Lewis, J. M., S. Lakshmivarahan, J. Hu, R. Edwards, R. A. Maddox, R. L. Thompson, and S. F. Corfidi, 2016: Ensemble forecasting of return flow over the Gulf of Mexico. *Electronic J. Severe Storms Meteor.*, **11** (4), 1–26.

Electronic Journal of
SEVERE STORMS METEOROLOGY

Ensemble Forecasting of Return Flow over the Gulf of Mexico

JOHN M. LEWIS,[*] S. LAKSHMIVARAHAN,[+] JUNJUN HU,[+] ROGER EDWARDS,[#] ROBERT A. MADDOX,[@] RICHARD L. THOMPSON,[#] AND STEPHEN F. CORFIDI[#]

[*] *National Severe Storms Laboratory, Norman, Oklahoma, and Desert Research Institute, Reno, Nevada*
[+] *School of Computer Science, University of Oklahoma, Norman, Oklahoma*
[#] *NOAA/NWS, Storm Prediction Center, Norman, Oklahoma*
[@] *Tucson, Arizona*

(Submitted 14 February 2016; in final form 08 August 2016)

ABSTRACT

Errors in operational forecasts of return flow events (RFEs) over the Gulf of Mexico have dictated the search for sources of these errors. Based on earlier studies, likely candidates for these errors are: incorrect parameterization of turbulent transfer processes at the air-sea interface, uncertain vertical motion above the mixed layer, and incorrect initial conditions. We investigate these possible sources of error by performing numerical experiments with a Monte Carlo ensemble prediction model applied to a well-observed case in February 1988. In essence, we examine uncertainty in prediction due to uncertainty in the model's elements of control. A mixed-layer model with roughly 50 elements of control is used to determine forecast uncertainty due to initial conditions alone, boundary conditions alone, parameterization alone, as well as the full complement of uncertainty in these elements of control. The uncertainty is calculated at points along a predetermined outflow trajectory that originates over shelf waters in the northeastern Gulf, passes north of the Yucatan Peninsula, and terminates in the west-central Gulf—all points along the trajectory are characterized by convective heating at the sea-air interface. Results from the numerical experiments led to the following results: 1) parameterization of physical processes exerts the greatest influence on forecast uncertainty, and 2) the water-vapor mass in the mixed-layer column is uncertain by a factor of two at the trajectory's terminal point. The latter result confirms forecasters' long-held view that vapor return is the most suspect product in operational prediction of RFEs. In addition to these numerical experiments with the 1988 case, a recent RFE is examined in the context of operational model performance at the National Center for Environmental Prediction (NCEP). The paper ends with discussion of steps to be taken that hold promise for improved operational prediction of RFEs over the Gulf of Mexico.

1. Introduction

In late fall and winter, a rhythmic cycle of cold air penetrations into the Gulf of Mexico (GoM) takes place. Whether these penetrations are deep, intermediate or shallow, they are generally followed by return of modified air to land in response to circulation around an eastward-moving cold anticyclone. Keith Henry, late professor of meteorology at Texas A & M University aptly termed this large-scale process "return flow" (Henry 1979 a,b). Four to five of these return-flow events (RFEs) occur every month between November and March. And, as noted in climatological studies of RFEs (Crisp and Lewis 1992), variations on this scenario are myriad—indefinite but large—and this is central to the challenge of accurately forecasting the phenomenon.

Project GUFMEX (Gulf of Mexico Experiment) in 1988 and 1991 was driven by the desire to identify factors that led to poor operational forecasts of the RFEs. Difficulties

Corresponding author address: J. M. Lewis, National Severe Storms Laboratory, Norman, OK, 73072, and Desert Research Institute, Reno, NV, 55912. e-mail: jlewis@dri.edu

were conjectured to stem from one or more of the following factors (Lewis et al. 1989; Lewis 1993):

1) Absence of routine upper-air observations over the Gulf,
2) Absence of dewpoint (moisture) measurements on tethered buoys over the Gulf's shelf water,
3) Errors in sea-surface temperature (SST) due to aged data in response to cloud cover, and
4) Inaccuracy in model parameterizations of moisture and heat fluxes at the sea-air interface.

The operational forecasts of water-vapor mixing ratio have been especially troublesome. Among those who have investigated this aspect of the problem are Janish and Lyons (1992), Weiss (1992), Thompson et al. (1994), Edwards and Weiss (1996), and Manikin et al. (2000, 2001, 2002). The consequence of poor forecast guidance is extreme where forecasts range from sea fog and stratus cloud when vapor content is low, to shallow cumulus with light showers for intermediate values of the moisture, to cumulonimbus and associated severe weather for large-magnitude vapor mixing ratios.

The online supplement to this paper brings the reader up-to-date on recurring problems with operational forecasts of RFEs at the National Center for Environmental Prediction (NCEP)/ Environmental Modeling Center (EMC). For the case examined (15 March to 1 April 2015), both the Global Forecast System (GFS) and the North American Mesoscale (NAM) model clearly continue to exhibit problems with vertical structure of both moisture and temperature. Somewhat surprisingly, the longer-range forecasts (72 h) appear to be more faithful to observed structures than those at shorter ranges (at least for this single RFE). In the absence of a more complete validation study, a qualitative assessment of operational model performance in these RFE situations has been addressed in the paper's main body (section 6).

The primacy of the anticyclone—its path, its strength and fate, is central to the RFE phenomenon. Palmén and Newton (1951) were the first to use upper-air observations to investigate this anticyclone and the cold air outflow over the Gulf. They were intent on understanding the exchange of momentum between mid-latitudes and the tropics in support of a mechanism that could sustain the subtropical jet steam—a general circulation problem (Riehl 1988). Yet, their study had synoptic implications as later studied by Bosart and his graduate students (Dallavalle and Bosart 1975, and Boyle and Bosart 1983). Bosart's retrospective view of this research follows (Bosart, 2016, personal communication): "These two papers taught me that the behavior of cold anticyclones that drive cold air southward into the Gulf of Mexico is very much dependent upon both the configuration and evolution of the large-scale flow as well as subtleties on how the evolving and southward-moving cold air masses would interact with the Gulf of Mexico (e. g., surface sensible and latent heat fluxes)".

Indeed, the subtleties of air-sea interaction in these RFEs calls for a study that accounts for the uncertainty in controls that govern numerical forecasts. In an effort to uncover the sources of errors in RFE forecasts, an ensemble forecast based on the Monte Carlo (MC) methodology is proposed. In this study, we will create ensemble members that result from perturbation to: 1) initial conditions alone, 2) boundary conditions alone, 3) parameterizations alone, and 4) the full set of controls. In this way, we hope to determine the controls that have most bearing on forecast uncertainty. An important consideration in the linkage between forecast error (as found through comparison of forecast with quality observations such as root-mean-square differences) and uncertainty in prediction follows: *Forecast uncertainty cannot be strictly viewed as forecast error*, but *it can be viewed as a measure of spread or variance in the forecast through divergence of analogous states.* The rate of divergence of analogue states, and those processes that most influence the rate of divergence, point to likely culprits in generating forecast error.

The forecast will be restricted to the "outflow phase" of the RFE where buoyancy is the dominant factor in the governing dynamical equations. Focus on the outflow is justified for two reasons: 1) the operational forecasters at NOAA's Storm Prediction Center (SPC) collectively feel that errors in moisture return are primarily due to errors in fluxes at the air-sea boundary during outflow; and 2) the classic mixed-layer model (Ball 1960; Lilly 1968), a low-order model, can be used to make the forecasts and can easily accommodate many ensemble members. The larger the number of members in the ensemble, the more accurate the statistics of the MC forecast (Hammersly and Handscomb 1964). And importantly, the mixed-layer model is faithful to the structure of heating and moistening during outflow (Liu et al. 1992).

The case we study takes place in late February 1988 during GUFMEX. The case is representative of deep penetrating RFEs, and justification for reaching so far back in history rests on the excellent set of upper-air observations collected during this event. We first summarize the observed synoptic features of the February 1988 RFE and follow with presentation of the dynamical constraints—mixed-layer equations—used in the ensemble forecast. The heart of the research resides in the numerical experiments with the ensemble model. The paper ends with focus on key results from the experiments and their implications on prediction of the RFE.

2. Observations and the RFE trajectory

Cold and dry maritime Polar (*mP*) air initially moved over the western Gulf at 0600 UTC 20 February 1988. Roughly two days later (0000 UTC 22 February), modified air returned to the Texas coast and gradually spread eastward. This return flow along the Gulf's northern coastal plain persisted until ≈0600 UTC 24 February. Thus, the time frame for the RFE was approximately 4 days. The sequence of Geostationary Operational Environmental Satellite (GOES) images in Fig. 1 gives a large-scale perspective of the cyclone-anticyclone couplets and fronts associated with this late February 1988 RFE. In the top panel, the deep cold-air penetration across the GoM is evident. The cold front associated with the penetration stalled over the Straits of Florida while a second front bore down on the Gulf. Return-flow air fueled moderate rain showers from southeast Texas to New England.

We pay particular attention to a plume of cold, dry air that spread over the Gulf just southeast of New Orleans, LA. This cold-air outflow that extends to the Yucatan and Straits of Florida by 1200 UTC 21 February is shown in Fig. 2 with an MSLP map and surface winds superimposed. This air mass executed an expansive 3-day trajectory over the Gulf as shown in Fig. 3. The SST analysis underlying the trajectory is the operational product from National Hurricane Center (NHC) valid on 23 February, 1988. The surface winds used in constructing the trajectory come from Fig. 2 and similar analyses from the NHC archive. Trajectory construction follows steps in Saucier (1955).

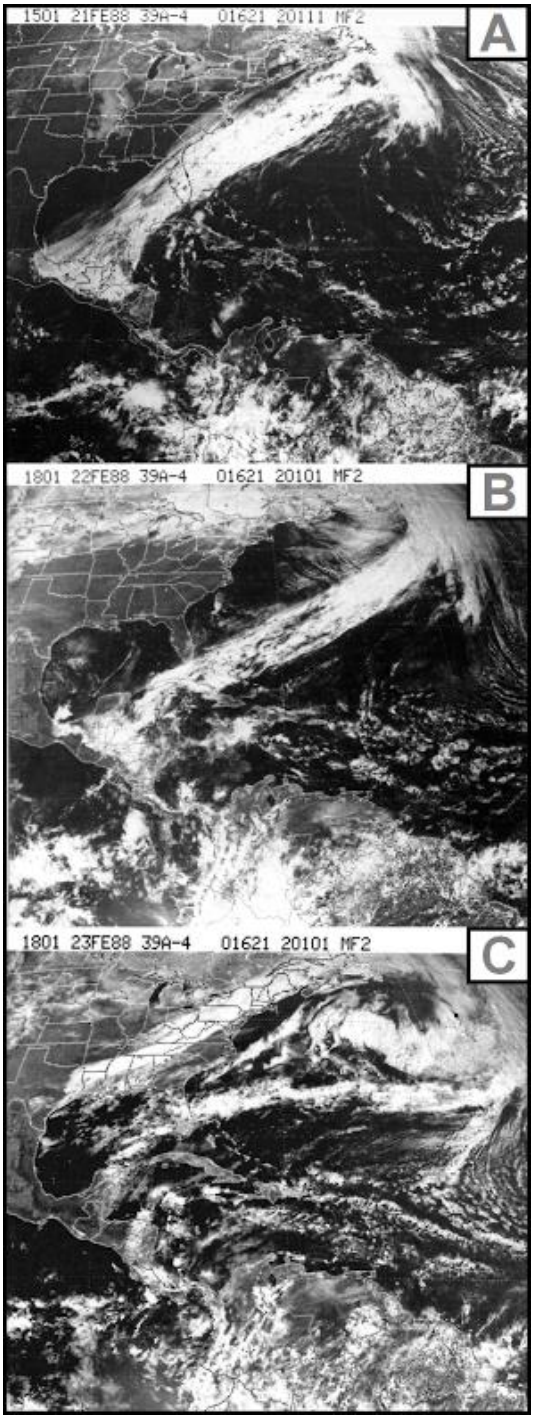

Figure 1: Scans of GOES visible-wavelength satellite photos for the Gulf of Mexico (left center), western Atlantic Ocean, eastern North America and northern South America: a) 1501 UTC 21 February 1988, b) 1801 UTC 22 February 1988, and c) 1801 UTC 23 February.

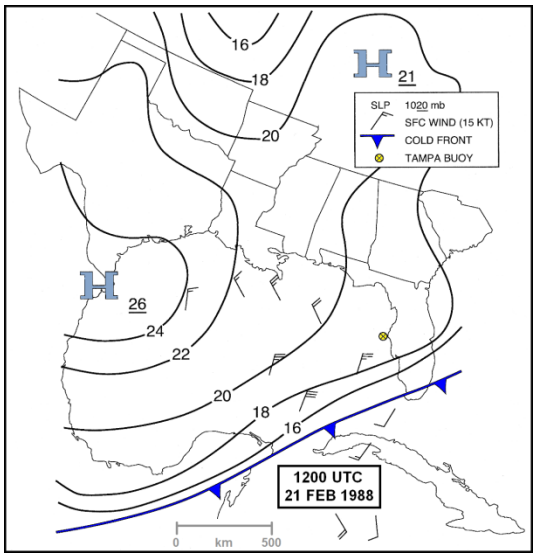

Figure 2: NHC operational surface analysis valid at 1200 UTC 21 February 1988. The two leading digits ("10") in MSLP have been omitted. *Click image to enlarge.*

The key to trajectory construction hinges on the location of U. S. Coast Guard ship *Salvia* at 1728 UTC 21 February (time approximated by 1800 UTC 21 February). The location of this observation is denoted by a bulls-eye (a black dot within a yellow circle) aside the Roman numeral "I". The upper-air observation taken from *Salvia* at this time serves as the initial condition for the forecast to follow. And since the forecast will be an ensemble forecast, the *Salvia* observation is assumed to be the mean value of the initial condition to which perturbations will be added. And to emphasize the fact that the observation at this time serves as an initial condition, the notation "t=0" is placed aside this location.

