A variational regularization of Abel transform for GPS radio occultation

Tae-Kwon Wee

University Corporation for Atmospheric Research, Boulder, Colorado, USA

Correspondence to: Tae-Kwon Wee (wee@ucar.edu)

5 **Abstract.** In the Radio Occultation (RO), the refractivity is generally obtained from the inverse Abel transform of measured bending angle, often called Abel inversion (AI). While concise and straightforward to apply, AI is susceptible to the error present in the bending angle. Aiming at reducing the adverse effects of the measurement error, this study proposes a new method for determining the refractivity through a variational regularization (VR). The method approximates the inversion of the forward Abel transform by an optimization problem in which the regularized solution describes the measurement as

10 closely as possible within the measurement's considered accuracy. The optimal problem is then solved iteratively by means of the adjoint technique. VR incorporates the prior information about measurement characteristics and desired behaviour of the solution into the regularization via error covariance matrices. In contrast to variational data assimilations, VR holds the control variable in the measurement space. This makes VR particularly effective by allowing the method to benefit from the posterior height determination and to deal with model's error in the impact parameter. The advantages are elaborated using a

15 purposely corrupted synthetic sounding and with known true solution. The competency of VR relative to AI is validated with a large number of actual RO soundings. The comparison with nearby radiosonde observations shows that VR is considerably smaller than AI in both random and systematic errors. It is concluded based on the results presented in this study that VR offers a definite advantage over AI in the quality of refractivity.

1 Introduction

20 The Abel transform pairs (Abel, 1826) are widely used to reconstruct radially (or spherically) symmetric physical parameters from their line-of-sight projections in a variety of disciplines in engineering and science. In the Global Positioning System (GPS) Radio Occultation (RO) technique, the inverse Abel transform (often called Abel inversion) of measured bending angle in particular has become a cornerstone serving as the standard means of deriving the refractive index. Knowledge of refractive-index structure in the atmosphere is valuable for numerous applications relevant to weather and climate.

25 The beauty of Abel inversion (AI hereafter) lies in mathematical exactness and conciseness. The shortcoming on the other hand is that it integrates flaws as well as signal components in the measurement. Therefore, the analytical inversion works well for measurements with exceptionally rigorous accuracy and GPS RO is believed to provide such a measurement. Nonetheless, RO measurements are subject to non-negligible errors arising from diverse sources to which AI is sensitive. More fundamentally, the premise of AI (i.e., spherical symmetry of the atmosphere) is never strictly fulfilled. Thus, the

spherical asymmetry in the atmosphere is one of sources of the measurement error. A major concern in AI is that the measurement error is accumulated and propagated in vertical direction without being controlled. Consequently, a single corrupted piece of the measurement affects not just the location of the particular datum but a wide area that the Abel kernel dictates, and thereby deteriorates the solution even in the region that received RO signals are clean. Previous validation

5   studies suggest that the RO refractive index derived through AI is of high quality, which however does not necessarily mean that AI yields the best possible refractive index in the presence of measurement error.

An alternative to the analytical inversion is finding the inverse solution of the forward Abel transform (FAT) numerically. In our setting, FAT simulates the bending angle given the refractive index. Thus, the numerical inversion (note that it differs from numerical implementations of the analytical AI) attempts to find the refractive index that maps into the measured

10   bending angle through the FAT. Unfortunately, inverse problems of the kind of FAT are generally known to be ill-posed in the presence of measurement error, e.g., the solution may not exist or may not be unique. Therefore, it is challenging or even infeasible at times to obtain useful solutions by applying straightforward algebraic methods to the underlying discrete operators.

Regularization methods such as the Tikhonov regularization (Tikhonov, 1963) are widely used to tackle the ill-posedness in

15   inverse problems (Schuster et al., 2012, and references therein). Based on the notion that it is practically undesirable to fit every detail of noisy measurements in solving the inverse problem numerically, the approach considers an optimization problem in which the solution describes the measurements closely, while the solution's behaviour is regularized as per a priori knowledge. The solution is sought by minimizing a joint function that consists of the data fidelity term (which gauges the discrepancy between measurement and its model counterpart) and the penalty term (which regularizes the solution).

20   Besides warranting that the inverse problem remains well posed, regularization methods are known useful for reducing adverse effects of measurement error.

Regularization methods have proven useful for a variety of inverse problems in different forms. However, it is seldom straightforward to apply those to real-world problems. First of all, the computational cost, which rises rapidly as the number of model parameters increases, can be expensive when applied to GPS RO. That is because the limb-viewing geometry

25   offers high-resolution measurements; and, one typically attempts to retrieve a profile of refractive index in as much detail as the data permit. More importantly, the regularization must be customized for individual applications in order to be effective. For instance, different observing systems differ in measurement characteristics and solution's desired behaviour. Thus, the specifics of an optimal regularization differ from one problem to another and those for the Abel transform in GPS RO are underexplored.

30   In this study, we propose and study a variational regularization (VR hereafter) for the Abel transform in GPS RO. The degree of benefit attainable from using regularization methods depends on the quality of measurement. That is, while comparable to AI in the solution for error-free measurements, regularization methods are more promising for noisier measurements. In order to attain the benefit maximally, however, it is essential to estimate the uncertainties of RO measurement and a priori realistically and take them into account properly. In this study, the data uncertainties are estimated

Atmospheric
Measurement
Techniques
Discussions

by a reliable method, and they are incorporated into the proposed regularization method through error covariance matrices. This differs from classical regularization methods, which in general do not make use of the data uncertainty. The quality and vertical coverage of RO measurements differ significantly across occultation events varying with the occultation geometry, atmospheric conditions encountered, and other environmental factors. The variability causes a well-known difficulty in

5   classical regularization methods as to determining the regularization parameter, which is a constant that controls the tradeoff between the data fidelity term and the penalty term. The proposed method on the other hand is highly adaptive to those variations because the error covariances offer a proper scaling for the data fidelity and penalty terms. Moreover, the proposed method reduces the computational cost significantly, solving the underlying inverse problem iteratively by means of the adjoint method, which is a very efficient way of calculating the gradients of the cost function with respect to all

10 control parameters at once.

The quality of AI-derived refractive index also depends on how the AI is numerically implemented. However, varying implementations might not cause significant difference in the refractive index because the data resolution of RO bending angle is sufficiently high so that the numerical error due to the discretization of AI is limited in the magnitude. Moreover, RO data processing centres worldwide are now using polished algorithms aiming at attaining the best data quality, which

15 includes well-tuned implementation of the discrete AI. The additional quality improvement in the refractive index obtainable from further polishing the discrete AI is thus likely to be marginal at most. This necessitates a breakthrough that can improve the quality of refractive index to a level beyond AI can tender. This study explores the possibility with the proposed method of variational regularization. Section 2 describes the methods of obtaining refractive index from measured bending angle. Section 3 demonstrates the performance of the proposed method against AI with a synthetic sounding. Some issues in

20 practical application of the Abel transform are also discussed. In Sect. 4, a real-data validation with respect to two sets of radiosonde observations is given. We conclude with a short summary of our results in Sect. 5.

2 Method

2.1 Abel inversion

The total transpired phase path of a radio wave that propagates through the atmosphere between a transmitter and its receiver

25 can be described by an integral equation (e.g., Wallio and Grossi, 1972):

$$\Phi(x) = 2 \int_x^\infty \frac{n(r) r \, dr}{\sqrt{r^2 - x^2}}, \tag{1}$$

where Φ is the phase path; r is the radius from the Earth's curvature center; and, n is the refractive index. A change of variable can show that Eq. (1) is equivalent to standard form of the Abel transform (Bracewell, 1978). A geometrical interpretation of Eq. (1) is that the Abel transform of $n(r)$, right-hand side (RHS) term, is the projection of $n(r)$ onto the

30 space of the traverse coordinate, x, which indicates the nearest approach of the line of sight to the Earth's curvature centre. Being an Abel transform, the analytical inverse of Eq. (1) also exists (Ahmad and Tyler, 1998):

$$n(r) = -\frac{1}{\pi} \int_r^\infty \frac{d\Phi}{dx} \frac{dx}{\sqrt{x^2 - r^2}} \qquad (2)$$

The inverse Abel transform, Eq. (2), provides a straightforward solution to the reconstruction of the refractive index given measurements of phase path. However, the inverse transform is of limited usefulness for practical applications with real-world data because the differentiation operator in RHS exacerbates small perturbations in $\Phi(x)$ (e.g., phase noises or

5   artifacts introduced by arbitrary noise mitigations), causing the derived $n(r)$ unstable.

Another limitation of Eqs. (1)-(2) is that they are valid for a thin, spherically symmetric atmosphere in which the ray's path can be adequately approximated by the straight line that connects the transmitter and receiver. A variant of Eq. (1) suitable for dense, optically stratified media (Fjeldbo et al., 1971) is:

$$\alpha(a) = -2a \int_{r_o}^\infty \frac{d\ln(n)}{dr'} \frac{dr'}{\sqrt{(nr')^2 - a^2}}, \qquad (3)$$

10   where α is the ray's bending angle; and, the impact parameter, a, is an invariant measure (known as the Snell's constant) in the Bouger's law (the Snell's law generalized for spherical symmetry). The impact parameter is defined as: $a = nr \sin\psi = n_o r_o$, where ψ is local zenith angle of the ray and n_o is the refractive index at the tangent point r_o. With a change of variable, $x = nr$, Eq. (3) can be rewritten as:

$$\alpha(a) = -2a \int_a^\infty \frac{d\ln(n)}{dx} \frac{dx}{\sqrt{x^2 - a^2}} \qquad (4)$$

15   The corresponding inverse transform is:

$$\ln(n) = \frac{1}{\pi} \int_a^\infty \frac{\alpha(x) dx}{\sqrt{x^2 - a^2}} \qquad (5)$$

Equation (5) has the advantage of not having the differentiation operator in RHS, in addition to accounting for refracting ray paths. Consequently, Eq. (5) is the AI commonly used in GPS RO. A caution is that GPS RO does not measure the bending angle directly. Accordingly, the bending angle must be estimated in some ways in connection with the Doppler shift, which

20   again relates to the phase derivative in either time or frequency domain. Therefore, the AI using Eq. (5) is not entirely free from the differentiation operator. Instead, the procedure of retrieving the refractive index from the phase path, equivalent to Eq. (2), is split into two sequential steps: bending angle estimation and subsequent AI. The differentiation operator is put to use before or within the bending angle estimation, although it does not appear explicitly in Eq. (5).

