Nordic Hydrology, 23, 1992, 315-332

Improvements of Runoff Models What Way to Go?

Lotta Andersson

Dept. of Water and Environmental Studies,
Linköping University, Sweden

Model performance before and after the introduction of some alternative routines for calculation of evaporation, snow accumulation and melt with the PULSE/HBV runoff model were compared. The results showed that improvements were, in the best cases, small. Sometimes model fits deteriorated as a result of increased model complexity. On the basis of these, and from other experiences of attempts of model improvements, the success potentials for various efforts of model sophistication are discussed. It is hypothesised that model improvement cannot be achieved by increasing the complexity of some sub-routines, without considering the problems that are linked to spatial resolution of driving variables and the spatial distribution of physiographic parameters. It is suggested that physically based and conceptual schools of modelling can meet in a landscape mosaic context, with development of distributed models, based on information generally available from maps, remote-sensing images and meteorological stations.

Introduction

Predictions of how precipitation causes response in the form of river discharge involve much complexity. Our ability to simulate the different processes that determine the partitioning of water within a catchment is limited not only by our knowledge of the processes involved, but also by the heterogeneity both of relevant landscape factors, and by precipitation and other meteorological factors.

Opinions of the success of a model must be linked to the purpose of the simulation. There are at least three fields of application for runoff-models:

I) Practical interest of management of water resources and estimations of risks
II) Test of hypothesis regarding the scientific understanding of flow processes
III) Forecasts of the effects of future changes of climate, land use and management

For I) the success of a model can be determined from the fit between measured and simulated river flow. The type of river flow events that is of practical interest for a specific study should determine the selection of fitness criterion.

However, a good fit between measured and modelled runoff does not necessarily mean that the used model gives a physically correct description of the involved processes. Usually, a number of different combinations of parameter values or different choices of model sub-routines can yield similar results. For studies within fields II) and III), which are often combined with modelling of chemical or biological variables, it is necessary that the model works right for the right reason (Klêmes 1986). The validation of the model must therefore include more than the integrated response in the form of river flow. Internal validation is, however, an area that still is often neglected also by modelers who work with models that are aimed to be distributed and physically based (Beven 1989a).

In the early stage of model development, large improvements can be achieved by increasing model complexity. Quite soon, model runs with somewhat more sophisticated sub-routines will, however, give similar results as given by a simpler model structure. When increasing the complexity of conceptual models even more, the correlation between measured and simulated variables will sometimes decrease (Calder *et al.* 1983, WMO 1986, Vehviläinen 1986, Andersson and Harding 1991) see Fig. 1. Andersson and Harding (1991) observed that one explanation for this was that a simple formulation can encompass a number of different processes which means that improvement in only one area can destroy a fortuitous performance. However, models that aim to describe all involved processes in a physically correct way have not been shown to improve correlations between simulated and measured variables. The effort that has to be put into calibrations increases, since the numbers of parameters are increased. It is not always possible to use physical reasoning to find proper parameter values. The difficulties to find satisfactory model calibrations for a physically based model (SHE) is, *e.g.* discussed by Lumadjeng (1989). This can usually be attributed to incapability to determine the spatial heterogeneity of parameters and driving variables (Beven 1989b), and to lack of considerations of feedback mechanisms that makes it incorrect to use assumptions that are correct on a laboratory or plot scale for a whole catchment (Morton 1982). This paper describes experiences from some attempts to improve the HBV/PULSE runoff-model (Bergström 1975, Bergström *et al.* 1985) by modifying the evapora-

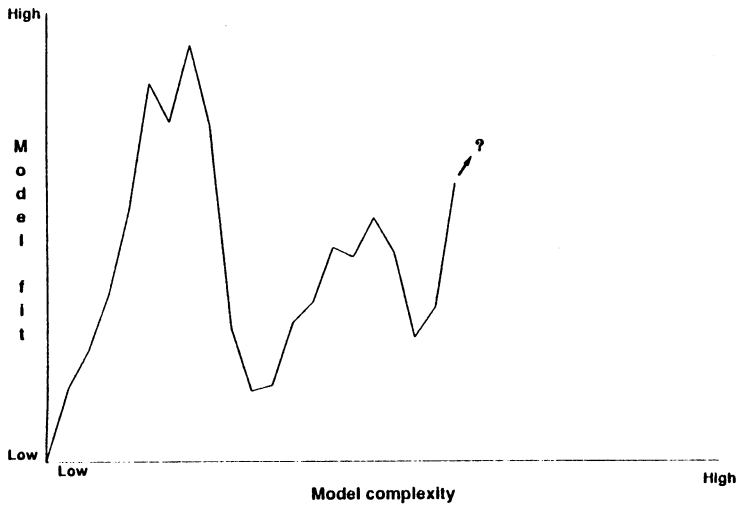

Fig. 1. Typical relations between model complexity and performance for a runoff model, showing a fast increase of fit until further attempts to sophisticate the model fails due to inability to find appropriate parameter and variable values, or due to not considered feedback-mechanisms. The curve is hypothetical, and does not refer to any specified models. The questionmark illustrates that it might never be possible to achieve fits superior to those given by a simpler model.

tion and snow routines. The achieved improvements were small, and sometimes increased complexity decreased the agreement between modelled and measured river flow. The probable reasons for this and some ideas of when, and in that case how, efforts should be made towards further development of existing models are given. The paper can bee seen as a contribution to the discussion of principles of modelling, initiated by Bergström (1991).

