Development of a 2DVAR for Combined Retrieval of Ionosphere and Atmosphere: A NASA-funded Project

Tae-Kwon Wee and Bill Kuo
COSMIC Project Office

Ionosphere-Atmosphere Coordination Workshop

Research Goals

What is 2DVAR? Variational Analysis Scheme in the *occultation plane* using a ray tracing observation operator

To serve as a testbed for more comprehensive approaches:

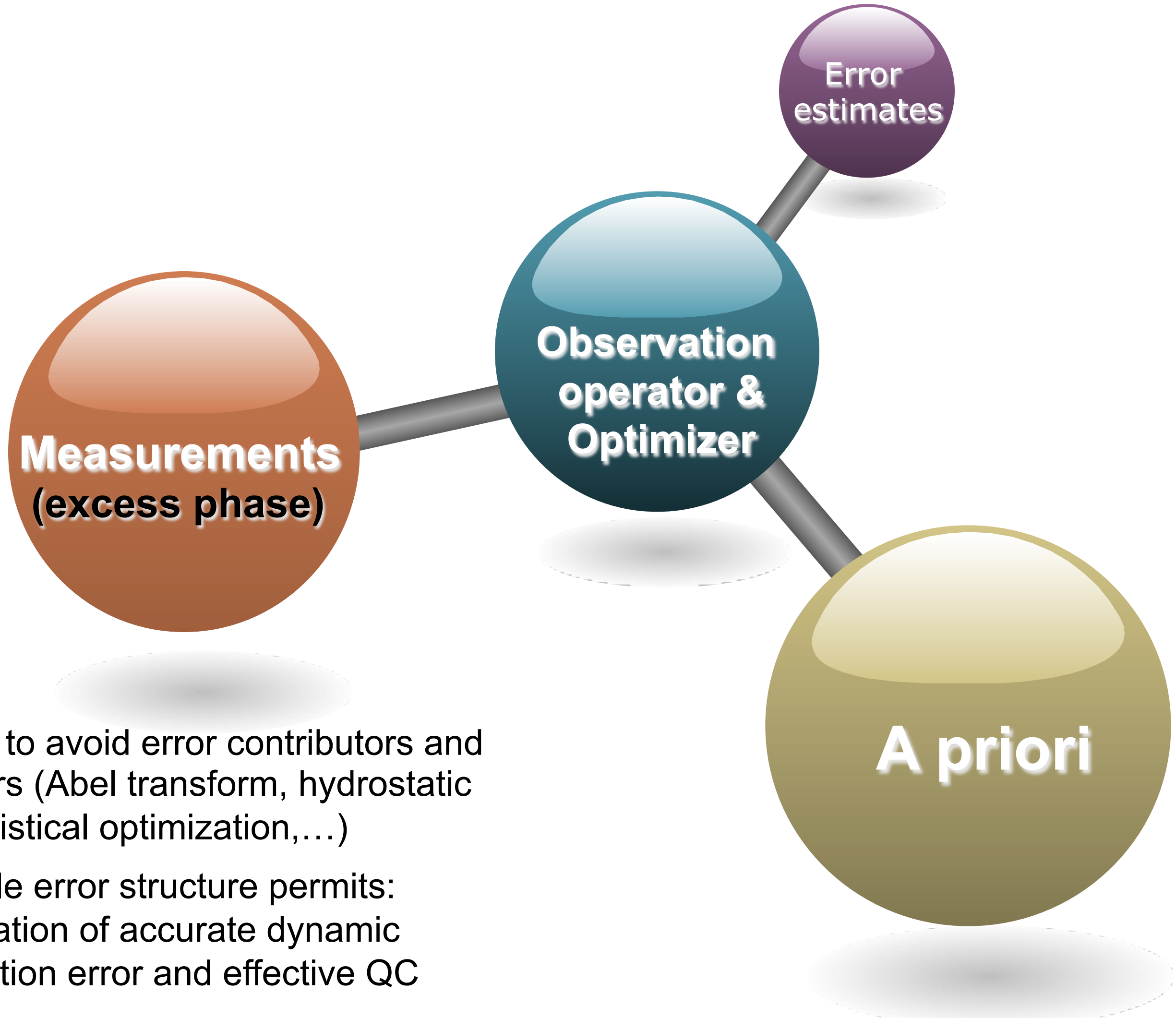
- Ionosphere + neutral atmosphere
- RO + In situ (e.g. ground-based GPS)
- 3DVAR, 4DVAR, EnKF

To provide an affordable, rigorous, replicable environment for RO-related studies

To improve understanding of issues that are difficult to be addressed in 1-d framework:

- Spherical asymmetry
- Super-refraction & atmospheric ducting
- Surface reflection
- Atmospheric multipath

Individual Components of 2DVAR



In order to avoid error contributors and collectors (Abel transform, hydrostatic eq., statistical optimization,…)

Its simple error structure permits: specification of accurate dynamic observation error and effective QC

Ray tracing along complete GPS-LEO links

The primary observable in GPR RO is phase path, the refractive index integrated along the ray path. A complete ray tracing that links GPS-LEO satellites replicates the measurement.

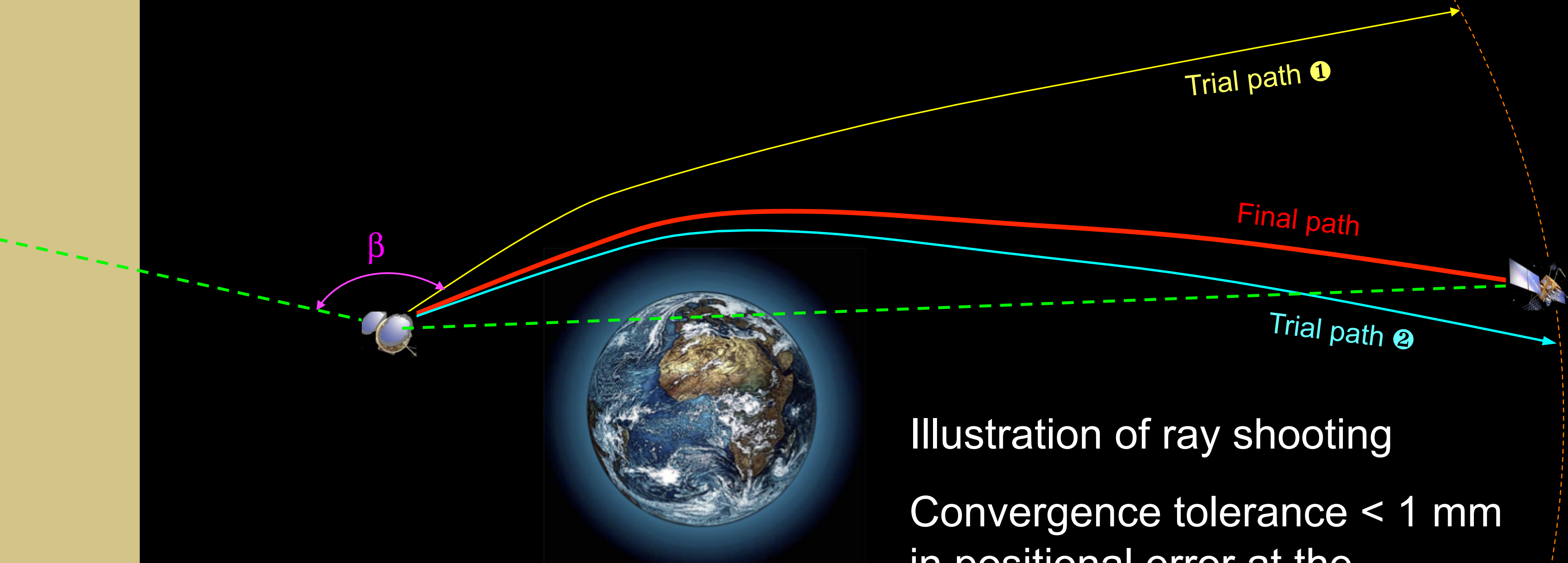
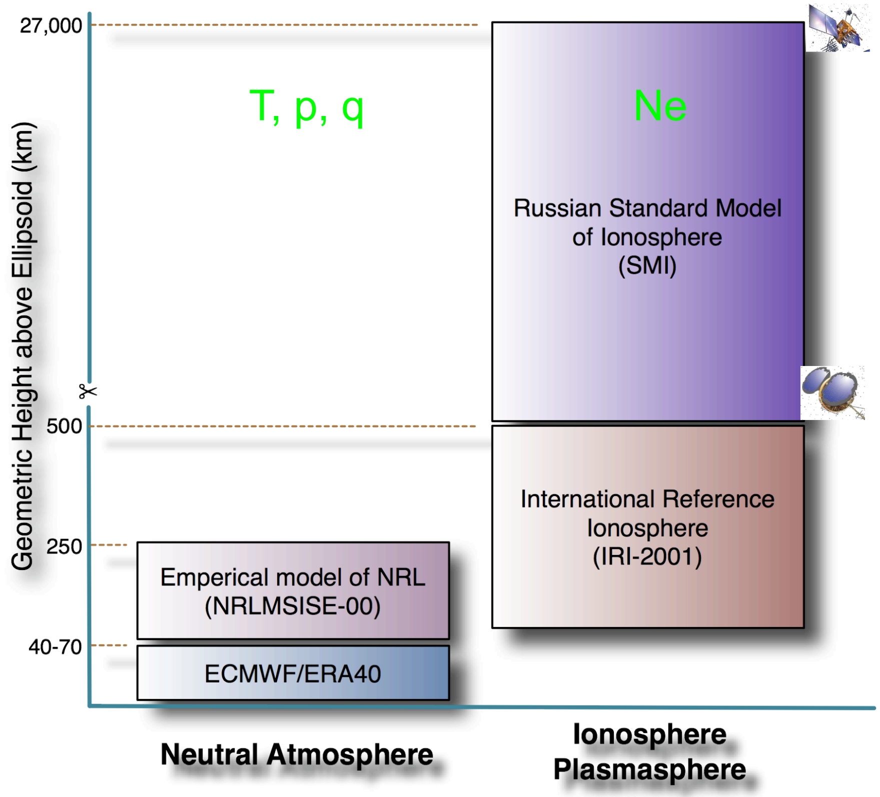
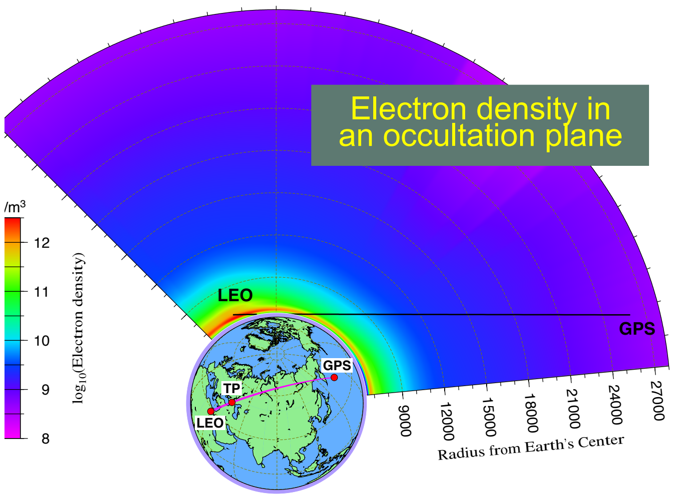