While real-world measurements cannot be flawless, AI considers them to be complete. Consequently, Eq. (5) integrates not

25   only signal components but also the noises contained in the measured bending angle. Indeed, they are integrated in the same manner without being distinguished from each other. Hence, the quality of derived refractive index is strongly dependent on the signal-to-noise ratio (SNR) of the bending angle. For an effective use of Eq. (5), it is thus crucial to control the measurement noise properly. In practical applications, one may rely on external means of noise mitigation while preparing the bending angle for Eq. (5). The limitation is that the information about the measurement noise apprehended by the pre-

4

mitigation is simply thrown away. Hypothetically, the information could have been utilized in some ways while obtaining the refractive index.

In the presence of measurement error, the AI using Eq. (5) generally does not yield the best possible refractive index; and, corrupted measurement often causes Eq. (5) ineffective or unreliable. An alternative to AI is seeking for the inverse solution

5   of FAT numerically. However, doing so is challenging because the inverse problem of the form of Eq. (4) is generally considered to be ill-posed violating the so-called Hadamard's stability criterion (Hadamard, 1902). The standard approach to addressing the ill-posedness in inverse problems is regularization methods, described in the following.

2.2 Tikhonov regularization

Regularization methods have been used for a variety of inverse problems in different forms. Detailed description of the

10   methods is beyond the scope of this study. Here, a sketch of the Tikhonov regularization (TR hereafter) is provided in the context of GPS RO data processing. Amongst the regularization methods, TR is perhaps the most widely used method and is indeed the very method that opened up the concept of regularization. The general purpose of TR is to solve ill-posed inverse problems in which the forward operator $H(\mathbf{x}) = \mathbf{y}$ defines a mapping $H: X \to Y$, between the solution (model) space X and the data (measurement) space Y. Here, $\mathbf{x}$ is the state vector consisting of model parameters and $\mathbf{y}$ is the vector of modeled

15   observation. In our setting, the forward operator is Eq. (4), and, $\mathbf{x}$ and $\mathbf{y}$ hold the refractive index and bending angle, respectively. The method solves a minimization problem of a real-valued function J such that the minimizer, $\mathbf{x}$, is a suitable approximate solution:

$$\min_{\mathbf{x} \in X} J(\mathbf{x}) = F(\mathbf{x}) + \xi S(\mathbf{x}), \tag{6}$$

where F is the data fidelity term; S is the penalty (also called regularization) function; and, ξ is the regularization parameter.

20   The data fidelity term F measures the misfit between measured and modelled observations, whereas the penalty term ξS weighs the degree of irregularity in the solution. The regularization parameter ξ controls the relative contribution of the two terms to J . A widely used form of F is the squared L^2 norm, $F(\mathbf{x}) = \|H(\mathbf{x}) - \mathbf{y}_o\|^2$, where $\mathbf{y}_o$ contains measured bending angle. A popular choice for S aimed at noise reduction is a seminorm, $S(\mathbf{x}) = \|\nabla \mathbf{x}\|^2$, where ∇ is the gradient operator. Determining the tradeoff between the two terms, a decreasing ξ steers Eq. (6) toward (4). Therefore, the solution of TR

25   approaches AI as ξS vanishes.

Noises in measured bending angle give rise to high-frequency artifacts in the retrieved refractive index. Although the S shown above suppresses such high-frequency modes effectively, it is difficult to choose a proper ξ that can optimally control the degree of solution's smoothness. The difficulty is not different from what one may encounter designing an optimal filter. Moreover, the particular S does not account for any height dependency of the refractive-index gradient, which is very strong

30   in the real atmosphere. For instance, ∇n near the Earth's surface can be a few orders greater than that in the stratosphere. Another reason that hampers the settling on ξ is that S usually has considerable latitudinal and seasonal variations. Likewise,

the F shown above tacitly assumes that the uncertainty of measured bending angle is unvarying in space and time. In reality, the quality of RO measurement is influenced by environmental factors. For instance, the tracking of GPS signal performs less optimally under certain unfavourable atmospheric conditions. The influence of the environmental factors varies with height, latitude, and season, etc., as do the atmospheric structure and processes. The regularization parameter ξ is supposed

5   to take into account the spatiotemporal variations because neither S nor F does. However, ξ is typically a constant (i.e., scalar) and it does not have any structural information inside regarding the spatiotemporal variations.

Given that only noisy measurements are available in the real world, TR is generally advantageous over AI. For instance, TR allows one to merge noise control into the inversion as shown by the penalty function S. Nonetheless, the TR in the form described above is most useful for general observations and media whose characteristics are unknown. In order to make the

10   regularization more effective, it is desirable to maximally incorporate all significant prior information that is specific to the problem of interest. In our study, the prior information is about the character of RO measurement, and about the desired structure and behaviour of the refractive index to be retrieved. In addition, regularization methods can be costly when applied to actual RO data. It is thus crucial for practical applications to reduce the cost to a feasible level. We aim at addressing these issues through the variational regularization proposed in this study, as described below.

15   **2.3 Proposed method of variational regularization**

The variational regularization (VR hereafter) proposed in this study searches for the solution, $\mathbf{x}$, by minimizing a cost function defined as follows:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}\{\mathbf{y}_o - H(\mathbf{x})\}^{\mathrm{T}} \mathbf{R}^{-1} \{\mathbf{y}_o - H(\mathbf{x})\}, \tag{7}$$

where $\mathbf{x}_b$ is a priori of $\mathbf{x}$; $\mathbf{B}$ and $\mathbf{R}$ are error covariance matrices of $\mathbf{x}_b$ and $\mathbf{y}_o$, respectively; and, superscripts "T" and "-1"

20   indicate transpose and inverse of a matrix, respectively. Conceptually, the method seeks the optimal solution that replicates the observation as closely as possible (compelled by the second RHS term) in the vicinity of a priori state (constrained by the first RHS term). The background (first RHS) term corresponds to the penalty term of TR in the role and the observation (second RHS) term is equivalent to the data fidelity term. An obvious distinction from TR in the formulation is that VR makes use of the error covariance matrices that conveys the prior information about the uncertainty of $\mathbf{y}_o$ and $\mathbf{x}_b$. In our

25   study, $\mathbf{B}$ and $\mathbf{R}$ are built for each month on a $5° \times 5°$ latitude-longitude grid. This allows VR to account for spatiotemporal variation of the data uncertainty. Specifically, the normalization of $\mathbf{x} - \mathbf{x}_b$ and $\mathbf{y}_o - H(\mathbf{x})$ with their respective error, which is the square root of the diagonal elements of $\mathbf{B}$ and $\mathbf{R}$, eliminates the dependency of RHS terms on height, latitude, season, and so forth. In addition, the off-diagonal elements of the matrices permit VR to take into account spatial error correlations, which is essential for the background term.

30   The forward observation operator used here is Eq. (4), which maps the state variable (refractive index) onto the observation (bending angle). It must be emphasized that the state variable (control variable to be precise) of VR resides in the impact

Atmospheric
Measurement
Techniques
Open Access
Discussions

parameter space where the observation exists, as can be conjectured from Eq. (4). This differs from meteorological data assimilations the state variables of which are usually the same as those of the prediction model: temperature, moisture, pressure, wind, and so on. In the data assimilations, the location of the state-vector elements is known only in relation to the model's native grids. For the reason, in order to place measured bending angles in the model space and vice versa, it is

5 necessary to compute the model-side impact parameter, $x = nr$. The problem is that the observation-side impact parameter, $a = nr \sin \psi$, is hypothetical in that it changes with the ray's direction even at the same point in the Euclidean space. Also, x is not a definite measure of the position, relying on the never-perfect model's refractive index. Hence, the modelled impact parameter always contains some errors. Because data assimilations place measured bending angles in the model space by using the erroneous x as the coordinate, the observation's position in the model space cannot be determined precisely not

10 because of its own uncertainty but because of the model's error in the refractive index. The positional error is an extra error source, leading to a suboptimal use of the observation. This is known as errors-in-variables (EIV) problem in which the error in the independent (or coordinate) variable causes an apparent error in the dependent variable (observation).

In the proposed method, the EIV error is attributed to the uncertainty of the background refractive index because the model space coincides with the observation space. In VR, the solution's geometric height keeps changing implicitly during the

15 iterative minimization of the cost function, whereas the impact parameter assigned to the solution remains fixed. Upon the completion of the minimization, the solution's geometric height z is determined as: $z = r - R_c = an^{-1} - R_c$, where R_c is the local curvature radius of the Earth. Therefore, the solution's height in the proposed method is undecided until the solution is acquired, which is the same as in AI. The posterior height determination (PHD hereafter) reduces a substantial portion of retrieval error when viewed in the height coordinate. More details on the topic are presented in the next section.

20 In VR, the background term is based on the distance between the state vector and the background, $\mathbf{x} - \mathbf{x}_b$, meaning that the solution is expected to be around $\mathbf{x}_b$. In that sense, the quality of the background matters, although the final solution does not depend on every detail of the background. A good background furnishes faster convergence to the solution and assists in attaining the desired global minimum of the cost function. A factor of particular importance in VR is the adequacy of error covariance matrices. That is, the error covariance matrices must be factual, representing well the actual error characteristics

25 of the background and observation that are used for VR. For instance, while a climatology is usable as the background, it is not easy to estimate or define the relevant data uncertainty precisely. In that regard, short-term forecast of numerical weather prediction (NWP) models is one of the best sources of the background. The forecasts are routinely available and of a reasonably good quality. More importantly, the forecast error is largely uncorrelated with observation error. This allows one to construct realistic $\mathbf{B}$ and $\mathbf{R}$ by comparing the forecast and observation for a large number of samples. In this study, the

30 approach proposed by Hollingsworth and Lönnberg (1986) is applied to closely located (< 3 h and < 300 km) pairs of RO soundings and the forecasts of the European Centre for Medium-Range Weather Forecasts (ECMWF) interpolated to the location of RO soundings. The method separates forecast and observation errors from the variance of $\mathbf{y}_o - H(\mathbf{x})$, under the assumption that only forecast errors are spatially correlated with each other. The number of RO pairs used in this study is

about 1.8 million collected for a period of ten years (February 2006 – March 2016).

A practical difficulty in solving inverse problems of a large size is the computational cost. It is unfeasible to perturb individual elements of the control vector arbitrarily in all directions and sizes, and then search for the very combination of the perturbations that leads to the minimum of the cost function. In order to reduce the cost, the adjoint technique (Lewis and

5 Derber, 1985; Le Dimet and Talagrand, 1986) is employed in this study. The method efficiently computes the steepest gradient of the cost function with respect to all elements of the control vector at once, which is needed for the optimization algorithm. In this study, a quasi-Newton limited-memory algorithm for large-scale optimization (Zhu et al., 1997) is used. The proposed method makes use of the control-variable transform (Parrish and Derber, 1992), which reduces the conditioning number of **B**. The square root of **B** needed for the transform is computed with the method proposed by Kaiser

10 (1972).

