NVE

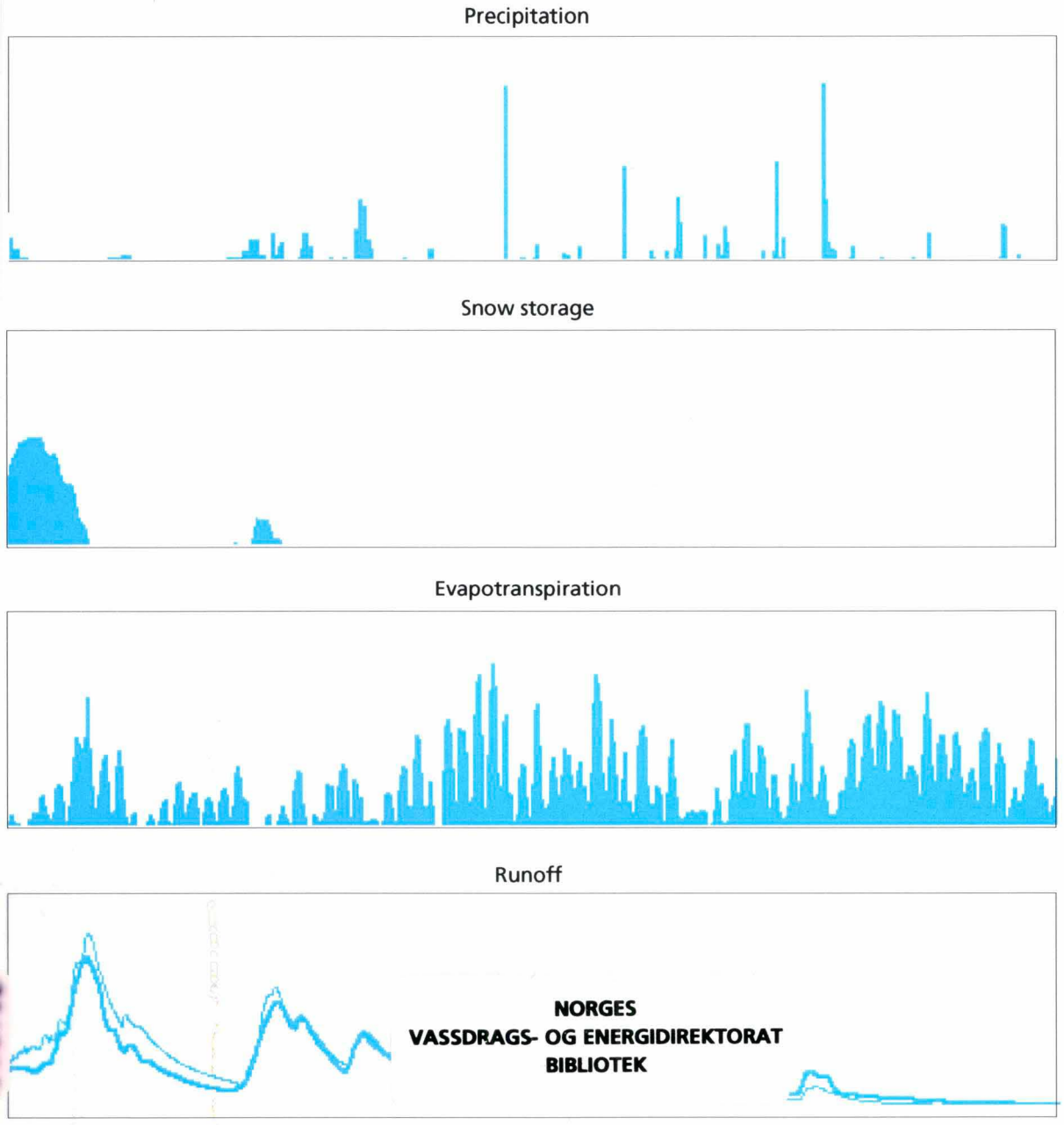
Real time updating of hydrological forecasting models

Methods and information sources

Stein Beldring

17
2000

DOCUMENT



Real time updating of hydrological forecasting models

Methods and information sources

Norges vassdrags- og energidirektorat

2000

Dokument nr 17

Real time updating of hydrological forecasting models

Utgitt av: Norges vassdrags- og energidirektorat

Forfatter: Stein Beldring

Trykk: NVEs hustrykkeri
Opplag: 50
Forside: Stein Beldring
ISSN: 1501-2840

Emneord: Real time updating; Hydrological forecasting; Streamflow; Kalman filter; Data assimilation

Norges vassdrags- og energidirektorat
Middelthuns gate 29
Postboks 5091 Majorstua
0301 OSLO

Telefon: 22 95 95 95
Telefaks: 22 95 90 00
Internett: www.nve.no

November 2000

Contents

Preface

One of the main areas for application of precipitation-runoff models is for short term forecasting of streamflow. These forecasts form the basis for flood warning, for operation of water resources systems such as reservoirs, and for planning hydropower sales. Improved meteorological forecasts, technological developments in the estimation of rainfall by radar, systems for real time acquistion and transmission of data, and more powerful digital computers has stimulated the development and use of hydrological models for forecasting purposes during the last decades. A large number of models of varying complexity exists. However, as all models seek to simplify the complexity of the hydrological processes by focusing on the fundamental aspects of the system under consideration, the models are not fully realistic representations of real world phenomena. Although these simplifications are necessary in order to use hydrological models for operational purposes, the model results are not always as reliable as desired. This problem is further accentuated by uncertainty in the forecasts of the meteorological forcing data used to drive the hydrological models. Due to the large economical and social consequences of decisions concerning flood warning and operation of water resources systems, there is a need to improve the hydrological forecasts. Although better meteorological forecasts and more realistic representations of hydrological processes may result in improvements, there is also a need for techniques for adjusting or updating the hydrological forecasting models based on observations of state variables or the forecast variable. The topics of this report are concerned with the latter issue.

This report is based on a trial lecture for the degree dr.scient. at the Department of Geophysics, University of Oslo.

Oslo, November 2000

Kjell Repp
Director of Hydrology Department

Svein Taksdal
Section manager

Abstract

This report presents a review of methods for adjusting or updating hydrological forecasting models. Although most of the examples which are presented are concerned with forecasting streamflow, the methods may also be used for forecasting other hydrological processes. Different techniques for updating hydrological forecasting models based on measurements of state variables or the forecast variable are presented. Data assimilation methods have the potential for updating spatially distributed environmental models based on optimal use of information from different sources. Some examples of data assimilation applications in hydrology are presented, and the possibility of using these techniques for updating hydrological models are discussed. Information sources which may be used during updating of hydrological forecasting models are considered.

Sammendrag

Denne rapporten presenterer en oversikt over metoder som kan benyttes for å korrigere eller oppdatere resultatene av prognoser beregnet med hydrologiske modeller. Selv om eksemplene som presenteres i hovedsak dreier seg om vannføring, kan metodene også benyttes ved fremskriving av andre hydrologisk prosesser. Teknikker for oppdatering basert på observasjoner av tilstandsvariabler eller den prognoserte variabelen presenteres. Data-assimilering kan benyttes for å oppdatere romlig fordelte modeller av fysiske prosesser basert på optimal bruk av tilgjengelig informasjon. Noen eksempler på bruk av data-assimilering i hydrologi presenteres, og muligheten for å bruke slike teknikker for å oppdatere hydrologiske prognose-modeller diskuteres. Informasjonskilder som kan benyttes ved oppdatering av hydrologiske prognosemodeller presenteres.

1. Introduction

Floods and droughts kill more people and cause more damage than any other natural disasters. Due to the social and economical consequences of these extreme events, authorities and the public must be warned about the possibility of their occurence as soon as possible. Although flood control structures such as reservoirs and dikes may be constructed, this is usually not sufficient to prevent the harm caused by floods. Furthermore, human effects upon the hydrological cycle, e.g. introduction of radiatively active gases in the atmosphere and physical alterations of the land surface have an impact on the climate and the hydrological cycle which may result in an increase in the frequency of extreme events in the future (Eagleson, 1994). Thus, there is a need for systems capable of forecasting the behaviour of hydrological processes.

Improved meteorological forecasts, technological developments in the estimation of rainfall by radar, systems for real time acquistion and transmission of data, digital computing and mathematical modeling has stimulated the use of hydrological models for real time flood forecasting during the last decades (O'Connel, 1991). A large number of models of varying complexity exists. However, due to errors in the model structures and the meteorological input data, there is a need for adjusting or updating the forecasting models (O'Connel and Clarke, 1981). This is demonstrated by Figure 1 which shows results from simulating and forecasting streamflow in the 304 km^2 Eggedal catchment in south-east Norway during a flood in October 2000. The simulated flood peak is less than 50 % of the observed peak and it occurs 24 hours too late. This report considers various methods which may be used for updating purposes.

Real time processing of data is necessary when it is required to forecast changes in streamflow or in the state of a catchment in response to external forcing within some small upper limit of response time. This is necessary when the management of a river basin demands that a continuous watch must be kept on the natural events occuring in the catchment (and possibly on man's activities affecting water movement). The elements of the hydrological cycle and the state of water stores (natural or artificial) must be monitored so that any changes can be assessed, and the necessary action taken to preserve acceptable conditions in the catchment and prevent loss of life and damage to property. The topics of this report are related to real time forecasting of streamflow using mathematical models of hydrological processes, however, hydrological forecasting in more general terms is considered first.

2. Hydrological forecasting

Hydrological forecasting means estimating the conditions in a catchment at a specific future time, or during an interval of time. It is distinguished from prediction, which is the estimation of future conditions, without reference to a specific time. Examples of forecasts are the discharge at a particular cross-section of a river at noon tomorrow or

