A Nonsymmetric Logit Model and Grouped Predictand Category Development

BOB GLAHN
NOAA/National Weather Service/Office of Science and Technology/Meteorological Development Laboratory,
Silver Spring, Maryland

(Manuscript received 28 September 2013, in final form 24 April 2014)

ABSTRACT

Logistic regression is an alternative to regression estimation of event probabilities (REEP) and other techniques for estimating weather event probabilities based on NWP output or other predictors. Logistic regression has the advantage over REEP in that the probability estimates are constrained between zero and unity, whereas REEP can “overshoot” these values. It may be a detriment in some applications that the curves developed, one for each of several predictand categories (events), are symmetric. This paper shows how the logit curve can easily be made nonsymmetric as a function of a predictor, and thereby possibly achieve a better fit to the data. As with REEP, the probabilities estimated by logistic regression for each of several categories of a variable may not be consistent. For instance, the probability of snow > 2 in. may exceed the probability of snow > 1 in. Such inconsistencies can be avoided by developing a single equation involving all predictand categories and including another predictor that is a function of the predictand. This effectively, for a single predictor, produces parallel curves separated along the predictor axis but imposes restrictions on the equations and probabilities produced from them. The relationship between the predictor(s) and the predictand must be considered in determining the functional form. With only one predictor, defining the function is relatively straightforward. However, with multiple predictors, the process is more problematic. This paper demonstrates an alternative to imposing a functional form by using binary predictors. This formulation also achieves the goal of producing consistent forecasts and generalizes more readily to multiple predictors.

1. Introduction

Numerical weather prediction (NWP) models have been steadily improving since their operational implementation over 60 years ago. At the time statistical postprocessing of NWP output was introduced over 40 years ago, the NWP models had serious deficiencies, and much effort went into choosing a set of several, typically 10–12, predictors that together captured the information in the NWP model to produce “down to earth” weather forecasts. This included the production of probabilistic guidance forecasts; linear regression applied to forecasting events for that purpose was called regression estimation of event probabilities (REEP) by Bob Miller (1968). REEP was easy to use, and predictor selection, either forward, backward, or stepwise—whatever the developer preferred—from a large number of potential predictors, typically over 100, was straightforward. Such postprocessing was especially useful because NWP did not at first provide forecasts of the elements desired, such as 2-m temperature or 10-m wind, and especially probability forecasts.

But the situation has changed. NWP many times does provide forecasts of the exact “surface” weather elements desired, even though not usually at the specific locations desired. However, conditional biases many times exist in NWP forecasts, biases that can be partially corrected by postprocessing. Today, NWP forecasts are much more accurate, and the 2-m temperature from a model is a very good predictor in some statistical postprocessing techniques for producing 2-m temperature forecasts, so much so that regression equations do not need as many predictors as formerly, and the data samples can be smaller because overfitting is a lesser problem. This makes other techniques that do not have an analytic algorithm for arriving at the “best” solution and that have to rely on more extensive computer processing more attractive.

Probabilistic forecasting has been promoted for years, but it has been slow to catch on. Some reasons for this are given by Glahn (2013). Nevertheless, quite reliable
objective probabilistic forecasts are routinely available from the National Weather Service (NWS) for several surface weather elements over the United States, including precipitation amount (Antolik 2000; Charba and Samplatsky 2009a, 2011), snow amount (Cosgrove and Sfanos 2004), temperature and dewpoint (Glahn et al. 2009), ceiling height and visibility (Ghirardelli and Glahn 2010), thunderstorms (Shaffer and Gilbert 2008; Charba and Samplatsky 2009b), convection (Charba et al. 2011), and precipitation type (Shaffer 2010). Other countries also have well-developed postprocessing systems, some of which produce probabilistic information (e.g., Wilson and Vallée 2002). Other organizations produce probability forecasts that may be made available to the public or specific customers; for instance, the University of Washington is active in providing forecasts for the Pacific Northwest (e.g., Kleiber et al. 2011).

