

Fuzzy Verification: Giving Partial Credit to Erroneous Forecasts

Beth Ebert

Bureau of Meteorology Research Centre

Melbourne, Australia

e.ebert@bom.gov.au

Motivation

A disadvantage of traditional point-based verification is that it gives no credit to “near misses”. That is, forecasts that are close are considered to be just as wrong as forecasts that are quite far away from the observations.

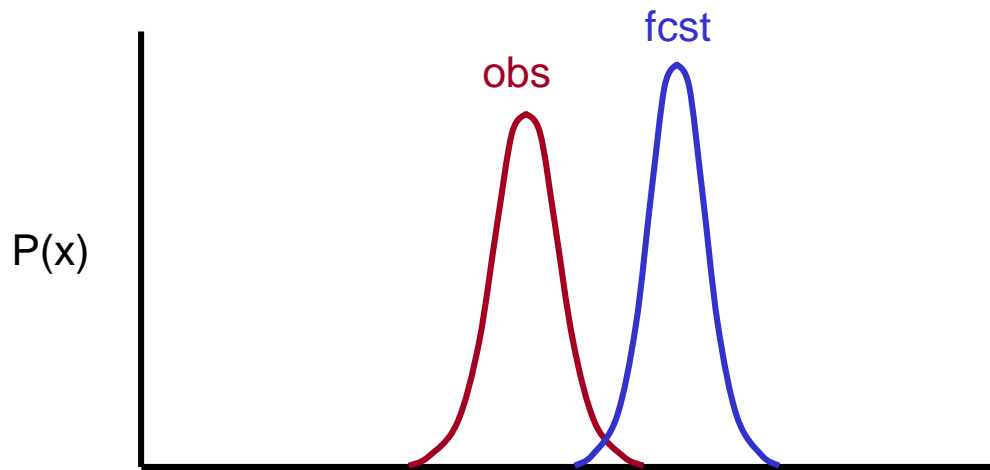
If the observation has some uncertainty associated with it, a “near miss” has some probability of actually being a “hit”. Using the same logic, a “hit” in the traditional view has some probability of being a “false alarm” when the observation is uncertain.

Concept

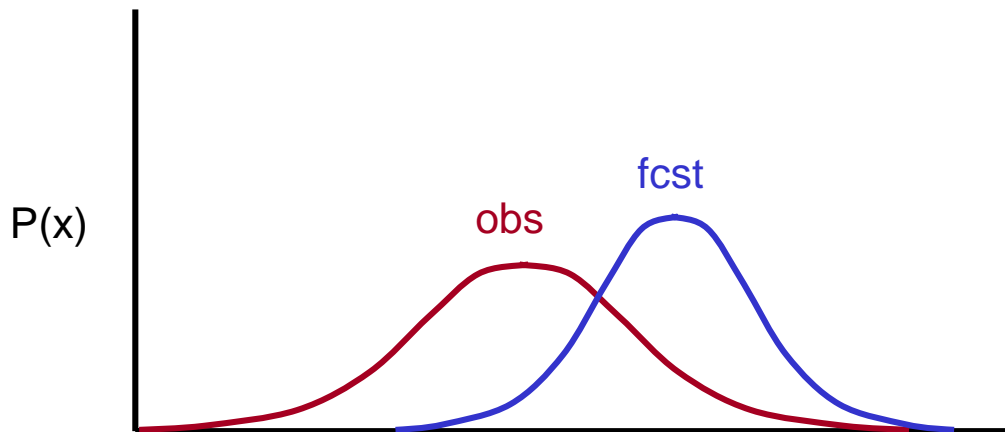
Fuzzy verification attempts to take into account the uncertainties in the observations and/or the forecasts, to give partial credit to forecast success and failure instead of only 0's and 1's.

