Improving Simulated Soil Moisture Fields through Assimilation of AMSR-E Soil Moisture Retrievals with an Ensemble Kalman Filter and a Mass Conservation Constraint

Bailing Li

Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland, and Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland

David Toll

Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD

Xiwu Zhan

NOAA NESDIS Center for Satellite Applications and Research, Camp Springs, MD

Brian Cosgrove

NOAA/NWS Office of Hydrologic Development, Silver Spring, MD

Corresponding author and address:

Bailing Li

Hydrological Sciences Branch, code 614.3

NASA Goddard Space Flight Center

Greenbelt, MD 20771.

Tel: 301-286-6020

E-mail: bailing.li@nasa.gov

32

Abstract

34   Model simulated soil moisture fields are often biased due to errors in input parameters

35   and deficiencies in model physics. Satellite derived soil moisture estimates, if retrieved

36   appropriately, represent the spatial mean of soil moisture in a footprint area, and can be used to

37   reduce model bias (at locations near the surface) through data assimilation techniques. While

38   assimilating the retrievals can reduce model bias, it can also destroy the mass balance enforced

39   by the model governing equation because water is removed from or added to the soil by the

40   assimilation algorithm. In addition, studies have shown that assimilation of surface observations

41   can adversely impact soil moisture estimates in the lower soil layers due to imperfect model

42   physics, even though the bias near the surface is decreased. In this study, an ensemble Kalman

43   filter (EnKF) with a mass conservation updating scheme was developed to assimilate the actual

44   value of Advanced Microwave Scanning Radiometer (AMSR-E) soil moisture retrievals to

45   improve the mean of simulated soil moisture fields by the Noah land surface model.

46   Assimilation results using the conventional and the mass conservation updating scheme in the

47   Little Washita watershed of Oklahoma showed that, while both updating schemes reduced the

48   bias in the shallow root zone, the mass conservation scheme provided better estimates in the

49   deeper profile. The mass conservation scheme also yielded physically consistent estimates of

50   fluxes and maintained the water budget. Impacts of model physics on the assimilation results are

51   discussed.

52

53

1. Introduction

Soil moisture plays an important role in the energy and water exchange between the

atmosphere and the land surface, as well as in agricultural applications and water resource

management. Model simulated soil moisture fields are often biased due to uncertainties in model

input parameters and model physics. The existence of model bias can be seen in several model

inter-comparison studies which showed that model estimated soil moisture is significantly

different from each other, even when identical forcing data are used (*Wood et al*., 1998; *Mitchell*

et al., 2004). Recognizing the significant disparity between the models, *Mitchell et al*. (2004)

concluded that there was a 'stringent need for good absolute states of soil moisture.' Reducing

the bias in model estimated soil moisture fields has been shown to have a positive impact on

other physical processes. *Dirmeyer* (2000) demonstrated that the rainfall patterns and the near

surface air temperature can be improved by using a mean soil moisture data set derived from a

global soil moisture data bank.

Satellite derived soil moisture retrievals represent the spatially averaged soil moisture in a

footprint area (*Njoku et al*., 2003). If retrieved appropriately, they can be used to improve the

spatial mean of the modeled soil moisture field as well as the temporal mean through continuous

assimilations in time. While interests in assimilating satellite retrieved soil moisture estimates

began more than a decade ago (*Houser et al*, 1998; *Walker et al*., 2001; *Margulis et al*., 2002),

recent studies have focused on using the anomaly information extracted from the sensor data by

removing the mean of the observations priori to assimilation to improve model's anomaly

74   detection (*Reichle et al.*, 2007; *Crow and Zhan*, 2007; *Bolten et al.*, 2008; *Draper et al.*, 2009).

75   While assimilation of anomalies does not directly address if models are unbiased which is

76   required for optimal estimators (*Kalnay*, 2003), it preserves the water budget of forecasts.

77   An alternative to the offline bias-removal technique, as those used in the above studies, is

78   to estimate the forecast bias online by adding a bias state in the filtering process (*Keppenne et al.*

79   2005; *De Lannoy et al.* 2007a and 2007b). *De Lannoy et al.* (2007a and 2007b) compared the

80   performance of several online bias correction techniques with the standard EnKF using the CLM

81   land model and profile soil moisture observations. Their results showed that the online bias

82   correction techniques, on average, yielded slightly more reductions in root mean square error.

83   One major obstacle for applying this approach in assimilating satellite retrieved soil moisture is

84   that observations are only available at the surface which makes it very challenge to estimate the

85   bias state in the deeper profile. When bias is not correctly estimated, assimilation may lead to

86   unbalanced water budget as the assimilation may change the mean of simulated soil moisture

87   fields. Lack of water budget closure is a weak point for many data assimilation systems as

88   pointed out by *Pan and Wood* (2006), perhaps more so for land surface models whose major goal

89   is to partition the total water budget, precipitation, into different physical processes such as

90   evapo-transpiration (ET) and runoff.

91   When sensor data are less biased (relative to the truth) than model estimates, they can be

92   used to reduce uncertainty in model estimates. Recognizing this potential , studies have been

93   conducted to assimilate actual values of satellite data without using any bias correction

94   techniques (*Houser et al.*, 1998; *Walker et al.*, 2001; *Margulis et al.*, 2002; *Ni-Meister et al.*,

95   2006). While the bias reduction at the surface was achieved in these studies, improvements in

96   the deeper soil layers did not always occur. *Houser et al*. (1998) and *Walker et al*. (2001)

97   showed that assimilation of surface observations actually adversely impacted the soil moisture

98   state in the lower soil layers. The representation of the hydrological condition in the lower soil

99   zone is often a weak point in land surface models due to lack of knowledge and observations. If

100   model physics is flawed, it may adversely impact the outcome of data assimilation, especially for

101   an EnKF which relies on model physics to calculate the Kalman gain matrix dynamically

102   (*Keppenne et al*., 2000).

103   The objective of this study is to assimilate the actual value of AMSR-E soil moisture

104   retrievals into the Noah land surface model to improve the mean of simulated soil moisture fields

105   using an EnKF. To overcome the potential bias issue associated with both the model and the

106   AMSR-E retrieval, a mass conservation updating scheme was developed to allow the upper soil

107   layers updated using the conventional EnKF while the lower layers are updated with an equation

108   that conserves mass of the forecast. This study differs from those recent studies on AMSR-E

109   data assimilation (*Reichle et al*., 2007; *Crow and Zhan*, 2007; *Bolten et al*., 2008; *Draper et al.*,

110   2009) in that AMSR-E retrievals were not pre-processed priori to assimilation while, in the other

111   studies, the mean of retrievals were removed through matching the cumulative distribution

112   functions (*Reichle and Koster, 2004; Drusch et al., 2005*). By assimilating the actual value of

113   AMSR-E soil moisture, the objective of this study is to reduce forecast bias and estimation

114   errors, rather than to improve anomaly detections (e.g., *Reichle et al*., 2007). In section 2, the

115   experiment site, data, and the Noah model are briefly described. Details of the mass

116   conservation assimilation method along with the conventional EnKF are described in section 3.

117   Assimilation results including all land surface fluxes and water budgets are presented in section

118   4. Impacts of model physics on model simulation and assimilation results and the limitations of

119   the mass conservation scheme are discussed in section 5.

120   **2. Experiment site, data and model**

121   **2.1 Study area and ground validation data**

122   The Little Washita watershed, located in southwestern Oklahoma, was chosen as the

123   study site primarily for its abundance of in situ soil moisture measurements. With an area of 611

124   square kilometers, the watershed is one of the two Micronet sites maintained by the U.S.D.A.

125   Agriculture Research Service (ARS) for hydrological and meteorological observations

126   (http://ars.mesonet.org). Figure 1 shows the watershed boundary and the locations of the ARS

127   stations. At each station, hourly soil moisture and temperature measurements are taken at 5, 25

128   and 45 cm depths below the surface, in addition to surface measurements such as precipitation.

129   Figure 1 also shows the only Soil Climate Analysis Network (SCAN) station located within the

130   watershed (*Schaefer et al*., 2007; http://www.wcc.nrcs.usda.gov/scan/). The SCAN site

131   complements the ARS stations in that it provides soil moisture measurements at the 100 cm

132   depth which were used to verify simulated soil moisture in the deeper soil profile. Daily stream

133   flow data recorded at the watershed by USGS (see Figure 1 for the location of stream site

134   07327550) were used for validating model predicted runoff. Latent heat measurements from the

135   Southern Great Plain (SGP) main station (http://public.ornl.gov/ameriflux) were used for

136   validating the simulated latent heat. Although SGP, which is approximately 200 km north of

137   Little Washita, is not located near the watershed, it is the nearest site where flux data are

138   publically available.

139   **2.2 AMSR-E retrievals**

140   The AMSR-E soil moisture product produced by the NOAA's National Environmental

141   Satellite, Data and Information Service (NESDIS) was used in this study. The soil moisture

142   retrievals, based on the X-band brightness temperature measurements, were obtained through the

143   inversion of the Single Channel Retrieval algorithm with the MODIS vegetation water content as

144   an auxiliary variable (*Zhan et al.*, 2008; *Jackson*, 1993). *Zhan et al*. (2008) showed that this

145   version of AMSR-E generally has larger dynamic ranges than the official AMSR-E product

146   (*Njoku et al.*, 2003) even though both products show strong temporal correlations (*Crow and*

147   *Zhan*, 2007). The spatial resolution of AMSR-E retrievals is about 25 by 25 km after re-

148   sampling from its original sensor data (*Njoku et al.*, 2003). The experiment site contains about 5

149   to 6 AMSR-E pixels at any observation time. On average, there are 1~2 retrievals per day at any

150   given location and both ascending and descending data were assimilated at the retrieval time,

151   except in areas of dense vegetation or frozen grounds.

152   The sensing depth of the AMSR instrument is believed to be about 1-2 cm from the

153   surface for the frequency range of AMSR-E (*Njoku et al*., 2003). This depth is shallower than

154   the ARS surface measurement (5 cm) and the center of Noah's surface layer. However, without

155   reliable methods to extrapolate the AMSR-E estimates, it was assumed that the AMSR-E soil

156   moisture retrieval is representative of soil moisture in the top 5 cm soil and therefore was

157 assimilated into Noah's top layer directly. This approximation could bring some bias into the

158 retrievals used for data assimilation.

