Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management (Proceedings of Symposium HS2004 at IUGG2007, Perugia, July 2007). IAHS Publ. 313, 2007.

Representativeness of point soil moisture observations, upscaling and assimilation

GABRIËLLE J. M. DE LANNOY[1], VALENTIJN R. N. PAUWELS[1], PAUL R. HOUSER[2], TIMOTHY GISH[3] & NIKO E. C. VERHOEST[1]

1 *Laboratory of Hydrology and Water Management, Ghent University, Coupure links 653, B-9000 Ghent, Belgium*
gabrielle.delannoy@ugent.be

2 *George Mason University & Center for Research on Environment and Water, 4041 Powder Mill Road, Suite 302, Calverton, Maryland 20705-3106, USA*

3 *Hydrology and Remote Sensing Laboratory, USDA-ARS, BARC-West Beltsville, Maryland 20705-2350, USA*

Abstract To estimate the temporal evolution of the spatial mean soil moisture in the Optimizing Production Inputs for Economic and Environmental Enhancement (OPE^3) field, the relationship between point measurements and the average behaviour of field-scale soil moisture has been investigated. In a simple variational assimilation experiment with the Community Land Model (CLM2.0), it has been shown that the soil moisture information from a representative site was much more appropriate for estimating the spatial mean soil moisture profile than the information from other sites. The best results for the re-analysis as well as for the prediction of the spatial mean soil moisture were obtained through the assimilation of observations from probes with time-mean differences between their recorded point values and the spatial mean values close to zero. Further improved results can be obtained by upscaling the point data, e.g. after matching the point observations cumulative density function to that of the spatial mean soil moisture.

Key words soil moisture; stability; assimilation; scaling

INTRODUCTION

Large area hydrological modelling typically requires a coarse grid cell resolution to be computationally feasible. Observational information to calibrate and initialize these models should be available at a similar resolution to be directly useful, or point measurements should at least be representative for the spatial area they are assumed to cover. In general, there is a mismatch in scale between the classical point measurements in field experiments and the spatially averaged estimates from land surface modelling. A similar issue rises for the calibration and validation of remote sensing products by ground point measurements.

In order to obtain spatial information from point measurements, these measurements are typically interpolated. Crow *et al*. (2005) updated spatial mean soil moisture model predictions by a weighted average difference between point observations and local point model predictions and obtained better spatial soil moisture estimates than averaging either point observations only or trusting a spatially averaged model prediction. Grayson & Western (1998) investigated the existence of certain locations in catchments that consistently show the mean areal soil moisture behaviour, to determine areal estimates of soil moisture based on point measurements.

Even when a representative site can be found to estimate the spatial mean soil moisture, there will still be a scale mismatch between the large grid cell model predictions and the point observations. Reichle & Koster (2004) and Drusch *et al.* (2005) proposed observation operators for upscaling the observations before assimilation. De Lannoy *et al.* (2006b) identified representative sites in an intensively instrumented agricultural field (Optimizing Production Inputs for Economic and Environmental Enhancement, OPE^3) near Washington DC and compared different upscaling techniques, both in the time and frequency domain. This contribution is an extension of the latter study, which will be referred to as DL06 in the remainder.

The objective of the study at hand is the variational assimilation of soil moisture point measurements from the OPE^3 site into the Community Land Model (CLM2.0) to obtain an improved OPE^3 spatial mean soil moisture estimate. The relative merits of assimilation of representative site observations *versus* assimilation of observations from any other site were studied.

In the next section, the data and the land surface model used to represent the OPE^3 field are described. Then, some features of the representative sites and the assimilation method are explained. The following section investigates whether assimilation of soil moisture from rank stable sites in a land surface model is advantageous for re-analysis of field-averaged soil moisture, and how this assimilation influences the predictive modelling results. Finally, the conclusions from this study are summarized.

