Digital Forecast Process: Comments on the Day4-7 Forecast

IFPS Science Steering Team
March 2007

I. Introduction

There are many facets of the Digital Forecast Process (DFP) that need to undergo a transformation to support the current forecast suite and to provide a solid foundation for the next generation of products and services. The IFPS Science Steering Team (ISST) has been seeking out forecast process inadequacies in an effort to identify opportunities that will support a unified philosophy and common approach to producing the forecast for days 4 through 7 (hereafter Day4-7). Much of the discussion revolves around proper application of numerical guidance, including aspects related to frequency and latency. Furthermore, the ISST sees this as an opportunity to shift the workload burden from the Day4-7 forecast to time ranges characterized by impact weather forecasting (namely Days 1-3). Ultimately, this includes exploring aspects of CONOPS structures that take advantage of temporal splitting and cluster support functions in the gridded forecast suite during the Day4-7 timeframe.

The Day4-7 forecast suite presents several challenges that include numerical guidance support, grid population methods, coordinated collaboration, verification, and desired performance relative to effort expenditure. The ISST recognizes that Day4-7 operations at the WFO are governed by office staffing structure, local management decisions, and regional policy. However, there are common elemental practices that should govern the Day4-7 forecast process irrespective of heterogeneous application.

II. Evaluation

Examination of multi-season (for this study a 16 month average from September 2005 to December 2006 is used) verification statistics from the NDFD point verification system (see bestpractices.nws.noaa.gov/contents/ndfd-stats/verification for more information) reveal some fundamental characteristics of the forecast process. First, the value added by the latest cycle of guidance and forecaster improvement is perishable (Figure 1 – illustrates the steep slope in the MaxT forecast MAE curve as a function of time range). The reason for this steepness is illustrated by the large increase in skill due to greater predictability between successive model cycles, as measured by anomaly correlation decay measurements (Figure 2), which exhibits the greatest cycle-to-cycle increases in skill during the Day4-7 timeframe. Second, the act of not acting upon the latest suite of guidance results in a substantially poorer forecast. This is clearly depicted by the “Night” forecast from HPC in Figure 1. Namely, HPC does not produce a forecast overnight; and these results indicate how the forecast skill degrades through omission. Third, it is difficult to identify opportunities for improvement from national scale statistics, given the variability of forecast process applications from region to region.

The MaxT verification data (Figure 1 and 3a) suggests that forecasts during the first two to three days are re-evaluated every primary cycle and that forecasters are improving upon the latest cycle of supplied guidance (e.g., Day Shift NDFD (NDFD Day) to 12z GFS MOS (MOS Day)) in the first 48-60 hours (Figure 3a). This improvement rapidly decays around hour 72 – likely a result of the migration from the Day4-7 portion of the previous forecast due to the absence of MaxT guidance at the 72 hour lead time from the

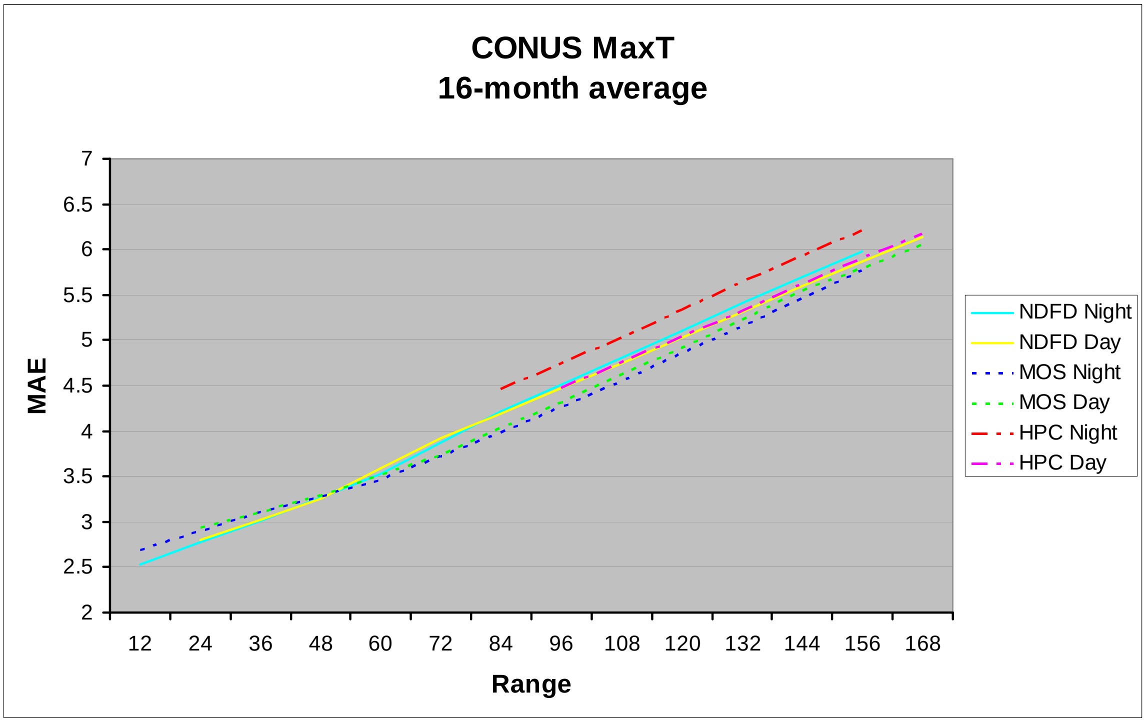


Figure 1: Sixteen month average (Sept 2005-Dec 2006) Maximum Temperature Mean Absolute Error (MAE) at all CONUS GFS MOS points as function of forecast range for the 00z (Day) and 12z (Night) NDFD verification snapshots of available data from NDFD, GFS MOS, and HPC at the collection times.

traditional guidance bulletins (FWC, MAV, MET). The performance in the Day4-7 timeframe improves slightly for the Day Shift forecast; but remains nearly steady for the Night Shift.

There are differing practices in addressing the Day4-7 forecast with frequency ranges from once per day (a practice sanctioned by several regions), twice daily tied to the primary cycles of the GFS and attendant post-processed guidance, and even four times daily with each cycle of the GFS and Global Ensemble Forecast System (GEFS). The most common practice is once daily, with the option of making minor adjustments on the "off" cycle (e.g., Eastern Region – hereafter ER – targets the Day Shift while Southern Region – hereafter SR – the Night Shift). Many times this portion of the forecast is addressed during a quiet portion of the day, which usually translates into assessing guidance from a previous cycle (e.g., evaluating the 00z guidance during the 14-17z timeframe – while the 12z cycle is already arriving). Consequently, for completeness, a comparison is performed against the previous cycle of guidance (Figure 3b – e.g., Day Shift NDFD (NDFD Day) to 00z GFS MOS (MOS Night)). Forecasters are certainly making improvements at all times over the previous cycle; but it is unclear as to what is contributing to the signal, without investigating the forecast practices employed to produce the values. There is an opportunity within this dataset to peer into internal forecast practices by evaluating the regional statistics. During this evaluation period, there were significant variations in regional policies related to grid production cycles. The most uncomplicated comparison of grid practices is between ER and SR. Where ER addresses the Day4-7 forecast during the 14-17z time window once daily using the 15z HPC guidance (based upon the 00z cycle NWP suite) as a common starting point; and SR addresses the Day4-7 forecast during the 06-09z time window once daily using the 00z

