Ensemble Prediction of Atmospheric Refractivity Conditions for EM Propagation

QINGYUN ZHAO, TRACY HAACK, JUSTIN MCLAY, AND CAROLYN REYNOLDS

Marine Meteorology Division, Naval Research Laboratory, Monterey, California

(Manuscript received 7 January 2016, in final form 11 July 2016)

ABSTRACT

An ensemble forecast system has been developed at the Naval Research Laboratory to improve the analyses and forecasts of atmospheric refractivity for electromagnetic (EM) propagation with the intention of accounting for uncertainties in model forecast errors. Algorithms for a matrix of ensemble statistics have been developed to analyze the probability, location, intensity, and structure of ducting of various types. Major parameters of ducting layers and their ensemble statistics are calculated from the ensemble forecasts. Their relationships to the large-scale and mesoscale environment are also investigated. The Wallops Island field experiment from late April to early May 2000 is selected to evaluate the system. During the spring season, this coastal region maintains a strong sea surface temperature gradient between cold shelf waters and the warm Gulf Stream, where the boundaries between land, the coastal water, and the Gulf Stream have a strong influence on marine boundary layer structures and the formation of ducting layers. Sounding profiles during the field experiment are used in the study to further understand the structures of the ducting layers and also to validate the ensemble forecast system. While some advantages of the ensemble system over the deterministic forecast for atmospheric refractivity prediction in the boundary layer are studied and demonstrated in this study, the weaknesses of the current ensemble system are revealed for future improvement of the system.

1. Introduction

In the last couple of decades, efforts have been made to model the atmospheric refractivity and electromagnetic (EM) wave propagation conditions in the boundary layer using high-resolution numerical weather prediction (NWP) models (Burk and Thompson 1997; Thompson et al. 1997; Atkinson et al. 2001; Atkinson and Zhu 2005, 2006; Haack et al. 2010). While some successes have been achieved in capturing the main features—such as size, location, and internal structure—of the trapping layers and ducts in these model simulations, challenges are also apparent in modeling the magnitudes of duct depth and strength as well as their development. In addition to small grid spacing required to resolve the fine structures of moisture and temperature in the lower boundary layer, the study by Haack et al. (2010) also indicates that accurate and dynamically balanced initial fields, appropriate large-scale forcing, and the coupling of the atmospheric boundary layer and sea surface temperature (SST) are the critical aspects of the model system necessary for accurately forecasting the atmospheric refractivity and ducting. The requirements for the model system to accurately simulate the atmospheric EM propagation conditions are even higher in coastal regions where the land–sea boundary further complicates the dynamic and thermodynamic structures of the atmospheric boundary layer (Grisogono et al. 1998; Atkinson and Zhu 2006).

In the last 20 years or so, ensemble prediction systems have gained prominence and become operational at major NWP centers around the world (Palmer et al. 1992; Toth and Kalnay 1993; Tracton and Kalnay 1993; Du et al. 1997; Houtekamer et al. 1996; McLay et al. 2008; Reynolds et al. 2011). By generating a series of forecasts, the ensemble approach represents the atmospheric state in the form of probabilistic forecasts rather than a single deterministic forecast. This is a comprehensive undertaking that involves quantifying and sampling the error probability density functions (PDFs) of the atmospheric state in both the analyses and forecasts, evolving the samples within the NWP model to generate a forecast PDF, and processing the forecast PDF (McLay et al. 2008). This approach provides a means to quantify propagation of uncertainty in
forecasting in the form of probabilistic forecasts and probability density functions (Abaza et al. 2013). For atmospheric refractivity modeling, the ensemble approach looks attractive because of its ability to account for the uncertainties in the model's initial state and forecasts, in the large-scale forcing, and in the lower boundary conditions. It is known that, in the lowest part of the atmospheric boundary layer, interactions between the atmosphere and the surface play a critical role in defining atmospheric refractivity profiles that determine the atmospheric EM propagation conditions. This is especially true over the ocean surface, where SST and sea surface fluxes show significant impacts on the atmospheric boundary layer structures (Doyle and Warner 1993; Haack et al. 2010; Thompson and Haack 2011). The lack of observations of the atmospheric wind, temperature, and moisture profiles above the ocean surface with a very high vertical resolution of up to a few meters makes it difficult to accurately describe the structures of the lower boundary layer in the initial fields of a high-resolution NWP model for refractivity modeling. Inaccurate initial conditions plus model forecast errors make it very challenging to predict the occurrence, location, strength, and variation of trapping layers and ducts from a single deterministic forecast with an acceptable forecast accuracy. Ensemble forecasts, on the other hand, seek to represent the uncertainties in model initial state and large-scale forcing by appropriately perturbing the initial and lateral boundary conditions. The uncertainties in model forecast errors may be accounted for through multimodel ensembles, perturbed dynamics or physics, or stochastic forcing. Similarly, ensemble refractivity forecasts differ from deterministic forecasts in that they represent the occurrence, location, and strength of trapping layers and ducts in terms of probability with uncertainties in the forecasts quantified by ensemble spread, an ensemble parameter that measures the width of the forecast probability distribution.