Assume: A given forecast can be represented by a probability distribution function (PDF)

The corresponding observation can also be represented by a PDF



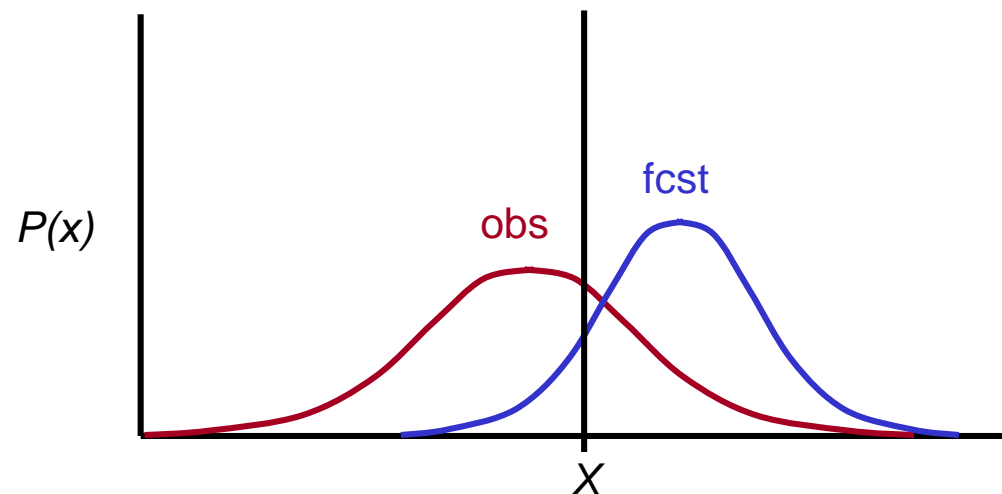
Strong separation between observed and forecast PDFs. Forecast has little chance of being correct.



Large uncertainty in both observation and forecast. Reasonable chance that the forecast may be correct.

Fuzzy categorical statistics

When computing categorical statistics, the observation and forecast are assigned to one or the other side of the yes/no threshold. When the threshold value X intercepts the PDF, as illustrated below, then there is some likelihood of the forecast or observation falling on either side of the threshold.



If $F_o(x)$ is the cumulative probability of observing a value of x , then $F_o(X)$ is the probability that the observed value is less than X , and $1-F_o(X)$ is the probability that the observed value is greater than X .

Similarly, $F_f(X)$ is the probability that the forecast value is less than X , and $1-F_f(X)$ is the probability that the forecast value is greater than X .

Combining the probabilities, the contingency table for this forecast-observation pair would be

		Forecast		
		yes	no	
Observed	yes	$(1 - F_o(X)) (1 - F_f(X))$	$(1 - F_o(X)) F_f(X)$	$1 - F_o(X)$
	no	$F_o(X) (1 - F_f(X))$	$F_o(X) F_f(X)$	$F_o(X)$
		$1 - F_f(X)$	$F_f(X)$	1

For the example given previously, the **fuzzy** contingency table would look something like

		Forecast		
		yes	no	
Observed	yes	0.255	0.045	0.30
	no	0.595	0.105	0.70
		0.85	0.15	1.0

The **traditional** way of scoring this forecast, not accounting for uncertainties in the observation and forecast, would be

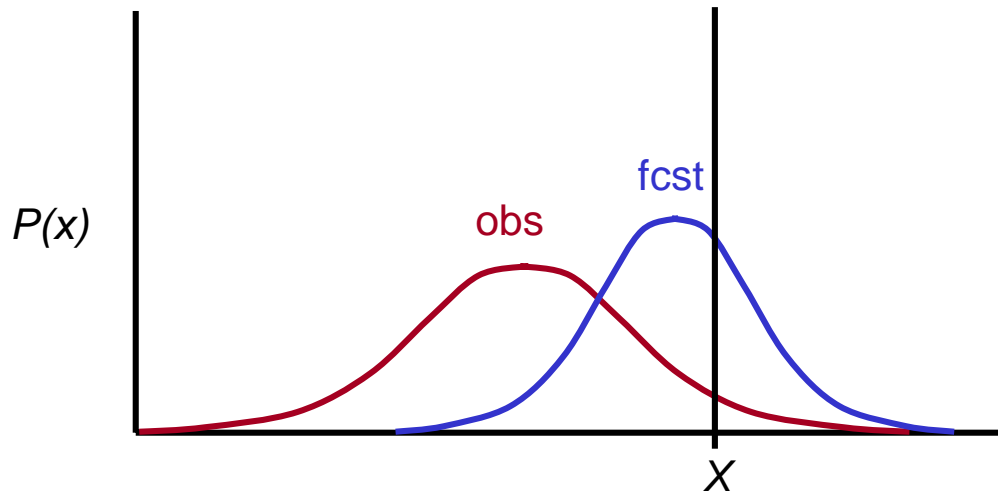
		Forecast		
		yes	no	
Observed	yes	0	0	0
	no	1	0	1
		1	0	1

