COSMIC Radio Occultation Processing: Cross-Center Comparison and Validation

MICHAEL E. GORBUNOV AND A. V. SHMAKOV
Obukhov Institute for Atmospheric Physics, Moscow, Russia

STEPHEN S. LEROY
School of Engineering and Applied Science, Harvard University, Cambridge, Massachusetts

KENT B. LAURITSEN
Danish Meteorological Institute, Copenhagen, Denmark

(Manuscript received 29 April 2010, in final form 1 February 2011)

ABSTRACT

A radio occultation data processing system (OCC) was developed for numerical weather prediction and climate benchmarking. The data processing algorithms use the well-established Fourier integral operator–based methods, which ensure a high accuracy of retrievals. The system as a whole, or in its parts, is currently used at the Global Navigation Satellite System Receiver for Atmospheric Sounding (GRAS) Satellite Application Facility at the Danish Meteorological Institute, German Weather Service, and Wegener Center for Climate and Global Change. A statistical comparison of the inversions of the Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) data by the system herein, University Corporation for Atmospheric Research (UCAR) data products, and ECMWF analyses is presented. Forty days of 2007 and 2008 were processed (from 5 days in the middle of each season) for the comparison of OCC and ECMWF, and 20 days of April 2009 were processed for the comparison of OCC, UCAR, and ECMWF. The OCC and UCAR inversions are consistent. For the tropics, the systematic difference between OCC and UCAR in the retrieved refractivity in the 2–30-km height interval does not exceed 0.1%; in particular, in the 9–25-km interval it does not exceed 0.03%. Below 1 km in the tropics the OCC – UCAR bias reaches 0.2%, which is explained by different cutoff and filtering schemes implemented in the two systems. The structure of the systematic OCC – ECMWF difference below 4 km changes in 2007, 2008, and 2009, which is explained by changes in the ECMWF analyses and assimilation schemes. It is estimated that in the 4–30-km height range the OCC occultation processing system obtains refractivities with a bias not exceeding 0.2%. The random error ranges from 0.3%–0.5% in the upper troposphere–lower stratosphere to about 2% below 4 km. The estimate of the bias below 4 km can currently be done with an accuracy of 0.5%–1% resulting from the structural uncertainty of the radio occultation (RO) data reflecting the insufficient knowledge of the atmospheric small-scale structures and instrumental errors. The OCC – UCAR bias is below the level of the structural uncertainty.

1. Introduction

Radio occultation (RO) of the microwave signals transmitted by the global positioning system (GPS) has been shown to be an accurate and precise sounding technique of Earth’s atmosphere (Ware et al. 1996; Kursinski et al. 1997; Kursinski et al. 2000; Kuo et al. 2004). Currently, RO data are widely used for numerical weather prediction in schemes of variational assimilation (Zou et al. 2000; Liu et al. 2001; Liu and Zou 2003; Healy et al. 2005; Healy and Thépaut 2006; Healy et al. 2007; Cucurull et al. 2007; Aparicio and Deblonde 2008; Cucurull et al. 2008; Wee et al. 2008; Pingel and Rhodin 2009). The Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) mission (Rocken et al. 2000) opened a new perspective for the application of RO sounding. This observing system consists of six satellites that, on average, obtain 2000–2500 globally distributed occultation events per day. The
open-loop mode of the receiver allows for the tracking of lower-tropospheric signals (Sokolovskiy et al. 2006a,b, 2007).

The application of RO data for the investigation of climate became the subject of many discussions (Ringer and Healy 2008; Lösch et al. 2008; Ho et al. 2009). The radio occultation sounding technique is expected to overcome the shortcomings of the climate data records (National Research Council 2004; Leroy et al. 2006a; von Engeln 2006) that rely on the stability of the calibration and require extended overlaps between successive instruments. RO technique differs fundamentally from other remote sounding techniques in that it is an active measurement of the phase of an electromagnetic signal. Its most basic observed quantity is one of time, and so its fundamental calibration must be linked to the international definition of the second (Leroy et al. 2006b), which is itself accurate to a few parts in $10^{15}$. By contrast, radiance standards, if they exist, are accurate to $\approx 0.1\%$. Performance analyses of GPS radio occultation as a climate monitoring system shows remarkable accuracy, especially in the upper troposphere and lower stratosphere (Steiner and Kirchengast 2005).

Despite their very high nominal calibration accuracy, RO data are subject to other kinds of random errors and biases. In the upper stratosphere, the retrieval of refractivity is subject to systematic errors related to the initialization of the Abel transform, which converts profiles of bending angle to profiles of refractivity. Most modern retrieval algorithms use a form of optimal estimation, wherein a climatological model of bending angles in the upper stratosphere and mesosphere is merged with observed bending angles in order to render the Abelian transform less noisy in the upper stratosphere (Sokolovskiy and Hunt 1996; Gorbunov 2002b). Still, if the background climatology is biased, then the bias may propagate down to 10 km.

