Soil–Vegetation–Atmosphere Transfer Schemes and Large-Scale Hydrological Models (Proceedings of a symposium held during the Sixth IAHS Scientific Assembly at Maastricht, The Netherlands, July 2001). IAHS Publ. no. 270, 2001.

Assimilation of soil moisture in a hydrological model for flood forecasting

A. WEISSE, C. MICHEL, D. AUBERT & C. LOUMAGNE

Cemagref, BP 44, F-92163 Antony Cedex, France

e-mail: anne.weisse@cemagref.fr

Abstract This paper introduces a parameter updating procedure of the "variational" type that can be combined with any conceptual rainfall–runoff model for flood forecasting purposes. The main feature of this method is that it carries out updating by reference not only to recent streamflow observations, as do classic procedures, but also to soil moisture measurements, which can be retrieved either from TDR probes or from satellite remote sensing systems. The aim of the research was to assess the usefulness of this additional soil moisture information. The application of this new methodology has been carried out on a sub-basin of the River Seine at Paris. The first results are presented and, for this sub-basin, soil moisture seems not to be very useful to increase the efficiency of the updating procedure. But, due to the reduced number of recorded flood events, no definite conclusion can be reached.

Key words assimilation; soil moisture; flood forecasting

INTRODUCTION

Flood forecasting is one of the unsolved problems of operational hydrology (Garrote & Bras, 1995). A wide range of methods has been proposed to overcome this challenge, from simple error correction to the use of artificial neural networks (Zealand *et al.*, 1999). Because rainfall–runoff models are far from being perfect, hydrologists need to put such models in better compliance with the current observations prior to use in forecasting mode. This operation has been termed updating in hydrology and assimilation in meteorology.

O'Connell & Clarke (1981) and, more recently, Refsgaard (1997) reported on the four different methodologies used for model updating. These methodologies depend on what is considered to be the main cause of discrepancy between observed and computed streamflow values. First, one can correct the errors in rainfall data and therefore make corrections of the inputs to improve the model accuracy (input updating). Another point of view is to recognize that the model suffers from deficiencies that must be remedied. In this case, two additional procedures can be proposed: one is to alter the current state of the model (state updating); the other is to better adjust the model to current conditions by changing its parameter values (parameter updating). The fourth procedure considers the actual streamflow as the sum of the model output and an error term, which has to be modelled to allow for a prediction about its short-term realizations (error correction). This updating is especially used with black box models. In addition, these four procedures can be mixed together.

The inability of a model to produce correct streamflow values generally translates into parameter uncertainty. Parameter calibration is the means used by a model structure to adjust to a given set of data. Therefore, parameter updating seems to be a natural way to amend errors in streamflow value. In this paper, a specific methodology of parameter updating is presented. We show the way it can assimilate external information such as soil moisture. Soil moisture could be derived from Earth observation (EO) data using microwave spaceborne synthetic aperture radar (SAR) images (Quesney *et al.*, 2000). However, these data are too scarce for operational use (the satellite passes over the basin, in the same configuration, only every 35 days). Time domain reflectometry (TDR) probes provide measurements of soil moisture at a daily time step, which is more accurate for a hydrological purpose. The latter data are used in this study.

We first present the parameter updating methodology, and then the way soil moisture is assimilated. The first results over a sub-basin of the River Seine are then shown.

FLOOD FORECASTING: A PARAMETER UPDATING METHODOLOGY

The method presented here is a parameter updating one that has been first tested for the Seine sub-catchments (Yang & Michel, 2000). One original feature of the method is that the potential defect of the model, as regards high discharge values, is not dealt with within the model calibration. Intervening at the stage of model calibration is an approach that has been developed in other areas, e.g. in flood frequency estimation (Lamb, 1999). Here, a prior analysis of errors impairing flood discharges is carried out through a regression relationship:

$$Q_j = \alpha + \beta C_j + \varepsilon = C_{A,j} + \varepsilon \tag{1}$$

where Q is the observed flow rate, C is the flow rate calculated by the model, j refers to any day when C_j is above the threshold defining a flood, and ε is the error term of the regression. The terms α and β are additional parameters that take into account the potential lack of accuracy of the model for high discharge rates. They are determined prior to the updating procedure.

The additional originality of this approach lies in the way it deals with the main problem of parameter updating methodology, i.e. when action is taken to correct model behaviour, it is generally too late. The solution adopted here is to retrace backwards for a time (denoted D) long enough for parameter updating to produce a substantial effect at the present time. A "baseline" set of parameters, obtained by calibration over a long period, is assumed to represent the long-term behaviour of the basin. Every set of updating computations will always refer to this basic functioning.

