Re-Evaluating Estimates of Temporally-Downscaled Climatological Precipitation Probabilities

MATTHEW HIRSCH
National Weather Service, Phoenix, Arizona

ABSTRACT

Climatological probabilities of precipitation (PoP) are utilized by operational meteorologists in the National Weather Service (NWS) and can be a valuable forecast tool. A dataset containing digital estimates of climatological PoPs on a 2.5 km X 2.5 km grid is currently available to forecasters in the NWS' Western Region. Temporal downscaling of the PoPs from 24-h to 12-h is achieved with a linear equation, though prior research suggests an exponential relationship may be more valid.

A new exponential equation is proposed, and the resulting 12-h climatological PoP values are compared to those currently used operationally. During the Cool Season, differences are negligible except in the high probability ranges common across the Pacific Northwest, where linear estimates are likely underestimating 12-h PoPs. Differences during the Warm Season are smaller but affect a larger number of gridpoints, particularly in the probability ranges most prevalent across the Southwestern United States. The operational values using the linear equation are likely overestimating 12-h PoPs in this region, though discrepancies are no higher than 3%.

1. Introduction

Research dating back to the 1970s has addressed the calculation of PoP as a function of sub-period probabilities. For example, Hughes and Sangster (1979, hereafter HS) initially used linear regression to combine 12-h sub-period probabilities into 24-h probabilities, though deficiencies were found with this method. HS ultimately tried a different approach, heuristically employing an exponent to account for the degree of dependence in the sub-periods. This method yielded successful results, but Wilks (1990) determined the HS method over-estimated probabilities in the lower half of the probability range, and he modified the equation by creating a new exponential dependency parameter. The modified Hughes-Sangster-Wilks (hereafter HSW) dependency parameter showed additional improvement in forecast reliability across several climate regions in the continental United States, and for nearly every month of the year. Krzysztofowicz (1999) investigated estimates of the sub-period probabilities from a theoretical perspective, and found they were comparable to those calculated using the HSW parameter.

Once the degree of dependency between two periods has been established, combining probabilities is computationally straightforward. However, manipulating longer-duration PoPs in reverse to derive sub-period PoPs, or temporal downscaling has not been attempted. An infinite number of possible combinations of sub-period probabilities exist, even with prior knowledge of the degree of dependency of the sub-periods. However, by assuming the sub-period probabilities are equal, estimates of the 12-h PoPs using the HSW parameter can be calculated. This assumption is applied to the NWS Western Region gridded (2.5 km X 2.5 km) database of 12-h and 24-h climatological (hereafter, NWS Climo) PoP values (Petrescu 2006a). No diurnal component exists in the PoPs; namely the 12-h daytime PoP is equal to the 12-h nighttime PoP. Although some benefit may be gleaned by calculating separate daytime and nighttime 12-h PoPs through empirical computations of climatological data, this study focuses instead on a refinement of the 24-h Climo PoP dataset. The 24-h Climo PoPs are temporally downscaled with the HSW parameter to derive 12-h PoPs, which are then re-evaluated against current operational values.

2. Relationship between 24-h and 12-h PoPs

The additive rule of probability for two events is given by equation (1) (Wilks 1995):

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\} \quad (1)$$

In the context of PoP, the terms $\Pr\{A\}$and $\Pr\{B\}$ are independent events and represent the PoP of the 12-h sub-periods. The intersection $\Pr\{A \cap B\}$ represents the occurrence of precipitation in both sub-periods, and the union $\Pr\{A \cup B\}$ represents precipitation occurrence in either or both sub-periods. However, precipitation occurrence is, in general, dependent, not independent (D. S. Wilks 2012, personal communication); that is, the occurrence (or nonoccurrence) of precipitation in one sub-period often has an effect on the occurrence (or nonoccurrence) of precipitation in another sub-period.

HS created a new statistical model of $\Pr\{A \cap B\}$ which incorporated dependency of the sub-period probabilities. The dependence parameter, k, is assigned to the higher of the two probabilities of $\Pr\{A\}$ and $\Pr\{B\}$ in equation (2).

$$\Pr\{A \cap B\} = [\Pr\{A\}]^k \Pr\{B\} \quad (2)$$

The HS dependence parameter, k, ranges from 0 (completely dependent) to 1 (independent). At one extreme, if $k = 0$, equation (2) simplifies to $\Pr\{A \cap B\} = \Pr\{B\}$. At the other extreme if $k = 1$, equation (2) becomes the more familiar, $\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B\}$ and equation (1) becomes $\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A\}\Pr\{B\}$. HS determined realistic values of k (Table 1) for both the Warm Season (April-September) and Cool Season (October-March) in the Upper Midwest by calibrating 12-h forecasts against 24-h observations.

TABLE 1. HS VALUES OF DEPENDENCE PARAMETER, k

Cool Season (October-March)	**0.55**
Warm Season (April-September)	**0.70**

The values in Table 1 intuitively make sense. One would expect the sub-period PoPs to become more independent during the Warm Season, as convective precipitation becomes more prevalent. Wilks (1990) validated these k values over the continental United States, but noted slight over-forecasting for smaller probabilities. This bias was eliminated by creating a new value of k, k^*, shown in equation (3), which increases exponentially with increasing probability.

$$k^* = k(1 - e^{-7\,\Pr\{B\}}) \quad (3)$$

In other words, more dependency exists between sub-periods for lower probabilities. By instead using k^*, Wilks found forecast reliability was improved for nearly all months and across all climate regions.

For simplicity's sake, $\Pr\{A \cup B\}$ can be redefined as the 24-h PoP, or P_{24}. If Climo PoP has no diurnal component, the sub-period probabilities are equal, $\Pr\{A\} = \Pr\{B\}$ = 12-h PoP, or P_{12}.

By combining terms, equation (1) then becomes:

$$P_{24} = 2P_{12} - P_{12}(P_{12}^{k\left(1-e^{-7P_{12}}\right)}) \quad (4)$$

Equation (4) forms the basis of the relationship between 24-h and 12-h PoPs.

If the probabilities of the two sub-periods are assumed to be equal and strongly dependent (i.e. $k = 0$), equation (4) simplifies to:

$$P_{24} = P_{12} \quad (5)$$

3. Assessing values of P_{12} and P_{24} graphically

In Figure 1, P_{24} has been plotted as the abscissa and P_{12} as the ordinate for ease of interpretation, despite the fact P_{24} is defined as a function of P_{12} in (4). Line A represents completely dependent (i.e. (5)) probabilities. Line C represents values for independent sub-periods.

Independent PoPs are analogous to individual coin flips. The outcome of one flip does not influence the outcome of the next flip. If a coin is flipped twice, what is the probability of "Heads" occurring (in at least one of the flips)? There are four possible outcomes: {Heads,Heads}, {Heads,Tails}, {Tails,Heads}, {Tails,Tails}. By simply counting the events in which "Heads" appears, the probability can be determined as ¾ or 75%. In terms of PoPs, if a given point is defined as (P_{24}, P_{12}) in Figure 1, this appears as the point (75%, 50%) on Line C.