DATA AND MODEL DESCRIPTION

OPE^3 field

The OPE^3 project (http://hydrolab.arsusda.gov/ope3/) is an interdisciplinary research project which was started in 1998 and is managed by the Beltsville Agricultural Research Center (BARC) – Agricultural Research Service (ARS) of the United States Department of Agriculture (USDA). The project is conducted on a corn field of 21 ha, subdivided into four sub-fields, named A, B, C and D from north to south. The site is situated in Prince Georges County, Maryland, USA, and it is part of the Anacostia watershed.

Data

Soil moisture measurements Figure 1 gives an overview of the field and the lay-out of the soil moisture measuring sensors. In each sub-watershed there are 12 capacitance probes (EnviroSCAN, SENTEK Pty Ltd, South Australia). The observations were aggregated to hourly time steps for comparison with model results. H-probes have sensors at 10, 30 and 80 cm. L- and M-probes have sensors at 10, 30, 50, 120, 150 and 180 cm. L-probes have an additional sensor at 80 cm depth. During the study period, 1 May 2001 through 30 April 2002, only 36 of the 48 probes were operational. A detailed analysis of the four-dimensional soil moisture data (Gish *et al.*, 2002; De Lannoy *et al.*, 2006c) revealed a complex subsurface hydrology, mainly caused by an irregular shaped clay layer underlying the top soil layers.

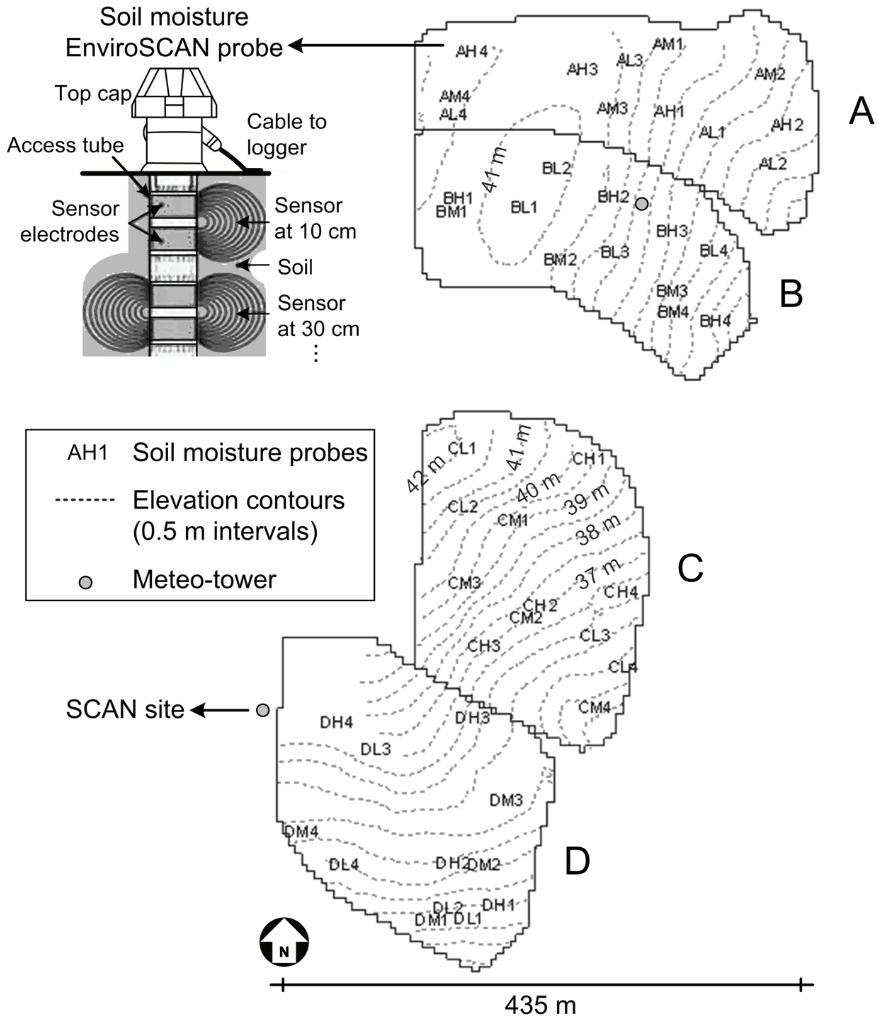

Fig. 1 Digital elevation model and location of the soil moisture probes in the OPE[3] field.

