RICHER: A DATA ENRICHMENT SUBROUTINE

E. R. Hamerly
University of Washington

ABSTRACT

RICHER is a sophisticated interpolation routine which allows the user to generate data for points he does not have. The following activities are at best marginally legitimate.

A. Regenerate a dataset at a different step size (monthly points instead of bimonthly)
B. Estimate missing data due to equipment failure
C. Estimate state vectors for times that are not increments of the simulation step.

INTRODUCTION

RICHER was written upon request for George Hendry and Dave Culver to allow better estimation of heat budgets in the four lakes study. In exchange for RICHER, the information bank received two additional useful tools:

1. GRID -- a subroutine which squares data to a rectangular grid by estimating the grid value from the 6 closest points
2. ISOPLET -- (author George Hendry) a program to draw temperature isopleths over the annual cycle

DIGDAT, GRID, and ISOPLET form an analysis chain for producing contour maps of lake bottoms from strip charts of transect runs allowing better estimation of the amount of water at different depths in the lakes. RICHER and ISOPLET furnish an analysis for producing more accurate isopleths of the annual temperature regimes at different depths.

THE ALGORITHM

RICHER selects three data points at a time \((\alpha_i, \beta_i), (\alpha_{i-1}, \beta_{i-1}), (\alpha_{i-2}, \beta_{i-2})\), where \(i=3, \ldots, n\) and \(n\) is the number of data points. The three data points being considered will be referred to as \((x_1, y_1), (x_2, y_2),\) and \((x_3, y_3)\). The point \((\alpha_{i-3}, \beta_{i-3})\) is referred to as \((x_0, y_0)\).
RICHER checks that $y_2$ is strictly between $y_1$ and $y_3$. If it is not, then $(x_3, y_3)$ is reset so that a parabola passing through the three points will have its extremum at $(x_2, y_2)$.

$$(x_3, y_3) = (x_1 + 2(x_2 - x_1), y_1) \quad (1)$$

This does not alter $(\alpha_2, \beta_2)$ because $(x_3, y_3)$ are stored in separate working areas. RICHER passes the parabola $y = \sigma_0 + \sigma_1x + \sigma_2x^2$ through the three points. $\sigma_0$, $\sigma_1$, and $\sigma_2$ are calculated from the simultaneous equations

$\sigma_0 + x_1\sigma_1 + x_1^2\sigma_2 = y_1$

$\sigma_0 + x_2\sigma_1 + x_2^2\sigma_2 = y_2 \quad (2)$

$\sigma_0 + x_3\sigma_1 + x_3^2\sigma_2 = y_3$

The parabola passes through $(x_1, y_1)$, $(x_2, y_2)$ and $(x_3, y_3)$.

Let $X$ be the matrix of coefficients for $C$

$$X = \begin{bmatrix} 1, x_1, x_1^2 \\ 1, x_2, x_2^2 \\ 1, x_3, x_3^2 \end{bmatrix}, \quad C = \begin{bmatrix} \sigma_0 \\ \sigma_1 \\ \sigma_2 \end{bmatrix}$$

$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$.

Equation (2) is represented by $XC = Y$. $C$ is found by $C = X^{-1}Y$.

A second parabola $y = b_0 + b_1x + b_2x^2$ is fitted through $(x_1, y_1)$ and $(x_2, y_2)$ with the constraint that $y_1'$ has the same value as $y_2' = \alpha_1 + 2\alpha_2x_2$ where $\alpha_1$ and $\alpha_2$ are the values found for $\sigma_1$ and $\sigma_2$ on points $(x_0, y_0)$, $(x_1, y_1)$, and $(x_2, y_2)$. $(y_1'$ is the slope of the line at $x_1$).

$b_0$, $b_1$, and $b_2$ are calculated from the equations

$$b_0 + x_1b_1 + x_1^2b_2 = y_1$$

$$b_0 + x_2b_1 + x_2^2b_2 = y_3$$

$$b_1 + 2x_1b_2 = \alpha_1 + 2\alpha_1\alpha_2 \quad (3)$$

Here

$$X = \begin{bmatrix} 1, x_1, x_1^2 \\ 1, x_2, x_2^2 \\ 0, 1, 2x_1 \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \alpha_1 + 2x_1\alpha_2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

Equation set (3) becomes $XB = Y$ and $B$ is found by $B = X^{-1}Y$.

