Assessing the Impact of Simulated COSMIC GPS Radio Occultation Data on Weather Analysis over the Antarctic: A Case Study

L. CUCURULL AND Y.-H. KUO
COSMIC Project, University Corporation for Atmospheric Research, Boulder, Colorado

D. BARKER AND S. R. H. RIZVI
National Center for Atmospheric Research–University Corporation for Atmospheric Research, Boulder, Colorado

(Manuscript received 15 August 2005, in final form 3 February 2006)

ABSTRACT

The Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) mission was launched in April 2006. As part of its mission, COSMIC will provide approximately 2500–3000 global positioning system (GPS) radio occultation (RO) soundings per day distributed uniformly around the globe. In this study, a series of sensitivity experiments are conducted to assess the potential impact of COSMIC GPS RO data on the regional weather analysis over the Antarctic. Soundings of refractivity are assimilated into the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Meso-scale Model using its three-dimensional variational data assimilation system. First, the sensitivity of the analysis to the background error statistics and balance constraints is analyzed. Then the effects of the data distribution and the observational error of the simulated refractivity observations are examined. In this study, the simulated soundings are based on a realistic set of orbit parameters of the COSMIC constellation. Analysis of the assimilation results indicates the significant potential impact of COSMIC data on regional analyses over the Antarctic. In the one case studied here, the root-mean-square differences between the background and observed values are reduced by 12% in the horizontal wind component, 17% in the temperature variable, 8% in the specific humidity, and 22% in the pressure field when COSMIC GPS RO data are assimilated into the system by using a 6-h assimilation time window. These preliminary results suggest that COSMIC GPS RO data can have a significant impact on operational numerical weather analysis in the Antarctic.

1. Introduction

In contrast to the ground-based GPS receivers, which are located on the earth’s surface, GPS radio occultation (RO) data are provided by receivers on board low-earth orbiting (LEO) satellites (Ware et al. 1996). As the radio signals, transmitted by the GPS satellites, pass through the atmosphere, they are refracted due to the density gradients along the path. As a LEO satellite sets or rises behind the earth’s limb relative to the GPS satellite, the onboard GPS receiver takes measurements of the phase and amplitude of the GPS signals. These measurements can be used to derive vertical profiles of the bending angle with the precise knowledge of the positions and velocities of the GPS and LEO satellites. Under the assumption of local spherical symmetry, refractivity profiles can be derived from bending angle profiles through Abel inversion (Hajj et al. 1994). The derived refractivity soundings can be directly assimilated into an operational data assimilation system (Eyre 1994; Healy et al. 2005). Since RO soundings have an accuracy compatible with that of traditional radiosondes in terms of refractivity (Kursinski et al. 1996; Rocken et al. 1997; Kuo et al. 2004), it is expected that GPS RO data will have a significant impact on operational weather analysis and prediction (Zou et al. 1995; Kuo et al. 1998; Zou et al. 1999; Anthes et al. 2000; Kuo et al. 2000; Zou et al. 2000). Even though GPS RO does not provide direct information on u- and v-wind components—as opposed to traditional radiosondes—a dense network of GPS observations will give
some insight on the winds. This will happen because of mass–wind balance, in regions where the geostrophic equilibrium applies, and provided there are enough data. In April the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) mission launched six small satellites, each carrying a GPS occultation receiver. As part of its mission, COSMIC will provide approximately 2500–3000 RO soundings per day uniformly distributed around the globe in near-real time. These data will be available to the scientific community around 20 min after collection time. Since the data will be collected every 100 min, data will be available between 20 and 120 min after observation time.

The objective of this study is to assess the sensitivity of the three-dimensional variational data assimilation (3DVAR) system to the analysis of GPS RO observations in order to evaluate the potential impact of the COSMIC mission on weather analysis over the Antarctic. For this purpose, we conducted a series of sensitivity experiments where we simulated observations from a hypothetical satellite or network configuration. Using simulated observations instead of real data provides a clear analysis of the model response to the ingestion of a given observation (Wee and Kuo 2004).

The case used in this study is a cyclogenesis event that took place over the Antarctic during 13–14 October 1995. This is motivated by the fact that the RO observations are expected to have a significant impact over areas such as the Antarctic and Southern Ocean where the lack of traditional meteorological information results in inferior forecast skill as compared with midlatitudes over the Northern Hemisphere (Kuo et al. 2002). One key goal of COSMIC is to demonstrate improvements in the performance of NWP models with the use of GPS RO data, especially in polar and oceanic regions.

