

The spatial scale of model errors and assimilated retrievals in a terrestrial water storage assimilation system

B. A. Forman,¹ and R. H. Reichle²

Received 29 August 2012; revised 24 September 2013; accepted 7 October 2013; published 18 November 2013.

[1] Synthetic satellite observations (or retrievals) of terrestrial water storage (TWS) in the Mackenzie River basin located in northwestern Canada were assimilated into the Catchment land surface model to evaluate the impact (i) assimilating TWS retrievals at subbasin ($\sim 10^5$ km²) or basin ($\sim 10^6$ km²) scales and (ii) incorrectly specifying the model error correlation length that is used for the perturbation of model forcing and prognostic variables in the ensemble-based assimilation system. Specifically, a total of 16 experiments were conducted over a 9 year study period using different combinations of the spatial scale of the assimilated TWS retrievals and the horizontal model error correlation length. In general, assimilation of the TWS retrievals at the subbasin scale ($\sim 2.7 \times 10^5$ km² on average) yielded the best agreement relative to the synthetic truth. Greater improvement in TWS and snow water equivalent, in general, was witnessed as the (designed) horizontal model error correlation length increased. Conversely, subsurface soil water, evaporation, and runoff estimates typically improved (or remained unchanged) as the horizontal model error correlation length decreased. As the scale of the assimilated TWS retrieval decreased, more mass was effectively transferred from snow water equivalent into the subsurface, thereby dampening the hydrologic runoff response in the study area and correcting for improper model physics related to the runoff routing scheme. In general, TWS retrievals should be assimilated at the smallest spatial scale for which the observation errors can be considered uncorrelated while the specification of the horizontal error correlation length scale is of secondary importance.

Citation: Forman, B. A., and R. H. Reichle (2013), The spatial scale of model errors and assimilated retrievals in a terrestrial water storage assimilation system, *Water Resour. Res.*, 49, 7457–7468, doi:10.1002/2012WR012885.

1. Introduction and Background

[2] Space-based gravimetric retrievals of terrestrial water storage (TWS) changes provide a unique capability to better diagnose the Earth's hydrologic cycle [Yeh *et al.*, 2006; Rodell *et al.*, 2009; Syed *et al.*, 2009; Strassberg *et al.*, 2009; Tang *et al.*, 2010; Wang *et al.*, 2011]. Gravimetric measurements remotely sense TWS changes via detection of gravitational anomalies associated with the aggregation (or dissipation) of mass near the Earth's surface [Wahr *et al.*, 2004; Tapley *et al.*, 2004]. The ability to “feel” rather than “see” terrestrial water throughout the water column enables an assessment of TWS that microwave, infrared, or visible spectrum remote sensing measurements cannot provide. Despite some inherent advantages, however, space-based gravimetric retrievals

are limited by their coarse ($\geq 150,000$ km²) spatial and (approximately monthly) temporal resolution. In addition, gravimetric retrievals lack the vertical resolution necessary to deduce whether the measured mass is associated with snow on the surface, soil moisture in the subsurface, or with any other hydrologic storage component that contributes to TWS.

[3] Recent studies merged space-based gravimetric retrievals with land surface models in order to improve hydrologic state estimates via application of a statistical conditioning procedure [Zaitchik *et al.*, 2008; Su *et al.*, 2010; Forman *et al.*, 2012; Houborg *et al.*, 2012; Li *et al.*, 2012]. These studies employed retrievals from the gravity recovery and climate experiment (GRACE) within a data assimilation (DA) framework. While these studies demonstrated the added value associated with utilizing GRACE TWS retrievals as part of a DA procedure, the conclusions made were limited by the amount of available ground-based observations for use during validation. For example, Su *et al.* [2010] and Forman *et al.* [2012] showed that snow water equivalent (SWE) estimates could be improved across portions of North America, but neither study could draw definitive conclusions regarding groundwater or soil moisture due to the sparsity of available in situ observations. Analogously, Zaitchik *et al.* [2008] and Houborg *et al.* [2012] discussed groundwater and soil moisture

¹Department of Civil and Environmental Engineering, University of Maryland, College Park, Maryland, USA.

²Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

Corresponding author: B. A. Forman, Department of Civil and Environmental Engineering, University of Maryland, 1173 Glenn Martin Hall, College Park, MD 20742-3021, USA. (baforman@umd.edu)

estimates in portions of the continental United States, but due to the sparsity of in situ observations in these areas, point-scale observations that are, in general, only representative of an area covering a few km² (or less) were often compared to model-derived estimates that covered thousands of km² (or more).

[4] Additionally, none of these previous studies investigated in detail the impacts of the horizontal model error correlation length or the spatial averaging scale at which the GRACE TWS retrievals were assimilated. The spatial averaging scale was selected somewhat subjectively without any particular guidance on how best to assimilate TWS retrievals as a function of spatial scale. Most TWS retrieval products result from an amalgam of GRACE overpasses [Wahr *et al.*, 2004] and are typically provided on a 1° × 1° grid (~10⁴ km²). At the 1° × 1° scale, however, TWS retrievals contain significant spatial error correlations associated with truncation [Chambers, 2006] and leakage [Landerer and Swenson, 2012] errors. Conversely, the spatial scale that is sufficiently resolved by GRACE (i.e., scale at which spatial error correlations are irrelevant) is ~10⁵ km² [Rowlands *et al.*, 2005; Famiglietti and Rodell, 2013]. Since almost all land data assimilation systems (including the system used here) assume that observation errors are uncorrelated, assimilation of TWS retrievals on a 1° × 1° grid is nontrivial and generally avoided in order to reduce computing infrastructure requirements. Therefore, 1° × 1° grid retrievals are often aggregated prior to data assimilation to a spatial scale at which spatial error correlations are negligible. For example, Zaitchik *et al.* [2008] assimilated GRACE TWS retrievals at scales greater than ~10⁵ km².

[5] While it may be argued that retrievals should be assimilated at the subbasin (~10⁵ km²) scale rather than at the basin (~10⁶ km²) scale, the presence of nonlinearities in the system does not make this immediately obvious. Furthermore, the DA system requires input *model* error spatial correlation scales that are difficult to specify. This study is designed to provide guidance to the land data assimilation community on the treatment of spatial error correlations in TWS retrievals and in the modeling system in the context of data assimilation. The performance goal is to yield the best possible estimates at the model (~10³ km²) scale using GRACE TWS retrievals at a monthly, subbasin scale with the caveat that GRACE TWS retrievals can only add information at a monthly, subbasin scale or greater (that is, ≥10⁵ km²), consistent with their temporal and spatial resolution. Science questions addressed here include:

[6] 1. Does the spatial scale of the GRACE TWS retrievals impact DA performance at the fine-scale model resolution? Is there a preferred spatial scale to which TWS retrievals should be aggregated prior to assimilation?

[7] 2. Does the input horizontal *model* error correlation scale impact the ability of the DA routine to transfer information from the coarse-scale (~10⁵ km²) retrievals to the fine-scale (~10³ km²) model space? If so, what is the most applicable model error correlation length scale that should be used?

[8] 3. Can a DA procedure effectively add vertical *and* horizontal resolution to TWS retrievals via application of a fine-scale prognostic model?

[9] Section 2 outlines the methodology and study domain used here, including the setup and validation of the

synthetic twin experiment. Section 3 highlights the results from the open-loop and assimilation experiments, including a discussion of the relevant components of the hydrologic cycle. Section 4 discusses the conclusions of the study, provides answers to the science questions listed above, and considers implications for future studies that assimilate gravimetric TWS retrievals.

2. Methodology

[10] The following describes the setup for the synthetic twin experiment used in this study. Only the essential details of the synthetic twin experiment are discussed here. Further information regarding the assimilation system can be found in section 3d of Zaitchik *et al.* [2008] as well as in Forman *et al.* [2012].

