Advances in land data assimilation at NASA/GSFC

Rolf Reichle*  
M. Bosilovich¹, J. Case², W. Crow³, R. Koster¹,  
S. Kumar¹, Q. Liu¹, C. Peters-Lidard¹, S. Mahanama¹,  
M. Rodell¹, J. Santanello¹, B. Zaitchik⁴

¹GSFC, ²MSFC, ³USDA, ⁴JHU

*Email: Rolf.Reichle@nasa.gov  
*Phone: +1-301-614-5693
Summary of activities

**Satellite observations**
- Soil moisture
- Snow
- Land surface temperature (LST, a.k.a. “skin” temperature)
- Terrestrial water storage (TWS)

**Algorithms**
- EnKF and ensemble smoothing
- Dynamic bias correction
- Adaptive estimation of error parameters

**Systems**
- GEOS-5 LDAS (Catchment model; EnKF, bias, adaptive)
- Land Information System (LIS)
  - multiple land models (Catchment, Noah, CLM, HTESSEL…)
  - includes GMAO EnKF and bias estimation
- coupled to WRF
- parameter estimation tools

So far, mostly “**uni-variate**” and “**off-line**” (land-only).
**Ensemble Kalman filter (EnKF)**

Propagation \( \mathbf{t}_{k-1} \) to \( \mathbf{t}_k \):

\[
\mathbf{x}_k^{i-} = f(\mathbf{x}_{k-1}^{i+}) + \mathbf{e}_k^i
\]

\( \mathbf{e} \) = model error

Update at \( \mathbf{t}_k \):

\[
\mathbf{x}_k^{i+} = \mathbf{x}_k^{i-} + \mathbf{K}_k (\mathbf{y}_k^i - \mathbf{x}_k^{i-})
\]

for each ensemble member \( i = 1 \ldots N \)

\[
\mathbf{K}_k = \mathbf{P}_k (\mathbf{P}_k + \mathbf{R}_k)^{-1}
\]

with \( \mathbf{P}_k \) from ensemble spread

Nonlinear ensemble propagation approximates model errors.

Apply small perturbations to each ensemble member (model forcings and states) at every time step.

Dynamic bias estimation.

Adaptive estimation of error parameters.

Developed in GEOS-5 LDAS and integrated into LIS.

Andreadis and Lettenmaier (2005); Durand and Margulis (2007); Kumar et al. (2008a, 2008b, 2009); Pan and Wood (2006); Reichle et al. (2002a, 2002b, 2007, 2008a, 2008b, 2009); Reichle and Koster (2003, 2004, 2005); De Lannoy et al. (2007); Crow and Reichle (2008); Zaitchik et al. (2008); Zhou et al. (2006)
Outline

**Soil moisture**
- SMAP Level 4 Products
- Multi-model soil moisture assimilation
- Adaptive filtering

**Land surface temperature**
- Bias

**Snow data and terrestrial water storage**
- Smoothing
- Multi-scale assimilation
- Vertical and horizontal disaggregation

**LIS examples**
- Soil moisture and sea-breeze
- Boundary layer mixing diagrams
- Parameter estimation
Outline

**Soil moisture**
- SMAP Level 4 Products
- Multi-model soil moisture assimilation
- Adaptive filtering

**Land surface temperature**
- Bias

**Snow data and terrestrial water storage**
- Smoothing
  - Multi-scale assimilation
  - Vertical and horizontal disaggregation

**LIS examples**
- Soil moisture and sea-breeze
- Boundary layer mixing diagrams
- Parameter estimation
**NASA Soil-Moisture-Active-Passive (SMAP) mission**

First of NRC Earth Science Decadal Survey missions

**Science objectives**
- Global land surface water, energy, and carbon fluxes.
- Enhance weather and climate forecast skill.
- Improve flood prediction and drought monitoring.

