Forecasting Regional Chance of Occurrence through Aggregation of MOS PoPs

GEORGE S. YOUNG, SREECE D. GOLDBERGER, JOHANNES VERLINDE, and CHRIS J. HANLON
The Pennsylvania State University, University Park, Pennsylvania

(Manuscript received 15 July 2014; review completed 4 December 2014)

ABSTRACT

Regional sets of point probabilities [e.g., model output statistics (MOS) probability of precipitation (PoP) forecasts] can be used to forecast both the expected areal coverage and the regional chance of occurrence. The former is just a data quality check given the statistical equivalence of the spatial average of point probabilities and the expected value of areal coverage. The latter, however, is a new product with utility to those users whose actions depend upon the occurrence of the event anywhere within their region of operation. While the method is demonstrated here using MOS PoP forecasts, potential applications include any weather variable for which point probability forecasts are available operationally. The forecasting of thunderstorm chance of occurrence for fire weather planning is one such application.

For this demonstration, MOS PoP forecasts are matched with the corresponding National Centers for Environmental Prediction Stage IV precipitation analyses. Comparison of regional average PoP with fractional area coverage verifies their equivalence, but reveals a MOS station location bias to the drier lowlands in the inter-mountain West. Regional chance of occurrence is forecast via logistic regression with the mean and standard deviation of the region's MOS PoP forecasts as predictors. Hindcast results show significant skill, but the regression equations vary by both season and location.

1. Introduction

Many societally important actions depend on the regional occurrence of a particular type of weather. Decision to undertake these actions in the face of uncertainty thus requires a forecast of the regional chance of occurrence (RCO). Forecast problems of this sort arise when there is an area-deployable resource or a specific region of responsibility. For example, wildfire managers need to know the probability of lightning within their region of responsibility (Rorig et al. 2007). Likewise, weather researchers deploying aircraft and other limited resources require probability forecasts for the occurrence of the event of interest within the region of their field experiment (Hanlon et al. 2013, 2014). The same sort of forecast problem arises in area threat avoidance applications such as enroute aviation, flash flooding, and lightning safety. Going further from traditional applications, equivalent problems occur in crop pest forecasting (Hirschi et al. 2012).

Traditional model output statistics (MOS) use regression equations to convert raw model output into

forecasts of user-relevant weather parameters (Glahn and Lowry 1972). Operational MOS uses regression to predict the probability of an event at a particular location, probability of precipitation (PoP) being one of the most commonly used examples (Glahn and Lowry 1972). This point probability approach has the advantage that training data are available without remote sensing. Moreover, it is the relevant predictand for many applications. For area applications such as those discussed above, the expected fractional area coverage (FAC) is mathematically equivalent to the regional average of point probabilities, so it can also be derived from operational MOS guidance. It should be noted, however, that this is only the expected value in the statistical sense of time average. This definition of the term expected value will be used throughout the subsequent discussion. Day-to-day regional averages of MOS PoPs may differ greatly from the actual fractional area coverage.

The chance of occurrence of the phenomena of interest anywhere within a region is a completely different quantity. For example, with popcorn con-

vection one can imagine situations where the chance of convection occurring somewhere within the region is high while the fractional area coverage is low. Thus, a separate forecast system is needed for regional chance of occurrence. The primary challenge to building an RCO forecast system is the need for regional observations of the predictand. This is often easiest to obtain via remote sensing, for example using radar to provide training data for an RCO precipitation forecast.

An RCO forecast system can be developed in one of two ways: either as a stand-alone regression equation using raw model output as predictors or as a piggyback on operational MOS. Building a stand-alone RCO system is a challenge of similar scope to the development of MOS itself, requiring access to a large archive of raw model output and considerable insight into appropriate predictors. In contrast, building a piggyback RCO system merely requires access to a large archive of MOS probability forecasts and a statistically sensible approach to aggregating them. In short, piggybacking on MOS point probability forecasts allows the RCO forecaster to exploit the data archives and expertise of the National Weather Service's Meteorological Development Laboratory. This is the approach taken here.

2. Data and methods

In order to demonstrate the potential of MOS piggyback modeling of FAC and RCO across varied climates, four regions are tested: North centered in southwestern Minnesota, South centered in north-central Texas, East centered in southwestern West Virginia, and West centered in northeastern Utah. These regions sample the northern and southern Great Plains, the Appalachian Mountains, and the interior of the mountainous West. For each of these regions a nested set of boxes (100-, 200-, 400-, and 800-km square) is tested. The complete model building described below was conducted for each box size in each region (Fig. 1). The number of MOS PoP stations in each box is given in Table 1. Although the number of stations in the smallest boxes is small, these boxes are included in the study to investigate the feasibility of using the method for small domain sizes.