2.3 The Noah land surface model, forcing and input parameters

160 The Noah land surface model (version 2.7.1) is used operationally at the NOAA's

161 National Centers for Environmental Prediction for coupled weather and climate modeling. The

162 soil moisture simulation in Noah is based on the one-dimensional Richards equation (*Chen et al.*,

163 1996; *Ek et al.*, 2003):

$$\frac{\partial \theta(z,t)}{\partial t} = \frac{\partial}{\partial z}(D(\theta)\frac{\partial \theta(z,t)}{\partial z} + K(\theta)) + P - R - E \qquad (1)$$

164 where θ is the soil moisture content; K is the hydraulic conductivity; D is the water diffusivity,

165 which is defined as $K\partial\psi/\partial\theta$, where ψ is the matric potential; P is the precipitation; R is the

166 surface runoff; E is the ET; z is the vertical dimension with upward as the positive direction; t is

167 the time.

168 Following the operational version of Noah (*Ek et al.*, 2003), four soil layers with

169 thicknesses of 10, 30, 60 and 100 cm were used in this experiment. The top two layers, a thin

170 surface layer and the shallow root zone, generally show stronger and faster interactions with the

171 atmospheric forcing. The third and the fourth layers represent the deeper root zone and water

172 storage, respectively.

173   Equation (1) cannot be solved without a boundary condition at the 200 cm lower

174   boundary:

$$q\,|_{z=200cm}=-K \tag{2}$$

175   where q is the subsurface runoff or base flow. Equation (2) is also referred to as the free

176   drainage condition, meaning gravity is the only force pushing water downward (so the negative

177   sign) and no upward diffusive movement is allowed across the lower boundary (*Jury et al.*,

178   1991). The use of free drainage is very common in land surface models because it does not

179   require any knowledge about the soil moisture state or flux in the subsurface which is impossible

180   to obtain for large-scale modeling.

181   Noah uses the *Campbell* (1974) model to describe the nonlinear relationship between the

182   conductivity and soil moisture:

$$K = K_s\left(\frac{\theta}{\theta_s}\right)^{2b+3} \tag{3}$$

183   where K_s is the saturated conductivity; θ_s is the saturated water content; b is a fitting parameter.

184   The U.S. general soil texture classes (STATSGO) and a look-up table, based on a unified soil

185   hydraulic parameter set (*Mitchell et al.*, 2004), were used to provide soil hydraulic parameters

186   needed for solving equation (1). Hydraulic conductivity usually exhibits the property of a log-

187   normal distribution and is positive skewed (*Cosby et al.*, 1984). As a result, the subsurface

188   runoff calculated using equation (2) is non-Gaussian which can lead to unrealistic ensemble

189   mean values in an EnKF when larger ensemble spreads occurred in the lowest soil layer (*Ryu et

190   al.*, 2009; *De Lannoy et al.*, 2007a).

191   Model simulations were carried out in the NASA's Land Information System (LIS,

192   version 5.0) which is a software interface between various land surface models and forcing/static

193   parameter fields (*Kumar et al*., 2006). LIS is also equipped with a one-dimensional EnKF

194   (*Kumar et al*., 2008) which will be described in the next section. The Noah model was

195   integrated on a 0.01 degree grid so that spatial variability was well represented in model

196   estimated soil moisture and flux at the watershed. To avoid the model spin up issue (*Rodell et*

197   *al.*, 2005; *Cosgrove et al*., 2003a), the initial soil moisture conditions used were extracted from

198   the output of Global Land Data Assimilation (GLDAS)/Noah model which have been

199   continuously integrated since 1979 (*Rodell et al*., 2004).

200   Model simulations were driven by forcing data (including precipitation, radiation, wind,

201   and temperature fields) from the NOAA/NCEP Global Data Assimilation System (GDAS,

202   *Derber et al*., 1991; *Rodell et al*., 2004). Basin-averaged monthly GDAS and ARS precipitation

203   for the simulation period (2006-2007) are compared in Figure 2 and their annual precipitation

204   amounts are listed in Table 1. Despite some underestimated and overestimated events in the

205   GDAS forcing data, both data sets showed that 2006 is a drier year than 2007.

206   **3. Data assimilation methods**

207   In this section, the conventional EnKF and the mass conservation EnKF scheme are

208   described. EnKF is a widely used technique for assimilating observations into numerical models

209   to improve model estimates (e.g. *Evensen and van Leeuwen*, 1996; *Crow and Wood*, 2003;

210   *Keppenne et al*., 2000; *Pan and Wood*, 2006; *Reichle et al*. 2007). EnKF is especially suited for

211   a non-linear system since the error covariance, used for passing observation information from

10

212   data-rich zones to data-poor zones, is calculated through an ensemble of model states (*Evensen*

213   *and van Leeuwen*, 1996). An EnKF usually consists of two steps: the forecast step where an

214   ensemble of model forecasts are obtained and propagated forward in time with perturbations

215   added for forcing and state variables, and the update step where an analysis is obtained using an

216   update equation when observations become available. The model forecast can be expressed as:

$$\mathbf{X}_t^f = M(\mathbf{X}_{t-1}^a, F, U) \tag{4}$$

217   where **X** is the vector containing the four state variables of soil moisture of Noah; M represents

218   the Noah model; F represents all the forcing fields such as precipitation and radiation; U

219   represents static input parameters such as soil hydraulic parameters; and t indicates the time step.

220   The superscript (^{f}) indicates results for the forecast and (^{a}) for the analysis. Although not

221   explicitly noted, equation (4) and the following update equations are valid for each ensemble

222   member. The conventional EnKF updating scheme for obtaining the analysis can be written as:

$$\mathbf{X}_t^a = \mathbf{X}_t^f + \underline{\mathbf{K}}(v_t - \mathbf{H}\mathbf{X}_t^f) \tag{5}$$

223   Where $\underline{\mathbf{K}}$ is the Kalman gain matrix computed from the ensemble statistics of the model

224   simulated soil moisture fields (*Keppenne*, 2000); v is the observation (AMSR-E retrievals in this

225   study); **H** is the observation operator that relates the observation to the model state and is [1, 0, 0,

226   0] in this study because the observation is the same type as the model state and is only available

227   at the surface layer. The AMSR-E retrievals were used without downscaling, i.e., all model grid

228   points within the footprint of the satellite were given the same retrieved soil moisture value,

229   which is equivalent to a priori partition of the large scale retrieval to the finer scale with the same

230   value assigned to each grid cell. This approach allows for direct and efficient assimilation of

231   satellite retrievals using the current infrastructure of LIS. It is justified for the purpose of this

232   study which is to improve the spatial mean of simulated soil moisture fields and will not be an

233   issue for larger scale simulations where model resolutions can be made to match that of AMSR-

234   E. Since observations are only available at the surface, the innovation, $(v-\mathbf{H}\mathbf{X}^f)$, is a scalar. The

235   $\underline{\mathbf{K}}$ matrix propagates the innovation downwards to obtain the increment, $\underline{\mathbf{K}}(v-\mathbf{H}\mathbf{X}^f)$, for all lower

236   layers.

237   When both the model and the observation are unbiased, the mean of the innovation (and

238   increments) is zero. When either or both of them are biased, the analysis $(\mathbf{X}^a)$ obtained through

239   equation (5) may not possess the same mean as the forecast $(\mathbf{X}^f)$ which is enforced by the mass

240   balance Richards equation. The CDF matching technique used by previous studies (e.g., *Reichle*

241   *et al*. 2007) renders the mean of the retrievals equal to that of the model and therefore preserves

242   the mean of the forecast. The tradeoff of this scaling approach is that it discards the mean value

243   of retrievals which may be useful in improving the mean of model estimates.

244   In order to assimilate the actual value of retrievals which may not have the same mean as

245   model estimates, the loss of water mass (relative to the forecast) needs to be handled in the

246   updating scheme. *Pan and Wood* (2006) used a two-step constrained Kalman filter to

247   redistribute the mass imbalance caused by assimilating multiple types of observations (ET,

248   stream flow and soil moisture). When only the surface soil moisture observation is available for

249   assimilation, the redistribution of mass imbalance can be carried out within the four soil layers.

250   Specifically, while the top two layers are updated using equation (5), a different updating scheme

251   can be used for the lower two layers:

$$\mathbf{Y}_t^a = \mathbf{Y}_t^f - \sum_{k=1}^{2}(\Delta C_k d_k)/(d_3 + d_4) \qquad (6)$$

252   where $\mathbf{Y}$ contains the soil moisture state of the lower two layers; d represents the thickness of

253   each soil layer; subscripts $(_k)$, $(_3)$, and $(_4)$ indicate the soil layer; ΔC_k represents the increment,

254   i.e., water (in soil moisture content) lost or gained, for the top two layers when they are updated

255   using equation (5). Equation (6) redistributes the mass imbalance (amount of water) incurred in

256   updating the top two layers to the lower layers and therefore, guarantees that the total water

257   storage remains the same for each ensemble member after the ensemble update. The division of

258   layer thicknesses in equation (6) is to convert the amount of water to volumetric soil moisture

259   content to match the unit of the state variable. Equation (6) is performed each time when the

260   upper two layers are updated so that the column water of the analysis remains the same as the

261   forecast (but with a different soil moisture profile). By maintaining the water storage within a

262   soil column, the mass conservation scheme also preserves the long-term water budget of the

263   control run (without any data assimilation) since ET and runoff are calculated based on the

264   column water storage and perturbations added to the forcing and state variables are unbiased.

265   Because of the enforcement of mass conservation of the control run, this scheme (equation (5)

266   for top two layers and (6) for the two lower layers) is referred to as the mass conservation

267   updating scheme. Note that no assumption was made about the observation and the model, both

268   of which can be biased, in deriving equation (6).

269   In addition to preserving mass, equation (6) avoids updating the lower layers with the

270   conventional EnKF which has been shown to yield undesired increments due to inappropriate

271   model physics (*Houser et al.*, 1998; *Walker et al.*, 2001). Preserving water mass does not

272   necessarily lead to improved soil moisture estimates in the lower layers, but equation (6) keeps

273   the increments small due to the larger thickness of lower two layers relative to the upper two

274   layers, and thus minimizes any potential adverse impacts.

