WATER RESOURCES RESEARCH, VOL. 34, NO. 12, PAGES 3405–3420, DECEMBER 1998

Integration of soil moisture remote sensing and hydrologic modeling using data assimilation

Paul R. Houser,[1,2] W. James Shuttleworth,[1] James S. Famiglietti,[3]
Hoshin V. Gupta,[1] Kamran H. Syed,[4] and David C. Goodrich[4]

Abstract. The feasibility of synthesizing distributed fields of soil moisture by the novel application of four-dimensional data assimilation (4DDA) applied in a hydrological model is explored. Six 160-km^2 push broom microwave radiometer (PBMR) images gathered over the Walnut Gulch experimental watershed in southeast Arizona were assimilated into the Topmodel-based Land-Atmosphere Transfer Scheme (TOPLATS) using several alternative assimilation procedures. Modification of traditional assimilation methods was required to use these high-density PBMR observations. The images were found to contain horizontal correlations that imply length scales of several tens of kilometers, thus allowing information to be advected beyond the area of the image. Information on surface soil moisture also was assimilated into the subsurface using knowledge of the surface-subsurface correlation. Newtonian nudging assimilation procedures are preferable to other techniques because they nearly preserve the observed patterns within the sampled region but also yield plausible patterns in unmeasured regions and allow information to be advected in time.

1. Introduction

Soil moisture is most often described as the water in the root zone that can interact with the atmosphere through evapotranspiration and precipitation. Because soil moisture links the hydrologic cycle and the energy budget of land surfaces by regulating latent heat fluxes, accurate assessment of the spatial and temporal variation of soil moisture is important for the study, understanding, and management of surface biogeophysical processes. Given the crucial role of soil moisture in land surface processes, it should be monitored with the same accuracy and frequency as other important environmental variables. However, because in situ soil moisture measurements are generally expensive and often problematic, no large-area soil moisture networks exist to measure soil moisture at the high frequency, multiple depths, and fine spatial resolution that is required for various applications.

Remote sensing of soil moisture is limited by errors introduced by soil type, landscape roughness, vegetation cover, and inadequate coverage in both space and time. Alternatively, many reliable hydrologic models are available for calculating soil moisture, but these are prone to error in both structure and parameterization. It has been suggested [*Wei*, 1995] that the best, operational soil moisture estimates might be obtained through a synthesis between remote-sensing data and hydrologic modeling. Remote-sensing data, when combined with numerical simulation and other data, should provide estimates of soil moisture with higher spatial and temporal resolution and less error than either remotely sensed data or model simulations separately.

The focus of this research is a prototype system that uses four-dimensional data assimilation (4DDA) methods applied in a macroscale land hydrology model to generate soil moisture fields on regular space and time intervals. The resulting model is a first step toward a new generation of meteorological models that have the capability to assimilate both atmospheric and hydrologic observations for an improved understanding of the interaction between weather and hydrological processes.

2. Methods

2.1. TOPLATS

The Topmodel [*Beven and Kirby*, 1979]–based Land Atmosphere Transfer Scheme (TOPLATS) [*Famiglietti and Wood*, 1994] was used in this study. However, any high-quality spatially distributed soil-vegetation-atmosphere transfer scheme (SVATS) could have been used. The TOPLATS incorporates simple representations of atmospheric forcing, vertical soil moisture transport, plant-controlled transpiration, interception, evaporation, infiltration, surface runoff, and sensible and ground heat fluxes. The subsurface soil column was originally partitioned into root and transmission zone layers in the TOPLATS. However, passive microwave remote sensing is sensitive to moisture only near the soil surface [*Jackson*, 1993], so a third surface soil layer was added for this study (J. S. Famiglietti et al., manuscript in preparation, 1998).

2.2. Data Assimilation Methods

The TOPLATS was modified to allow the assimilation of soil moisture and other state variables. For ease of explanation, the following description assumes assimilation of observed surface soil moisture θ_o. However, with modifications specific to the state variable, the following description can be used to assimilate other variables, such as surface temperature.

After each TOPLATS time step, coincidental observations

[1]Department of Hydrology and Water Resources, University of Arizona, Tucson.
[2]Now at NASA Goddard Space Flight Center, Greenbelt, Maryland.
[3]Department of Geological Sciences, University of Texas at Austin.
[4]Agricultural Research Service, U.S. Department of Agriculture, Tucson, Arizona.

Copyright 1998 by the American Geophysical Union.

Paper number 1998WR900001.
0043-1397/98/1998WR900001$09.00

or observations within a predefined temporal window of the current model time step are sought. It was found that when there is a precipitation event between the time of an observation and the model time, assimilation of that observation is undesirable because of the very different surface conditions caused by the precipitation event.

If the observations fall outside the TOPLATS soil moisture range, defined by the saturated soil moisture θ_s and the residual soil moisture θ_r they are reassigned as follows:

$$\theta_o = \theta_r + 0.1\% \qquad \theta_o \leq \theta_r \tag{1}$$

$$\theta_o = \theta_s - 0.1\% \qquad \theta_o \geq \theta_s \tag{2}$$

It was found that this applied to less than 5% of the PBMR observations. If a large number of observations fall outside the models' allowed range, it is most probably an indication that the models' allowed soil moisture range is incorrect or that there is an observational error.

A control and a direct insertion simulation are used as the basis for evaluating the data assimilation runs. A control simulation (i.e., the simulation without data assimilation) can be considered an extreme case, in which it is assumed that the observations contain no information. The other extreme is direct insertion, where it is assumed that the model contains no information. In this case, the model prediction of surface soil moisture, θ_{sz}, is replaced with a soil moisture observation θ_o whenever an observation is available, i.e.,

$$\theta_{sz} = \theta_o \qquad \text{if } \theta_o \text{ observed} \tag{3}$$

With direct insertion, no data are assimilated outside the region where observations are available; therefore any advection of information is accomplished only via the model physics in subsequent model integrations.

2.3. Statistical Correction Assimilation Method

In the "statistical correction" data assimilation technique, the modeled surface soil moisture mean and standard deviation are adjusted to match the observed mean and standard deviation. The method assumes that the statistics of the observations are perfect, which is arguably more reasonable than assuming that each observation is perfect, as in direct insertion. It also assumes that the patterns predicted by the model are correct but that the predicted surface soil moisture statistics contain bias. As with direct insertion, advection of information into deeper soil layers is accomplished solely through the model physics.

The statistical correction method was implemented as follows. The mean $\bar{\theta}_o$ and standard deviation σ_θ^o of the soil moisture observations and model predictions were determined, and the model's standard deviation σ_θ was adjusted to the observation standard deviation as follows:

$$\theta_i = \frac{\sigma_\theta^o}{\sigma_\theta}\theta_i \qquad \text{for all } i \tag{4}$$

where θ_i was the field of model soil moisture. The mean of the model-adjusted field of soil moisture states was redetermined following the adjustment of the standard deviation, and the mean of the model field, $\bar{\theta}$, was adjusted to match the observation field mean; thus

$$\theta_i = \theta_i - (\bar{\theta} - \bar{\theta}_o) \qquad \text{for all } i \tag{5}$$

2.4. Newtonian Nudging Assimilation Method

Newtonian nudging relaxes the model state toward the observed state in a manner that resembles the process of Newtonian cooling. This relaxation is performed by adding a term to the prognostic equation that is proportional to the difference between the two states. These small forcing terms gradually correct the model fields, which are assumed to remain in approximate equilibrium at each time step [*Stauffer and Seaman*, 1990]. In this way the model can be nudged toward regularly spaced observations, or toward randomly spaced observations during a period of time and space.

2.5. Nudging Towards a Gridded Analysis

A gridded analysis is an interpolation of observation data to the model grid. When this is available, the "nudging to the analysis" method described by *Stauffer and Seaman* [1990] is used:

$$\frac{\partial \theta}{\partial t} = F(\theta, x, t) + G_\theta W_\theta(x, t)\varepsilon_\theta(x)(\theta_o' - \theta) \tag{6}$$

The model's forcing terms are represented by F, θ_o' is the observation at the model grid, and t is time. G_θ is the nudging factor that determines the magnitude of the nudging term relative to all other model processes, while the four-dimensional weighting function W_θ specifies its spatial and temporal variation. The analysis quality factor ε varies between 0 and 1 and is based on the quality and distribution of the observations. Equation (6) is implemented for all three TOPLATS soil layers, with the weighting factor decreasing with depth and time using the four-dimensional weighting function described later. The horizontal weighting function is always unity, because the only observation considered was that lying directly on the model grid that is being nudged.

2.6. Nudging to Individual Observations

When observations are randomly spaced, the technique of "nudging to individual observations," as described by *Stauffer and Seaman* [1990], is used. In this method the difference between the simulated and observed state is determined at the location of the observation then interpolated back onto the model grid. The model's predictive equation is therefore

$$\frac{\partial \theta}{\partial t} = F(\theta, x, t) + G_\theta \frac{\left[\sum_{i=1}^{N} W_i^2(x, t)\gamma_i(\theta_o - \theta')_i\right]}{\sum_{i=1}^{N} W_i(x, t)} \tag{7}$$

where the subscript i denotes the ith observation of N that lies within a preset radius, θ_o is the locally observed value of θ, and θ' is the model's prognostic variable interpolated in three dimensions to the location of the observation. The observational quality factor γ varies from 0 to 1 and accounts for characteristic errors in measurement systems and representativeness. The four-dimensional weighting function (described below) accounts for the separation distance, both spatially and temporally, of the ith observation from a given grid point at a given time. Again, this nudging technique was applied vertically to all model soil layers.