Using "I" as a starting point, a fifteen-hour (15 h) back trajectory is first constructed. The air parcel ends up at a point near New Orleans denoted by a red dot and the affixed time 0300 UTC 21 February. Relative to the initial time, this point is located at t = –15 h. Starting from this point, a sequence of curved lines connecting red dots represents the air parcel's movement over 6-h periods. The first curved line is dashed to indicate some movement over land. Subsequent curved lines connecting the red dots are solid indicating movement over water. Note that point "I" falls midway between two of these red dots. Then from position "I", a 57-h forward trajectory is constructed that ends at the location denoted by the time 0300 UTC 24 February. The entire trajectory from 0300 21 February to 0300 24 February covers a period of 72 h (three days). The choice of a 3-day trajectory is dictated by the action of a convective boundary layer—namely, an air column that exhibits heating from below (further discussion of the dynamics will follow in section 3).

For this RFE, *Salvia* followed a southerly-directed path about 50–100 km west of the trajectory shown in Fig. 3. A set of *Salvia's* upper-air soundings along this nearby path and covering a period of 18 h is displayed in Fig. 4. The temperature profile we use for our initial condition is highlighted in red (at 1728 UTC 21 February). The mixed-layer profiles from *Salvia* are not shown, but the profile at 1728 UTC 21 February exhibits a nearly constant value of 4.5 g kg^{-1} over a depth of 0.9 km. The warming and deepening of the mixed layer encountered by *Salvia* is most evident in this sequence of upper-air soundings shown in Fig. 4.

In support of GUFMEX, the NOAA P-3 aircraft made upper-air observations of temperature, pressure and humidity along steep ascent-descent pathways between 10 000 and 1000 ft (32808—3281 m) ASL. For this RFE, a set of observations was collected over the warm Loop Current shown in Fig. 3. Two of these upper-air observations fell close to the trajectory and are located at positions II (at 0000 UTC February 22, t = 6 h) and III (0300 UTC 22 February, t = 9 h)—again denoted by bulls-eyes. Location II falls on the trajectory and is midway between two of the red dots.

The thermodynamic profiles at locations II and III are displayed in Figs. 5 and 6, respectively. These profiles make it clear that it is nearly impossible to estimate the jumps in temperature (σ) and mixing ratio (μ)—accordingly we make no estimates for these variables. A list of these variables (see symbols list in section 3 and Fig. 7) from *Salvia* and the P-3 are displayed in Table 1. The difference in depth of the mixed layer is difficult to estimate from the P-3 soundings. It is reasonable to state that there is little difference in these heights but the height at the earlier time appears to be slightly greater.

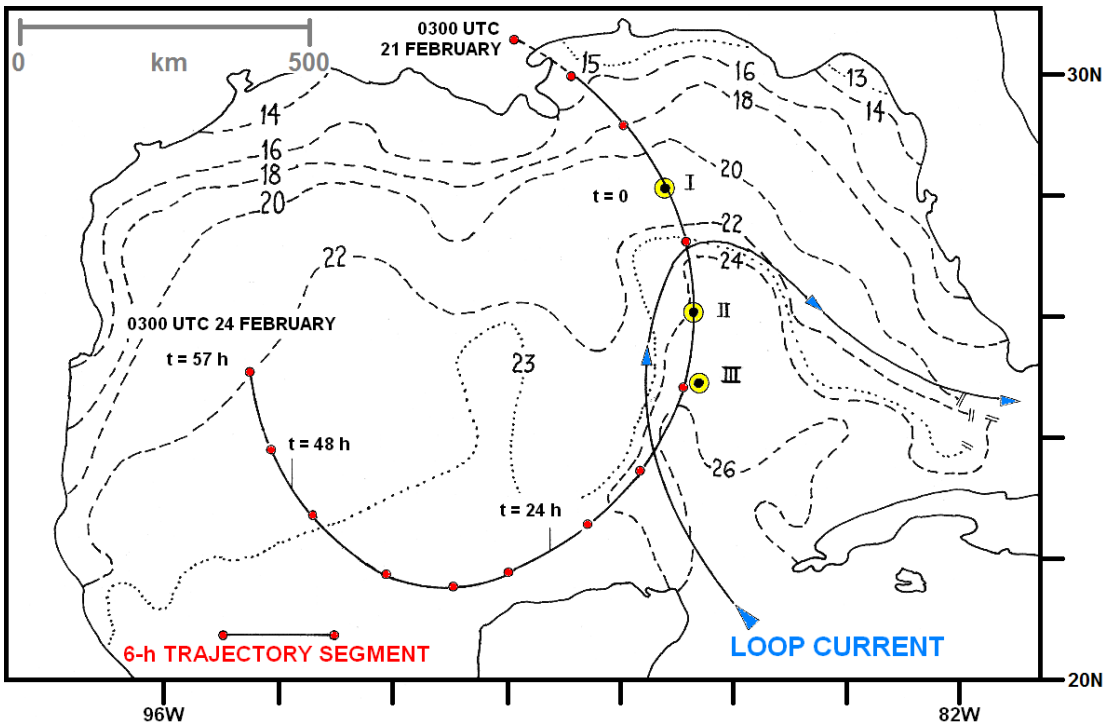

Figure 3: Trajectory (solid curve, red dots every 6 h) of surface air over the Gulf of Mexico, starting near New Orleans at 0300 UTC 21 February and terminating over the western Gulf at 0300 UTC 24 February 1988. Sea-surface temperature isopleths (°C) from 23 February are dashed and dotted. Loop Current axis is represented by a solid curve with blue triangles. Yellow-filled symbols as labeled represent I) an upper-air sounding from the deck of the *Salvia*; II and III) aircraft-based upper-air observations discussed in text.

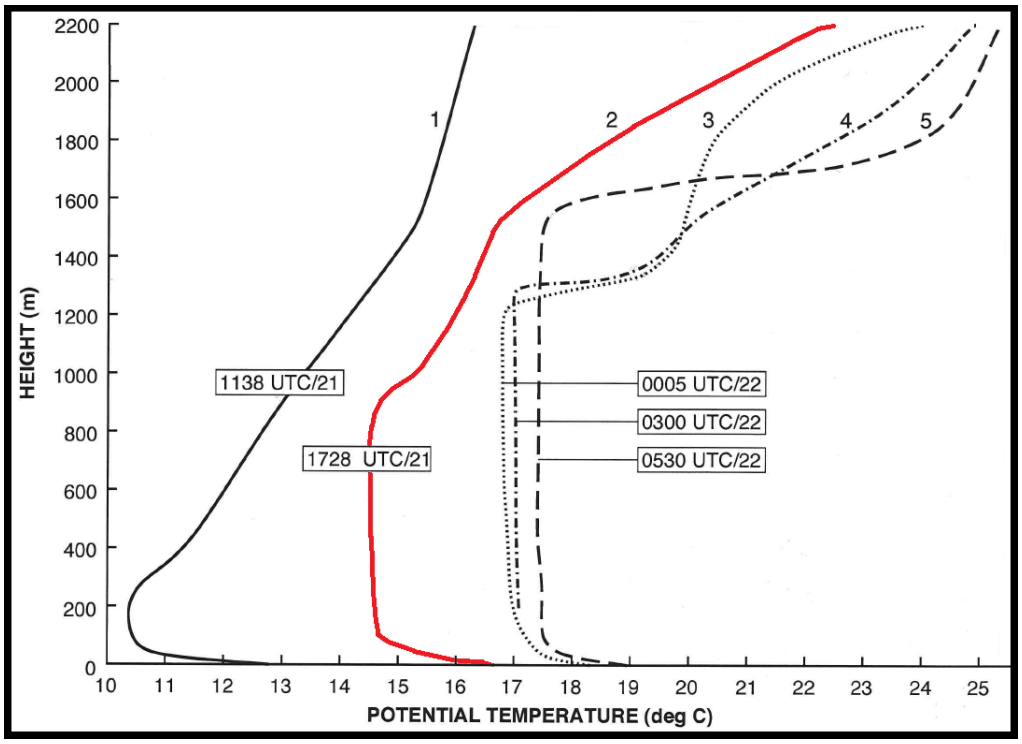

Figure 4: Potential temperature profiles measured from the *Salvia* at UTC times and dates labeled. Profile 2 (red curve) corresponds to the initial condition for the model in section 3.

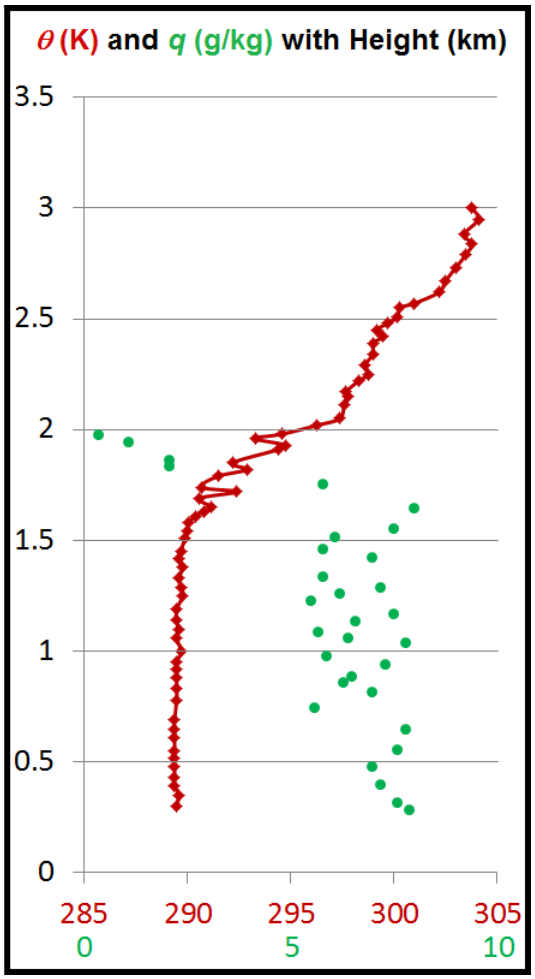

Figure 5: Vertical profiles measured at ≈0000 UTC 22 February 1988 from the NOAA P-3 aircraft: potential temperature θ (K, red) and mixing ratio q (g kg^{-1}, green), corresponding to location II in Fig. 3, using the scales shown.

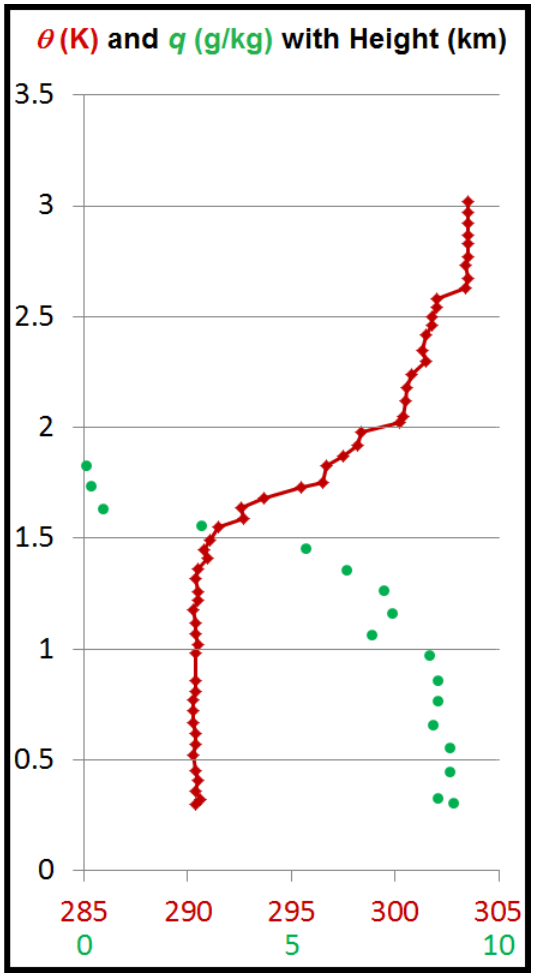

Figure 6: As in Fig. 5, but for 0300 UTC 22 February 1988 and location III in Fig. 3.

Table 1: Upper-air observations from *Salvia* (red curve in Figure 4) and P-3 (profiles in Figures 5 and 6). Variables are described in section 3.

