Sensitivity of Forward-Modeled Bending Angles to Vertical Interpolation of Refractivity for Radio Occultation Data Assimilation

SHAY GILPIN, RICHARD ANTHES, AND SERGEY SOKOLOVSKIY

University Corporation for Atmospheric Research, Boulder, Colorado

(Manuscript received 22 June 2018, in final form 25 October 2018)

ABSTRACT

Assimilation of radio occultation (RO) observations into numerical weather prediction (NWP) models has improved forecasts, where RO is typically one of the top five observational systems contributing to forecast accuracy. By measuring the phase delay of radio waves traversing Earth’s atmosphere between global positioning system (GPS) and low-Earth orbiting satellites, RO obtains quasi-vertical profiles of bending angles (BA) of the radio waves’ trajectories. BA are the RO observation most often assimilated into NWP models, but since they are not computed or analyzed in the models, they must be computed from model data using a forward model. First, model refractivity $N$ is computed from variables specified on the model’s vertical levels, then using the Abel integral, BA are computed from $N$. The forward model requires vertical differentiation of $N$, and accurate differentiation requires vertical interpolation of $N$ between model levels. The interpolation results in errors, which then propagate through the forward model to produce BA errors. In this study, we investigate the sensitivity of forward-modeled BA to five different methods of vertical interpolation of $N$ between model levels to determine a method that minimizes interpolation errors. We use RO-observed $N$ to isolate the interpolation errors and determine an accurate method that can be applied to any NWP model. Of the five methods investigated, the log-spline interpolation reduces $N$ and BA errors the most, regardless of the vertical resolution of the model grid.

1. Introduction

Radio occultation (RO) provides measurements of Earth’s atmosphere with characteristics that are useful for assimilation into numerical weather prediction (NWP) models. By measuring the phase delay of radio waves traversing quasi-horizontally through Earth’s atmosphere between global positioning system (GPS) and low-Earth orbiting (LEO) satellites, quasi-vertical profiles of bending angles (BA) and refractivity $N$ can be derived with little to no instrument bias, meaning no necessary bias correction upon assimilation (Healy and Thépaut 2006). Because of the global coverage obtained from the satellites, RO provides observations in locations where other measurements are sparse, resulting in forecast improvements globally, especially in undersampled regions such as the Southern Hemisphere (Anthes et al. 2008; Cucurull and Derber 2008).

Since the first discussion of RO assimilation by Eyre (1994), several NWP centers now assimilate RO observations into their operational models. Either BA or $N$ profiles can be assimilated; Eyre (2008) details their advantages and disadvantages. Generally, BA are the most common RO observations assimilated into NWP systems. Many studies, (e.g., Zou et al. 2004; Healy and Thépaut 2006; Aparicio and Deblonde 2008; Cucurull and Derber 2008; Healy 2008; Cucurull 2010; Poli et al. 2010; Anlauf et al. 2011) have shown improvements in forecast accuracy upon assimilation of RO. Though RO observations constitute only a small amount of the assimilated data in NWP models, RO is typically one of the top five observational systems contributing to forecast accuracy (Healy et al. 2012).

Because models do not predict $N$ or BA directly, a forward model, or observation operator, must be used to
compute \( N \) or BA from the model variables temperature, pressure, and water vapor. Janjić et al. (2018) discuss forward models and their errors in the context of representativeness errors in NWP data assimilation. The model \( N \) or BA are then compared to the observed \( N \) or BA, and the model temperature and water vapor are adjusted to decrease the difference between the model and observed \( N \) or BA. The amount of adjustment (analysis increment) is based on the magnitude of this difference and the specified (assumed) error variance of the model and observations. Errors associated with the forward model contribute to the total observation error and the difference between model and observed \( N \) or BA. The forward model errors, in turn, include vertical interpolation errors as discussed below. Thus, it is important to know what the vertical interpolation errors are and to minimize them to the extent possible.

There are other approaches to RO data assimilation in addition to those that assimilate BA or \( N \). RO observations can be assimilated using either local (typically 1D) or nonlocal (typically 2D or 3D) forward operators, both of which have advantages and disadvantages. Local operators tend to be simpler but can lose accuracy by omitting horizontal gradients in \( N \). Nonlocal operators capture horizontal gradients in \( N \) using ray-tracing techniques but require NWP information at multiple locations and tend to be more computationally expensive (Ma et al. 2009; Healy 2014). An overview of different data assimilation approaches can be found in Kuo et al. (2000) and Syndergaard et al. (2006). The total forward model errors have several different contributions depending on the choice of data, methods of assimilation, and the model itself. The topic of this study is the vertical interpolation errors associated with the forward modeling of BA for RO assimilation, just one contributor to the total forward model errors.

During BA assimilation, models are required to compute BA from model variables to compare with the observed BA. First, \( N \) is computed from model temperature (\( T; K \)), pressure (\( p; \) hPa), and water vapor pressure (\( e; \) hPa), which in a neutral atmosphere is given by the following relationship (Smith and Weintraub 1953):

\[
N = 77.6 \frac{p}{T} + 3.73 \times 10^{5} \frac{e}{T^2}. \tag{1}
\]

Refractivity \( N \) is converted into the refractive index \( n = 1 + N \times 10^{-6} \), which, under the assumption of spherical symmetry, is used to compute BA as a function of impact parameter \( \alpha(a) \) using the Abel integral (Fjeldbo et al. 1971; Melbourne et al. 1994; Kursinski et al. 1997):

\[
\alpha(a) = -2 \int_a^\infty \frac{d \ln(n)/dx}{\sqrt{x^2 - a^2}} dx, \tag{2}
\]

where \( x = nr \), also known as the refractional radius, and \( r \) is the radius counted from the local center of curvature of the geoid.

