4A.8 THE PERFORMANCE OF SEVEN WRF PLANETARY BOUNDARY LAYER / SURFACE LAYER PHYSICS SCHEMES OVER COMPLEX TERRAIN

Richard S. Penc, Jeffrey A. Smith*,
John W. Raby, and Robert E. Dumais Jr.
US Army Research Laboratory, White Sands Missile Range, New Mexico

1 INTRODUCTION

The US Army Research Laboratory (ARL) has been performing long-term research into application of the Advanced Research version of the Weather Research and Forecast (WRF-ARW) model (Skamarock et al. 2008) for battlefield short-range forecasting in a field-deployed location. The purpose of these forecasts is to provide weather support for mission planning and execution and to augment coarser resolution models for specific applications required by field deployed units.

There are three primary goals in this research. The ultimate goal is to deploy WRF in a forward location and produce timely and useful 0- to 3-h and perhaps 0- to 6-h forecasts (nowcasts) tailored to the individual end user. Secondly, we need to increase the resolution so that user needs are met. Specifically, we require a grid spacing of roughly 1 km to resolve approximately 5-km-scale atmospheric phenomena that are necessary to resolve the detailed flow field over complex terrain.

Thirdly, there is a need to provide a measure of forecast uncertainty. Since WRE-N system operators are unlikely to have a meteorological background, the need to express forecast confidence is essential. There are a number of ways to achieve this. One is to use a time-lagged ensemble (Lu et al. 2007), which involves setting up and running the model in a rapid update cycle mode and compiling statistics and variances based on sequential model output. Another method to achieve the goal of developing a measure of uncertainty involves using physics-based ensembles (Stensrud et al. 2000). This method involves making several model runs with different physics packages selected and calculating model ensemble statistics from the model runs.

The goal of this aspect of the current research is to examine the performance of a number of physics packages to see which performs best in a complex terrain environment. We selected a southern California domain that has been widely used and documented at ARL (Dumais et al. 2009, 2013; Dyer et al. 2015, 2016; Foley et al. 2015). The inner nests of this domain include coastal, transitional, mountainous, marine, urban, agricultural, forested, and desert environments. To focus on and examine the verification of surface and near-surface properties and focus on the diurnal variation of the convective boundary layer (BL) during weak synoptic forcing, we examined a number of planetary BL (PBL) options using the latest version of WRF-ARW (v.3.8.1, when this research began).

The primary goal in the current research is to determine which of the BL/SL parameterization schemes works best for generalized forward deployments where the deployed location is not known in advance. Because the modeling system is preconfigured, one scheme will be set up for use in the deployed location, regardless of location. Our modeling domain was also chosen to represent a number of geographic regions focusing on complex terrain. This goal is challenging from the standpoint of not only capturing a complex range of physical processes, land use, and elevations, but also accounting for the steep gradients in elevation, which can potentially make achieving modeling stability criteria more difficult. Our evaluation of these 7 schemes is based largely on surface meteorological parameters, including temperature, dew point, and wind speed. In addition, we briefly examine the model derived PBL depth, a parameter important in our application. Talagrand diagrams were constructed to examine the feasibility of using a BL physics based ensemble to potentially derive measures of model uncertainty.

* *Corresponding author address:* Jeffrey A. Smith, US ARL, WSMR, NM, 88002, e-mail: jeffrey.a.smith1.civ@mail.mil

2 MODEL CONFIGURATION

The numerical model used in this analysis, the WRF-ARW (Skamarock 2008), is a community weather forecast model designed and supported by NCAR. For the purposes of this study, WRF v3.8.1 was selected. This study uses a triple nest configuration of WRF-ARW centered approximately near San Diego, California (Figure 1). The domain includes a variety of geography, land use, and topography. The marine environment, coastal environment, large valley, gently sloping desert, and mountainous terrain are all represented. Regarding land use, there are urban, suburban, agricultural, grassland, arid, and mountainous areas with a variety of subtropical and mid-latitude vegetation including grassland and forest. The domains are shown in Fig. 6. WRE-N was employed an outer 9-km nest of 175 × 175 grid points (1566 × 1566 km) (D1), 242 × 241 grid points on the middle 3-km nest (720 × 720km) (D2), and 127 × 127 grid points on the 1-km inner nest (126 × 126 km) (D3). The model top was selected to be 10mb. Although we concentrated on analysis of the 09-10 Feb 2012 event, additional simulations were also run for the other 4 days in our dataset. We chose to concentrate on this case since the primary intent of this study was to examine the model's ability to capture the development of the daytime CBL and transition to the nocturnal BL.

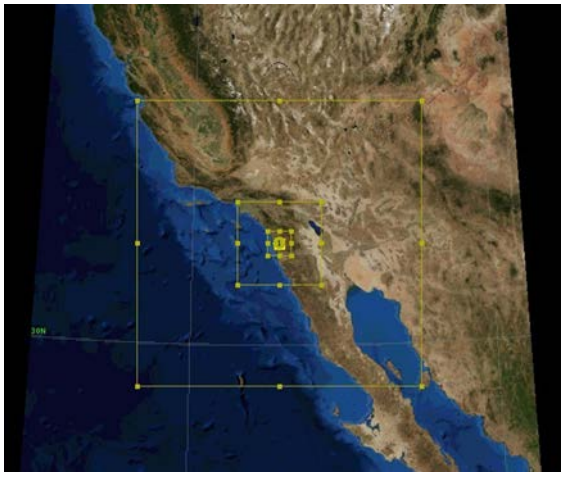

Figure 1. Location of the triple nested model domain used in these simulations.

The model specifications common to all 7 experiments using WRF-ARW as employed in this study are shown in Table 1. The FDDA option used in these model runs is based on observation nudging (Liu et al. 2005; Deng et al. 2009). This option is much less computationally expensive than traditional 4-D variational data assimilation (Huang et al. 2009) or ensemble Kalman filtering (Zupanski et al. 2008). Table 2 lists some of the data assimilated into the first 6h of the 24h forecast.