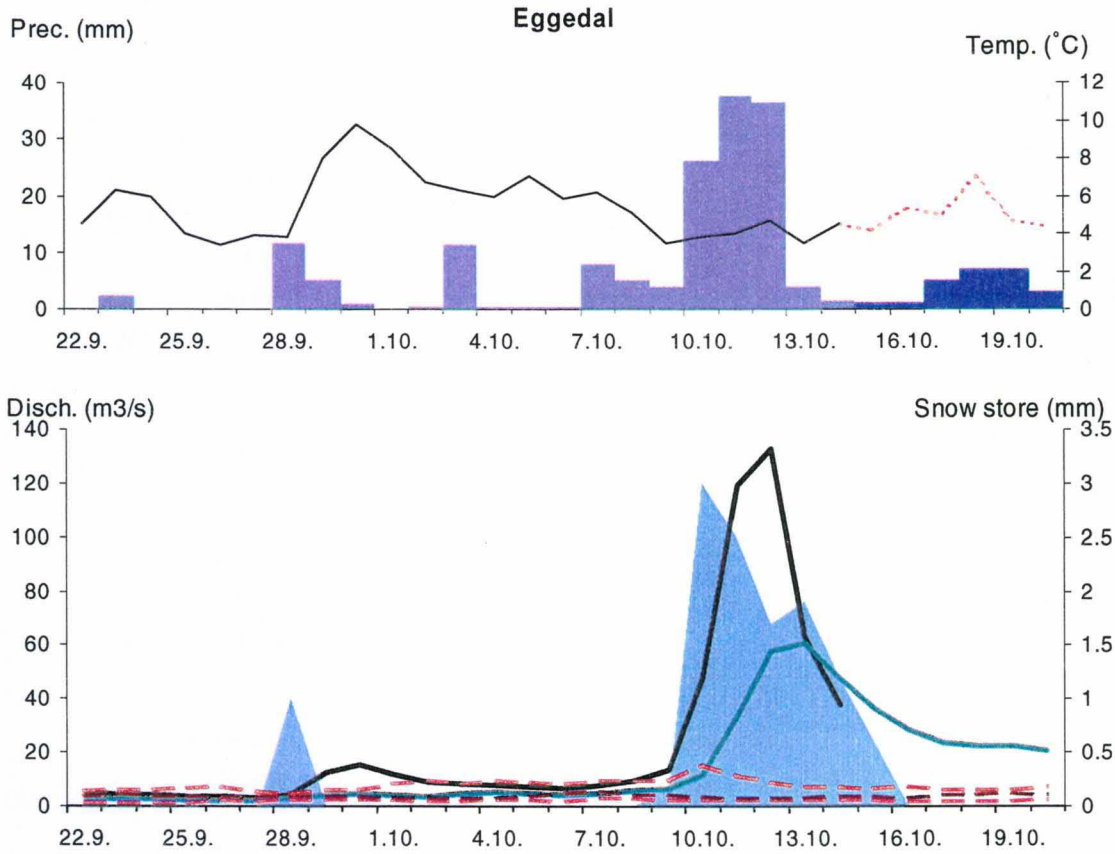


Figure 1. Observed (black) and simulated (green) streamflows from the 304 km^2 Eggedal catchment in south-east Norway during September and October 2000. The light blue shaded area shows the simulated snow store (snow water equivalent). The dottet lines represent the sample quartiles. The forecast was issued on 14 October.

the inflow to a reservoir during the next two months. Conversely, the magnitude of the annual maximum flood with exceeedance probability 0.01 is predicted. The accuracy of a forecast generally decreases with increasing lead time, i.e. the forecast period or forewarning period. For very long lead times, the distinction between a forecast and a prediction vanishes. Real time forecasts are necessary for warning of extreme events, e.g. floods, for operation of water resources systems such as reservoirs or for planning hydropower sales. In the event of a toxic spillage into a stream, real time forecasting of the arrival of the pollution at a water intake can ensure that abstraction is halted before the arrival time.

A hydrological forecast is characterized by: (i) the forecast variable, i.e. the hydrological element being forecasted; (ii) the lead time; (iii) the computational method; (iv) the purpose of the forecast; (v) the form of presentation, e.g. single

value, total hydrograph or probability distribution; and (v) the means of information to authorities and the public (WMO, 1994).

The dynamics of hydrological processes are driven by astronomical and meteorological factors, but the changes these factors bring about in the hydrological conditions of a catchment do not occur instantaneously. For example, the duration of runoff caused by precipitation is often many times longer than the rainfall itself, and a lag time exists between temperature rise and melting of snow on one hand, and the runoff in a river on the other. The relatively slow rate at which many hydrological processes develop and the fact that they lag behind meteorological processes make it possible to forecast some elements of the hydrological cycle. Techniques for forecasting range from the use of simple empirical formulae and correlations to the use of complex mathematical models representing all phases of the water balance of a catchment.

Although this report is concerned with forecasting streamflow from a catchment in response to water input from rain or snowmelt, forecasting of many different hydrological processes and events are possible. In principle, any process involved in the hydrological cycle may be subject to forecasting if we know:

1. The initial factors existing at the time the forecast is made. These may be estimated based on hydrological and meteorological measurements or calculations.

2. Future factors which influence the hydrological processes under consideration during the forecast lead time. The most important future natural factors in streamflow forecasting are the weather conditions, which can be taken explicitly into account only if a weather forecast is available.

The following are some examples of hydrological elements in rivers, lakes, reservoirs and the subsurface for which forecasting is of practical interest (WMO, 1994):

- Volume of runoff in various periods of time (month, season, year).
- Discharge or stage hydrograph.
- Peak flood stage or discharge and its time of occurence.
- Extent of flooded area and its variation in space and time.
- Maximum water level in lakes and the date this will be reached.
- Water level in navigable rivers during various seasons.
- Height of waves created by wind on lakes and reservoirs.
- Water quality elements such as temperature and transport of dissolved substances.
- Date of formation of floating ice in the autumn.
- Date of freezing of the surface of lakes and reservoirs.

- Thickness of ice cover.
- Date on which ice begins to break up in spring.
- Date on which ice disappears completely from lakes and reservoirs.
- Minimum aquifer level and time of occurrence.
- Maximum aquifer level and time of occurence.
- Water in the root zone available to agriculture .
- Dispersion and flow of toxic substances contained in the groundwater.

There are differences in forecasting methods, depending on the forecast lead time. Short term forecasts have lead times less or equal to two days, medium term forecasts have lead times between two days and ten days, long term forecasts have lead times larger than ten days, while seasonal forecasts have lead times of several months. For longer periods than a year, prediction techniques are more applicable. Short term forecasts are most often used for the purpose of flood warning and for real time operation of water resources systems, e.g. hydropower production systems. Real time data processing is required when the hydrological forecasting model must be able to receive and immediately process constantly changing data to give results very quickly. This is necessary in a situation where forecasts of the discharge of a river in the near future must be provided.

When the forecasting lead time is longer than the time of concentration, i.e. the time required for water input at the farthest point on the catchment boundary to reach the forecast point, some of the water that is to be included in the forecast will be input as rain or snowmelt in the near future. A procedure which includes both precipitation forecasts and precipitation-runoff modeling is therefore needed. When the forecasting lead time is shorter than the time of concentration, and the time of concentration is domianted by the hydrological response of the landscape surrounding the channel system, it is not necessary to include precipitation forecasts in the hydrological forecasting procedure. However, streamflow forecasts should be based on observed rainfall from a network of rain gauges whose data are transmitted to the forecast center. Such forecasts must incorporate a precipitation-runoff model. At the other extreme, if the forecast lead time is shorter than the time of concentration, and the total time of concentration is dominated by the transport of the flood wave through the channel system, streamflow forecasts can be based on routing of observed flows down a river channel. This is the situation for large river systems such as the Mississippi River and its major tributaries (Lettenmaier and Wood, 1993).

In Scandinavia the need for flood warning and reliable forecasts of the inflow to the reservoirs of the hydropower systems has stimulated the operational use of hydrological models. The forecasts are either short term, with a few days lead time, or long term, covering the whole snowmelt season (several months). For short term forecasting meteorological forecasts are combined with hydrological models. Long term forecasts may be based on present hydrological conditions, and a number of

simulations with climate data from the same season earlier years. The output of the model is subject to statistical analysis to determine the probability of future events (Bergström, 1995).

Long term forecasts are most often used for water resources management, e.g. allocation of irrigation water or implementing mitigating measures such as water conservation during droughts. Long term forecasts suffer from lack of reliable meteorological forecasts. Other methods must be applied instead:

- Index variables that can be related to the forecast variable using multiple regression. These index variables may for example be rainfall or temperature during different seasons.
- Time-series models.
- Probability forecasting, i.e. running a model with different meteorological scenarios as input in order to construct a probability distribution of the outcomes.

3. Hydrological forecasting models

Real time forecasting of streamflow is usually performed with a precipitation-runoff model, i.e. an algorithm or computer code consisting of a set of mathematical and logical expressions that define quantitative relationships between flow generating factors (meteorological input) and flow characteristics (output). The precipitation-runoff model may include a description of snow accumulation and snow melt and a river routing procedure. Furthermore an updating procedure is required to correct the forecasts. Compehensive data processing systems are also important for operation of real time forecasting systems.

The mathematical and logical expressions used to describe the hydrological system contain variables and parameters. Model parameters remain constant over time or vary in a manner which may be described using physical principles or empirical relationships. Parameters either represent physically measurable properties of a watershed, or are used to describe hydrological processes. A variable may represent: (i) the state of the different storages in the hydrological system as approximated by the forecasting model; (ii) the input signal which drives the model; or (iii) the output from the model. Variables vary with time.

There are a range of models for forecasting of stream discharge based on precipitation inputs. At one extreme are the purely empirical, lumped black-box models. These models identify a relationship between rainfall input and streamflow output without attempting to describe any of the mechanisms whereby this process takes place. Relatively simple models for the representation of the rainfall-runoff process have successfully been used for forecasting purposes. For example, if the response of the

catchment is assumed linear with respect to effective rainfall, methods based on convolution of the impulse response function of the catchment with the effective rainfall can be used to forecast stream discharge. Effective rainfall is the total rainfall less any losses due to infiltration, interception, and surface ponding. The instantaneous unit hydrograph model is an example of this approach. The major problem with these empirical relationships is that they are subject to serious error when it becomes necessary to rely upon them under conditions not previously experienced. Models that somehow consider the interactions between hydrological processes taking place within the catchment are believed to be more reliable under these conditions.

At the other extreme are the distributed physically based models which attempt to describe the spatial distribution of storage and flow of water using equations based on principles from fluid mechanics and thermodynamics. The resulting system of partial differential equations has to be solved numerically at all points on a three-dimensional grid representation of the catchment. Examples of these models are the Institute of Hydrology Distributed Model (IHDM) (Calver and Wood, 1995), the Système Hydrologique Européen (SHE) model (Abbot *et al.*, 1985a, 1985b). Theoretically, the main advantage of distributed physically based models is that they represent more accurately the heterogeneity in space and time of the various hydrological processes. Of course, this comes at the expense of a large number of parameters, most of which are related to better representation of the physics involved. A critique expressed against these models concern the many parameters values which can be modified during the calibration process. Another problem with these model is the simplified representation of sub-grid processes. It is not uncommon for distributed models to assume that model parameters and state variables are uniform within the grid cells. This is unrealistic, considering the non-linearities of the processes involved and the heterogeneity of natural systems. These considerations caused Beven (1989) to considered models which are usually claimed to be distributed physically based as in fact being lumped conceptual models, just with many more parameters. Mroczkowski *et al.* (1997) defined a model as conceptual if one or more of its parameters had to be determined by calibration using observed catchment response.

Between these extremes we find the majority of the hydrological simulation models in use, the conceptual models, which are often semi-distributed. Examples of conceptual models are the HBV-model (Bergström, 1995) or the U.S. NWSRFS-model (Burnash, 1995). Rather than using the relevant equations of mass, energy and momentum, these models use simplified, but plausible conceptual representations to describe the components of the rainfall-runoff process. They represent the various water stores (interception, soil moisture, groundwater) and flux terms (infiltration, evapotranspiration, snowmelt, interflow, groundwater baseflow) with varying levels of complexity. These representations involve several interlinked reservoirs and simple budgeting procedures which ensure that at all times a complete mass balance is maintained between inputs, outputs and storage changes. The main purpose of these models is to describe the conversion of precipitation falling on a watershed to streamflow at the catchment outlet, with the eventual operational forecast in mind.

