APPLICATION OF A NEIGHBORHOOD POP FORECASTING APPROACH TO A COUNTY WARNING AREA

Christopher J. Schaffer*
National Weather Service, Southeast River Forecast Center

1. Introduction

Many studies in the past decade have used a neighborhood approach, which considers an area of grid points around an individual grid point in order to better account for uncertainty (Theis et al. 2005; Ebert 2008; Roberts and Lean 2008; Ajevek et al. 2009; Ebert 2009; Gilleand et al. 2009; Schwartz et al. 2010; Ruiz et al. 2011; Schaffer et al. 2011; Johnson and Wang 2012; Bouellegue et al. 2013; Kochasic 2013). While some studies use a neighborhood approach to verify forecasts, others use a neighborhood to produce forecasts.

Schaffer et al. (2011) applied a post-processing technique (originally proposed by Gallus and Segal [2004]) in the form of a two-parameter neighborhood approach to create a probability of precipitation (POP) from a gridded area of quantitative precipitation forecasts (QPFs). They showed that the resultant POPs were more skillful than the POPs of more traditional forecasting approaches. This two-parameter neighborhood approach was further tested by Kochasic (2013), who also showed that this two-parameter neighborhood approach could outperform more traditional methods of forecasting POPs despite its relatively simple nature.

The goal of this study is to test this two-parameter neighborhood approach over a National Weather Service County Warning Area (CWA) and compare its POP forecasts to those of the more sophisticated Model Output Statistics (MOS). Brier scores, bias values, and Relative Operating Curve (ROC) areas will be calculated for each approach's forecasts in order to judge their performance.

2. Methodology

The neighborhood forecasting approach in this study uses a single POP lookup table, created through the methodology originally described in Schaffer et al. (2011). Based on the findings of Schaffer et al. (2011), an 11x11 grid point neighborhood was used with a grid spacing of 20 km. To make the POP table, QPFs from the 2007 and 2008 Hazardous Weather Testbed Spring Experiment datasets (Kong et al. 2007; Xue et al. 2008) were verified against NCEP Stage IV precipitation estimates (Baldwin and Mitchell 1997) while considering two parameters: the average precipitation amount (assigned to one of seven QPF bins) within a neighborhood, and the number of points within this neighborhood with a QPF greater than or equal to 0.01 inch. For each possible combination of these two parameters, the correct alarm ratio was calculated, and these ratios collectively became POPs in the lookup table. The correct alarm ratio can be defined as h/f, where f is the number of neighborhoods with a particular combination of the two parameters, and h is a subset of neighborhoods in f that had precipitation observed at the center point of the neighborhood.[1]

For example, consider QPFs in a theoretical 3x3 grid point neighborhood (Fig. 1). In this neighborhood, six of the nine grid points have QPFs greater than or equal to 0.01 inch, and the average QPF in the neighborhood is between 0.05 inch and 0.10 inch. When considering all days and time periods within a data set, there will be a number of neighborhoods with the same two-parameter combination, and of that number, a fraction will have precipitation reported at the center grid point. This ratio can be expressed as a POP through the correct alarm ratio. For instance, if 5976 neighborhoods had an average QPF between 0.05 inch and 0.10 inch while six of the nine points had QPFs greater than or equal to 0.01 inch, and 2566 of these neighborhoods had precipitation observed at their center points, the correct alarm ratio would be 2566/5976=42.9%. This POP would be one of many within the POP table created by considering all possible combinations of the two parameters. Table 1 is an example of such a POP table, including the above POP. Note that the POP table for the 11x11 grid point neighborhood tested in this study would be much larger, because the neighborhood would contain 121 grid points instead of the 9 used in this example, increasing the number of rows.

Unlike Schaffer et al. (2011), 12-hour time periods were used in this study, rather than 6-hour time periods. This change was made because the National Weather Service verifies 12-hour POPs. Also, because the approach would be tested over gridded QPFs from April 2012 to April 2013, both the 2007 and 2008 Hazardous Weather Testbed Spring Experiment datasets were used to train the POP table. In Schaffer et al. (2011), only the 2008 dataset was used to train the neighborhood approach, so it could be tested against the 2007 dataset.