REEP has been used for producing most of the centrally implemented postprocessed probabilistic forecasts (see citations in the previous paragraph). When there are multiple events, defined by applying thresholds to a quasi-continuous variable such as precipitation amount, REEP has two particular deficiencies. First, the forecasts from a REEP equation are not limited to the 0–1 range. The usual procedure is to truncate the forecasts to that interval. This procedure is not pleasing, especially because there may be many forecasts of exactly zero or one, which implies more certainty than warranted. However, the operational impact may not be important because many times the forecasts are rounded to 5% or 10% for distribution. The other deficiency is that the regression lines for the different categories defined by the thresholds may not be parallel when only one predictor is involved, or the multidimensional planes may intersect when two or more predictors are involved. This means that the forecasts can be inconsistent, especially when several predictors are involved. For instance, the probability of snow > 2 in. may be greater than the probability of snow > 1 in. If this happens, then some ad hoc procedure must be invoked to make the forecasts consistent.

The first deficiency of REEP can be dealt with by using logistic regression for which the output is constrained to the zero/one interval. The result is a symmetrical curve with one predictor, or a symmetrical surface in case of multiple predictors. However, the second deficiency still exists—the forecasts can be inconsistent. To deal with this, Wilks (2009, p. 362) proposed an “extension” where instead of a relationship being developed for each predictand category (i.e., event) separately, only one equation be produced “by including the predictand threshold itself as one of the regression predictors.” The resulting curves do not cross, and the forecasts are therefore consistent. The use of this extension has been reported by Roulin and Vannitsem (2012). This paper discusses the extension option, proposes a new alternative, and presents the advantages and disadvantages of each.

Logistic regression produces a symmetrical curve, symmetric on the predictor axis (in the case of a single predictor) about the point where the probability of the dependent variable is 0.5. Especially if the predictor covers a range of values where the curve asymptotes to both zero (from above) or unity (from below), the fit may not be particularly good at one or both ends. For instance, if there are many more data points at one end than the other, the fit may not be as good where the data density is less. This paper shows how symmetry can be avoided by including another predictor in the equation as a function of the original predictor, thereby making the curve nonsymmetrical. This possibility has not, to my knowledge, been discussed previously in the meteorological literature.1

Both the nonsymmetry and the consistency of forecasts from the logit model are explored here by using a dataset for the months of October–March for the years 2011–12 (developmental data) and 2012–13 (test data). The predictand was 12-h amounts of precipitation in inches, the units normally used in the United States for precipitation measurement and forecasting, observed at 24 stations at 0000 UTC in the Pacific Northwest (see the appendix). This area was chosen because of its relatively high frequency of precipitation. The primary predictor was 12-h precipitation amount in millimeters from the National Centers for Environmental Prediction’s (NCEP) Global Forecast System (GFS) at the 24-h projection verifying at the same time as the observations. Some use was also made of the model-produced relative humidity at 850 hPa as a predictor.

In addition, a larger dataset is employed to more fully demonstrate the utility of the nonsymmetric feature. This dataset is being used within the Meteorological Development Laboratory to create a conditional probabilistic snowfall product. The forecasts will be the probability of 24-h categorical amounts, conditional on snowfall occurring within the 24-h period. This dataset consists of eight seasons of data, 1 September 2002–31 May 2009, for 5021 locations within the conterminous United States. The predictand is the observed snowfall amount at each observation point, and the single

1 According to Agresti (2002), the term “logit” was introduced by Berkson (1944), although the logistic transformation was used earlier by Bartlett (1937). Walker and Duncan (1967, p. 169) note that “restriction to a symmetric curve is not essential” but do not use the approach proposed here.
predictor is a variable derived from the output of the GFS model. This predictor was designed especially for predicting snowfall amount, and has been described by Cosgrove and Sfanos (2004).

2. Nonsymmetric logit regression

The logistic equation is simply

\[ P = \frac{\exp[f(x)]}{1 + \exp[f(x)]}, \]

where \( f(x) \) is a function, usually linear, of the predictor variables. When only one predictor is involved, call it “\( q \)” for the cube root of model quantitative precipitation forecast (qpf), the function is

\[ f(x) = a + bq. \] (1)  

The cube root transformation has been found to be a better predictor of the observed precipitation than the original variable (Sloughter et al. 2007; Bentzien and Friederichs 2012), not differing much from the fourth root used by Hamill et al. (2004).\(^2\) The constant and coefficient are usually arrived at by the iterative method of maximum likelihood (Wilks 2011, 238–242), with an analytic solution not being possible.

