Climate, Soil, and Vegetation

7. A Derived Distribution of Annual Water Yield

PETER S. EAGLESON

Department of Civil Engineering, Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

The average annual soil moisture balance, as derived from the mechanics of storm and interstorm soil moisture movement and from the statistics of the climatic variables, is used to define the average annual soil moisture. This soil moisture is used in the equation for average annual yield to give a first-order approximation of the annual precipitation yield function. This function is used to transform the cumulative distribution function (cdf) of annual precipitation into the cdf of annual yield, and application is made in a subhumid and in an arid climate. The derived yield frequency function is seen to be sensitive to the soil and vegetal properties. Proper selection of these parameters brings close agreement with observed streamflow-frequency and suggests the model’s utility for parameterizing drainage basins with respect to effective average soil and vegetal properties.

INTRODUCTION

Planning for water resource development requires, among other things, estimation of the average return period of extreme annual water yields, both high and low. This is done most commonly through extrapolation of some assumed probability distribution (such as the normal, log normal, or extreme value) which has been fitted to the available set of data either for the river basin in question or for one which has been judged to be ‘hydrologically similar.’

There are several difficulties with this approach. First, in developing countries in particular, the length of streamflow record is seldom great enough to provide a good estimate of the parameters of a fitted distribution, thus extrapolation to extreme events introduces very large possible errors with all the attendant cost and risk penalties.

Second, in developed countries where there may be an adequate record for fitting and extrapolating an assumed distribution, the activities of man have often created (or can be expected to create over the useful life of the project) changes in the precipitation versus yield relation through urbanization, deforestation, irrigation, etc. In such nonstationary situations the estimation of future behavior based upon observations of the past is also error-prone and provides no mechanism for assessing the impact of man-induced changes.

Third, the transference of observations from another location either to supplement a meager record or to capture the characteristic behavior of a future state of urbanization requires the ability to recognize hydrologically similar regions. But what are the similarity criteria governing this transferability? Until now [Eagleson, 1978c], these have not been known.

Fourth, when fitting probability distributions to observations, we have little physical basis for our choice of distribution type, yet the shape of the rare event ‘tails’ is critically dependent upon this choice.

Most of these deficiencies can be corrected, at considerable cost, however, through the use of computer-based simulation of the precipitation-streamflow behavior of the basin. Such simulations accept sequences of appropriately variable precipitation and through the detailed formulation of physical behavior generate sequences of varying streamflow. This very powerful technique has its place in our bag of tricks but is often hard to justify for reasons of time and cost.

In this work a new analytical method of addressing water yield problems is developed and demonstrated. It answers the first difficulty by using the hitherto neglected statistical properties of the individual storms to derive an improved estimate of the cumulative distribution function (cdf) of annual yield from only a few years of observation. It responds to the second difficulty by deriving the water yield statistics in terms of observable properties of the catchment’s climate-soil-vegetation system. Therefore changes in risk due to man-induced alteration of the system can be explored directly. The third difficulty is addressed elsewhere through the derivation of the dimensionless parameters governing the dynamic similarity of the water budget [Eagleson, 1978c]. By deriving the probability distributions of the individual water budget components from physically based equations the empirical fitting of distributions is shifted back to the independent variables where generality of the chosen distributions is more likely. This responds to the fourth difficulty as well as can be done with the present understanding of the precipitation-generating processes.

PRECIPITATION-YIELD FUNCTION

Annual water yield $Y_A$ is defined as the sum of two water balance components, surface runoff $R_s$ and groundwater runoff $R_g$. In the absence of changes in moisture storage it is given alternatively by the difference between annual precipitation $P_A$ and evapotranspiration $E_T$. In equation form we have

$$Y_A = R_s + R_g \quad (1)$$

or

$$Y_A = P_A - E_T \quad (2)$$

The form of the variation of these water balance constituents with change in the primary climatic variable $P_A$ is sketched in the upper portion of Figure 1 for the simplified case in which all evapotranspiration is an abstraction from infiltration $I_A$. By implying zero surface retention capacity $h_o$ this assumption gives the auxiliary conservation equation,

$$I_A = P_A - R_s = E_T + R_s \quad (3)$$

which may be used to deduce the form of the precipitation-yield function as shown in the lower portion of Figure 1.

This important $Y_A$ versus $P_A$ relationship has been a feature of much of the rainfall-runoff research of the last 75 years [Raffier, 1903; Lee, 1942; Troxell et al., 1954].
It long ago became clear, however, that even for a given and
basin, $Y_A$ is not a unique function of $P_A$ alone as is implied by
Figure 1. It is important to yield production not only how
much rainfall arrives in the year but also the intensity $i$, dura-
tion $t_r$, and spacing $t_o$ of the storms which deliver it and the
length of the rainy season $r$ during which they occur. Of equal
importance are the temperature $T_a$ affecting soil moisture
movement and the rate of potential evaporation $e_s$. Given
the soil and vegetation, the precipitation-yield function takes
the form

$$Y_A = g(i, t_r, t_o, r, T_a, e_s)$$

(4)

With appropriate aggregation it may be possible to write the
independent variables of (4) in terms of their annual values to
obtain

$$Y_A = g_i(P_A, E_{P_A}, T_K(1); \text{parameters})$$

(5)

where $E_{P_A}$ is the annual potential evapotranspiration from
soil moisture, $K(1)$ is the saturated effective hydraulic con-
ductivity of the soil, and the parameters define the soil and
vegetation.