Atmospheric forcings The meteorological data required for the CLM modelling were measured in field B of the OPE[3] corn field. Data from two towers outside the field were used to complete the data set, as discussed by De Lannoy *et al*. (2006a).

Model description

The Community Land Model (CLM2.0) was used as the land surface model to simulate independent soil moisture profiles, without any interaction between cells (Dai *et al*., 2003). CLM2.0 has one vegetation layer, a user-defined number (by default 10) of soil layers, and up to five snow layers (depending on the snow depth). The depths of the different model soil nodes were set to 2.5, 5, 10, 20, 30, 50, 80, 120, 150 and 180 cm. For the model applications in this study, the runs with CLM2.0 were identical to those discussed in De Lannoy *et al*. (2006a) for the purpose of calibration and initialization for the observed individual soil moisture profiles. The focus of the current study, however, is to validate the simulated soil moisture profiles as estimation for the observed spatial mean (instead of point) profile. The estimated soil moisture profiles are now treated as the outcome of a weak contraint variational assimilation scheme, but are essentially identical to the results obtained by De Lannoy *et al.* (2006a) for calibration.

REPRESENTATIVE SITES IN THE OPE^3 FIELD

With $SM_{j,i}$ the soil moisture for sensor j, at each time step i, and $\overline{SM_i}$ the spatially averaged soil moisture, the relative and absolute difference $d_{j,i}$ of the soil moisture content for a sensor are calculated respectively by:

$$d_{j,i} = \frac{SM_{j,i} - \overline{SM}_i}{\overline{SM}_i} \quad \text{and} \quad d_{j,i} = SM_{j,i} - \overline{SM}_i \tag{1}$$

The time averaged difference, $\langle d_j \rangle$, gives an indication of how much the sensor deviates from the spatial mean during the entire time period under consideration. Independent of the kind of difference calculated, a representative site can be identified as one for which $\langle d_j \rangle$ is close to 0. The standard deviation in the (relative/absolute) difference $d_{j,i}$ in time for each a sensor j, $stdv(d_j)$, is of primary importance, as a small standard deviation implies that a sensor shows a similar temporal evolution in soil moisture as the areal soil moisture. A sensor with this feature is called time or rank stable and can be used as representative for the areal soil moisture, if the offset $\langle d_j \rangle$ between the areal soil moisture and the soil moisture at the sensor is known. Of course, the most attractive representative site would be one for which both $\langle d_j \rangle$ and $stdv(d_j)$ are close to 0.

In the study of DL06 it was found that sites which were generally wetter or drier than the average OPE^3 soil moisture during the summer period, mostly kept this property during winter time, but due to the complex geohydrology of the field, it was not possible to link this finding to terrain features. However, the temporal variability in the differences $stdv(d_j)$ for the individual sensors was strongly dependent on the studied period.

ASSIMILATION

To estimate the spatial mean soil moisture of the OPE^3 field, the point observations that could best be used to constrain the model results from the CLM2.0 were studied. Therefore, for all individual probes, the observational information was assimilated such that an optimal state trajectory was found by adjusting the parameters and the initial state for a one-dimensional (profile) model run over a grid cell which covered the entire OPE^3 field. The resulting state estimates for the spatial mean soil moisture profiles were basically a re-analysis product, i.e. the best possible state estimate during some period in the past obtained by a combination of background model results and historical measurements. Also, the capability of the optimally parameterized and initialized model to predict the spatial mean soil moisture was investigated. The 24 hours of observations on 3 May 2001 and all observations from 2 September 2001 to

1 October 2001, were used for assimilation, because during this period no clear evidence of preferential lateral flow was observed (which the model cannot simulate).

