

A NON-SYMMETRIC LOGIT MODEL
AND GROUPED PREDICTAND CATEGORY DEVELOPMENT

Bob Glahn*
Meteorological Development Laboratory
Office of Science and Technology
National Weather Service, NOAA
Silver Spring, Maryland

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 to 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 given the moniker REEP for Regression Estimation of Event Probabilities 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 technique for producing 2-m temperature forecasts, so much so that regression equations don't need as many predictors as formerly, and the data samples can be smaller because overfitting is a lesser problem. This makes other techniques that don't 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 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 Vallee 2003). 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 references 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 zero-one 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 inches may be greater than the probability of snow > 1 inch. If this happens, 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) 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

* *Corresponding author address:* Bob Glahn, Meteorological Development Laboratory, National Weather Service Scientist Emeritus, 1325 East West Highway, Silver Spring, MD 20910; email: harry.glahn@noaa.gov

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 non-symmetrical¹. This possibility has not, to my knowledge, been discussed previously in the meteorological literature.

Both the non-symmetry and the consistency of forecasts from the logit model are explored here by using a dataset for the months of October through March for the years 2011-2012 (developmental data) and 2012-2013 (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 appendix). This area was chosen because of its relatively high frequency of precipitation. The primary predictor was 12-h precipitation amount in mm 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.

2. NON-SYMMETRIC LOGIT REGRESSION

The logistic equation is simply

$$P = \frac{\text{Exp}[f(x)]}{1 + \text{Exp}[f(x)]}$$

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

$$f(x) = a + bq. \quad (1)$$

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) note that "Restriction to a symmetric curve is not essential." but do not use the approach proposed here.

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, pp. 238-242), an analytic solution not being possible.

A logit equation was developed for each of six categories of precipitation amount, > Trace (i.e., measurable), ≥ 0.10 , ≥ 0.25 , ≥ 0.50 , ≥ 0.75 , and ≥ 1.0 inches; these are here called categories 1 through 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 non-symmetric about this point by including a second predictor q' where

$$q' = q - q^*$$

when $q > q^*$, and zero otherwise. The resulting equation

$$f(x) = a_{ns} + b_{ns}q + c_{ns}q' \quad (2)$$

produces the curves also shown in Fig. 1 (dotted lines). The addition of an additional 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 breakpoint q^* . For piecewise linear, the breakpoint would be at some meaningful value;³ to make the logit non-symmetrical, the breakpoint can be anywhere the user wants to place it, but would likely be at or near the 50% point. If the 50% point is used, it could be determined by deriving a one-predictor equation and noting the 50% point, which in terms of Eq. (1) is equal to $-b/a$.

Fig. 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, 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 to 0.23 inches). There is a relative frequency of precipitation > Trace of 0.045 for the first bin of model precipitation which consists of zero amount, each bin

² The use of the cube root transformation for normalization purposes goes at least as far back as 1953 (Stidd 1953).

³ Specialized piecewise linear development has been applied in fitting climate temperature trends where the line is constant up to a point, then angles upward (for increasing temperatures with time). Livezey 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.

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, the forecasts from them would be inconsistent (e.g., the probability of 0.25 inches of precipitation being greater than the probability 0.10 inches).

The non-symmetrical curves do not deviate a great deal from the corresponding symmetrical ones for the first four predictands, the largest departure being for the ≥ 0.1 inch category, where the probabilities deviate as much as 5% in the mid probability 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 non-symmetry. 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 non-symmetrical, and the two curves nearly coincide (see Fig. 1). For the amount ≥ 0.1 inches, the non-symmetric 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,⁴ and show decreasing relative frequencies with increasing model qpf. For amounts ≥ 0.75 and ≥ 1.0 inches (solid squares in Fig. 1), the second predictor is likely a detriment, and it produced unexpected and unrealistic results. The data, as summarized in the observed relative frequencies of ≥ 1.0 in, do show a downward trend for higher model forecast amounts. For the symmetric case, the low quantitative precipitation forecasts and corresponding low observed amounts largely determine the curve at higher amounts, whereas the non-symmetric curves tend to fit the higher amounts.

The Brier scores [one half the Score P defined by Brier(1950)] for the symmetrical and non-symmetrical 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 non-symmetric curves are very small, the non-symmetric

fit gave lower (better) scores for the amounts ≥ 0.10 and ≥ 0.25 on the 2012-2013 season of independent data.

The relative humidity at 850 hPa was added as a second predictor to qpf for the symmetric curve, then also as a third predictor after qpf and the non-symmetric predictor for the non-symmetric case. The additions in reduction of variance (and decreases in Brier score) were small, and of the same order as adding the second non-symmetric predictor to qpf. For the second and third categories, the non-symmetric predictor was more important than adding relative humidity to qpf. In other words, it was more important to make the curve non-symmetric 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 non-symmetric curves for the category ≥ 0.10 in, the category where the symmetric and non-symmetric 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 number of cases.) Reliability is somewhat better for the non-symmetric 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 non-symmetrical curve did not exhibit those characteristics.