The main purpose of this study is to explore and demonstrate the advantages of ensemble forecasts in atmospheric refractivity modeling. Specifically, we want to address two questions. First, how does the ensemble system perform in modeling the atmospheric boundary layer structures in comparison with the single deterministic forecast? Second, are the ensemble statistics from the probabilistic forecasts useful in predicting and understanding the formation and structures of the trapping layers and ducts in the atmospheric boundary layer? Do they offer any additional information? The second objective of this investigation is to reveal the possible weaknesses of the current ensemble system we use for atmospheric refractivity prediction and seek future improvements. To do this, the U.S. Navy’s Coupled Ocean–Atmosphere Mesoscale Prediction System (COAMPS; Hodur 1997) is employed in this study as the atmospheric forecast model in both the ensemble and deterministic predictions of the atmospheric refractivity. The COAMPS-based ensemble Kalman filter (EnKF) recently developed at the Navy Research Laboratory (NRL) is used as the data assimilation system for initializing the COAMPS ensemble (Zhao et al. 2013). Algorithms for a matrix of ensemble statistics to analyze the possibility, location, intensity, and structure of ducting layers of various types from the ensemble forecasts have also been developed. The Wallops Island field experiment from late April to early May 2000 (Haack et al. 2010) is selected to perform the study. Vertical profiles of temperature, pressure, and relative humidity of the lower-atmospheric boundary layer observed by an instrumented helicopter (HELO) during the field experiment are used as observations in the investigation.

As will be shown later in the paper, the ducts investigated in this study are surface based and elevated ducts over the coastal waters observed by HELO profiles. Evaporation ducts are also surface ducts, but it is difficult to resolve this feature with the vertical resolution of the model. In addition, the HELO was unable to fly low enough to capture the evaporation duct, and therefore there is no consideration of evaporation duct prediction in this study.

Descriptions about the COAMPS ensemble system as well as the products from the matrix of ensemble statistics for atmospheric refractivity are given in the next section, accompanied by a detailed description of the model setup and experimental design. In section 3, results from the experimental studies will be evaluated and discussed. Discussions will focus on three major issues: (i) comparison between ensemble and deterministic forecasts in atmospheric refractivity modeling, (ii) applications of ensemble statistics in predicting atmospheric trapping layers and ducts, and (iii) investigation of the relationship between the ducting layers and the atmospheric fields. Section 4 will summarize the study and provide conclusions from the investigation.

2. System description and experimental design

COAMPS is the Navy’s globally relocatable, high-resolution nonhydrostatic NWP model with multiple-nesting

---

1 COAMPS is a registered trademark of the Naval Research Laboratory.
capability developed at NRL. The model uses a terrain following sigma-\( z \) coordinate in the vertical direction (Hodur 1997). In this study, we used three nested domains with horizontal resolutions of 36, 12, and 4 km, respectively, in which the inner domain with 4-km resolution was used for the refractivity modeling study. The standard setup of the model uses 40 levels in the vertical direction. In this study, however, we used 71 vertical levels to increase the vertical resolution in the boundary layer with the lowest model level 4 m above the surface. The same model configuration is used for both COAMPS ensemble and deterministic forecasts.

The Navy Operational Global Atmospheric Prediction System (NOGAPS; Hogan and Rosmond 1991; Peng et al. 2004) ensemble forecasts are used as lateral boundary conditions for the COAMPS ensemble. They are also used as the first guesses to cold start the COAMPS ensemble. The NOGAPS ensembles obtain initial perturbations from the latitude-banded ensemble transform (ET) technique. The ET is an initial perturbation technique that generates balanced and well-conditioned initial perturbations from ensemble forecast perturbations with the best available guess of the error covariance from a three-dimensional variational (3DVar) analysis system. Readers are referred to McLay et al. (2008) and McLay et al. (2010) for technical details of the ET initial perturbation method. This setup enables the global ensemble to transmit flow-dependent large-scale initial and boundary condition uncertainty information to the regional ensemble. After the cold start, the COAMPS-based EnKF is employed as the data assimilation system for COAMPS warm start with the background fields from the COAMPS ensemble forecasts. The EnKF used here is adopted from the research version of the EnKF developed by The Pennsylvania State University and the National Center for Atmospheric Research for the Weather Research and Forecasting Model (Zhang et al. 2006, 2009). This ensemble data assimilation system uses the ensemble square root technique (Whitaker and Hamill 2002). The COAMPS-based EnKF was implemented at NRL first as a research tool for COAMPS ensemble data assimilation and forecasting. Numerous major changes and improvements have been made to the EnKF since then to make it more suitable for not only research but also operational applications such as rapid environmental assessment and nowcasting (Zhao et al. 2013). The ensemble forecasts are updated every 12 h with the assimilation of conventional meteorological observations plus some satellite data using the EnKF, followed by a 24-h ensemble forecast. Figure 10 in Zhao et al. (2013) provides a reference about the general performance of the ensemble forecasts as a function of forecast lead time relative to deterministic forecasts as verified against rawinsonde observations. It should be mentioned that there is no need for a spinup period since the NOGAPS ensemble has been run for a long period of time. Because of the lack of a coupled atmospheric–ocean ensemble system, the SST analyses used by Haack et al. (2010) in the single deterministic forecasts are used in the COAMPS ensemble forecasts without perturbations. The reason for not perturbing the SST stochastically is that a random perturbation in SST may cause increased inconsistency between the perturbed SST and the COAMPS ensemble members that follow the NOGAPS ensemble initially from a cold start and are produced later through the EnKF warm start. The influence of the unperturbed SST on the ensemble forecasts of the atmospheric boundary layer will be discussed in next section.

A postprocessor for COAMPS ensemble refractivity forecasts has been developed. The processor contains two parts. In the first part, the atmospheric refractivity \( N \) and the modified atmospheric refractivity \( M \) for radio waves are calculated from the model forecasts of atmospheric temperature \( T \), pressure \( p \) (hPa), and water vapor pressure \( e \) (hPa) for each ensemble member using the equations (Atkinson et al. 2001)

\[
N = \frac{77.6}{T} \left( p + 4810^e \right) \quad \text{and} \quad (1)
\]

\[
M = N + \frac{z}{R} \times 10^6, \quad (2)
\]

in which \( z \) (m) is the height above sea level and \( R \) (m) is the mean radius of Earth. Given an \( M \) profile, a ducting layer is indicated by \( dM/dz < 0 \) at one or more adjacent vertical levels for each horizontal grid point of each member. If one or more ducting layers are found in an ensemble member profile, five parameters for each of the ducting layers are calculated: duct-top height, duct-base height, depth, duct strength, and M excess (a parameter indicating whether a duct exists and the degree to which it is a surface-based duct). An example of a surface duct and its five parameters are illustrated in Fig. 1.