The weighted average of the two parabolas is formed

$$y = \frac{[(x_2 - x) (b_0 + b_1x + b_2x^2) + (x - x_1) (\sigma_0 + \sigma_1x + \sigma_2x^2)]}{(x_2 - x_1)} \quad (4)$$

Close inspection will show equation (4) is not a parabola, but a third degree polynomial

$$y = d_0 + d_1x + d_2x^2 + d_3x^3 \quad (5)$$
where \( d_0 = \frac{(x_2b_0 - x_1c_0)}{(x_2 - x_1)} \), \( d_1 = \frac{(x_2b_1 - b_0 + c_0 - x_1c_1)}{(x_2 - x_1)} \), \
\( d_2 = \frac{(x_2b_2 - b_1 + c_1 - x_1c_2)}{(x_2 - x_1)} \), \( d_3 = \frac{(c_2 - b_2)}{(x_2 - x_1)} \).

Equation (4) is not valid to use between \((x_1, y_1)\) and \((x_2, y_2)\) if a relative extremum occurs between \(x_1\) and \(x_2\). This happens if:

\[
d_1 + 2d_2x + 3d_3x^2 = 0
\]

Equation (6) has at most two solutions

\[
x = \left\{ \begin{array}{ll}
-e_1 - \frac{(e_1^2 - 4e_2e_0)^{1/2}}{2e_2} & \\
-e_1 + \frac{(e_1^2 - 4e_2e_0)^{1/2}}{2e_2}
\end{array} \right.
\]

where \( e_0 = d_1 \), \( e_1 = 2d_2 \), and \( e_2 = 3d_3 \). Further refinement of the curve is needed if either value of \( x \) in equation (7) is strictly between \(x_1\) and \(x_2\).

Equation (5) has inflection points when \(2d_2 + 6d_3x = 0\) or \(x = -d_2/2d_3\).

An inflection between \(x_1\) and \(x_2\) is only desired if the line tangent to equation (5) at \((x_1, y_1)\) does not intersect the line tangent to equation (5) at \((x_2, y_2)\) within the rectangle \((x_1, y_1), (x_2, y_1), (x_2, y_2), (x_1, y_2)\).
The line tangent to equation (5) at the point \((e, \gamma)\) will have the equation

\[
y'c = \frac{(y - \gamma)}{(x - e)}
\]

where \(y'c\) is the slope of equation (5) at \((e, \gamma)\). Equation (8) can be written \(g_0 x + y = g_0\), where \(g_0 = \gamma + y'c\) and \(g_1 = y'c\). Two equations are formed where \((x_I, y_I)\) is the intersection. \(h_1 x_I + y_I = h_0\) and \(k_1 x_I + y_I = k_0\), or

\[
\chi = \begin{bmatrix} h_1, 1 \\ k_1, 1 \end{bmatrix}, H = \begin{bmatrix} h_0 \\ k_0 \end{bmatrix}, P = \begin{bmatrix} x_I \\ y_I \end{bmatrix}\] or

\[
XP = H, P = \chi^{-1}H. \]

Equation (5) is adequate if \(x_I\) is not strictly between \(x_1\) and \(x_2\) or if \(y_I\) is not strictly between \(y_1\) and \(y_2\). Equation (5) furnishes an adequate solution when \(y_2\) is not strictly between \(y_1\) and \(y_3\). More refinement may be needed when \(y_1\), \(y_2\), and \(y_3\) are strictly ordered. Equation (5) is such that \(y'x_1 = y_2 - y_1 / x_2 - x_1\) and \(y'x_2 = y_3 - y_2 / x_2 - x_1\). \(y_1\), \(y_2\), and \(y_3\) being strictly ordered require that \(y'x_1\) and \(y'x_2\) be of the same sign. An elliptical segment can be passed through \((x_1, y_1)\) and \((x_2, y_2)\) such that the slope at \(x_1\) is \(y_1'\) and the slope at \(x_2\) is \(y_2'\).

