Predictions in Ungauged Basins: Promise and Progress (Proceedings of symposium S7 held during the Seventh IAHS Scientific Assembly at Foz do Iguaçu, Brazil, April 2005). IAHS Publ. 303, 2006.

Assimilation of remotely sensed soil saturation levels in conceptual rainfall–runoff models

PATRICK MATGEN[1], JEAN-BAPTISTE HENRY[2], LUCIEN HOFFMANN[1] & LAURENT PFISTER[1]

1 *Public Research Centre, Gabriel Lippmann, 41 rue du Brill, L-4422 Belvaux, Luxembourg*
matgen@lippmann.lu

2 *Service Régional de Traitement d'Image et de Télédétection, Bd. Sébastien Brant, BP 10413, F-67412 Illkirch Cedex, France*

Abstract Owing to the nonlinearity of the rainfall–infiltration–runoff relationship, soil water content in the river basin represents a key environmental variable to be monitored for flood management purposes. In this study an attempt was made to sequentially assimilate into a simple lumped conceptual rainfall–runoff model an estimate of the soil saturation level. The estimate was obtained from: (a) field measurements of water table depth; and (b) backscattering of the radar signal emitted by active microwave sensors on board ERS-1. The assimilation scheme is based on an extended Kalman filter as both simulated and observed soil saturation states are prone to errors. The magnitude of the internal state updating thus depends on the ratio of errors on the observations and the model. The analysis of a series of ERS-1 SAR images showed that hydrologically relevant information could be retrieved from radar imagery by averaging the backscattering coefficient over clusters of pixels for which the sensitivity towards changing moisture conditions is significant. The assimilation procedure is performed on the experimental Alzette River basin (1175 km^2). Improvements of model performance through data assimilation demonstrate the usefulness of field measurements and remote sensing observations in flood forecasting applications.

Key words data assimilation; flood forecasting; Kalman filter; Synthetic Aperture Radar

INTRODUCTION

There is a growing interest in the capabilities of Earth Observation (EO) data for improving the effectiveness of operational flood forecasting systems. Thus, both rainfall–runoff and flood propagation models take benefit from the availability of spatially distributed EO data, especially in ungauged basins. Among other applications, remote sensing observations can also be used as parametric input data, as initial condition data and as time-varying hydrological state and flux data (Walker, 2005). However, there is still significant uncertainty over the reliability of spaceborne microwave sensors such as Synthetic Aperture Radar (SAR) to provide accurate soil moisture data (Schmugge *et al.*, 2002). In a recent OECD report on the opportunities and challenges facing the use of space applications in flood management (OECD, 2005), it is stated that the measuring of soil moisture by spaceborne sensors may not be available on an acceptable operational basis for another 10 years. Wagner & Pathe (2005) also suggested that in order to advance soil moisture retrieval at the field scale, new field experiments with novel technologies are needed. At the basin scale, significant improvements in reliability and repeatability of data extraction may eventually rely on passive microwave techniques. The apparent failure of SAR to provide accurate soil moisture values can be explained by the fact that the system parameters (wavelength, polarization, incidence angle) of the currently available spaceborne active microwave sensors are by no means ideal for soil moisture retrieval. The apparent lack of suitable remote sensing methods for soil moisture estimation represents a serious data gap in flood forecasting systems because the monitoring of this key environmental variable would help assessing the readiness of a river basin to generate storm runoff during rainfall events. On a more positive note, previous studies (Houser *et al.*, 1998; Francois *et al.*, 2003) have shown that, despite the aforementioned limitations of available microwave techniques, average soil moisture observations at basin scale could still be successfully used to improve hydrological model-based discharge predictions. This has been achieved through the integration of remotely sensed soil moisture information in a variety of rainfall–runoff models, ranging from very simple lumped conceptual (Aubert *et al.*, 2003), to spatially distributed physically-based models (Pauwels *et al.*, 2001). The scope of this study is therefore to use remote sensing observations and ground-based measurements of water table depth to retrieve representative estimates of the basin averaged saturation degree and to integrate the observed soil saturation indices of soil saturation into a conceptual rainfall–runoff model in an attempt to enhance the reliability of flood forecasts of the Alzette River at Ettelbrück station (Grand-Duchy of Luxembourg, Europe).