* *Corresponding author address*: Christopher J. Schaffer, NWS Southeast River Forecast Center, 4 Falcon Dr, Peachtree City, GA 30269; email: christopher.schaffer@noaa.gov

[1] In terms of the standard verification contingency table, the correct alarm ratio is the number of hits divided by the number of hits and false alarms

The Spring Experiments in 2007 and 2008 used ten-member ensembles, so the POP tables for each of the ten members were averaged together, similar to what was done in the final approach tested in Schaffer et al. (2011). This approach was shown to produce a more skillful POP than those of the individual members.

Four locations within the Goodland Weather Forecast Office (WFO) CWA were chosen for testing: Goodland, KS (GLD), Burlington, CO (ITR), Hill City, KS (HLC), and McCook, NE (MCK). These sites were chosen based on availability of ASOS precipitation observations, which were used for verification of the POPs. At each of the four sites, the neighborhood approach (abbreviated NBH) was applied to gridded QPFs from the GFS, NAM, SREF, and GFS MOS Guidance. The average of these four POPs was verified and compared to the POPs of NAM MOS (MET) and GFS MOS (MEX).

3. Results

3.1 Brier scores

Brier scores had a tendency to worsen (increase) most quickly for MET and least quickly for MEX, with the rate of increase for the neighborhood approach (abbreviated NBH hereafter) generally in between the two (Fig. 2). Considering forecasts through 60 hours, NBH had a lower (better) Brier score than MET and/or MEX for just over half of the time periods considered (11 out of 20 periods). For these forecasts through 2.5 days, NBH was competitive at Goodland, Burlington, and McCook, but performed worse than MOS for all time periods at Hill City. When considering forecasts past 60 hours (NBH and MEX), MEX outperforms NBH for all but 3 time periods. NBH's good performance for early time periods and poor performance for the later periods may be a consequence of how the NBH was trained. The Spring Experiments only had data through 30 hours, and this absence of training data past 30 hours could have caused NBH's POPs to be better suited for short term, rather than long term, forecasts. This could be tested and possibly remedied by training on a longer duration data set, and also by creating time-specific POP tables.

3.2 Bias values

Like the Brier scores, NBH had better Bias values than MET and/or MEX for just over half of the time periods considered prior to 60 hours (12 out of 20 periods). Unlike the Brier scores, however, bias values were almost always better for NBH compared to MEX for forecasts past 60 hours (Fig. 3). Of these latter 20 time periods, 18 time periods had NBH bias values that were closer to 1.00 (the ideal bias value) when compared to the MEX bias values, suggesting that NBH's overestimates better compensated for underestimates relative to MOS. With better NBH biases past 60 hours, but worse Brier scores, we have evidence to say that forecast errors for NBH may be more varied than those of MEX. More specifically, these bias values indicate that MEX is more likely than NBH to overestimate chances for precipitation, though considering the associated Brier scores, those accumulated errors are less than those of NBH.

3.3 ROC areas

ROC areas for MEX were almost always better than those of MET and NBH, showing that MEX is doing a better job than MET and NBH when discerning different forecast scenarios. Comparing MET and NBH (Fig. 4), NBH had a better (higher) ROC area for 13 of the 20 time periods. NBH tended to provide better ROC areas for the latter time periods, with MET tending to do better for the earlier periods. While NBH had better ROC areas for all periods at Hill City, KS, of the remaining 8 periods where NBH was doing better, 7 of the 8 occurred with forecasts after 24 hours.

4. Conclusion

Brier scores, bias values, and ROC areas were computed for four locations over the WFO Goodland forecast area for a neighborhood forecasting approach, MET, and MEX. The neighborhood approach was found to be competitive with the MET and MEX, often providing a more skillful forecast than at least one of the two. The neighborhood approach showed a competitive Brier score for over half of the time periods prior to 60 hours, though it did not do as well as MEX after 60 hours. On the other hand, bias values past 60 hours were almost always better for NBH compared to MEX, and bias values for the neighborhood approach were also competitive with MEX and MET prior to 60 hours. Finally, the neighborhood approach yielded a better ROC area over half of the time compared to the MET.