A logit equation was developed for each of the six categories of precipitation amount: >trace (i.e., measurable), \( \geq 0.10, \geq 0.25, \geq 0.50, \geq 0.75, \) and \( \geq 1.0 \) in.; these are here called categories 1–6, respectively. The probabilities for each of these categories produced by the six equations are shown in Fig. 1 as a function of \( q \) (solid lines).

The symmetry along the predictor axis for each curve is about the point \( q^* \), where \( P = 0.5 \). The curves can be made nonsymmetric about this point by including a second predictor \( q' \), where

\[ q' = q - q^* \]

does not produce the curves also shown in Fig. 1 (dotted lines). The addition of a predictor defined in this way is called piecewise linear regression in a linear regression setting (Neter and Wasserman 1974, p. 313), and provides for a different slope on different sides of the break point \( q^* \).

\(^2\)The use of the cube root transformation for normalization purposes goes at least as far back as 1953 (Stidd 1953).
meaningful value; to make the logit nonsymmetrical, the break point can be anywhere the user wants to place it, but it would likely be at or near the 50% point when the curves nearly span the probability range, 0–1, as they do in Fig. 1. If the 50% point is used, then it could be determined by deriving a one-predictor equation and noting the 50% point, which in terms of Eq. (1) is equal to \(-b/a\).

Figure 1 also gives the distribution of relative frequency of forecast precipitation amount for bins of \(q\) (solid black line). There is a large spike at zero and then a broad maximum at values between 0.6 and 1.8 (corresponding to between about 0.2 and 6.0 mm = about 0.01–0.23 in.). There is a relative frequency of precipitation \(>\) trace of 0.045 for the first bin of model precipitation that consists of zero amount, each bin representing 0.073 of the transformed forecast precipitation. For this sample of data, the symmetric logit curves do not cross in any detrimental way for the predictand categories defined within the range of the data in the sample; that is, where the crossing occurs, it is so close to zero or unity that it has no practical consequence. If they were to cross, then the forecasts from them would be inconsistent (e.g., the probability of \(>0.25\) in. of precipitation being greater than the probability \(>0.10\) in.).

The nonsymmetrical curves do not deviate a great deal from the corresponding symmetrical ones for the first four predictands, the largest departure being for the \(\geq 0.1\) in. category, where the probabilities deviate as much as 5% in the midprobability range. However, the fifth category shows a departure for the larger predictor values, and the last curve actually turns down and would asymptote to zero rather than unity. This latter result is due to the small number of cases at high values of the predictor, and the abnormal situation where more model forecast precipitation is positively related to lower observed amounts. This shows the care that would have to be taken in using a second predictor to produce nonsymmetry. It also demonstrates the potential that might be helpful in situations where the relationship between predictor and predictand is U shaped.

For precipitation \(>\) trace, there is no obvious need for the curve to be nonsymmetrical, and the two curves nearly coincide (see Fig. 1). For the amount \(\geq 0.1\) in., the nonsymmetric curve fits a bit better in the midrange and also at both the upper and lower ends. For amounts \(>0.75\) in., the relative frequencies associated with the higher bins become rather erratic on this quite small sample of data, and show decreasing relative frequencies with increasing model qpf. For amounts \(\geq 0.75\) and \(>1.0\) in., the second predictor was a detriment, and it produced unexpected and unrealistic results. The data, as summarized in the observed relative frequencies of \(\geq 1.0\) in., do show a downward trend for higher model forecast amounts (solid squares in Fig. 1). For the symmetric case, the low quantitative precipitation forecasts and corresponding low observed amounts largely determine the curve at higher amounts, whereas the nonsymmetric curves tend to fit the higher amounts. While the downturn of the nonsymmetric curve for category 6 is not realistic, it shows the potential for dealing with a nonmonotonic relationship between a predictand and predictor.

The Brier scores [one-half the score \(P\) defined by Brier (1950)] for the symmetrical and nonsymmetrical curves differ little, differing usually in the fourth decimal place. It is noted that maximizing according to the log likelihood does not necessarily give the best Brier score, but the differences are trivial. Although the differences in Brier scores between the symmetric and nonsymmetric curves are very small, the nonsymmetric fit gave lower (better) scores for the amounts \(\geq 0.10\) and \(\geq 0.25\) on the 2012–13 season of independent data.