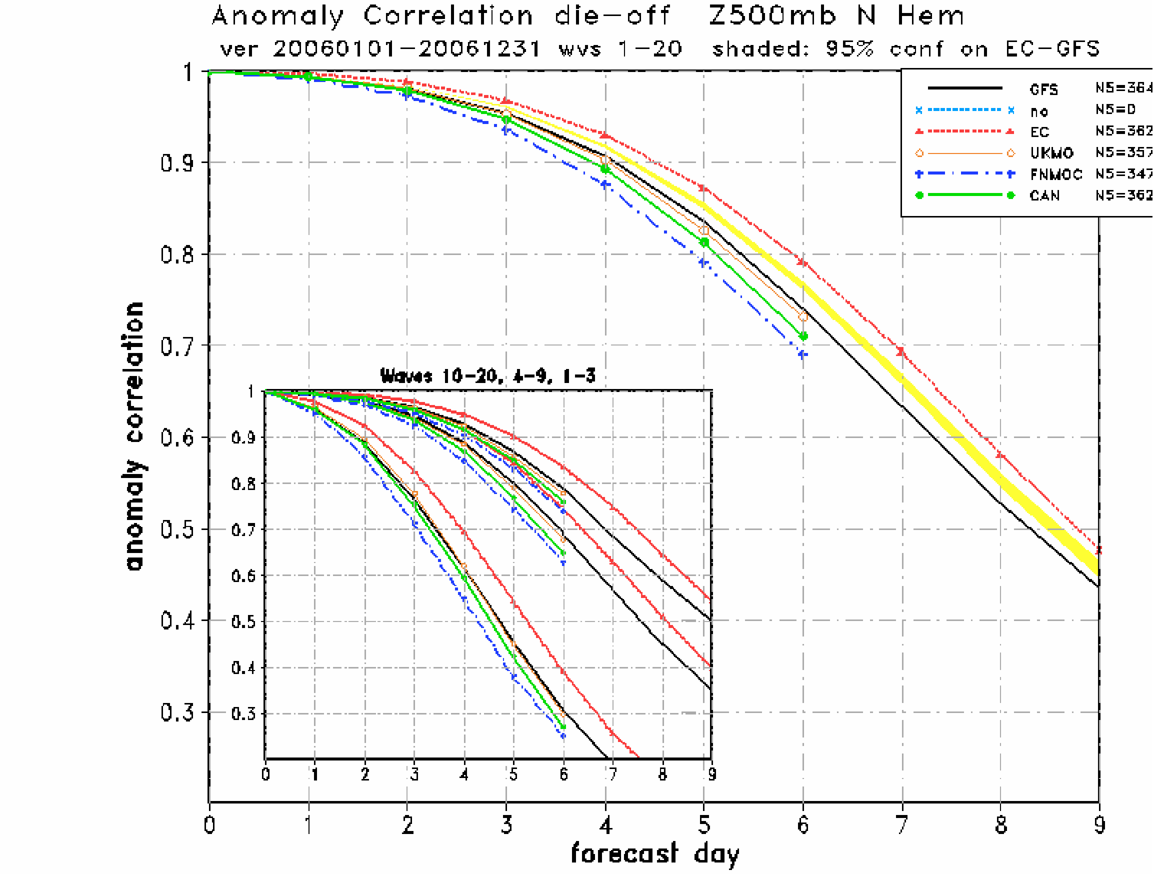
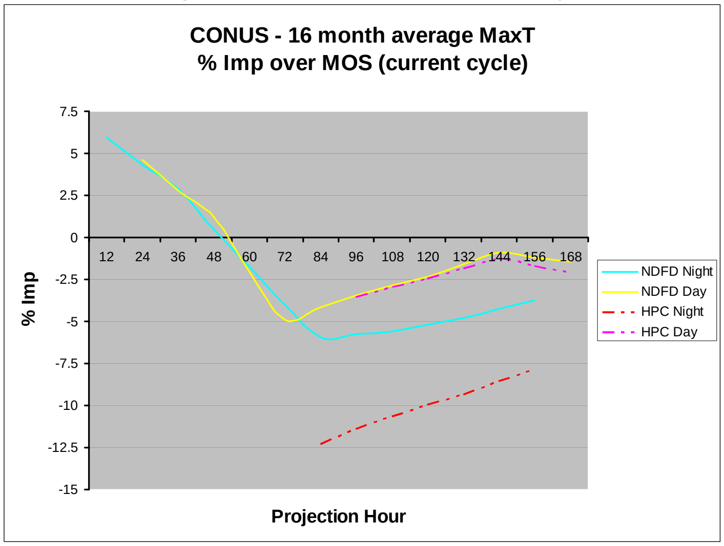


Figure 2: Annual (2006) 500mb Height Anomaly Correlation decay curves for the Northern Hemisphere from commonly use Global Models – GFS, ECMWF, UKMET, NOGAPS, CMC Global. The yellow line is the 95% confidence interval of the difference between the ECMWF and GFS. ****Note that the ECMWF has a half to three-quarter day increase in skill over the GFS****

GFS MOS as a common starting point. This comparison lends a clean evaluation of when to apply the available guidance to the forecast.

Regional comparisons of forecast errors for MaxT (Figure 4) show that the NDFD performance parallels that of the input guidance from GFS MOS and HPC in general; however region differences in relative performance are evident. This is more clearly depicted by comparisons to the available guidance (Figure 5). SR is consistently performing at or above the level of the available GFS based guidance, while ER statistics indicate performance lagging that of GFS MOS beyond 60 hours. Recognizing the HPC and ER practice of using latent guidance, an evaluation against the previous cycle is performed (Figure 6). During the Day4-7 period, ER and HPC are improving upon the guidance that they are using – demonstrating value added to the forecast guidance. Nevertheless, this improvement is less than that offered by the introduction of new guidance (i.e., the percent improvement of the next model cycle exceeds that of the NDFD – Figure 6). This further exemplifies the perishable nature of the forecast during time ranges when the predictability increases substantially from one cycle to the next. Consequently, in order for the value of forecaster input to be realized it must be in phase with the "current" cycle of model guidance, regardless of prescribed practice (e.g., evaluation once or twice per day).



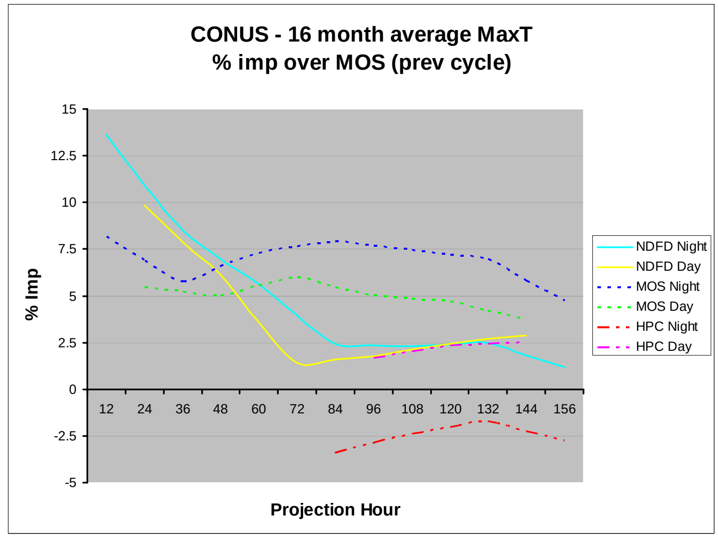


Figure 3: Sept 2005 – Dec 2006 (a) CONUS MAE % Improvement over the current MOS cycle (e.g., day shift NDFD (NDFD Day) compared with 12z GFS MOS (MOS Day)). (b) CONUS MAE % Improvement over the previous MOS cycle (e.g., day shift NDFD (NDFD Day) compared with 00z GFS MOS (MOS Night)). The percent improvement between successive model cycles is depicted by the addition of MOS to the comparison.

Furthermore, during the evaluation period, forecasts available from ER offices at the 00z snapshot (Day Shift forecast) were based on the previous 00z cycle (1 cycle latent) and those available at the 12z snapshot (Night Shift forecast) still on the previous 00z cycle (2 cycles latent). Additionally, the 12z GFS MOS performance was inferior to that of 00z GFS MOS during much of the evaluation period. The reduced skill of the 12z guidance has been addressed by MDL (new equations implemented June 20, 2006). While the 12z is still a slightly poorer performer, it now has comparable error characteristics to that of the 00z cycle. Therefore, error calculations relative to 12z GFS MOS (MOS Day) should now parallel that of 00z GFS MOS relative statistics. Finally, anomaly correlation statistics for 2006 (not shown) show a slight advantage to the 12z GFS over the other 3 cycles – with the performance order being 12z, 00z, 18z, 06z. Granted, the GFS family continues to lag the skill of the ECMWF by approximately ¾ day. Regardless of the relative statistics, it should be noted that the absolute performance is still quite poor (MAE ~6-7°F at day 7) indicating poor predictability resulting in the reduction of skill at those forecast lead times.