Statistical model development requires training data for both the predictors and predictand. For this study the regional mean and standard deviation of the 6-h Global Forecast System (GFS) MOS PoP forecast (Ruth et al. 2009) were used as predictors. The mean

Table 1. The number of MOS PoP stations in each of the four regional boxes.

	100 km	200 km	400 km	800 km
North	7	15	51	130
South	10	12	23	83
East	2	10	31	119
West	4	6	11	49

Figure 1. Box locations for each of the four regions: North, South, East, and West. The boxes for each region are not concentric, but rather the smaller boxes are positioned within the larger boxes in such a manner as to maximize the number of included MOS sites. *Click image for an external version; this applies to all figures hereafter.*

was selected because of its direct relationship with FAC and its expected impact on the RCO. The standard deviation was included to capture those situations where precipitation was more likely in one part of the region than another. While the maximum has some advantages over standard deviation, particularly as it nears 100%, it may not be representative of the true regional maximum if the station count is low. Because this is often the case for the smallest regions studied here, the combination of average and standard deviation provides a more robust estimate of the maximum PoP within the region. The same argument indicates that it also would be unwise to use quantiles to capture the shape of the regional MOS PoP distribution. Sample size considerations also influenced the choice of GFS MOS. MOS PoP forecasts from other models could have been used instead, but the GFS provided a large training set for development of our piggyback models.

The predictands are FAC by precipitation and occurrence of precipitation anywhere within the region. The former ranges from 0 to 100% while the latter is either 0 or 100%. These statistics are computed from NCEP Stage IV precipitation analyses (Lin and Mitchell 2005) for the period 1 January 2002 through 31 December 2013. This multi-sensor precipitation estimate has a 4-km resolution. The 6-h accumulations were used in order to match the valid periods

for the GFS MOS PoP forecasts. All lead times from 12 through 72 h were used, but each lead time was modeled separately so as to allow for the increase in uncertainty with lead time. FAC was computed by dividing the number of grid points within a box that received precipitation by the total number of grid points in the box. This approach is accurate to the extent that the Stage IV dataset's polar stereographic map factor is constant across a box. The MOS forecasts were obtained from www.mdl.nws.noaa.gov/~mos/archives/.

Because PoP is the probability that precipitation will occur at any given point in the forecast area, the spatial average of the PoPs is equivalent to the expected value of FAC. The two statistics will be equal in practice, however, only if the MOS PoP sites provide an unbiased sample of the regional climatology (i.e., if the average across MOS stations of their climatological average PoP value is the same as that for the region as a whole). Moreover, as mentioned in the Introduction, the equivalence in expected value does not imply day-to-day equality, only equality of long-term averages.

In contrast to FAC, the relationship between RCO and mean and standard deviation of MOS PoP is not theoretically obvious. The limiting cases do, however, shed some light on the subject. For a uniformly precipitation free region, both mean and standard deviation of PoP will be zero as will the RCO. At the other end of the range, a region certain to be covered by precipitation will have an average PoP of 100% and PoP standard deviation of 0%. Between these two extremes, a plot of PoP standard deviation against average PoP shows considerable spread, indicating that PoP standard deviation provides information not captured in average PoP. Thus, both average PoP and PoP standard deviation were tested as predictors of RCO.

Because the expected value of FAC is just the average PoP, a simple linear relationship between the two parameters was fit using ordinary least squares regression. When built with a sufficiently large set of training cases (14 868 for this study), the linear regression coefficients linking regional average PoP to observed FAC provides a check on the representativeness of the MOS sites. In cases of unrepresentative site locations, this linear regression equation provides a means of converting regional average MOS PoP into regional FAC. The more complex nature of the RCO relationship, and its hard bounds between 0 and 100%, suggest logistic regression is more appropriate. Indeed,

logistic regression is the simplest statistical model that meets these requirements.