As a remark, it should be mentioned that regions such as the Antarctic provide a unique framework that prevents generalizing the results to other areas of the earth. First, the information contained in observations of refractivity in the polar regions relates mostly to air density (i.e., mass, and hence temperature if hydrostatic equilibrium is assumed). Although there is also a bit of information in the humidity of midlatitude summer and tropical regions, the information content in the humidity is difficult to capture in polar regions, where the fractional contribution of water vapor to the total refractivity is small. Consequently, assimilating refractivity in polar regions brings in (mostly) mass–temperature information in a data assimilation system, without suffering from the ambiguity that would arise if the same observations were assimilated in other regions of the globe where water vapor becomes a key player with regard to its contribution of refractivity. Second, the mass–wind balance is easily modeled and handled by today’s data assimilation systems in these regions. Assimilating mass information in these regions amounts then to assimilating information about the atmospheric circulation (winds), which is easier to capture and to satisfactorily show results in an experiment than the more complex and nonlinear mechanisms such as those related to the proper assimilation of humidity information (consequences on convection/rain being difficult to apprehend) or assimilation of mass information only in regions not controlled by the mass–wind balance (i.e., Tropics).

The structure of the paper is as follows. Section 2 describes the algorithm used for the assimilation of the observations and the different series of experiments. Results on the sensitivity of the model analysis to variations of a hypothetical network of refractivity profiles are given in section 3. Section 4 analyzes the results from an experiment with a GPS RO distribution derived from a realistic configuration of the COSMIC mission. Finally, the main conclusions are presented in section 5.

2. Methodology

a. Three-dimensional variational assimilation system

The assimilation system is based on the three-dimensional variational data assimilation (3DVAR) algorithm in an incremental formulation (Courtier et al. 1994). Such an algorithm has been developed for the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5) in recent years (Barker et al. 2004). Briefly described, this is a model-space-based multivariate incremental analysis system capable of assimilating a wide range of conventional and remote sensing data. The cost function includes a background and an observational term. Following Lorenc et al. (2000), the background error (BE) covariance matrix is designed so as to project onto vertical modes allowing the horizontal length scales to be dependent on the vertical mode. The vertical modes are obtained from the decomposition in EOFs of statistical model forecast error profiles. These profiles were generated by application of the National Meteorological Center (NMC) method (Parrish and Derber 1992).

A weak constraint is applied to the analysis through the choice of the analysis or control variables. In our application, the MM5 model variables—wind, pressure, temperature, and water vapor mixing ratio—are de-
rived from streamfunction, velocity potential, unbalanced pressure, and relative humidity (or specific humidity) increments. This choice of control variables follows Lorenc et al. (2000) and was motivated by their relative independence, so that error correlations between analysis variables can be neglected in the background statistics. The square root of the BE covariance matrix is used as preconditioning.

Even if there are already cases where other satellite meteorological observations of the polar regions, we have not considered them in our study as our focus was on evaluating the impact of the GPS RO alone.

Assimilation of refractivity soundings $N$ is performed by the addition of a new term in the cost function to the already existing background ($J_b$) term:

$$J(x') = J_b + J_N.$$  (1)

This new term $J_N$ is defined as

$$J_N(x') = \frac{1}{2} (y'' - Hx')^T R^{-1} (y'' - Hx'),$$  (2)

where $x'$ is the vector of the analysis increments defined by $x' = x^b + x'$ and $y''$ is the refractivity observation increment ($y'' = H[x^o]$). The vector $x^b$ is the background state (first guess) vector, $y^o$ is the vector of the observations of refractivity, $x^o$ is the sought analysis, and $R$ is the covariance matrix of the refractivity observation errors. Since observational errors are assumed to be uncorrelated, the matrix $R$ is simply the diagonal with the refractivity observational error variances as elements. The diagonal specification of the observation error covariance matrix is a strong constraint in our assimilation system. Consequently, we have assumed the observational error covariance matrix to be diagonal (see section 3c). Here, $H$ is the linear approximation (operator) of the nonlinear operator $H$ used in the calculation of the observation increments, and it maps the model variable increments to refractivity at the location of the observations and includes both variable transformation and spatial interpolation. The nonlinear observational operator is the model simulation of the refractivity sounding and is calculated based on the profiles of pressure, temperature, and water vapor partial pressure (Smith and Weintraub 1953):

$$N = 77.6 \left( \frac{P}{T} \right) + 3.73 \times 10^5 \left( \frac{P_v}{T^2} \right),$$  (3)

where $P$ is the total atmospheric pressure (mb), $T$ is the atmospheric temperature (K), and $P_v$ is the partial pressure of water vapor (mb). In most contexts the first term in (3) is considerably larger than the second. (The wet term becomes important in the troposphere for temperatures higher than 240 K and contributes up to 30% of the total refractivity in the tropical boundary layer.) There are two different approaches to conducting the space interpolation. The first one is to interpolate the atmospheric profiles of temperature, pressure, and water vapor pressure from the model grid points to the location of the observation and then evaluate the refractivity. The second approach is to estimate the refractivity soundings at the model grid space and then interpolate them to the location of the radio occultation profile. Since calculating the refractivity from interpolated values makes the tangent linear model more complex, we adopted the second approach.