Table 1: Common configuration used in the WRF simulations for this study.

Namelist Parameter	Option Selected
Shortwave Radiation	Dudhia Scheme
Longwave Radiation	RRTM
Explicit moist microphysics	Thompson
Cumulus parameterization	Kain-Fritsch 9km only, explicit 1, 3km
PBL scheme	Varies (elsewhere)
Surface layer	Paired with PBL scheme
Land Surface Scheme	NOAH
Time step to grid ratio (s/km)	3:1
Horizontal subgrid diffusion	Second-order on coordinate surfaces
Subgrid turbulence closure	Horizontal Smagorinsky first order

Table 2: Selected data assimilation switches as configured for this study.

setting name	setting value
use_tamdar	yes
use_madis_mesonet	yes
use_madis_profiler_npn	yes
use_madis_acars	yes
use_madis_maritime	yes
use_madis_metar	yes
use_madis_raob	yes
use_madis_sao	yes
use_madis_satwin	no
use_madis_satwind1h	no
geog_data_res	2m, 30s, 30s

The 7 PBL/SL parameterization combinations selected for this study are listed in Table 3. In each of these, the default SL scheme was coupled with the PBL scheme being tested because they are generally accepted and the most widely used by researchers. Where there was a matching SL, we used that option. For the others, where multiple options existed for SL choice, we used the revised MM5 SL scheme. That scheme is reported to work with many of the PBL options.

Table 3: PBL/SL combinations evaluated.

Case no.	PBL/SL option	PBL scheme	SL scheme
1	5 / 5	Mellor-Yamada Nakanishi and Niino (MYNN)	Nakanishi and Niino PBL's SL scheme
2	11 / 1	Shin-Hong Scheme (SH)	Revised MM5 SL scheme
3	2 / 2	Mellor-Yamada-Janjic Scheme (MYJ)	Eta similarity SL scheme
4	1 / 1	Yonsei University Scheme (YSU)	Revised MM5 SL scheme
5	8 / 1	Bougeault-Lacarrère PBL (BouLac)	Revised MM5 SL scheme
6	4 / 4	Quasi-Normal Scale Elimination (QNSE)	QNSE PBL scheme's SL option
7	7 / 1	Asymmetric Convective Model (ACM2)	Revised MM5 SL scheme

3 RESULTS

Model bias and RMSE were calculated for each of the 7 members in the ensemble for each hour of the simulation, including the data assimilation (hours 1–6), the nowcast period (hours 7–12), and the extended forecast (hours 13–24). Over the model domain, these periods roughly correspond to morning, afternoon, and nighttime hours, respectively. This corresponds to, respectively, 1300 UTC 09 February 2012 through 1800 UTC 09 February (0500–1000 Pacific Standard Time [PST]), 1900 UTC 09 February 2012 through 0000 UTC 10 February 2012 (1100–1600 PST), and 0100 UTC 10 February 2012 through 1200 UTC 10 February 2012 (1700–0400 PST). In addition, the overall model statistics were computed (hours 1–24) and reflect the period 1300 UTC 09 February 2012 through 1200 UTC 10 February 2012.

All of the analyses we present apply to the innermost D3 (1-km) domain. The temperature bias is shown in Fig. 2. During the data assimilation period (1200–1800 UTC) all of the schemes show a small forecast bias, typically less than 1 K. While the ACM2, BouLac, MYJ, SH, and YSU schemes perform very similarly, the MYNN and QNSE schemes diverge, showing a negative bias, underestimating the surface temperature. The former schemes slightly overestimate the surface temperature at 2m for the hours 1400 through 1600 UTC. After hour 4 of the simulation, all of the schemes converge and tend toward underestimating the surface temperature.

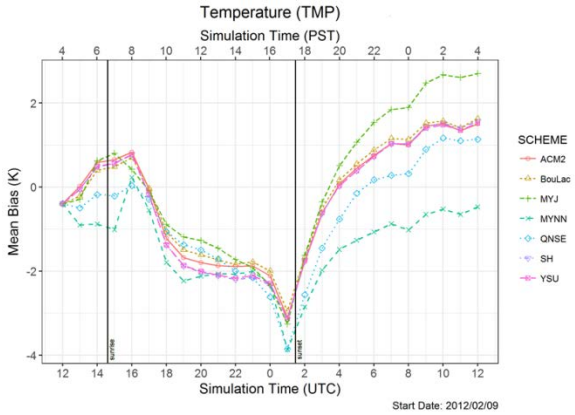

Figure 2: Temperature mean bias for the 7 members.

The bias for the DPT is shown in Fig. 3. For the majority of the forecast cycle, the DPT (at the surface) is underestimated by the model as shown by the negative bias, with the exception being between 0000 and 0200 UTC (1600–1800 PST). During these 3 h, which occur in late afternoon leading into early evening, the DPT is overestimated. Throughout the simulation, there is larger spread among the individual schemes than for the temperature bias.

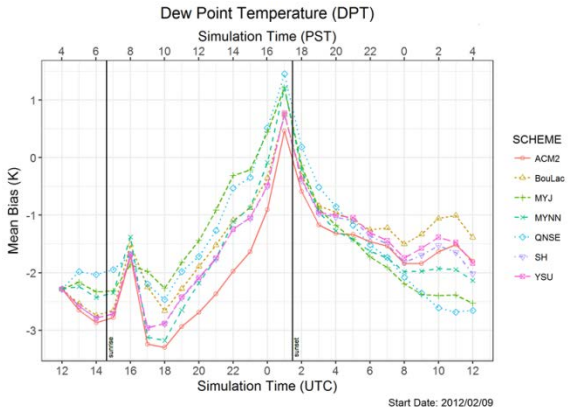

Figure 3: Dew point mean bias for the 7 members.