2.1. Data Assimilation Framework

[11] The DA framework employs an ensemble-based smoothing approach that consists of a two-step process as illustrated in Figure 1. During Step 1, the nonlinear, prognostic land surface model propagates the model states forward in time from 1 month to the next using an ensemble of realizations with prescribed model errors (see section 2.3, for more details). Assumed model errors are represented by perturbations that are applied to both model states and forcings. The prior model states are then updated using synthetic retrievals available for a given time period of interest (Step 2). A linear update equation is employed to update the model states using a weighting factor based on the uncertainty in the prior states and the retrievals. The increments obtained from the update are then added to the model states during a second model integration for this month. As illustrated in the right-hand side of Figure 1, the second round of model integration begins anew. During this second integration, the monthly analysis increments are divided by the number of days within the given month and then added on a daily basis. The monthly resolution of the TWS retrievals is what ultimately dictates the calculation of the analysis increments at the monthly timescale. Our approach to disaggregate each increment evenly across the days of the month reflects this underlying resolution because the resulting application of the analysis increments at the daily is consistent with the original, monthly averaged TWS changes observed by the GRACE satellite system. After the second round of model integration, the update procedure is repeated for the next month using the next set of available retrievals. The application of the analysis increments in the ensemble smoother formulation used here differs from the more traditional ensemble Kalman smoother that was discussed in Dunne and Entekhabi [2005]. These differences are discussed in detail in section 3d of Zaitchik *et al.* [2008].

2.2. Study Domain

[12] The study domain is the Mackenzie River basin (MRB) located in northwestern Canada (Figure 2). As a whole, MRB is ~1.8 × 10⁶ km² in drainage area (~1.6 × 10⁶ km² for land areas only; see Table 1) with the main branch of the Mackenzie River running from the highlands in the southwestern corner of the domain northward toward the Arctic Ocean. The snow classification scheme of Sturm *et al.* [1995] suggests that MRB snow is dominated by

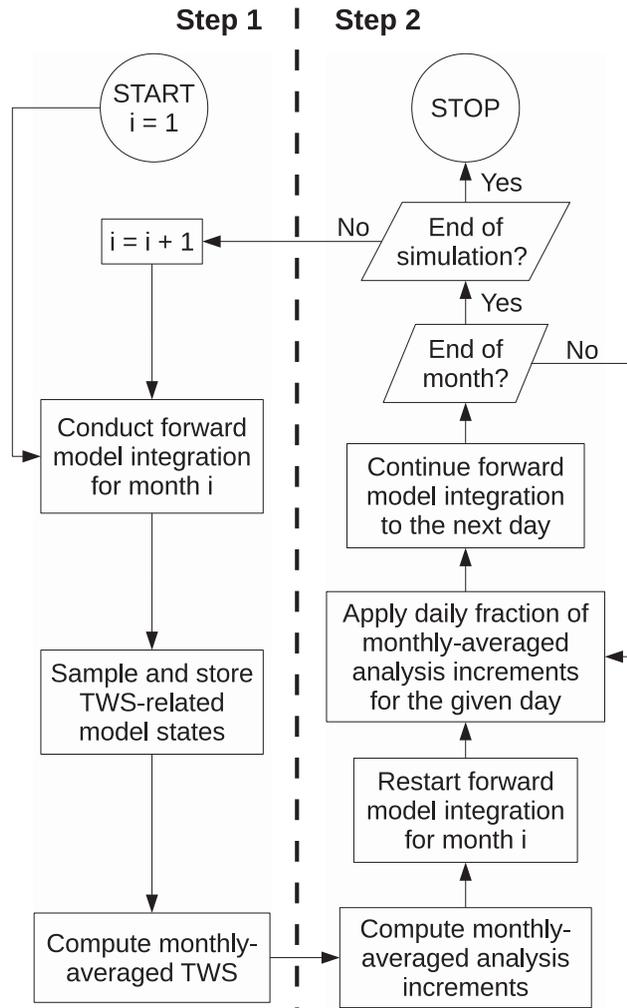


Figure 1. Simplified flowchart of ensemble-based smoothing application reproduced from *Forman et al.* [2012].

taiga-type snow with smaller areas of tundra and alpine-type snow found in the northwest and southern regions, respectively (not shown).

[13] Figure 2 shows the MRB discretized into six (6) subbasins for which details are provided in Table 1. Discretization was based on topographic control and adhered to the topology of the river network. The smallest subbasin, Peel, is 200,000 km², which is larger than the minimum area of roughly 150,000 km² that can be resolved by GRACE at midlatitudes [Rowlands et al., 2005; Swenson et al., 2006]. Synthetic retrieval preprocessing included a monthly averaged TWS estimate for each of these subbasins, which is discussed in more detail in section 2.4.1.

2.3. Prognostic Land Surface Model

[14] The prognostic model used in this application is the Catchment land surface model (Catchment) developed by *Koster et al.* [2000]. Individual watersheds in the Catchment model are discretized into “tiles” where the average size of a tile in the study domain is approximately 3000 km². The Catchment model employs a catchment deficit prognostic variable that accounts for shallow groundwater

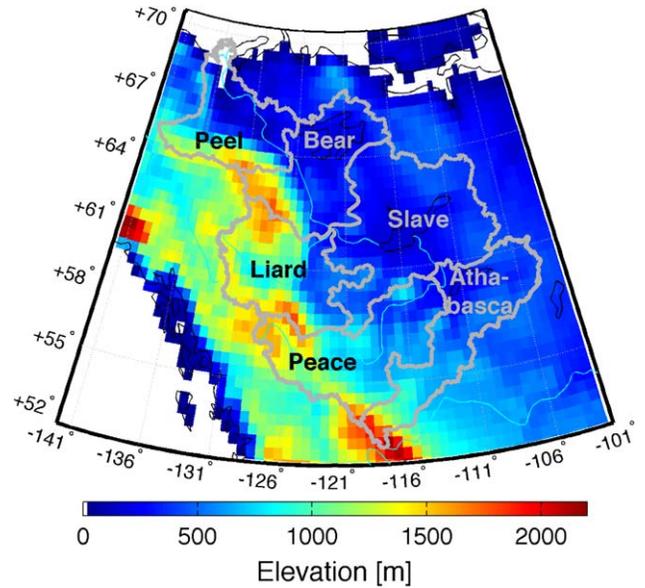


Figure 2. Map of Mackenzie River basin including the delineation of six (6) subbasins according to river network topology.

and soil moisture. The model also accounts for nonequilibrium conditions in the unsaturated zone and explicitly models subtile scale soil moisture variability and its effect on hydrological processes such as runoff and evaporation. The inclusion of a three-layer snow model [Stieglitz et al., 2001] provides additional capability in the estimation of TWS in areas where snow is a significant contributor to the hydrologic cycle. The observation operator, \mathbf{H} , maps the model states into retrieval space via spatiotemporal aggregation of the model estimates in horizontal “tile” space (subdiurnal; 10³ km²) and vertical integration of the unconfined water table, root zone soil moisture, surface soil moisture, SWE, and vegetative canopy interception model states [Forman et al., 2012] into subbasin or basin scale TWS estimates that range from $\sim 10^5$ to $\sim 10^6$ km². Even though riverine and lake storage can be a significant contributor to

Table 1. Subbasin Delineation and Assimilated Retrieval Discretization for the MRB (Land Areas Only) Along With Applied GRACE TWS Retrieval Error Covariance, \mathbf{R}

Subbasin Name	Land Area (10 ⁵ km ²)	\mathbf{R} (mm ²)
<i>1 Assimilated Retrieval</i>		
Entire Mackenzie	16.1	8 ²
<i>2 Assimilated Retrievals</i>		
Slave + Peace + Athabasca	9.3	12 ²
Liard + Bear + Peel	6.8	12 ²
<i>4 Assimilated Retrievals</i>		
Liard	2.8	20 ²
Peace + Athabasca	5.7	14 ²
Slave	3.6	20 ²
Bear + Peel	4.1	14 ²
<i>6 Assimilated Retrievals</i>		
Liard	2.8	20 ²
Peace	3.2	20 ²
Athabasca	2.6	20 ²
Slave	3.6	20 ²
Bear	2.1	20 ²
Peel	2.0	20 ²