The 3DVAR solution $x'$ is obtained for the analysis increment $x'$ that minimizes the total cost function. It is, therefore, the model space vector that best fits—in a least squares sense—simultaneously the background vector and the observations of the refractivity. This fit is measured by the quadratic distance weighted by the background and observational error covariance matrices. The limited-memory quasi-Newtonian method is used to solve the minimization of the cost function (Liu and Nocedal 1989).

b. Experiment design

In this study, we focus on a cyclogenesis event that took place over the Ross Sea between 13 and 16 October 1995. During this period, several cyclones developed over the area.

Profiles of the refractivity were generated from the high-resolution simulation of the cyclogenesis event using the MM5 model (Dudhia 1993), usually known as the NATURE run. The goal of the NATURE run is to realistically simulate the true atmosphere, so that the simulated refractivity profiles would have characteristics similar to the actual occultation sounding profiles. The NATURE run was initialized at 0000 UTC 13 October 1995 and was integrated for 72 h. The grid resolution was 30 km with 50 sigma levels in the vertical and the model was run with sophisticated physics packages. The initial and boundary conditions were provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) global analysis. We then extracted all four of the points in both latitude and longitude from the NATURE run (effectively reducing the resolution from 30 to 120 km) to generate the simulated refractivity profiles. The vertical resolution of the profiles was reduced from 50 to 20 sigma levels. The GPS refractivity soundings from this set then serve as the truth from which the simulated observations are derived for our studies. The assimilation domain is prescribed with a $69 \times 69$ horizontal grid (latitude $\times$ longitude) with 120-km horizontal resolution and 20 sigma
levels in the vertical. All the assimilation runs start at 0000 UTC 14 October 1995. This allows the NATURE run to develop realistic mesoscale atmospheric structures (after 24 h of integration), before it is used to generate simulated observations.

To initialize our CONTROL experiment, we first performed a 12-h forecast run starting at 1200 UTC 13 October 1995 using the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP-NCAR) reanalysis. In this case, the grid resolution was lower than the one used for the NATURE run, with 20 levels in the vertical and a grid spacing of 120 km. A simpler physics package was used for the CONTROL. For a detailed description of the NATURE and CONTROL runs used here, please refer to Wee and Kuo (2004).

We conducted a series of analysis experiments to investigate the sensitivity of the results to the following aspects of the assimilation system: background statistics, balance constraints, measurement error, and data density.

1) BACKGROUND STATISTICS

We tested the role of the BE covariance matrix used in the 3DVAR algorithm by using two different sets of BE statistics. The first set of BE statistics was derived from one month (May) of real-time MM5 forecasts run at NCAR on a 101 × 181 horizontal grid (latitude × longitude) with 210-km horizontal resolution and at 21 vertical sigma levels. These default (DEFAULT) global statistics are supplied as a starting point for users to learn how to run the MM5 3DVAR system. The intention is that users would then calculate and tune “optimal” BE covariances for their particular domain. The BE statistics are calculated according to the NMC method, that is, accumulated differences between 24 – 12 h forecasts are first converted to “control variable space” and averaged in time and space to produce a climatological estimate of forecast error covariances.

In the MM5 3DVAR system, before projection onto the vertical error modes (computed in the NMC method calculation), the three-dimensional fields are normalized by the square root of the eigenvalue of the relevant vertical mode. These normalized fields are passed through a series of recursive filters, which creates the smoothing effect of a convolution with a covariance matrix. In the early “beta release” of the MM5 3DVAR used here, a first-order (exponential smoother) filter is repeatedly applied (Hayden and Purser 1995). The basic assumption under the application of the filter is that horizontal model forecast error correlations are homogeneous and isotropic.

In addition, a second statistics matrix (PERF) is computed, representing a smoothed “perfect” error covariance matrix derived from the difference between the CONTROL and NATURE (i.e., the “truth”) fields. In reality, the BE covariance matrix can never be evaluated exactly because the true state of the atmosphere is unknown. Even given the true forecast error, approximations to the BE covariance matrix are necessary in order to render the 3DVAR solution computationally feasible, for example, neglecting error correlations between control variables, domain-averaged eigenvectors, and isotropic/homogeneous horizontal error correlations. The impact of the different BE covariance matrices is analyzed by assimilating “perfect” (i.e., zero error) profiles of refractivity at all grid points. In another set of experiments, observational errors are added to these perfect profiles.

2) BALANCE CONSTRAINTS

Atmospheric fields obey dynamic and thermodynamic relationships and thus they do not evolve independently (i.e., their errors are correlated). Consequently, it is expected that the assimilation of soundings of refractivity (related to temperature, pressure, and humidity) will also have an effect on other model variables. As a second series of experiments, the increments of the analysis variables are evaluated through the balance constraints specified in the BE covariance matrix.

3) MEASUREMENT ERRORS

The observations are by no means perfect. Each measurement is subject to two types of errors. One is the instrument error and the other is the representativeness error with respect to the atmospheric state (or conditions). The former depends on the instrument performance and the ambient conditions in which the instrument operates. The latter is more difficult to quantify and is related to errors in the observation operator (3) and to the differences in the scales of the atmospheric circulation systems measured by the instrument and the observing network, and those resolvable by the gridpoint analysis. The observational errors and their statistics must be properly taken into consideration prior to the assimilation of the data. For the third set of the experiments, we assess the impact of the observational errors on the results of the analysis studies.