Because the ratio of stratiform to convective precipitation varies on an annual cycle, there is the potential for the RCO relationship to change with time of year. Other seasonal factors might be at work as well. Therefore, the RCO model for each region is developed for three sets of periods: 1) four meteorological seasons of December–February (DJF), March–May (MAM), June–August (JJA), September–November (SON), 2) cold and warm seasons (October–March and April–September), and 3) the entire year taken together. Skill of the various models will be compared using the Brier skill score (BSS; Wilks 2006, p. 259) to see if it is indeed important to take seasonal effects into account. The reference forecast used in the BSS was the climatological forecast derived from the Stage IV precipitation data because no more sophisticated competing forecast was available. BSS should, thus, be interpreted as a measure of improvement over climatology, with values above zero indicating that the forecast system being tested is more skillful than a climatological forecast. Relative skill of the seasonal and annual models was also evaluated using the Brier score (BS; Wilks 2006, p. 284) to see if it is important to take seasonal effects into account. All of these tests were conducted on the training dataset, so despite the large number of cases and the small number (three) of parameters being fit, the skill estimates may be biased high.

3. Analysis and discussion

a. Fractional area coverage

Linear regression reveals that the expected value of FAC is quite close to the average PoP in the North, South, and East regions. Slopes range from 0.8 to 1.2, increasing smoothly with lead time and do not vary much with area size (Fig. 2). Thus, in these regions, the average PoP is a good estimate of the expected value of FAC. An increasing slope means that it takes less average PoP to generate the same FAC. This trend may reflect the decrease in sharpness (i.e., the long-term MOS PoP distribution is closer to the climatological odds of precipitation) with increasing lead time, thereby requiring a larger coefficient to achieve the same response to variation in PoP. In contrast, in the West region the slope ranges from 1.38 to 2.19 (i.e., roughly 1.7 times greater than for the other regions) and appears to decrease somewhat with increasing region size. This means that average PoP

Figure 2. Slope of the regression line linking mean PoP to FAC. Each subplot corresponds to one of the four regions: North, South, East, and West. Line width increases with box size: 100, 200, 400, and 800 km.

underestimates FAC in the mountainous West, probably because the MOS sites are primarily in dry valleys so, while locally accurate, the PoP forecasts would not be regionally representative.

Intercept values for all regions are rather small, -3% to $+2\%$, and, except in the North, decrease with lead time, presumably to keep the FAC climatology the same in the face of increasing slope. The intercept is a small correction on what is basically an equivalence relationship between FAC and average PoP. There is, however, considerable scatter around this one-to-one line when FAC is plotted against average PoP (not shown), especially at longer lead times as one would expect given the relationship between PoP skill and lead time. Thus, predicting FAC from average PoP is an error-prone process even with this regression model.

b. Regional chance of occurrence

Logistic regression models for predicting the RCO from the average and standard deviation of PoP can be evaluated using p values, BS (i.e., mean squared error, MSE), and BSS (i.e., one minus the ratio of the model's BS to that of a climatological forecast), and understood in terms of their intercept and predictor coefficients (Wilks 2006, p. 285). Large p values for one of these predictors indicate doubt about their ability to inform RCO. A small BS or large BSS, on the other hand, indicates that these two predictors suffice to yield skillful forecasts of RCO. More

specifically, since these are tests that were conducted on the training data, a BSS skill greater than zero indicates these predictors provide a better fit to the data than does climatology.

1) p VALUES

Because the logistic regression equation for RCO includes an intercept term, as well as coefficients for the average and standard deviation of PoP, there are three p values to examine, each varying with region, box size, and lead time. These will be discussed first for the annual model and then for the various seasonal models. Because the p values depart significantly from zero for so few lead times, these plots are not shown. Rather the p values will be used to determine which lead times to plot on the coefficient plots (i.e., only those with p values less than 1%). Thus, only those coefficients for which the null hypothesis (that their value is zero) can be rejected with high confidence are plotted.

The annual model's intercept is statistically significant (p value of 1% or less) for the North and South regions at all box sizes and lead times. The same holds true for the East and West regions, except for the largest (800 km) box at lead times beyond 36 and 42 h, respectively. The annual model's coefficient for mean PoP is significant for all box sizes and lead times in the West region and in the North, South, and East regions for all box sizes and lead times except 800 km for less than 18 (North), 24 (South) and 36 h (East). Results are similar for the annual model's coefficient for the standard deviation of PoP, which is significant for all box sizes and lead times in the North, South, and West regions and for all box sizes except 100 km in the East. These statistical results indicate that it is almost always best to retain both the average and standard deviation of PoP as predictors in an annual logistic regression model for RCO. As mentioned during the discussion of predictor selection, it makes sense that both mean PoP (because of its relation to FAC) and standard deviation of PoP (a measure of spread of distribution) inform RCO. For situations with small standard deviation, mean PoP is a good indicator of overall PoP in the region, whereas large values of standard deviation indicate the likelihood of large PoP somewhere in the region.