To perform the assimilation, the discrepancy between observations and simulations, given their respective uncertainties, was minimized by searching the optimal initial state and model parameters. In analogy with the quadratic cost function J (L2-norm minimization, i.e. minimizing the sum of the squared vector element differences) used for 4D-variational assimilation (Le Dimet & Talagrand, 1986) under the assumption of a perfect model, i.e.:

$$J(\mathrm{x}_0) = (\hat{\mathrm{x}}_0^b - \mathrm{x}_0)^T \mathbf{P}_0^{b^{-1}} (\hat{\mathrm{x}}_0^b - \mathrm{x}_0) + \sum_{i=0}^{N-1} (\mathrm{y}_i - \mathbf{H}_i \mathrm{x}_i)^T \mathbf{R}_i^{-1} (\mathrm{y}_i - \mathbf{H}_i \mathrm{x}_i) \qquad (2)$$

a cost function J_1 for an imperfect model was defined as:

$$J_1(\mathrm{x}, \boldsymbol{\theta}) = 100 \sum_{i=a}^{a+Nic-1} (\mathrm{y}_i - \mathbf{H}_i \mathrm{x}_i)^T (\mathrm{y}_i - \mathbf{H}_i \mathrm{x}_i) + \sum_{i=b}^{b+N-1} (\mathrm{y}_i - \mathbf{H}_i \mathrm{x}_i)^T (\mathrm{y}_i - \mathbf{H}_i \mathrm{x}_i) \qquad (3)$$

which is a weak-constraint objective function, with the vector x_i the state vector, y_i the observation vector and $\mathbf{H}_i$ a linear operator (containing 1- and 0-values only), which maps the state vector to the observation space. The state vector $\mathbf{x}_i$ depends on the parameter vector $\boldsymbol{\theta}$ (including soil physical constants, vegetation parameters, etc.), and the time steps i are discrete hourly time steps.

The first term in equations (2) and (3) refers to the background penalty, which penalizes the deviation of the control vector from some *a priori* information about the initial state. In equation (2), $\hat{\mathbf{x}}_0^{\mathrm{b}}$ stands for the *a priori* initial state estimate or background state. The *a priori* state error or background covariance $\mathbf{P}_0^{\mathrm{b}}$ gives an indication of the uncertainty in the a priori initial state estimate. In equation (2), the control state vector $\mathbf{x}_0$ is optimized for the initial state only, because the model is assumed perfect and $\mathbf{x}_i$ directly depends on $\mathbf{x}_0$. In equation (3), the state vector $\mathbf{x}$ is optimized over all time steps i by optimizing the parameters and the initial state. In equation (3), the values of the observations $y_{l,i} \in \mathbf{y}_i$ at each measurement layer l ($l = 1, \dots, L$) in the profile were used as *a priori* estimated initial state variables at Nic hourly time steps with $Nic = 24$ for 1 day of initial conditions. Time step a is the first hour on 3 May 2001.

The second term in equations (2) and (3) refers to the observation penalty and is calculated for all N time steps for which observations $y_{l,i}$ in vector $\mathbf{y}_i$ are available. $N = 24 \times 30 = 720$ for the 1 month calibration period in September with time step b the first time step on 2 September 2001. The uncertainty of the observations is captured in the measurement error covariance matrix $\mathbf{R}_i$ in equation (2).