Attempts to Improve the HBV/PULSE Model

The HBV model is a conceptual runoff model that has been extensively used in Scandinavia and other countries for operational forecasts of river flow and design floods. It was developed during the 1970s and no major changes of the model structure have taken place since then. A version of the model, called PULSE was, however, developed during the 1980s (Bergström *et al.* 1985). This model version was developed with the aim to enable simulations of hydrochemical dynamics of streamflow.

In general, the experiences from the operational forecasts with the HBV model have been satisfactory. During some periods, however, the model has rather systematically over or underestimated river flows. Since the HBV model uses stan-

dard values of monthly potential evaporation, and a fixed value of the degree-day factor, it has been hypothesised that these failures could be linked to the fact that the effects of extreme meteorological conditions are not satisfactorily described. Studies have therefore been undertaken to test if snow and evaporation routines that consider these extremes, would improve model performance (Andersson 1989, Andersson 1990, Arnér 1991, Lindström and Bergström 1992).

Andersson (1989, 1990) used the PULSE model, whereas Arnér (1991) and Lindström and Bergström (1992) used the HBV model. Both models use a daily time step and are driven by daily values of precipitation and average air temperature, plus standard monthly values of Penman potential evaporation. The differences between the models are mainly in the description of the saturated zone, where the PULSE model is developed to give more realistic description of the dynamics of shallow groundwater. There are also differences in the distribution of elevation and vegetation zones.

The Evaporation Subroutine

The evaporation subroutine in the HBV and PULSE models is driven by monthly standard values of the potential evaporation, calculated by Penman's equation. The estimated soil-moisture deficit limits the actual evaporation.

During 1987, summer and autumn flow were underestimated for many rivers in Sweden. This year was characterised by low air temperatures and few sunshine hours. It was therefore hypothesised that the use of standard values of monthly potential evaporation caused an overestimation of evaporation during this period, when the atmospheric demand of vapour probably was lower than average.

During spring and early summer, low soil temperatures can limit the evaporation (Turner and Jarvis 1975). Neither this, nor the fact that part of the incoming radiation is used for warming up of the soil, or that the development of vegetation during spring is directly dependent on the air temperature is considered by solely using the Penman equation, linked to soil moisture deficits. Another hypothesis was therefore that consideration of the air temperature would improve the evaporation routine.

Methods Used for the Evaporation-Routine Comparison

The fit between measured and simulated runoff when using the traditional evaporation routine, was compared to fits achieved when six alternative routines were used. The models were tested for three river basins (Fig. 2). The roman numbers refer to Table 1.

The original routine (EI) was driven by monthly average (1961-1978) Penman values (Eriksson 1981). For April-September an albedo of 0.12 was used, whereas 0.50 was used for the winter months. Usually, the average of monthly means calculated with data from two meteorological stations within or nearby the catchment were used.

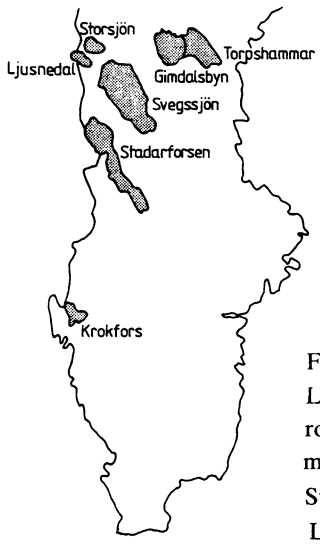

Fig. 2.
Location of river basins used in the comparison of evaporation routines (Andersson 1989; Stadarforsen, Svegsjön, Torpshammar), in the comparison of snow routines (Andersson 1990; Stadarforsen, Svegsjön) (Arnér 1991; Gimdalsbyn,Krokfors, Ljusnedal, Stadarforsen, Storsjön, Svegsjön).

One of the tested routines used the same methodology as the original, but was driven by Penman values calculated for individual months (EII).

Three routines reduced the evaporation when air temperatures were low. The two simplest were based on daily average air temperatures Eq. (1) (Andersson and Harding 1991), which were combined with monthly averages of potential evaporation (EIII), or with values calculated for individual months (EIV)

$$\begin{aligned} T_i < 0 \quad & E_{\max_i} = 0 \\ 0 < T_i < 10 \quad & E_{\max_i} = Ep_i (T/10) \\ T_i > 10 \quad & E_{\max_i} = Ep_i \end{aligned} \tag{1}$$

where

i – day i,
T – daily average of air temperature (°C)
$E_{\max}$ – maximum evaporation (mm)
Ep – Penman's potential evaporation (mm)

The third of the temperature-related routines (Andersson 1988) considered the accumulated climatic conditions during spring Eq. (2). In this routine (EV), Ep was multiplied by an index that was set to zero until five-days average air temperature ($T\text{acc}_i$) exceeded 5°C. After 28 days with accumulated temperature above 5°C the index value of one was reached

$$\begin{aligned} T_{\text{acc}_i} > 5 \quad & \text{Index}_i = \text{Index}_{i-1} + (1/28) \\ T_{\text{acc}_i} < 5 \quad & \text{Index}_i = \text{Index}_{i-1} \end{aligned} \tag{2}$$

Table 1 – Explained variance (R^2) (upper values) and accumulated differences (lower values) between measured and simulated runoff obtained with models using different evapotranspiration routines.

