"APPLICATION OF THE GANDIN-MURPHY EQUITABLE SKILL SCORE TO NUMERICAL FORECASTS OF QUANTITATIVE PRECIPITATION"

JOSEPH P. GERRITY, JR.
DEVELOPMENT DIVISION

JUNE 1991

THIS IS AN UNREVIEWED MANUSCRIPT, PRIMARILY INTENDED FOR INFORMAL EXCHANGE OF INFORMATION AMONG NMC STAFF MEMBERS
Q1 is the probability that the event is forecast to occur; and Q2 is the probability that the event is forecast not to occur.

Also use the notations:

\[ Q_1 P_1 = \text{Prob(event forecast and event observed)} \]
\[ Q_1 P_2 = \text{Prob(event forecast and event not observed)} \]
\[ Q_2 P_1 = \text{Prob(event not forecast and event observed)} \]
\[ Q_2 P_2 = \text{Prob(event not forecast and event not observed)} \]

In terms of these probabilities the expected form of the contingency table is:

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Q1*P1</td>
<td>Q2*P1</td>
<td>P1</td>
</tr>
<tr>
<td>Observed No</td>
<td>Q1*P2</td>
<td>Q2*P2</td>
<td>P2</td>
</tr>
<tr>
<td>Sum</td>
<td>Q1</td>
<td>Q2</td>
<td>1.</td>
</tr>
</tbody>
</table>

A "random" prediction is one in which no correlation exists between forecast and observation; the forecast and observation are statistically independent. In such a forecast process, one may write:

\[ P_1 Q_1 = (P_1)(Q_1), \quad P_2 Q_2 = (P_2)(Q_2), \quad P_1 Q_1 = (P_1)(Q_1) \text{ and } P_2 Q_2 = (P_2)(Q_2) \]

The entries in the contingency table can be displayed as an array \( C \),

\[
\begin{pmatrix}
  c_{11} & c_{12} \\
  c_{21} & c_{22}
\end{pmatrix}
\]

where

\[ c_{11} = P_1 Q_1 \quad c_{12} = P_1 Q_2 \]
\[ c_{21} = P_2 Q_1 \quad c_{22} = P_2 Q_2 \]

and then the expected value of the score \( S \) may be expressed as

\[ E(S) = \text{Trace}(W^T C) \]

The equitable score was designed so that its "expected value" \( E(S) \) vanishes for the three "zero skill" forecasts:

#1. \( Q_1 = 1 \); always forecast Yes
#2. \( Q_2 = 1 \); always forecast No
#3. Random prediction

and has the value unity for a perfect forecast,

#4. \( H_1 = 0 \) and \( H_2 = N_0 \)
These criteria impose the conditions expressed by the equations,

\[ w_{11} P_1 + w_{21} (1 - P_1) = 0 \quad (2a) \]
\[ w_{12} P_1 + w_{22} (1 - P_1) = 0 \quad (2b) \]
\[ Q_1 (w_{11} P_1 + w_{21} (1 - P_1)) + (1 - Q_1) (w_{12} P_1 + w_{22} (1 - P_1)) = 0 \quad (2c) \]
\[ w_{11} P_1 + w_{22} (1 - P_1) = 1 \quad (2d) \]

for all \( P_1 \) and \( Q_1 \).

As shown by Gandin and Murphy, the conditions are satisfied when:

\[ w_{11} = \frac{(1 - P_1)}{P_1} \quad (3a) \]
\[ w_{22} = \frac{P_1}{1 - P_1} \quad (3b) \]
\[ w_{12} = w_{21} = -1 \quad (3c) \]

The vanishing of the expected value of \( S \) is not affected by a forecasting strategy intended to give preference to either the occurrence or non-occurrence of the event. For any random forecasting system, irrespective of the "bias" (ratio of \( Q_1 \) to \( P_1 \)), the random forecast gives the equitable score

\[ S_R = (Q_1 P_1) (P_2 / P_1) + (Q_2 P_2) (P_1 / P_2) - (Q_1 P_2) - (Q_2 P_1) = 0 \quad (4) \]

provided the weights (eq. 3) are computed using the probabilities \((P_1, P_2)\) reflected in the sample underlying the contingency table.

On the other hand, the value of the threat score computed for the random forecast contingency table gives,

\[ T S_R = \frac{(Q_1)(P_1)}{Q_1 + P_1 - (Q_1)(P_1)} \quad (5) \]

Introducing the bias \((B = Q_1 / P_1,)\) the threat score for a random forecast is,

\[ T S_R = \frac{B P_1}{(1 + B) - B P_1} \quad (6) \]
which does not vanish. This implies (at least implicitly) that a
random forecast possesses skill; further, this "skill" can be
enhanced by using a biased estimate of the probability of occur-
rence of the event. As a consequence, both the Bias and the
threat score need to be considered when assessing the performance
of a forecaster or a forecasting system.

As a further clarification, recall that the day-to-day eval-
uation of the threat score can yield significant variations due
to the fluctuation of O/N about its expected value P1. This can
be seen in calculations of the correlation between the threat
score and O/N, and has long been known to favor the threat score
achieved by subjective forecasters who have the good fortune to
work during periods of above normal occurrence of precipitation.
This property of the threat score is compounded by the influence
of a biased estimate of P1 on the score's non-zero, expected
value for a random forecast.

If the equitable score is computed for a sample of fore-
casts, produced by a forecaster or forecasting system, it is
important that the probabilities used in defining the weights be
representative of the sample. Alternatively, the samples be suf-
ciently large that climatological probabilities are appropri-
ate. Observing this precaution, it is to be anticipated that
biased forecasting strategies will be ineffectual.

Gandin and Murphy outlined a comprehensive, but complex,
method for determining the equitable score weights for multiple
class contingency tables. A relatively simple method for gener-
ating the weights was found and applied to three and four class
tables (sections 2 and 3, respectively.)

The results of computation of equitable score and other
statistics for a set of May 1991 precipitation forecasts by the
ETA model (Mesinger, Janjic and Black, 1990) are contained in
section 4. A comparison between quantitative precipitation fore-
casts (June 1991) by the Nested Grid Model and the ETA model is
made in section 5. Additional remarks on the equitable score
appear in section 6; the paper ends with a summary and statement
of some expectations for future work.
In this section, an equitable skill score is constructed for a contingency table with three classes of the event. The selected classes exhaust the set of possible outcomes and are ordinally related. Consider for concreteness the prediction of quantitative precipitation in one of three classes: Light, Moderate, Heavy. Suppose the contingency table has the entries:

<table>
<thead>
<tr>
<th>Forecast</th>
<th>light</th>
<th>moderate</th>
<th>heavy</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>light</td>
<td>n11</td>
<td>n12</td>
<td>n13</td>
</tr>
<tr>
<td></td>
<td>moderate</td>
<td>n21</td>
<td>n22</td>
<td>n23</td>
</tr>
<tr>
<td></td>
<td>heavy</td>
<td>n31</td>
<td>n32</td>
<td>n33</td>
</tr>
<tr>
<td></td>
<td>sum</td>
<td>q1</td>
<td>q2</td>
<td>q3</td>
</tr>
</tbody>
</table>

To take advantage of Gandin and Murphy's two class equitable score weight matrix, partition the three-class contingency table into two, two-class tables:

Table 1.