**Platform and instruments**
- L-band (1.4 GHz) synthetic aperture radar (active) and radiometer (passive) with 6-m rotating antenna
  - **Orbit:** Sun-synchronous
  - **Swath width:** 1000 km
  - **Resolution:** 1-3 km (radar), 40 km (radiometer)
  - **Revisit:** 2-3 days
  - **Duration:** 2015-18
  - **Sensing depth:** ~5 cm

**Latent heat flux depends on soil moisture**

![Normalized evaporation vs. soil moisture graph]

(Cahill et al., 1999)

**Soil freeze-thaw drives boreal carbon balance**

![Soil temperature and evaporation graph]

(Frolking et al., 1996)

** Only surface soil moisture!**
### SMAP Baseline Science Data Products

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
<th>Resolution</th>
<th>Latency*</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1B_S0_LoRes</td>
<td>Low Resolution Radar Backscatter ($\sigma^o$)</td>
<td>~ 30 km</td>
<td>12 hours</td>
</tr>
<tr>
<td>L1C_S0_HiRes</td>
<td>High Resolution Radar Backscatter ($\sigma^o$)</td>
<td>~ 1-3 km</td>
<td>12 hours</td>
</tr>
<tr>
<td>L1B_TB</td>
<td>Radiometer Brightness Temperature ($T_B$)</td>
<td>~ 40 km</td>
<td>12 hours</td>
</tr>
<tr>
<td>L1C_TB</td>
<td>Radiometer Brightness Temperature ($T_B$)</td>
<td>~ 40 km</td>
<td>12 hours</td>
</tr>
<tr>
<td>L3_F/T_HiRes</td>
<td>Freeze/Thaw State</td>
<td>~ 3 km</td>
<td>24 hours</td>
</tr>
<tr>
<td>L3_SM_HiRes</td>
<td>Radar Soil Moisture (internal product)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>L3_SM_40km</td>
<td>Radiometer Soil Moisture</td>
<td>~ 40 km</td>
<td>24 hours</td>
</tr>
<tr>
<td>L3_SM_A/P</td>
<td>Radar/Radiometer Soil Moisture</td>
<td>~ 10 km</td>
<td>24 hours</td>
</tr>
<tr>
<td><strong>L4_SM</strong></td>
<td>Surface &amp; <strong>Root-zone</strong> Soil Moisture</td>
<td>~ 10 km</td>
<td>7 days</td>
</tr>
<tr>
<td>L4_C</td>
<td>Carbon Net Ecosystem Exchange</td>
<td>~ 10 km</td>
<td>14 days</td>
</tr>
</tbody>
</table>

GSFC *develops* L4_SM algorithm and *generates* L4_SM and L4_C products. L4_SM builds on experience with AMSR-E soil moisture assimilation.
Assimilate **AMSR-E** surface soil moisture (2002-08) into NASA Catchment model.

**Anomaly RMSE**

<table>
<thead>
<tr>
<th></th>
<th>AMSR-E [m³/m³]</th>
<th>Model [m³/m³]</th>
<th>Assim. [m³/m³]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface s.m.</td>
<td>0.049</td>
<td>0.051</td>
<td>0.048</td>
</tr>
<tr>
<td>Root zone s.m.</td>
<td>n/a</td>
<td>0.039</td>
<td>0.036</td>
</tr>
</tbody>
</table>

**Anomaly R**

<table>
<thead>
<tr>
<th></th>
<th>AMSR-E</th>
<th>Model</th>
<th>Assim.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface s.m.</td>
<td>.42±.01</td>
<td>.38±.01</td>
<td>.47±.01</td>
</tr>
<tr>
<td>Root zone s.m.</td>
<td>n/a</td>
<td>.37±.01</td>
<td>.45±.01</td>
</tr>
</tbody>
</table>

Root zone critical for applications but **not** observed by satellite.

- Assimilation product agrees better with ground data than satellite or model alone.
- Modest increase may be close to maximum possible with *imperfect* in situ data.
- Higher quality SMAP obs will provide better improvements.

Validate with USDA SCAN stations (only 36 of 103 suitable for validation).

Soil-Moisture-Active-Passive (SMAP) mission design

Q: How uncertain can retrievals be and still add useful information in the assimilation system?  
A: Synthetic data assimilation experiments.

Skill measured in terms of R (=anomaly time series correlation coefficient against synthetic truth).
Each plus sign indicates result of one 19-year assimilation integration over Red-Arkansas domain.

Results
• Assimilation of (even poor) soil moisture retrievals adds skill (relative to model product).
• Published AMSR-E and SMMR assimilation products consistent with expected skill levels.
• Derive error budget analysis for SMAP.

