Implementation of a global-scale operational data assimilation system for satellite-based soil moisture retrievals

J. Bolten\textsuperscript{a}, W. Crow\textsuperscript{b}, X. Zhan\textsuperscript{b}, C. Reynolds\textsuperscript{c}

\textsuperscript{a}USDA-ARS Hydrology and Remotes Sensing Lab, Bldg. 007
BARC-West, Beltsville, MD 20705

\textsuperscript{b}NOAA NESDIS Center for Satellite Applications and Research,
Camp Springs, MD 20746

\textsuperscript{c}USDA-FAS Production Estimate and Crop Assessment Division
Washington, DC 20250

ABSTRACT

Timely and accurate monitoring of global weather anomalies and drought conditions is essential for assessing global crop conditions. Soil moisture observations are particularly important for crop yield fluctuations provided by the US Department of Agriculture (USDA) Production Estimation and Crop Assessment Division (PECAD). The current system utilized by PECAD estimates soil moisture from a 2-layer water balance model based on precipitation and temperature data from World Meteorological Organization (WMO) and US Air Force Weather Agency (AFWA). The accuracy of this system is highly dependent on the data sources used; particularly the accuracy, consistency, and spatial and temporal coverage of the land and climatic data input into the models. However, many regions of the globe lack observations at the temporal and spatial resolutions required by PECAD. This study incorporates NASA’s soil moisture remote sensing product provided by the EOS Advanced Microwave Scanning Radiometer (AMSR-E) into the U.S. Department of Agriculture Crop Assessment and Data Retrieval (CADRE) decision support system. A quasi-global-scale operational data assimilation system has been designed and implemented to provide CADRE a daily product of integrated AMSR-E soil moisture observations with the PECAD two-layer soil moisture model forecasts. A methodology of the system design and a brief evaluation of the system performance over the Conterminous United States (CONUS) is presented.

Keywords: soil moisture, remote sensing, passive, microwave, data assimilation

1. INTRODUCTION

The U.S. Department of Agriculture (USDA) Production Estimates and Crop Assessment Division (PECAD) is responsible for providing monthly global crop estimates. These global crop estimates are used to increase agricultural efficiency, influence global commodity market access and provide early warning of production changes and agricultural drought monitoring. In addition to crop forecasting for food security, global agricultural monitoring has also been shown to be useful for estimating the effects of climate change on global food supplies and agriculturally-driven environmental changes\textsuperscript{1-3}.

In an effort to determine anomalous conditions indicating times of water stress or flooding, PECAD analysts compare current global conditions against a database of archived satellite imagery and crop yields. Estimates from PECAD are derived from a merging of many data sources including satellite and ground observations, and more than 20 years of climatology and crop behavior data over key agricultural areas. To most efficiently manage the data sources, PECAD has developed a series of analytical tools, crop models, and hazard calendars within a Crop Condition Data Retrieval and Evaluation (CADRE) Data Base Management System (DBMS). The goal of CADRE is to provide timely and accurate estimates of global crop conditions for use in up-to-date commodity intelligence reports.
A crucial requirement of these global crop yield forecasts is the regional characterization of surface and sub-surface soil moisture. However, due to the spatial heterogeneity and dynamic nature of precipitation events and soil wetness, accurate estimation of regional land surface-atmosphere interactions based sparse ground measurements is difficult. Temporal resolution is particularly important for predicting adequate surface wetting and drying between precipitation events and is closely integrated with CADRE. This work aims at improving the PECAD surface and sub-surface soil moisture estimates by assimilating satellite-retrieved soil moisture estimates into the CADRE two-layer soil moisture within an Ensemble Kalman Filter (EnKF) framework.

Soil moisture observations from the Eos Advanced Microwave Scanning Radiometer (AMSR-E) are integrated into the PECAD soil moisture model via a data assimilation system. Surface soil moisture dynamics observed by AMSR-E update the root zone through the vertical soil moisture coupling of the 2-layer soil moisture model. The improved temporal resolution and spatial coverage of the satellite-based EOS Advanced Microwave Scanning Radiometer (AMSR-E) over station data and model outputs used by PECAD is expected to provide a better characterization of surface wetness and enable more accurate crop monitoring in key agricultural areas.