Although they use words like groundwater store or soil moisture store to describe their reservoirs, one should be careful with considering them as more than abstract representations. For flood forecasting purposes mainly conceptual and black-box models are in use today (Refsgaard, 1997).

River routing models may be classified as either hydraulic (distributed) or hydrological (lumped). In hydraulic routing models the flows and water levels are computed as a function of time simultaneously at several cross sections along the watercourse using the hydrodynamic equations of unsteady flow (the Saint Venant equations) or their dynamic wave or kinematic wave approximations. Hydrological routing is based on continuity considerations for storage of water in reservoirs or river reaches and require less data than hydraulic routing. (Lettenmaier and Wood, 1993).

4. Updating of hydrological forecasting models

As no model can perfectly represent the real system, simulation errors occur in any model application. The performance of a hydrological forecasting model is affected by: (i) errors due to imperfect model structure, limited number of calibration data and changing system conditions; (ii) the meteorological observations used to run the model up to the initial state at the start of the forecasting period; (iii) the accuracy and reliability of meteorological forecasts; and (iv) errors in streamflow measurement, e.g. rating curve errors. In real time forecasting applications, it is therefore necessary to minimize these errors using so-called feedback or updating procedures. This should be done in a manner which avoids the propagation of the errors to future forecasts. This can be done by periodically updating the model, i.e. to provide new information from observations of system output (streamflow) or state variables. The parameters or state variables of the model must be adjusted so that the required agreement between the forecasted and the observed streamflow is obtained. Prior to updating, the hydrological forecasting model must be run with observed meteorological variables as input. The state variables of the model at the time of the issue of the forecast have then been determined by the basic model simulations. During updating, these model states may be adjusted. Real time updating means updating a model with new information within some small upper limit of response time. This is necessary in order to perform real time forecasting.

What is generally called updating can be divided into five categories:

1. Updating the computed streamflow by measured values when they become known, i.e. correction based on the difference between measured and forecasted discharges. This procedure does not change model states or parameters.

2. Adaptive forecasting, i.e. adjustment of model states or parameters in the process of forecasting based on observed streamflow.

3. Adjustment of observed temperature, precipitation and other input variables prior to the time the forecast is made. The purpose is to update model states.

4. Updating of state variables based on observations; e.g. snow covered area, snow water equivalent, soil moisture storage and groundwater storage.

5. The parameters of a conceptual model may be adjusted. Manual adjustment of parameters is usually not realistic in a real time system. Therefore this procedure requires an automatic parameter optimization algorithm and an objective function derived from all data up to the present. Due to dependence between parameters, the modification of one parameter generally requires modification of the other parameters as well.

The purpose of the the four last procedures is to correct the model's state variables or parameters in order to have the best possible initial state for the next forecast. It is difficult to justify that the parameters of conceptual models which have been determined by a calibration procedure using historical data should be updated based on the relatively few data during the updating period. Periodic recalibration of a conceptual hydrological model used for forecasting is usually not performed. Updating by recalibration of model parameters is mostly confined to black-box models where no clear distinction exists between states variables and model parameters (Refsgaard, 1997).

The basic model simulations are essential for accurate forecasts, and the better the basic simulations are, the better the updating routines generally function. A continuous need for large corrections may indicate that there are problems with the model structure, the calibration of the model or the meteorological input data. This puts emphasis on the importance of thorougly calibrating and validating the hydrological simulation models before applying them in operational real time forecasting.

4.1 Forecast error correction

The streamflow forecast is adjusted using an estimation of the forecast error, which is the difference between the true streamflow and the forecasted streamflow. This forecast error is approximated by the difference between the observed streamflow and the forecasted streamflow. The forecast errors are usually found to be correlated, giving rise to the possibility of forecasting future values of these errors. The persistence in the forecast errors can be used to develop a forecasting procedure that consists of an initial forecast based on the hydrological model plus a correction based on the errors between the forecast and the measurement (Lettenmaier and Wood, 1993):

$$q_f(t) = m(p, \theta, s) + e_f(t)$$

$q_f(t)$ - forecasted streamflow at time t

$e_f(t)$ - estimated forecast error for time t based on the previous forecast errors

$m\ (p,\theta,s)$ - the forecast provided by the hydrological model

p - meteorological input

θ - model parameters

s - state variables

A time-series model such as an autoregressive model or autoregressive/moving-average model can be constructed for determination of the forecast error. Using a number of calibration storms, a time series of forecast errors is constructed. The statistical model is developed by using identification and parameter estimation techniques from time-series-analysis. This model is then used to estimate the forecast error. A typical form of the forecast error model is a lag-one autoregressive process, which results in an estimated forecast error $e_f(t)$ at time t:

$$e_f(t) = \rho e_f(t-1) + \varepsilon_t$$

$e_f(t)$ - forecast error at time t, stationary

ρ - lag-one autocorrelation coefficient of the forecast errors without updating

ε_t - zero mean purely random process (white noise)

A white noise process consists of a sequence of uncorrelated random variables.

When an operational forecast is made, the streamflow for the previous period has been measured and the forecast error calculated. Using the measured inputs (meteorological variables required by the model), an initial forecast based on the hydrological model is made. Using the calculated forecast errors from the previous time periods, a forecast error is estimated using the statistical model. The final forecast is the sum of the forecast based on the hydrological model and the forecast error estimate. This is a simple, but effective way of updating forecasts. If the parameters of the time-series model are estimated after each time step when a new residual becomes available, this procedure is an adaptive forecasting technique, which is the subject of the next section.

This method for updating of short term forecasts is most effective when the forecast errors are highly correlated, e.g. during recession in large catchments. Small catchments with rapid response to rain or snowmelt may not be suitable for this method. This technique can be evaluated for a particular application by computing the forecast error autocorrelation function. If the forecast errors are uncorrelated, this is an indication that the forecasting procedure is extracting as much information as possible from the measurements, and no further improvements can be achieved through updating. Different forecasting models can be evaluated by comparing the variances and autocorrelation functions of the forecast error time series. It is desireable to have low forecast error variances and small autocorrelations.

This was one of the methods applied within the "Dee weather radar and real time hydrological forecasting project", which had the purpose of exploring the potential for using technological developments for more efficient operation of a water resources system (Green, 1979a, 1979b). Subcatchment rainfall-runoff models were calibrated and linked to a main channel flow routing model; the former were simple in structure and based on an input-storage-output relationship. A time-series model was used to forecast the discrepancy (noise) between observed and forecasted discharge. The "simulation mode" forecasts from the linked model was updated through the noise model. This precipitation-runoff model also had the possibility of updating its forecasts using the Kalman filter, a technique which to be considered later.

Lindström and Carlsson (2000) used forecast error correction with a lag-one autoregressive model for updating streamflows forecasted by the HBV-model. The method generally improved the forecasts.

The Hydrology Department, Norwegian Water Resource and Energy Directorate have applied this procedure for estimating the forecast errors of the HBV-model (Langsrud *et al.*, 1998; Skaugen, 1999). A first-order autoregressive process is used to model the forecast error using the logarithms of observed and forecasted streamflows. Figure 2 shows that the residuals of the autoregressive process in the 1625 km^2 Knappom catchment in south-east Norway depend on temperature. This requires that the variance of the purely random process in the autoregressive model is temperature dependent. A threshold of zero degrees is used. Figure 2 also shows that the mean value of the residuals is zero. Recent results from applying this method during the flood in south-east Norway in October 2000 have been promising.

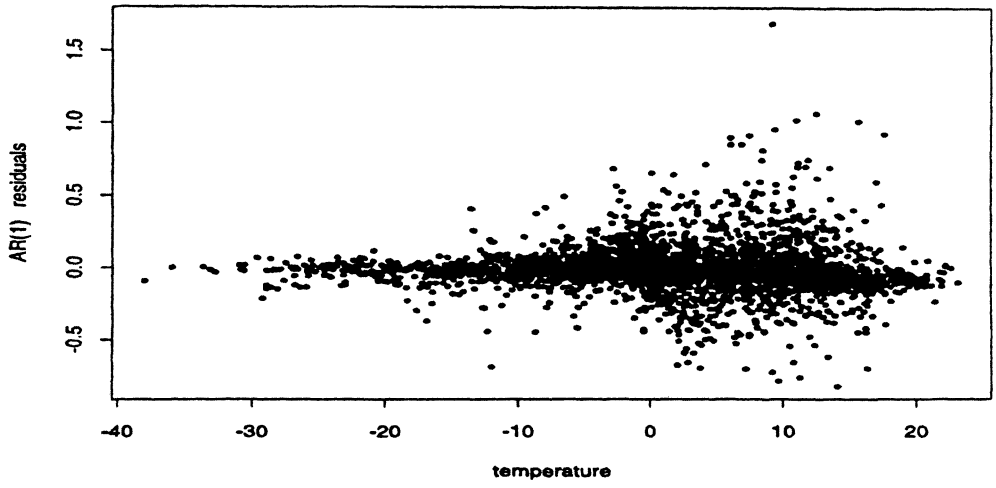


Figure 2. Residuals of first order autoregressive model of forecast errors vs. temperature for the 1625 km^2 Knappom catchment in south-east Norway (from Langsrud *et al.*, 1998).

4.2 Adaptive forecasting using measured streamflow values

Forecasting models can be used in one of two modes: they can take in measured inputs like rainfall and temperature and make streamflow forecasts (simulation mode), or they can use both measured inputs and measured streamflows when making the streamflow forecasts (adaptive mode). If the forecast lead time is longer than the concentration time of the catchment, a meteorological forecast is also required. Conceptual models may have a complex structure which prevents the use of adaptive forecasting techniques. In simulation mode, recalculation of the model's states or parameters is usually not performed (O'Connel and Clarke, 1981).

Adaptive hydrological forecasting requires that:

1. The discharge and meteorological driving variables of a catchment are available at short time intervals.

2. A model which describes the transformation of meteorological input previous to time t to runoff at time t has been identified:

$$q_t = m(p_t, \theta, s) + \varepsilon_t$$

q_t - runoff at time t

p_t - meteorological input up to time t

$m(p_t, \theta, s)$ - the hydrological model

θ - vector of model parameters

s - vector of model states

ε_t - error term due to inadequacy of model and input data

3. Estimates of model states or parameters at time t have been calculated from observations of precipitation and runoff up to the time t.

4. The model is used to make forecasts q_{t+1}, q_{t+2}, ... of runoff during future time intervals $t + \Delta t$, $t + 2\Delta t$, ...

5. The estimates of model states or parameters are corrected at the end of the time interval $t + \Delta t$ when p_{t+1} and q_{t+1} become available.

6. The forecasts of future runoff q_{t+2}, q_{t+3}, ... are corrected using the new estimates of states or parameters and possibly a new meteorological forecast which must also be available at the end of the time interval $t + \Delta t$.