Since observations were used as a first guess for the *a priori* state estimate, it was reasonable for the background uncertainty and the observational uncertainty to be equal, i.e. $\mathbf{P}_0^{\mathrm{b}} = \mathbf{R}$. Therefore, this factor was not needed in the minimization of J_1. However, a deviation from the initial background was enlarged by a factor 100, to assure a good initialization. The objective function J_1 is simply an expression of the squared error between the observations $y_{l,i} \in \mathbf{y}_i$ and the model predictions $x_{l,i} \in \mathbf{x}_i$, which could also be written as:

$$J_1(\mathrm{x}, \boldsymbol{\theta}) = \sum_{l=1}^{L} \left\{ 100 \sum_{i=a}^{a+Nic-1} (y_{l,i} - x_{l,i})^2 + \sum_{i=b}^{b+N-1} (y_{l,i} - x_{l,i})^2 \right\} \tag{4}$$

Some additional observation penalties were included to find the optimal state and parameters in a multi-objective framework. These observation penalties are given by:

$$J_2(\mathrm{x}, \boldsymbol{\theta}) = \frac{1}{L} \sum_{l=1}^{L} \left[\sqrt{\frac{1}{N} \sum_{i=b}^{b+N-1} (y_{l,i} - x_{l,i})^2} \right] \tag{5}$$

$$J_3(\mathrm{x}, \boldsymbol{\theta}) = \frac{1}{L} \sum_{l=1}^{L} \left[1 - \frac{\sum_{i=b}^{b+N-1} (y_{l,i} - x_{l,i})^2}{\sum_{i=b}^{b+N-1} (y_{l,i} - < y_l >)^2} \right] \tag{6}$$

$$J_4(\mathrm{x}, \boldsymbol{\theta}) = \frac{1}{L} \sum_{l=1}^{L} \left[\frac{\sum_{i=b}^{b+N-1} (y_{l,i} - < y_l >)(x_{l,i} - < x_l >)}{\sqrt{\sum_{i=b}^{b+N-1} (y_{l,i} - < y_l >)^2 \sum_{i=b}^{b+N-1} (x_{l,i} - < x_l >)^2}} \right] \tag{7}$$

$$J_5(\mathrm{x}, \boldsymbol{\theta}) = \frac{1}{L} \sum_{l=1}^{L} [| < y_l > - < x_l > |] \tag{8}$$

with $< \cdot >$ the time average and $| \cdot |$ the absolute value operator. These measures are expressions for the Root Mean Square Error (*RMSE*), Nash-Sutcliffe criterion (*N-S*), correlation (*R*), and absolute mean difference (*BIAS*), respectively. The consecutive and iterative minimization of the different cost functions were performed by a random search, in which Monte Carlo model runs for a large number of different initial states and parameters were evaluated. The procedure is identical to the multi-objective optimization discussed in De Lannoy *et al.* (2006a). This optimization was performed for each probe individually. Since the three types (H-, M- and L-probes) have different observation layers (L = 3, 6 and 7, respectively), the soil moisture profiles were not estimated with the same amount of constraining information and will therefore be treated separately for validation.

VALIDATION AND DISCUSSION

Re-analysis

To assess the benefit of soil moisture assimilation from representative probes for spatial mean soil moisture re-analysis, the assimilation was validated by the spatially averaged observed soil moisture over the field during the period from 2 September 2001 to 1 October 2001, which covers the period used in the observation penalties. It was expected that the measures of goodness-of-fit for the estimated spatial mean soil moisture would be minimal when representative soil moisture observations were assimilated. Figure 2 shows the profile-integrated validation measures of goodness-of-fit obtained by comparing the model results to the observations from the individual probes and to the observed spatial mean. The measures include the *RMSE*, *N-S*, *R*, *BIAS* or in other words the above formulas of J_2, J_3, J_4 and J_5, respectively. However, to validate the estimated spatial mean, the observational information for individual probes was replaced by the observed spatial mean soil moisture, i.e. $y_{l,i}$ for an