The analysis of Brier scores with time suggests that training the approach on later forecast periods could improve the approach's performance. Instead of using a single POP table, POP tables could be created for each 12-hour period in an attempt to provide a more skillful forecast. With this in mind, more sophisticated applications of a neighborhood approach to forecasting may provide POPs even more competitive with the commonly used MOS POPs.

Acknowledgments. The author thanks Jeremy Martin for his computational assistance and helpful discussions.

REFERENCES

Ahijevych, D., E. Gilleland, B. G. Brown, E. E. Ebert, 2009: Application of spatial verification methods to idealized and NWP-gridded precipitation forecasts. *Wea. Forecasting*, **24**, 1485–1497.

Baldwin, M. E., and K. E. Mitchell, 1997: The NCEP hourly multisensory U.S. precipitation analysis for operations and GCIP research. Preprints, 13th Conf. on Hydrology, Long Beach, CA, Amer. Meteor. Soc., 54–55.

Bouallegue, Z. B., S. E. Theis, C. Gebhardt, 2013: Enhancing COSMO-DE ensemble forecasts by inexpensive techniques. *Meteor. Z.*, **22**, 49-59.

Ebert, E. E., 2008: Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. *Meteor. Appl*., **15**, 51–64.

——, 2009: Neighborhood verification: A strategy for rewarding close forecasts. *Wea. Forecasting*, **24**, 1498–1510.

Gallus,W. A., Jr., and M. Segal, 2004: Does increased predicted warm-season rainfall indicate enhanced likelihood of rain occurrence? *Wea. Forecasting*, **19**, 1127–1135.

Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Wea. Forecasting*, **24**, 1416–1430.

Johnson, A., X. Wang, 2012: Verification and calibration of neighborhood and object-based probabilistic precipitation forecasts from a multimodel convection-allowing ensemble. *Mon. Wea. Rev.*, **140**, 3054–3077.

Kochasic, M. C., 2013: Further evaluation of improving probabilistic precipitation forecasts through the QPF-POP neighborhood relationship. Thesis, Department of Geological and Atmospheric Sciences, Iowa State University, 76 pp.

Kong, F., and Coauthors, 2007: Preliminary analysis on the realtime storm-scale ensemble forecasts produced as a part of the NOAAHazardous Weather Testbed 2007 Spring Experiment. Preprints, 22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction, Park City, UT, Amer. Meteor. Soc., 3B.2. [Available online at http://ams.confex.com/ams/pdfpapers/124667.pdf.]

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev*., **136**, 78–97.

Ruiz, J. J., C. Saulo, E. Kalnay, 2011: How sensitive are probabilistic precipitation forecasts to the choice of calibration algorithms and the ensemble generation method? Part II: sensitivity to ensemble generation method. *Meteor. Appl*., **19**, 314-324.

Schaffer, C. J., W. A. Gallus, Jr., and M. Segal, 2011: Improving probabilistic ensemble forecasts of convection through the application of QPF-POP relationships. *Wea. Forecasting*, **26**, 319–336.

Schwartz, C. S., and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280.

Theis, S. E., A. Hense, and U. Damrath, 2005: Probabilistic precipitation forecasts from a deterministic model: A pragmatic approach. *Meteor. Appl*., **12**, 257–268.

Xue, M., and Coauthors, 2008: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2008 Spring Experiment. Preprints, 24th Conf. on Severe Local Storms, Savannah, GA, Amer. Meteor. Soc., 12.2. [Available online at http://ams.confex.com/ams/pdfpapers/142036.pdf.]

0.00	0.00	0.07
0.00	0.05	0.12
0.01	0.09	0.18

Figure 1. An example of QPFs distributed across a 3x3 grid point neighborhood.

Table 1. An example of a POP table for a 3x3 grid point neighborhood. Each column represents a neighborhood average QPF (assigned to one of seven bins), and the rows indicate the number of grid points in the neighborhood with QPF >= 0.01 inch.