This sample of data did not overall demonstrate a clear advantage of using a non-symmetrical 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 breakpoint for the non-symmetric predictor used here). For category 2 where there was the most difference between the symmetric and non-symmetric solutions, the latter was slightly more accurate and reliable on the independent data.

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 two-fold 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

⁴ It is noted that developers at the Meteorological Development Laboratory have characteristically required at least two seasons of data for development, but for this demonstration, only one season was used.

consumed in defining a workable metric for group membership. The regression equation,

$$f(x) = a_g + b_g q + c_g g \quad (3)$$

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_g 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, 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, the units 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 term 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 don’t matter; when the separation term $c_g g$ is merged into the constant a_g the same two equations will result no matter the values of g .

Figure 3 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 mm, 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 mm, 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. 3.

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 one and two is larger than between the curves for categories five and six because $1.822 - 1.119 = 0.703$ is larger than $3.348 - 3.108 = 0.240$.

Figure 4 is similar to Fig. 3 in that it shows the curves for individual and grouped development, but with non-symmetric curves. The non-symmetric 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 5 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. 3) 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 were 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 that takes the value of 1 for group membership and 0 otherwise (Neter and Wasserman 1974, p. 309). 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 or not 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 6 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 ≥ 1 inch 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 for the coded predictand membership; 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, little difference is to be expected. However, this would not always be the case and additional predictors might render the constants non-optimal.

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 non-symmetric, 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 non-symmetric fit to the data, although the resulting reductions of variance were very little better than for the symmetric curves; even so, the non-symmetric curves were slightly better in terms of reduction of error⁵ on the independent test sample and the non-symmetric curve was more reliable for category 2 (Fig. 2). The non-symmetric curves for the two larger amount categories had unexpected behavior, a result of two 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 non-symmetric logit when the predictand is not a continuously increasing or decreasing function of the predictor(s).

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 mm, 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 percent" 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

⁵ Reduction of error is a term introduced by Lorenz (1959, p 84), and 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.

membership. This alternative requires no knowledge of 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 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, and Jung-Sun Im for 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

Station Identifier	Name
KAST	Astoria, Ore.
KAWO	Arlington Airport, Wash.
KBFI	Boeing Field, Wash.
KBLI	Bellingham, Wash.
KCLM	Fort Angeles, Wash.
KEUG	Eugene, Ore.
KFHR	Friday Harbor, Wash.
KHIO	Hillsboro/Portland, Ore.
KHQM	Hoquiam, Wash.
KMMV	McMinnville, Ore.
KOLM	Olympia, Wash.
KPAE	Everett, Wash.
KPDX	Portland, Ore.
KRNT	Renton, Wash.
KSEA	Seattle-Tacoma, Wash.
KSHN	Shelton, Wash.
KSLE	Salem, Ore.
KSPB	Scappoose, Ore.
KTCM	McChord AFB, Wash.
KTIW	Tacoma, Wash.
KTTD	Portland/Troutdal, Ore.
KUAO	Aurora State, Ore.
KUIL	Quillayute, Wash.
KVUO	Vancouver, Wash.

REFERENCES

- Agresti, A., 2002: *Categorical Data Analysis*, 2nd ed. Wiley.
- Antolik, M.S., 2000: An overview of the National Weather Service's centralized statistical quantitative precipitation forecasts. *J. Hydrology*, **239**, 306-337.
- Bartlett, M. S., 1937: Some examples of statistical methods of research in agriculture and applied biology. *J. Royal Stat. Soc. (supplement)*, **4**(2), 137-183.
- Bentzien, S., and P. Friederichs, 2012: Generating and calibrating probabilistic quantitative precipitation forecasts from the high-resolution NWP model COSMO-DE. *Wea. Forecasting*, **27**, 988-1002.
- Berkson, J., 1944: Application of the logistic function to bio-assay. *J. Amer. Statistical Assn.*, **39**, 357-365.
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1-3.
- Charba, J. P., and F. G. Samplatsky, 2009a: Hi-res gridded MOS 6-h QPF guidance. Preprints, 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction, Omaha, NE, American Meteorological Society, **17B2**.
- _____, and _____, 2009b: Operational 2-h thunderstorm guidance forecasts to 24 hours on a 20-km grid. Preprints, 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction, Omaha, NE, American Meteorological Society, **15B5**.
- _____, and _____, 2011: High-resolution GFS-based MOS quantitative precipitation forecasts on a 4-km grid. *Mon. Wea. Rev.*, **139**, 39-68.
- _____, _____, and P. E. Shaffer, 2011: Experimental LAMP 2-h convection guidance forecasts on a 20-km grid. Preprints, 24th Conference on Weather Analysis and Forecasting/20th Conference on Numerical Weather Prediction, Seattle, WA, American Meteorological Society, **J19.3**.
- Cosgrove, R. L., and B. Sfanos, 2004: Producing MOS snowfall amounts from cooperative observer reports. Preprints, 20th Conf. On Weather Analysis and Forecasting, Seattle, WA, American Meteorological Society, **6.3**.
- Dallavalle, J. P., J. S. Jensenius, and W. H. Klein, 1980: Improved surface temperature guidance from the limited-area fine mesh model. Preprints, Eighth Conference on Weather Forecasting and Analysis, Denver, CO, American Meteorological Society, 2-8.