STUDY AREA AND AVAILABLE DATA

At its outlet in Ettelbrück, the Alzette River basin has a drainage area of 1175 km^2. Hourly discharge data are available from 1996 to the present. These were used to calibrate the rainfall–runoff model. Also, the daily streamflow data for the period 1993–1995 were used to assess the performance of the integration scheme. The basin averaged rainfall was based on the hourly rainfall data of 12 raingauges located within the drainage basin. Potential evapotranspiration was determined with the Penman-Monteith equation using the daily meteorological data from the synoptic station at Luxembourg airport. Also, the meteorological data are necessary to guarantee that on the days of satellite overpass the radar signal return is not influenced by frozen soils or high wind velocities on flooded areas, lakes or ponds. Based on the water table depth data at 10 piezometric stations scattered throughout the basin's alluvial plain, the basin wetness was estimated using the soil saturation index. The index is explained in more detail hereafter. The EO database comprises 13 ERS-1 and ERS-2 images, acquired on descending pass, with nine of them during the ERS-1 *Ice Phase*, from 20 November 1993 until 23 February 1994. During this phase the usual repeat cycle of 35 days was shortened and ERS-1 operated with a repeat cycle of 3 days. The high revisit frequency represents an idealized scenario within which radar images are readily made available for observing environmental variables from the space at a fine temporal resolution level.

ESTIMATION OF THE BASIN SATURATION STATE

The sequential assimilation of observations represents a possible step towards improving the functional accuracy of hydrological models. In the present study two data sets were assimilated into the rainfall–runoff model in order to control the hydrological model. These data sets are a flood plain based soil saturation index obtained from the piezometric measurements, and a radar backscattering coefficient averaged over selected pixels. In lumped models the integrated data sets need to be representative of the entire catchment at a given time.

Field data

The time variation of water content in the first few centimetres of the basin's soil is loosely connected to the time variation of the water budget over the entire basin. Fluctuations of the water table depth in the flood plain are more representative in terms of time variation of the basin hydraulic state (Pfister *et al.*, 2003) because the water table levels reflect the behaviour of a characteristic fraction of the basin. However, unlike more aggregate components such as river discharge, individual point measurements of water table depth do not fully account for the basin saturation state. A spatial averaging procedure is therefore required for a thorough basin scale estimation. Based on the recorded minimum and maximum water table depth at available piezometers during the time period 1993–1994, a regional mean soil saturation index (SSI) is computed (Matgen *et al.*, 2005b). When the water table reaches its minimum depth during the January 1994 flood, the SSI is 100% i.e. it is assumed that the soil is completely saturated. With the alluvial groundwater depletion during dry weather phases, the SSI decreases linearly until the measured water table depth reaches its maximum value (SSI = 0%). The simultaneous evolution of the computed water budget at a catchment scale and the mean SSI, as well as the strong correlation between storm runoff coefficients and SSI for recorded rainfall–runoff events, support the assumption that the estimation of a flood plain based SSI provides valuable information on the expected runoff generation during storm events.

Remote sensing data

Matgen *et al.* (2005b) developed a methodology based on Principal Components Analysis (PCA) and k-means clustering (Ramos, 2001) to convert radar backscattering signal into hydrologically relevant information (i.e. the regional flood plain based SSI). Because runoff generation is strongly controlled by deeper layers, especially in regions with a temperate oceanic climate, the processing of the ERS SAR scenes is limited to the flood plain image subset. This is due to the fact that in the shallow groundwater area there exists a strong bond between the water table depth and the water content in the first few centimetres of the upper soil layer (Chen & Hu, 2004). To separate the soil moisture contribution to the backscattering signal from the influence of other physical factors, it is important to process a series of images with homogeneous topography and stable land use. Within the study area the topography is characterized by small elevation changes with a permanent low-height pasture dominating homogeneous land use. Hence, during the winter months, the first Principal Components are dominated by the variance related to areas with changing back-scattering responses and are thus controlled by subsurface and surface water dynamics. The methodology allows outlining pixel clusters (50 × 50 m microplots) that show a backscattering behaviour that is related to moisture conditions. Also, instead of using the subjectively delimited soil parcels or calculating mean backscattering coefficients over large areas, the proposed segmentation approach automatically provides distinctive hydrological response units. By averaging the radar signal over the pixels of categorized classes, the robustness of the empirical linear relationships between radar backscattering and ground-based estimates of soil saturation is being strengthened. This methodology is similar to the one proposed by Quesney *et al.* (2000) who estimated the soil moisture at basin scale by averaging the radar signal only over "sensitive targets", i.e. pixels with significant sensitivity to surface soil moisture.