Illustration of ray shooting

Convergence tolerance < 1 mm in positional error at the receiving satellite

A priori information

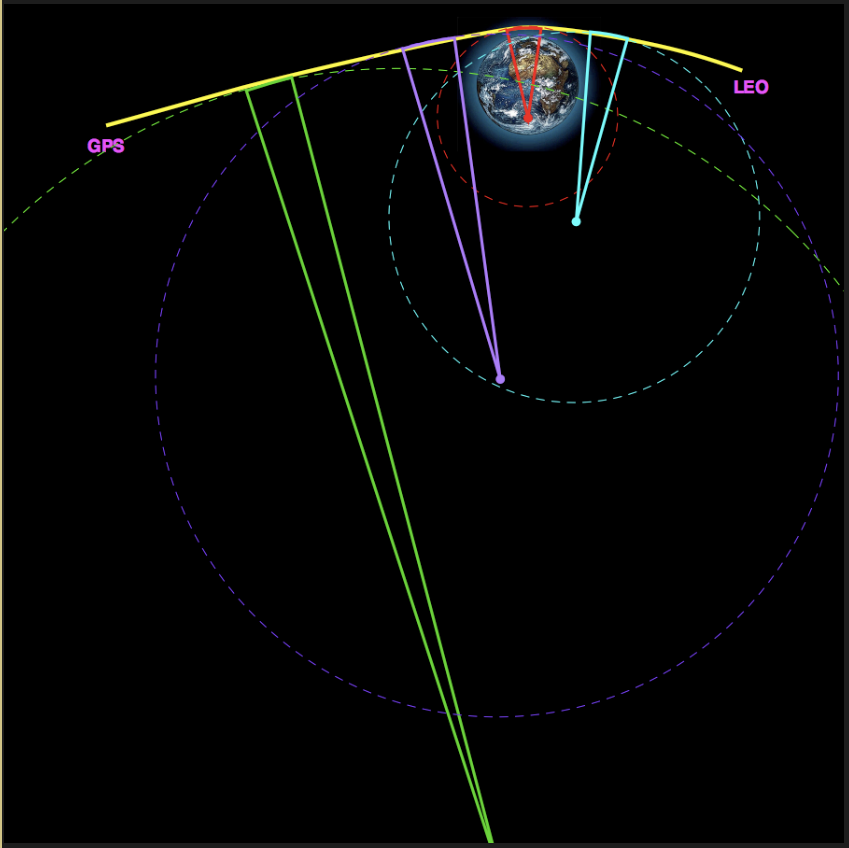




Empirical models for ionosphere and plasmasphere are meant to assist in extracting information from RO data, in perspective of “better-than-nothing”.

It is assumed that the information content of RO data is so rich that the demand for precise a priori is less.

The Curved Ray Tracing

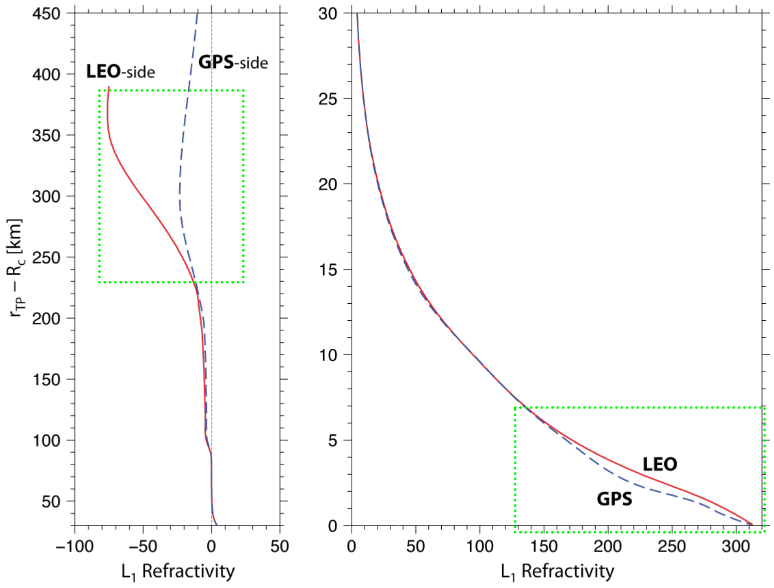


Example of osculating circles used to model a ray path linking GPS-LEO

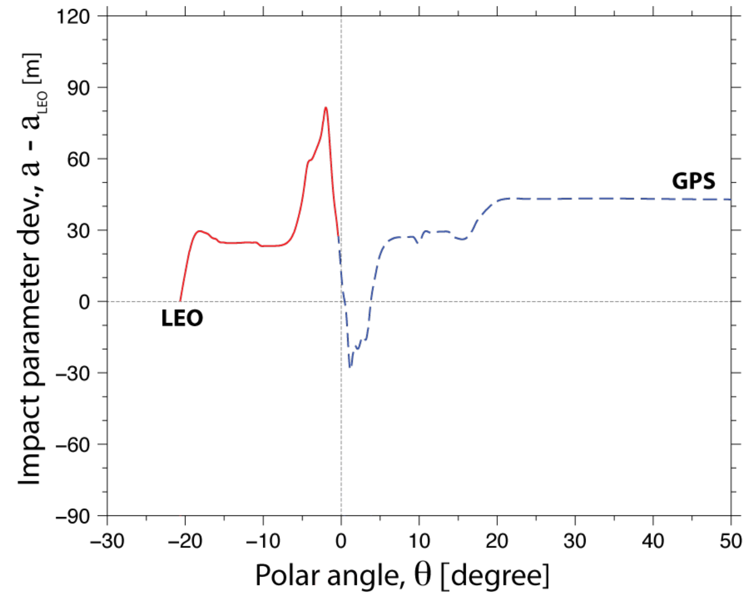
- We do not have any data for high-resolution (10-100 meter) refractivity field
- The ionosphere/atmosphere is continuous and varies slowly, and so is/does ray's curvature
- However, atmosphere is strongly stratified and thus a ray never follows a straight line
- It is possible to compute ray's curvature locally. Ray's bending effect can be accounted for in modeling ray path
- CRT (Wee et al., 2010, JGR) uses parameterized curves replacing straight lines
- CRT permits significantly longer time steps