- Ghirardelli, J. E., and B. Glahn, 2010: The Meteorological Development Laboratory's Aviation Weather Prediction System. *Wea. Forecasting*, **25**, 1027-1051.
- Glahn, B., 2013: Statistics in hydrometeorological forecasting—contributions and challenges. Preprints, *Symposium on the Role of Statistical Methods in Weather and Climate Prediction*, Austin, TX, American Meteorological Society, **1.1**.
- _____, M. Peroutka, J. Weidenfeld, J. Wagner, G. Zylstra, and B. Schuknecht, 2009: MOS uncertainty estimates in an ensemble framework. *Mon. Wea. Rev.*, **137**, 246-268.
- Hamill, T. M., J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, **132**, 1434-1447.
- Kleiber, W., A. E. Raftery, J. Baars, T. Gneiting, C. F. Mass, and E. Gritti, 2011: Locally calibrated probabilistic temperature forecasting using gostatistical model averaging and local Bayesian model averaging. *Mon. Wea. Rev.*, **139**, 2630-2649.
- Livezey, R. E., K. Y. Vinnikov, M. M. Timofeyeva, R. Tinker, and H. M. van den Dool, 2007: Estimation and extrapolation of climate normals and climatic trends. *J. Appl. Meteor. Climatol.*, **46**, 1759-1776.
- Lorenz, E. N., 1959: Prospects for statistical forecasting. Final Report, Statistical Forecasting project, Contract No. AF 19 (604)-1566, Massachusetts Institute of Technology, Department of Meteorology, 185 pp.
- Miller, R. G., 1968: Regression estimation of event probabilities *Tech Rep. No. 1*, U.S. Weather Bureau Contract Cwb-10704, The Travelers Research center, Inc., Hartford, CN, 153 pp.
- Neter, J., and W. Wasserman, 1974: *Applied Linear Statistical Models*. Richard D. Irwin, Inc., 842 pp.
- Palko, J. M., J. S. Jensenius, Jr., and J. P. Dallavalle, 1985: Development of medium range forecast equations for maximum/minimum temperature and probability of precipitation. *TDL Office Note* **85-7**, National Weather Service, NOAA, U.S. Department of Commerce, 14 pp.
- Roulin, E., and S. Vannitsem, 2012: Postprocessing of ensemble precipitation predictions with extended logistic regression based on hindcasts. *Mon. Wea. Rev.*, **140**, 874-888; Corrigendum, p. 2018.
- Shaffer, P. E., 2010: Logit transforms in forecasting precipitation type. Preprints, *20th Conference on Probability and Statistics in the Atmospheric Sciences*, Atlanta, GA, American Meteorological Society, **P222**.
- _____, and K. Gilbert, 2008: Developing GFS-based MOS thunderstorm guidance for Alaska. Preprints, *Third Conference on Meteorological Applications of Lightning Data*, New Orleans, LA, American Meteorological Society, **P2.9**.
- Sloughter, J. M., A. E. Raftery, T. Gneiting, and C. Fraley, 2007: Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Wea. Rev.*, **135**, 3209-3220.
- Stidd, C. K., 1953: Cube-root-normal precipitation distributions. *Transactions of the Amer. Geophysical Union*, **34**, No. 1, National Research Council of the National Academy of Sciences, Washington, D. C., 31-35.
- Walker, S. H., and D. B. Duncan, 1967: Estimation of the probability of an event as a function of several independent variables. *Biometrika*, **54**, 167-179
- Wilks, D. S., 2009: Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteor. Applications*, **16**, 361-368.
- _____, 2011: *Statistical Methods in the Atmospheric Sciences*, Third Edition. Academic Press/ Elsevier. 676 pp.
- Wilson, L. J., and M. Vallee, 2002: The Canadian Updateable Model Output Statistics (UMOS) system: Design and development tests. *Wea. Forecasting*, **17**, 206-222.

