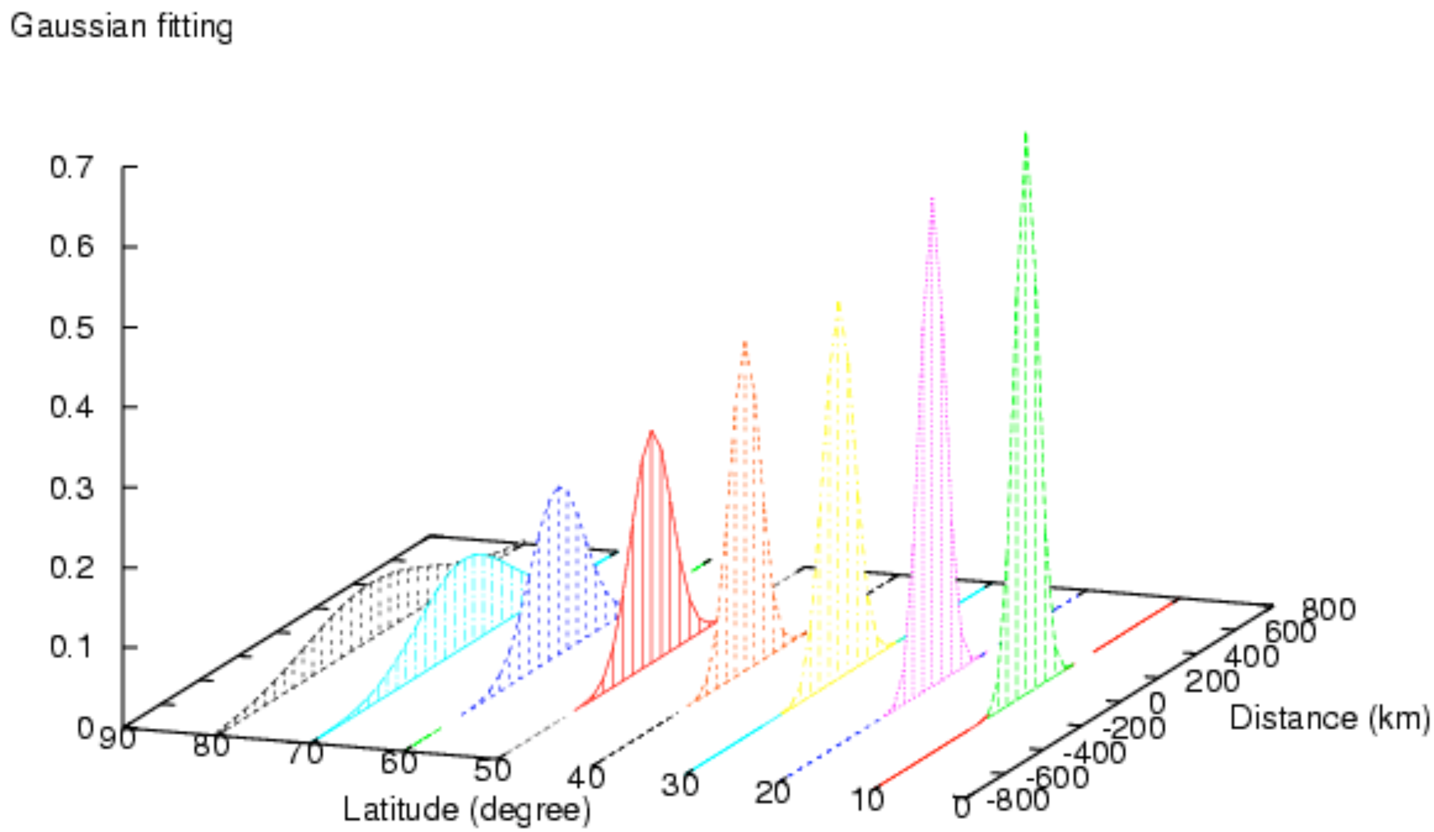
Horizontal: 1^ox1^o,

Vertical: 1 km

Tangent point: 0^oE

Tangent dir.: Equ.

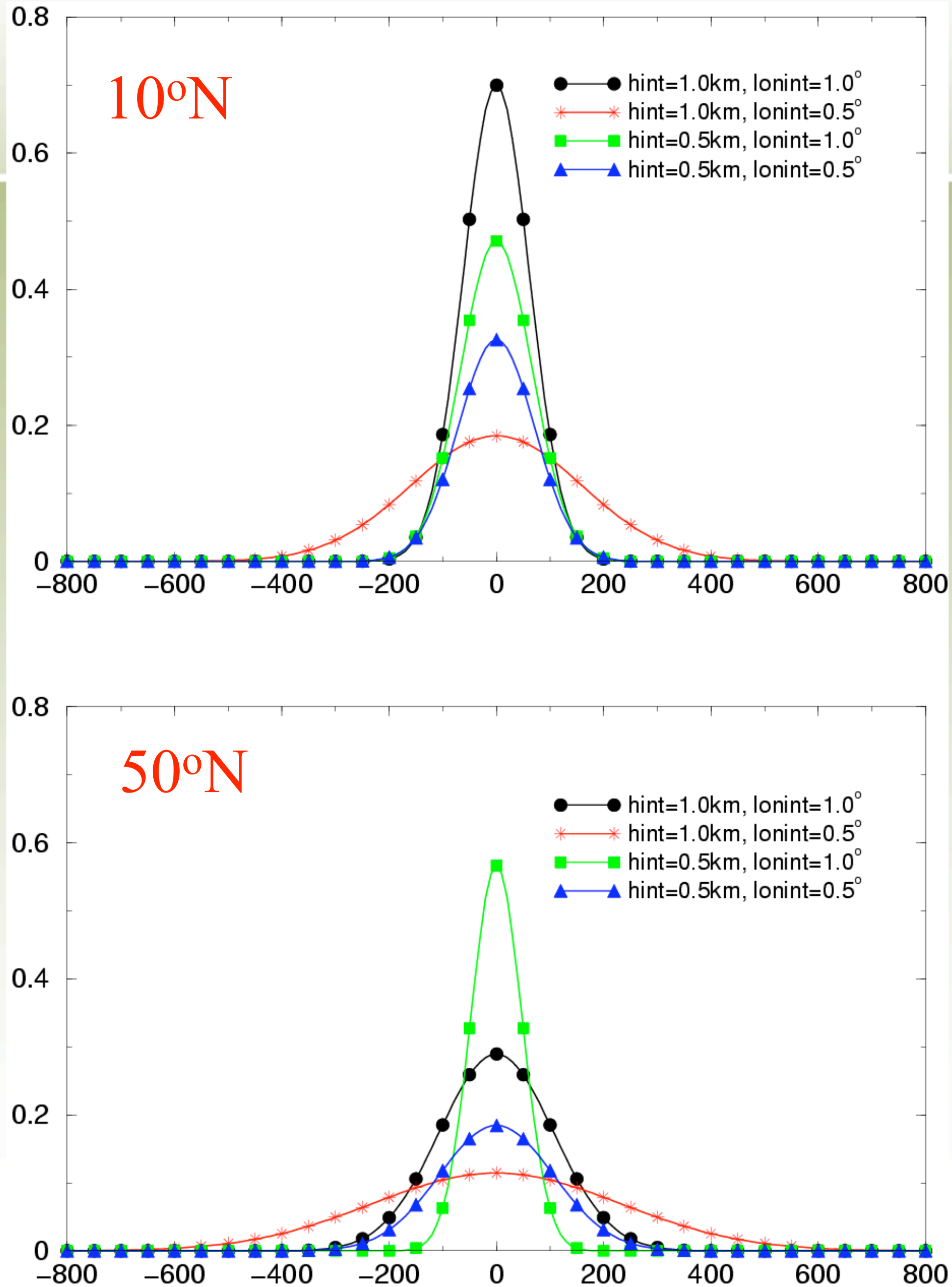
Latitudinal dependence of the kernal function