SMAP L4_SM uncertainty estimates

Interpreting the OSSE for SMAP yields:

Assimilation of SMAP obs will provide improvements (over model) of 0.01 m³/m³ for surface and 0.005 m³/m³ for root-zone soil moisture.

We expect the L4_SM product to meet the 0.04 m³/m³ error requirement.

The above numbers probably underestimates the skill improvement for regions with less reliable precipitation data (compared to the US).

| Skill scenario | L3_SM\(^1,3\) (A/P) | Model\(^2,3\) | L4_SM\(^3\) | \(|\Delta|\) |
|----------------|---------------------|--------------|--------------|---------------|
| **Surface soil moisture** | | | | |
| High | 0.028 | 0.046 | 0.035 | 0.012 |
| Low | 0.037 | 0.051 | 0.038 | 0.012 |
| **Root zone soil moisture** | | | | |
| High | n/a | 0.036 | 0.031 | 0.005 |
| Low | n/a | 0.038 | 0.031 | 0.007 |

\(^1\)Source: SMAP measurement requirements.
\(^2\)Source: USDA/SCAN results.
\(^3\)Source: OSSE results.

\(\Delta\) ≡ \left| \text{Model} - \text{L4_SM} \right| \quad \text{(skill contribution of SMAP to model products)}

Anomalies ≡ mean seasonal cycle removed
Multi-model soil moisture assimilation

How does land model formulation impact assimilation estimates of root zone soil moisture?

Normalized ROOT ZONE soil moisture improvement from assimilation of surface soil moisture

<table>
<thead>
<tr>
<th>Model</th>
<th>Catch</th>
<th>Mos</th>
<th>Noa</th>
<th>CLM</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catch</td>
<td>0.71</td>
<td>0.54</td>
<td>0.36</td>
<td>0.38</td>
<td>0.50</td>
</tr>
<tr>
<td>Mos</td>
<td>0.55</td>
<td>0.69</td>
<td>0.31</td>
<td>0.33</td>
<td>0.47</td>
</tr>
<tr>
<td>Noa</td>
<td>0.43</td>
<td>0.43</td>
<td>0.36</td>
<td>0.26</td>
<td>0.37</td>
</tr>
<tr>
<td>CLM</td>
<td>0.11</td>
<td>0.21</td>
<td>0.10</td>
<td>0.45</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Catchment or MOSAIC “truth” easier to estimate than Noah or CLM “truth”.

Stronger coupling between surface and root zone provides more “efficient” assimilation of surface observations.

Binning the spatially distributed results of all fraternal twin experiments according to VCS values yields:

**Stronger coupling between surface and root zone leads to more efficient assimilation.**

The slight asymmetry (across the diagonal) suggests that it is prudent to overestimate the VCS in the assimilation model.

NIC = normalized information contribution

VCS = vertical coupling strength
Impact of model and obs error inputs on assimilation skill

Each “+” symbol represents one 19-year assim. experiment over the Red-Arkansas with a unique combination of input model and observation error parameters.

RMSE of assimilation estimates v. truth for:

Surface soil moisture $m^3/m^3$

$\sqrt{P(Q_{\text{true}})} = 0.05$, OL = 0.035

$Q = \text{model error (including errors in precip, radiation, and soil moisture tendencies)}$

$P = P(Q) = \text{soil moisture error variance}$

Reichle et al., doi:10.1029/2007WR006357
Impact of model and obs error inputs on assimilation skill

RMSE of assimilation estimates v. truth for:

Surface soil moisture m$^3$/m$^3$

$\sqrt{P(Q_{true})}$

$\sqrt{R_{true}} = 0.05$, OL = 0.035

$\sqrt{P(0)}$

$\sqrt{R(0)}$

- “True” input error covariances yield minimum estimation errors.
- Wrong model and obs. error covariance inputs degrade assimilation estimates.
- In most cases, assimilation still better than open loop (OL).

Reichle et al., doi:10.1029/2007WR006357
Diagnostics of filter performance and adaptive filtering

Find true Q, R by enumeration?
• RMSE plots require “truth” (not usually available).
• Too expensive computationally.