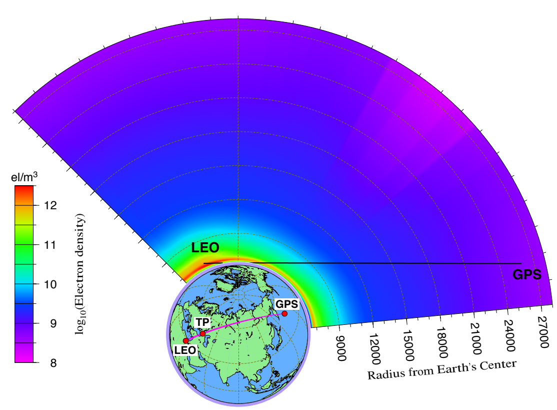
Trace along a ray path (asymmetry)

Refractivity

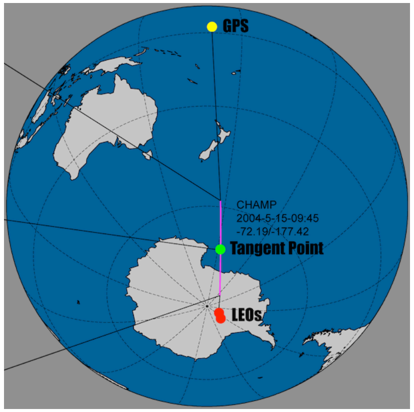


Known as ray invariant



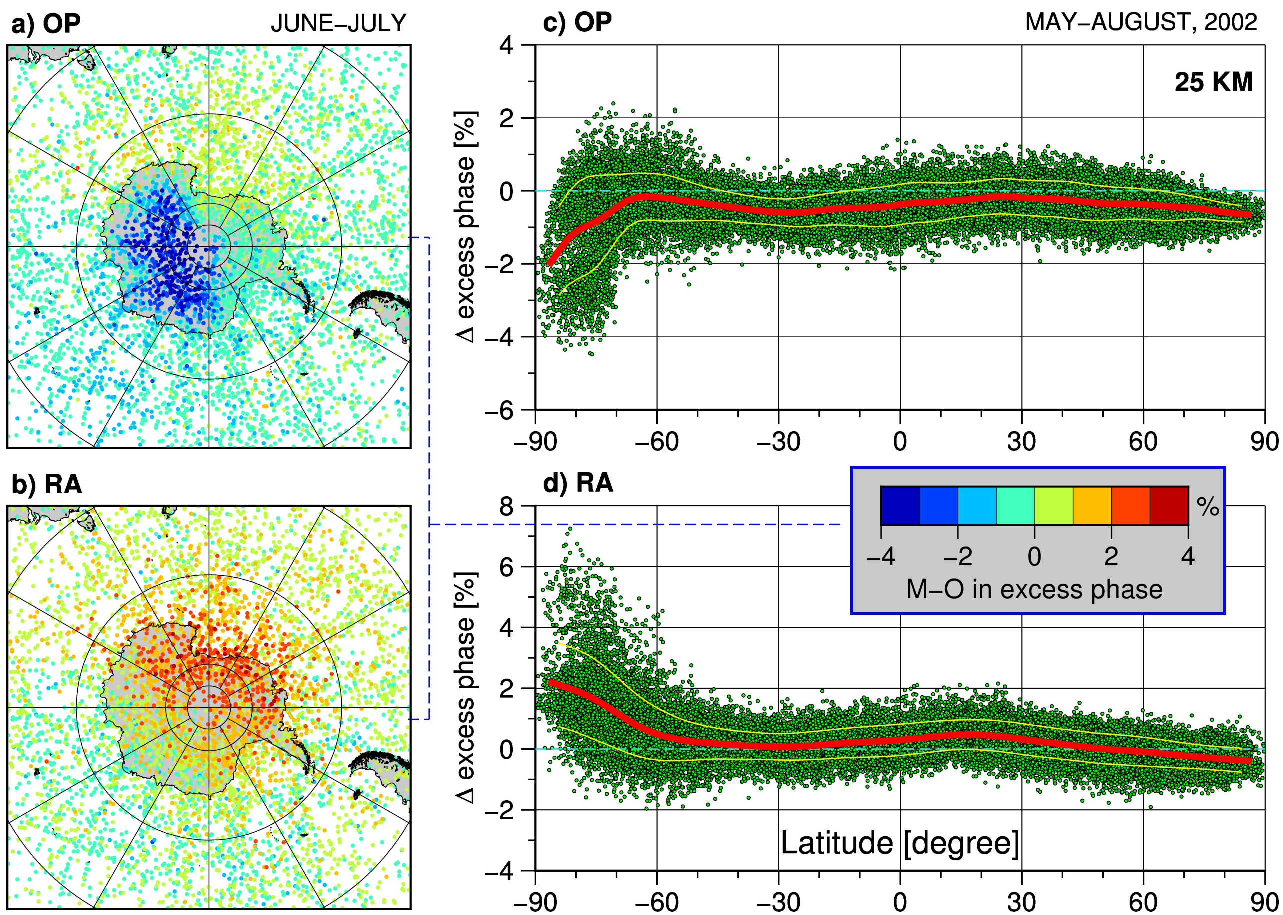


- Standard RO processing is done in 1-d framework, which is based on the assumption of spherical asymmetry.
- Spherical asymmetry is notable in the ionosphere as well as in the lower troposphere.
- It exists even without small-scale disturbances whenever a ray path is aligned latitudinally.