In the lower troposphere, the retrieval of refractivity is subject to the complicated radio occultation signal dynamics caused by water vapor. Open-loop tracking of radio signals has been implemented to address this problem (Sokolovskiy et al. 2006b), but systematic error in refractivity still remains (Ao et al. 2003a; Ao 2007). Steiner and Kirchengast (2005) also show that climatology of refractivity in the lower troposphere must be further biased because of receiver tracking errors. During the setting of radio occultations the GPS receiver tends to lose tracking at higher altitudes, and rising occultations tend to initiate at higher altitudes in the presence of stronger ducting, thus biasing radio occultation–based climatologies toward states of lower ducting and drier atmospheres. Despite this tendency, it is not obvious that systematic errors in a retrieved quantity (refractivity) derived from Système International d’Unités (SI)–traceable observations (atmosphere-induced time delay) are a hindrance to climate monitoring, because perfectly systematic errors should not contribute to errors in trends in that retrieved quantity.

Biases of the order of 0.1% may constitute a serious problem in climate monitoring. Therefore, such biases need to be understood at least. An effective way of exposing biases is through a cross-center intercomparison of the results of processing the excess phase data. A study wherein the starting dataset was the raw phases and orbits (before clock corrections) was the Radio Occultation Sensor Evaluation (ROSE) study (Ao et al. 2003b). A comparison of processing of data from Challenging Minisatellite Payload (CHAMP) by two different centers [GeoForschungsZentrum (GFZ) and University Corporation for Atmospheric Research (UCAR)] and a study of the corresponding structural uncertainty (difference introduced by the choice of processing methodology) was presented by von Engeln (2006). Another similar project was the Radio Occultation Processing Intercomparison Campaign [ROPIC; see information online at http://www.grassaf.org (see also Lösch et al. 2009)], whose purpose was to validate the processing chains from various RO centers worldwide. A fourth project is the RO Trends study (Ho et al. 2009), in which different RO processing centers aimed at quantifying the structural uncertainty in GPS RO profiles of refractivity through the analysis of 5 yr of climatologies derived from RO data from the German CHAMP satellite.

The observed systematic difference between the retrieved quantities like refractivity and those from analyses of weather prediction centers are a sum of systematic errors of the RO data and those of the analyses (there are no data without some random and systematic errors). The general rule is that each systematic error cannot be attributed to some single factor; it is a sum of systematic errors from many different sources. These sources may be, in turn, subdivided into two categories—instrumental and algorithmic errors—and distinguishing between them is difficult. The comparison of the results of the processing of phase excess data by different centers may reveal the distribution of the errors with respect to different error sources. Such a comparison is especially effective when performed on levels of different quantities, such as raw bending angles (prior to the statistical optimization) and refractivities. Refractivity is a very convenient quantity, because it can be directly computed from the fields of temperature, pressure, and humidity supplied by weather prediction centers. Also, retrieval of refractivity from RO data does not rely upon the assumption of a perfectly dry atmosphere. By comparing refractivity retrieved using different upper
boundary condition initialization schemes on different
days we can discover some biases inherent to different
RO retrieval systems and separate them from other fac-
tors that are intrinsic to the raw data.

In this paper we present the results of such cross-
center intercomparison. The initial data for this study
were the following UCAR data products: 1) files con-
taining calibrated excess phase and signal-to-noise ratio
(SNR) from the COSMIC constellation (atmPhs files); 2) files containing GPS navigation bits used for 50-Hz de-
modulation (gpsBit files); and 3) files containing the at-
mospheric profiles (atmPrf files) retrieved by UCAR
from atmPhs and gpsBit files. We compare the atmo-
spheric profiles of refractivity retrieved by our system
from the same atmPhs and gpsBit data used by UCAR, to
UCAR retrievals and the analyses of European Centre
for Medium-Range Forecasts (ECMWF). The ECMWF
analyses are only used as a reference. We do not discuss
any possible biases of ECMWF data.

Our system for processing GPS RO data has been
developed by the first author (MEG) starting from his
participation in the GPS Meteorology Network (GPSMet)
project. Currently, it is capable of handling GPS radio
occultation data from all the known missions, includ-
ing GPSMet, CHAMP, the Satellite de Aplicaciones
Cientificas–C (SAC–C), COSMIC, and the Global Navi-
gation Satellite System Receiver for Atmospheric Sound-
ing (GRAS)/Meteorological Operation (MetOp). The
system consists of the main program for processing radio
occultation data, as well as different diagnostic, visuali-
sation, and simulation tools. The system was tested many
times by processing artificial RO data obtained by forward simulation (Gorbunov and Kornblueh 2001, 2003).
Different versions of the data processing system are cur-
tently used at the GRAS Satellite Application Facility
at the Danish Meteorological Institute, German Weather
Service, and the Wegener Center for Climate and Global
Change [not for GPS RO data published so far (e.g.,
Foelsche et al. 2008a, b), but for performance simula-
tion only]. In particular, many of its modules are shared
with the End-to-End Global Navigation Satellite System
(GNSS) Occultation Performance Simulator (EGOPS;
Kirchengast et al. 2007) and used as the basis for the
development of some parts of the Radio Occultation
Processing Package (ROPP; online at http://www.grassaf.
org). The Hydrometeorological Center of Russia also
uses our system as a part of the data assimilation system.
In the following we will refer to our processing system as
“OCC.”

