Sustainability of Water Resources under Increasing Uncertainty (Proceedings of the Rabat Symposium S1, April 1997). IAHS Publ. no. 240, 1997.

B

Water resource predictions from meteorological probability forecasts

THOMAS E. CROLEY II

Great Lakes Environmental Research Laboratory, 2205 Commonwealth Boulevard, Ann Arbor, Michigan 48105-1593, USA

Abstract NOAA recently began providing monthly seasonal climate outlooks of expected air temperature and precipitation probabilities. Users can interpret climate outlooks to assess the risk to water resources of extreme meteorological conditions and of variability in general. One important approach uses historical meteorology record segments with hydrological, limnological and other models to simulate hydrological possibilities for the future (preserving observed spatial and temporal relationships). The meteorological possibilities are weighted to be compatible with climate outlook probabilities. The corresponding weighted hydrological possibilities are used to infer water resource probabilities and other parameters. The weights are determined by constructing boundary equations for the weights to match climate outlooks, setting the relative importance of each equation in case incompatibilities arise and solving them for physically relevant values. Their solution becomes an optimization problem for the general case. An example illustrates the concepts and method.

MAKING PROBABILISTIC OUTLOOKS

The National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center now provides a monthly climate outlook, consisting of a 1-month outlook for the next month and thirteen 3-month outlooks, going into the future in overlapping fashion in 1-month steps. Each outlook estimates probabilities of average air temperature and total precipitation falling within the lower, middle and upper thirds of observations from 1961-1990. Users of these climate outlooks can interpret the forecast probabilities in terms of the impacts on themselves through operational hydrology approaches. Possibilities for the future are identified that resemble past meteorology (preserving observed spatial and temporal relationships) yet are compatible with the climate outlooks. One operational hydrology approach considers historical meteorology as a possibility for the future by segmenting the historical record and using each segment with models to simulate a hydrological possibility for the future (Fig. 1). Each segment of the historical record then has associated time series of meteorological and hydrological variables, representing a possible scenario for the future. The approach can then consider the resulting set of possible future scenarios as a statistical sample and infer probabilities and other parameters associated with both meteorology and hydrology through statistical estimation from this sample (Croley, 1996; Day, 1985; Smith *et al.*, 1992).

The operational hydrology approach uses statistical sampling tools as if the set of possible future scenarios were a single random sample (i.e. the scenarios are independent of each other and equally likely). This means that the relative frequencies of selected events are fixed at values different (generally) from those specified in climate

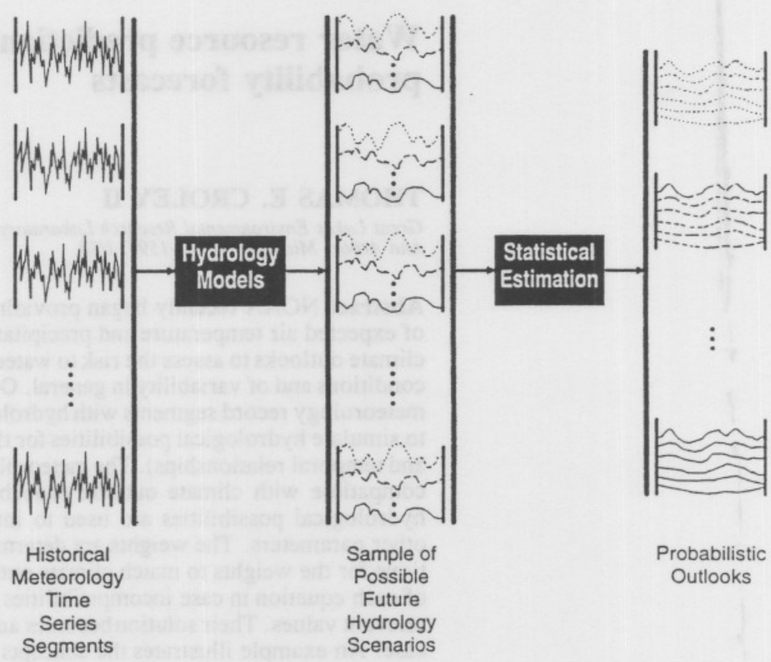

Fig. 1 An operational hydrology approach.