3 Test with a synthetic sounding

It is hypothesized in this study that analytical AI is not necessarily the best way of retrieving refractive index from bending angles with error. We deliberate that meticulous utilization of the prior information via the variational regularization might be beneficial for the retrieval. Meanwhile, a common issue that arises while verifying hypotheses with real-world data is that

15 the verifying data accompany their own error, which often impedes drawing decisive conclusions on the verification. In order to overcome the difficulty, we begin the verification with a synthetic profile of bending angle. The synthetic case provides the known true solution against which the results of inversion methods can be compared with reduced ambiguity. We intend to consider large-amplitude errors so as to emphasize their influence on the solution's quality and to assess the relative robustness of the inversion methods to the erratic observations. This section is also purported as an extended

20 description of the methods but with a tangible example.

3.1 Data generation

The tracking of RO signal is often challenging, in particular in the lower tropical troposphere where sharp refractivity gradients frequently exist. As an example, a high-resolution radiosonde sounding, observed at a tropical site in Nauru (0.52° S, 166.93° E) at 12:00 UT on 3 March 2011, shows a complicated structure in the refractivity mainly due to abrupt small-

25 scale moisture variations across multiple inversion layers (Fig. 1). The station was one of the Global Climate Observing System (GCOS) Reference Upper-Air Network (GRUAN) sites until closed in August 2013. The refractive index in the neutral atmosphere can be approximated as (Smith and Weintraub, 1953): $N \equiv 10^6(n-1) = k_1 \frac{p}{T} + k_2 \frac{p_w}{T^2}$, where N is the refractivity; T is temperature in K; p is (total) pressure in hPa; p_w is water vapor pressure in hPa; and, k_1=77.6 hPa K^{-1} and k_2=3.73·10^5 K^2 hPa^{-1} are coefficients.

30 The discretization of Abel transforms causes significant numerical error when the data resolution germane to the discrete transform is poor. In order to reduce the numerical error, we used the highest data resolution available to us for the

Atmospheric
Measurement
Techniques
Discussions
Open Access
EGU

radiosonde sounding. Despite an important factor for practical applications, the numerical error is not the focus of this study. Thus, no extra effort (e.g., use of higher-order numerical schemes) is made in this regard. The sampling rate is one second in time, which corresponds to about 4 meters in height interval. The bending angle simulated with the radiosonde data and using Eq. (4) also presents rapid variations (green line in Fig. 2a). This bending angle is assumed to be absolutely accurate

5 (error-free) and is referred to as the perfect (true) observation. The measurement error of the radiosonde data carried forward into the bending angle is considered as legitimate small-scale variations in the true atmosphere.

We then generate a synthetic observation by adding suppositional measurement errors to the perfect observation. The measurement error in bending angle ε is assumed to follow a first-order autoregressive process and modeled as: $\varepsilon_k = \sigma_k \mu_k = \sigma_k \left(\rho_{k,k-1} \mu_{k-1} + \eta_k \right)$, where σ is the typical (climatic) size of the measurement error at the time and location of ε;

10 μ and η are random normal (zero-mean and unit-variance) variables; subscripts indicate the height indices of measurement samples (top-down order with increasing k); and, $\rho_{k,k-1}$ is the error correlation between the two height levels. The error correlation is modeled with the Gaussian function:

$$\rho_{k,k-1} = e^{-\frac{(a_k - a_{k-1})^2}{2L^2}}, \tag{8}$$

where a is again the impact parameter and L is the length scale of error correlation. Assuming that the vertical error

15 correlation for the bending angle is very weak, we set L to 10 m. The black line in Fig. 2a is the error-added, "measured" bending angle.

3.2 Abel inversion of low-pass filtered bending angle

Noises in measured bending angle negatively affects the quality of refractive index unless properly mitigated. It is thus customary to apply data smoothing to the bending angle prior to AI. In order to mimic the practical application of AI, we use

20 a low-pass filtered bending angle (red line in Fig. 2b) for which the fourth-order Butterworth filter (Butterworth, 1930) with a cutoff wavelength of 200 m is used. The step following is obtaining refractivity soundings through AI from the true, measured, and filtered bending angles. It must be mentioned that the resulting true refractivity (i.e., the one derived from the true bending angle via AI) differs slightly from the refractivity used to generate the true bending angle (shown in Fig. 1c). The discrepancy stems from numerical approximations made for the discrete version of Abel inversion.

25 Figure 2c compares the refractivity error, which is the difference from the true refractivity. The red line indicates the refractivity error when the filtered bending angle is used for AI. The result from "raw" bending angle (black line) is shown as the reference against which the effect of the low-pass filtering can be evaluated. A common problem with any low-pass filtering is that the degree of smoothing is hard to control. An excessive smoothing leads to a loss of observational information, whereas a minor filtering causes insufficient noise attenuation. On top of that, the noises are often non-

30 stationary, meaning that the noise spectra vary with height in accordance with vertically varying atmospheric structure. For instance, the low-pass filtering used in this study tends to reduce the refractivity error below 2 km and above 6 km, where

the true bending angle varies rather slowly. The error reduction is relative to the case of unfiltered bending angle (i.e., red line versus black in Fig. 2c). On the contrary, the filtering increases the refractivity error around the local peaks of the bending angle in 2-6 km height range. As different occultation events encounter different atmospheric conditions, it is impractical to design a customized low-pass filter for each occultation, which is adaptive to the local noise spectrum that

5   varies with the height of a given sounding. Another limitation of the sequential approach (filtering followed by AI) is one-way flow of the information in the process. That is, AI does not pass any information about the effect of attenuated noise back to the filtering. In the regularization methods on the other hand the penalty term, which acts like a filtering, invokes a reverse communication since the term is minimized iteratively and jointly with the data fidelity term. In VR, the feedback to the control vector is given through the adjoint of observation operator.

10   **3.3 The errors-in-variables problem**

The EIV problem occurs when the independent variable (position in space or time) of measurements is not known perfectly. Suppose that a particular type of radiosonde system has an offset error in the height. In that case, a flawless temperature sensor of the system is bound to produce a temperature bias, which is the height offset multiplied by the local temperature lapse rate, when monitored at a fixed height. The EIV error also emerges while comparing a measurement with its model, if

15   the independent variable of the measurement is none of the model's coordinates. This is exactly the case for RO bending angle, the location of which is defined in terms of the impact parameter. For the comparison with NWP data, for instance, it is necessary to compute the model-side impact parameter. As explained earlier in Sect. 2.3, the model's refractive index is never perfect and so the resulting impact parameter unavoidably contains errors. The impact parameter error then increases the discrepancy between measured and modeled bending angles, which is interpreted either as model error or measurement

20   error depending on the context. The issue is closely relevant to the data assimilation of RO bending angle, at least for the methods that the control variables are defined in the model space. The physical-space statistical analysis (Cohn et al., 1998) can be an exception but is not considered in this study because model-space methods are currently in prevailing use for operational weather forecasting.

The EIV error in bending angle ε_α can be estimated as: $\varepsilon_\alpha = \frac{d\alpha}{da} \cdot \varepsilon_a$; $\varepsilon_a = r\varepsilon_n$, where ε_n is model's error in the refractive

25   index. In this study, the error estimate of 12-h forecasts of the ECMWF in the refractivity (σ_N^b) at the time and location of the synthetic sounding (obtained through the procedure described in Sect. 2.3) is considered as the representative of ε_n (Fig. 3a). The corresponding ε_a is shown in Fig. 3b; the black line in Fig. 3c indicates ε_α, where the bending angle gradient is based on the true observation. For the particular synthetic sounding used in this study, ε_a is overall of comparable magnitude with the typical observation error in the bending angle (σ_α^o) (dashed red line in Fig. 3c; also described in Sect. 2.3). A point to

30   notice though is that ε_α is exceedingly lager than σ_α^o at the heights of sharp bending angle gradient. Around those heights a closer fit to the observation deteriorates the solution because not the observation but the modeled impact parameter is faulty. In model's standpoint, the EIV problem increases the observation error of bending angle. This in turn causes the observation

Atmospheric
Measurement
Techniques
Discussions

to be underutilized in the data assimilation. In contrast, the proposed variational regularization reproduces the true refractivity if perfect observations are given and the method is able to fit them exactly. That is because the proposed method holds the solution and observation in the same space. Simply speaking, the complete reconstruction of bending angle in the impact parameter space is the sufficient condition for the perfect recovery of refractivity.

5 **3.4 The proposed variational regularization**

Figure 4a shows the trace of the cost function of VR with iteration. The background term J_b (blue line) is set to zero at the start, meaning that the solution is initially identical with the background. While J_b gradually increases in the magnitude with iteration as the solution deviates more and more from the background, the observation term J_o (red line) decreases rapidly as the mapped solution approaches the observation. The iteration continues as long as the total cost function J_T (black line)

10 keeps decreasing and a norm of $\nabla_{\mathbf{x}} J_T$ exceeds a prescribed threshold. After 15 iterations all of the cost functions are nearly flat.

Figure 4b compares errors in the refractivity. The background error σ_N^b (blue line) is sizable at a number of heights. That is mainly because the vertical resolution of the 12h ECMWF forecast used here (91 levels) is insufficient to represent all of small-scale details of the true refractivity, especially the local peaks (Fig. 1c). The proposed method (red line) yields a

15 refractivity error considerably smaller than σ_N^b. The black line in Fig. 4b is the refractivity error resulting from the AI with the "measured" (rather than the smoothed) bending angle, which is the same as shown in Fig. 1c and overlaid for comparison. The proposed method is smaller than the AI in the refractivity error almost everywhere. As described earlier, the low-pass filtering shown in Fig. 2b is unable to reduce the refractivity error (Fig. 2c) and differing choices of the cutoff wavelength do not make any notable difference in the result. The pre-filtering is ineffective in the error reduction at least for

20 the particular synthetic sounding used in this study, where the true bending angle contains high frequencies that cannot be isolated from the noises. It is difficult to further reduce the remaining errors in the solution of VR since the measured bending angle is severely corrupted by the noises.