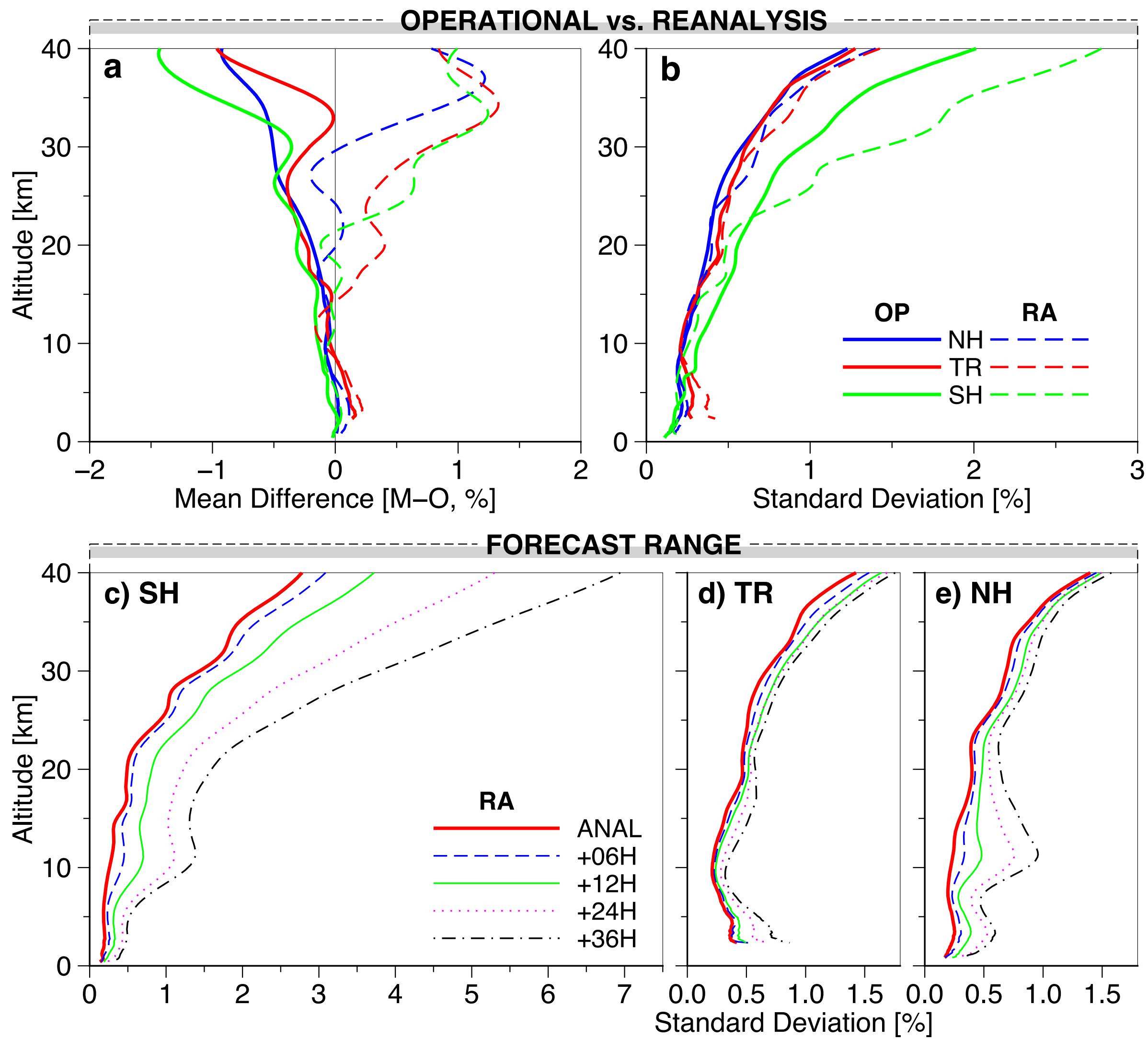
Real-data assessment of CRT

Shown is model-observation in excess phase [%]
OP: Operational ECMWF analysis, RA: ERA40 reanalysis



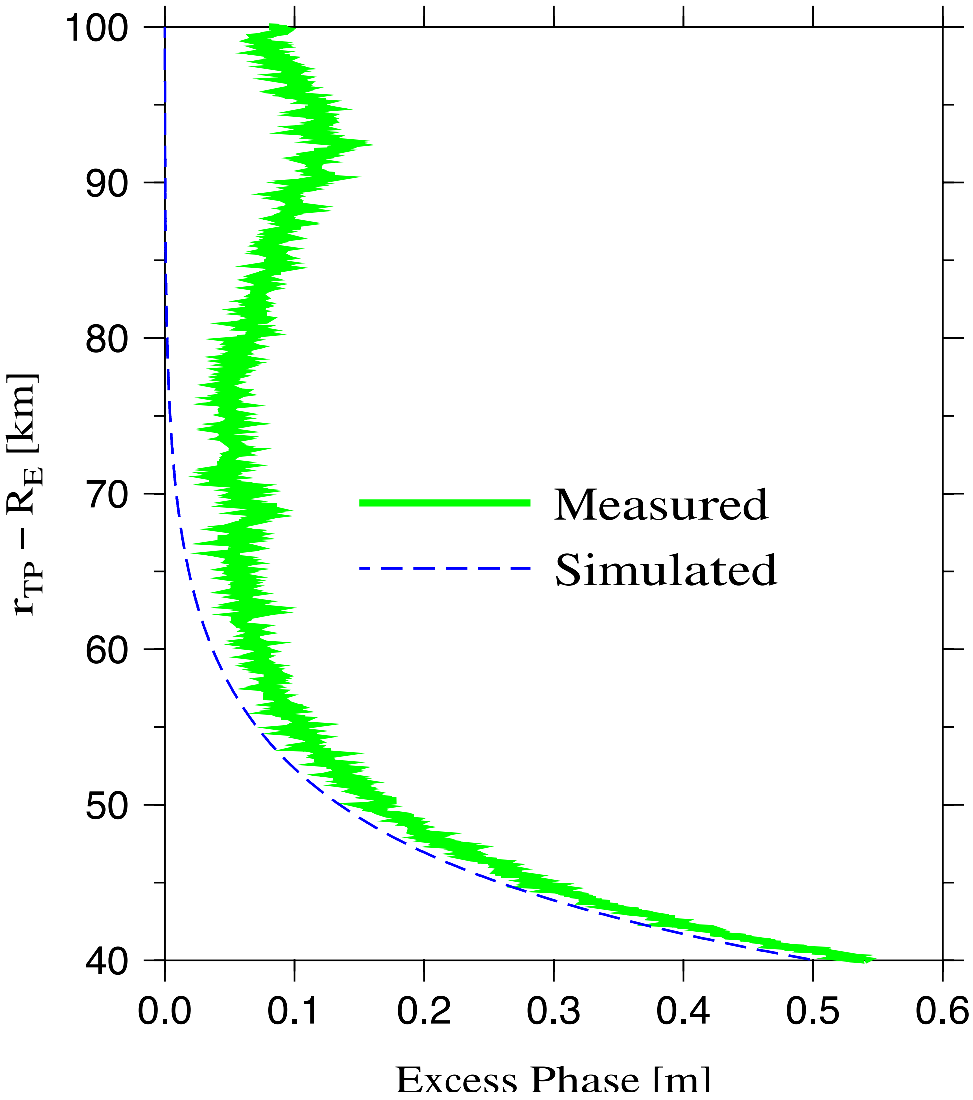
Wee and Kuo [2014, AMTD]

Use of CRT for NWP verification



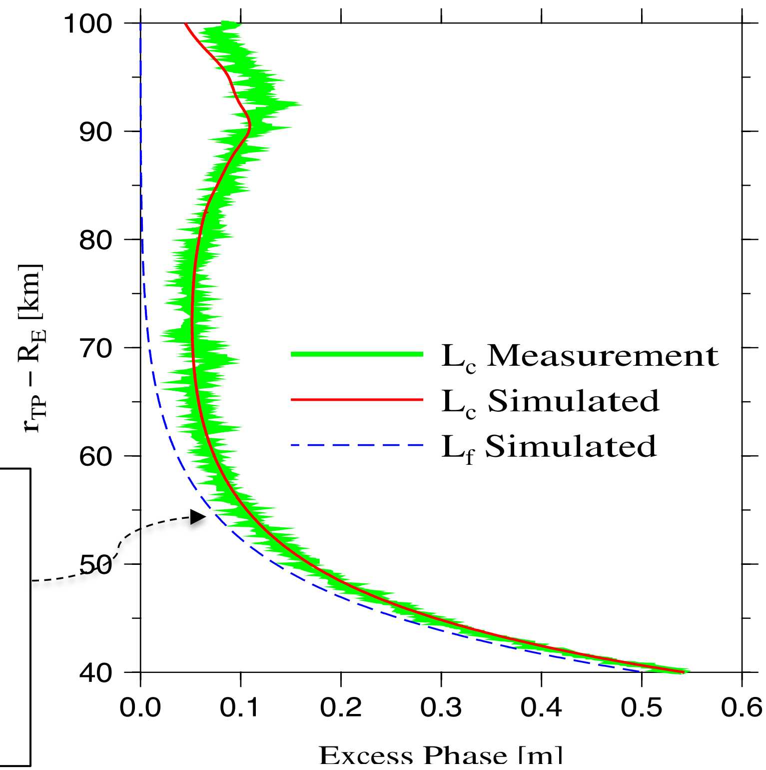
Error growth with forecast lead time

Not everything went that well from the beginning. Sometimes, …



Initially, the simulated phase showed systematic deviations from measurement.
It turns out that this was caused by not including ionospheric effects in the refractive index

ION-free vs. ION-corrected



Ionosphere-corrected (L_c)

- What RO actually measures in the presence of ionosphere

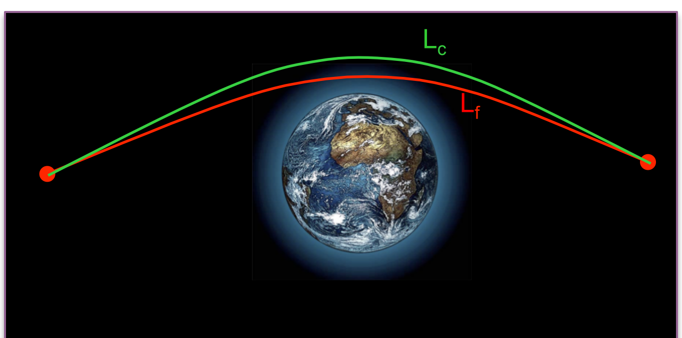
$$L_1 = \rho + \frac{I}{f_1^2} + \varepsilon_1 \quad ; \quad L_2 = \rho + \frac{I}{f_2^2} + \varepsilon_2$$

$$L_c \equiv \frac{f_1^2 L_1 - f_2^2 L_2}{f_1^2 - f_2^2}$$

- L1 & L2 are ray traced separately and then linearly combined

Ionosphere-free (L_f)