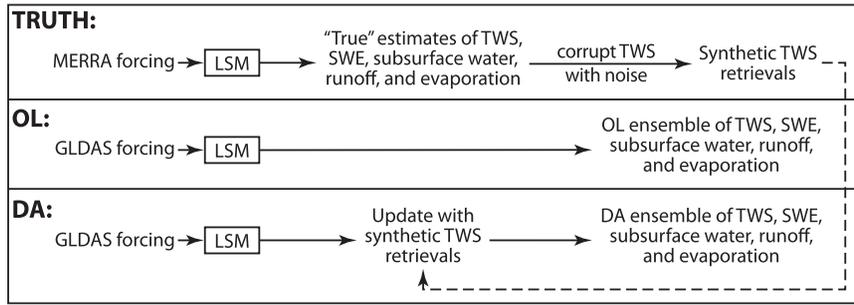


Figure 3. Conceptual framework for the synthetic twin experiment, including forcings, land surface model (LSM), synthetic TWS retrieval generation, and synthetic TWS retrieval assimilation.

TWS, the Catchment model does not currently account for storage changes within surface water impoundments. Surface meteorological forcing fields used as boundary conditions for Catchment are discussed in detail in sections 2.4.1 and 2.4.2.

[15] Model spin-up and initialization consisted of a two-step approach. The first step involved a five-time repeat of a 4 year (i.e., 1 May 2001 to 1 May 2005) cycle for a single replicate without model perturbations in order to yield a climatologically reasonable estimate of TWS on 1 May 2001. The second step involved running the model as an open-loop (OL) ensemble (see section 2.4.2, for discussion of model perturbations) from 1 May 2001 to 1 September 2001 in order to produce an adequate amount of uncertainty (spread) within the OL ensemble and to spin up the cross correlations between different state variables. From 1 September 2001 to 1 September 2010, the model was run in either OL mode or DA mode. Finally, an ensemble size of 16 was selected based on the convergence of the TWS standard deviation of the prior ensemble [Forman *et al.*, 2012].

2.4. Synthetic Twin Experiment

[16] The twin experiment started with a model integration that served as the “truth” and effectively represents nature. This simulation served as the basis for synthetic truth estimates of relevant hydrologic states and fluxes as well as for the generation of noisy synthetic TWS retrievals used during assimilation (section 2.4.1). “Designed” errors were then imposed in the modeling system by replacing the forcing data set used in the synthetic truth simulation with an alternative forcing data set (see discussion in section 2.4.2). “Assumed” errors were subsequently generated via random perturbations about the forcing and prognostic variables, which were intended to *represent* the “designed” errors in the modeling and assimilation system (see discussion in section 2.4.2). Ensemble open-loop (OL) simulations were conducted without the assimilation of synthetic TWS retrievals and subsequently compared to ensemble DA simulations where synthetic TWS retrievals were employed. A schematic for the conceptual framework of this synthetic twin experiment is provided in Figure 3.

2.4.1. Synthetic Truth and Synthetic Retrievals

[17] Generation of the synthetic truth involved a single replicate simulation by Catchment. The Goddard Earth Observing System Version 5.2.0 Modern-era retrospective analysis for research and applications (MERRA) [Rienecker *et al.*, 2011] product was used to force the land

surface model, which is provided at an hourly temporal resolution and a $1/2^\circ \times 2/3^\circ$ (latitude/longitude) spatial resolution. Hydrologic states (i.e., groundwater, soil moisture, and SWE) and fluxes (i.e., runoff and evaporation) estimated from the synthetic truth simulation were used during the validation activities to assess DA performance, which is discussed in more detail in sections 2.4.2 and 2.4.3.

[18] The synthetic truth simulation was also used to generate the synthetic TWS retrievals for application within the DA routine. Generation of these retrievals involved a two-step process: (1) truth-derived time series of TWS were computed using the linear observation operator, \mathbf{H} , described in section 2.3, and (2) the TWS time series were then corrupted with a prescribed amount of zero-mean, additive Gaussian noise that is representative of TWS retrieval error. It is assumed that the monthly averaged TWS retrieval errors are temporally uncorrelated [Zaitchik *et al.*, 2008; Su *et al.*, 2010; Forman *et al.*, 2012] and that the retrieval errors are horizontally uncorrelated for all scales of assimilated retrieval discretization used in this study. Estimated horizontal error correlations were computed as $\rho(b_i, b_j) = \exp\{-d(b_i, b_j)/\lambda_{o,ret}\}$ where $d(b_i, b_j)$ is the distance between points b_i and b_j and $\lambda_{o,ret}$ is the retrieval error correlation length scale. The exponential decorrelation function is based on recommendations found on the NASA Tellus website at <http://gracetellus.jpl.nasa.gov/data/gracemonthlymassgridsland/>. Assuming a retrieval error correlation length scale of $\lambda_{o,ret} = 300$ km and the measured distances between the centroids of the six subbasins listed in Table 1, we find that spatial error correlations between TWS retrievals for neighboring subbasins are less than 0.15.

[19] Most TWS retrieval products are provided on a discretized grid that must be aggregated up to the Mackenzie River basin as a whole or to one of its subbasins prior to utilization by the DA framework. As TWS retrievals are aggregated, the accuracy of the assimilated retrievals generally improves. In other words, TWS retrievals are a balance between accuracy and spatial resolution [Landerer and Swenson, 2012]. Assuming that retrieval errors are spatially uncorrelated at the finest subbasin scale discretization (Table 1) and that the subbasins are approximately equal in area, the error variance of an aggregated, coarse-scale retrieval is:

$$\sigma_c^2 = \frac{1}{B^2} \sum_{b \in B} \sigma_{f,b}^2, \quad (1)$$

where σ_c^2 and σ_f^2 are the retrieval error variances at the coarse and fine scale, respectively, and B is the total number of fine-scale (subbasin) retrievals aggregated into the coarser scale retrieval. The increase in error variance with decreasing spatial scale does not mean TWS retrievals contain less information, per se, but rather results from the aggregation via equation (1). If we assume the retrieval error variance for each of the finest scale subbasins (Table 1) is $\mathbf{R} = 20^2 \text{ mm}^2$, which is identical to the retrieval error variance used in *Zaitchik et al.* [2008] and *Su et al.* [2010], then the right-most column in Table 1 lists the retrieval error variance applied as a function of retrieval scale.

[20] While it may be argued that retrievals should be assimilated at the subbasin ($\sim 10^5 \text{ km}^2$) scale rather than at the basin ($\sim 10^6 \text{ km}^2$) scale, the presence of nonlinearities in the system does not make this immediately obvious. The *analysis* error variance is guaranteed to decrease as the spatial scale of the assimilated retrievals decreases *only* if the dynamic model is linear, the errors are Gaussian, the errors are mutually independent, and a number of other constraints are met. However, these assumptions are rarely satisfied in hydrologic data assimilation. It is therefore not clear at which scale the TWS retrievals should be assimilated, which makes the guidance provided in this manuscript valuable.

2.4.2. Data Assimilation Experimental Setup

[21] As mentioned previously, designed error in the modeling system is imposed through the use of imperfect forcing fields from the global land data assimilation system (GLDAS) [*Rodell et al.*, 2004] as utilized in the heritage NASA Global Modeling and Assimilation Office seasonal forecasting system 3 hourly temporal and $2.0^\circ \times 2.5^\circ$ (latitude/longitude) spatial resolution. Significant climatological (9 year) differences were found for precipitation and downwelling shortwave radiation between the GLDAS and MERRA forcing data sets within each of the six individual subbasins. Therefore, a bias correction strategy was employed such that the GLDAS precipitation and downwelling shortwave radiation were rescaled to match corresponding climatological values from MERRA. Climatological differences in downwelling longwave radiation were negligible (i.e., $\sim 0.5\%$ or less) hence no longwave bias correction strategy was deemed necessary.