2.7. Temporal and Spatial Weighting Functions

The Newtonian nudging weighting function W at time t and location x for each observation I is a combination of the hor-

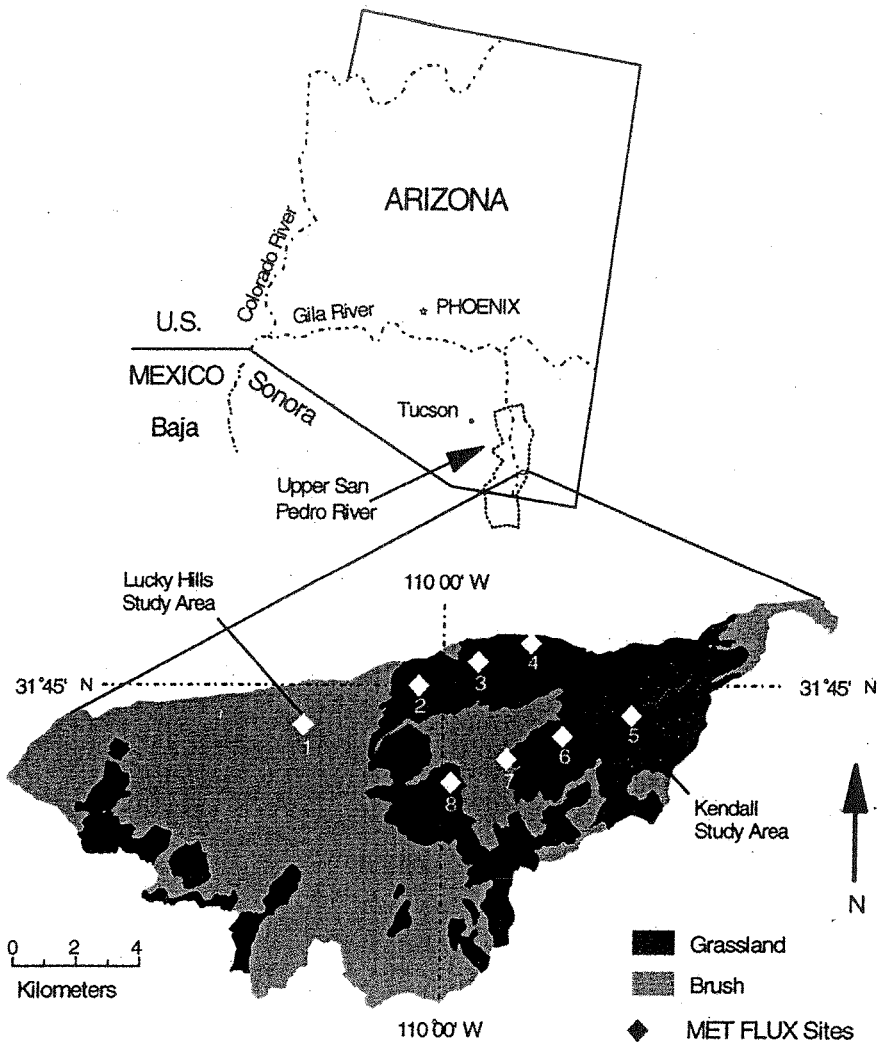

Figure 1. Schematic illustrating the location and boundaries of the USDA-ARS Walnut Gulch Experimental Watershed in southeastern Arizona. Grass- or brush-dominated ecosystems are mapped, and the locations of the Lucky Hills and Kendall sites are shown [*Kustas and Goodrich*, 1994].

izontal weighting function w_{xy}, the vertical weighting function w_z, and the temporal weighting function w_t; thus

$$W(x, t) \equiv w_{xy} w_z w_t \tag{8}$$

For the analysis-nudging technique, where observations are available for each grid in the remotely sensed area, the horizontal weighting function w_{xy} is unity. In the case of the observation-nudging technique, the horizontal weighting function can be defined by a Cressman-type horizontal weighting function as

$$w_{xy} = \frac{R^2 - D^2}{R^2 + D^2}, \qquad 0 \le D \le R \tag{9}$$

$$w_{xy} = 0, \qquad D > R \tag{10}$$

where R is the radius of influence, and D is the distance from the ith observation to the grid point. The vertical weighting function, w_z, is also a distance weighting function, following *Seaman* [1990]; thus

$$w_z = 1 - \frac{|z_{\text{obs}} - z|}{R_z}, \qquad |z_{\text{obs}} - z| \le R_z \tag{11}$$

$$w_z = 0, \qquad |z_{\text{obs}} - z| > R_z \tag{12}$$

where R_z is the vertical radius of influence, and z_{obs} is the vertical position of the ith observation. The temporal weighting function is defined as

$$w_t = 1, \qquad |t - t_o| < \frac{\tau}{4} \tag{13}$$

$$w_t = \frac{(\tau - |t - t_o|)}{\tau/4}, \qquad \frac{\tau}{4} \le |t - t_o| \le \tau \tag{14}$$

$$w_t = 0, \qquad |t - t_o| > \tau \tag{15}$$

where t is the model-relative time, t_0 is the model-relative time of the ith observation, and τ is the half period of a predetermined observation-influencing time window.

Table 1. Spatially and Temporally Constant TOPLATS Parameters for Walnut Gulch

Parameter	Source	Value
	Vegetation Parameters	
Vegetation height, m	*Humes et al.* [1994]	0.22
Leaf area index	*Daughtry et al.* [1991]	1.31
Minimum stomatal resistance, s/m	calibrated (P. R. Houser et al., manuscript in preparation, 1998)	574
Initial water storage in canopy, m	assumed dry at start of simulation	0
Unstressed soil moisture, %	assumed halfway between θ_s and θ_r	20
Wilting point soil moisture, %	assumed θ_r (cacti rarely wilt)	2.2
Vegetation fraction	*Kustas et al.* [1994]	0.42
Albedo, wet vegetation	*Dickinson et al.* [1993]	0.2
Albedo, dry vegetation	*Dickinson et al.* [1993]	0.25
Albedo, bare soil	*Dickinson et al.* [1993]	0.33
Root activity factor	calibrated (P. R. Houser et al., manuscript in preparation, 1998)	348,025
Root density, m/m^3	calibrated (P. R. Houser et al., manuscript in preparation, 1998)	86.5
Root resistivity, s/m	calibrated (P. R. Houser et al., manuscript in preparation, 1998)	4×10^{11}
Critical leaf water potential, m	calibrated (P. R. Houser et al., manuscript in preparation, 1998)	−500
	Soil Parameters	
Surface zone depth, m	*Jackson* [1993]	0.1
Initial surface soil moisture, %	Monsoon '90 Database [*Kustas and Goodrich*, 1994]	10.0
Root zone depth, m	D. J. Breckenfeld (unpublished document, 1993)	0.8
Initial root soil moisture, %	Monsoon '90 database [*Kustas and Goodrich*, 1994]	17.0
Maximum rate of capillary rise, m/s	default [*Famiglietti*, 1992]	0.1
Initial transmission soil moisture, %	Monsoon '90 database [*Kustas and Goodrich*, 1994]	17.0
Sand content, %	calibrated (P. R. Houser et al., manuscript in preparation, 1998]	14.4
Clay content, %	calibrated (P. R. Houser et al., manuscript in preparation, 1998)	8.1
Bulk density, g/cm^3	*Kustas and Goodrich* [1994]	1.6
Residual soil moisture, %	calibrated (P. R. Houser et al., manuscript in preparation, 1998)	2.2
Saturated soil moisture, %	calibrated (P. R. Houser et al., manuscript in preparation, 1998)	30
Saturated hydraulic conductivity, m/s	calibrated (P. R. Houser et al., manuscript in preparation, 1998)	8.7×10^{-6}
Bare soil roughness length, m	assumed	0.001
Bare soil zero plane displacement, m	assumed	0.0
	Topmodel Parameters	
K_s exponential decay parameter	*Famiglietti* [1992]	7.0
Water table depth, m	*Kustas and Goodrich* [1994]	100.0
	Energy Balance Parameters	
Soil moisture for PET calculation, %	calibrated (P. R. Houser et al., manuscript in preparation, 1998)	0.46
Diurnal heat penetration, m	calibrated (P. R. Houser et al., manuscript in preparation, 1998)	0.33
Temperature of deep soil layer, °K	calibrated (P. R. Houser et al., manuscript in preparation, 1998)	287.6

These horizontal weighting functions are well behaved in areas of low observation density, but they degrade when applied in regions of high observation density typical of remote-sensing observations. By adding more information in the form of observations at a distance, this weighting function decreases the total magnitude of the nudging. As information is added with increasing numbers of observations, the weight should increase. The simple solution adopted here was to prevent the total nudging weight in (7) from falling below the largest individual observation weighting function, $W(x, t)$. Information from more distant observations is still used based on its individual weight, but the total weight is set to correspond to that at the closest observation. The modified predictive equation is

$$\frac{\partial \theta}{\partial t} = F(\theta, x, t) + G_\theta W(x, t)_{\max} \frac{\left[\sum_{i=1}^{N} W_i(x, t)\gamma_i(\theta_o - \theta')_i\right]}{\sum_{i=1}^{N} W_i(x, t)} \tag{16}$$

where $W(x, t)_{\max}$ is the maximum weight calculated for any single observation in the set of observations.

2.8. Statistical Interpolation Assimilation Method

Statistical interpolation is a minimum variance method that is closely related to kriging. The technique can be traced back to *Kolmogorov* [1941] and has been in use by most major western meteorological services since the mid-1970s.

Statistical interpolation is implemented in all three TOPLATS soil layers as follows [*Daley*, 1991]:

$$\theta_a(r_i) = \theta_b(r_i) + \sum_{k=1}^{K} W_{ik}[\theta_o(r_k) - \theta_b(r_k)] \tag{17}$$

where K is the number of observation points, W_{ik} is the weight function, $\theta(r)$ is the soil moisture analysis variable, r is the three-dimensional spatial coordinates, $\theta_a(r_i)$ is the analyzed value of θ at the analysis grid point r, $\theta_b(r_i)$ is the background or first-guess value of θ at r_i, and $\theta_o(r_k)$ and $\theta_b(r_K)$ are the observed and background values, respectively, at the observation station r_k. Again, in this study no interpolation is required to obtain the above fields because of the correspondence between the model and observation grids.