4.3 Adaptive forecasting using black-box transfer function models

Black-box transfer function models use empirical relations between rainfall and runoff. An example is the following linear relationship between runoff at time t, runoff in earlier time periods and rainfall depths.

$$q_t = \alpha_1 q_{t-1} + \alpha_2 q_{t-2} + \ldots\ldots + \alpha_r q_{t-r} + \beta_0 p_{t-b} + \beta_1 p_{t-b-1} + \ldots\ldots + \beta_{s-1} p_{t-b-s+1} + \varepsilon_t$$

$\{q\}$ - runoff in earlier time periods

$\{p\}$ - precipitation in current and earlier time periods

$\{\alpha\}$ - a set of r parameters applied to the previous r runoff values

$\{\beta\}$ - a set of s parameters applied to the previous rainfalls delayed by b time intervals

ε_t - error term representing the effect of model inadequacy and possibly noise in the measurements of q_t and p_t

This model has self-correcting ability inherent in its structure, because the observed values of the forecast variable are used as input variables to the model. It also represents explicitly the lagged response of streamflow to rainfall, and can be used for real time forecasting. However, the linearity inherent in this description of the rainfall-runoff process is a shortcoming. An approach to overcome this problem is to incorporate an auxiliary variable, which reflects antecedent moisture conditions in the catchment. This requires an additional parameter to be multiplied with this variable.

If all $\{\alpha\}$ reduce to zero, q_t is storm runoff and $\{p\}$ is effective rainfall, this is the unit-hydrograph model. The existence of baseflow in the absence of rainfall, implies that the memory in this model would have to be very large in order to avoid streamflow from becoming zero during recession periods. If all $\{\beta\}$ reduce to zero, this is an autoregressive model. Although a high degree of interdependence usually exists between streamflow over short time intervals, there is no means whereby the natural lag between rainfall and runoff can be utilized, nor can this model anticipate the occurence of a peak discharge. Nevertheless, this model does have a self-correcting ability if real time streamflow measurements are available.

The statistical properties of ε_t and the measurement noise in p_t generally prevent applying least squares techniques to the estimation of the parameters in this model. Provided the error terms are uncorrelated, least square estimates will be consistent and asymptotically efficient. If the error terms are autocorrelated, least square estimates will be inconsistent (O'Connel and Clarke, 1981). The Instrumental Variable-Approximate Maximum Likelihood algorithm is another approach for estimating the parameters of this model (Young, 1986). This algorithm assumes that observed discharge is a sum of a deterministic component and a stochastic component. The estimation of the transfer function parameters and the noise model parameters are decoupled, giving two separate estimation problems which are solved

using least square estimation techniques. This procedure can be used for updating the transfer function model in adaptive forecasting mode.

The transfer function model can also be expressed in the state-space form required for using the Kalman filter technique for adaptive forecasting.

4.4 Adaptive forecasting using the Kalman filter

If the forecast model can be cast in a sufficiently simple form, it provides a basis for adaptive forecasting using the Kalman filter. This is a formal strategy for adjusting model output based on the difference between the observed and modeled output. Updating with the Kalman filter increases the complexity of forecasting, but can significantly improve forecast accuracy. This method is referred to as filtering because the errors in the measurements and the model are filtered out through the algorithm used. The Kalman filter can be applied to linear models with errors in both the model and the measurements, e.g. time-series models. However, the hydrological cycle is highly non-linear, implying that non-linear models, e.g. conceptual models are better suited to forecast streamflow (Wood and O'Connel, 1985).

The behaviour of a hydrological system can be described by the evolution of a set of state variables. In practice, the individual state variables cannot be determined exactly by direct measurements. Instead, we find that the measurements we make of system output are functions of the state variables and these measurements contain disturbances. The purpose of the Kalman filter is to estimate the state of the system from measurements of system output which contain random errors. To use the Kalman filter, it is therefore necessary to describe the internal state of the hydrological system in terms of a state vector. By this is meant a list of numbers that expresses in quantitative form the effect of all external influences on the system before the present moment, so that the future evolution of the catchment can be exactly given from the knowledge of the present state and the future inputs.

Given a model of the dynamics of the hydrological system that contains noise and measurements of the output of the hydrological system that also contains noise, the Kalman filter balances the model and output error terms to provide an optimal estimator for the model state which minimizes the forecast error variance. The following equations are a simplified description of the Kalman filter. The model is formulated within the required state-space framework:

System equation: $\boldsymbol{x}_{t+1} = \boldsymbol{F}_t \boldsymbol{x}_t + \boldsymbol{B}_t \boldsymbol{u}_t + \boldsymbol{w}_t$ (eq. 1)

Measurement equation: $\boldsymbol{y}_t = \boldsymbol{H}_t \boldsymbol{x}_t + \boldsymbol{v}_t$ (eq. 2)

State update equation: $\boldsymbol{x}_{t|t} = \boldsymbol{x}_{t|t-1} + \boldsymbol{K}_t \boldsymbol{v}_t$ (eq. 3)

$\boldsymbol{x}_t$ - (n x 1) vector of system states at time t

$\boldsymbol{x}_{t|t}$ - updated state vector at time t

$\boldsymbol{x}_{t|t-1}$ - forecasted state vector at time t

$\boldsymbol{y}_t$ - (m x 1) vector of system outputs at time t

$\boldsymbol{u}_t$ - (r x 1) vector of system inputs at time t

$\boldsymbol{F}_t$ - (n x n) state transition matrix

$\boldsymbol{B}_t$ - (n x r) weight matrix

$\boldsymbol{H}_t$ - (m x n) measurement matrix

$\boldsymbol{K}_t$ - (n x m) Kalman gain matrix

$\boldsymbol{v}_t$ - (m x 1) measurement noise vector

$\boldsymbol{w}_t$ - (n x 1) system noise vector

The above Kalman filter formalism is under the assumption that the system noise process and the measurement noise process are uncorrelated and that they are Gaussian white noise sequences. For a Gaussian process the joint probability distribution of the values at discrete timesteps is multivariate normal. It is completely determined by its mean, variance and covariance. The matrices $\boldsymbol{F}_t$, $\boldsymbol{B}_t$ and $\boldsymbol{H}_t$ may be time invariant or time variant.

The Kalman gain is used to update the vector of system states at time t. The correction of the system state given by the Kalman gain is small when the measurement noise is large, and increases when the precision in the observations of the system output improves. The correction increases when the system noise increases.

Suppose that we have observed the system output process $\boldsymbol{y}$ up to the time $(t\text{-}1)$, and that on the basis of these observations we have computed an optimal estimate of the system state vector $\boldsymbol{x}_{t-1|t-1}$ at time $(t\text{-}1)$. Using only this information, the best estimate of the system state at time t is provided by equation (1):

$$\boldsymbol{x}_{t|t-1} = \boldsymbol{F}_{t-1}\, \boldsymbol{x}_{t-1|t-1} + \boldsymbol{B}_{t-1}\, \boldsymbol{u}_{t-1}$$

The best forecast of the system output process $\boldsymbol{y}_t$ at time t is given by equation (2):

$$\boldsymbol{y}_{t|t-1} = \boldsymbol{H}_t\, \boldsymbol{x}_{t|t-1}$$

We have no other way of estimating $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$ except to use their mean values of zero. When the observation of $\boldsymbol{y}_t$ becomes available we can compare this with the forecasted value and update our estimate of $\boldsymbol{x}_t$ by taking a linear combination of the previous estimate $\boldsymbol{x}_{t|t-1}$ and the forecast error. This is exactly what equation (3) does:

$$\boldsymbol{x}_{t|t} = \boldsymbol{x}_{t|t-1} + \boldsymbol{K}_t\, \boldsymbol{v}_t = \boldsymbol{x}_{t|t-1} + \boldsymbol{K}_t\, (\boldsymbol{y}_t - \boldsymbol{H}_t\, \boldsymbol{x}_{t|t-1})$$

The Kalman gain matrix $\boldsymbol{K}_t$ may be regarded as a weighting factor. In order to ensure that the filter is of minimum variance, it is necessary to update $\boldsymbol{K}_t$ for each time step using information about the measurement noise vector $\boldsymbol{v}_t$ and the updated state vector $\boldsymbol{x}_{t|t}$. The information used is the covariance matrix of each of these vectors and its transpose.

There are problems with using the Kalman filter in hydrological forecasting, due to the non-linear nature of hydrological processes. The state-space model formulations which are necessary to apply the method leads to non-linear estimation problems. However, these can be solved using the extended Kalman filter. The extended Kalman filter approach is to apply the linear Kalman filter to a linearization of the non-linear problem. This is performed by continuously updating a linearization of the equations around the current state estimates. The non-linear functions describing system states and system output are linearized using Taylor series expansions retaining only linear terms. The optimal properties of the linear Kalman filter are lost, and the properties of the resulting state estimates are difficult to establish.

Aam *et al.* (1977) applied the extended Kalman filter for updating the HBV-model. The continuous diffierential equations of the model were linearized around the current state of the system using Taylor series expansions which were truncated after one term. Discrete solutions to the differential equations for each time interval were then obtained. Jørgensen *et al.* (1982) performed updating of the conceptual NAM-model using the extended Kalman filter. The elements of the state transition matrix $\boldsymbol{F}_t$ were made differentiable by a smoothing procedure. The state vector also included one of the model parameters, the time constant of a linear store used for channel routing.

Refsgaard (1997) presented some results from the WMO project "Simulated real time intercomparison of hydrological models". A comparison between the two conceptual models NAMS11 and NAMKAL using two updating procedures was performed. Both models comprise the NAM conceptual precipitation-runoff model, but contain different routing modules. NAMS11 used a hydraulic routing procedure and updating based on forecast error correction. NAMKAL used the NAM-model formulated in state-space form and a hydrological routing procedure with two linear reservoirs. Updating was performed using the extended Kalman filter approach. Both models used the same parameters for the NAM-model. Data from two cacthments, one small and one large were used. As the NAMKAL model lacked an explicit river routing procedure, it could not be expected to give good results for the large catchment (Bird Creek, USA, 2344 km^2), while river routing was not expected to be of any significance in the small catchment (Orgeval, France, 104 km^2). The differences in forecast lead times also reflect differences in cathment response time. The results of the simulations were as expected:

- In the small catchment the basic simulations were practically identical. Updating short term forecasts by the extended Kalman filter performed only slightly better than forecast error prediction.

- In the large catchment, the river routing was crucial and the NAMS11 which used a hydraulic routing procedure performed significantly better than NAMKAL.

The results indicated that an extended Kalman filter may be a better updating procedure in cases where the basic simulation by the hydrological model is good, but it is no guarantee for good forecasts in cases where the hydrological model provides a

poor simulation. Figure 3 shows results of running the two models in simulation mode and adaptive mode for the Bird Creek catchment.