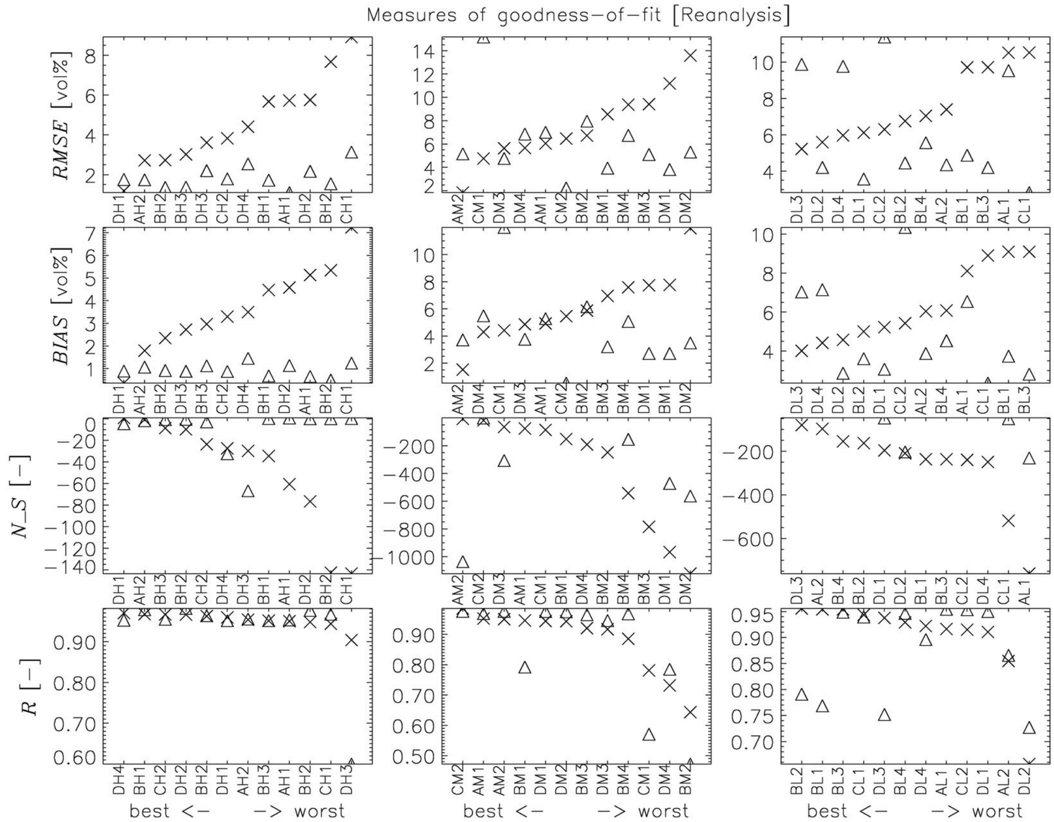

Fig. 2 Sorting of the different probes for their performance to provide useful observational information to estimate the spatial mean during the period 2 September 2001 to 1 October 2001. The × symbols are for the validation to the observed spatial mean, while the Δ symbols are for the validation to the individual probe observations. For the *N-S* criterion, the lowest values for validation of the individual probes are not included for M- and L-probes.

individual sensor is replaced by $\bar{y}_{l,i}$ for the spatial mean soil moisture at depth l. For this study, only the overall profile-integrated performance was studied without detailed analysis for the individual layers. The different probes were sorted according to their performance to provide useful observational information for the spatial mean soil moisture estimate. From Fig. 2 it is clear that the best re-analysis results (best agreement with the observed spatial mean soil moisture) for the *RMSE*, *BIAS* and *N-S* were found for probes DH1, AM2 and DL3 for the H-, M- and L-probes, respectively. These probes had $<d_j>$ values close to zero for most soil layers in the first half year and could be identified as representative sites (DL06). Furthermore, the best results were also obtained for those probes with very low values for $stdv(d_j)$ at all depths. As the assimilated soil moisture values from the individual probes deviated from the spatial mean and as the $stdv(d_j)$ showed high values at some depths, the estimated spatial mean profile performance decreased.