The third new feature of the method lies in the way updating is accomplished. The aim is to reduce, possibly make equal to zero, the last error in streamflow value beyond the correction implied by equation (1). Reducing this ultimate error translates into a single equation that would have to be solved with respect to several parameters, a situation that typically leads to nondetermination. The solution adopted here was to keep the parameter updating the most likely, i.e. the closest to the baseline set. Actually, in the absence of additional information, the best thing to do is to alter all parameters around the baseline set, in the smallest way with respect to their uncertainties.

The fourth and last feature of the proposed method is a hedging measure. It may be possible that some sources of discrepancy remain that cannot be remedied by using general discharge correction (equation (1)) plus updating parameter shifts. For example, if the rainfall input is wrong, it would be unwise to force parameters to depart too much from the baseline set. To this end, a limit has been placed on the allowed departure: we simply chose to stop the procedure after a reasonable number of iterations (50) has been reached.

A procedure has been derived to implement the proposed methodology (see Fig. 1). At every time step when updating is needed, one retraces for a set number of time steps ($D = 60$ days) and an iterative process is initiated. During one stage of this process, all parameters are shifted, each one in turn, by $+\Delta x$ and then by $-\Delta x$, Δx being proportional to each parameter standard deviation. Successively, for each parameter, the model is run with the two possibilities $x + \Delta x$ and $x - \Delta x$. If, for one of these two traces, there is a reduction of the gap between the last computed discharge and the observed discharge, the corresponding modification in the parameter value ($+\Delta x$ or $-\Delta x$) is considered valid for the following step. When all parameter modifications have been tested, a new round of modifications is carried out, and so on. The iterative process is stopped when either the last error is reduced to an insignificant amount, or when a pre-set maximum number (50) of iterations has been reached. Here, Δx has been taken to be equal to a tenth of the standard deviation of the model parameter, x, as calculated with the linear approximation method (Mein & Brown, 1978; Troutman, 1985). Streamflows computed with such altered parameters are denoted $C_{B,j}$.

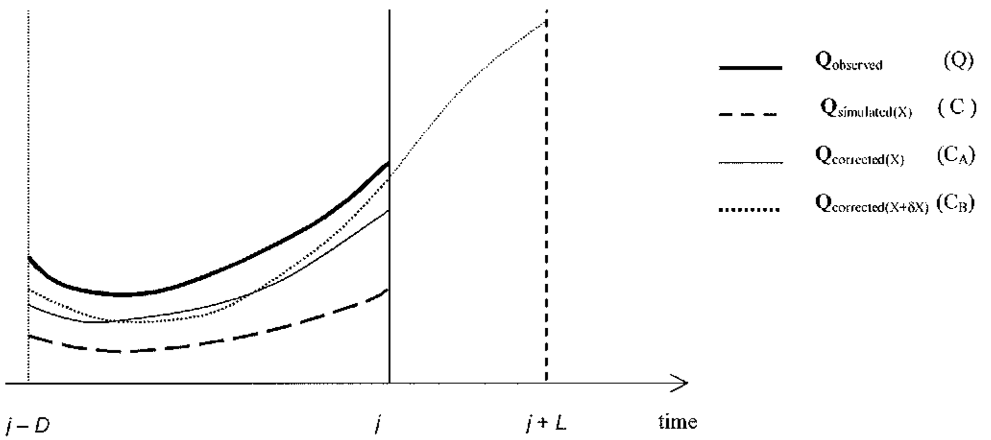

Fig. 1 Schematic representation of the parameter updating procedure.

This methodology has been tested in Yang & Michel (2000) against data from four basins upstream of Paris using the GR4J model (Edijatno *et al.*, 1999). The results show that the parameter updating method performed better than the traditional method of correcting errors using an auto-regressive model of order 1 (AR(1) model—Lettenmaier & Wood, 1992). Improvements, as measured by the persistence criterion (equation (5), later in this text), varied between 3 and 73%.