Use diagnostics that are available within the assimilation system.

Filter update: \[ x^+ = x^- + K(y - x^-) \]

K = \( P (P + R)^{-1} \) = Kalman gain

Dissociation: \[ E[(y - x^-)(y - x^-)^T] = P + R \]

innovations \( \equiv \) obs – model prediction
(internal diagnostic)

state err cov + obs err cov (controlled by inputs)

Example: Average “obs. minus model prediction” distance is much larger than assumed input uncertainties
Adaptive v. non-adaptive EnKF

Contours: Surface soil moisture RMSE of assimilation estimates v. truth

- Adaptive filter yields improved assimilation estimates for initially wrong model and observation error inputs (except for $R_0=0$).

Reichle et al., doi:10.1029/2007WR006357
Outline

Soil moisture
- SMAP Level 4 Products
- Multi-model soil moisture assimilation
- Adaptive filtering

Land surface temperature
- Bias

Snow data and terrestrial water storage
- Smoothing
- Multi-scale assimilation
- Vertical and horizontal disaggregation

LIS examples
- Soil moisture and sea-breeze
- Boundary layer mixing diagrams
- Parameter estimation
Model v. satellite land surface temperature (LST)

July 2004 LST: GEOS-5 DAS minus MODIS
[Bosilovich et al, NASA/GMAO, Mar 2008]
Strategies for LST assimilation

Kalman filter state update:
\[ x^+ = x^- + K_x(y - Hx^-) \]
\[ K_x = P_x H^T (HP_x H^T + R)^{-1} \]
Bias update (2\textsuperscript{nd} Kalman filter):
\[ b^+ = b^- - K_b(y - H(x^- - b^-)) \]
Assume:
\[ P_b \sim \lambda P_x \Rightarrow K_b = \lambda K_x \]

STRATEGIES

1. A priori scaling
Assimilate *anomalies* (after removing climatological bias prior to data assimilation; broken down by season and time-of-day).

2. Bias estimation.
*Dynamically* estimate bias (Dee, Da Silva, Bosilovich).

Simple assumption allows use of *regular Kalman filter machinery* to update bias.

Bias estimate is effectively time average of increments.

---

LST lacks memory.

*Bosilovich et al., JMSJ 2007*
Land surface temperature (LST) assimilation

Assimilate ISCCP LST into off-line land models: Catchment (CLSM) & Noah.

Validate against CEOP obs. (48 stations; 2003-2004).

“Model” LST much better than ISCCP. Assimilation reduces anomaly RMSE by ~0.3 K. Bias estimation necessary. Model formulation impacts assimilation strategy.

$\text{Anomalies} \equiv \text{mean seasonal cycle removed}$

$LST: \ \text{Land surface temp.}$
$LH: \ \text{Latent heat flux}$
$SH: \ \text{Sensible heat flux}$
$GH: \ \text{Ground heat flux}$
A few days at MGS in Tibet…

Dynamic bias correction without a priori scaling can force the land models out of their “comfort zones” and leads to unrealistic flux estimates.
Soil moisture
- SMAP Level 4 Products
- Multi-model soil moisture assimilation
- Adaptive filtering

Land surface temperature
- Bias

Snow data and terrestrial water storage
- Smoothing
- Multi-scale assimilation
- Vertical and horizontal disaggregation

LIS examples
- Soil moisture and sea-breeze
- Boundary layer mixing diagrams
- Parameter estimation
Model fills spatial and temporal data gaps, provides continuity and quality control.

Assimilation output
- agrees better with IMS snow cover (top middle)
- contains more information (~hourly SWE) than MODIS (~daily snow cover)

Use MODIS snow cover to update model snow water equivalent (SWE)
Forward-looking “pull” algorithm (smoother):
- Assess MODIS snow cover 24-72 hours ahead
- Adjust air temperature (rain v. snowfall, snow melting v. frozen)
Highly accurate measurement of distance between twin satellites

Gravity anomaly

“Fast” signal (weekly to monthly; after correction for atmospheric pressure)

Terrestrial water storage (TWS) anomaly

Water Storage Anomaly (cm)
1.) Run high-resolution land model forecast for one month.