	<0.01	**0.01-0.05**	**0.05-0.10**	**0.10-0.25**	**0.25-0.50**	**0.50-1.0**	**>1.0**
0	7.0	-	-	-	-	-	-
1	23.5	30.7	21.1	0.0	-	-	-
2	26.0	31.6	30.0	31.0	-	-	-
3	26.5	33.3	38.7	36.7	38.9	0.0	-
4	23.9	33.7	38.7	42.9	44.7	66.7	-
5	25.5	34.4	42.5	44.4	49.0	50.0	100.0
6	22.2	36.4	42.9	48.2	50.0	54.4	66.7
7	-	38.6	45.6	48.6	53.5	52.7	63.6
8	-	38.9	48.8	51.2	57.0	58.3	55.4
9	-	38.7	51.6	64.4	75.2	78.3	77.5

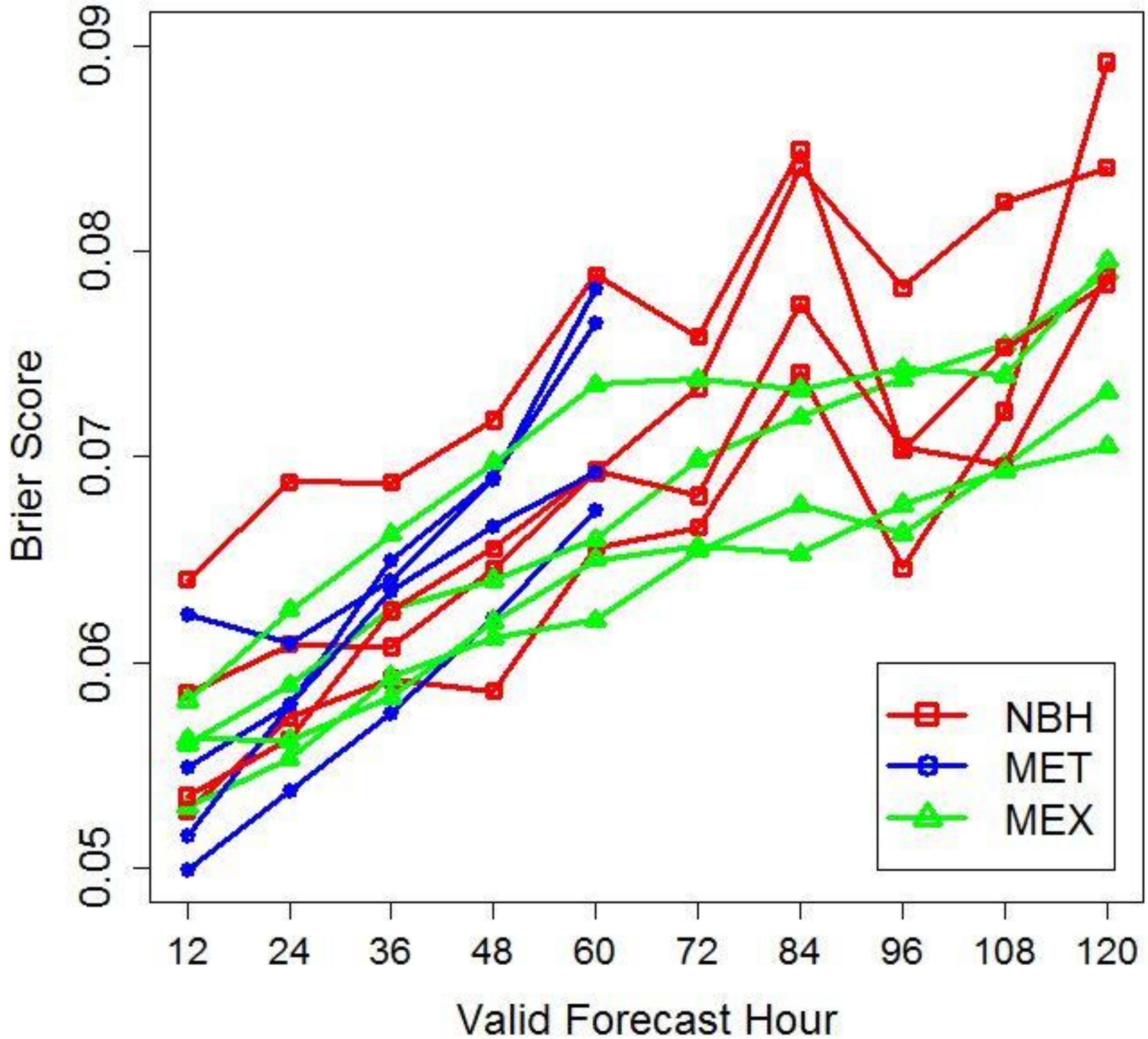

Figure 2. Brier scores over time for the four sites.