Minimum temperature forecasts (Figures A1-A3) generally do not demonstrate improvement over the available guidance at all time ranges – with SR showing the "best" performance. Moreover, there are many instances where NDFD minimum temperatures under-perform in comparison to the previous cycle guidance indicating that there is a general lack of recognition of how and when to deviate from guidance. The same forecast process tendencies isolated from the MaxT verification are also evident in the MinT dataset. Once again, the greatest influence to improved performance is utilizing guidance from the current model cycle.

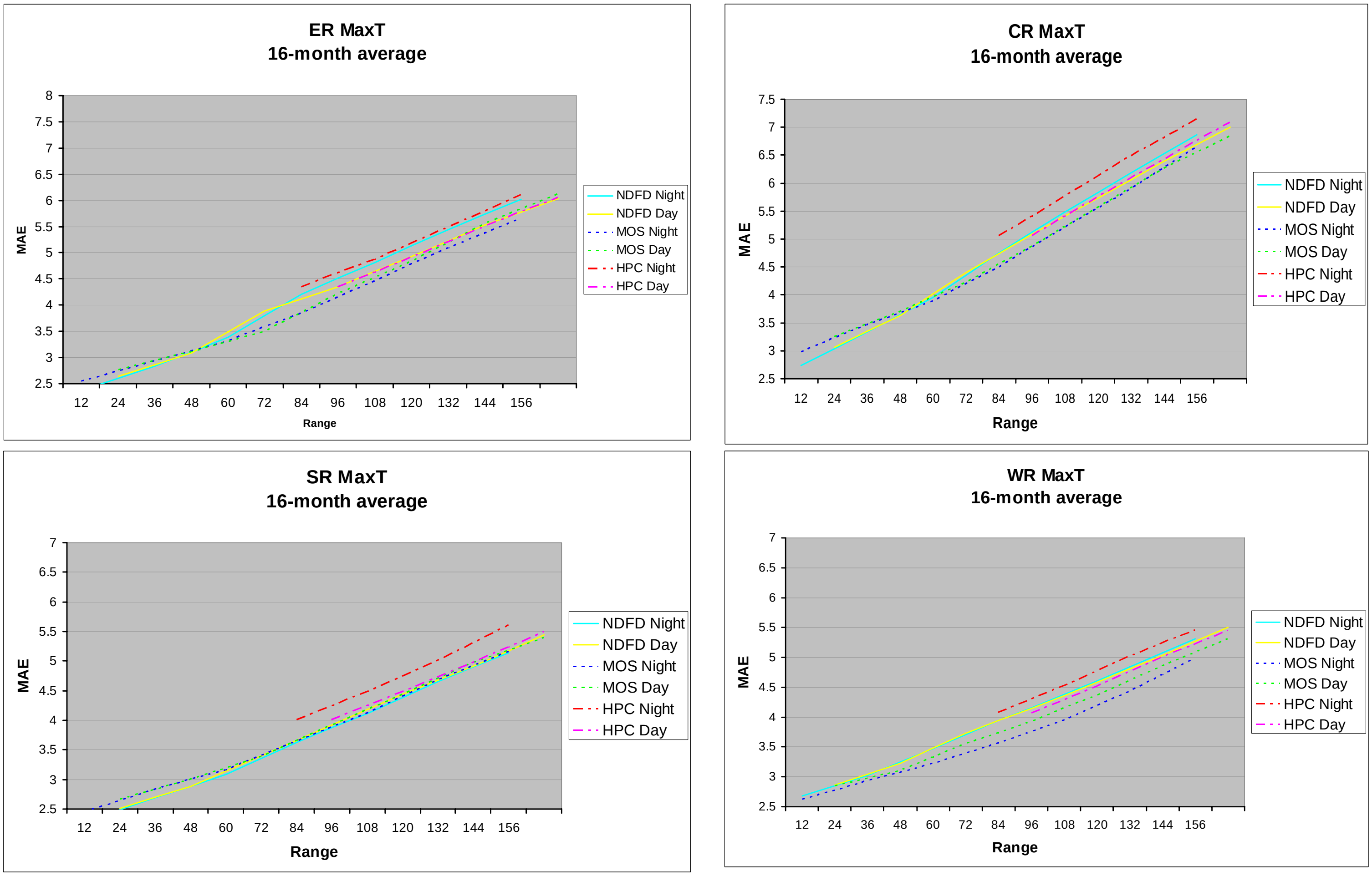


Figure 4: Sixteen month average for the four CONUS regions (Sept 2005-Dec 2006) MaxT MAE at all CONUS GFS MOS points as function of forecast range for NDFD, GFS MOS, and HPC.

For completeness, PoP statistics are included (Figures A4-A7). The NDFD PoP forecasts perform substantially worse than guidance in the day 1-3 timeframe (Figures A5 and A6), while offering some skill over guidance in the extended forecast time frame, granted at a reduced level of skill. Some of this reduction in skill can be attributed to the policy driven practices of over using 14% (as seen most notably in the PoP resolution for ER and CR) and a high frequency of 0% forecasts lending to more precipitating events occurring with a 0% chance which is very penal in the Brier Score calculation.

Overall, the use bulk statistics is a very limited means of identifying forecast characteristics. For instance, the propensity of a particular source to produce significant errors is not discernible due to relative infrequency of such events. However, those forecasts have a significant impact on service. Additionally, the spatial quality of the forecasts is not evaluated using these bulk error statistics. As an example, GFS MOS is point guidance and must be assimilated into a usable gridded format. This is typically done by means of adjusting the GFS downscaled output (smartInit) through an objective analysis process (serp via the MatchGuidance smart tool suite).