- Hypothetic, obtainable when the ionosphere were absent
- What we intend to obtain from an ideal ionospheric correction



- The presence of ionosphere changes ray path.
- Neutral atmospheric effect experienced along the path (close to L_c) differs from L_f,
- L_c differs from L_f , even if the perfect correction along the path were possible
- When ionospheric influence is great, even ionospheric climatology is useful
- In general, the L_c - L_f difference is small

Inclusion of the ionosphere is essential for studying atmospheric effect; however,

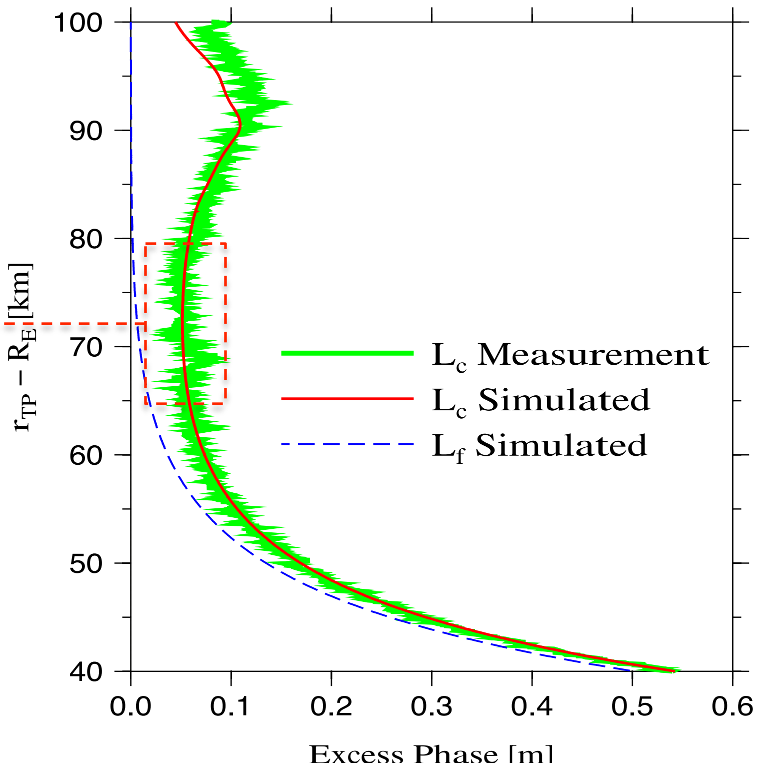
Measured L_c is so noisy, compared to the difference from modeled L_c

Temporal differentiation (i.e., Doppler) doubles the noise level, making it difficult to detect the difference even compared to L_f

Rigorous approach and precise error estimation are required to relate the difference in L_c to that in L_f

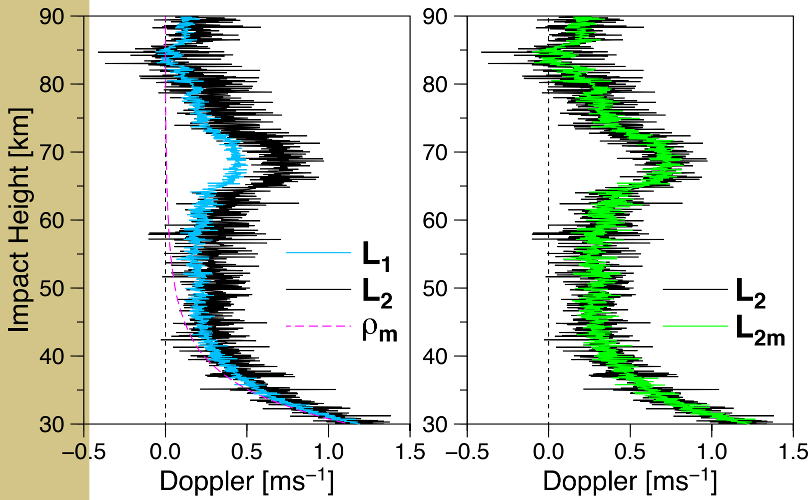
In order to separate the error resulting from imperfect ionospheric correction (from caused by spherical asymmetry), the following studies are conducted:

- Noise separation and signal detection via dynamic filtering and singular spectrum analysis
- Dynamic error estimation
- Noise-Aware combination
- Variational combination