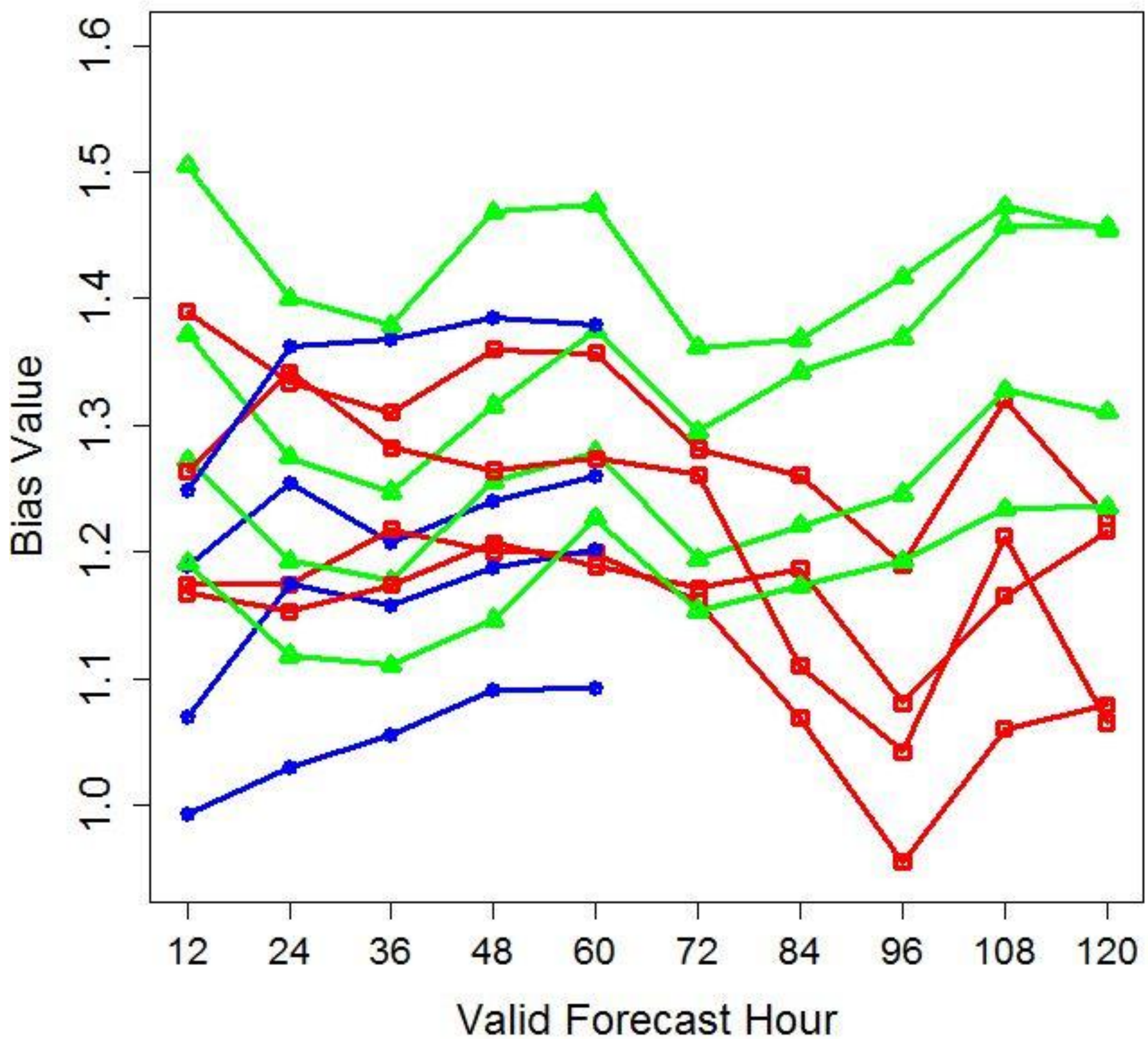

Figure 3. Bias values over time for the four sites, using the legend in Fig. 2.