Computing BA from model \( N \) using the Abel integral [Eq. (2)] requires vertical differentiation of the refractive index \( n \), where accurate differentiation requires vertical interpolation of \( N \) between model levels. Accurate interpolation is necessary because \( N \) is specified on a relatively low-vertical-resolution model grid (compared to the resolution needed to accurately approximate the Abel integral). At the relatively coarse resolutions of current models, small-scale atmospheric structures (which are not oversampled) near the model grid size, as well as the quasi-exponential variation of \( N \) with height, can create unacceptably large errors in the forward-modeled BA without accurate interpolation.

Different NWP centers use different methods of interpolation: the NOAA National Centers for Environmental Prediction (NCEP), for example, uses cubic interpolation obtained by blending two adjacent quadratic Lagrangian polynomials, thus ensuring continuity of the first derivatives (Cucurull et al. 2013). The European Centre for Medium-Range Weather Forecasts (ECMWF) initially applied log-linear interpolation (Healy and Thépaut 2006). However, upon noticing biases resulting from this interpolation, Burrows et al. (2014) suggested using an improved hybrid log-linear–polynomial interpolation, which reduced the forward-modeled BA biases. Anlauf et al. (2011) describe combining spline interpolation below 30 km with log-linear interpolation above 30 km for BA assimilation in the Deutscher Wetterdienst (DWD) operational global model.

In this study, we investigate the sensitivity of forward-modeled BA to vertical interpolation of \( N \) between model levels to determine a method that minimizes BA errors associated with the interpolation. To isolate the errors caused by vertical interpolation, we use RO-observed \( N \) to represent \( N \) between model levels as the “truth”; we do not perform data assimilation experiments. The goal is to determine an interpolation method that is optimal for both \( N \) and BA through the full vertical assimilation region (surface to tops of at most 60 km) and can be applied to any model regardless of the model’s vertical resolution. We test five different interpolation methods: linear, log-linear, log-cubic, cubic-Bessel, and log cubic-spline (hereafter log-spline).

In this study, we ignore the tangent point drift of the RO (Poli and Joiner 2004; Foelsch et al. 2011;
et al. 2011), thus assuming the observed profile is vertical rather than slightly slanted. The total tangent point drift in a typical occultation is about 100–250 km (Burrows 2015). The filtered and interpolated values of refractivity on the high-resolution grid are determined mainly by neighboring points, and the typical difference in horizontal position for these points is less than 10 km. Given the long horizontal footprint of an RO of ~200 km (Kursinski et al. 1997; appendix A of Anthes et al. 2000), the adjustments in the filtered and interpolated bending angles associated with correcting for these differences in horizontal location would be small and would not alter the vertical structure of the profiles in any way that was significant to affect the interpolation results, either in individual profiles or in a statistical sense. The errors caused by assuming a vertical profile rather than slightly slanted profiles may be viewed as representativeness errors (Foelsche et al. 2011).

The first section describes the datasets, methodology, filtering, and forward model. The following section illustrates the effects of interpolation on an individual profile, followed by statistics and comparisons with error profiles used by two NWP centers. The last section discusses conclusions and implications, followed by three appendixes that provide additional details on the interpolation methods, filtering, and results.

2. Methods

To isolate the errors caused by vertical interpolation during the forward modeling of BA, we began with a set of RO-observed $N$ profiles. We then generated “truth” profiles of $N$ and BA, which were used in the analysis of the interpolation errors. The high-resolution observed $N$ profiles were provided by the COSMIC Data Analysis and Archive Center (CDAAC; https://cdaac-www.cosmic.ucar.edu) in the “atmPrf” files. We used all available processed, high-quality $N$ profiles from the COSMIC-1 mission located in the tropical west Pacific near Guam (between 10.5°–16.2°N and 146°E–180°) for the month of January 2008 (114 profiles total). We chose this location and a month in the deep tropics because of the high vertical variability and extreme (high and low) values of water vapor in this region. Rieckh et al. (2017, 2018) and Anthes and Rieckh (2018) characterized the variability and errors associated with RO in this same region and compared RO with other datasets. Upon our investigation into the sample size, we conclude that the sample size does not have a significant effect on our results. We conducted the same comparisons with 57 of the 114 profiles and found the results to be very similar to those presented in this paper. The mean and standard deviation of these BA, $N$, temperature, and water vapor pressure profiles are given in Fig. 1.2

a. General overview

Figure 2 describes the testing process. We started with a set of observed $N$ profiles given on a high-resolution, nonuniform mean sea level height grid. We linearly interpolated the observed $N$ profiles to a 20-m uniform mean sea level height grid, referred to as the high-resolution grid (step 1), and then filtered the $N$ profiles on the uniform grid (step 2) using the Savitzky–Golay low-pass filter (Savitzky and Golay 1964; see section 2b for more details on filtering). These filtered $N$ profiles serve as our $N$ “truth” on the high-resolution grid. To generate our simulated model $N$, we linearly interpolated the filtered $N$ to a lower-resolution model grid (step 3). We used two model grids of contrasting vertical resolutions to test the sensitivity of interpolation to the model vertical resolution. Both grids were derived from ECMWF model grids. The first is a high-resolution grid of 86 vertical levels up to about 58 km, and the second is a low-resolution grid with 21 levels up to about 48 km. It is important to note that there were no model data involved in this study; these model grids were chosen as arbitrary examples and only serve as mean sea level height grids during step 3.

