Land data assimilation and seasonal climate prediction

Rolf H. Reichle[1,2] and Randal D. Koster[1]

[1]*NASA Goddard Space Flight Center, Global Modeling and Assimilation Office, Greenbelt, MD, USA.*

[2]*University of Maryland, Baltimore County, Goddard Earth Sciences and Technology Center, Baltimore, MD, USA*
(reichle@gmao.gsfc.nasa.gov)

1. Introduction

Sub-seasonal and seasonal prediction of summer precipitation and screen-level temperature over mid-latitude land must rely on predictability associated with the slowly varying land boundary conditions, in particular root zone soil moisture. Our goal is therefore to derive a realistic representation of global land surface conditions in order to initialize the General Circulation Model that produces the seasonal forecast. This may be partly achieved by forcing a land surface model with observed fields of precipitation and radiation (rather than with the typically poor surface meteorology produced by the GCM) up to the start time of the forecast, and transforming the resulting land surface initial conditions to be consistent with the GCM's climatology for forecast initialization. Such a system is now in place for the seasonal forecasts produced at the NASA Global Modeling and Assimilation Office (GMAO).

In addition to observations of land surface forcing fields (such as precipitation and radiation), there are limited satellite observations of land surface states, notably surface soil moisture. Such state observations can be retrieved from satellite measurements of microwave radiation emitted by the land surface. Land data assimilation techniques can be used to merge these satellite observations of land surface states with model-derived states, where the latter incorporate the information contained in the observed land surface forcings as well as our best knowledge of land surface dynamics as formulated in the land model. Estimates of land surface conditions produced by land data assimilation should, if derived properly, optimally combine all available information. In practice, strong biases and large uncertainties in models and observations pose severe challenges to our ability to derive such estimates. We demonstrate that despite the complications, estimates derived from data assimilation can improve our knowledge of soil moisture (*Reichle and Koster*, 2004b) and have the potential to improve seasonal forecasts.

2. Sub-seasonal climate prediction over land

In the tropics, skill in seasonal forecasting derives primarily from observations and predictions of sea surface temperature (SST) and their feedback onto the atmosphere. Over mid-latitude land, SST forcing has limited impact on predictions of summer precipitation and surface temperature, but some predictability at sub-seasonal to seasonal time scales may derive from slowly varying land boundary conditions. In particular, the memory associated with root zone soil moisture, in combination with feedback of soil moisture on precipitation through evaporation, may provide predictability at time scales from 2 weeks to 2 months (*Koster et al.*, 2004a; *Koster et al.*, 2004b).

At the NASA-GMAO, near real-time seasonal forecasts are routinely produced with the GMAO CGCMv1, a fully coupled atmosphere-ocean-land general circulation model. Since April 2004, we use realistic soil moisture and ground temperature initial conditions that are derived off-line using observed surface meteorological forcing data from the Global Land Data Assimilation System (GLDAS) (*Rodell et al.*, 2003) for all 18 ensemble members. Also since April 2004, 12 out of 18 ensemble members are initialized with anomalies derived from atmospheric data assimilation systems at the National Center for Environmental Prediction (NCEP). The initial atmospheric conditions of the remaining 6 ensemble members are taken from "AMIP"-style integrations that rely on SST only.