which makes no allowance for the small probability (0.255) that the forecast was actually correct.

In this case fuzzy verification scores *better* than traditional verification.

Just as fuzzy verification gives forecasts partial credit for getting a forecast “sort of correct”, it also penalizes a forecast that would be perfect in the traditional view for getting it “sort of wrong”.

For example, if the threshold of interest had been greater than both the observed and forecast peaks, then the fuzzy verification results would have been *worse* than the traditionally derived results.



Forecast

		Forecast		
		yes	no	
Observed	yes	0.035	0.065	0.10
	no	0.315	0.585	0.90
		0.35	0.65	1.0

fuzzy

Forecast

		Forecast		
		yes	no	
Observed	yes	0	0	0
	no	0	1	1
		0	1	1

traditional

Brier score and fuzzy verification

$$BS(X) = \frac{1}{N} \sum_{i=1}^N [(1 - F_f(X)) - (1 - F_o(X))]^2$$

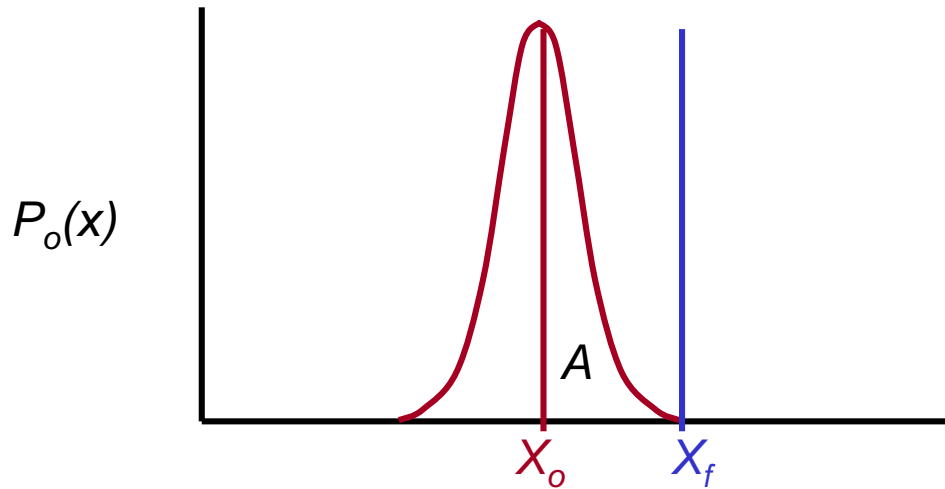
The Brier score lends itself well to this framework. Unlike the traditional Brier score that uses only values of 0 or 1 to represent the probability of an observation exceeding a threshold X , the fuzzy version uses the PDF of the observation to specify that probability.