In step 4, we interpolated the filtered $N$ on the model grid (serving as our simulated model $N$) from the model grid back to the 20-m high-resolution grid. We tested five interpolation methods: linear, log-linear, log-cubic, cubic-Bessel, and log-spline (cubic spline). Appendix A summarizes these interpolation methods. There are two reasons for interpolating to the high-resolution grid. First, while some interpolations [e.g., log-polynomial in the $x$ space of Eq. (2)] allow analytical evaluation of Eq. (2) and calculation of $a(a)$ on an arbitrary impact parameter grid, other interpolations do not. The use of a high-resolution grid for simple and accurate numerical evaluation of $a(a)$ allows for the calculation of $a(a)$ for any interpolation method. The second reason is specific

2 The $N$ profiles provided by CDAAC are computed from the optimized BA, where the observed BA is additionally smoothed and gradually mixed with the first guess BA above 30–40 km. This results in the suppression of atmospheric variations and thus an underestimation of the interpolation errors. However, fractionally, this underestimation is limited, because the first guess realistically reproduces the mean exponential decrease of $N$ and BA with height. On the contrary, the use of nonoptimized BA would result in an overestimation of the interpolation errors in the stratosphere. Fractionally, this overestimation would exponentially increase with height because the mean $N$ and BA exponentially decrease, while the magnitude of the observational noise in BA remains about constant.
to this study. Since the RO-observed $N$ are smoothed and used as the “truth” (reference), they must be specified on a grid denser than the model grid for evaluation of the interpolation errors. The uniformity of the grid is convenient for the application of filtering. (Note that a uniform height grid corresponds to a nonuniform grid in $x$ space.) Therefore, for our purpose of comparing interpolation methods, we evaluated the Abel integral using the high-resolution grid at step 4. This ensures all interpolation methods were evaluated in a consistent manner using the same forward model.

We computed BA during two phases. The first set was computed from the filtered $N$ (step 2 BA) and serves as our BA truth. The second set was computed from the filtered, interpolated $N$ (step 4 BA) and was used for the evaluation of the forward-modeled BA errors. All resulting BA were given on a 20-m impact parameter grid in steps 2 and 4.

We analyzed the interpolation errors by computing normalized differences (errors):

$$X_{\text{diff}} = \frac{X_{\text{interp}} - X_{\text{truth}}}{X_{\text{truth}}} \quad (X = N \text{ or BA}).$$

\(b\). Filtering

Models are unable to resolve waves with wavelengths less than 2 times the grid interval, $2\Delta z$ (Wallington 1962; Shapiro 1970; Raymond and Garder 1991). Therefore, since RO observations generally have a higher vertical resolution (about 100m) than most NWP models, RO profiles require filtering to remove harmonics with wavelengths less than $2\Delta z$. If not removed, these small-scale features would result in representativeness errors (Lohmann 2007), which would be aliased into longer
wavelengths resolvable by the model. During data assimilation, filtering may be performed implicitly, or additional filtering may be needed depending on the details of the data assimilation. In the classical approach, the model is fitted to first guesses and observations being projected into observational space by an observational operator [forward model; see Healy (2001)]. If the model state vector is projected onto the observational grid, this fitting, which is equivalent to an implicit low-pass filtering, may be sufficient. However, projection onto a coarse grid increases the need for additional low-pass filtering of the observational vector.

Since we did not perform data assimilation, but instead generated simulated model $N$ using RO-observed $N$ (step 3, Fig. 2), the RO profiles had to be explicitly filtered to the model resolution. The choice of filter window for this filtering step may be subject to different considerations. First, this window should not be smaller than the Nyquist cutoff (i.e., $2\Delta z$). Similarly, the minimum number of grid points to define a wave resolvable by models is two (Kalnay 2003); therefore, $2\Delta z$ might be a natural choice for filter window. However, $2\Delta z$ waves are poorly resolved and can generate larger overall errors in the model itself (Shapiro 1970; Kalnay 2003; Rutt et al. 2006). Therefore, some models include filters or other damping mechanisms such as vertical diffusion to suppress other features (such as gravity waves) with vertical periods smaller than $4\Delta z$ [noting that features with $3\Delta z$ wavelengths can result in nonlinear instabilities; Orszag (1971)].

With these considerations in mind, we performed filtering using both $2\Delta z$ and $4\Delta z$ filter windows. The $2\Delta z$ window may cause undersmoothing, whereas the $4\Delta z$ window may cause oversmoothing; therefore, when estimating the interpolation errors, the results from the $2\Delta z$ and $4\Delta z$ filter windows can be considered as upper and lower error bounds.

We used the Savitzky–Golay low-pass filter (Press et al. 1992) to filter all observed $N$ profiles on the 20-m uniform grid (step 2, Fig. 2). We applied three passes of the same filter using a quadratic fitting polynomial. More details about the filtering methods are given in appendix B. Since the model grid interval $\Delta z$ varies with height, and filtering with a variable window is computationally challenging, we performed the filtering as follows:

1) We divided the model grid into three regions: 1) 0–20 km, 2) 20–40 km, and 3) 40 km and above. For each region, we computed the average grid interval $\Delta z$.

2) For each region, we defined the filter window as 2 (or 4) times the average grid interval (hereafter referred to as $2\Delta z$ or $4\Delta z$). We filtered the full $N(z)$ profiles individually with each window, resulting in the smoothed profiles $N_1(z)$, $N_2(z)$, and $N_3(z)$.