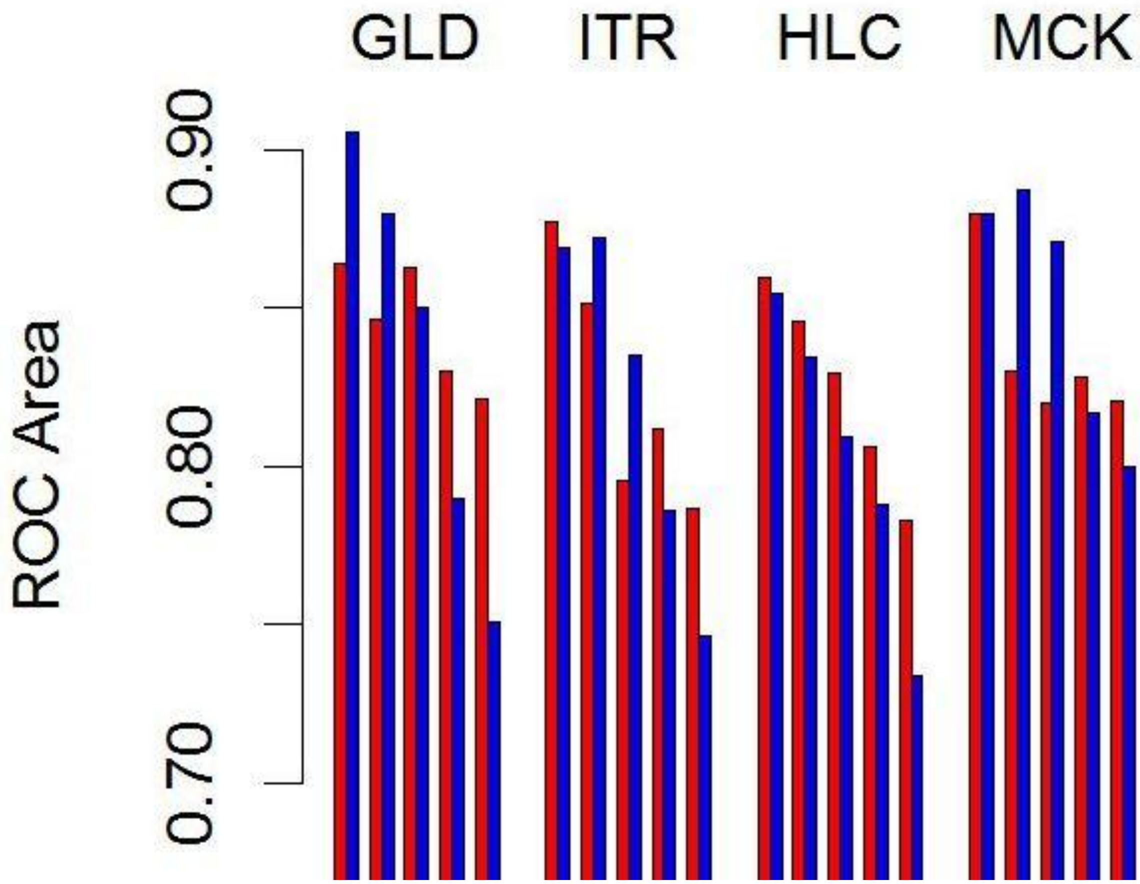

Figure 4. ROC areas for NBH (red) and MET (blue) for the four sites.