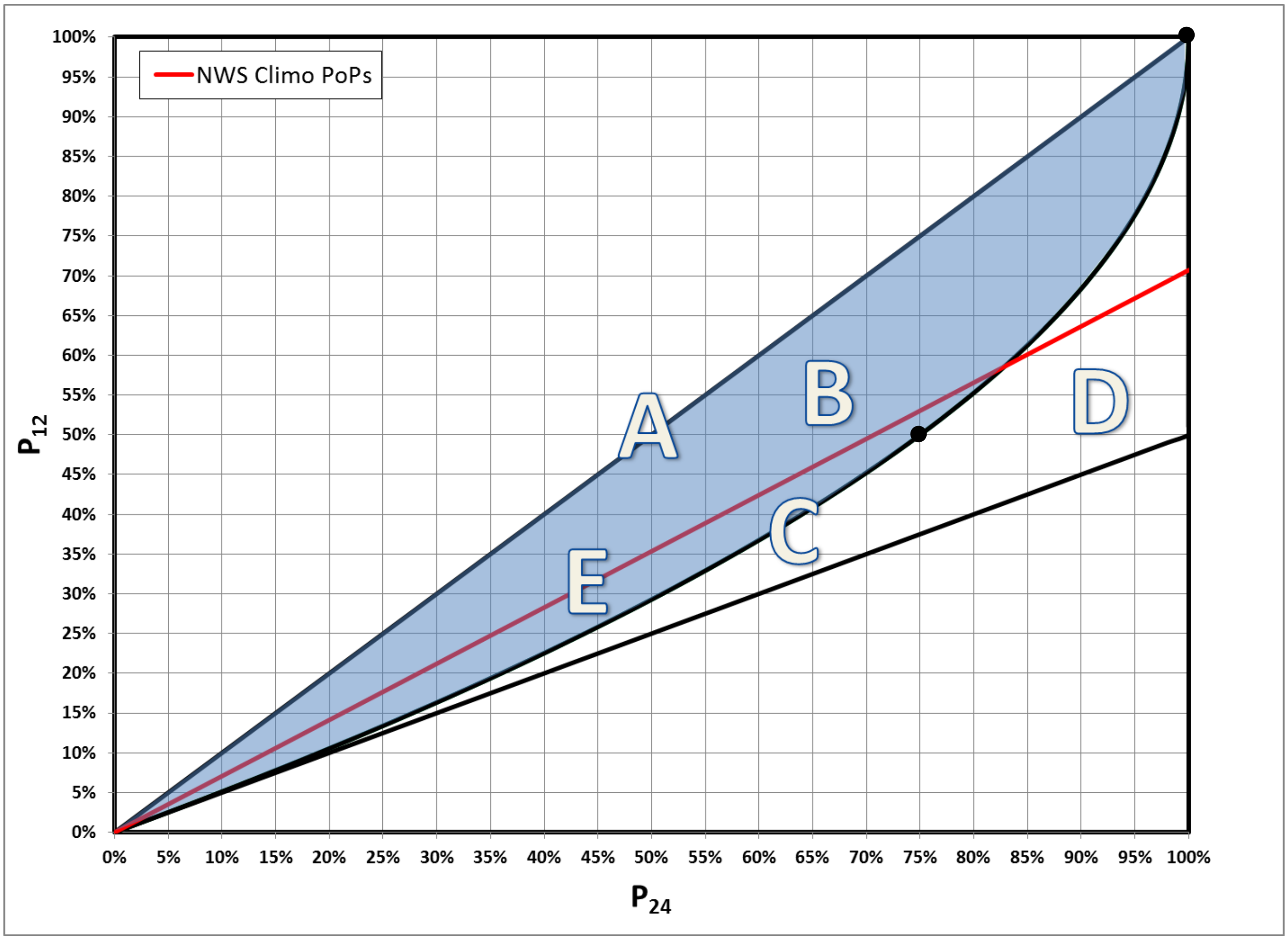

FIG. 1. Illustration of probability spectrum. Line A assumes complete dependence (i.e. $\boldsymbol{P_{12}} = \boldsymbol{P_{24}}$). Line C assumes complete independence (with no HSW scaling for $\boldsymbol{k}$). The point (75%, 50%) lies along Line C and the point (100%, 100%) lies along both Line A and Line C. Region B is the region between Line A and Line C and represents all allowable precipitation probabilities. Region D denotes negatively-dependent probabilities not applicable to PoPs. Line E represents the current operational NWS Climo PoPs.

PoPs are generally either independent or dependent, but not negatively dependent. Negative dependency arises when the probabilities exceed the independent probabilities (Krzysztofowicz 1999). Probabilities with the assumption of independence were generally determined to result in over-forecasting (Hughes (1980); Wilks (1990)). With negative dependency consequently ruled out, the range of possible probabilities has been reduced to those in the highlighted area (Region B) of Figure 1. The actual spectrum of probabilities lies somewhere between the linearly dependent line (Line A) and the curved independent line (Line C).

4. Results and Discussion – A comparison of NWS Climo PoPs and Exponential PoPs with varying degrees of dependency

a. Results

The 12-h values in the NWS Climo PoP data set were derived using the PRISM precipitation data set (Daly et al. 2008) and ultimately through linear regression of derived 24-h (daily) PoPs. The relationship between P_{24} and P_{12} appears within a text utility included in Petrescu (2006b) as:

$$P_{12} = \sqrt{1/2}(P_{24}) \quad \text{or} \quad P_{12} \sim .71(P_{24}) \quad (6)$$

When plotted graphically in Figure 1, P_{12} is a straight line (Line E) which tracks through the positively dependent region and undesirably into the negatively dependent region (Region D) when P_{24} reaches approximately 83%.

Probabilities for (4) using the HSW dependency parameter have been plotted in Figure 2a for $k = 0.55$ (Cool Season) and in Figure 2b for $k = 0.70$ (Warm Season). Both exponential curves in Figure 2 fall squarely within Region B of Figure 1.

The Cool Season values are within 1% of the NWS Climo PoPs when P_{24} is less than ~55% (and corresponding P_{12} is less than ~40%). However, differences in the techniques increase rapidly thereafter. An undesirable probability occurs when testing the extreme case of P_{24} = 100%. NWS Climo PoP P_{12} equals 71%, though this is clearly illogical. A P_{24} can only equal 100% if and only if both sub-period probabilities are also equal to 100%, not 71%. This appears as the point (100%, 100%) in Figure 1.

Larger differences are observed for the lower probability range when more sub-period independence is present and k is increased to 0.70 (Warm Season) in Figure 2b. Within the lower probability range, the differences in the two techniques reach a maximum of ~2.5% near $P_{24} = 40\%$ (and P_{12} for NWS Climo PoP is ~27.5%).