[22] Since each of the six subbasins was bias corrected individually, aggregation of the individual subbasins will also be bias corrected by construct. However, even though the total amount of precipitation or shortwave radiation that impacts a given subbasin is identical between MERRA and GLDAS as a result of the bias correction strategy, significant differences in seasonality, synoptic scale variability, and flux intensity still exist. The nonlinear hydrologic response associated with differences in precipitation and shortwave radiation timing and intensity results in significant differences in subbasin TWS. Remaining differences in TWS between the synthetic truth and the OL simulations should ideally be mitigated via assimilation of the synthetic TWS retrievals.

[23] Synthetic TWS retrievals are assimilated into ensemble model integrations. For the ensemble integrations, perturbations to select model states and forcings were prescribed to represent assumed model errors. Both multiplicative and additive perturbations were specified as listed in

Table 2. Parameters for Perturbations to Meteorological Forcing Inputs and Model Prognostic Variables^a

Perturbation	Type	Standard Deviation	Units	AR(1) (day)
Precipitation	M	0.5		3
Shortwave radiation	M	0.5		3
Longwave radiation	A	50	W m^{-2}	3
Snow water equivalent ^b	M	0.0004		1
Catchment deficit	A	0.05	mm	1
Surface excess	A	0.02	mm	1

^aThe horizontal error correlation length scale, λ , is included in the experimental design, but is omitted from the table.

^bPerturbations made to all three (3) snow layers; M = Multiplicative; A = Additive; AR(1) = first-order auto-regressive temporal correlation.

Table 2. Model state perturbations were applied at each model time step (i.e., every 20 min) and model forcing perturbations were applied at each forcing time step (i.e., every 3 h). Temporal correlations were imposed using a first-order auto-regressive model within the perturbed fields as discussed in *Reichle et al.* [2008]. Horizontal model error correlation lengths, λ , were defined for different ensemble integrations from the set $\Lambda = \{1^\circ, 2^\circ, 3^\circ, 4^\circ\}$. The value of λ represents the e-folding distance given an exponential horizontal error model. The set Λ bounds the “true” design error correlation length for TWS (error computed as GLDAS minus MERRA), which is $\lambda_o \approx 3.0^\circ$ in the MRB as estimated from computed variograms that reached a near-asymptotic value less than $1/e$ but greater than zero [*Mela and Louie*, 2001; *Forman and Margulis*, 2010]. It is therefore reasonable to expect a priori that the assimilation will perform best when $\lambda \approx 3^\circ$ since the model representation of spatial error correlations would most closely match those of the “true” design errors. In addition, values of $\lambda > 4^\circ$ were examined, but are excluded from discussion here because these results provided relatively little insight beyond those already presented in set Λ .

[24] For the computation of the analysis increments during the ensemble update step, spurious long-range error correlations were suppressed within the background (sample) error covariance terms via element-wise multiplication of a compact support function as outlined in *Gaspari and Cohn* [1999]. The methods of *Gaspari and Cohn* [1999] impose an error covariance localization. It is worth stating explicitly that the compact support function is only used during computation of the analysis increments and is not applied during generation of the forcing and prognostic variable perturbations. The isotropic, compact support scale was set to 2.5λ such that the analysis at a given location is only impacted by retrievals within a radius of 2.5λ . In other words, correlated errors greater than 2.5 times the e-folding distance are suppressed. Cross correlations between forcing perturbations were accounted for according to *Reichle et al.* [2007].

[25] A total of 16 DA experiments were conducted. This corresponds to the permutation of horizontal model error correlation lengths, $\Lambda = \{1^\circ, 2^\circ, 3^\circ, 4^\circ\}$, with the set of assimilated retrieval discretization, $ARD = \{1, 2, 4, 6\}$. ARD is representative of the assimilated retrieval scale as highlighted in Table 1.

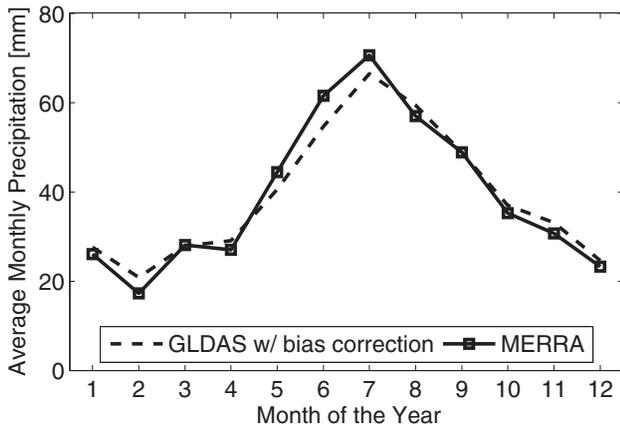


Figure 4. Monthly averaged, climatological precipitation in the MRB for the current study period.

2.4.3. Validation Approach

[26] Validation included the computation of several different metrics: (1) bias, (2) root mean squared error (*RMSE*), and (3) anomaly time series correlation coefficient, or anomaly *R*. All statistics were computed from the ensemble mean estimates and were area-weighted in model tile space such that larger tiles had a greater impact on the computed metric. Bias was computed as the model (i.e., OL or DA) minus the truth. The anomaly *R* value was computed by first determining the climatological seasonal cycle over the course of the simulation period for both the modeled ensemble mean and the synthetic truth. Next, the anomaly time series were computed by subtracting the climatological seasonal cycle from the corresponding original time series prior to the computation of the anomaly *R* values. Statistically significant differences in anomaly *R* values between the OL and DA experiments were determined using the Hotelling-Williams test [Steiger, 1980; Forman *et al.*, 2012].

3. Results and Discussion

3.1. Open-Loop Results

[27] In order to better understand the DA results, it is worthwhile to first examine the OL results. As mentioned previously, the OL simulations utilized GLDAS forcing whereas the truth utilized MERRA forcing. A bias correction procedure was implemented in order to ensure the total amount of incoming mass (precipitation) and energy (short-wave and longwave downwelling radiation) over the 9 year study period in each of the six subbasins was identical (see section 2.4.2).

[28] One significant difference between the GLDAS and MERRA forcing is illustrated in Figure 4. MERRA contains more precipitation than GLDAS during the peak (summer) rainfall season whereas GLDAS contains more precipitation than MERRA during the winter, which often falls as snow. More precipitation during the peak summer season, in general, equates to stronger, more intense rainfall events. This behavior is further enhanced by the fact that the spatial and temporal resolution of MERRA is much finer than that of GLDAS, which generally results in the presence of more intense rainfall events in MERRA

whereas the same rainfall event is effectively “smoothed out” in space and time in GLDAS. Collectively, these two attributes result in a more rapid hydrologic response when MERRA forcing is utilized, which produces a larger fraction of runoff relative to infiltration.

[29] Figure 5 illustrates this behavior. The left-most column shows the computed bias between the simulated (i.e., GLDAS-based) and true (i.e., MERRA-based) output. Similarly, the middle and right-most column shows the *RMSE* and anomaly *R*, respectively. Each row of subplots corresponds to a particular hydrologic state or flux. Each individual subplot contains the results from the OL simulations along with the DA simulations as a function of assimilated retrieval discretization, *ARD*. The different lines within each individual subplot correspond to different values of the horizontal model error correlation length, λ . Focusing on just the computed bias for only the OL simulations, the difference in hydrologic response associated with the GLDAS and MERRA forcings is readily apparent. The more intense precipitation events in MERRA, particularly during the summer, result in more runoff in the truth relative to OL, which ultimately produces a negative bias in runoff in the OL simulations. Less runoff associated with less intense rainfall implies more opportunity for water retention within the basin, which is in agreement with the positive TWS bias in the OL simulations. This is further shown in the subsurface water bias, which clearly highlights the behavior that less intense rainfall events in GLDAS allow for a significantly larger fraction of the water to infiltrate into the subsurface, which yields a positive bias in the subsurface relative to the truth.

