Parsimonious monthly rainfall–runoff models for humid basins with different input requirements

C.-Y. Xu* & G. L. Vandewiele

Laboratory of Hydrology, Free University Brussels, Pleinlaan 2, B-1050 Brussels, Belgium

(Received 1 February 1994; accepted 29 September 1994)

The availability of data is a major problem for the widespread application of rainfall–runoff models. In order to assist the practicing hydrologists in the selection of suitable models, four levels of monthly water balance models are developed requiring different sets of input data. Essentially four types of models are considered according to the inputs necessary, which are respectively: precipitation and potential evapotranspiration; precipitation, temperature and humidity; precipitation and temperature; precipitation alone. The outputs from the models are monthly river flow and other water balance components. The models are examined on 91 small and medium sized basins, of which 85 are situated in Belgium and 6 in southern China, and they are capable of reproducing both the magnitude and timing of monthly and seasonal runoff, as well as changes in soil moisture conditions. Results are satisfactory and there are rarely any differences between the model types, although the amount of information used is very different. Calibration is by automatic optimization without any subjective element. Application in forecasting and Monte Carlo simulation are exemplified.

INTRODUCTION

Lumped rainfall–runoff models compute the runoff of a river (the output variable) on the basis of a number of measured variables like precipitation, potential evapotranspiration, air humidity, and temperature (the input variables). The input variables are areal quantities, to be computed by the Thiessen method or other techniques, so that they represent the entire river basin.

The usefulness of such models is at least twofold: they can produce forecasts of runoff and they can be used for generation of long runoff series. The latter application goes as follows. In many practical cases runoff data are short (some 10 or 20 years); too short for computing return periods of more or less rare hydrological events like droughts. Nevertheless a reliable rainfall–runoff model can be calibrated on such relatively short data. The input data on the contrary are often much longer (typically several tens of years). The rainfall–runoff model thus can simulate several tens of years of runoff, which can be described in its turn by a time series model.

With the latter model it is then possible to generate a Monte Carlo simulation of runoff of arbitrary length (1000 years for example).

Both runoff forecasts and simulation can be used in design and control of water resources systems. In water supply projects (irrigation, drinking water, hydro-electricity), the bulk of the problems can be solved with the help of monthly models, since those models translate the seasonal behaviour of the river, and since features like peak flow and other short term properties are less important. Therefore, monthly models are important in those applications, as they are easier to formulate and to calibrate than, say, daily models. The time step of one month has been chosen in the present paper in view of the above mentioned considerations, and not as a 'natural' time step related to basin size.

Thornthwaite and Mather,13 Palmer1 and Thomas11 defined monthly models, which are primarily meant to be water balance models for agricultural use. Alley1 defined two variants of Thornthwaite and Mather's models and studied the performances of those five models. He calibrated the model by optimization. Vandewiele et al.15 introduced a class of new models and compared the latter with Alley's models by application to basins in Belgium and China. These new models turned out to be much better, especially in explaining the seasonality of runoff. A complete discussion with many technical details and examples has been published in a booklet by Vandewiele et al.16

The aim of the present paper is to introduce a set of
monthly rainfall–runoff (or water balance) models which are operational in the sense that they are easily applicable by engineers in charge of water resources projects. To be operational a model has to fulfill at least two conditions: first the necessary data have to be easily obtainable, and second the calibration procedure must be easy.

Four model types with different input requirements are presented here:

- Type 1: precipitation and potential evapotranspiration.
- Type 2: precipitation, temperature and relative air humidity.
- Type 3: precipitation and temperature.
- Type 4: precipitation only.

With respect to the condition of easy calibration, the procedure used is automatic optimization. Therefore the models have to be parsimonious as regards the number of parameters (unknown constants to be estimated, which are characteristic of the basin).

The present models are applicable in humid basins (up to about 5000 km$^2$) without important snow and frost. Apparently semi-arid basins have to be modeled in another way (Vandewiele et al.,$^{15}$ Xu$^{17}$). Artificial abstraction or introduction of water in the basin, and important influences of lakes have to be taken into account separately.

Type 1 models have been introduced and thoroughly discussed by Vandewiele et al.$^{15,16}$ In particular statistical analysis has been treated, including the split sample technique which has been applied for demonstrating the model’s ability towards extrapolation and its insensitivity to calibration period. Moreover, the new models have been compared to a number of already existing ones, and the former have been demonstrated to perform much better, in particular in explaining seasonality. In the present paper these topics are not repeated except when necessary for a clear comprehension of the subject matter; in particular no further comparison with other models is made.

All four model types are first defined by their equations. Then some topics in statistical analysis are explained. The principal results from a representative sample of 26 basins (out of 91 basins) are then tabulated and discussed. Finally it is shown how to apply the models in order to obtain forecasts and return periods of water shortages. The sections on Type 1 model equations and on statistical analysis are taken over from Vandewiele et al.$^{15}$ whereas Types 2, 3 and 4 model equations are new together with their comparative results.