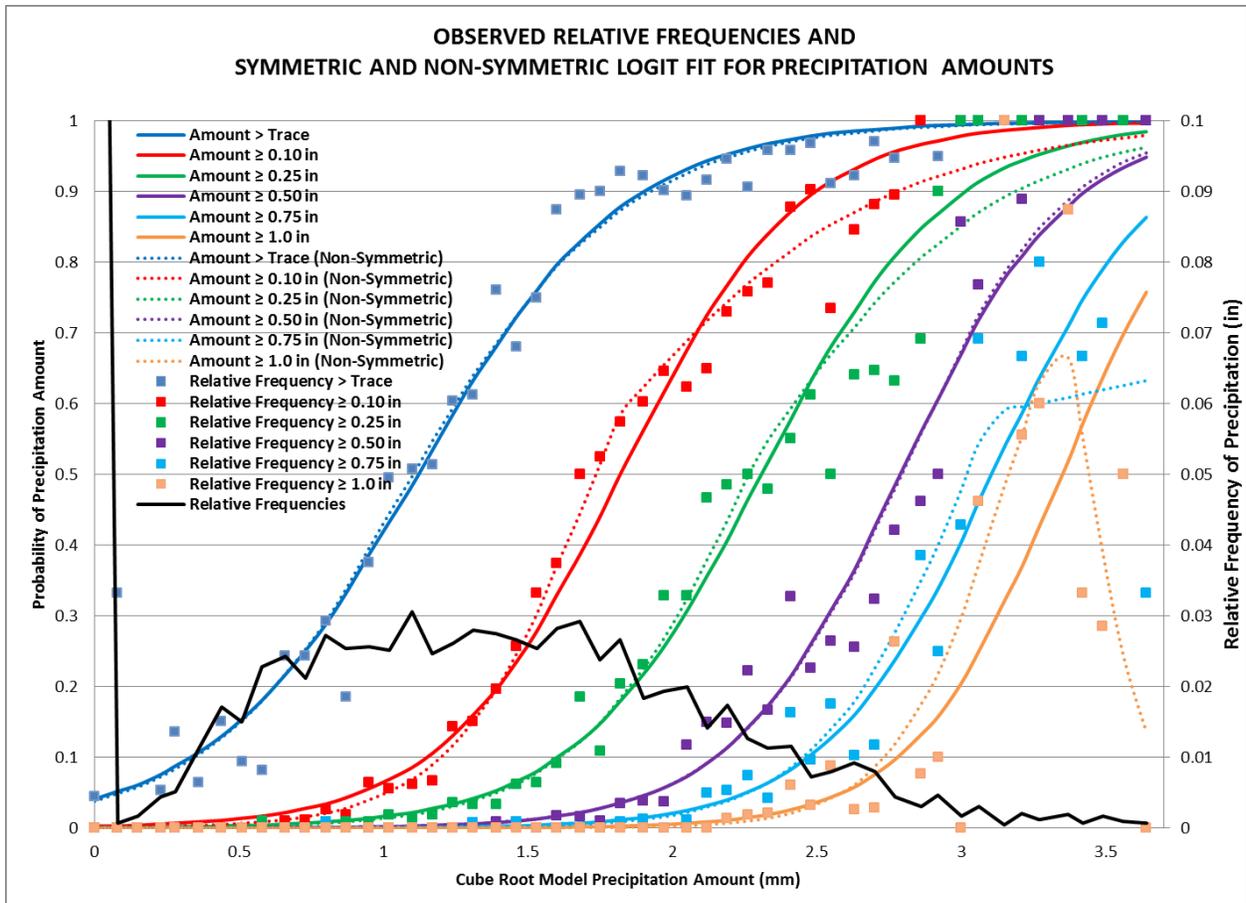


Figure 1. Observed relative frequencies (solid squares) for each of the six categories of qpf in bins 0.073 wide along the predictor axis, the cube root of model qpf in mm. The sample data range from 0 through 3.64 in those units. The solid curves are the symmetric fits and the dotted curves are the non-symmetric fits. The black line is the forecast relative frequency of precipitation (right axis) in the same bins.