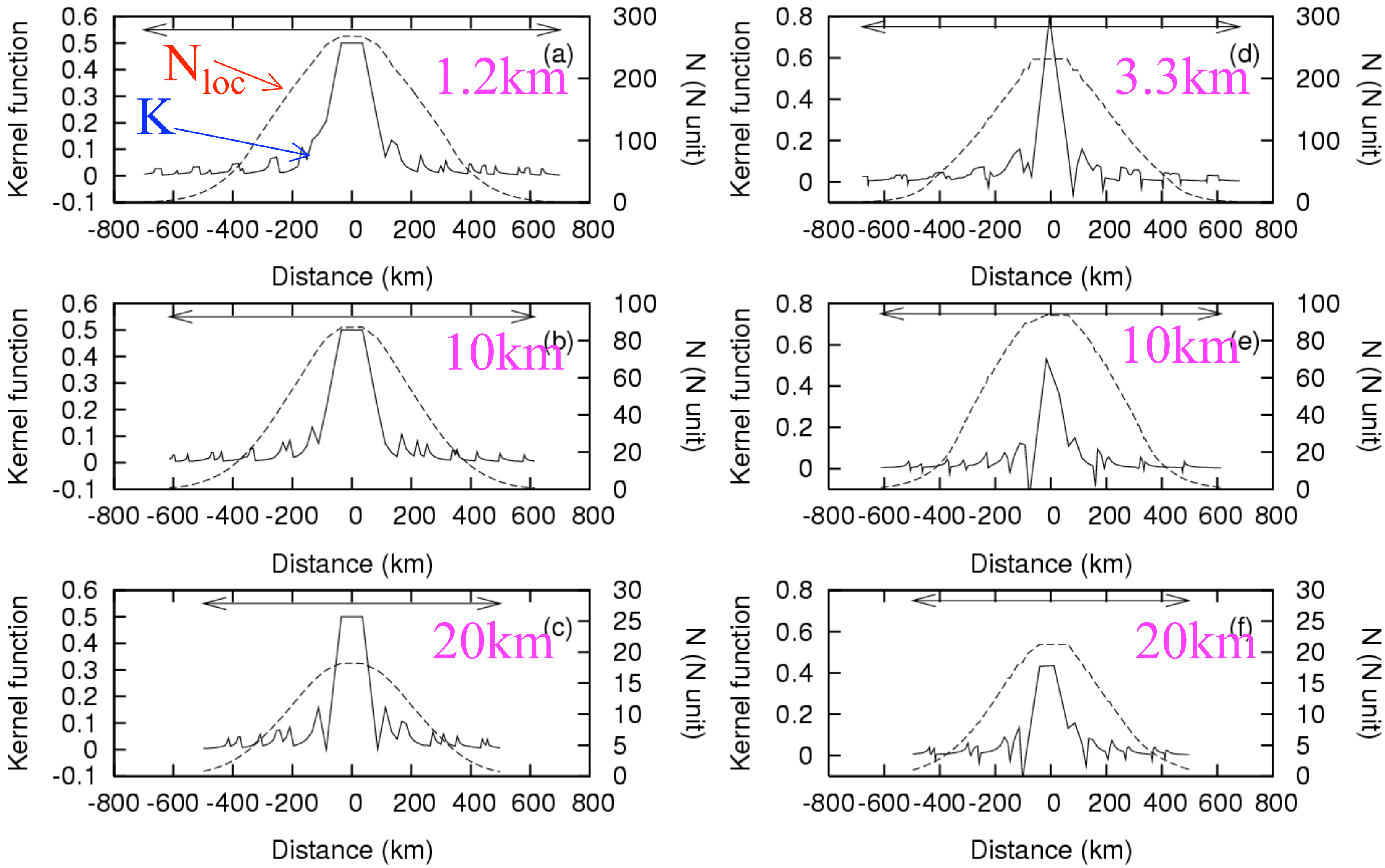
Kernal function

Resolution dependence

Tangent point: 1 km



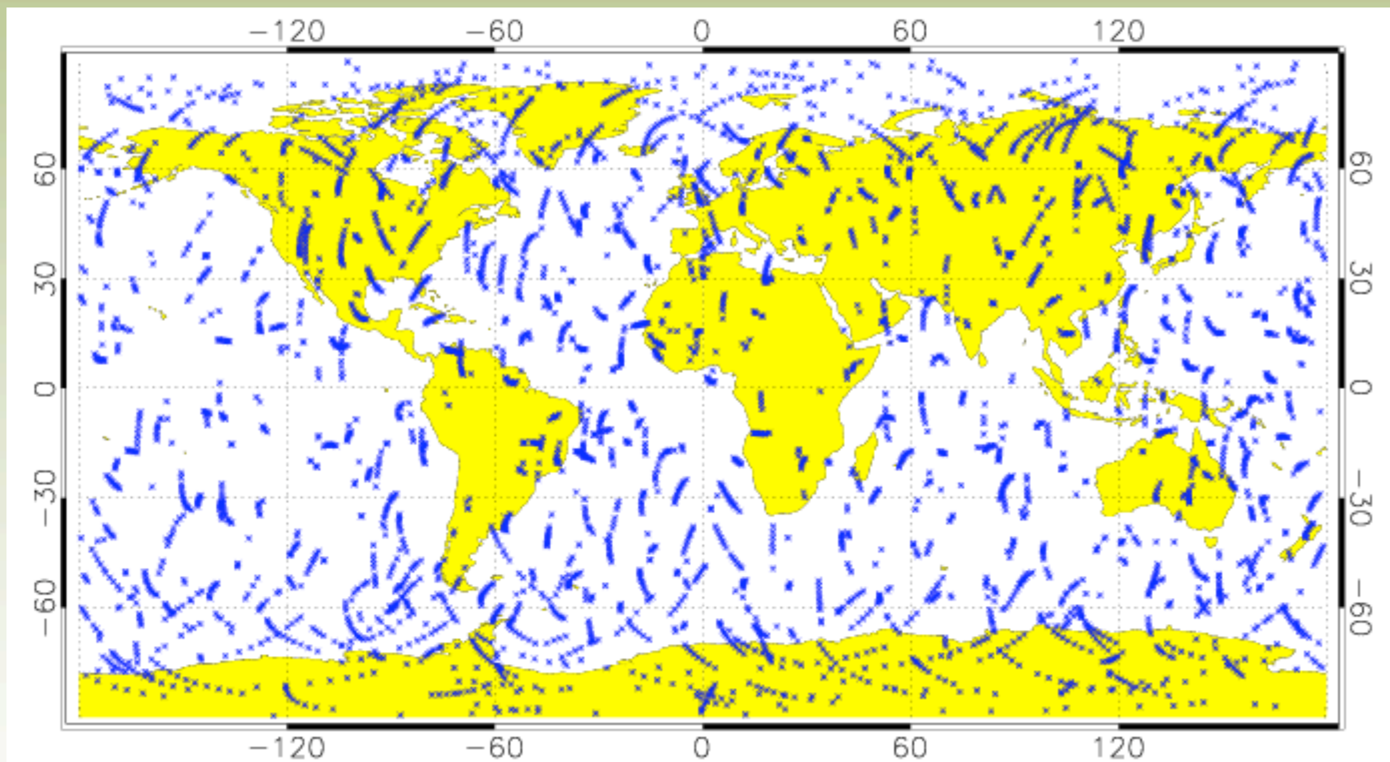
Kernel function using NCEP grid (T170L42)