If we can identify the function $g_i( )$ of (5), we have not
only the useful relation between yield and climate, but also
(and perhaps more importantly) we have the analytical basis
for deriving the distribution of $Y_A$ from known distributions of
$P_A, E_{P_A}$, and $T_a$.

The average annual water balance [Eagleson, 1978c] will
provide the basis for our approximation of the function $g_i( )$.

WATER BALANCE EQUATION

The average annual balance of soil moisture has been for-
mulated dimensionlessly [Eagleson, 1978c] for regions in
which snowfall is a minor contributor to annual precipitation.
This is, for non-zero surface runoff

$$[1 - e^{-e_s - T_w(\sigma + 1)})] = E[Y_A]/m_{P_A} J(E, M, k_o)$$

$$+ \frac{m_o K(1)}{m_{P_A}} \frac{s_c}{m_{P_A}} - T_w (6)$$

(The term to the left of the equal sign is infiltration, the first
term to the right is evapotranspiration from soil moisture, and
the last two terms are the groundwater runoff (the first is
groundwater recharge and the last is groundwater loss).)

We choose to express the average annual yield in the form
given by (2). By expanding, this becomes

$$E[Y_A]/m_{P_A} = 1 - \frac{E[E_{P_A}^*]}{m_{P_A}} J(E, M, k_o) - E[E_{P_A}^*]/m_{P_A}$$

(7)

(The term to the left of the equal sign is yield, the first term to
the right is precipitation, the second is evapotranspiration from
soil moisture, and the last is evaporation from surface
retention.) In the preceding two equations,

$$G = \left[\alpha K(1)/2\right][1 + s_c^2] - \alpha w$$

(8)

$$\sigma = \left[\frac{5m^+ K(1)\Psi(1)(1 - s_c)^2}{6\pi b_m} \right]^{1/3}$$

(9)

$$E = \frac{2\beta K(1)\Psi(1)\phi_e}{\pi m e_p^2} t_{s_a + 2}$$

(10)

$$w/K(1) = \left[1 + \frac{3/2}{m e + 1}\right] \left[\frac{\Psi(1)}{Z}\right]^{m}$$

(11)

and

$$E[E_{P_A}^*] = m_m m_e[1 - M(1 - k_o)]e_p - m_e E[E_{P_A}]$$

(12)

where the average interstorm surface retention loss is approxi-
mately

$$\beta E[E_{P_A}]/e_p = (1 - M) \left[1 - e^{-e_s/\phi_e} \frac{\Gamma(\xi, \lambda_k h_0)}{\Gamma(\xi)} \right.$$

$$- \left[1 + \frac{\beta_h e_p}{\lambda_k h_0} \right]^{-1} \frac{\gamma(e_s, \lambda_k h_0 + \beta h/e_p)}{\Gamma(\xi)}$$

$$+ M k_o \left[1 - e^{-e_s/\phi_e} \frac{\Gamma(\xi, \lambda_k h_0)}{\Gamma(\xi)} \right.$$

$$- \left[1 + \frac{\beta_h e_p}{\lambda_k h_0} \right]^{-1} \frac{\gamma(e_s, \lambda_k h_0 + \beta h/e_p)}{\Gamma(\xi)}$$

(13)

and the soil moisture evapotranspiration function is given by

$$J(E, M, k_o) = 1 - [(1 - M)/(1 - M + M k_o)]$$

$$\cdot \left[1 + M k_o + (2B)^{1/2} e^{-e_k}$$

$$- [M k_o + (2C)^{1/2} e^{-e_k}$$

$$- (2E)^{1/2} (\gamma(\xi, CE) - \gamma(\xi, BE))]\right.$$  

(14)

and

$$B = \frac{1 - M}{1 + M k_o - w/e_p} + \frac{M k_o + (1 - M)w/e_p}{2(1 + M k_o - w/e_p)}$$

(15)

and

$$C = k(M k_o - w/e_p)^2$$

(16)

Assuming that the vegetal cover has reached its 'growth
equilibrium' density $M_0$ for the given species [Eagleson,
1978c], we can find $M = M_0$ (and thus replace it in (14)) by

$$\delta s / \delta M = 0 \quad M = M_0$$

(17)