In the second part of the postprocessor, ensemble statistics for each of the five ducting-layer parameters are computed. These statistics include maximum, minimum, ensemble mean, and ensemble spread. The ensemble spread is measured by ensemble standard deviation \( \sigma \) given by (Wilks 2006)

\[
\sigma = \sqrt{\frac{1}{K-1} \sum_{k=1}^{K} (x_k - \bar{x})^2}, \quad \text{where} \quad (3)
\]
is the ensemble mean and $K$ is the number of ensemble members with one or more ducting layers. Note that $K$ is the same as the ensemble size only if all ensemble members predict at least one ducting layer. Otherwise it is smaller than the ensemble size. If there is more than one ducting layer in a profile for one ensemble member, the strongest duct layer is selected in the calculations of the ensemble statistics. In addition, the maximum number and the ensemble mean of the number of ducting layers are also computed. The most useful statistical product of the ensemble refractivity forecasts, however, is the duct probability defined by

$$P_d = \frac{K}{N_e},$$

and the probability of a surface-based duct given by $P_s = \frac{K_s}{N_e}$, where $N_e$ is ensemble size. Table 1 gives the matrix of the statistical products for the ensemble refractivity forecasts from the postprocessor.

There could be a potential issue in the calculation of the ensemble statistics of the ducting-layer parameters that use the strongest duct if multiple ducting layers exist in an ensemble profile. Suppose there are two ducting layers in the forecasts. Ensemble member A may predict layer 1 as the strongest, while member B may show layer 2 stronger than layer 1. In this case, the calculated ensemble statistics of ducting-layer parameters would contain numbers from the two ducting layers that are physically inconsistent in the vertical direction. To address this issue, we also compute the ensemble statistics based on the vertical locations of the ducting layer. For example, ensemble statistics are also computed for the lowest ducting layer that is important for surface-based sensor EM propagation.

The Wallops Island field experiment, conducted from late April to early May 2000, is selected as the measurement dataset for this study. This coastal area encompasses the boundary regions between land, cold coastal shelf water, and warm Gulf Stream offshore water. All of these factors strongly influence the marine boundary layer structures and the formation of ducting

![Diagram of a surface duct and its five parameters](https://example.com/diagram.png)

**Figure 1.** Illustration of a surface duct and its five parameters (after Atkinson and Zhu 2005).

<table>
<thead>
<tr>
<th>Duct probability</th>
<th>Products Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duct probability</td>
<td>Number of ensemble members with at least one ducting layer of any type divided by the ensemble size</td>
</tr>
<tr>
<td>Surface duct probability</td>
<td>Number of ensemble members with surface duct divided by the ensemble size</td>
</tr>
<tr>
<td>Number of duct layers</td>
<td>The max duct layers from any one member of the ensemble</td>
</tr>
<tr>
<td>Max duct layers</td>
<td>Ensemble mean of duct layers from all members with duct</td>
</tr>
<tr>
<td>Mean duct layers</td>
<td>The duct strength from the ensemble member that predicts the strongest duct</td>
</tr>
<tr>
<td>Max strength</td>
<td>The duct strength from the ensemble member that predicts the weakest duct</td>
</tr>
<tr>
<td>Min strength</td>
<td>Ensemble mean of duct strength</td>
</tr>
<tr>
<td>Mean strength</td>
<td>Standard deviation of duct strength</td>
</tr>
<tr>
<td>Ensemble spread</td>
<td>Max $M$ excess</td>
</tr>
<tr>
<td>Min $M$ excess</td>
<td>The $M$ excess from the ensemble member that predicts a duct with the smallest $M$ excess</td>
</tr>
<tr>
<td>Mean $M$ excess</td>
<td>Ensemble mean of $M$ excess</td>
</tr>
<tr>
<td>Ensemble spread</td>
<td>Standard deviation of $M$ excess</td>
</tr>
<tr>
<td>Duct depth</td>
<td>Max depth</td>
</tr>
<tr>
<td>Max depth</td>
<td>The duct depth from the ensemble member that predicts the shallowest duct</td>
</tr>
<tr>
<td>Min depth</td>
<td>Ensemble mean of duct depth</td>
</tr>
<tr>
<td>Mean depth</td>
<td>Standard deviation of duct depth</td>
</tr>
<tr>
<td>Ensemble spread</td>
<td>Max top height</td>
</tr>
<tr>
<td>Max top height</td>
<td>The duct-top height from the ensemble member that predicts the duct with lowest top</td>
</tr>
<tr>
<td>Min top height</td>
<td>Ensemble mean of duct-top height</td>
</tr>
<tr>
<td>Mean top height</td>
<td>Standard deviation of duct-top height</td>
</tr>
</tbody>
</table>
layers over water. Haack et al. (2010) provide a detailed description of the field experiment. The main reason for selecting this field experiment is that the atmospheric boundary layer was well measured by 190 HELO profiles of $T$, $p$, and relative humidity (RH) over the coastal water obtained from an instrumented helicopter during 7 days of the field experiment. The $M$ profiles are computed from the HELO data and are used as observations to evaluate the ensemble forecasts. Figure 2 shows the 4-km COAMPS domain over the Wallops Island study area. The black asterisks on the map indicate the HELO locations. At some locations, multiple HELO profiles were collected. The vertical resolution of these profiles varies from usually less than 1 m near the ocean surface to a few meters at higher levels. Relative to the 71 vertical levels of COAMPS, the HELO profiles have much higher vertical resolutions. Another reason for selecting the Wallops Island case for this study is the detailed investigation of the ducting-layer formation associated with the atmospheric boundary layer structures, with COAMPS simulation conducted by Haack et al. (2010) that can be used as a reference for this investigation.