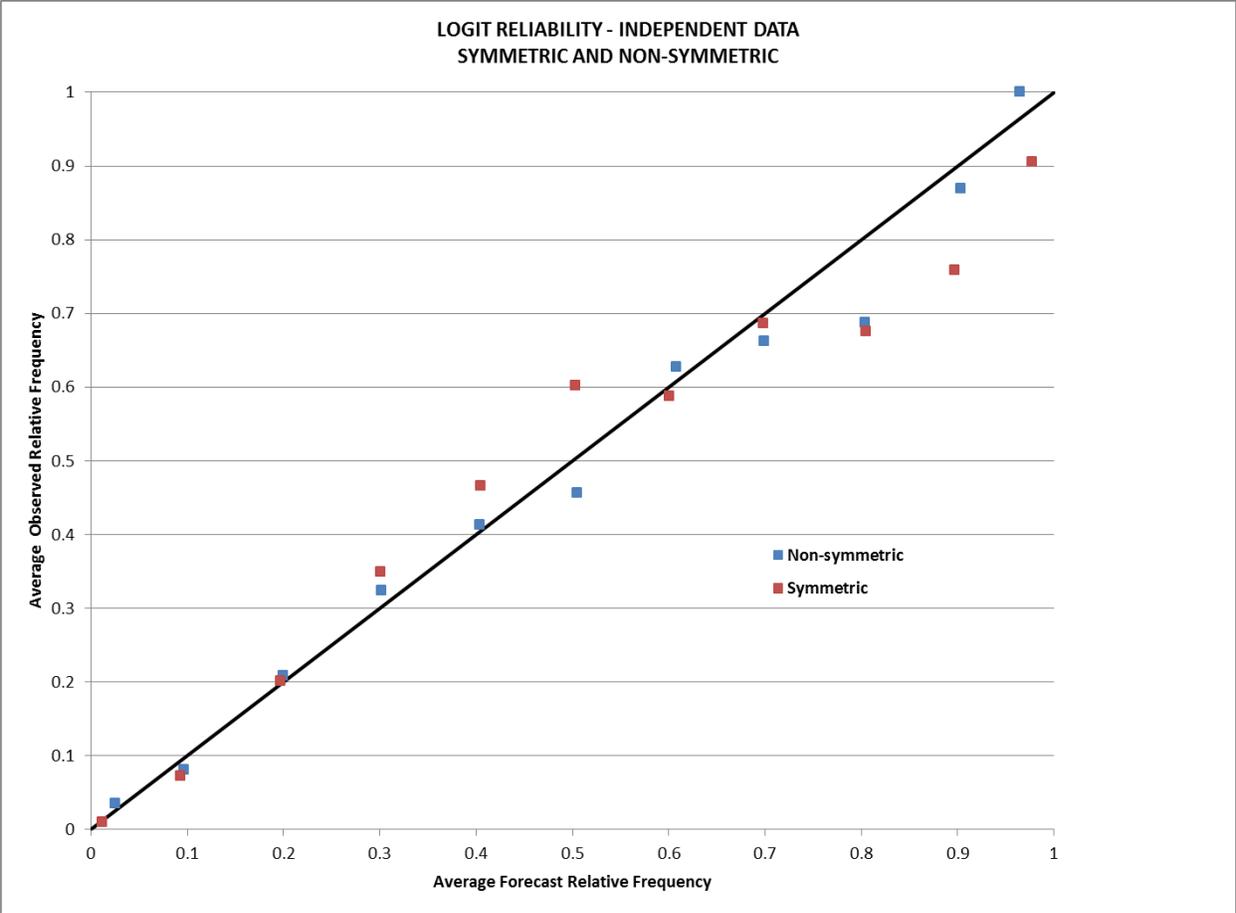


Figure. 2. The reliability of the symmetric and non-symmetric logit curves for the category ≥ 0.10 in on the 2012-2013 season of independent data.