OCC can be characterized as a robust system capa-
ble of selecting data that are suitable for further pro-
cessing and sorting out corrupted data and producing
diagnostics. Here we only provide an overview of the
system, pointing the reader to the previous publica-
tions for reference (Gorbunov and Lauritsen 2004,
2006; Gorbunov et al. 2006).

The paper is organized as follows: section 2 contains
the general description of our data processing chain—
preprocessing of measurements in L1 (1.57542 GHz) and
L2 (1.22760 GHz) frequency channels, quality control,
geometric and wave optical data processing, ionospheric
correction and noise reduction, and bending angle in-
verson. Section 3 discusses the main topic of the paper:
OCC–UCAR–ECMWF statistical comparisons on differ-
ent levels and for the four different periods. In section
4, we offer our conclusions.

2. Data processing and quality control
a. L1 preprocessing

The measurements at L1 (frequency $f_1 = 1.575$ 42 GHz)
are performed in the closed- and open-loop receiver
modes on COSMIC. The closed-loop mode is used for
tracking signals corresponding to rays with perigee heights
above $\sim 10$ km. This height varies for different occult-
tations, depending on the complexity of the local at-
mosphere. The open-loop mode is ideal for tracking
signals with strong variations of the phase, which are
characteristic of the tropospheric propagation. The open-
loop mode is used for ray perigee heights below $\sim 10$ km.
The lowest part of open-loop data is measured deeply in
the shadow zone and contains wide-band noise. This
part is usually cut off to suppress the influence of the
noise that can result in a positive bias (Sokolovskiy et al.
2009a; Sokolovskiy et al. 2010). In this study we use
a simple cutoff procedure (Sokolovskiy et al. 2009a;
Sokolovskiy et al. 2010): for each pair of GPS–low Earth
orbiting (LEO) satellite positions we compute the
straight-line perigee height and discard the data below
some threshold. In this study we mostly use the cutoff
height of $\sim 250$ km. For comparison we made a run with
the cutoff height of $\sim 150$ km.

The processing of open-loop mode signals requires
demodulation of the navigation message, a 50-Hz mod-
ulation of the L1 signal by a sequence of 0 and $\pi$ phase
shifts, and subsequent re-accumulation of the atmospher-
induced phase delay. Our procedure follows the guide-
lines of the removal of the navigation message outlined
by Sokolovskiy et al. (2006b, 2009b). The frequency of the
signal is down converted by subtracting an atmospheric
phase model from the phase, and then the navigation
modulation is subtracted from the phase, the phase is re-
accumulated by adding an integer number of cycles $2\pi$
to each sample so as to minimize the phase variation
between 50-Hz samples, and, finally, the original phase
variation is restored with the addition of the phase model. Where the time series of bits defining the navigation message is absent or marked as unreliable we use the internal demodulation (Sokolovskiy et al. 2009b). This algorithm imitates a two-quadrant detector that is insensitive to modulation of the navigation message but cannot effectively track the signal in presence of severe multipath conditions (Beyerle et al. 2003). The atmospheric phase model is based on the Mass Spectrometer Incoherent Scatter (MSIS) climatology complemented by a simple model of humidity: we assume a constant relative humidity of 80% below a height of 15 km, and humidity is set to 0 above 15 km. If the maximum bending angle projected by this model is insufficient to cover the deepest parts of the satellite trajectories, then we extend this model below Earth’s surface by the exponential extrapolation. As shown by Sokolovskiy et al. (2006b), even a rough model of the atmosphere allows for the prediction of the Doppler frequency within 10–15 Hz. Because the sampling rate of COSMIC data is 50 Hz, the ±25-Hz interval around the phase model will cover the spectrum of the signal.

In most cases the timing of the navigation bits is accurate. Still, rare cases remain in which the navigation bit sequence needs to be shifted by one sample. To correct the timing we apply the correlation approach (Ziedan 2006) and form the following two functions:

\[
F_i^\varphi = \frac{[(\pi/2 + \varphi_{1,i} - \varphi_{1,i-1}) \bmod 2\pi] - \pi}{\pi/2},
\]

\[
F_i^B = 2|B_i - B_{i-1}| - 1,
\]

where \(\varphi_{1,i}\) are the L1 phases, including the navigation bit modulation, and \(B_i\) are the navigation bits. The first function \(F_i^\varphi\) searches out the timing of the modulation flips of the navigation message and works well when the variation of the atmosphere-induced phase delay is small. The function should approximately equal +1 when the consecutive elements in the time series of bits comprising the navigation message switch from 0 to 1 (or vice versa) and -1 when the consecutive elements are the same. The second function \(F_i^B\) has the same property but operates on the navigation message recorded by a receiver on the ground. We compute the correlation function \(C^B(\Delta t) = \langle F_i^B F_{i+\Delta t}^B \rangle\) averaged with respect to index \(i\) over the time interval of 4 s below the upper height of the open-loop mode region. The correlation maximum position gives the temporal shift of the navigation bits sequence with respect to the timing of the open-loop data.