Looking at the surface wind speed (Fig. 4) bias we again see differing behavior for the 3 periods: assimilation, nowcast, and extended forecast. The model overestimates the surface wind during data assimilation by as much as 2 ms^{-1} but then quickly settles down to near zero

bias by 4 h into the simulation. There is more spread in the model bias by the nowcast period, with the greatest bias occurring with the QNSE scheme. QNSE overestimates the wind speed by >1 ms^{-1} from hours 9–10 of the simulation. In contrast, the ACM2 scheme underestimates the wind by as much as 1 ms^{-1} around 8 h into the simulation. As in all of the previous analyses, the model bias contrast between the 7 schemes is most pronounced for the latter half of the simulation time. For wind speed bias, the best performers at nighttime are YSU, SH, and ACM2, with near zero bias for the extended forecast period. The worst performer for the extended forecast is QNSE, followed by MYJ, with a mean bias error of approximately 1 ms^{-1}, overestimating the surface wind. Note that the physics differences appear to be most pronounced during nighttime.

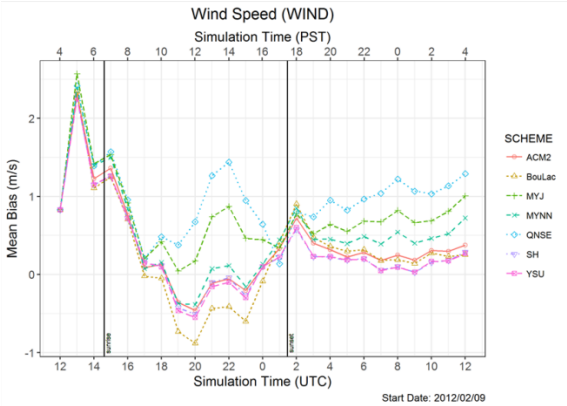

Figure 4: Wind speed mean bias for the 7 members.

We also performed an analysis of the model RMSE. The surface (2m) temperature RMSE for each of the members is shown in Fig. 5. The RMSE averages approximately 2.6 K at the beginning of the simulation time. The MYNN scheme shows the largest RMSE for the initial 4 h while QNSE shows the lowest. After 1400 UTC, the RMSE decreases to around 2 K for all of the members. For the nowcast period (6–12 h simulation time), the RMSE increases fairly linearly from about 1.5 to 3.5 K, with the highest RMSE associated with the MYNN and QNSE schemes. After 15 h, the RMSE of the members typically varies from 2.5 to 3.5 K. BouLac, ACM2, SH, and YSU follow each other closely toward the ending hours of the simulation in the extended forecast. Since SH is based on YSU, the similarity between these 2 schemes is not unexpected. The closeness of these schemes reflects the similarity in the formulation of the physics for the nighttime case, and these times reflect the nighttime scenario over the model domain. The increase of RMSE toward the end of the simulation is not unexpected since the errors tend to be cumulative.

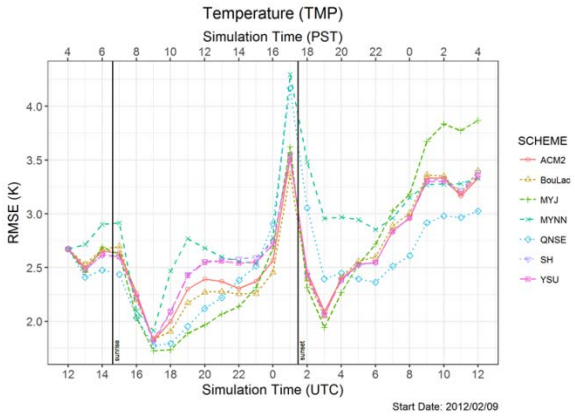

Figure 5: Temperature RMSE for the 7 members

Figure 6 shows the DPT RMSE for each of the schemes tested. The mean RMSE at the point of initialization is approximately 3.3 K, which is significant. As data assimilation progresses, the RMSE increases to approximately 3.8 K, with an increased spread among the schemes tested. During the nowcast period the RMSE decreases from approximately 4.0 to 2.5 K, corresponding to the development of the daytime CBL, and decreasing by late afternoon (1600–1800 PST). There is little spread among the model members. The most notable spread among the model members occurs during the early part of the nowcast period. MYJ, QNSE, and BouLac perform the best during this time with the lowest RMSE of all the members. MYNN and ACM2 have the highest RMSE of the 7 members during the period extending from the end of data assimilation through the early part of the 6-h nowcast period. There is overall no clearly superior scheme to choose from.

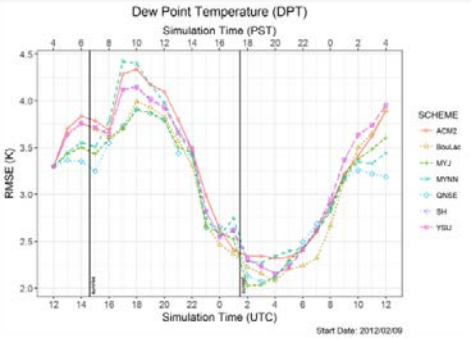

Figure 6: Dew point RMSE for the 7 members.

The wind speed RMSE for the 7 schemes is shown in Fig. 7. At model initialization, the RMSE is 1.6 ms^{-1}. RMSE then increases to over 3 ms^{-1} in the first simulation hour, then settles to about 2.5 ms^{-1} the next hour. The RMSE continues to decrease for the next 5 h to around 1.5 ms^{-1} where it remains for the remainder of the model simulation time. There is little difference between the individual schemes during the first 5 h. After that, the spread between individual BL schemes increases. During the nowcast period, which corresponds to daytime, the QNSE scheme shows the greatest RMSE, followed by BouLac and MYJ. The remaining schemes (YSU, SH, YSU, and MYNN) have the lowest RMSE. The latter half of the simulation time, during the formation and maintenance of the nighttime BL, shows the greatest spread. MYJ, SH, BouLac, and ACM2, show the least RMSE, ranging from 1.2 to 1.5 ms^{-1}. The greatest variation between schemes occurs during the nighttime. This is consistent with the analyses of temperature and RH errors presented earlier. Differences between the various BL/SL formulations appear to be greater at nighttime rather than daytime.