outlooks. Only by restructuring the set of possible future scenarios can we obtain relative frequencies of selected events that match climate outlooks. This restructuring violates the assumption of independent and equally likely scenarios (no random sample) from the point of view of the historical record (*a priori* information). However, the restructured set can be viewed as a random sample (*a posteriori* information) of scenarios conditioned on climate outlooks. There are many methods for restructuring the set of possible future scenarios (Croley, 1996; Day, 1985; Ingram *et al.*, 1995; Smith *et al.*, 1992).

BUILDING A STRUCTURED SET

Consider structuring a set of possible future scenarios that gives relative frequencies of average air temperature and total precipitation (over various times in the scenarios) satisfying *a priori* settings of climate outlooks. We can arbitrarily construct a very large structured set of size N by adding (duplicating) each of the available scenarios (in the original set of n possible future scenarios); each scenario numbered i, ($i = 1, \ldots, n$) is duplicated r_i times. By judiciously choosing these duplication numbers, ($r_1, r_2, \ldots, r_n$), it is possible to force the relative frequency of any arbitrarily defined "group" of scenarios in the structured set to any desired value. For example (see Fig. 2), suppose only five of 50 (10%) 12-month scenarios beginning in April have a total April precipitation exceeding 80 mm and our *a priori* setting (from a climate outlook) for this exceedance is 20%. We could repeat each of these five scenarios nine times and repeat the other 45 scenarios four times to build a structured set as in Fig. 2. This structured set of size 225 (= 5 × 9 + 45 × 4) would then have a relative frequency of 20% of total April precipitation exceeding 80 mm (5 × 9 / 225 = 0.2). For sufficiently large N, we can approximate *a priori* settings at any precision by using integer-valued duplication numbers, r_i. Note that the n duplication numbers sum to N.

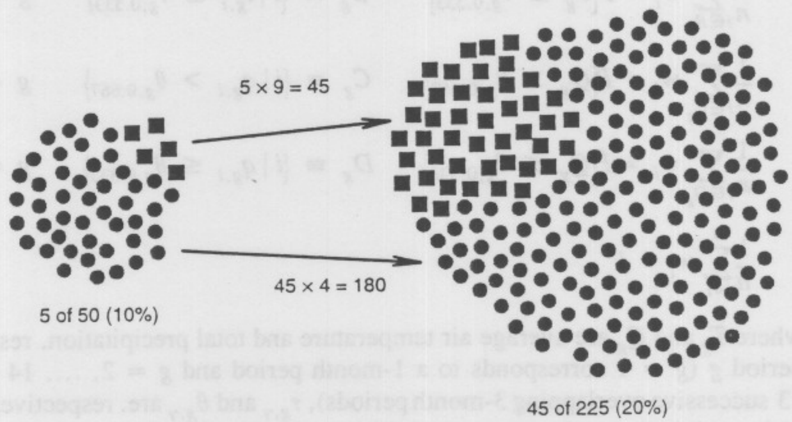

Fig. 2 Example restructuring of an operational hydrology "sample".

By treating the N scenarios in the very large structured set as a statistical sample, we can estimate probabilities and calculate other parameters for all variables. In particular, we can estimate the probability, that any variable X is less than or equal to a value x, by its relative frequency:

$$\hat{P}[X \le x] = \sum_{k \in \Omega} \frac{1}{N} \qquad \Omega \equiv \left\{k \mid x_k^N \le x\right\} \tag{1}$$