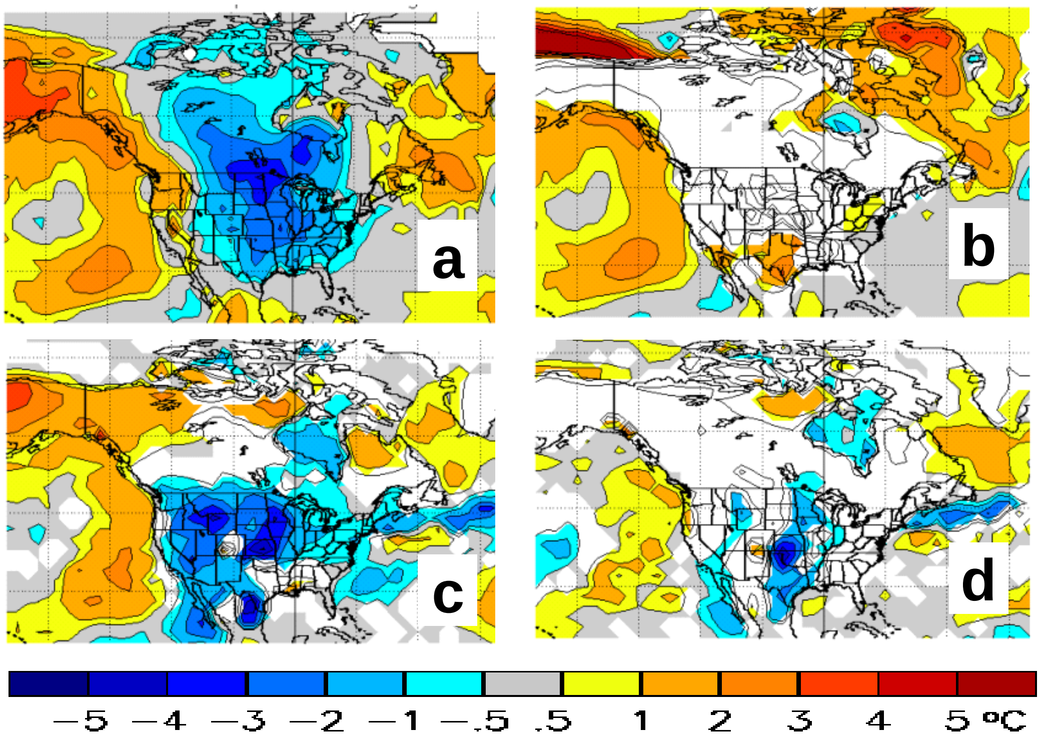

Figure 1 Screen-level air temperature for Aug 2004 in North America: (a) CAMS observations (b) 9-member ensemble mean "AMIP" integration based on SST only (c) 18-member ensemble mean forecast initialized Aug 1, 2004 with anomalies from GLDAS land and NCEP atmosphere (d) same as (c) but initialized Jul 1, 2004.

Figure 1 shows one- and two-month forecasts of screen-level temperature over North America for August 2004, along with validating observations from the Climate Analysis and Monitoring System (CAMS) at the Climate Prediction Center. For comparison, a "forecast" that is based solely on SST is also shown. It is taken out of a long-term "AMIP"-style integration of a 9-member ensemble that uses observed SST at the ocean-atmosphere boundary but contains no observational information for the land or the atmosphere. Because of the small number of real-time forecasts produced this far, it is not yet possible to validate the skill in a statistically meaningful way. However, Figure 1 clearly demonstrates the strong impact on the forecast of initializing the land surface (and the atmosphere) with more realistic land surface conditions when compared to the old system that relied solely on SST. The more statistically complete hindcast experiment of *Koster et al.* (2004b) suggests increased skill associated with land initialization, skill that is independent of that from atmospheric initialization.

3. Soil moisture data assimilation

While there has been considerable progress in the methodological development of soil moisture data assimilation (*Walker and Houser*, 2001; *Margulis*, 2002; *Reichle et al.*, 2002; *Crow and Wood*, 2003, *Seuffert et al.*, 2003), there is little experience with the assimilation of a multi-year, global dataset of surface soil moisture retrievals. Here, we assimilate global soil moisture retrievals from the Scanning Multichannel Microwave Radiometer (SMMR) into the NASA Catchment land surface model (*Koster et al.*, 2000) for the period 1979-87. Through validation against ground measurements, we demonstrate that assimilation of

SMMR data yields improved soil moisture estimates – better than those obtained with the model or from the satellite alone.

Owe et al. (2001) recently developed a novel retrieval algorithm for soil moisture from passive microwave measurements and produced a nine-year, global soil moisture dataset from Scanning Multichannel Microwave Radiometer (SMMR) observations for the period 1978-87 (*De Jeu*, 2003). For the period 1979-93, *Berg et al.* (2003) developed a high-quality, global dataset of surface meteorological fields based on reanalysis data and corrected with observations as much as possible. This dataset is used to force the NASA Catchment land surface model. Finally, ground-based soil moisture data for the SMMR time period are available for select locations in Eurasia and North America from the Global Soil Moisture Data Bank (GSMDB) (*Robock et al.*, 2000).