3. Results

a. Ensemble refractivity forecasts and comparison with deterministic predictions

COAMPS ensemble forecasts with 32 ensemble members for the Wallops Island field experiment are conducted from 1200 UTC 28 April to 1200 UTC 4 May 2000. The ensemble size choice of 32 is based on the fact that there are 32 NOGAPS ensemble members available for use in this study. From the study by Zhao et al. (2013), COAMPS deterministic forecasts are used as the lateral boundary conditions for the deterministic COAMPS for the whole experiment period. They are also used at the beginning of the experiment to cold start COAMPS. The same SST analyses used by the COAMPS deterministic forecasts are also used by the COAMPS ensemble.

Figure 3 shows an example of the modified refractivity profiles from both the ensemble and deterministic forecasts valid at 1500 UTC 1 May 2000 near a HELO profile location at $(37.814^\circ N, 75.401^\circ W)$. The $M$ profile calculated from the HELO data at 1506 UTC is also shown. As seen in Fig. 3, a strong surface duct was observed with the trapping-layer top at about 45 m. The ensemble forecasts represent the ducting layer very well with nearly the same duct-top height. Only a couple of ensemble members fail to predict the duct. The ensemble mean matches the observed $M$ profile very well at model levels above 22 m. At model levels at and below 22 m, however, the ensemble mean departs from the observations. The ensemble perturbations are also reduced. This result could be attributed to the effect of the ocean surface, especially the SST, on the atmospheric fields in the lowest boundary layer. In the ensemble forecasts, the atmospheric fields are perturbed for each member while the unperturbed SST analysis is used by all members. The unmatched ensemble temperature and the SST will likely create a thin layer inside which the ensemble forecasts are affected and the ensemble perturbations are also reduced. This result is also evidenced by the reduced ensemble spread of the $M$ profiles (measured by the difference between the two dashed curves in Fig. 3 representing $\tau + \sigma$ and $\tau - \sigma$ on each side of the ensemble mean) near the ocean surface.

The $M$ profiles from both the ensemble mean and deterministic forecasts over the 7-day period are verified.
against HELO observations. A total of 105 HELO profiles are selected to match the output every 3 h from the model forecasts. Verification is done for each ensemble forecast by selecting the nearest model grid point to each HELO location, the nearest HELO observation height to each model level, and the shortest forecast lead time closest to the mean time of each observed profile. Vertical profiles of root-mean-square (RMS) errors and biases (model – observation) are calculated for both the ensemble mean and deterministic forecasts and are given in Fig. 4. Note that this method verifies the $M$ profiles regardless whether a duct is present at the verification location. Similar to what we see in Fig. 3, the ensemble forecasts perform much better than the deterministic forecast at model levels above 22 m. The horizontal bar at each model level of the verification in Fig. 4a indicates the threshold of statistical significance (at 95% probability level) of the RMS error differences (only bars on the side between the two curves are plotted to make the figure more legible). At model levels between 45 and 262 m, the reduction in RMS errors is statistically significant. At and below the 22-m level, however, the deterministic forecasts slightly outperform the ensemble in terms of RMS errors, with the RMS error differences close to the statistical significance thresholds. This may again indicate the need for ensemble SST perturbations consistent with the ensemble atmospheric perturbations. At levels above 262 m, the number of HELO observations decreases with height and confidence level also decreases as a consequence. It is also worth noting that the ensemble mean gives much lower biases than the deterministic forecast (Fig. 4), especially near the ocean surface. This result is because the ensemble mean has warm–dry biases at some HELO locations and cold–moist biases at other locations. When averaging over the 105 HELO locations, the biases at those locations cancel and the total bias is reduced.

Daily verification of the model forecasts of the $M$ field against the 105 HELO profiles are also conducted to study the performance of the ensemble and deterministic forecasts on each day over the 7-day period. Instead of calculating RMS errors and biases at each model level as shown in Fig. 4, the scores are computed over the three-dimensional (3D) domain for each day and are displayed in Fig. 5. The 3D domain here means all the model vertical levels with HELO data available at the 105 HELO locations within the model inner nested grid (all model levels above the highest HELO data at a HELO location are not used in the calculation).
Also, the average scores for the whole 7-day period are given in Table 2. The ensemble forecasts perform better than the deterministic on 5 of the 7 days in terms of RMS errors. This result is also seen by the better overall scores in Table 2. It is also interesting to note that, except on the first day, the deterministic forecasts always have a positive bias \( M(\text{forecast}) - M(\text{observation}) > 0 \), which is likely related to a cold and moist bias, a result that is also reflected by the large values of positive bias for the deterministic forecasts in Fig. 4b. In contrast, the ensemble forecasts have a positive bias on some days and a negative bias on others. The two days on which the deterministic forecast outperforms the ensemble mean are 30 April and 3 May. To further understand what happened on 3 May, the \( M \) profiles from the ensemble and deterministic forecasts as well as the HELO observations at 1822 UTC 3 May 2000 at the HELO location of 37.809°N, 75.434°W are shown in Fig. 6. Obviously, all ensemble members failed to predict the deep surface duct as revealed by the HELO data. Although the deterministic forecast did not represent the ducting layer very well either, its \( M \) profile is closer to the HELO data near the surface. In contrast, most of the ensemble members underpredict \( M \) by a large amount below heights of 100 m, which again may be related to the use of unperturbed ocean surface boundary conditions for the ensemble.