2. PECAD WATER BALANCE MODEL

The PECAD DBMS combines many data sources including over 3000 ground observations from the World Meteorological Organization and climatological estimates provided by the Air Force Weather Agency (AFWA). Daily estimates of minimum and maximum temperature and precipitation are applied to a modified Palmer two-layer soil moisture model which calculates the daily amount of soil moisture withdrawn by evapotranspiration and replenished by precipitation. FAS has improved upon the original Palmer model by replacing the Thornthwaite evapotranspiration equation with the modified FAO Penman-Monteith equation as described in Allen. Moisture is depleted from the both layers at a fraction of the potential evapotranspiration rate dependant upon surface and root zone capacity and moisture content. Excess water is lost from the system (i.e., no runoff is implied). The model assumes a top soil layer of 2.5 cm and a second layer max depth of 1 m or less, dependent upon total water holding capacity derived from the FAO Digital Soil Map of the World, soil texture and depth of the soil column. PECAD applies the Palmer model at daily time steps within a stereographic projection with approximately 47 km horizontal grid spacing at 60 degree latitude (i.e., 1/8th mesh). A 3-day composite of the PECAD soil moisture product is shown in Figure 1.
3. AMSR-E

This work applies daily soil moisture estimates from the satellite-based AMSR-E instrument. AMSR-E is a conically-scanning microwave radiometer designed specifically for detecting soil moisture within the top 3 cm of soil depth. It was launched in 2002 on board the Aqua satellite, and provides full global coverage every 2-3 days. The NASA-delivered soil moisture product provided by AMSR-E is produced from brightness temperatures at 10.65 and 18.7 GHz. Observations of soil moisture are calculated from Polarization Ratios (PR) at 10.7 and 18.7 GHz, and three empirical coefficients used to compute a vegetation/roughness parameter for each grid cell. Deviations from an 18.7 GHz PR baseline value for each grid cell are used to calculate daily soil moisture estimates for each grid cell. Estimates of soil moisture are provided at a re-gridded global cylindrical 25 km Equal-Area Scalable Earth Grid (EASE-Grid) cell spacing. Within this application, the soil moisture estimates provided by AMSR-E are assumed represent a soil depth comparable to the first layer soil moisture used by the PE CAD DBMS (i.e., 2.54 cm). Figure 2 illustrates the daily global coverage of the re-scaled Level-3 soil moisture product.

![Image of AMSR-E soil moisture](image)

Figure 2. A 3-day composite of AMSR-E soil moisture.

4. ENSEMBLE KALMAN FILTER

The increased availability of a multitude of remote sensing products has been shown to be beneficial for improving meteorological, oceanographic, and land surface predictions. Data assimilation techniques use auto-recursive analyses to optimally merge model estimates with state observations. The reduction in model uncertainty is achieved by taking advantage of model state temporal constancy restraints and model physical properties.

Data assimilation is based on the availability of an observation $y$ that can be related to the state vector $x$ via a known observation operator $H$

$$y_k = H_k (x_k) + v_k$$

(1)
where \( v_k \) represents a random perturbation of observation. Such perturbations are assumed to be Gaussian with a known covariance of \( R \).

In this study, a 1-dimensional Ensemble Kalman filter (EnKF) is applied. The EnKF is a nonlinear extension of the standard Kalman filter and has been successfully applied to land surface forecasting problems\(^{11}\). Within the filter, sequential ensembles of stochastically perturbed model trajectories are corrected towards an observation of model state when available. The error covariance of both the forecasted observations - or \( H_k(\mathbf{x}_k) - (\mathbf{CM}_k) \) and the cross-correlation between these observations and each forecasted state variable \((\mathbf{CYM}_k)\) are calculated by sampling across the an ensemble created by the Monte Carlo realization of (3).