After removal of the navigation bits from open-loop data, we apply the half-cycle re-accumulation to the L1 that is phase recorded in closed-loop mode. There are several instances of occultations with extra half-cycles between samples in the closed-loop mode, which is equivalent to 25-Hz offset of the Doppler frequency. The half-cycle phase accumulation removes these extra half-cycles, restoring the correct phase variation.

b. L2 preprocessing and quality control

Processing of L2 (frequency \(f_2 = 1.22760\) GHz) phase excess differs from processing L1 data. The L2 channel does not track in open-loop mode, and missing L2 measurements are replaced by a phase model. In the closed-loop mode L2 data are much noisier than L1 data. We only use L2 above some cutoff height, which is defined by the L2 noise level, signal tracking loss, and the receiver mode. Below this height we extrapolate the L2 signal to account for the ionospheric contribution to stratospheric sounding. Here we describe our scheme for reducing the noise of L2 signal and extrapolating it.

The L2 amplitude is replaced by the smooth geometric optical (GO) amplitude simulated based on the MSIS model. L2 data are subjected to the radio holographic (RH) filter in the time domain as described in Gorbunov and Lauritsen (2006), and then the L2 phase is filtered with a window 0.25 km in impact height. We compute a badness estimate—our quality control indicator—using the procedure described in Gorbunov et al. (2006), with some modifications. Using radio holographic analysis (Igarashi et al. 2000; Gorbunov et al. 2004), we retrieve the smoothed profiles of the L1–L2 impact parameter \(P_{1,2}(t)\); the bending angle \(\xi_{1,2}(t)\), corresponding to the relative Doppler frequency shift \(D_{1,2}(t)\); the phase excess \(\delta p_{1,2}(t)\); and the impact parameter spectral widths \(\Delta p_{1,2}(t)\) (Gorbunov and Lauritsen 2006). We form the following L2 badness estimate:

\[
Q^{L2}(t) = \left[\frac{\langle P_t - P_{L2}(t) \rangle}{\Delta P_A} + \frac{\delta p_{1,2}(t)^2}{\Delta P_D}\right],
\]

where \(\Delta P_A = 0.2\) km and \(\Delta P_A = 0.15\) km (these values having been determined empirically). The badness parameter penalizes data for which L1 and L2 impact parameters differ from each other by more than \(\Delta P_A\), and where the L2 signal has a spectral width larger than \(\Delta P_D\). The former criterion determines where L2 most likely lost closed-loop tracking, and the latter criterion penalizes L2 data that are overly noisy. The discrepancy between L1 and L2 [represented here by the first term in Eq. (3)] is also employed in the UCAR and GFZ data processing systems for the determination of the L2 cutoff height (Kuo et al. 2004; Beyerle et al. 2004). The L2 cutoff point is defined as the highest point of the closed-loop record where \(Q^{L2}(t)\) exceeds threshold \(Q_0\) empirically determined to be 15, but not higher than.
the straight-line perigee height of 20 km. If everywhere \(Q^{L2}(t) < Q_0\), then the cutoff point is defined as the lowest point of the closed-loop record.

Below the cutoff point we extrapolate the L2 signal in order to account for ionospheric contribution to tropospheric sounding. We estimate the ionospheric difference of the bending angles \(\Delta E_f\) by averaging \(\tau_1(p) - \tau_1(p)\) over the 2-km interval above the cutoff point. The L2 bending angles below the cutoff point are estimated as \(\tau_1(p) + \Delta E_f\), and from them we compute the extrapolated smoothed L2 phase excess \(\tilde{S}_2^\text{ext}(t)\). The ionospheric difference of the phase excess is estimated as \(\tilde{S}_2 = \tilde{S}_2^\text{ext}(t) - \tilde{S}_2(t)\). The corrected L2 phase excess \(\tilde{S}_2^\text{cor}(t)\) below the cutoff point is computed as follows: the finite differences of \(S_2^\text{cor}(t)\) are formed as a linear combination of the finite differences of \(S_1 + \Delta S_1\) and the finite differences of \(S_2\) with weights \(W(t)\) and \([1 - W(t)]\), respectively (Gorbunov et al. 2006), where the weighting function \(W(t)\) smoothly increases from 0 to 1 in the vicinity of the cutoff point. The L2 amplitude is also multiplied with \([1 - W(t)]\) and it is used for the determination of the cutoff point within the CT algorithm.

**c. Geometric optical retrieval of bending angles**

The GO retrieval of the bending angle uses the standard relationship between the Doppler frequency (i.e., the time derivative of phase), the impact parameter, and the bending angle (Ware et al. 1996; Kursinski et al. 1996). The computation of the Doppler frequency is based on the numerical differentiation of the noisy phase excess, which thus has to be filtered. Because phase excess is a convex function (i.e., has positive second derivative), some filters may result in a systematic error. To reduce this effect, we subtract a trend from the phase excess. The trend is computed by means of spline regression with a limited number of nodes. We use 30 nodes, which is significantly smaller than the typical number of samples in one occultation (usually there are not less than 3000). The resulting spline can be differentiated analytically, and the residual phase excess is differentiated numerically. Our data processing system has two choices of the filter type: the filter described by Ustinov (1990) and the polynomial regression. The latter was used in this study. Bending angles and impact parameters are obtained from the derivative of the phase excess as described in Gorbunov and Kornblueh (2001). In this study we use a filter width of 2 km, which is consistent with the UCAR settings.