<table>
<thead>
<tr>
<th>Forecast</th>
<th>light or moderate or heavy</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>light</td>
<td>n11</td>
</tr>
<tr>
<td></td>
<td>moderate or heavy</td>
<td>(n21+n31)</td>
</tr>
<tr>
<td></td>
<td>sum</td>
<td>q1</td>
</tr>
</tbody>
</table>

Table 2.

<table>
<thead>
<tr>
<th>Forecast</th>
<th>light or moderate</th>
<th>heavy</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>light or moderate</td>
<td>n11+n12+n21+n22</td>
<td>(n13+n23)</td>
</tr>
<tr>
<td></td>
<td>heavy</td>
<td>(n31+n32)</td>
<td>n33</td>
</tr>
<tr>
<td></td>
<td>sum</td>
<td>(q1+q2)</td>
<td>q3</td>
</tr>
</tbody>
</table>

Using the definitions:

\[
P_1 = E\left(\frac{p_1}{N}\right), P_2 = E\left(\frac{p_2}{N}\right), P_3 = E\left(\frac{p_3}{N}\right) \quad (7a)
\]

\[
Q_1 = E\left(\frac{q_1}{N}\right), Q_2 = E\left(\frac{q_2}{N}\right), Q_3 = E\left(\frac{q_3}{N}\right) \quad (7b)
\]

the equitable weight matrix associated with 2-class Table 1 is

\[
\begin{pmatrix}
   w_{11}^{[1]} & w_{12}^{[1]} \\
   w_{21}^{[1]} & w_{22}^{[1]}
\end{pmatrix}
\]
with

\[ w_{11}^{[1]} = \frac{(P2 + P3)}{P1} \quad (8a) \]
\[ w_{12}^{[1]} = \frac{P1}{(P2 + P3)} \quad (8b) \]
\[ w_{12}^{[1]} = w_{21}^{[1]} = -1 \quad (8c) \]

and that associated with Table 2 is

\[
\begin{pmatrix}
  w_{11}^{[2]} & w_{12}^{[2]} \\
  w_{21}^{[2]} & w_{22}^{[2]}
\end{pmatrix}
\]

with

\[ w_{11}^{[2]} = \frac{P3}{(P1 + P2)} \quad (9a) \]
\[ w_{22}^{[2]} = \frac{(P1 + P2)}{P3} \quad (9b) \]
\[ w_{12}^{[2]} = w_{21}^{[2]} = -1 \quad (9c) \]

Evaluating the equitable scores for tables 1 and 2, and denoting them by \( S1 \) and \( S2 \), yields:

\[ S1 = n_{11}w_{11}^{[1]} + (n_{22} + n_{23} + n_{32} + n_{33})w_{12}^{[1]} + (n_{12} + n_{13})w_{11}^{[1]} + (n_{21} + n_{31})w_{21}^{[1]} \]

\[ S2 = n_{33}w_{33}^{[2]} + (n_{11} + n_{12} + n_{21} + n_{22})w_{12}^{[2]} + (n_{13} + n_{23})w_{11}^{[2]} + (n_{32} + n_{31})w_{21}^{[2]} \]

In terms of \( S1 \) and \( S2 \), an equitable score \( S \), for the three class table, may be written

\[ S = \alpha S1 + (1 - \alpha) S2, \quad (10) \]

where \( \alpha \) is arbitrary, but chosen to be \(.5 \) for specificity. The weight matrix implied by this definition of an equitable score may be worked out by noting the multipliers of each element of the three class table arising in the equation defining the scores \( S1 \) and \( S2 \). The result is the weight matrix

\[
\begin{pmatrix}
  \tilde{w}_{11} & \tilde{w}_{12} & \tilde{w}_{13} \\
  \tilde{w}_{21} & \tilde{w}_{22} & \tilde{w}_{23} \\
  \tilde{w}_{31} & \tilde{w}_{32} & \tilde{w}_{33}
\end{pmatrix}
\]

in which

\[ \tilde{w}_{11} = \frac{1}{2}(w_{11}^{[1]} + w_{11}^{[2]}) \quad (12a) \]
\[ \tilde{w}_{12} = \tilde{w}_{21} = \frac{1}{2}(w_{12}^{[1]} + w_{12}^{[2]}) \]
\[ \tilde{w}_{22} = \frac{1}{2}(w_{22}^{[1]} + w_{22}^{[2]}) \]
\[ \tilde{w}_{33} = \frac{1}{2}(w_{33}^{[1]} + w_{33}^{[2]}) \]

\[ \tilde{w}_{23} = \frac{1}{2}(w_{23}^{[1]} + w_{23}^{[2]}) \]
\[ \tilde{w}_{32} = \frac{1}{2}(w_{32}^{[1]} + w_{32}^{[2]}) \]
\[ \tilde{w}_{31} = \frac{1}{2}(w_{31}^{[1]} + w_{31}^{[2]}) \]
\[ \hat{\omega}_{22} = \frac{1}{2}(w_{12}^{[1]} + w_{12}^{[2]}) \]
\[ \hat{\omega}_{33} = \frac{1}{2}(w_{22}^{[1]} + w_{22}^{[2]}) \]
\[ \hat{\omega}_{13} = \hat{\omega}_{31} = \frac{1}{2}(w_{12}^{[1]} + w_{12}^{[2]}) \] (12b)

\[ \hat{\omega}_{23} = \hat{\omega}_{32} = \frac{1}{2}(w_{22}^{[1]} + w_{22}^{[2]}) \] (12c)

or in terms of the probabilities:

\[ \hat{\omega}_{11} = \frac{(P2 + P3)}{2(P1 + P2)} + \frac{(P3)}{2(P1 + P2)} \]
\[ \hat{\omega}_{12} = \hat{\omega}_{21} = -\frac{1}{2} + \frac{(P3)}{2(P1 + P2)} \] (13a)

\[ \hat{\omega}_{22} = \frac{(P1)}{2(P2 + P3)} + \frac{(P3)}{2(P1 + P2)} \]
\[ \hat{\omega}_{23} = \hat{\omega}_{32} = -\frac{1}{2} + \frac{(P1)}{2(P2 + P3)} \] (13b)

\[ \hat{\omega}_{33} = \frac{(P1 + P2)}{2(P2 + P3)} + \frac{(P1)}{2(P2 + P3)} \]
\[ \hat{\omega}_{13} = \hat{\omega}_{31} = -1 \] (13c)

To prove that the score \( S \) is equitable, it is necessary to show that it satisfies the following conditions:

1. \( S = 0 \); if \( Q1=1 \), or \( Q2=1 \), or \( Q3=1 \).