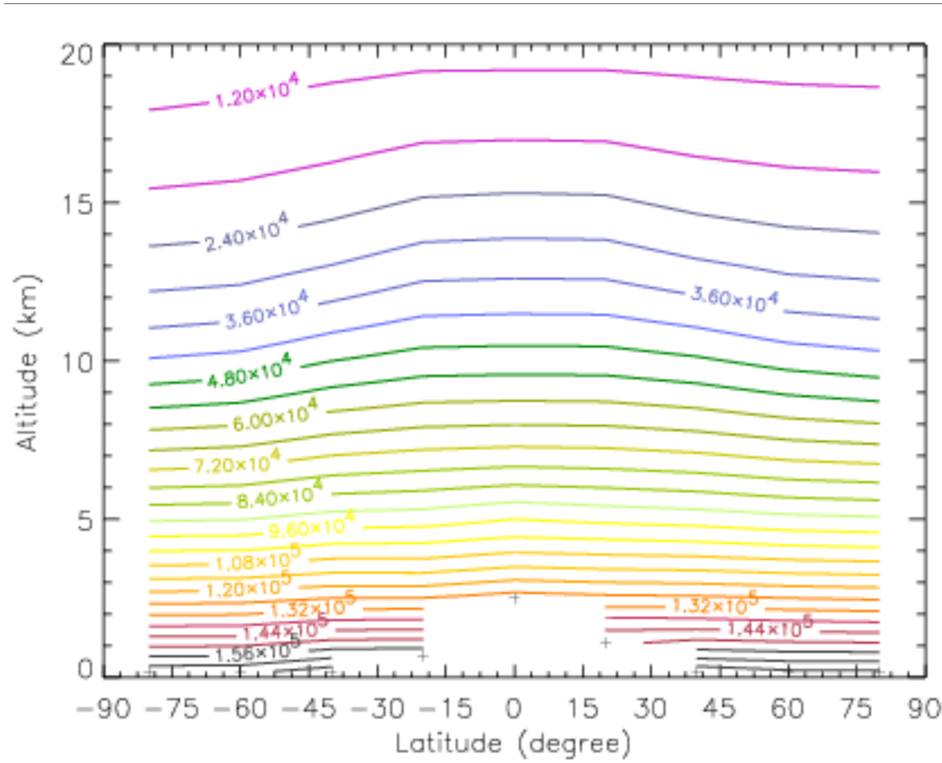


RO (62.39ºW, 54.29ºS)

RO (158.59ºE, 3.92ºS)

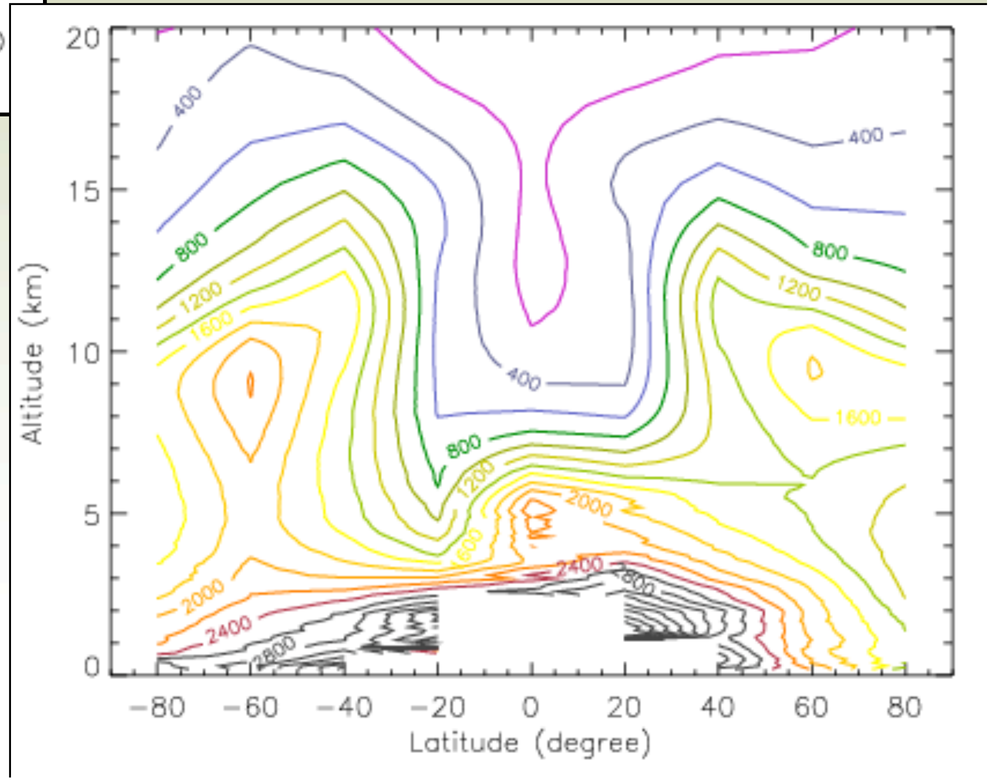
Global Distributions of 4240 CHAMP ROs in May 2002