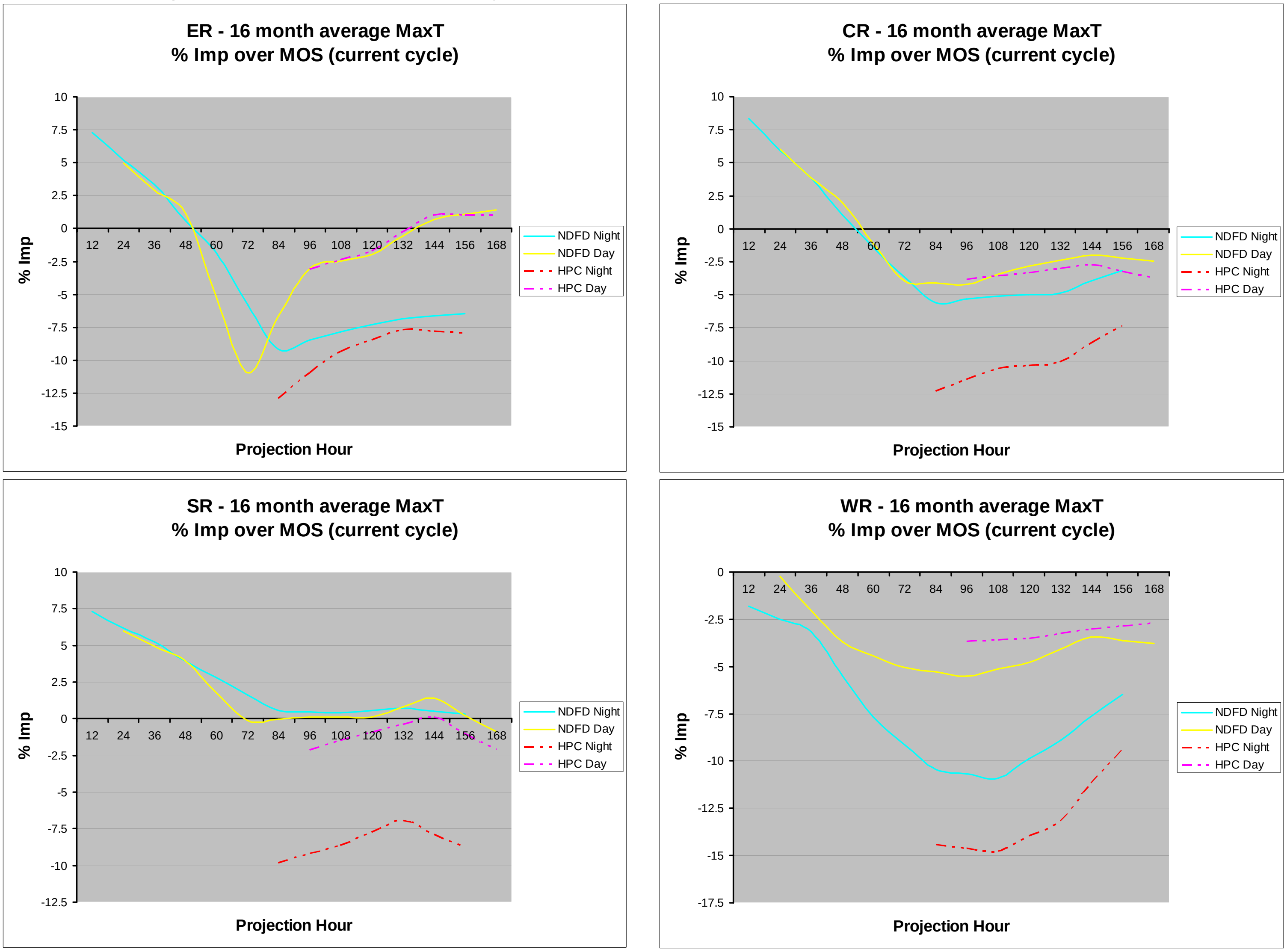


Figure 5: Sept 2005 – Dec 2006 Regional MaxT MAE % Improvement over the current MOS cycle (e.g., day shift NDFD (NDFD Day) compared with 12z GFS MOS (MOS Day)).

Finally, the HPC dataset also possesses appreciable weaknesses for direct application in the production of gridded forecasts. There are many assumptions and systematic techniques that are used irrespective of atmospheric condition. The techniques are convenient for producing national scale guidance from a limited set of controls; but in many instances lack the required geoclimatic detail for direct application in the WFO gridded forecast production. The net result is a substantial amount of grid modification by the forecaster to include the missing geoclimatic data – an important source of information for the local customer – an unintended result of using this guidance source.

III. Summary and Recommendations

The ISST recognizes that employing a standard methodology for the Day4-7 forecast will be difficult given time zone differences and heterogeneous staffing levels and shift distributions. However, there are common philosophies that can be incorporated across regional boundaries; most notably the concept of evaluating and acting upon the latest

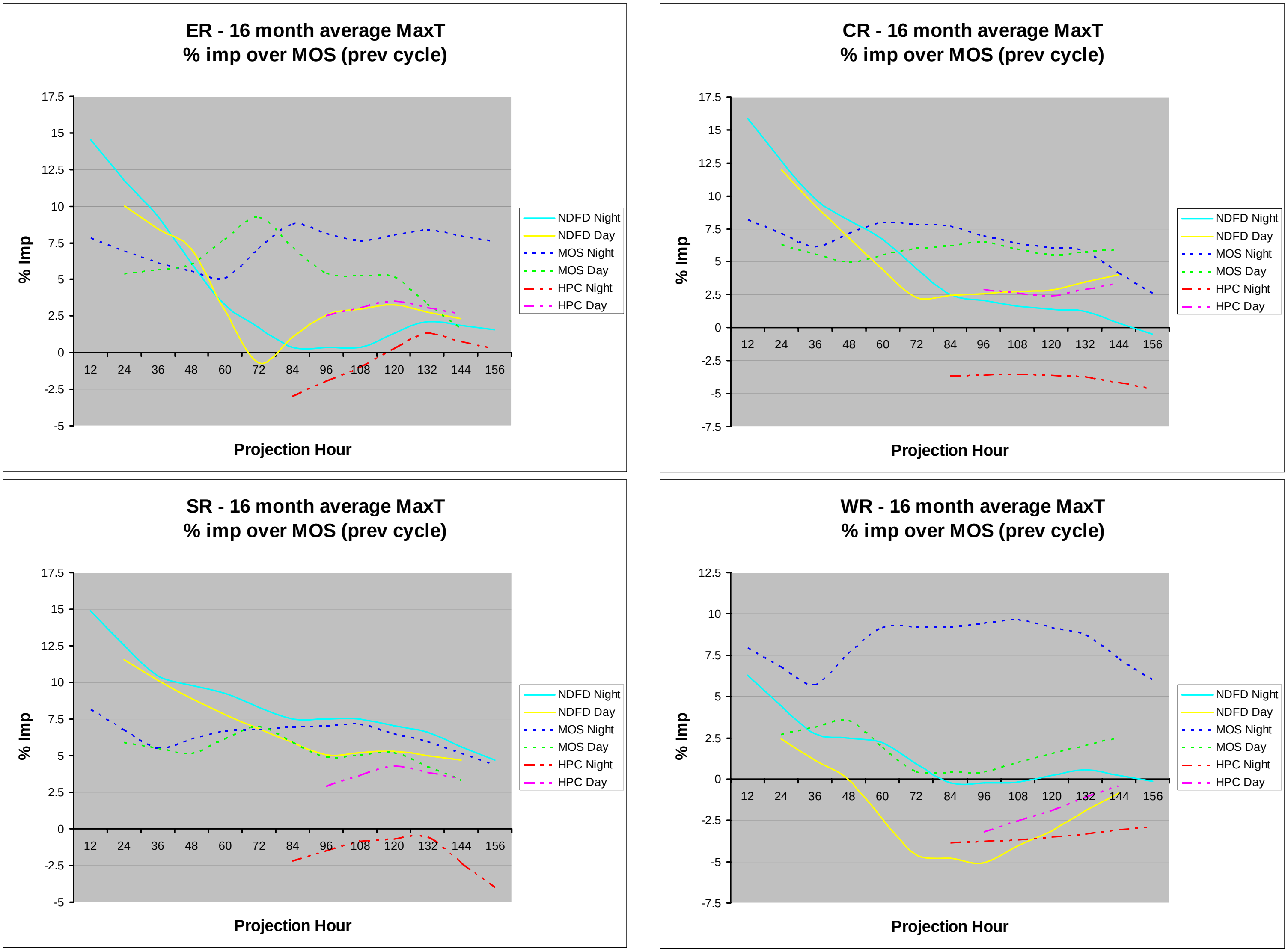


Figure 6: Sept 2005 – Dec 2006 Regional MaxT MAE % Improvement over the previous MOS cycle (e.g., day shift NDFD (NDFD Day) compared with 00z GFS MOS (MOS Night)). The percent improvement between successive model cycles is depicted by the addition of MOS to the comparison.

cycle of guidance. If there is a conscious decision to routinely base the forecast on the previous model cycle (i.e., dictated by policy), this is ill advised given that the value added by forecasters (both at the WFOs and HPC) is less than the increase in skill offered by the next incoming cycle. Furthermore, there is little evidence that suggests that the 00z GFS guidance suite is far superior to the 12z suite, to the contrary height anomaly statistics indicate the opposite. Therefore, suggesting a single cycle as preferred based upon historical context (i.e., MRF) is also ill advised. Rather, an intelligent application of both 00z and 12z guidance is necessary to consistently add value to the Day4-7 forecast. The ISST certainly does not support any policy that encourages forecasters to use latent information for the sake of convenience. Such approaches do not attempt to provide the best possible service. Rather, we support initiatives that foster the intelligent use of all the available guidance (including information from other national centers) which can be applied directly to the gridded forecast in an expedited manner.

The ISST is not advocating a single source for populating the Day4-7 forecast. Rather, we are encouraging a more thorough scientific discourse in an effort to maximize performance while reducing workload. It is in this context that the agency leverages its technological and human resources effectively. This is essentially the mantra "Forecast Smarter". However, for this to be successful forecasters must have access to real-time validation of guidance sources to make value judgments on guidance inputs.