Forecasting

In current hydrological studies, one is mostly interested in the forecasting capability of the model. Therefore, the model forecasts (without any assimilation) were studied after

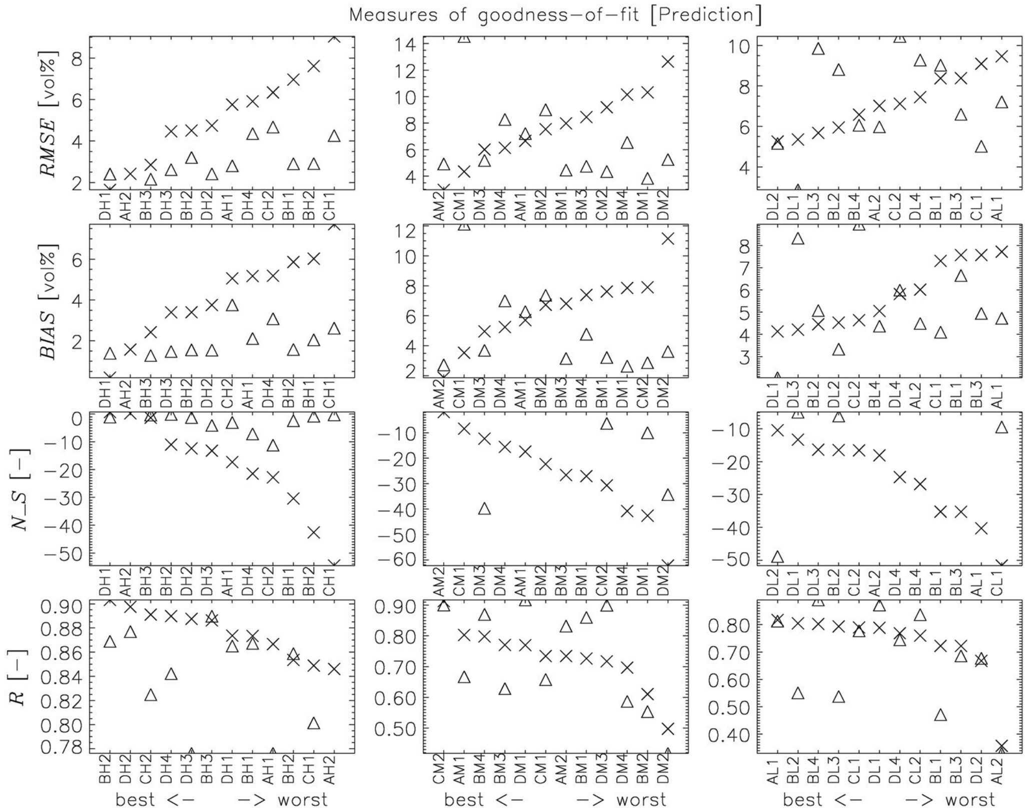

Fig. 3 Similar to Fig. 2, but for the period 2 October 2001 to 30 April 2002.

the period of assimilation in September. Field averaged soil moisture observations from 3 October 2001 to 30 April 2002, were used to validate whether spatially averaged soil moisture was better predicted by a model which was calibrated and initialized (through assimilation) for a representative soil moisture measuring site.

Figure 3 shows that the probes were sorted in a very similar way as in Figure 2, with some exceptions for the L-probes and for the correlation measure *R*. Again the best performance was found for the assimilation of observations from probes with a low $<d_j>$ value, except for *R* and for some probes for which the assimilation was not very successful (i.e. the model was not able to properly represent the point profile observations), but for which the spatial mean corresponded well with the resulting state estimates. The best performances for the prediction of spatial mean soil moisture were registered for those probes which were representative in the second half year (DL06) and a similar order of probes was obtained to that obtained for the re-analysis, because the ranking of the $<d_j>$ values was similar to that in the first half year (DL06).

Upscaling

Through matching the individual point observations cumulative density function (cdf) to that of the spatial mean observations, observation operators were derived to upscale the observations at each single location in the OPE^3 field (DL06). These relationships were used to upscale the individual point measurements and these data were then used for assimilation. As expected, both the re-analysis and forecasting results could be further improved by this upscaling. After cdf-matching it was also possible to use point data that showed a clear bias compared to the spatial mean observations.