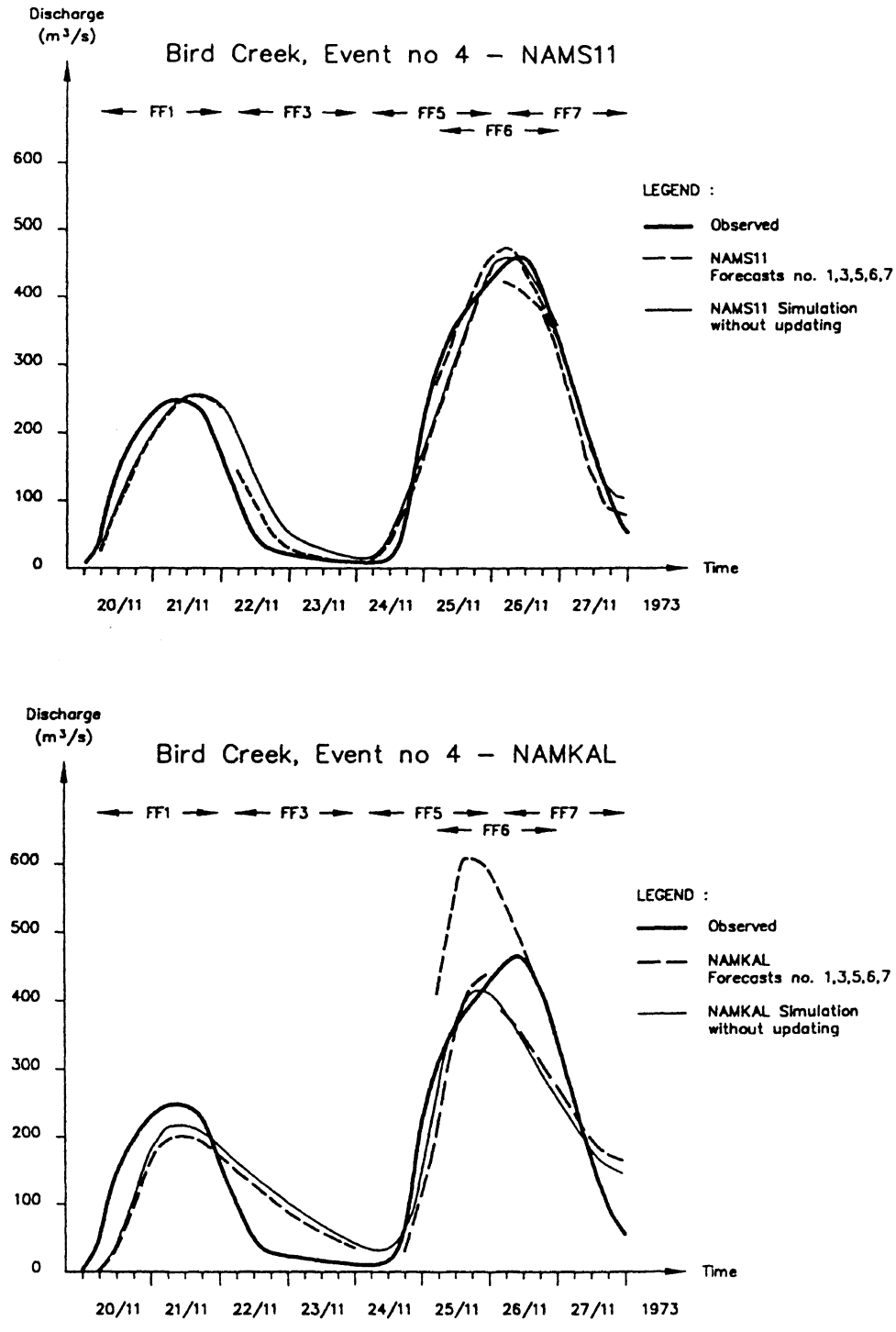


Figure 3. Simulation and forecast results for the NAMS11 and NAMKAL models for the 2344 km^2 Bird Creek catchment, USA (from Refsgaard, 1997).

Georgakakos (1986a, 1986b) coupled a precipitation forecasting model to a modified version of the U.S. NWSRFS conceptual model to produce streamflow forecasts. Variables such as soil moisture storage and streamflow were updated using an extended Kalman filter as state estimator. The model was applied to the Bird Creek catchment both with and without the updating procedure. Results showed that the nonupdating forecasting model produced time to peaks which were on the average 10 hours later than the observed peaks and peak flows which were on the average 37 percent lower than observed peaks. Incorporating real time data and updating improved these statistics so that the time to peak was forecast within 0.5 hour on average, and the average error in the peak discharge was reduced to 22 percent. A study where a non-linear storage model was applied to a 370 km^2 catchment in Japan was described by Lettenmaier and Wood (1993). The forecast error was reduced when updating with the extended Kalman filter was performed. The autocorrelation of the forecast errors showed that errors persisted longer when no updating was done, i.e. the errors were highly autocorrelated.

4.5 Correction of input variables in order to change model states

The following method may be used for adjusting the output of a hydrological model during forecasting. It does not require any changes in the model structure nor in the algorithms used in the model. Rather, this approach adjusts the input data to the model (precipitation and temperature) prior to the forecast and, consequently, the state variables of the model in such a way as to reproduce more closely the current and previous flows. These adjusted states are then used as initial values of the model at the start of the forecast period. This is a highly subjective method, and in order to be successful it requires considerable skill with regard to the model structure and the characteristics of the catchment being modeled. This method is described by Bergström *et al.* (1978). An automatic procedure for updating the HBV-model during forecasts of streamflow according to this method has been developed at the Swedish Meteorological and Hydrological Institute. This automatic method is also complemented with direct updating of the upper zone storage (Lindström and Carlsson, 2000).

4.6 Correction of state variables based on observations

Forecast adjustments may be accomplished by using measurements of state variables for comparison with the values generated by the model. For example, measurements of the water equivalent of the snow cover may be used to improve the streamflow forecasts of a conceptual model. This method requires considerable experience with the model and the catchment under consideration. Direct substitution of field measurements for values generated by the model would be incorrect as simplifications in the model structure results in its values being non-realistic (non-

physical). An example of problems caused by updating the snow water equivalent in the 218 km^2 Teita bru catchment in western Norway is shown in Figure 4. The snow store was reduced to zero at the end of September (before the new snow accumulation season started). The difference in the model's ability to match observed peaks before and after reducing the snow store is obvious. The problem may possibly be attributed to the following two causes; (i) snow may still be present in the catchment which covers altitudes in the range 130 - 1850 m a.s.l.; and (ii) glacier mass balance may not be described correctly. The last point is important as 17 % of this catchment is covered by glaciers.

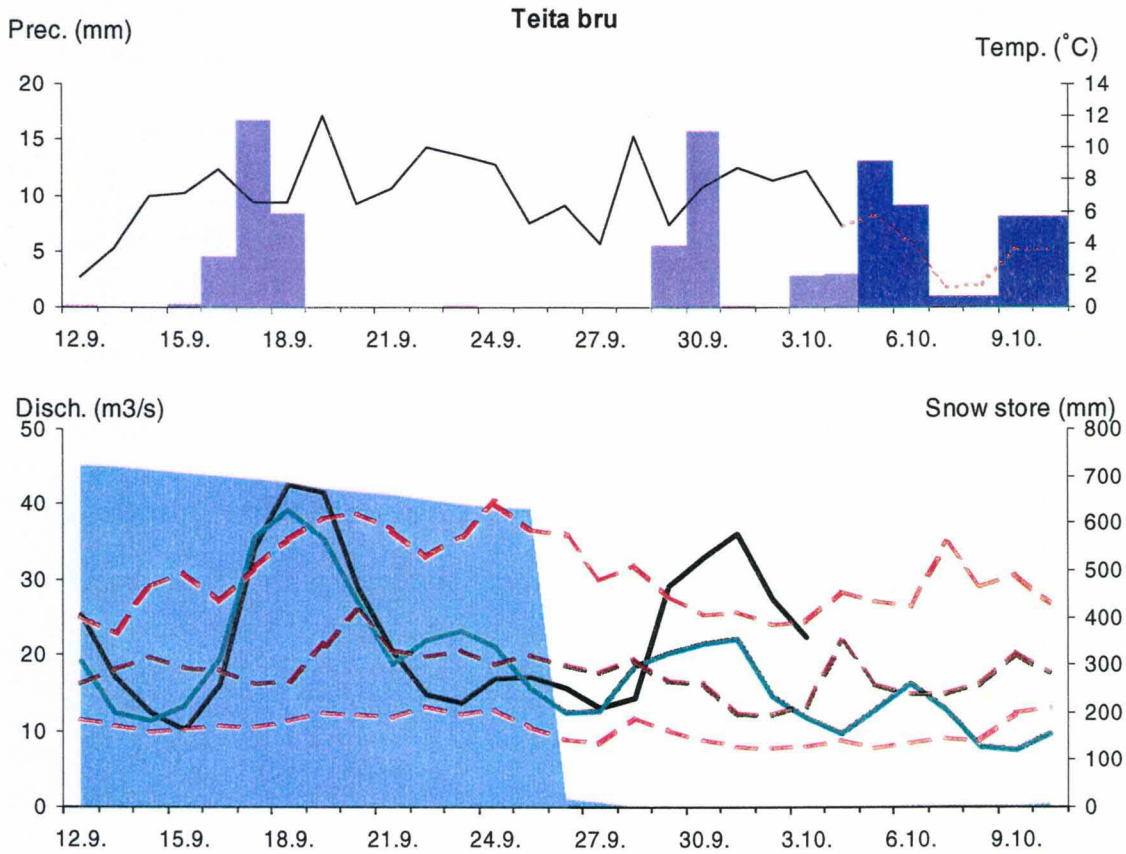


Figure 4. Observed (black) and simulated (green) streamflows from the 218 km^2 Teita bru catchment in western Norway during September and October 2000. The light blue shaded area shows the simulated snow store (snow water equivalent). The dottet lines represent the sample quartiles. The forecast was issued on 4 October.

The Snowmelt Runoff Model (Rango, 1995) is designed to simulate and forecast daily streamflow in catchments where snowmelt is a major runoff factor. It requires remote sensing input in the form of basin or zonal snow cover extent. Updating of the

snow cover is performed by substituting simulated snow cover by observed data from remote sensing or terrestrial observations. Depletion curves of snow covered area are determined from periodical snow cover mapping. These depletion curves are extrapolated and used to determine daily snow cover during the forecast lead time. The extrapolation procedure replaces the time scale of the depletion curves by cumulative daily snowmelt depths as computed by the model. The decline of the modified depletion curves depends on the initial accumulation of snow and not on the climatic conditions, as is the case for the conventional depletion curves.

Häggström (1994) studied the correlation between snow covered area estimated from NOAA-AVHRR data and the water equivalent of the snow store of the HBV-model using data from several Swedish catchments during six years. The purpose was to establish a linear relation, which could be used for updating the model during the snowmelt season. In all but one of the catchments was the correlation considered too low for using satellite data for updating the model. The nearly linear relationship observed in this one catchment was explained by the fact that it had a much larger area and larger altitude differences than the others.

Turpin *et al.* (1999) and Johansson *et al.* (2000) investigated a method for updating the snow store simulated by the HBV-model during real time hydrological forecasting. The model divided the catchment into several elevation zones, with a spatial distribution of snow water equivalent in each zone. Remote sensing data from satellites with optical and microwave sensors were used to determine snow covered area. Each satellite pixel was classified as snow covered or snow free. This method relies on the assumption that the vertical distribution of snow water equivalent within a catchment is correctly described by the model. The model error is therefore the estimation of the total snow store. The observed snow covered area is compared to the snow covered area simulated by the model for a period before or after the date of the satellite image. When a date is found with the same snow covered area as in the remote sensing image, the model state of that date is transferred to the model state of the date of the image. If the observed snow covered area was underestimated by the model, a model state from a previous date was selected and vice versa. The model was calibrated against observed runoff and the distribution of snow covered area, then the updating procedure was applied during one snowmelt season. The updating procedure had no impact on the forecasts when the snow distribution simulated by the model was in agreement with the observed snow distribution. When the snow store simulated by the model was too small, the updating procedure improved the runoff forecasts.