**MODEL EQUATIONS**

**General model structure**

One considers monthly areal precipitation $p_t$ and eventually other input variables to be the inputs transformed by a so-called filter into computed discharge $d_t$, which should be the monthly observed runoff $q_t$, except for a deviation $u_t$ ($t$ is time in months). The input series are the ‘observed’ factors. Clearly, discharge is also influenced by other phenomena, the ‘unobserved’ factors, such as measurement errors, Thiessen errors, the nonhomogeneity of rainfall during the month, model imperfections, etc. Accordingly, $q_t$ is considered to be a random variable resulting from a deterministic function (rainfall–runoff filter) of the inputs on the one hand and of a random deviation $u_t$ on the other. This is represented in Fig. 1.

For statistical analysis it is convenient to have homoscedastic deviations (i.e. common variance $\sigma^2$ for all deviations $u_t$). Vandewiele et al.$^{16}$ show that a square root transformation of flow solves this problem. Therefore it is supposed that

$$\sqrt{q_t} = \sqrt{d_t} + u_t$$  \hspace{1cm} (1)$$

with

$$u_t \sim N(0, \sigma^2)$$ \hspace{1cm} (2)$$

i.e. $u_t$ is normally distributed with zero expectation and common variance $\sigma^2$, the so-called model variance. Moreover deviations are assumed to be independent, i.e. that for all $t$

$$E(u_t u_{t-1}) = 0$$ \hspace{1cm} (3)$$

where $E$ is the expectation operator.

The independence of the $u_t$ has been discussed by Vandewiele et al.$^{15,16}$, and it turned out to be a good hypothesis as compared with other transformations.

**General filter structure**

From past and present values of the input series a new time series $m_t$ is computed which represents the state of

![Fig. 1. General structure of rainfall–runoff models.](image)
the catchment at the end of month $t$, and is to be interpreted as a moisture index, soil moisture content or storage, and it summarizes the memory of the catchment. This is expressed by the balance equation

$$m_t = m_{t-1} + p_t - r_t - d_t$$

(4)

where $r_t$ is actual evapotranspiration during month $t$. All quantities in the balance equation are expressed in mm depth.

A distinction is made between slow flow $s_t$ and fast flow $f_t$ such that

$$d_t = s_t + f_t$$

(5)

Slow and fast discharges can more or less be interpreted as baseflow and direct runoff, although they are not identical to them, since it is impossible to distinguish properly with a monthly model between baseflow and direct runoff.

Moreover, $r_t$, $s_t$ and $f_t$ are time invariant functions of $p_t$, $m_t$ and eventually actual values of other input variables: $r_t = r(p_t, m_{t-1})$, other inputs in month $t$; $s_t = s(p_t, m_{t-1})$, other inputs in month $t$; $f_t = f(p_t, m_{t-1})$, other inputs in month $t$). Rainfall–runoff filters differ by their functions $r(\cdot)$, $s(\cdot)$ and $f(\cdot)$.

Also $m_t$ itself might be included in these functions for evident reasons. Vandewiele et al.\textsuperscript{15} discussed this possibility and concluded that it does not improve the quality of the model. It is not considered anymore in the present paper.

**Evapotranspiration equations for Type 1 models**

In Type 1 models potential evapotranspiration (PET) is measured by pan evaporation or computed by a formula like Penman\'s, based on the measurement of several variables. So PET is an input series. Many proposals have been formulated for the computation of monthly actual evapotranspiration, $r_t$ (Roberts,\textsuperscript{9} Dyck,\textsuperscript{1} Vandewiele et al.,\textsuperscript{15} Xu\textsuperscript{17}).

For computing monthly actual evapotranspiration $r_t$, two quantities (among others) are important: the monthly potential evapotranspiration $e_t$ (evaporation for short), and the available water $w_t$ during month $t$ defined as

$$w_t = p_t + m_{t-1}^+$$

(6)

where $m_{t-1}^+ = \max(m_{t-1}, 0)$ is the available storage.

For evident reasons, a good evapotranspiration equation must be such that

$$r_t \text{ increases with } e_t \text{ and } w_t$$

(7)

$$r_t = 0 \text{ when } w_t = 0 \text{ or } e_t = 0$$

(8)

$$r_t \leq e_t \text{ and } r_t \leq w_t$$

(9)

$$r_t = e_t \text{ when } w_t \rightarrow \infty$$

(10)

Two equations proved to be efficient with the data used in the paper. The first one is

$$r_t = \min\left\{e_t(1 - a_1^{w_t/\max(e_t, 1)}, w_t)\right\}$$

(11)

where the symbol $a_1$ is a parameter (an unknown constant to be estimated), which is a characteristic of the river basin under study. This parameter is constrained by $0 \leq a_1 \leq 1$ because of the conditions (7) to (10). Equation (11) is represented in Fig. 2. The expression $\max(e_t, 1)$ is used in the exponent instead of $e_t$, itself, in order to avoid occasional division by zero.