The contingency table for \( Q1=1 \) is:

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>P1</td>
<td>0</td>
<td>0</td>
<td>P1</td>
</tr>
<tr>
<td>Observed</td>
<td>C2</td>
<td>P2</td>
<td>0</td>
<td>P2</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>P3</td>
<td>0</td>
<td>P3</td>
</tr>
<tr>
<td>Sum</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

The contingency table for \( Q2=1 \) is:

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>0</td>
<td>P1</td>
<td>0</td>
<td>P1</td>
</tr>
<tr>
<td>Observed</td>
<td>C2</td>
<td>0</td>
<td>P2</td>
<td>P2</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>0</td>
<td>P3</td>
<td>P3</td>
</tr>
<tr>
<td>Sum</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

The contingency table for \( Q3=1 \) is:

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>0</td>
<td>0</td>
<td>P1</td>
<td>P1</td>
</tr>
<tr>
<td>Observed</td>
<td>C2</td>
<td>0</td>
<td>0</td>
<td>P2</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>0</td>
<td>P3</td>
<td>P3</td>
</tr>
<tr>
<td>Sum</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
2. $S = 0.$ ; if the forecast is randomly generated with arbitrary $Q$'s; in this case the contingency table takes the expected form:

<table>
<thead>
<tr>
<th>Forecast</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Q1</td>
<td>P1</td>
<td>Q2</td>
<td>P1</td>
</tr>
<tr>
<td>C2</td>
<td>Q1</td>
<td>P2</td>
<td>Q2</td>
<td>P2</td>
</tr>
<tr>
<td>C3</td>
<td>Q1</td>
<td>P3</td>
<td>Q2</td>
<td>P3</td>
</tr>
<tr>
<td>Sum</td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
<td>1.</td>
</tr>
</tbody>
</table>

3. $S = 1.$ ; the forecast is perfect, i.e. the expected three class contingency table has the form:

<table>
<thead>
<tr>
<th>Forecast</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>P1</td>
<td>0.</td>
<td>0.</td>
<td>P1</td>
</tr>
<tr>
<td>C2</td>
<td>0.</td>
<td>P2</td>
<td>0.</td>
<td>P2</td>
</tr>
<tr>
<td>C3</td>
<td>0.</td>
<td>0.</td>
<td>P3</td>
<td>P3</td>
</tr>
<tr>
<td>Sum</td>
<td>P1</td>
<td>P2</td>
<td>P3</td>
<td>1.</td>
</tr>
</tbody>
</table>

For the case in which event 1 is always forecast, i.e. $Q_1 = 1,$ the score $S$ is

$$S = P_1 \omega_{11} + P_2 \omega_{21} + P_3 \omega_{31}$$

$$= .5 \left[ P_1 (w_{11}^{[1]} + w_{11}^{[2]}) + P_2 (w_{21}^{[1]} + w_{21}^{[2]}) + P_3 (w_{31}^{[1]} + w_{31}^{[2]}) \right]$$

$$= .5 \left[ (P_1 w_{11}^{[1]} + (P_2 + P_3) w_{21}^{[1]}) + ((P_1 + P_2) w_{11}^{[2]} + P_3 w_{21}^{[2]}) \right]$$

Since $S_1$ and $S_2$ are equitable scores, it follows that the two factors enclosed in curly braces each vanish. Hence, $S$ also vanishes when $Q_1 = 1.$, as required for $S$ to be an equitable score. In the interest of space, we omit the similar proofs for $Q_2 = 1$ and $Q_3 = 1.$

The satisfaction of the condition 2. (random forecast) follows directly from the satisfaction of the three parts of condition 1. This is seen when the equitable score $S$, using the random forecast contingency table, is written:

$$S = Q_1 \left( P_1 \omega_{11} + P_2 \omega_{21} + P_3 \omega_{31} \right)$$

$$+ Q_2 \left( P_1 \omega_{12} + P_2 \omega_{22} + P_3 \omega_{32} \right)$$

$$+ Q_3 \left( P_1 \omega_{13} + P_2 \omega_{23} + P_3 \omega_{33} \right)$$

(15)
The coefficient of $Q_1$ in equation (15) is the combination of terms which was shown to vanish in equation (14). The coefficients of $Q_2$ and $Q_3$ are similarly the combination of terms which must vanish for the two cases, $Q_2=1$, $Q_3=1$.

Finally, the proof that $S$ has the value unity for a perfect forecast follows:

$$S = (P_1w_{11} + P_2w_{22} + P_3w_{33})$$

$$= S\left\{P_1(w^{[1]}_{11} + w^{[2]}_{11}) + P_2(w^{[1]}_{22} + w^{[2]}_{22}) + P_3(w^{[1]}_{33} + w^{[2]}_{33})\right\}$$

(16)

From the "equitability" of $S_1$ one has,

$$S_1 = P_1w^{[1]}_{11} + (P_2 + P_3)w^{[1]}_{22} = 1$$

(17)

and from the "equitability" of $S_2$ one has,

$$S_2 = (P_1 + P_2)w^{[2]}_{11} + P_3w^{[2]}_{22} = 1$$

(18)

These relations when introduced in the expression (16) yield unity for $S$, as was to be demonstrated.

It may be useful to give some examples of the three-class equitable score weighting matrix as determined by the probability of occurrence of each of the three classes - $P_1$, $P_2$, and $P_3$.

#1. $P_1 = P_2 = P_3 = 1/3$

<table>
<thead>
<tr>
<th></th>
<th>1.25</th>
<th>-.25</th>
<th>-1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.25</td>
<td>.50</td>
<td>- .25</td>
<td></td>
</tr>
<tr>
<td>-1.00</td>
<td>-.25</td>
<td>1.25</td>
<td></td>
</tr>
</tbody>
</table>

Notice that #1. gives the scoring matrix in Gandin and Murphy (their eq. 26.)

#2. $P_1=.5$ $P_2=P_3=.25$

<table>
<thead>
<tr>
<th></th>
<th>.67</th>
<th>-.33</th>
<th>-1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.33</td>
<td>.67</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>-1.00</td>
<td>0.00</td>
<td>2.00</td>
<td></td>
</tr>
</tbody>
</table>

It is interesting to note that in #2. no weight (positive or negative) is given for two of the nine possible combinations of forecast and observation.

#3. $P_1=.1$ $P_2=.3$ $P_3=.6$

<table>
<thead>
<tr>
<th></th>
<th>5.25</th>
<th>.25</th>
<th>-1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.25</td>
<td>.81</td>
<td>-.44</td>
<td></td>
</tr>
<tr>
<td>-1.00</td>
<td>-.44</td>
<td>.39</td>
<td></td>
</tr>
</tbody>
</table>
Cases #3. and #4. are examples involving unbalanced choices of the classes in terms of their probability of occurrence. Note that it is possible for a "near miss" of the "rare" event to receive a more positive weight than a "hit" of a relatively much higher probability event. That this is not at odds with the basic concept of an equitable score follows from the discussion in section 4. of Gandin and Murphy.