Time (UTC)	Forecast Time (h)	θ (°C)	h (km)	σ (°C)	q (g kg^{-1})	μ (g kg^{-1})
1800 UTC 21 February	t = 0	14.50	0.90	0.50	4.50	−1.50
0000 UTC 22 February	t = 6	16.50	1.60	—	6.50	—
0300 UTC 22 February	t = 9	17.50	1.55	—	8.00	—

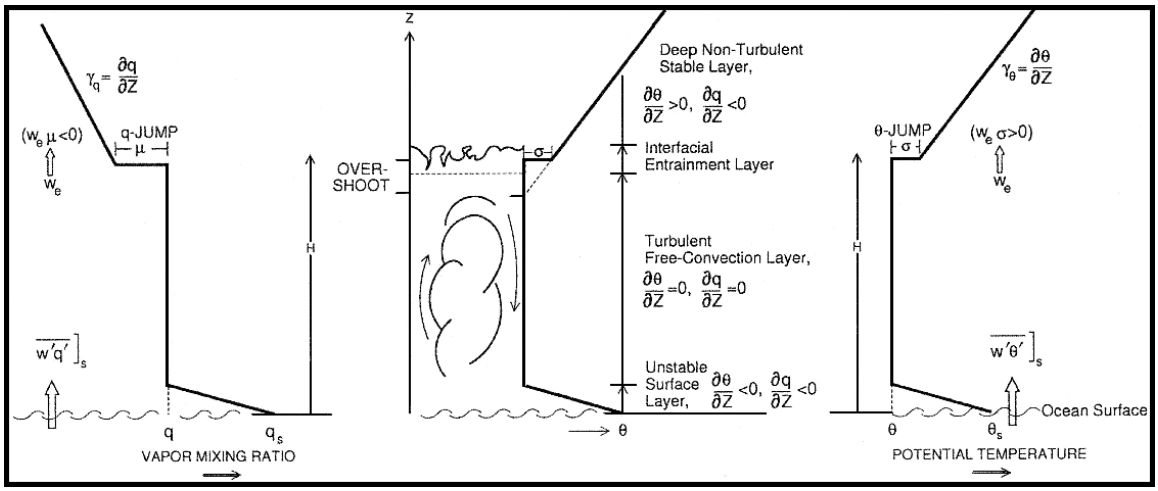

Figure 7: Schematic diagram of idealized vertical profiles of potential temperature θ and vapor mixing ratio q, through layers as labeled, from Lewis (2007). *Click image to enlarge.*

The changes in profiles I, II, and III can be interpreted as Lagrangian changes along the trajectory if the flow is steady over the timespan it takes air to move from location I to location III (the 9-h period from 1800 UTC 21 February to 0300 UTC 22 February). However, the flow field between locations I and III is not steady during this period. The degree of steadiness or absence of steadiness is estimated by analyzing surface winds from the NHC archive. The analysis of average surface winds (direction/speed) in this section yielded the following results: 360°/20 kt (10 m s^{-1}, at 1800 UTC 21 February), 010°/15 kt (8 m s^{-1}, at 0000 UTC 22 February), and 030°/15 kt (8 m s^{-1}, at 0300 UTC 22 February). This veering of wind is a direct result of the eastward movement of the anticyclone across the Gulf. In this absence of steadiness, sequential changes in the profiles I, II and III are only approximations to the Lagrangian changes along the trajectory. Despite this fact, there is minimal degradation to the objective of evaluating sensitivity of RFE predictability to the model's initial, boundary, and physical parameters.

3. The mixed-layer model

The upper-air observations from *Salvia* and the P-3 generally exhibit the presence of a mixed layer with near-constancy of potential temperature and mixing ratio that is overlain by a stable layer representative of the vertical structure of the air mass before it moved over the Gulf. The idealization of an abrupt discontinuity in structure at the interface of the mixed layer and the overlying stable air is absent. The fine-scale structure at this interface, especially from P-3 observations, indicates more of a diffuse interfacial entrainment layer several hundred meters thick. Nevertheless, the gradual deepening, warming and moistening of this mixing layer—especially noted from the time sequence of *Salvia's* upper-air observations—justifies our use of a classic mixed-layer model to study modification of the air mass in the RFE.

A schematic diagram displaying processes in the atmospheric mixed layer is shown in Fig. 7. The tacit assumption is that the column of air remains intact as it moves over the sea surface—that is, differential speed and direction of the wind are sufficiently small such that the column remains erect.

The symbols in Fig. 7 represent the following variables and parameters:

Forecast variables:

- θ: Potential temperature in the mixed layer
- h: Height of the mixed layer
- σ: Potential-temperature jump atop the mixed layer
- q: Vapor mixing ratio in the mixed layer
- μ: Mixing-ratio jump atop the mixed layer

Physical and empirical parameters:

- w: Large-scale subsidence
- w_e: Turbulent (entrainment) velocity
- C_θ: Bulk exchange coefficient for heat
- C_q: Bulk exchange coefficient for moisture
- V_s: Translation speed of the column

γ_θ: Lapse rate of potential temperature in the stable layer
γ_q: Lapse rate of vapor mixing ratio in the stable layer
κ: Entrainment coefficient
$w'\,\theta'_s$: Transport of heat into the mixed layer from below
$w'\,q'_s$: Transport of moisture into the mixed layer from below

Boundary conditions:

θ_s: Potential temperature at the air-sea interface
q_s: Saturated vapor mixing ratio at the air-sea interface

The governing equations for the mixed-layer model take the form:

$$\frac{d\theta}{dt} = C_\theta V_s[(1+k)(\theta_s - \theta)]h^{-1} \quad (1)$$

$$\frac{dh}{dt} = \kappa C_\theta V_s(\theta_s - \theta)\sigma^{-1} + w \quad (2)$$

$$\frac{d\sigma}{dt} = \gamma_\theta \frac{dh}{dt} - \frac{d\theta}{dt} - \gamma_\theta w \quad (3)$$

$$\frac{dq}{dt} = C_q V_s \left[(q_s - q) + \frac{\mu\kappa}{\sigma}(\theta_s - \theta)\right] h^{-1} \quad (4)$$

$$\frac{d\mu}{dt} = \gamma_q \frac{dh}{dt} - \frac{dq}{dt} - \gamma_q w \quad (5)$$

The jumps σ and μ are classified as "secondary variables" since their evolution is dependent on $\frac{d\theta}{dt}$ and $\frac{dh}{dt}$. A detailed development of these equations is found in Lewis (2007)—a development that follows the pioneering work of Lilly (1968, 1987) for the case of a cloud-free mixed layer.

As might be expected, the simplicity of the low-order Lagrangian model brings both strengths and weaknesses. Its strength rests on faithfulness to the growth of the mixed layer in the presence of buoyancy associated with the convective boundary layer. Further, its low order allows exploration of ensemble forecasting with a multitude of members, and that brings credibility to the statistics. However, its "one-dimensional" character (changes along a fixed trajectory) has disadvantages. Especially noticeable is the absence of influence from the surrounding atmosphere. Examination of HYSPLIT model trajectories in the online supplement for the late-March 2015 case makes it clear that the vertical structure of the thermodynamic profiles along the trajectory are three-dimensionally forced. In the HYSPLIT trajectories shown, the thermodynamic structure of return flow at Lake Charles (KLCH) has an influx of air at the 500-m level that came from a descending trajectory different from the trajectory of near-surface air (Fig. A5, online supplement). This contributes to the strength of the inversion above the mixed layer. Earlier studies as well as this study of the RFE through use of the Lagrangian model invariably have forecasted temperature jumps much weaker than observed, likely due to the absence of 3-dimensionality in the model.

4. Base-state solutions

The control elements for the mixed-layer model consist of initial conditions (forecast variables in the model), the boundary conditions (the saturated water-vapor mixing ratio q_s at the air-sea interface and SST), and the several parameters in the model. The mean values of initial conditions and parameters are given in Tables 2 and 3, respectively. Mean values of the boundary conditions are found in Table 4. The spacing (and number) of these boundary conditions is variable along the trajectory in order to capture detail in the SST field. The control vector for the mixed-layer model has 45 elements: 5 initial conditions, 6 parameters, and 34 boundary conditions—17 SST values and 17 q_s values.

The base-state solutions for the mixed-layer variables are shown in Figs. 8 and 9. The upper-air observations from *Salvia* and the P-3 are superimposed over the solutions shown in Figs. 8 and 9 and appear as + signs. The solutions for h and q are biased low (~10%) while the θ forecast is close to observations.

Buoyancy at the air-sea interface is driven by the temperature difference (θ_s–θ) and this difference becomes negligibly small as t→48 h (Fig. 8a). Accordingly, the mixed-layer dynamics cease to govern return flow beyond this point. After t = 48 h, the governing dynamics are associated with a veering wind as the next frontal system approaches and a boundary layer that is transitioning from neutral to stable. At 0000 UTC 24 February 1988 (t = 54 h), the sounding at Brownsville (KBRO), about 250 km west of the terminal point of the trajectory, registered a veering wind from the surface to 700 hPa—from 120 degrees/12 kt (6 m s^{-1}) at the surface to 255 degrees/16 kt (8 m s^{-1}) at 700 hPa.

In opposition to the steadiness of θ near t = 48 h, the water-vapor mixing ratio q continues to increase after this point in time. In the presence of weak buoyancy between t = 48 and 57 h, the vertical gradient of q is ≈3.5 g kg^{-1} along this section of the trajectory. Thus, the flux of heat at the air-sea interface is vanishingly small after t = 48 h, but the flux of vapor at the interface remains significant to the end of the trajectory (Fig. 8c). The height of the mixed layer decreases significantly after t = 36 h in response to the decrease in buoyancy and the persistence of subsidence.

Table 2: Mixed-layer model initial conditions (IC).

a) Mean values of the IC:

Time	θ (°C)	h (km)	σ (°C)	q (g kg^{-1})	μ (g kg^{-1})
1800 UTC 21 February	14.5	0.90	0.50	4.50	−1.50

b) Standard deviations of the IC:

Time	θ (°C)	h (km)	σ (°C)	q (g kg^{-1})	μ (g kg^{-1})
1800 UTC 21 February	1.0	0.075	0.20	0.50	0.50

Table 3: Mixed-layer model parameters (defined in section 3).

Parameter	Mean Values	Range
w (cm s^{-1})	−0.50	(−0.10) – (− 0.90)
κ (non-dimensional)	0.25	0.20–0.30
$V_s C_\theta$ (m s^{-1})	1.25×10^{-2}	1.00×10^{-2} –1.50×10^{-2}
$V_s C_q$ (m s^{-1})	1.25×10^{-2}	1.00×10^{-2} –1.50×10^{-2}
γ_θ (°C km^{-1})	6.0	5.0–7.0
γ_q (g kg^{-1} km^{-1})	−2 .0	(−1.0) – (−3.0)

Table 4: Mixed-layer boundary conditions (BC) as defined in section 3, by time and model hour t. Temporally corresponding values provided for SST and q_s.

Time	t: model time (h)	SST (°C)	q_s (g kg^{-1})
1800 UTC Feb 21	0	20.8	14.92
1900 UTC Feb 21	1	21.4	15.48
2000 UTC Feb 21	2	22.0	16.06
2100 UTC Feb 21	3	23.0	17.06
0000 UTC Feb 22	6	24.0	18.12
0300 UTC Feb 22	9	25.0	19.24
0400 UTC Feb 22	10	26.0	20.42
0500 UTC Feb 22	11	26.1	20.54
0600 UTC Feb 22	12	26.1	20.54
1200 UTC Feb 22	18	24.2	18.34
1800 UTC Feb 22	24	23.5	17.59
0000 UTC Feb 23	30	24.2	18.34
0600 UTC Feb 23	36	23.1	17.17
1200 UTC Feb 23	42	23.1	17.17
1800 UTC Feb 23	48	22.7	16.76
0000 UTC Feb 24	54	22.2	16.26
0300 UTC Feb24	57	22.0	16.06

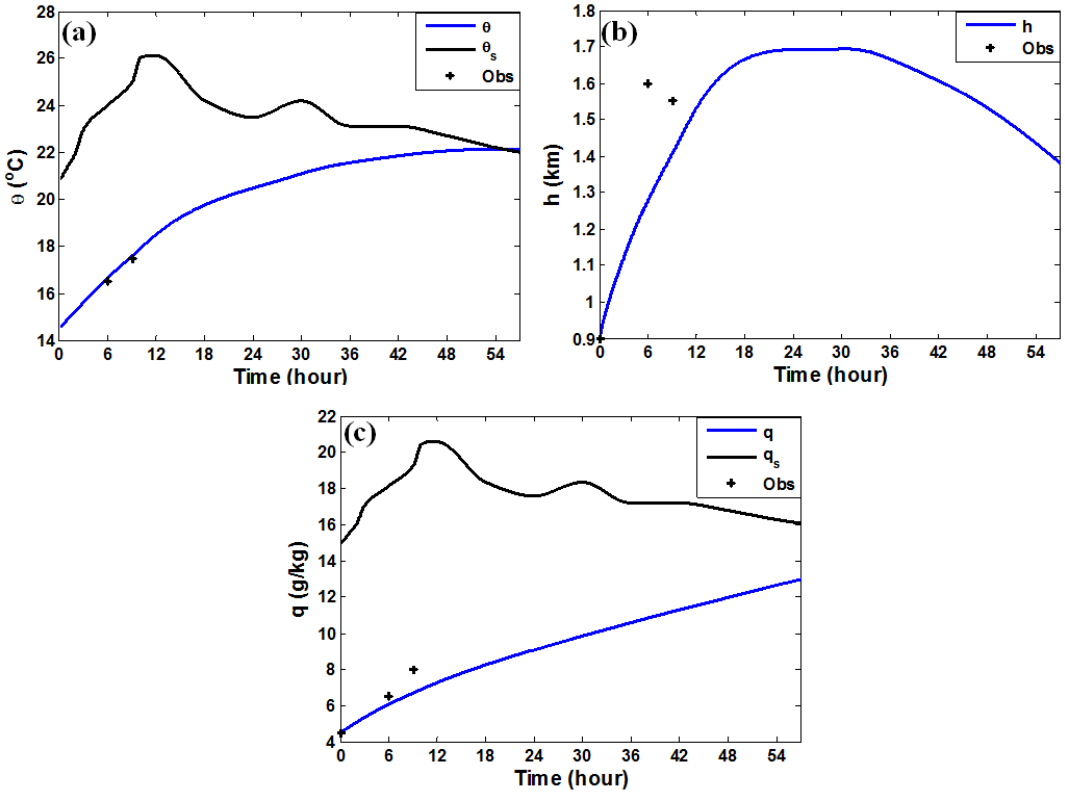

Figure 8: Base state for: a) potential temperature θ (blue) and saturated θ_s (black) in °C, b) mixed-layer height h (km) and c) water vapor mixing ratio q (blue) with saturated q_s (black) in g kg^{-1}). Observed values are marked with + in each panel. *Click image to enlarge.*