Seasonal logistic regression models for RCO yield similar conclusions about predictor significance. The intercept is significant for both two-season and four-season models for all box sizes except 800 km. The

coefficient for average PoP is significant in all of these models except for the 800-km box at various lead times (depending on region). Compared with the annual models, the seasonal models, especially the four-season models, have more lead times for which the average PoP is not a statistically significant predictor in the 800-km box (14% for annual, 30% for two-season, and 46% for four-season models). This finding suggests that the average PoP in the annual models is capturing seasonal variability that is captured instead by the intercept in the seasonal models, as discussed below in the sections on BSS and regression models. The coefficient for the standard deviation of PoP fails to reach statistical significance for some box-size and lead-time combinations (4% for annual, 8% for two-season, and 12% for four-season models). Thus, most (88–96%) of the box-size and lead-time combinations do yield statistically significant coefficients for the standard deviation of PoP. Given that both predictors are significant in most of the models, they will be retained in all of the seasonal logistic regression models for RCO. In the figures discussed below, the coefficients are only plotted for each projection if they are significant.

2) BRIER SKILL SCORE

The BSSs (Figs. 3–9) of both the annual and seasonal logistic regression models for RCO are large enough to indicate potentially useful skill, but by no means perfect forecasts. That is, while BSS is greater than zero for all combinations of region, box size, and lead time, it never reaches 1.0. Given that the underlying MOS PoP is itself an imperfect predictor (Ruth et al. 2009), this result is not surprising.

Table 2 shows the maximum BSS as a function of region and box size for the annual model results displayed in Fig. 3. Recall from the discussion above that BSS was computed on the training dataset, so these values may be biased high. The maximum BSS occurs in the first two lead times for all but one region/box size combination (South/800 km), as would be expected from the decline of MOS PoP skill with lead time. The maximum BSS falls off with increasing box size, except in the South region where it is nearly constant across box sizes and the North where it only falls off at 800 km. The decrease in BSS for RCO with increasing region size is expected as climatology becomes increasingly hard to beat as the region becomes larger. This occurs because the climatological chance of precipitation occurring somewhere in a

Figure 3. BSS for the RCO forecast plotted as a function of lead time for the annual model. Each subplot corresponds to one of the four regions: North, South, East, and West. Line width increases with box size: 100, 200, 400, and 800 km.

Figure 4. Same as Fig. 3, but for the warm-season model, April–September.

Table 2. The maximum across lead times of BSS for RCO from the annual model as a function of region and box size. Numbers in parentheses are the lead times (h) at which the maximum occurs.

	100 km	200 km	400 km	800 km
North	0.45 (12)	0.45 (12)	0.41 (18)	0.32 (18)
South	0.42 (12)	0.43 (12)	0.43 (18)	0.42 (24)
East	0.43 (12)	0.39 (18)	0.36 (18)	0.26 (18)
West	0.40 (12)	0.38 (12)	0.33 (12)	0.24 (12)

region at any given time increases as the region size approaches that of a synoptic weather system. This effect is particularly clear in the limit, where the

climatological RCO would approach 100% long before box size reaches global.

The BSS of the two-season logistic regression models for RCO varies in a pattern consistent with convection being more challenging to forecast than stratiform precipitation. Warm-season skill generally is a bit less than cold-season skill for all regions, box sizes, and lead times (compare Figs. 4 and 5). Warm-season trends in skill with box size are similar to those of the annual models. The only exception is that the 800-km box in the South exhibits less skill than the other box sizes, rather than similar skill as is the case in the annual model. The cold-season skill trends with box size are similar to those of the annual models. The warm- versus cold-season skill difference is most pronounced in the South and East, and least pronounced in the North. These regional differences reflect the differing annual cycles in these regions, with the North's weather remaining dominated by synoptic cyclones and fronts throughout the year while that of the South and East are dominated by tropical and subtropical convective systems in the summer, but synoptic cyclones and fronts in the winter.