where x_k^N is the value of variable X for the kth scenario in the very large structured set of N scenarios. [Read the set notation in equation (1) as "Ω is all values of k such that $x_k^N \le x$"]. Actually, there are only n different values of X (x_i^n, $i = 1, \ldots, n$) since these n values were duplicated, each by a number, r_i, to create the N values in the very large structured set. We can rewrite equation (1) in terms of the original set of possible future scenarios, for any variable X:

$$\begin{aligned}\hat{P}[X \le x] &= \sum_{i \in \Omega} \frac{r_i}{N} \qquad \Omega \equiv \left\{i \mid x_i^n \le x\right\} \\ &= \frac{1}{n}\sum_{i \in \Omega} w_i \qquad \Omega \equiv \left\{i \mid x_i^n \le x\right\}\end{aligned} \tag{2}$$

where $w_i = (n/N)r_i$. Note that the n weights sum to n. If all $w_i = 1$, then equation (2) gives contemporary (unstructured) estimates from the original set, treated as a statistical sample. Croley (1996) derives other statistics.

CONSIDERING MULTIPLE OUTLOOKS

Now consider matching relative frequencies in equation (2) to the multiple *a priori* settings of the NOAA Climate Outlook:

$$\frac{1}{n}\sum_{i\in A_g} w_i = \hat{P}[T_g > \tau_{g,0.667}] \qquad A_g \equiv \{i \mid t_{g,i} > \tau_{g,0.667}\} \qquad g = 1,\ldots,14$$

$$\frac{1}{n}\sum_{i\in B_g} w_i = \hat{P}[T_g \le \tau_{g,0.333}] \qquad B_g \equiv \{i \mid t_{g,i} \le \tau_{g,0.333}\} \qquad g = 1,\ldots,14$$

$$\frac{1}{n}\sum_{i\in C_g} w_i = \hat{P}[Q_g > \theta_{g,0.667}] \qquad C_g \equiv \{i \mid q_{g,i} > \theta_{g,0.667}\} \qquad g = 1,\ldots,14 \qquad (3)$$

$$\frac{1}{n}\sum_{i\in D_g} w_i = \hat{P}[Q_g \le \theta_{g,0.333}] \qquad D_g \equiv \{i \mid q_{g,i} \le \theta_{g,0.333}\} \qquad g = 1,\ldots,14$$

$$\frac{1}{n}\sum_{i=1}^{n} w_i = 1$$

where T_g and Q_g are average air temperature and total precipitation, respectively, over period g ($g = 1$ corresponds to a 1-month period and $g = 2, \ldots, 14$ corresponds to 13 successive overlapping 3-month periods), $\tau_{g,\gamma}$ and $\theta_{g,\gamma}$ are, respectively, temperature and precipitation reference γ-probability quantiles for period g, and $t_{g,i}$ and $q_{g,i}$ are average air temperature and total precipitation, respectively, over period g of scenario i. By definition, the reference γ-probability quantiles are estimated from the 1961-1990 historical record for each period g. To illustrate equations (3), consider the April 1996 climate outlook: there is a 1-month April outlook ($g = 1$ or "Apr") and thirteen 3-month outlooks successively lagged by 1 month each ($g = 2$ or "April-May-June" or "AMJ," and $g = 3, \ldots, 14$ or "MJJ," ..., "AMJ," respectively). The last of equations (3) corresponds to the requirement that relative frequencies sum to unity.

Rewriting equations (3):

$$\sum_{i=1}^{n} a_{k,i} w_i = e_k \qquad k = 1,\ldots,57 \qquad (4)$$

where $a_{k,i}$ has the value of 0 or 1 corresponding to the exclusion or inclusion, respectively, of each variable in the above sets and e_k corresponds to the climate outlook relative frequency settings specified above (e.g. $e_k = n\hat{P}[T_k > \tau_{k,0.667}], k = 1, \ldots, 14$).

Ordinarily, all of the climate outlooks may not be used; then simply write equation (4) as:

$$\sum_{i=1}^{n} a_{k,i} w_i = e_k \qquad k = 1,\ldots,m \qquad (5)$$

where $m \le 57$ and the appropriate equations, corresponding to the unused outlooks, are omitted. We must solve equations (5) simultaneously to find the weights.