The relative humidity at 850 hPa was added as a second predictor to qpf for the symmetric curve, and then also as a third predictor after qpf and the nonsymmetric predictor for the nonsymmetric case. The additions in reduction of variance (and decreases in Brier score) were small, and of the same order as adding the second nonsymmetric predictor to qpf. For the second and third categories, the nonsymmetric predictor was more important than adding relative humidity to qpf. In other words, it was more important to make the curve nonsymmetric than to add the second predictor that was individually well related to observed precipitation.

Figure 2 shows the reliability on the independent sample of both the symmetric and nonsymmetric curves for the category \(\geq 0.10\) in., the category where the symmetric and nonsymmetric solutions were the most different (see Fig. 1). (Categories 5 and 6 corresponding to the larger amounts were not considered for verification because they were obviously too dependent on the small

---

3 Specialized piecewise linear development has been applied in fitting climate temperature trends where the line is constant up to a point and then angles upward (for increasing temperatures with time). Livezy et al. (2007) call this the “hinge” model, the hinge being specified at the point in time where many climate records indicate an increase in temperature to be starting.

4 It is noted that developers at the Meteorological Development Laboratory have characteristically required at least two seasons of data for development; however, for this demonstration, only one season was used.
number of cases.) Reliability is somewhat better for the nonsymmetric fit in the areas where there was a difference between the curves, and in the direction expected; the symmetrical curve forecast too high probabilities in the upper range and too low at $P = 0.5$, whereas the nonsymmetrical curve did not exhibit those characteristics. Although categories 2 and 3 showed a somewhat better nonsymmetric fit to the data, the resulting reductions of variance were very little better than for the symmetric curves; even so, the nonsymmetrical curves were slightly better in terms of reduction of error$^5$ on the independent test sample.

This sample of data did not overall demonstrate a clear advantage of using a nonsymmetrical curve, and was meant only to demonstrate the method of producing one. In fact, for the two higher categories, the results were clearly not meaningful because of the rarity of events above the $P = 50\%$ threshold of the symmetrical curve (the break point for the nonsymmetric predictor used here). For category 2 where there was the most difference between the symmetric and nonsymmetric solutions, the latter was slightly more accurate and reliable on the independent data.

The second sample of data shows a definite improvement of the nonsymmetrical curve over the symmetrical one. Figure 3 shows logit curves for snowfall amounts $\geq 1$ and $\geq 2$ in. for the 24-h period 18–42 h after run time of 0000 UTC. For both amounts, the symmetrical case overpredicts at low values of the predictor and also at the high values, whereas the nonsymmetric case shows a much better fit. The break points $q^*$ were chosen at the maximum underpredicting in the symmetrical case. There can be no doubt this is real; 8 years of data for 5021 locations were included, and the sample size was 395 285. Forecasts produced from the nonsymmetric curve would be highly reliable, while those produced from the symmetric curve would not be.

It is not suggested that a nonsymmetric fit is always better, but it may be in some cases. It is also not

---

$^5$ Reduction of error is a term introduced by Lorenz (1959, p. 84), and it is the same as the reduction of variance except the mean square error is around the dependent sample mean rather than around the test sample mean. This compares the error of the forecast with the mean forecast that could have been made for the test sample.
suggested that there are not other fitting methods that may do well. However, the nonsymmetric feature is very easy to use, and as was indicated in the qpf case where it produced a better fit for category 2, the additional predictability was greater than adding another predictor that was by itself a good predictor.