These satellite, ground-based, and model soil moisture data are independent, and each has its own set of limitations. State-of-the-art land surface models produce widely different soil moisture output even when integrated with identical meteorological forcing inputs (*Entin et al.*, 1999). Errors in C-band surface soil moisture retrievals are generally high, and modest amounts of vegetation obscure the soil moisture signal, which in any case is limited to the top centimeter of the soil. Ground-based measurements – used for validation – are sparse and not necessarily representative of large-scale soil moisture. At this time, errors in global soil moisture observation and modeling are so large that there is no universally accepted climatology (*Reichle et al.*, 2004). Consequently, we scale the satellite observations to the model's climatology before assimilating the data into the model (*Reichle and Koster*, 2004a). For seasonal climate prediction, knowledge of soil moisture anomalies is, in any case, more important than knowledge of absolute soil moisture.

In a data assimilation system, the model-generated soil moisture is corrected toward the observational estimate, with the degree of correction determined by the levels of error associated with each. The assimilation system used here is based on the Ensemble Kalman filter (EnKF) (*Reichle et al.*, 2002). The EnKF is well suited to the nonlinear and intermittent character of land surface processes. The key feature of the EnKF is that error estimates of the model-generated results are dynamically derived from an ensemble of model integrations. Each member of the ensemble experiences slightly perturbed instances of the observed precipitation fields (representing errors in the precipitation data) and is also subject to randomly generated noise that is directly added to the soil moisture states (representing errors in model physics and parameters). In this paper, we use the one-dimensional version of the EnKF. Preliminary results with the three-dimensional EnKF (*Reichle and Koster*, 2003) show a further improvement in surface soil moisture but also some deficiencies in root zone soil moisture. Calibration of the latter is very complex and work is still in progress.

In the next section, we analyze "raw" time series of monthly mean soil moisture as well as anomaly time series. The latter are obtained by subtracting the monthly climatology of each dataset (i.e., the average for each calendar month) from the raw time series. In other words, the raw time series include the seasonal cycle, while the anomaly time series describe only deviations from the average seasonal cycle. Our analysis will be focused on time series correlations between the various data sets rather than on root-mean-square errors, because there is not enough evidence about which climatology is more correct (*Reichle et al.*, 2004).

4. Results from global assimilation of SMMR soil moisture retrievals

Clues about the global performance of the assimilation algorithm can be extracted from its innovations sequence (the difference between SMMR retrievals and their corresponding model forecasts during the assimilation integration). If the filter is operating according to its underlying assumptions – that various linearizations hold, and that model and observation errors are uncorrelated and normally distributed – the sum of the model error covariance (diagnosed from the ensemble spread) and the measurement error covariance should equal the sample covariance of the innovations sequence. In other words, we can easily check the assumptions underlying the assimilation process by checking whether the innovations sequence has the expected mean and variance (*Reichle et al.*, 2002).

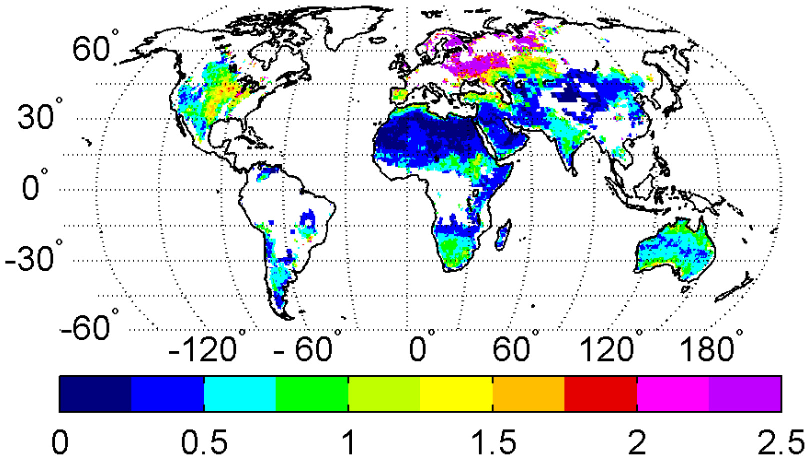

Figure 2 Variance of normalized innovations [-] (from Reichle and Koster, *2004b).*