INTRODUCTION OF A NEW TYPE OF MEASUREMENT VALUE: SOIL MOISTURE DATA

A new constraint in the method

The method that has been summarized above provides a framework that can accommodate other sources of information. There are a lot of solutions for the problem of cancelling the last error in streamflow value. Here, the hypothesis is that soil moisture can be the information that needs to be exploited to work out the best parameter changes looked for in this updating procedure. The present research is aimed at analysing the different ways that soil moisture can be accounted for, and highlighting the one that offers the best prospects. The most obvious solution would be to work out a combination criterion, which would put together the last errors in streamflow and soil moisture values as determined by the model prior to updating the system. However, this would be a change in the objective function to be minimized. Such a change could blur the comparison between the original method and the present one. A more prudent approach is to only add a constraint such as:

$$\frac{1}{n}\sqrt{\sum_{j=1}^{n}\left(Ho_j - Hc_j\right)^2} < \frac{\sigma}{k} \qquad (2)$$

where *Ho* is the observed soil moisture, *Hc* is the modelled soil moisture (explicitly or implicitly), n is the chosen number of soil moisture values taken into account, σ is the standard error of *Hc* values, and k is a parameter that allows the severity of the constraint to be varied so as to find out the importance that can be attributed to moisture observations. If k is very small (e.g. $k < 0.1$), the constraint fades away, and if k becomes higher (e.g. $k > 1$), the constraint is tighter: k is a measure of the interest of complying with observations of soil moisture.

Another issue is whether or not it is necessary to use a model which simulates the evolution of soil moisture at different depths. One can argue that, if the model employed does not feature a store that can be identified closely to the observed soil moisture, there would be no possibility of benefiting from such measurements. On the other hand, it can be argued that, if the soil moisture is an output of the model, all the information drawn from soil moisture observations will be directed at improving this specific output at the expense of improving streamflow values generated by the model. To this end, two models will be tested. The first model, GRHUM (modèle du Génie Rural avec simulation de l'HUMidité—Loumagne *et al.*, 1996), has been especially developed to introduce a two-layer soil reservoir that simulates the surface and sub-surface soil moisture. The second one, GR4J (modèle du Génie Rural à 4 paramètres Journalier—Edijatno *et al.*, 1999) has no explicit counterpart for soil moisture measurements.

General outlines of the composite procedure

Soil moisture values that are taken into account in equation (2) are first estimated, when the GRHUM model is used, with an equation similar to equation (1):

$$Ho_j = \gamma + \delta Hc_j + \eta \tag{3}$$

where η is the error term with a standard deviation equal to σ. When the GR4J model is used, observed soil moisture values are correlated with a store content value, for example, with the content (S) of the soil reservoir, if it is the best variable correlated to observed soil moisture. The determination of such a regression equation is restricted to the days j when $C_{A,j}$ is above a threshold equal to two times the mean annual streamflow. This threshold is lower than the one used for equation (1) for two reasons: firstly, due to the short period of observations, lowering the threshold allows one to have a larger sample; secondly, the correlation between streamflow and soil moisture is not obvious and it would be wiser to widen the sample of soil moisture data.

With the new variable Hc, one may outline the new procedure. At time step j, shifts concerning all parameters are carried out successively and a new set is retained, until the next iteration, under two conditions:

(a) the new set of parameters reduces the expression:

$$|Q_j - C_{B,j}| \tag{4}$$

(b) the new set of parameters satisfies the condition described by constraint (2). If not satisfied, this condition suggests to reject the corresponding parameter set. Although this parameter set could reduce the last error in streamflow, it is considered unsuitable because of the too large difference between the simulated and observed soil moisture.

This second item is the only change from the method described in Yang & Michel (2000). Increasing k from an insensitive low value to a larger one allows for a progressive enforcing of the condition (2) up to a value that would make the most from this new type of data. If, in the process, the method performance is not improved, this would be an indication that the additional information is not relevant for its use in flood forecasting.

ASSESSMENT OF THE COMPOSITE PROCEDURE

First, one needs a criterion to decide what is a flood that deserves a forecasting procedure. It has been decided to treat all events during which the flow rate keeps in excess of a threshold equal to four times the mean annual flow (not to be confused with the threshold used to calibrate equation (3)). Even when the crossing above the threshold happens for only one time step, the forecasting operation is carried out for this step and the result is tallied together with those from all other events. Both models used in association with the forecasting procedure were calibrated against the longest series of data available, thus yielding the baseline parameter set and, after a sensitivity analysis, the standard deviation of each parameter. The parameters of the model and the correction parameters (α, β, γ, δ...) are calibrated against data from the whole period of available data.

For each model, for each day a flow rate is forecast, the squared error of the forecast flow rate is recorded and summed up to form the persistence criterion (Wallis & Todini, 1975):

$$G = 100\left[1 - \frac{\sum_j \left(Q_{j+L} - C_{B,j+L}\right)^2}{\sum_j \left(Q_{j+L} - Q_j\right)^2}\right] \tag{5}$$

where Q_{j+L} is the observed flow rate at time $j+L$, $C_{B,j+L}$ is the forecast flow rate for time $j+L$ made at time j (L is the lead time), and the summation is done on all values of j such that either Q_{j+L} or $C_{A,j+L}$ (where C_A is the flow rate computed using the baseline set of parameters plus adjustment provided by equation (1)) is larger than the previously mentioned flood threshold equal to four times the mean annual flow rate.