Moreover, comprehensive post-processed guidance from a more advanced ensemble system is required. Currently, the ensemble prediction systems do not possess a level of reliability to be used routinely in forecast operations; and the post-processing of ensemble output is very rudimentary and is not immediately applicable to the production of digital forecasts.

The intended end state of the forecast process – especially in the Day4-7 timeframe – is one of managing/manipulating forecast input via more advanced selection and synthesis procedures, rather than making manual edits to a single predefined downscaled source. This change in process places more emphasis on meteorological analysis and diagnosis and less on brute force grid manipulation. **The net result of our efforts should be the identification and communication of significant meteorological events in this planning time frame, not on whether the WFOs can add an average 0.25-0.5 degree improvement on a forecast that averages an error of 6^{o}F.** This philosophy is essentially a hallmark of the proposed CONOPS; and can be incorporated into the operational setting well before the new structure of operations. However, direction and support is required at all levels of the agency to ensure success, because there is much to be done in the operational support system – especially in the avenues of guidance production, verification, and forecast production software.

Unfortunately, the Day4-7 forecast process has dominated much of the digital forecast process discussion recently. More attention is now being devoted to adding skillful precision to the near term forecast. However, the addition of skillful detail requires a significant amount of attention by forecasters, of which the current Day4-7 practices detract. There is a positive feedback loop via situational awareness between a relatively accurate/representative Day4-7 forecast and successful forecasts in the near term. Consequently, the success of providing accurate mesoscale forecasts in the near term hinges upon an efficient and accurate Day4-7 forecast.

Appendix A

Additional Verification Data (MinT / PoP)

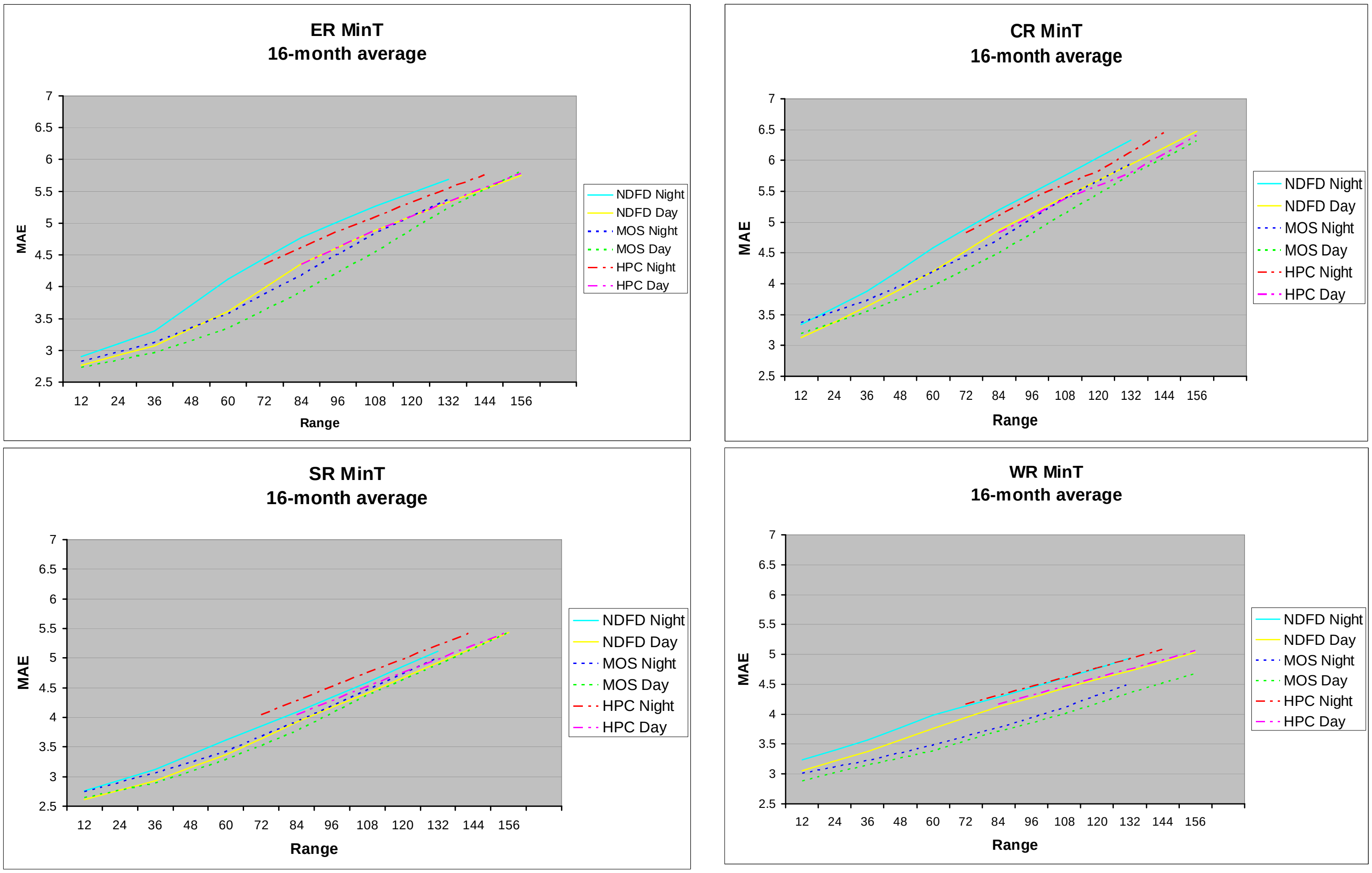


Figure A1: Sixteen month average for the four CONUS regions (Sept 2005-Dec 2006) Minimum Temperature MAE at all CONUS GFS MOS points as function of forecast range for NDFD, GFS MOS, and HPC.