Because of the bias reduction applied before the assimilation, the mean of the innovations is statistically indistinguishable from zero. A supplemental analysis shows that not scaling the SMMR data a priori leads to a mean that is about one standard deviation away from the expected mean of zero. This provides further evidence of the absolute necessity of including bias removal as part of the assimilation system. Next, Figure 2 shows global maps of the variance of the innovations sequence after normalization with its expected standard deviation. The global average variance of the normalized innovations sequence is around 0.7; it thus falls short of the expected unit variance. Moreover, there are strong variations across the globe. The innovations variance slightly exceeds one in the eastern half of North America, and it is closer to two in mid-latitude Eurasia. For the rest of the globe, the innovations show too small a variance. The imperfect variance is explained in part by nonlinearities in the model and in the observation operator. It also relates, however, to an imperfect representation of the model error characteristics in the ensemble generation. It is probably not a coincidence that the variance deficiency is prominent in relatively dry climates, and that excess variance is found in wetter climates. Therefore, it might be possible to use the innovations variance to tune filter parameters (such as model error variances) before repeating the assimilation integration. Alternatively, adaptive tuning methods could be tried (*Dee*, 1995).

Without any such tuning of the filter, we will now show that the assimilation of SMMR retrievals already yields modest but significant improvements in the estimation of soil moisture. For this validation, we use in situ observations from up to 77 locations in North America and Eurasia that have sufficient GSMDB and SMMR data for our analysis (*Reichle et al.*, 2004). First, Figure 3 shows the 1979-87 average seasonal cycles for surface and root zone soil moisture at one representative location in Illinois. For clarity, we adjusted the annual mean of the SMMR, model, and assimilation data to match the annual mean of the GSMDB data. At this location, the phase of the model data lags the phase of the ground data by about one month. The SMMR data, on the other hand, show a better phase agreement with the ground data than the model, although the SMMR data are not available year-round because the vegetation is too dense in the summer months. SMMR data are also only available for the surface layer. The assimilation of just a few months of SMMR surface soil moisture retrievals per year shifts the spring dry-down and fall wet-up by about one half month towards the phase of the annual cycle of the ground data. Most importantly, this improvement applies equally to the root zone.

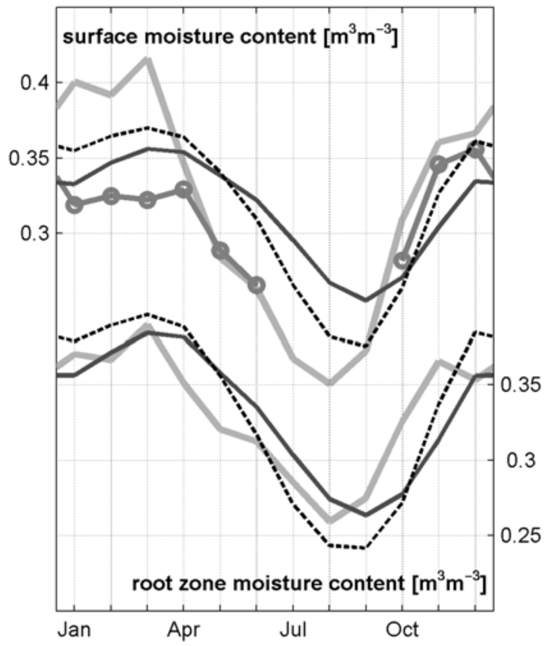

Figure 3 Time-average seasonal cycle of (Top group of lines, left axis) surface and (Bottom group of lines, right axis) root zone soil moisture at a representative location in Illinois (89.5W, 38.6N): (Light gray) GSMDB, (Dark grey with circles) SMMR, (Black solid) model, (Black dashed) EnKF (from Reichle and Koster, 2004b).