As to the rainfall rates of the forecast period, perfect foreknowledge was adopted, i.e. actual rainfalls were used. Clearly, this is not an operational option, but it was adopted because it was deemed essential to best assess the effect of introducing soil moisture measurements without blurring the picture by additional sources of error.

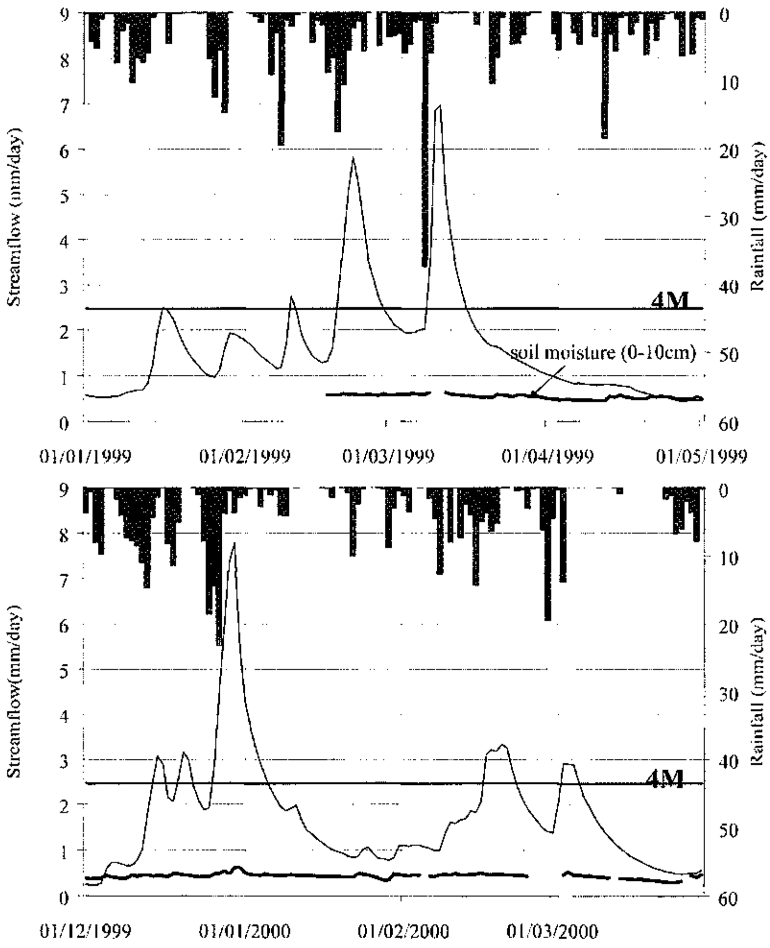

Fig. 2 Rainfall, runoff and soil moisture for two periods encompassing the nine flood events (Serein basin, France).

Table 1 Results for equations (1) and (3).

Model	Regression equation	Determination coefficient	St. deviation
GR4J	$Q_j = 0.587 + 0.841 \times C_j$	$R^2 = 0.53$	$\sigma = 0.835$
GRHUM	$Q_j = 0.581 + 0.849 \times C_j$	$R^2 = 0.53$	$\sigma = 0.807$
GR4J	$Ho_j = 0.159 + 0.0017 \times S_j$	$R^2 = 0.10$	$\sigma = 0.057$
GRHUM	$Ho_j = 0.076 + 0.0025 \times Hc_j$	$R^2 = 0.23$	$\sigma = 0.034$

Table 2 Persistence coefficient for the gradual introduction of soil moisture information.

k	GR4	GRHUM
0.2	92.4	93.4
0.33	92.4	91.2
0.5	92.4	91.0
0.67	92.0	91.0
1	91.0	91.0
2	89.5	90.8

FIRST RESULTS

The assimilation methodology has been tested using data from the River Serein at Chablis during the period 1999–2000. The middle part of the 1999–2000 period encompasses about one year of continuous soil moisture data (given by a TDR probe). For this period we observed nine flood events (see Fig. 2). Considering the time-to-peak of the unit hydrographs featured in the GR4J and GRHUM models, the lead time (L) was chosen to be equal to 2 days. The models have been calibrated on the period 1971–1998 (with 1971 as warm-up year), during which raingauges did not change significantly, remaining between eight (1972–1977) and nine (1978–2000). The optimization criterion used with a step-by-step method was the Nash criterion applied to daily streamflows. In calibration mode, the models give a fairly good level of accuracy with a Nash criterion of 87.8% for GR4J and 87.9% for GRHUM.