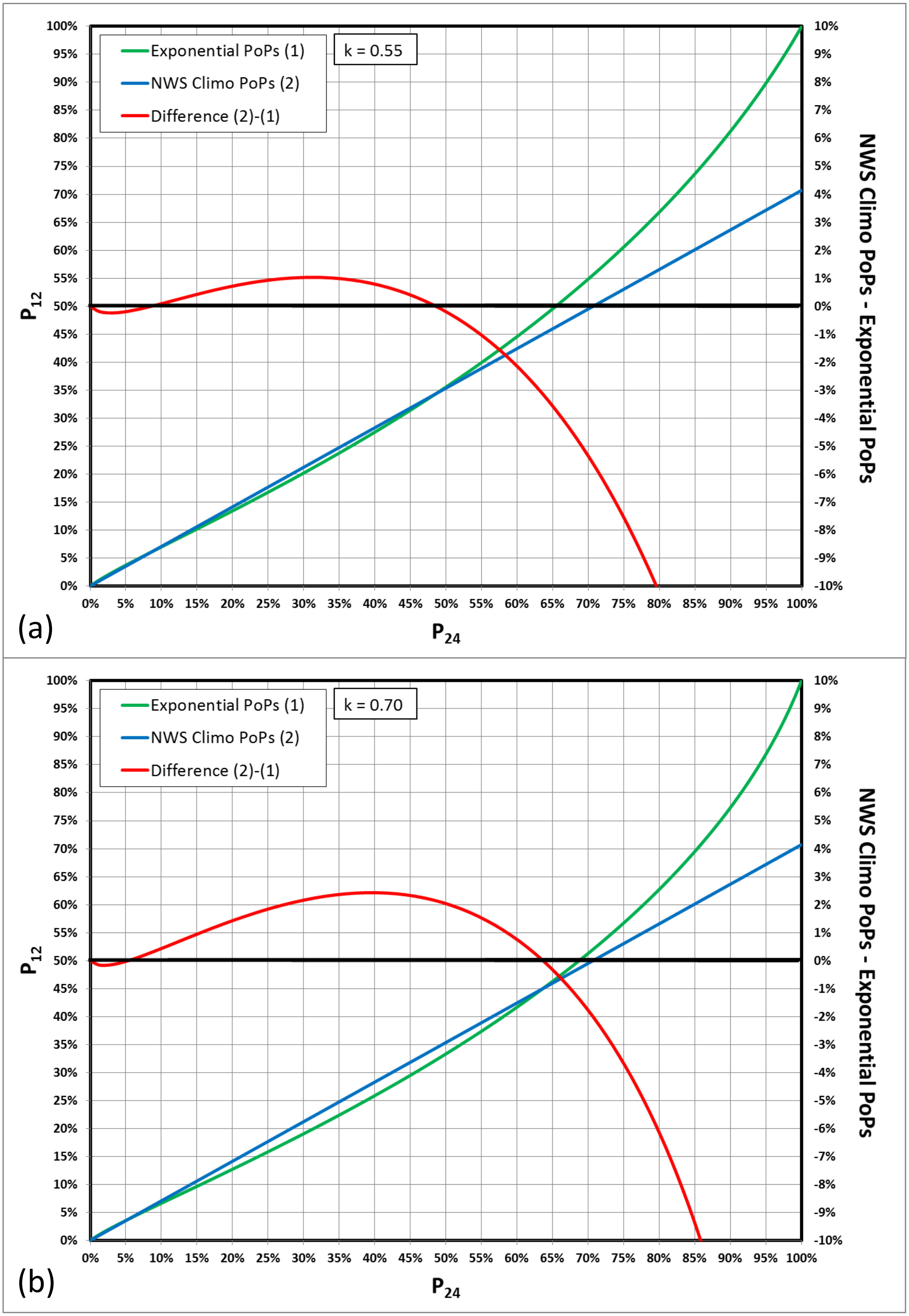

FIG. 2. P_{12} plotted as a function of P_{24}. The blue line represents the current operational NWS Climo PoPs. The green line (Exponential PoPs) represents the new PoPs using Equation 4 and a HSW dependency parameter of (a) k = 0.55 and (b) k = 0.70. The red line represents the difference between Exponential PoPs and NWS Climo PoPs; differences are plotted on the secondary axis on the right side of the graphs and are truncated at -10% and 10% to emphasize the smaller values.

b. Discussion

NWS Climo PoPs are remarkably similar to the Exponential PoPs, particularly for the Cool Season (Figure 2a) and for lower probabilities. The results in Figure 2 also indicate that NWS Climo PoPs in the upper end of the probability spectrum have likely been underestimated by failing to account for the exponential rate of increase in the probabilities. These higher PoPs are most prevalent across the Pacific Northwest during the Cool Season. For example, a P_{24} of 70% using the NWS Climo PoP equation (6) equates to a P_{12} of roughly 50%. The estimates using the HSW parameter indicate a P_{12} near 55% may be more accurate. It is important to note these larger differences (generally above P_{24} = 55%) affect only a small portion (roughly 13%) of first-order and COOP stations in the Western United States (Figure 3). Figure 4a shows P_{24} for January, taken from Petrescu (2006a). A visual estimation of Figure 4a would suggest roughly 5% to 15% of the gridpoints across the Western United States have P_{24} above 55% with corresponding values of P_{12} that are likely underestimated. Nevertheless, NWS forecasters should feel confident incorporating the Climo PoP dataset into their forecast methodology, particularly during the Cool Season and when lower PoPs are forecast, despite the shortcomings inherent in linear regression and the simplicity of equation (6).

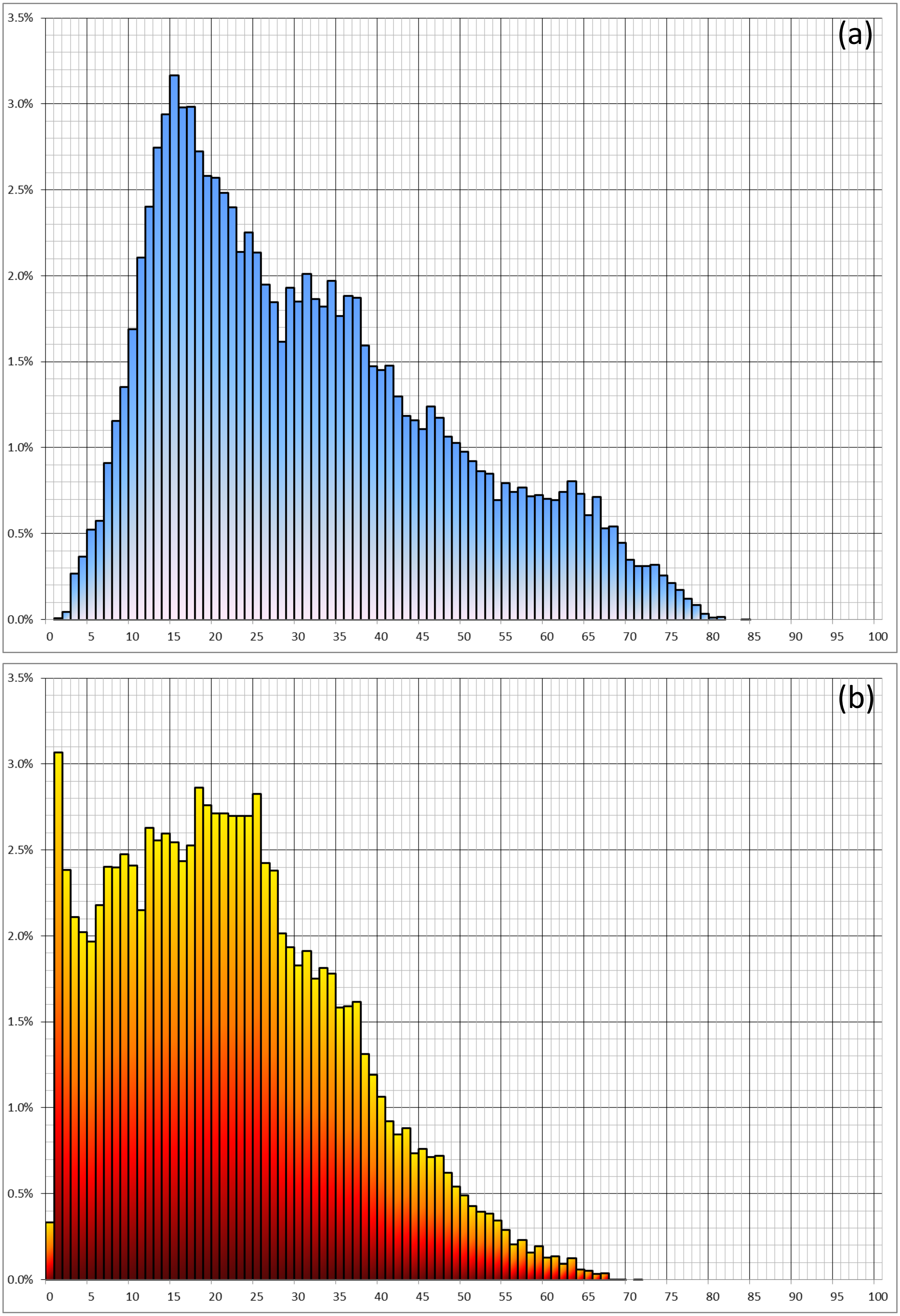

FIG. 3. Relative frequencies (along the y-axis) of all bi-monthly $\boldsymbol{P_{24}}$ (along the x-axis in %) for first-order, COOP, and SNOTEL stations in the Western United States calculated from Petrescu (2006b) for the (a) Cool Season and (b) Warm Season