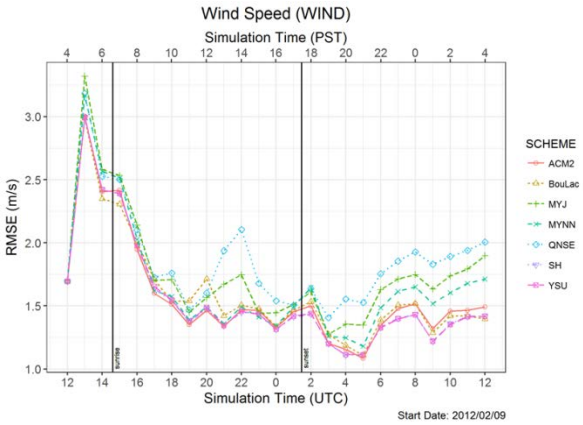

Fig. 7: Wind speed RMSE for the 7 members.

The RMSE and bias calculations for the entire forecast period (nowcast and extended forecast) period are summarized in Table 4. When we eliminate the redundancy in the wind fields (u, v components and wind speed), we found that the QNSE and BOU schemes performed marginally better in the number of measures of forecast quality (TEMP, DPT, WSP). Additionally, the SH scheme performed very similarly to the YSU scheme that it was based on, indicating that the scale independent nature of the scheme appears not to be important at the grid spacing we used (1 km). However, overall, the individual differences between schemes were small.

To choose the best overall scheme, we assigned rankings (from 1 through 7) for each of the 3 independent variables (T, DPT, and wind) and averaged the RMSE rankings for the nowcast, extended forecast, and complete forecast periods, for temperature, dew point, and wind speed. Since the bias calculation potentially includes large swings in the error that may be offset by one in the other direction, we evaluated the performance of the schemes using only the RMSE.

The results are shown in Table 5. We included rankings for T, DPT, and wind speed. For the nowcast period (daytime), the BouLac scheme performs best. For the extended forecast (nighttime) YSU, BouLac, and QNSE tied for best performance. Overall, for the entire forecast period, BouLac performs best for our data. The ensemble mean RMSE for temperature varies between 2.4 and 3.0 K, DPT RMSE ranges from 2.8 to 3.4 K and for wind velocity is nearly constant at approximately 1.5 ms^{-1}. The standard deviation of the RMSE is quite small, indicating the small differences between the various PBL schemes we tested.

4 TALAGRAND DIAGRAMS

To provide an estimate of the utility of the 7 member PBL/SL scheme combinations for producing a valid ensemble that could provide useful model statistics and quantification, Talagrand diagrams (Hamill 2001; Wilks 2011), also known as rank histograms, were produced for this study. The shape and distributions shown by these diagrams represent a measure of the validity of an ensemble of forecasts and tell us about the relationship of the forecasts to the observed data.

The 2m temperature histogram is shown in Fig. 8. Very few data are represented in bins 1–7, and a majority of the measurements occur in bin 8. This type of plot is produced when the observed value is higher than nearly all of the ensemble forecasts. Since very little data appear in the other bins, the WRF model consistently underestimates the late afternoon surface temperatures for the 7 PBL/SL schemes considered in this analysis. The observed temperature is consistently higher than the ensemble members and independent of which scheme we choose for the PBL/SL physics.

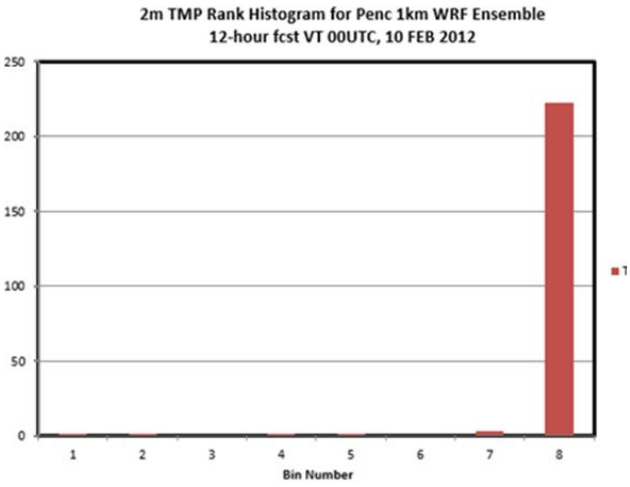

Fig. 8 WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, surface (2m) temperature.

The Talagrand diagram for DPT (Fig. 9) and WSP (Fig. 10) are notably different. Here we obtained a classic U-shaped plot, which is produced when there are a sufficient number of low and high biases but the individual members do not spread out enough to provide an adequate statistical measure of spread within the ensemble. Although there is some spread in the ensemble members, the spread is small, resulting in relatively few samples in the intermediate bins and a majority lying at the extreme bins. While this represents an improvement in the quality of the spread produced by the ensemble, it remains inadequate, as the idealized plot would be relatively flat through all 8 bins (7 members plus the observations).

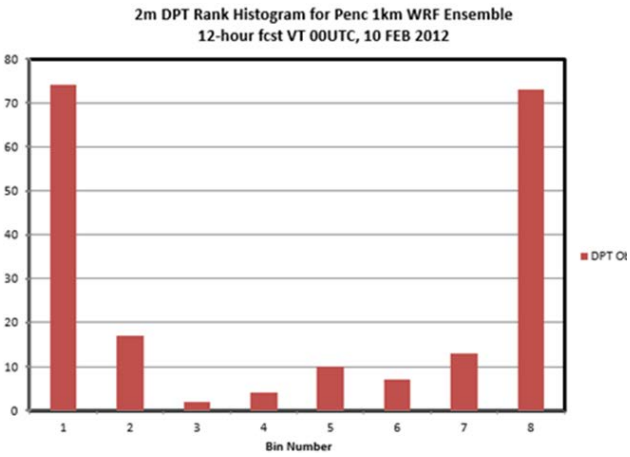

Fig. 9 WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, surface (2m) DPT.