Uncertainty weighted error

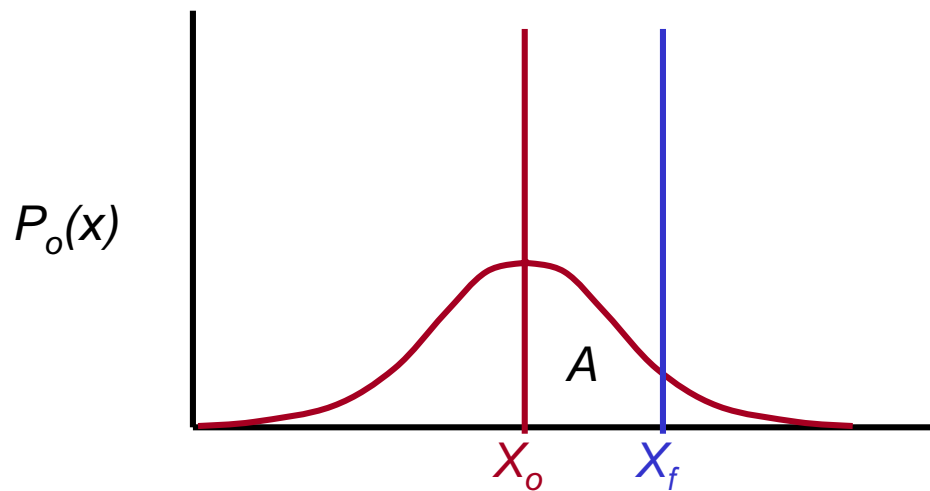
Motivation: When measuring the error in a forecast, the estimated forecast error should get greater credence if the observation has high certainty than if the observation is less certain. When accumulating forecast error statistics, it may be desirable to weight the final result toward the samples in which we have greater confidence.

Assume: Forecast is deterministic (for now)

Uncertainty in the observation can be approximated by a Gaussian PDF.



Forecast certainly incorrect.
Strong confidence in value of estimated forecast error.



Forecast might be correct.
Less confidence in value of estimated forecast error.

The area A under the PDF bounded by the expected value of the observation (X_o) and the forecast (X_f) is

$$A = |F_o(X_f) - F_o(X_o)|$$

$$\sim |F_o(X_f) - 0.5| \quad \text{if PDF is } \sim \text{Gaussian}$$

This is strongly related to Linear Error in Probability Space (LEPS), but instead of using the climatological PDF of observations, here we are using an *instantaneous* PDF and measuring the error in probability space with respect to the expected value of the particular observation.

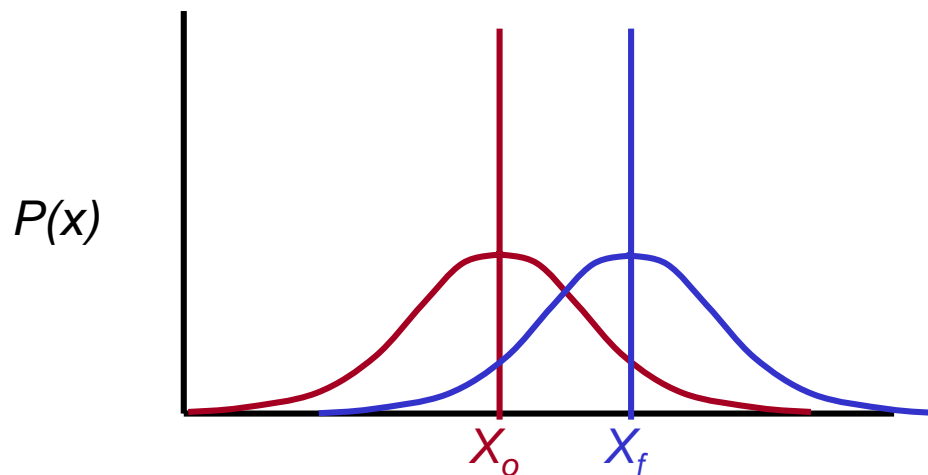
A weight that varies between 0 and 1 and reflects the uncertainty in the observation is the ratio of the area under the PDF bounded by the expected value of the observation (X_o) and the forecast (X_f), to the total possible area (0.5 in the case of Gaussian).

$$\text{weight} = A/0.5 = |F_o(X_f) - 0.5| / 0.5$$

The *error* could be $X_f - X_o$ or $|X_f - X_o|$ or $(X_f - X_o)^2$ or whatever metric is desired.

Extension to probabilistic forecasts:

Suppose that the forecast also contains uncertainty. Then the level of uncertainty associated with the measured error $X_f - X_o$ depends on both the uncertainty in the observations and the uncertainty in the forecast.