The procedure requires knowledge of the bias between observed and simulated daily streamflows (equation (1)) and the similar correction applied to simulated soil moisture values and corresponding to equation (3). Regressions have been established using data from the period 1971–1998 (see Table 1). For the establishment of those regressions, the samples used are not identical because points are selected according to a criterion involving C_j values. Therefore, R^2 and σ are not to be compared between the two models.

As explained above, we will progressively introduce information about soil moisture by means of a constraint becoming tighter at each successive test, which corresponds to a k value (equation (2)) becoming higher. Successive trials are summarized in Table 2.

The main result is that assimilation of soil moisture for these data (only nine floods) seems not to be very useful to increase the efficiency of the updating procedure for both models. The introduction of soil moisture information appears to debase the results more quickly when using the GRHUM model.

Because only nine events, totalling 47 days, are involved in the updating process, a definite conclusion could not be reached until more data are retrieved from the catchment site.

CONCLUSION AND PERSPECTIVES

A new parameter updating methodology has been demonstrated to be used in flood forecasting. This method is versatile enough to exploit any other information such as soil moisture. Thus, we proposed a methodology to assimilate soil moisture into a hydrological model. It appears, based on the first results presented above related to the Serein basin, that trying to comply with observed soil moisture impaired the performance of the parameter updating procedure. Due to the paucity of available data, a definite conclusion could not be reached until more data are retrieved from the catchment site. However these first results with *in situ* measurements need also to be validated. This method needs to be applied to other sub-basins and when the entire EO database is available. The use of only single point TDR measurements could also be questioned and the conclusion needs to be confirmed by a more extensive soil moisture survey.

Acknowledgements The framework of the presented study is a three years EU funded project (Contract no. ENV4-CT98-0740 (DG12-ESCY)). The authors would like to thank all the partners of this project (Cemagref, France; CETP/CNRS, France; the University of Valencia, Spain; the Universidade Independente, Portugal; CEH Wallingford, UK; Associaçao de Regantes e Beneficiarios de Silves, Portugal; and Institut Interdépartemental des Barrages Réservoirs du Bassin de la Seine (IIBRBS), France) and especially the Cemagref team for its support and help during the field campaigns.

REFERENCES

Edijatno, Nascimento, N. O., Yang, X., Makhlouf, Z. & Michel, C. (1999) GR3J: a daily watershed model with three free parameters. *Hydrol. Sci. J.* **44**(2), 263–277.

Garrote, L. & Bras, R. J. (1995) A distributed model for real-time flood forecasting using digital elevation models. *J. Hydrol.* **167**, 279–306.

Lamb, R. (1999) Calibration of a conceptual rainfall–runoff model for flood frequency estimation by continuous simulation. *Wat. Resour. Res.* **35**(10), 3103–3114.

Lettenmaier, D. P. & Wood, E. F. (1992) Hydrologic forecasting. Chapter 26 in: *Handbook of Hydrology* (ed. by D. R. Maidment). McGraw-Hill, New York, USA.

Loumagne, C., Chkir, N., Normand, M., Ottlé, C. & Vidal-Madjar, D. (1996) Introduction of soil/vegetation/atmosphere continuum in a conceptual rainfall–runoff model. *Hydrol. Sci. J.* **41**(6), 889–902.

Mein, R. G. & Brown, B. M. (1978) Sensitivity of optimized parameters in watershed models. *Wat. Resour. Res.* **14**(2), 299–303.

O'Connell, P. E. & Clarke, R. T. (1981) Adaptive hydrological forecasting—a review. *Hydrol. Sci. Bull.* **26**(2), 179–205.

Quesney, A., Le Hégarat-Mascle, S., Taconet, O., Vidal-Madjar, D., Wigneron, J. P., Loumagne, C. & Normand, M. (2000) Estimation of watershed soil moisture index from ERS/SAR data. *Remote Sens. Environ.* **72**, 290–303.

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

Troutman, B. M. (1985) Errors and parameter estimation in precipitation–runoff models. 1. Theory. *Wat. Resour. Res.* **21**(8), 1195–1213.

Wallis, J. R. & Todini, E. (1975) Comment upon the residual mass curve coefficient. *J. Hydrol.* **24**, 201–205.

Yang, X. & Michel, C. (2000) Flood forecasting with a watershed model: a new method of parameter updating. *Hydrol. Sci. J.* **45**(4), 1–10.

Zealand, C. M., Burn, D. H. & Simonovic, S. P. (1999) Short term streamflow forecasting using artificial neural networks. *J. Hydrol.* **214**, 32–48.