The verification of \( M \) profiles is insightful because it reflects accuracies of the atmospheric temperature, moisture, and pressure fields in the model forecasts. For ducting-layer prediction verification, however, the scores from \( M \)-profile verification alone may not be sufficient or may give misleading information. Figure 7 shows two good examples for which the deterministic forecasts have better RMS errors and biases scores than the ensemble mean. If we look at the duct forecasts, however, the deterministic forecast misses the ducting layer completely in the May 4 case and overpredicts the duct strength in the April 30 case. Most ensemble members in the first case produce a ducting layer with similar structure as observed (the ensemble mean also shows a surface duct although the strength is much weaker and the duct-top height is slightly lower than the one in the HELO observations). In the second case, the ensemble mean shows similar duct strength as observed. Apparently, in both cases the ensemble performs better than the deterministic for ducting-layer and duct parameter forecasts, although the \( M \) profile from the deterministic forecast is closer to the observations than most ensemble members. These two examples highlight the insufficient information given by a statistical assessment of \( M \) and indicate the need to verify the ducting-layer parameters (top height, strength, thickness, \( M \) excess, etc.) directly. But technically, standard statistical verification of duct parameters is difficult because atmospheric ducting is a binary yes/no process. At some locations, the model may show a ducting layer but the observations do not, or vice versa. And, there can be
inconsistencies associated with matching duct layers in the statistics when there are multiple ducts. For this reason, we present an alternative method for ducting-layer verification. Instead of verifying the duct parameters, we verify the vertical gradient of the $M$ profile, $\gamma = dM/dz$ since $\gamma$ is a continuous field that exists at all grid points. Most importantly, it is the parameter that determines the structure of ducting layers and hence the atmospheric EM propagation. To perform this verification, using the vertical grid spacing of the 71 model levels, $\gamma$ profiles from both the deterministic and ensemble forecasts at the 105 HELO locations over the 7-day period are calculated using $\gamma_k = (M_{k+1} - M_k)/(z_{k+1} - z_k)$, where $k$ is the model vertical level index with $k = 1$ at the surface, and are compared with the HELO observations. The $\gamma$ profiles of RMS errors and biases are shown in Fig. 8. The thresholds of statistical significance for RMS error differences at each model level are also given in Fig. 8a. The total RMS error and bias over the 3D domain (including all the model levels in Fig. 8) are given in Table 3. Similar to $M$ profiles, the $\gamma$ profiles from the ensemble verify better than those from the deterministic forecasts in the lower boundary layer except at the lowest model level. At model levels 10, 22, and 75 m, the reduction in RMS errors are statistically significant. Figures 8 and 4 show that the ensemble system has advantages over the deterministic approach in predicting not only the atmospheric refractivity fields in the lower boundary layer but also their vertical gradient. With the fully coupled atmosphere–ocean ensemble system currently under development at NRL, we expect the prediction of the $M$ field and its vertical gradient near the ocean surface to also improve.

### b. Ensemble statistical products and probabilistic forecasts of ducting layers

In addition to the improved prediction of the atmospheric refractivity and its vertical gradient in the lower-atmospheric boundary layer as shown above, the

| Table 2. Mean RMS errors and biases of the ensemble mean and deterministic forecasts of modified refractivity over the period of the Wallops Island field experiment (28 Apr–4 May 2000) verified against the 105 HELO profiles averaged over all heights, times, and locations. The sample size for the verification is 960. |
|-----------------|-----------------|
|                | Ensemble mean   | Deterministic |
| Bias           | 0.8752          | 4.9712         |
| RMS error      | 7.9182          | 9.4746         |

Fig. 5. (a) Daily mean RMS errors and (b) biases of the ensemble mean (red bars) and deterministic (blue bars) forecasts of modified refractivity, inclusive of all heights, for the Wallops Island field experiment period (28 Apr–4 May 2000) verified against HELO profiles. The numbers at the bottom are the sample size for each day.
ensemble forecasts also provide an array of statistical products (see Table 1) that make probabilistic forecasts of ducting layers possible. Among them, the probability of duct occurrence and probability of surface-based duct occurrence are given along with statistical parameters describing the duct characteristics. Unlike the deterministic forecast that has a yes/no binary flag for duct occurrence at a given location, the ensembles yield the probability of a duct occurrence, as defined by (5). Note that a duct probability of 50% equals the detection of a ducting layer by the ensemble median. As an example, Fig. 9a shows a horizontal map of duct probability from the 9-h ensemble forecasts valid at 2100 UTC 4 May 2000. With strong onshore surface flow, the ensemble forecasts show more than 70% chance for one or more ducting layers to develop over most of the model domain except along the coastline, where the land–sea boundary may enhance vertical mixing and hence reduce the probability for a ducting layer to occur, and in the northeast corner where the water temperature is cold and the air is quite dry. Figure 9b gives an enlarged map of the area around the Wallops Island (marked by the gray box at the center of Fig. 9a). The black plus sign in Fig. 9b indicates a HELO location where the observed $M$ profile at 2102 UTC is given in Fig. 10 along with $M$ profiles from the ensemble and deterministic forecasts valid at 2100 UTC at the nearest grid point. As seen in Fig. 10, the HELO observations show a strong surface duct with the duct top at about 75 m. Both the ensemble mean and deterministic forecasts have no duct, although the $M$ profile from the ensemble mean looks closer to the observations than the deterministic forecast. If we look at the $M$ profiles from the individual ensemble members, however, we find that many members have a weak duct in their forecasts. It is also interesting to notice that the $M$ profile from one ensemble member (to the most left of the ensemble group) shows very similar vertical structure as the observed. The number of
ensemble members with a duct forecast at this HELO location is also seen in the duct probability map in Fig. 9b where the duct probability at the plus sign location is about 58%, which means more than half of the ensemble members contain a ducting layer.