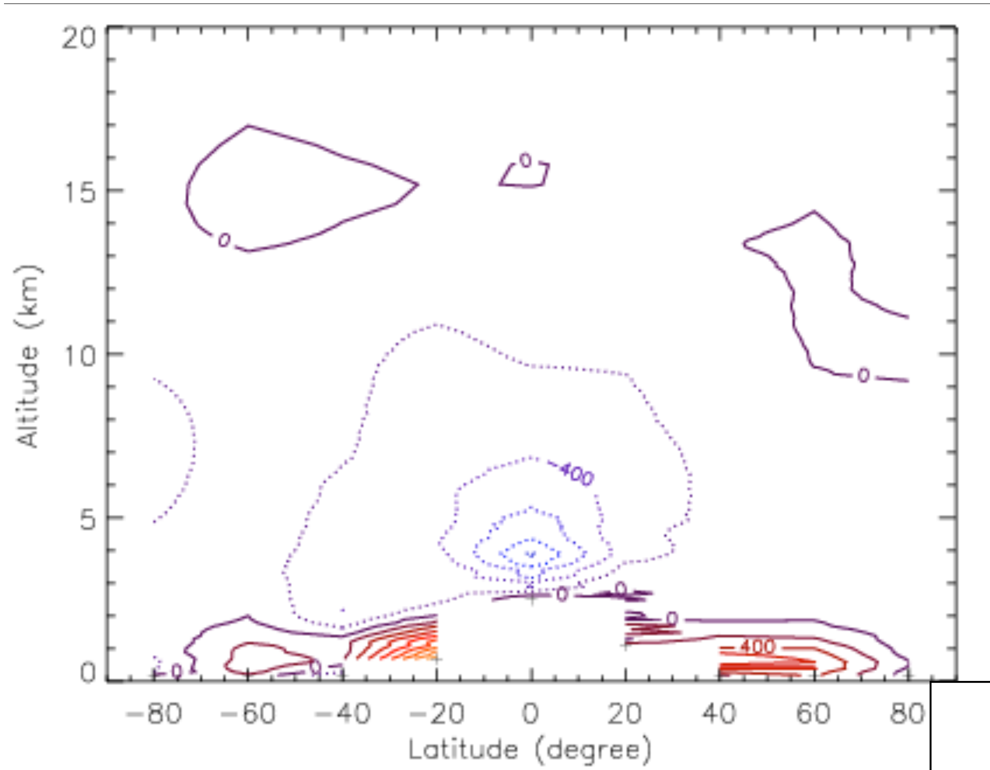




Mean of L^{obs} (km N)

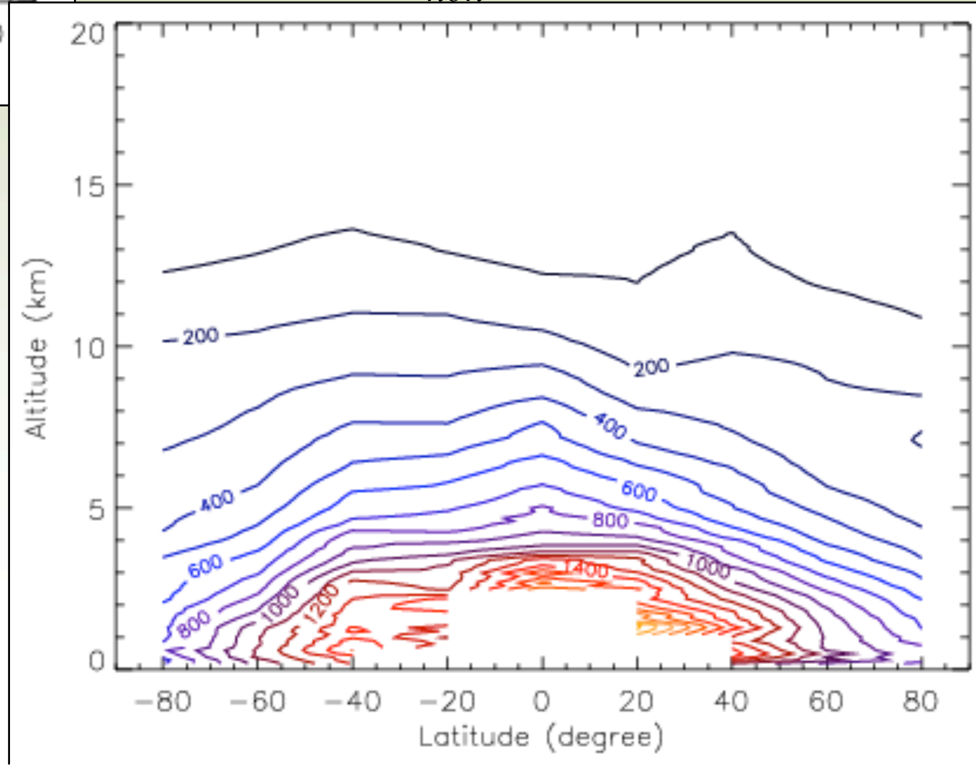
rms of L^{obs} (km N)

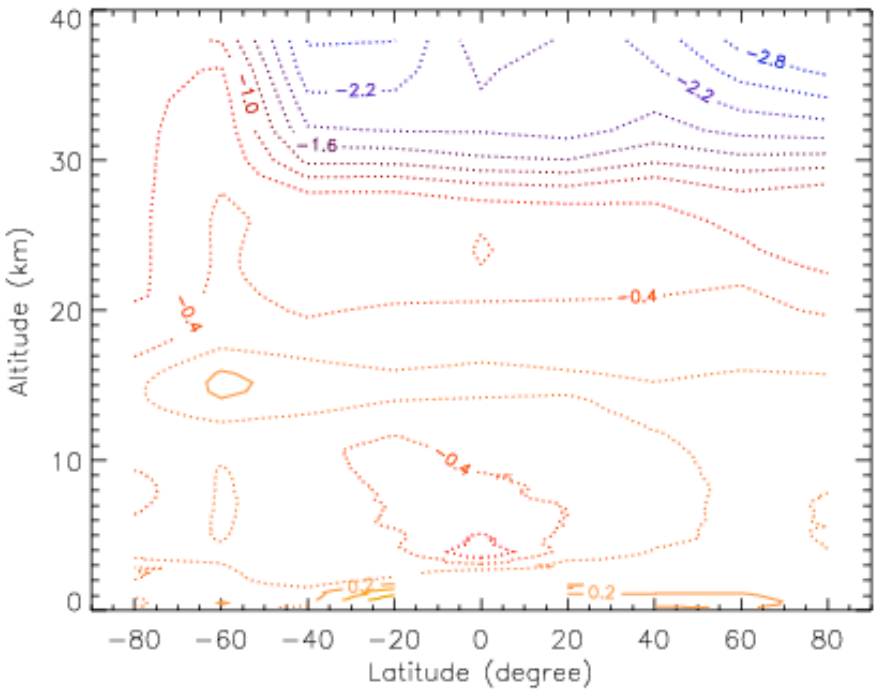




Mean of $(L_{non}^{guess} - L^{obs})$ (km N)