**d. Wave optical retrieval of bending angles**

Our wave optical (WO) retrieval of bending angles has three choices: back propagation (BP) (Gorbunov and Gurvich 1998), canonical transform of the first type (CT) applied after BP (Gorbunov 2002a), and canonical transform of the second type (CT2) applied directly to the measurements of the complex field along the LEO orbit (Gorbunov and Lauritsen 2004). The latter method combines the numerical efficiency of the full-spectrum inversion (FSI) method (Jensen et al. 2003) and accuracy approaching that of the Phase Matching (PM) method (Jensen et al. 2004), which provides the most accurate solution. In this study we use CT2, and WO processing is activated below the height of 25 km.

The WO processing includes the RH filtering in the impact parameter domain (Gorbunov and Lauritsen 2006). As shown by Sokolovskiy et al. (2009a) and Sokolovskiy et al. (2010), RH filtering results in a negative shift of the retrieved bending angles, which is most significant in the tropics and is negligible in the polar regions. This is a consequence of two competing effects: RH suppresses the noise coming from higher altitudes and shifts the center of gravity of the spectra in the impact parameter domain in the direction of larger bending angles. RH also suppresses the stochastic part of the useful wide-band signal, corresponding to large bending angles. Currently we can neither distinguish between these two effects, nor estimate which one is stronger. This situation is termed “structural uncertainty.” We use the following data processing setup. The wave field transformed into the impact parameter domain (Gorbunov and Lauritsen 2004) is first subjected to the RH filtering (Gorbunov and Lauritsen 2006). The CT2 amplitude of the RH-filtered field is then used for the shadow border determination, and the CT2 phase of the unfiltered field is used for the bending angle retrieval, which is, therefore, not affected by the RH filter.

The L1 cutoff height is determined as the height of maximum correlation of the L1 CT2 amplitude as a function of height against a step function shifted in height. RH filtering reduces the sensitivity to noise and improves the quality of the shadow border determination. For L2, the cutoff height determined from the L2 CT2 amplitude corresponds to the L2 cutoff height defined above. Below the L2 cutoff height, we extrapolate the L2 CT2 bending angle in the same way as in GO processing.

If WO processing is activated, the bending angles from GO (above 25 km) and WO (below 25 km) processing steps are merged, producing profiles of L1 and L2 bending angles \(e_{1,2}(p)\). In the framework of WO processing, we also compute the RH error variances \(\langle \delta \hat{\gamma}_{W0}(p) \rangle^2\) of the bending angles. The standard linear combination of L1 and L2 (Vorob’ev and Krasil’nikova 1994) together with the error variance estimate is suitable for direct assimilation of bending angles at the numerical weather prediction centers.
e. Ionospheric correction and noise reduction

The ionospheric correction and noise reduction follows the optimal linear combination (OLC) technique described by Gorbunov (2002b). The computation of the background profile of the bending angle has the following options: 1) bending angles are computed by the Abel integral from local refractivity profile from the MSIS climatology, 2) there is a global search of the MSIS bending angle profile that best approximates the high-altitude bending angles (Gobiet and Kirchengast 2004) (originally suggested by S. Syndergaard 1999, personal communication), and 3) bending angles are obtained by Abel integral from the local refractivity profile extracted from global fields of analyses of ECMWF or National Centers for Environmental Prediction (NCEP).

The ECMWF refractivity is interpolated from the global fields at the approximate ray perigee locations, which are computed for each data sample. For locating the ray perigees we use the local MSIS refractivity and bending angles. Using MSIS refractivity rather than a more realistic analysis does not cause a serious error in a determination of ray perigee locations, and, because it is smooth and free of superrefraction, it requires no special handling.

The upper height of ECMWF global fields reaches 70–80 km. Because the background bending angle profile is necessary up to a larger height (150 km in our system), we extrapolate the analysis by adding MSIS refractivity profiles multiplied by a fitting coefficient. The coefficient is chosen such that the integration of the hydrostatic equation should result in temperatures coinciding with ECMWF temperatures in the height interval of ECMWF fields.

When using the MSIS climatology, either locally or from a global search and match, the profile of the bending angle is subjected to a one-parameter (Gorbunov 2002b) or two-parameter (Lohmann 2007) fitting. Given the model profile $\epsilon_M(p)$, it is replaced by $a\epsilon_M(p)$ or $a[\epsilon_M(p)]^\beta$. For a one-parameter fitting, the coefficient $\alpha$ is determined from the regression of $\epsilon_M(p)$ with respect to the linear combination $\epsilon_{OLC}(p) = [\epsilon_1(p)f_1^2 - \epsilon_2(p)f_2^2]/(f_1^2 - f_2^2)$ of L1 and L2 bending angle profiles. For a two-parameter fitting, the coefficients $\alpha$ and $\beta$ are determined from the regression of $\ln \epsilon_M(p)$ with respect to $\ln \epsilon_{OLC}(p)$. The linear combination $\epsilon_{OLC}(p)$, while it eliminates the influence of the ionosphere to first order, still contains the ionospheric noise. It is referred to as the raw bending angle profile. In this study we used two-parameter fitting in the 20–70-km height interval.