In vegetal systems which are water limited and have reached
their 'evolutionary equilibrium' species mix [Eagleson, 1978c]
we may find the equilibrium plant coefficient $k_{e_o}$ by

$$\delta(M k_o) / \delta k_o = 0 \quad k_o = k_{e_o}$$

(18)
where (for the preceding equations)
- \( a \) - average storm intensity, equal to \( m_r \), centimeters per second;
- \( K(1) \) - saturated effective hydraulic conductivity of soil, centimeters per second;
- \( s_0 \) - time and space average soil moisture concentration in surface boundary layer;
- \( c \) - soil pore disconnectedness index;
- \( w \) - apparent capillary rise velocity from water table, centimeters per second;
- \( n \) - effective porosity of soil;
- \( r_t \) - average storm depth, equal to \( m_r \), centimeters;
- \( m_p \) - pore size distribution index;
- \( d \) - diffusivity index;
- \( Z \) - depth to water table, centimeters;
- \( G \) - gravitational infiltration parameter;
- \( a \) - capillary infiltration parameter;
- \( E \) - exfiltration parameter;
- \( E^* \) - annual potential evapotranspiration from soil moisture, centimeters;
- \( m_y \) - average number of storms per rainy season;
- \( /5 \) - average time between storms, equal to \( m_{tr} \), days;
- \( / \) - dimensionless sorption diffusivity;
- \( d \) - diffusivity index;
- \( \Delta \) - dimensionless desorption diffusivity;
- \( M \) - vegetal canopy density;
- \( M_0 \) - equilibrium vegetal canopy density, equal to \( P_A/m_{mv} \);
- \( k_0 \) - plant coefficient;
- \( k_{so} \) - equilibrium plant coefficient where water (not nutrition or light) is limiting;
- \( \lambda \) - parameter of gamma distribution of storm depths, equal to \( \lambda/m_H \), cm\(^{-1}\);
- \( h_0 \) - surface retention capacity, centimeters.

Equation (6) allows us to solve for the soil moisture \( s_0 \) which satisfies the annual average soil moisture balance. This solution, which cannot be performed analytically, gives

\[
\begin{align*}
   s_0 &= s_0(m_r, E[E_p^*], m_r K(1); \text{parameters}) \quad (19)
\end{align*}
\]

Substituting (19) into (7) eliminates \( s_0 \) from the latter equation to give the relation

\[
\begin{align*}
   E[Y_A] &= g_1(m_{p}, E[E_p^*], m_r K(1); \text{parameters}) \quad (20)
\end{align*}
\]

Our problem now becomes one of relating the two functions \( g(\ ) \) of (5) and \( g(\ ) \) of (20).

**First-Order Approximation to Annual Water Balance**

Defining a climatic mean \( m_r \) at which \( P_A = m_{p}, E_p^* = E[E_p^*] \), and \( r K(1) = m_r K(1) \) and expanding (5) about this point in a multidimensional Taylor expansion, we have [Hilbrand, 1949, p. 353]

\[
\begin{align*}
   Y_A &= g_1(m_{p}, E[E_p^*], m_r K(1); \text{parameters}) + \sum_{k=1}^{n} \left( P - m_{p} \right) \left( \frac{\partial g_1}{\partial P_{A}} \right) + \left( E_p^* - E[E_p^*] \right) \left( \frac{\partial g_1}{\partial E_p^*} \right) + \left( r - m_{r} \right) \left( \frac{\partial g_1}{\partial r} \right) g_1(m_{p}) \quad (21)
\end{align*}
\]

Taking expected values of (21) term-by-term and omitting higher-order terms, we have the ‘second-order approximation’ to \( E[Y_A] \) [Benjamin and Cornell, 1970, p. 184]:

\[
\begin{align*}
   E[Y_A] &= g_1(m_{p}, E[E_p^*], m_r K(1); \text{parameters}) + \sum_{k=1}^{n} \left( P - m_{p} \right) \left( \frac{\partial g_1}{\partial P_{A}} \right) Var[P_{A}] + \left( E_p^* - E[E_p^*] \right) \left( \frac{\partial g_1}{\partial E_p^*} \right) Var[E_p^*] + \left( r - m_{r} \right) \left( \frac{\partial g_1}{\partial r} \right) \left( \frac{\partial g_1}{\partial r} \right) Cov[P_{A}, E_p^*] \quad (22)
\end{align*}
\]

If we assume that all variances, covariances, and curvatures are small, (22) reduces to the ‘first-order approximation’ of \( E[Y_A] \):

\[
\begin{align*}
   E[Y_A] &= g_1(m_{p}, E[E_p^*], m_r K(1); \text{parameters}) \quad (23)
\end{align*}
\]

Thus to the first order,

\[
\begin{align*}
   g_1(\ ) &= g_1(\ ) \quad (24)
\end{align*}
\]

and we may obtain the desired (5) by replacing the average annual values in (19) and (20) by the annual values. That is, to the first order,

\[
\begin{align*}
   Y_A &= g_1(P_A, E_p^*, r K(1); \text{parameters}) \quad (25)
\end{align*}
\]

**Distribution of Annual Yield**

We are now in a position to use \( g(\ ) \) as defined by (6) and (7) to derive the distribution of \( Y_A \) from the distribution of the independent variables \( P_A, E_p^* \), and \( r K(1) \). Here we will deal only with the simplest case, that in which all variation in \( Y_A \) comes from variation in \( P_A \). Therefore we will consider \( E_p^* \) and \( r K(1) \) to be fixed at their long term average values \( E[E_p^*] \) and \( m_r K(1) \), respectively. Equation (25) then becomes

\[
\begin{align*}
   Y_A &= g_1(P_A, E[E_p^*], m_r K(1); \text{parameters}) \quad (26)
\end{align*}
\]

or, simply

\[
\begin{align*}
   Y_A &= g_1(P_A) \quad (27)
\end{align*}
\]