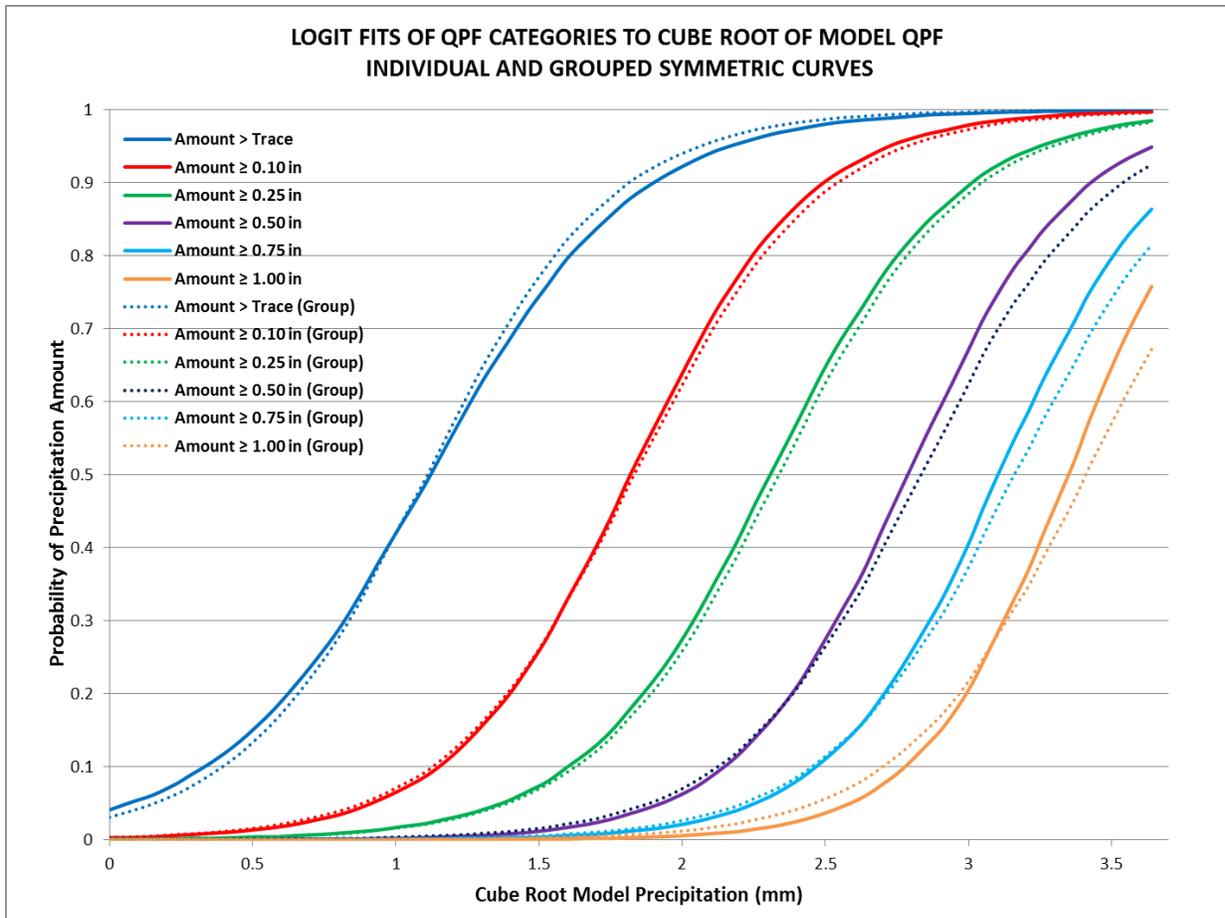


Figure 3. The symmetrical logit fits to the six categories of observed precipitation independently derived (solid lines, which are the same as in Fig. 1) and derived by grouping categories (dotted lines), as a function of the cube root of the model qpf in mm. The separation constants for grouped development are the 50% points for separate category development.