4.7 Updating the parameters of a conceptual model

The runoff coeficcient in the Snowmelt Runoff Model (Rango, 1995) takes care of the losses, that is the difference between the available water volume (snowmelt and rainfall) and the outflow from the catchment. On a long term basis, it should correspond to the ratio between runoff and precipitation. At the start of the snowmelt season, losses are small because they are limited to evaporation from the snow

surface and the runoff coeffcient is large. Towards the end of the snowmelt season evapotranspiration and interception losses increase as vegetation grows, and the runoff coefficient decreases. In addition, the runoff coefficient is usually different for rain and snowmelt events. This parameter must be updated if runoff forecasts are not successful.

Some parameters change for natural reasons during the course of time. For example, the albedo of the snow surface changes according to the properties of the snow (wetness, impurities, particle size, density, surface roughness). However, the cause of this change may be attributed to known factors such as metamorphism, contamination of the snow surface by dust, forest litter and other material, and the occurence of melt, rain and snow events. This rate of change should therefore be described using physical principles or empirical relationships, which is not an updating procedure.

5. Data assimilation

An issue of great practical and theoretical interest is objective methods for updating of spatially distributed models, based on optimal use of information from different sources. This can be accomplished through a process known as data assimilation. This is a technique that is most closely associated with atmospheric data analysis where it is used to generate a consistent four-dimensional picture of the state of the atmosphere taking advantage of the great variety of atmospheric observations. The state of the atmosphere is characterized by its composition, its thermodynamical state as specified by pressure, temperature and specific humidity, and its three-dimensional velocity field. A complete description of the state of the atmosphere should also include the global distribution of other variables, such as cloudiness, aerosols, precipitation and so on, because they influence the large-scale behaviour of the atmosphere.

Numerical weather forecast models use the physical constraints imposed by the governing equations of the hydrodynamics and thermodynamics of the atmosphere to forecast its future states. These governing equations also provide a sound basis for filling in the spaces between observations with model data. The purpose of data assimilation is to provide the proper initial conditions for the numerical weather forecast models. However, there are always insufficient data to completely specify the state of the atmosphere. Therefore, a model integrated from such imperfect initial conditions will eventually diverge from the true state of the atmosphere because of the imperfections in the initial state and the inherent model faults. Atmospheric data are therefore assimilated into the model at their true time, forcing the model closer to the true atmospheric state. Based on the state of the atmosphere at a previous time, a first guess of the state of the atmosphere at the current time is made using the numerical forecasting model. It is updated with the most recent observations. The updating procedure interpolates the forecasted state variables from the model grid to the observation stations and the difference between the two fields is obtained. The

increments required to correct the model states are determined through an objective analysis technique. The numerical model incorporates this information to arrive at a physically consistent description of the state of the atmosphere. This state is used as the initial value for a new forecast, and the whole process is repeated. Initialization procedures are also necessary to satisfy an appropriate balance between mass and wind fields (Daley, 1991).

Frequent adjustments are necessary to avoid divergence between the model and the true atmosphere. During continuous data assimilation, the observations are assimilated at the same rate as they are observed, but somewhat behind real time to allow for data to be communicated and prossessed. The atmospheric state simulated by the numerical model is continuously adjusted to fit the new observations. A routine forecast can therefore be initiated at any time.

The data assimilation process results in more realistic model states, and if the model structure and parameters are properly specified, these more realistic states results in better forecasts. Data assimilation techniques are therefore also of interest in hydrology. As the interactions between precipitation, evapotranspiration, infiltration, subsurface moisture conditions and runoff production are described by conceptual and physically based hydrological models, some important hydrological data assimilation problems include characterization of the temporal and spatial distribution of snow cover, soil moisture content and depth to the groundwater table. Better information on the space and time structure of these variables could enhance our ability to forecast hydrological processes. Unfortunately, traditional sources of hydrological data such as groundwater piezometers have typically been too limited and too widely scattered to provide much insight about spatial variability. This situation may be changing as new monitoring technologies, including remote sensing, become available. These technologies may eventually enable us to construct pictures of hydrological variables which vary constinuously over both time and space. Data assimilation will provide the key to including these data in hydrological forecasting models. When using data assimilation methods in hydrology, one should be aware that mass and energy are not conserved. If the hydrological model has too much or too little water, the data assimilation process either destroys or creates water in order to make model states more realistic.

As the objective of data assimilation is to provide physically consistent estimates of spatially distributed environmental variables, it should in principle be possible to apply these techniques for updating spatially distributed hydrological models. Since the state variables are connected to points or elements in the landscape it is possible to couple observations from single stations and from remote sensing directly to the model. Furthermore, as spatially distributed estimates are defined everywhere, not just at the measurement points, it should be possbile to advect information from data-rich to data-poor regions using process descriptions compatible with physical laws such as conservation of mass, momentum and energy. In order for the hydrological model to be updated through a data assimilation cycle, it should therefore apply physically realistic representations of catchment processes to describe the evolution

of the state variables in time and space. Refsgaard (1997) used the term data assimilation to characterize any procedure which takes measurements of the system's state or output into consideration when preparing the forecasts. Meteorologist use this term to designate an objective procedure for updating an atmospheric model based on observations of the physical fields describing the state of the atmosphere. It may be difficult to classify different techniques for updating the state variables of a hydrological model. However, in order to use the term data assimilation, an objective method for updating model states should be used.

One of the key state variables in land surface hydrology is soil moisture in the root zone, as its amount determines whether infiltrated water will be returned to the atmosphere as evapotranspiration or flow laterally to the nearest stream as overland flow or subsurface flow. As the soil moisture content in the root zone rises, the runoff event caused by a certain amount of rain or snowmelt will generally be larger. Although soil moisture modeling is firmly grounded in theory, it relies on parameter estimates which are not adequately measured even at the fine model scales. Hence, model soil moisture estimates are imperfect and often drift away from reality during simulations. This may lead to a poor ability to simulate runoff, in particular during dry periods. In situ soil moisture measurements are generally expensive if they are to cover large areas. However, remote sensing with visible, infrared or microwave sensors may be used for estimating soil moisture over large areas regularly. Although success has been attained over the past decades in estimating soil moisture using passive and active microwave sensors, progress has been slow. Some factors impeding progress include landscape roughness, vegetation cover and the penetration depth of microwave radiation. Another problem is the disparity in scale between remote sensing data resolution, the scale of hydrological processes, and the model grid scale. As a result, currently there is no comprehensive method for assimilating remote soil moisture observations within a surface hydrology model at watershed or larger scales.

Houser *et al.* (1998) used four-dimensional data assimilation methods to correct surface zone soil moisture fields in TOPLATS (Franks *et al.*, 1997), a soil-vegetation-atmosphere transfer scheme based on the TOPMODEL concept (Beven *et al.*, 1995). TOPLATS incorporates simple representations of land-surface atmosphere interactions, ground heat flux, three subsurface stores, surface runoff and subsurface flow. The study was performed in148 km^2 experimental watershed in Arizona. Microwave brightness temperature data were collected from an aircraft flying at 600 m above the ground. A linear relationship was used to invert these data to surface zone soil moisture. This relationship was established using gravimetric soil moisture data. The meteorological forcings of the model were assumed spatially constant, except precipitation which was spatially distributed. Information on surface zone soil moisture were also assimilated into the subsurface using knowledge of the surface-subsurface correlation. Soil moisture data collected each day at several depths were used for model verification.

The data assimilation methods used were:

1. Statistical correction data assimilation. The modeled surface soil moisture mean and standard deviation are adjusted to match the observed mean and standar deviation. This method assumes that the statistics of the observations are correct, which is more reasonable than assuming that each observation is perfect.

2. Newtonion nudging data assimilation. This method pushes the model states toward the observed data by adding a term to the prognostic equation that is proportional to the difference between the two states. This method gradually corrects the model fields.

3. Statistical interpolation data assimilation. This is a minimum variance method that is closely related to kriging. The degree of spatial variability of a regionalized variable can be expressed by the semivariogram. If measurements have been made at scattered points and the form of the semivariogram is known, it is possible to estimate the value of the variable at any location.

The two latter methods were implemented in all three soil layers of the TOPLATS model using horisontal and vertical weighting functions. The data assimilation methods resulted in improved simulations of surface soil moisture compared to running the model without data assimilation, whereas in the root zone the results fell into two groups corresponding to the methods with and without the capability for vertical assimilation of information. Another result from this study was that the simple methods were available to perform almost as well as complex techniques if enough data are available. The ability of data assimilation methods to advect information from data-rich regions to data-poor regions depends on the assumption that there is some natural spatial structure or correlation in the data. A geostatistical analysis showed that the spatial correlation structure of the microwave radiometer-derived soil moisture varied with time since the previous rainfall event. This is a problem as data assimilation methods generally assume that correlation structures are temporally invariant. This was considered a topic of further research.

Ottlé and Vidal-Madjar (1994) performed assimilation of soil moisture inferred from infrared remote sensing in a hydrological model. The objective of this study was to improve the simulation of the water balance. A distributed conceptual hydrological model which describes surface and subsurface flow with three soil layers (upper 10 cm, root zone from 0.1 m to 1 m, and groundwater zone) was applied to a 120 by 120 km^2 study area in south-west France (HAPEX-MOBILHY). The model domain was discretized in 1208 squares based on a soil classification. The sides of the squares ranged between 1.25 and 5 km. The direction of flow from each grid cell was determined by topography. The model domain was further divided into 84 homogeneous precipitation zones for which the same daily amount of rainfall was assigned. These meteorological zones were grouped into six evaporation zones around meteorological stations where evaporation was measured routinely. These measurements have been shown to be well correlated with aircraft flux measurements

integrated over a larger region. The model was run with a daily time step and calibrated with nine years of data to determine the storage capacity, infiltration rate and hydraulic diffusivity for each soil type.

Surface soil moisture was estimated from remote sensing using NOAA-AVHRR satellite data. From knowledge of precipitation and potential evaporation, the hydrological model simulated the evolution of soil moisture in the surface layer of the catchments in the study area. When a clear sky satellite image was available, an inversion procedure was used to calculate the surface radiative temperature with the hydrological model's estimate of surface soil moisture as initial values. This surface humidity was modified until the difference between the modeled surface radiative temperature and the one observed by the satellite was less than 1.5 K. The difference between the surface soil moisture estimated in this way and the initial one simulated by the hydrological model was then averaged over each precipitation area and the difference was introduced into the hydrological model, which was run until a new cloudless image was available.