Table 1 provides a stronger, global-scale demonstration of improvement associated with assimilation. Listed are time series correlation coefficients (including 95% confidence intervals) computed from monthly mean time series and averaged over all locations with sufficient data in North America and Eurasia. For surface soil moisture, the satellite and model data show about the same skill in reproducing the in situ data, with correlation coefficients of 0.44 and 0.43, respectively (0.32 and 0.36, respectively, for anomalies.) Merging the SMMR retrievals with the model through data assimilation leads to a statistically significant increase in the correlation coefficients to 0.50 for surface soil moisture (0.43 for anomalies). Note that even if the assimilation data were perfect, correlations could still be much less than one due to the mismatch of scale between the assimilation data and the GSMDB data. In other words, the seemingly modest increase of correlation to 0.50 could be quite large relative to the maximum increase possible given the point-scale character of the validation data. In any case, the increases seen are statistically significant, suggesting that the satellite and model data contain some independent information that the assimilation algorithm is able to combine into superior estimates.

	N	SMMR	Model	EnKF
Sfmc	77	.44 ± .03	.43 ± .03	.50 ± .03
Sfmc anomalies	66	.32 ± .03	.36 ± .03	.43 ± .03
Rzmc	59	-	.46 ± .03	.50 ± .03
Rzmc anomalies	33	-	.32 ± .05	.35 ± .05

Table 1 Average time series correlation coefficients with GSMDB surface and root zone soil moisture content (sfmc and rzmc, respectively) for SMMR, model, and assimilation estimates with 95% confidence intervals. N denotes the number of locations with sufficient data.

The model's skill for root zone soil moisture is comparable to its skill at the surface (Table 1), with correlation coefficients of 0.46 (0.32 for anomalies). Merging the surface information contained in the SMMR retrievals via data assimilation also leads to a small increase in the correlation coefficients for the root zone soil moisture to 0.50 (0.35 for anomalies). While going in the right direction, the improvement in the root zone is not statistically significant. Improvement of root zone soil moisture through assimilation of surface retrievals hinges on many factors, making it difficult to pinpoint a strategy for refining the assimilation system. Again, the ground measurements taken at point scale must be a reasonable representation of root zone soil moisture at the catchment scale, or else our measure of improvement is invalidated. (This argument applies equally to the surface layer.) Second, the model must accurately describe the propagation of the surface information into the deeper soil. Third, the model error parameters of the assimilation system that co-determine the strength of the coupling between the surface and the root zone must also be realistic. Unfortunately, it is currently impossible to test these assumptions at the global scale with any confidence. Again, though, despite these limitations, the assimilation of SMMR retrievals does yield improved estimates of soil moisture conditions, with at least a suggestion of an improvement in the root zone data.

5. Conclusions

Since April 2004, the NASA-GMAO seasonal forecasting system relies on land initial conditions that are derived from observations of precipitation and radiation. Along with a major change in atmospheric initialization, the new land initialization method has a significant impact on sub-seasonal forecasts of screen-level temperature and precipitation at mid-latitudes during summer.

The global assimilation of SMMR satellite retrievals of soil moisture into the NASA Catchment land surface model using the EnKF was also examined. We find that the assimilation of the satellite information improves the average annual cycle of surface and root zone soil moisture at locations with GSMDB ground data. The assimilation also produces small but significant improvements in time series correlations with ground data for surface soil moisture and its anomalies. Correlations for root zone soil moisture are also improved, though not with statistical significance.

Global analysis of the innovations sequence reveals that the assimilation algorithm only partially performs within its underlying assumptions. While the innovations have the expected zero mean property by design, they typically have too much or too little variance in different parts of the globe. Such a deficiency is not surprising for a first application of a global assimilation system to satellite data for a state (soil moisture) controlled by poorly understood non-linear processes. In future work, information from the innovations sequence can be used to design spatially distributed model error parameters, potentially in an adaptive framework, that might improve the performance of the assimilation algorithm. Finally, modern-era data such as C-band retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System, L-band retrievals from the planned Hydrospheric States mission (*Entekhabi et al.*, 2004), and satellite-supported surface meteorological observations of higher quality should further our knowledge of global soil moisture fields.