ASSIMILATION PROCEDURE

Rainfall–runoff Model

The hydrological model used in this study is an 11-parameter lumped conceptual model (Fig. 1), which simulates hourly discharge using rainfall (*Rtot*) and potential evapotranspiration (*ETP*) as input data. The conceptual model can be considered as a modified version of the HBV-96 model (Lindström *et al.*, 1997). The conceptual model structure of the soil reservoir was adapted so as to allow for the integration of observed soil saturation state variables. The soil reservoir module is characterized by the following parameters namely maximal storage capacity, *Smax* [mm], a parameter of nonlinearity, *b* [–], controlling the infiltration capacity, a parameter *lp* [–], giving the fraction of *Smax* below which ETP is constrained by $S(t)$ and the maximum percolation rate, *p* [mm day^{-1}]. *Center* and *alp* are shape parameters that describe the percolation function. The modified soil reservoir module is drained by evaporation and deep percolation fluxes. The rainfall is divided into two terms; the first part fills the soil reservoir module, while the second part consists of the net rainfall that fills the two routing reservoirs (linear baseflow reservoir and nonlinear fast runoff reservoir). The model output largely depends on the basin saturation state since the fraction of the total rainfall that fills the soil reservoir module and the fraction of net rainfall that fills the fast runoff reservoir module both depend on the level of the soil reservoir. The proposed assimilation scheme can be implemented in a large range of lumped conceptual models. The objective function that was used to calibrate the model is the Nash-Sutcliffe criterion. The SCE-UA algorithm (Duan *et al.*, 1992) was adopted to find the set of parameter that gives the highest performance measure for the time period 1996–1998 of hourly streamflow measurements. The Nash-Sutcliffe criterion calculated on the streamflows is 93.2% for this simulation. Figure 2 shows the time variation of the level of the soil reservoir as well as the agreement of fit between simulated and observed discharge.