Example #5. corresponds to a case in which one class of the event is extremely unlikely. It seems somewhat unreasonable to produce such a selection of classifications; nonetheless, the equitable score weights seem reasonable, and are designed so that the forecaster is appropriately rewarded for attempts to forecast the rare event. The reward is not restricted solely to a precise hit, but is also provided for an imperfect, but suggestive, forecast.
The four class table is assumed to involve a mutually exclusive and exhaustive partition of the predicted event into four classes; it is also assumed that the classes - C1, C2, C3, C4 - bear an ordered arrangement. For application to quantitative precipitation forecasting these conditions are natural, for example:

\[
\begin{align*}
\text{C1} & \quad \text{precip} < .01 \\
\text{C2} & \quad .01 \leq \text{precip} < .25 \\
\text{C3} & \quad .25 \leq \text{precip} < .50 \\
\text{C4} & \quad \text{precip} \geq .50
\end{align*}
\]

A four-class contingency table will have the form:

<table>
<thead>
<tr>
<th>Forecast</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>C1</td>
<td>n11</td>
<td>n12</td>
<td>n13</td>
<td>n14</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>n21</td>
<td>n22</td>
<td>n23</td>
<td>n24</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>n31</td>
<td>n32</td>
<td>n33</td>
<td>n34</td>
</tr>
<tr>
<td></td>
<td>C4</td>
<td>n41</td>
<td>n42</td>
<td>n43</td>
<td>n44</td>
</tr>
<tr>
<td>sum</td>
<td>q1</td>
<td>q2</td>
<td>q3</td>
<td>q4</td>
<td>N</td>
</tr>
</tbody>
</table>

Derivation of an equitable skill score weight matrix proceeds as in section 3. First, the four-class table is partitioned into three, three-class tables:

**Table 1**

<table>
<thead>
<tr>
<th>Forecast</th>
<th>C1</th>
<th>C2</th>
<th>(C3 or C4)</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>C1</td>
<td>n11</td>
<td>n12</td>
<td>(n13+n14)</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>n21</td>
<td>n22</td>
<td>(n23+n24)</td>
</tr>
<tr>
<td></td>
<td>(C3 or C4)</td>
<td>(n31+n41)</td>
<td>(n32+n42)</td>
<td>(n33+n43+n34+n44)</td>
</tr>
<tr>
<td>sum</td>
<td>q1</td>
<td>q2</td>
<td>(q3+q4)</td>
<td>N</td>
</tr>
</tbody>
</table>

**Table 2**

<table>
<thead>
<tr>
<th>Forecast</th>
<th>(C1 or C2)</th>
<th>C4</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>(C1 or C2)</td>
<td>(n11+n12+n21+n22)</td>
<td>(n13+n23)</td>
</tr>
<tr>
<td></td>
<td>(C2 or C3)</td>
<td>(n21+n31)</td>
<td>(n22+n32+23+n33)</td>
</tr>
<tr>
<td></td>
<td>C4</td>
<td>n41</td>
<td>(n42+n43)</td>
</tr>
<tr>
<td>sum</td>
<td>q1</td>
<td>(q2+q3)</td>
<td>q4</td>
</tr>
</tbody>
</table>

**Table 3**

<table>
<thead>
<tr>
<th>Forecast</th>
<th>(C1 or C2)</th>
<th>C4</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>(C1 or C2)</td>
<td>(n11+n13+n21+n22)</td>
<td>(n14+n23)</td>
</tr>
<tr>
<td></td>
<td>(C3 or C4)</td>
<td>(n31+n41)</td>
<td>(n32+n42)</td>
</tr>
<tr>
<td></td>
<td>C4</td>
<td>n43</td>
<td>n44</td>
</tr>
<tr>
<td>sum</td>
<td>(q1+q2)</td>
<td>q3</td>
<td>q4</td>
</tr>
</tbody>
</table>
Define the probabilities:

\[ P_1 = E\left(\frac{p_1}{N}\right), P_2 = E\left(\frac{p_2}{N}\right), P_3 = E\left(\frac{p_3}{N}\right), P_4 = E\left(\frac{p_4}{N}\right) \]

\[ Q_1 = E\left(\frac{q_1}{N}\right), Q_2 = E\left(\frac{q_2}{N}\right), Q_3 = E\left(\frac{q_3}{N}\right), Q_4 = E\left(\frac{q_4}{N}\right) \]

where \( E\{ \} \) means the expected value.

Using the results in section 2, equitable skill scores, \( S_1, S_2 \) and \( S_3 \), may be written for each of the three, three class tables. By setting the equitable score for the four-class table equal to the arithmetic average of \( S_1, S_2 \) and \( S_3 \), the elements of the four-class weight matrix,

\[
\begin{pmatrix}
\omega_{11} & \omega_{12} & \omega_{13} & \omega_{14} \\
\omega_{12} & \omega_{22} & \omega_{23} & \omega_{24} \\
\omega_{13} & \omega_{23} & \omega_{33} & \omega_{34} \\
\omega_{14} & \omega_{24} & \omega_{34} & \omega_{44}
\end{pmatrix}
\]

may be determined, in terms of the set of weights of the three-class tables,

\[
\begin{align*}
\omega_{11} &= \frac{(w_{11}^{[1]} + w_{12}^{[1]} + w_{13}^{[1]})}{3} \\
\omega_{33} &= \frac{(w_{22}^{[1]} + w_{22}^{[2]} + w_{22}^{[3]})}{3} \\
\omega_{12} &= \frac{(w_{11}^{[1]} + w_{12}^{[1]} + w_{12}^{[2]})}{3} \\
\omega_{14} &= \frac{(w_{13}^{[1]} + w_{13}^{[2]} + w_{13}^{[3]})}{3} \\
\omega_{24} &= \frac{(w_{13}^{[2]} + w_{23}^{[2]} + w_{23}^{[3]})}{3}
\end{align*}
\]

or, in terms of the probabilities, \( p_1, p_2, p_3 \) and \( p_4 \):

\[
\begin{align*}
\omega_{11} &= \frac{1}{3} \left( \frac{1 - p_1}{p_1} + \frac{1 - (p_1 + p_2)}{p_1 + p_2} + \frac{p_4}{1 - p_4} \right) \\
\omega_{22} &= \frac{1}{3} \left( \frac{p_1}{1 - p_1} + \frac{1 - (p_1 + p_2)}{p_1 + p_2} + \frac{p_4}{1 - p_4} \right) \\
\omega_{33} &= \frac{1}{3} \left( \frac{p_1}{1 - p_1} + \frac{1 - (p_3 + p_4)}{p_3 + p_4} + \frac{p_4}{1 - p_4} \right)
\end{align*}
\]
\[ \omega_{44} = \frac{1}{3} \left( \frac{P1}{1-P1} + \frac{1-(P3+P4)}{(P3+P4)} + \frac{1-P4}{P4} \right) \]

\[ \omega_{12} = \omega_{21} = \frac{1}{3} \left( \frac{1}{1-P1} + \frac{1-(P1+P2)}{P1+P2} + \frac{P4}{1-P4} \right) \]

\[ \omega_{13} = \omega_{31} = \frac{1}{3} \left( -2 + \frac{P4}{1-P4} \right) \]

\[ \omega_{14} = \omega_{41} = -1. \]

\[ \omega_{23} = \omega_{32} = \frac{1}{3} \left( \frac{P1}{1-P1} + \frac{P4}{1-P4} -1. \right) \]

\[ \omega_{24} = \omega_{42} = \frac{1}{3} \left( \frac{P1}{1-P1} -2. \right) \]

\[ \omega_{34} = \omega_{43} = \frac{1}{3} \left( \frac{P1}{1-P1} + \frac{P1+P2}{(P3+P4)} -1. \right) \]

The proof of the equitability of this four class weight matrix follows the same method used for the three class weight matrix in section 2, and doesn't warrant the use of additional space here. (A BASIC language program is given in an appendix which the reader may use to satisfy any misgivings on this point.)