3) We combined (merged) the individually filtered $N$ profiles [$N_i(z)$] by using a transition function $w(z)$, which changes linearly from 1 to 0 between 0 and 20 km:

$$N(z) = N_1(z)w(z-20\text{ km}) + N_2(z)\times[1-w(z-20\text{ km})]w(z-40\text{ km}) + N_3(z)[1-w(z-40\text{ km})].$$

We divided the model grid into three regions to ensure the $N$ profiles were not over- or underfiltered at certain heights due to the nonuniformity of the model grid intervals. We chose the three height regions because the average grid interval was representative of the individual grid intervals in each region. We filtered the $N$ profiles individually in the whole height interval in order to eliminate end effects when merging them at 20 and 40 km. The transition function eliminated discontinuities in the combined (merged) $N$ profiles at 20 and 40 km where the filter window shifts. Figure 3 gives an example of a $N$ profile computed from $N$ using this filtering process, comparing both $2\Delta z$ (red) and $4\Delta z$ (blue) windows.

c. Bending angle forward model

The computation of model $\beta(a)$ as a function of impact parameter $a(x)$ from model $N$ as a function of mean sea level height $N(z)$ requires a series of steps before the Abel integral [Eq. (2)] can be applied. The basic geometry is illustrated in Fig. 4. Assuming the atmosphere is spherically symmetric, $N$ can be expressed as a function of radius $r$:

$$r = z + r_c + \Delta r_{geo},$$

where $z$ is the model mean sea level height, $r_c$ is the local radius of curvature of the reference ellipsoid under the occultation point, and $\Delta r_{geo}$ is the geoid undulation (i.e., the height of the geoid above the reference ellipsoid; the latter two quantities are defined uniquely for each occultation; Syndergaard 1998).

Refractivity $N$ is then converted into the refractive index $n(r) = 1 + N \times 10^{-6}$. For further simplification, the refractive index can be considered as a function of the refractive radius $x$, where $x = m(r)$. Defining $m(x) = \ln[n(x)]$, we can simplify the Abel integral [Eq. (2)] to

$$a(a) = -2a\int_a^{\infty} \frac{dm(x)}{\sqrt{x^2-a^2}} dx.$$  

Certain functions used for interpolation allow analytical evaluation of the Abel integral [Eqs. (2), (6)],
while others do not, and \( N \) (and therefore \( n \) and \( m \)) must be sampled on a high-resolution computational grid for numerical evaluation using a simple representation at each integration step. Since we are performing comparisons of different interpolation methods, we evaluate the Abel integral by using a high-resolution grid and the same numerical representation at each step for all interpolation methods; this allows for minimization of the computational differences.

We can approximate Eq. (6) in terms of finite sums:

\[
\alpha(a) = -2a \sum_i \int_{x_i}^{x_{i+1}} \frac{dm}{dx} \frac{dx}{\sqrt{x^2 - a^2}} \\
\simeq -2a \sum_i \frac{1}{\sqrt{x_i + a}} \int_{x_i}^{x_{i+1}} \frac{m'(x) dx}{\sqrt{x - a}}, \quad (7)
\]

where \( m'(x) \) is the approximation of the derivative of \( m \) on the interval \((x_i, x_{i+1})\). Since \( m(x) \) is oversampled due to the vertical interpolation of \( N \) to the high-resolution 20-m grid, it is sufficient to approximate \( m'(x) \) linearly, defining

\[
A_i = \frac{m(x_{i+1}) - m(x_i)}{x_{i+1} - x_i}. \quad (8)
\]

This results in a simplified expression of the forward model, which is used to compute BA in this study:

\[
\alpha(a) = -4a \sum_i \frac{A_i}{\sqrt{x_i + a}} \left( \sqrt{x_{i+1} - a} - \sqrt{x_i - a} \right), \\
x_i = r_i [1 + N(r_i) \times 10^{-6}]. \quad (9)
\]
The BA computed using Eq. (9) are then converted from functions of impact parameter \( a \) to impact height, where impact height is defined as

\[
\text{impact height} = a - r_c - \Delta r_{\text{geo}}.
\]  

(10)

During BA assimilation, the model \( N \) profiles must be extrapolated above the model tops in accordance with the upper limit of the integrals in Eqs. (2) and (6). In this study, no extrapolation of \( N \) is performed. Since we compute BA from smoothed (step 2, Fig. 2) and interpolated (step 4, Fig. 2) \( N \) using the same forward model [Eq. (9)], any errors associated with the lack of extrapolation are removed when their difference is computed [Eq. (3)].

Further details and other examples of the forward modeling of BA can be found in Eyre (1994), Healy and Thépaut (2006), and Cucurull et al. (2013).

3. Results

a. Example profile

Since the vertical derivative of \( N \) is necessary for the forward modeling of BA, each interpolation method is evaluated by not only how well it reproduces \( N \), but also how well it reproduces the vertical derivative. We computed an approximation of the vertical derivative of \( N \), \( dN/dz \), using a centered difference (computed on the high-resolution 20-m grid). Some characteristics of the interpolation errors are different in the troposphere and stratosphere; thus, we provide an example and discussion for each region.

Figure 5 shows \( N \), \( dN/dz \), and BA profiles computed for a single RO profile between 4 and 6 km using both the \( 2\Delta z \) and \( 4\Delta z \) filter windows. We chose to illustrate the linear, log-linear, and log-spline interpolations because they showed the most contrast in the differences between the interpolated and “true” profiles. The log-cubic and cubic-Bessel interpolations produce results similar to the log-spline interpolation.