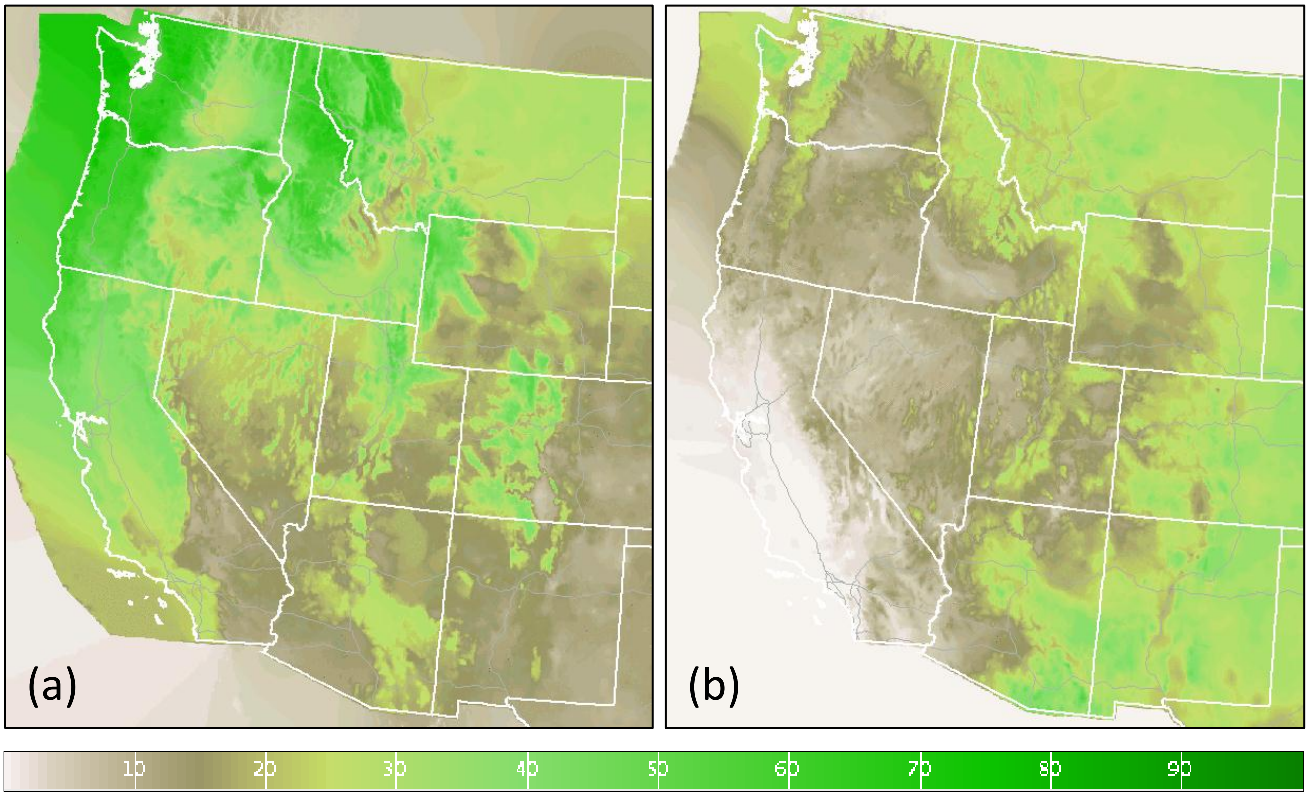

FIG. 4. NWS Climo PoP $\boldsymbol{P_{24}}$ in (%) over the Western United States for (a) January and (b) July, taken from Petrescu (2006a)

During the Warm Season ($k = 0.70$), lower probabilities are more common across the Western United States. NWS Climo PoPs have likely overestimated P_{12} when P_{24} is less than ~63%, though the differences remain below 3%. Differences in the techniques are most pronounced between P_{24} of 30% and 50% (and corresponding $P_{12}$of ~22% and ~35% for NWS Climo PoPs). This affects a larger portion of first-order, COOP and SNOTEL stations (approximately 25% of stations across the Western United States) than during the Cool Season. Figure 4b shows P_{24} for July taken from Petrescu (2006a), with the highest probabilities located in Arizona and New Mexico, associated with the influx of moisture and convection during the North American Monsoon.

It would be fair to assume values of k vary spatially by region, city, and even at higher resolutions such as a grid point. Similarly, they could vary temporally by season, month and even date. HS did verify the spatial variability for cities in the North-Central United States. Wilks (1990) nevertheless found increased forecast reliability by using a constant k across the entire continental United States and only varying the k parameter between the Cool Season and Warm Season. Additional work may certainly reveal a wide range of k values. It is plausible the Pacific Northwest does actually exhibit more sub-period dependence in the Cool Season, when stratiform precipitation events are more common. If this were true, the actual relationship between P_{24} and P_{12} would be closer to Line A in Figure 1, exacerbating the underestimate of the NWS Climo PoPs, even at low probabilities. For example, a k of 0.20 would result in underestimates of 3% at P_{24}=25% and near 10% at a P_{24}=55%. Conversely, Warm Season overestimates would be exacerbated as more independence (i.e. k=0.85) is introduced. The effect of changing k to more extreme values is illustrated in Figure 5. Similar diagrams for PoPs with HS and HSW-derived k values can be seen at:

http://www.weather.gov/psr/climopop

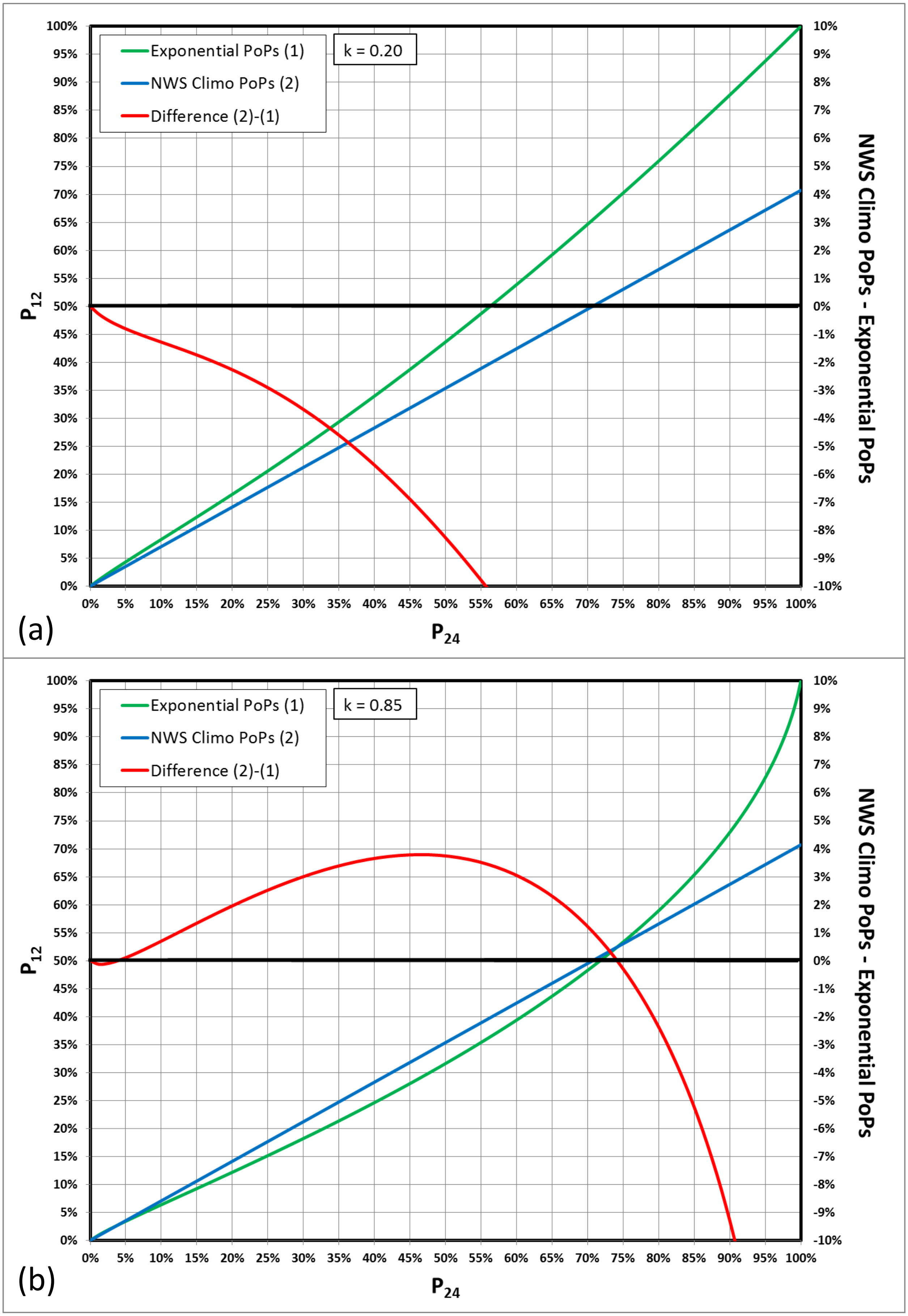

FIG. 5. As in Figure 2 but for (a) $\boldsymbol{k}$ = 0.20 and (b) $\boldsymbol{k}$ = 0.85.

5. Proposed change to NWS Climo PoPs

Not unlike the original HS experiment using linear regression, improved results were achieved by recognizing an exponential relationship with regard to PoPs. It is suggested that future updates of the NWS Climo PoPs use exponential curves, such as those derived using the aforementioned HSW Cold and Warm Season dependency parameters. This could presumably lead to more accurate precipitation forecasts while also substantiating regional verification statistics.