The area under the PDF of the observation, bounded by X_o and X_f , is (as before)

$$A_o = |F_o(X_f) - 0.5|$$

Similarly, the area under the PDF of the forecast, bounded by X_o and X_f , is

$$A_f = |F_f(X_o) - 0.5|$$

The combined weight when both the observation and forecast have PDFs is (after some algebra)

$$A_o/0.5 * A_f/0.5 = \\ [F_o(X_f) + F_f(X_o) - 2 F_o(X_f) F_f(X_o) - 0.5] / 0.5$$

Estimating the PDFs

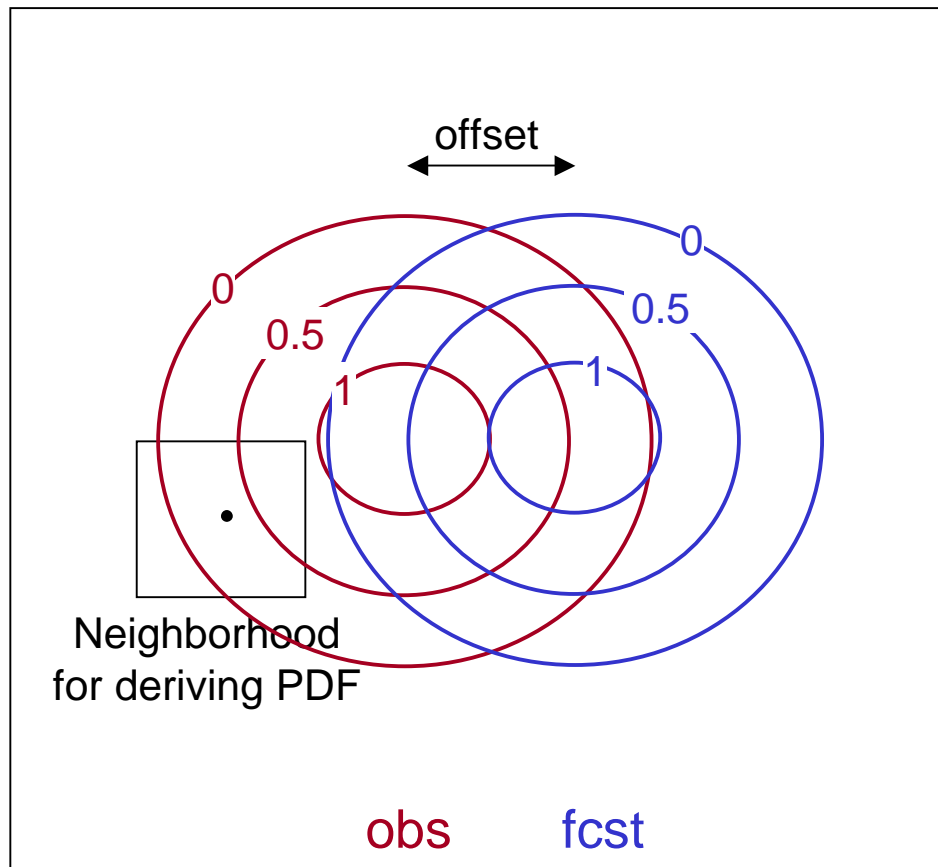
PDF for an **observation** could be estimated from:

- expected measurement (instrument) errors
- variances used in data assimilation
- kriging or other objective analysis scheme
- near neighbours in time or space

PDF for a **forecast** could be estimated from:

- ensemble members
- statistical relationships derived from historical data
- near neighbours in time or space