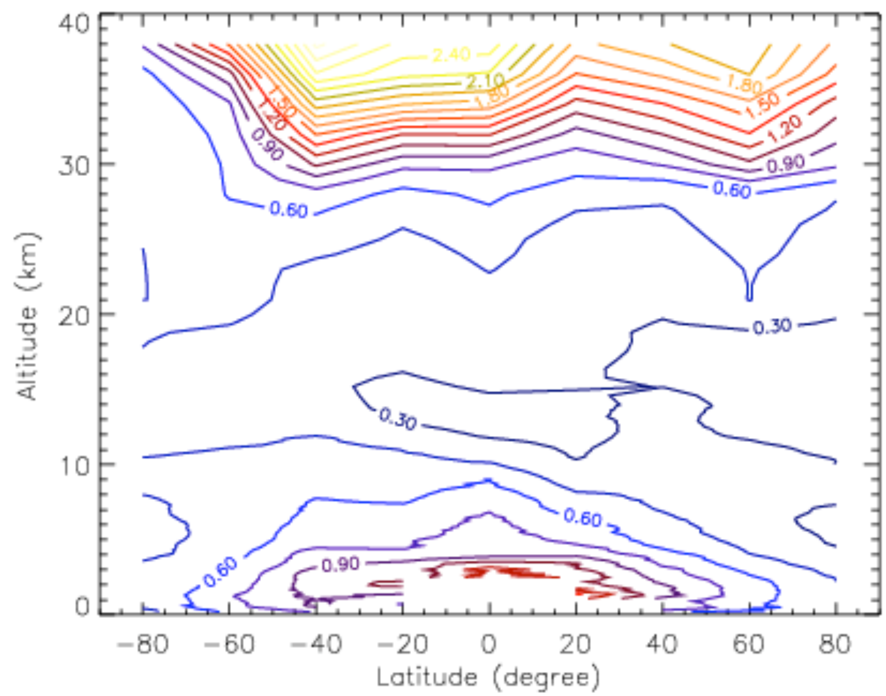
rms of $(L_{non}^{guess} - L^{obs})$ (km N)

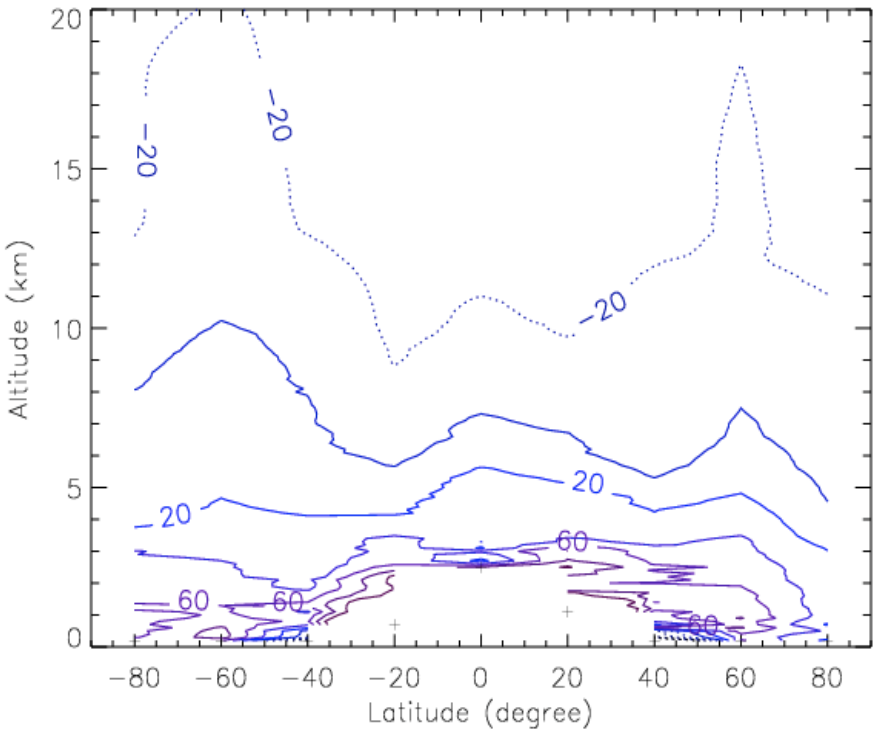




Mean of $(L_{non}^{guess} - L^{obs})/L^{obs}$ (%)

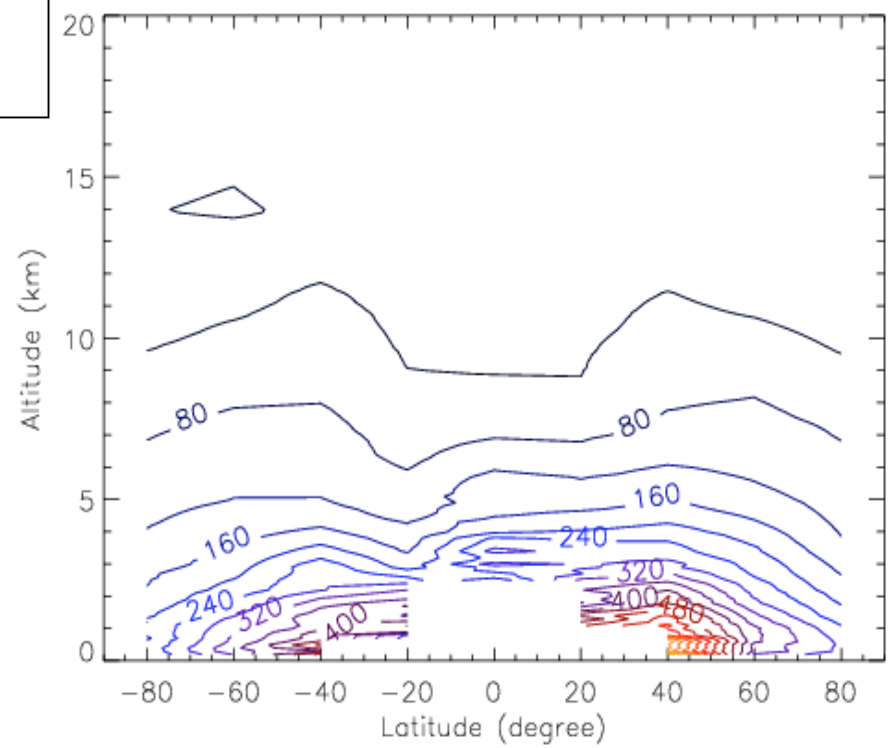
rms of $(L_{non}^{guess} - L^{obs})/L^{obs}$ (%)