The interpolation errors are larger with the \( 2\Delta z \) filter window (Figs. 5a–c). The small differences between the interpolated \( N \) profiles and the true \( N \) profile in Fig. 5a generate large errors in the resulting \( dN/dz \) profiles (Fig. 5b). Both linear and log-linear interpolation poorly reproduce \( dN/dz \) (producing step functions, characteristic of linear interpolation). Log-spline interpolation, though generating a smooth \( dN/dz \) profile, does not reproduce the local extrema in \( dN/dz \) (particularly between 5 and 6 km), effectively smoothing them out. The errors in \( dN/dz \) map into the BA errors in Fig. 5c. In this height interval, linear and log-linear interpolation produce similar (almost indistinguishable) BA errors, whereas log-spline interpolation produces smaller, but noticeable, BA errors.

Increasing the filter window to \( 4\Delta z \) (Figs. 5d–f) decreases the magnitude of the interpolation errors, mostly for linear and log-linear interpolations. The structures of the errors caused by linear and log-linear interpolation remain consistent across the filter windows, particularly in the \( dN/dz \) and BA profiles. The errors caused by log-spline interpolation decrease the most when the filter window increases, especially in Figs. 5e and 5f, where log-spline produces \( dN/dz \) and BA profiles that are nearly identical to the true profiles.

Figure 6 shows \( N \), \( dN/dz \), and BA using linear, log-linear, and log-spline interpolations for the \( 2\Delta z \) window between 30 and 32 km for the same RO profile and model grid used in Fig. 5. The \( 4\Delta z \) results in this region are similar to the \( 2\Delta z \) results in Fig. 6 and are not shown.
The structures of \( N \) (not clearly distinguishable), \( dN/dz \), and forward-modeled BA errors caused by linear interpolation are similar to those in the troposphere, whereas the errors caused by log-linear interpolation are much smaller than those caused by linear interpolation. Log-spline interpolation reduces errors in \( N \), \( dN/dz \), and BA relative to both linear and log-linear interpolations, generally producing smooth and accurate \( N \), \( dN/dz \), and BA profiles in this region.

The differences and similarities between the interpolation errors for different interpolation methods, height regions (i.e., troposphere and stratosphere), and filter windows can be explained by the structure of the \( N \) profile. Overall, \( N \) profiles are close to exponential (with a scale of about 10 km) with “superimposed” variations of different scales. In the troposphere, those superimposed variations are the main contributors to the interpolation errors, while the mean exponential behavior of \( N \) does not matter since the grid interval is much smaller than the exponential scale. This explains the small difference between linear and log-linear interpolations and the better performance of log-spline interpolation. This also explains the reduction of the interpolation errors when increasing the filter window from \( 2\Delta z \) to \( 4\Delta z \).
In the stratosphere, the increase of the model grid intervals increases the contribution of the mean exponential shape of $N$ to the interpolation errors and decreases the contribution of the “superimposed” variations. The latter is due to two factors: (i) increased smoothing of $N$ in our processing and (ii) the use of $N$ retrieved with optimization of BA in the RO processing (mentioned earlier). This results in the better performance of log-linear interpolation, compared to linear interpolation in the stratosphere. Small difference between the interpolation errors for the $2\Delta z$ and $4\Delta z$ filter windows is because our filtering does not substantially change the mean exponential shape of $N$.

**b. Statistics**

We computed the mean and root-mean-square (RMS) of the normalized $N$ and BA differences [Eq. (3)] to generate vertical profiles of errors for each interpolation method. Hereafter, these are referred to as mean and RMS percent differences.

1) **MEAN PERCENT DIFFERENCES**

The mean percent differences illustrate the biases each interpolation method introduces to both the $N$ and BA profiles. Here, we show the mean percent differences in $N$ and BA for the $2\Delta z$ filter window and discuss the biases caused by the five interpolation methods investigated.

Figure 7 shows the mean percent $N$ differences using the $2\Delta z$ filter window for both model grids. The results using the $4\Delta z$ filter window (not shown) are similar to the results presented in Fig. 7 (the only difference being that the magnitudes of the errors for the $4\Delta z$ filter window are smaller in the troposphere). The linearly interpolated $N$ (gray) has a positive bias on both model grids, consistent with the convexity of the mean profile and model grid interval (less pronounced below 2 km on the high-resolution grid due to the smaller grid intervals). Errors caused by log-linear interpolation (pink) are relatively large below 30 km and decrease above 30 km on both model grids (similar to the results in Figs. 5, 6). Log-cubic (green) and cubic-Bessel (orange) interpolation similarly reduce errors relative to linear and log-linear interpolation; however, log-spline (blue) interpolation produces the smallest errors on both model grids.

The $N$ errors in Fig. 7 propagate into the BA errors in Fig. 8 as $N$ is forward modeled into BA. Again, linear (gray) and log-linear (pink) interpolation generate the largest BA errors. The errors caused by log-linear interpolation decrease above 30 km (just as with $N$). When interpolating from either model grid, log-spline interpolation (blue) reduces BA errors the most overall.

The interpolation errors for all methods on the high-resolution grid are much larger in the lower troposphere, compared to the upper troposphere and lower stratosphere (Figs. 7b, 8b). The high moisture content with finescale structures and strong vertical gradients produces large vertical variability in $N$ and BA in the lower troposphere. The high-resolution model grid preserves some of this variability, even after filtering, resulting in the increase of interpolation errors. This increase is somewhat more pronounced for log-spline due to its nonlocal nature, though log-spline interpolation errors still remain.
the smallest. Note that all errors decrease again near
surface due to the significant decrease of the model grid
spacing, which allows better resolution of the small-scale
structures.