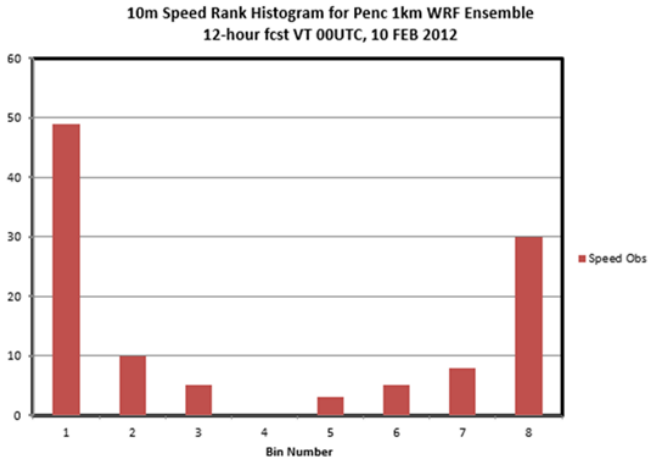

Fig. 10 WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, 10m wind speed.

The characteristic U-shaped patterns are consistent with an ensemble whose members do not have sufficient spread, contain a mixture of low and high biases, and it indicates that the use of PBL/SL schemes alone does not constitute a "good ensemble". A "good ensemble" has sufficient spread among its members such that the distribution is within the envelope of expected values for a given probability distribution defined by the validation observation(s). While we have shown only the results from the 00Z analysis, analyses from other times in this run set are similar. Therefore, we conclude that using an ensemble based solely on the choice of PBL/SL scheme combination, is, in itself, insufficient in creating a valid ensemble for the purpose of nowcasting with this version of WRF, and the current dataset (i.e., a single case day), geography, and synoptic conditions.

5 PBL DEPTH

Another useful meteorological quantity aside from surface values of temperature, moisture, and wind is the PBL depth, which is useful for some meteorological sensors since this measure gives an indication where enhanced turbulence due to convective mixing is expected, and the mixing depth for dilution of materials released into the atmosphere. Optical systems are also highly affected by atmospheric turbulence, including mixing and the formation of BL clouds. Turbulence also affects radio propagation and the

performance and safety of unmanned aerial vehicles, so knowledge of the PBL depth defines the layer where degraded performance may be expected. We unfortunately had very few data on which to validate our model PBL depth forecasts. Only one routine upper air observation station, at Miramar Marine Corps Air Station (KNKX), is located within the inner (D3) domain (Figure 11).

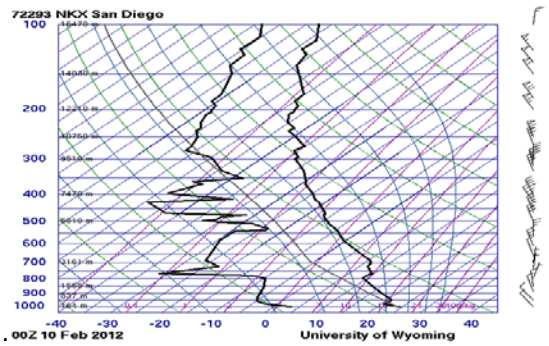

Fig.11 Atmospheric sounding at KNKX, San Diego (top) 0000 UTC 10 February 2012 (Courtesy of the University of Wyoming: http://weather.uwyo.edu/upperair/sounding.html.)

The diurnal variation in modeled PBL depth at a point near the center of the inner grid (32.9° latitude, 117.1° longitude) is shown in Fig. 12. Qualitatively, the evolution of the PBL depth is consistent with what is expected in the evolution of the PBL over land. However, it is not possible to follow the growth and decay with the available upper air observational data, largely because we have only one available site to use in the analysis and only point measurements (soundings) at regularly scheduled times (i.e., 0000 UTC and 1200 UTC at KNKX). However, the general behavior is captured by the model.

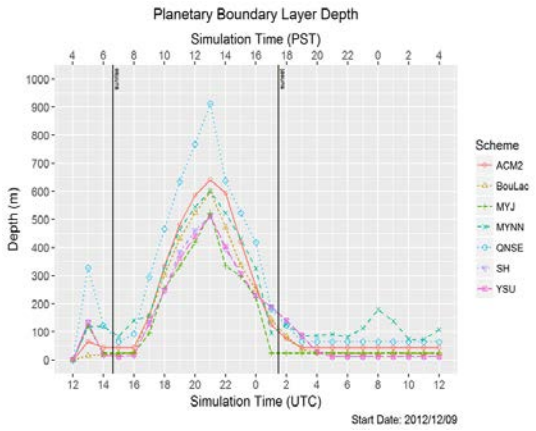

Fig. 12 Temporal evolution of PBL depth for the 7 PBL schemes tested over the center point of innermost (D3, 1-km) domain

Note the QNSE scheme produced a PBL depth significantly deeper than the others. Because the sounding was taken near the interface between the marine and continental air, the analysis is complicated. When comparing PBL depths forecast by different BL schemes it is important to be aware that the methods used to diagnose the PBL depth vary among the schemes. Thus, it is possible for 2 schemes to forecast the same temperature and moisture profiles and yet diagnose a different PBL depth. Differences in PBL depth among experiments are then a combination of differences in the forecast structure of the atmosphere and differences in the methods used to diagnose the PBL depth. Two general methods are used to estimate the depth of the PBL: one based on the Richardson number and the other based on the vertical profile of temperature. The thresholds for detection vary among the individual schemes. Reen et al. (2014) explored this issue in more detail. A more detailed analysis will follow.