Routine No.	(E I)	(E II)	(E III)	(E IV)	(E V)	(E VI)	(E VII)
Potential evapo-transpiration	Monthly standard Penman	Monthly individual Penman	Monthly standard Penman	Monthly individual Penman	Monthly standard Penman	Monthly standard Penman	Thornth-wait daily values
Tempe-rature effect	No	No	Daily tempera-ture sen-sitive	Daily tempera-ture sen-sitive	5 days accumu-lated tem-perature sensitive	Reduces Penman values by 60% in Jan-May, and by 30% in June	No
Stadar-forsen (1962-87)	0.819 26	0.817 22	0.823 23	0.818 19	0.805 18	0.809 23	0.805 21
Sveg (1965-87)	0.812 30	0.813 27	0.816 26	0.814 25	0.822 25	0.807 30	0.797 30
Torps-hammar (1970-87)	0.803 20	0.794 21	0.823 19	0.814 25	0.798 25	0.796 18	0.813 18
Average	0.812 26	0.809 23	0.821 23	0.816 21	0.804 21	0.804 24	0.798 23

Another routine (EVI), assumed a constant over-estimate of evaporation during spring and early summer Eq. (3)

$$\begin{aligned} &\text{1st January - 31st May} && E_{\max_i} = 0.40\, Ep_i \\ &\text{1st June - 30th June} && E_{\max_i} = 0.70\, Ep_i \\ &\text{1st July - 31st December} && E_{\max_i} = Ep_i \end{aligned} \qquad (3)$$

The last routine (EVII) was based on Thornthwaite's equation (1939). In this empirical formula, potential evaporation was described as a function of the daily average air temperature, the geographical position, and the time of the year.

All the temperature-related routines gave lower average potential evaporation compared to the original routine. In order to achieve similar accumulated runoff

volumes over the calibration period for all model versions, the precipitation-correction factors used for the temperature dependent routines were about 5% lower than those used by the original routine. Otherwise, the parameters determined from calibrations made with the original model were used for all of the tested model versions.

A more recent attempt to improve the evaporation routine (Lindström and Bergström 1992), initiated as a response to the results from Anderson's study (1989), was based on mean daily air temperatures and long-term averages as follows

$$E_i = (1+C_t(T_i-T_n))E_m \quad (4)$$

where

E = potential evaporation,
C_t = empirical model parameter
T_m = monthly long-time average air temperature
E_m = monthly long-time average Penman potential evaporation

Upper and lower boundaries for the evaporation were assumed according to

$$0 < E_i < 2E_m \quad (5)$$

Results from the Evaporation-Routine Comparison

In Fig. 3, the variability of monthly averages of daily Penman potential evaporation is shown. The quartile-intervals were small, reaching a maximum of about 0.4 mm during June. Variations between individual days were larger. In June daily differences of up to 4 millimetres could be found. In the studied runoff-model it is, however, the accumulated effect that evaporation has on the soil-moisture that determines how the evaporation routine will affect the calculated runoff. This type of model is therefore insensitive even to drastic daily variations of potential evaporation.

A comparison of R^2 – values (explained variance) Nash and Sutcliffe (1970) and accumulated volume errors between measured and simulated volume of runoff was made in order to test the influence of the different evaporation routines on the model performance. These objective criteria were combined with visual judgements of the influence of different routines. It was then shown that the effect of the choice of routine in general was difficult to detect by eye although, in some cases differences of the objective criteria were large.

The inclusion of a temperature effect gave a marginal increase of the R^2 – value and a decrease of the volume error. The use of Penman potential values calculated for individual months, instead of standard values, slightly deteriorated the R^2-value but decreased the volume error (Table 1).

It was mainly during years with relatively cold springs, like 1987, that visible

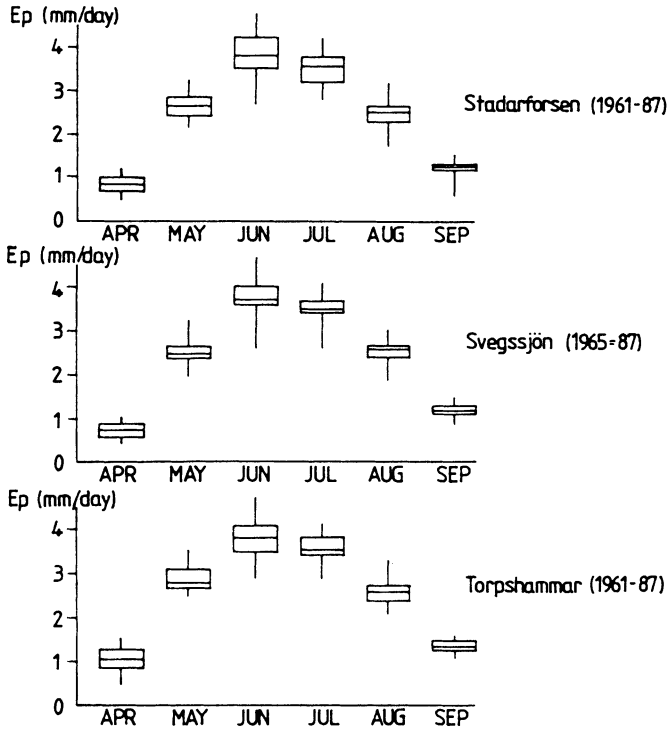

Fig. 3. Interanual variability of monthly averages of potential evaporation *(Ep)* for time periods shown in brackets, expressed as median, quartiles, minimum- and maximum values.

improvements occurred when the temperature effect was considered. The use of individual monthly averages instead of standard values of Penman potential evaporation did only give visible improvement during one year, and then only for one of the studied river basins (Fig. 4).