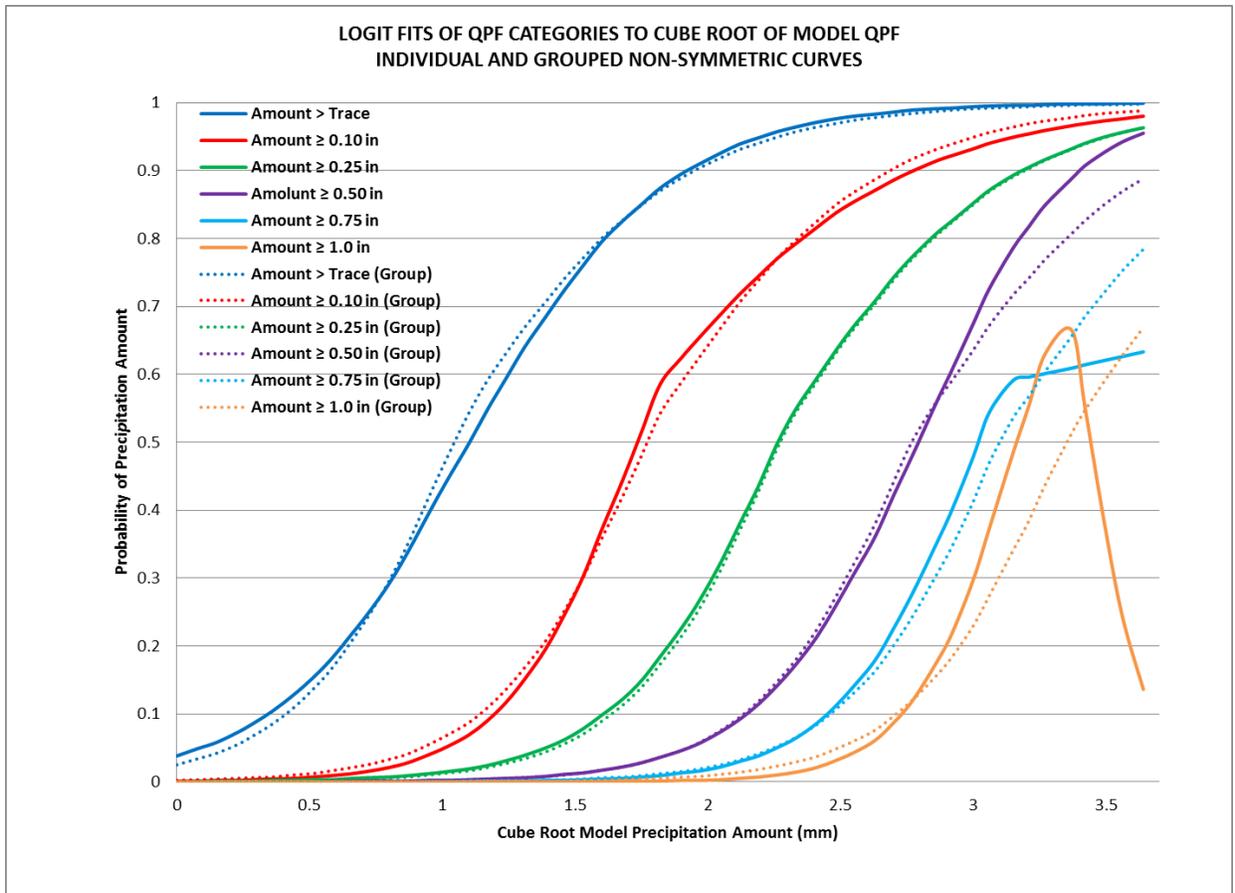


Figure 4. The same as Fig 3, except for non-symmetrical logit fits. The solid lines are the same as the dotted lines in Fig. 1.

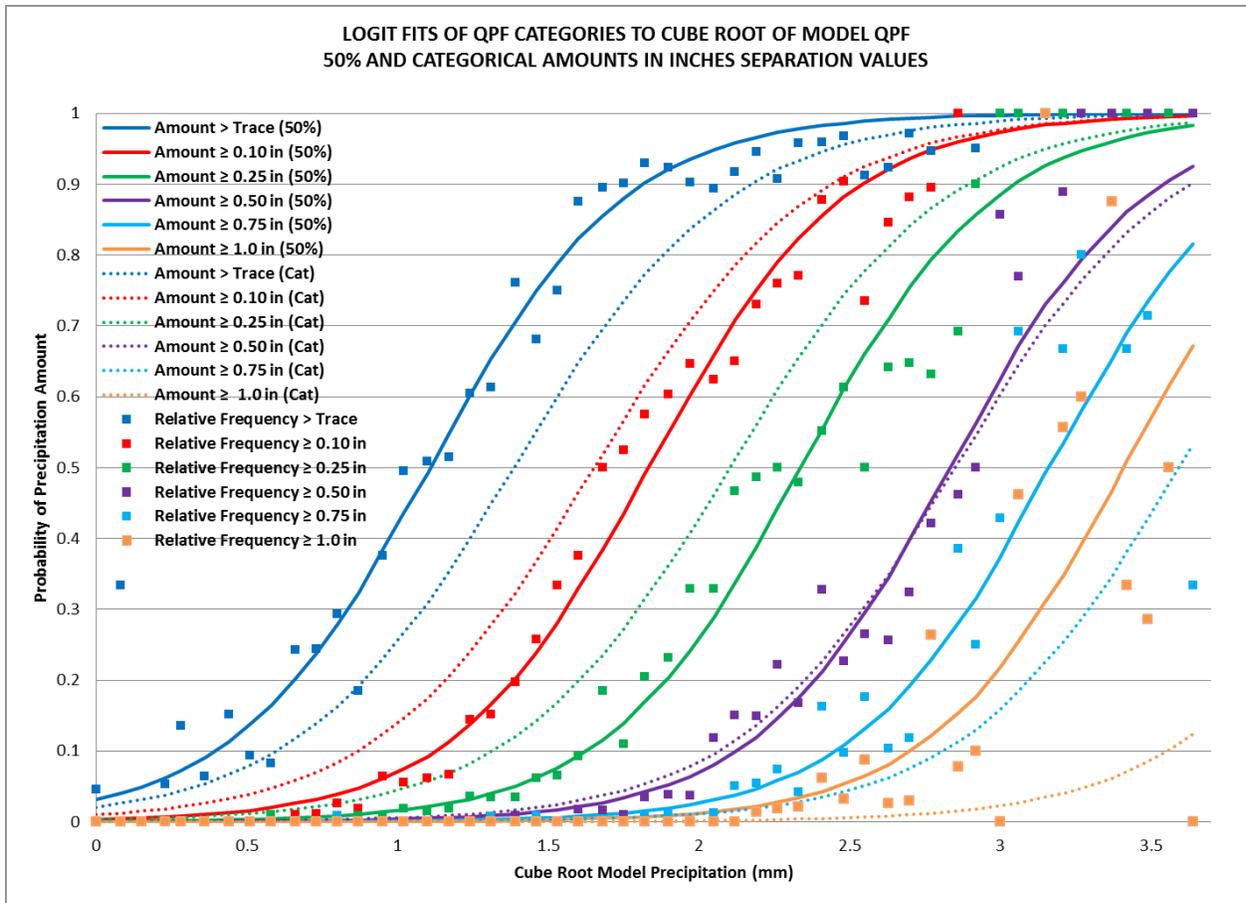


Figure 5. The symmetrical logit fits with separation constants defined by individual development 50% points (solid curves, the same as dotted lines in Fig. 3), and those representing category definition in inches (dotted lines).