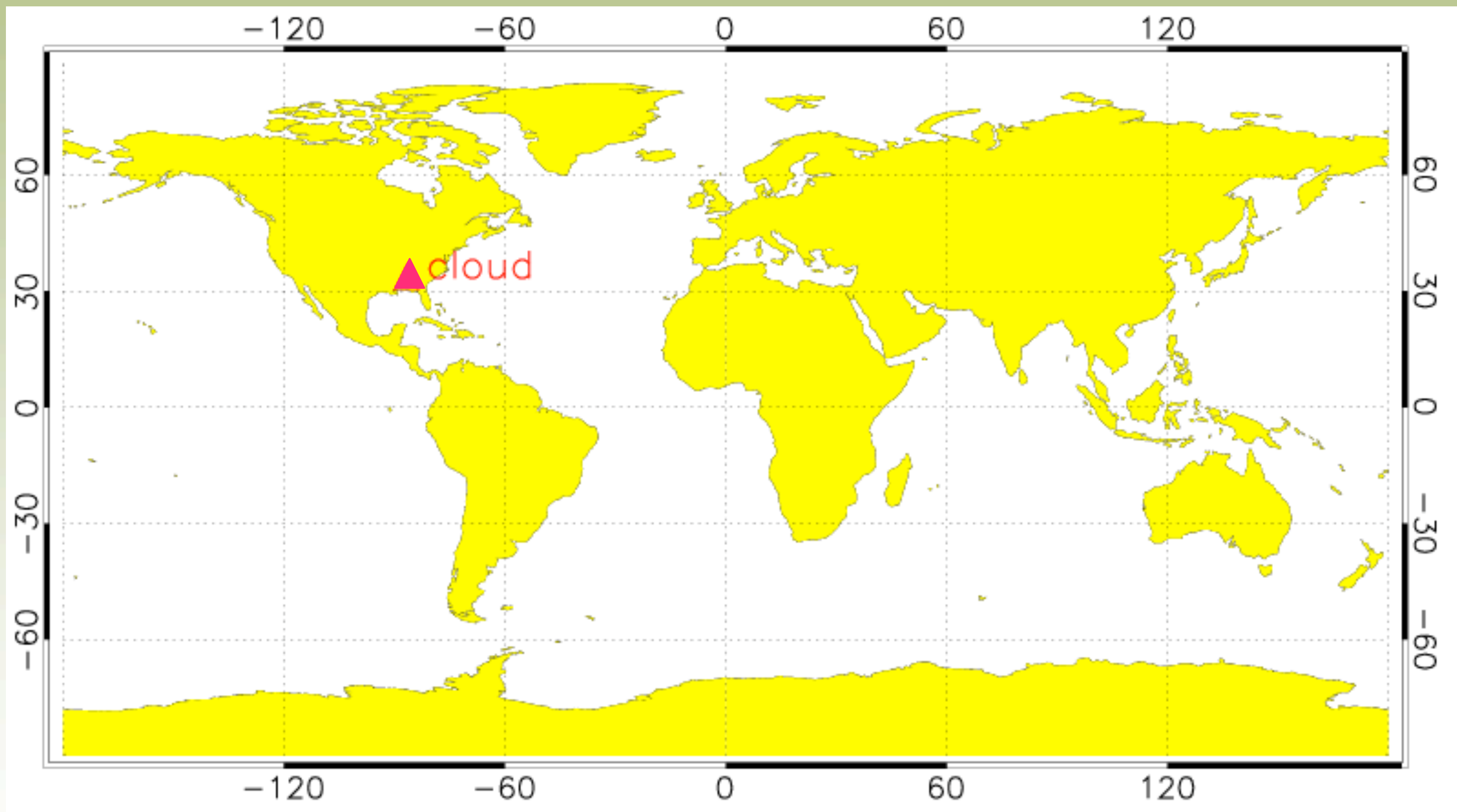


Mean of $(L_{non}^{guess} - L_{loc}^{guess})$ (km N)

rms of $(L_{non}^{guess} - L_{loc}^{guess})$ (km N)



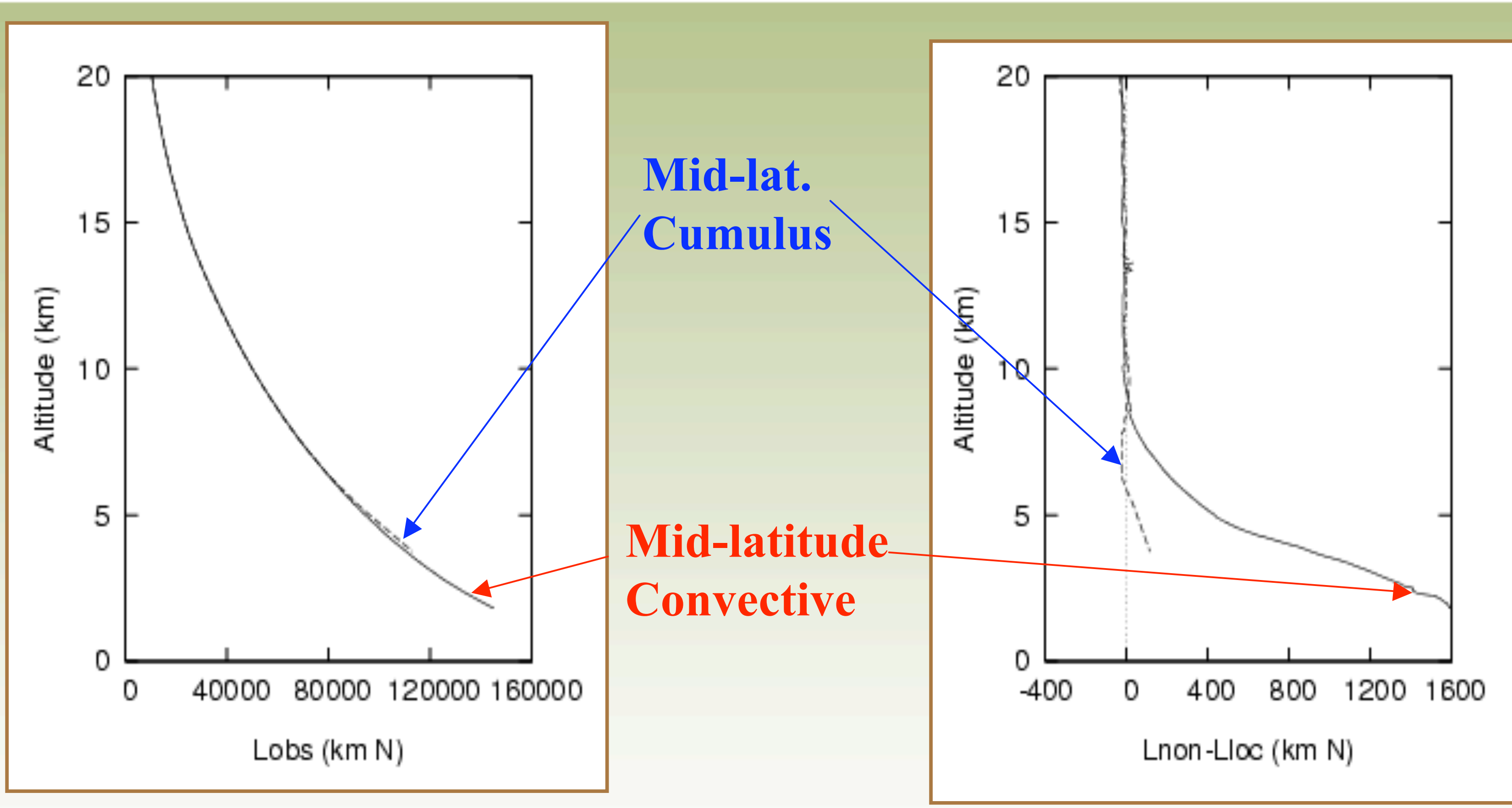
Location of Two Collocated CHAMP ROs within Cloud Systems