To quantify the performance of the ensemble and deterministic forecasts in predicting ducting-layer events, the Brier scores of duct occurrence forecast are calculated using the equation

$$B_s = \frac{1}{M} \sum_{m=1}^{M} (f_d - o_d)^2,$$

where $M$ is the number of observations used for the verification, $o_d$ is the HELO observations of duct occurrence (no duct = 0; duct = 1), and $f_d$ is the model forecast of duct occurrence. Three Brier scores are calculated: (i) for the ensemble duct probability forecast using (5), where $f_d$ varies between 0 and 1; (ii) for the ensemble mean forecast (using the ensemble mean as a single forecast), where $f_d$ is either 0 or 1 meaning no duct or duct; and (iii) for the deterministic forecast, where $f_d$ is also binary. The scores from the verification against the 105 HELO observations are given in Table 4. As seen in the table, the ensemble probability has the best score while the score from the deterministic forecast is the worst. The score from the ensemble mean is in the middle. To better understand these results, the values of $f_d$ at the 105 HELO locations are plotted in Fig. 11 for the ensemble probability (the green curve), the ensemble mean (the blue curve), and the deterministic (the red curve) forecasts. For verification, the values of $o_d$ are also plotted. As seen in Fig. 11, most duct events are predicted by all model forecasts. The ensemble probability, however, has the most duct events (with duct probability > 0.5). The main difference between the ensemble probability and ensemble mean is at HELO locations 99–103 where the ensemble mean misses the duct event while the product probability shows a value larger than 0.75. That result suggests that most ensemble members predict the duct event, but perhaps at different heights leading to large forecast errors in $dM/dz$ and the tendency to remove the ducting layers in the ensemble mean. The deterministic forecast has the lowest score because it misses most of the duct events during two periods, identified by HELO locations 9–25 and 81–93.

These results demonstrate the usefulness of duct probability forecast. Ensemble mean is an important ensemble forecast parameter, but it is not necessarily a physically consistent solution. Sometimes it may be less relevant and insufficient for EM propagation considerations because it can dampen the vertical gradient structures through the smoothing effect of averaging. This is especially true when some individual ensemble members have very large forecast errors or ducts at different heights. In the case shown in Figs. 9 and 10 when the ensemble mean misses the deep ducting-layer below 75 m, the number of the ensemble members with a ducting layer (even a weak one) in their forecasts is still a useful indication of the probability of a duct.

Figure 12a gives another example of duct probability at 2100 UTC 30 April 2000 from the ensemble forecasts. This is a case in which surface offshore flow dominates the coastal areas and high probability (>90%) of ducting...
covers a large portion of the domain except in the area far away from the coastal water where SST changes from the cold shelf coastal water to warm Gulf Stream, at a few spots inland, and along the coastline. Figures 12b–e give the ensemble means of duct strength, $M$ excess, duct-top height, and duct depth, respectively, at the same time. The white contours in these figures indicate the ensemble spread of these fields. By comparing the duct probability in Fig. 12a with the ensemble means in Figs. 12b–e, we find that ducts over land are generally thin (<20 m in thickness) and very weak (<1 $M$ unit in strength). Over the coastal waters, however, there exists a large area of stronger ducting extending seaward about 100 km from the coastline. This duct area also covers most parts of the Chesapeake Bay and the Delaware Bay. The ensemble spread indicates differences among the ensemble members in predicting the strength, $M$ excess, thickness, and top height of the ducting layers. The similar patterns of the ensemble mean between the duct strength and the $M$ excess and between the duct-top height and duct depth near the coastline imply that most ensemble members predict a surface duct there. In areas away from the coastline, the $M$ excess has negative values, indicating elevated ducts. The example in Fig. 12 further illustrates the usefulness of the ensemble ducting products, when used conjointly, in the probabilistic forecasts of the occurrence, location, type, base and top height, and strength of ducting layers. The ensemble spread also represents the uncertainties in the model forecasts where ensemble members largely differ from each other. The largest values appear in the duct-top height (Fig. 12d) and duct depth (Fig. 12e) forecasts along the coast, possibly because of the complexities of the terrain and variations in the coastline.

c. Environmental impacts on ducting-layer development

It is known that a ducting layer forms under certain circumstances of the combined large-scale environment, mesoscale forcing, the atmospheric boundary layer structures, and the surface conditions. From (1) and (2), we obtain the vertical gradient of $M$ at a grid point

$$
\gamma = \frac{dM}{dz} = \frac{10^p}{R} + \frac{77.6}{T^2} \left[ \frac{4810}{\overline{T}} \frac{de}{dz} - \frac{g}{R_d} \frac{p}{T} \right] - \left( \frac{9620}{T} \right) \frac{dT}{dz},
$$

(7)

where $g = 9.8$ (m s$^{-2}$) and $R_d = 287.0$ (m$^2$ s$^{-2}$ K$^{-1}$) are the gravity acceleration and gas constant, respectively. Units for other variables are the same as in (1) and (2).

<table>
<thead>
<tr>
<th>Table 3. As in Table 2, but for $\gamma = dM/dz$.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Bias</td>
</tr>
<tr>
<td>RMS error</td>
</tr>
</tbody>
</table>
Hydrostatic assumption is used in (7). Obviously, atmospheric temperature, pressure, and moisture and their vertical gradients all have impacts on the value of $g$, which, in turn, determines the development and structures of a ducting layer. In addition to these parameters, previous studies (Thompson et al. 1997; Burk and Thompson 1997; Atkinson et al. 2001; Haack et al. 2010) show the horizontal winds and surface conditions also influence ducting layers. By using the ensemble refractivity forecasts, we can further investigate the influence of these atmospheric parameters on ducting layers by quantifying relationships between $g$ and other

---

**FIG. 9.** (a) Duct probability (%) colors from the 9-h ensemble forecasts valid at 2100 UTC 4 May 2000. White arrows are the horizontal wind vectors at the lowest model level ($z = 4 \text{ m}$). (b) The enlarged map of the central area in (a) marked by the gray box. The black plus sign indicates the location of the HELO profile shown in Fig. 10, below.