3.5 Posterior height determination

In our experience, RO refractivity tends to agree with correlative data better than the bending angle (used to derive the

25 refractivity) suggests. For instance, the comparison of RO data with independent verifying data (e.g., short-term NWP forecast or high-resolution radiosonde observation) can be made separately in bending angle and refractivity. The comparison provides $\Delta\alpha = \alpha^o - \alpha^m$, where α^o and α^m indicate the observed (i.e., RO) and modelled (i.e., the verifying data) bending angle, respectively. Likewise, the comparison in the refractivity gives $\Delta N = N^o - N^m$. Meanwhile, it is possible to closely approximate $\Delta\alpha$ and ΔN with one another. For instance, $\Delta\alpha$ can be propagated into the refractivity by:

30 $\Delta N_\alpha = \mathbf{H}^T \Delta\alpha$, where $\mathbf{H}^T$ denotes the adjoint of forward Abel transform. What is puzzling us is that ΔN is generally smaller than ΔN_α in magnitude. The same is observed in the comparison of error estimates (e.g., those described in Sect. 2.3). For

11

example, the bending angle error propagated into the refractivity by means of the Monte Carlo approach is large than the refractivity error that is estimated separately. It must be noted that in both cases mentioned above, the conversion between bending angle and refractivity is based on the linear approximation, meaning that the perturbations do not change the base state including the location. Data assimilations are similar in that the location of observation and state variables must be kept

5 unchanged during the minimization of the cost function. Otherwise, the definition of the cost function fails to remain consistent in the course of the minimization. For instance, some observations that are near the NWP model's boundaries could come in and out the model domain during the iteration, leading to an ill-defined cost function and subsequent failure of the minimization.

As described in Sect. 2.3, the control variable in VR belongs to the impact parameter space and so its location in the

10 geometric height is initially unknown. After the solution is found at the minimum of the cost function, VR determines the height based on the definition of the impact parameter at the tangent point: $r = an^{-1}$, where n the optimal refractive index (see the description of PHD in Sect. 2.3 for more details). This is identical with the way that AI determines heights of the refractivity. On the other hand, the height of refractivity in the aforementioned examples (propagation of perturbations and data assimilation) is predefined in the model space and does not change afterwards. The difference in the way that the height

15 is determined may sound trivial, but PHD has a substantive effect to the interpreted quality of the solution as illustrated in Fig. 5.

The thick solid line in Fig. 5a is the profile of true refractivity. The solution will appear somewhere on a horizontal line at the true height of a given impact parameter, deviating arbitrarily from the true refractivity. Here, we consider the horizontal line with open-headed arrows around 2.8 km and next to $N(a)$. The solution's error is indicated by the distance from the true

20 refractivity (i.e., the open circle in the middle of the line). Now suppose that the solution is smaller than the true refractivity at the given impact parameter. Because $r = an^{-1}$, PHD places the solution at a higher location in the height coordinate than the true height. Likewise, a positive refractivity error pushes down the solution in the height. As a result, the trajectory of possible solutions is slanted as shown by the dashed line next to $N(r|a)$.

An example inside the small box is shown in Fig. 5b for a detailed depiction, where the solution (denoted **P**) is smaller than

25 the true refractivity (denoted **T**) by an error $\delta N < 0$. It can be shown that the height displacement of the solution due to PHD is: $\delta r \simeq -r \delta n = -10^{-6} r \delta N$. The solution is thus placed at a higher location (r') than the true radius (r). What important here is that the solution's error in the height coordinate is to be perceived as the difference from $N(r')$ rather than from $N(a)$, because the solution's true height will never be known in practice. For the particular example shown in Fig. 5b, the solution's apparent error $\overline{|\mathbf{P'T'}|}$ is smaller than that the true error $\overline{|\mathbf{PT}|} = \delta N$, where overbar denotes the line connecting two

30 points and vertical bars indicate the line length. By linearizing $a = nr$, the slope of $\overline{|\mathbf{P'T}|}$ can be shown as: $\frac{dN(r|a)}{dr} = -\frac{10^6 n}{r} \simeq -157 \text{ km}^{-1}$, where $r|a$ stands for the conditional radius of the solution given the impact parameter.

The slope is indeed the critical refractivity gradient at which the ray's curvature radius is equal to the Earth's radius and is thus the threshold for the occurrence of atmospheric ducting. By inspection, the solution's true error is $\delta N = G_c \delta r$ and the

Atmospheric
Measurement
Techniques
Discussions

apparent error is $\delta N' = \overline{|\mathbf{P'T'}|} = (G_c - G)\delta r$, where G_c is the critical gradient and G is the local refractivity gradient. As $\frac{\delta N'}{\delta N} = 1 - \frac{G}{G_c}$, $\delta N'$ is smaller than δN in the absolute size if $2G_c < G < 0$. In general, PHD reduces the apparent error since G is known to be about -40 km^{-1} in the standard atmosphere (United States, 1946). An example for which the apparent error is larger than the true error is shown at the point denoted as **B** in Fig. 5a, where the true refractivity increases with height. It

5   must be underscored here that not only the derived refractivity but also the height assigned to it possesses error; and, the errors are negatively correlated. Therefore, the derived refractivity is also subject to the EIV problem. However, in practical circumstances (e.g., when a comparison to other data is made), the errors are attributed entirely to the refractivity and the height is assumed to be free of error.

Figure 6 illustrates the response of $\frac{\delta N'}{\delta N}$ to varying G. The black solid line in Fig. 6a is the true refractivity sounding in the

10   impact parameter coordinate and black dashed line indicates a hypothetical solution in the same coordinate, which is set to be 5 % larger than the true solution everywhere. In the example, the gradient of the true refractivity, G, ranges from -135 km^{-1} to 22 km^{-1} (Fig. 6b). The red lines in Fig. 6a are the same refractivity soundings but seen in the height coordinate. Again, the solid and dashed lines indicate the true refractivity and the solution, respectively. As shown in Fig. 6c, $\delta N'$ is smaller than 5 % (δN) except for around 3.25 km where G is positive. In particular, $\delta N'$ is significantly smaller than δN near 2.3 km

15   where $\delta N' \approx 0$ as $G \approx G_c$. The tracking of RO signals that are affected by super-refraction, which occurs when G is negatively large, is known to be more challenging. This usually leads to noisy bending angle. In the heights that $G \leq G_c$, therefore, the quality of RO refractivity is not expected to be the best. Surprisingly, Fig. 6c shows an opposing result: AI and VR have the smallest refractivity error when $G \approx G_c$. That is because PHD purges all the apparent error no matter how big the true error is. On the other hand, PHD increases the apparent error in case of strong sub-refraction ($G > 0$), which is often

20   observed at the immediate underside of local refractivity peaks. Therefore, the apparent refractivity error depends on the optical structure in the atmosphere as well as quality of the bending angle data. Syndergaard (1999) described the reduction of refractivity error resulting from PHD but without relating it to G_c.

4 Test with real data

In this section, we apply AI and VR to actual RO events and thenceforward compare the results with nearby radiosonde

25   soundings. In doing so, we use two sets of radiosonde data in order to complement each other's weakness. In the following, the data sets used here (including RO, NWP, and radiosonde) are briefly described and next the validation with respect to the radiosonde data is presented.

4.1 Data

The GPS RO data used here are made available from the Constellation Observing System for Meteorology, Ionosphere, and

30   Climate (COSMIC) mission and are processed by the Data Analysis and Archive Center (CDAAC) at the University

Corporation for Atmospheric Research (UCAR). The six COSMIC satellites have been producing 1,000-2,500 globally distributed soundings each day since the launch in April 2006 (Anthes et al., 2008). Kuo et al. (2004) and Schreiner et al. (2011) describe the CDAAC's data processing. VR uses the COSMIC neutral atmospheric bending angle as the input, as does CDAAC's AI. This is to ensure the consistency between the methods in the observation. It must be mentioned though

5   that the particular bending angle, pre-filtered by CDAAC for its AI, is not necessarily the best choice for VR. For instance, a lighter low-pass filtering can be sufficient or better. In addition, we take CDAAC's refractivity as the solution of AI, instead of carrying out an AI ourselves. This is to avoid any potential uncertainty that might arise from the difference in the practical implementation of AI. Here, we simply assume that the CDAAC's refractivity is the optimal solution obtainable via AI and from the CDAAC's bending angle. The COSMIC data used here (version 2013.3520) are available online at http://cdaac-

10   www.cosmic.ucar.edu/cdaac/products.html.

The first guess provided for VR is the operational forecasts made by the European Centre for Medium-Range Weather Forecasts (ECMWF), which are on a reduced Gaussian grid (~25 km spacing in latitude and longitude) with 91 vertical levels between the earth's surface and 80 km. Although the model's spatial resolution has been increasing with time (e.g., ~9 km and 137 levels as of March 2017), we used the fixed resolution throughout our study period for the sake of convenience.

15   We used the empirical model of the US Naval Research Laboratory (NRL), MSIS (Hedin, 1991), above the ECMWF model's top and up to 2,000 km. This is to reduce the systematic errors in high altitudes, which are notable when the upper bound of Abel transform is not high enough.

One of radiosonde data sets used in this study is the *Automatic Data Processing (ADP)* upper air observation provided by the Data Support Section (DSS) of UCAR (available online at http://rda.ucar.edu/datasets/ds337.0). The data are the global six-

20   hourly upper air reports routinely collected by the National Centers for Environmental Prediction (NCEP) for operational uses through the Global Telecommunications System (GTS). The reports consist of messages that are prepared using a set of World Meteorological Organization (WMO) alphanumerical TEMP (upper air soundings) codes, e.g., FM-35 (land stations) and FM-36 (ship-based). The codes were designed to keep the messages as short as possible whilst retaining all noteworthy features observed during the balloon's ascent. As a result, TEMP codes support a limited vertical resolution,

25   allowing reports only on the standard (mandatory) pressure levels and significant levels (if there is any).

The other set of radiosonde observation is the high vertical resolution data from the radiosonde stations operated by the National Oceanic Atmospheric Administration (NOAA), available online at ftp://ftp.ncdc.noaa.gov/pub/data/ua/rrs-data/. Beginning in 2005, NOAA began transitioning from radiotheodolite balloon tracking to GPS tracking. The data from this new system, called Radiosonde Replacement System (RRS), are recorded at 1-second (corresponds to 4-meter in height)

30   resolution, permitting good representation of small-scale atmospheric structures. A particular advantage of this data set regarding the comparison with RO data is that it provides balloon's height at every recorded moment of the flight. The operational radiosonde data (ORD hereafter) on the other hand have height reports only on the standard pressure levels. Thus, the heights on significant levels must be estimated based on the measured values of pressure (p), temperature (T), and humidity (U). The reconstructed heights are generally suitable, but at times have larger errors due to the poor data resolution

Atmospheric
Measurement
Techniques
Open Access
Discussions
EGU

as well as measurement errors in pTU. The pTU-height error is interpreted as a refractivity error when the radiosonde data are compared to RO refractivity because the comparison is to be made in the height coordinate. By using the high vertical resolution radiosonde data (HVRRD), the pTU-height error on significant levels can be avoided.