The second equation, based on Romanenko\textsuperscript{10} (see also Hounam\textsuperscript{3}), is

$$r_t = \min\left\{w_t(1 - e^{-a_2(e_t)}), e_t\right\}$$

(12)

with parameter $a_2$ constrained by $a_2 \geq 0$. Equation (12) is represented in Fig. 3. Both equations fulfill conditions (7) to (10).

An important difference from the point of view of interpretation between formulas (11) and (12) is that in (12) one has $r_t = w_t$ when $e_t \rightarrow \infty$ whereas this is not the case in eqn (11). This condition in fact requires that all available water is used up in evapotranspiration when the available energy (measured by $e_t$) is very great.

**Evapotranspiration equations for Type 2 models**

Type 2 models are designed for the situation where the PET data is not available. This is very often the case in areas away from the principal meteorological stations. A first idea is to replace the input series $e_t$ by a simple function of temperature $c_t$ and relative humidity $h_t$, which are two dominating quantities influencing $e_t$. The function used is

$$e_t = a_4(c_t^+)^2(100 - h_t)$$

(13)

where $a_4$ is a positive valued parameter, which is a
characteristic of the basin under study. Areal air temperature (°C) is taken here.

One remark concerning eqn (13) is that this equation is not intended to compute physically relevant potential evapotranspiration; the resulting \( e_t \) series here is just an auxiliary series internal in the model. The inclusion of this equation in the model was in keeping with the original purpose of the study of developing monthly water balance models that required a minimum of readily available input data. No special claim is made concerning the correctness of the internal workings of the model, as is apparent from the simplicity of the model. This is acceptable, since the objective here is to compute monthly streamflow series; so the \( e_t \) series estimated by eqn (13) may lose its true physical meaning to a certain extent.

To obtain actual evapotranspiration, eqn (13) then is introduced in eqns (11) or (12). As a consequence eqn (13) is to be considered as a kind of driving force for the computation of actual evapotranspiration.

**Evapotranspiration equations for Type 3 models**

In many regions of the world evaporation data or other meteorological data are not available, except perhaps temperature data. It was to meet the need of those users who have very limited data that Type 3 models were built. An idea is to remove humidity data and to express \( e_t \) as a simple function of the most important factor, viz. temperature

\[
e_t = a_4(c_t^3)^2
\]  

(14)

The same remark as for eqn (13) is valid. Equation (14) is introduced in (11) or (12), and (14) has the role of a driving function.

**Evapotranspiration equations for Type 4 models**

On the basis of the following three considerations type 4 models were constructed.

(i) The streamflow is controlled primarily by variations in precipitation. As the time period lengthens, say, over a month or a year, the relationship between rainfall and runoff becomes simpler, and the determination of water yield from a catchment can be accomplished satisfactorily by using relationships between totals of monthly or annual rainfall and runoff.

(ii) Rainfall data is perhaps more available than any other meteorological data all over the world.

(iii) The PET time series varies more regularly than precipitation and in most regions of the world it is nearly a periodic function with a more or less sinusoidal behavior on a monthly time scale.

This behavior leads to the idea to model \( e_t \) by a truncated Fourier series

\[
e_t = \left( a_4 + a_5 \sin \left( \frac{2\pi}{12}(t - a_6) \right) \right)^+\\
\]

(15)

where again \( t \) is time in months. The plus sign at the end is necessary for avoiding negative values of \( e_t \) which otherwise may occur in rare cases. Again \( a_4, a_5 \) and \( a_6 \) are characteristics of the basin, and just as before the numerical values of the \( e_t \) series do not necessarily have a precise physical meaning. To obtain actual evapotranspiration the driving force (15) is introduced in eqn (11) or (12).

Finally in a trial to diminish still further the amount of information necessary for constructing a monthly water balance model, some input variables like \( e_t, c_t \), and
$h_i$ can be replaced by their mean values for each month. In this way, twelve values of potential evapotranspiration, temperature and humidity data are used throughout, resulting in purely periodic input time series. In this way three new model types are obtained:

- **Type 1M**: series $e_i$ in Type 1 is replaced by monthly means $ar{e}_i$.
- **Type 2M**: series $c_i$ and $h_i$ in Type 2 are replaced by monthly means $ar{c}_i$ and $ar{h}_i$.
- **Type 3M**: series $c_i$ in Type 3 is replaced by monthly means $ar{c}_i$.

In the last three model types only a few years of $e_i$, $c_i$ and $h_i$ data are needed for computing their monthly mean values.

### Slow flow equations

Being similar to baseflow, slow flow depends essentially on the storage in the catchment during the month considered.

The general form of the slow flow equation is

$$s_t = a_2 (m_{t-1}^+)^b_2$$  \hspace{1cm} (16)

where $a_2$ and $b_2$ are non-negative parameters. In practice however $a_2$ and $b_2$ are highly correlated. This results in very difficult calibration and high imprecision of the estimates. Therefore $b_2$ was given three standard values $b_2 = 1/2$ or $1$ or $2$ of which at least one value suited a given river basin. Thus $b_2$ is a discrete parameter, whereas $a_2$ is a continuous one.