Comparisons with observed data showed that the data assimilation procedure resulted in improved simulations of runoff and evapotranspiration as compared to running the model without correcting its surface soil moisture. The major difference was that an increase in soil moisture content during dry periods in summer lead to increased evapotranspiration and runoff. Small flows during precipitation recovery at the end of the summer were better simulated using the data assimilation procedure. Furthermore, flood peaks were also higher during the assimilation runs. Running the model without data assimilation generally underestimated runoff during periods of low flow and peak flow. Figure 5 shows that flux measurements of monthly evaporation at the six meteorological stations were better described by the model runs using data assimilation, especially for the months of July and August.

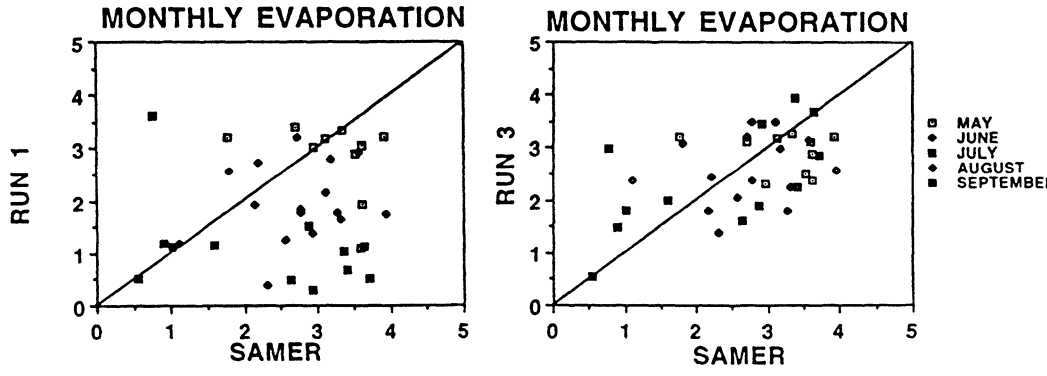


Figure 5. Monthly evaporation simulated without data assimilation (RUN 1) and with data assimilation (RUN 3) vs. measurements at six meteorological stations (SAMER) (from Ottlé and Vidal-Madjar, 1994).

The studies by Houser *et al.* (1998) and Ottlé and Vidal-Madjar (1994) have demonstrated that data assimilation techniques may be used to provide more realistic estimates of model states and model fluxes. Of course, one cannot expect to completely describe real world variability with a data assimilation technique that relies on a limited amount of scattered data. Assimilated state estimates will inevitably be smoother than the actual states. Nevertheless, it is still useful to try to extract as much information as possible from available measurements. This must be done in a manner that is both physically reasonable and computationally efficient.

Hydrological processes with characteristic length scales larger than the computational elements of a distributed model are represented explicitly by element to element variations, whereas processes with length scales smaller than the element size are represented implicitly. Data assimulation methods require distributed hydrological models which provide realistic estimates of state variables within their computational elements. An essential research question is therefore how to acccount for heterogeneities at spatial scales smaller than the model element size. One of the keys to more efficient data assimilation will therefore be the development of better methods for describing the spatial and temporal variability in hydrological processes. Furthermore, the data collection programs must provide the data required for updating the models (McLaughlin, 1995).

6. Information sources

The data required for developing the forecasting method include conventional hydrological and meteorological time series that are necessary for testing and evaluating the trial forecasts and the models. It also includes the data required to describe catchment characteristics, e.g. sub-catchment areas, land use, soil and vegetation characteristics, topography, channel dimensions and slopes. The reliability of a forecasting procedure can be directly related to the quality, amount and spatial distribution of the data used to develop the procedure and to their consistency. Care must be taken to ensure that there is not a bias between the data used to develop the forecasting procedure and the data used for operational forecasting. For this reason, the consistency of the records is as important as the quantity of data.

Procedures for updating of hydrological forecasting models are based on data from a hydrological observation network and possibly information about the operation of water resources systems, e.g. reservoirs and flood protection works. A hydrological data network is a group of data collection activities that are designed and operated to address a set of objectives. The information needed to perform real time updating of hydrological forecasting models is required to be timely and reliable. The hydrological forecasting network must therefore include procedures for acquistion, transmission and processing of data in real time. Although these systems may be considered as a component of the sources of information, this topic is not considered.

The informations sources used for real time updating of hydrological forecasting models includes the hydrological variables required by the forecasting method to specify as accurately as possible the state of the catchment prior to the issue of the forecast. It also includes measurements of the forecast variable that may be used to monitor the performance of the forecast and update the forecasting model. Data requirements ideally depend upon the forecasting and updating procedures that is to be used, and the hydrological characteristics of the catchment. In practice, data availability may place restrictions on the choice of the forecasting method. However, steps should always be taken to meet the ideal data requirements.

The procedures for real time updating of hydrological forecasting models described previously must be provided with some of the following data:

- River stage or discharge.
- Metorological variables which may be used to update model states at the beginning of the forecasting period. These are the variables used to drive the model, e.g. precipitation or temperature, however, observed data must be altered in order to a arrive at a different model state. Since these data are not measured, but manipulated in a subjective manner, measurements of meteorological variables is not considered.
- Measurement of the states of the different storages of the catchment which may be used for updating the state variables of the forecasting model, e.g. snow covered area or snow water equivalent, soil moisture data, groundwater data.

River stage or discharge must be measured at the point where the forecast of streamflow is made. The stream gauging stations should preferrably provide a continuous streamflow record. Since continuous streamflow records are usually not available unless acoustic or electromagnetic methods are used, records of discharge must be computed from the relationship between stage and discharge, as defined by periodic discharge measurements and a continuous record of stage. The selection of a particular site for the gauging station should be guided by several criteria. The most important are that the downstream hydraulic controls, which control the water level at the measurement site and the stage-discharge relationship, are stable and sensitive to changes in discharge. The hydraulic control should also be unaffected by variable backwater conditions caused by tidal effects, downstream lakes, or inflowing tributaries.

Hydrological forecasting models which includes a description of the distribution of state variables in a catchment can benefit greatly from information about snow covered area, snow water equivalent, soil moisture or groundwater conditions. In situ measurements of point processes will generally not be able to provide data with a spatial extent or resolution which is consistent with the characteristic scales of hydrological processes, unless the catchment is small. However, hydrological forecasting is usually performed for areas which are too large for manual observations to provide the data required for updating a forecasting model.

Remote sensing from the ground, from aeroplanes and from satellites may have the greatest potential for supplying data with the temporal and spatial coverage and resolution required for updating the state variables of hydrological forecasting models. Remote sensing technology is used to obtain information about a phenomenon through the analyses of data acquired by a sensor that is not in direct contact with the target of the investigation (Ritchie and Rango, 1996). Measurements of the reflected or emitted electromagnetic spectrum from the earth's surface are used to infer the properties of the landscape. A simple example is aerial photography in the visual wavelengths. The minimum area which can be mapped with an adequate accuracy depends on the spatial resolution of the sensor. Furthermore, remote sensing can provide observations from otherwise inaccessible terrain. Remote sensing is also useful for providing input to forecasting models, e.g. areal rainfall or movement of storms.

Some examples of remote sensing observations of hydrological state variables are:

- Microwave measurements of surface soil moisture by aircraft or satellites. These techniques depend on a relationship between soil moisture and another soil property, e.g. the dielectric constant. The passive microwave techniques use radiometers to measure the thermal emission of microwaves from the ground. The intensity of this radiation is proportional to the brightness temperature, which is the product of surface temperature and emissivity. The latter depends on the dielectric constant of the soil. The active microwave techniques (radar) emit radiation from an artificial source and measure the intensity of reflected radiation. The reflectivity of the soil depends on the dielectric constant. The surface roughness and the vegetation may influence soil emissivity and reflectivity. These factors reduce the sensitivy of the methods and should be accounted for.

- Mapping the spatial extent of groundwater discharge areas (where the entire soil profile is saturated). Several physically based models, e.g. TOPMODEL rely on a correct simulation of this area in order to forecast the response of a catchment to rain or snowmelt events. Franks *et al.* (1998) mapped the extent of saturated areas by combining data from active microwave instruments with an analysis of topography.

- Snow covered area mapped by microwave, visual or near-infrared sensors mounted on aircrafts or satelittes. Snow covered areas appear considerably brighter than snow-free surfaces.

- Snow water equivalent mapped by active and passive microwave sensors.

- Radar surveys used for measuring snow water equivalent based on the propagation time of the radar wave.

- Gamma radiation surveying of snow water equivalent. This method measures the attenuation by a layer of snow of the gamma radiation emitted from natural radioactive elements in the ground.

7. Concluding remarks

Although data assimilation methods in principle may be used to update the state variables of hydrological forecasting models, this cannot be expected to be succesful before the models and the data collection program are able to represent the natural spatial variability of hydrological processes. The theoretical development of distributed hydrological models has relied on the possibilty of using effective parameter values at the scale of the computational elements of these models. Effective parameters are assumed to take into account all of the local scale heterogeneity of soil, vegetation, topography, surface roughness, water stress, and meteorological input that influence the integrated discharge and evapotranspiration fluxes. However, it has been shown that the use of effective parameters with even the most "physically based" distributed models cannot be expected to produce accurate simulations of these fluxes. This is basically because of the non-linearities of the processes involved, together with the heterogeneity of the natural systems (Beven, 1989). Thus there is a need for a theory that will allow the correct description of state variables and the partitioning of rainfall or snowmelt at the spatial scale of interest. Typical spatial scales in precipitation-runoff models are the hillslope scale (~100 m), the catchment scale (~10 km), and the regional scale (~1000 km).

The spatial scale of a hydrological process refers to a characteristic length related to its natural spatial variability. This scale may be defined by the spatial extent, period or correlation length of the process, depending on its nature. Ideally, hydrological processes should be observed at the scale they occur. However, this is not always possible, as the spatial extent of the data set, the spacing between the samples (resolution) and the support (integration volume) of the samples is restricted by practical and economical constraints. Similar considerations apply to the modeling scale. The triplet spacing, extent and support of the observation and model scales can be thought of as a filter relative to the scale of natural variability (Blöschl and Sivapalan, 1995; Blöschl, 1999). For example, for a transect of time domain reflectometry (TDR) soil moisture samples in a catchment, this scale triplet may have typical values of 10 m spacing between the samples, 200 m extent (the length of the transect), and 10 cm support (the diameter of influence of the TDR measurements). Similarly, for a finite-difference model of spatial hydrological processes in the same catchment, the scale triplet may have typical values of 50 m spacing between the model nodes, 1000 m extent (the length of the model domain), and 50 m support (the size of the computational elements) (Western and Blöschl, 1999). Processes with larger spatial scale than the spatial extent of observations appear as trends in the data, whereas processes smaller than the resolution appear as noise. For example, it is well known that the highest frequency that can be detected by a data set is the Nyquist frequency (Blöschl and Sivapalan, 1995).

The conclusion from these considerations are that unless the process, measurement and model scales are consistent, updating of forecasting models based on observations of hydrological processes cannot be expected to be successful. This problem does not apply to streamflow which is an areally integrated value. At the

present, the largest possibility for updating hydrological forecasting models may therefore be through an adjustment method based on observed streamflow at the forecast point.