We performed a limited qualitative analysis of the variation in the PBL depth by the model over the innermost (D3) domain (Figs. 13–16). Each figure shows the D3 PBL depth at 0000 UTC (1600 PST) 10 February 2012 for the 7 schemes; we also include contours of the terrain height (in meters).

Figure 13 shows the horizontal variation of the PBL over the innermost model domain at 0000 UTC 10 February for the MYNN PBL scheme. As expected, the lowest PBL depths are found over the cool waters since the daytime surface temperature increase is greater over the land than over the water. Where surface temperature is the highest, we find the deepest PBLs. Areas removed from the coast show greater PBL depths. The enhanced surface mixing caused by disturbed air flow over the higher terrain also results in greater PBL depths. The model appears to be reasonably reproducing the basic physics that controls the depth of the PBL. That is, the surface heating and mixing combine to contribute to greater PBL depths with the expected geographical distribution. The maximum depths presented by the model data are in the vicinity of 1.0 to 1.2 km, which is reasonable for the surface heating expected during the wintertime over Southern California. Near the coastlines, and over water, the PBL depths are from a few tens of meters to a couple hundred meters. These values are again consistent with the formation of a stable marine BL, consistent with the cold waters of the eastern Pacific (Angevine 2006). The small-scale structure evident over the eastern half of the domain reflects

the effects of the higher terrain (and importantly, gradients) in generating mixing from the surface to drive the PBL and the enhanced entrainment of air from above the capping inversion in these situations. While 5 of the 7 schemes we looked at (Fig 13, 14) produced reasonably similar PBL depths over the domain, the ACM2 and QNSE schemes (Fig 15, 16) produced notably deeper PBLs inland in the vicinity of the higher terrain.

The deepest PBL depths are associated with the strongest heating (Figure 17) and secondarily, with the terrain height. Furthermore, the PBL was deeper further away from the marine influences, as would be expected.

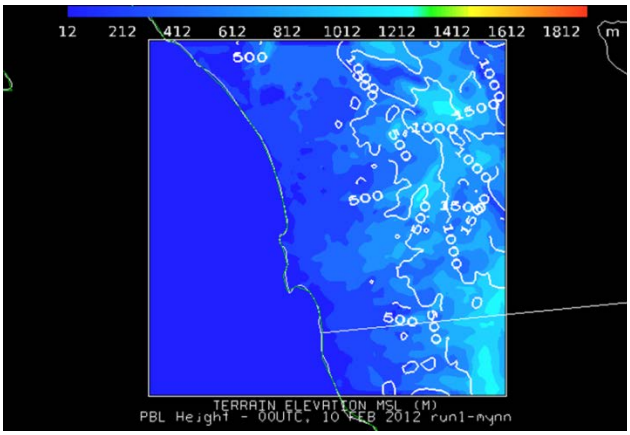

Fig. 13 PBL depth (meters) valid at 0000 UTC 10 February 2012 for the MYNN PBL option

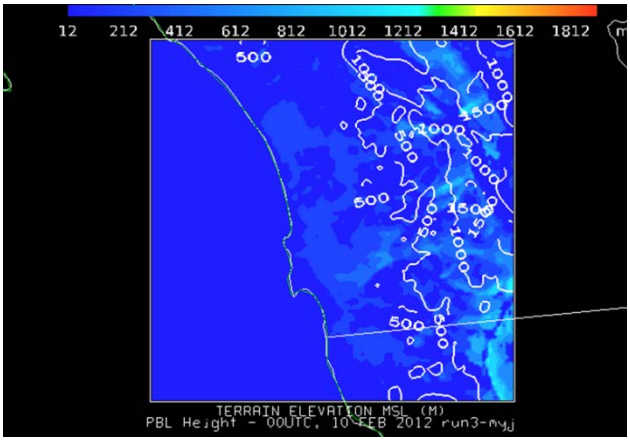

Fig. 14 PBL depth (meters) valid at 0000 UTC 10 February 2012 for the MYJ PBL option

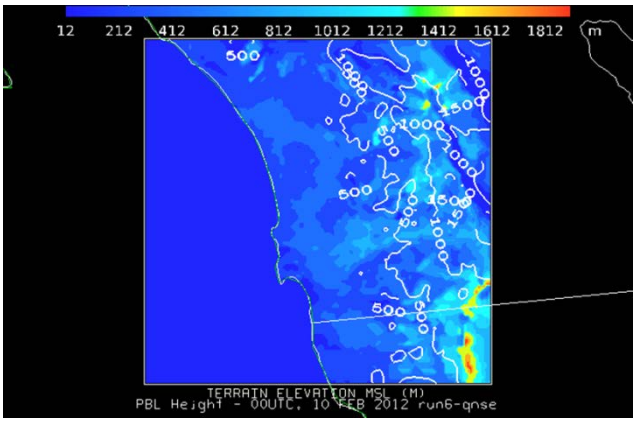

Fig. 15 PBL depth (meters) valid at 0000 UTC 10 February 2012 for the QNSE PBL option

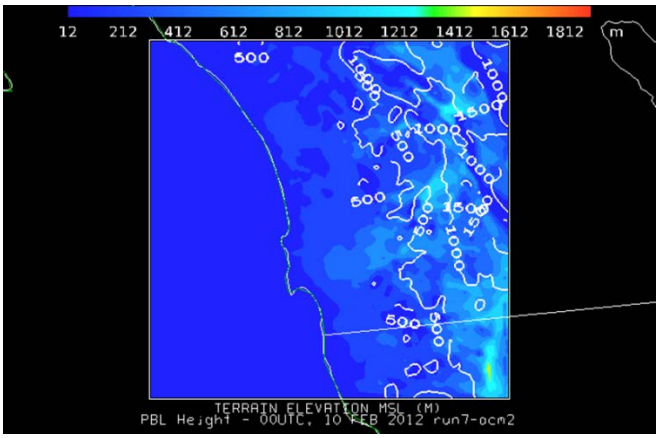

Fig. 16 PBL depth (meters) valid at 0000 UTC 10 February 2012 for the ACM2 PBL option