Analysis of (6) [Eagleson, 1978c] has shown the yield to increase monotonically with \( P_A \), thus (27) is also monotonically increasing. Under such conditions [Benjamin and Cornell, 1970, p. 133] we can relate the cumulative distribution functions (i.e., the cdf) of \( Y_A \) and \( P_A \) by

\[
\begin{align*}
   F_{Y_A}(z) &= F_{P_A}[g_1^{-1}(z)] \quad (28)
\end{align*}
\]

in which \( g_1^{-1}(z) \) represents (27) solved for \( P_A \). That is,

\[
\begin{align*}
   P_A &= g_1^{-1}(Y_A) \quad (29)
\end{align*}
\]

The cdf of annual precipitation is obtained by fitting the set of annual observations if the period of record is adequate. When the record is short but includes observations of the individual storms, we may derive the cdf of \( P_A \) from the properties of the observed storm sequence.

Considering the arrival of storms to be a Poisson process and fitting the storm depths with a gamma distribution, Eagleson [1978a] has derived the cdf of annual precipitation (for constant season length) as
Fig. 2. Frequency of annual precipitation at Clinton, Massachusetts.

\[
\text{Prob} \left[ \frac{P_{A}}{m_{P_A}} < x \right] = e^{-\omega m_{r}} \left\{ 1 + \sum_{i=1}^{\nu} \left( \frac{\omega m_{r}}{\nu!} \right)^i P[\nu x, \omega m_{r} x] \right\}
\]

where

- \( \omega \) mean storm arrival rate, days\(^{-1}\);
- \( \omega m_{r} \) mean number of storms per rainy season, equal to \( m_{r} \);
- \( \nu \) storm counting variable;
- \( \kappa \) parameter of gamma distribution of storm depths;
- \( P[\alpha, x] \) Pearson's incomplete gamma function [Abramowitz and Stegun, 1970, p. 262], equal to \( \gamma[\alpha, x]/\Gamma(\alpha) \).

Using (28), we get the desired cdf of annual yield.

\[
\text{Prob} \left[ \frac{Y_A}{m_{P_A}} < z \right] = e^{-\omega m_{r}} \left\{ 1 + \sum_{i=1}^{\nu} \left( \frac{\omega m_{r}}{\nu!} \right)^i P[\nu x, \omega m_{r} x] \right\}
\]

Fig. 3. Frequency of annual basin yield with equilibrium vegetal cover (southern branch of the Nashua River at Clinton, Massachusetts; \( A = 174 \text{ km}^2 \)).
MODEL VERIFICATION

By assuming that there is to be no import or export of water through mechanisms not considered in deriving (6) this one-dimensional ('point') relationship will be applied to entire natural watersheds. In this application we are, in effect, integrating (7) over the spatially variable soil and climate parameters and variables. We will assume that the groundwater and surface water basins share a common mouth and that yield is adequately approximated by total streamflow. This provides effective spatial integration for the left-hand side of (7). The right-hand side must therefore also be in terms of integrated or 'effective' parameters which may differ from those measured at a single point in the watershed.

We have assembled the available annual streamflow data for two catchments in contrasting climates.

Southern branch of Nashua River near Clinton, Massachusetts. This 108-mi² catchment is located north of Worcester, Massachusetts, and about 50 mi northwest of Boston. Annual precipitation and annual yield observations were published by the U.S. Geological Survey [1940] for the 30-year period of 1904-1933 and are presented here in Figures 2 and 3, respectively, to a normal probability scale using the plotting position of Thomas [1948].

Five years of storm observations at Boston were analyzed by Eagleson [1978a] to obtain the parameters listed in Table 1. These were used in (30) to derive the cdf of \( P_A/m_{PA} \), which is shown by the solid line in Figure 2. Because of its extraordinary fit to the observations this derived distribution will be used here as the basis for deriving the cdf of annual yield.

Santa Paula Creek near Santa Paula, California. This 40-mi² catchment is located in Santa Paula Canyon in Ventura County about 50 mi northwest of downtown Los Angeles. Annual precipitation data were collected by the Ventura County Flood Control District and are presented in Figure 4.

<table>
<thead>
<tr>
<th>TABLE 1. Independent Climate Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>----------------------------</td>
</tr>
<tr>
<td>( m_{PA}, \text{cm} )</td>
</tr>
<tr>
<td>( T_{\infty}, \text{°C} )</td>
</tr>
<tr>
<td>( m_{PA}, \text{d} )</td>
</tr>
<tr>
<td>( m_{PA}, \text{d} )</td>
</tr>
<tr>
<td>( m_{PA}, \text{d} )</td>
</tr>
<tr>
<td>( \kappa )</td>
</tr>
<tr>
<td>( \epsilon ), cm/d</td>
</tr>
</tbody>
</table>

*From method of moments.
†From visual best fit.