Mid-latitude Cumulus: (273ºE, 33ºN), 0633 UTC 31 May 2002

Mid-lat. Convective: (274ºE, 34ºN), 0645 UTC 29 May 2002

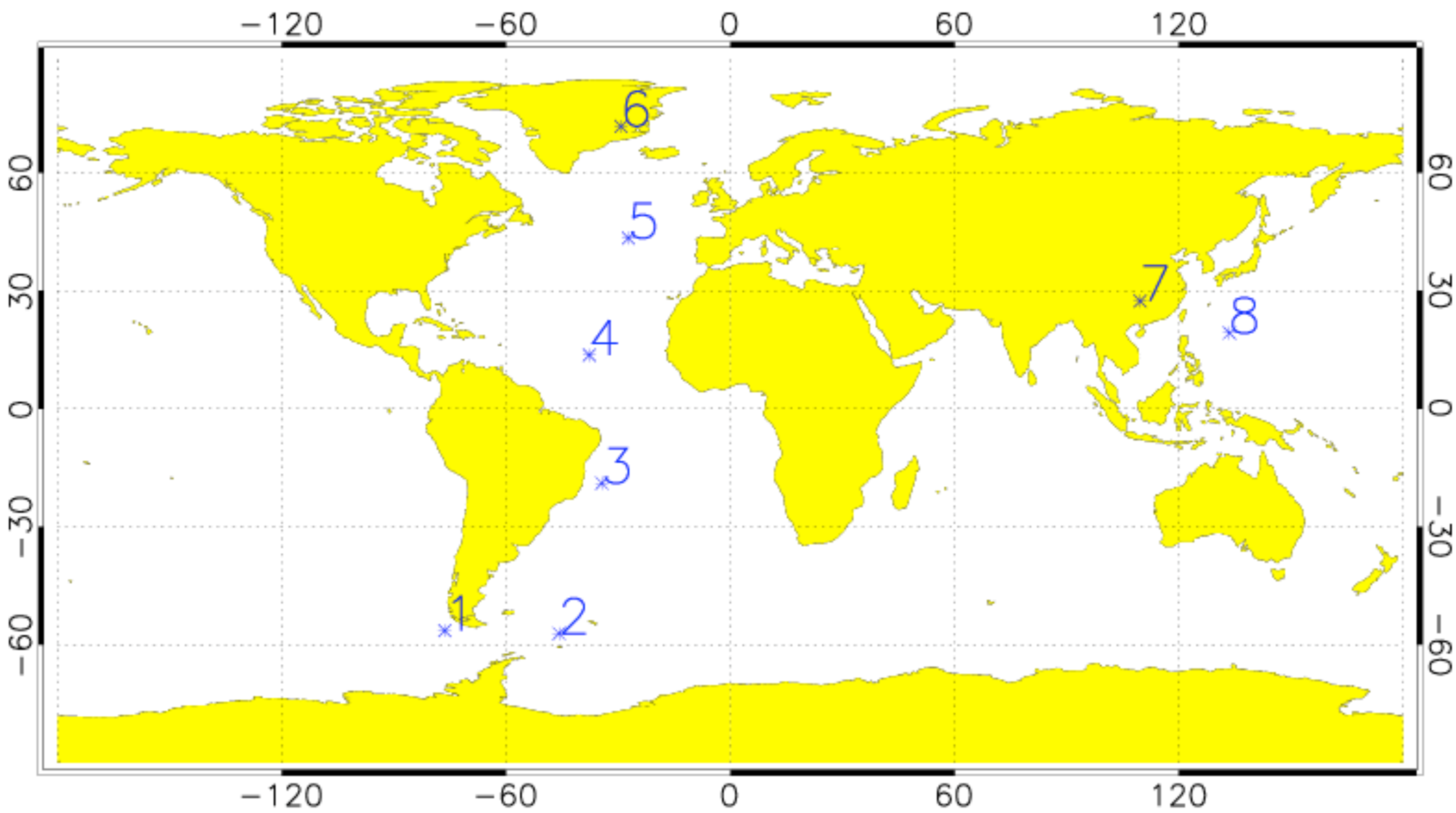
Two Individual Soundings



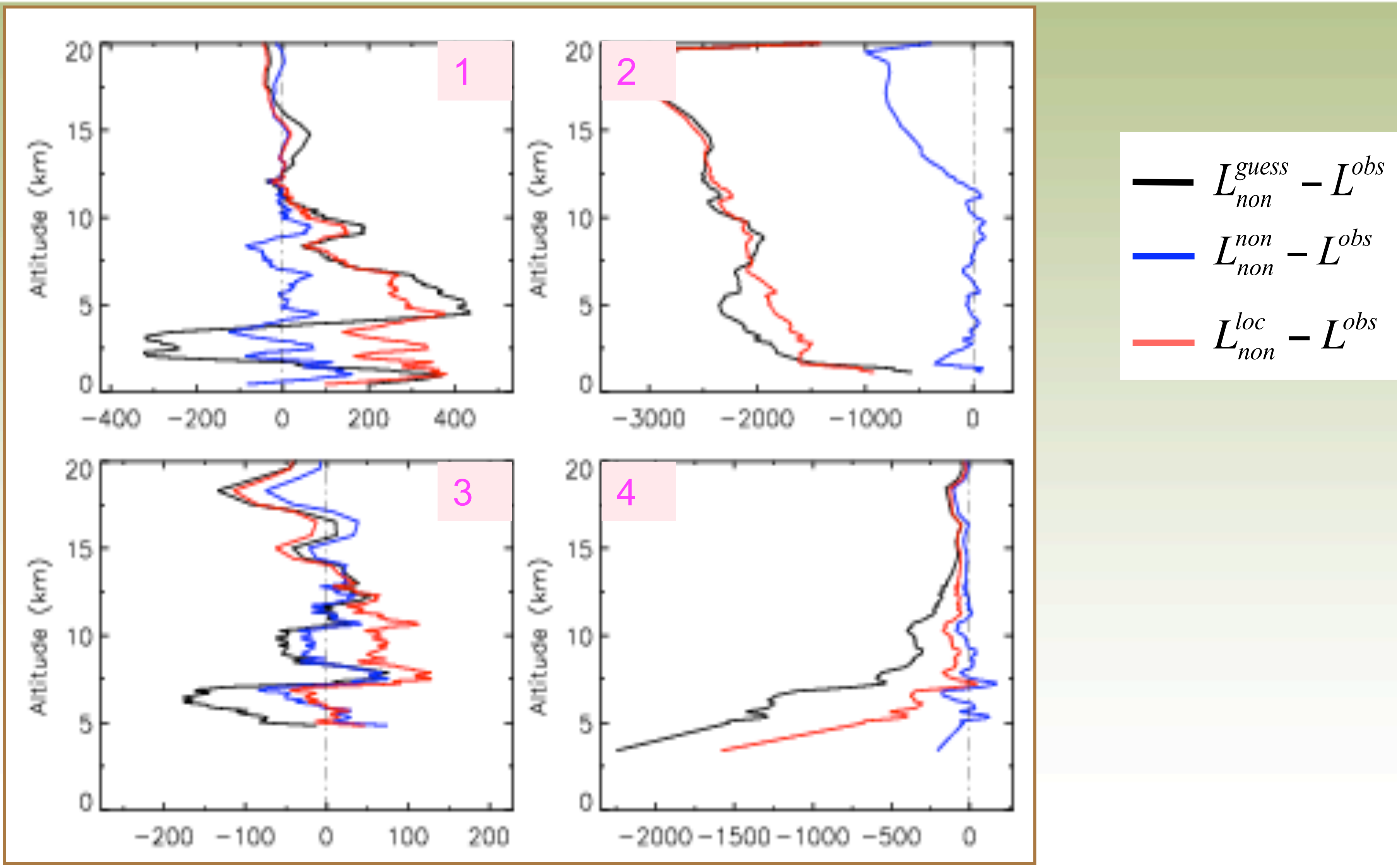
L^{obs} (km N)