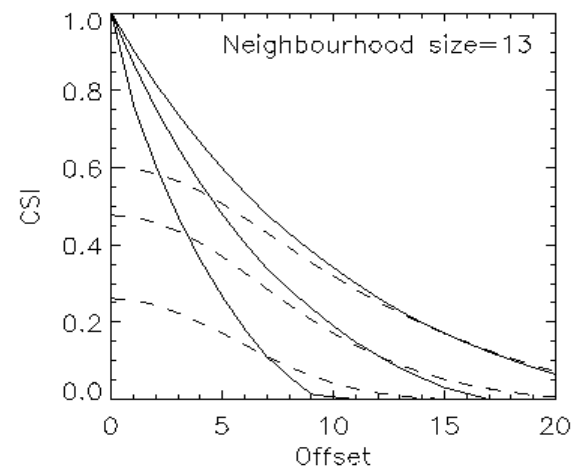
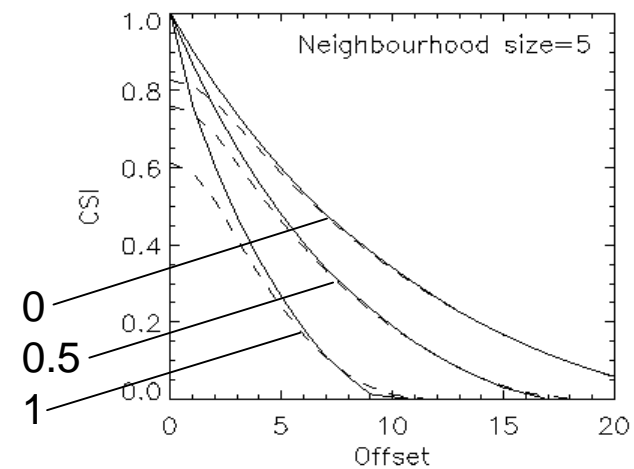
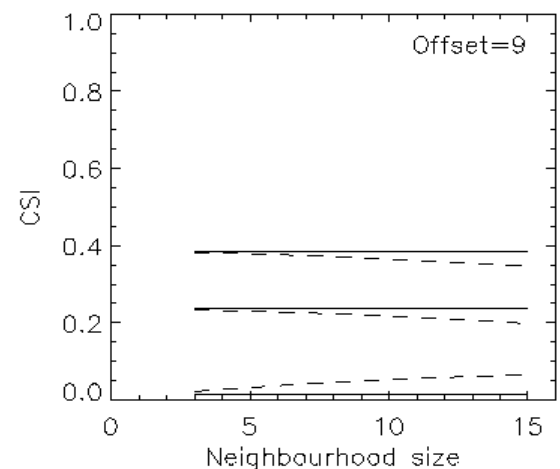
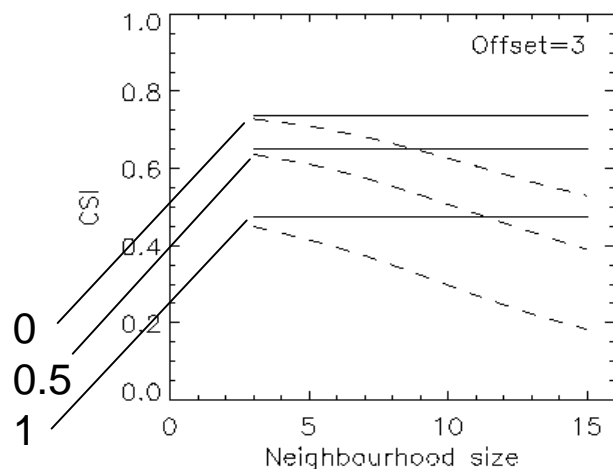
Idealized experiment