4) DATA DENSITY

Finally, we evaluate the impact of the data density by assimilating a reduced set of refractivity observations.
distributed over the domain. In particular, we would like to assess the potential impact of the COSMIC constellation. For this purpose, we distribute occultation soundings based on a realistic set of orbit parameters of the COSMIC constellation.

A summary of the different data analysis experiments conducted in this study is given in Table 1: CONTROL (no data are assimilated), DEFAULT (accurate observations of refractivity are assimilated by using DEFAULT BE), PERF (accurate observations of refractivity are assimilated by using PERF BE), DEFAULTE (observations of refractivity with their corresponding errors are assimilated by using DEFAULT BE), PERFE (observations of refractivity with their corresponding errors are assimilated by using PERF BE), and COSME (observations of the refractivity at the location of the COSMIC measurements by using DEFAULT BE).

3. Results

a. Impact of the background errors covariance matrix

To study the impact of the different background error statistics on the analysis results, we first conducted a single observation test. In this experiment one single observation of the refractivity located at 300 mb (vertical level 13), and with a value of 115.45 units of $N$ was assimilated into the system. The corresponding background value of the refractivity at the location of the observation was 114.45 units of $N$. Such an experiment allows us to analyze the radius of influence of a single piece of observation. That is, if the observed value of the refractivity and its corresponding background estimate differ by a unit of refractivity at the location of the observation, the single observation test indicates how far from the station the analysis is going to be influenced by the assimilation of that observation. In the PERF and DEFAULT experiments, the length scale of the BE covariance matrix was empirically tuned to better represent the covariance structure over the area of interest (see Hollingsworth and Lonnberg 1986). In the PERF experiment, the length scale was reduced by 50% while a larger reduction value (75%) was used in DEFAULT. These values were chosen in order to reduce the length scale of the covariance matrix to a more realistic structure (Zou et al. 1999). As PERF is actually an approximation to the truth, a tuning factor was also necessary in this case.

Figure 1 shows the temperature increment (analysis – CONTROL) of the single observation test for the PERF and DEFAULT experiments. Both the horizontal structures at 300 mb (PERF, Fig. 1a; DEFAULT, Fig. 1c) and the vertical scales (PERF, Fig. 1b; DEFAULT, Fig. 1d) for temperature increments are represented. From the figure, the radius of influence of a single observation of the refractivity presents a similar structure in the horizontal direction for both BE covariance matrices tested in this study as they have been tuned to do so. Experiment PERF shows a larger increment around the location of the observation (approximately $-0.6$ K) when comparing with DEFAULT, which shows more uniform values of the temperature increments (between $-0.1$ and $-0.3$ K). These results are consistent with the background errors associated with the different experiments. The influence of the assimilation of the single observation shows a larger vertical influence in PERF. In this case, the increment of the temperature varies between $-0.6$ and $-0.1$ K (between 400 and 200 mb). The increment of the temperature around 300 mb is also found in DEFAULT. However, the peak of $-0.6$ K surrounding the observation in PERF is not observed in the other case, which shows a smoother structure.

During the assimilation, 3DVAR uses observational information to correct for model deficiencies (charac-
Fig. 1. Increments of the temperature variable as a result of the assimilation of a single observation of refractivity at 300 mb. The horizontal increment at 300 mb is shown for the different background error covariance matrices: (a) PERF and (c) DEFAULT. Vertical cross sections as a function of the pressure level at 85°N are shown for (b) PERF and (d) DEFAULT. The surface is indicated in the figures as a solid line.

We next investigate the assimilation of a larger number of observations. We assume that our measurements are very accurate [root-mean-square (rms) error of 0.01 units of refractivity] and are collocated at every grid point. A total amount of 95 220 observations of refractivity are assimilated into the system. In both cases the cost function converges to the same value at the end of the minimization (not shown), which indicates that...
when having very small observation errors and high density data, $J$ is dominated by $J_S$ and the background term has less influence. The mean difference and the rms error in terms of refractivity for the two experiments are shown in Table 2. The convergence is achieved when the gradient of the cost function is reduced by three orders of magnitude. After the assimilation of the very accurate observations, the rms error of the analysis decreases from 2.70 units of refractivity (1.34% difference) in CONTROL down to 0.24 (0.43% difference) in PERF and 0.26 (0.44% difference) in DEFAULT.