### Fast flow equations

Fast flow depends on precipitation $p_t$, on other meteorological conditions as measured by $e_i$, on the state of the basin as measured by the storage and on the physical characteristics of the basin, which are taken into account by the introduction of parameters.

A useful quantity is the 'active' precipitation defined as

$$n_t = p_t - e_i (1 - e^{-p_t/\max(e_i)})$$  \hspace{1cm} (17)

where $e_i$ is PET or one of the quantities defined by eqns (13), (14) and (15).

In many cases it seems to translate the influence of precipitation and meteorological conditions. The $n_t$ function is represented in Fig. 4. This figure shows that the ‘active’ precipitation $n_t$ decreases when potential evapotranspiration increases. On the other hand $n_t$ increases with $p_t$ and is equal to total precipitation $p_t$ minus $e_i$ for heavy rainfall (high $p_t$).

A most efficient equation for fast runoff proved to be

$$f_t = a_3 (m_{t-1}^+)^b_3 n_t$$  \hspace{1cm} (18)

where $a_3$ and $b_3$ are non-negative parameters. For the same reason as in the case of slow flow $b_2$ was given four standard values: $b_2 = 0$ or $1/2$ or $1$ or $2$.

The equation (18) can be seen as a translation of the variable source area concept: the greater the storage $m_{t-1}$, the wetter the catchment, the greater the ‘source’ of fast runoff, the greater the part of the ‘active’ rainfall running off rapidly.

A summary of particular filter structures is shown in Table 1. Together with formulas (1), (2) and (3) the set of models is completely defined.

<table>
<thead>
<tr>
<th>Table 1. Particular filter structures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Potential evapotranspiration or driving force</strong></td>
</tr>
<tr>
<td><strong>Type 1 model</strong> $e_i = \text{pan evaporation or Penman's PET data (mm/month)}$</td>
</tr>
<tr>
<td><strong>Type 2 model</strong> $e_i = a_4 (c_i + 1)^2 (100 - h_i)$ with $a_4 \geq 0$ where $c_i = \text{monthly temperature data (°C)}$, $h_i = \text{monthly relative humidity data (%)}$</td>
</tr>
<tr>
<td><strong>Type 3 model</strong> $e_i = a_4 (c_i + 1)^2$ with $a_4 \geq 0$ where $c_i = \text{monthly temperature data (°C)}$</td>
</tr>
<tr>
<td><strong>Type 4 model</strong> $e_i = \left{ a_4 + a_5 \sin \left[ \frac{2\pi}{12} (t - a_6) \right] \right}^+$</td>
</tr>
</tbody>
</table>

| **Actual evapotranspiration** |
| $r_i = \min \left\{ e_i (1 - e^{-p_t/\max(e_i)}), w_i \right\}$ with $0 \leq a_1 \leq 1$ or $r_i = \min \left\{ w_i (1 - e^{-p_t/\max(e_i)}), e_i \right\}$ with $a_1 \geq 0$ where $w_i = m_{t-1}^+ + p_t$, available water |

| **Slow runoff** |
| $s_t = a_2 (m_{t-1}^+)^b_2$ with $a_2 \geq 0$ and $b_2 = 1/2$ or $1$ or $2$ |

| **Fast runoff** |
| $f_t = a_3 (m_{t-1}^+)^b_3 n_t$ with $a_3 \geq 0$ and $b_3 = 0$ or $1/2$ or $1$ or $2$ where $n_t = p_t - e_i (1 - e^{-p_t/\max(e_i)})$, active precipitation |

| **Total runoff** |
| $d_t = s_t + f_t$ |

| **Water balance equation** |
| $m_t = m_{t-1} + p_t - r_t - d_t$ |
the evapotranspiration eqn (11) or (12), one finds what will be called henceforth a particular filter. Since there are three possible values of $b_2$ and four possible values of $b_3$, and two possible evapotranspiration equations, there are $2 \times 3 \times 4 = 24$ possible filters.

**STATISTICAL ANALYSIS**

Statistical analysis reduces to nonlinear regression analysis. In the present section only a limited number of items are discussed. A full account is given by Vandewiele et al.\textsuperscript{15,16} and Xu.\textsuperscript{17}

**Estimation**

Because of the hypotheses in formulas (1), (2) and (3), maximizing the loglikelihood with respect to the continuous parameters is equivalent to minimizing the sum of squares

$$SSQ = \sum_i (\sqrt{q_i} - \sqrt{d_i})^2 \quad (19)$$

where the sum is extended over all months for which output $q_i$ as well as input data are available. The runoff sequence may show data gaps, but not the input data series, because of the recursive nature of the balance eqn (4).