6. References

Berg, A. A., J. S. Famiglietti, J. P.Walker, and P. R. Houser (2003), The impact of bias correction to reanalysis products on simulations of North American soil moisture and hydrological fluxes, *J. Geophys. Res.*, **108** (D16), 4490, DOI 10.1029/2002JD003334.

Crow, W. T., and E. F. Wood (2003), The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97, *Adv. Water Resour.*, **26**, 137-149.

Dee, D. P. (1995), On-line estimation of error covariance parameters for atmospheric data assimilation, *Mon. Wea. Rev., 123*, 1128-1145.

De Jeu, R. A. M. (2003), Retrieval of land surface parameters using passive microwave remote sensing, Ph.D. diss., Dep. of Geo-Environ. Sci., Vrije Univ., Amsterdam, Netherlands.

Entekhabi, D., et al. (2004), The Hydrosphere State (HYDROS) mission concept: an Earth system pathfinder for global mapping of soil moisture and land freeze/thaw, *Trans. Geosci. Remote Sens.,* in press.

Entin, J. K., A. Robock, K. Y. Vinnikov, V. Zabelin, S. Liu, A. Namkai, and T. Adyasuren (1999), Evaluation of Global Soil Wetness Project soil moisture simulations, *J. Meteorol. Soc. Jpn.*, **77**, 183-198.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model, 1: Model structure, *J. Geophys. Res*., **105**, 24,809-24,822.

Koster, R. D., P. A. Dirmeyer, Z. Guo and 22 Coauthors (2004a), Regions of Strong Coupling Between Soil Moisture and Precipitation, *Science*, **305**, Issue 5687, 1138-1140 , 20 August 2004.

Koster, R. D., M. J. Suarez, P. Liu, U. Jambor, A. A. Berg, M. Kistler, R. H. Reichle, M. Rodell, and J. Famiglietti (2004b), Realistic initialization of land surface states: Impacts on subseasonal forecast skill, *J. Hydrometeorol*., in press.

Margulis, S. A., D. McLaughlin, D. Entekhabi, and S. Dunne (2002), Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 field experiment, *Water Resour. Res.*, **38**, 1299, doi:10.1029/2001WR001114.

Owe, M., R. de Jeu, and J. Walker (2001), A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index, *IEEE Trans. Geosci. Remote Sens*., **39**, 1643-1654.

Reichle, R. H., and R. D. Koster (2003), Assessing the impact of horizontal error correlations in background fields on soil moisture estimation, *J. Hydrometeorol*., **4,** 1229-1242.

Reichle, R. H., and R. D. Koster (2004a), Bias reduction in short records of satellite soil moisture, *Geophys. Res. Let.*, **31**, doi:10.1029/2004GL020938.

Reichle, R. H., and R. D. Koster (2004b), Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model, submitted to Geophys. Res. Let., 7 Oct. 2004.

Reichle, R. H., J. P. Walker, R. D. Koster, and P. R. Houser (2002), Extended versus Ensemble Kalman Filtering for Land Data Assimilation, *J. Hydrometeorol.,* **3**, 728-740.

Reichle, R. H., R. D. Koster, J. Dong, and A. A. Berg (2004), Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation, *J. Hydrometeorol*., **5**, 430-442.

Robock, A., K. Y. Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai (2000), The global soil moisture data bank, *Bull. Am. Meteorol. Soc.*, **81**, 1281-1299.

Rodell, M., et al. (2003), The Global Land Data Assimilation System, *Bull. Am. Meteorol. Soc.*, **85**, 381-394, DOI 10.1175/BAMS-85-3-381.

Seuffert, G., H. Wilker, P. Viterbo, J.-F. Mahfouf, M. Drusch, and J.-C. Calvet (2003), Soil moisture analysis combining screen-level parameters and microwave brightness temperature: A test with field data, *Geophys. Res. Lett.*, **30**, 1498, doi:10.1029/2003GL017128.

Walker, J. P., and P. R. Houser (2001), A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations, *J. Geophys. Res.*, **106**, 11,761-11,774.