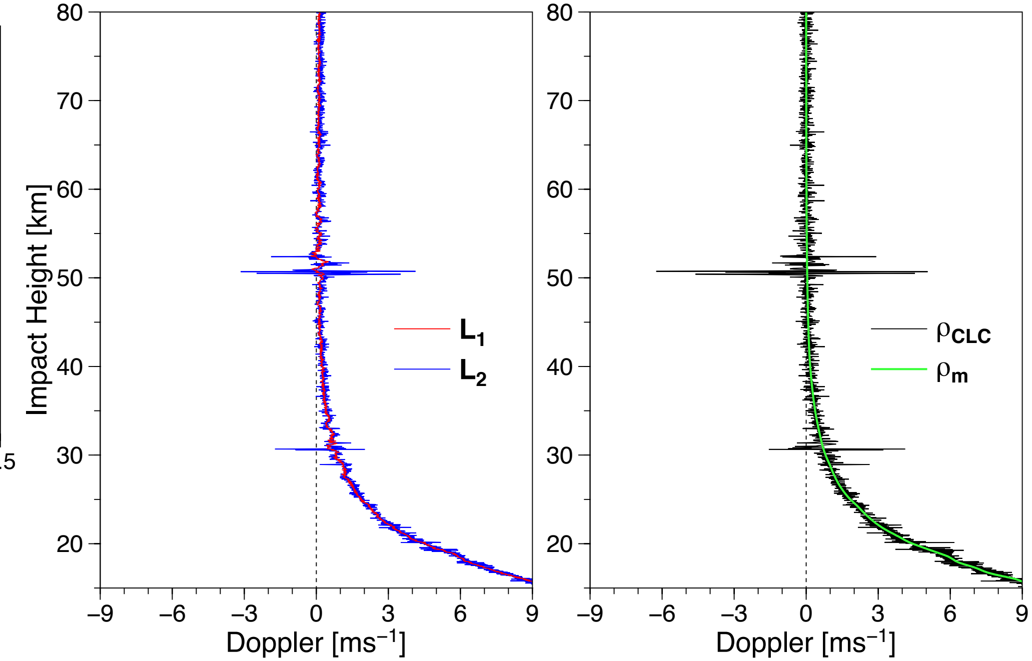


Quality Control: L1-L2 is not always ionospheric effect

Doppler: $\Delta\Psi(i) = \Psi(i+1) - \Psi(i)$

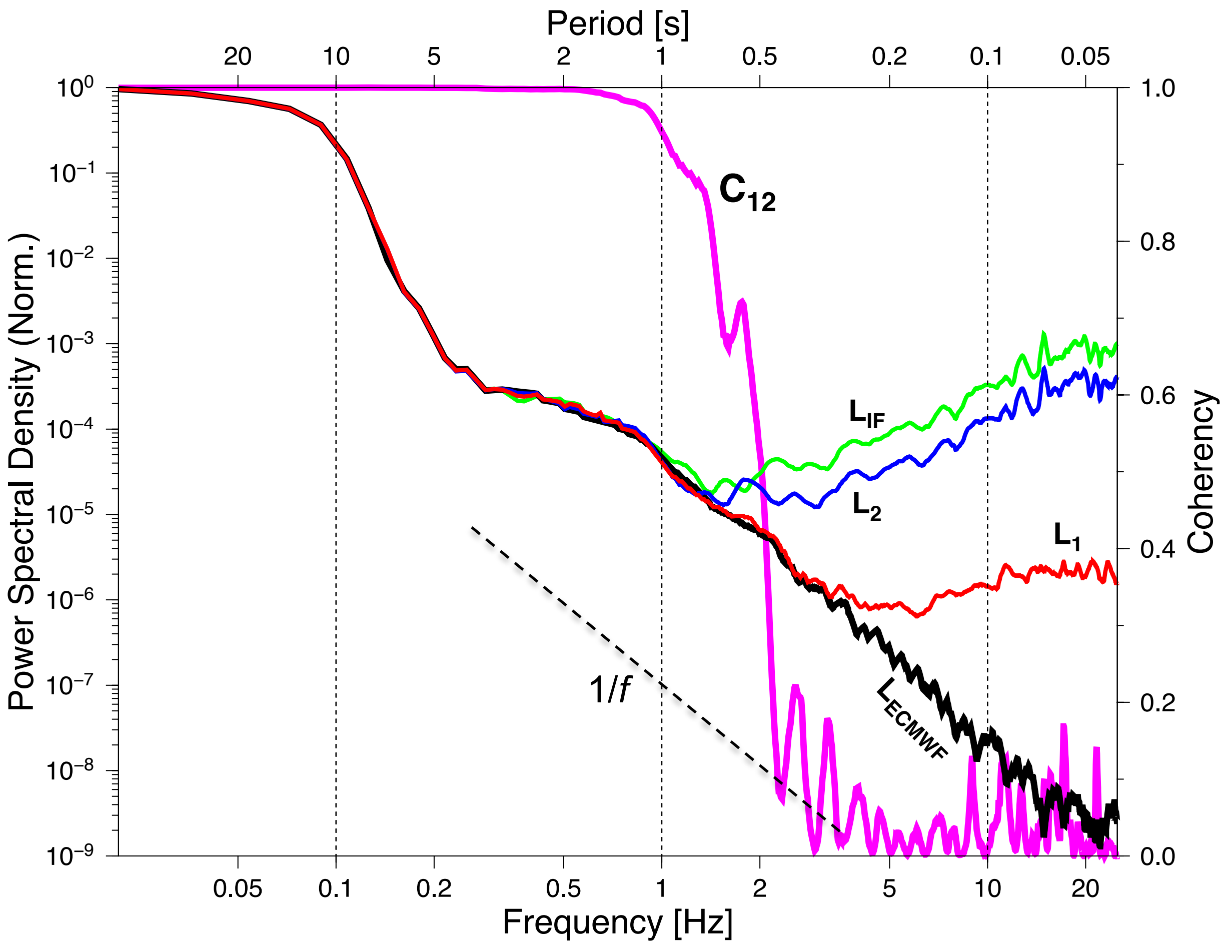


$$L_{2m} \equiv \rho_m + \frac{f_1^2}{f_2^2}(L_1 - \rho_m)$$



- By having two frequencies available, we can tell a certain feature is realistic or not with an increased level of confidence.
- This could have been another strength of dual frequency, in addition to the ionospheric correction, if L1 and L2 were about the same in their quality
- In reality, L2 is so noisy that L1 can be used for QC of L2, instead of cross-validation
- Inter-channel coherence is an important measure to detect signal component and to isolate noises

Dynamic filtering & Inter-channel coherence



L1-L2 Coherence

$$C_{xy} = \frac{|G_{xy}|^2}{G_{xx} G_{yy}}$$

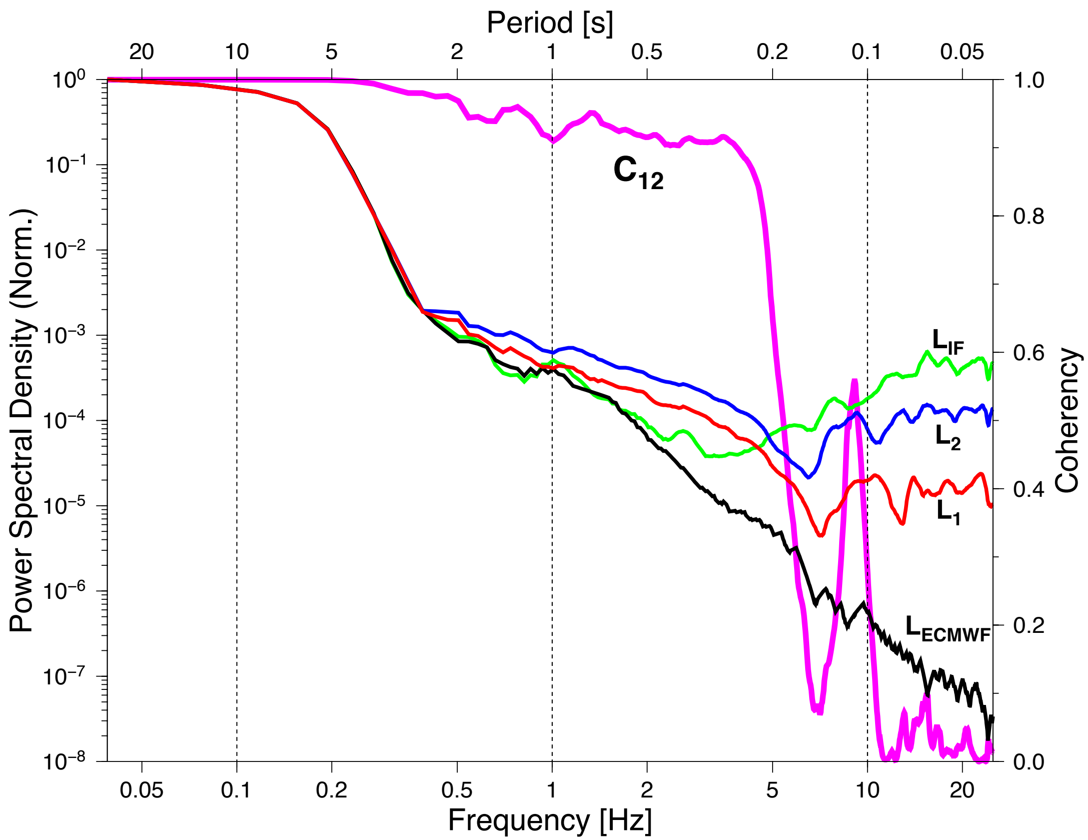
G_{xy} : cross-spectral density

G_{xx} : autospectral density

Multitaper PSD allows us to detect noise floor (3 Hz for L1 and 1.5 Hz for L2) for individual soundings.

Coherence in the range of noise is low.

Coherence is a key measure for signal detection



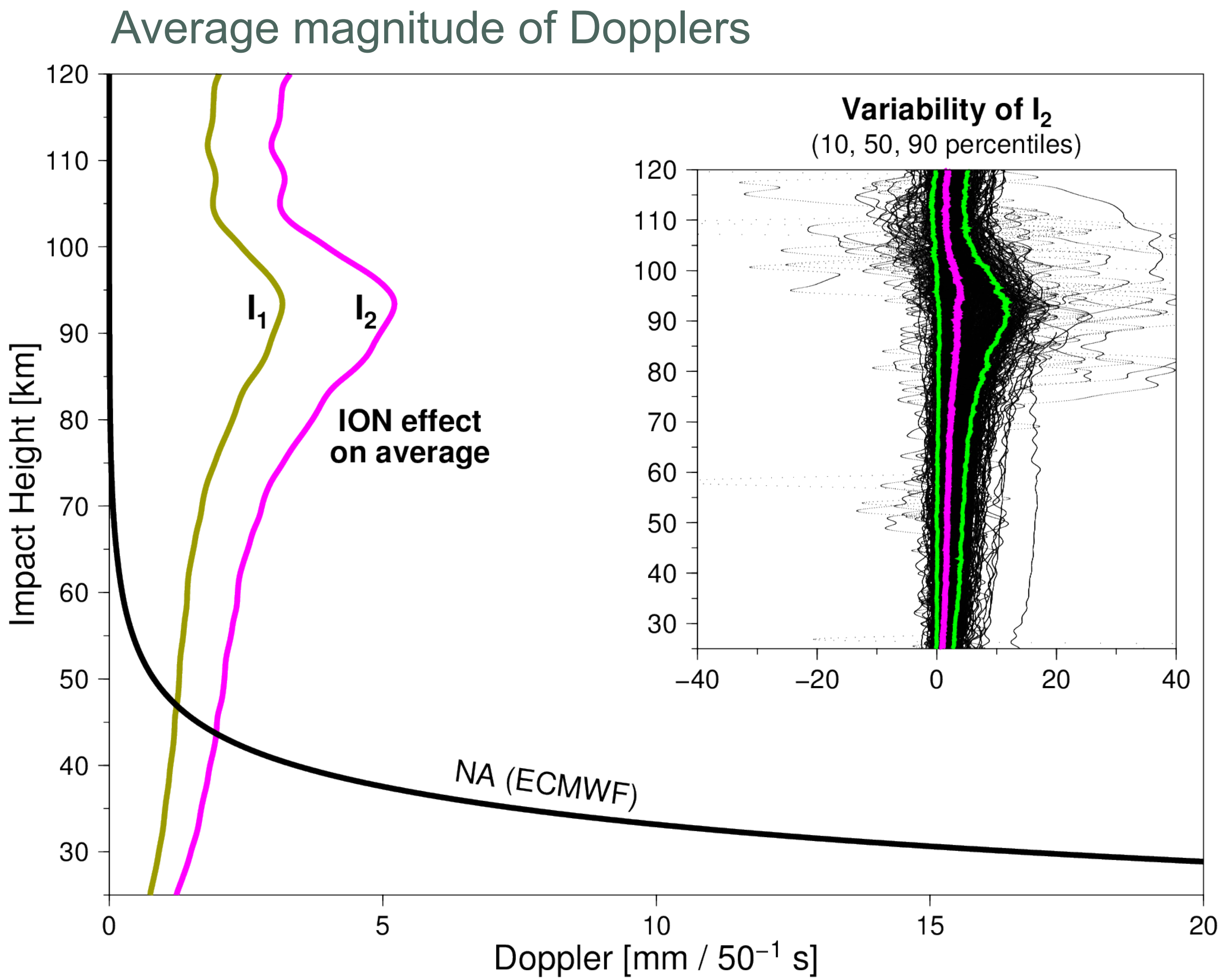
L1 & L2 deviate from model and from each other starting from 0.4 Hz