275   The ensemble of model states was generated by adding zero-mean perturbations (errors)

276   to the forcing fields and state variables to represent random errors in them. Following *Reichle et*

277   *al.* (2007), precipitation, long and short wave radiation fields which have the largest impact on

278   soil moisture were perturbed using the same parameters given by *Reichle et al.* (2007) as the

279   same forcing data were used in both studies. Perturbations for precipitation and shortwave

280   radiation were assumed to be multiplicative and additive for longwave radiation. The

281   perturbation frequency for these forcing fields was 5.5 hours.

282   Perturbations were also added to soil moisture variables to account for errors in the input

283   parameters such as soil hydraulic conductivity and model physics using parameters listed in

284   Table 2. Smaller perturbations (in volumetric soil moisture content) were given to the lower two

285   layers because of their larger thickness and the fact that perturbations added in the top two layers

286   can travel downward through the dynamics of the Richards equation. In addition, the issue with

287   the calculation of ensemble mean base flow due to the skewness of the hydraulic conductivity

288   function (*De Lannoy et al.,* 2007a; *Ryu et al.,* 2009) also requires smaller perturbations in the

289   lower layers to ensure physically consistent ensemble runoff. All soil moisture variables were

290   assumed to have additive zero-mean Gaussian errors with vertical cross-correlations among four

291   layers given in Table 2. The perturbation frequency for soil moisture was 24 hours. Noah Soil

292   moisture moves very slowly in drier conditions, which is why the longer perturbation frequency

293   was used to avoid ensemble bias. Despite all zero-mean perturbations, ensemble bias could still

294   exist in the ensemble soil moisture field due to the nonlinear relationship among various

295   processes and the strong influence of model physics. Parameters in Table 2 were chosen because

296   they yielded unbiased ensemble (without data assimilation) soil moisture fields relative to a

297   single member control run. The 3% AMSR-E error (*Njoku et al*., 2003) was used in the filter to

298   account for errors in the observation. The same filter parameters were used for both updating

299   schemes.

4. Results

301   Three simulation runs were performed at the Little Washita watershed for the 2006 to

302   2007 period. The control run (Control), which represents the baseline performance of the Noah

303   model, was driven by the GDAS forcing and all the parameter fields in their unperturbed states.

304   The other two simulations featured assimilations of AMSR-E soil moisture retrievals using the

305   conventional (DA) and mass conservation (DA MassCon) updating schemes.

306   Given the objective of this study which is to improve the mean of model estimates, basin

307   averaged daily bias and root mean square errors (rmse) were used to evaluate the assimilation

308   results. All statistics were calculated with respect to the ground validation data described in

309   section 2.

4.1 Soil moisture

Figure 3 shows the comparison of soil moisture in the four Noah soil layers from the
three simulations. The upper left panel also includes basin averaged AMSR-E soil moisture
retrievals and ARS measurements at the 5 cm depth. Overall, the AMSR-E soil moisture
compares well with ARS by capturing the seasonal change and the mean value of in situ
measurements. The daily variation of AMSR-E is small due to the twice per day (maximum)
retrieval interval. Control also captured the wetting and drying cycles of the surface soil
moisture, exhibiting strong correlation with the ARS measurements. However, it consistently
overestimated the surface soil moisture throughout the simulation period, even in the period from
December 2006 to June 2007 when GDAS underestimated the precipitation (see Figure 2). The
same overestimation was also observed (not shown) when the model was driven by the North
America Land Data Assimilation System (NLDAS, *Cosgrove et al*., 2003b) forcing data which
yielded nearly unbiased monthly precipitation estimates against ARS measurements (not shown).
These results indicate that the bias at the surface was not initiated by errors in the precipitation
forcing data. Figure 3 also shows that the overestimation by Noah was more severe in winter
periods when ET and precipitation were low, which limits the likelihood that incorrect runoff
and ET algorithms may have left excessive water at the surface. Flux results that will be
discussed in section 4.2 also do not show any negative bias. *Li and Rodell* (2011) compared
NLDAS/Noah with SCAN soil moisture for the continental US and found the similar
overestimation in the western US. The likely cause for this persistent overestimation in such a
large area may be the static parameters such as soil hydraulic conductivity. The vertical drainage
of soil moisture in Noah is controlled by the nonlinear function of soil hydraulic conductivity as

332 shown in equations (1) and (3). The parameters in equation (3) were obtained through linear

333 regressions (*Cosby et al*., 1984) which may not capture all the nonlinear behaviors of hydraulic

334 conductivity. If the hydraulic conductivity value is lower than expected in the drier range of soil

335 moisture, it would explain why Noah failed to drain soil moisture quickly in Little Washita and

336 the western US (*Li and Rodell*, 2011).

337   DA and DA MassCon both greatly reduced the overestimation of Control in the surface

338 layer. The degree of correction is not even, especially in very wet conditions where the

339 assimilation failed to nudge the soil moisture towards the AMSR-E retrievals. This is because

340 the perturbation parameters for the filter had to be tuned to work with the driest condition in

341 order to avoid ensemble bias. If overly perturbed to fit the need of wetter conditions, ensemble

342 bias would appear in drier periods because some of the ensemble members would hit the lower

343 bound of soil moisture (*Reichle and Koster*, 2002).

344   Figure 3 also shows that both updating schemes decreased soil moisture in layer 2.

345 However, for layers 3 and 4, the two schemes acted differently. DA lowered the soil moisture in

346 layers 3 and 4 as it did with the top two layers. DA MassCon increased the soil moisture in the

347 lower layers because it captured the amount of water removed from the top two layers in the

348 lower layers. As shown in Figure 4 which compares the simulations with in-situ soil moisture

349 measurements at various measuring depths, DA and DA MassCon both improved over Control at

350 25 and 45 cm by lowering the soil moisture accordingly. But only DA MassCon improved over

351 Control at 100 cm while DA degraded the estimate by further lowering the soil moisture.

352 Statistics in Table 3 show that DA performed slightly better than DA MassCon in the upper three

353   observation levels. But only DA MassCon achieved improvements in all four levels. Table 3

354   also lists the statistics for the AMSR-E retrievals which are nearly unbiased relative to in situ

355   measurements, although the mass conservation scheme does not require retrievals to be unbiased.

356   The improvement made by DA MassCon at 100 cm may be debatable due to the co-

357   existence of the overestimation in the upper layers and the underestimation of soil moisture in

358   the lower layers, which is found to be true for the Noah model in the western US (*Li and Rodell*,

359   2011). *Houser et al*. (1998) also showed the similar model behavior with a different model.

360   When the overestimation and underestimation do not occur concurrently, the mass conservation

361   algorithm may not lead to improved soil moisture estimates in the lower soil profile, but it does

362   not cause significant changes (relative to Control) to the lower soil moisture states, as seen in

363   Figures 3, because of the smaller increments given by equation (6).

364   The conventional updating scheme generated increments with the same sign for all

365   layers that significantly decreased soil moisture in the lower profile, a result not supported by in

366   situ measurements at 100 cm. *Ni-Meister et al*. (2006) and *Houser et al*. (1998) also showed that

367   the sign of increments for the lower layers was the same as the surface layer. At least for Noah,

368   the fact all increments have the same sign is due to the free drainage condition which has to

369   adjust the soil moisture in the lower layers according to changes in the upper layers in order to

370   maintain the downward flow direction (see equation (2)). Negatively cross-correlated soil

371   moisture perturbations between the upper layers and lower layers were also tested for the

372   conventional EnKF (not shown). They did not change the sign of the increments but slightly

373   lowered their magnitudes, with the soil moisture estimates in the lower two layers slightly wetter

374 than those shown in Figures 3 and 4 but still much worse (drier) than Control when compared to

375 in situ measurements. Cross-correlations of perturbations only partially influence the outcome of

376 the increments which also strongly depend on model physics. As model physics largely

377 determines the mean behavior of soil moisture, it is difficult for the zero-mean perturbations

378 alone to overcome the large difference between DA and Control seen in Figure 4 (lower panel).

379 Significantly increasing perturbations for soil moisture is not permitted because it will lead to

380 ensemble bias in soil moisture and base flow.

381   Similar to surface overestimation, the underestimation by Control in the lower profile

382 cannot be explained by any precipitation forcing errors. In fact, the underestimation is caused by

383 the free drainage condition which drains excessive water away and prevents moisture moving up

384 from below the land surface. The incorrectness of the free drainage condition is why the surface

385 overestimation did not occur at 100 cm. The underestimation has also been observed with other

386 models which employ the same boundary condition (*Zeng and Decker*, 2009; *Houser et al.*,

387 1998). *Li and Rodell* (2011) further found the underestimation is true for NLDAS/Noah in the

388 entire continental US when Noah was compared with SCAN soil moisture. This deficiency in

389 model physics is why the conventional EnKF cannot obtain increments favorable for

390 improvements in the lower profile. Presumably, the underestimation of soil moisture in the

391 lower soil moisture profile could also be balanced out with base flow which would require either

392 deeper soil moisture measurements or observations of base flow to create the innovation. With

393 only the surface soil moisture observation available, the mass conservation scheme elects to

394 improve the soil moisture fields first and let the mass conservation to constrain flux estimates.

395   Figure 5 features the contour plot of the annual mean surface soil moisture at the

396   watershed. Control revealed that 2006 is drier than 2007, confirming the earlier analysis

397   regarding GDAS precipitation (Figure 2 and Table 1). NOAA AMSR-E retrievals also captured

398   the difference in annual precipitation, as DA and DA MassCon all show the wetter soil moisture

399   condition in 2007. DA MassCon yielded slightly higher soil moisture estimates than DA

400   because the former has a wetter lower soil profile (see Figure 3) which pushed the surface soil

401   moisture slightly higher via the capillary force. This is why DA achieved slightly better statistics

402   than DA MassCon for the upper three observation levels shown in Table3. However, for the root

403   zone soil moisture (consisting of the upper three Noah soil layers), Figure 6 shows that while DA

404   MassCon and Control yielded the wetter soil moisture condition in 2007, DA struggled to show

405   this variation in annual precipitation, further confirming the failure of DA in updating the lower

406   layers.