The weight function W_{ik} is determined by least squares minimization of (17), with the assumptions that $\theta_b(r_k)$, $\theta_b(r_i)$, and $\theta_o(r_k)$ are unbiased; that there is no correlation between

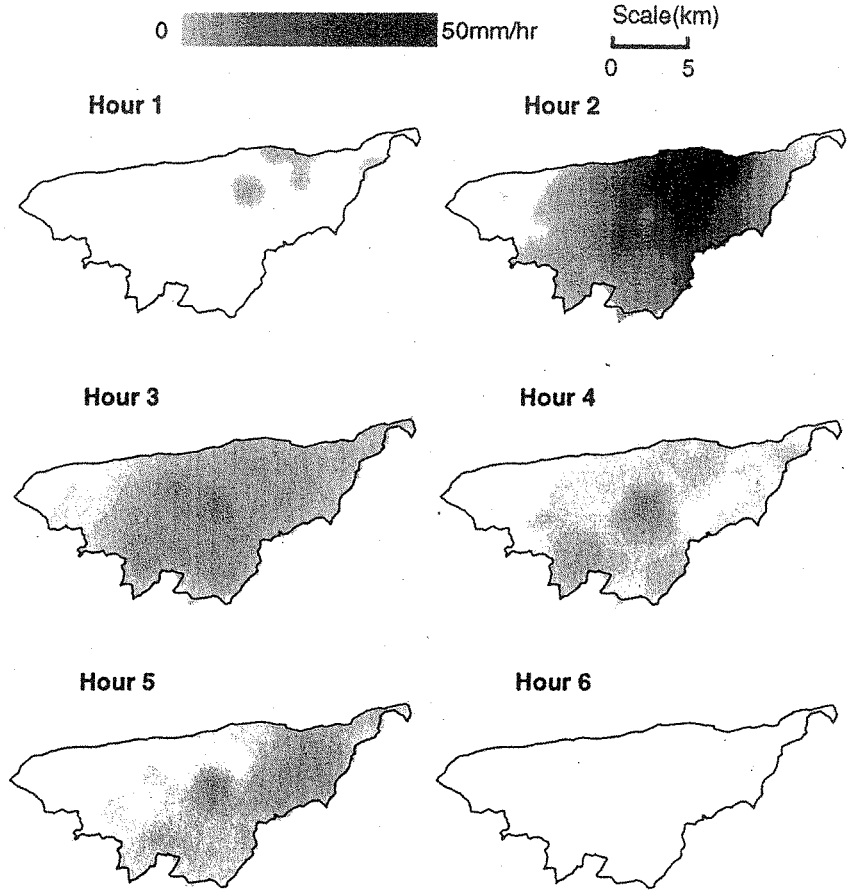

Figure 2. Multiquadric interpolated precipitation for the Walnut Gulch Experimental Watershed for the 6-hour storm of day 213, 1990.

the model and observation error; that the error correlations are homogeneous, isotropic, and time invariant; and that the background error correlation ρ_b is horizontally and vertically separable (i.e., $\rho_b = \rho_{bxy}\rho_{bz}$) [*Daley*, 1991]. Thus

$$\sum_{l=1}^{K} W_{il}[\rho_{bxy}(r_l - r_k) + \varepsilon_o^2\rho_o(r_l - r_k)] = \rho_{bxy}(r_i - r_k)\rho_{bz}(z_i - z_k) \tag{18}$$

where ρ_o is the observation error correlation matrix, ρ_{bxy} is the background horizontal error correlation matrix, and ρ_{bz} is the vertical error correlation matrix. The most satisfactory way of estimating ρ_o and ρ_b is to use observations from a dense homogenous observation network and corresponding model predictions [*Schlatter*, 1975]. Calculations of background error correlation matrices are much more efficient if correlations are specified using symmetric, positive definite correlation models [*Buell*, 1972]. The system of linear equations given in (18) is solved using a simple and efficient Cholesky decomposition. Each PBMR image contains over 35,000 observations, which requires solving a system of 35,000 linear equations for each model grid point, each time an observation was available. Clearly, the computational resources needed for this task are unreasonable; hence a simplified method is required.

Two alternative methods for simplifying this problem through reduction of observations were explored. In the first, a subset of observations is randomly selected. The closest observation which is the observation with the highest information content is always included in this subset, thus approximating the analysis made with all available observations. The second method uses "superobservations," these being average values of observation groups. The model domain is divided into sections, and all observations found in those sections are averaged to define the superobservation. In this study, either 100 random samples or 100 superobservations are used when applying the statistical interpolation method.

2.9. Walnut Gulch–Monsoon '90 Observations

The Walnut Gulch Experimental Watershed (31°43′N, 110°41′W) near Tombstone, Arizona, is operated by the Southwest Watershed Research Center (SWRC), Agriculture Research Service (ARS), U.S. Department of Agriculture (USDA). The catchment is an instrumented area comprising the upper 148 km^2 of the Walnut Gulch drainage basin in an alluvial fan portion of the San Pedro watershed in southeastern Arizona (Figure 1). Eighty-five recording rain gauges, 11 primary watershed runoff-measuring flumes, and micrometeorological observations make the Walnut Gulch Experimental Watershed a valuable research location. During Monsoon '90 (July 23 through August 10, 1990), eight micrometeorological–energy flux (Metflux) instruments provided continuous measurement of local meteorological conditions and the surface energy balance, and extensive remote-sensing observations were made; see *Kustas and Goodrich* [1994] for details.

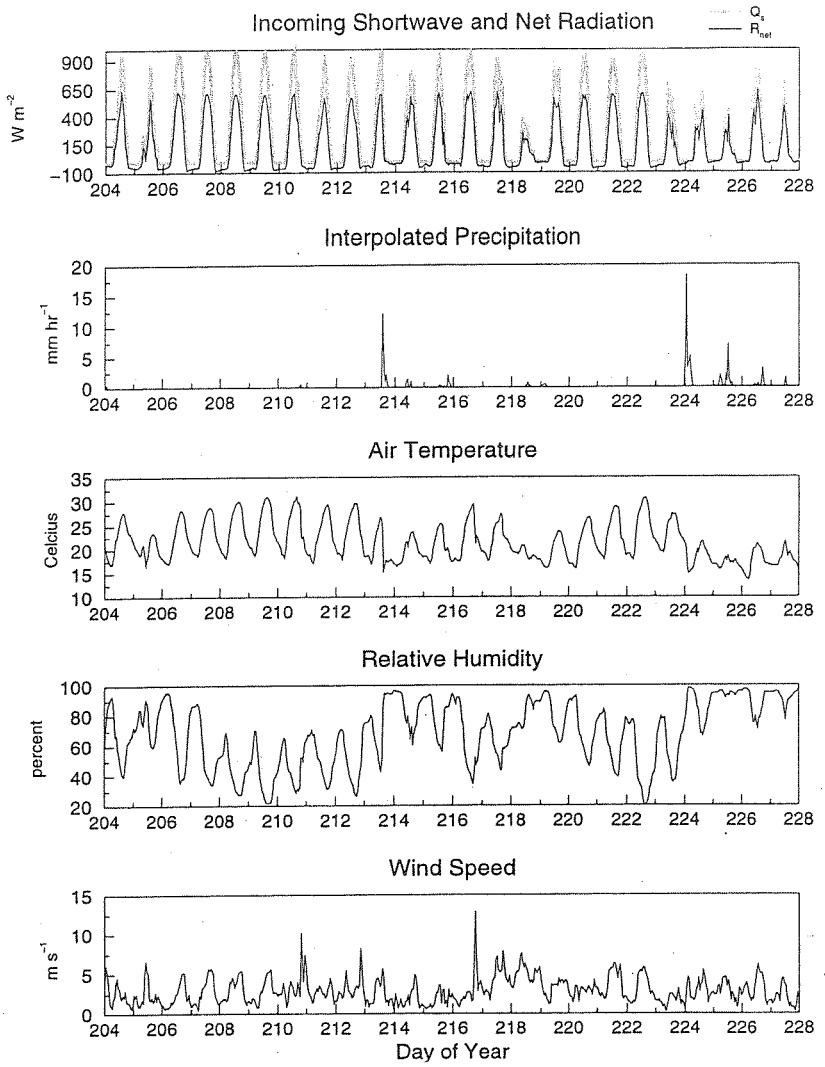

Figure 3. Watershed average Monsoon '90 meteorological observations used as forcing for the TOPLATS model.

2.10. TOPLATS Model Domain

The entire Walnut Gulch Experimental Watershed was modeled in a spatially distributed manner at a 40-m resolution. Predictions were made hourly from July 22 to August 15 (day of year 204–228), 1990. The model was driven with spatially variable topography and precipitation because these have a significant impact on soil moisture. All other soil and vegetation characteristics were spatially constant.

2.11. TOPLATS Parameters

The TOPLATS parameterization was based largely on observations made during Monsoon '90 [*Daughtry et al.*, 1991; *Kustas et al.*, 1994; *Humes et al.*, 1994; *Kustas and Goodrich*, 1994; D. J. Breckenfeld, unpublished document, 1993]. However, many model parameters were not observed and had to be estimated or specified by model calibration (P. R. Houser et al., manuscript in preparation, 1998). A summary of the TOPLATS parameter values used in this study is given in Table 1.

2.12. Temporal Forcing

Precipitation is the most important spatial forcing variable in semi-arid regions owing to its highly variable, convective nature; thus much effort was devoted to deriving spatially distributed precipitation data sets for the Monsoon '90 experiment. A multiquadric-biharmonic interpolation algorithm [*Syed*, 1994] was used to produce spatially distributed precipitation values for the entire model domain from the available rain gauge data. An example of this interpolated precipitation is shown in Figure 2. The multiquadric-biharmonic interpolation method was selected because it proved superior to other common interpolation algorithms [*Syed*, 1994]. All other meteorological forcing (Figure 3) was assumed to be spatially constant and derived from averaging observations at the eight Metflux stations in place during the experiment [*Kustas and Goodrich*, 1994].

2.13. Soil Moisture State Observations for Assimilation

The National Aeronautics and Space Administration (NASA) L band push broom microwave radiometer (PBMR), which operates at a frequency of 1.42 Ghz (21-cm wavelength), was flown on a NASA C-130 aircraft at an altitude of 600 m above the ground. The PBMR has four horizontally polarized beams pointing at ±8° and ±24° from nadir. Each beam has a full width at half maximum power of ~16°, which yields a