The updating step of the EnKF utilizes this error covariance information to optimally update forecasts in response to observations, based on the calculation of the Kalman gain as

\[
K_k = \frac{\mathbf{CYM}_k}{\mathbf{CM}_k + R}
\]

(2)

and the application of the Kalman filter updating equation individually to each realization within the ensemble

\[
\mathbf{x}_k^{+} = \mathbf{x}_k^{-} + K_k \left[ y_k - H_k(\mathbf{x}_k^{-}) + v_k \right].
\]

(3)

where “-” and “+” notation is used to signify state estimates made before and after updating in response to observations at time \( k \). The EnKF state estimate at time \( k \) is given by simply taking the mean of this updated ensemble.

Our particular implementation of the EnKF integrates soil moisture observations from AMSR-E with the modified Palmer two-layer soil moisture model described in Section 3 by applying a 1-dimensional EnKF at daily time-steps when AMSR-E observations are available. The model operator \( H_k \) represents the 2-Layer Palmer introduced in Section 3. Since AMSR-E observations are pre-processed into surface soil moisture estimates (assumed to be consistent with the top layer of the Palmer model), our observation operator is simply \( H = (1,0) \).

Before AMSR-E soil moisture retrievals can be assimilated, the modeled (PECAD) and observed (AMSR-E) data must be scaled to a common climatology to reduce potential biases and differences in dynamic range that commonly exist between modeled and observed soil moisture products. By removing time-invariant biases from the observation data, the two datasets can be optimally merged, allowing a more efficient assimilation strategy. We performed a retrospective analysis of archived AMSR-E and PECAD datasets from June 2002 to June 2007 to establish a representative climatology for both AMSR-E and PECAD soil moisture estimates. Based on these climatologies, a cumulative distribution function (CDF) matching algorithm was employed to transform individual AMSR-E retrievals such that their transformed climatology is comparable to that of PECAD surface soil moisture estimates, as shown by Reichle and Koster\(^{12}\). The climatologically re-scaled AMSR-E data were then introduced as observations to the Ensemble Kalman Filter (EnKF) using sequential observations of AMSR-E and AFWA climatological data.

### 4.1 Filter tuning

The accuracy of remotely sensed passive microwave observations vary greatly over different land cover types due to signal attenuation by vegetation and increased scattering over rough terrain\(^{13}\). At the wavelengths used by AMSR-E, the accuracy of observed soil moisture is significantly degraded over areas of vegetation water content greater than approximately 8-10 kg/m\(^2\)\(^{14}\). To reduce errors introduced into the assimilation system from incorrect AMSR-E observations over heavily vegetated areas, we exploit this relation by adjusting the magnitude of the errors applied to the state observations in relation to vegetation type.

Assuming Gaussian noise added to equations 1 and 3, filter innovations (i.e., the difference between the observations and updated state) should also display Gaussian moments and be un-correlated in time. Adjusting the model or observation
covariance will impact the Kalman gain and effectively adjust the innovation magnitude and direction. These filter diagnostics can be used to test whether the filter is working properly and to optimize the filter. In this way, a diagnostic calibration of the filter was performed for three ranges of vegetation. A land cover mask was used to adjust the observation error for 1) evergreen and deciduous forests 2) wooded grasslands and mixed forests, and 3) bare soil, grasses and open shrub land. The mean value of observation error that satisfied the innovation statistics for each land cover type was then applied to all global areas possessing a similar land cover type. This method ensures that the filter is placing the proper relative weight on model predictions and remote sensing observations when calculating an analysis soil moisture product with minimized error. The filter has been applied globally after applying the tuned observation errors based on land cover. A 3 day composite of the integrated product for the same time period as Figures 1 and 2 is shown in Figure 3.

5. EXPERIMENT DESCRIPTION AND RESULTS

For evaluation of the data assimilation system, we focus on a 5 year (06/19/2002 – 06/19/2007) analysis of the assimilated product over the conterminous United States. A data denial framework is employed to test the filter performance by comparing three model runs: one forcing the model with ‘bad’ error-prone precipitation, one forcing the model with ‘bad’ error-prone precipitation and assimilating AMSR-E soil moisture, and another forced with ‘good’ benchmark precipitation.