The routine, suggested by Lindström and Bergström (1992) proved to be successful. Since they used the HBV-model, it was not possible to make direct comparisons with the results from Anderson's temperature-routine, used in connection with the PULSE-model (1989). This was mainly due to different spatial distributions of the snow-calculations. In most cases obtained R^2-values were similar. As an example model runs for Torpshammar 1987 are shown in Fig. 4 (R^2 = 0.86 for Lindström & Bergström's model, compared to 0.85 for Andersson's). The possibility to consider temperature effects also when temperatures are above 15°C, in the routine suggested by Lindström and Bergström, does, however, seem to be more successful in mid- and late summer, which indicates that the temperature effect not only is linked to the temperature-dependent development stage of the vegetation, but also to the atmospheric vapour demand.

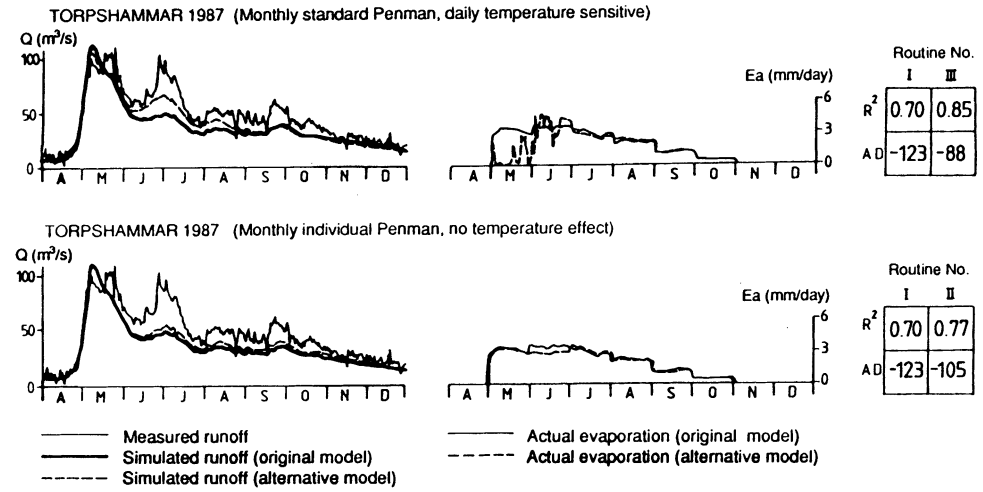

Fig. 4. The only spring event for which the use of individual Penman *Ep* gave a visible improvement of calculated runoff. The larger improvement, given by the simple temperature routine is also shown. Explained variance (R^2) and accumulated difference *(AD)* between measured and simulated runoff is shown.

The Snow Routine

Predictions of spring flow are of vital importance for the regulation of water power dams. The operational forecasts, based on the HBV-model, have in general been satisfactory. Sometimes, the prediction of the start of the spring flow has, however, been unsuccessful. It was hypothesised that this could be due to the fact that the air temperature, at certain meteorological conditions, is not a good indicator of the amount of energy available for melting. Another hypothesis was that the use of daily average air temperatures could underestimate melt during periods when melting takes place during part of the day, although the average air temperature was below the threshold temperature. This is usually linked to cloud-free conditions, when temperature fluctuations are large.

The snow routine in the HBV and PULSE models is based on the degree-day method Eq. (6), driven by daily mean air temperatures. Usually, weighted averages from a number of climatic stations are used. Different degree-day factors are used for forest and open land. The model considers retention Eq. (7) and refreezing Eq. (8) of rain- and meltwater in the snowpack. A threshold temperature determines if precipitation is considered to consist of snow or rain

$$T_i > TT, \qquad \text{MELT}_i = CF(T_i - TT) \text{ (mm)} \tag{6}$$

where

TT – threshold temperature for melting (°C)
CF – degree-day factor

$$\text{RETENTION}_i \equiv CWH_i * SP_i \text{ (mm)} \quad (7)$$

where

RETENTION – the ability of the snowpack to retain rain- and meltwater (mm),
CWH – parameter,
SP – snowpack (mm)

$$\text{REFREEZE}_i \equiv CF * T_i' * CFR \text{(mm)} \quad (8)$$

where

REFREEZE – refreezing of rain- and meltwater in the snowpack (mm),
CFR – parameter

Methods Used for the Snow-Routine Comparison

Measured flows were compared to simulated, using the original snow-routine (S I) and routines based on the energy balance (S II) and on daily maximum and minimum air temperatures (S III, S IV). The roman numbers refer to Table 2.

In order to test if the use of an energy-balanced snow-melt routine improved the model, daily values of the degree-day factor were calculated. The used model (Killingtveit 1988) considered daily means of cloudiness, air temperature, relative humidity, vapour pressure and wind speed. Latitude, time of the year, and the age of the snow are also considered. The needed meteorological variables were only available from synoptic stations. Therefore, fewer climatic stations could be used for the energy balance, than for the air temperature based degree-day method. The daily values of the degree-day factor were achieved by dividing the calculated snow-melt energy (mm/day) by the average daily air temperature. The energy balance was set up to be representative for open land. The degree-day factors for forest were obtained by considering the relative difference between the factors calibrated for forest and open land with the original model structure.

The hypothesis that the use of daily average air temperatures underestimated melt during periods when the average air temperature was below the threshold parameter for melting, but the temperature was above this threshold during parts of the day, was tested by a routine that used daily maximum- and minimum air temperatures (S III).