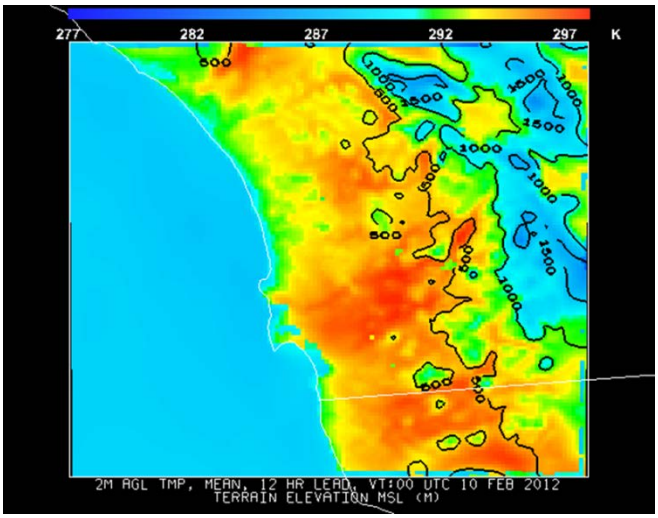

Fig. 17 Ensemble mean surface (2m AGL) temperature (Kelvin) over the innermost model domain (D3, 1-km grid) at 0000 UTC 10 February 2012

The PBL depth is the focus of continued research because of its importance and potential application.

6 SUMMARY AND CONCLUSIONS

We tested 7 PBL/SL parameterization schemes using a 9-/3-/1-km triple nest grid configuration centered over San Diego, California, under quiescent conditions in late winter. There was little spread in the RMSE/bias statistics for the schemes we tested. The WRF model had its greatest difficulty in capturing the transition between daytime and nighttime boundary layers for our test day. The greatest variation between schemes is with the nocturnal BL. For the nowcast period, the best overall performer was BouLac, followed by MYJ. For the extended forecast period, the YSU, BouLac, and QNSE schemes were tied for best performance. For the overall forecast period (nowcast, extended) BouLac was best, followed by QNSE. The standard deviation of the ensemble created using the various PBL schemes was small despite the fact that the schemes we tested were devised using local, nonlocal, and hybrid approaches.

While the SH scheme was developed primarily to address the scale awareness problem (Shin and Hong, 2013), we did not see an improvement in using this scheme with a 1-km grid spacing, in particular with the YSU scheme, which shares common lineage.

Our study examined only a single day, under quiescent conditions. In addition to the 7 simulations we performed here, we produced data for 4 additional days with more active synoptic situations, and this is the subject of ongoing analysis. In addition to the BL/SL schemes, we are looking at the other physics categories as well as the initialization data in an attempt to quantify model spread and uncertainty.

Independent statistical analysis by Smith et al. (2018) is in agreement with our determination of the lack of variance between the schemes, and attributes only 3%–4% of total model variance to the PBL schemes. That calculation involves, at the highest level, an extension of a common data analytics approach called ANOVA (for analysis of variance), which uses a parameter eta (η). Bakeman (2005) found that a generalized eta squared, η_G, was superior. Eta squared (η_G^2) was calculated using the lsr package in R (Navarro 2015), allowing us to estimate the relative contribution to model error (uncertainty) due to the BL physics packages.

Talagrand (Hamill, 2000) diagrams were also constructed and show very little dispersion between the 7 members. Because we did not obtain an optimal flat response in our Talagrand plots, but rather a classic “U shaped” distribution characteristic of inadequate spread between the ensemble members for all but temperature, the use of only a physics-based ensemble (in this case BL/SL combinations) is inadequate for deriving probabilistic information and forecast uncertainty. Research looking into the relative contributions to model uncertainty due to the other physics schemes and initialization data is addressed in the DoE approach described by Smith et al. (2018).

ACKNOWLEDGMENTS

We gratefully recognize the assistance of Dr. Brian Reen, as well as the assistance provided by Ms. Leelinda Dawson who helped automate various parts of the data reduction/post processing efforts. The principle author would also like to thank Dave Knapp at ARL for the rewarding opportunity to work with ARL in the Atmospheric Modeling Branch, and to Dr. Andre Pattantyus for the fine job of being put on the spot to present this paper in such short notice in my absence.

Table 4. Mean bias and RMSE (K for temperatures, ms^{-1} for winds) for 1-km ensemble members (D3) (1 km) for 1200 UTC 9 February through 1200 UTC 10 February (complete forecast, 6–24 h). Best performers in each statistic are colored in red.

Bias/RMSE	MYNN	SH	MYJ	YSU	BOU	QNSE	ACM2
T	–1.64/3.04	–0.50/2.78	0.12/2.77	–0.50/2.77	–0.31/2.68	–0.84/2.66	–0.42/2.71
DPT	–1.37/2.99	–1.30/3.05	–1.23/2.91	–1.24/3.03	–1.06/2.90	–1.21/2.89	–1.56/3.05
RH	–0.52/11.2	–3.23/11.6	–5.09/12.5	–3.03/11.6	–3.03/11.2	–2.65/11.8	–3.97/11.4
u comp	0.10/1.54	0.30/1.42	0.03/1.68	0.31/1.42	0.22/1.46	–0.11/1.87	0.22/1.46
v comp	0.18/1.58	0.20/1.53	0.10/1.62	0.20/1.52	0.20/1.54	0.10/1.69	0.25/1.55
Wind	0.31/1.51	0.07/1.35	0.61/1.59	0.06/1.35	0.05/1.42	0.92/1.73	0.15/1.38