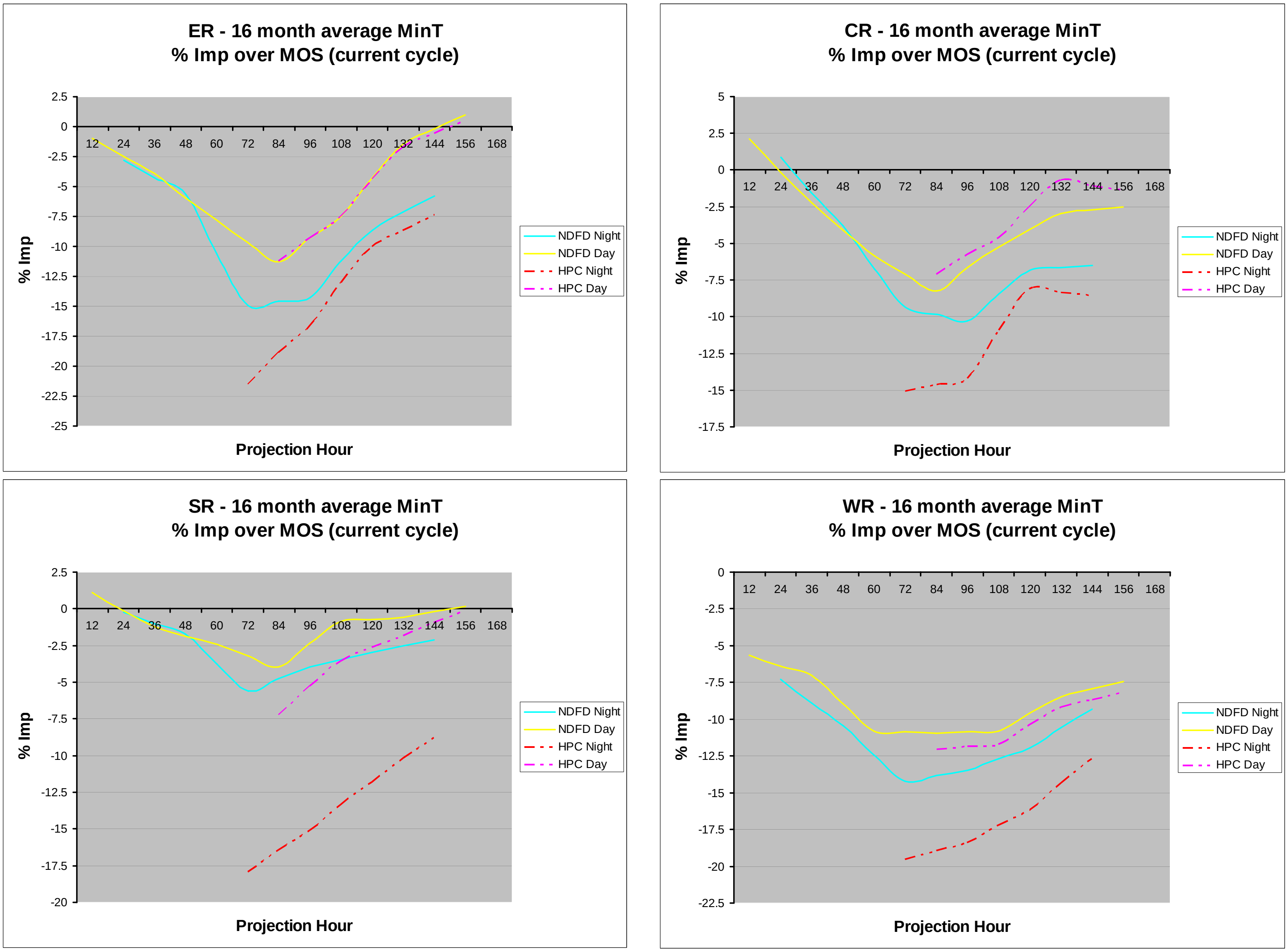


Figure A2: Sept 2005 – Dec 2006) Regional MinT MAE % Improvement over the current MOS cycle (e.g., day shift NDFD (NDFD Day) compared with 12z GFS MOS (MOS Day)).

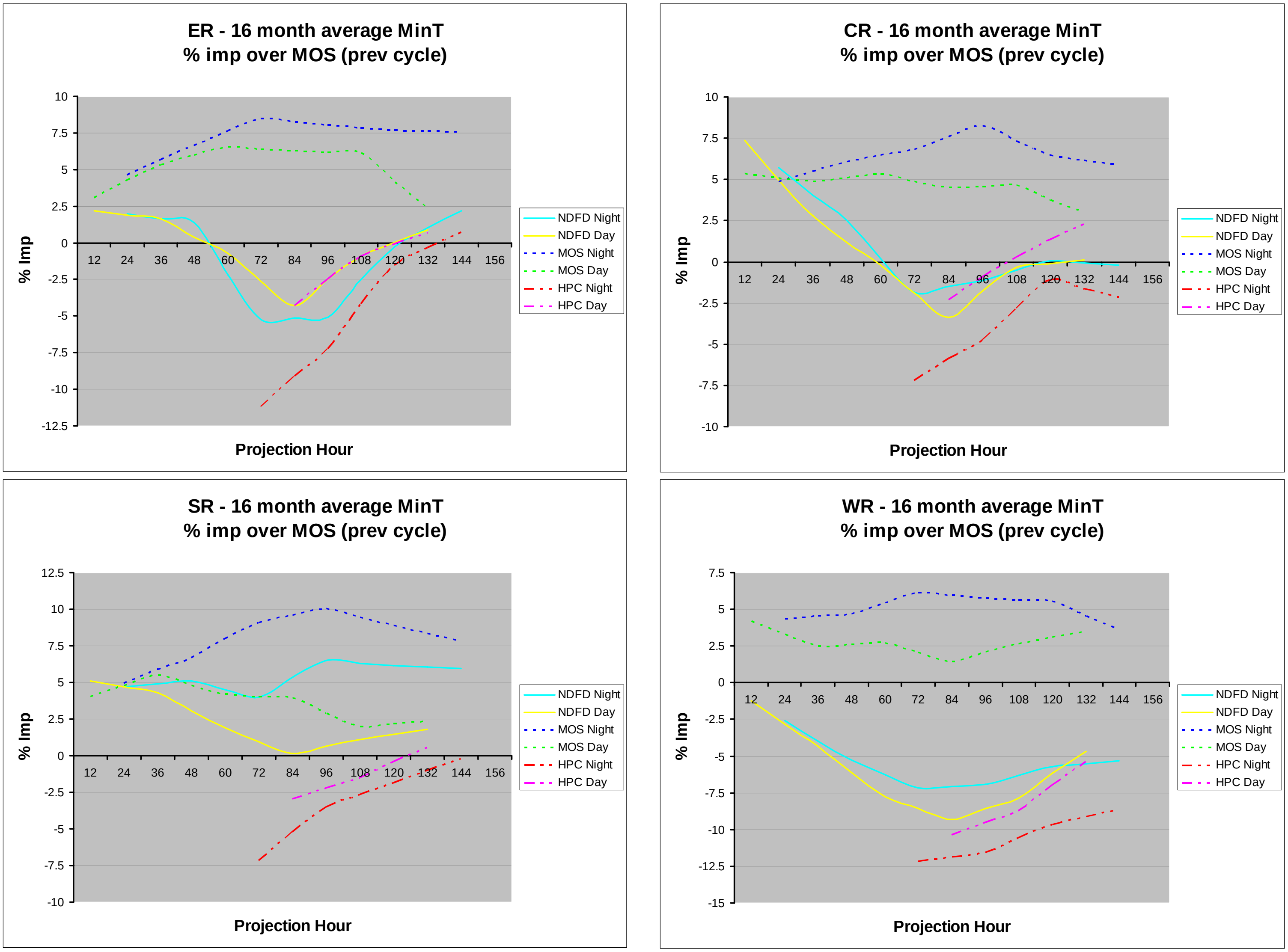


Figure A3: Sept 2005 – Dec 2006 Regional MinT MAE % Improvement over the previous MOS cycle (e.g., day shift NDFD (NDFD Day) compared with 00z GFS MOS (MOS Night)). The percent improvement between successive model cycles is depicted by the addition of MOS to the comparison.