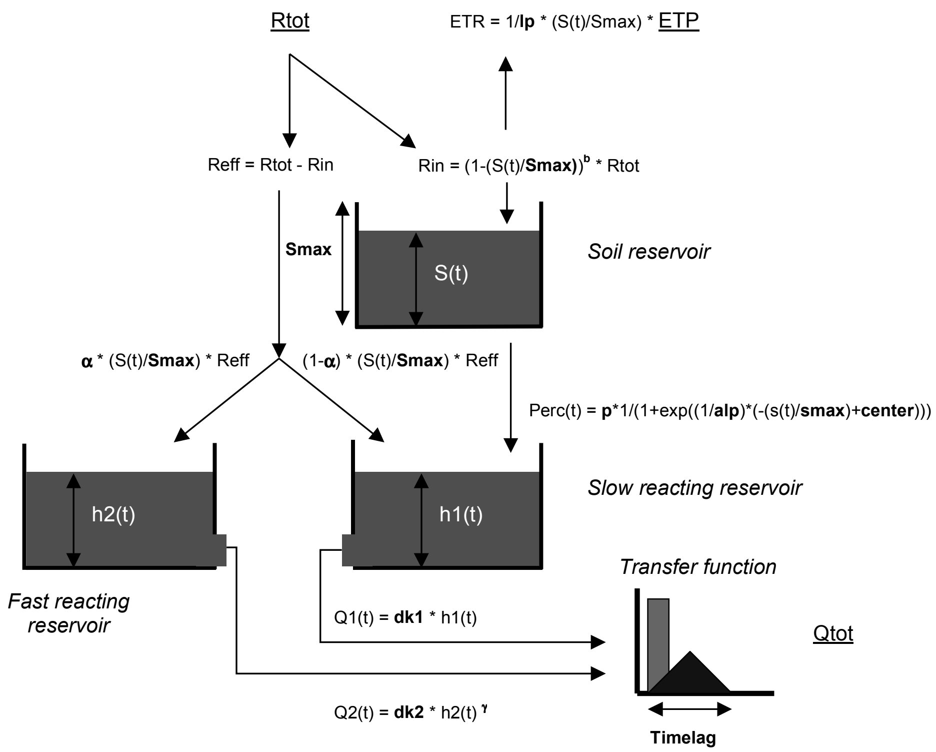

Fig. 1 Schematic representation of the lumped conceptual rainfall–runoff model.

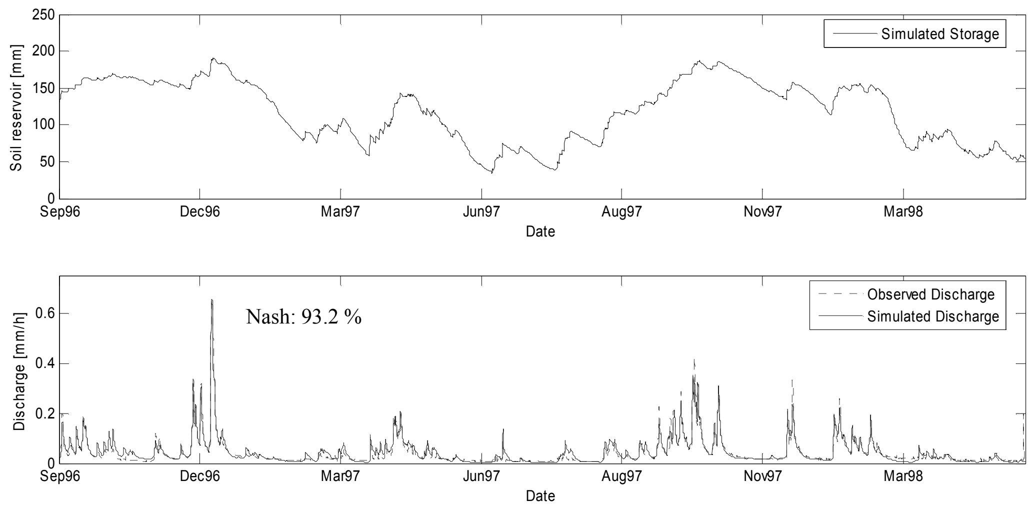

Fig 2 Simulation of the soil saturation level and discharge over the calibration period September 1996—April 1998.

Sequential assimilation

The state of the model that represents the storage water in the soil reservoir (i.e. $S(t)$) can be updated with field measurements and remote sensing observations. The method is based on the assumption that a better simulation of the model states at day j will improve the accuracy of the model states at days $j + 1$, $j + 2$, etc. (Aubert *et al.*, 2003). The model gives a state estimate with high temporal resolution, but the values are altered by the accumulation of errors. The field measurements and remote sensing observations give alternative estimates that may be more accurate, but due to commensurability errors it seems impossible to accurately estimate how representative these estimates of the basin-averaged saturation level are. Thus, the representation error is typically the main error source that needs to be considered (Sorensen & Madsen, 2004). After each simulation step, the difference between observed and simulated values is partially corrected. Therefore, the uncertainties of the observations and simulations are taken into account by calculating a correction factor using the extended Kalman filter (Francois *et al.*, 2003). The magnitude of the correction depends on the ratio between observation and model errors. Hence, the best estimate of the exact saturation state is based on the available information from the two sources of information, namely the modelled and the observed saturation levels. Thus, the ratio (Qr) of the model variance over the observation variance is considered as an additional calibration parameter due to the fact that the data at hand does not allow the fixing of observation and modelling errors *a priori*. A control simulation without any data assimilation can be considered as the baseline run i.e. it is assumed that the real-time observations contain no valuable information. On the contrary, if a direct integration takes place ("hard updating"), it is assumed that the model contains no information at all. Besides improving the reliability of the model forecast through a better estimation of the antecedent moisture conditions, the sequential assimilation allows for the increase of the internal consistency of the conceptual rainfall–runoff model. Pauwels *et al.* (2001) stated that one of the main reasons for adopting data assimilation techniques is to reduce the need for model calibration and to reduce the effect of uncertainty in the model output as a result of the equifinality of different sets of parameters. Francois *et al.* (2003) demonstrated that the sequential assimilation of SAR data could also correct for some errors in the input data (precipitation and evapotranspiration).