The OLC procedure produces the neutral bending angle $\epsilon(p)$ and its error variances $(\sigma_{\epsilon_{OLC}}(p))^2)$. Typically, at heights above 40–45 km the combined observation–background bending angle profile is almost entirely composed of background information. Finally, we create a merged profile of bending angle consisting of $\epsilon(p)$ in the height range where RO data are present and extended up to 150 km by the background profile $\alpha[\epsilon_M(p)]^\beta$. The transition from data to a background profile is handled by optimal estimation. We also obtain the profile of the error variances $[\sigma_{\epsilon_{OLC}}(p)]^2 + [\sigma_{\epsilon_M}(p)]^2$. The profiles of bending angles and error variances are interpolated to a homogeneous grid of impact parameters. In this study the step of the grid was 0.1 km.

f. Inversion of bending angles

The retrieval of refractivity and dry temperatures follows the standard procedure based on the Abel inversion (Rocken et al. 1997). Using the error variance of the bending angle, we also retrieve the error variances of refractivity and dry temperature (Gorbunov and Kirchengast 2005).

g. UCAR data processing system

The UCAR data processing chain is described by Kuo et al. (2004), Lohmann (2007), Sokolovskiy et al. (2009b), and Ho et al. (2009). The UCAR initialization of the bending angle inversion uses a static climate model developed at NCAR based on Met Office analysis, the Halogen Occultation Experiment (HALOE) and Microwave Limb Sounder (MLS) climatologies (Randel et al. 2002; Ho et al. 2009). The UCAR quality control (QC) discards an occultation event if 1) the noise level on raw L2 or smoothed L1–L2 phase rates exceed 6 or 0.5 cm per sample above 20 km, respectively, 2) the top-fit height for the background profile is less than 40 km, and 3) the retrieved refractivity differs fractionally from the NCAR climatology by more than 50% between 10 and 40 km. The top-fit height is determined as the maximum height where 1) the fractional difference between the observational bending angle and a smoothed one (with a 2-s window) does not exceed 60% and 2) the difference in slopes of logarithms of the observational and background bending angle profiles does not exceed 5% (from Ho et al. 2009). For the wave optical retrieval of bending angles, UCAR first implemented GO back propagation to a circular orbit (Lohmann 2007) followed by the standard FSI. Later, the phase matching (PM) technique (Jensen et al. 2004) was implemented (Sokolovskiy et al. 2009a; Sokolovskiy et al. 2010). It was found that these two algorithms provide results coinciding up to 0.1%.

3. OCC–UCAR–ECMWF statistical comparison

In our comparison we use two versions of UCAR data products: 2007.3200 and 2009.2650. In version 2009.2650,
improvements in the filtering schemes were made to reduce a positive bias of about 0.1% above 20 km, and a positive bias reaching 0.5% below 6 km. Version 2009.2650 data are currently available for dates starting on 1 April 2009.

Figure 1 shows the cross comparison of OCC, ECMWF, and UCAR (version 2007.3200) data. We use the COSMIC observations for 20 days of 2007, 5 days in the middle of every season, that is, 1–5 January, 1–5 April, 1–5 July, and 1–5 October. We compare the refractivities retrieved by the OCC and UCAR systems and the refractivities evaluated from the ECMWF analyses. The comparison only includes the data that passed both UCAR’s and our quality control procedures. The number of events included in the statistics is 29 443, the total number of UCAR profiles is 31 848, and the total number of COSMIC observations is 53 059. The events included in the statistics belong to the intersection of the subsets of the total set of events that can be processed by OCC and UCAR, respectively. Therefore, the difference between the number of events processed by these two systems does not exceed 10%. The cross-center difference of events processed may reach 30%–50% (Ho et al. 2009). Here we used the initialization based on MSIS with global search.

The ECMWF analyses have a horizontal resolution of 1° × 1° and 91 vertical levels up to a height of about 80 km (0.01 hPa). The ECMWF fields are not independent from the COSMIC observations, because the latter have been assimilated operationally at ECMWF since December 2006. For our study, however, this does not have any implications, because the ECMWF fields are only used as a reference.

In the 10–20-km height interval OCC and UCAR results agree best. The difference of the retrieved refractivities is about 0.2%, which is very close to the estimate of the RO precision (Schreiner et al. 2007). Below 6 km, the OCC – UCAR systematic difference is negative and its absolute value has a maximum at a height of 1 km. The difference is largest in the tropics, where it reaches −0.6%. The OCC – ECMWF systematic difference in the 6–30-km height range does not
exceed 0.15%. In the 2–6-km height range this difference is positive and reaches a value of 0.6% at a height of 2.3 km. Below 2 km the difference becomes negative, and in the tropics it reaches −2.2%. This bias structure does not influence the OCC – UCAR difference, which is a very smooth function of the height without any irregularities near 2 km. The OCC – UCAR difference is mostly caused by the older version of the filter that was implemented in version 2007.3200 of the UCAR processing system. This positive bias has been eliminated in version 2009.2650. The systematic differences between OCC, UCAR, and ECMWF are smaller in midlatitude and the smallest in the polar latitudes.