407   Figures 5 and 6 also show that the spatial variability of Control was generally preserved

408   by the assimilation schemes even though the AMSR-E retrievals were assimilated directly

409   without any spatial downscaling. Note that spatial variability of soil moisture may be lost

410   slightly at the assimilation time, but it recovered quickly afterwards because of the high

411   resolution soil and vegetation parameters.

412   **4.2 Flux**

413   One of the important roles of any land surface model is to simulate water and energy

414   fluxes based on soil moisture fields. Improvements on soil moisture do not necessarily lead to

415   improvements in the calculation of flux because of imperfect model physics and complex

20

416 relationship among various processes. Therefore, it is important to examine all components of

417 model estimates to prevent unexpected flux estimates.

418   Figure 7 shows the simulated latent heat fluxes in comparison with SGP observations.

419 The differences among the three simulations are relatively small, with total ET for the two-year

420 period estimated at 1362, 1013 and 1147 mm, for Control, DA and DA MassCon, respectively.

421 Part of the reason is that the ET algorithm in Noah is more sensitive to the vegetation greenness

422 fraction than soil moisture (*Chen et al.*, 1996). In addition, the watershed is mostly covered by

423 vegetations with shallow root zone depths such as shrubs and grasses which do not strongly

424 depend on the soil moisture state in lower profile where DA differs from DA MassCon the most.

425 Nevertheless, Table 3 shows that Control yielded the largest bias (positive) in latent heat

426 estimation. DA reduced the bias but DA MassCon produced the smallest bias. Although the

427 improvement by DA MassCon and DA should not be overstated given that the SGP is not

428 located near the watershed, the impact of the different soil moisture fields on the latent heat

429 estimation is demonstrated.

430   Noah employs the Simple Water Balance (SWB) model by *Schaake et al*. (1996) to

431 partition the precipitation into surface runoff and infiltration. Soil moisture deficits in the entire

432 profile and precipitation intensity are accounted for in the implementation of SWB in Noah (*Ek

433 et al.*, 2003). Figure 8 (upper panel) shows that the three simulations yielded very similar

434 surface runoff. DA, which produced the driest soil moisture profile, as expected, yielded the

435 lowest surface runoff. The insensitivity of surface runoff to soil moisture is probably due to the

436   fact that there were no prolonged precipitation periods and that the soil in the basin remained

437   relatively dry which left enough room for infiltration.

438   On the other hand, the assimilation of AMSR-E has a much larger impact on base flow as

439   shown in Figure 8 (lower panel). As mentioned early, Noah uses equation (2) to calculate base

440   flow which has a monotonic relationship with the soil moisture in layer 4. As a result, DA

441   yielded the lowest base flow while DA MassCon generated the largest base flow. Compared to

442   Control, DA MassCon significantly increased base flow in winter months when more corrections

443   were made to the soil moisture fields. Notice that DA generated significantly smaller amounts of

444   base flow in 2007 than in 2006 which is relatively drier. Frequent rainfalls in 2007, which

445   restored bias in the surface, means more water was removed and not captured by DA. Overall,

446   the assimilation results support the findings by *Li et al*. (2009) who concluded that the initial soil

447   moisture condition has a larger impact on base flow while precipitation uncertainty has a larger

448   impact on surface runoff.

449   Additional information beyond surface runoff and base flow is required to compute the

450   stream flow for the watershed. In the western US where significant groundwater recharges may

451   occur, simple summation of base flow with surface runoff will lead to overestimation of stream

452   flow. Figure 9 (upper panel) shows the comparison of simulated total runoff (surface runoff plus

453   base flow) with the USGS stream flow data. The predicted total runoffs are much higher than

454   gauged values, except the result by DA in 2007. Based on the study by *Schaller and Fan* (2009),

455   about 30% of total runoff in the Little Washita area contributes to the stream flow. Using this

456   information, the simulated stream flow, which was taken as 30% of the total runoff, was plotted

457   in the lower panel of Figure 9. The stream flow estimation by Control and DA MassCon now

458   compare reasonably well with the gauge data. The significant overestimation of precipitation by

459   GDAS in August 2006 and the underestimations in June 2007 are noticeable in the predicted

460   stream flow. Bias in forcing data cannot be corrected through soil moisture data assimilation

461   since the forcing was assumed to be unbiased. While these comparisons do not constitute

462   accurate validations (which is why no statistics were calculated for stream flow), they illustrate

463   the potential impact of AMSR-E retrievals on runoff by different algorithms. Note that the

464   estimated monthly total runoff and stream flow in Figure 9 are simple aggregations of the

465   estimated surface and subsurface runoff at all the grid points within the basin. No routing

466   algorithm or time delay was used in producing them, which can be justified given the relatively

467   smaller basin size and the large time scale.

468

469   **4.3 Water budget**

470   As mentioned early, a unique challenge in assimilating remotely sensed data is that the

471   observation is only available for a thin surface layer. An assimilation method, which may look

472   successful based on the verification of soil moisture near the surface, may fail in the lower soil

473   zone. For instance, DA could have been declared a success based on the verification of soil

474   moisture in the shallow root zone and the latent heat. Yet, it degraded soil moisture estimates in

475   the deeper soil profile that led to the deterioration of base flow and failure to show annual

476   precipitation changes in the root zone. Lack of both complete observations and a full set of soil

477   moisture constraints is the root cause for this inconsistent conclusion. To avoid this problem a

478   quality check, independent of any soil moisture measurement, is needed.

479   Water budget checks represent one way to ensure the assimilation results are physically

480   consistent across all the processes. The water budget here is defined as the sum of ET, surface

481   runoff, base flow and the net change in column water, which is essentially the precipitation

482   amount. Since the forcing perturbations were assumed unbiased, the assimilation runs should

483   produce the same water budget as Control in time scales much longer than the perturbation

484   frequency. To assess the overall performance of the two assimilation methods, monthly GDAS

485   precipitation and water budgets from the three simulations are displayed in Figure 10. While the

486   difference between GDAS precipitation and the water budget of Control is due to numerical

487   errors associated with the discretization of the Richards equation, the difference between Control

488   and the two data assimilation runs can only be attributed to the Kalman filters. The failure of

489   DA is clearly evident because it does not have water budget closure in every month. Failure to

490   capture mass loss from the top two layers and the inappropriate update in the lower layers

491   contribute to the loss water budget. On the other hand, DA MassCon, in general, achieved

492   monthly water balance throughout the two-year period. Some ensemble bias still existed in DA

493   MassCon in January and February of 2006 when the soil was so dry that the perturbations used

494   in the filter were probably slightly larger than needed.

495   **5. Discussion**

496   This study demonstrates that modeled soil moisture fields are significantly biased due to

497   errors in static parameters such as soil hydraulic conductivity and inappropriate model physics

498   such as the free drainage condition. Satellite derived soil moisture data, if retrieved

499   appropriately, i.e., less biased, can be used to reduce the bias. However, the difference between

500   the mean of model estimates and that of sensor data can also lead to mass imbalance when the

501   bias is corrected near the surface. In addition, since satellite retrievals only represent information

502   in the top few centimeters of the soil, effectively passing the surface information to the deeper

503   soil layers without causing adverse impacts poses additional challenges.

504   The mass conservation updating scheme developed in this study preserves the water

505   budget of the forecast (as well as the control) by transferring the mass imbalance incurred in

506   updating the top two layers to the lower layers. The development of this scheme was largely

507   based on analyses of model simulation results and considerations of model physics. As reasoned

508   in the result section, the overestimation at the surface is likely caused by the lower than expected

509   hydraulic conductivity values, given the persistent occurrence of overestimation, especially in

510   winter periods when precipitation and ET were very low. For this reason, moving the surface

511   overestimation to the lower layers via mass conservation is to mitigate the inaccuracy of model

512   parameters. If the surface bias was indeed caused by inadequacy in ET and surface runoff

513   algorithms, redistributing mass imbalance back to these fluxes would require direct observations

514   of these variables and extensive knowledge about how each process contributes to the bias (*Pan*

515   *and Wood*, 2006). Precipitation errors could also cause bias in soil moisture. As shown in

516   Figure 10 and stated in section 3, the mass conservation scheme does not allow the water budget

517   to change with the assimilation of the retrievals. While it is tempting to adjust water budget

518   based on surface soil moisture observations, the retrievals alone simply do not provide sufficient

519   information for changing the water budget and may lead to erroneous results. Using Figure 3 as

520   an example, given the consistent positive bias in Control at the surface, if the water budget were

521   allowed to change with soil moisture data assimilation, the assimilation would always lead to

522   reduced water budgets, which would contradict with the precipitation validation in Figure 2

523   where GDAS actually overestimated precipitation or was nearly unbiased in some months. This

524   is part of the reason why DA failed to achieve monthly water balance. With the limited

525   information provided by the surface soil moisture observation, the mass conservation scheme

526   focuses on improving the soil moisture fields first. Regardless of what initiated the bias at the

527   surface, the mass conservation scheme will improve soil moisture estimates in the upper layers if

528   the observations are less biased than the model, but the improvements in the lower layers may be

529   model dependent and region dependent. But with the mass conservation constraint, the

530   assimilated lower soil moisture states are much more reliable (closer to Control) than those given

531   by the conventional EnKF which may yield significantly degraded results due to inappropriate

532   model physics.

533   Although it was found that updating the top two layers is more appropriate at Little

534   Washita, studies in different climate conditions are needed to examine how far the surface

535   measurements can influence the deeper soil layers through the conventional EnKF without

536   causing adverse impacts. A general form of equation (6) for a model with N soil layers and the

537   upper L layers are to be assimilated using the conventional EnKF is:

$$\mathbf{Y}_t^a = \mathbf{Y}_t^f - \sum_{k=1}^{L} (\Delta C_k d_k) / \sum_{j=L+1}^{N} d_j \tag{7}$$

538   and **Y** now contains the soil moisture states for layers L+1 to N. In general, the fewer upper

539   layers that are assimilated using the conventional EnKF, the less impact the assimilation has on

540   the rest of soil layers and fluxes.