Performing log-linear interpolation (or equivalently
assuming that \( N \) varies exponentially between model
levels) produces significant biases in \( N \) and BA rela-
tive to higher-order interpolation methods (e.g., the
cubic interpolations: log-cubic, log-spline, and cubic-
Bessel). Our results are similar to those addressed by

Burows et al. (2014), where biases caused by log-linear
interpolation are significant when model grid intervals
are large, particularly in the stratosphere. Our results
also indicate that log-linear interpolation produces sig-
nificant \( N \) and BA biases around the tropopause on the
high-resolution model grid (about 18–22 km; Figs. 7b,
8b), even though the model resolution is relatively high
(average model grid interval of 0.56 km between 18 and
22 km). This bias at the tropopause can be reduced by
applying a higher-order interpolation method (such as

\[ \text{Refraction Mean Percent Differences (2\( \Delta z \) filter window)} \]

\[ \text{Bending Angle Mean Percent Differences (2\( \Delta z \) filter window)} \]

[Fig. 7. Mean percent \( N \) differences for each interpolation method using the 2\( \Delta z \) filter window. Profiles of the mean interpolation errors for the (a),(b) high- and (c),(d) low-resolution model grids. The model levels are plotted for reference.]

[Fig. 8. As in Fig. 7, but for BA.]
log-spline), which is an alternate method of reducing log-linear interpolation biases complementary to those proposed by Burrows et al. (2014).

The magnitude of the $N$ and BA errors in Figs. 7 and 8 depends heavily on the vertical resolution of the model grid. The errors for the low-resolution model grid are 2–10 times larger than the errors for the high-resolution model grid.

2) RMS PERCENT DIFFERENCES AND COMPARISONS WITH NWP ERROR PROFILES

Some NWP centers (e.g., the Met Office and ECMWF) use an estimated standard deviation of the observation plus forward model error as a function of height to define the assumed error during $N$ or BA assimilation. We show the Met Office and ECMWF profiles as examples to help determine the significance of the interpolation errors relative to the total assumed error during $N$ or BA assimilation. These examples in no way affect our computations. The interpolated errors should be smaller than the total assumed error; if they equal or exceed the assumed errors, then either the assumed errors should be increased or (preferably) the interpolation errors should be reduced.

Figure 9 compares the Met Office $N$ error profile (Rennie 2010) with the RMS percent $N$ differences using the $2\Delta z$ filter window for both the high- and low-resolution model grids. The Met Office $N$ error profile (black, dashed) from Rennie (2010) is plotted for comparison. Model levels are plotted for reference.

![Figure 9. RMS percent $N$ differences for each interpolation method using the $2\Delta z$ filter window. Profiles of interpolation errors for (a) high- and (b) low-resolution model grids. The Met Office $N$ error profile (black, dashed) from Rennie (2010) is plotted for comparison. Model levels are plotted for reference.](image-url)
in which only linear and log-linear interpolations produce errors that are greater than the model error profile (Fig. 10b). Note that the errors resulting from the 2Δz and 4Δz filter windows are considered upper and lower bounds for the interpolation errors.

The BA percentage errors (Fig. 11) are similar to the N errors in Fig. 9, though larger in magnitude. Here, we compare BA errors to the error profile used at ECMWF (Healy et al. 2017). Generally, on the high-resolution model grid, all interpolation methods produce BA errors that are less than the model error profile (Fig. 11a). An exception occurs above 18 km, where linear (gray) and log-linear (pink) interpolation generate BA errors that are greater than the model error profile, unlike what we find in N (Fig. 9a). The BA errors are much more significant for the low-resolution model grid (Fig. 11b). As with N, all five interpolation methods produce errors that are greater than the model error profile when applying the 2Δz filter window (occurring between 15 and 30 km), which does not occur when applying the 4Δz filter window (Fig. 12b). With the 4Δz filter window, the BA errors caused by log-cubic (green) and log-spline (blue) interpolation are consistently less than the model error profile, reaching at most two-thirds of the assumed model BA error.

The increase in N and BA errors between about 16 and 20 km for the high-resolution model grid (see Figs. 9–12a) is a result of the sharp tropopause characteristic in the tropics. Though the mean tropopause height for this set of RO profiles occurs around 17.5 km, there are small convexity changes in N occurring right above the tropopause (see Fig. 1) that the high-resolution model grid can resolve. Upon analysis of individual N and BA profiles, the interpolation methods had difficulty reproducing both the convexity change above the tropopause and the tropopause itself, resulting in large N and BA errors between about 16 and 20 km, just as we see in the RMS percent differences.

There is a similar increase in N (between 2 and 3 km) and BA (between about 3 and 5 km) errors for both model grids (see Figs. 9–12) that is likely caused by large gradients in N associated with the top of the boundary layer. The large spread in the mean temperature, water vapor pressure, and N profiles in Fig. 1 indicates the
presence of the boundary layer around 2–3 km (about 3–5-km impact height) and corresponds to the increase in $N$ and BA errors in this region. The large $N$ and BA errors caused by linear and log-linear interpolations show how poorly these interpolations resolve the boundary layer, whereas log-cubic, cubic-Bessel, and log-spline interpolations can better resolve the boundary layer, as reflected in the smaller $N$ and BA errors when using these three interpolations.

To analyze the BA errors further, we calculated the percentage of interpolated BA with absolute values of interpolation errors greater than the model error, where the model BA error is given by Healy et al. (2017). For a given BA value computed from interpolated $N$, if the absolute value of the BA error is larger than the assumed model error, the BA value is counted and the percentage of all forward-modeled BA per interpolation method is computed and given in Table 1.