The performance of four-season RCO models (Figs. 6–9) differs in a number of ways from that of two-season and annual models. The four-season models exhibit much less skill decline with lead time than occurs with two-season and annual models. Likewise, the increase in skill with decreasing box size is even more consistent with four-season models than with two-season or annual models. However, unlike the two-season models, 100-km boxes are not consistently more skilled in DJF versus JJA. In fact, 100-km boxes in MAM and SON are more skilled than in DJF and JJA, a surprising example of transition seasons being easier to forecast. Most interesting is that skill increases with box size for four-season models (consistently) rather than decreasing with box size as it does for most two-season and annual models. At 800 km the greatest skill is in summer, followed by spring/fall and then winter.

These findings have several implications. RCO prediction from point PoP is easier in the cold season than in the warm season—not surprising given the challenges posed by convection. Four-season models give very different dependence of skill on lead time, season, and box size than do two-season models. Four-season models derived for large boxes are consistently more skilled than those for small boxes. Because the enhanced skill of the four-season models derives from having a separate model fitted to each season's clima-

Figure 5. Same as Fig. 3, but for the cold-season model, October–March.

Figure 6. Same as Fig. 3, but for the winter model, December–February.

tology, this increase in skill with box size may imply a seasonal dependence in the regression relation for RCO for large box sizes that is absent for small box sizes. One possible explanation is that there is a greater chance of partial coverage by seasonally dependent (e.g., convective) mechanisms in big boxes, so that big-box standard deviation has different origins with different seasons. This difference would be reflected in seasonal variation in the regression coefficient for standard deviation (see the next subsection).

As expected, skill decreases with lead time for all regions, box sizes, and seasons. The four-season models lose skill more slowly with lead time than do two-season and annual models. This difference implies that

Figure 7. Same as Fig. 3, but for the spring model, March–May.

Figure 8. Same as Fig. 3, but for the summer model, June–August.

seasonal effects get stronger with lead time. This result can be interpreted in terms of the contribution of numerical weather prediction (NWP) model skill and climatology to the RCO forecasts. For four-season models climatology has more skill, due to the more limited climatological range. Thus, the skill of MOS PoP forecasts and the resulting RCO forecasts decrease less rapidly as NWP model skill falls off with lead time. For all models smallest boxes lose skill fastest, presumably due to timing and track errors in the underlying GFS model. In contrast, large boxes are less sensitive to location errors because they allow room for a precipitation event to just shift position within the box in (at least some cases).

The comparison of skill between annual and seasonal models is somewhat simpler. Because both

Figure 9. Same as Fig. 3, but for the fall model, September–November.

would be compared to the same climatological forecast in BSS, the same insight can be gained by simply comparing the BS of the annual model with the average BS of the seasonal models (Figs. 10, 11, and 12). Recall from the discussion above that BS was computed on the training dataset, so the individual values may be biased low. Improvements in BS resulting from using seasonal models were fairly steady with lead time. The two-season models typically provide less than 10% improvement in BS, although this value varies considerably with region and box size. In contrast, four-season models yielded slightly greater than 50% improvement in BS for all regions and box sizes. Clearly, four-season modeling of RCO is the best of these choices. The optimal seasonal breakdown, however, remains an open question as only the conventional meteorological seasons were tested.

3) REGRESSION MODELS

The RCO logistic regression coefficients and their variation with region, box size, and lead time provide insight into the origin of the models' forecast skill. Because the logistic regression equation is S-shaped (real number inputs, but outputs restricted to the range of 0 to 100%), interpreting these results is not as intuitive as is the case for linear regression. Two rules of thumb suffice though: 1) decreasing the intercept pushed the probability curve to the right, requiring larger values of slope multiplied by predictor to achieve the same odds; and 2) positive slope means odds increase as the predictor value increases.

Figure 10. BS for the RCO forecast plotted as a function of lead time for the annual model. Each subplot corresponds to one of the four regions: North, South, East, and West. Line width increases with box size: 100, 200, 400, and 800 km.

Figure 11. Same as Fig. 10, but for the two-season model.

The intercept turns out to be predominately negative and nearly constant with lead time for all regions (Fig. 13). It decreases in magnitude with increasing box size, nearing zero at 800 km for North and becoming somewhat positive at 800 km for East and West. Except for the West the intercept is smaller in magnitude in the warm-season model and larger in magnitude in the cold model than in annual models. For four-season models (not shown) winter has the largest intercept magnitude, summer the smallest, and spring and fall intermediate values. The trend of intercept towards zero with increasing box size can be interpreted using the two rules of thumb given above.