Some examples are given below for equitable weights for four class contingency tables with the indicated probabilities of occurrence for each class: P1, P2, P3 and P4.

#1. P1=P2=P3=P4=.25:

\[
\begin{array}{cccc}
1.44 & 0.11 & -0.56 & -1.00 \\
0.11 & 0.56 & -0.11 & -0.56 \\
-0.56 & -0.11 & 0.56 & 0.11 \\
-1.00 & -0.56 & 0.11 & 1.44 \\
\end{array}
\]

#2. P1=P4=.10, P2=P3=.40:

\[
\begin{array}{cccc}
3.37 & 0.04 & -0.63 & -1.00 \\
0.04 & 0.41 & -0.26 & -0.63 \\
-0.63 & -0.26 & 0.41 & 0.04 \\
-1.00 & -0.63 & 0.04 & 3.37 \\
\end{array}
\]

#3. P1=.10, P2=.20, P3=.30, P4=.40:

\[
\begin{array}{cccc}
4.00 & 0.67 & -0.44 & -1.00 \\
\end{array}
\]
Examples #4 and #5 are given because they illustrate how the four class scoring matrix "approaches" that of its three class counterpart when one of the four classes has a probability of occurring that approaches zero.
APPLICATION TO ETA FORECASTS

The sample of ETA model forecasts of 24 hour accumulation of precipitation was produced during the period April 28 to May 31, 1991. Only the first period forecasts, hours 0 to 24, and only forecasts initiated at 1200 GMT were used. The verification data were the observations made by both cooperative and regular components of the NWS River Forecast Network. The observations cover the day, beginning and ending at 1200 GMT. Unfortunately, only observations of occurrence of precipitation are reported by the cooperative component of the network; non-occurrence is not reported.

The observed data are applied to the forecast model's gridpoints, by assigning to the gridpoint the arithmetic mean of the observations lying within the cell surrounding the gridpoint. The ETA model uses a grid which has separation interval of approximately 80 km. Only gridpoints where there is a possibility of an observation being reported are included in the verification; there are 1060 such points over the contiguous United States.

Four classes were used in the verification,

\[
\begin{align*}
C1 : & \quad R < 0.01" \\
C2 : & \quad 0.01 \leq R < 0.50" \\
C3 : & \quad 0.50" \leq R < 1.00" \\
C4 : & \quad 1.00" \leq R
\end{align*}
\]

where \( R \) stands for the amount of precipitation, forecast or observed. The collection of data was missed on three days; so the sample involves just 29 days.

Note that the four classes are partitioned by three thresholds - 0.01", 0.50" and 1.00". It can be shown (section 6) that the equitable score for the four class contingency table may be constructed from the arithmetic average of the equitable scores for each of three two class tables, one for each of the thresholds. The data for each day and each threshold are given in appendix II.

The contingency tables constructed from the average of the daily entries in each cell of the table are:

<table>
<thead>
<tr>
<th></th>
<th>FORECAST</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>239.5</td>
<td>155</td>
</tr>
<tr>
<td>OBSERVED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO</td>
<td>142.5</td>
<td>523</td>
</tr>
<tr>
<td>SUM</td>
<td>382</td>
<td>678</td>
</tr>
</tbody>
</table>

Contingency table for .01" threshold
<table>
<thead>
<tr>
<th>FORECAST</th>
<th>YES</th>
<th>NO</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>35.</td>
<td>52.4</td>
<td>87.4</td>
</tr>
<tr>
<td>OBSERVED</td>
<td>45.8</td>
<td>926.8</td>
<td>972.6</td>
</tr>
<tr>
<td>SUM</td>
<td>80.8</td>
<td>979.2</td>
<td>1060.</td>
</tr>
</tbody>
</table>

Contingency table for .50" threshold

<table>
<thead>
<tr>
<th>FORECAST</th>
<th>YES</th>
<th>NO</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>7.8</td>
<td>18.9</td>
<td>26.7</td>
</tr>
<tr>
<td>OBSERVED</td>
<td>18.0</td>
<td>1015.3</td>
<td>1033.3</td>
</tr>
<tr>
<td>SUM</td>
<td>25.8</td>
<td>1034.2</td>
<td>1060.</td>
</tr>
</tbody>
</table>

Contingency table for 1.00" threshold

For the .01" threshold the probability of occurrence is .372, so the equitable score weights are 1.69 and .59 for the successful forecasts of occurrence and non-occurrence, respectively. The resulting equitable score is .39.

For the 0.5" threshold the probability of occurrence is .082, so the equitable score weights are 11.1 and .09 for the successful forecasts of occurrence and non-occurrence, respectively. The resulting equitable score is .35.

For the 1.0" threshold the probability of occurrence is .025, so the equitable score weights are 38.7 and .025 for the successful forecasts of occurrence and non-occurrence, respectively. The resulting equitable score is .27.

Thus the equitable score for the full four class table is 0.34.

Other statistics for this data set are:

| Threshold | TS | Bias | P(F|O) | P(O|F) | GTS | TSS |
|-----------|----|------|------|-------|-----|-----|
| .01"      | .44 | .97  | .61  | .63   | .14 | .29 |
| .50"      | .26 | .92  | .40  | .43   | .20 | .23 |
| 1.00"     | .17 | .96  | .29  | .30   | .16 | .16 |
| Average   | .29 | .95  | .43  | .45   | .17 | .23 |

where in addition to the previously defined threat score (TS) and bias: P(F|O) is the conditional probability that the event was forecast given that it was observed; P(O|F) is the conditional probability that the event was observed given that it was forecast; GTS is the Gilbert threat score (Schaefer, 1990) defined by,

\[ GTS = \frac{(H1 - E\{H1\})}{F + O - (H1 - E\{H1\})} \]
in which \( E\{H1\} \) is the number of successful forecasts of the occurrence of the event to be expected due to chance. The calculation of \( E\{H1\} \) is done using the formula,

\[
E\{H1\} = \frac{F \cdot O}{N},
\]

and TSS is the skill score defined in terms of the threat score by the relation,

\[
TSS = \frac{TS - TS_r}{1 - TS_r},
\]

where

\[
TS_r = \frac{E\{H1\}}{F + O - E\{H1\}}
\]

is the threat score for a random forecast.