The daily range of air temperature was assumed to be linearly distributed between the maximum- and minimum values. Snow melt and rainfall were assumed to occur only during the portion of the day when a threshold temperature was exceeded. The temperature variable T, used in the degree-day melting equation (Eq. (6)) was calculated as the average between the daily maximum air-temperature and the threshold temperature. Snowfall and refreezing of water in the snow-

pack were assumed only to take place during the part of the day that corresponds to the portion of the temperature interval that falls below the threshold temperature. The temperature T used in the refreezing equation (Eq. (8)) was calculated as the average between the minimum and the threshold temperature.

In an alternative version of the minimum-maximum snow routine (S IV), skewness between the average of minimum and maximum air temperatures and the mean, calculated from daily measurements was considered in the calculations of melt and refreezing ((Eqs. (9)-(10)).

$$\mathrm{MELT}_i = CF\ (\mathrm{PART}_i^r * T_i^{t>tt}) - (\mathrm{PART}_i^r * T\,\mathrm{SKEW}_i) \quad (9)$$

$$\mathrm{REFREEZE}_i = CF * CFR\,(\mathrm{PART}_i^s * T_i^{t<tt}) + (\mathrm{PART}_i^s * T\,\mathrm{SKEW}_i) \quad (10)$$

where

PART^r – part of the linear distribution of max-min temperatures that is above the threshold temperature,

$T^{t>tt}$ – mean of TT and maximum temperature,

$T\mathrm{SKEW}$ – average of daily maximum and minimum temperatures, subtracted by the accurate estimate of daily mean air temperature,

$\mathrm{PART}^s \equiv 1 - \mathrm{PART}^r$,

$T^{t<tt}$ – mean of TT and minimum temperature.

Except for the degree-day factors in the minimum-maximum snow routines. Andersson (1990) used the parameters calibrated with the original model for all model runs, whereas Arnér (1991) calibrated the different model versions individually.

The objective criteria used were the explained variance R^2 and estimates of the discrepancy between the real and the modelled start of the spring flow.

It was shown that the choice of starting time for the spring flow is crucial for the determination of the most successful model (Arnér 1990). Andersson (1990) defined the start of the spring flow as the time when the flow was raised to a value that corresponded to three times the base flow during the actual spring. The reason for considering each years base flow was that small rises from an unusually high base flow should not be misinterpreted as the start of the spring flow. The variation of the base flows during different years were, however, small. Arnér (1991) used three criteria: 6 times average low flow, average high flow divided by three, and one criterion that considered the gradient of the hydrograph. The criterion with the best correspondence to visual estimates of the start of the spring flow was shown to vary between the studied river basins.

The snow routines were identical for the PULSE and HBV models. Different results can, however, be expected due to dissimilarities in the division into elevation zones. The HBV model used a number of subcatchments, which were individually divided into elevation zones. The PULSE model used the average elevations in a maximum of six subcatchments.

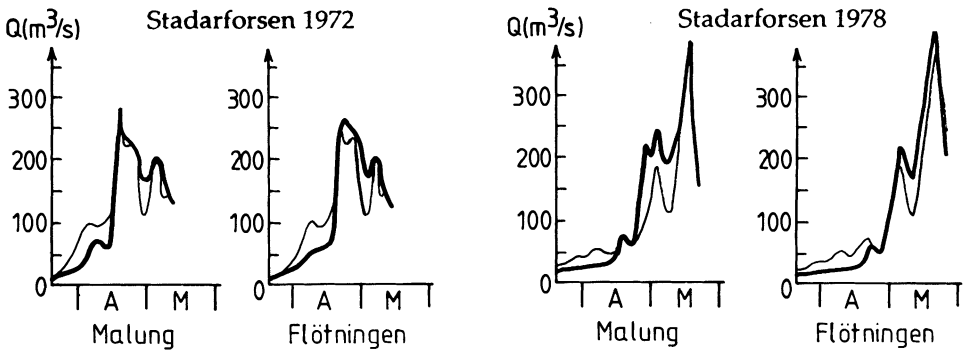

Fig. 5. Measured (thin line) and simulated (bold line) start of the springflow. Simulations made with energy-balanced degree-day factors, using driving variables from two different climatological stations within the river basin.

Results from the Snow-Routine Comparison

The PULSE model was used in a preparatory study (Andersson 1990) to simulate 14 spring flows in two river basins (Fig. 2). Approximately half of the chosen spring flows were from events when the original model gave good estimates, and half when the estimate of the start of the spring flows were unsuccessful.

The energy-balance estimated degree-day factors were shown to be very sensitive to the local meteorological conditions (Fig. 5). The most significant factor was wind speed. In times with wind speeds above 5 m/s, degree-day factors could reach values above 10 mm melting/°C* day.

A comparison of the result with the different snow-routines is shown in Table 2 to which the roman numbers below refer. This study showed a small improvement, when substituting the fixed degree-day factors (S I) with energy-balance calculated factors (S II). The dependence of point measurements of a number of meteorological factors that are more spatially heterogeneous than air temperature, did, however sometimes cause severe over- or underestimates of the melting.

A more significant improvement was achieved when using the minimum-maximum air-temperature based snow-routines (S III, S IV). The simplest version, that did not consider skewness was the most successful (S III). In Fig. 6, examples of simulations with different routines are shown.

The rather promising results with the simplest of the snow-routines that considered the daily maximum- and minimum air temperatures gave hope that this routine could substitute the original one, based on average air temperatures, in the operational forecasts of the start of the spring flow. It was, however, concluded that this method first should be tested for some more river basins and for longer time periods.

This was done by Arnér (1991), who compared this routine and the original snow-routine during 88 spring flows in six river basins (Fig. 2), using the HBV-

Table 2 – Explained variance (R^2-values) during the snow-melt period, and time gap between the days when the measured and modelled flow was exceeding three times the antecedent baseflow (days).