3. Consistency of probability forecasts for different categories

The crossing of lines, in the linear case, or of curves in the logistic case for different predictand categories, can be a problem, and Wilks (2009, 2011) has proposed an extension to the logit as a solution. It consists of developing only one equation, essentially repeating the sample for each predictand category but including another predictor that represents group membership. This has a twofold advantage. First, if the predictor value representing group membership is some analytic function of the predictand, then the single equation can be applied to yield the probability of any desired event defined by thresholding the quasi-continuous predictand. Second, the forecasts will be consistent, as the curves will not cross. Also, there would be fewer coefficients to fit, although the degrees of freedom saved may be consumed in defining a workable metric for group membership. The regression equation

$$f(x) = a + b_g x + c_g g$$  \hspace{1cm} (3)

is of exactly the same form as Eq. (2), except the second predictor “g” is the separation function (a specific value for each predictand category). I call it a separation function (or constant for each category) because it just separates the curves along the predictor axis. Because \(c_g g\) in Eq. (3) is a constant for a particular value of the predictand, after development, it can be added to the overall constant \(a\) to give a unique equation for each predictand category.

Wilks (2009) calls this “simultaneous” development. I use the term grouped development, because the term simultaneous development was coined by the Techniques Development Laboratory years ago and has been used repeatedly (e.g., Dallavalle et al. 1980, pp. 2, 5; Palko et al. 1985, p. 2; Wilks 2011, p. 266) to describe the situation when there is more than one predictand and the same predictors are chosen for all the equations. For instance, if one were screening (forward selection) predictors and several predictands were involved, then the process would be simultaneous but not necessarily grouped.

When the predictand is the same variable as the predictor, quantitative precipitation in this experiment, defining the separation constants is trivial—just use the same values as the predictand category definitions for
producing the events for the equation. However, any transformations need to be considered. For instance, when using the cube root of precipitation amount in millimeters as the predictor, the best separation constants may be in those terms also. Actually, any linear transformation of the separation constants will give the same result; that is, values of $g$ could be multiplied by and/or added to a constant and the same solution would result. The constant and coefficient would be different to compensate for the scaling of $g$.

However, when the predictand and predictor are different meteorological variables, the choice of separation constants is not obvious, and in actuality the relationship between predictor and predictand must be considered because the separation constants just shift the curves along the predictor axis. The separation constants are in effect providing group membership, and by giving them specific values, a metric is imposed [Neter and Wasserman (1974, p. 319) discuss this in a linear regression framework] that might not be reasonable. For good results, the values of the constants are necessarily related to the predictor, although they indicate predictand group membership. These are essentially values that define group membership in terms of the predictor. Actually, it is the differences in value of the separation constants that are important. If there are only two categories, the values do not matter; when the separation term $c_{gg}$ is merged into the constant $a_g$, the same two equations will result no matter the values of $g$.

Figure 4 shows the symmetric logit curves fit individually and also those fit by grouping the categories. The separation constants for grouped development are values of the predictor, scaled in the cube root of precipitation amount in millimeters, at which the predictor produces a 50% probability when symmetric curves are fit individually. Specifically, the six values are 1.119, 1.822, 2.309, 2.789, 3.108, and 3.348. These values are probably about the optimum that can be found. With this imposed metric, the curves fit separately and by grouping broadly agree, separating somewhat at the higher categories where the individual fitting may not be warranted. This is the easy situation, in which the single predictor and predictand are the same variable, although as stated earlier, the separation constants should be a linear function in the predictor units. If one converts the predictand thresholds to the cube root in millimeters, the values are 0.633, 1.364, 1.852, 2.333, 2.671, and 2.939. Note that the differences between consecutive
values in these two sets of separation constants are very similar. Curves from these separation constants are nearly identical (not shown) to those produced from the 50% values and shown by dotted curves in Fig. 4.

It can be seen that the curves developed by grouping are moved along the predictor axis, and the separation of the curves is that imposed by the metric. For instance, the separation between curves for categories 1 and 2 is larger than between the curves for categories 5 and 6 because $1.822 - 1.119 = 0.703$ is larger than $3.348 - 3.108 = 0.240$.

Figure 5 is similar to Fig. 4 in that it shows the curves for individual and grouped development but with nonsymmetric curves. The nonsymmetric nature is much less noticeable with grouped development. Of importance is that for the grouped development, the curves for the fifth and sixth categories are now meteorologically reasonable, even though the numbers of cases are low and the data are not fit well at the high end (see Fig. 1 and discussion).