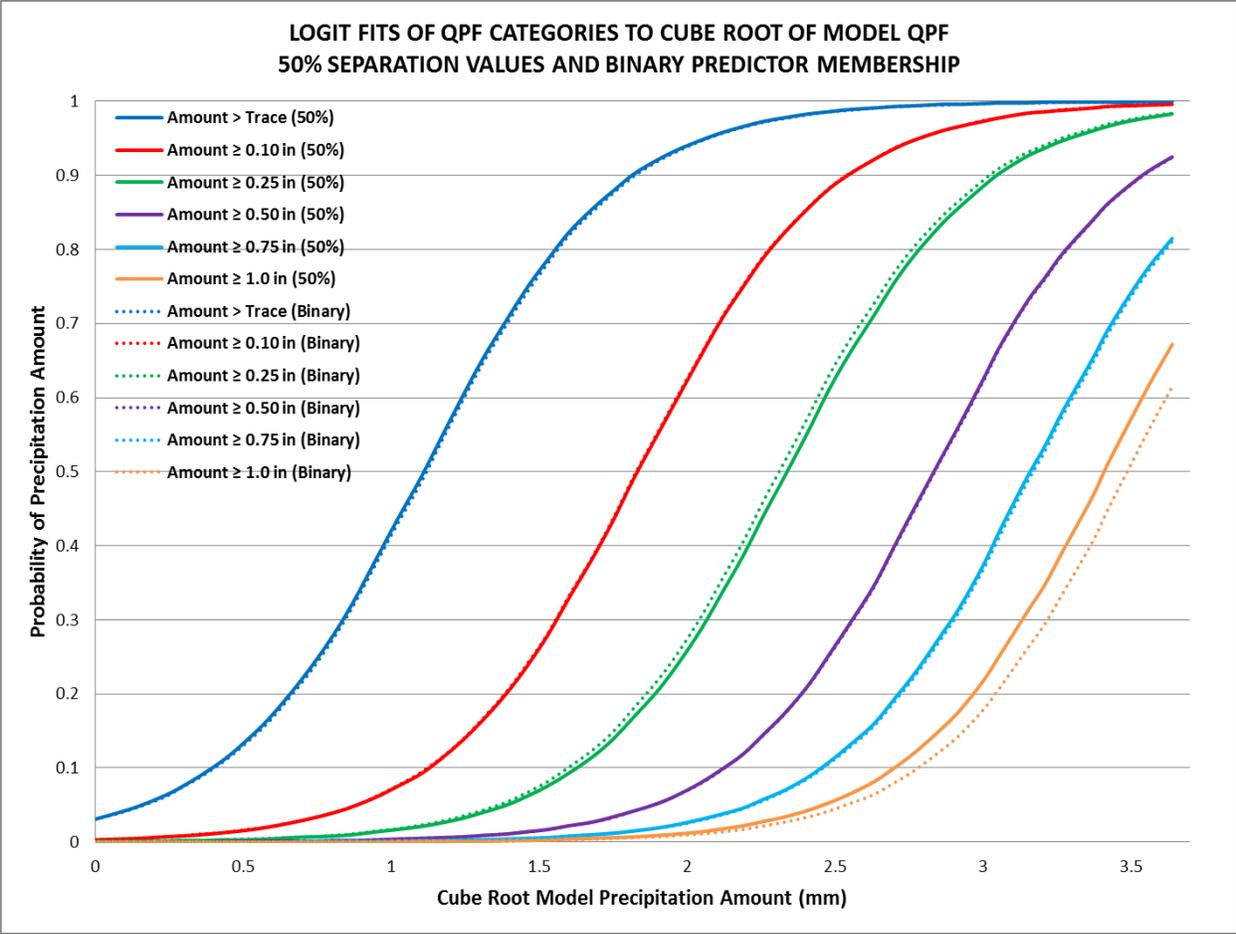


Figure 6. The symmetrical logit fits with separation constants defined by individual development 50% points (solid lines, the same as the dotted lines in Fig. 3), and the fits with binary predictors defining category membership (dotted lines).