In both the PERF and DEFAULT experiments, the analysis is unbiased in terms of refractivity, which indicates the good behavior of the assimilation system as the observations are not biased. Before the assimilation of the observations, the mean difference was 0.88 units of refractivity (0.38% difference).

b. Increments of analysis variables

The refractivity is strongly related to the temperature ($T$), pressure ($P$), and humidity ($Q$) fields and consequently its assimilation into an analysis system should not only modify the analyzed field of the refractivity but also the other related variables—$T$, $P$ and $Q$—through the adjoint of the observation operator. In addition, the variables that describe the state of the atmosphere satisfy exact or approximate physical relationships or constraints. Thus, the state variables to be analyzed are interrelated and cannot evolve completely independently. As a consequence, the other analyzed atmospheric variables, which are not directly related to the refractivity (such as the wind field) should in turn be updated due to the existing balance constraints. These interrelations satisfied by the state variables are specified in the background error statistics. In the following, the impact of the assimilation of the refractivity in terms of the wind, temperature, pressure, and humidity is investigated.

In the previous section it was found that the assimilation of very accurate observations of the refractivity resulted in a significant improvement (~90%) of the analyzed field of the refractivity over the grid. Improvements are also found for the other variables after the assimilation of the refractivity (Table 2). Such improvement though is found to be much smaller than those of the refractivity field. For instance, the horizontal wind improves in both cases, as indicated in Table 2. As expected, the larger improvement is found in PERF (19%) and slightly less is found in DEFAULT (16%). In both cases, the mean difference remain almost the same after the assimilation of the observations of refractivity. As we can observe in Table 2, the rms error of the temperature field is significantly reduced in PERF (26%) and a smaller improvement (17%) is found for DEFAULT. The mean difference in temperature is completely removed in PERF after the assimilation of refractivity. An overall improvement is also found for the moisture and pressure fields. When analyzing the humidity, the mean difference is completely removed after the assimilation and this is independent of the BE covariance matrix used. Once again, the larger improvement in terms of rms if found in
PERF (37%) and a slightly smaller value (26%) is found for DEFAULT. The assimilation of the refractivity also improves the pressure field. As expected, the improvement is found to be larger in PERF (47%) and slightly less in DEFAULT (39%). In both experiments, the mean difference is also reduced after the assimilation, with the smaller value found in PERF.

In summary, when a large number of very accurate observations of the refractivity are assimilated into the analysis system, an overall improvement is found not only in the refractivity field, but also in all of the atmospheric variables. The improvements are significant for variables that are directly related to refractivity (temperature, moisture, and pressure). Even though wind fields are not directly related to refractivity, some moderate improvements are found. As one could expect, the improvement is found to be larger when the most accurate background statistics is used in the variational algorithm. The fact that the BE statistics used in PERF contains a lot of approximations to the real description of the atmosphere explains why the “true” atmospheric state is not totally recovered in PERF. It is important to note that Table 2 shows the relative size of the error reduction due to using the correct statistics (PERF versus DEFAULT) versus the error still remaining due to the various approximations employed in our BE mode as discussed in the previous section (i.e., the refractivity fit is nearly perfect, but the model fit—due purely to BE statistics—is still far from optimal).

Figure 4 shows the mean error and the standard deviation of the different state variables for the different experiments (CONTROL, PERF, and DEFAULT) as a function of the sigma vertical level. The horizontal winds show an overall improvement at all levels with respect to the standard deviation error of the CONTROL (continuous line). The improvement is particularly worth noting at the jet stream level. For instance, at vertical level 14, the error of the horizontal u wind (v wind) is reduced from around 8.0 (7.8) m s$^{-1}$ before the assimilation down to around 5.6 (6.2) m s$^{-1}$ in DEFAULT (dashed line), and 5.6 (5.8) m s$^{-1}$ in PERF (dotted–dashed line). As expected, experiment PERF in general performs better than the DEFAULT experiment, which uses less accurate (and more realistic) background statistics. Looking at the standard deviation error, DEFAULT seems to perform slightly better than PERF only between levels 9 and 13. In general, the bias with respect to the NATURE run is not affected by the assimilation process (a small reduction of the mean error is only slightly apparent in the v-wind component in DEFAULT).

The improvement in the wind field in DEFAULT after refractivity assimilation is in general close to that of PERF except at the highest vertical levels. At these levels, only PERF seems to slightly reduce the error of the wind field. The use of the DEFAULT background statistics results in an increase of the error. For instance, the total rms error of the u (v) wind at the top of the model in DEFAULT increases from 1.8 (1.6) m s$^{-1}$ prior to the assimilation to 3.2 (3.6) m s$^{-1}$ after the assimilation of the refractivity. Similar results are found for the pressure variable. Errors in the upper boundary condition can create bad MM5 forecasts that are represented in the DEFAULT statistics (H. Wei et al. 2004, unpublished manuscript). The 3DVAR scheme is therefore trying to correct these perceived errors (which may not exist in the true error). This is an example of problems being caused by using statistics from one application being applied to another. Further evidence for this is seen in the structure of the eigenvectors for the wind components (not shown), where there is a large nonzero eigenvector at the top in the DEFAULT, but not in the PERF, experiments. On the other hand, PERF has no such problem.