$L^{guess}_{non} - L^{guess}_{loc}$ (km N)

Distributions of 8 Selected CHAMP ROs during a 6-h Time Window (03-09 UTC 31 May 2002)

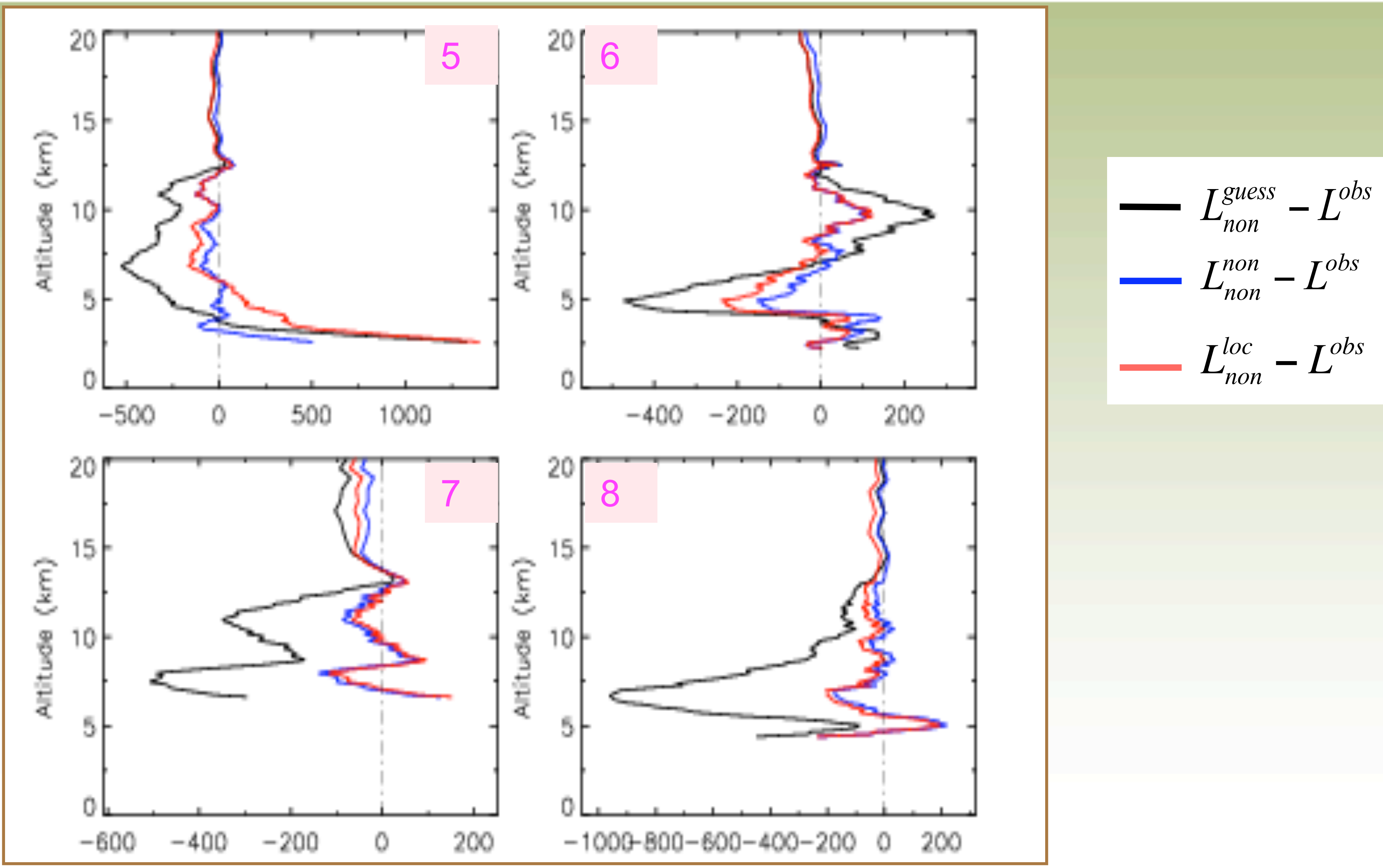


Differences between Non-Local and Local Refractivity Assimilation



Unit : km N

Differences between Non-Local and Local Refractivity Assimilation



Unit : km N

Summary for Part I
(Quality Control)

1. Outliers are identified based on a biweight estimate applied sequentially to GPS retrievals themselves and then to deviations of GPS retrievals from the NCEP analyses.
2. A range check and a symmetric check are incorporated, removing the negative bias near the surface.
3. After QC, GPS RO data show a more coherent pattern compared with NCEP analyses, a more symmetric distribution of a probability density function, much reduced variance in upper levels, and a nearly diagonal vertical correlation distribution at all heights.
4. Although the proposed QC involves very little physical considerations, it is shown that most RO profiles removed by our QC are also identified as of bad quality based on the values of the RO characteristic parameters.
5. Compared with CDAAC's QC, our QC checks RO observations separately at different height levels, but not simply removes either the entire profile or nothing at all based on each RO's characteristic parameters.

Summary for Part II

(Assimilation of α versus N_{loc})

- **Assimilation of bending angle through a ray-tracing technique results a large influence radius than the local refractivity assimilation.**
- **Assimilation of bending angle results in a better fit of model to GPS refractivity than the local refractivity assimilation in terms of the mean standard deviation.**
- **Differences in analyses resulting from assimilating bending angle and local refractivity are weather-dependent. Large mean differences are found at the cloud height for specific humidity and near and above the tropopause for temperature.**

Summary of Part III

(A Non-Local Refractivity Assimilation)

- **The kernel function distribution depends on model resolution, vertical resolution of observations, and geographic location of ROs.**

- **Differences between non-local and local refractivity are largest in the tropics than middle and high latitudes. In the vertical, large differences are found in the low troposphere and stratosphere.**

- **Preliminary results on 8 selected CHAMP ROs suggest that a more significant difference between local and non-local GPS data assimilation results in Southern Hemisphere than in Northern Hemisphere.**

Ongoing Work

- Test QC impact on GPS RO data assimilation results.
- Continue GPS non-local refractivity assimilation research:
  - ✓ Where and when is a non-local refractivity assimilation needed?
  - ✓ Assess impact of local and non-local refractivity assimilation on both global and mesoscale analyses

Acknowledgments

NSF under project no. ATM-9812729.

Integrated Program Office of NOAA National Polar-Orbiting Operational Environmental Satellite System under SMC/CIPN project order no. Q000C1737600086.