Figure 12. Same as Fig. 10, but for the four-season model.

Figure 13. The intercept in the logistic regression equation for the RCO plotted as a function of lead time for the annual model. Each subplot corresponds to one of the four regions: North, South, East, and West. Line width increases with box size: 100, 200, 400, and 800 km. Intercepts with p values greater than 1% are not plotted.

A less negative intercept means less need for the positive predictor values in order to achieve high RCO. Thus, this trend may reflect the fact that the climatological likelihood of precipitation occurring somewhere within the region at any given time increases with box size. Taken to the extreme, the intercept would be strongly positive for a global box and the PoP statistics irrelevant because it is always precipitating somewhere.

The coefficient for average PoP is always positive in the annual models, indicating that increasing mean PoP yields increasing RCO as expected (Fig. 14).

Figure 14. The coefficient for the average PoP in the logistic regression equation for the RCO plotted as a function of lead time for the annual model. Each subplot corresponds to one of the four regions: North, South, East, and West. Line width increases with box size: 100, 200, 400, and 800 km. Coefficients with p values greater than 1% are not plotted.

There is, however, considerable regional variation in this coefficient, particularly for the larger boxes. Variation with lead time depends on box size, slightly decreasing in magnitude with lead time for 100 and 200 km for all regions, flat in the West and South, and decreasing in the North and East for 400 and 800 km. The two-season models show similar mean PoP coefficient values, except in the West where the warm-season value is higher than the cold-season value. The coefficient increases with lead time for the warm and cold seasons for 400 and 800 km in most regions. The four-season models show the same effects. The mean POP coefficients do not vary much with season.

The coefficient for the standard deviation of PoP is positive for almost all regions, box sizes, lead times, and seasons (Fig. 15), the sole exception in the annual model being the two lead times that reached statistical significance for the 100-km box in the East region. The sign of this coefficient follows from RCO being at least as large as the largest PoP, so RCO exceeds average PoP if the standard deviation of PoP is nonzero. The magnitude of this effect varies substantially between regions. Moreover, the seasonal variation in magnitude is not consistent between regions (not shown). The effect is often larger for larger boxes, particularly at shorter lead times. The lead-time trend in the coefficient for the standard deviation of PoP varies with region, season, and box size. The largest

Figure 15. The coefficient for the standard deviation of PoP in the logistic regression equation for the RCO plotted as a function of lead time for the annual model. Each subplot corresponds to one of the four regions: North, South, East, and West. Line width increases with box size: 100, 200, 400, and 800 km. Coefficients with p values greater than 1% are not plotted.

effect is for large boxes at short lead times. These results explain the BSS differences between seasonal and annual models. Capturing the seasonal variation in this coefficient is important to minimizing RCO forecast error.

The variation of the two logistic regression coefficients with lead time shows larger regions have a stronger relationship between RCO and average PoP, especially at longer lead times (i.e., the same average PoP yields a higher RCO the farther out you go, but not so strongly for small boxes). This result could indicate small skill of average PoP for small boxes (an expected result). In some regions and seasons though, at short lead times, small boxes can have a stronger relationship between mean RCO and average PoP than do large boxes. Perhaps for those region/season combinations climatology dominates, although it is counterintuitive why it should do so at short lead times but not long lead times. The answer may be that, for short lead times in big boxes, variability takes on much greater importance. Large variability coupled with low average PoP means someplace in the region has high PoP and will almost certainly receive precipitation, but only at short lead times when the forecast is fairly certain, and only for boxes big enough to get a good sample of the standard deviation of PoP.

4. Conclusions

Can regional chance of precipitation occurrence be forecast from MOS PoP? The answer is yes, not perfectly but with skill much better than climatology. The tests here were conducted on the training data, however, so skill estimates may be optimistic despite the large number of training cases (~3750 for four-season models, ~7500 for two-season models, and ~15 000 for annual models) relative to the number of parameters fit (three). Future work should include testing on independent data. On the training data, this skill extends to lead times of at least 72 h. The best strategy is to use both the average and standard deviation of PoP as predictors, retain an intercept, and build a separate logistic regression model for each of the four seasons.