[30] Based on Figure 4, the OL SWE bias in Figure 5 appears counterintuitive. However, upon closer inspection of the spatial distribution of the precipitation forcing in MERRA versus GLDAS, these results are justifiable. An examination of the spatial distribution of the precipitation in each subbasin over the 9 year study period (results not shown) indicates a significantly larger amount of snow accumulation along the southwestern border of the MRB. Probably due to its higher spatial resolution, MERRA generates more snowfall in the mountainous regions of the Liard and Peace subbasins where, in general, SWE is greatest. Even though the total amount of snowfall, in general, is less in the truth than in the OL across the MRB as a whole, the enhanced accumulation of SWE in the areas where SWE is greatest results in a snowpack that takes longer to completely melt and ablate. This extension of the snow season, so to speak, in the subbasins where SWE is the greatest yields the negative bias for SWE as shown in Figure 5.

[31] The final outcome worth highlighting in the OL simulations is related to evaporative flux. Increased infiltration associated with GLDAS forcing, in general, increases the amount of surface and root zone soil moisture available for use in evaporative processes. The positive bias in the subsurface of the OL simulations is clearly reflected in the positive bias in the evaporation. Additionally, there are small variations in the OL results depending on the horizontal model error correlation length, λ , used during prognostic and forcing variable perturbation. This is not only seen in the evaporation results, but also in the other states and fluxes displayed in Figure 5. The differences associated with different λ values are small relative to the differences

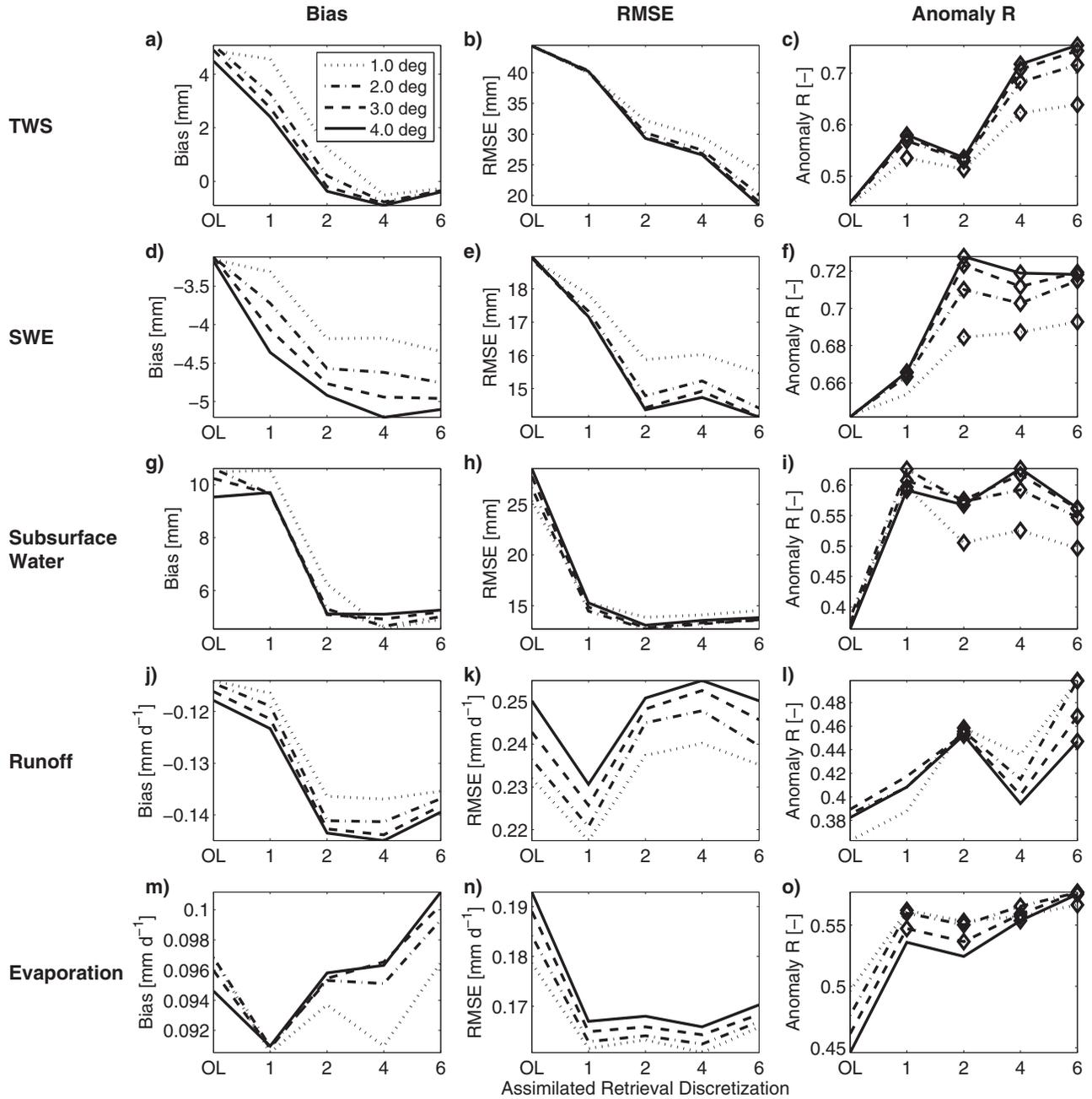


Figure 5. Subbasin area-weighted statistics for the Mackenzie basin. The different lines represent different horizontal model error correlation lengths as shown by the legend in the top-left subplot. Black diamonds represent statistically significant anomaly R improvements (relative to the OL simulation) for a majority of the six subbasins that comprise the Mackenzie as a whole.

between utilization of the MERRA versus GLDAS forcing and are relatively insignificant. These small differences arise in part due to the finite ensemble size, I , used in this application. In the limit as $I \rightarrow \infty$, the differences in OL results associated with different λ values should approach zero.

3.2. Data Assimilation Results

3.2.1. Terrestrial Water Storage

[32] In this section, we proceed with the investigation of the impacts on DA results of the assumed horizontal model

error correlation length, λ , and the assimilated retrieval discretization, ARD . Starting with TWS shown along the top row of Figure 5, we see significant improvements as the assimilated retrieval discretization increases (i.e., as finer-scale TWS retrievals are assimilated). The bias is reduced to near-zero when the synthetic retrievals are assimilated at the scale of two or more subbasins. The $RMSE$ decreases monotonically as the spatial scale of the assimilated retrievals decreases from an OL value greater than 40 mm to a value less than 20 mm at $ARD = 6$. This reduction in uncertainty occurs even though the error variance in the

assimilated retrievals increases as the retrieval scale is reduced (Table 1). Note that this is not a trivial result in the presence of model nonlinearities. In the presence of model nonlinearities, the aggregation of $1^\circ \times 1^\circ$ retrievals to the subbasin ($\sim 10^5 \text{ km}^2$) scale rather than the basin ($\sim 10^6 \text{ km}^2$) scale allows the DA framework to capture more information. (As discussed earlier in the context of equation (1), this result considers the change in the retrieval error variance with the spatial aggregation.)