Criterion function (19) is the only one which corresponds to eqns (1) to (3) according to sound statistical principles. As already remarked before, the square root flow transformation as well as the independence of the deviations have been checked, as reported by Vandewiele et al.\textsuperscript{15} As a consequence (19) is the only valid criterion and no subjective elements enter into the optimization of the model parameters.

The minimization of (19) was performed with the help of the VA05A computer package (Hopper\textsuperscript{3}). The method is a compromise among three different algorithms for minimizing a sum of squares, namely Newton–Raphson, Steepest Descent, and Marquardt.\textsuperscript{3} The quality of optimization was checked by plotting the sum of squares (eqn (19)) versus each of the filter parameters. In that way it was possible to see whether a global minimum was reached. This is done for every model–basin combination. Illustrations of this procedure can be found in Vandewiele et al.\textsuperscript{15,16}

Finally the VA05A-program has been supplemented with a program called E0X4F (NAG Fortran Subroutine Library\textsuperscript{4}). The results relating to the statistical regression formulation such as variance-covariance matrix, correlation matrix, variance and confidence intervals for the individual parameters can then be calculated. Several models have to be tried in order to find the best values of the discrete parameters and the best evapotranspiration equation.

The model standard deviation is estimated by

$$\sigma = \sqrt{\frac{\text{minimum SSQ}}{N - K}}$$

where $N$ is the number of terms in eqn (19), and $K$ is the number of filter parameters (regression coefficients).

**Quality of model performance**

Evidently model standard deviation $\sigma$ is an inverse measure of model quality. In order to measure model quality by a dimensionless quantity, the coefficient of variation of model runoff is used. Vandewiele et al.\textsuperscript{15} show that for mean runoff the latter is equal to

$$mcv = \frac{\sigma \sqrt{4d + 2\sigma^2}}{d + \sigma^2}$$

where $d$ denotes overall mean computed runoff. For small $\sigma$ (the practical case) this quantity is approximately proportional to $\sigma$.

In case a basin has a high coefficient of variation of observed runoff, it can be expected that mcv will be high also. Therefore a quality measure comparable for different basins is

$$Q = \frac{ocv}{mcv} \quad (20)$$

where ocv is defined as

$$ocv = \frac{s}{\bar{q}}$$

with

$$\bar{q} = \frac{1}{N} \sum q_i \quad s = \sqrt{\frac{1}{N} \sum (q_i - \bar{q})^2}$$

Since mcv is the denominator in eqn (20), the quantity $Q$ is proportional to model quality. Because of the 'norming' by ocv, it can be expected that, between different basins, the quantity $Q$ will not vary much. This is brought out by subsequent results.

Many other quality measures are defined in the literature. It is not to be expected that the conclusions of the present comparative study are much influenced by the choice of the quality measure.

**Residual seasonality test**

A test which turns out to be very severe is the residual seasonality test, which checks whether the seasonal character of runoff has been completely explained by the model. Therefore the residuals are grouped in four seasons and it is tested whether their arithmetic mean is significantly different from zero. If yes the runoffs in that season have not been adequately described by the model. If no the model is satisfactory as for this aspect. This test is performed at a 5% significance level.
Table 2. Information on 91 basins

<table>
<thead>
<tr>
<th>Region</th>
<th>Northern Belgium</th>
<th>Southern Belgium</th>
<th>Southern China</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of basins</td>
<td>65</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>Area (km²)</td>
<td>16–3190</td>
<td>68–3626</td>
<td>385–2000</td>
</tr>
<tr>
<td>Precipitation</td>
<td>60–70</td>
<td>80–100</td>
<td>120–160</td>
</tr>
<tr>
<td>(mm/month)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potential</td>
<td>50–60</td>
<td>50–60</td>
<td>70–85</td>
</tr>
<tr>
<td>evapotranspiration (mm/month)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Runoff coefficient (%)</td>
<td>18–54</td>
<td>25–59</td>
<td>54–76</td>
</tr>
</tbody>
</table>

Coefficient of variation on monthly scale

| Precipitation | 0.50–0.60       | 0.50–0.60       | 0.75–0.95      |
| Potential     | 0.70–0.75       | 0.70–0.75       | 0.80–0.95      |
| evapotranspiration |        |                  |                |
| Runoff        | 0.26–1.18       | 0.57–0.99       | 0.90–1.03      |

RESULTS AND DISCUSSION

The models are applied to 85 Belgian catchments and 6 Chinese catchments representing a wide range of climatic conditions and areas. A summary of relevant catchment information is shown in Table 2. For example, a result obtained with a Type I model is shown in Fig. 5.