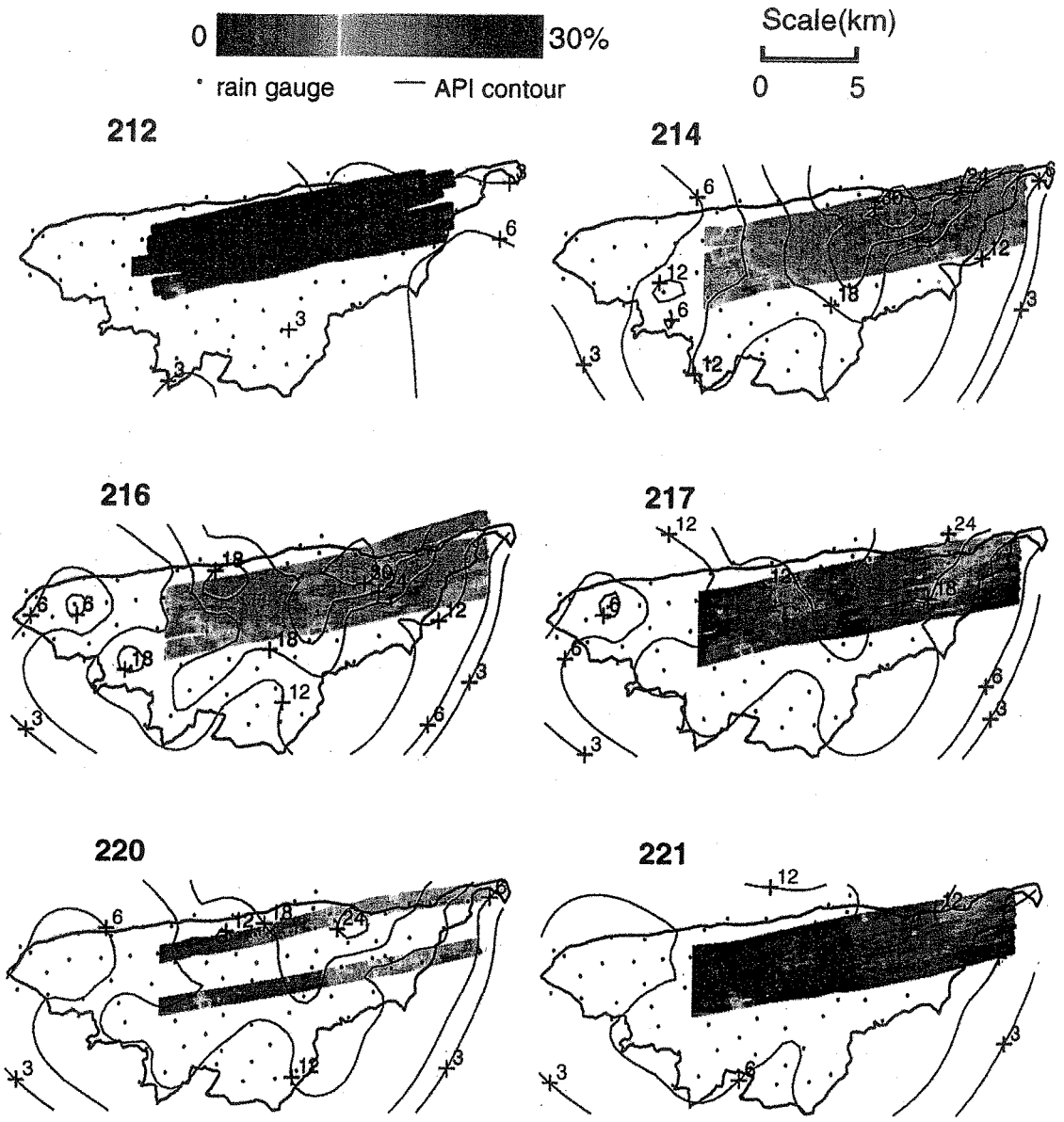

Plate 1. Push broom microwave radiometer derived soil moisture for Monsoon '90. The antecedent precipitation index (API) is also shown, which generally has a good correlation with soil moisture.

180-m instantaneous field of view. The outer beams were corrected for incidence angle effects, and interpolated to a 40-m universal transverse Mercator (UTM) grid [*Schmugge et al.*, 1994].

Six days (days 212, 214, 216, 217, 220, and 221) of microwave brightness temperature data were collected with the PBMR over an 8 × 20 km area in the northeastern portion of the watershed during the Monsoon '90 field campaign [*Schmugge et al.*, 1994]. The period was very dry prior to the first flight (day 212), which was followed by 5 cm of rain falling over most of the study area on day 213. This produced a significant decrease in brightness temperature (50°–60°K) on day 214. The successive flights on days 216, 217, 220, and 221 showed the effects of some smaller rain storms and drydown of the area. The changes in brightness temperature were well correlated with rainfall with a correlation coefficient $R^2 = 0.9$ and in situ soil moisture with $R^2 = 0.8$ [*Schmugge et al.*, 1994]. The linear relationships established between microwave brightness temperature and gravimetric soil moisture at each Metflux site by *Schmugge et al.* [1994] were used with an inverse distance weighting scheme to invert microwave brightness temperature to soil moisture (Plate 1).

2.14. Observations for Model Verification

Some additional observations were available for model validation and calibration in this study. During Monsoon '90, three replicate gravimetric surface soil moisture samples were collected each day at the eight Metflux sites [*Schmugge et al.*, 1994]. Resistance sensors collected continuous time series of soil moisture at 2.5 cm and 5 cm below the surface at all eight Metflux sites [*Amer et al.*, 1994]. Because such sensors are generally difficult to calibrate and tend to drift, they were recalibrated each day against gravimetric measurements for the purpose of this study. Time domain reflectrometry (TDR)

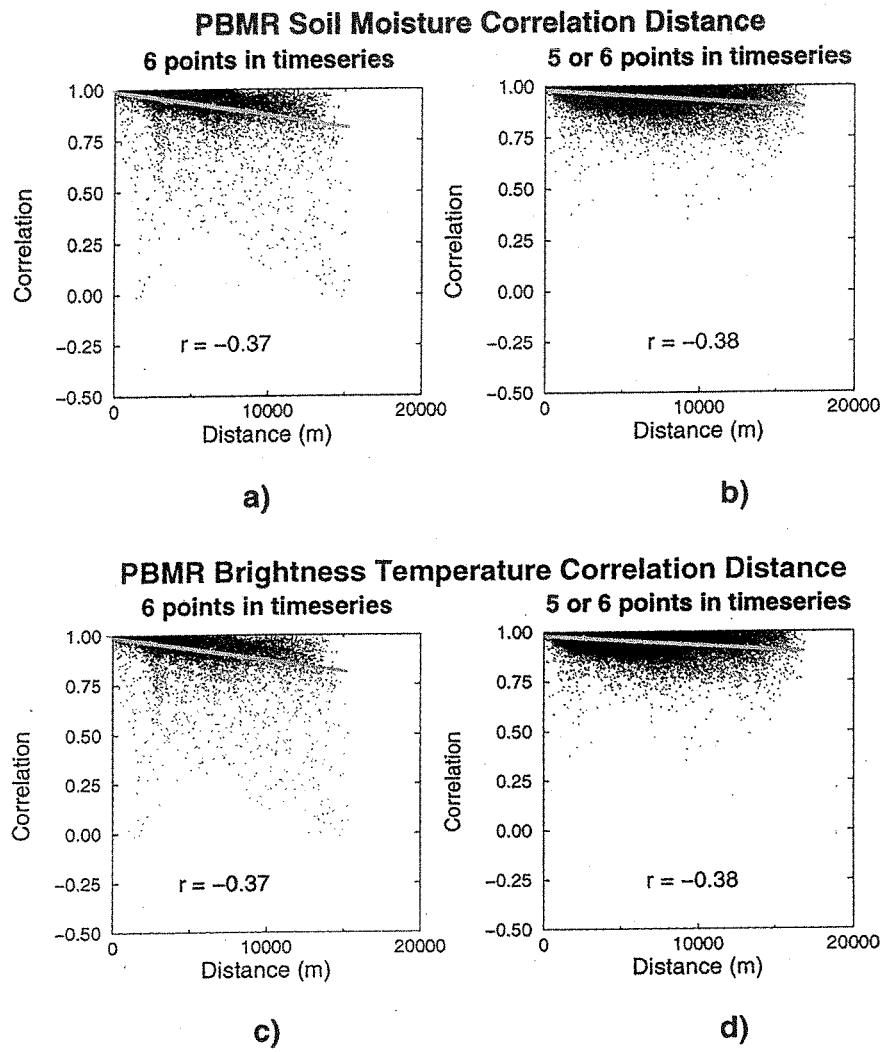

Figure 4. Correlation between time series of PBMR data collected at different locations, plotted versus the distance separating the time series locations. Correlation distances for (a) PBMR soil moisture six-point time series; (b) PBMR soil moisture five- and six-point time series; (c) PBMR brightness temperature six-point time series; and (d) PBMR brightness temperature five- and six-point time series. Data presented are a random selection of 0.01% selection of the total.

measurements were also made at daily intervals and at multiple depths down to 0.5 m at two of the Metflux sites [*Kustas and Goodrich*, 1994].

3. Results

3.1. Correlation Structure Analysis

The ability of data assimilation methodologies 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. The correlation of PBMR brightness temperature and soil moisture time series versus distance is shown in Figure 4. This figure shows that the inversion from brightness temperature to soil moisture has little impact on this correlation, and that longer time series improves the definition of the spatial structure. Figure 4 also shows that the PBMR time series are highly spatially correlated. In fact, it seems that the correlation structure present in these remotely sensed data extends well past the 15-km-long measurement area and can be projected to extend to ~60 km, assuming linear correlation structures. Similar linear spatial correlation structures were observed for the time series of sensible, latent, and soil heat flux. This supports the validity of forcing the TOPLATS model with watershed average meteorological forcing data and of calibrating the TOPLATS model with watershed average flux observations.

A geostatistical analysis of the PBMR-derived soil moisture data showed that their correlation structure varies with time. The first PBMR observations taken during a dry day show a disorganized spatial pattern in the variogram. After the large day 213 precipitation event, the variogram becomes organized; then this organization degrades with soil dry down over the next few days (Figure 5). The variograms shown in Figure 5 are based on instantaneous observations, which are a fundamentally different measure of spatial structure than the time series correlation distances shown in Figure 4. Data assimilation methodologies generally assume that correlation structures are temporally invariant; hence further investigation of the chang-