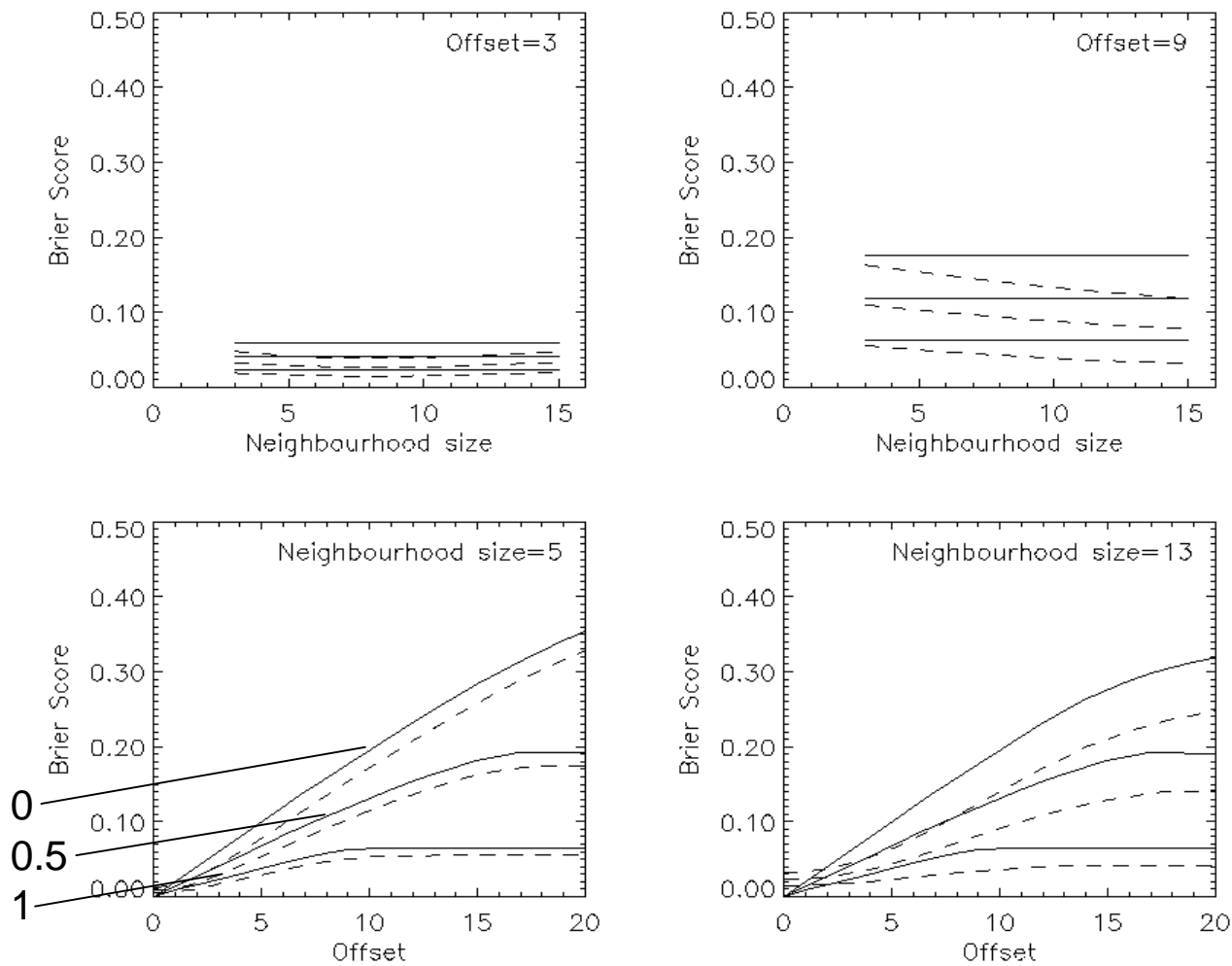
Deterministic forecast values, uncertain observations with PDFs derived from values of near neighbors. Accumulate statistics over the domain.

Compare fuzzy verification results to traditional results when uncertainty and offset are allowed to vary.