The seven model Types 1 to 4 and 1M to 3M were optimized on these 91 catchments. Table 3 shows the results for a representative set of 26 catchments. Two critical aspects are retained in Table 3: the quality Q as defined in eqn (20) and the number of seasons (out of four seasons) with significant residuals. Table 3 shows that these seven model types perform equally well except in a very few cases. Model quality Q (first number in the table) does not differ much among different models for the same basin; the difference is never greater than 20%, whereas only in 5 basins (out of 26) the difference exceeds 10%. This has to be compared with the half width of the confidence interval of Q at 95% confidence level, which is about 10%. Seasonality is very well explained since only five seasons (out of four seasons per year × 158 basin and model combinations = 632 tests) show significant residuals. That means that only 5/632 = 0.8% of the tests are negative. This is to be compared with the 5% significance level, at which the tests were performed.

Other model checks turn out to give favorable results in nearly all cases: model parameters are significantly different from zero; parameter correlations are small; estimated mean runoff is nearly equal to observed mean runoff; residual autocorrelation is very small; residuals do not show trends or heteroseasticity; extrapolation is satisfactory; model parameters are insensitive to the choice of the calibration period. These tests were performed according to the methodology explained in Vandewiele et al.15

In view of the possible use of the soil moisture storage series in irrigation problems or as part of a drought index or index of productivity within a drainage basin, it is interesting to check whether the $m_i$-series in different model types are more or less the same for the same basin. This would reinforce the physical background of the models. The same may be said about the actual evapotranspiration $r_i$-series, being another term in the water balance.

This has been checked for all model types and several basins; the results show that both $m_i$ and $r_i$ series are reasonably consistent in different model types for the same basin. An example is shown in Fig. 6. This means that the state variables simulated by these new models are much more reasonable than what has been obtained by Alley¹ in his comparative study of five monthly water balance models, where he concluded that ‘state variables simulated by different models may be quite different. ...This is an unrealistic simulation of end of month soil moisture storage’.

AN APPLICATION TO RETURN PERIODS OF WATER SHORTAGES

A direct application of the models to water project design is as follows. On a given river a dam is to be built in order to meet a given constant water demand. The problem is to decide on the capacity of the dam. A useful aid to this decision is the knowledge of the return period of water shortages as a function of dam capacity.

![Graph](image-url)  
Fig. 5. Observed and estimated monthly runoff of Xingzi River at Fenghuangshan station (1967–1984) in the humid region of southern China.
Table 3. Model quality $Q$ (eqn (20)) and number of seasons with significant residuals from different model types