While it is true that inspection of the full four class contingency table is necessary to comprehend the full character of the joint distribution of forecasts and observations, it is desirable to produce a statistic that digests the full set of information and provides an "equitable" assessment of performance. From the statistical data, presented above, the GTS, Skill and Equitable scores seem to possess the desirable property of small, inter+threshold variability, when compared to the TS and the two conditional probabilities. This is so because they take into account the difference in difficulty associated with forecasts at the different thresholds. Scores with this property lend themselves to graphical presentation without the tendency for over-emphasis on the lower thresholds, that is evident in plots of the Threat Score versus threshold.
To gain insight into the applicability of the equitable score to comparison of model forecasts, the equitable score and other statistics were computed for a small sample of quantitative precipitation forecasts made, during June 1991, by two different regional models, the NGM (Phillips, 1976) and the ETA (Messinger, Black and Janjic, 1988.) Because the models use different computational grids, there is a difference in the number of points used in the verification, but the verification area, the contiguous United States, is essentially the same. The source of verification data is the same as outlined in the previous section.

The contingency tables are based on the four classes used in section 4, (C1: R<.01", C2: .01"<=R<.5". C3: .5"<=R<1." and C4: R>1."). The sample consists of sixteen daily forecasts made between 7 June and 23 June (the 22nd is missing.)

Since the sample is so small, it is evident that no conclusions may be drawn from this data regarding the merits of one model versus the other. The two models do have somewhat different characteristics, and it is of interest to learn if the equitable score provides a way to account rationally for these differences.

In presenting the average contingency tables, the entries are given as decimal fractions of the total number of verification points. This avoids the confusion which might arise from the different number of verification points (1261 for the NGM; 1060 for the ETA model.)

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Yes</th>
<th>No</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>.24</td>
<td>.06</td>
<td>.30</td>
</tr>
<tr>
<td>No</td>
<td>.23</td>
<td>.47</td>
<td>.70</td>
</tr>
<tr>
<td>Sum</td>
<td>.47</td>
<td>.53</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Contingency table for .01" NGM forecasts.
<table>
<thead>
<tr>
<th>Observed</th>
<th>Yes</th>
<th>0.18</th>
<th>0.14</th>
<th>0.32</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>0.12</td>
<td>0.56</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>Sum</td>
<td>0.30</td>
<td>0.70</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Contingency table for .01" ETA forecasts.

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Yes</th>
<th>No</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>Yes</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>0.06</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>Sum</td>
<td>0.08</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Contingency table for .5" NGM forecasts.

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Yes</th>
<th>No</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>Yes</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>0.04</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>Sum</td>
<td>0.06</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Contingency table for .5" ETA forecasts.

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Yes</th>
<th>No</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>Yes</td>
<td>0.002</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>0.008</td>
<td>0.972</td>
</tr>
<tr>
<td></td>
<td>Sum</td>
<td>0.010</td>
<td>0.990</td>
</tr>
</tbody>
</table>

Contingency table for 1." NGM forecasts.

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Yes</th>
<th>No</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>Yes</td>
<td>0.003</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>0.020</td>
<td>0.960</td>
</tr>
<tr>
<td></td>
<td>Sum</td>
<td>0.023</td>
<td>0.977</td>
</tr>
</tbody>
</table>

Contingency table for 1." ETA forecasts.

The probability of occurrence of each of the four classes differs slightly between the ETA and NGM data sets. For the ETA model data set, the probabilities of occurrence in the four classes are .68, .25, .05 and .02. For the NGM data set, the probabilities are: .70, .23, .05 and .02.
The equitable score was calculated based on the three two-class tables associated with the thresholds separating the four classes. The resulting four-class equitable score for the NGM is 0.29 and for the ETA model is 0.25.

It was surprising that the ETA model's score for this period in June was so much less than the .34 score found in the May data set, analyzed in section 4. Examining the contributions to the four-class score made by the three thresholds shows that the degradation in score is caused by performance at the larger amounts of precipitation. This is shown in the following table:

<table>
<thead>
<tr>
<th>Threshold</th>
<th>.01&quot;</th>
<th>.5&quot;</th>
<th>1.&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>May</td>
<td>.39</td>
<td>.35</td>
<td>.27</td>
</tr>
<tr>
<td>June</td>
<td>.40</td>
<td>.22</td>
<td>.11</td>
</tr>
</tbody>
</table>

Table Contributions to the ETA model's Equitable score from the three thresholds.

The probability of occurrence of precipitation exceeding each of these thresholds during the two data collections is tabulated below.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>.01&quot;</th>
<th>.5&quot;</th>
<th>1.&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>May</td>
<td>.37</td>
<td>.082</td>
<td>.025</td>
</tr>
<tr>
<td>June</td>
<td>.31</td>
<td>.071</td>
<td>.019</td>
</tr>
</tbody>
</table>

Table Probability of occurrence of precipitation exceeding the three thresholds.

The forecast model's frequency of forecasting precipitation amounts exceeding the several thresholds may also be expressed in terms of probability measures, as in the following table:

<table>
<thead>
<tr>
<th>Threshold</th>
<th>.01&quot;</th>
<th>.5&quot;</th>
<th>1.&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>May</td>
<td>.36</td>
<td>.075</td>
<td>.024</td>
</tr>
<tr>
<td>June</td>
<td>.30</td>
<td>.063</td>
<td>.022</td>
</tr>
</tbody>
</table>

Table Probability of prediction of precipitation exceeding the three thresholds.

It does not seem likely that the small variations in observed and predicted probabilities of occurrence would distort the computations of the equitable score. This leaves the conclusion that the accuracy of the ETA model's forecasts, during the June data period, were not affected by unusual, systematic biases but by other factors, that the precipitation verification statistics cannot elucidate.
Attention may be turned to a comparison of the equitable score with three other statistical estimates of performance. These are the previously defined threat score (TS), bias (B), conditional probabilities (P(F|O) and P(O|F)), Gilbert threat score (GTS), threat skill score (TSS), and the Gandin-Murphy equitable skill score (GMS.) Also tabulated is the threat score for a random forecast (TSR.) Although most of these scores can be calculated day-by-day, only values for the average contingency table will be considered.

| Threshold | TS  | B   | P(F|O) | P(O|F) | GTS  | TSS  | GMS  | TSR  |
|-----------|-----|-----|-------|-------|------|------|------|------|
| .01"      | .42 | .94 | .57   | .61   | .16  | .29  | .40  | .18  |
| .50"      | .16 | .87 | .26   | .30   | .12  | .14  | .22  | .04  |
| 1.00"     | .06 | 1.15| .13   | .11   | .05  | .05  | .11  | .01  |

Table Statistics for ETA model precipitation forecasts at three thresholds for June data set.

| Threshold | TS  | B   | P(F|O) | P(O|F) | GTS  | TSS  | GMS  | TSR  |
|-----------|-----|-----|-------|-------|------|------|------|------|
| .01"      | .44 | 1.57| .79   | .50   | .14  | .28  | .55  | .22  |
| .50"      | .18 | 1.09| .31   | .29   | .13  | .14  | .26  | .04  |
| 1.00"     | .06 | .47 | .08   | .16   | .05  | .05  | .07  | .01  |

Table Statistics for NGM model precipitation forecasts at three thresholds for June data set.