541   The difficulty of using the surface observation to improve root zone soil moisture has

542   also been reported by *Walker et al*. (2001) and *Houser et al*. (1998) who showed that soil

543   moisture estimates in the lower soil zone deteriorated with the assimilation of surface

544   observations. As pointed out by *Walker et al*. (2001) that data assimilation could only achieve

545   what model physics is capable of delivering, the failure of the conventional EnKF in updating the

546   lower layers, as shown in this study, is a result of inappropriate model physics. As shown in

547   Figures 3 and 4, Noah (the Control run) failed to capture the trend of increasing wetness with

548   depth as observed by in situ measurements. As a result, the conventional EnKF was not able to

549   yield increments favoring the improvements in the lower profile. Even for assimilation methods

550   that do not depend on model physics such as the least square and variational method (*Kalnay*,

551   2003), there is probably a limit to how far the surface information can be extrapolated to improve

552   the soil moisture state in deeper soil zones. *Houser et al.* (1998) showed that a nudging and a

553   statistical interpolation method also caused similar detrimental effects on the lower soil layer

554   when surface observations were assimilated. Lack of observations in the entire profile to

555   constrain the increments is the root cause for these difficulties. The mass conservation scheme

556   avoids the interference of imperfect model physics for the lower layers by using a model-

557   independent updating equation that also preserves the mass of the forecast.

27

558   The reduction of bias in both upper and lower layers by DA MassCon changed the soil

559   moisture profile which is more aligned with in situ observations than with model physics

560   (Control). As mentioned in the introduction, simulated soil moisture fields from different

561   models exhibit significant disparities which have greatly affected their applications in some

562   areas. *Mo* (2008) showed that the correlations of model-based drought indices are so low in the

563   western US that they are not reliable for drought monitoring. Assimilating the actual value of

564   AMSR-E retrievals into these models can reduce the uncertainty associated with model physics

565   and should lead to more consistent soil moisture fields and thus, more reliable model-based

566   drought indices.

567   With the current framework of LIS, parameter uncertainties are implicitly represented in

568   errors added to soil moisture variables, which is a common practice in many studies (*Reichle et*

569   *al*., 2007; *De Lannoy et al*., 2007a). Alternatively, parameters uncertainty can be represented

570   through directly perturbing parameters (*Margulis et al*., 2002; *Ng et al*., 2009; *Qin et al.*, 2009).

571   The assimilation results should remain similar as they are determined by the relative error of the

572   observation versus that of the model and constrained by the observation and the control run.

573   Perturbing parameters can also be used to simultaneously retrieve model parameters as shown by

574   *Qin et al*. (2009) who retrieved surface soil moisture and soil texture parameters using a particle

575   filtering technique. Their study showed that changes in initial conditions can lead to completely

576   different retrieved parameter values. Lack of constraints in parameters, particularly the

577   knowledge about their mean value, may be responsible for this behavior. For the case studied

578   here, the hydraulic conductivity is likely biased relative to the truth and its uncertainty can hardly

579   be represented by a zero-mean Gaussian process. Bias, in either parameters or state variables, is

580   an important issue that needs to be considered when assimilating real observations.

581   With the biased model shown in this study, the estimates by DA and DA MassCon were

582   not optimized, i.e., the estimation error was not minimized (*Kalnay*, 2003). The same is true if

583   the retrievals are scaled, priori to assimilation, using model climatology (*Reichle and Koster,*

584   *2004; Drusch et al., 2005*) because the model estimates (Control) were still biased. For the

585   example presented here, more reductions in the estimation error for the surface layer can be

586   obtained by directly inserting the AMSR-E retrievals into the model, as Table 3 shows that

587   AMSR-E retrievals have the smallest bias against the ARS measurements. However, retrievals

588   may not always be better than modeled estimates, in which case data assimilation will yield

589   better estimates than direct insertion. In addition, direct insertion is not as effective as an EnKF

590   in reducing estimation errors in root zone soil moisture because data assimilation techniques can

591   force the surface observation to impact the adjacent soil layer while direction insertion, relying

592   on model physics, may not be effective in passing the information downward (*Crow and Wood,*

593   2003).

594

601

602

603

References

605 Bolten, J., W.T. Crow, X. Zhan, C. Reynolds and T.J. Jackson, 2008: Assimilation of a satellite-

606 based soil moisture product in a two-layer water balance model for a global crop production

607 decision support system, *Data Assimilation for Atmospheric, Oceanic, and Hydrologic*

608 *Applications*, S.K. Park (Ed.), Springer-Verlag, 449-465, doi:10.1007/978-3-540-71056-1.

609 Campbell, G. S., 1974: A simple method for determining unsaturated conductivity from moisture

610 retention data, *Soil Sci*., **117**, 311-314.

611 Chen, F., K. Mitchell, J. Schaake, Y. Xue, H. Pan, V. Koren, Q. Duan, M. Ek, and A. Betts,

612 1996: Modeling of land surface evaporation by four schemes and comparison with 1GFIFE

613 observations, *J. Geophys. Res*., **101**(D3), 7251-7268.

614 Cosby, B. J., G. M. Hornberger, R. B. Clapp, and T. R. Ginn, 1984: A statistical exploration of

615 the relationships of soil moisture characteristics to the physical properties of soils, *Water Resour.*

616 *Res*., **20**(6), 682-690.

617 Cosgrove, B. A., D. Lohmann, K. E. Mitchell, P. R. Houser, E. Wood, J. Schaake, A. Robock, C.

618 Marshall, J. Sheffield, Q. Duan, L. Luo, R. Higgins, R. Pinker, D. Tarpley, and J. Meng, 2003a:

619 Real-time and retrospective forcing in the North American Land Data Assimilation

620 System(NLDAS) project, *J. Geophys. Res*., **108**(D22), 8842, doi:10.1029/2002JD003118.

621

622 Cosgrove, B. A., D. Lohman, K. E. Mitchell, P. R. Houser, E. F. Wood, J. C. Schaake, A.

623 Robock, J. Sheffield, Q. Duan, L. Luo, R. W. Higgins, R. T. Pinker and J. D. Tarpley, 2003b:

624 Land surface model spin-up behavior in the North American Land Data Assimilation System

625 (NLDAS), *J. Geophys. Res.*, **108**(D22), 8845, doi:10.1029/2002JD003316.

626

627 Crow, W. and E. Wood, 2003: The assimilation of remotely sensed soil brightness temperature

628 imagery into a land surface model using Ensemble Kalman Filtering: a case study based on

629 ESTAR measurements during GSP97, *Adv. Water Res*, **26**, 137-149.

630 Crow, W.T., and X. Zhan, 2007: Continental-scale evaluation of remotely sensed soil moisture

631 products, *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*, **4**(3), 451-455.

632 De Lannoy, G. J. M., R. H. Reichle, P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest,

633 2007a: Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter,

634 *Water Resour. Res.*, **43**, W09410, doi:10.1029/2006WR005449.

635 De Lannoy, G.J.M**.**, Reichle R.H., Houser P.R., Pauwels V.R.N., Verhoest N.E.C., 2007b:

636 Correction to `Correcting for forecast bias in soil moisture assimilation with the Ensemble

637 Kalman Filter', *Water Resources Research*, **43**(10), W10799, doi:10.1029/2007WR006542.

638 Derber, J. C., D. F. Parrish, and S. J. Lord, 1991: The new global operational analysis system at

639 the National Meteorological Center, *Wea. Forecasting,* **6,** 538–547.

640

641   Dirmeyer, P. A., 2000: Using a Global Soil Wetness Dataset to Improve Seasonal Climate

642   Simulation, *J. Climate*, **13**, 2900-2922.

643

644   Draper, C. S., J.-F. Mahfouf, and J. P. Walker, 2009: An EKF assimilation of AMSR-E soil

645   moisture into the ISBA land surface scheme, *J. Geophys. Res*., **114**, D20104,

646   doi:10.1029/2008JD011650.

647   Drusch, M., E. F. Wood, and H. Gao (2005), Observation operators for the direct assimilation of

648   TRMM microwave imager retrieved soil moisture, *Geophys. Res. Lett*., **32**, L15403,

649   doi:10.1029/2005GL023623.

650   Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D.

651   Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for

652   Environmental Prediction operational mesoscale Eta model, *J. Geophys. Res*., **108**(D22),

653   doi:10.1029/2002JD003296.

654   Evensen, G. and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas

655   current using the ensemble Kalman filter with a quasigeostrophic model, *Monthly Weather*

656   *Review*, **124**, 85-96.

657   Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed, and D. C.

658   Goodrich, 1998: Integration of soil moisture remote sensing and hyrologic modeling using data

659   assimilation, *Water Resour. Res*., **34**(12), 3405-3420.

660   Jackson, T.J., 1993: Measuring surface soil moisture using passive microwave remote sensing,

661   *Hydrol. Process.*, **7**(2), 139–152.

662   Jury, W. A., W. R. Gardner, W. H. Gardner, 1991: *Soil Physics*, John Wiley & Sons, 328 pp.

663

664   Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge

665   University Press, 341 pp.

666

667   Keppenne, C. L., 2000: Data Assimilation into a Primitive-Equation Model with a Parallel

668   Ensemble Kalman Filter, *Mon. Wea. Rev.*, **128**, 1971-1981.

669

670   Keppenne, C. L., M. M. Rienecker, N. P. Kurkowski, and D. A. Adamec, 2005: Ensemble

671   Kalman filter assimilation of temperature and altimeter data with bias correction and application

672   to seasonal prediction, *Nonlinear Processes in Geophysics*, **12**, 491–503.

673

674   Kumar, S. V., C. D. Peters-Lidard, Y. Tian, P. R. Houser, J. Geiger, S. Olden, L. Lighty, J. L.

675   Eastman, B. Doty, P. Dirmeyer, J. Adams, K. Mitchell, E. F. Wood and J. Sheffield, 2006: Land

676   Information System - An Interoperable Framework for High Resolution Land Surface Modeling.

677   *Environmental Modelling & Software,* Vol. **21**, 1402-1415.

678

679   Kumar, S.V., R.H. Reichle, C. D. Peters-Lidard, R. D. Koster, X. Zhan, W. T. Crow, J. B.

680   Eylander, and P. R. Houser, 2008: A Land Surface Data Assimilation Framework using the Land

681   Information System: Description and Applications, *Advances in Water Resources*, **31**(11), 1419-

682   1432.

683

684 Li, H., L. Luo, E. F. Wood, and J. Schaake, 2009: The role of initial conditions and forcing

685 uncertainties in seasonal hydrologic forecasting, *J. Geophys. Res*., **114**, D04114,

686 doi:10.1029/2008JD010969.