8. References

Aam, S., Fossdal, M., Wingård, B., Killingtveit, Å., and Fjeld, M. (1977) Hydrologisk model for drift av kraftverk, *Elektrisitetsforsyningens forskningsinstitutt, Norges vassdrags- og elektrisitetsvesen, Institutt for vassbygging, Selskapet for industriell og teknisk forskning ved Norges tekniske høgskole, Report,* 151 pp.

Abbot, M.B., Bathurst, J.C., Cunge, J.A., O'Connel, P.E., and Rasmussen, J. (1986a) An introduction to the European Hydrological System – Système Hydrologique Européen, "SHE", 1: History and philosophy of a physically-based, distributed modelling system, *Journal of Hydrology, 87*, 45-59.

Abbot, M.B., Bathurst, J.C., Cunge, J.A., O'Connel, P.E., and Rasmussen, J. (1986b) An introduction to the European Hydrological System – Système Hydrologique Européen, "SHE", 2: Structure of a physically-based, distributed modelling system, *Journal of Hydrology, 87*, 61-77.

Bergström, S. (1995) The HBV model, In: Singh, V.P. (Ed.), *Computer Models of Watershed Hydrology*, Water Resources Publications, Highlands Ranch, 443-476.

Bergström, S., Persson, M., and Sundqvist, B. (1978) Operational hydrological forecasting by conceptual models, *Swedish Meteorological and Hydrological Institute, HB Report no. 32*, 21 pp.

Blöschl, G. (1999) Scaling issues in snow hydrology, *Hydrological Processes, 13*, 2149-2175.

Blöschl, G., and Sivapalan, M. (1995) Scale issues in hydrological modelling: a review, In: Kalma, J.D., and Sivapalan, M. (Eds.), *Scale Issues in Hydrological Modeling*, Wiley, Chichester, 9-48.

Burnash, R.J.C. (1995) The NWS River Forecast System - catchment modeling, In: Singh, V.P. (Ed.), *Computer Models of Watershed Hydrology*, Water Resources Publications, Highlands Ranch, 311-366.

Beven, K. (1989) Changing ideas in hydrology - the case of physically based models, *Journal of Hydrology, 105*, 157-172.

Beven, K.J., Lamb, R., Quinn, P., Romanowicz, R., and Freer, J. (1995) TOPMODEL, In: Singh, V.P. (Ed.), *Computer Models of Watershed Hydrology*, Water Resources Publications, Highlands Ranch, 627-668.

Calver, A., and Wood, W.L. (1995) The Institute of Hydrology Distributed Model, In: Singh, V.P. (Ed.), *Computer Models of Watershed Hydrology*, Water Resources Publications, Highlands Ranch, 595-626.

Daley, R. (1991) *Atmospheric Data Analysis*, Cambridge University Press, Cambidge, 457 pp.

Eagleson, P.S. (1994) The evolution of modern hydrology (from watershed to continent in 30 years), *Advances in Water Resources, 17*, 3-18.

Franks, S.W., Beven, K.J., Quinn, P.F., and Wright, I.R. (1997) On the sensitivity of soil-vegetation-atmosphere transfer (SVAT) schemes: equifinality and the problem of robust calibration, *Agricultural and Forest Meteorology, 86,* 63-75.

Franks, S.W., Gineste, P., Beven, K., and Merot, P. (1998) On constraining the predictions of a saturated model: The incorporation of fuzzy estimates of saturated areas in the calibration process, *Water Resources Research, 34*, 787-797.

Georgakakos, K.P. (1986a) A generalized stochastic hydrometeorological model for flood and flash-flood forecasting. 1. Formulation, *Water Resources Research, 22,* 2083-2095.

Georgakakos, K.P. (1986b) A generalized stochastic hydrometeorological model for flood and flash-flood forecasting. 2. Case studies, *Water Resources Research, 22*, 2096-2106.

Green, C.S. (1979a) Real time flow forecasting on the River Dee, *Institute of Hydrology, Report*, 45 pp.

Green, C.S. (1979b) An improved subcatchment model for the River Dee, *Institute of Hydrology, Report no. 58*, 45 pp.

Houser, P.R., Shuttleworth, J.W., Famiglietti, J.S., Gupta, H.V., Syed, K.H., and Goodrich, D.C. (1998) Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, *Water Resources Research, 34*, 3405-3420.

Häggström, M. (1994) Snökartering i svenska fjällområdet med NOAA-satelittbilder, *Swedish Meteorological and Hydrological Institute, SMHI Hydrologi no. 57*, 57 pp.

Johansson, B., Caves, C., Ferguson, R., and Turpin, O. (2000) Remote sensing data on snow cover in the HBV model, Proceedings of the Nordic Hydrological Conference 2000, *Nordic Hydrological Programme, Report no. 46*, 339-346.

Jørgensen, G., Refsgaard, J.C., and Rosbjerg, D. (1982) Updating procedures for real time operation of hydrological models, Proceedings of the Nordic Hydrological Conference 1982, *Nordic Hydrological Programme, Report no. 2*, 738-747.

Langsrud, Ø., Frigessi, A., and Høst, G. (1998) Pure model error of the HBV-model, *HYDRA, Notat nr. 4/1998, 28 pp.*

Lettenmaier, D.P., and Wood, E.F. (1993) Hydrologic forecasting, In: Maidment, D.R. (Ed.), *Handbook of Hydrology*, McGraw-Hill, New York, 26.1-26.30.

Lindström, G., and Carlsson, B. (2000) Evaluation of updating methods for short range forecasting by the HBV model, Proceedings of the Nordic Hydrological Conference 2000, *Nordic Hydrological Programme, Report no. 46*, 380-386.

McLaughlin, D. (1995) Recent developments in hydrologic data assimilation, *Reviews of Geophysics, 33*, 977-984

Mroczkowski, M., Raper, G.P., and Kuczera, G. (1997) The quest for more powerful validation of conceptual catchment models, *Water Resources Research, 33*, 2325-2335.

O'Connel, P.E. (1991) A historical perspective, In: Bowles, D.S., and O'Connel, P.E. (Eds.), *Recent Advances in the Modeling of Hydrologic Systems*, NATO ASI Series C: Mathematical and Physical Sciences, Vol. 345, 3-30.

O'Connel, P.E., and Clarke, R.T. (1981) Adaptive hydrological forecasting - a review, *Hydrological Sciences Bulletin, 26(2),* 179-205.

Ottlé, C, and Vidal-Madjar, D. (1994) Assimilation of soil moisture inferred from infrared remote sensing in a hydrological model over the HABEX-MOBILHY region, *Journal of Hydrology, 158*, 241-264.

Rango, A. (1995) The Snowmelt Runoff model (SRM), In: Singh, V.P. (Ed.), *Computer Models of Watershed Hydrology*, Water Resources Publications, Highlands Ranch, 477-520.

Refsgaard, J.C. (1997) Validation and intercomparison of different updating procedures for real-time forecasting, *Nordic Hydrology, 28*, 65-84.

Ritchie, J.C., and Rango, A. (1996) Remote sensing applications to hydrology: introduction, *Hydrological Sciences Journal, 41*, 429-431.

Skaugen, T. (1999) Kvantifisering av usikkerhet i meteorologiske og hydrologiske prognoser, Sammendrag av delprosjekter, *HYDRA, Notat nr. 3/1999*, 11 pp.

Turpin, O., Ferguson, R., and Johansson, B. (1999) Use of remote sensing to test and update simulated snow cover in hydrological models, *Hydrological Processes, 13*, 2067-2077.

Western, A.W., and Blöschl, G. (1999) On the spatial scaling of soil moisture, *Journal of Hydrology, 217*, 203-224.

WMO (1994) *Guide to Hydrological Practices*, World Meteorological Organization, WMO-no. 168, 735 pp.

Wood, E.F., and O'Connel, P.E. (1985) Real-time forecasting, In: Anderson, M.G. and Burt, T.P. (Eds.), *Hydrological Forecasting*, Wiley, Chichester, 505-558.

Young, P.C. (1986) Time-series methods and recursive estimation in hydrological system analysis, In: Kraijenhoff, D.A. and Moll, J.R. (Eds.), *River Flow Modelling and Forecasting, Water Science and Technology Library*, Reidel, Dordrecht, 129-180.

Denne serien utgis av Norges vassdrags- og energidirektorat (NVE)

Utgitt i Dokumentserien i 2000

Nr. 1 Rune V. Engeset (red.): NOSIT - utvikling av NVEs operasjonelle snøinformasjonstjeneste (77 s)

Nr. 2 Inger Sætrang (red):Statistikk over overføringstariffer (nettleie) i Regional- og distribusjonsnettet 2000 (55 s.)

Nr. 3 Bjarne Kjøllmoen, Hans Christian Olsen og Roger Sværd: Langfjordjøkelen i Vest-Finnmark Glasiohydrologiske undersøkelser (56 s.)

Nr. 4 Turid-Anne Drageset: Flomberegning for Otta og Gudbrandsdalslågen -Flomsonekartprosjektet (40 s.)

Nr. 5 Erik Holmqvist: Flomberegning for Trysilvassdraget, Nybergsund (311.Z) - Flomsonekartprosjektet (20 s.)

Nr. 6 Turid-Anne Drageset: Flomberegning for Jølstra - Flomsonekartprosjektet (30 s.)

Nr. 7 Arne Hamarsland og Tore Olav Sandnæs: Vassdragsforvaltning i Japan - hva kan vi lære? (30 s.)

Nr. 8 Inger Sætrang: Oversikt over vedtak. Tariffer og vilkår for overføring av kraft i 1999 (12 s.)

Nr. 9 Turid-Anne Drageset: Flomberegning for Daleelva i Høyanger (079.Z) - Flomsonekartprosjektet (28 s.)

Nr. 10 Lars-Evan Pettersson: Flomberegning for Glommavassdraget oppstrøms Vorma (002.E-T) Flomsonekartprosjektet (46 s.)

Nr. 11 Henriette Hansen, Lars Roald: Flomsonekartprosjektet. Ekstremvannsanalyse i sjø ved utvalgte stasjoner (39 s.)

Nr. 12 Arne Hamarsland: Biotoptiltak for laks og ørret i British Colombia og Washington (21 s.)

Nr. 13 Lars-Evan Pettersson: Flomberegning for Saltdalsvassdraget (163.Z) Flomsonekartprosjektet (25 s.)

Nr 14 Miriam Jackson: Svartisen Subglacial Laboratory (27 s.)

Nr. 15 Lars-Evan Pettersson: Flomberegning for Gaulavassdraget (122.Z) Flomsonekartprosjektet (22 s.)

Nr. 16 Lars-Evan Pettersson: Flomberegning for Tokkeåi ved Dalen (016.BD-BL) Flomsonekartprosjektet (20 s.)

Nr. 17 Stein Beldring: Real time updating of hydrological forecasting models Methods and information sources (37 s.)

Norges vassdrags- og energidirektorat

Norges vassdrags- og energidirektorat
Middelthunsgate 29
Postboks 5091 Majorstua
0301 Oslo

Telefon: 22 95 95 95
Telefaks: 22 95 90 00
Internett: www.nve.no