Table 5: Ranking for RMSE statistics by scheme (1=best, 7=worst)

Hours	Field	MEAN	STDEV	MYNN	SH	MYJ	YSU	BouLac	QNSE	ACM2
06 through 12 (nowcast)	Temp	2.43	0.18	7	6	1	5	2	3	4
	DPT	3.38	0.09	5	5	3	4	1	1	7
	WIND	1.49	0.12	4	1	6	1	5	7	1
	COMP			7	5	3	2	1	4	5
12 through 24 (extended)	Temp	2.95	0.15	7	3	6	2	3	1	3
	DPT	2.77	0.07	4	7	2	6	2	1	5
	WIND	1.47	0.16	5	1	6	1	4	7	3
	COMP			7	5	6	1	1	1	4
06 through 24 (complete)	Temp	2.77	0.13	7	6	4	4	2	1	3
	DPT	2.97	0.07	4	6	3	5	2	1	6
	WIND	1.47	0.14	5	1	6	1	4	7	3
	COMP			7	5	6	3	1	2	4

REFERENCES

Angevine WM, Hare JE, Fairall CW, Wolfe DE, Hill RJ, Brewer WA, White AB. Structure and formation of the highly stable marine boundary layer over the Gulf of Maine. J Geophys Res: Atmospheres. 2006;111(23):2156–2202.

Bakeman R. Recommended effect size statistics for repeated measures designs. Behavior Research Methods. 2005;37(3):379–384.

Deng A, Stauffer D, Gaudet B, Dudhia J, Bruyere C, Wu W, Vandenberghe F, Liu Y, Bourgeois A. Update on WRF-ARW end-to-end multi-scale FDDA system. Proceedings of 10th National Center for Atmospheric Research (NCAR) WRF Users' Workshop. 2009 June 23–26; Boulder, CO. p. 14. .

Dumais R, Kirby S, Flanigan R. Implementation of the WRF 4-dimensional data assimilation method of observation nudging for use as an ARL weather running estimate-nowcast. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2013 June. Report No.: ARL-TR-6485. p. 24.

Dumais R, Passner J, Flanigan R, Sauter B, Kirby S. High resolution WRF-ARW studies at the US Army Research Laboratory for use in short-range forecast operations. Short range forecast applications: P2.4. Proceedings of the 23rd Conference on Weather Analysis and Forecasting/19th Conference on NWP; 2009 June 1–5; Omaha, NE.

Dyer J, Zarzar C. US Army Research Laboratory (ARL)/Mississippi State University (MSU) project: atmospheric modeling and decision aids: field-based numerical weather simulations and analysis support tools period of performance: WRF on a laptop-user's guides. White Sands (NM): Battlefield Environment Division, Army Research Laboratory (US); 2015 Sep 29. p. 31.

Dyer J, Zarzar C, Dumais R, Raby J, Smith JA. Defining the influence of horizontal grid spacing on ensemble uncertainty within a regional modeling framework. Wea Forecast. 2016;31:1997–2017.

Foley T, Smith J, Raby J, Reen B, Penc R. Developing sub-domain verification methods using GIS tools. Presented at the Esri User Conference; 2015 July 21; San Diego, CA.

Hamill TM. Interpretation of rank histograms for verifying ensemble forecasts. Mon Wea Rev. 2000;129:550–560.

Huang X, Xiao Q, Barker D, Zhang X, Michalakes J, Huang W, Henderson T, Bray J, Chen Y, Ma Z, et al. Four-dimensional variational data assimilation for WRF: formulation and preliminary results. Mon Wea Rev. 2009;137:299–314.

Liu Y, Bourgeois A, Warner T, Swerdlin S, Hacker J. Implementation of observation nudging based

FDDA into WRF for supporting ATEC test operations. 6th National Center for Atmospheric Research (NCAR) WRF/15th MM5 Users' Workshop; 2005; Boulder, CO.

Lu C, Yuan H, Schwartz B, Benjamin S. Short-range numerical weather prediction using time-lagged ensembles. Wea Forecast. 2007;22:580–595.

Navarro DJ. Learning statistics with R: a tutorial for psychology students and other beginners. Version 0.5.: University of Adelaide; 2015.

Penc, R. S., J. A. Smith, J. W. Raby, R. E. Dumais, Jr., B. P. Reen, and L. P. Dawson, 2018b: Intercomparison of 7 Planetary Boundary-Layer/Surface-Layer Physics Schemes over Complex Terrain for Battlefield Situational Awareness. Technical Report ARL-TR-8353.

Reen BP, Dumais RE Jr, Passner JE. Mitigating excessive drying from the use of observations in mesoscale modeling. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2014a Jan. Report No.: ARL-TR-6775. p. 33.

Reen BP, Schmehl KJ, Young GS, Lee JA, Haupt SE, Stauffer DR. Uncertainty in contaminant concentration fields resulting from planetary boundary layer depth uncertainty. J Appl Meteor Climatol. 2014b;53:2610–2626.

Skamarock W, Klemp J, Dudhia J, Gill D, Barker D, Duda M, Huang X, Wang W, Powers J. A description of the advanced research WRF Version 3. Boulder (CO): National Center for Atmospheric Research (NCAR); 2008 June. Report No.: TN–475+STR.

Smith JA, Penc RS, Raby JW. Statistical design of experiments in numerical weather prediction: emerging results. Presented at the 98th Annual American Meteorological Society Meeting/25th Conference on Probability and Statistics; 2018 Jan 7–11; Austin, TX. Paper No.: 6.1.

Stensrud DJ. Parameterization schemes: keys to understanding numerical weather prediction models. New York (NY): Cambridge University Press; 2007.

Shin H, Hong S. Analysis of resolved and parameterized vertical transports in convective boundary layers at gray-zone resolutions. J Atmos Sci. 2013;70:3248–3261.

Wilks DS. On the reliability of the rank histogram. Mon Wea Rev. 2011;139:311–316.