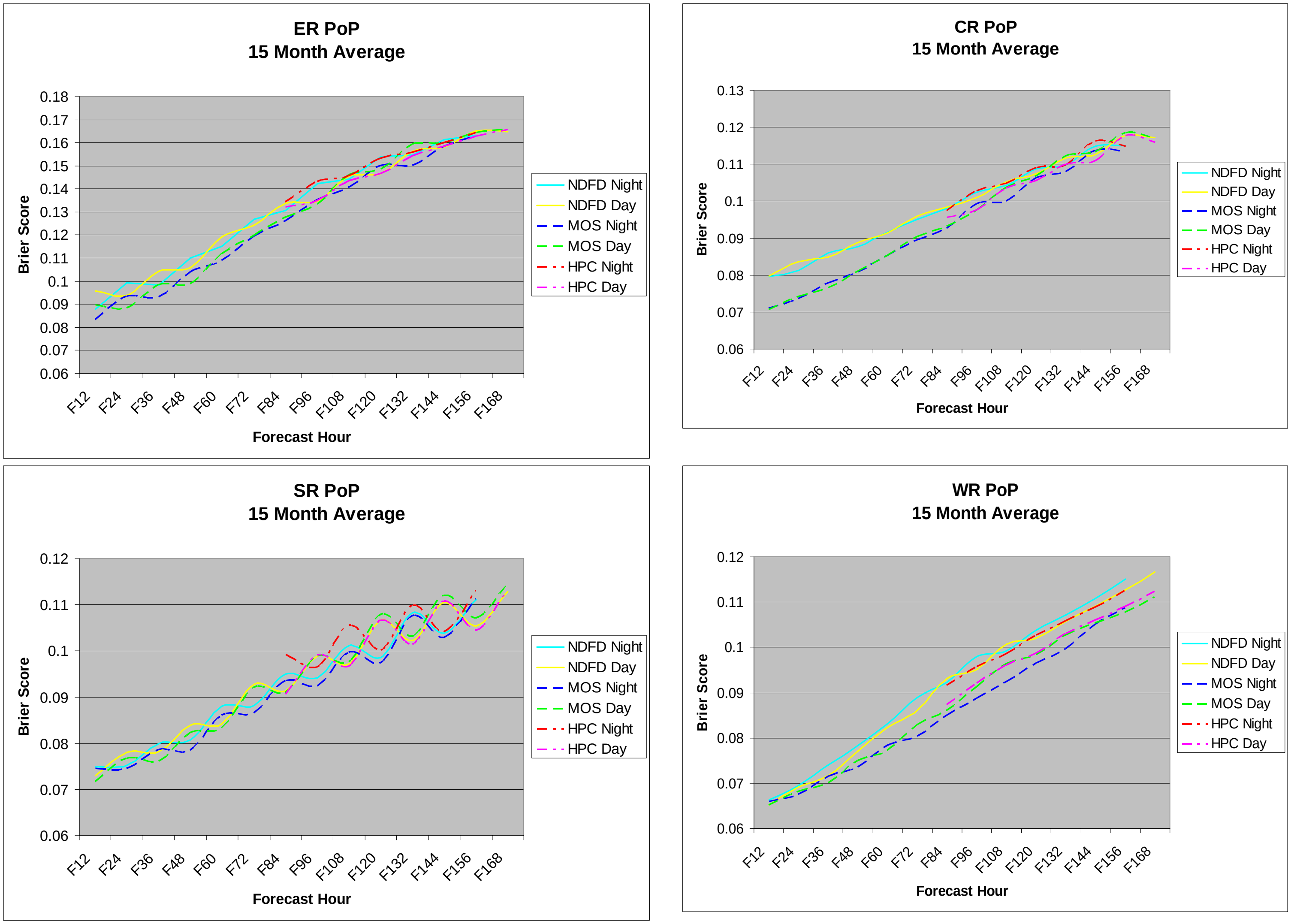


Figure A4: Fifteen month average for the four CONUS regions (Sept 2005-Dec 2006) Probability of Precipitation Brier Score at all CONUS GFS MOS points as function of forecast range for NDFD, GFS MOS, and HPC.

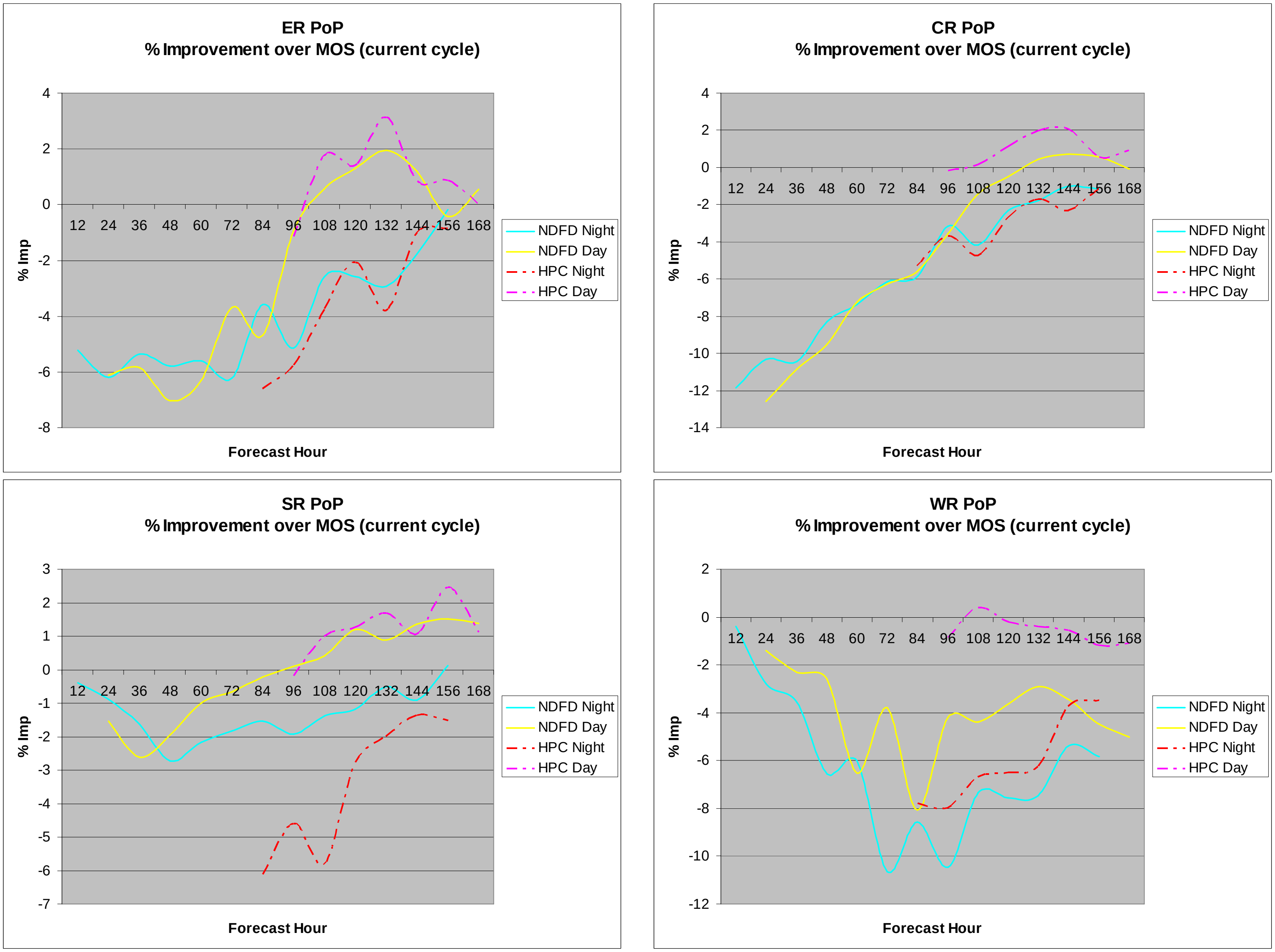


Figure A5: Sept 2005 – Dec 2006 Regional PoP Brier Score Improvement over the current MOS cycle (e.g., day shift NDFD (NDFD Day) compared with 12z GFS MOS (MOS Day)).