Since it is impractical to solve (4) for P_{12}, the exponential values can be approximated using the 5th degree polynomial equations in Table 2 below. Both equations have exceptionally high R^2 values (>0.9999). Although it may appear a 5th degree polynomial would be more computationally intensive than a linear equation, no discernible difference in processing was observed during testing, and gridded PoPs (at 2.5 km resolution) were able to be produced nearly instantaneously.

TABLE 2. SUGGESTED POLYNOMIAL EQUATIONS FOR P_{12}

Warm Season	$\boldsymbol{P_{12} = 2.91558(P_{24})^5 - 6.12826(P_{24})^4 + 4.99407(P_{24})^3 - 1.58656(P_{24})^2 + 0.79742(P_{24})}$
Cool Season	$P_{12} = 1.28965(P_{24})^5 - 2.79811(P_{24})^4 + 2.62294(P_{24})^3 - 0.88096(P_{24})^2 + 0.76517(P_{24})$

Figure 6 shows gridded values of P_{24} for a December day across western Washington. It is during this time of year when Climo PoPs are highest across the Pacific Northwest. Within this domain, 82% appears to be the highest P_{24}, along the Olympic Peninsula in the northwestern part of the state. Climo PoPs are temporally downscaled to 12-h in Figure 7 using the linear equation (6). The proposed adjustments using the polynomial equations can be seen in Figure 8. The modification becomes more apparent in Figure 9, which is a difference of the values in Figures 7 and 8. Nearly all gridpoints west of the Cascades exhibit an increase, with the largest differences (more than 10% higher) in the areas with the highest values of P_{24}. Recall from Figure 2a that the differences in the two techniques increase with increasing P_{24}.

Figure 10 shows gridded values of P_{12} for a day in July across Arizona using the current NWS Climo PoPs. The proposed adjustments are shown in Figure 11. Figure 12 shows the difference between the two grids. The largest differences (2% to 3% lower) are noted across the Mogollon Rim as expected, where probabilities are in the range which exhibited the largest differences in Figure 2b.

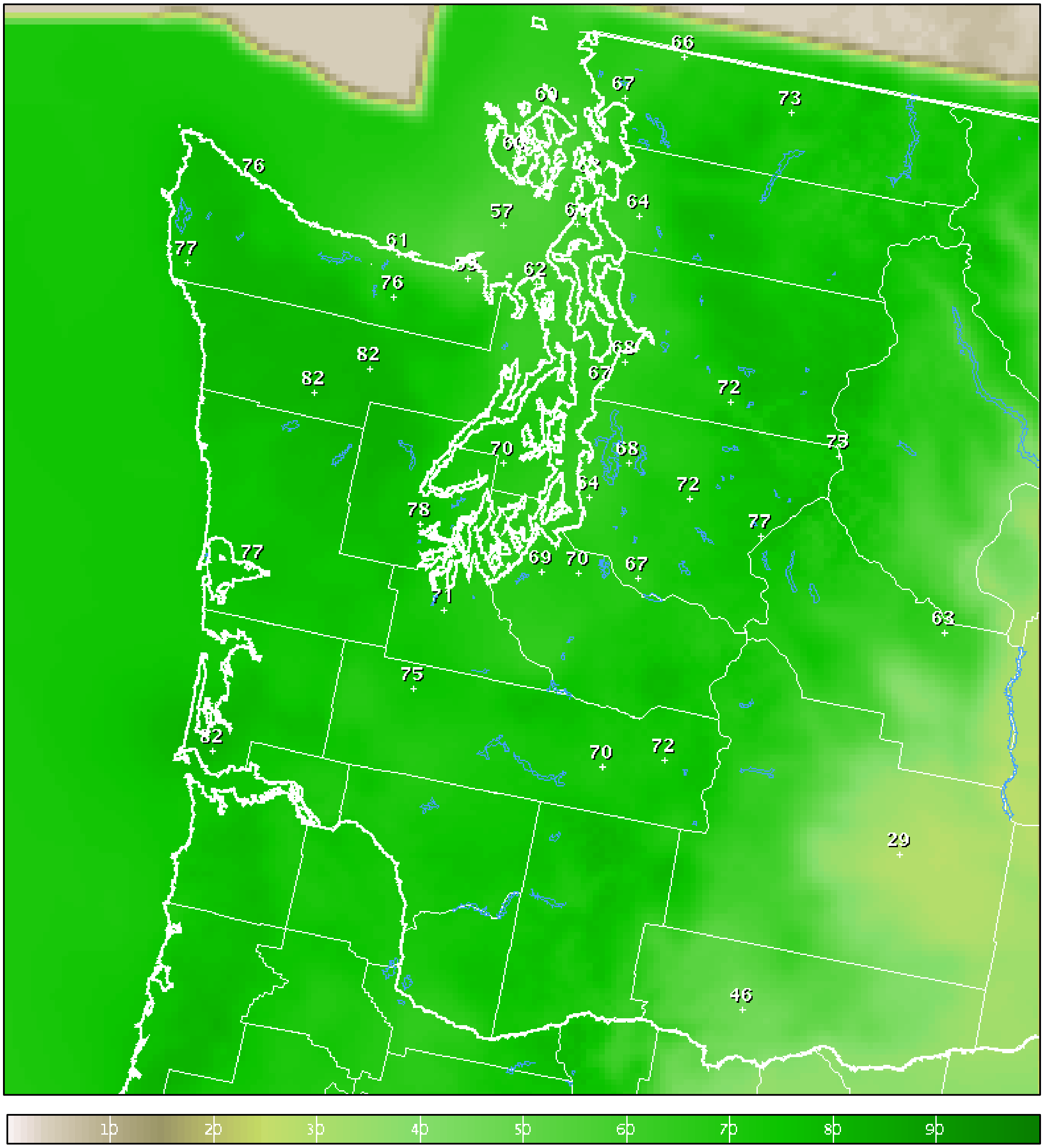

FIG. 6. NWS Climo PoP P_{24} (in %) for December 6th across western Washington. Numbers represent PoPs for selected cities and familiar observation sites. A maximum 24-h PoP of 82% is located across the northwest part of the state near Olympic National Park. Values along the extreme northern edge of the graphic should be disregarded.