687

688 Li, B. and M. Rodell, 2011: Spatial variability and its scale dependency of observed and modeled

689 soil moisture fields, submitted to *Water Resour. Res*.

690 Margulis, S. A, D. McLaughlin, D. Entekhabi, and S. Dunne, 2002: Land data assimilation and

691 estimation of soil moisture using measurements from the Southern Great Plains 1997 Field

692 Experiment, *Water Resour. Res*., **38**(12), 1299, doi:10.1029/2001WR001114.

693

694 Mitchell, K. E., D. Lohmann, P. R. Houser, E. F. Wood, J. C. Schaake, A. Robock, B. A.

695 Cosgrove, J. Sheffield, Q. Duan, L. Luo, R. W. Higgins, R. T. Pinker, J.D. Tarpley, D. P.

696 Lettenmaier, C. H. Marshall, J. K. Entin, M. Pan, W. Shi, V. Koren, J. Meng, B. H. Ramsay, and

697 A. A. Bailey, 2004: The multi-institution North American Land Data Assimilation System

698 (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed

699 hydrological modeling system, *J. Geophys. Res*., **109**, D07S90, doi:10.1029/2003JD003823.

700 Mo, K. C., 2008: Model-based drought indices over the United States, *J. Hydrometeor*., **9**, 1212-
701 1230.

702

703 Ng, G.-H. C., D. McLaughlin, D. Entekhabi, and B. Scanlon, 2009: Using data assimilation to

704 identify diffuse recharge mechanisms from chemical and physical data in the unsaturated zone,

705 Water Resour. Res., 45, W09409, doi:10.1029/2009WR007831.

706

707 Ni-Meister, W., P. R. Houser, and J. P. Walker, 2006: Soil moisture initialization for climate

708 prediction: Assimilation of scanning multifrequency microwave radiometer soil moisture data

709 into a land surface model, *J. Geophys. Res.*, **111**, D20102, doi:10.1029/2006JD007190.

710

711 Njoku, E. G., T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem, 2003: Soil moisture

712 retrieval from AMSR-E, *IEEE Trans. Geosci. Rem. Sens*., **41**, pp. 215-229.

713

714 Pan, M. and E. F. Wood, 2006: Data assimilation for estimating the terrestrial water budget using

715 a constrained ensemble Kalman filter, *J. Hydrometeor*., **7**(3), 534-547.

716

717 Qin, J., S. Liang, K. Yang, I. Kaihotsu, R. Liu, and T. Koike, 2009: Simultaneous estimation of

718 both soil moisture and model parameters using particle filtering method through the assimilation

719 of microwave signal, *J. Geophys. Res*., **114**, D15103, doi:10.1029/2008JD011358.

720

721 Reichle, R. H. and R. Koster, 2002: Land data assimilation with the ensemble Kalman filter:

722 assessing model error parameters using innovations, *Proceedings of XIV International*

723 *Conference on Computational methods in water resources*, Delft, Neitherlands.

724

725   Reichle, R. H., and R. Koster (2004), Bias reduction in short records of satellite soil moisture,

726   *Geophys. Res. Lett*., **31**, L19501, doi:10.1029/2004GL020938.

727

728   Reichle, R. H., R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe, 2007:

729   Comparison and assimilation of global soil moisture retrievals from the Advanced Microware

730   Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel

731   Microwave Radiometer (SMMR), *J. Geophy. Res.*, **112**, D09108, doi:10.1029/2006JD008033.

732

733   Rodell, M., P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B.

734   Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, D. Lohmann, and D. Toll,

735   2004: The Global Land Data Assimilation System, *Bull. Amer. Meteor. Soc*., **85** (3), 381394.

736

737   Rodell, M., P.R. Houser, A.A. Berg, and J.S. Famiglietti, 2005: Evaluation of 10 Methods for

738   Initializing a Land Surface Model, *J. Hydromet.*, **6**(2), 146-155.

739   Ryu, D., W.T. Crow, X. Zhan and T.J. Jackson, 2009: Correcting unintended perturbation biases

740   in hydrologic data assimilation using Ensemble Kalman filter, *Journal of Hydrometeorology,*

741   **10**(3), 734-750, doi: 10.1175/2008JHM1038.1.

742

743 Schaake, J. C., V. Koren, Q. Duan, K. Mitchell, and F. Chen, 1996: Simple water balance model

744 for estimating runoff at different spatial and temporal scales, *J. Geophys. Res.*, **101**(D3), 7461-

745 7475.

746

747 Schaefer, G. L, M. H. Cosh, and T. J. Jackson, 2007: The USDA Natural Resources

748 Conservation Service Soil Climate Analysis Network (SCAN), *J. of Atmospheric and Oceanic*

749 *Technology*, **24**(12), 2073-2077.

750

751 Schaller, M. F., and Y. Fan, 2009: River basins as groundwater exporters and importers:

752 Implications for water cycle and climate modeling, *J. Geophy. Res.*, **114**, D04103,

753 doi:10.1029/2008JD010636.

754

755 Walker, J. P., G. R. Willgoose, and J. D. Kalma, 2001: One-dimensional soil moisture profile

756 retrieval by assimilation of near-surface measurements: a simplified soil moisture model and

757 field application, *J. Hydrometeor.*, **2**, 356-373.

758

759 Wood, E. F. and coauthors, 1998: The Project for Intercomparison of Land-surface

760 Parameterization Schemes _PILPS/ Phase 2_c/ Red–Arkansas River basin experiment:

761 1. Experiment description and summary intercomparisons, *Global and Planetary Change*, **19**.

762 115–135.

763   Zeng, X. and M. Decker, 2009: Improving the numerical solution of soil moisture-based

764   Richards equation for land models with a deep or shallow water table, *J. Hydrometeor.*, **10**, 308-

765   319.

766   Zhan, X., J. Liu, T.J. Jackson, J. Meng, F. Weng, K. Mitchell, 2008: Merging satellite soil

767   moisture retrievals and model simulations for a blended consistent operational global soil

768   moisture data product. *SPIE Annual Conference on Atmospheric and Environmental Remote*

769   *Sensing Data Processing and Utilization IV: Readiness for GEOSS II*. San Diego, CA. 10-14

770   August 2008.

771
772
773

Figures

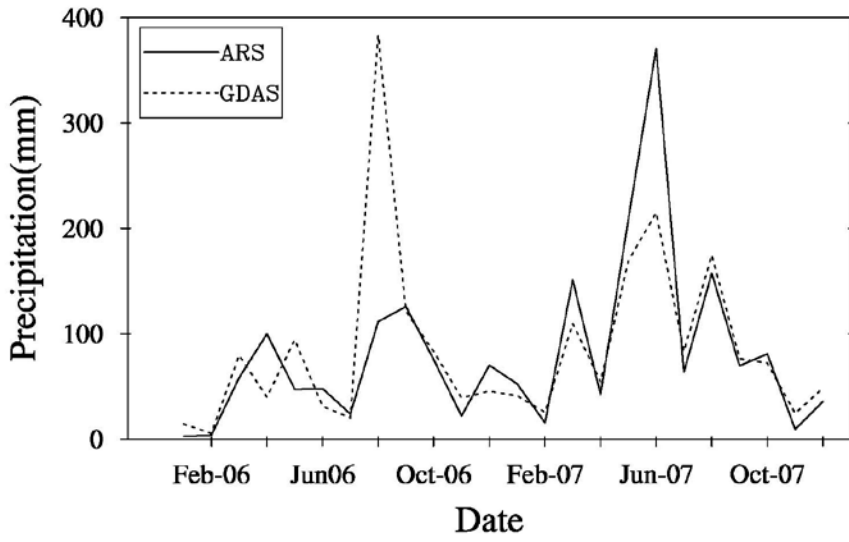Figure 1: The Little Washita watershed and locations of ARS, SCAN and USGS stations.

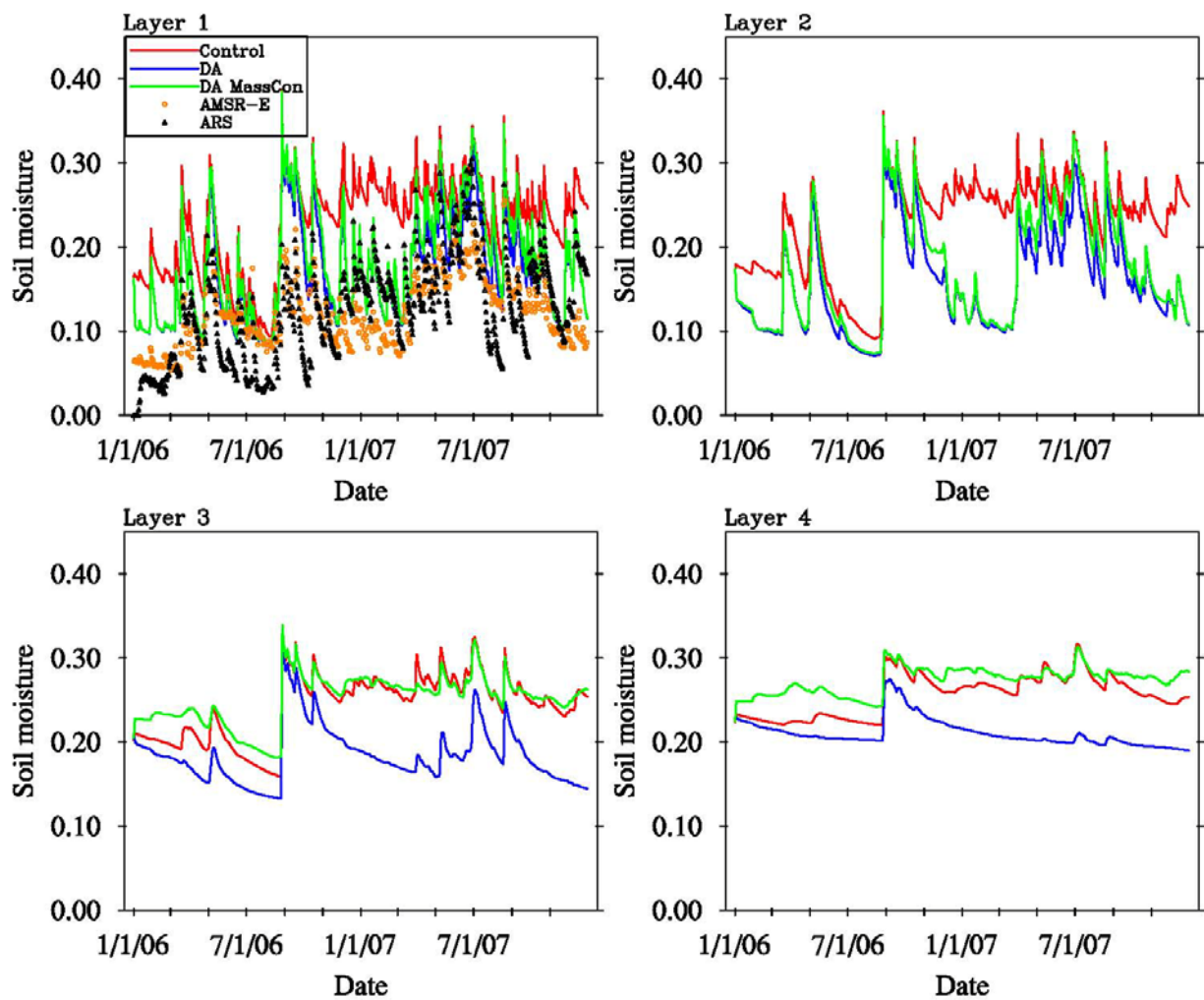Figure 2: Basin averaged monthly ARS and GDAS precipitation.