Figure 6 shows the symmetric curves developed with separation constants derived from the single development 50% points as solid lines (the same solid curves as shown in Fig. 4) and separation constants equal to the category definitions in inches as dotted lines. Specifically, the latter separation constants were 0.01, 0.10, 0.25, 0.50, 0.75, and 1.0. Note that these separation constants were chosen in the same way as those discussed above, except that the conversion to millimeters and the cube root transformation was not made. The solid lines fit the data rather well, but the dotted lines do not. This demonstrates that the separation constants must be chosen with considerable care.

4. An alternative for grouped development

Rather than specifying separation constants, a dummy (binary) variable can be provided for each category of the predictand except one, which takes the value of 1 for group membership and 0 otherwise. Including this predictor for every category would render the cross product matrix singular. This does require fitting more coefficients and does not have the advantage of being able to obtain a probability for any value of the predictand, but it imposes no restrictions on the separation of the curves along the predictor axis. It also needs no adjustment if more than one predictor is used. When a metric is imposed in grouped development, the predictor–predictand relationship needs to be considered. After selection of the first predictor, whether a second one is useful will depend on the metric, which may not be good for that second predictor.
Initially, for this alternative and this sample of data, the equation contains a constant, a coefficient for the cube root qpf predictor, and five coefficients for the five binaries. Because each of the binary terms is a constant for its predictand category, its coefficient can be combined with the overall constant to form a new constant for the corresponding predictand category, leaving a two-term equation for each category, as can be done for the grouped development discussed in the previous section.

Figure 7 compares the curves for the six categories when the 50% thresholds were used as separation constants with those developed by using binaries for category membership. It can be seen that there is very little difference, except for the $\geq 1\text{-in.}$ category. The use of binary predictors does not specify the amount of separation of the curves and also maintains the same slope of each equation to guarantee consistency among categories.

This formulation does not require any knowledge of the relationship between the predictor, in this case the cube root of model qpf, and the predictand. The equation developed with 50% separation constants had three constants developed: one for the overall constant, one for the qpf predictor, and one each for the five binaries; the reduction of variance was 0.5804. The equation developed with category membership designated by a binary predictor had seven constants developed: one for the overall constant, one for the qpf predictor, and one each for the five binaries; the reduction of variance was 0.5805, different only in the noise level. Single-category development requires 12 constants be developed, two for each of six categories.

For this sample of data, the curves were extremely similar, the only strong departure was for category 6 for which the number of cases was small. As indicated earlier, if the separation constants for the predictand are well related to the predictor, then little difference is to be expected. However, this would not always be the case and additional predictors might render the constants nonoptimal.

5. Summary and conclusions

I have used a cool season of model forecasts and matching observations of 12-h precipitation amounts to demonstrate two aspects of logistic regression. First, I have shown how the symmetric logit curve can be made nonsymmetric, and to possibly fit the data better. There is no a priori reason the curve should be symmetric. I developed logistic relationships between six categories of observed precipitation amount and the cube root of the model precipitation amount. Two of these categories showed a somewhat better nonsymmetric fit to the data, although the resulting reductions of variance were very little better than for the symmetric curves; even so, the
nonsymmetric curves were slightly better in terms of reduction of error on the independent test sample and the nonsymmetric curve was more reliable for category 2 (Fig. 2). The nonsymmetric curves for the two larger amount categories had unexpected behavior, a result of too few cases of the larger amounts to fit that portion of the curve acceptably. Even so, the curve for the larger amount shows the potential for the nonsymmetric logit when the predictand is not a continuously increasing or decreasing function of the predictor(s).

While this sample demonstrated the method and showed some promise, a second much larger sample showed clearly that a nonsymmetric curve can be an advantage (see Fig. 3).

Second, I have shown that the “extension” that Wilks proposed imposes a metric on the predictand, and one has to have, or assume, some knowledge of the relationship between the predictor and predictand to craft a good metric. If the predictor and predictand are of the same variable (e.g., precipitation amount), then the metric is straightforward, although any transformations have to be considered. For instance, if the predictor is the cube root of precipitation amount in millimeters, then the predictand category membership should be defined in a similar way. This extension places the curves along the predictor axis relative to each other, and the metric is essentially a series of “separation constants” for the curves. If the relationship between the predictand and predictor is not known well, then “50%” separation constants can be found by developing a separate equation for each category, and noting the 50% points—the place along the predictor axis where the curve crosses $P = 50\%$. These values are possibly the best that can be obtained to define group membership in the extended mode. However, this does not necessarily define a “function” that could be used to evaluate the expanded equation for any specific predictand category, except those used in the development—one of the two reasons for the grouped development.