4.2 Validation with operational radiosonde data

5   ORD has a larger number of soundings and a superior geographical coverage than HVRRD the stations of which belong to a small subset of ORD sites. Hence, ORD provides more soundings that are closely located (collocated hereafter) with RO soundings for a given period of time. It also allows the collocated soundings to be sampled at various locations and under diverse atmospheric conditions. Focusing on tropical regions, we used 24,328 collocated ORD-COSMIC pairs in latitudes between 35°S and 35°N for the period from 17 February 2007 to 7 November 2010 (Fig. 7). Prior to the search of collocated

10   (< 2 h and < 200 km) pairs, ORD soundings that are dubious in quality are discarded through a series of data screening as done by Wee and Kuo (2014). Also, COSMIC soundings flagged as bad by CDAAC are dropped.

A challenging factor in GPS RO over the tropics is the sharp refractivity gradient in the lower troposphere, which could cause atmospheric ducting. The variation of phase path in ducting layers can be rapid enough to surpass the receiver's bandwidth. Also, the strong multipath effects pertinent to the anomalous wave propagation are unfavorable to the signal

15   tracking. In addition to increasing random errors, the atmospheric ducting leads to significant negative biases in the retrieved refractivity all the way down to the surface from the top of the ducting layer (Sokolovski, 2003; Ao et al., 2003). On the whole the ducting event occurs rather infrequently and is apt to happen over specific small areas. Once happens, however, it can give rise to exceptionally large errors in the derived refractivity. Consequently, the retrieved refractivity in the heights affected by the ducting tends to appear as outliers in the comparison with ORD. The difficulty is that while the influence of

20   an intense elevated ducting reaches down to the surface, the derived refractivity does not always deviate discernibly large from ORD. Reduced by PHD, the refractivity error might be small in some height below the ducting layer. In some cases, a large error of ORD can counteract the RO error. Moreover, the refractivity error can be the largest at the surface rather than in the vicinity of the ducting layer. This is because AI accumulates the RO error vertically throughout the height range below the ducting layer. All these make the outlier detection futile. The effect of the anomalous wave propagation to AI is very

25   complicated and is beyond the scope of this study.

The behavior of retrieved refractivity under the ducting layer is erratic and it makes the comparison with ORD unreliable. We therefore attempt to detect the occurrence of ducting in each RO event. Once the ducting is identified, we exclude the heights below the ducting layer from the comparison with ORD. We consider this step as a quality control. To do so, we search for ducting layers in the background refractivity ($G < -150 \text{ km}^{-1}$, slightly relaxed from G_c) starting from 7 km and

30   downward. When a ducting layer is detected for the first time, the bending angles below the layer's top are discarded and unused for VR. We do not use the refractivity retrieved via AI for the detection since it cannot have a gradient smaller than G_c by definition. Although it is the best not to use the bending angles below the ducting layer for AI, we decided to stick to the CDACC's refractivity and discard the data below the ducting layer instead. In order to reduce the dependency on the

background refractivity, the detection is repeated with the refractivity of ORD. Needless to say, both AI and VR are compared to ORD only within the height range that is common to them. In order to make the comparison with ORD robust to outliers, the lower and upper 1 % of AI soundings in the difference from ORD is discarded at each height level. The VR and background refractivity that correspond to the discarded AI are also rejected.

5   Figure 8a shows the difference from ORD in the mean refractivity. The solid black line (denoted AI) indicates the CDAAC's refractivity; the lines denoted as FCST indicate 12h (solid blue) and 24h (dashed green) ECMWF forecasts; and, the refractivity produced by VR is also shown for which the 12h (solid red) and 24h (dashed magenta) forecasts are used as the background. AI results in a distinct negative bias below 3 km, which increases with decreasing height reaching -1.5 % near the surface. This means that the above-mentioned quality control in regard to the ducting layer is not perfect although it has

10   reduced the maximum bias from -3.5 % (not shown). This might be due to the limitation of the forecasts and ORD for the purpose or there could be other RO-side sources of the systematic error. AI shows small positive deviations from ORD above 4.5 km, which are about 0.2 % at 6.5 km. On the other hand, the ECMWF's 12h and 24h forecasts in that height range deviate very little from ORD. In the lowest 2 km, the forecasts show positive systematic deviations that increase with the forecast lead time. Considering that moisture is the dominant contributor to the variability of refractivity in the height range,

15   the ECMWF's forecast model is likely to have a wet bias in the planetary boundary layer, at least with respect to ORD. VR is similar with the background above 2 km, but shifts toward AI in lower heights. In general, VR lies in between AI and FCST. This is not surprising because VR combines RO and FCST by some means, and AI might be a good indicator of RO in the systematic error. Although overly simplistic, the assumption here is that the bias of RO bending angle can be translated into refractivity bias via the observation operator and the subsequent refractivity bias is close to that shown by AI.

20   As is naturally expected, the relative weighting given to RO and FCST is though crucial for VR to reduce the bias.
In the standard deviation from ORD (Fig. 8b), AI is slightly larger than 24h forecast in the lowest 2 km and is comparable elsewhere. As expected, the random forecast error increases with the lead time. Encouragingly, VR is smaller than the background (FCST) throughout the height range regardless of the forecast lead time. This indicates that VR functions as designed attaining an error variance smaller than those of observation (represented by AI) and a priori. Nevertheless, it is

25   noticeable that AI, FCST, and VR are similar in magnitude and structure; and, the difference among them is very small, less than 0.5 % at most. This is because the bulk of the standard deviation is due to the spatiotemporal distance between COSMIC and ORD soundings, in addition to the representativeness error. Although the collocation threshold used in this study is reasonably tight, the significant horizontal inhomogeneity (especially in the moisture) over the tropical region causes the two nearby soundings to differ considerably from each other. The difference is particularly large if the small-scale

30   features in those two soundings are out of phase. The systematic difference on the other hand is insensitive to the distance.
The vertical resolution of verifying observation is also relevant to this issue. As mentioned earlier, ORD has a low resolution and the data points are distributed irregularly in height. Moreover, the depths between adjacent data levels differ substantially from one sounding to another depending on the number and location of significant levels. The significant levels are where the observed atmospheric structure turns or changes abruptly. The rapid change around the significant levels, in

conjunction with the limited data resolution, results in sizable error if one attempts to interpolate ORD to a regular height grid. Therefore, all other data (AI, FCST, and VR) are interpolated to the heights of ORD in this study. Afterwards, the data points are binned according to the pressure of each ORD height for the statistical comparison. For standard pressure levels, all data samples of the same pressure are grouped together and assigned to the pressure level. In this case, the data samples

5   refer to the refractivity values from all soundings interpolated to the height of the pressure level of collocated ORD. For significant levels on the other hand, bins are allocated in the middle of adjacent standard pressure levels, accommodating all data samples whose pressure are between the neighboring standard levels. For instance, all significant levels between 700 hPa and 850 hPa belong to the same bin. For the reason, the bins have a saw-toothed distribution of data counts as shown in Fig. 8c. The difficulty here is that the samples in a bin are different in their height. Given that the refractivity varies

10   exponentially with height to a good degree, the height discrepancy increases the samples' spread (standard deviation), which could have been reduced greatly if all the samples were taken at the same height. Namely, a portion of the standard deviation shown in Fig. 8b is attributable to the vertical variation of the true refractivity rather than to actual error. A correction accounting for the inter-sample height difference might be possible, but no attempt in the direction is made in this study.

4.3 Validation with high vertical resolution radisosonde data

15   It is possible to overcome the difficulties due to the limited vertical resolution of ORD by using HVRRD. In this comparison, all data (HVRRD, AI, FCST, and VR) are interpolated to a regular height grid. Thanks to the high vertical resolution of HVRRD, the interpolation of HVRRD does not produce large error unless the reported heights are corrupted. Besides, by placing all data samples to the same height, the uncertainty due to the inter-sample height discrepancy is eliminated. Figure 9 shows the location of 92 HVRRD stations used in this study. The stations are classified into three

20   regions: the tropics (15 sites, red filled circle), US (64, green open diamond), and Arctic (Alaska) (13, blue filled square). With the same distance threshold used for ORD, 30,796 collocated COSMIC-HVRRD pairs (2,602 in the tropics; 25,128 US; and, 3,064 Arctic) are found for the period from 17 February 2007 to 31 August 2015 (insets in Figs. 10a-c). A longer time period than the comparison with ORD is needed here since the number of HVRRD stations is small. For COSMIC data, two slightly different data versions are used: 2013.3520 until 30 April 2014 and 2014.2860 afterwards.

25   As mentioned above, all data including HVRRD are interpolated to a regular grid of 50-meter interval and thus the comparison does not need the data binning used for ORD. Without the blurring due to the binning, the comparison yields a more detailed vertical structure of the statistics. Another distinction from the comparison with ORD is that the data rejection below the top of the highest ducting layer detected from radiosonde refractivity is omitted here. The reason is that the differentiation between very shallow layers (due to the high resolution) intensifies measurement noises and so the vertical

30   refractivity gradient computed with HVRRD oscillates rapidly. Unless a heavy pre-filtering is applied, the refractivity gradient is unreliable to be used for the ducting-layer detection. In addition to the strength of refractivity gradient, the effect of ducting also depends on the depth of the ducting layer. That is, some paper-thin ducting layers can be ignored. The

radiosonde-based detection of ducting layer is omitted because the choice of pre-filtering, numerical differentiation, and threshold depth is difficult to be made properly and is arbitrary at best. The background-based data detection is still applied.

Figures 10a-c show the systematic differences from HVRRD in the three regions. In the tropics, the overall feature of AI is similar with that shown by the comparison with ORD (Fig. 8a) such as large negative bias near the surface and moderate
5   positive bias in 6-10 km range. Without the spread due to the data binning, the comparison with HVRRD shows larger maxima of the bias in magnitude. The 24h ECMWF forecast (denoted as FCST) is biased negatively below 5 km and positively in 5-10 km range, which is quite different from the comparison with ORD. In addition, VR is analogous to AI deviating significantly from FCST above 2 km, whereas it is close to FCST in the comparison with ORD. After looking into the difference between the comparisons with ORD and HVRRD, it is found that VR gives a heavier weighting to RO
10   observations on average at the HVRRD stations than at the tropical ORD stations. At the tropical HVRRD stations, RO observations are assumed to be more accurate, while the background soundings are less reliable. The difference in the weighting relates to the geological coverage. The HVRRD stations are smaller in the number and located in small specific areas. For instance, about 10 out of 15 HVRRD stations are in latitudes lower than 20°, whereas the majority of ORD sites are located in the subtropics (Fig. 7).
15   Compared to AI, VR shows a smaller bias in the lowest 2 km and in the stratospheric heights. The error reduction is pronounced in the low tropical troposphere (Fig. 10a) and Arctic stratosphere (Fig. 10c). In general, however, the difference between AI and VR in the bias is small except for in the lowest 2 km. Although all of AI, FCST, and VR deviate significantly from HVRRD, they show similar patterns from mid-troposphere to 20 km in all "latitudinal" regions. Especially, AI and FCST are very close to each other over the Arctic even though they are largely independent. In principle,
20   AI and FCST can be weakly correlated in the error since the ECMWF assimilates COSMIC data while producing the initial condition of the forecast. If it is the case, however, HVRRD will exert a greater influence on FCST because the radiosonde data, which are also assimilated, are routinely available at the same location (of FCST) at every 12 hours, whereas COSMIC data are obtainable intermittently at random locations. The nearness between AI and FCST thus indicates that their difference from HVRRD is mostly caused by the systematic error of HVRRD. Despite the uncertainty of HVRRD, VR is
25   generally smaller than AI in the systematic difference from HVRRD.