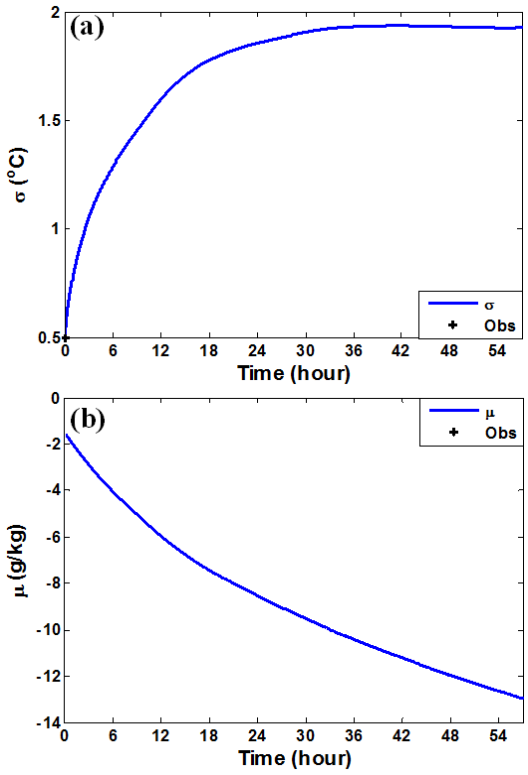

Figure 9 (left): As in Fig. 8 but for a) potential temperature jump σ (°C) and b) water vapor mixing ratio jump μ (g kg^{-1}). Initial conditions (observations) are indicated by "+" signs. *Click image to enlarge.*

5. Ensemble forecasting of θ, h, q, and qh

a. Background

The roots of ensemble forecasting stem from the work of mathematicians Stanislaw Ulam, John von Neumann, and Nicholas Metropolis at Los Alamos in the late 1940s (Metropolis and Ulam 1949; Metropolis 1987). For physics problems that could not be solved analytically, a stochastic-dynamic approach was invented that relied on the generation of random perturbations to elements of control—elements of the control vector that served as input to the governing equations. These mathematicians coined the phrase "Monte Carlo (MC) Method" for this process in deference to the famous casino in Monaco. With advances in computer power following WWII, especially the commercial

availability of massively parallel computational machinery in the 1980s, ensemble weather forecasting became a reality at the European Centre for Medium-Range Weather Forecasts (ECMWF) in the late 1980s (ECMWF 2012–2013) and National Center for Environmental Prediction (NCEP) since the early 1990s (Toth and Kalnay 1997). Ensemble forecasting has now become mainstream at operational weather prediction centers worldwide (Hirschberg et al. 2011).

We use the MC method to create ensemble forecasts for the February 1988 RFE. As mentioned earlier, the control vector for the mixed-layer model has 45 elements. Each element of the control vector is represented by either a normal probability density function (IC's and BC's) or a uniform probability density function (parameters). The mean values and standard deviations for the normal distributions are specified *a priori*, and the means and ranges for the uniform distributions are also specified *a priori*. The uniform distribution is used for parameters to avoid the random choice of physically unrealistic parameters—negative values for κ and γ_θ, and positive values for w and γ_q. The standard deviations for the IC's are found in Table 2b and the ranges for the parameters are found in Table 3. The standard deviations for the SST and saturated mixing ratios are 1.0°C and 1.0 g kg^{-1}, respectively.

A set of N control vectors is created by randomly choosing elements from the appropriate distributions. These control vectors yield a set of N forecasts. The second-order statistics (means and covariances) from this set of forecasts serve to define forecast uncertainty. And as stated in the Introduction, there is a strong linkage between forecast error and forecast uncertainty.

The mean values of initial conditions come from *Salvia's* upper-air sounding at 1800 UTC 21 February (at the point labeled I in Fig. 3). The IC uncertainties in primary variables (θ, h, q) are roughly ±10% of the mean values. The uncertainty in secondary variables (σ, μ) is larger than ±10% of the mean values and this reflects difficulty in assessing magnitudes of these idealized discontinuities. The mean values of parameters come from observations (γ_θ , γ_q, w, V_s) and the literature (κ, C_θ, C_q)—κ from Lilly (1968, 1987) and C_θ, C_q from Kondo (1975). For subsidence w, we have relied on the analysis of 900-hPa vertical motion over the Gulf at 1800 UTC 21 February (t = 0), as shown in Fig. 10 [also Fig. 10 in Lewis and Crisp (1992)]. The average value of subsidence over the trajectory was found by moving the analyzed vertical motion pattern eastward in concert with movement of the weakening surface anticyclone.

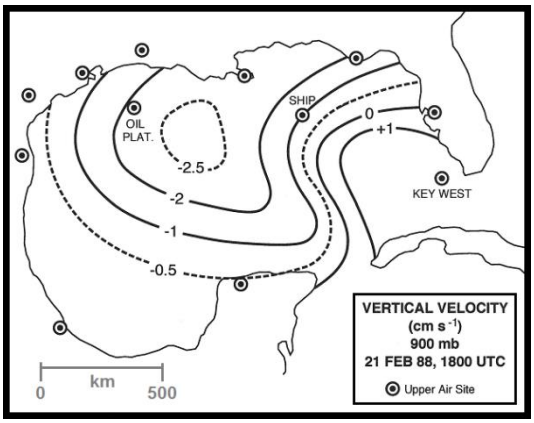

Figure 10: Field of 900-hPa vertical motion is superimposed over the upper-air stations that were used to make the calculation of vertical motion by the kinematic method. Marine upper-air sites (oil platform, ship and Key West, FL) are labeled. *Click image to enlarge.*

The model's small dimension and modest computational demands allow determination of forecast uncertainty due to IC alone, parameters alone, BC alone, and the full complement (FC) of uncertainties. We have the luxury of comparing ensemble statistics for an enormous number of members (N = 200 000) down to the dimension of the control (N ~50). A scalar that can be used to measure the "steadiness" of the ensemble statistics is the sum of the model variances (technically the trace of the model covariance matrix). Table 5 registers the trace as a function of N for both 24- and 48-h forecasts. The trace for N = 200 000 is taken as "truth", and the traces for other values of N are displayed in this table. The degree of non-steadiness or error in the statistics is measured by err_N (a difference of absolute values of the traces divided by truth). It becomes clear that the statistics become stable (steady) as the number of samples increases, and that the error is significant when N is the order of the control (~50–100 members). In short, the ensembles could be created with 1000–2000 members without much loss of accuracy. Nevertheless, results shown in this paper use N = 200 000.

Table 5: The number of ensemble members (N), trace of covariance matrix (tr_N), and the percentage of the error (err_N). The latter is determined using

$$err_N = \frac{abs(tr_N - tr_{200000})}{tr_{200000}} \times 100$$

a) t = 48 h:

N	tr_N	err_N (%)
200 000	3.1303	0
100 000	3.1538	0.7512
50 000	3.1509	0.6572
20 000	3.1599	0.9466
10 000	3.1711	1.3040
5000	3.0885	1.3356
2000	3.1907	1.9296
1000	3.2290	3.1515
100	3.9030	24.6848
50	4.1047	31.1293

b) t = 24 h:

N	tr_N	err_N (%)
200 000	1.7944	0
100 000	1.7977	0.1843
50 000	1.8044	0.5597
20 000	1.7925	0.1083
10 000	1.8115	0.9533
5000	1.7688	1.4239
2000	1.8196	1.4032
1000	1.8589	3.5930
100	2.0629	14.9658
50	2.1764	21.2904

b. Ensemble forecasts of θ, h *and* q

The ensemble forecasts of θ, h and q are shown in Figs. 11–13. Within each figure, the ensemble forecasts are configured as follows: a) upper-left corner—uncertainty in initial conditions alone, b) upper-right corner—uncertainty in boundary conditions alone, c) lower-left corner—uncertainty in parameters alone, and d) lower right-corner—uncertainty in all element of control, the full complement (FC) uncertainty. The dark curved lines in the panels of each figure depict the evolution of mean values over the 48-h period. The range about the mean values appears as the turquoise-colored ribbons. Within the range at any time, the probability density function has a bell-shaped structure similar to the normal probability density function (not shown). The number of outcomes within one and two standard deviations of the mean is typically within a few percent of those limits associated with the normal distribution. For the normal distribution, 68 (96) % of the outcomes fall within 1 (2) standard deviation(s) about the mean. The standard deviation curves associated with each variable and for each category of control are shown in Figs. 14–16.

Due to the nonlinearity of the mixed-layer model, the probability distribution of the FC [panel (d) in the figures] cannot generally be derived from knowledge of the probability distributions for the ICs [panel (a)], the BCs [panel (b)], and parameterizations [panel (c)]. However, at the initial time, the probability distribution for FC is identical to the probability distribution associated with IC alone.

The salient features of these ensemble forecasts follow:

- The forecast FC uncertainty exhibits the largest spread at all times for each of the variables and this uncertainty is primarily due to parameter uncertainty except in the case of the temperature ensemble where it is due to IC uncertainty (Fig. 11).
- The IC uncertainty for h and q retains influence well into the forecast period, and in fact its influence increases with time for h and increases slightly after t = 24 h for q, whereas IC uncertainty gradually decreases for θ to the point where its influence is minimal at t = 48 h.
- For each variable, the ensemble forecast spread due to BC uncertainty alone is generally the smallest of all categories despite its gradual increase with time.

c. Forecast of the combined variable qh

The variable qh is proportional to the mass of water vapor in the mixed-layer column. The mass per unit area is given by ρqh where ρ is the density of air. If q = 10 g kg^{-1} and h = 1 km, then the mass of water in this column is 1 g cm^{-2}. This is equivalent to a precipitable water (PW) depth of 1 cm, or 10 mm in units normally used for this variable (Glickman 2000). Figure 17 displays the ensemble evolution of this variable following the same format used in Figs. 11–16. The graphs of standard deviations due to

partitioned uncertainties appear in Fig. 18. And as expected from the statistics for *q* and *h* discussed earlier, the uncertainty in the ensemble forecast of *qh* is primarily due to parameter uncertainty.

d. Elaboration on FC statistics

In Table 6, magnitudes of the variables that bound the probability density within ±1 standard deviation of the mean values for the FC forecast at t = 24 h and t = 48 h are recorded. First, note that the probability of ensemble members falling within one standard deviation of the mean values range from 0.63 to 0.69—not far from the idealized value of 0.68 associated with the normal distribution. Next, note that the ratio of the upper bound to the lower bound is relatively large for *h* and *qh*—values of 1.83 and 1.65 at t = 48 h, respectively. In terms of PW, these bounding values of *qh* are 13.62 and 22.48 mm.