However, L_c stays close to ECMWF and the coherence in the range is high

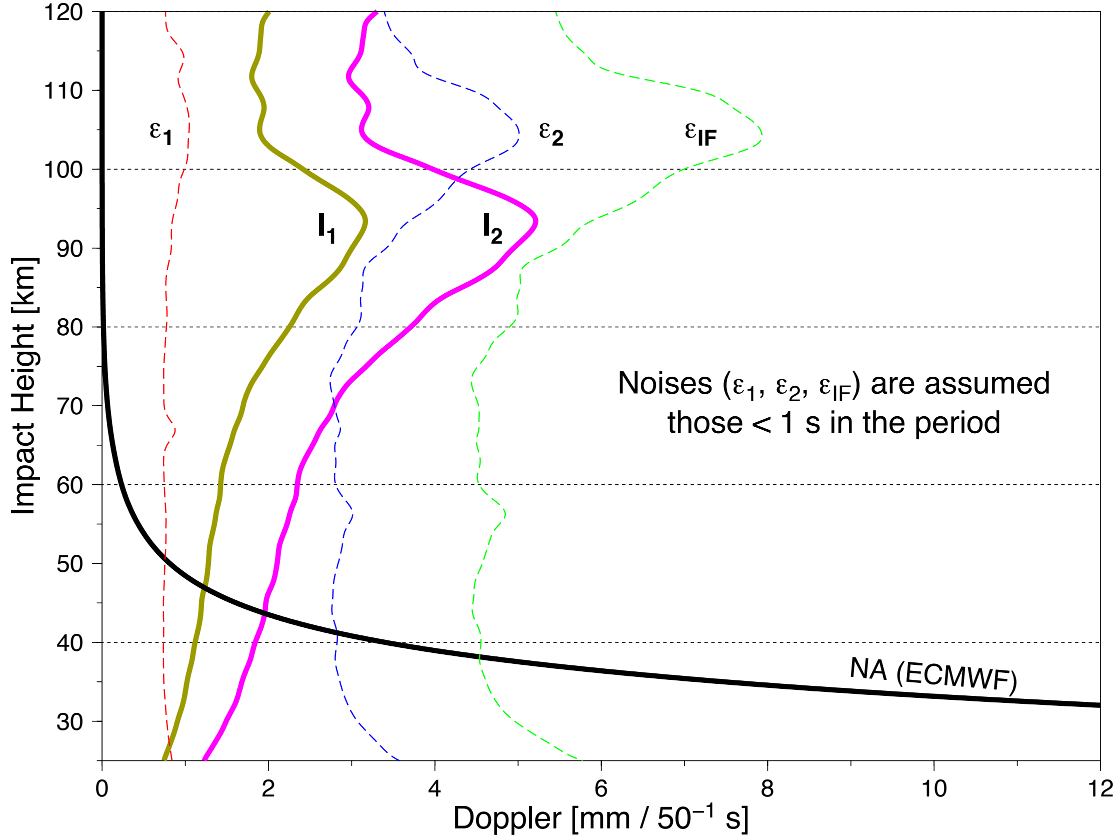
The early deviation of L1 & L2 from ECMWF is due to Ionospheric effect

Inter-channel coherence is thus a reliable measure for signal detection

Information content of Doppler



SNR



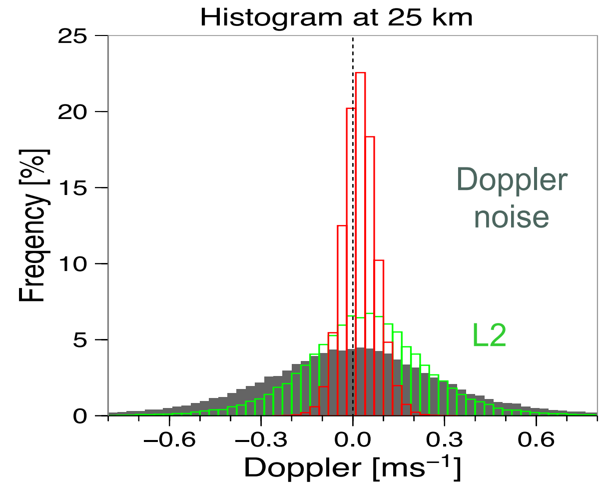
Above 40 km, L_c noise exceeds the absolute magnitude of neutral atmospheric effect, and is greater than ionospheric effect throughout the entire height range. Hopeless.

L1 seems still useful up to 50 km even for neutral atmosphere. How can we make good use of this?

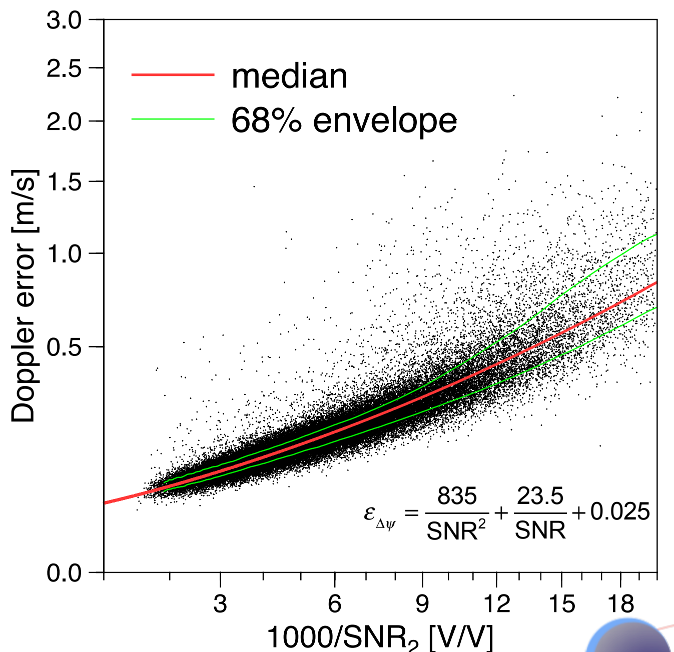
Dynamic observation error

Data assimilation and retrieval algorithms require accurate estimates of observation error. However:

- These method often use static observation error
- Noise level of RO data varies significantly from one occultation to another, and L2 noise, relative to L1 noise, also changes greatly
- **Use of a static observation error is thus suboptimal**
- Dynamic filtering, SSA, and signal modeling provide insight into measurement noise and allows us to characterize them
- Dynamic (occultation-specific) observation error has great potential to strengthen any uncertainty-based state estimations (e.g., data assimilation methods and retrieval schemes)



An example of dynamic error



Noise-Aware Combination (NAC)