The difference among AI, FCST, and VR is more pronounced in the standard deviation (Figs. 10d-f). In the comparison with ORD (Fig. 8b), AI is no better than (24h) FCST especially in the lowest 2 km. In the standard deviation from HVRRD, On the other hand, AI is smaller than FCST in the tropospheric heights of all regions. In the stratosphere above 13 km over the US and Alaska, however, AI is slightly larger than FCST. It is found through the error estimation described in Sect. 2.3 that
30   the stratospheric degradation of AI is due to a rapid, unsmooth transition of bending angle estimation, which is from the geometrical optics method applied above 20 km to a wave optics method below. Nonetheless, VR agrees better with HVRRD than any of FCST and AI does in all heights regardless of the geographical area. As a result of the high vertical resolution, HVRRD has small-scale vertical variations in the refractivity, which might not be easy to be observed

concurrently by other observing systems. Therefore, the evident error reduction attained by VR in both measures (i.e., systematic and random errors) is very impressive.

5 Summary and concluding remarks

The refractivity soundings provided by GPS RO have been used widely for weather and climate research. Typically, the
5 refractivity is obtained from the inverse Abel transform (Abel inversion) of measured bending angle (measurement). The foremost problem of Abel inversion (AI) is that the error contained in the measurement is carried forward into the refractivity without being controlled. The measurement error includes artifacts introduced by arbitrary noise mitigations that are applied prior to AI. It is challenging to improve the noise mitigation because the separation between signal component and noise is never easy. In order to address the issue, we seek for an approach that accounts for the measurement error while
10 obtaining the refractivity. Our approach is motivated by two facts: Firstly, model has been used to interpret measurements and to evaluate their fidelity in many geophysical applications; and secondly, the error characteristics of RO measurement and the model are fairly well understood at least statistically. After considerable deliberation we come up with an idea that it is possible and synergetic to combine noise attenuation and the inversion together into an estimation problem. Based on the hypothesis, we have proposed a new method for determining atmospheric refractivity from RO bending angle through a
15 variational regularization (VR) of the forward Abel transform.

The proposed method considers the numerical inversion of the forward Abel transform. Instead of solving the inverse problem directly, however, VR turns it into an optimization problem in which the solution (refractivity) is required to describe the measurement as closely as possible while the solution's behavior is regularized as per priori knowledge. The optimization problem is solved via the adjoint technique, which is a very efficient way of calculating the gradients of the
20 cost function with respect to all control parameters at once. The fundamental difference between AI and VR is that while the former considers the measurement to be complete, the latter seeks for a filtered solution for which the fitting to the measurement is used as a weak constraint. The proposed method is a generalization of classical regularization methods. The essential feature of VR in the formulation that differentiates the method from classical regularizations is the use of error covariance matrices, which resolves the issues challenging in the classical methods, e.g., as to determining the regularization
25 parameter and as to dealing with the spatiotemporal variation of system parameters. The covariance matrix of measurement error and that of background (a priori estimate of the solution) error play crucial roles in the regularization. The background error covariance matrix, for instance, spreads information from each measurement sample to neighboring locations through the spatial error correlation. As a result, VR cherishes the observational information that is coherent across multiple data samples and prohibits the solution from being too sensitive to the nearest sample. The filtering effect of the error correlation
30 thus encourages VR to produce a smooth solution. It must be emphasized that the smoothness differs from that achievable through low-pass filtering because the solution must attain, in addition to being smooth, the statistical optimality delineated by the error covariance matrices. The optimality accounts for the spatiotemporal variation of the measurement error. Thus,

the weightings given to individual measurement samples differ. AI on the other hand treats all measurement samples equally in their trustworthiness.

We have tested the robustness of the proposed method using a synthetic sounding with error. The known true solution in the setting resolves a long-standing problem in real-data tests, which is the ambiguity in the validation stemming from the

5 uncertainty of verifying data. We have demonstrated the limitation of AI with the focus on the difficulty in the pre-filtering of measured bending angle. It is shown that the filtering is particularly challenging for large-amplitude, non-stationary measurement errors. We have described an important source of measurement error in case of model-based inversion or data assimilation, which is the errors-in-variables (EIV) problem. The EIV error is induced by model's error in the coordinate variable (i.e., impact parameter). It is shown that at the heights that the bending angle varies rapidly, the size of EIV error

10 can be exceedingly larger than the typical magnitude of measurement error. We have explained that while the EIV error makes the data assimilation of bending angle less effective, it can be reduced by placing the control variable of the optimization problem in the impact parameter space. Another point highlighted in the test is the posterior height determination (PHD). In both AI and VR, the geometric height of refractivity is determined according to the definition of the impact parameter but only after the refractivity is decided in the impact parameter space. We have described in detail with

15 examples and illustrations that PHD can reduce the refractivity error substantially. We have utilized the synthetic case in order to articulate why we had decided to define the control variable of VR in the observation space, which is to reduce EIV error and to take advantage of PHD.

We have applied the proposed method to actual COSMIC events and validated the result by comparing it with nearby radiosonde soundings. We used two radiosonde data sets, the operationally collected TEMP-format global data and the high

20 vertical resolution data collected at stations operated by NOAA, in order to complement each other's weakness. The former is lower in the resolution but has a superior geographical coverage, and vice versa. The validation shows that VR is suggestively smaller than AI and the background (ECMWF forecasts), in the standard deviation of refractivity from the radiosonde data. Both radiosonde data sets equally show the smaller standard deviation of VR in all heights and latitudes. VR also agrees better with the radiosonde data than AI does in the mean of refractivity, especially in the lowest 2 km. We

25 have seen in some heights that VR is slightly larger than the background in the mean difference. Although not certain for sure, the likely cause is the systematic error of the radiosonde data.

Based on the results presented herein it is concluded that VR is a considerable improvement over AI in the quality of refractivity. This suggests that VR is able to enhance the data value of RO bending angle with the aid of prior information. Our study has an important implication for the data assimilation of GPS RO data. Recent years, most of global NWP centers

30 prefer bending angle to refractivity for data assimilation. Although there are good supporting reasons for the preference, this study finds that the assimilation of bending angle has drawbacks that are often disregarded in previous studies. It appears that VR is very promising as it reduces the EIV error and benefits from the PHD. It is straightforward to assimilate the refractivity produced by VR. In order to take the full advantage of VR and to ensure the consistency with the background, however, it will be desirable to incorporate the regularization into data assimilation methods. The recent version of COSMIC

Atmospheric
Measurement
Techniques
Discussions

one-dimensional variational retrieval method (1D-Var), for instance, conducts VR on the fly prior to the actual retrieval. Alternatively, stand-alone VR that shares the background with data assimilation can be carried out as a RO data pre-processing so as to reduce the computational complexity.

Acknowledgements

References

Abel, N. H.: Auflösung einer mechanischen Aufgabe, J. Reine. Angew. Math., 1, 153-157, doi: 10.1515/crll.1826.1.153, 10   1826.

Ahmad, B. and Tyler, G. L.: The two-dimensional resolution kernel associated with retrieval of ionospheric and atmospheric refractivity profiles by Abelian inversion of radio occultation phase data, Radio Sci., 33(1), 129-142, 1998.

Anthes, R. A., Ector, D., Hunt, D. C., Kuo, Y.-H., Rocken, C., Schreiner, W. S., Sokolovskiy, S. V., Syndergaard, S., Wee, T.-K., Zeng, Z., Bernhardt, P. A., Dymond, K. F., Chen, Y., Liu, H., Manning, K., Randel, W. J., Trenberth, K. E., Cucurull, 15   L., Healy, S. B., Ho, S.-P., McCormick, C., Meehan, T. K., Thompson, D. C., and Yen, N. L.: The COSMIC/FORMOSAT-3 mission: early results, B. Am. Meteorol. Soc., 89, 313–333, doi:10.1175/BAMS-89-3-313, 2008.

Ao, C. O., Meehan, T. K., Hajj, G. A., Mannucci, A. J., and Beyerle, G.: Lower-troposphere refractivity bias in GPS occultation retrievals, J. Geophys. Res., 108, 4577, doi:10.1029/2002JD003216, 2003.

Bracewell, R. N.: The Fourier transform and its applications, The Fourier transform and its applications, McGraw-Hill, New 20   York, 444 pp., 1978.

Butterworth, S.: On the theory of filter amplifiers, Wireless Engineer, 7, 536-541, 1930.

Cohn, S. E., Da Silva, A., Guo, J., Sienkiewicz, M., and Lamich., D.: Assessing the effects of data selection with the DAO physical-space statistical analysis system, Mon. Wea. Rev., 126, 2913-2926, 1998.

Fjeldbo, G., Kliore, G. A., and Eshleman, V. R.: The neutral atmosphere of Venus as studied with the Mariner V radio 25   occultation experiments, Astron. J., 76, 123-140, 1971.

Hadamard, J.: Sur les problèmes aux dérivées partielles et leur signification physique, Princeton Univ. Bull., 13, 49-52, 1902.

Hedin, A. E.: Extension of the MSIS thermosphere model into the middle and lower atmosphere, J. Geophys. Res., 96(A2), 1159-1172, 1991.

Hollingsworth, A. and Lönnberg, P.: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, Tellus, 38(A), 111-136, 1986.

Kaiser, H. F.: The JK method: a procedure for finding the eigenvectors and eigenvalues of a real symmetric matrix, Comput. J., 15(3), 271-273, 1972.

5 Kuo, Y.-H., Wee, T.-K., Sokolovskiy, S., Rocken, C., Schreiner, W., Hunt, D., and Anthes, R. A.: Inversion and error estimation of GPS radio occultation data, J. Meteorol. Soc. Jpn., 82, 507–531, 2004.

Le Dimet, F. X. and Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological observations, Tellus, 38(A), 97-110, 1986.

