Land hydrology data assimilation - Are we on the right track?

Sujay V. Kumar

Hydrological Sciences Laboratory,
NASA Goddard Space Flight Center, Greenbelt, MD
source: twitter, inspired by xkcd.com

Types of Hydrology Papers

- Let’s resurrect this paradigm from the 1970s
- Everybody is measuring soil moisture incorrectly, including us
- On some topic I’ve been thinking about a lot
- LOL, our field site burned down
- We couldn’t answer our question, does that make it a paradox?
- We pointed a radar at something
- Spatial heterogeneity means we’ll never know anything about anything
- We changed one model parameter from our last paper and predictive skill improved 2%
- Spatial heterogeneity doesn’t matter if you just do more math
- There’s no such thing as a physics based model
- Isn’t every model physics based when you think about it?
- My watershed is the weirdest

(Types of Hydrology Papers)

Overall, the assimilation of soil moisture and snow datasets was found to provide marginal improvements over the open-loop configuration. Though the improvements in soil moisture fields through soil moisture data assimilation were barely at the statistically significant levels, these small improvements were found to translate into subsequent small improvements in simulated streamflow. The assimilation of

<table>
<thead>
<tr>
<th>R (-)</th>
<th>OL</th>
<th>DA_α</th>
<th>DA_β</th>
<th>DA_α+β</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.58</td>
<td>0.63</td>
<td>0.66</td>
<td>0.70</td>
<td>±0.04</td>
<td></td>
</tr>
<tr>
<td>0.59</td>
<td>0.61</td>
<td>0.53</td>
<td>0.60</td>
<td>±0.14</td>
<td></td>
</tr>
<tr>
<td>0.84</td>
<td>0.86</td>
<td>0.92</td>
<td>0.93</td>
<td>±0.07</td>
<td></td>
</tr>
<tr>
<td>0.66</td>
<td>0.81</td>
<td>0.52</td>
<td>0.62</td>
<td>±0.17</td>
<td></td>
</tr>
</tbody>
</table>

The quality of the model. In situations where the atmospheric variables and the biophysical parameters of the model are well characterised, the data assimilation has a limited impact on the simulations. In data poor areas, the assimila-

where the SNOTEL sites are located, however, AMSR-E retrievals are typically biased low and assimilation without prior scaling leads to degraded SWE estimates. Anomaly SWE assimilation could not improve the interannual SWE variations in the assimilation results because the AMSR-E retrievals lack realistic interannual variability in deep snowpacks. SCF assimilation has only a marginal impact at the SNOTEL locations because these sites experience extended periods of near-complete snow cover. Across all sites, SCF

source: twitter, inspired by xkcd.com
Systematic errors are dominant in land surface hydrology. Global soil moisture climatology is unknown; Interpretation issues galore. Assimilation systems rely on rescaled estimates to get around this limitation.

Simulated “soil moisture” does not have an unambiguous meaning. It is a strongly model-specific quantity, essentially an “index” of the moisture states, with a dynamic range defined by the specific evaporation and runoff formulations utilized by the given model.

Given these differences, the direct transfer of one model’s soil moisture into another and in applications is, in fact, inappropriate.

The remote sensing of soil moisture via satellite holds particular promise to address the lack of in situ observations across the globe, but much work is still needed to address the proper use of these data for model improvement. In some existing cases, satellite methods seem to return their own forms of soil moisture indices rather than direct soil moisture estimates.

Koster et al. (2009)
Rescaling approaches problematic, they lead to information loss (> 11%)

Can introduce statistical errors if unmodeled features are present
The common approach to evaluation is to compare against in-situ datasets, though there may be little information in retrievals about point-scale processes.

Most evaluation metrics (similar to assimilation strategies) are focused on anomaly metrics, ignoring biases that can be sources of important signals.

Improving ‘observability’ of modeled and remotely sensed soil moisture is important for better representation of land surface heterogeneity and management impacts.
What if you relax the ‘observability’ assumptions?

An alternate approach to CDF matching where the SMAP anomalies are directly assimilated.

The anomaly DA provides comparable performance to CDF-matching and is more effective in incorporating unmodeled features such as irrigation.
Are we meeting the promising of improving unobserved processes with data assimilation?

<table>
<thead>
<tr>
<th></th>
<th>FLUXNET</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE (Wm$^{-2}$)</td>
</tr>
<tr>
<td>Qle</td>
<td></td>
</tr>
<tr>
<td>GLDAS</td>
<td>24.7 ± 0.3</td>
</tr>
<tr>
<td>NLDAS (Noah v2.7.1)</td>
<td>19.3 ± 0.3</td>
</tr>
<tr>
<td>NLDAS (Noah v3.2)</td>
<td>27.6 ± 0.3</td>
</tr>
<tr>
<td>NLDAS (Noah v3.2) + NASA DA</td>
<td>29.4 ± 0.3</td>
</tr>
<tr>
<td>NLDAS (Noah v3.2) + LPRM DA</td>
<td>25.6 ± 0.3</td>
</tr>
</tbody>
</table>

from Peters-Lidard et al. 2011

Improvements in ET from SMAP soil moisture assimilation (red indicates improvements, blue suggests degradations)

Impacts on ET from soil moisture DA is small. Comparatively, assimilating vegetation optical depth from SMAP leads to much larger improvements in ET.
Not many studies examining LST assimilation (and demonstrating improvements) despite the multi-decadal availability of remote sensing observations.

Unavailability of relevant model prognostics

Large biases between model and observations

from Reichle et al. 2010
Are we focused on the right type of measurements to exploit?

Snow cover DA has a positive impact on snow simulation, with less degradations. The level of improvements from snow cover DA is also smaller compared to those from albedo DA.

From Kumar et al. 2020
Focus on representing the ubiquitous land management impacts could lead to high impact results from land DA

Vegetation assimilation (LAI DA) improves water and carbon fluxes, captures features of irrigated and agricultural areas, improves drought estimation over irrigated areas

Agricultural areas of the U.S.

Warm colors indicate improvements, cool colors indicate degradations

from Kumar et al. 2019
Capturing the impact of vegetation disturbances (2019-2020 Australian bushfires)

SMAP VOD captures the vegetation disturbances from the 2019-2020 bushfires

from Kumar et al. 2021
Impact of bushfire disturbances on the water cycle

SMAP VOD retrievals are assimilated into NoahMP LSM to examine the impact on evaporative and runoff fluxes.

Removal of vegetation leads to increased runoff, increased bare soil evaporation, reduced transpiration.
Multivariate constraints may be needed to get the right answers!

Example - Heavily irrigated areas with ground water pumping

TWS decreases, LAI increases; Assuming a single source measurement is unlikely to capture these processes and may give misleading results.
Final thoughts

Get away from the 2% improvement syndrome; Need to improve the observability of models and retrievals (reduce the reliance on rescaling methods)

Do not ignore the elephant in the room; DA approaches need to focus on methods and metrics that allow the reduction of systematic errors.

Model development needs to infuse DA requirements; Without the right set of prognostics, DA will always be suboptimal.

Multivariate remote sensing data infusion is critical for providing insights of human management; Let’s not miss the boat.
Over the arid location, soil moisture is the primary control on ET. Changes in ET are more connected to bare soil evaporation. Soil moisture DA drives the changes in ET.

Over the wet location, VOD DA changes LAI and transpiration, which drives the changes in ET.