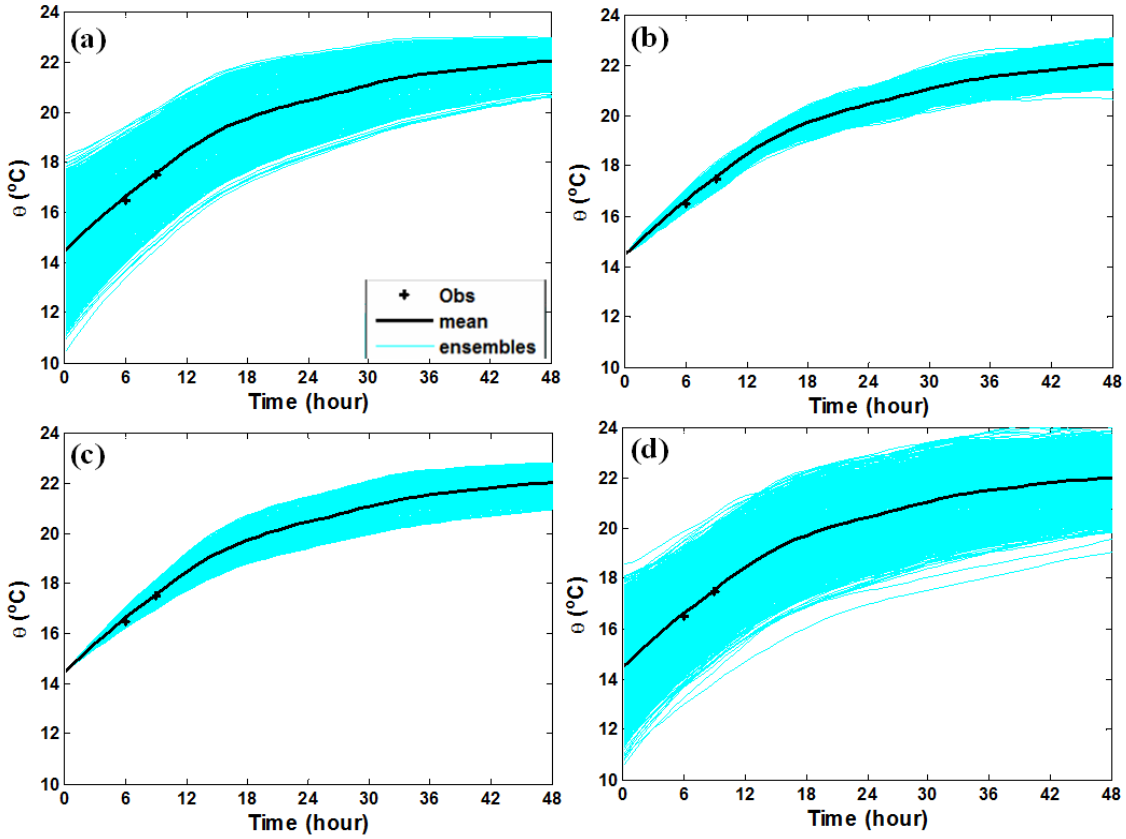

Figure 11: Ensemble predictions of potential temperature θ based on perturbations to the following elements of control: a) initial conditions alone, b) boundary conditions alone, c) parameters alone, and d) all elements of control (full complement of control). Ensemble averages appear as solid curves in each panel, and observations are indicated by "+" signs. *Click image to enlarge.*

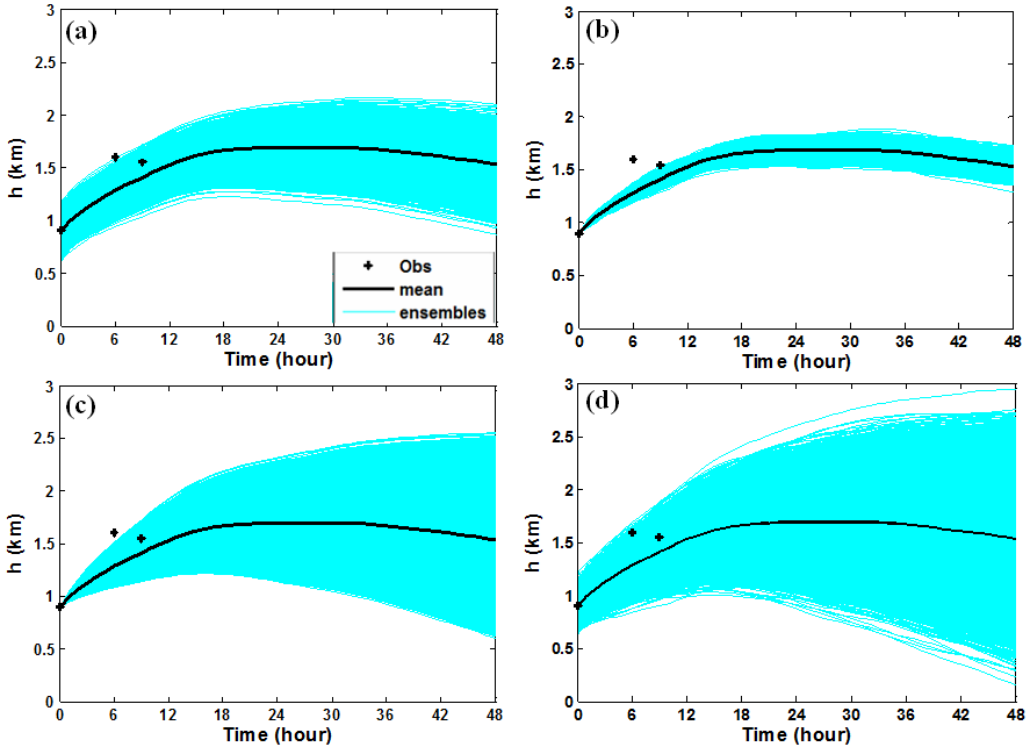

Figure 12: As in Fig. 11 but for mixed-layer height *h*. *Click image to enlarge.*

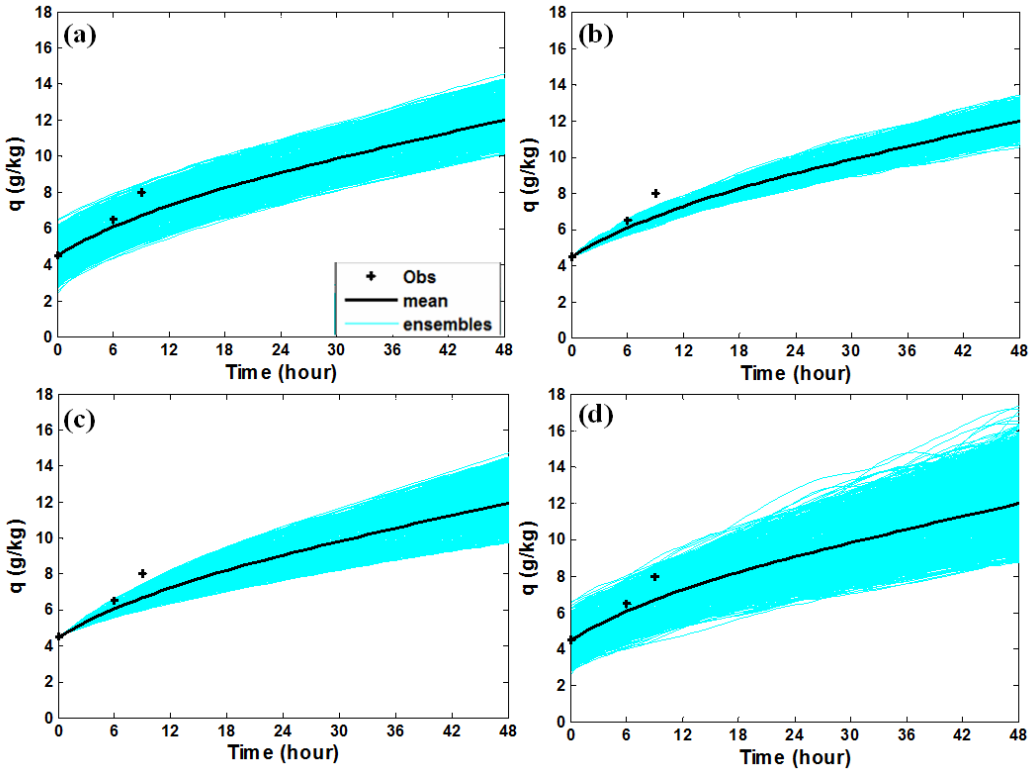

Figure 13: As in Fig. 11 but for water-vapor mixing ratio *q*. *Click image to enlarge.*

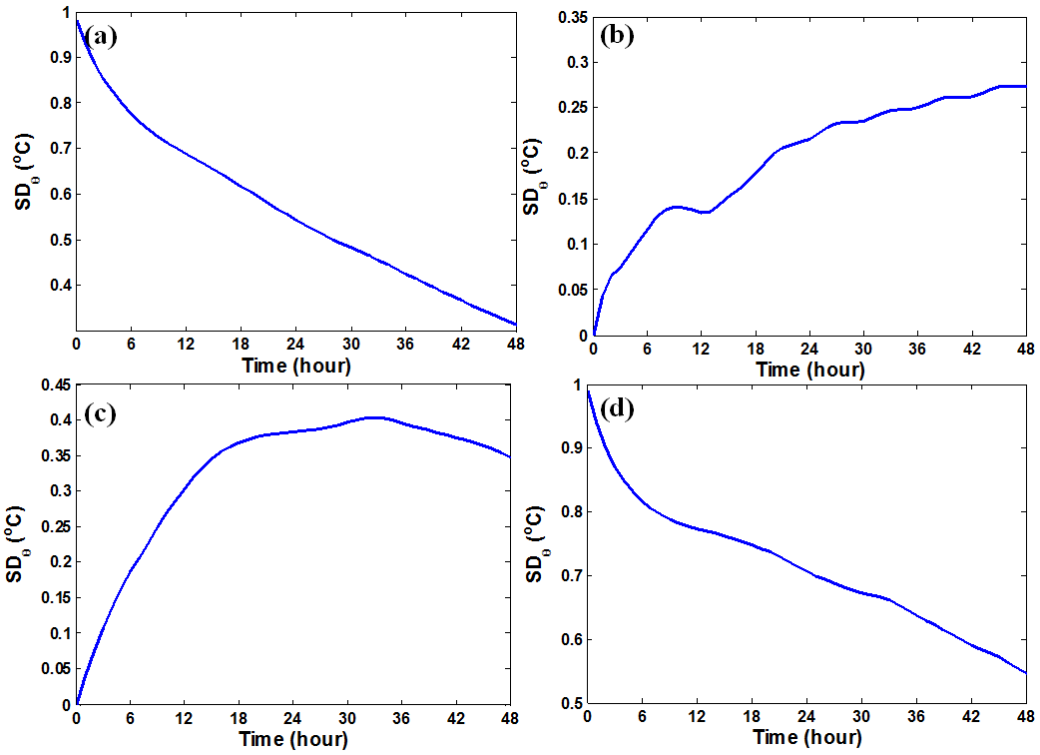

Figure 14: Standard deviation of the ensemble prediction for potential temperature θ when perturbations are due to: a) initial conditions alone, b) boundary conditions alone, c) parameters alone, and d) all elements of control (full complement of control). *Click image to enlarge.*

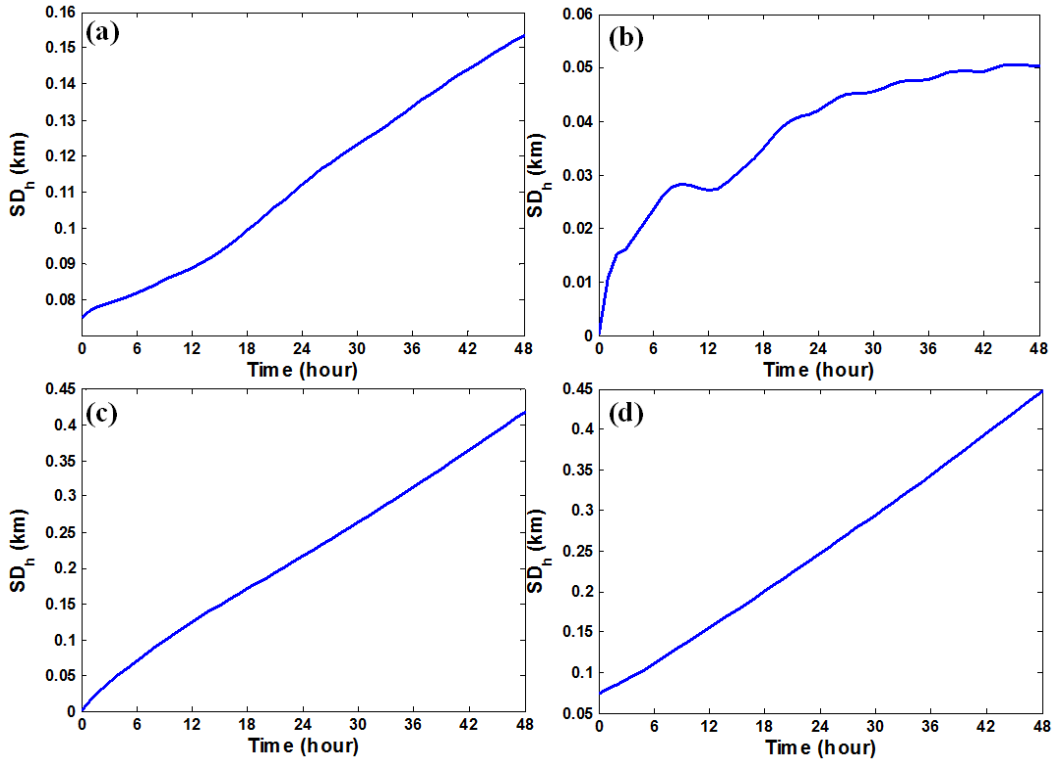

Figure 15: As in Fig. 14 but for standard deviation of the mixed-layer height *h*. *Click image to enlarge.*