Observation models

The aforementioned interpretation of the SSI and the radar backscattering coefficient suggests that these observations are related to the level of the soil reservoir of the conceptual model. As these relationships are not direct, the first step in data assimilation consists of the establishment of the empirical relationship between the level of the soil reservoir and both the measured SSI value and each hydrological response unit's average backscattering signal. A linear relationship is chosen in order to relate the SSI to the model state whenever a measurement is available (Matgen *et al.*, 2005a). Linear relationships are also found between the simulated saturation level of the soil reservoir and the mean backscattering of the three hydrological response classes that proved to be significantly related to the SSI (Fig. 3). Linear regression is acceptable as shown by the coefficients of determination ranging between 0.83 and 0.87. In order to separate the dielectric effect (related to soil moisture) from the specular effect (related to the specular backscattering on the water surface), the

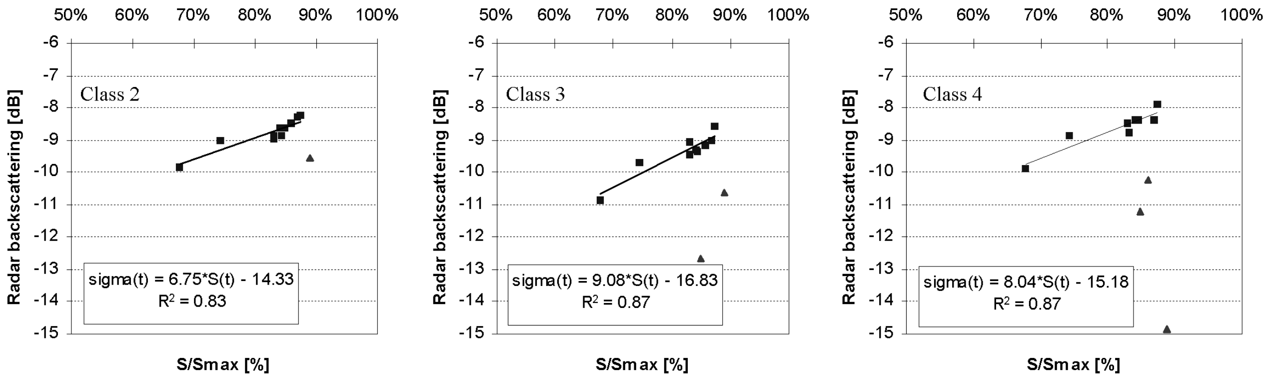

Fig. 3 Relationship between the state of the soil reservoir (S) and each hydrological response unit's mean radar backscattering coefficient.

radar backscattering values from SAR images showing large scale flooding within the study area, were excluded from the regression analysis.