When comparing with the CONTROL pressure field, the variational assimilation produces an improvement of the analyzed pressure field. This improvement is significant with the use of both BE statistics although it is found to be larger in PERF (the reduction of the standard deviation error in PERF versus DEFAULT is significant along the whole profile). In this case, the reduction of the total rms error is achieved at all sigma levels (the slightly increase of the bias when compared to DEFAULT at the lower sigma levels is compensated by the reduction of the variance at these levels). As found when analyzing the horizontal wind, there is a slight increase in the error at the upper levels in DEFAULT. Figure 4 also shows an improvement in the humidity field. Both the mean difference and the standard deviation error are reduced with the assimilation algorithm.

### Table 2. Mean difference and rms error (in parentheses) before and after the assimilation of very accurate observations of the refractivity for the different experiments.

<table>
<thead>
<tr>
<th>Expt</th>
<th>N</th>
<th>U (m s$^{-1}$)</th>
<th>V (m s$^{-1}$)</th>
<th>T (K)</th>
<th>Q (g kg$^{-1}$)</th>
<th>P (mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTROL</td>
<td>0.88 (2.70)</td>
<td>−0.15 (5.33)</td>
<td>0.53 (5.74)</td>
<td>−0.21 (2.29)</td>
<td>0.07 (0.38)</td>
<td>0.46 (3.13)</td>
</tr>
<tr>
<td>PERF</td>
<td>0.00 (0.24)</td>
<td>−0.16 (4.39)</td>
<td>0.54 (4.53)</td>
<td>0.00 (1.69)</td>
<td>0.00 (0.24)</td>
<td>0.16 (1.65)</td>
</tr>
<tr>
<td>DEFAULT</td>
<td>0.00 (0.26)</td>
<td>−0.13 (4.60)</td>
<td>0.55 (4.71)</td>
<td>0.16 (1.89)</td>
<td>0.00 (0.28)</td>
<td>0.19 (1.91)</td>
</tr>
</tbody>
</table>
As the radio occultation observations are expected to have an important impact in terms of temperature (due to a direct relationship between $N$ and $T$) we now analyze with more detail the impact of the assimilation of the refractivity on the temperature analysis. As found for the pressure and horizontal wind fields, the assimilation of a large number of accurate observations of refractivity results in an improvement to the temperature profile at all vertical levels, except at the upper levels, which, as discussed above, may be related to the model-top boundary condition error in DEFAULT and to a regression coefficient problem at the top of the model in PERF as we only used one day to estimate these coefficients. As shown in Fig. 4, both the mean difference and standard deviation error are reduced after the assimilation. In general, the improvement, in terms of rms, is found to be larger in PERF, but DEFAULT performs slightly better between levels 4 and 8. The moisture also does slightly better with DEFAULT in these levels. The large standard deviations corresponding to the background error for the temperature in PERF (Fig. 2) will result in a significant projection of the observational increments onto this variable. Looking at the temperature standard deviation and covariance structure for the background error in Figs. 2 and 3 (covariance shown for vertical level 13),

![Fig. 4. Mean error and standard deviation for the $u$ wind ($m\,s^{-1}$), $v$ wind ($m\,s^{-1}$), pressure (mb), humidity (g kg$^{-1}$), and temperature (K) profiles for the different experiments: PERF (dotted–dashed) and DEFAULT (dashed). The error for the CONTROL run is shown as a continuous line.](image-url)
and at the observational operator in (3), it is expected that the increments of the temperature will be very significant. This is noticeable in Fig. 4, as both the bias and standard deviation are reduced with respect to the CONTROL. A slight deterioration is found in both PERF and DEFAULT at the very lowest sigma levels. This behavior is not found when analyzing the other variables and might be related to a regression coefficient problem.

Even though both of the BE covariance matrices result in a similar reduction of the total rms error of the temperature field in most vertical levels, some differences can be observed when analyzing the horizontal structures of these values at a given vertical level. For instance, at a pressure level of 200 mb, PERF and DEFAULT present a similar temperature rms error of 1.2 K. This value accounts for an improvement of 50% as compared with the error from the CONTROL experiment. However, different structures can be observed in the horizontal when dealing with the different covariance matrices. This can be observed in Fig. 5, where the temperature differences of the CONTROL, DEFAULT, and PERF experiments with respect to the NATURE run are represented at a pressure level of 200 mb. The different structures between the CONTROL and the “true” temperature before the assimilation of the observations of the refractivity are shown in Fig. 5a. Several areas over the domain present absolute temperature differences greater than 4 K, and both positive and negative regions can be observed. Some of these differences are reduced in DEFAULT (Fig. 5b), and the increment field presents a smoother structure (the absolute different is now less than 4 K). See, for instance, the reduction of the absolute bias in the eastern part of the domain. Although the overall field improvement in DEFAULT and PERF is the same in terms of rms, PERF (Fig. 5c) presents a different horizontal structure. Some of the positive bias

![Figure 5](image-url)
found in the northern part of the domain in DEFAULTE has been significantly reduced. However, a more pronounced negative bias in the southeast and a positive bias area in the central-eastern part of the region are still present.