Figure 3: Time series of basin averaged daily soil moisture from Control, DA, and DA MassCon

for Noah soil layers 1 to 4. Simulated soil moisture at layer 1 is compared to basin averaged

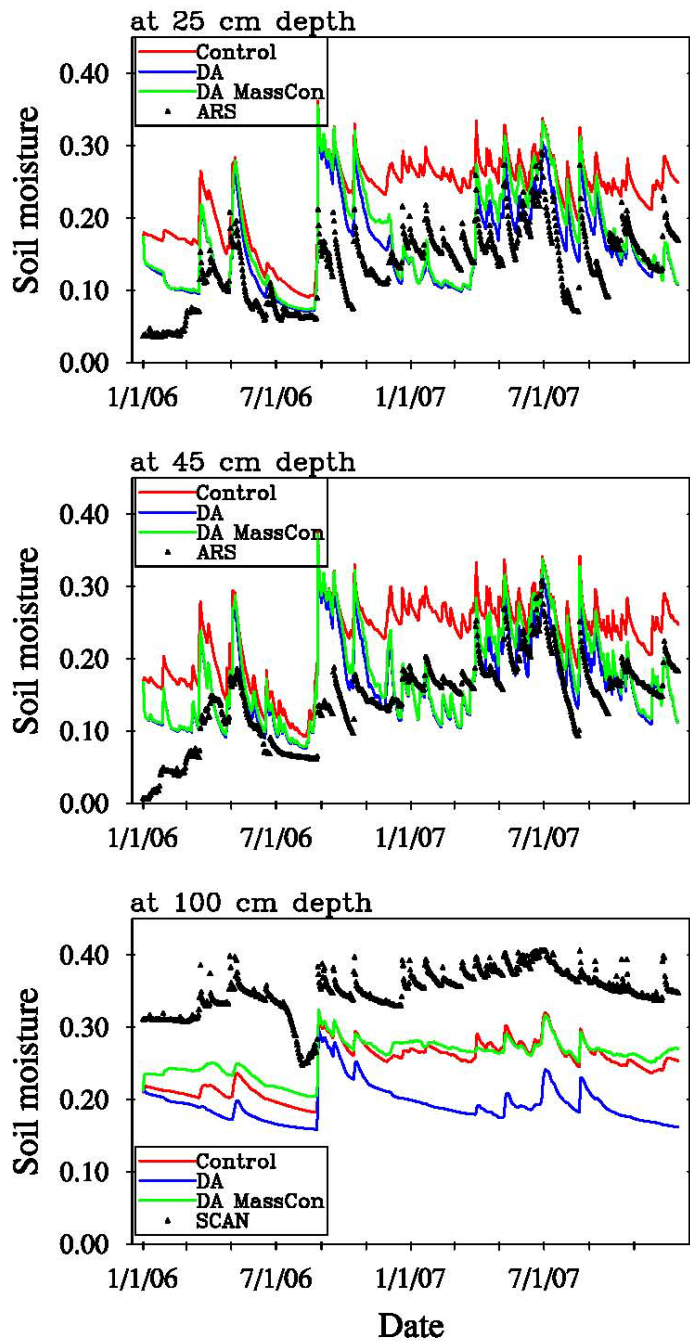ARS measurements at the 5 cm depth and the AMSR-E retrievals.

Figure 4: Comparison of basin averaged daily soil moisture from Control, DA, and DA

MassCon, interpolated at 25, 45 and 100 cm depths, with measurements from ARS stations and

the SCAN site.

Figure 5: Mean annual surface soil moisture (soil layer 1) from Control, DA and DA MassCon in

2006 and 2007.

Figure 6: Mean root zone (upper 100 cm) soil moisture from Control, DA and DA MassCon in

2006 and 2007.

Figure 7: Comparison of daily latent heat from Control, DA, and DA MassCon versus the SGP

flux data. The daily latent heat estimates are averaged values from 6 am to 6 pm local time for

both the SGP measurements and Noah estimates.

Figure 8: Monthly surface and base flow (mm) from Control, DA and DA MassCon.

Figure 9: Comparison of monthly total runoff and stream flow from Control, DA and DA

MassCon versus USGS gauge data.

Figure 10: Monthly GDAS precipitation and water budget for Control, DA and DA MassCon.

798

799

800

801
802

803

804   Figure 1: The Little Washita watershed and the locations of ARS, SCAN and USGS stations.

805

806

807

808   Figure 2: Basin averaged monthly ARS and GDAS precipitation.
809
810

811
812 Figure 3: Time series of basin averaged daily soil moisture from Control, DA, and DA MassCon
813 for Noah soil layers 1 to 4. Simulated soil moisture at layer 1 is compared to basin averaged
814 ARS measurements at the 5 cm depth and the AMSR-E retrievals.

815
816 Figure 4: Comparison of basin averaged daily soil moisture from Control, DA, and DA
817 MassCon, interpolated at 25, 45 and 100 cm depths, with measurements from ARS stations and
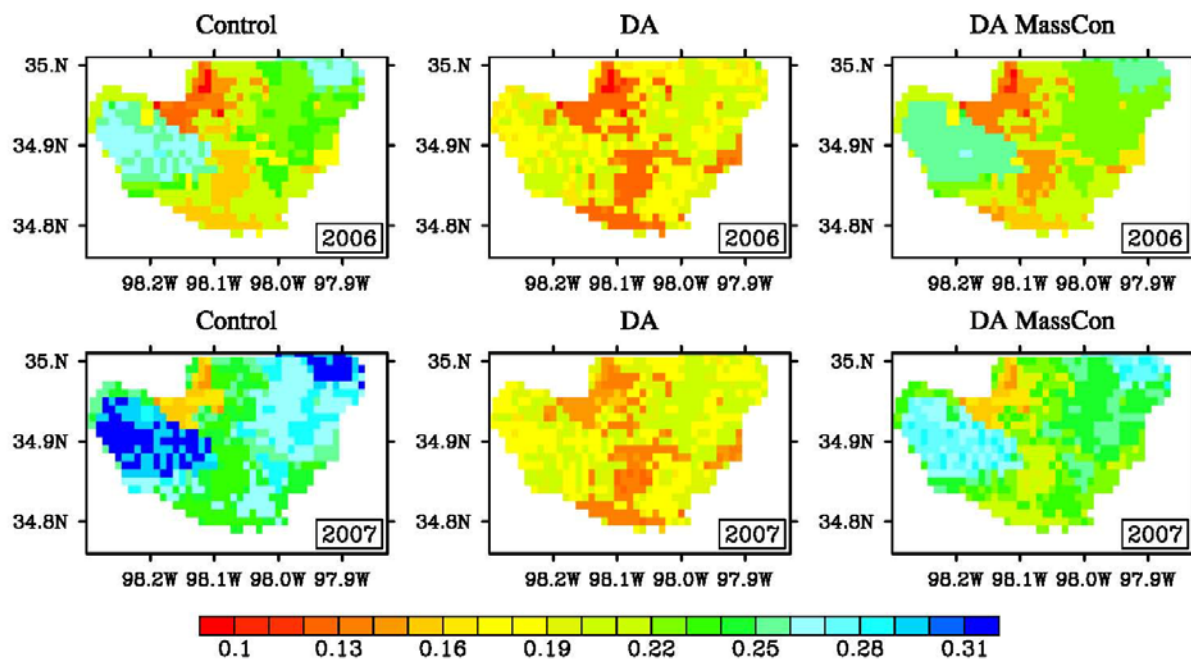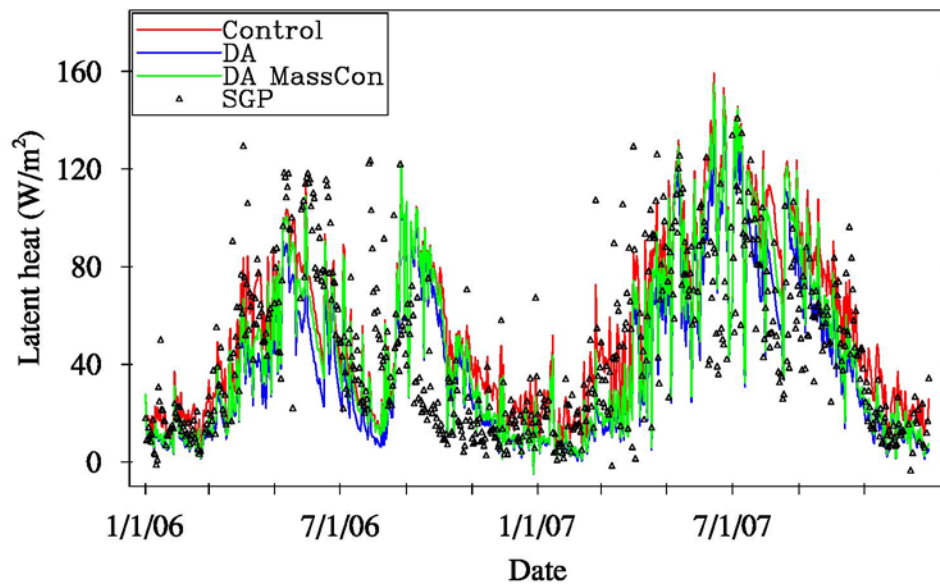818 the SCAN site.
819

820

821

Figure 5: Mean annual surface soil moisture (soil layer 1) from Control, DA and DA MassCon in 2006 and 2007.