**FIG. 10.** As in Fig. 7, but valid at 2100 UTC 4 May 2000 at the same HELO location (the black plus sign in Fig. 9b). The HELO data were observed at 2102 UTC.
model fields. Here, we calculated the three-dimensional correlation coefficients between the $\gamma$ at the model level of $z = 10$ m and horizontal grid point near 38°N, 75°W (the black dot in Fig. 13) and the three-dimensional model fields of $T$, $q_v$, $u$, and $v$ from the ensemble forecasts using the equation

$$
\rho_k = \frac{1}{N-1} \sum_{i=1}^{N} \left( \gamma_i - \overline{\gamma} \right) \left( x_{ik} - \overline{x}_k \right) \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left( \gamma_i - \overline{\gamma} \right)^2 \frac{1}{N-1} \sum_{i=1}^{N} \left( x_{ik} - \overline{x}_k \right)^2},
$$

where $x$ denotes the model state variables ($T$, $q_v$, $u$, or $v$), $i$ indicates the $i$th ensemble member, $N$ is the ensemble size, and $k$ is the model grid index in the three-dimensional model domain. The reason for selecting $\gamma$ at $z = 10$-m level is that this is a critical height for the development of a surface duct, and it is less affected than the lowest model level ($z = 4$ m) by the unperturbed SST used by the ensemble forecast system as mentioned earlier. The calculation of (8) is straightforward. The $\gamma$ forecasts at the given location (the black dot in Fig. 13) from all ensemble members are used to compute $\overline{\gamma}$, the ensemble mean of $\gamma$. The same thing can be done for the state variable $x$ at the grid point $k$. Then the correlation coefficient between $\gamma$ at the given location and $x$ at the grid point $k$ is computed using (8). By changing grid index $k$, the correlation map inside the model domain between $\gamma$ at the given location and $x$ at all grid points is obtained.

Figure 13 shows the maps of the calculated horizontal correlation coefficients from the ensemble forecasts valid at 2100 UTC 30 April 2000 on the model level of $z = 10$ m. The black dots on the maps indicate the $\gamma$ location. It is seen from Fig. 12b that the model predicted a ducting layer at the $\gamma$ location with the duct strength of about $6$ M units at that time. A low correlation is found between $\gamma$ and $T$, less than $\pm 0.4$, and also between $\gamma$ and the horizontal winds ($u$ or $v$), less than $\pm 0.5$. In contrast, the correlation between $\gamma$ and $q_v$ is high and much higher than those for $T$, $u$, and $v$ in the area surrounding the $\gamma$ location. The maximum correlation is about $0.75$ at the $\gamma$ location. This indicates that the moisture field has a strong impact on the development of the duct layer in that area. Interestingly, the correlation contours in some areas on the ocean side surrounding the $\gamma$ location in Figs. 13a–d roughly follow the horizontal wind bars (this is more apparent on the east and southeast sides of the $\gamma$ location where strong northerly surface winds exist). Along the coastline, the density of the contours is notably high, which means the horizontal gradient of the correlation coefficient in the direction across the coastline is large. The contour patterns in Fig. 13 suggests that the horizontal winds do not impact the development of ducting layers directly as
FIG. 12. (a) As in Fig. 9a, but for the forecast valid at 2100 UTC 30 Apr 2000. (b) Duct strength (M units), (c) M excess (M units), (d) duct-top height (m), and (e) duct depth (m) at the same time. In (b)–(e), the contours are the ensemble spreads.
indicated by the low correlations, but they may define the areas in which the refractivity is influenced by the moisture field through advective processes.

Figure 14 gives the same correlation coefficients in the vertical direction at the $\gamma$ location. We should mention that these correlation calculations are point to point. No horizontal averaging over the neighborhood of the $\gamma$ location is used. Similar to what we have seen on the horizontal maps in Fig. 13, $q_y$ is the only model field that shows a high correlation with $\gamma$. This high correlation starts from the ocean surface with a value of about 0.72, increases slightly to a maximum of about 0.8 at $z = 22$ m, stays at about the same value until $z = 205$ m, and then decreases with height. At the height of $z = 645$ m, the correlation is still above 0.5, indicating that the moisture over a deep layer may have large influences on atmospheric refractivity structure near the surface. This result is consistent with (7), where the term $[p + 9620(e/T)]$ is usually smaller than 4810 by a factor of 3 or 4 at sea level (where $p$ is close to 1000 hPa) with normal temperature (K) and water vapor pressure (hPa) values, and hence $de/dz$ dominates the $\gamma$ calculation unless the vertical temperature gradient $dT/dz$ is extremely large. In addition, moisture field can have small-scale features. Winds and local turbulence can change moisture distribution through horizontal advection and/or vertical mixing and hence impact refractivity profiles.

4. Summary and discussions

In this study the COAMPS ensemble system is evaluated to demonstrate the usefulness of the system in atmospheric refractivity forecasts and also to show the challenges in predicting boundary layer ducting caused...
by the unpredictability of the atmosphere and uncertainty due to model errors associated with physical parameterizations. The Wallops Island field experiment conducted from late April to early May 2000 is used to validate the ensemble system. The ensemble forecasts are compared with COAMPS deterministic forecasts and verified against the observations from HELO profiles obtained during the field experiment. The ensemble statistic products from the ensemble postprocessor are also presented and demonstrated in probabilistic forecasts of duct occurrence, location, and ensemble mean duct strength, height, and thickness. The vertical structure is also evaluated using the HELO data. Applications of the ensemble forecasts and their products in investigating the possible impacts of the atmospheric environment on the development and structures of ducting layers are illustrated.