TABLE 2. Independent Parameters of Representative Soils

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Clay</th>
<th>Clay Loam</th>
<th>Silty Loam</th>
<th>Sandy Loam</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k(1) ), cm²</td>
<td>( 1 \times 10^{-10} )</td>
<td>( 2.8 \times 10^{-19} )</td>
<td>( 1.2 \times 10^{-12} )</td>
<td>( 2.5 \times 10^{-9} )</td>
</tr>
<tr>
<td>( n )</td>
<td>0.45</td>
<td>0.35</td>
<td>0.35</td>
<td>0.25</td>
</tr>
<tr>
<td>( c )</td>
<td>12</td>
<td>10</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>
while streamflow measurements were made by the U.S. Geological Survey at station 11-1135 and are plotted in Figure 5.

Five years of storm data from a Ventura County Flood Control District precipitation recorder at Ferndale Ranch were analyzed by Eagleson [1978a] to obtain the parameters listed in Table 1. These parameters were used in (30) to derive the cdf of $P_A/m_{PA}$, shown by the broken lines in Figure 4. We note that in this case the observations are not fitted very well in the tails by either of the derived distributions, perhaps because of the smallness of the storm sample in only 5 years.

In both cases the rate of potential evaporation was estimated [Eagleson, 1978f] by using mean annual temperature, humidity, wind speed, and cloud cover data from the nearest U.S. Weather Service station reporting these values. The remaining dependent climate parameters were calculated by definition [Eagleson, 1978c] by using the independent values of Table 1.

To calculate the parameters of (6) and (7), it is also necessary to know the independent soil parameters $n$, $k(1)$, and $c$, the water table depth $Z$, the vegetation parameters $M$ and $k_v$, and the surface retention capacity $h_0$. No direct observations of these parameters were available, consequently, 'verification' can consist only of a comparison of the observed yield frequency with that predicted by (31) for various reasonable sets of parameter values.

1. The water table effect is likely to be negligible, thus we will assume, in all cases, that $w = 0 (Z = \infty)$.

2. From an extensive study of natural soils by Holtan et al. [1968] we have selected four consistent sets of independent soil parameters chosen to span the range of observations. These values are listed in Table 2. The corresponding dependent soil parameters were calculated according to their definitions [Eagleson, 1978b].

3. Topographic maps of the watersheds indicate the presence of vegetation in both cases. However, the use of (17), the 'growth equilibrium' hypothesis, obviates the need to estimate the vegetal canopy density $M$. Under this hypothesis [Eagleson, 1978c], $M$ takes on the value $M_0$, which is defined as that at which, for $P_A = m_{PA}$ and the given $k_v$, the average soil moisture $s_o$ is a maximum. In humid climates where the coefficient of variation of $P_n$ is small we expect the natural vegetation to consist primarily of perennial varieties and do not expect a significant year-to-year change in $M$ toward a new equilibrium as the soil moisture level fluctuates. In arid climates, however, where the coefficient of variation of $P_A$ is large, we expect the natural vegetation to contain a significant
percentage of annual grasses whose seeds either germinate or lie dormant, depending upon the precipitation in a given year. In this case we may expect to see a significant fluctuation in \( M \) from year to year. Initially, to demonstrate the sensitivity of \( Y_A \) to soil properties alone, we will keep \( M = M_0 \) in both watersheds regardless of the value of \( P_A/m_P \). These values as determined by (17) are listed in Table 3 for each climate-soil combination.

4. If we assume no limitation in nutrient or light supply, the 'evolutionary equilibrium' hypothesis for natural vegetal systems \([\text{Eagleson, 1978c}]\) determines the equilibrium plant coefficient \( k_v = k_v_0 \) from that growth-equilibrium state at which the rate of soil moisture utilization \( M_0 k_v \) is a maximum. The values of \( k_v_0 \) as given by (17) and (18) for each climate-soil combination are shown in Table 3.

5. The literature on interception losses suggests an estimate of 1 mm for the surface retention capacity \( h_0 \).

Writing (6) in the forms

\[
P_A = \frac{E[E_A] J(E, M, k_v) + m_\gamma K(1) s_\gamma^c - T_w}{1 - e^{-G - 2I'(\sigma + 1)\sigma^{-\sigma}}} \tag{32}
\]

for

\[
E[E_A]/m_P < e^{-G - 2I'(\sigma + 1)\sigma^{-\sigma}}
\]

and otherwise,

\[
P_A = E[E_A] J(E, M, k_v) + E[E_A] + m_\gamma K(1) s_\gamma^c - T_w \tag{33}
\]

we can, for each chosen parameter set, assume values of \( s_\gamma \) and calculate the corresponding \( P_A \). Normalization by \( m_P \) gives the value \( x \), which when used with (30) or with an alternate expression for the cdf of \( P_A/m_P \), determines the associated probability. To demonstrate the forecasting of annual yield frequency from short records of rainstorms, we will begin by using the cdf of \( P_A/m_P \) derived from 5 years of storm data.

Writing (7) in the forms

\[
Y_A = P_A - E[E_A] J(E, M, k_v) - E[E_A] \tag{34}
\]

for

\[
E[E_A]/m_P < e^{-G - 2I'(\sigma + 1)\sigma^{-\sigma}}
\]

and otherwise,

\[
Y_A = m_\gamma K(1) s_\gamma^c - T_w \tag{35}
\]

yields the value of \( Y_A \) and hence of \( Y_A/m_P \) corresponding to the above probability.