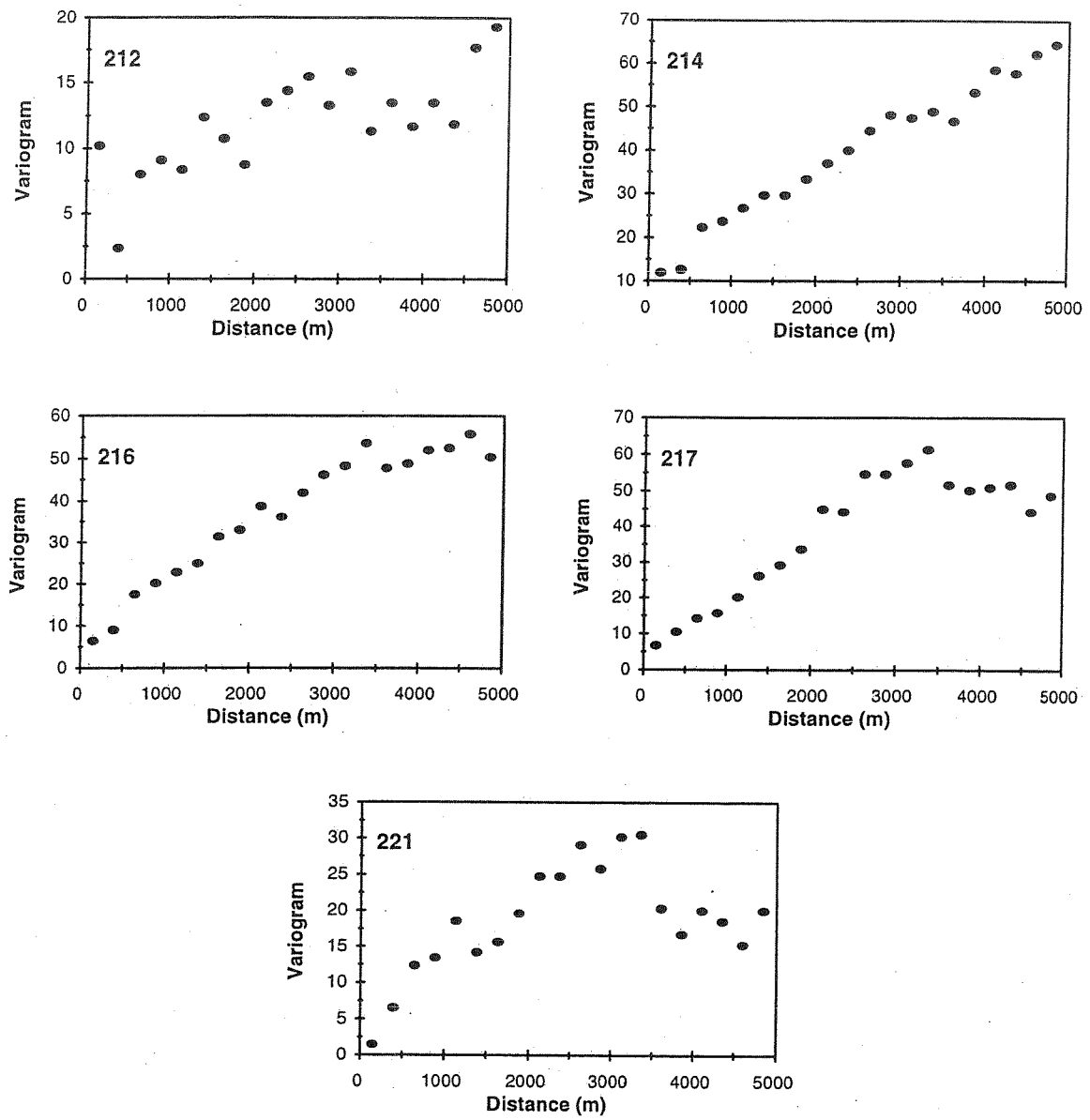

Figure 5. Variograms for a 1% selection of PBMR brightness temperatures.

ing soil moisture correlation structure is warranted when appropriate longer-term soil moisture observations become available.

Implementation of statistical interpolation requires the specification of the background and observation error correlation. The observation error correlation is the horizontal correlation between various observed time series with correlated error, minus observations without correlated error. Usually, for observations taken with separate, stationary instruments, observation errors are uncorrelated [*Daley*, 1991]. For remotely sensed observations measured by the same instrument, however, the observation errors are correlated and cannot be ignored. The time series of in situ gravimetric observations G were used to assess the error correlation in the PBMR soil moisture observations P and the TOPLATS soil moisture control run predictions T_C. An analysis of the correlation of the $(G\text{-}P)$ and $(G\text{-}T_C)$ time series showed no error correlation in the PBMR data or TOPLATS predictions.

The background error correlation (i.e., the distance correlation of the $(P\text{-}T_S)$ time series) shown in Figure 6 is the primary mechanism for advecting observational information into data-sparse regions in the statistical interpolation method. It is clear from this figure that one additional measurement in the time series has a large positive impact on the ability to identify the correlation structure; hence a longer time series would produce a more identifiable structure. Because of the dense nature of remote-sensing observations, the total number of pairs can exceed 600 million for the PBMR-TOPLATS data set; thus a 0.01% random subsample was used in Figure 6.

A standard background error correlation function [*Thiebaux*, 1976]:

$$\rho_b(r) = \left[\cos(cr) + \frac{\sin(cr)}{Lc}\right] \exp\left(-\frac{r}{L}\right) \qquad (19)$$

was fit to the 6-period background horizontal error correlation using a simplex technique to derive the constants c and L. This

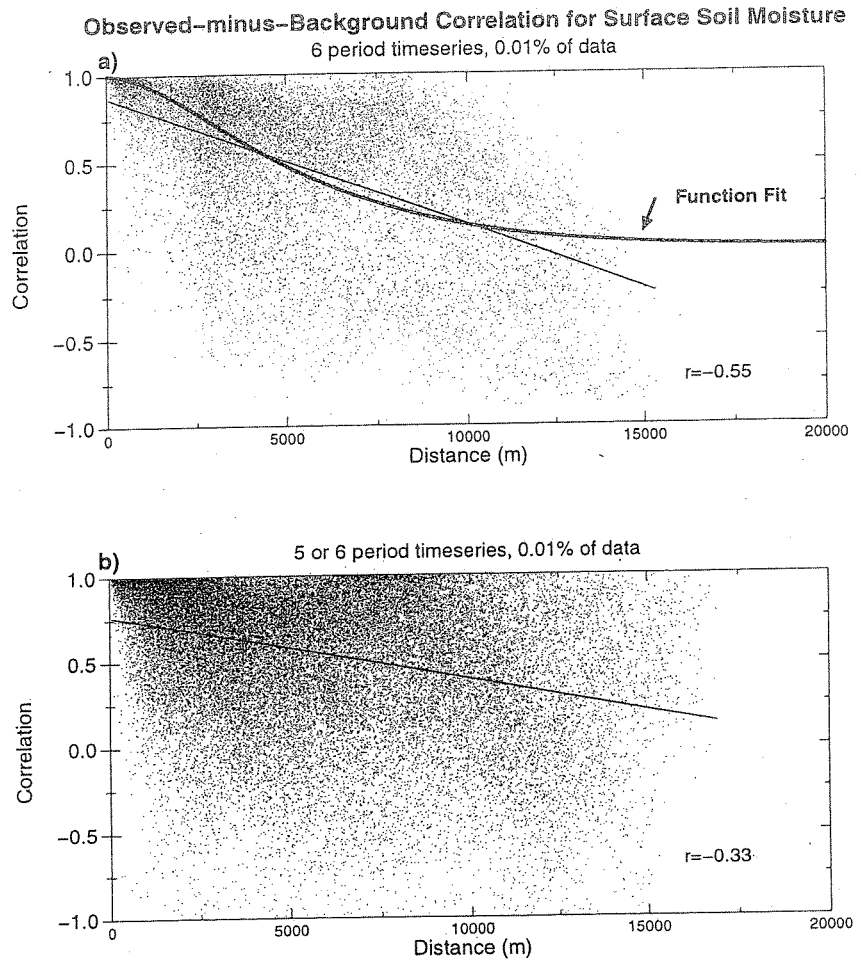

Figure 6. Surface soil moisture background error correlation versus distance for (a) six-point time series, and (b) five-point time series. Plots represent a 0.01% selection of the total data.

fitted curve is shown in Figure 6, and its parameters are given in Table 2.

Statistical interpolation allows for vertical assimilation of data, but to do so requires specification of a vertical, background error correlation function. Because explicit definition of this function was not possible from PBMR observations, in situ profile TDR observations were used as a surrogate to formulate this function for the root zone (Table 3).

Following *Stauffer and Seaman* [1990], most of the Newtonian nudging parameters, such as the analysis and observation quality factors, were set to unity. The parameters that determine the spatial radius of influence were set to the values determined for the statistical interpolation technique (Table 4). The nudging factor is usually less than $1/\Delta t$, where Δt is the model's time step length [*Stauffer and Seaman*, 1990]; 10% of this value worked well in this study.

Table 2. Optimized Background Horizontal Error Correlation Function Parameter Values

	Parameter	
	c	L
Random observation reanalysis	1.1×10^{-7}	3277.1
Superobservation reanalysis	8.9×10^{-8}	12715.4

3.2. Assimilation Results

Watershed average time series of surface and root zone soil moisture derived using the various assimilation strategies are shown in Figure 7. The transmission zone soil moisture is not shown because it varies less than 1% in practice. Nudging to the gridded PBMR observations inside the observed area yields an undesirable discontinuity at this boundary; therefore nudging to randomly spaced observations was performed both inside and outside the observed area. Without calibration the TOPLATS was unable to simulate surface soil moisture dynamics, and even after a comprehensive multiobjective parameter calibration, the model overestimated surface zone soil moisture and was unable to achieve observed poststorm drydown. All of the data assimilation methods significantly and similarly improved the simulation of surface zone soil mois-

Table 3. Vertical Background Error Correlation

		Reanalysis	
Metflux Site	Original	100 Random Observations	100 Superobservations
Lucky Hills (site 1)	0.4	0.26	0.27
Kendall (site 5)	0.62	0.43	0.40

Surface minus root zone, $z = 0.7$ m.

Table 4. Parameters for Soil Moisture Newtonian Nudging in TOPLATS

Parameter	Value	Source
Average depth of observation	0.05 m	*Schmugge et al.* [1994]
Average depth of surface zone	can be variable	TOPLATS parameters
Average depth of root zone	can be variable	TOPLATS parameters
Average depth of transmission zone	can be variable	TOPLATS parameters
Temporal radius of influence	12 hours	subjective optimization
Horizontal radius of influence	10,000 m	SI correlation analysis
Vertical radius of influence	0.5 m	SI correlation analysis
Surface zone nudging constant	0.2	subjective optimization
Root zone nudging constant	0.1	subjective optimization
Transmission zone nudging constant	0.0	subjective optimization
Analysis quality factors	1.0	*Stauffer and Seaman* [1990]
Observation quality control factors	1.0	*Stauffer and Seaman* [1990]