When interpolating from the high-resolution model grid, regardless of the filter window, the percentage of interpolated BA values with errors larger than the model error is small, particularly for log-spline interpolation. The percentages become more significant when interpolating from the low-resolution model grid. Almost half of the forward-modeled BA generated by linearly interpolated $N$ contain errors that are greater than the assumed model error. There is improvement when applying the other four interpolation methods; however, the amount of improvement depends on the filter window. When the $N$ profiles are undersmoothed (i.e., the $2\Delta z$ filter window), log-linear, cubic-Bessel, log-cubic, and log-spline interpolations produce similar percentages of BA with errors that are greater than the model error. In contrast, when profiles are oversmoothed (i.e., the $4\Delta z$ filter window), log-cubic and log-spline interpolations reduce BA errors the most, producing significantly smaller BA errors than log-linear and cubic-Bessel interpolations.

The relationship between the vertical resolution of the model grid and the vertical structures represented on the model grid heavily influences the accuracy of the interpolation. Our results suggest that when complex structures are sampled on a low-resolution grid (but are still resolvable; i.e., the $2\Delta z$ results on the
low-resolution model grid), log-linear, log-cubic, cubic-Bessel, or log-spline interpolation suffice in reducing forward-modeled BA errors. When the model resolution increases (even as the complexity of the vertical structures increases as well, but still remain resolvable), log-spline proves to be the most accurate interpolation method.

4. Conclusions

We have shown that vertical interpolation of refractivity \( N \) between model levels during the forward modeling of bending angles (BA) results in errors that vary with each interpolation method. By using RO-observed \( N \) as input and assumed “truth,” we isolated the vertical interpolation errors in both \( N \) and BA resulting from five different interpolation methods: linear, log-linear, log-cubic, cubic-Bessel, and log-spline. We used two model grids of contrasting vertical resolutions to assess the magnitude of the interpolation errors relative to the vertical resolution of the model grid.

Overall, linear and log-linear interpolations result in the largest \( N \) and BA errors when interpolating from both high- and low-resolution model grids. The magnitude of the errors caused by log-linear interpolation decreases above 30 km due to the exponential behavior of \( N \) in the stratosphere. Also, log-linear interpolation errors reach a local maximum around the tropopause; these errors can be substantially reduced by using higher-order interpolation methods such as log-cubic, cubic-Bessel, and log-spline interpolations. Of the five

<table>
<thead>
<tr>
<th>Method</th>
<th>High-resolution model grid</th>
<th>Low-resolution model grid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>4.709%</td>
<td>3.664%</td>
</tr>
<tr>
<td>Log-linear</td>
<td>0.945%</td>
<td>0.412%</td>
</tr>
<tr>
<td>Cubic-Bessel</td>
<td>0.264%</td>
<td>0.015%</td>
</tr>
<tr>
<td>Log-cubic</td>
<td>0.211%</td>
<td>0.002%</td>
</tr>
<tr>
<td>Log-spline</td>
<td>0.037%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

These results highlight the importance of using appropriate interpolation methods to minimize errors in the forward modeling of bending angles.
methods, the log-spline interpolation produces the smallest mean and RMS $N$ and BA errors on both model grids, and these errors generally remain within the assumed $N$ and BA error models used at the Met Office and ECMWF, respectively.

The assumed (specified) error profile in the data assimilation system is intended to represent all errors, including interpolation errors. Ideally, interpolation errors should be much smaller than the specified error profile. If the interpolation errors contribute significantly to the total specified errors, their reduction would permit a reduced specified error profile, resulting in increased impact of the RO observations in NWP. The interpolation errors of all of our tested methods associated with relatively high-vertical-resolution model grids are small and generally remain within the total assumed errors. For models with low vertical resolution, the interpolation method becomes more important, and interpolation errors can exceed the specified errors in some modeling systems.

**Acknowledgments.** This work was supported by the National Science Foundation under Cooperative Agreement AGS-1522830. The authors thank Eric DeWeaver (NSF) and Jack Kaye (NASA) for their support of this research through this grant. The first author was supported in part by the Significant Opportunities in Atmospheric Research and Science (SOARS) program, NSF Grant AGS-1641177, and by the Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) program at UCAR, which is sponsored by the National Space Office in Taiwan, NSF, NASA, NOAA, and the U.S. Air Force. The authors thank Therese Rieckh, Sean Healy, Ian Culverwell, and Stig Syndergaard for their input throughout this project. Thank you to CDAAC for the RO data sets. The authors declare that they have no conflict of interest. The authors thank Mark Leidner and two anonymous reviewers for their constructive comments and suggestions.

**APPENDIX A**

**Description of the Interpolation Methods**

We considered five interpolation methods in this study: linear, log-linear, log-cubic, cubic-Bessel, and log-spline. Here, we describe each interpolation method briefly.

Linear interpolation is the simplest interpolation, requiring only two data points and interpolating by determining the straight line between the two data points. Three of the five interpolations are logarithmic interpolations: log-linear, log-cubic, and log-spline. In each of these methods, the natural logarithm of the dependent variable ($N$) is taken (converting $N$ into log-space), the interpolation is performed in log-space, and then the interpolated function is projected out of log-space using the exponential function. Since $N$, on average, behaves exponentially in the vertical, converting $N$ into log-space before interpolation should result in more accurate interpolation, especially for linear interpolation. Linear, cubic, and cubic-Bessel interpolations are local interpolations [an interpolated value depends on a limited number (two to four) data points], while spline interpolation is nonlocal (an interpolated value depends on all data points in the profile).