The above procedure is followed repetitively for each of the four representative soils at both Clinton and Santa Paula, and the results are shown by the solid lines in Figures 3 and 5, respectively. In the Santa Paula case, all derived curves are for \( K = 0.25 \).

As an aid in the interpretation of these illustrations, Figure 6 presents the variation of annual bare soil evapotranspiration with the principal soil properties at Clinton and at Santa Paula for the special case \( h_0 = 0 \). In discussing the qualitative differences between these two climates, \([\text{Eagleson, 1978c}]\) noted that in the subhumid case the atmospheric capacity to absorb moisture was limited (relative to the supply), and hence variations in the soil moisture supply caused either by precipitation changes or by variations in soil permeability caused little or no change in the actual evapotranspiration (except for impermeable soils). In the arid climate, however, there is an excess of atmospheric moisture capacity (relative to the supply), and thus the actual evapotranspiration is extremely sensitive to changes in either the precipitation or the soil properties.

Therefore under the most humid conditions, where evapotranspiration is insensitive to \( P_A \), equation (2) gives a nearly linear yield function \( g(P_A) \). Such linearity will preserve the form of the precipitation cdf in transformation to the cdf of yield. Under more arid conditions, where \( Y_A/m_P \) is small, \( E_r \) is sensitive to changes in \( P_A \), and the yield function will become quite nonlinear. In this case, transformation of the precipitation cdf to that of yield will introduce significant distortion. This can be seen by comparing the variance of the yield observations of Figures 3 and 5 with those for the respective precipitation as given in Figures 2 and 4.

In the absence of vegetation the insensitivity of bare soil evapotranspiration to soil properties in the subhumid case is shown by the insensitivity of \( ETA/P_A \) to \( h_0 \) at Clinton and Santa Paula, respectively.
would cause near coincidence of the yield functions at high values of $Y_A/m_P$, for all soils of low $c$. With vegetation, however, differences in properties of the equilibrium cover separate the yields for the permeable soils, as can be seen for the sandy loam and the silty loam at Clinton in Figure 3. For the clay and clay loam soils at this location the high values of $c$ restrict the soil moisture movement, thus modulating the system response in a way which reduces the yield variance while it increases its mean. Of the four soils tested under hypothesized vegetal equilibrium, the silty loam gives the best agreement with observation. This agreement is remarkably good over the full range of the observations. Later we will explore suboptimal vegetal covers for this soil and climate.

In the arid Santa Paula climate (Figure 5) there is a significant difference in the yield variances for the various soil-vegetation combinations, with the less permeable soils again showing the greatest reduction of the precipitation variance and the greatest increase in average yield. In this case, none of the equilibrium soil-vegetation combinations give a very good fit to the observations, but once again the silty loam is the best. Below we will explore suboptimal vegetal covers for this soil and climate.

In both cases it should be noted that the mean yield for the most permeable soil, the sandy loam, exceeds that of the next most permeable soil, the silty loam. This reversal in the general trend toward higher mean yield for less permeable soils was noted in an earlier paper [Eagleson, 1978c] and results from the weak capillary forces in the very permeable soils being unable to hold soil moisture against gravity. The bulk of infiltrated water thus becomes groundwater yield instead of evapotranspiration.

Recognizing that these two vegetal systems may not have reached equilibrium due to limitations by nutrition, light, or some other ecological factor, we will explore suboptimal $M-k_v$ combinations. Reasoning that an energy limitation is most likely (i.e., nutrition or light) and that such a limitation will affect primarily the evolutionary process, we will relax the requirement that $k_v = k_{v0}$, while we retain the short-term growth equilibrium condition $M = M_0$.

Considering the Clinton situation first, we choose two suboptimal values of $k_v$, $k_v = 0.7$ and $k_v = 1.0$, bracketing the equilibrium value $k_v = 0.9$ for the silty loam soil. The condition of maximum soil moisture for these values of $k_v$ determines the respective growth equilibrium values $M_0 = 1$ and $M_0 = 0.8$. The dimensionless yield frequency curves for these suboptimal values are presented in comparison with the optimal values and with the observations in Figure 7. The condition $M_0 = 0.8$, $k_v = 1.0$ brings good agreement between this silty loam soil and the observations. The index of biomass productivity $M_0k_v$ corresponding to this best fit is 0.8 in comparison with the potential (i.e., optimal) value $M_0k_v = 0.9$. We tentatively conclude that the Clinton system is only slightly suboptimal. The limitation is probably one of light.

Turning now to the Santa Paula climate and the silty loam soil, we choose the two suboptimal conditions $M_0 = 0.6$, $k_v = 0.7$ and $M_0 = 0.38$, $k_v = 1.0$. The dimensionless yield frequency curves for these values are presented in comparison with the optimal values and with the observations in Figure 8. All derived curves are for $\alpha = 0.25$. We notice that the silty loam soil with the vegetal condition $M_0 = 0.38$, $k_v = 1$ provides a good fit to the observations for the wet years (i.e., high $Y_A/m_P$) but underpredicts the yield for the dry years. This raises the question of the applicability of a long-term average $M = M_0$ in a climate with high annual variability.