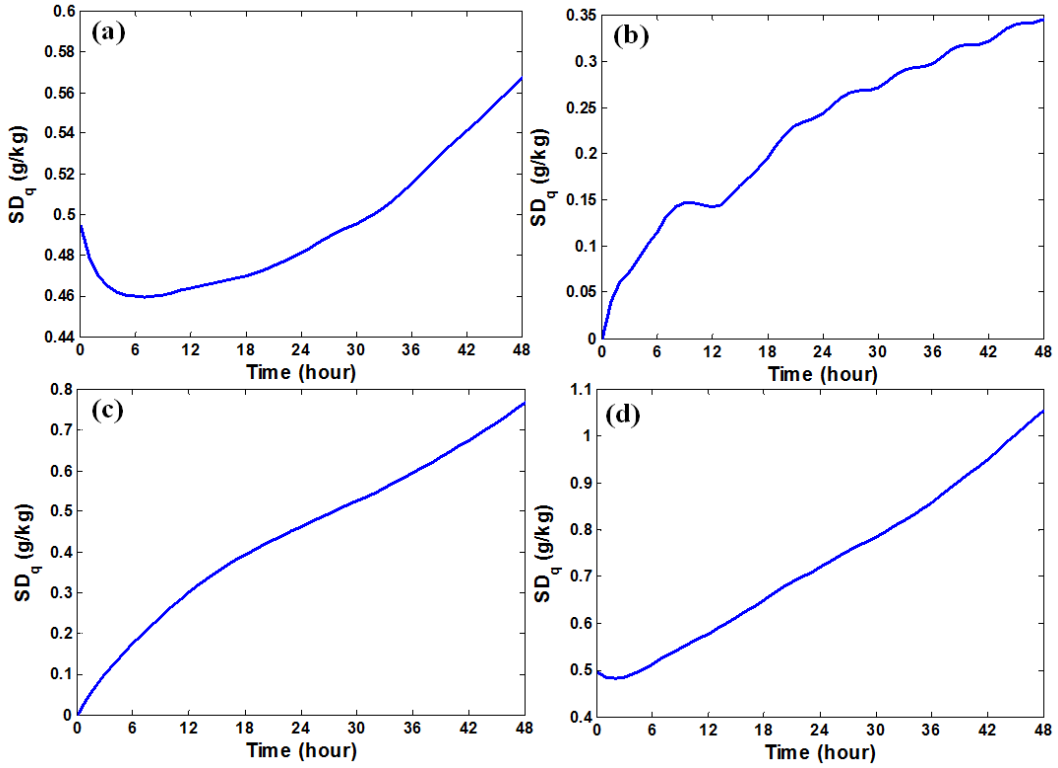

Figure 16: As in Fig. 14 but for standard deviation of water-vapor mixing ratio *q*. *Click image to enlarge.*

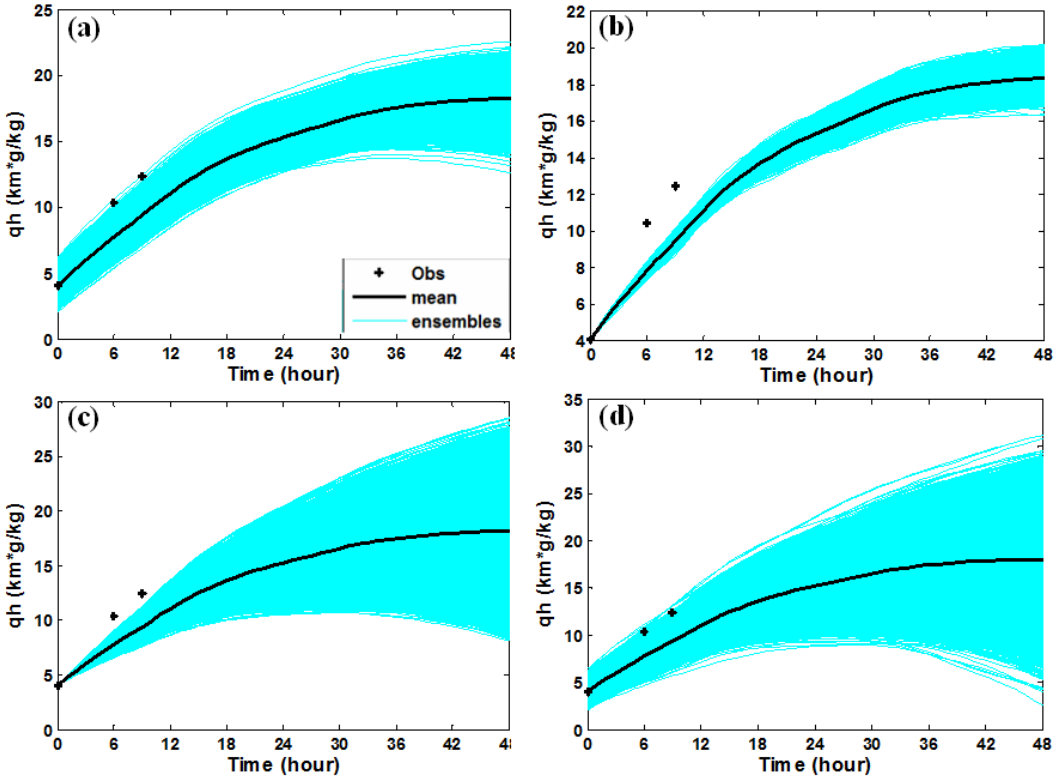

Figure 17: As in Fig. 11 but for mass of water vapor in the mixed-layer column *qh*. *Click image to enlarge.*

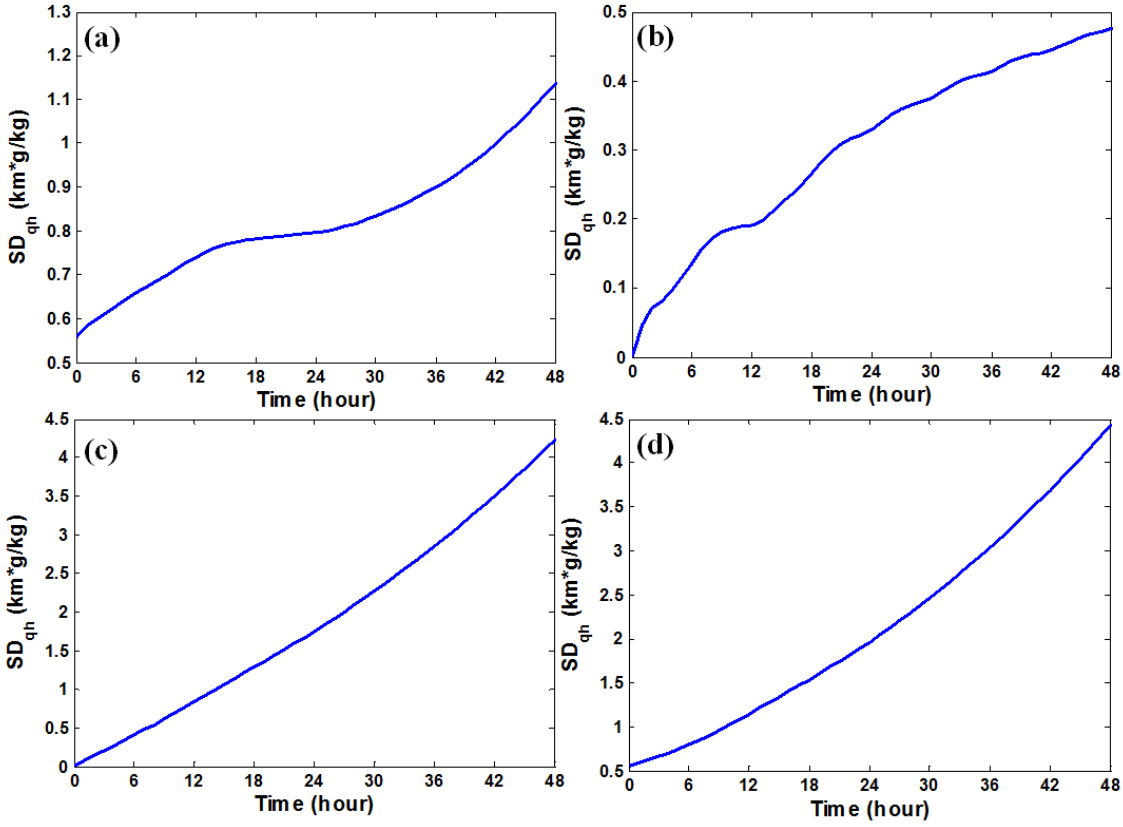

Figure 18: As in Fig. 14 but for standard deviation of water-vapor mass in the mixed-layer column *qh*. *Click image to enlarge.*

6. Qualitative assessment of operational model performance

In this contribution, a single research-model ensemble forecast of an RFE has been investigated. Further, the deterministic operational forecast of a recent RFE has been compared with observations—in essence, an isolated validation study. Results from both lines of research justify a more complete validation study. However, in the absence of such a comprehensive study, some advantage will accrue from qualitative assessment of operational model performance by SPC forecasters—forecasters who are tasked with using operational model guidance. A collective statement on this issue from SPC forecasters (Corfidi, Edwards and Thompson) follows:

Quantitative assessment of operational numerical weather prediction (NWP) models in RFE situations has not been accomplished at the SPC. Nevertheless, our collective experiences allow us to make statements of a qualitative nature about model performance. The North American Mesoscale (NAM) forecast system and Global Forecast System (GFS) are the two primary U.S.-based deterministic models used to predict RFEs as they relate to the potential for severe local storms. These models are verified operationally on a relatively coarse set of metrics that are mostly (but not entirely) outside the near-surface layer, such as comparisons with quality-controlled observed soundings of wind, temperature, moisture, and geopotential height at mandatory levels aloft (NCEP 2016).

Based on our RFE forecasting experience that spans the period from the 1980s to the present time, improved horizontal grid spacing and vertical resolution have had mixed results. The NAM, in particular, often remains overly aggressive with moistening the returning planetary boundary layer (PBL) in early-stage RFEs. This is a longstanding concern that goes back to the late 1990s when the NAM's predecessor, the Eta model, exhibited these same difficulties. An example of the Eta bias is shown in the 24-h and 48-h Eta forecast of a RFE in April 1999 (Fig. 19).

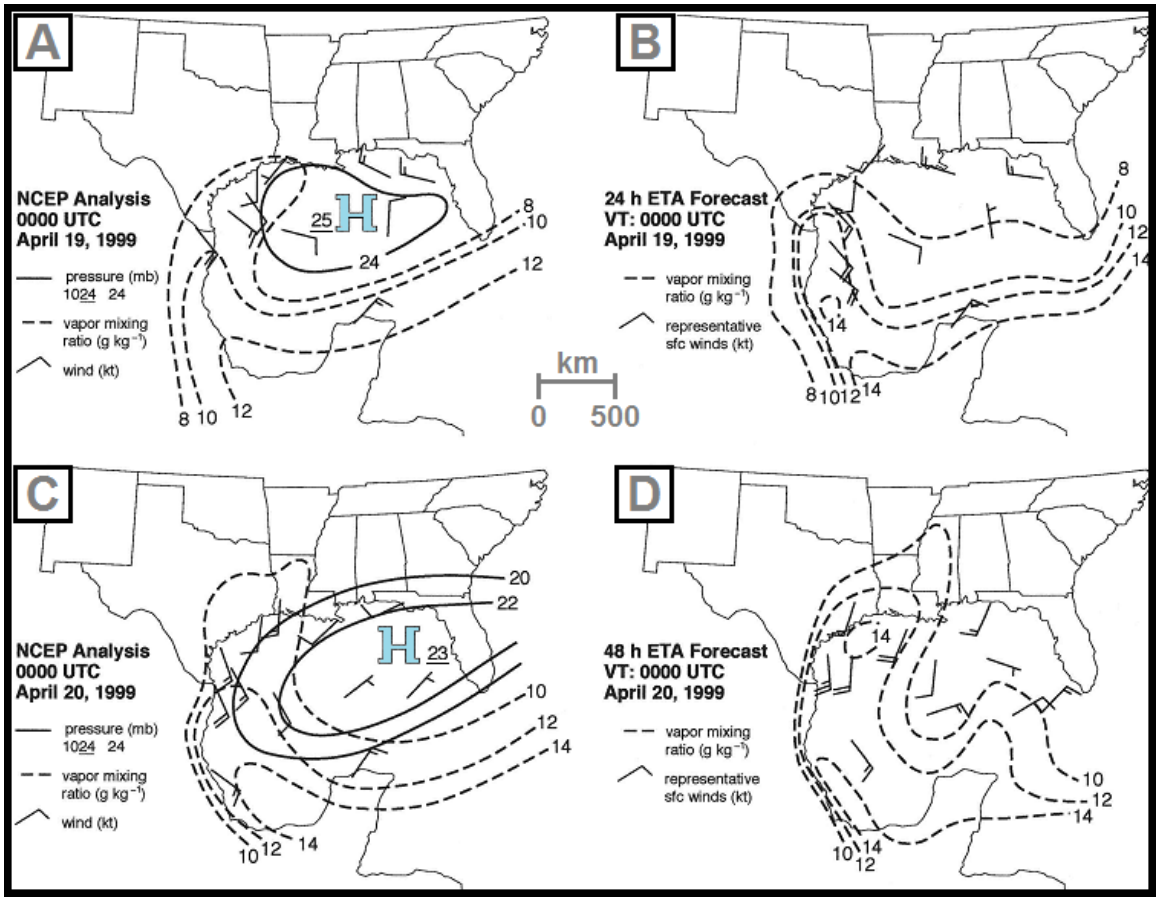

Figure 19: 24- and 48-h Eta model forecasts (right) of surface mixing ratio and winds shown alongside NCEP verification analyses with MSLP (left) for: a, b) 0000 UTC 19 April 1999; c, d) 0000 UTC 20 April 1999. Forecasted winds are selected from the full array of winds at model grid points. *Click image to enlarge.*

On the other hand, GFS forecasting of moisture can be highly variable and difficult to calibrate subjectively from one RFE to another. We suspect that some of the difficulty is related to PBL parameterization (See discussion below related to recent research by Cohen et al. 2015).