Routine No.	(S I)		(S II)		(S III)		(S IV)	
Routine	Original snow routine		Energy balance		Max-min, linear		Max-min, skewed	
	R^2	days	R^2	days	R^2	days	R^2	days
Stadarforsen								
1965	0.69	−3	0.26	−2	0.73	−12	0.82	−11
1967	0.94	−29	0.89	−19	0.92	−16	0.94	−3
1972	0.62	+2	0.82	+3	0.89	+1	0.83	0
1974	0.15	+14	0.08	+12	0.88	0	0.63	+10
1978	0.84	+3	0.86	+1	0.82	−1	0.91	−1
1980	0.30	+8	0.57	+2	0.84	+1	0.58	+3
1983	0.77	+1	0.90	0	0.94	−1	0.90	0
Svegsjön								
1969	0.83	+3	0.81	+1	0.70	0	0.77	+1
1970	0.98	0	0.78	−4	0.89	−1	0.78	−3
1973	0.52	−18	0.54	−18	0.31	−12	0.23	−18
1974	0.83	+11	0.63	+12	0.88	−1	0.86	+1
1976	0.90	+1	0.82	−1	0.65	+1	0.59	0
1977	0.41	−1	0.86	−2	0.85	−6	0.69	−6
1983	0.16	+4	0.79	+2	0.91	0	0.77	0
$\bar{x}$ Stadarforsen	0.62	9	0.63	6	0.86	5	0.80	4
$\bar{x}$ Svegsjön	0.66	5	0.75	6	0.74	3	0.67	4
$\bar{x}$ average	0.64	7	0.69	6	0.80	4	0.74	4

model. A weighted average, using the best criterion for each basin, showed that approximately 68% of the simulated starting days for the spring flow deviated with less than two days from the true starting day when using the original model. When using the alternative snow-routine, 76% of the simulations fell within this interval. It was thus confirmed that the minimum-maximum air-temperature snow-model slightly improved the predictions of the start of the spring flow, although it was less significant than in the pilot study made by Andersson (1990). One reason for this was that the largest improvements in both studies were found for the Stadarforsen basin, which was included in both studies. This was the only basin for which also Arnér found that the maximum-minimum air-temperature based routine significantly improved the simulations. For the other river basin, used in both of the studies (Svegsjön), the division of the basin in sub-catchments and elevation zones differed significantly. This might be the reason why Andersson showed a slight