As was discussed earlier, in arid climates we might expect $M$ to change from year to year as annual grasses either germinate or remain dormant in response to fluctuations in precipitation. We can approximate this to the first order by considering the vegetal system to reach a new growth equilibrium each year in which the $M_0$ is defined as that which produces maximum $s_0$ (at a given $k_v$) at each annual value of $P_A/m_P$.

In the manner outlined earlier we select a $P_A/m_P$ and solve the water balance equation for $s_0 = s(M)$. Differentiating
EAGLESON: CLIMATE, SOIL, AND VEGETATION

Fig. 8. Frequency of annual basin yield with suboptimal vegetal cover; $\kappa = 0.25$ (Santa Paula Creek near Santa Paula, California; $A = 64$ km$^2$).

Fig. 9. Potential sensitivity of equilibrium vegetal canopy density to changes in annual precipitation.
tionally) and is much less productive biologically than is the Clinton system (\(M_0k_0 = 0.38\) versus \(M_0k_0 = 0.8\)).

Also shown on Figure 8, as a basis for judging sensitivity, is the bare soil (\(M = 0\)) yield frequency for the silty loam soil and \(k = 0.25\). Notice that for this soil the yield is lower in the absence of vegetation than it is with vegetal cover. This has been explained elsewhere [Eagleson, 1978c] and will occur in all cases. It results from the evapotranspiration formulation [Eagleson, 1978b, d] which reduces the exfiltration velocity from bare soil due to moisture extraction by plant roots. This puts maximum \(s_o\) and hence maximum yield at a nonzero value of \(M\).

When assessing the value of this method for parameterizing the soil and vegetation, we must remember two things. First, we began this test with four sets of representative soil characteristics which we did not alter in the fitting process. Even better agreement could be obtained in each case if we varied the soil parameters \(k(1), c,\) or \(n\) around the values used here. Second, we must not let inadequacies in the derived cdf of \(P_A/m_P\) obscure the basic accuracy of the water balance model relating yield to precipitation.

To eliminate this second uncertainty, we will examine the sensitivity of the cdf of annual yield to the underlying cdf of annual precipitation for the Santa Paula case. In Figure 4, three cdf of Santa Paula precipitation are presented in comparison with the observations. The one being used until now is that derived from (30) with \(k = 0.25\) and \(m_P = 15.7\), as determined by 'best visual fit' to 5 years of storm data. Using the objective method of moments fitting procedure to the same 5 years of storm data gives \(k = 0.37\) and \(m_P = 15.7\), which, with the use of (30), gives the other broken line in Figure 4. As was noted earlier, neither of these derived distributions fits the annual observations well in the tails. A third, empirically determined, distribution which does fit the annual observations well over their full range is the log normal distribution, which is shown as the solid line (best visual fit) in Figure 4.

If we allow \(M_0\) to vary with \(P_A\) as described above, each of these three cdf of \(P_A/m_P\) is used to derive the respective cdf of \(Y_A/m_Y\) and gives the results of Figure 10. The log normal cdf of \(P_A/m_P\) transforms with this water balance model into a remarkably accurate cdf of annual yield. The others are inaccurate in the tails; this is attributable to the derived cdf of annual precipitation.

**SUMMARY AND CONCLUSIONS**

The average annual soil moisture water balance equation has been used to define the average annual soil moisture in the surface boundary layer. This soil moisture is used in the equation for average annual basin water yield to give a first-order approximation to the function relating annual yield to annual precipitation.

This annual yield function is used to transform the cdf of annual precipitation into the cdf of annual yield. The latter is evaluated for two disparate climates, that of Clinton, Massa-
the records of observed storm characteristics and a range of representative soil properties. Because no observations of the vegetal properties \( M \) and \( k_r \) were available, values were selected according to an equilibrium hypothesis based upon natural selection.

In the subhumid climate (Clinton), one of the representative soils and the associated equilibrium vegetal cover gives a yield frequency curve based upon only 5 years of storm precipitation data which is in close agreement with the historical record. This indicates the following are true in this case.

1. The climate is humid enough (i.e., sufficient events per year) that the annual precipitation statistics are well defined by such a short period of observation.
2. The actual soil is probably relatively permeable, since for such soils in humid climates the evapotranspiration and hence the yield is insensitive to soil properties.
3. The natural vegetal system is essentially water limited, since the product of its fitted \( M \) and \( k_r \) is close to the hypothesized optimum value for water-limited systems.
4. The water balance model describes the behavior of this system very well.

In the arid climate (Santa Paula), it is necessary to use a variable vegetal canopy density and the observed cdf of annual precipitation in order to get close agreement between the historical and the derived cdf of annual yield for one of the representative soils. This indicates the following are true in this case.

1. The climate has too few events per year to base a derived cdf of annual precipitation upon storm observations over only 5 years, or the Poisson arrival process does not, for some reason, represent the real situation adequately. This could be tested by using derived cdf based on successively longer observational periods.
2. The natural vegetal cover fluctuates considerably from year to year in response to variations in annual precipitation.
3. The natural vegetal system is limited, probably by nutrition, since the product of its fitted \( M \) and \( k_r \) is significantly less than its optimum water-limited value.
4. The water balance model describes the behavior of this system very well.