[33] As the horizontal *model* error correlation length is increased from $\lambda = 1^\circ$ to $\lambda = 4^\circ$, we see additional improvements, but these improvements are small compared to the improvements with decreasing the retrieval scale. Collectively, continual improvements in bias and *RMSE* of the posterior TWS estimate are witnessed as the scale of the assimilated TWS retrievals approaches the finest subbasin scale tested here. In this study, it is assumed that the minimum spatial scale that can be sufficiently resolved by GRACE in the MRB is approximately $200,000 \text{ km}^2$ (see Table 1), which occurs when spatial error correlations between subbasins are insignificant. Note that this estimate is not universal and depends on basin shape, basin topography, proximity to ablating glaciers, proximity to oceans, proximity to regions with postglacial rebound, the half-width of the Gaussian smoothing kernel used as part of the TWS retrieval algorithm, and a host of other factors.

[34] Improvements to the anomaly *R* are also greater, in general, as the retrieval scale is decreased and λ is increased. In other words, the best TWS results are achieved when assimilating TWS retrievals for the six subbasins rather than for a TWS retrieval of the basin as a whole. Anomaly *R* increased from $R \approx 0.45$ for the OL simulations to $R \geq 0.70$ at $ARD = 6$ when using $\lambda \geq 2^\circ$. The black diamonds in the anomaly *R* subplot denote improvements within a majority of the MRB subbasins that are statistically different from the OL at the 5% level based on the Hotelling-Williams significance test [Steiger, 1980; Forman et al., 2012].

[35] In general, the greatest gains in TWS estimation via assimilation of gravimetric retrievals are achieved when applying the smallest retrieval scale that can be sufficiently resolved by the retrievals (i.e., the smallest spatial scale at which the retrievals are spatially uncorrelated). A secondary gain is achieved when applying $\lambda \geq 2^\circ$. Further, the majority of these gains are statistically significant. It is worthwhile noting here that TWS assimilation can result in statistically significant *degradation* in some of the subbasins for some of the hydrologic fluxes (e.g., runoff discussed in section 3.2.3), but that the black diamonds in Figure 5 represent statistically significant *improvements* within a *majority* of the MRB subbasins.

3.2.2. Snow Water Equivalent and Subsurface Water Storage

[36] As shown in the second row of Figure 5, assimilation of TWS retrievals has the tendency to remove a small amount of SWE. The negative bias in SWE is made more negative (albeit by only 2 mm or less) as a result of the DA procedure. This is generally due to the removal of SWE at or near peak accumulation during the snow season [Forman et al., 2012]. However, even though the SWE bias becomes slightly worse, the *RMSE* is improved by $\sim 20\%$ for $ARD \geq 2$ and values of $\lambda \geq 2^\circ$. That is, even though the

bias is degraded via the DA procedure, the *RMSE* is reduced, which is indicative of increased confidence in the posterior estimate. In addition, the anomaly *R* is significantly improved from the OL values of $R \approx 0.64$ to $R \geq 0.70$ for $ARD \geq 2$ and $\lambda \geq 2^\circ$.

[37] The greatest improvements from DA were witnessed in the subsurface water storage (i.e., soil moisture and groundwater) relative to the surface state variables and surface fluxes. The third row of Figure 5 shows the bias is reduced by half from $\sim 10 \text{ mm}$ in the OL simulations to $\sim 5 \text{ mm}$ in the DA simulations when $ARD \geq 2$. The bias remains relatively unchanged for $ARD = 1$ even though the *RMSE* is dramatically reduced from $\sim 30 \text{ mm}$ in the OL simulations to $\sim 15 \text{ mm}$ or less at all values of ARD in the DA simulations. This is evidence that the DA procedure is improving the subsurface storage estimate no matter the scale of the assimilated retrievals or the horizontal model error correlation length used, but that the greatest gains are achieved in both bias and *RMSE* when $ARD \geq 2$. In much the same manner as witnessed with the *RMSE* results, the anomaly *R* results are improved significantly at all values of ARD . These results, in conjunction with the SWE results, suggest the DA procedure simultaneously improves surface and subsurface state estimates, thereby implicitly adding vertical resolution to the TWS retrievals as part of the DA procedure. Additional discussion regarding the partitioning of the TWS retrievals between the SWE and subsurface components is found in section 3.2.4.

3.2.3. Runoff and Evaporation

[38] In addition to improvements in hydrologic state estimation, it is useful to analyze the effects of the state updates on hydrologic flux estimation. We emphasize here that the fluxes are not directly updated in the DA framework, but do change as a result of the state vector update. The fourth row of Figure 5 shows the OL and DA results for runoff. As previously discussed in section 3.1, differences in precipitation timing and intensity between MERRA and GLDAS yield differences in the amount of runoff associated with different hydrologic responses. The DA update further exacerbates the negative bias in the runoff for two reasons: (1) SWE is removed, on average, across the MRB, which means less SWE is available to produce snow melt runoff, and (2) subsurface water is removed, much of which is in the form of soil moisture, which results in more infiltration and hence less runoff during hydrologic partitioning at the land surface.

[39] Runoff bias becomes more negative with increasing ARD . *RMSE*, on the other hand, remains unchanged as a function of ARD (except for $\lambda = 1$) and can even be degraded (albeit by only 5%) relative to the OL. This behavior is, in large part, due to the degradation of runoff bias that is ultimately reflected in the *RMSE*. The anomaly *R*, in general, improves with increasing ARD , but only a handful of occurrences are statistically different at the 5% level. Even though the impact of TWS assimilation on hydrologic runoff estimation is relatively small compared to the effects on hydrologic state estimation, it is worthwhile highlighting the result that the best performance from the DA routine on runoff estimation occurred when λ was smallest, which is contrary to the state estimation results where, in general, the best performance occurred when λ was greatest.

[40] As mentioned in section 3.1, the positive bias in the subsurface water results in a positive bias in evaporation. As shown in the bottom row of Figure 5, assimilation of TWS retrievals had little or no effect on the evaporation bias and was relatively insensitive to the horizontal model error correlation length. *RMSE*, on the other hand, was reduced by a small amount regardless of the spatial scale of the assimilated retrievals. The anomaly *R* values were improved via assimilation and were significantly different from the OL results in most instances. However, as with the *RMSE* results, the anomaly *R* results were relatively insensitive to *ARD*. In an analogous manner as found with runoff, the best performing DA results occurred when λ was smallest (i.e., $\lambda = 1^\circ$) as opposed to the state estimates that were mostly improved when λ was largest (i.e., $\lambda = 4^\circ$). The exact reason for this behavior remains elusive, but appears to be consistent for both of the fluxes investigated in this study. It is possible that the energy-limited regime found in much of the Mackenzie River basin exerts less control over changes in evaporative flux resulting from TWS assimilation than for a more water-limited regime such as that found in semiarid areas.

3.2.4. Analysis Increments

[41] An evaluation of the analysis increments (i.e., $\mathbf{x}_\tau^+ - \mathbf{x}_\tau^-$) provides useful information about the behavior of the TWS assimilation routine [Forman et al., 2012]. It allows for mass to be tracked during exchanges between relevant components of the hydrologic cycle as well as for any water imbalance generated during the filter update to be investigated and better understood. In this study, the temporally averaged, basin-averaged analysis increment for the entire MRB, I_{MRB} , was computed as

$$I_{MRB} = \sum_{\tau \in T} \sum_{b \in B} w_b \cdot (\langle \mathbf{x}_{b,\tau}^+ \rangle - \langle \mathbf{x}_{b,\tau}^- \rangle), \quad (2)$$

where $\langle \mathbf{x}_{b,\tau}^+ \rangle$ is the ensemble mean of the posterior estimate (of SWE or subsurface water) for a given subbasin b and month τ , $\langle \mathbf{x}_{b,\tau}^- \rangle$ is the same but for the prior estimate, w_b is a factor for area-weighted averaging ($\sum_{b=1}^B w_b = 1.0$), and T is the total number of months in each simulation. A positive value for I_{MRB} suggests mass has been added to the system as a whole whereas a negative value for I_{MRB} suggests mass has been removed.