An alternate option for using the grouped development method is to define a binary predictor for each predictand category except one to denote group membership. This alternative requires no knowledge of the predictor–predictand relationship and imposes no arbitrary metric on group membership. Like the 50% option, this does not lead to an equation that can be used for any specific category. However, the curves for the categories are consistent, and by defining several categories the specific values defined may be sufficient for an operational product; any finer definition could be adequately handled by interpolation.
Reductions of variance on the developmental data and reductions of error on the year of test data are so close for the different formulations that no practical difference could be discerned. The grouped development has an advantage over individual development on each category, not only because the resulting forecasts are consistent, but because each more rare category has a relationship, built on all the data in the sample, that is reasonable. It is likely the binary option generalizes to more predictors better than the imposition of a metric that to some degree needs to be devised with the specific predictor(s) and predictand in mind.

Acknowledgments. I am grateful for the assistance of David Rudack and Phillip Shafer, who provided me the data for this study; and to Tamarah Curtis for expert advice in preparing the figures and text. I also thank Jerome Charba, David Rudack, Jung-Sun Im, and three anonymous reviewers for their valuable comments. The views expressed here are those of the author and do not necessarily represent those of the NWS or of any other governmental agency.

APPENDIX

The 24 Stations Used in the Study

<table>
<thead>
<tr>
<th>Station identifier</th>
<th>Name</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>KAST</td>
<td>Astoria, OR</td>
<td></td>
</tr>
<tr>
<td>KAWO</td>
<td>Arlington Airport, WA</td>
<td></td>
</tr>
<tr>
<td>KBF1</td>
<td>Boeing Field, WA</td>
<td></td>
</tr>
<tr>
<td>KBLI</td>
<td>Bellingham, WA</td>
<td></td>
</tr>
<tr>
<td>KCLM</td>
<td>Fort Angeles, WA</td>
<td></td>
</tr>
<tr>
<td>KEUG</td>
<td>Eugene, OR</td>
<td></td>
</tr>
<tr>
<td>KFHR</td>
<td>Friday Harbor, WA</td>
<td></td>
</tr>
<tr>
<td>KHIO</td>
<td>Hillsboro/Portland, OR</td>
<td></td>
</tr>
<tr>
<td>KHQM</td>
<td>Hoquiam, WA</td>
<td></td>
</tr>
<tr>
<td>KMMV</td>
<td>McMinnville, OR</td>
<td></td>
</tr>
<tr>
<td>KOLM</td>
<td>Olympia, WA</td>
<td></td>
</tr>
<tr>
<td>KPAE</td>
<td>Everett, WA</td>
<td></td>
</tr>
<tr>
<td>KPDX</td>
<td>Portland, OR</td>
<td></td>
</tr>
<tr>
<td>KRNT</td>
<td>Renton, WA</td>
<td></td>
</tr>
<tr>
<td>KSEA</td>
<td>Seattle–Tacoma, WA</td>
<td></td>
</tr>
<tr>
<td>KSHN</td>
<td>Shelton, WA</td>
<td></td>
</tr>
<tr>
<td>KSLE</td>
<td>Salem, OR</td>
<td></td>
</tr>
<tr>
<td>KSPB</td>
<td>Scappoose, OR</td>
<td></td>
</tr>
<tr>
<td>KTCM</td>
<td>McChord Air Force Base, WA</td>
<td></td>
</tr>
<tr>
<td>KTW</td>
<td>Tacoma, WA</td>
<td></td>
</tr>
<tr>
<td>KTTD</td>
<td>Portland/Trountal, OR</td>
<td></td>
</tr>
<tr>
<td>KUAO</td>
<td>Aurora State, OR</td>
<td></td>
</tr>
<tr>
<td>KUIL</td>
<td>Quillayute, WA</td>
<td></td>
</tr>
<tr>
<td>KVUO</td>
<td>Vancouver, WA</td>
<td></td>
</tr>
</tbody>
</table>

REFERENCES