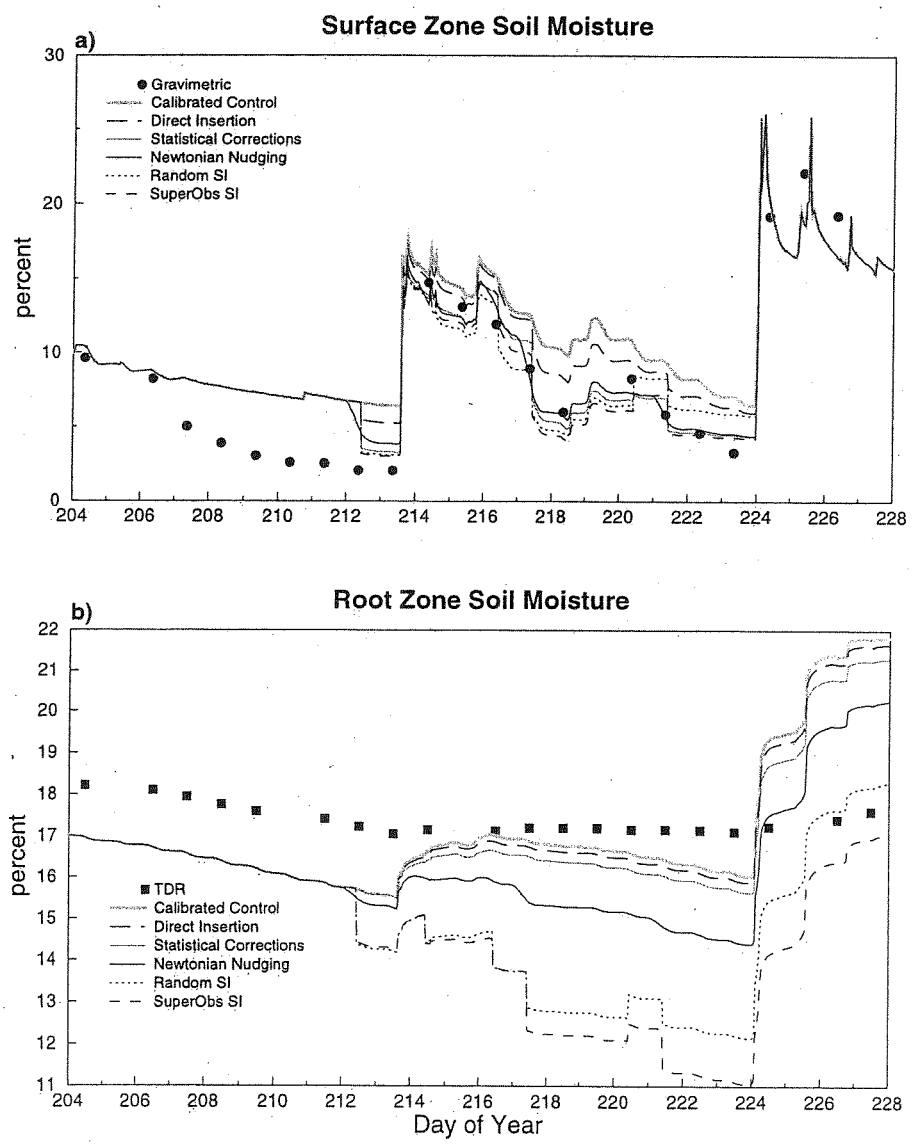

Figure 7. (a) Comparison of TOPLATS watershed average surface zone soil moisture time series from various assimilation simulations. (b) Comparison of TOPLATS watershed average root zone soil moisture time series from various assimilation simulations. (Identical model trajectories often result in one line that represents multiple simulations.)

Table 5. Watershed Average Water Balance Information for the TOPLATS Data Assimilation Simulations

	Simulation				
Variable	Control	Updating	Stat. Corr.	Nudging	SuperObs SI
Precipitation	93.44	93.44	93.44	93.44	93.44
Capillary Rise	0.00	0.00	0.00	0.00	0.00
ET	41.86	39.97	36.76	37.32	36.19
Runoff	6.21	5.68	5.15	5.08	5.11
Recharge	0.00	0.00	0.00	0.00	0.00
Surface ΔS	5.69	5.69	5.69	5.69	5.69
Root ΔS	33.51	32.31	29.85	22.60	0.40
Trans ΔS	6.14	5.75	5.09	3.49	1.91
Leaf ΔS	0.03	0.03	0.03	0.03	0.03
In-out	45.37	47.79	51.53	51.03	52.14
Total ΔS	45.37	43.78	40.66	31.81	8.03
Error	0.00	4.01	10.87	19.22	44.11

Capillary rise refers to the upward flux of water from the saturated surface, ET is evapotranspiration, recharge is the drainage to the saturated zone, surface ΔS is the sum of the changes in storage in the surface zone, leaf ΔS is the sum of the changes in storage in the interception store, in-out is the total input less the total output, total ΔS is the total change in storage, and error is in-out less total ΔS. All values are in millimeters.

ture, with the exception of direct insertion, which is unable to impose an entire watershed correction and was therefore unable to adjust the model trajectory sufficiently. Nudging had the clear advantage of providing smoother temporal adjustments; however, this characteristic also inhibits the application of this method in real time. All simulations produced identical surface zone soil moisture simulations after the storm on day 224 because this storm saturated the surface zone, causing all past surface zone forcing to be forgotten; however, this process does not occur in the models' root zone, where memory of past assimilation is preserved. This sequence of events is not unrealistic; rather, it suggests a time interval at which soil moisture observations are needed for data assimilation, this interval being less than or equal to the time between storm events.

In the root zone the modeled time series fell into two distinct groups corresponding to methods with and without the capability for vertical assimilation of information. Among the latter group, nudging assimilation performs a more conservative correction compared with statistical interpolation. None of the methods produced time series that match the root zone observations. However, it is important to bear in mind that with only two root zone observations, the root zone spatial variability is not adequately sampled.

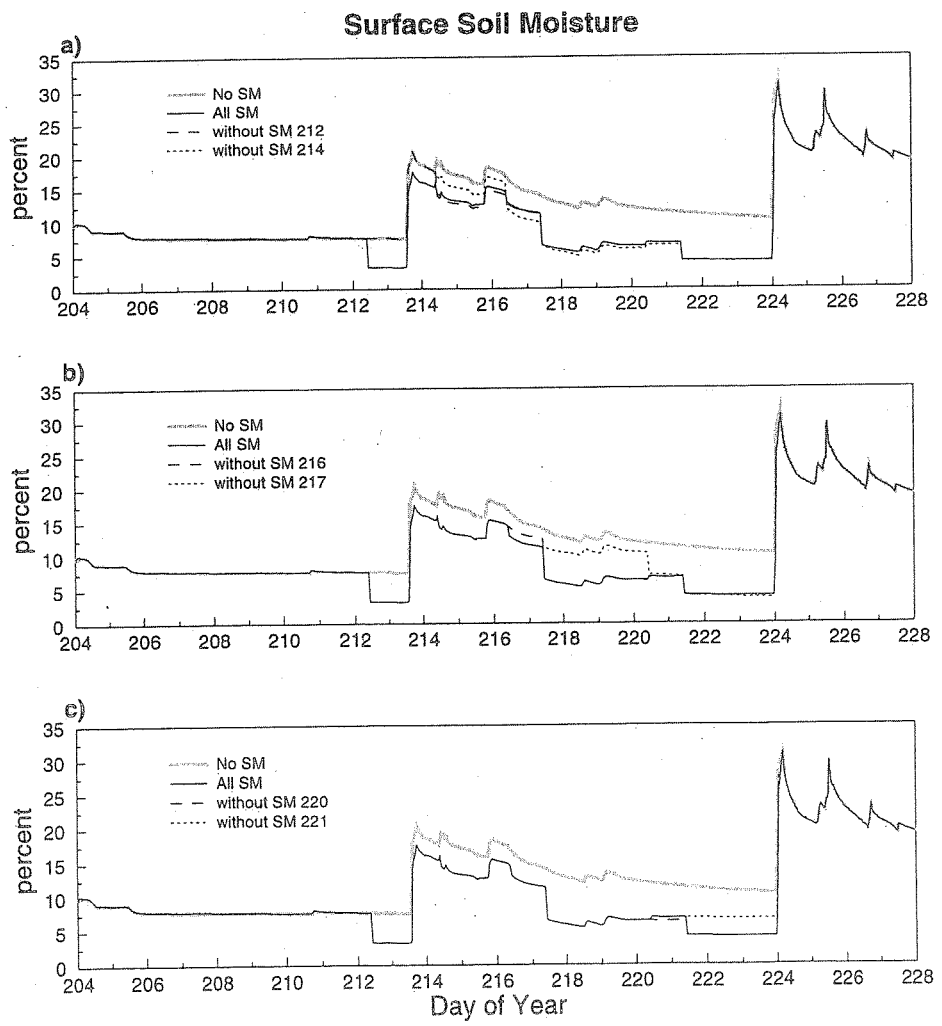

Figure 8. Sensitivity of assimilation information on watershed average surface soil moisture prediction using statistical corrections assimilation.