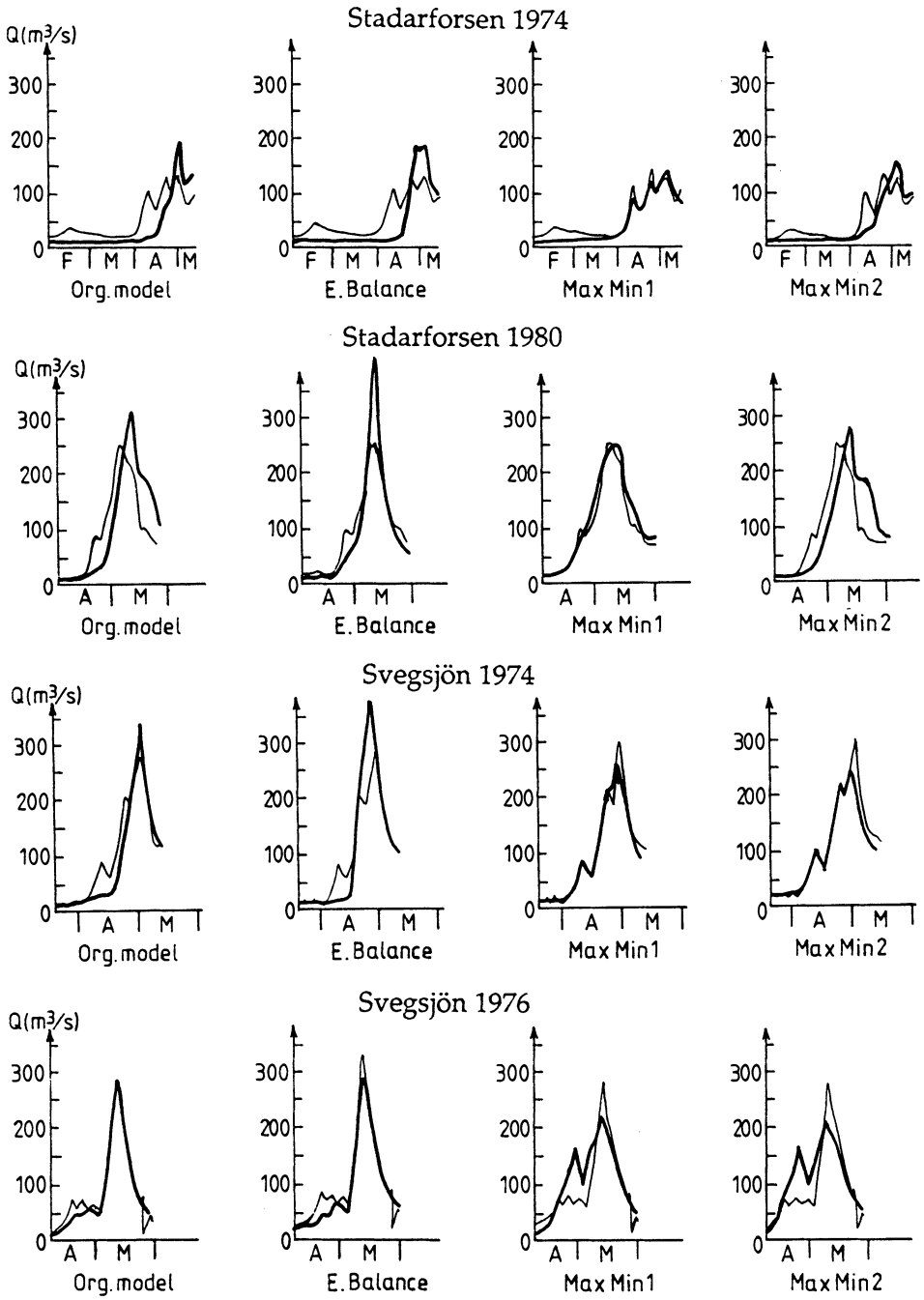

Fig. 6. Examples of model runs using the original snow routine (S I), energy balance (S II), or a routine based on linear (S III), or skewed (S IV) consideration of maximum and minimum of daily air temperatures.

improvement, whereas Arnér's study showed a slight deterioration when the original snow routine was substituted. Arnér (1991) found that R^2-values (estimated for the whole calibration period, *i.e.* not only the melting periods) became somewhat lower when using the minimum-maximum air-temperature based model. His general conclusion was, however, that the model performance could be considered equivalent.

This conclusion drawn from the present study thus could not justify a change of snow-routine in the operational forecasts.

Discussion

The general conclusion from the presented studies, and from most work that aims at improving existing models by attempting to increase the physical realism, is that the correspondence with measured variables decreases, or in the best cases is unchanged. An exception to this is the modelling of events that are due to extreme meteorological conditions. If a model has been developed and used during a period when such events did not exist, a sudden failure of the model can be shown. An example of this is the evaporation-routine of the HBV/PULSE-models, based on long-time averages of potential evaporation, which was shown to give too low evaporation during rainy and cold springs and summers. An empirical consideration of the air-temperature was, however, shown to be superior to using potential evaporation values calculated for individual months. This implies that this physical equation does not give a reliable estimate of the decrease of evaporation in times of temperature anomalies.

It is of vital importance to assess why most of these attempts fail and where the effort on model development should be directed. It is also necessary to keep in mind why a certain improvement is desired. If the spatial distribution of internal variables, or transit times, are of interest, increase in complexity, but no improvement, or a slight decrease in the fit between measured and modelled runoff can be seen as a success if it can be proven that the variables of interest, and their spatial distribution is modelled more successfully, *i.e.* the model has become more physically realistic.

The reason behind the relative success of the degree-day snow-melt routine is probably that it is based on daily average air-temperatures which is easier to interpolate in space than many other melt-related variables. When using a full energy-balance, many site-specific parameters, such as wind-speed and radiation, have to be interpolated for a large region. The number of meteorological stations with all the necessary meteorological parameters is much lower than the number of stations with only air-temperature, which makes it even more difficult to make a representative spatial interpolation. It is also probable that daily extreme values, which were used in two of the tested routines, are more site-specific than daily average air temperatures.

A recommendation could therefore be to look for driving variables that are space stable, or are generally available with a good spatial resolution.

Clearly, the choice of climatic stations and spatial interpolation routines for climatological variables are in most cases more important than the choice of which specific model subroutines to use (Andersson 1990, Arnér 1991). It is not relevant to use physically based equations for large regions if only point measurements from a limited number of sites are available. The most promising results will, probably, include incorporation of remote sensing data that makes it possible to directly obtain spatially distributed variables. This was, *e.g.* shown by Brandt (1991) who improved the forecasts of spring flows from areas with a sparse precipitation-measurement network by using air-borne georadar for updating of the modelled snow-pack.

Another important factor is the time-resolution of driving variables. In the Nordic countries, we are accustomed to work with daily time-steps, both for large basins and small catchments. It is fairly evident that a finer time resolution for both precipitation and temperature would be more suitable for small catchments, but, at least for snow accumulation and melt, it would probably be worth to continue the attempts with a better time resolution also for large basins.

One way to get around the lack of input variables is to try to build in some sort of indirect consideration with regard to extreme meteorological conditions (available with good time-resolution from maybe only one meteorological station in the basin) into the available simple models (driven by easily available parameters available with a lower time resolution but from a denser network). The difficulties in interpolating the information from a single station make the hope of success rather low, but it could be used to indicate the occurrence and direction of a risk of a misjudgement of the original model due to extreme meteorological conditions.

Another conclusion that could be drawn from the attempts of model improvement is that a certain change of model structure can improve the model performance for some basins whereas it is unchanged or deteriorated for other basins. Improvements can also occur only for certain time periods. It is therefore important to test the new model for a large set of basins and for long time series before drawing conclusions of a general model improvement. Detection of time or site specific improvements can be used to increase the knowledge of the mechanisms behind spatial and temporal variability of hydrological processes. This can be done by defining the physiographic differences between basins that reacted differently on a change in a model structure or by investigation of which meteorological conditions that prevailed during periods when the introduction of a new subroutine improved the model fit.

In summary, runoff-model development should be directed towards physically sound assumption of links between physiographic factors, climatic variables and hydrological processes. The spatial distribution of the used parameters must, however, be generally available from maps or remote sensing images. To get a fruitful

development of the field of runoff modelling, we should try to join the best from the physically based and conceptual schools of modelling in a landscape mosaic context. The two schools are, actually not that different, since both depend on model calibration, and both types of models probably (at least after calibration) have routines where a simple formulation encompasses a number of different processes. Those using and developing physically based models should ask themselves why simpler models often work better and try to scale up their knowledge into the landscape scale. One way to do this is by determining probability indexes from overlays of relevant landscape factors. The conceptual school of modelling could probably contribute a lot to this work since it is based on using easily available spatial and temporal data, and indirect relationships that are relevant in the spatial scale used by the model.