Generally, $m \ne n$ and some of the equations may be either redundant or non-intersecting and must be eliminated. (If $m > n$, then $m - n$ of the equations must be either redundant or non-intersecting. This corresponds to not being able to simultaneously satisfy all climate outlooks with fewer scenarios than there are outlook boundary conditions). Selection of some for elimination is facilitated by assigning each of equations (5) a priority reflecting its importance to the user. [The highest priority is given to the last equation in (5), guaranteeing that all relative frequencies sum to unity]. Each equation, in priority order starting with the next-to-highest priority, is compared to the set of all higher-priority equations and eliminated if it is redundant or does not intersect the set. By starting with the higher priorities, we ensure that each equation is compared with a known valid set of equations and that we keep higher-priority equations in preference to lower-priority ones. Thus we can always reduce equations (5) so that $m \leq n$. If $m = n$, equations (5) can be solved via Gauss-Jordan elimination as a system of linear equations for the weights, w_i, since the equations are now independent and intersecting (in n-space).

For $m < n$, there are multiple solutions to equations (5) and identification of the "best" requires the specification of a measure for comparing them. One such measure is the sum of squared differences of the weights from unity. Solutions that give smaller values of this measure can be judged "better" than those that do not (and the resulting very large structured set of scenarios is more similar to the original set of scenarios in this sense). Other measures are also possible, including those using other functions expressing deviation of the weights from a goal, or measures defined on the resulting joint probability distribution function estimates (looking at similarity in joint distributions between the very large structured set and the original set).

We can formulate an optimization problem to minimize the above deviation of weights from unity in selecting a solution to equations (5):

$$\begin{aligned} &\min \sum_{i=1}^{n} (w_i - 1)^2 \\ &\text{s.t.} \sum_{i=1}^{n} a_{k,i} w_i = e_k \qquad k = 1, \ldots, m \end{aligned} \tag{6}$$

The solution of equations (6) may give positive, zero, or negative weights, but only non-negative weights make physical sense and we must further constrain the optimization to non-negative weights. However, non-negativity constraints can result in infeasibility (there is no solution). In this case, additional lowest priority equations must be eliminated from equations (5) to allow a non-negative solution. Croley (1996) suggests two systematic procedures for finding non-negative weights. The first method guarantees that only strictly positive weights will result; this means that all possible future scenarios are used (no scenario is weighted by zero and effectively eliminated) in estimating probabilities and other parameters. The second method disallows some of the possible future scenarios (by allowing zero weights) but satisfies more *a priori* settings [more of the equations in (5)] in the solution.

EXAMPLE CONSIDERATION OF MULTIPLE OUTLOOKS

The NOAA Climate Outlook for April 1996 (made 14 March 1996) over the Lake Ontario Basin is given in Table 1. The highlighted entries in Table 1 are used arbitrarily, in priority of their appearance, to make a hydrological outlook for Lake Ontario

Table 1 NOAA April 1996 climate outlook for the Lake Ontario basin.

Period, g	$\hat{P}\left[T_g \le \tau_{g,0.333}\right]$	$\hat{P}\left[T_g > \tau_{g,0.667}\right]$	$\hat{P}\left[Q_g \le \theta_{g,0.333}\right]$	$\hat{P}\left[Q_g > \theta_{g,0.667}\right]$
Apr '96	33	33	33	33
AMJ '96	28	38	33	33
MJJ '96	29	37	33	33
JJA '96	30	36	33	33
JAS '96	33	33	33	33
ASO '96	33	33	33	33
SON '96	38	28	33	33
OND '96	33	33	33	33
NDJ '96	33	33	33	33
DJF '96	33	33	33	33
JFM '97	33	33	33	33
FMA '97	33	33	33	33
MAM '97	33	33	33	33
AMJ '97	28	38	33	33

[a]Probabilities are expressed as percentages and two-digit round-off is used here. Highlighted entries are chosen arbitrarily for the example.