<table>
<thead>
<tr>
<th>Basin</th>
<th>Area (km$^2$)</th>
<th>1</th>
<th>1M</th>
<th>2</th>
<th>2M</th>
<th>3</th>
<th>3M</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Northern Belgium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D4011</td>
<td>243</td>
<td>2.77-0</td>
<td>2.74-0</td>
<td>2.78-0</td>
<td>2.83-0</td>
<td>2.65-0</td>
<td>2.75-0</td>
<td>2.79-0</td>
</tr>
<tr>
<td>A516</td>
<td>46</td>
<td>2.49-0</td>
<td>2.52-0</td>
<td>2.40-0</td>
<td>2.57-0</td>
<td>2.40-0</td>
<td>2.52-0</td>
<td>2.60-0</td>
</tr>
<tr>
<td>A32</td>
<td>19</td>
<td>2.36-0</td>
<td>2.32-0</td>
<td>2.34-0</td>
<td>2.33-0</td>
<td>2.29-0</td>
<td>2.28-0</td>
<td>2.33-0</td>
</tr>
<tr>
<td>A565</td>
<td>167</td>
<td>2.26-0</td>
<td>2.13-0</td>
<td>2.33-0</td>
<td>2.27-0</td>
<td>2.36-0</td>
<td>2.52-0</td>
<td>2.16-0</td>
</tr>
<tr>
<td>A24/2</td>
<td>44.8</td>
<td>2.10-0</td>
<td>2.04-0</td>
<td>2.09-0</td>
<td>2.08-0</td>
<td>2.08-0</td>
<td>2.08-0</td>
<td>2.08-0</td>
</tr>
<tr>
<td>D0371</td>
<td>66.4</td>
<td>3.27-0</td>
<td>3.03-0</td>
<td>2.92-0</td>
<td>2.87-0</td>
<td>2.93-0</td>
<td>2.89-0</td>
<td>3.01-0</td>
</tr>
<tr>
<td>D0731</td>
<td>468</td>
<td>2.69-0</td>
<td>2.54-0</td>
<td>2.64-0</td>
<td>2.60-0</td>
<td>2.65-0</td>
<td>2.60-0</td>
<td>2.56-0</td>
</tr>
<tr>
<td>D1221</td>
<td>2163</td>
<td>2.56-0</td>
<td>2.55-0</td>
<td>2.63-0</td>
<td>2.66-0</td>
<td>2.59-0</td>
<td>2.61-0</td>
<td>2.64-0</td>
</tr>
<tr>
<td>D1521</td>
<td>810.5</td>
<td>3.18-0</td>
<td>3.16-0</td>
<td>2.95-0</td>
<td>3.12-0</td>
<td>2.78-0</td>
<td>3.07-0</td>
<td>3.29-0</td>
</tr>
<tr>
<td>D1551</td>
<td>208.4</td>
<td>2.45-1</td>
<td>2.44-0</td>
<td>2.31-0</td>
<td>2.40-0</td>
<td>2.32-0</td>
<td>2.41-0</td>
<td>2.42-0</td>
</tr>
<tr>
<td><strong>Southern Belgium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM</td>
<td>1044</td>
<td>3.02-0</td>
<td>2.93-0</td>
<td>2.96-0</td>
<td>2.96-0</td>
<td>2.97-0</td>
<td>3.01-0</td>
<td>3.01-0</td>
</tr>
<tr>
<td>HO</td>
<td>68</td>
<td>2.26-0</td>
<td>2.27-0</td>
<td>2.33-0</td>
<td>2.36-0</td>
<td>2.34-0</td>
<td>2.29-0</td>
<td>2.42-0</td>
</tr>
<tr>
<td>LG</td>
<td>1314</td>
<td>2.89-0</td>
<td>2.74-0</td>
<td>2.77-0</td>
<td>2.75-0</td>
<td>2.80-0</td>
<td>2.77-0</td>
<td>2.84-0</td>
</tr>
<tr>
<td>OU</td>
<td>1597</td>
<td>3.51-0</td>
<td>3.44-0</td>
<td>3.21-0</td>
<td>3.09-0</td>
<td>3.16-0</td>
<td>3.27-0</td>
<td>3.59-0</td>
</tr>
<tr>
<td>SE</td>
<td>1235</td>
<td>2.77-0</td>
<td>2.75-0</td>
<td>2.57-0</td>
<td>2.64-0</td>
<td>2.55-0</td>
<td>2.62-0</td>
<td>2.86-0</td>
</tr>
<tr>
<td>VI</td>
<td>554</td>
<td>3.03-0</td>
<td>2.98-0</td>
<td>2.87-0</td>
<td>2.99-0</td>
<td>2.87-0</td>
<td>3.00-0</td>
<td>3.08-0</td>
</tr>
<tr>
<td>TP5811</td>
<td>3626</td>
<td>2.99-0</td>
<td>2.91-0</td>
<td>2.86-0</td>
<td>2.79-0</td>
<td>2.91-0</td>
<td>2.96-0</td>
<td>2.96-0</td>
</tr>
<tr>
<td>TP774</td>
<td>187</td>
<td>2.54-0</td>
<td>2.57-0</td>
<td>2.50-0</td>
<td>2.58-0</td>
<td>2.45-0</td>
<td>2.57-0</td>
<td>2.56-0</td>
</tr>
<tr>
<td>TP8181</td>
<td>161</td>
<td>3.00-0</td>
<td>2.98-0</td>
<td>2.93-0</td>
<td>3.05-0</td>
<td>2.92-0</td>
<td>3.06-0</td>
<td>3.09-0</td>
</tr>
<tr>
<td>TP8161</td>
<td>127</td>
<td>2.77-0</td>
<td>2.71-0</td>
<td>2.76-0</td>
<td>2.73-0</td>
<td>2.81-0</td>
<td>2.77-0</td>
<td>2.80-0</td>
</tr>
<tr>
<td><strong>Southern China</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEJ1</td>
<td>2000</td>
<td>4.38-1</td>
<td>4.34-0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.59-0</td>
</tr>
<tr>
<td>CHL2</td>
<td>595</td>
<td>4.47-1</td>
<td>4.41-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.28-0</td>
</tr>
<tr>
<td>CFH3</td>
<td>1556</td>
<td>3.87-0</td>
<td>3.81-0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.72-0</td>
</tr>
<tr>
<td>CXG4</td>
<td>1881</td>
<td>3.52-0</td>
<td>3.53-0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.02-1</td>
</tr>
<tr>
<td>CJZ5</td>
<td>385</td>
<td>3.41-0</td>
<td>3.53-0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.62-0</td>
</tr>
<tr>
<td>CST6</td>
<td>1357</td>
<td>4.35-0</td>
<td>4.36-0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.36-0</td>
</tr>
</tbody>
</table>

$^a$In columns, first number is $Q$, second number is the number of seasons with significant residuals.

A fundamental requirement for the study is a long monthly runoff series. Because the observed runoff series is usually short, say, less than 20 years, a long synthetic flow record has to be generated. This can be done in several ways. In the present example a 158-year rainfall record in a nearby rainfall station is used after correction by a factor characteristic for each month. The corrected monthly rainfall series was shown to be statistically indistinguishable from observed areal rainfall in the catchment. By means of a calibrated Type 4 model a 158-year monthly runoff series was then computed, including Monte Carlo simulation of the deviations.

In case a much shorter rainfall series is available a pure runoff model (time series model) has to be used, as explained in the introduction. The Thomas-Fiering model (Thomas & Fiering) can be used or the Non Gaussian Multicomponent model (Zebene & Vandewiele), which has been shown to produce better results.