The cubic-Bessel interpolation is a special case of piecewise-cubic Hermite interpolation. Piecewise-cubic Hermite interpolation requires knowledge of the derivative at each data point, which is most often unknown or difficult to determine (Conte and de Boor 1980). In our cubic-Bessel interpolation, we approximate the derivative with a second-order polynomial through the data point (using one known neighboring value on either side). This makes our interpolation equivalent to that used at NCEP (Cucurull et al. 2013) in the internal points. At the end points, the derivatives are determined by a third-order polynomial passing through the four closest values to the end (S. Syndergaard 2017, personal communication).

The fifth interpolation is log cubic-spline (referred to as log-spline in the paper), where a cubic-spline interpolation is performed in log-space. Splines are advantageous because the resulting function is $C^2$ (i.e., continuously differentiable up to the second derivative), which is useful since the BA forward model relies on the derivative of $N$. There are instances where splines can behave poorly and result in unrealistic values (typically, this happens when interpolating undersampled functions). Upon investigation (not shown), this did not occur during our interpolation tests. We also conducted tests comparing log
cubic-spline interpolation with cubic-spline interpolation with no logarithm performed (not shown), and these tests showed very minor differences between the resulting \( N \) and BA profiles. However, the log cubic-spline interpolation resulted in slightly smaller errors and was therefore chosen for the main results.

Finally, it should be noted that with the use of a dense computational grid, which allows accurate calculation of BA using a simple discrete representation [Eq. (9)], interpolation of \( N \) can be performed either in height \( = r - r_c - \Delta r_{\text{geo}} \) or in refractional height \( = rm(r) - r_c - \Delta r_{\text{geo}} \). In this study, we perform interpolation in height. However, accurate (analytical) calculation of BA by using log-polynomial interpolation (without a dense computational grid) would require interpolation in refractional height.

Fig. C1. Mean \( N \) differences for each interpolation method and both filter windows [(a)–(d) \( 2 \Delta z \); (e)–(h) \( 4 \Delta z \)]. (a),(b),(e),(f) Error profiles for the high-resolution model grid. (c),(d),(g),(h) Error profiles for the low-resolution grid. Model levels are plotted for reference.
Further Discussion on Filtering

In this study, in order to smooth the RO refractivity \( N(z) \) specified on a uniform grid, we use the sliding window polynomial regression, sometimes termed the Savitzky–Golay filter (Press et al. 1992). We use a second-order polynomial by fitting it to the data in the least squares sense in the sliding window of fixed length \( w \). The continuous representation is given as

\[
\int_{z-w/2}^{z+w/2} \left[ Az'^2 + Bz' + C - N(z') \right]^2 dz' = \text{min}. \quad (B1)
\]

The smoothed function equals \( \bar{N}(z) = Az^2 + Bz + C \). Coefficients \( A \), \( B \), and \( C \) are calculated for \( z \) when \( z \) is inside the interval \([z_{\text{min}} + w/2, z_{\text{max}} - w/2]\), and they are
calculated for the endpoints of this interval when \( z \) is outside the interval (the latter results in certain end effects). We used the FORTRAN subroutine “savgol” from Press et al. (1992).

It is common to coarsely characterize the smoothing property (or smoothing scale) of a sliding window filter by the length of the window \( w \). However, the filter can be more accurately characterized by the response function, shown in Fig. B1 (red line). The \( x \) axis shows the period, normed by the window length \( w \), of the sinusoidal function of unit amplitude on input to the filter. The \( y \) axis shows the amplitude of the sinusoid on output of the filter. It is seen that the sinusoid with period 1 is damped by a factor of \( \approx 0.76 \). Damping factors of 0.5 and 0.7 (sometimes used for the conventional definition of the filter cutoff period) correspond to periods of \( \approx 0.79 \) and \( \approx 0.94 \) (i.e., slightly smaller than the sliding window). However, the response function has local maxima (side
lobes) at smaller periods. For example, harmonics with periods of $\sim 0.45$ will propagate into the filtered function with a damping factor of $\sim 0.24$, which is not small (this may be considered as an aliasing effect). One way to suppress the side lobes in the response function (to reduce aliasing effects) is by applying multiple applications of the same filter. For example, after three applications of the second-order polynomial regression (Fig. B1, green line), the side lobes in the response function are significantly repressed. Damping factors of 0.5 and 0.7 now correspond to periods of $\sim 1.05$ and $\sim 1.25$ (i.e., slightly larger than the sliding window). Formally, it is possible to redefine the smoothing scale of the filter. However, in this study, we conventionally consider $w$ as the smoothing scale by just keeping in mind the shape of response function shown in Fig. B1, which has damping factor of $\sim 0.44$ for harmonics with period $w$ and negligible side lobes at smaller periods.

**Fig. C4.** As in Fig. C3, but for BA.
APPENDIX C

Nonnormalized Mean and RMS Differences

To complement the normalized differences computed in sections 3b(1) and 3b(2), we include the mean and RMS of the nonnormalized differences for both filter windows here (see Figs. C1–C4). Since $N$ and $BA$ values are much larger in magnitude in the troposphere relative to the stratosphere, the $N$ and $BA$ errors in the troposphere are emphasized in these nonnormalized differences.

REFERENCES


——, 2008: An introduction to GPS radio occultation and its use in numerical weather prediction. GRAS SAF Workshop on Applications of GPS Radio Occultation Measurements, Reading, United Kingdom, ECMWF, 10 pp.