On the ensemble forecasting side, the Short-Range Ensemble Forecasting (SREF) system has gained nearly a decade of operational use for all aspects of severe-storm forecasting (e.g., Stensrud et al. 1999; Du et al. 2006; Guyer and Bright 2008). SPC also offers a menu of SREF-based forecast products online, specific to severe and fire weather. However, SREF member composition has changed several times, and as with the NAM and GFS, no systematic RFE verification has been done on any iteration of the SREF package.

Whether we consider deterministic or ensemble forecasting, a possible candidate in the poor forecasts of the moisture return in RFEs rests with the PBL parameterization. The growth of the boundary layer, especially as it reacts to the action of buoyancy, has been a major concern of researchers at NCEP/EMC since the late 1980s. The recent study by Cohen et al. (2015) sheds light on this major concern. As summarized by Cohen (A. Cohen, 2016, personal communication):

"A primary difference between the way in which the PBL is parameterized in the NAM and GFS models lies in the realm of local versus non-local mixing (Stensrud 2009). These two different techniques of representing vertical mixing have implications on simulated thermodynamic and kinematic profiles. The NAM model uses a local scheme at its core, and this could partly explain the shallower PBL and

PBL-moisture positive biases that we note during peak heating over the Great Plains. The local scheme invoked by the NAM model is the Mellor-Yamada-Janjić scheme (Janjić 1990, 1994). On the other hand, the GFS model uses a hybrid non-local/local scheme, which incorporates both local and non-local mixing processes in its formulation that could offset corresponding biases common to the local and non-local schemes, separately."

Cohen et al. (2015) is especially relevant to our concern with poor forecasts of moisture return in RFEs since it deals with cool-season boundary layers and severe storm potential.

7. Discussions and conclusions

Despite the Lagrangian mixed-layer model's relative simplicity compared to operational NWP models, it faithfully warmed, moistened, and deepened the convective boundary layer as cold, dry air moved out over the Gulf of Mexico's Loop Current in the RFE we studied. By limiting the forecast to changes along a single trajectory, the number of control elements in the model is small—in this case ~50 elements. Ensemble forecasts based on the Monte Carlo method categorically determined forecast uncertainty due to IC's alone, BC's alone, and parameters alone, as well as the combined influence of all elements of control. The ensemble experiments left little doubt that the uncertainty in the physical and empirical parameters had the most influence on forecast uncertainty. Surprisingly, IC uncertainty contributed more to forecast uncertainty than BC uncertainty.

The frequency distribution of *qh*, a variable proportional to PW, exhibited a considerable range bounding the ±1 standard deviation about the mean near the end of RFE's outflow phase. The bounding values of PW at t = 48 h were 13.6 and 22.5 mm. Based on the information in the final row of Table 6, a PW value in this range has a 63% probability of occurrence. These results point to the difficulty facing the operational deterministic prediction of PW where slight changes in model parameterization packages and initial conditions can lead to significantly different states.

Our simulation uncertainties and ranges of parameters in the well-mixed and contact layers indicate a large range of possible CAPE values as the convective boundary layer advects inland. From operational experience, this translates to uncertainties in assigning subjective outlook probabilities to severe-storm threats near the coast, particularly those 1–2 days or more beforehand that are strongly influenced by operational numerical models. Another important aspect of RFEs that has not been

Table 6: Variables at the boundaries of $\mu \pm 1\sigma$ (mean ±1 standard deviation). The probability of ensemble members falling within these boundaries is given in the last column labeled *prob* and the ratio of $(\mu+\sigma)/(\mu-\sigma)$ is given in the column labeled *ratio*.

a) t = 24h:

	μ	σ	$\mu-\sigma$	$\mu+\sigma$	*ratio*	*prob*
θ (°C)	20.44	0.71	19.73	21.15	1.07	0.68
h (km)	1.69	0.25	1.44	1.94	1.35	0.66
q (g kg^{-1})	9.06	0.72	8.34	9.78	1.17	0.68
hq (km g kg^{-1})	15.23	1.97	13.26	17.20	1.30	0.66

b) t = 48 h:

	μ	σ	$\mu-\sigma$	$\mu+\sigma$	*ratio*	*prob*
θ (°C)	22.00	0.55	21.45	22.55	1.05	0.68
h (km)	1.53	0.45	1.08	1.98	1.83	0.64
q (g kg^{-1})	11.98	1.05	10.93	13.03	1.19	0.69
hq (km g kg^{-1})	18.05	4.43	13.62	22.48	1.65	0.63

addressed in this paper is identification of synoptic-scale processes that have bearing on the depth of penetration and duration of the RFE. Crisp and Lewis (1992) gave some attention to this problem, but that investigation was limited to duration as a function of airmass type.

With parameter uncertainty looming large in forecast uncertainty of RFEs, it is prudent to test the parameterization packages in the operational models—especially the turbulent exchanges of moisture and heat at the air-sea interface. As stated in the conclusions of Lewis (2007), an ideal place to test the boundary-layer package is over the shelf water where the atmospheric instability is typically great (as exhibited in the *Salvia* sounding). In these conditions, the mixing-length hypothesis gives reasonable estimates of exchanges based on vertical differences in temperature and humidity adjoining the sea-air interface (Priestley 1959). These can then be compared with exchanges found from the operational models.

It is unlikely that routine rawinsonde observations will ever be taken over the Gulf. Meteorological and oceanographic observations obtained through remote sensing, especially from satellites, have taken center stage in data collection on both regional and global scales. It is currently possible to obtain integrated PW estimates from both microwave and infrared sensors aboard satellites[1]. Total columnar PW derived from GPS satellite signals (Bevis et al. 1992) is also routinely available to operational forecasters from several sites surrounding the Gulf in the U. S. and Merida, Mexico, along with an oil rig south of Louisiana. GPS PW data have been assimilated successfully into non-operational mesoscale models for improving precipitation predictions (e.g., Falvey and Beavan 2002). Since the full-atmospheric GPS PW data do not parse with certainty into specific vertical subsets, assumptions must be made about the fraction of total PW contained in the mixed layer near the surface.

The reader is advised to examine the latest PW products from NOAA's cooperative institute at Colorado State University—Cooperative Institute for Research in the Atmosphere (CIRA)—and the Space Science and Engineering Center (SSEC) at University of Wisconsin-Madison. At CIRA, Stan Kidder has developed a blended PW product from the Advanced Microwave Sounding Unit, GPS (TPW)—ground-based Suomi-net sensors, and the Special Sensor Microwave Imager (Figs. A3 and A6 in the online supplement). These fields can be found at the following website: http://www.ospo.noaa.gov/Products/bTPW/Overview.html. The SSEC product was developed by Gary Wade and stems from the GOES sounder that replaced the VAS (VISSR Atmospheric Sounder) in 1994. Fields from the SSEC product can be found at the following website: http://cimss.ssec.wisc.edu/goes/rt/viewdata.php?product=pwa_us.

[1]An overview of blended, satellite-based PW products is located at the following website: http://www.ospo.noaa.gov/Products/bTPW/Overview.html

ACKNOWLEDGMENTS

We are grateful to the crews aboard ship (USCG *Salvia*), plane (NOAA P-3), and oil platform (Mobil 617) that took observations during the February 1988 RFE. Jeanne Schneider, chief scientist aboard the P-3, supplied us with the thermodynamic profiles collected aboard this aircraft. Analysts and forecasters at the National Hurricane Center (NHC) were most helpful in timely deliverance of operational products. Steve Baig (retired NHC oceanographer) deserves special credit for making his superb SST analyses available to us.

Ariel Cohen's succinct summary of planetary boundary layer (PBL) schemes used in NCEP operations—a major component of his doctoral dissertation at University of Oklahoma—added important details to this study. Others who deserve our thanks include NCEP research meteorologist Zavisa Janjić and Mike Kaplan, Professor at University of Nevada-Reno. Joan O'Bannon, former graphics artist at NSSL, used her superb skill in taking hand-drawn maps and figures and creating electronic versions of them. Her handiwork is displayed in Figs. 2 and 7.

Finally, the astute comments and suggestions from the three formal reviewers (Lance Bosart, David Bright, and William Gallus) were followed and this led to an improved presentation of material in this paper.

APPENDIX

The accompanying Supplemental Material contains observations and operational numerical-model forecasts for an illustrative RFE that occurred during the 25–30 March 2015 period.

REFERENCES

Ball, F. K., 1960: Control of inversion height by surface heating. *Quart. J. Roy. Meteor. Soc.*, **86**, 483–494.

Bevis, M., S. Businger, T. A. Herring, C. Rocken, R. A. Anthes, and R. H. Ware, 1992. GPS meteorology: Remote sensing of atmospheric water vapor using the Global Positioning System. *J. Geophys. Res.*, **97**, 15787–15801.

Boyle, J., and L. Bosart, 1983: A cyclone/anticyclone couplet over North America: An example of anticyclone evolution. *Mon. Wea. Rev*., **111**, 1025–1045.

Cohen, A. E., S. M. Cavallo, M. C. Coniglio, and H. E. Brooks, 2015: A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern U.S. cold season severe weather environments. *Wea. Forecasting*, **30**, 591–612.

Crisp, C. A., and J. M. Lewis, 1992: Return flow over the Gulf of Mexico. Part I: A classificatory approach with a global historical perspective. *J. Appl. Meteor*., **31**, 868–881.

Dallavalle, P., and L. Bosart, 1975: A synoptic investigation of anticyclogenesis accompanying North American polar air outbreaks. *Mon. Wea. Rev*., **103**, 941–956.

Draxler, R. R. and G. D. Rolph, 2003 (cited 2015): HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) model. NOAA Air Resources Laboratory, Silver Spring, MD. [Available online via http://www.arl.noaa.gov/HYSPLIT_info.php.]

Du, J., J. McQueen, G. DiMego, Z. Toth, D. Jovic, B. Zhou, and H. Chuang, 2006: New dimension of NCEP Short-Range Ensemble Forecasting (SREF) system: Inclusion of WRF members. Preprints, *WMO Expert Team Meeting on Ensemble Prediction Systems*, Exeter, United Kingdom, 5 pp. [Available online at https://www.wmo.int/pages/prog/www/DPFS/Meetings/ET-EPS_Exeter2006/Doc6%285%29.pdf.]

ECMWF, 2012–2013: 20th anniversary of operational ensemble weather prediction at European Centre for Medium-Range Weather Forecasts. Newsletter 134. [Available online at www.ecmwf.int/publications/cms/get/ecmwfnews/305.]

Edwards, R., and S. J. Weiss, 1996: Comparisons between Gulf of Mexico sea surface temperature anomalies and southern U. S. severe thunderstorm frequency in the cool season. Preprints, *18th Conf. on Severe Local Storms*, San Francisco, CA, Amer. Meteor. Soc., 317–320.

Falvey, M., and J. Beavan, 2002: The impact of GPS precipitable water assimilation on mesoscale model retrievals of orographic rainfall during SALPEX '96. *Mon. Wea. Rev.*, **130**, 2874–2888.

Glickman, T. S., 2000: *Glossary of Meteorology*. 2nd Ed.. Amer. Meteor. Soc., 591.

Guyer, J. L., and D. R. Bright, 2008: Utility of Short-range Ensemble Forecast (SREF) guidance for forecasting the development of severe convection. Preprints, *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 13A.1.

Hammersley, J. M., and D. C. Handscomb, 1964: *Monte Carlo Methods*. Methuen, 169 pp.

Henry, W., 1979a: Some aspects of the fate of cold fronts in the Gulf of Mexico. *Mon. Wea. Rev.*, **107**, 1078–1082.

——, 1979b: An arbitrary method of separating tropical air from "return flow" polar air. *Natl. Wea. Dig*., **4**, 22–26.

Hirschberg, P. A., and Coauthors, 2011: A weather and climate enterprise strategic implementation plan for generating and communicating forecast uncertainty information. *Bull. Amer. Meteor. Soc*., ***92***, 1651–1666.

Janish, P. R., and S. W. Lyons, 1992: NGM performance during cold-air outbreaks and periods of return flow over the Gulf of Mexico with emphasis on moisture-field evolution. *J. Appl. Meteor*., **31**, 995–1017.

Janjić, Z. I., 1990: The step-mountain coordinate: Physical package. *Mon. Wea. Rev.*, **118**, 1429–1443.

——, 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945.

Kondo, J., 1975: Air-sea bulk transfer coefficients in diabatic conditions. *Bound.-Layer Meteor.*, **9**, 91–112.

Lewis, J. M., 1993: challenges and advantages of collecting upper-air data over the Gulf of Mexico. *Marine Tech. Soc. J.*, **27**, 55–65.

——, 2007: Use of a mixed layer model to investigate problems in operational prediction of return flow. *Mon. Wea. Rev.*, **135**, 2610–2628.

——, C. M. Hayden, R. Merrill, and J. M. Schneider, 1989: GUFMEX: A study of return flow in the Gulf of Mexico. *Bull. Amer. Meteor. Soc.*, **70**, 24–29.

——, and C. A. Crisp, 1992: Return flow in the Gulf of Mexico. Part II: Variability in return-flow thermodynamics inferred from trajectories over the Gulf. *J. Appl. Meteor.*, **31**, 882–898.