It is shown that the yield from natural catchments is reduced in the absence of vegetation as a consequence of the associated reduction in average soil moisture. This has implications for the study of deforestation and overgrazing.

The results give confidence not only in the yield model developed in this and preceding papers [Eagleson, 1978a, b, c, d, e] but also in the feasibility of yield frequency estimation from relatively short records of precipitation events. The demonstrated sensitivity and accuracy suggest the utility of this model for "parameterization" (i.e., selection of average effective values), with respect to soil and vegetation parameters, of entire drainage basins. This is a necessary but troublesome requirement of land surface models for use in climate modeling.

The results also suggest an application for judging whether a natural vegetal system is limited by water or by another factor such as nutrition or light and for comparing the potential and actual biological productivity of such systems.

**Notation**
- \( A \) drainage area, square kilometers.
- \( c \) pore disconnectedness index.
- \( d \) diffusivity index.
- \( E \) exfiltration parameter.
- \( E_{pa} \) annual potential evapotranspiration from soil moisture, centimeters.
- \( E_{Ra} \) annual evapotranspiration, centimeters.
- \( e_{so} \) potential soil surface evaporation rate, centimeters per second.
- \( \bar{e}_{so} \) time average rate of potential soil surface evaporation, centimeters per second.
- \( G \) gravitational infiltration parameter, equal to \( \alpha A_o \).
- \( h_o \) surface retention capacity, centimeters.
- \( I_A \) annual infiltration, centimeters.
- \( K(1) \) saturated hydraulic conductivity, centimeters per second.
- \( k_o \) potential transpiration efficiency.
- \( k(1) \) saturated intrinsic permeability, square centimeters.
- \( M \) vegetated fraction of surface.
- \( m \) pore-size-distribution index.
- \( m_c \) mean climatic conditions, equal to \( P_A = m_{pa} \).
- \( m_{pa} \) average annual precipitation, centimeters.
- \( m_{st} \) mean time between storms, days.
- \( m_{sd} \) mean storm duration, days.
- \( m_m \) mean number of storms per year.
- \( m_r \) mean length of rainy season, days.
- \( n \) medium effective porosity, which equals volume of active voids divided by total volume, counting variable.
- \( P_A \) annual precipitation, centimeters.
- \( R_{ga} \) annual groundwater runoff, centimeters.
- \( R_s \) annual surface runoff, centimeters.
- \( s_o \) time and spatial average soil moisture concentration in surface boundary layer.
- \( T \) one year, seconds.
- \( T_a \) temperature of near surface air, degrees Centigrade.
- \( T_o \) normal annual temperature, degrees Centigrade.
- \( w \) upward apparent pore fluid velocity representing capillary rise from the water table, centimeters per second.
- \( x \) value of normalized annual precipitation.
- \( Y_A \) annual yield, centimeters.
- \( Z \) depth to water table, centimeters.
- \( z \) value of dimensionless yield.
- \( \alpha \) reciprocal of average rainstorm intensity, equal to \( m_{-1} \), seconds per centimeter.
- \( \beta \) reciprocal of average time between storms, equal to \( m_{-1} \), days \(^{-1} \).
- \( \delta \) reciprocal of average storm duration, equal to \( m_{-1} \), days \(^{-1} \).
- \( \eta \) reciprocal of mean storm depth, equal to \( m_{-1} \), cm \(^{-1} \).
- \( \kappa \) parameter of gamma distribution of storm depth.
- \( \lambda \) parameter of Gamma distribution of storm depths, equal to \( \kappa / m_{st} \), cm \(^{-1} \).
- \( \nu \) counting variable for number of storms.
- \( \sigma \) capillary infiltration parameter.
- \( \tau \) length of rainy season, days.
- \( \phi_o \) dimensionless desorption diffusivity.
- \( \phi_s \) dimensionless sorption diffusivity.
- \( \Psi(1) \) saturated soil matrix potential, centimeters (suction).
\( \omega \) storm arrival rate, days\(^{-1}\).

\( \text{Cov} [a, b] \) covariance of \([a, b]\).

\( E[\cdot] \) expected value of \([\cdot]\).

\( F[\cdot] \) cumulative distribution function.

\( g(\cdot) \) functional notation.

\( J(\cdot) \) evapotranspiration function.

\( P[\cdot] \) Pearson's incomplete gamma function.

\( \text{Var}[\cdot] \) variance of \([\cdot]\).

\( \Gamma(\cdot) \) gamma function.

\( \gamma[a, x] \) incomplete gamma function.

\( \hat{\cdot} \) estimate of \( \cdot \).

Acknowledgments. This work was performed in part while the author was a visiting associate in the Environmental Quality Laboratory of the California Institute of Technology on sabbatical leave from MIT. Preparation of algorithms for fitting the yield-frequency model to observations was conducted at MIT by Pedro Restrepo, research assistant in the Department of Civil Engineering, under NSF grant ENG 76-11236. Publication has been made possible by a grant from the Sloan Basic Research Fund of MIT.

REFERENCES


(Received August 23, 1977; revised January 9, 1978; accepted February 16, 1978.)