RESULTS AND DISCUSSION

In order to measure the efficiency of the assimilation procedure, the simulated discharges of the Alzette River are compared to the observed discharges with a confidence level in the model compared to the observations varying from $Qr = 10^{-6}$ (no assimilation) to $Qr = 10^{6}$ (forced mode). During model evaluation, two types of performance measures are considered, namely the Nash-Sutcliffe criterion and the efficiency criterion with a lead time of prediction of 1 h (Aubert *et al.*, 2003). The latter compares the mean accuracy of the 1 h forecasts of the model with and without assimilation. As the model performs reasonably well even without assimilation, the improvement margin of the Nash-Sutcliffe criterion is rather small and the effect of the assimilation is less visible (Aubert *et al.*, 2003). For each source of information, the results of the error ratio giving the highest efficiency is plotted in Fig. 4. It is shown that the updating of the level of the soil reservoir with observations obtained from field measurements and radar imagery allows for the improvement in the high flow simulation exercise. The Nash-Sutcliffe performance measure increases by 1.7% and 1.1%, respectively, and the maximum effectiveness criterion is 7% and 4.5%, respectively. The updating based on the piezometric measurements leads to higher performances than the assimilation of the estimates derived from EO data. This result could be explained by the higher temporal resolution of the piezometric measurements. Hence, during the ERS-1 *Ice Phase*, the integration of estimates of soil saturation slightly increased the performances of the model. Because of the high correlation between the observations and the simulated levels of the soil reservoir, the magnitude of the corrections being undertaken is rather low and no substantial improvement could be achieved. On the whole period, the three soil saturation estimates of the soil saturation are consistent and thus, despite the slight increase of the Nash-Sutcliffe performance measure and the effectiveness criterion, the added value provided by the two sources of observations appears to be not highly significant. Furthermore, it can be noted that the hard updating of the remotely sensed saturation levels provided the best

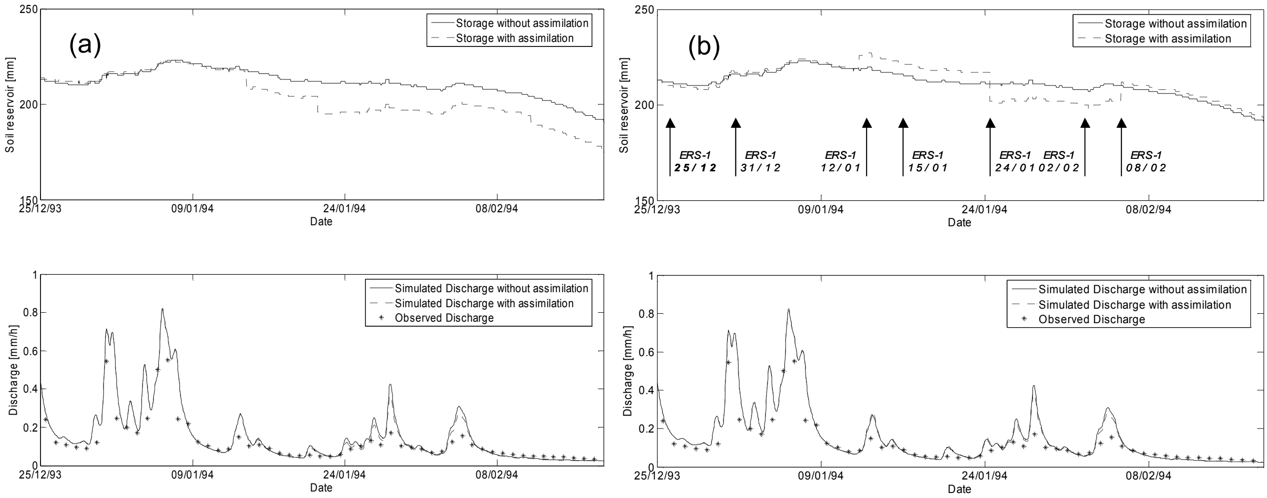

Fig. 4 Simulated soil saturation levels and stream flows with and without data assimilation: (a) assimilation of a flood plain based soil saturation level obtained with piezometric measurements, and (b) assimilation of an average backscattering coefficient obtained over selected pixels with active microwave sensing.

results, whereas some intermediate value of *Qr* gave the best results with the field measurements of SSI. In the forced mode the model is over-corrected and the assimilation of the SSI degrades the model performance (efficiency criterion of –5%). This result suggests that the confidence given to the SSI should not be absolute. The lower confidence level might be related to the limited number of observation points within the study area. Finally, the level of confidence given to the observation data as related to the level of the soil reservoir has to be chosen very carefully since dramatic decreases of the model performance may result from over-corrections of the model.

CONCLUSION

The present study showed that hydrologically relevant information for watershed management was derived from SAR imagery through the use of sensitive targets, i.e. specific hydrological response units with significantly different radar responses under changing moisture conditions. The relevance of the time variation of average backscattering coefficients is further demonstrated by the strong relationship between radar responses and the level of soil reservoir in a previously calibrated hydrological model. Estimates of soil saturation that were obtained with field measurements of water table depth and remote sensing observations were successfully used to improve discharge predictions through data assimilation. However, further improvement in reliability and repeatability of data extraction from SAR imagery is required if it is to be routinely used to update the states or parameters of operational flood forecasting models.