For this analysis, the data denial approach is based on these three separate model runs over a five-year duration at daily time steps within CONUS. The individual model runs were forced by 1) reliable precipitation (i.e. benchmark loop) 2) error-prone precipitation (i.e. open loop) and 3) error-prone precipitation and the EnKF assimilation of AMSR-E soil moisture retrievals (i.e. EnKF loop). We use the AFWA gauge-corrected precipitation data for our benchmark loop. The un-corrected real-time precipitation product provided by the Tropical Rainfall Measuring Mission (TRMM), 3B40RT, is used for the open loop and EnKF loops. In this way, the application of the EnKF to assimilate remotely-sensed soil moisture retrievals in case 3 can be evaluated based on how efficiently it transforms the low-accuracy results in case 2.
Figure 5 shows spatial maps of the forecasted first layer soil moisture over CONUS on 07/21/2004 for all three schemes discussed above (open loop in Figure 5a, benchmark run in Figure 5c and the AMSR-E/error-prone/EnKF run in Figure 5d). Also shown is a spatial map of the scaled AMSR-E soil moisture product for the same time period (Figure 5b). It is apparent when comparing 5a and 5d that the assimilation of AMSR-E soil moisture via the EnKF adds spatial heterogeneity to the open loop case. The added heterogeneity allows the EnKF case to better approximate soil moisture patterns in the benchmark case (Figure 5a) by adding soil moisture in areas (e.g. the south-central and northeastern CONUS) where the error-prone product underestimates antecedent precipitation magnitudes. This demonstrates that the AMSR-E soil moisture retrievals are able to effectively compensate modeled soil moisture for the impact of poorly observed rainfall patterns. It is worth noting that over many important – but data poor - agricultural regions of the world, such compensation is critically important due to the lack of ground-based rainfall observations.

![Data denial experiment flowchart.](image)

A time-series analysis of the data denial strategy is presented in Figure 6. Shown is soil moisture from the surface and root zones for one year duration over cropland area (i.e., longitude=-110°, latitude=32°). The EnKF (red) loop compares well with the benchmark (blue) realization compared to the open (green) realization. It can be seen from the figure that the AMSR-E observations have a significant influence on both the surface and root zone soil moisture layers during both wetting and drying events. This location is a good representation of a PECAD target area and efficiently demonstrates the added value of the assimilation system over the open loop forced with error-prone precipitation.

PECAD is continually assessing and updating their data base management system to provide timely and accurate estimates of global crop conditions for use in up-to-date commodity intelligence reports. Significant progress has been made in the application of formally integrated models of atmosphere and land-based water cycle to monitor regional land surface processes, namely soil moisture and modeled crops such as corn and soy. This work has demonstrated a currently running data assimilation system that is operating in near-real time to provide the PECAD Crop Condition Data Retrieval and Evaluation (CADRE) Data Base Management System (DBMS) with an additional soil moisture product of improved spatial-temporal scale. The improved spatial structure of wetness patterns demonstrated in Figures 3, 5d and 6 suggest that the system holds promise for improved characterization of surface wetness conditions at the regional scale and enable more accurate monitoring of soil moisture changes in key agricultural areas.
Future work includes extending our diagnostic calibration of the filter to areas of the globe lacking in sufficient meteorological observations where the PECAD system is more susceptible to agricultural forecasting errors. Further testing will also include additional comparisons to heavily monitored in situ soil moisture ground networks and a more quantitative (and longer term) analysis of data denial results in Figure 6 to clarify the added impact of AMSR-E soil moisture data assimilation on the accuracy of PECAD soil moisture estimates.

Figure 5. Data denial experiment results over CONUS for 07/21/2004. A) Benchmark loop B) re-scaled AMSR-E Soil moisture product C) Open loop and D) EnKF results.

Figure 6. Time-series of data denial experiment from 07/21/2003 – 07/21/2004 over cropland area. Longitude =-110°, Latitude= 32°.
ACKNOWLEDGEMENTS

This research was supported by funding from NASA’s Applied Sciences Program and grant NNS06AA051 entitled “Integrating NASA’s Global Soil Moisture Remote Sensing and Modeling Data into the USDA’s Global Crop Production Decision Support System”.

REFERENCES