Lewis, J. M. and Derber, J. C.: The use of adjoint equations to solve a variational adjustment problem with advective

10 constraints, Tellus, 37(A), 309-322. doi:10.1111/j.1600-0870.1985.tb00430.x, 1985.

Parrish, D. F. and Derber, J. C.: The National Meteorological Center's spectral statistical-interpolation analysis system, *Mon. Wea. Rev.*, 120, 1747-1763, doi:10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2, 1992.

Schreiner, W., Sokolovskiy, S., Hunt, D., Rocken, C., and Kuo, Y.-H.: Analysis of GPS radio occultation data from the FORMOSAT-3/COSMIC and Metop/GRAS missions at CDAAC, Atmos. Meas. Tech., 4, 2255–2272, doi:10.5194/amt-4-

15 2255-2011, 2011.

Schuster, T., Kaltenbacher, B., Hofmann, B., and Kazimierski, K. S.: Regularization methods in Banach spaces, Walter de Gruyter, Berlin, 294 pp., 2012.

Smith, E. K. and Weintraub, S.: The constants in the equation of atmospheric refractive index at radio frequencies, Proc. IRE, 41, 1035-1037, 1953.

20 Sokolovskiy, S.: Effect of superrefraction on inversions of radio occultation signals in the lower troposphere, Radio Sci., 38(3), 1058, doi:10.1029/2002RS002728, 2003.

Tikhonov, A. N.: Solution of incorrecly formulated problems and the regularization method, Soviet Math. Dokl., 4(4), 1035-1038, 1963.

Syndergaard, S.: Retrieval analysis and methodologies in atmospheric limb sounding using the GNSS occultation technique,

25 DMI Scientific Report 99-6, Danish Meteorological Institute, Copenhagen, 86 pp., 1999.

United States (Office of Scientific Research and Development, National Defense Research Committee, and Committee on Propagation): The propagation of radio waves through the standard atmosphere, Columbia Univ. Press, Washington, D. C., 206 pp., available at: https://archive.org/details/propagationofrad03unit/, 1946.

Wallio, H. A. and Grossi, M. D.: A comparison of techniques for inversion of radio-ray phase data in presence of ray

30 bending, NASA Technical Note D-7098, Hampton, VA, 34 pp., 1972.

Wee, T.-K. and Kuo, Y.-H.: Advanced stratospheric data processing of radio occultation with a variational combination for multifrequency GNSS signals, J. Geophys. Res., 119, 11,011-11,039, doi:10.1002/2014JD022204, 2014.

Zhu, C., Byrd, R. H., and Nocedal, J.: L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization, ACM Trans. Math. Softw., 23(4), 550-560, 1997.
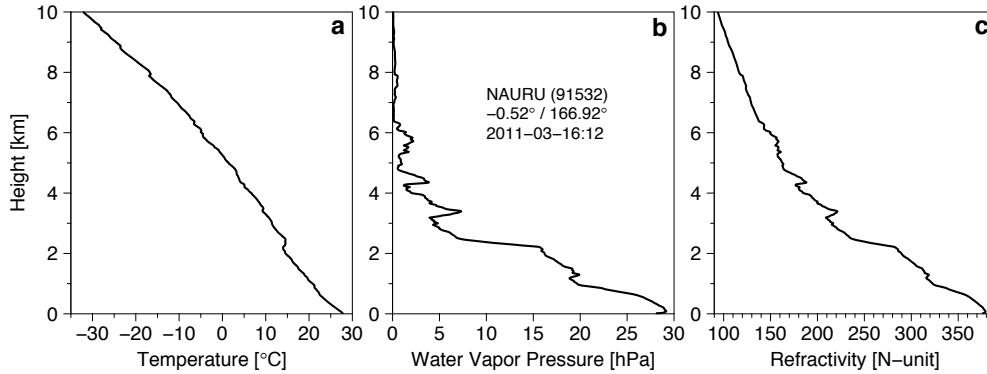
Atmospheric
Measurement
Techniques
Discussions

Figure 1. High-resolution radiosonde sounding at a tropical site in Nauru (0.52° S, 166.93° E) at 12:00 UT on 3 March 2011: (a) temperature (° C); (b) water vapor pressure (hPa); and, (c) refractivity (N-unit).
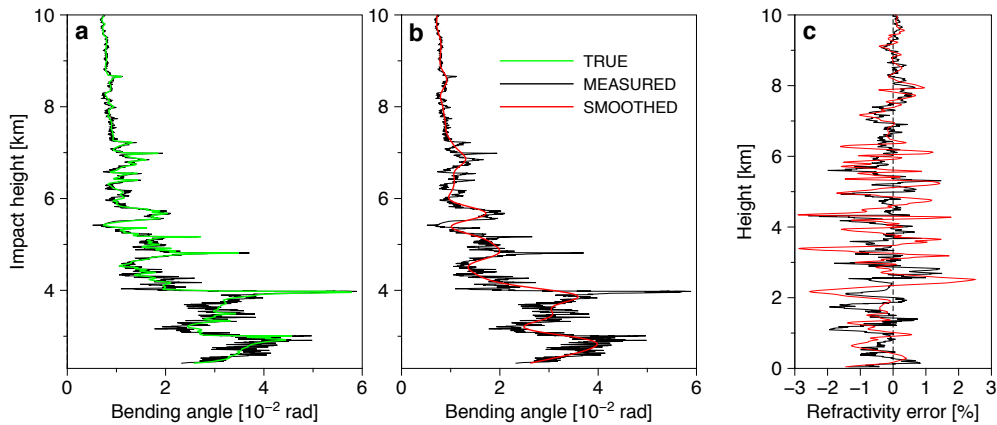

Figure 2. Simulated bending angles and their errors propagated into the refractivity via Abel inversion. (a-b) Green line (denoted TRUE) indicates the bending angle modeled with the exact refractivity profile shown in Fig. 1c; black line (MEASURED) represents the "measured" bending angle for which assumed measurement errors are added to the perfect bending angle; red line (SMOOTHED) indicates a low-pass filtered version of the MEASURED. Note that the black line in (b) is the same as in (a) and is duplicated for comparison. (c) Errors in the refractivity; the refractivity profiles are obtained from the bending angle of matching color via Abel inversion and the error is defined as the difference from the perfect refractivity, which is the one derived from the TRUE bending angle. See text for more details.
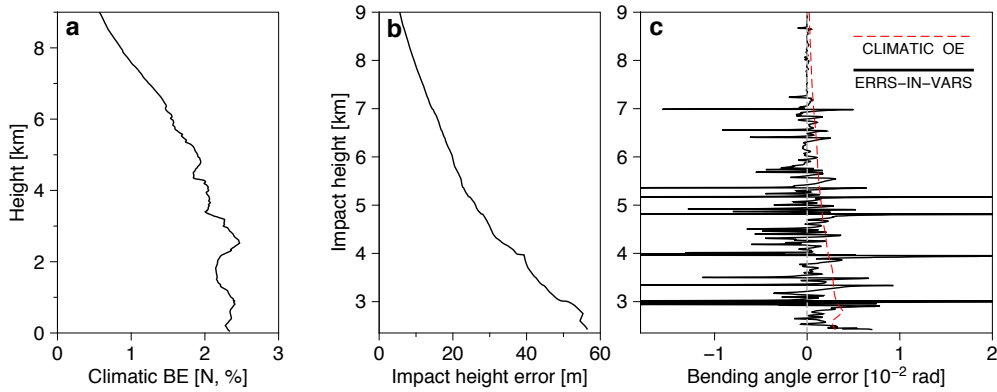
Figure 3. Errors involved in the test with a synthetic sounding: (a) statistical error estimate of a priori (12h forecast) refractivity at the time and location of the sounding; (b) the impact height error resulting from a priori refractivity error; (c) statistical error estimate of bending angle observation (dashed red line) and the bending angle error due to the errors-in-variables problem (solid black), which is the modeling error of the synthetic bending angle stemming from a priori refractivity error.
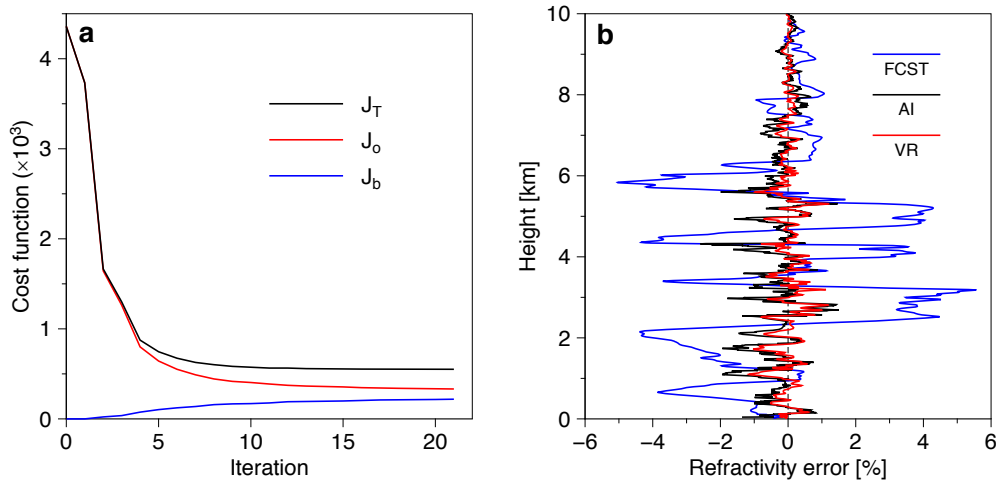
Figure 4. (a) Change of the cost function with iteration in the variational regularization: observational term (denoted J_o, red line); background term (J_b, blue); and, the total cost function ($J_T = J_o + J_{ob}$, black). (b) Refractivity error of the background (blue), of the Abel inversion (black), and of the variational regularization (red).