c. Impact of the observational error

In the previous sections, observations of the refractivity have been assumed to be very accurate. However, in practice, observations are never perfect. To investigate the validity of our previous results in a more realistic situation, the results from Kuo et al. (2004) have been adopted in this study. Based on the Hollingsworth and Lonnberg (1986) method, these authors found that the radio occultation observational errors at the mid-latitudes are nearly constant from 500 to 30 mb with a value of 0.3%. They increase from 0.3% at 30 mb to 1.2% at 10 mb, and from 0.3% at 500 mb to about 0.75% near the surface. A larger value of about 3% at the surface was found over the Tropics. The errors are assumed to be random and uncorrelated with the model variables. A Gaussian distribution for the observational error is assumed. The impact of the distribution of the observations is left to the following section, and we assimilate here observations located at all grid points (at 120-km resolution on regular intervals).

The minimization was stopped when the norm of the gradient was reduced by two orders of magnitude. To evaluate the impact of the observational error in the default BE covariance matrix relative to the perfect one, we conducted the PERFE and DEFAULTE experiments (see Table 1).

The minimization of the cost function in experiments PERFE and DEFAULTE gives similar performance and it takes around 35 iterations to converge. Table 3 summarizes the mean difference and the rms error for the different variables and experiments. The assimilation of the less accurate observations of the refractivity still results in an improvement of all meteorological fields when comparing with the error associated with the CONTROL. As one would expect, this improvement is not as significant as was found with the use of very accurate observations. In PERFE, the temperature improves by 22% (26% in PERF), humidity by 29% (37% in PERF), pressure variable by 27% (47% in PERF), and the horizontal wind field roughly improves by the same amount as before (17% instead of 19%). In DEFAULTE, the pressure, temperature, and horizontal wind fields, on average, present similar improvements as found with the use of very accurate soundings of refractivity. The inclusion of observation errors in DEFAULTE only results in slight degradations of the analyses results. This is due to the fact that the observational error of the refractivity is still much smaller than its background error counterpart. If high-density soundings are available over the domain, the prescription of a realistic error to the observations of refractivity only slightly modifies the results due to the lower accuracy of the background field when simulating the refractivity soundings.

A close fitting to the observed refractivity field after the minimization would impact the accuracy achieved for the other atmospheric variables. It is not always better to demand a very close fit of the analysis to the observations of a variable, which are being assimilated into the system (soundings of the refractivity in this case) as this can deteriorate the other analyzed variables. If there is an overfit of noisy observations, then we will degrade the whole analysis, including that of the observed variable. For instance, from our results, we found that demanding the analyzed refractivity field to closely match the refractivity observation resulted in a deterioration of the other state variables.

4. Impact of the COSMIC mission

In the previous section, we have analyzed different aspects of the analysis of a large number of observations (one per grid point). However, this is never the case for a real GPS radio occultation constellation, where refractivity observations are not as dense and are unevenly distributed in time and space. To quantify the potential impact of COSMIC, we analyze in this section the impact of the assimilation of profiles of the refractivity based on a realistic set of orbit parameters of the COSMIC constellation. Two different time windows (3 and 6 h) for the assimilation of the observations at the analysis time are considered. This accounts for 1.5% (3.0%) of the total number of observations used in DEFAULTE when a 3-h (6 h) assimilation time window is used. In COSME (see Table 1), the DEFAULT
BE covariance is used and the observations are assimilated with its corresponding error (Kuo et al. 2004; see section 3c).

A total of 72 (141) profiles of refractivity were assimilated into the objective analysis in a 3-h (6 h) time window starting at 0000 UTC 14 October 1995. (Here, the difference between a 3- and a 6-h time window resides in the number of observations assimilated in the system only, as the early version of MM5 3DVAR system used here does not allow any time distribution of the observations around the analysis date. That is, all the observations are assimilated at one time.) After the assimilation of the observations in a 3-h (6 h) window, the rms error of the refractivity was decreased by 8% (14%) and in both cases the bias between the analysis and the observations was also reduced.

As we saw in the previous sections, the model variables have been modified after the assimilation of the observations of the refractivity. Table 4 shows the mean difference and rms error of the analyzed fields in terms of refractivity, wind, temperature, humidity, and pressure for the different assimilation time windows. After the assimilation in COSME for 3 h (6 h), the rms of the different variables has been reduced by 7% (10%) for the u-wind component, 10% (14%) for the w-wind component, 11% (17%) for the temperature variable, 5% (8%) for the specific humidity, and 17% (22%) for the pressure field. As expected, results are slightly better with a wider time window (i.e., with the assimilation of a larger number of observations). Comparing the results from COSME and DEFAULTE (Table 3), we observe that when a lower number of observations is assimilated into the system (2820 observations in COSME at 6 h and 1440 in COSME at 3 h, versus 95220 in DEFAULTE), the impact of the assimilation is slightly reduced for wind (8% for the u- and w-wind components in COSME at 3 h; 4% for the u- and w-wind components in COSME at 6 h), and significantly reduced for pressure (28% in COSME at 3 h; 23% in COSME at 6 h) and specific humidity (36% in COSME at 3 h; 34% in COSME at 6 h). It is noticeable that the temperature field has roughly the same average rms error in both DEFAULTE and COSME at 6 h and that this value is only lightly degraded in COSME at 3 h (5%). The fact that the atmospheric fields are not largely degraded when a lower number of observations is used might be due to excessively large-scale background errors (and consequently the high-resolution information is smoothed out) and also that the meteorological situation under study is governed by large-scale atmospheric phenomena.