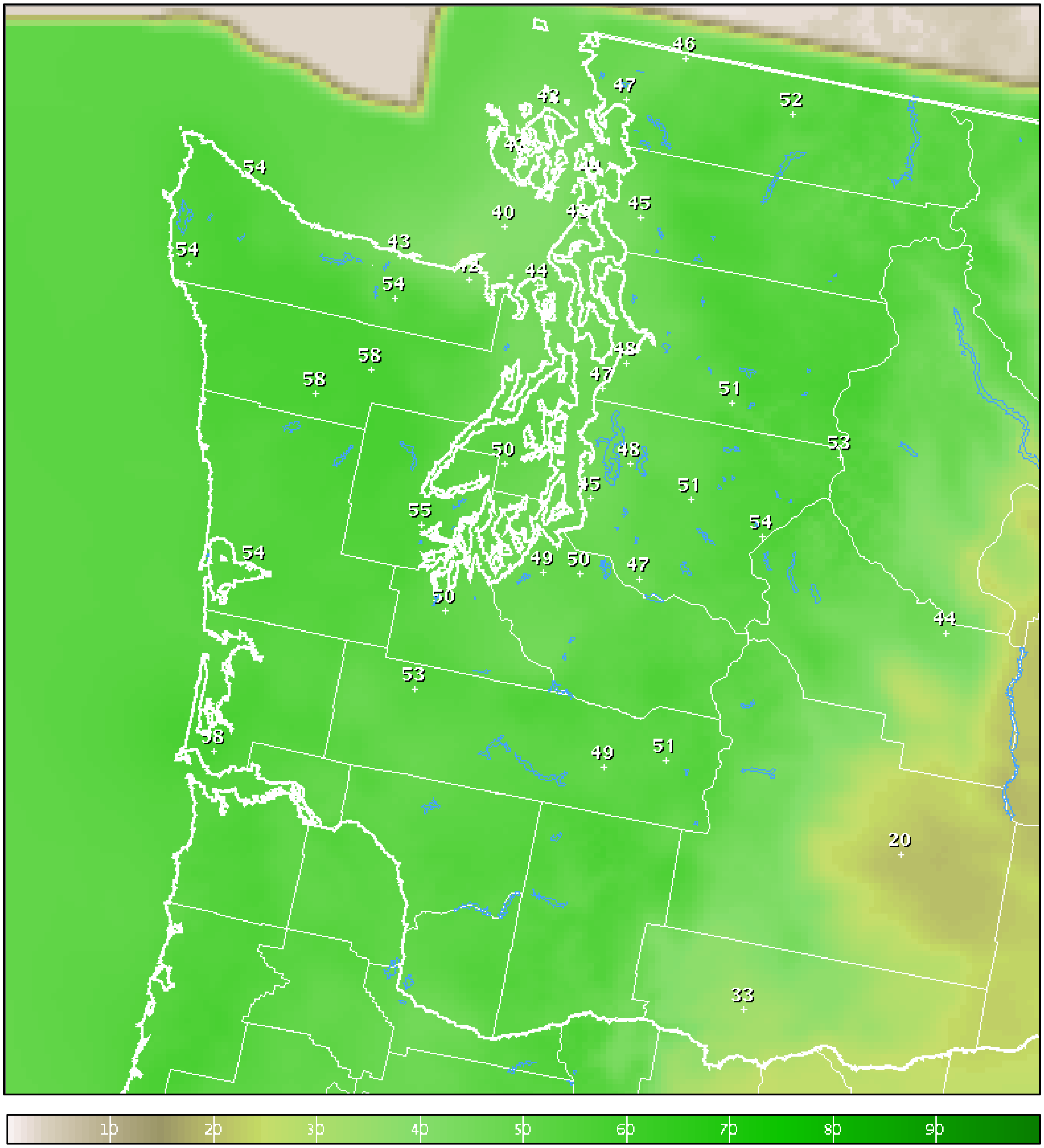

FIG. 7. As in Figure 6, but for P_{12} (in %).

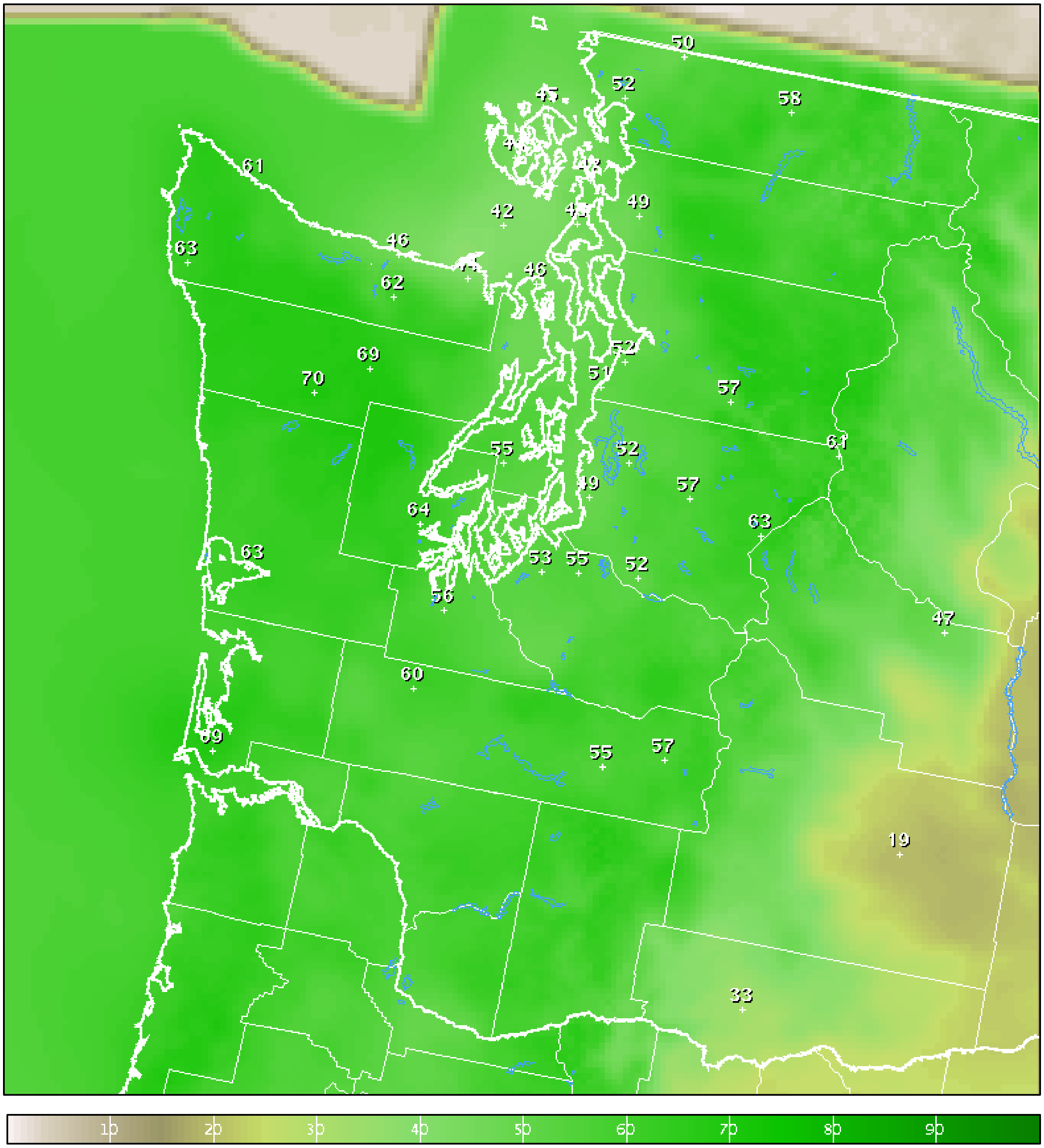

FIG. 8. As in Figure 7, but after the suggested adjustment.