Atmospheric
Measurement
Techniques
Open Access

Discussions
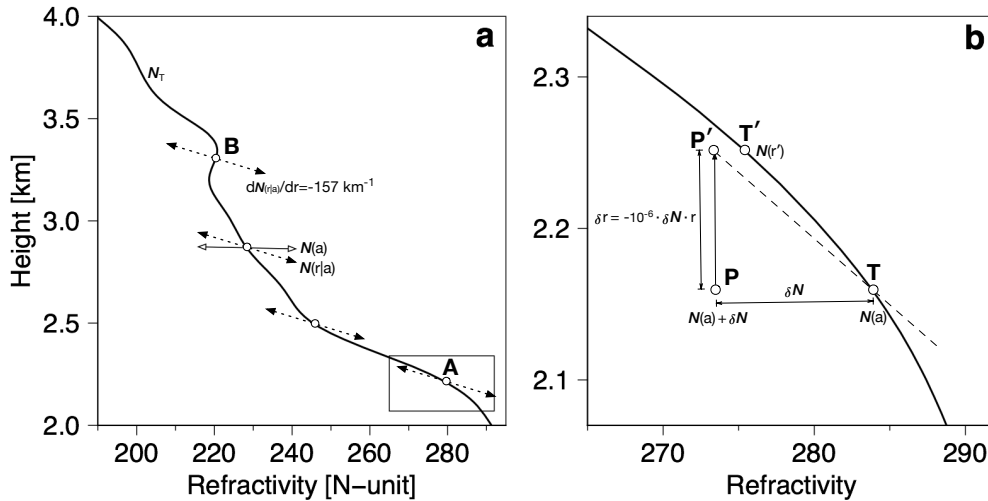
Figure 5: Illustration of solution's trajectory in the posterior height determination (PHD). The heavy solid curve in (a) is the true refractivity sounding. The solid line with open-headed arrows around 2.8 km, denoted $N(a)$, indicates the actual track of potential solutions for a given impact parameter (or true height). PHD relocates the solution, making the track to be interpreted as the dashed, slanted line next to $N(r|a)$. (b) shows an example of the relocation in the rectangular inset. In the example, the solution P is smaller than the true refractivity T; therefore, P is relocated to P' from the true radius r to a higher location r' since $r = an^{-1}$. This leads the apparent error $\overline{P'T'}$ to be smaller than the true error $\overline{PT}$. The slope of $\overline{P'T}$ is constant, -157 km^{-1}, regardless of the true height or the size of true error as shown by the dashed lines in (a). See text for more details.
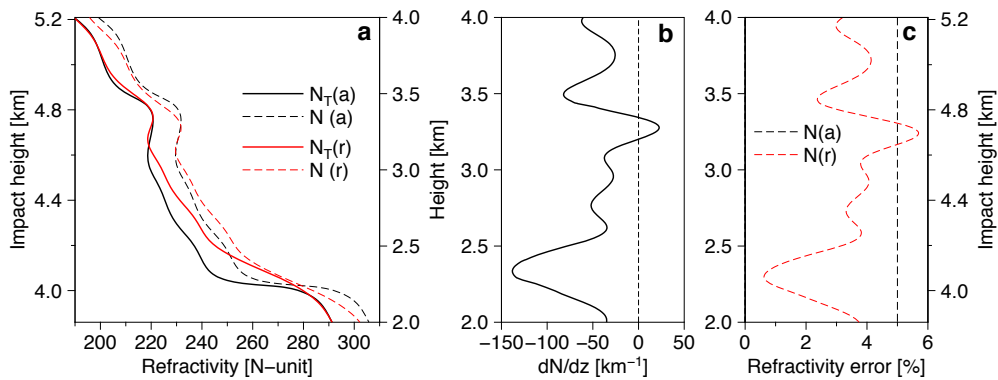
Figure 6: An illustrative example of true error versus apparent error. (a) A hypothetical solution (black dashed line), which is 5 % larger than the true refractivity (black solid) everywhere. Both reside in the impact parameter coordinate. The solution (red dashed) and true (red solid) refractivity in the height coordinate. (b) The vertical gradient of the true refractivity. (c) The true error (black dashed) and apparent error (red dashed), which are defined as the departure from the true refractivity in the same coordinate.

Atmospheric
Measurement
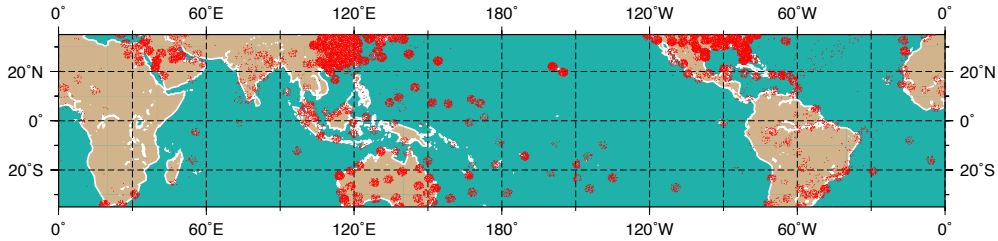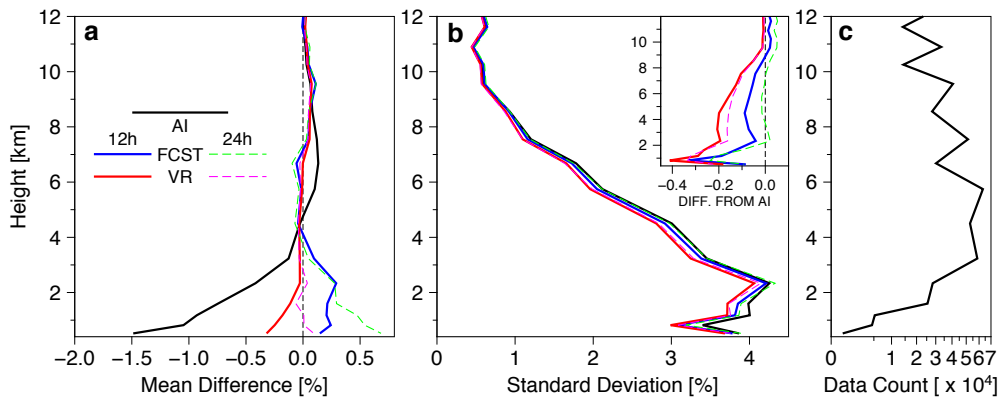Techniques
Open Access
Discussions

Figure 7: Location of COSMIC soundings (red dots) collocated with the operational radiosonde soundings in latitudes between 35° S and 35° N for the period from 17 February 2007 to 7 November 2010.

5 **Figure 8: Comparison with collocated operational radiosonde data (ORD). (a) Mean difference (from ORD) of: CDAAC's refractivity derived via Abel inversion (heavy solid black line, denoted AI); ECMWF's 12h forecast (blue solid) and 24h forecast (green dashed); and, VR-produced refractivity for which 12h (solid red) or 24h (dashed magenta) forecast is used as the first-guess. The unit used here is %, i.e., the percentage with respect to the mean of ORD. (b) Same as in (a) except for standard deviation; the inset shows the difference from AI in the standard deviation from ORD. (c) Number of samples used in the**
10 **comparison.**

Atmospheric
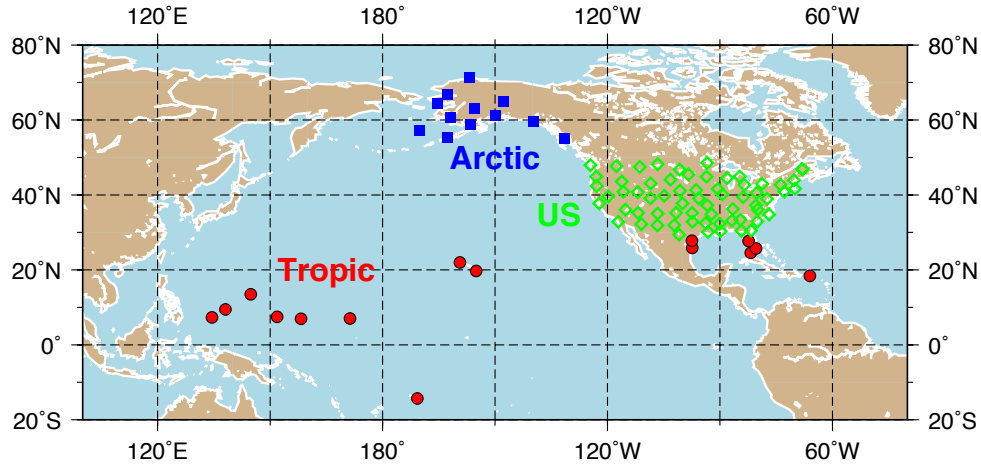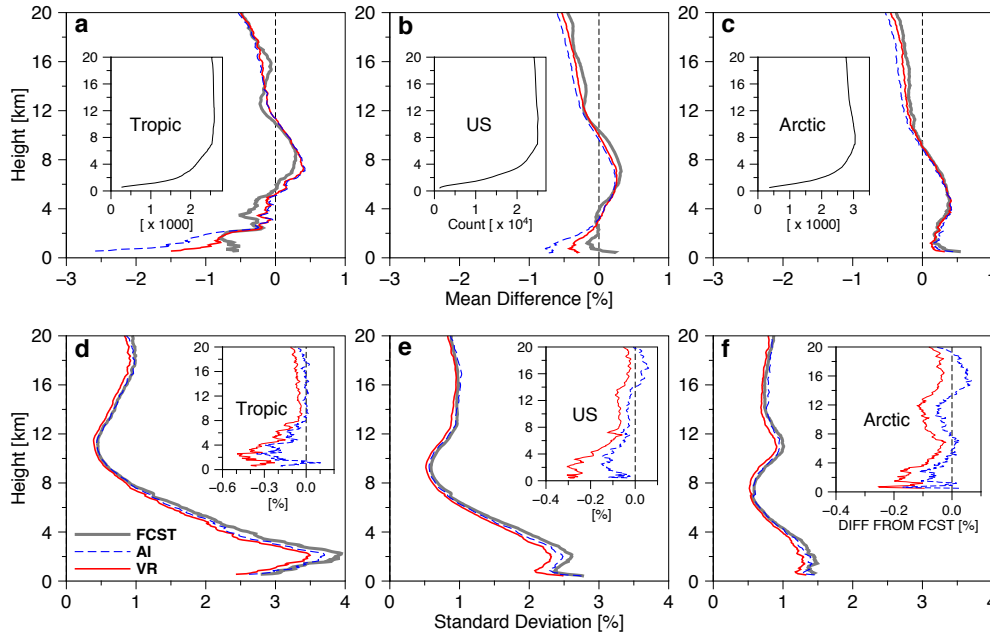Measurement
Techniques

Open Access

Discussions

EGU

Figure 9: Location of US-owned HVRRD stations used in this study. The stations are classified into three latitudinal regions: the tropics (15 sites, red filled circle); US (64, green open diamond); and, Arctic (Alaska) (13, blue filled square).

5 **Figure 10: Comparison with collocated HVRRD. (a-c) Mean difference (%) (from HVRRD) of: CDAAC's refractivity derived via Abel inversion (blue dashed line, denoted AI); ECMWF's 24h forecast (heavy solid grey, FCST); and, VR-produced refractivity (solid red) for which the FCST is used as the first-guess. Shown are the statistics over (a) the tropics, (b) US, and (c) Alaska. The inset shows the vertical data count. (d-f) The same as in (a-c) except for the standard deviation. The inset shows the difference from FCST in the standard deviation from HVRRD.**