CONCLUSIONS

The possible benefit of assimilation of representative sites for the estimation of spatial mean soil moisture through one-dimensional profile modelling over a single grid cell was explored. To this end, separate simulations were studied in which the observations from all different probes individually were assimilated variationally in a weak constraint framework. The initial state background was estimated by observations and the model state estimates were dependent on the estimation of the parameters. For the re-analysis of the spatial mean soil moisture as well as for the prediction, the goodness-of-fit was best after assimilation of observations from probes which had mean differences $<d_j>$ close to zero at most depths. This is logical because the ranking in $<d_j>$ was observed to be similar over different observation periods. This illustrates that the identification of representative sites is not only promising for calibration and validation of remote sensing observations, but also for modelling and more in particular for data assimilation studies. Preliminary results showed that upscaling of the point data to spatial soil moisture estimates, before assimilation into the model runs, further improved the spatial mean soil moisture estimated and allowed use of data from nearly any observation point (after upscaling) to obtain satisfactory results.

Acknowledgements The authors thank the BARC-ARS of the USDA for collecting and providing the data set. The HSB of NASA/GSFC is thanked for hosting the first author during part of the research. The research is supported by a PhD fund from the Bijzonder Onderzoeksfonds (BOF) of Ghent University.

REFERENCES

Crow, W. T., Ryu, D. & Famiglietti, J. S. (2005) Upscaling of field-scale soil moisture measurements using distributed land surface modeling. *Adv. Water Resour.* **28**, 1–14.

Dai, Y., Zeng, X., Dickinson, R., Baker, I., Bonan, G. B., Bosilovich, M. G., Denning, A. S., Dimeyer, P. A., Houser, P. R., Niu, G., Oleson, K. W., Schlosser, C. A. & Yang, Z.-L. (2003). The Common Land Model. *Bull. Am. Met. Soc.* **84**, 1013–1023.

De Lannoy, G. J. M., Houser, P. R., Pauwels, V. R. N. & Verhoest, N. E. C. (2006a) Assessment of model uncertainty for soil moisture through ensemble verification. *J. Geophys. Res.* **111**(D10), D10101.1–18, doi:10.1029/2005JD006367.

De Lannoy, G. J. M., Houser, P. R., Verhoest, N. E. C. & Pauwels, V. R. N. (2006b) Representativeness and upscaling of point soil moisture observations. *J. Hydrol.* (conditionally accepted).

De Lannoy, G. J. M., Verhoest, N. E. C., Houser, P. R., Gish, T. & Van Meirvenne, M. (2006c) Spatial and temporal characteristics of soil moisture in an intensively monitored agricultural field (OPE^3). *J. Hydrol.* **331**(3-4), 719–730.

Drusch, M., Wood, E. F. & Gao, H. (2005) Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture. *J. Geophys. Res.* **32**, L15403.1–4.

Gish, T. J., Dulaney, W. P., Kung, K.-J. S., Daughtry, C., Doolittle, J. A. & Miller, P. T. (2002) Evaluating use of ground-penetrating radar for identifying subsurface flow pathways. *Soil Sci. Soc. Am. J.* **66**(5), 1620–1629.

Grayson, R. B. & Western, A. W. (1998) Towards areal estimation of soil water content from point measurements: time and space stability of mean response. *J. Hydrol.* **207**(1–2), 68–82.

Le Dimet, F. X. & Talagrand, O. (1986) Variational algorithms for analysis and assimilation of meteorological observations. *Tellus A* **38**, 97–110.

Reichle, R. H. & Koster, R. (2004) Bias reduction in short records of satellite soil moisture. *Geophys. Res. Lett.* **31**, L19501.1–L19501.4.

Vachaud, G., De Silans Passerat, A., Balabanis, P. & Vauclin, M. (1985) Temporal stability of spatially measured soil water probability density function. *Soil Sci. Soc. Am. J.* **49**, 822–827.