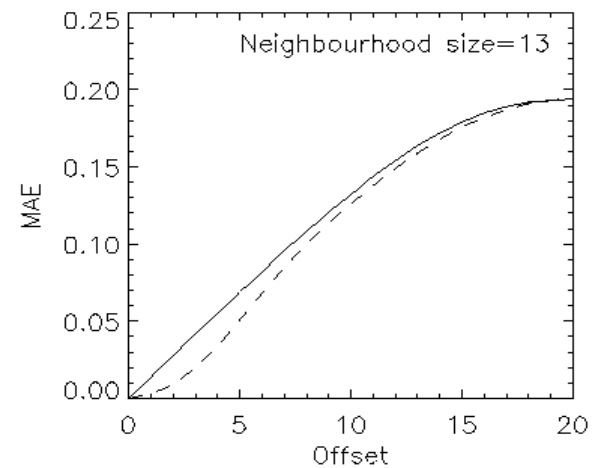
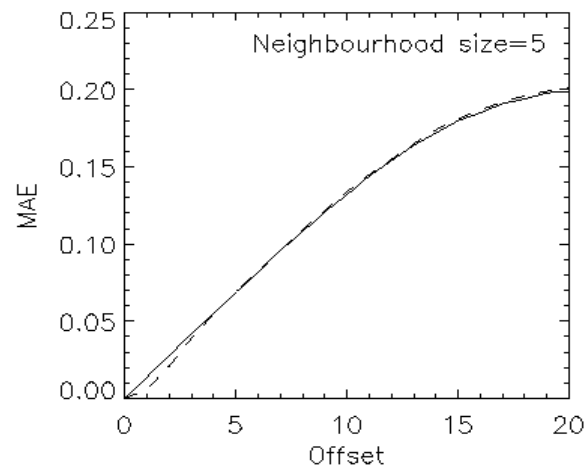
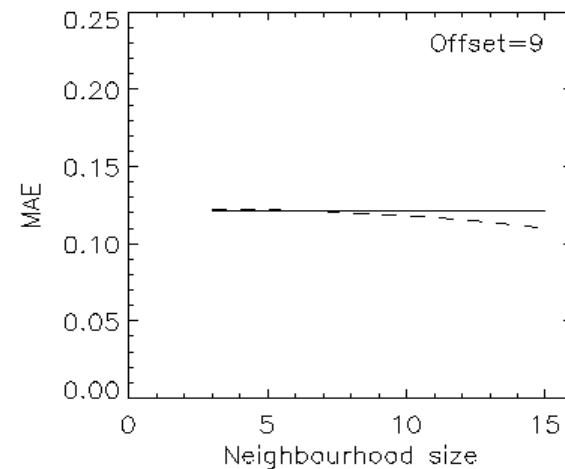
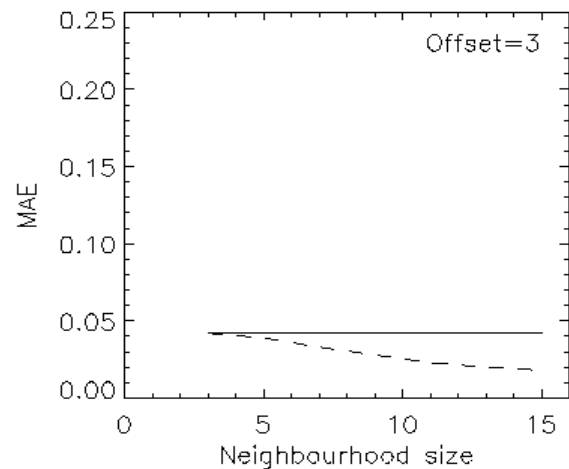
CSI for thresholds of 0, 0.5, and 1 as a function of forecast offset and observational uncertainty (neighborhood size) using traditional (solid) and fuzzy (dashed) verification.



Brier score for thresholds of 0, 0.5, and 1 as a function of forecast offset and observational uncertainty (neighborhood size) using traditional (solid) and fuzzy (dashed) verification.



Weighted MAE as a function of forecast offset and observational uncertainty (neighborhood size) using traditional (solid) and fuzzy (dashed) verification.



Results from this simple experiment show that:

- CSI using fuzzy verification was usually lower than using traditional point-to-point statistics. This means that the negative effects of correct forecasts having some probability of being incorrect outweighed the positive effects of wrong forecasts being partly correct.
- Brier scores were noticeably lower (better) when uncertainty in the observations was taken into account using fuzzy verification.
- Weighted MAE decreased with increasing uncertainty, and was usually smaller than the unweighted MAE.

Conclusions

The verification community has been talking for some time about the need for "probabilistic observations" to reflect the uncertainty in the verification data itself.

Fuzzy verification represents a step in this direction by treating individual observed and/or forecast values as probability density functions (PDFs), rather than discrete values. Familiar verification statistics can be interpreted in this framework to give partial credit where traditional views gave only 0's and 1's.

Future work

- Experiments using **actual data** are needed to test the validity and usefulness of these ideas.
- Fuzzy verification can be extended to other verification techniques such as **signal detection theory** and **error decomposition**.
- A challenging task will be to get **good estimates of the PDFs** associated with observations.