822

823

824

825

826

827

828

829

830

831

832

833

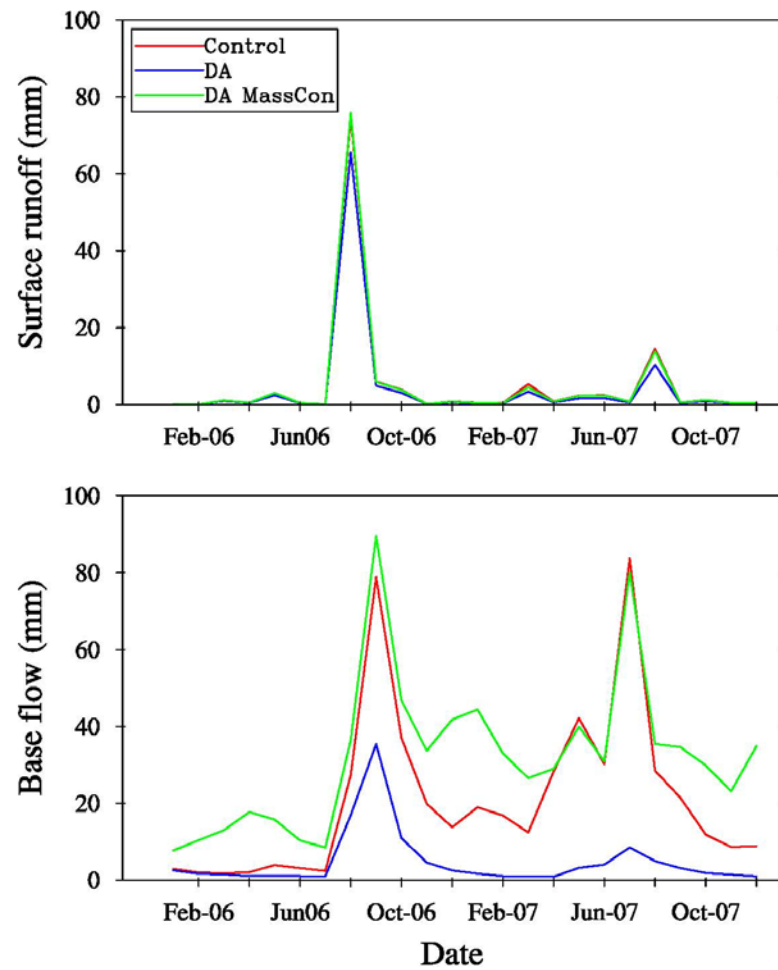Figure 6: Mean root zone (upper 100 cm) soil moisture from Control, DA and DA MassCon in 2006 and 2007.

838

839

Figure 7: Comparison of daily latent heat estimates from Control, DA, and DA MassCon versus
the SGP flux data. The daily latent heat values are averaged values from 6 am to 6 pm local time
for both the SGP measurements and Noah estimates.

843

844

845
846

847
848
849   Figure 8: Monthly surface and base flow (mm) from Control, DA and DA MassCon.
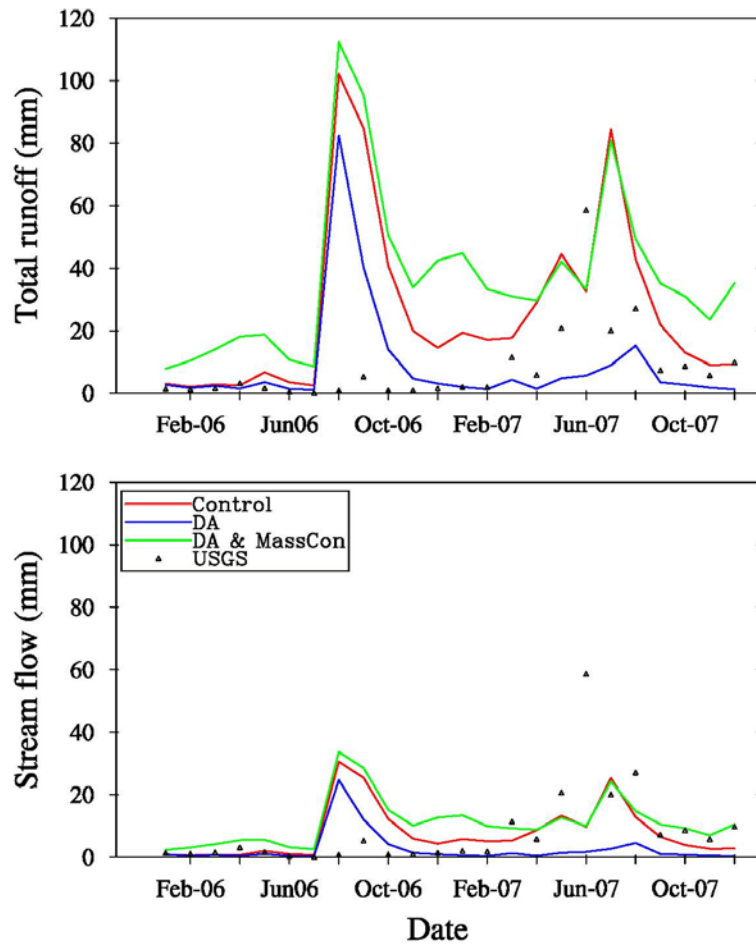850

851

852   Figure 9: Comparison of monthly total runoff and stream flow from Control, DA and DA
853   MassCon versus USGS gauge data.
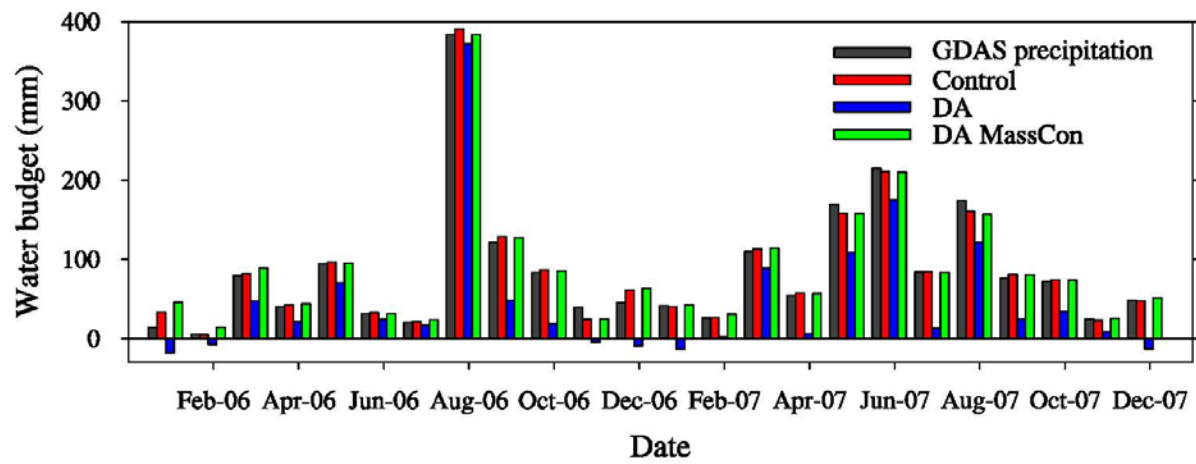
854

855

856
857

858

859

Figure 10: Monthly GDAS precipitation and water budget for Control, DA and DA MassCon.

861

862

863

864 **Tables:**
865

866 Table 1: Annual precipitation (mm) at Little Washita in 2006 and 2007 by ARS and GDAS.

867 Table 2: Perturbations given to the four soil moisture variables and their cross-correlations (the
868 last four columns).
869

870 Table 3: Basin averaged bias and root mean square error (rmse) of daily simulated soil moisture

871 at the 5, 25, 45 and 100 cm depths and latent heat for the two-year period. Statistics were

872 calculated with respect to daily values of ground measurements at ARS, SCAN and SGP.

873
874
875

876

877   Table 1: Annual ARS and GDAS precipitation (mm) at Little Washita in 2006 and 2007.

	2006	2007
ARS	690	1259
GDAS	959	1096

878

879

880

881   Table 2. Perturbations given to the four soil moisture variables and their cross-correlations (the
882   last four columns).
883

variable	standard deviation	correlation in time	cross-correlation with θ_1	cross-correlation with θ_2	cross correlation with θ_3	cross correlation with θ_4
θ_1	0.002	12 h	1.0	0.6	0.4	0.1
θ_2	5.0e-4	12 h	0.6	1.0	0.5	0.2
θ_3	6.0e-5	12 h	0.4	0.5	1.0	0.4
θ_4	6.0e-6	12 h	0.1	0.2	0.4	1.0

884

885

886 Table 3: Basin averaged bias and root mean square error (rmse) of daily simulated soil moisture
887 at the 5, 25, 45 and 100 cm depths, latent heat (W/m^2) and NOAA NESDIS AMSR-E soil
888 moisture retrievals for the two-year period. Statistics were calculated with respect to daily
889 values of ground measurements at ARS, SCAN and SGP.

	Control		DA		DA MassCon		AMSR-E	
	bias	rmse	bias	rmse	bias	rmse	bias	rmse
soil moisture (5 cm)	0.11	0.11	0.05	0.07	0.06	0.08	-0.002	0.05
soil moisture (25 cm)	0.10	0.10	0.03	0.06	0.04	0.07	-	-
soil moisture (45 cm)	0.09	0.09	0.02	0.05	0.03	0.06	-	-
soil moisture (100 cm)	-0.10	0.11	-0.16	0.16	-0.09	0.09	-	-
latent heat	8.27	28.87	-5.76	28.80	-0.33	28.38	-	-

890