Table 2 Boundary condition equations (5) for April 1996 outlook on Lake Ontario.

Period, g[a]	k[b]	Inclusion in interval, $a_{k,i}$, $i = 1, ..., 45$[c]	e_k[c]
Apr '96	2	100010110110100000001100001011000100011100110	0.33×45
Apr '96	3	001000001000010001100001100100100011000001000	0.33×45
Apr '96	4	100100101000110010000100010010011001100000111	0.33×45
Apr '96	5	001000010010000001101001100100100010010001000	0.33×45
AMJ '96	6	010111110101101010000010010011000100001100110	0.38×45
AMJ '96	7	001000001010010001110001101000100011000000001	0.28×45
AMJ '96	8	000000001001100000000100111010000001100001100	0.33×45
AMJ '96	9	011000010001001111100001000000110000000010010000	0.33×45
MJJ '96	10	010111010001001010000010010101000000001111010	0.37×45
MJJ '96	11	000000001010010001111100001010000000110000001	0.29×45
JJA '96	12	110011010001000000110000010100000000110011110	0.36×45
JJA '96	13	001100001010101011001001100001000010011000001	0.30×45
SON '96	14	100001100011110100001011010100000011010000000	0.28×45
SON '96	15	000000000000010000100001010101011000001010001	0.38×45
AMJ '97	16	101111101011010000000100100110001000011001100	0.38×45
AMJ '97	17	010000010100100011100011010001000111000000010	0.28×45
Entire	1	111111111111111111111111111111111111111111111	1.00×45

[a]Period as selected (highlighted) in Table 1.
[b]Period renumbered by priority (1 ≡ highest) as in equations (5).
[c]Coefficients in equations (5) defined for each selected period, k, of the climate outlook, and for each scenario, i, in the historical record.

Table 3 Climate outlook weights using all the *a priori* climate settings[a].

Year	Weight	Year	Weight	Year	Weight
1948	0.176736	1963	1.314862	1978	2.312796
1949	0.730660	1964	0.629071	1979	1.141716
1950	0.373923	1965	1.105104	1980	1.459480
1951	0.251514	1966	1.138449	1981	1.850892
1952	0.198247	1967	0.694093	1982	0.608515
1953	0.375382	1968	1.370861	1983	0.851183
1954	0.436866	1969	0.912931	1984	1.746106
1955	0.543422	1970	0.747359	1985	0.698122
1956	0.691685	1971	0.691137	1986	1.238027
1957	1.022952	1972	1.609610	1987	0.953608
1958	0	1973	0.847258	1988	2.081123
1959	0.458004	1974	0.861465	1989	2.310484
1960	0.866600	1975	2.087578	1990	0.582730
1961	1.069536	1976	2.169144	1991	0.424938
1962	2.230866	1977	0.542460	1992	0.592504

[a]Solution of equations (6) with Table 2 coefficients; all *a priori* settings in Table 1 are used.

beginning 21 March 1996. These 16 outlook settings are used with inspection of the forty-five 12-month time series, beginning in April from the available historical record of 1948-1993, to construct 17 equations represented by equations (5) in Table 2. Table 3 presents the solution of these equations, found by minimizing the deviation of weights from unity, as in equation (6), by utilizing all 16 climate outlook settings. Note from Table 3 that one weight was assigned a value of zero to enable this inclusion. This means that the scenario starting in April 1958 is unused in the ensuing probabilistic outlook.

Finally, an example probabilistic outlook for three variables, over the 12 months from April 1996 to March 1997, is given in Fig. 3. There were 45 values of each modelled monthly variable (runoff, lake surface temperature and lake evaporation), corresponding to the 45 scenarios used in the simulation, for each of the 12 full months and the 1 partial month (March). Each value was used with its respective weight from Table 3, as in equations (2), to compute various statistics for the probabilistic outlook each month.