[42] Figure 6 shows the results of the time-averaged, basin-averaged analysis increments for the SWE and subsurface components in the different DA experiments as a function of the horizontal model error correlation length, λ , and the spatial scale of the assimilated retrievals, *ARD*. Figure 6a shows the results as computed using equation (2). Figure 6b shows similar results but computed with an absolute value operator applied to the analysis increments inside the parentheses. As is shown in Figure 6a, the cumulative increments vary considerably as a function of *ARD* but are relatively insensitive to λ . As *ARD* increases, SWE is increasingly removed during the analysis update. Correspondingly, on a relative scale, less mass is removed from the subsurface. In other words, as the spatial scale of the assimilated TWS decreases, more mass is removed from the snowpack near peak accumulation while less mass is removed from the soil moisture and/or groundwater stores during spring runoff and ablation, which has the overall

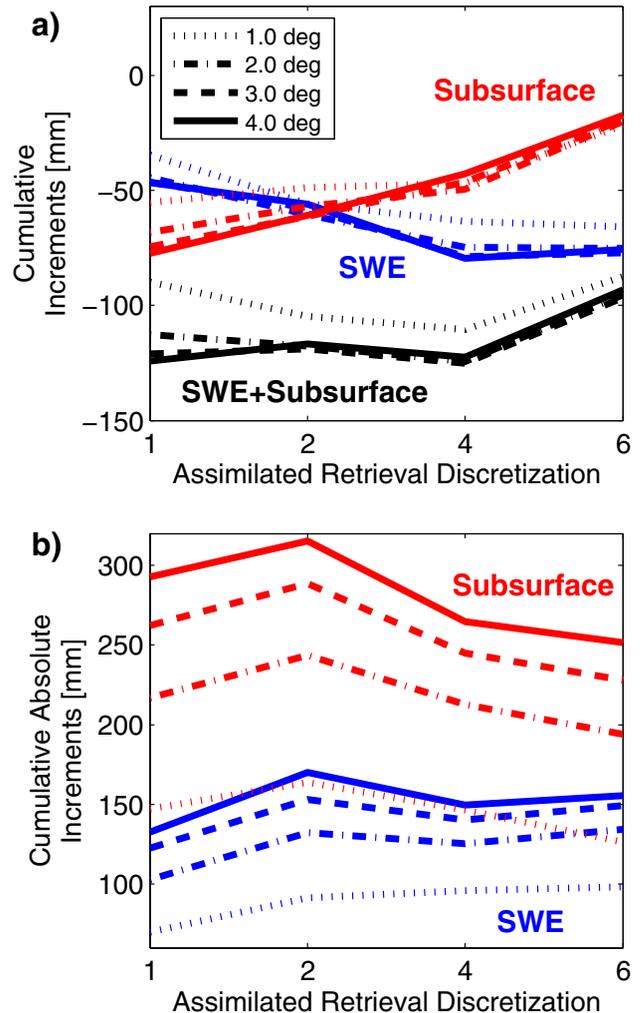


Figure 6. MRB-averaged (a) cumulative analysis increments and (b) cumulative *absolute* analysis increments over the 9 year simulation for (blue) SWE, (red) subsurface water, and (black) SWE + subsurface water for different values of the horizontal model error correlation length, λ , and assimilated retrieval discretization, *ARD*.

effect of dampening the runoff response during the melt season. Further, mass is collectively being removed from the system during the analysis as indicated by the negative values of the cumulative increments for all values of λ and *ARD*. The systematic removal of mass from the hydrologic system introduces a small water imbalance in the MRB, which was also witnessed in Forman et al. [2012]. The water imbalance, however, is smallest when assimilating TWS retrievals at the smallest spatial scale.

[43] Figure 6b provides additional evidence on the behavior of the update routine. When investigating the cumulative *absolute* analysis increments, two features are apparent: (1) the cumulative absolute value of the subsurface increments is significantly greater than the cumulative absolute value of the SWE increments for a given λ , and (2) the cumulative absolute value increases monotonically with increasing λ . Even though the magnitude of the cumulative increments for the subsurface illustrated in Figure 6a, particularly at $ARD \geq 4$, are smaller than those for SWE,

Figure 6b suggests that on average, for a given time step, the magnitude of the analysis increment is much greater for the subsurface water than for SWE and that a larger portion of the analysis increment is being applied to subsurface water. This seems reasonable given that the subsurface contains more water (i.e., a larger fraction of the TWS) than what is found in SWE and, in part, helps explain why TWS assimilation in the MRB generated relatively small changes in the posterior SWE estimate [Forman *et al.*, 2012].

[44] The increase in the cumulative *absolute* increments with increasing horizontal model error correlation length suggests that the multiplicative effect of the gain matrix (i.e., weighting factor based on the uncertainty in the prior states and the predicted retrievals) times the innovation (i.e., difference between the TWS retrieval and the predicted TWS estimate) increases as λ increases. This implies that an increase in the horizontal length scale of the errors (perturbations) applied to the prognostic and forcing variables increases the background error covariance while the retrieval error covariance, \mathbf{R} , remains unchanged, which ultimately produces a larger magnitude in the analysis increment. Hence, the result is more temporal variability in the analysis increments for both the subsurface water and SWE as λ is increased. This is evident in Figures 5b and 5e for $\lambda \leq 2^\circ$ and suggests that an underestimation of the horizontal error correlation length results in an underestimation of the background error covariance and hence an underestimation of the gain. This behavior quickly dissipates for $\lambda > 2^\circ$ (results for $\lambda > 4^\circ$ not shown). In addition, these results suggest TWS assimilation is relatively insensitive to the overestimation of λ , but that underestimation of the background error covariance (via improper selection of λ) adversely affects the analysis update and limits the amount of information exchange from the TWS retrievals into the conditioned model estimate.

3.2.5. Horizontal Downscaling

[45] The ability of the DA framework to effectively downscale the assimilated TWS retrievals to scales below the subbasin scale was investigated using spatial (pattern) correlations of TWS anomalies. TWS anomalies were first computed in fine-scale tile space ($\sim 10^3 \text{ km}^2$) for each model tile in the MRB. This was done separately for the truth simulation each OL simulation, and each DA simulation. The skill of each DA and OL simulation was then defined as the anomaly pattern correlation (as a function of time) relative to the truth. The anomaly pattern correlation was computed as the spatial correlation between the true anomaly and the OL or DA anomaly at each moment in time, from which the time-averaged anomaly pattern correlation could then be computed for the entire study period.

[46] Figure 7a shows the temporal average of the differences between DA skill and OL skill for each of the horizontal model error correlation lengths and assimilated retrieval scales. In general, the improvement in pattern correlation increased with increasing assimilated retrieval discretization, *ARD*. This is reasonable since a reduction in assimilated retrieval scale implicitly contains additional information regarding the spatial distribution of TWS across the domain. There is some evidence that the DA procedure effectively adds horizontal resolution to the retrievals. Namely, the DA results highlighted in Figure 7 outperform those of the OL for almost every combination of λ and

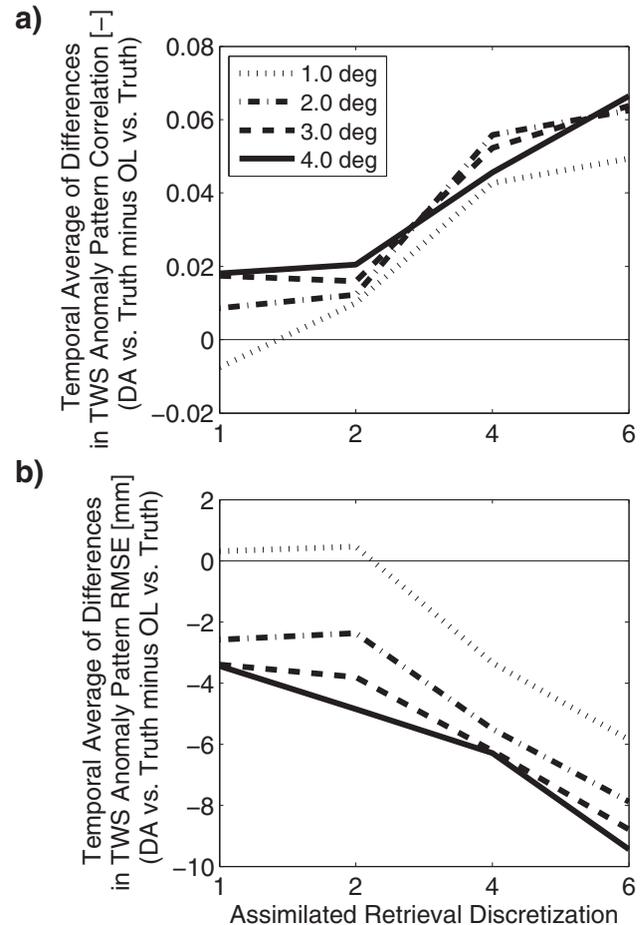


Figure 7. Temporally averaged differences (DA versus Truth minus OL versus Truth) in (a) TWS anomaly pattern correlation and (b) TWS anomaly pattern *RMSE* for the MRB for different assimilated retrieval scales and different horizontal model error correlation length scales.