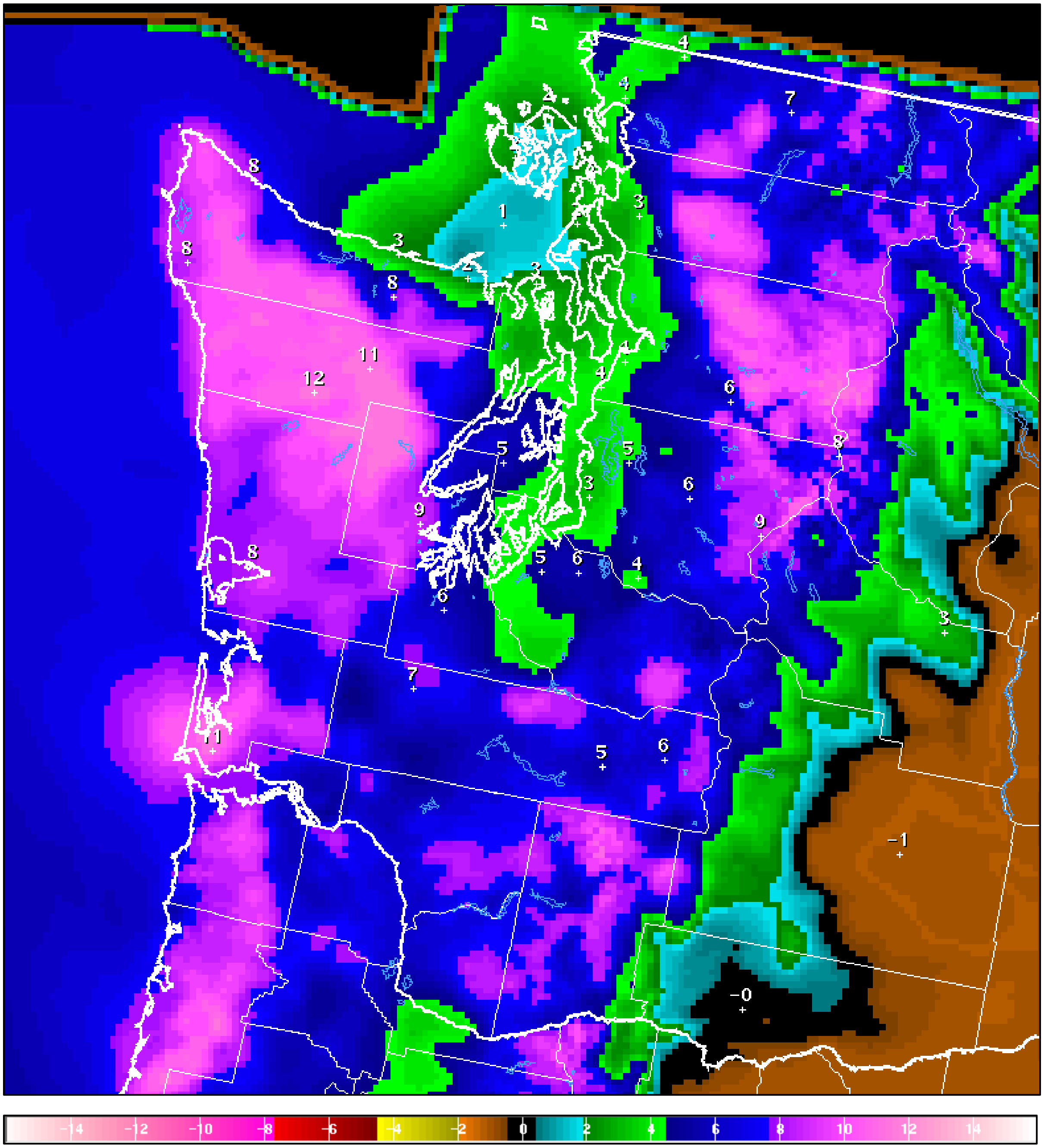

FIG. 9. Difference between the original NWS Climo PoPs and the suggested adjustment across Washington. Note the largest differences (greater than +10% in pink shading) are generally located across the Olympic Peninsula and along the windward side of the Cascades.

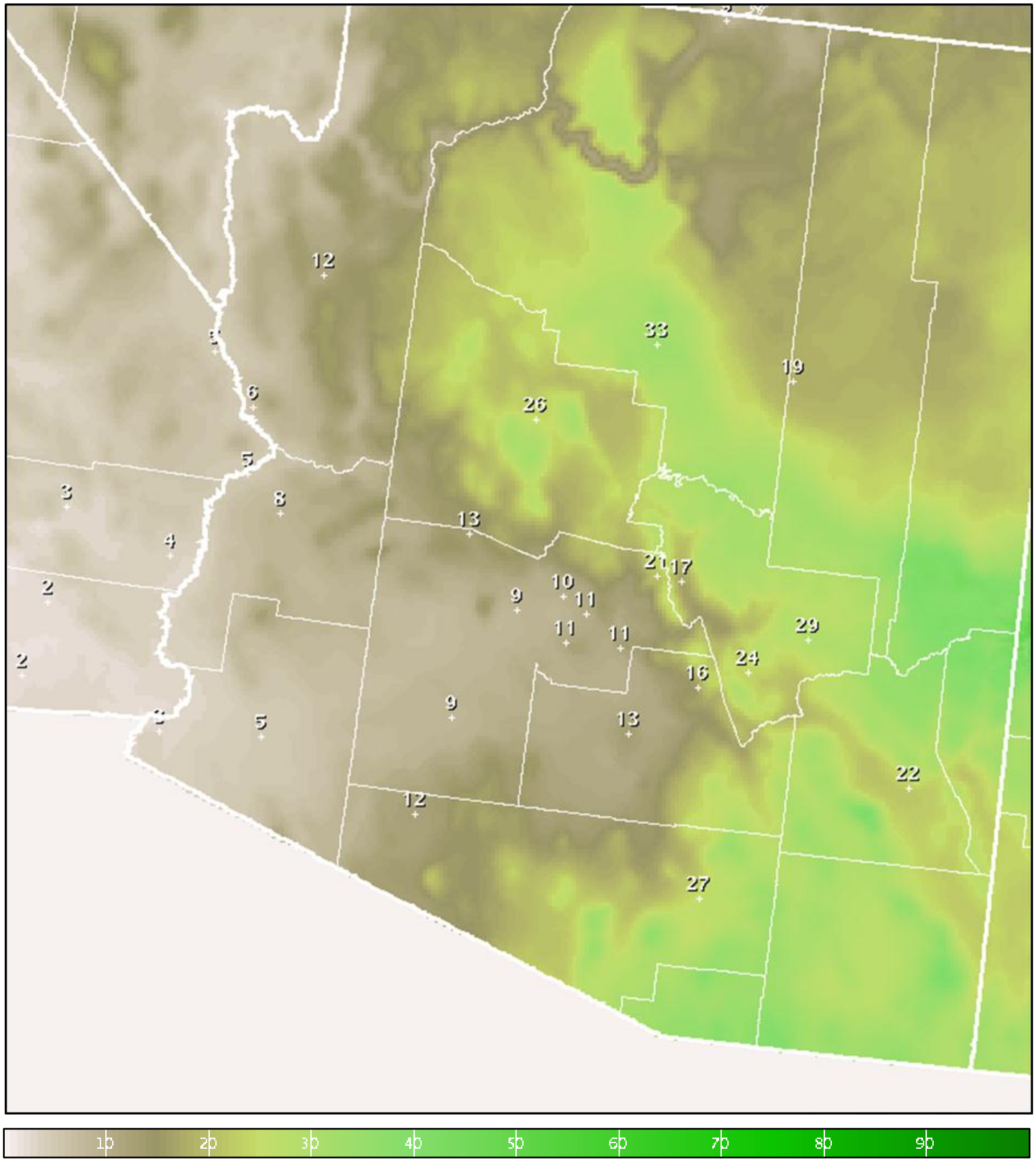

FIG. 10. NWS Climo PoP P_{12} (in %) for July 24th in Arizona.