General results are as follows:

- Skill was greatest for four-season models and least for annual models, highlighting the importance of seasonal differences in the relationship between the mean and standard deviation of PoP and RCO.
- Reduction of BS relative to a climatological forecast ranges from 90% for large boxes in the summer to 25–45% for small boxes in the annual models.
- BSS decreases with lead time, with the decrease being most pronounced in the annual model and least pronounced in the summer model.
- BSS decreases with box size for most annual and two-season models, but increases with box size for almost all four-season models.
- For the two-season models skill is typically greatest in the cold season. For the four-season models, however, the pattern is reversed for large boxes, which exhibit their greatest skill in summer.
- Despite low station counts, both skill and regression coefficients for the smallest boxes varied smoothly with lead time in a manner consistent with that of the larger boxes.

There are, of course, caveats:

- MOS sites are not always representative of the region (e.g., West), although the logistic regression equation can compensate to some extent.
- While the four-season models show much smaller BS than do the two-season and annual models, it is not yet clear which seasons are the best (e.g., the traditional meteorological seasons, others). Would

“regimes” (e.g., Greybush et al. 2008) be better than seasons? If so, how should they be defined?

- Operational systems to generate point probability forecasts for predictands other than precipitation, precipitation amount, convection, severe weather, and snowfall are rare. Thus, applying the piggyback approach will only be feasible for this limited set of high-impact weather phenomena.
- The PoP to RCO logistic regression model must be recomputed whenever MOS sharpness is significantly improved. Most MOS upgrades probably would not change sharpness (i.e., tendency to forecast values of near 0 or 100% instead of intermediate values near the climatology frequency) enough to demand this, but major improvements would.

Does the piggyback approach described here increase the utility of MOS PoP forecasts? We think it does. Linear regression can be used to safely compute the expected value of FAC from the regional average of PoP, although forecast uncertainty is moderately large as it is for MOS PoP itself. Likewise, logistic regression linking RCO to the average and standard deviation of PoP exhibits far more skill than climatology. The intercept term is generally significant though, so just equating RCO to the average PoP will not yield an accurate forecast. Likewise, the standard deviation of PoP is often a significant predictor of RCO, so that should be considered as well.

Implementing this method for other regions and predictands is straightforward. First, obtain a training set of point probability forecasts and verification data for regional occurrence. Then use logistic regression to link the average and standard deviation of the point probabilities to RCO. If enough training data are available, a more sophisticated (and thus data intensive) artificial intelligence scheme such as an artificial neural network or decision tree could replace the logistic regression. While obtaining an archive of point probability forecasts is a nontrivial undertaking, it is much less demanding in terms of storage space and download time than obtaining the NWP output needed to create RCO models from scratch.

Acknowledgments. Thanks to Chad Bahrmann for assistance with data acquisition. Special thanks to two anonymous reviewers who contributed both insight and clarity to this discussion. This work was supported by the National Science Foundation Atmospheric and Geospace Science Grants AGS-1063692.

REFERENCES

- Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211.
- Greybush, S. J., S. E. Haupt, and G. S. Young, 2008: The regime dependence of optimally weighted ensemble model consensus forecasts of surface temperature. *Wea. Forecasting*, **23**, 1146–1161, [CrossRef](#).
- Hanlon, C. J., J. B. Stefik, A. A. Small, J. Verlinde, and G. S. Young, 2013: Statistical decision analysis for flight decision support: The SPaTICus campaign. *J. Geophys. Res.*, **118**, 4679–4688, [CrossRef](#).
- _____, G. S. Young, J. Verlinde, A. A. Small, and S. Bose, 2014: Probabilistic forecasting for isolated thunderstorms using a genetic algorithm: The DC3 campaign. *J. Geophys. Res.*, **119**, 1–10, [CrossRef](#).
- Hirschi, M., C. Spirig, A. P. Weigel, P. Calanca, J. Samietz, and M. W. Rotach, 2012: Monthly weather forecasts in a pest forecasting context: Downscaling, recalibration, and skill improvement. *J. Appl. Meteor. Climatol.*, **51**, 1633–1638, [CrossRef](#).
- Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: Development and applications. Preprints, *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2. [Available online at ams.confex.com/ams/Annual2005/webprogram/Paper83847.html.]
- Rorig, M. L., S. J. McKay, S. A. Ferguson, and P. Werth, 2007: Model-generated predictions of dry thunderstorm potential. *J. Appl. Meteor. Climatol.*, **46**, 605–614, [CrossRef](#).
- Ruth, D. P., B. Glahn, V. Dagastaro, and K. Gilbert, 2009: The performance of MOS in the digital age. *Wea. Forecasting*, **24**, 504–519, [CrossRef](#).
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed., Academic Press, 648 pp.