Acknowledgements This study is supported by the Ministère Luxembourgeois de la Culture, de l'Enseignement Supérieur et de la Recherche. In addition, we want to thank Mr Jean-Paul Abadie at the French Space Agency for supporting this research project.

REFERENCES

Aubert, D., Loumagne, C. & Oudin, L. (2003) Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall–runoff model. *J. Hydrol.* **280**, 145–161.

Chen, X. & Hu, Q. (2004) Groundwater influences on soil moisture and surface evaporation. *J. Hydrol.* **297**, 285–300.

Duan, Q., Gupta, H. V. & Sorooshian, S. (1992) Effective and efficient global optimization for conceptual rainfall–runoff models. *Water Resour. Res.* **28**(4), 1015–1031.

Francois, C., Quesney, A. & Ottlé, C. (2003) Sequential assimilation of ERS-1 SAR data into a coupled land surface-hydrological model using an extended Kalman Filter. *J. Hydrometeorology* **4**, 473–487.

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

Lindström, G., Johansson, B., Persson, M., Gardelin, M., & Bergström, S. (1997) Development and test of the distributed HBV-96 hydrological model. *J. Hydrol.* **201**, 272–288.

Matgen, P., El Idrissi, A., Henry, J. B., Tholey, N., Hoffmann, L., De Fraipont, P. & Pfister, L. (2005a) Patterns of remotely sensed flood plain saturation and its use in runoff predictions. *Hydrol. Processes* (in press).

Matgen, P., Henry, J. B., Drogue, G., Fenicia, F., Van den Bos, R., De Fraipont, P., Hoffmann, L. & Pfister, L. (2005b) Towards an increased performance of flood forecasting through assimilation of remotely sensed soil saturation levels in conceptual rainfall–runoff models. In: *Remote Sensing of Environment* (Proc. 31st Int. Symp., 20–24 June 2005, St. Petersburg, Russia). CD-ROM and http://www.isprs.org/publications/related/ISRSE/html/welcome.html

OECD (2005) *Space 2030. Tackling Society's Challenges*. OECD Publications, Paris, France.

Pauwels, V. R. N., Hoeben, R., Verhoest, N. E. C. & De Troch, F. P. (2001) The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation. *J. Hydrol.* **251**, 88–102.

Pfister, L., Drogue, G., El Idrissi, A., Humbert, J., Iffly, J. F., Matgen, P. & Hoffmann, L. (2003) Predicting peak discharge through empirical relationships between rainfall, groundwater level and basin humidity in the Alzette River basin, Grand-Duchy of Luxembourg. *J. Hydrol. & Hydromechanics* **51**, 210–220.

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.

Ramos, M. C. (2001) Divisive and hierarchical clustering techniques to analyse variability of rainfall distribution patterns in a Mediterranean region. *Atmos. Res.* **57**, 123–138.

Schmugge, T. J., Kustas, W. P., Ritchie, J. C., Jackson, T. J. & A. Rango, A. (2002) Remote sensing in hydrology. *Adv. Water Resour.* **25**, 1367–1385.

Sorensen, J. V. T. & Madsen, H. (2004) Data assimilation in hydrodynamic modelling: on the treatment of nonlinearity and bias. *Stochast. Environ. Res.* **18**, 1–17.

Wagner, W. & Pathe, C. (2004) Has SAR failed in soil moisture retrieval? In: *Proc. 2004 ENVISAT & ERS Symposium* (ed. by H. Lacoste & L. Ouwehand L.), (6–10 September 2004, Salzburg, Austria) ESA SP-572, April 2005) CD-ROM and http://earth.esa.int/workshops/salzburg04/proceedings.html.

Walker, J. (2005) Towards remote sensing for hydrologic prediction in ungauged basins. Proceedings of the IAHS General Assembly, Foz do Iguacu, Brazil. CD-ROM.