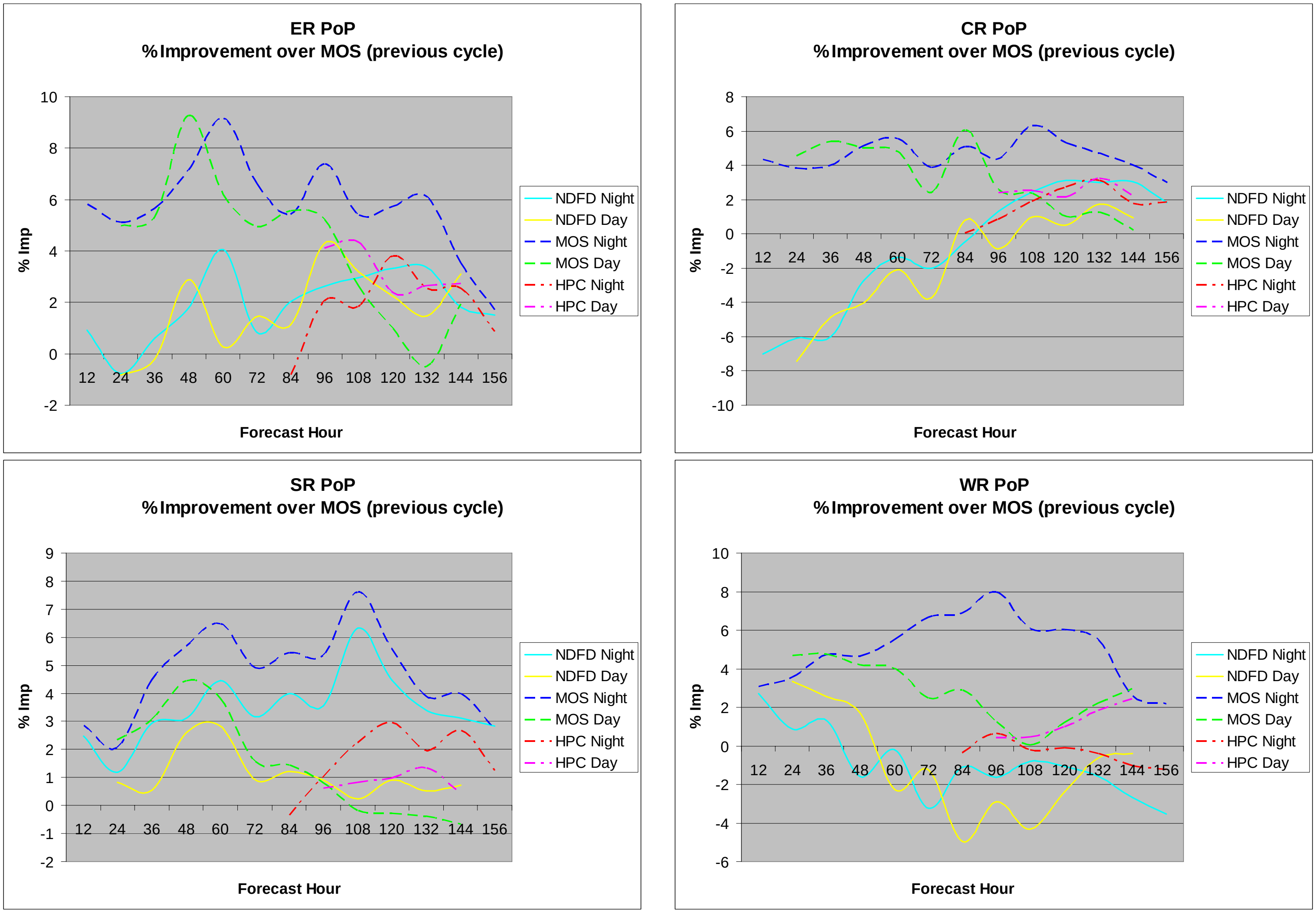


Figure A6: Sept 2005 – Dec 2006 Regional PoP Brier Score Improvement over the previous MOS cycle (e.g., day shift NDFD (NDFD Day) compared with 00z GFS MOS (MOS Night)). The percent improvement between successive model cycles is depicted by the addition of MOS to the comparison.

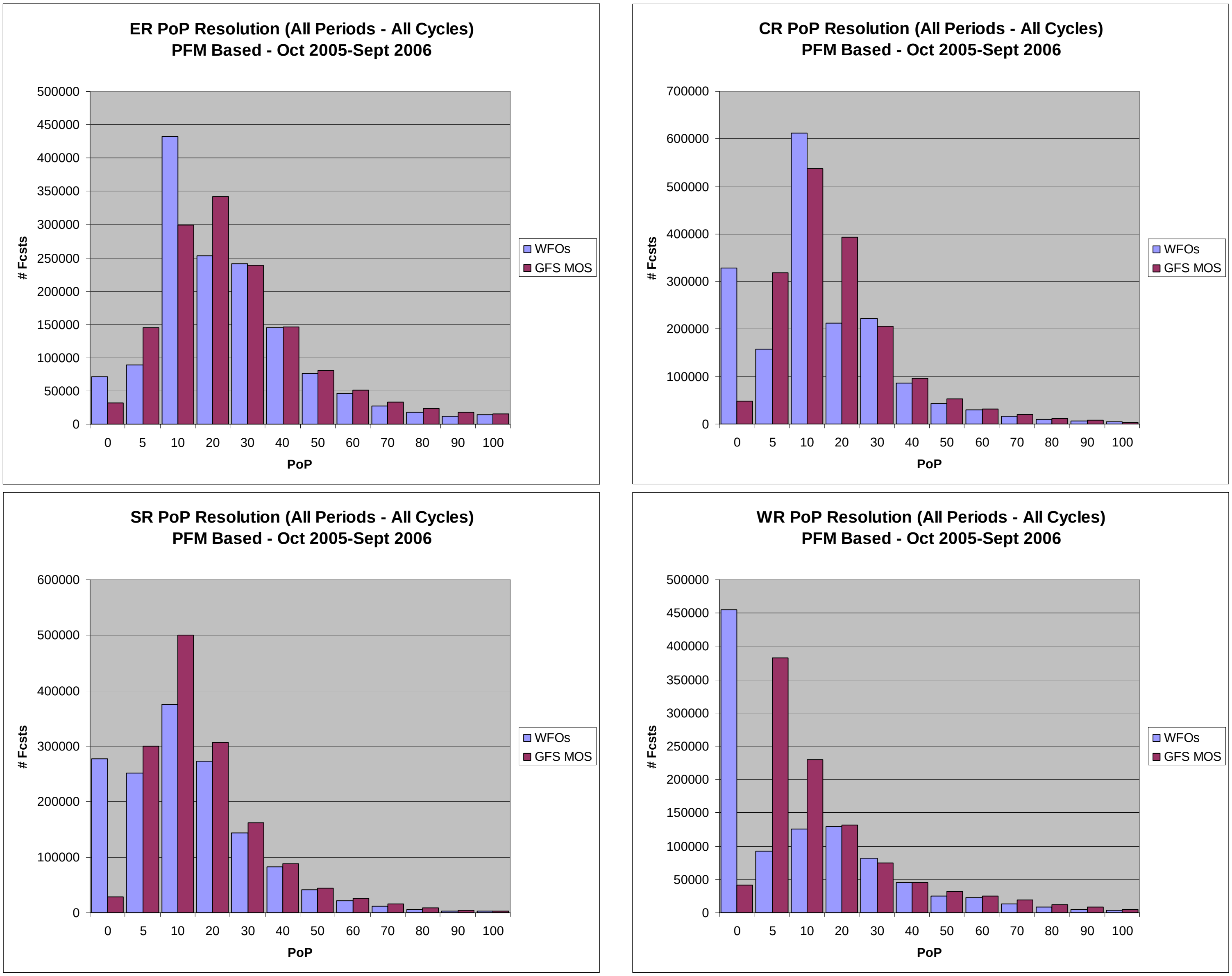


Figure A7: Regional PoP Resolution for Oct 2005 – Sept 2006 – Based upon the Point Forecast Matrix (PFM) verification statistics.

Appendix B

Reference Terms

CMC Global– Canadian Meteorological Center Global model
CONOPS – Concept of Operations Initiative
CONUS – Continermous United States
CR – Central Region
Day4-7 – Day 4-7 forecasts
DFP – Digital Forecast Process
ECMWF – European Center for Medium range Weather Forecast model
ER – Eastern Region
FWC – Nested Grid Model (NGM) MOS text bulletin
GEFS – Global Ensemble Forecast System
GFS – Global Forecast System
HPC – NCEP / Hydrometeorological Prediction Center
ISST – IFPS Science Steering Team
MAE – Mean Absolution Error
MAV – GFS MOS text bulletin
MaxT – Maximum Temperature
MinT – Minimum Temperature
MET – Eta MOS text bulletin
MOS – Model Output Statistics
MRF – Medium Range Forecast model (predecessor to the GFS)
NDFD – National Digital Forecast Database
NOGAPS – U.S. Navy Operational Global Atmospheric Prediction System
PoP – 12 hour Probability of Precipitation
SR – Southern Region
UKMET – United Kingdom Meteorological model
WFO – Weather Forecast Office
WR – Western Region

Appendix C

IFPS Science Steering Team Membership

Greg Mann – team lead	CR	WFO Detroit, MI
James Nelson – backup lead	AR	WFO Anchorage, AK
Dan St. Jean	ER	WFO Gray, ME
Reid Hawkins	ER	WFO Wilmington, NC
Karl Jungbluth	CR	WFO Des Moines, IA
Tom Salem	WR	WFO Glasgow, MT
Mike Staudenmaier	WR	WFO Flagstaff, AZ
Bill Ward	PR	SSD Pacific Region Headquarters
Amy McCullough	SR	WFO San Angelo, TX
Lee Anderson – facilitator	NWSHQ	OST

Former members contributing to this effort:

Jeffrey Medlin	WFO Mobile, AL
Ken Falk	WFO Shreveport, LA
Steven Keighton	WFO Blacksburg, VA