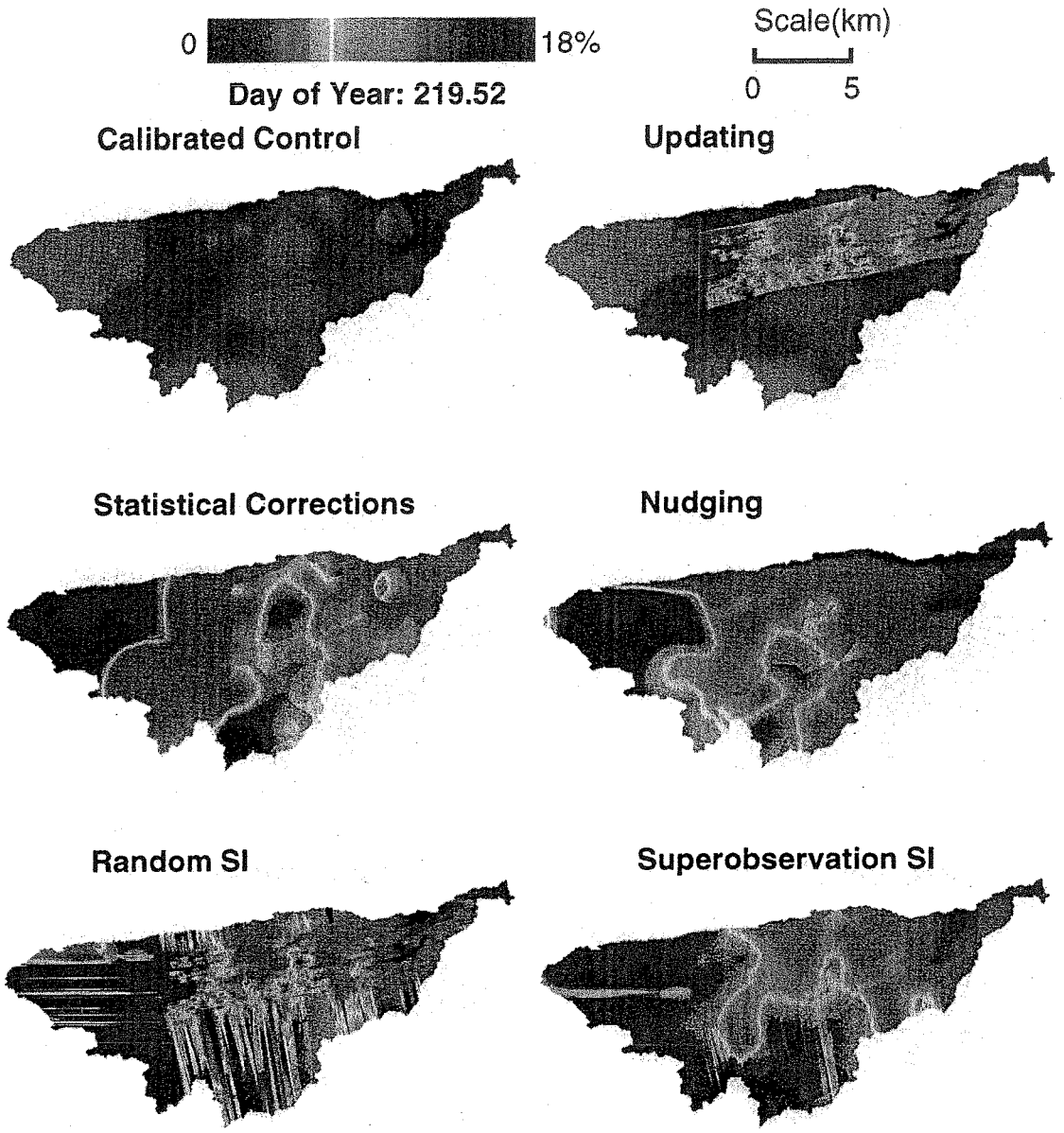

Plate 2. Comparison of TOPLATS spatial surface zone soil moisture on day 219.5 from various assimilation simulations.

The control run deviates most significantly from observations near the end of the drydown on day 219, so this time is selected to demonstrate the intercomparison between assimilation methods. The spatial patterns of model predicted surface and root zone soil moisture for the different assimilation methods are shown in Plates 2 and 3, respectively. The best spatial patterns are those without discontinuity at the edge of the observed area, without numerical artifacts, and with a similar nature to those produced by the model without assimilation.

Simple updating is unable to advect information horizontally, giving rise to an undesirable discontinuity in the calculated soil moisture field and preserving all the observational noise. Updating impacts root zone soil moisture very slightly through model physics and preserves the discontinuity in this zone. Data assimilation via statistical corrections is able to adjust the entire surface soil moisture field to observed levels, but only very slightly modifies the root zone field through model physics. It produces a soil moisture spatial field that does not contain discontinuities or retain the observed spatial pattern. Newtonian nudging also produces a spatial field of soil moisture without discontinuities and makes a larger impact on root zone soil moisture because of the explicit nudging in this zone.

Both the random and the superobservation statistical interpolation approaches result in an undesirable linear streaking feature that extends outward from the observed area, and both provide a relatively strong modification of root zone soil moisture. This streaking would be present had all the observations been used because they are an artifact of extrapolation using the statistical interpolation algorithm beyond the remotely sensed area. The streaking features may also be the result of a violation of the statistical interpolation unbiased background assumption. Statistical interpolation using superobservations

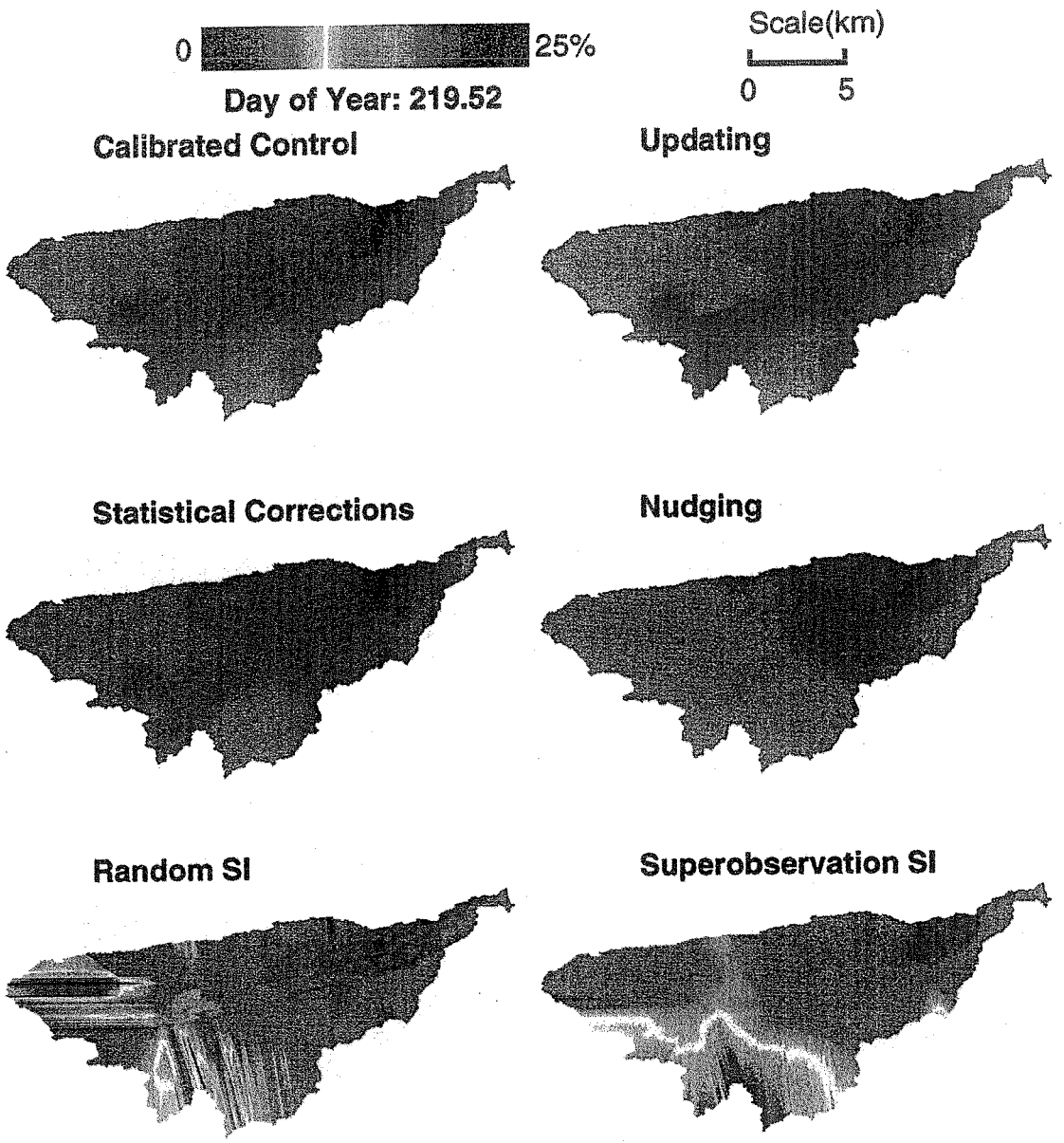

Plate 3. Comparison of TOPLATS spatial root zone soil moisture on day 219.5 from various assimilation simulations.

exhibits more desirable, smooth patterns which resembles those given by Newtonian nudging inside the remote-sensing area. Statistical interpolation has the advantage of using error correlation functions based on the characteristics of the observations and the model predictions. However, it also has the disadvantages of being excessively demanding on computer resources when addressed as a fully posed problem with remotely sensed data, and it lacks the benefits of temporal assimilation.

Overall, the Newtonian nudging method has the most desirable features for remotely sensed soil moisture data assimilation. It is the only true 4DDA method used in this study, and it produces relatively continuous soil moisture time series and reasonable spatial patterns. The primary drawback of Newtonian nudging is its large computational demand, which makes the extremely efficient statistical corrections algorithm look attractive. The statistical corrections method takes only slightly more computational time than a control simulation. However, it is applicable only in areas that have many remotely sensed observations that effectively sample the statistics of the state.

Data assimilation has the consequence that mass and energy are not conserved. If the model has too much or too little water, the data assimilation process either creates or destroys water in such a way as to make these states more realistic. The simulated water balances given by each of the simulations is shown in Table 5. As expected, the model conserves water in the control simulation, while more complex assimilation methods progressively modify the model water balance to a larger extent.

The data assimilation process results in model states that are more realistic and, if the model structure and parameters are properly specified, these more realistic states result in better