Both OCC and UCAR retrievals, when compared to those from ECMWF, indicate a negative bias below 2 km. In the UCAR data (version 2007.3200) it is combined
with a positive bias. The negative bias is a combination of systematic measurement errors and the systematic error of ECMWF analyses. Measurement errors may arise due to the systematic loss of cycles in the receiver when tracking low-level, noisy, and strongly beating tropical RO signals. It also may be the effect of stochastic superrefraction resulting from small-scale inhomogeneities (Sokolovskiy 2003). As shown by Gorbunov et al. (2010), strong horizontal gradients may violate the condition of the applicability of CT–FSI–PM methods and result in a negative bias of retrieved bending angles and refractivities.

Figure 2 shows the result of our numerical simulations with ECMWF fields. We used wave optics propagator (WOP) to simulate radio occultation data with realistic observation geometry taken from COSMIC occultation events. We did not simulate any measurement error, so these results provide an estimate of the effects of superrefraction and strong horizontal gradients only. In the tropics we observe a maximum negative bias of 0.75%, which is a substantial value, although it is much smaller than the negative bias of about 2% in COSMIC data. This relation between the real observations and simulations is explained by measurement error and atmospheric structures that are unresolved by ECMWF analyses, for example, structures associated with boundary layer inversions and stratus clouds.

Figure 3 shows the cross comparison of OCC, ECMWF, and UCAR (version 2007.3200) data for 2008. We use the COSMIC observations for the same 20 days of 2008 as above for 2007. The number of events included in the statistics is 31,750, the total number of UCAR profiles is 35,071, and the total number of COSMIC observations is 51,948. The negative systematic OCC – ECMWF difference is observed below 5 km. Still, the OCC – UCAR difference indicates the same stable pattern as that for 2007.

Figure 4 shows the cross comparison of OCC, UCAR (version 2009.2650), and ECMWF data for 2009. Here we use COSMIC observations for 1–20 April. The comparison is limited to these days for two reasons: 1) the new version of reprocessed UCAR products is only available as of 1 April 2009 and 2) we only have access to ECMWF analyses until 30 April 2009. The number of events included in the statistics is 37,599, the total number of UCAR profiles is 40,829, and the total number of COSMIC observations is 55,371. In version 2009.2650 of UCAR data products some improvements were made...
in the filtering schemes, which resulted in decreasing the positive biases in the lower troposphere and above 20 km (UCAR announcements are available at http://cosmic-io.cosmic.ucar.edu/cdaac/index.html).

For tropical occultations, the OCC – UCAR bias is positive in the 3–8-km height range, and it is negative below 3 km. The bias in the 2–30-km height interval does not exceed 0.1%, and in the 9–25-km height interval it does not exceed 0.03%. The positive bias has a local maximum of about 0.06% at 5–7 km and the maximum negative bias is about −0.2% at 500 m.

For the whole globe, the OCC – UCAR bias does not exceed 0.06% in the 2–30-km height interval, and in the 9–25-km height interval it does not exceed 0.01%. Below 2 km the maximum negative bias is about −0.14% at 800 m. Therefore, the OCC – UCAR bias has significantly decreased as compared to the previous version of UCAR data products.

Different data processing setups may result in biases of the order of 0.5%–1%, depending on filtering, cutoff schemes, and whether or not radio holographic filtering is used. These effects constitute the structural uncertainty of RO inversions caused by our ignorance of the statistical properties of the measurement noise and small-scale atmospheric structures (Sokolovskiy et al. 2009a; Sokolovskiy et al. 2010). Figure 5 presents an OCC – UCAR comparison for OCC runs with modified options. 1) We use the cutoff of the measurement records at the impact height of about −150 km, 2) we also use the RH filtering for the CT2 phase, and 3) we use the ECMWF initialization. The cutoff and RH options result in a negative bias below 6 km, which is in a good quantitative agreement with Sokolovskiy et al. (2009a) and Sokolovskiy et al. (2010). A comparison of Figs. 5 and 4 shows that the effect of different initialization is only observed above 35–40 km. The contribution of different algorithms for handling multipath must be negligible. The FSI–CT2–PM difference (even in the worst-case scenario) should be smaller than the observed OCC – UCAR difference, as follows from the estimates of the FSI and CT2 errors (Gorbunov and Lauritsen 2004).

Figure 6 shows the systematic OCC – ECMWF differences for 2007, 2008, and 2009, which were computed for the same sets of events as the OCC – UCAR statistics.
The largest biases occur below 4 km, with the dominant feature being a strongly negative bias of OCC with respect to ECMWF. This is commonly referred to as the “negative refractivity bias” and is a current topic of research. The negative bias has a peak of ~1.0% globally at the height of 500 m. It is always more pronounced in the tropics, peaking at ~2.5%, than at higher (and drier) latitudes. Also evident is a change in the OCC – ECMWF bias structure that occurs in late 2007. The OCC – UCAR difference shows a pattern that is independent of the variations of the OCC – ECMWF bias. After the change, the negative bias becomes weaker at 500 m and stronger at 1.5–3.5 km than before the change. The change is very likely associated with cycle 32r3 of ECMWF analysis (Bougeault 2008). In that cycle COSMIC RO was first assimilated down to the surface, and convection and entrainment physics were altered. The latter is probably of more importance that the former: RO data probably do not constrain the analysis strongly below 4 km because of the large observation error, and the enhanced entrainment before cycle 32r3 is expected to redistribute water vapor preferably at the top of the boundary layer than at the bottom. An enhancement of the refractivity at ~2 km would result.