The general equation for the ellipse is

\[
x_0 + l_1x + l_2y + l_3xy + l_4x^2 = y^2
\]

The ellipse passes through \((x_1, y_1)\) and \((x_2, y_2)\) requiring

\[
x_0 + x_1 l_1 + y_1 l_2 + x_1 y_1 l_3 + x_1^2 l_4 = y_1^2
\]

(10)

\[
x_0 + x_2 l_1 + y_2 l_2 + x_2 y_2 l_3 + x_2^2 l_4 = y_2^2
\]

(11)

A relation involving the slope \(y'\) of equation (9) can be found using implicit partial differentiation.

\[
x_1 + x_2 y' + l_3(y + xy') + 2x_4x = 2yy'
\]

(12)

The two additional equations are

\[
x_1 + y_1' l_2 + (y_1 + x_1 y_1') l_3 + 2x_1 l_4 = 2yy_1'
\]

(13)

\[
x_1 + y_2' l_2 + (y_2 + x_2 y_2') l_3 + 2x_2 l_4 = 2yy_2'
\]

(14)

An additional constraint is that the change in the slope is constant. Using implicit partial differentiation on equation (12),

\[
x_2 y'' + l_3(2y' + xy'') + 2x_4 = 2yy'' + 2yy''
\]

or

\[
x_2 y'' + l_3(2y' + xy'') + 2x_4 = 2(y'' + yy'')
\]

(15)

where \(y'' = (y_2' - y_1') / (x_2 - x_1)\).

Equations 10, 11, 13, 14, and 15 are five equations in 5 unknowns. A solution for \(x_0\), \(l_1\), \(l_2\), \(l_3\), and \(l_4\) can be found using the matrices.
\[
X = \begin{bmatrix}
1, x_1, y_1, x_1 y_1 \\
1, x_2, y_2, x_2 y_2 \\
0, 1, y_1, (y_1 + x_1 y_1') \\
0, 1, y_2, (y_2 + x_2 y_2') \\
0, 0, y''', (2y_1' + x_1 y'')
\end{bmatrix}, \quad Y = \begin{bmatrix}
y_1^2 \\
y_2^2 \\
2y_1 y_1' \\
2y_2 y_2' \\
2 (y_2'y_2' + y_2 y_2'')
\end{bmatrix}
\]

and \( L = \begin{bmatrix}
\ell_0 \\
\ell_1 \\
\ell_2 \\
\ell_3 \\
\ell_4
\end{bmatrix} \). The relation is \( XL = Y \).

\( \ell_0, \ell_1, \ell_2, \ell_3, \) and \( \ell_4 \) are found by \( L = X^{-1}Y \). The ellipse has been fitted. Now a solution for \( y \) in terms of \( x \) must be found.

Starting from equation (9)

\[
\ell_0 + \ell_1 x + \ell_2 y + \ell_3 xy + \ell_4 x^2 = y^2
\]

Add \((\ell_2 + \ell_3 x)^2/4\) to both sides completing the square. \( y - \frac{1}{2}(\ell_2 + \ell_3 x) = \pm[\ell_0 + \ell_1 x + \ell_4 x^2 + (\ell_2 + \ell_3 x)^2/4]^{1/2}, \quad y = \frac{1}{2}(\ell_2 + \ell_3 x) \pm [\ell_0 + \ell_1 x + \ell_4 x^2 + (\ell_2 + \ell_3 x)^2/4]^{1/2} \)

The choice of the function depends upon the second derivative of the ellipse. \( y'' = (y_2' - y_1')' / (x_2 - x_1) \). The larger value of \( y \) must be taken if the ellipse is curving downward and the smaller value of \( y \) must be taken if the ellipse is curving upward.
The call to subroutine RICHER is

```
CALL RICHER(XI,YI,NDI,XI,YI,NDI,XMIN,XSTEP)
```

where \((XI(I), YI(I)), I = 1, \ldots, NDI\) are the \(NDI\) ordered pairs to be enriched.

\((XI(J), YI(J)), J = 1, \ldots, ND\) are the \(ND\) ordered pairs generated by RICHER such that \(XI(J) = (J - 1) \times XSTEP + XMIN, J = 1, \ldots, ND\).

The deck structure using RICHER in a FORTRAN program:

```
JOBCARD,CM40000,T60,PO.
REQUEST,RICHER,YRN=P894,DI,IN,FILES=41.
FORTRAN
LOAD(RICHER)
LGO
(End of record card)
FORTRAN program
(End of record card)
Data cards
(End of job card)
```

RICHER was used to generate 1000 points from the 14 points shown on the following graph. The solid line was formed by connecting the generated points by straight line segments.