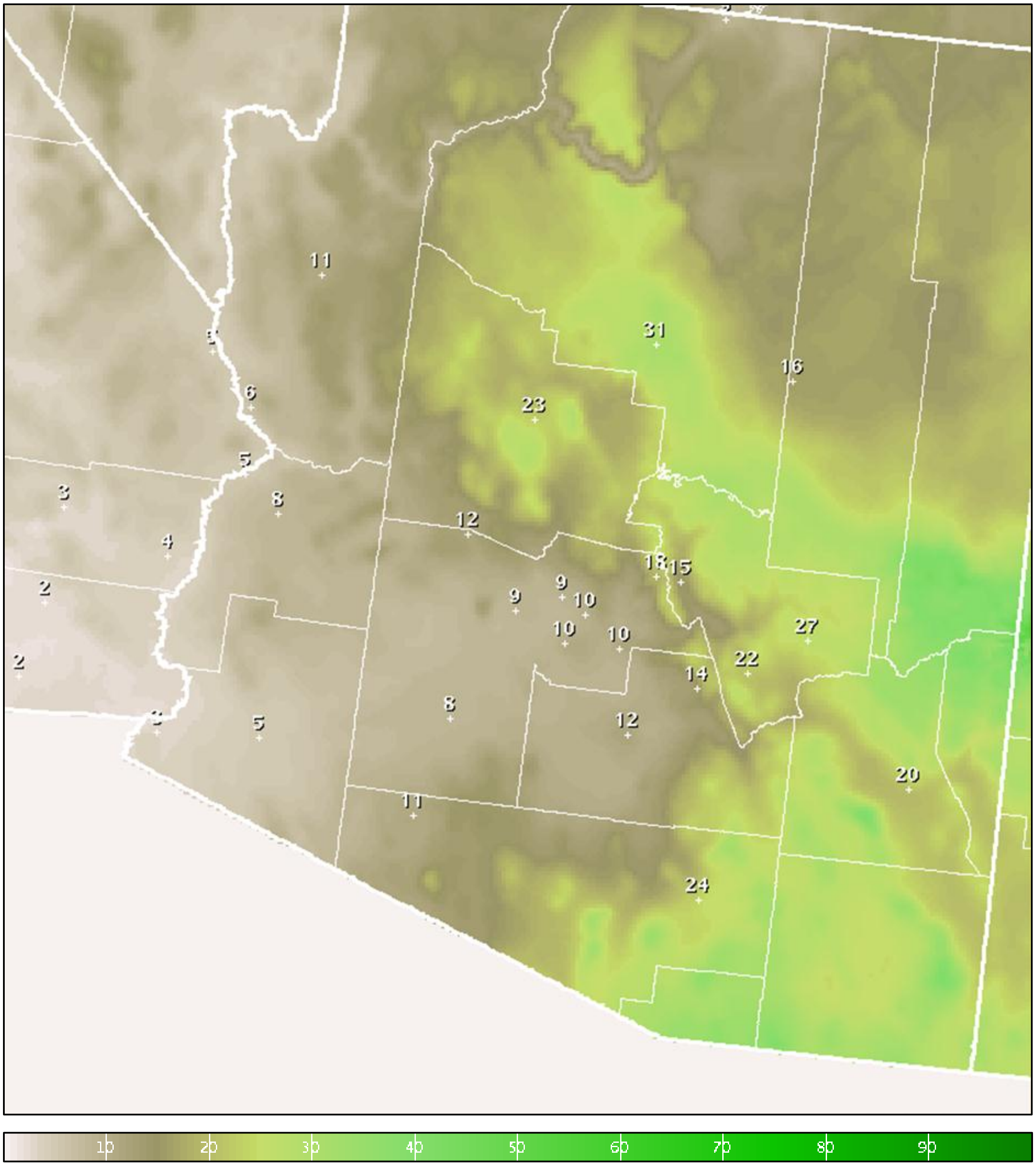

FIG. 11. As in Figure 10, but after the suggested adjustment.

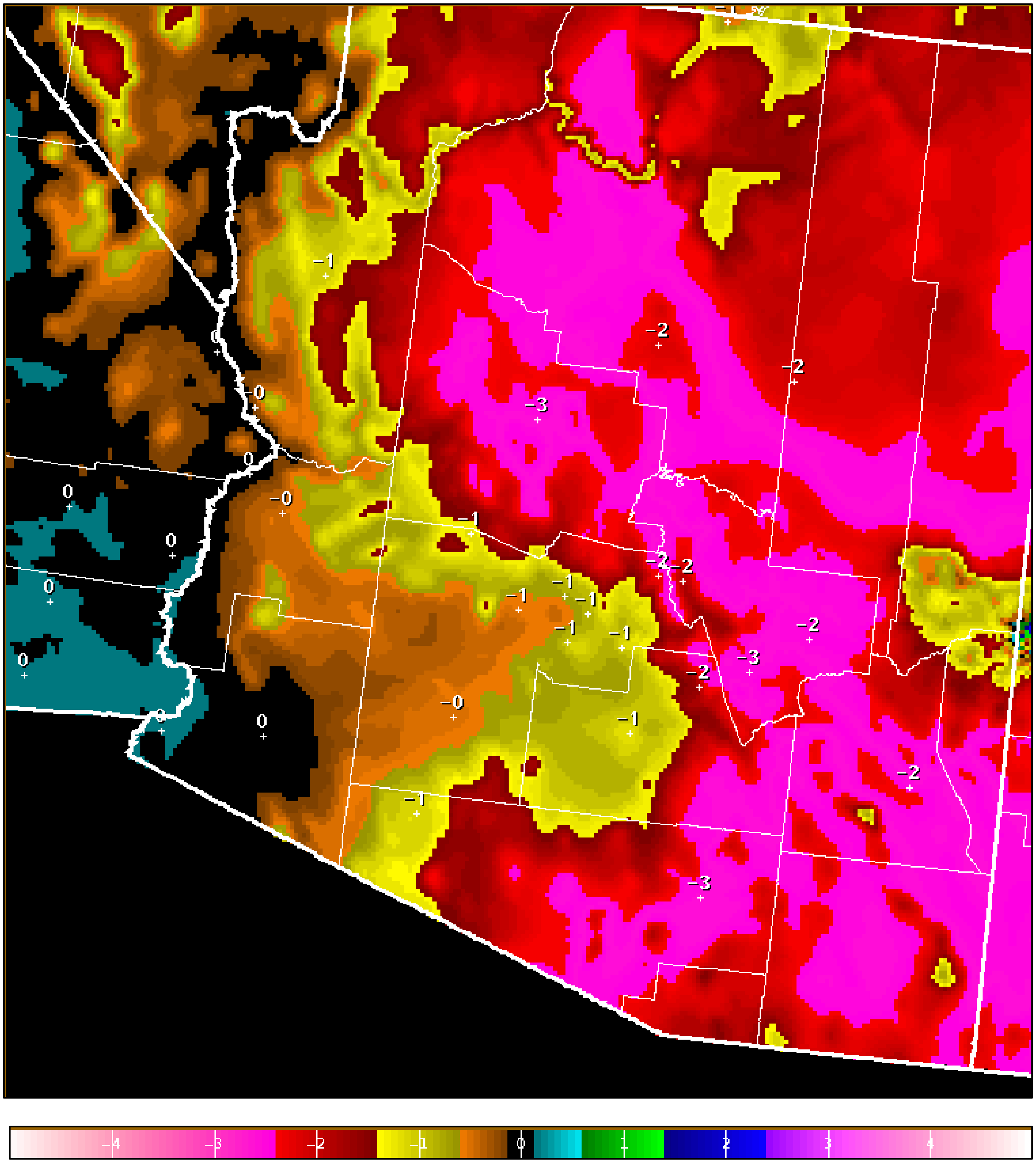

FIG. 12. Difference between the original NWS Climo PoPs and the suggested adjustment across Arizona. Note the largest differences (-2% to -3% in red and pink shading) generally located across the higher terrain, along the Mogollon Rim and across the southeastern part of the state. Differences are negligible in the lower deserts, where 12-h Climo PoPs are generally below 5%.

6. Conclusions

National Weather Service forecasters are required to forecast PoPs with durations less than 24 hours. When forecasting 12-h PoP, the NWS Climo PoP dataset is frequently employed by forecasters in the NWS' Western Region as a baseline or first guess, particularly in the extended period.

Gridded values of 24-h PoP in the Climo PoP dataset were temporally-downscaled to 12-h by assuming equivalent daytime and nighttime PoPs, and by utilizing an exponential equation derived from previous research. The recalculated values of 12-h PoP are generally close to the operationally-used linear estimates. Nevertheless, some differences were noted, specifically for higher probabilities during the Cool Season and for lower probabilities during the Warm Season. The proposed exponential curves represent a new climatological reference point by refining values at different spots along the probability spectrum. Although the adjustment is not critical, some benefit can be gained and implementation is recommended in a future Climo PoP update.

Acknowledgements. The author would like to thank Doug Green (NWS Phoenix, AZ) and Mike Staudenmaier (NWS WRH-STID) for their detailed reviews and suggestions. Gratitude is also extended to Gary Woodall (NWS Phoenix, AZ) for his helpful comments and to Dr. Dan Wilks (Earth & Atmospheric Sciences, Cornell University) for his guidance and statistical expertise. Paul Iñiguez (NWS Hanford, CA) assisted with the web page development and Gary Schneider (NWS Seattle, WA) supplied gridded Climo PoP data for the Pacific Northwest.

REFERENCES

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., and Pasteris, P.A., 2008: Physiographically-sensitive mapping of temperature and precipitation across the conterminous United States. *International Journal of Climatology*, **28**, 2031-2064.

Hughes, L. A., 1980: Probability Forecasting – Reasons, Procedures, Problems. NOAA Technical Memorandum NWS FCST 24, Meteorological Services Division, Silver Spring, MD, 84 pp.

Hughes, L. A., and W. E. Sangster, 1979: Combining precipitation probabilities. *Mon. Wea. Rev.,* **107,** 520–524.

Krzysztofowicz, R., 1999: Probabilities for a Period and Its Subperiods: Theoretical Relations for Forecasting. *Mon. Wea. Rev.*, **127,** 228–235.

Petrescu, E., cited 2006a: ClimoPoP Grids. [Available online at http://wr-s-sharepoint/WRSSD/digital_services/non_html/ClimoPoPGene.ppt]

Petrescu, E., cited 2006b: Regional Mod Number: WR06-008. [Available online at http://wr-s-sharepoint/projects/Shared%20Documents/wrModNotes/WR06-008.pdf]

Wilks, D. S., 1990: On the combination of forecast probabilities for consecutive precipitation periods. *Wea. Forecasting,* **5,** 640–650.

———, 1995: *Statistical Methods in the Atmospheric Sciences*. International Geophysics Series, Vol. 59, Academic Press, 467 pp.