This study successfully exhibits the overall better performance of the COAMPS ensemble over the deterministic forecasts in the prediction of atmospheric refractivity and its vertical gradient, especially evident during periods of onshore flow. It also explores the advantages of the ensemble forecast system, especially its ability in accounting for the uncertainties in model forecast errors in representing the finescale structures of the atmosphere in the lower-atmospheric boundary layer. At the lowest model level right above the ocean surface, however, the use of unperturbed SST analyses in the COAMPS ensemble in this study mismatches the perturbed atmospheric fields in the ensemble and therefore affects the ensemble forecasts near the ocean surface. This result indicates the need for perturbed SST in the ensemble forecasts. A fully coupled atmospheric-ocean ensemble system for such an application could offer further improvements.

In addition to the improved forecasts, the statistic products from the ensemble forecast postprocessor also provide additional information about the likelihood of occurrence in terms of percent chance of ducting-layer occurrence and the associated parameters. Unlike a deterministic forecast that gives a binary answer to ducting at a given location and time, the well-perturbed ensemble members give the possible solutions about the ducting layer in their forecasts. As illustrated by the example in Figs. 9 and 10, sometimes both the deterministic and ensemble mean may miss the duct at a given location, while the duct probability still indicates a high percent chance of a ducting layer. As seen in Table 4, the duct probability has a better Brier score in the prediction of the chance of the ducting-layer occurrence than both the ensemble mean and the deterministic forecast at all 105 HELO locations. These results demonstrate the usefulness of the duct probability in the prediction of ducting layers in terms of probabilistic forecast. As demonstrated by another example in Fig. 12, the duct probability maps should be used together with other ensemble statistic products on duct strength and base height, for example, to provide a reasonable forecast of the ducting layers and their characteristics, indicating the need to optimally use information from all the ensemble statistics in the probabilistic forecast of ducting layers.

While the occurrences of most duct events during the Wallops Island 2000 field experiment are well predicted by the model forecasts as shown in Fig. 11, quantifying the strength, vertical location, and depth of ducting layers in the forecasts still remains challenging. This can be seen in Table 3, where the RMS errors in $\gamma$ from both the ensemble and the deterministic forecasts are larger than the value separating trapping and normal refraction and close to that between trapping and subrefraction (see Fig. 1 in Haack et al. 2010). In addition to the imperfection in model boundary layer dynamics and physics, the lack of coupling with ocean and the lack of high-resolution observations in data assimilation that define the fine vertical structures of the atmospheric boundary layer near the surface in the model initial
fields are thought to be the major causes. Now that the fully coupled COAMPS is available to provide dynamically updated and coupled ocean surface conditions for atmospheric modeling and that the sensor-diagnosed effective modified refractive index observations are available for assimilation, we are working on developing techniques to address those challenging issues.

The application of ensemble forecasts in investigating the influence of the atmospheric environment on ducting-layer development and structures is also demonstrated. For the duct case discussed in section 3c, the moisture field has the most significant impact on the vertical gradient of refractivity near the surface, with high correlation values extending up to about 600 m. This example indicates the importance of moisture forecast for atmospheric refractivity prediction. The horizontal winds in this case do not show significant correlations, but do affect the areas of moisture impacting the refractivity gradient at a given location. This study also demonstrates the large influence of the surface conditions on either side of the coastline on the development of atmospheric refractivity conditions in coastal areas. All the above findings prove the usefulness of the ensemble forecast system in modeling ducting layers and indicate the importance of the environmental winds and surface conditions on the development of ducts over coastal water. In addition, the correlation analyses given in Figs. 13 and 14 are useful in providing guidance for assimilating refractivity measurements to update the atmospheric state.

It should be pointed out, however, that the 2000 Wallops Island field experiment was conducted more than 16 years ago. As mentioned earlier, the observational data of the ocean surface conditions and the atmospheric structure above the ocean surface from conventional and satellite measurements over the experiment area at that time were limited. Currently, we are working on testing the ensemble system with more recent field experiments in which more measurements from the remote sensing of the boundary layer between the ocean surface and the atmosphere are available for data assimilation. Another question that remains unaddressed in this study is how the ensemble spread affects the ensemble forecast skill of atmospheric refractivity. Quantifying the relationship between the ensemble spread and ensemble forecast skill has been an attractive research topic during the last 20 years or so, and many interesting studies have been reported (e.g., Whitaker and Loughe 1998). However, because of the limited sample size here, and the fact that a large number of forecast and observational data would be needed to statistically establish the relationship between the ensemble spread and the ensemble forecast skill in a meaningful way, establishing this relationship is beyond the scope of the current study. In this study, we have ensemble forecasts for only 7 days (14 forecasts) and about 105 HELO profiles from the field experiment that can be used to evaluate the model forecast skill. Such a small sample of data will likely yield a correlation between the ensemble spread and forecast skill that is not statistically meaningful and, if not treated appropriately, may be misleading. As we continue to run the ensemble system for more recent field experiments, more forecast and observational data are expected to be available. Such an investigation may be conducted separately as follow-on research in the future.

Acknowledgments. The authors thank Drs. Teddy Holt and Jim Hansen of the Marine Meteorology Division, Naval Research Laboratory. Their help and constructive discussions on COAMPS ensemble forecasts and the ensemble statistics are greatly appreciated. This research was supported by the Chief of Naval Research through the NRL Base Program, PE 0601153N, PE 0602435N, and ONR-sponsored research project N0001415WX00849. The computational resource was provided by the Department of Defense Supercomputing Resource Centers at the Navy Oceanographic Office and the Air Force Research Laboratory.

REFERENCES