EXTENSIONS

The determination of weights involves several choices made arbitrarily herein. For example, the weights could be determined directly from multiple climate outlooks as exemplified in Fig. 2 for a single climate outlook, but would involve restrictions on the multiple climate outlooks not considered here. The formulation of an optimization problem, used herein, allows for a more general approach in determining these weights in the face of multiple outlooks. However, this formulation also involves arbitrary choices, the largest of which is the selection of a relevant objective function. As mentioned earlier, other measures of relevance of the weights to a goal are possible and could require reformulating the solution methodology. An early approach, not reported herein, minimized the sum of squared differences between the relative frequencies

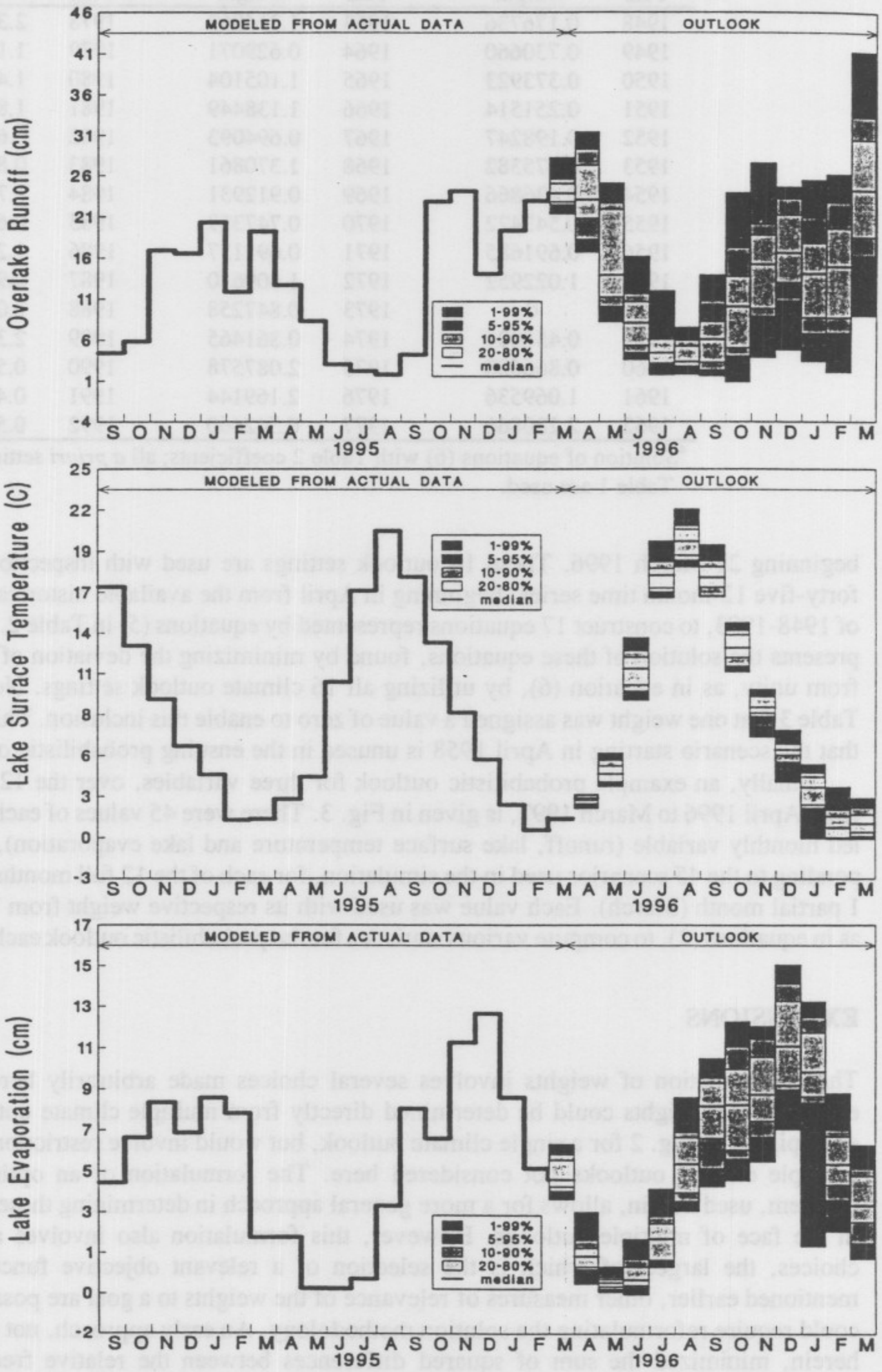

Fig. 3 Selected example probabilistic Lake Ontario hydrology outlooks from April 1996 Climate Outlook.

associated with the bivariate distribution of precipitation and temperature before and after application of the weights. The goal was to make the resulting joint distribution as similar as possible to that observed historically while making the marginal distributions match the climate outlooks. Unfortunately, the method was intractable for consideration of more than one climate outlook.

An important advantage associated with the computation of a weighted sample in the operational hydrology approach described herein is the independence of the weights and the hydrology models. After model simulations are made to build a set of possible future scenarios for analysis, several probabilistic outlooks can be generated with weights corresponding to the use of different climate outlooks, different methods of considering the climate outlooks and alternate selections of just which of the 14 outlooks (56 settings) to use that are available each month. In making these alternate analyses and weights (re)computations, it is unnecessary to redo the model simulations to rebuild the set. This is a real saving when the model simulations are extensive, as is the case with Great Lakes hydrological outlooks. This also enables efficient consideration of other ways to use the weights to make probabilistic outlooks. For example, the use of non-parametric statistics in