The characteristic difference between the two models' forecasts is most clearly manifested the Bias score and the associated threat score for a random forecast (TSR.) The NGM has a large bias for small amounts and a very small bias for large amounts of precipitation. The ETA model manifests the opposite trend but with a smaller variation.

The equitable score for the four class table judged the NGM performance to be superior the ETA. Yet, when the large contribution due to the heavily biased low threshold is noted, it seems that this judgment is questionable. Indeed, the Gilbert threat score when averaged over the three thresholds, gives the nod to the ETA models. The threat skill score suggests that when the variation in bias is taken into account both models have almost identical skill compared to random forecasts made using the associated model's estimate of the probability of occurrence in each class.

This data set is so small that only extremely tentative conclusions are warranted, but it seems reasonable to conclude that the equitable score will not be a panacea for the problem of
comparing the performance of models possessing significantly different bias characteristics. It also seems appropriate to include the threat skill score in future model comparisons.
It is one of the attractions of the threat score that it can be given a simple "geometric" interpretation in terms of the intersection and union of the forecast and observed data. For the interpretation of the "meaning" of the equitable score, the following analysis is helpful.

For the two class contingency table,

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Yes</th>
<th>No</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>H1</td>
<td>M1</td>
<td>O</td>
</tr>
<tr>
<td>No</td>
<td>M2</td>
<td>H2</td>
<td>N-O</td>
</tr>
<tr>
<td>Sum</td>
<td>F</td>
<td>N-F</td>
<td>N</td>
</tr>
</tbody>
</table>

the equitable score $S$ is written in terms of the entries $H1$, $H2$, $M1$ and $M2$, together with two weights that are defined in terms of the climatological expectancy of $O/N$, which may be denoted by $P$. The expression is,

$$ S = \left( \frac{1-P}{P} \right) \frac{H1}{N} + \left( \frac{P}{1-P} \right) \frac{H2}{N} - \frac{M1}{N} - \frac{M2}{N} $$

If the following notational simplifications are employed,

$$ h_1 = \frac{H1}{O} \quad h_2 = \frac{H2}{N-O} \quad \mu = \frac{O}{N} \quad \nu = \frac{N-O}{N} = 1 - \mu $$

and $M1$ and $M2$ are replaced by $O-H1$ and $N-O-H2$, $S$ may be written as

$$ S = \left( 1 + \frac{1-P}{P} \right) \mu h_1 + \left( 1 + \frac{P}{1-P} \right) \nu h_2 - 1 $$

It is evident that $P$ is the expected value of $\mu$ which can be written $P = E(\mu)$, with the corollary, $(1-P) = E(\nu)$. So the expression for the equitable score can be put into the form,

$$ S = \frac{\mu}{E{\mu}} h_1 + \frac{\nu}{E{\nu}} h_2 - 1. $$
When the data set has the property that \( \mu \) and \( \nu \) are close to their expected values, then the equitable score is approximately given by

\[
S = (h_1 + h_2 - 1)
\]

So provided the individual case does not depart too much from the climatological expectancy of the event's occurrence, the equitable score may be interpreted as the fraction of correct forecasts of the event's occurrence plus the fraction of correct forecasts of the event's non-occurrence minus unity! To optimize this score it is necessary to accurately predict both occurrence and non-occurrence of the event.

In section 4, it was noted that the evaluation of the equitable score for the four-class table reduced to the evaluation of three, two-class table scores -- one for each of the three thresholds present in the four-class table. The explanation for this can be seen in the following schematic:

\[
\begin{array}{cccc}
   & C1 & C2 & C3 & C4 \\
S_1 & S_2 & S_3 \\
1. C1@C2 & C3 & C4 & 2. C1 & C2@C3 & C4 & 3. C1 & C2 & C3@C4 \\
S_{1,1} & S_{2,1} & S_{3,1} \\
1a. C1@C2@C3 & C4 & 2a. C1@C2@C3 & C4 & 3a. C1@C2 & C3@C4 \\
S_{1,2} & S_{2,2} & S_{3,2} \\
1b. C1@C2 & C3@C4 & 2b. C1 & C3@C3@C4 & 3b. C1 & C2@C3@C4 \\
\end{array}
\]

Briefly, the score \( S \) for the four-class table reduces to the average of the three scores, \( S_{1,1}, S_{2,2}, S_{3,3} \), for the three itemized partitions of the four-class table into three-class tables. In turn 1a., 1b. 2a., 2b., 3a., and 3b. are partitions of the three-class tables into two-class tables with the associated scores, \( S_{1,1,1}, S_{1,2,1}, S_{2,2,1}, S_{3,1,1}, \) and \( S_{3,2,1} \).

But the set of six, two-class tables contains only three unique partitions, each of which is associated with one of the three thresholds of the original four-class table. Thus, in evaluating the equitable score \( S \), it suffices to evaluate \( S_{1,1}, S_{1,2}, \) and \( S_{1,3} \), as was done in section 4.
7 SUMMARY AND CONCLUSIONS

The equitable skill score proposed by Gandin and Murphy (1991) for evaluating the performance of categorical forecasts was applied to small samples of numerical model predictions of quantitative precipitation. The application of the score to multiple class contingency tables was found to reduce to the computation of scores for a set of two-class tables, one for each threshold in the multiclass table. This was demonstrated for three and four class tables, and is conjectured, but not proven, to be generally true.

It was expected that the equitable score's theoretical property of insensitivity to biased, random forecasting strategies might prove useful in comparing forecast models that possess different characteristics in the bias of their quantitative precipitation forecasts. In a reported test, this expectation was not completely fulfilled, and the alternative threat skill score statistic seemed to be somewhat better behaved.

It was hoped, perhaps naively, that the application of the equitable score to a multiclass table of quantitative precipitation forecasts and observations would lessen the need to examine scores for several thresholds of precipitation. Experience, gained in tests reported above, suggests that this hope will not be fulfilled. In part, because the computation of the score for the multiclass table can be carried-out considering the set of two-class partitions of the event associated with each of the several thresholds embedded in the multiclass table, it is unlikely that the contributions made to the final score by each partition can be overlooked.