Figure 7 shows the statistical comparison of raw bending angles obtained from COSMIC data by OCC and UCAR with raw bending angles obtained from ECMWF fields. The ECMWF raw bending angles were obtained by processing artificial radio occultation data as described above. Raw bending angles are noisy above 30 km, but they are independent from any a priori climatological information. Therefore, they may reveal systematic differences between COSMIC and ECMWF data above 30 km, if averaging is performed over a large enough data volume. In the 10–38-km height range we observe the systematic differences between COSMIC and ECMWF data above 30 km, if averaging is performed over a large enough data volume. In the 10–38-km height range we observe the systematic differences between COSMIC and ECMWF below 0.2%, except for polar region around 32 km, where the bias reaches 0.4%. Above 38 km, the bias increases up to values of 0.5%–1%.

The uncertainty of the bias is estimated as \( \sigma/\sqrt{N_d} \), where \( \sigma \) is the standard deviation and \( N_d \) is the number of data. Because the overall number of data is about
For the whole globe, the uncertainty of the bias is 0.006 \( N_d \), and for each latitude zone (for the tropics, middle, and polar latitudes) it is about 0.01 \( N_d \).

At 50 km, where the standard deviation reaches 20%, the uncertainty of the bias for the whole globe is about 0.12%, and for the individual latitude zones it is about 0.2%.

4. Conclusions

The comparison of OCC, UCAR, and ECMWF data shows that OCC and UCAR data are very consistent in the latitudes and height intervals where no multipath propagation is observed. This relates to both systematic and random differences. In the 2–30-km height interval the OCC – UCAR refractivity bias does not exceed 0.1%, in particular, in the 9–25-km interval it is below 0.03%. The systematic differences between OCC and UCAR on one side and ECMWF on the other side in this height range do not exceed 0.15%. The maximum systematic discrepancy between OCC and UCAR is below 2 km, where it is negative and reaches a value of −0.2%.

The largest systematic and random differences between OCC and UCAR on one side and ECMWF on the other side are observed in the troposphere in the tropics. These differences are known to be induced by the complicated structure of the fields of atmospheric parameters.

The overall systematic OCC – ECMWF difference is a combination of systematic errors of COSMIC data, biases of retrievals induced by the response of nonlinear processing algorithms to measurement noise, and systematic errors of ECMWF analyses. Our lack of knowledge of the statistical properties of noise, and, as a consequence, of the structure of different biases, results
in the uncertainty of the optimal choice of the processing parameters. Both the lowering of the cutoff height and the activation of radio holographic filtering, results in the suppression of some biases and the amplification of the others. This “structural uncertainty” of RO inversions is estimated at 0.5%–1%. The maximum systematic difference between OCC inversions and the latest version of UCAR data products is about 0.2%, which is below the uncertainty level.

The systematic OCC – ECMWF differences changed in late 2007 below 4 km, as indicated by the comparison of the data for 2007, 2008, and 2009. Before 2008, above the height of 1.4–1.7 km the biases are more positive, and below this height they are more negative. This discontinuity is likely a consequence of cycle 32r3 of the ECMWF analysis system, particularly the changes to convection and entrainment in the boundary layer.

We have developed an advanced and accurate processing system, suitable both for numerical weather prediction and climate monitoring. The accurate processing of the measurement data is achieved by the use of the well-established Fourier integral operator–based methods and filtering and cutoff schemes developed after analysis of many years of RO observations. Above 4-km altitude, biases in retrieved refractivity with respect to a weather analysis are of the order of 0.15%. The random error ranges from 0.3%–0.5% in the upper troposphere–lower stratosphere to about 2% below 4 km. The OCC retrieval system agrees very well with the UCAR system. The systematic differences between the two systems are below the limits of the structural uncertainty of RO inversions.

Acknowledgments. The authors acknowledge Taiwan’s National Space Organization (NSPO) and the University Corporation for Atmospheric Research (UCAR) for providing the COSMIC data and data products. The authors are grateful to Sergey Sokolovskiy (UCAR) for his valuable consultations upon processing COSMIC data. MEG and KBL acknowledge discussions with Hans Gleisner and Stig Syndergaard (Danish Meteorological Institute) and Barbara Pirscher, Johannes Fritzer, and Gottfried Kirchengast (Wegener Center). The authors acknowledge Sean Healy for the discussions about the ECMWF cycle changes. MEG and AVS are supported by Grant Ao, C. O., 2007: Effect of ducting on radio occultation measurements: An assessment based on high-resolution radiosonde soundings. Radio Sci., 42, RS2008, doi:10.1029/2006RS003485.


Bougeault, P., 2008: Our changing forecast model. ECMWF Newsletter, No. 114, ECMWF, Reading, United Kingdom, 1.