Table 4. Mean difference and rms error (in parentheses) after the assimilation of observations of the refractivity at the COSMIC location. Observations are assimilated with their corresponding errors.

<table>
<thead>
<tr>
<th>Expt</th>
<th>N</th>
<th>U (m s⁻¹)</th>
<th>V (m s⁻¹)</th>
<th>T (K)</th>
<th>Q (g kg⁻¹)</th>
<th>P (mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTROL</td>
<td>0.88 (2.70)</td>
<td>−0.15 (5.33)</td>
<td>0.53 (5.74)</td>
<td>−0.21 (2.29)</td>
<td>0.07 (0.38)</td>
<td>0.46 (3.13)</td>
</tr>
<tr>
<td>COSME at 3 h</td>
<td>−0.30 (2.47)</td>
<td>−0.14 (4.97)</td>
<td>0.53 (5.15)</td>
<td>−0.12 (2.03)</td>
<td>0.05 (0.36)</td>
<td>0.49 (2.60)</td>
</tr>
<tr>
<td>COSME at 6 h</td>
<td>−0.35 (2.32)</td>
<td>−0.13 (4.77)</td>
<td>0.54 (4.94)</td>
<td>−0.04 (1.90)</td>
<td>0.05 (0.35)</td>
<td>0.55 (2.43)</td>
</tr>
</tbody>
</table>

Figure 6 shows the bias and standard deviation for the errors of the different atmospheric variables in COSME at 3 and 6 h. The benefits of the assimilation of the radio occultation observations are notable at all vertical levels for the different analysis variables as compared to the CONTROL experiment (continuous line). As one would expect, the reduction of the rms error is more significant in COSME at 6 h (dotted-dashed line) than in COSME at 3 h (dashed line). However, as a lower number of observations with its corresponding error are used in the COSME experiments, the results are not as good as those found in the DEFAULTE experiment (Fig. 4). The fact that the three experiments (DEFAULTE, COSME at 3 h, and COSME at 6 h) present a similar error structure in the vertical is due to the fact that the different assimilation experiments use the same BE covariance matrix.

The use of a denser observation network should provide better information on the horizontal gradients and consequently a better constraint on the winds through the mass–wind balance. To investigate if some information is lost when a lower number of observations is being assimilated, Table 5 shows the mean difference and the rms error for COSME at 3 h and COSME at 6 h but computed only at the grid points where the observations are assimilated. As compared to DEFAULTE, the winds seem to slightly degrade in COSME at 6 h and a larger rms error if found in COSME at 3 h. Although the errors for temperature and humidity are similar between DEFAULTE and the COSME experiments, the pressure variable (and consequently the wind information) is not fully recovered when a lower number of observations is assimilated. This indicates that some horizontal structure is lost in a less dense network.

5. Summary and conclusions

To get ready for the arrival of the radio occultation data from the recently launched COSMIC project, we
have studied in this paper different aspects of the assimilation algorithm that need to be considered when assimilating radio occultation soundings into an NWP system. We have shown that the results of the assimilation are sensitive to the background error statistics. As already pointed out by other authors, an inadequate estimation of the background error covariances can deteriorate the accuracy of the analyzed variables. In this context, it is also necessary to specify accurate observation errors in order to extract the maximum information from the radio occultation data without degrading the other atmospheric fields.

When high-density observations are available, the benefits of the assimilation of the soundings of the refractivity are preserved, even when these measurements are prescribed with realistic measurement errors. This is due to the fact that higher density has the effect of reducing the effective observation error and the analysis is always fitting the observations very closely. In the one case studied here, we found an improvement
of around 17% in the horizontal wind component, 16% in the temperature, 39% in the specific humidity, and 40% in the pressure field. The results found in this work indicate that the COSMIC constellation could potentially lead to a significant improvement in weather analysis over the Antarctic. When GPS radio occultation observations are simulated using a realistic satellite constellation configuration of COSMIC, we find that the rms differences between the background and observed values are reduced by 8% (12%) in the horizontal wind component, 11% (17%) in the temperature, 5% (8%) in the specific humidity, and 17% (22%) in the pressure when 3 h (6 h) worth of COSMIC data are assimilated.

Acknowledgments. We thank Tae-Kwon Wee for kindly providing several data files used in the experiments and anonymous reviewer A for his or her very constructive remarks. This work was funded by the Constellation Observing System for Meteorology, Ionosphere, and Climate (UCAR/COSMIC) project.

REFERENCES