equation (2) restricts the range of any variable to that present in the historical record or in their hydrological transformations. An alternative that does not restrict range in this manner is to hypothesize a distribution family (e.g. normal, lognormal, log Pearson Type III) and to estimate its moments by utilizing sample statistics defined analogously to those in equation (2); see Croley (1996). The detractor for parametric estimation is hypothesizing the family of distributions to use.

Most significantly, the method allows joint consideration of multiple meteorological outlooks defined over different lengths and periods of time. It can be easily extended to incorporate consideration of 6- to 14-day outlooks, for which there is relatively greater skill, as well as other period outlooks.

Computer code is available to make all computations (outside of the hydrological modelling) for use by others in utilizing the NOAA Climate Prediction Center *Climate Outlook*. The code finds all necessary reference quantiles for using a climate outlook from a user-supplied file of historical daily air temperature and precipitation, sets up equations (5), formulates the optimization of equations (6) and performs sequential optimizations (either to use all historical data or to maximize use of *a priori* climate outlook settings). Both a stand-alone FORTRAN implementation, for use under a variety of operating systems and a specially-designed user interface Windows™ application are available. The latter allows understandable interpretation of the NOAA Climate Prediction Center's *Climate Outlooks* and assignment of relevant priorities.

Acknowledgement This paper is GLERL Contribution No. 1013.

REFERENCES

Croley, T. E., II (1996) Using NOAA's new climate outlooks in operational hydrology. *J. Hydrol. Engng* **1**, 93-102.

Day, G. N. (1985) Extended streamflow forecasting using NWSRFS. *J. Wat. Resour. Plan. Manage.* **111**, 157-170.

Ingram, J. J., Hudlow, M. D. & Fread, D. L. (1995) Hydrometeorological coupling for extended streamflow predictions. In: *Preprints of Conference on Hydrology*, 75th Annual Meeting of the American Meteorological Society (Dallas, Texas, January 1995), 186-191.

Smith, J. A., Day, G. N. & Kane, M. D. (1992) Nonparametric framework for long-range streamflow forecasting. *J. Wat. Resour. Plan. Manage.* **118**, 82-92.