It is expected that future efforts in the verification of precipitation forecasts will use the equitable score. It is of theoretical interest to prove that the computation of an equitable score for a general K-class table can be reduced to computing K-1 two-class table scores, and the corollary that the weight matrix for the K-class table may be built-up from the weights of the K-1 two class tables.
8 REFERENCES


APPENDIX I BASIC PROGRAM

REM BASIC LANGUAGE PROGRAM TO TEST EQUITABILITY OF 4-CLASS WEIGHT MATRIX

PRINT "INPUT P1"
INPUT P1
PRINT "INPUT P2"
INPUT P2
PRINT "INPUT P3"
INPUT P3
P4=1-(P1+P2+P3)
PRINT "PROBABILITIES OF EVENT CLASSES"
PRINT USING "P1=#.## P2=#.## P3=#.## P4=#.##";P1,P2,P3,P4
W11=(1/3)*( (1-P1)/P1 + (P3+P4)/(P1+P2) + P4/(1-P4) )
W22=(1/3)*( P1/(1-P1) + (P3+P4)/(P1+P2) + P4/(1-P4) )
W33=(1/3)*( P1/(1-P1) + (P1+P2)/(P3+P4) + P4/(1-P4) )
W44=(1/3)*( P1/(1-P1) + (1-P3-P4)/(P3+P4) + (1-P4)/P4 )
W12=(1/3)*(-1 + (P3+P4)/(P1+P2) + P4/(1-P4))
W21=W12
W13=(1/3)*(-2 + P4/(1-P4))
W31=W13
W14=(1/3)*(-3)
W23=(1/3)*( P1/(1-P1) + P4/(1-P4) -1)
W32=W23
W24=(1/3)*( P1/(1-P1) -2)
W42=W24
W34=(1/3)*( P1/(1-P1) + (P1+P2)/(P3+P4) -1)
W43=W34
PRINT "WEIGHT MATRIX"
PRINT USING "##.f## #fl.## ##.# ##.##";W11,W12,W13,W14
PRINT USING "##.## ##.## ##.## -##.#";W21,W22,W23,W24
PRINT USING "##.## ##.## ##.## ##.##";W31,W32,W33,W34
PRINT USING "##.# #. ## ##.## ##.##";W41,W42,W43,W44
REM NOW TEST FOR RANDOM CONTINGENCY TABLE
PRINT "INPUT Q1"
INPUT Q1
PRINT "INPUT Q2"
INPUT Q2
PRINT "INPUT Q3"
INPUT Q3
Q4=1-(Q1+Q2+Q3)
PRINT "FCST PROBS Q1=.## Q2=.## Q3=.## Q4=.##"
E11=P1*Q1 : E12=P1*Q2 : E13=P1*Q3 : E14=P1*Q4
E21=P2*Q1 : E22=P2*Q2 : E23=P2*Q3 : E24=P2*Q4
E31=P3*Q1 : E32=P3*Q2 : E33=P3*Q3 : E34=P3*Q4
E41=P4*Q1 : E42=P4*Q2 : E43=P4*Q3 : E44=P4*Q4
REM CALCULATE EQUITABLE SCORES
REM SK FOR ONLY FORECAST CLASS K
REM SR FOR RANDOM TABLE
S1=W11*E11+W21*E21+W31*E31+W41*E41
S2=W12*E12+W22*E22+W32*E32+W42*E42
460  S3=W13*E13+W23*E23+W33*E33+W43*E43
470  S4=W14*E14+W24*E24+W34*E34+W44*E44
480  S= S1+ S2+ S3+ S4
490  PRINT USING "S1=#.## S2=#.## S3=#.## S4=#.##
     S=#.##";S1,S2,S3,S4,S
500  REM TEST FOR PERFECT FORECAST
510  S=W11*P1 + W22*P2 + W33*P3 + W44*P4
520  PRINT USING "PERFECT FORECAST S=##.##";S
530  END
## APPENDIX II DATA FOR SECTION 4

<table>
<thead>
<tr>
<th></th>
<th>.01&quot;</th>
<th>.50&quot;</th>
<th>1.0&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HITS</td>
<td>O</td>
<td>F</td>
</tr>
<tr>
<td>APR28</td>
<td>292</td>
<td>403</td>
<td>502</td>
</tr>
<tr>
<td>APR29</td>
<td>382</td>
<td>494</td>
<td>553</td>
</tr>
<tr>
<td>APR30</td>
<td>309</td>
<td>453</td>
<td>433</td>
</tr>
<tr>
<td>MAY1</td>
<td>170</td>
<td>345</td>
<td>293</td>
</tr>
<tr>
<td>MAY2</td>
<td>341</td>
<td>461</td>
<td>514</td>
</tr>
<tr>
<td>MAY3</td>
<td>313</td>
<td>454</td>
<td>432</td>
</tr>
<tr>
<td>MAY4</td>
<td>336</td>
<td>442</td>
<td>418</td>
</tr>
<tr>
<td>MAY5</td>
<td>169</td>
<td>312</td>
<td>260</td>
</tr>
<tr>
<td>MAY6</td>
<td>173</td>
<td>339</td>
<td>289</td>
</tr>
<tr>
<td>MAY7</td>
<td>215</td>
<td>389</td>
<td>339</td>
</tr>
<tr>
<td>MAY8</td>
<td>183</td>
<td>329</td>
<td>361</td>
</tr>
<tr>
<td>MAY9</td>
<td>187</td>
<td>312</td>
<td>400</td>
</tr>
<tr>
<td>MAY10</td>
<td>160</td>
<td>285</td>
<td>355</td>
</tr>
<tr>
<td>MAY11</td>
<td>223</td>
<td>388</td>
<td>399</td>
</tr>
<tr>
<td>MAY12</td>
<td>159</td>
<td>343</td>
<td>298</td>
</tr>
<tr>
<td>MAY13</td>
<td>191</td>
<td>408</td>
<td>312</td>
</tr>
<tr>
<td>MAY14</td>
<td>229</td>
<td>403</td>
<td>367</td>
</tr>
<tr>
<td>MAY15</td>
<td>327</td>
<td>483</td>
<td>475</td>
</tr>
<tr>
<td>MAY16</td>
<td>321</td>
<td>454</td>
<td>507</td>
</tr>
<tr>
<td>MAY17</td>
<td>242</td>
<td>380</td>
<td>330</td>
</tr>
<tr>
<td>MAY18</td>
<td>237</td>
<td>372</td>
<td>346</td>
</tr>
<tr>
<td>MAY19</td>
<td>211</td>
<td>399</td>
<td>317</td>
</tr>
<tr>
<td>MAY20</td>
<td>211</td>
<td>420</td>
<td>367</td>
</tr>
<tr>
<td>MAY21</td>
<td>253</td>
<td>425</td>
<td>402</td>
</tr>
<tr>
<td>MAY22</td>
<td>262</td>
<td>423</td>
<td>407</td>
</tr>
<tr>
<td>MAY23</td>
<td>207</td>
<td>365</td>
<td>311</td>
</tr>
<tr>
<td>MAY24</td>
<td>217</td>
<td>441</td>
<td>366</td>
</tr>
<tr>
<td>MAY25</td>
<td>248</td>
<td>436</td>
<td>450</td>
</tr>
</tbody>
</table>