Zaitchik et al. (2008)
*J. Hydrometeorology*,
doi:10.1175/2007JHM951.1
1.) Run high-resolution land model forecast for one month
2.) Diagnose large-scale TWS on the 5th, 15th, and 25th, compute innovations ($\Delta Y$)

$\Delta Y = Y - M[X]$
1.) Run high-resolution land model forecast for one month
2.) Diagnose large-scale TWS on the 5th, 15th, and 25th, compute innovations (ΔY)
3.) Compute gain (K) and increments (ΔX)

\[
\Delta X = K \Delta Y
\]

\[
\Delta Y = Y - M[X]
\]

Zaitchik et al. (2008)
*J. Hydrometeorology*, doi:10.1175/2007JHM951.1
Ensemble Kalman smoother

1.) Run high-resolution land model forecast for one month

2.) Diagnose large-scale TWS on the 5th, 15th, and 25th, compute innovations ($\Delta Y$)

3.) Compute gain ($K$) and increments ($\Delta X$)

4.) Apply increments during second integration

5.) Repeat for next month…

$\Delta Y = Y - M[X]$

$\Delta X = K \Delta Y$

Zaitchik et al. (2008)
J. Hydrometeorology, doi:10.1175/2007JHM951.1
Assimilation of GRACE terrestrial water storage (TWS)

GRACE measures large-scale TWS
= groundwater
+ soil moisture
+ snow
+ surface water

Assimilation yields:
• fine-scale information subject to GRACE basin-scale constraints
• better runoff than model (not shown).


Assimilation disaggregates GRACE data into snow, soil moisture, and groundwater. Assimilation estimates of groundwater better than model estimates.

Validation against observed groundwater:

\[ \text{RMSE} = 23.5 \text{ mm} \]
\[ R^2 = 0.35 \]

\[ \text{RMSE} = 18.5 \text{ mm} \]
\[ R^2 = 0.49 \]

Assimilation disaggregates GRACE data into snow, soil moisture, and groundwater. Assimilation estimates of groundwater better than model estimates.

Outline

Soil moisture
- SMAP Level 4 Products
- Multi-model soil moisture assimilation
- Adaptive filtering

Land surface temperature
- Bias

Snow data and terrestrial water storage
- Smoothing
- Multi-scale assimilation
- Vertical and horizontal disaggregation

LIS examples
- Soil moisture and sea-breeze
- Boundary layer mixing diagrams
- Parameter estimation
More detail in LIS initial condition (as expected)

LIS/WRF drier over Northern FL & Southern GA

Difference in 12-h forecast of 2m air temp. (sea breeze)

LIS/WRF better than control (independent validation)

Diurnal evolution of 2m temperature and humidity reflects land surface (soil moisture) and atmospheric (boundary-layer depth) conditions and is a diagnostic of local land-atmosphere coupling.

The LIS-WRF mesoscale modeling system is a tool for testing several land surface models and PBL schemes in a consistent framework.

Soil moisture anomalies lead to significantly different signatures of heat and moisture evolution.
Heat and moisture budgets for different PBL-land model combinations can be derived from mixing diagrams and compared against observations.
Soil parameter estimation can improve soil moisture fields.
**Outlook**

**SUMMARY**

- **Assimilation products better than model or satellite data.**
- Obs. can be extrapolated and downscaled (space & time).
- Improvements are modest because the skill of land models (given observations-based forcings) is comparable to that of satellite observations.
- **Ensemble-based assimilation is appropriate for the problem.**
- **Bias is everywhere.**
- Validation is difficult for lack of in situ observations.
- Assimilation system contributes to mission design & products.
**Outlook**

**FUTURE PLANS**

- **Multi-variate** assimilation of soil moisture, land surface temperature, snow cover, and snow water equivalent.
- Customize system for SMAP, incl. novel technique for assimilation of freeze-thaw information.
- Integrate LDAS with GEOS-5 ADAS; assimilate LaRC near-real time LST.
- Investigate feedback of land analysis on atmospheric state in **coupled land-atmosphere analysis** system.
- Assimilate satellite-based vegetation observations.
- **Multi-variate “Integrated Earth System Analysis”** (atmosphere + ocean + land)
THANK YOU FOR YOUR ATTENTION!