Let $C_t =$ content of the reservoir at the end of month $t$ (m$^3$), $K =$ capacity of the reservoir (m$^3$), $D =$ monthly demand (m$^3$), and $q_t =$ inflow in the reservoir during month $t$ (m$^3$). Then, $C_t = \min \left[\frac{C_{t-1} + q_t - D}{K}\right]$. There is a shortage during month $t$ when $C_{t-1} + q_t - D < 0$. Count the number of years among the total number of years $N = 158$, during which there is at least one month with shortage. Let this number be $S$. Then, $R = N/S$ is an estimate of the corresponding return period, and $F = 1/R = S/N$ is the frequency in mean number of shortages per year and the probability of a shortage in a given year. The upper and lower boundaries of a 95% confidence interval are approximately (see e.g. Cramer$^2$)

$$F + \frac{1.96^2}{2N} \pm 1.96 \sqrt{\frac{F(1-F)}{N} + \frac{1.96^2}{4N^2}}$$

$$\frac{1 + \frac{1.96^2}{N}}{N}$$

As an example the Lesse River at Eprave station (419 km$^2$) in southern Belgium was considered with 14 years of concurrent runoff and areal rainfall data together with the 158-year rainfall data at Ukkel (near Brussels) at about 90 km distance. It was shown that those two monthly rainfall series are statistically indistinguishable, except for a correction factor for each month. The result is illustrated in Fig. 7. Such a diagram is an aid to the decision of the capacity to be built, since the building cost is a known increasing function of capacity.
APPLICATION TO FORECASTING

Forecasting even on a monthly scale can be used in real time control, e.g. in irrigation and hydropower generation.

The problem is to compute the point forecast of runoff $D_t$ of month $t$, knowing the input data (precipitation and evaporation) up to and including month $t - 1$. From the latter input data, storage $m_{t-1}$, i.e. the storage at the beginning of month $t$ is easily computed with a Type 1 model and one finds

$$D_t = S_t + F_t$$
$$S_t = a_2(m_{t-1}^+)^b$$
$$F_t = a_3(m_{t-1}^-)^b N_t$$
$$N_t = P_t - E_t (1 - e^{-P_t/E_t})$$

where capitals are forecasts. The forecasts $P_t$ and $E_t$ are taken equal to their monthly means. This procedure is based on the fact that $\{p_i\}$ and $\{e_i\}$ are uncorrelated series. Moreover, since residuals are uncorrelated, knowledge of the preceding $q_{t-1}$ does not contribute to the precision of the forecast, except by way of $m_{t-1}$, as expressed by the above equations.

In a similar way forecasts for greater lead times can be computed, and also interval forecasts.

The procedure is applied to two basins: Xingzi River at Fenghuangshan station (1 556 km$^2$) in southern China, and Ourthe River at Angleur station (3 626 km$^2$) in southern Belgium. The results are illustrated in Figs 8 and 9.
CONCLUSION

A set of monthly rainfall runoff models on basin scale (water balance models) have been defined. The models are applicable to small and medium sized basins (up to 5000 km²) in a humid climate without an important snowpack.

Vandewiele et al. demonstrated that these models are of better quality than a number of published models, especially in explaining seasonality of runoff.

These models are parsimonious with respect to the number of parameters used (unknown constants to be estimated, characteristic of the basin under study). As a consequence the calibration is easy and proceeds by automatic optimization excluding subjective elements.

Different model types were defined according to input data requirements:

- Type 1: precipitation and PET.
- Type 2: precipitation, temperature and relative humidity.
- Type 3: precipitation and temperature.
- Type 4: precipitation alone.

The most important conclusion is that PET, temperature and humidity data are not really necessary for computing runoff by these monthly water balance models, provided that more parameters are used. This is particularly important in regions where these data are lacking. On the other hand, models with PET as one of the inputs (Type 1 models) have only three filter parameters and are consequently more suitable for geographical regionalization studies, and when one tries to link parameter values with physical characteristics of the basins (Vandewiele et al.). In this way the models can be used in ungauged catchments.

Easily calibration and easily obtainable input variables are the conditions of an engineer in charge of water resources projects. Two examples of possible engineering applications are discussed: the computation of return periods of water shortage and the forecast of runoff.

ACKNOWLEDGMENTS

The authors thank Miss Ni-Lar-Win and Miss L. Munters who contributed to the first stages of research. Mr J. De Bruyne for help in programming and data handling, Dr G. Demaree and Dr F. Bulot (Royal Meteorological Institute, Brussels) and Prof. R. Verhoeven (State University Ghent, Belgium) for help in data gathering. The Ministry of the Flemish Region and the Government Office for Developing Countries (ABOS) partly funded the present research. The authors are indebted to Prof. A. Van der Beken (Interuniversity Postgraduate Program in Hydrology, IUPHY, Brussels) for his friendly collaboration and sympathy. Thanks are due to the reviewers for their many constructive remarks.

REFERENCES