NAC (Wee and Kuo, 2013, JGR) is an example of L1 and L2 discrimination

- Statistical optimization is needed to cope with noise amplification in the IF linear combination. Because IF combined data is so noisy, a heavy weighting must be given to a priori. This is a major error source
- L1 and L2 differ in the quality and so don't treat them equally
- Instead, discriminate them in the combination by giving a bigger weighting to L1 and smaller to L2
- **Combines L1 and L2 data with background, achieving minimum error variance (rather than amplifying noise)**

The observation eqs.

$$L_1 = \rho + \frac{I}{f_1^2} + \varepsilon_1 \quad ; \quad L_2 = \rho + \frac{I}{f_2^2} + \varepsilon_2$$

Variable transform using an N.A. model

$$\tilde{L}_1 = \frac{f_1}{f_2}\left(L_1 - \rho_m\right) = \frac{I}{f_1 f_2} + \frac{f_1}{f_2}\left(\varepsilon_1 + \varepsilon_m\right)$$

$$\tilde{L}_2 = \frac{f_2}{f_1}\left(L_2 - \rho_m\right) = \frac{I}{f_1 f_2} + \frac{f_2}{f_1}\left(\varepsilon_2 + \varepsilon_m\right)$$

Seek for the optimal combination

$$\frac{I^*}{f_1 f_2} = a\tilde{L}_1 + (1-a)\tilde{L}_2$$

When $\overline{\varepsilon_1\varepsilon_2} = \overline{\varepsilon_1\varepsilon_m} = \overline{\varepsilon_2\varepsilon_m} = 0$, the solution found to be

$$\frac{I^*}{f_1 f_2} = \frac{\left\{\left(\frac{f_2}{f_1}\right)^2\left(\varepsilon_2^2 + \varepsilon_m^2\right) - \varepsilon_m^2\right\}\tilde{L}_1 + \left\{\left(\frac{f_1}{f_2}\right)^2\left(\varepsilon_1^2 + \varepsilon_m^2\right) - \varepsilon_m^2\right\}\tilde{L}_2}{\left(\frac{f_1}{f_2}\right)^2\left(\varepsilon_1^2 + \varepsilon_m^2\right) - 2\varepsilon_m^2 + \left(\frac{f_2}{f_1}\right)^2\left(\varepsilon_2^2 + \varepsilon_m^2\right)}$$

Variational combination of multi-freq data

Seeks the optimal **x** that replicates the L1 and L2 measurements as closely as possible while satisfying other constraints imposed. Inferring atmospheric states from noisy measurements is ill-posed. However, adjusting model states to better replicate measurements is a well-posed problem. Wee and Kuo [2014, JGR, in press]

Background | **Observation**

$$J(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}-\mathbf{x}_b\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}_b\right) + \frac{1}{2}\left\{\mathbf{y}_o - H(\mathbf{x})\right\}^{\mathrm{T}}\mathbf{R}^{-1}\left\{\mathbf{y}_o - H(\mathbf{x})\right\}$$

$$\mathbf{x}_{2M} = (I_1, I_2, I_3, \ldots, I_M, \rho_1, \rho_2, \rho_3, \ldots, \rho_M)^T$$

$$\mathbf{B}_{2M\times 2M} = \begin{bmatrix} \mathbf{B}_I & \\ & \mathbf{B}_\rho \end{bmatrix} \quad \text{Combination}$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{n_e} & & & \\ & \mathbf{B}_T & & \\ & & \mathbf{B}_q & \\ & & & \mathbf{B}_{p_s} \end{bmatrix} \quad \text{1DVAR}$$

$$\mathbf{y}_{2N} = (L_1^1, L_2^1, L_3^1, \ldots, L_N^1, L_1^2, L_2^2, L_3^2, \ldots, L_N^2)^T$$

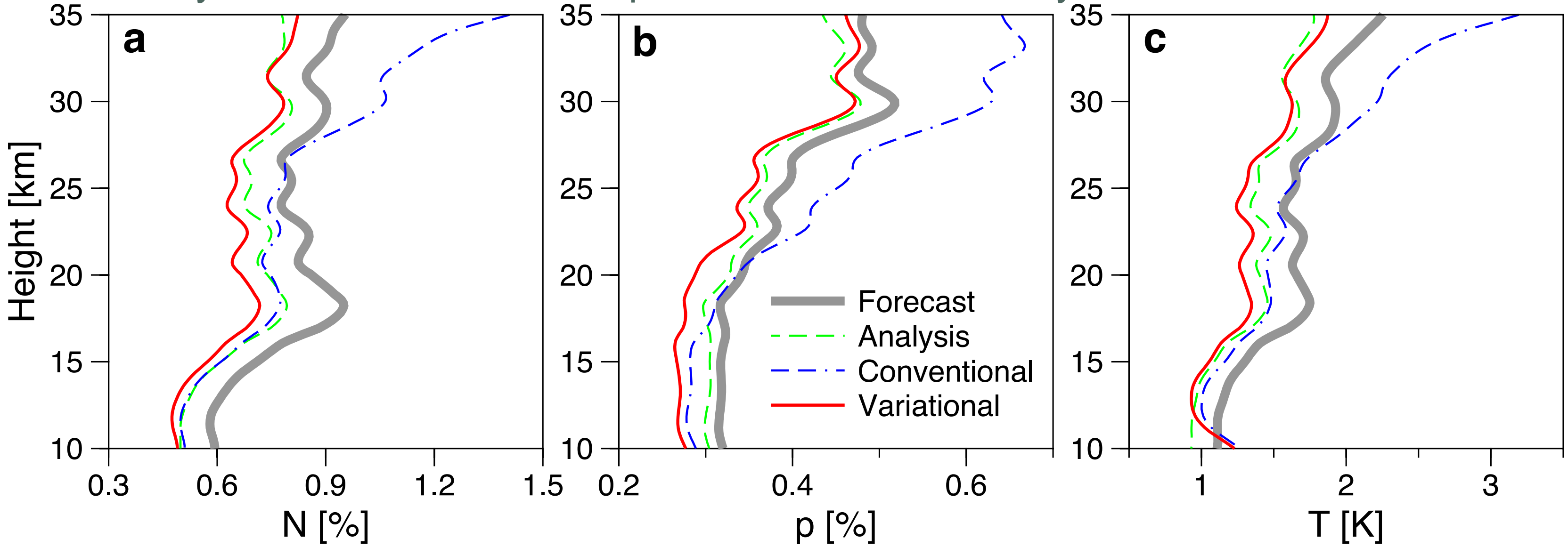
$$\mathbf{R}_{2N\times 2N} = \begin{bmatrix} \mathbf{R}_1 & \\ & \mathbf{R}_2 \end{bmatrix} \; ; \quad \varepsilon_\alpha \simeq \left(\frac{1}{r_L} + \frac{1}{r_G}\right)\frac{\varepsilon_{\Delta\psi}}{\Delta\theta}$$

$$H_1 = \rho + \frac{I}{f_1^2} \; ; \; H_2 = \rho + \frac{I}{f_2^2} \quad + \quad \text{interpolation}$$

- Additional frequencies (e.g. L5) can be easily added
- Classical linear combination for triple-frequency is a bit complicated and increases noise level even more
- Based on the principle of minimum variance, VAR further reduces error variance when extra data are added, virtually without any extra cost

Verification with radiosonde observations

Analysis & forecast refer to operational ECMWF analysis and 24-h forecast



Surprisingly, VAR in 1-d framework produces results better than ECMWF analysis, even if the verifying radiosonde data are assimilated into DA system; whereas RO data are independent from sonde data:

- In methodological point of view, 1DVAR can't compete with the state-of-the-art 4DVAR system of the ECMWF
- On top of that, ECMWF analysis was made by assimilating observations available from all other platforms, including the radiosonde data used here as verification

How VAR can outperform 4DVAR?

- L1 data are so good and VAR can make good use of it. VAR is able to to discriminate L1 and L2 in view of data quality without causing systematic error in the resulting ionospheric and atmospheric parameters. If everything works out, in principle, VAR can achieve an error variance lower than L1's
- By combining multiple frequencies with background directly (instead of using pre-combined data), noise amplification due to linear combinations is avoided.
- Bypassing statistical optimization, the error caused by the process is avoided.
- By using Phase and Doppler, reliable dynamic error can be provided. Effective QC is possible.
- Measurement errors are minimally correlated in time and space
- It is difficult to quantify the importance of ionospheric background for now, but even IRI seems useful