The effort in validating physically based model can often be criticized as being too small (especially concerning internal variables). For more conceptual models, it is, however, recommended to look less on R^2-values, and start to look more on the reasons why the hydrological response of different catchments varies.

It is most important, however, not to limit hydrological modelling to program codes and data-bases. A few visits to the studied catchments give more hydrological insight than years in front of a computer!

Acknowledgements

The attempts to improve the HBV/PULSE-model were carried out at the Swedish Meteorological and Hydrological Institute (SMHI), within two projects, financed by the Swedish Association of River Regulation Enterprises (VASO).

References

Andersson, L. (1988) Hydrological analysis of basin behaviour from soil moisture data, *Nordic Hydrol., Vol. 19*, pp. 1-18.

Andersson, L. (1989) Hänger det på avdunstningen? Försök att förbättra HBV/PULS-modellen (In Swedish – Does it depend on the evapotranspiration? Attempts to improve the HBV/PULSE model), *Vannet i Norden, Vol. 3*, pp. 14-27.

Andersson, L. (1990) Vårflodens start – försök att förbättra HBV/PULS-modellernas snörutin (In Swedish with English summary – The start of the spring-flow. Attempts to improve the snow routine of the HBV/PULSE-model), *Vannet i Norden, Vol. 4*, pp. 10-29.

Andersson, L., and Harding, R. J. (1991) Soil moisture deficit simulations with models of varying complexity for forest and grassland sites in Sweden and the UK, *Water Resour. Manage, Vol. 5*, pp. 25-46.

Arnér, E. (1991) Simulering av vårflöden med HBV-modellen (In Swedish with English summary – Simulations of springflows with the HBV-model), SMHI Rep. Hydrologi, No. 32, Norrköping, Sweden, 71 pp.

Bergström, S. (1975) The development of a snow routine for the HBV-2 model, *Nordic Hydrol. Vol. 6*, pp. 73-92.

Bergström, S., Carlsson, B., Sandberg, G., and Maxe, L. (1985) Integrated modelling of runoff, alkalinity and pH on a daily basis, *Nordic Hydrol., Vol. 16*, pp. 89-104.

Beven, K. J. (1989) Modelling interflow: The future. *In:* H. J. Morel-Seytox, *Unsaturated flow in hydrological modelling*, Reidl.

Beven, K. J. (1989) Changing ideas in hydrology – The case of physically-based models, *J. Hydrol., Vol. 105*, pp. 157-172.

Brandt, M. (1991) Snömätningar med georadar och snötaxering i övre Luleälven (In Swedish – Use of georadar for snow taxations and measurements in the upper Lule River) Report SMHI Hydrology. No. 33. 18 pp.

Calder, I. R., Harding, R. J., and Rosier, P. T. W. (1983) An objective assessment of soil-moisture deficit models, *J. Hydrol., Vol. 66*, pp. 329-355.

Eriksson, B. (1981) The potential evapotranspiration in Sweden (In Swedish with English summary), Report SMHI, RMK No. 28, Norrköping, Sweden, 40 pp.

Killingtveit, Å. (1988) Bruk av elektroniske regneark for hydrologiske analyser (In Norwegian – Use of spread-sheets for hydrological analyses), *In:* Proceedings Nordic Hydrological Conference, 1988 Rovaniemi, NHP Rep. No. 22, pp. 286-395.

Klêmes, V. (1986) Dilettantism in hydrology: Transition or density? *Wat. Resour. Res., Vol. 9*, pp. 177-188.

Lindström, G., and Bergström, S. (1992) Improving of the HBV and PULSE-models by use of temperature anomalies, *Vannet i Norden, Vol. 1*, pp. 16-25.

Lumadjeng, H. S. (1989) Modelling of hydrological processes. *In:* Gustard, A., Roald, L. A., Demuth, S., Lumadjeng, H. S., and Ross, R., *Flow Regimes from Experimental and Network Data* (FREND).

Morton, F. I. (1982) Integrated basin response – a problem of synthesis or a problem of analysis? *In:* Proc. Can. Hydrol. Symp. 1982, Nat. Res. Counc. Can., Ottawa, Ont, pp. 361-384.

Penman, H. L. (1948) Natural evapotranspiration from open water, bare soil and grass, Proc. R. Soc. London, Ser. A. 1993, pp. 120-146.

Thornthwaite, C. W., and Holzman, B. (1939) The determination of evaporation from land and water surfaces, *Month. Weath. Rev., Vol. 67*, pp. 4-11.

Turner, N. C., and Jarvis, P. G. (1975) Photo synthesis is Sitka spruce (*Pieca sitchensis* (Bong.) Carr.). IV: Response to soil temperature, *J. Appl. Ecol., Vol. 12*, pp. 561-576.

Vehviläinen, B. (1986) Modelling and forecasting snowmelt floods for operational forecasting in Finland. In: Modelling Snowmelt-induced processes, IAHS Publ. No. 155, Budapest, pp. 245-256.

WMO (1986) Intercomparison of models of snowmelt runoff, Operational Hydrology, Report No. 23, Geneva.

Received: 15 June, 1992
Accepted: 17 August, 1992

Address:
Department of Water and Environmental Studies,
Linköping University,
S-581 83 Linköping,
Sweden.