Lilly, D. K., 1968: Models of cloud-topped mixed layers under a strong inversion, *Quart. J. Roy. Meteor. Soc.*, **94**, 292–309.

——, 1987: Mixed layers and penetrative convection (unpublished lecture notes), School of Meteorology, University of Oklahoma, 15 pp. [Available from lead author.]

Liu, Q., J. M. Lewis, and J. M. Schneider, 1992: A study of cold-air modification over the Gulf of Mexico using *in situ* data and mixed-layer modeling. *J. Appl. Meteor.*, **31**, 909–924.

Manikin, G. S., K. E. Mitchell, S. J. Weiss, 2000: Eta model forecasts of return flow. Preprints, *20th Conf. on Severe Local Storms*, Orlando, FL, Amer. Meteor. Soc., 493–496.

——, ——,——, 2001: The handling of return flow in the Eta model. Preprints, *9th Conf. on Mesoscale Processes*, Fort Lauderdale, FL, Amer. Meteor. Soc., 96–99.

——, ——, B. S. Ferrier, and S. J. Weiss, 2002: Low level moisture in the Eta model: An update. Preprints, *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 615–618.

Metropolis, N., 1987: The beginning of the Monte Carlo Method. *Los Alamos Science* (Special Issue honoring Stanislaw Ulam), 125–130. [Available online at http://fas.org/sgp/othergov/doe/lanl/pubs/00326866.pdf.]

——, and S. Ulam, 1949: The Monte Carlo Method, *J. Amer. Statistical Assoc.*, **44**, 335–341.

NCEP, cited 2016: NCEP verification documentation, 4 pp. [Available online at http://www.nco.ncep.noaa.gov/sib/verification/NCEP_Verification_Documentation.pdf.]

Palmén, E., and C. Newton, 1951: On the three-dimensional motions in an outbreak of polar air. *J. Meteor.*, **8**, 25–39.

Priestley, C. H. B., 1959: *Turbulent Transfer in the Lower Atmosphere*. University of Chicago Press, 130 pp.

Riehl, H., 1988: General circulation studies in Chicago from the 1940's into the 1950's. *Palmén Memorial Symposium on Extratropical Cyclones*, Helsinki, Finland, Amer. Meteor. Soc., 4–5.

Stensrud, D. J., 2009: *Parameterization Schemes (Keys to Understanding Numerical Weather Prediction Models)*. Cambridge University Press, 480 pp.

——, H. E. Brooks, J. Du, M. S. Tracton, and E. Rogers, 1999: Using ensembles for short-range forecasting. *Mon. Wea. Rev.*, **127**, 433–446.

Thompson, R. L., J. M. Lewis, and R. A. Maddox, 1994: Autumnal return of tropical air to the Gulf of Mexico's coastal plain. *Wea. Forecasting*, **9**, 348–360.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.

Weiss, S. J., 1992: Some aspects of forecasting severe thunderstorms during cool-season return-flow episodes. *J. Appl. Meteor.*, **31**, 964–982.

REVIEWER COMMENTS

[Authors' responses in *blue italics.*]

REVIEWER A (William A. Gallus):

Initial Review:

Recommendation: Accept with minor revisions.

General comments: In general, I believe this paper is thorough and well-written, with its goal being to put some quantitative bounds on the types of errors that might be expected for any numerical weather prediction model as it simulates return flow from the Gulf. I have found some errors in the paper, and although they do need to be corrected (one involves an entire missing chunk of text), I believe overall the paper is relatively close to being in a form where it can be published. I will call the needed changes minor, and will list them below in the order that they appear in the paper.

[This] comment is broader and is meant to be something for you to think about (I'm not sure if it is absolutely necessary to do it). Your study uses a 1D model with a very large range of variations to create an enormous ensemble. I will admit that your technique is outside my area of expertise, and perhaps what you did is perfectly fine. However, at the end, I felt like the paper was missing something that would strengthen it—that being some discussion of how actual 3D models, preferably operational ones, did at simulating these variables for this event. If I understand your goal in the paper, it was to show how uncertainty in many parameters could interact in ways to create a very large range of uncertainty in important quantities like PW or CAPE by the time return flow reached shore. Your method seems reasonable, but the range seemed so large that it made me question if something like that really does happen today with our 3D models. Your online supplement shows performance of actual models for one event, I think, but uses a very old case and I believe some very old operational model output.

The case you focus on in the present study is also a very old case, and thus not much would be gained by adding in the operational guidance from 1988. However, if you could use archived data such as reanalysis data and re-run this case using the GFS or NAM that is used today, or perhaps even the WRF often used for research studies and pseudo-operationally, it would be very interesting to see what sort of errors occurred in those models for perhaps a 24h, 48h, and 72 h forecast for a region where the return flow comes ashore. The ideal situation would be if such extra material showed that indeed the errors were rather large, perhaps approaching your 1-standard-deviation range. If this material could be added, I would think the paper would pack more of a punch. However, if the 3D models end up showing very small errors, then I agree there may be nothing to gain, since presumably this does not invalidate your work, but would just be an unfortunate coincidence. As I said earlier, although I think this would be very helpful (and indeed I was expecting to see this), I realize it isn't completely necessary to make your point, so I will not insist on it. Instead, I just urge you to think about adding it.

[We] are not sure if your online supplement translated perfectly since the operational models we examined in the supplement applied to a case from last year (not a very old case). The results from the validation study in the online supplement indicated that the current models exhibit errors in low-level thermodynamic structure that are similar to earlier models such as the Eta (from the 1990s). You have a valid point about our limited study (a validation of a single case using current models and the prediction of uncertainty using an ensemble model from a very old case). In the presence of these limitations, we have strived to augment our study by including a new section (section 6 titled: Qualitative assessment of operational model performance) that examines trends in operational prediction of RFEs from the experiences of the three SPC forecasters involved in our paper. In this new section, we have also included a discussion of the PBL issues as they related to the operational models. Getting the fluxes correct which in turn determine the PW and height of the mixed layer in these RFEs situations is at the heart of the difficulties.

[Minor comments omitted...]
Second Review:

Reviewer recommendation: Accept.

REVIEWER B (David R. Bright):

Initial Review:

Reviewer recommendation: Accept with minor revisions.

General comment: The paper is acceptable with minor revisions and no further review is requested unless major changes are made in accordance with other reviews.

Summary: Overall this is a nice paper and I enjoyed reading it. The topic is interesting on several levels: modeling, forecasting, communicating uncertainty, ensemble and probabilistic prediction. Most of my comments are minor. The only section to give me pause was section 2, on the observations and Lagrangian trajectory. My concern within section 2 is not in its content, but rather in the timing and source of datasets which could be more clearly described and illustrated. I'll address my thoughts in the specific comments below. Otherwise, I appreciate the opportunity to review the manuscript, found the subject and scientific content interesting, and look forward to its publication.

Substantive specific comments: Not being a PBL modeling expert, I assume the results of this simple PBL model are directly extendable to more complex PBL schemes. Is that a correct statement?

Indeed, a more complex PBL scheme could be substituted for the simple "K-theory" scheme used in this model (K-theory: fluxes proportional to vertical differences in temperature and mixing ratio in the boundary layer). That being said, substitution of a more complex scheme would require input that is unavailable for our case study. Based on this inquiry of yours and similar questions from the other two reviewers, we added a new section 6 (qualitative assessment of operational model performance in RFE situations) before the discussion and conclusions. In this new section we address questions of this nature as well as questions about the operational ensemble model. Fortunately, we have access to recent information on the PBL used in operations at NCEP as collected by Ariel Cohen, a member of the SPC team.

I agree with your conclusions but you could also state the value of ensemble output in predicting severe storm[s] and QPF, especially for those 1–2+ days.

In the new section 6, we give some space to discussion of the ensemble output as it relates to severe storms. A very appropriate suggestion based on the absence of this information in the earlier version of the manuscript.

Supplemental Material: A short description of the PBL schemes in NAM and GFS would be nice…just a descriptive contrast comparing the complexity of current operational schemes to the slab ML model you used.

Again, this important issue is now discussed in our new section 6.

[Minor comments omitted...]

Second Review:

Reviewer recommendation: Accept.

REVIEWER C (Lance F. Bosart):

Initial Review:

Recommendation: Accept with major revisions.

Overview: I haven't thought much about the details of the return-flow problem in recent years, so I very much welcomed the opportunity to think about this problem again as I read through Lewis et al. (2016). The authors' conclusion that return flow forecast uncertainty is mostly driven by problems with model parameterization of physical processes and the distribution of water vapor in the return flow layer is well supported by their ensemble analysis. The authors' analysis is well supported observationally and numerically.

A larger-scale perspective: Reading through this paper also brought me back to work I did on North American anticyclones with my former graduate students too many years ago to count. In Dallavalle and Bosart (1975; A synoptic investigation of anticyclogenesis accompanying North American polar air outbreaks, *Mon. Wea. Rev.*, **103**, 941–957, DOI: http://dx.doi.org/10.1175/1520-0493(1975)103<0941:ASIOAA>2.0.CO;2), Paul and I attempted to distinguish between those polar anticyclones that penetrated the western Gulf of Mexico from those anticyclones that turned eastward along the Gulf coast.

In Boyle and Bosart (1983; A cyclone/anticyclone couplet over North America: An example of anticyclone evolution, *Mon. Wea. Rev.*, **111**, 1025–1045, http://journals.ametsoc.org/doi/pdf/10.1175/1520-0493%281983%29111%3C1025%3AACCONA%3E2.0.CO%3B2), Jim and I followed up on Dallavalle and Bosart (1975) and attempted to quantify the process whereby cold-core anticyclones transition to warm-core anticyclones. These two papers taught me that the behavior of cold anticyclones that drive cold air southward into the Gulf of Mexico is very much dependent upon both the configuration and evolution of the large-scale flow as well as subtleties on how the evolving and southward-moving cold air masses would interact with the Gulf of Mexico (e.g., surface sensible and latent heat fluxes).

Your retrospective views on the papers you published with your graduate students several decades ago are invaluable. In fact, so valuable, that we have reconstructed the introduction with your view as a centerpiece. These papers bring focus on the RFE anticyclone first studied with the aid of upper-air observations by Eric Palmén and Chester Newton in the early 1950s. Your statement, "…subtleties on how the evolving and southward-moving cold air masses would interact with the Gulf of Mexico (e. g., surface sensible and latent heat fluxes)", gave support to our effort to make ensemble forecasts with the intention of identifying the "subtleties". The Introduction has undergone major revision.

The earlier work I did with Dallavalle (has it really been 41 years?) and Boyle was extremely primitive by today's standards. The results of this paper are telling me that a science and operational opportunity exists to revisit these old problems from a fresh perspective using modern global gridded datasets in which ensemble forecasts can take center stage in an analysis of forecast uncertainty. This is certainly a subject ripe for further discussion at the upcoming HWT Spring Forecasting Experiment.

Your admonition to consider a re-visitation of these old problems using the modern global gridded datasets is absolutely on target. Even though it is out of our range for this paper, the suggestion is a good one. News of the HWT Spring Forecast Experiment is currently being disseminated, and of course our SPC leads who are co-authors of this paper are intimately aware of this experiment. A follow-on paper is most appropriate.

Medium-scale issue: A possible criticism of this paper is that the authors are doing an excellent job at documenting and understanding the trees, but at the expense of overlooking parts of the forest. I would like to recommend that the authors devote a little more attention to the various synoptic-scale processes that help to govern the strength and the depth of the return southerly flow following a cold-air outbreak. For example, how well are the models doing on forecasting the sea level pressure difference between, say, CRP and TPA which will govern the strength of the return southerly flow (and is related to Wes Junker's old rule of thumb that the area-averaged precipitation amount in inches can be related to the number of SLP contours analyzed every 4 hPa that cross the Gulf coast)?

Indeed, a most important component of a climatological study of RFEs. We have gone back and carefully examined Charlie Crisp's climatological study of RFEs (Crisp and Lewis, JAM, 1992). In retrospect, it would have been a valued component of Charlie's study to include "depth of penetration" as a function of the air mass categorizations. To Charlie's credit, he categorized events based on the air mass types (mP and cP and combinations), but he emphasized the duration of return flows under the action of the various air mass types—somewhat related to penetration depth but not exactly.

We have added a statement in the conclusions (third from last paragraph) that mentions this important aspect of RFEs that has not been addressed in this paper. The addition is found at the end of the third paragraph of section 7: Discussion and conclusions.

Likewise, if you take the SLP difference between, say, MSY and a NOAA buoy in the central Gulf of Mexico, you can derive good information on the overwater fetch of air parcels that cross the eastern Gulf coast and then recirculate northward. Shouldn't subtle differences in these trajectories, in conjunction with subtleties in the strength and depth of the return flow, matter in the grand scheme of things? Why not take advantage of the full physics ensemble runs of the operational GEFS and EPS to address some of the return flow issues?

Since we only included the simulation of the 1988 RFE and the single validation of a recent RFE in terms of NAM and GFS model performance, this suggestion and similar suggestions from the other two reviewers have led us to include a new section [section 6] *that qualitatively assesses operational model performance (including ensemble) with emphasis on the PBL packages.*

[Minor comments omitted...]

Second Review:

Reviewer recommendation: Accept.