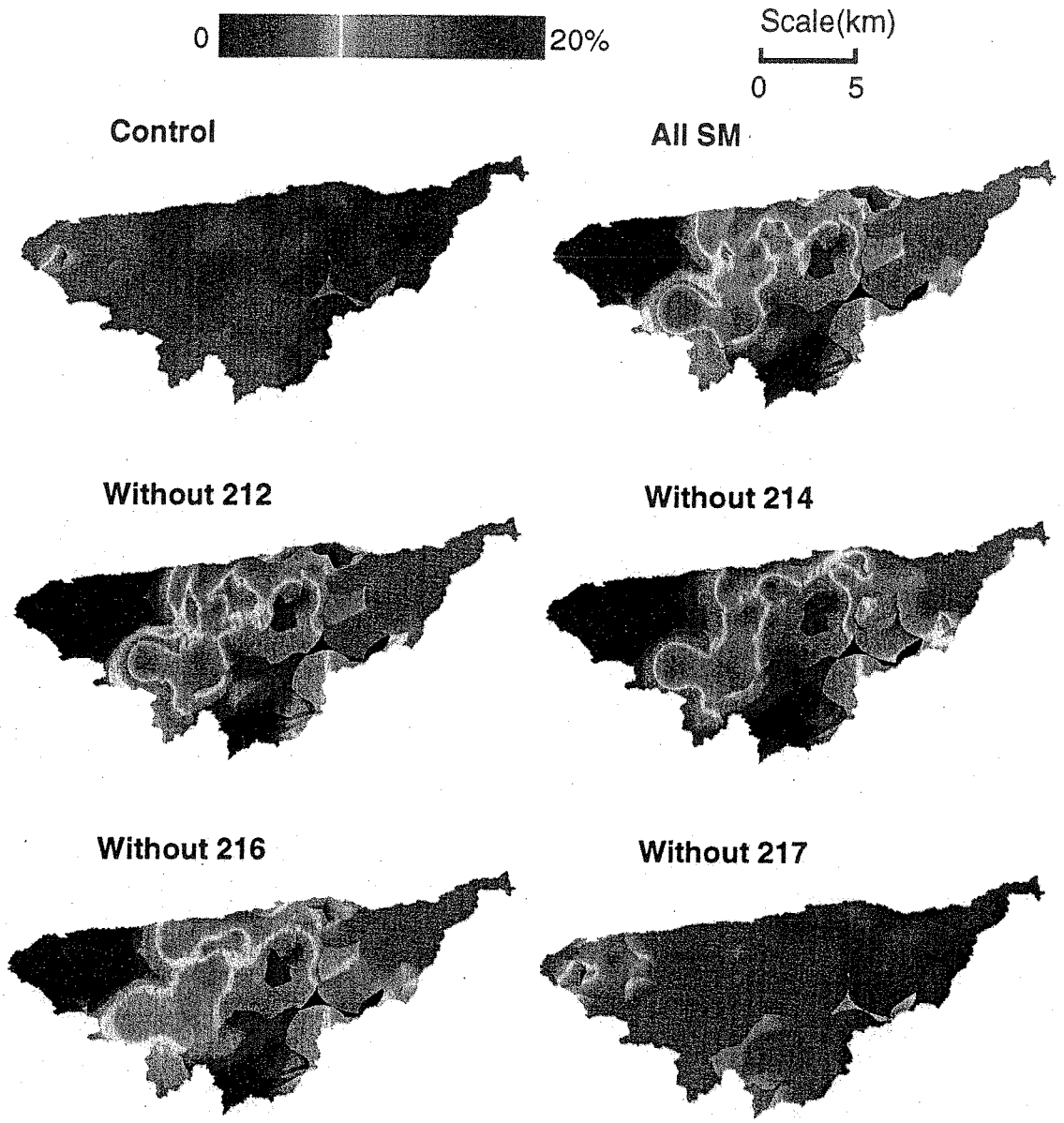

Plate 4. Sensitivity of assimilation information on spatial surface soil moisture prediction using statistical corrections.

model predictions, which was observed in this study. It is expected that these benefits would be more substantial over longer time periods.

3.3. Sensitivity to Assimilation Information

The sensitivity of model predictions to a single set of observations was explored using the statistical corrections assimilation method by comparing the control and the statistical correction simulations with all the observations to 12 statistical correction simulations, each of which excludes or includes one of the six PBMR observation images.

The watershed average surface soil moisture time series for the simulations that exclude one of the six PBMR images is shown in Figure 8. Spatial soil moisture patterns from these sensitivity runs are shown in Plate 4. The removal of any observation degrades the quality of overall patterns predicted. Because the model is not able to simulate dry down events adequately, the observations on day 217 (when the dry down from the day 213 storm completes) are of most importance for model correction.

4. Discussion

The results of this study emphasize the relative importance of hydrologic forcing in the Walnut Gulch watershed. Precipitation is known to be the dominant force in determining the spatial structure of hydrologic response in semi-arid regions. This was confirmed through the ability of a relatively simple model to predict spatial variability in soil moisture with precipitation as its only relevant spatially distributed input. The highly correlated time series of soil moisture and sensible, latent, and soil heat fluxes within Walnut Gulch support the conclusion that the land surface–atmosphere interactions of areas at least the size of Walnut Gulch can be represented reliably with a single set of forcing, flux measurements, and model parameters. If a linear relationship were assumed for

these correlations, it might be surmised that this concept holds for a distance of less than 60 km. The correlation structures vary in time, depending on the soil's state of drying, but its time series do have strong horizontal and vertical correlation structures.

There is a clear trade-off between using a complex data assimilation technique and the ability to use all the available data due to the large computational burdens of performing data assimilation at fine resolutions using dense data sets. On the basis of this study, it was found that, as the complexity of the data assimilation model increases, the size of the assimilated data set needs to decrease in order to maintain computational feasibility. Complex methods have the ability to extract more useful information from assimilated data, but simpler methods use more of the data to extract similar information. This trade-off allows simpler assimilation techniques to perform almost as well as complex techniques. In general, this argument suggests the use of assimilation methods that are of moderate complexity, are sound and computationally efficient, but use as much data as possible. If the information in the data can be efficiently compressed or filtered before its use in data assimilation, it may be more reasonable to use larger data sets in complex data assimilation strategies. Because hydrologic data assimilation requires hydrologic modeling predictions, it is limited by a similar trade-off between fine resolution and large area implementation. A statistically based assimilation may be a viable approach for use in large areas, but ultimately the trade-off between resolution and area will be determined by the application.

Several supplementary observations are essential for implementation of soil moisture data assimilation, the most important being meteorological forcing. Forcing averaged over large areas may be adequate, but detailed spatial patterns of precipitation are essential. Clearly, regular, remotely sensed soil moisture observations are required, but these must be supplemented by in situ surface and root zone observations across the operational domain to specify error correlations, to calibrate parameters, and to validate the model-calculated fields. Observations of soil and vegetation characteristics are likely needed for optimal model performance, while observations of surface water and energy fluxes are valuable for validating simulation results.

Acknowledgments. This work was supported, fostered, and carried out to completion at the University of Arizona, Department of Hydrology, with financial support provided through NASA grants NAGW-4165, NAGW-4087, and NAG5-3492#1 and through the award of a NASA Global Change Fellowship. C. White facilitated processing and checking of Walnut Gulch data sets, while C. White, J. Garatuza, A. Mohammed Arzin, R. Scott, T. Keefer, and C. Harlow assisted in field validation of this work. J. Broermann and D. Braithwaite provided patient and consistent computational assistance. The study would not have been possible without the data provided by the USDA–Southwest Watershed Research Center and without their assistance in interpreting these data from the Walnut Gulch Experimental Watershed. The authors would also like to thank T. Schmugge, T. Keefer, J. Toth, and an anonymous reviewer for their helpful comments on the manuscript.

References

Amer, S. A., T. O. Keefer, M. A. Weltz, D. C. Goodrich, and L. B. Bach., Soil moisture sensors for continuous monitoring, *Water Resour. Bull.*, *30*(1), 69–83, 1994.

Beven, K., and M. J. Kirby, A physically based, variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, *24*(1), 43–69, 1979.

Buell, C., Correlation functions for wind and geopotential on isobaric surfaces, *J. Appl. Meteorol.*, *11*, 51–59, 1972.

Daley, R., *Atmospheric Data Analysis*, 457 pp., Cambridge Univ. Press, New York, 1991.

Daughtry, C. S. T., M. A. Weltz, E. M. Perry, and W. P. Dulaney, Direct and indirect estimates of leaf area index, paper presented at Tenth Conference on Biometeorology and Aerobiology; Special Session on Hydrometeorology, Am. Meteorol. Soc., Salt Lake City, Utah, 1991.

Dickinson, R. E., A. Henderson-Sellers, and P. J. Kennedy, Biosphere-Atmosphere Transfer Scheme (BATS) version 1e as coupled to the NCAR Community Climate Model, *Tech. Note NCAR/TN-387+STR*, Natl. Cent. for Atmos. Res., Boulder, Colo., 1993.

Famiglietti, J. S., Aggregation and scaling of spatially-variable hydrological processes: Local, catchment-scale, and macroscale models of water and energy balance, dissertation, Dep. of Civ. Eng. and Oper. Res., Princeton Univ., Princeton, N. J., 1992.

Famiglietti, J. S., and E. F. Wood, Application of multiscale water and energy balance models on a tallgrass prairie, *Water Resour. Res.*, *30*(11), 3061–3078, 1994.

Humes, K. S., W. P. Kustas, and M. S. Moran, Use of remote sensing and reference site measurements to estimate instantaneous surface energy balance components over a semiarid rangeland watershed, *Water Resour. Res.*, *30*(5), 1363–1373, 1994.

Jackson, T. J., Measuring surface soil moisture using passive microwave remote sensing, *Hydrol. Processes*, *7*, 139–152, 1993.

Kolmogorov, A., Interpolated and extrapolated stationary random sequences (in Russian), *Izv. Akad. Nauk SSSR, Ser. Mat.*, *5*(2), 85–95, 1941.

Kustas, W. P., and D. C. Goodrich, Preface to the special section on Monsoon 90, *Water Resour. Res.*, *30*(5), 1211–1225, 1994.

Kustas, W. P., J. H. Blanford, D. I. Stannard, C. S. T. Daughtry, W. D. Nichols, and M. A. Weltz, Local energy flux estimates for unstable conditions using variance data in semiarid rangelands, *Water Resour. Res.*, *30*(5), 1351–1361, 1994.

Schlatter, T., Some experiments with a multivariate statistical objective analysis scheme, *Mon. Weather Rev.*, *103*, 246–257, 1975.

Schmugge, T., T. J. Jackson, W. P. Kustas, R. Roberts, R. Parry, D. C. Goodrich, S. A. Amer, and M. A. Weltz, Push broom microwave radiometer observations of surface soil moisture in Monsoon '90. *Water Resour. Res.*, *30*(5), 1321–1328, 1994.

Seaman, N. L., Newtonian nudging: A four-dimensional approach to data assimilation, paper presented at Mesoscale Data Assimilation, 1990 Summer Colloquium, Natl. Cent. for Atmos. Res., Boulder, Colo., June 6–July 3, 1990.

Stauffer, D. R., and N. L. Seaman, Use of four-dimensional data assimilation in a limited area mesoscale model, I, Experiments with synoptic-scale data, *Mon. Weather Rev.*, *118*, 1250–1277, 1990.

Syed, K. H., Spatial storm characteristics and basin response, M. S. thesis, 261 pp., Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson, 1994.

Thiebaux, H. J., Anisotropic correlation functions for objective analysis, *Mon. Weather Rev.*, *104*, 994–1002, 1976.

Wei, M. Y., Soil moisture: Report of a workshop held in Tiburon, California, 25–27 January 1994, *NASA Conf. Publ.*, *CP-3319*, 1995.

J. S. Famiglietti, Department of Geological Sciences, University of Texas at Austin, Austin, TX 78712.

D. C. Goodrich and K. H. Syed, Southwest Watershed Research Center, Agricultural Research Service, U.S. Department of Agriculture, Tucson, AZ 85719.

H. V. Gupta and W. J. Shuttleworth, Department of Hydrology and Water Resources, Building 11, Room 122, University of Arizona, P. O. Box 210011, Tucson, AZ 85721-0011. (shuttle@hwr.arizona.edu)

P. R. Houser, Hydrological Sciences Branch and Data Assimilation Office, Code 974, NASA Goddard Space Flight Center, Greenbelt, MD 20771. (houser@dao.gsfc.nasa.gov)

(Received May 14, 1997; revised August 23, 1998; accepted August 28, 1998.)