ARD tested. However, when compared against the OL, the changes are not statistically significant when using a Fisher *Z* transform with a 5% confidence interval. That is, most of the fine-scale horizontal resolution in TWS is the result of applying the prognostic model in tile space and does not result from the application of analysis increments as a function of λ or as a function of *ARD*.

[47] Figure 7b provides additional evidence as to the benefits of assimilating TWS retrievals at decreasing spatial scales. The anomaly *RMSE* shown is computed in a similar manner as that of the anomaly pattern correlation except that *RMSE* is computed rather than correlation coefficient. The reduction in TWS anomaly pattern *RMSE* is essentially monotonic such that increasing *ARD* results in decreasing anomaly pattern *RMSE*. These findings further suggest assimilation of TWS retrievals at the finest subbasin scale results in the greatest amount of improvement relative to the OL.

4. Conclusions

[48] A synthetic twin experiment was conducted during which synthetic retrievals of satellite-derived TWS in the

MRB were assimilated into the NASA Catchment model. In addition to the “true” and OL simulations, a total of 16 DA experiments were conducted. The experimental matrix was based on the permutation of different values of assimilated retrieval discretization ($ARD = \{1, 2, 4, 6\}$) in conjunction with different horizontal model error correlation lengths of the perturbations applied to model prognostic and forcing variables ($\Lambda = \{1^\circ, 2^\circ, 3^\circ, 4^\circ\}$). As the spatial scale of the assimilated TWS retrieval decreases, the corresponding retrieval error for each assimilated retrieval increases. As the λ scale increased, the background (sample) error covariance used in the gain enabled horizontally distributed updates across larger regions of space via the horizontal propagation of error information.

[49] Referring back to the science questions in section 1, we can say the following regarding the spatial scale of the assimilated TWS retrieval. The efficacy of the DA routine, in general, appears most enhanced when TWS retrievals are assimilated at the smallest spatial scale at which they can be reasonably resolved (that is, the smallest scale at which the observations are spatially uncorrelated). The smallest spatial scale that can be reasonably resolved will, of course, depend on geographic location. Theoretically, basins closer to the poles could be resolved at smaller scales due to the increased number of satellite overpasses associated with a polar-orbiting platform such as GRACE. However, other factors must also be considered such as proximity to oceans and the possibility of gravimetric signals in outlying areas “leaking” into gravimetric estimates related to terrestrial hydrology processes in a specific area of interest [Tapley *et al.*, 2004; Houborg *et al.*, 2012]. Additionally, due to the application of a Gaussian smoothing kernel during GRACE TWS product generation [Chambers, 2006; Swenson and Wahr, 2006], the scale of the TWS retrieval that can be reasonably resolved is presumably smaller for basins that are relatively circular in shape as opposed to those that are thin and elongated.

[50] As for the science question regarding horizontal error correlation scale, the results for the basin used in this study suggest that state estimation was clearly enhanced using a λ where the e -folding length was long enough to encompass the basin of interest as well as some of the neighboring regions outside of the basin of interest. Namely, state estimation was most improved at $3^\circ \leq \lambda \leq 4^\circ$, but this effect was secondary relative to changes in retrieval scale. Flux estimation using the DA procedure, however, yielded the best results when $\lambda \approx 1^\circ$. However, the state variable improvements had a deleterious (and nonlinear) effect on the diagnostic fluxes. In short, the DA routine attempts to correct for improper model physics in the runoff routing scheme by retaining more water within the subsurface of the basin. The resulting change in nonlinear hydrologic response and dynamic range of model-estimated TWS better matches that of the synthetic TWS retrievals. This also agrees with the findings of Forman *et al.* [2012] where “real-world” GRACE retrievals were assimilated in the Mackenzie basin. By retaining more mass in the subsurface, the DA routine exacerbated the negative bias in runoff (via increased subsurface storage) and also exacerbated the positive bias in evaporation (via increased subsurface storage and hence increased evaporative water availability). Given that the improve-

ments to the state estimates were much more significant at large values of λ compared to the relatively small improvements in flux estimation at small values of λ , it is preferable to focus on the state estimates. Hence, the greatest potential in improving model performance via assimilation of TWS was witnessed when using a λ that fully encompassed the spatial scale of the assimilated TWS retrievals. These findings corroborate the expectation that the best state estimation results occur when $3^\circ \leq \lambda \leq 4^\circ$, which most accurately represents the “true” design error of $\lambda_o \approx 3^\circ$.

[51] In terms of effectively downscaling the TWS retrievals in space by adding vertical resolution to the GRACE TWS retrievals as part of the assimilation procedure, it is clear that the DA routine can simultaneously improve the surface and subsurface state estimates through the assimilation of TWS. The enhanced vertical resolution in the updated (conditioned) estimate arises from the fine-scale model’s ability to better resolve TWS into surface and subsurface components. The DA routine appears capable of effectively adding horizontal resolution to the TWS retrievals via application of the fine-scale prognostic model, too, but horizontal downscaling is not significant at a level of $\alpha \leq 0.05$. That is, most of the downscaling capability is associated with the merger of the fine-scale, prognostic model with the TWS retrievals and has less to do with the value of λ , which essentially only impacts the analysis increments. Overall, the DA procedure added value to the model estimate by reducing model uncertainty while implicitly adding utility to the coarse-scale, column-integrated TWS retrievals via vertical downscaling. The DA procedure does not, however, add a significant amount of information to the model estimates at horizontal resolutions below the subbasin scale ($\sim 10^5 \text{ km}^2$).

[52] Lastly, it is important to point out some of the limitations of this study. This study only focused on a northern latitude basin where snow is a major contributor to the hydrologic cycle. What is not apparent from these results is how such a study might perform where snowfall is insignificant. An additional limitation is that these results are model-specific and may not transfer to assimilation systems that use other land surface models. Extending this synthetic study to other climatic regions and/or incorporating other land surface models and assimilation systems should be considered in future studies related to the assimilation of TWS. Despite these limitations, our findings have important implications for land data assimilation systems that extract information from satellite-based TWS retrievals such as the gravity recovery and climate experiment (GRACE) for the purpose of improving regional and continental-scale freshwater resource characterization.

[53] **Acknowledgments.** Partial funding provided by the NASA Postdoctoral Program (contract NNH06CC03B). Computing was supported by the NASA High End Computing Program. We thank the anonymous reviewers and the Associate Editor for their insights and feedback that led to an improved manuscript. Additional thanks go to Gabrielle De Lannoy for many useful conversations related to this work.

References

Chambers, D. P. (2006), Evaluation of new GRACE time-variable gravity data over the ocean, *Geophys. Res. Lett.*, *33*, L17603, doi:10.1029/2006GL027296.

- Dunne, S., and D. Entekhabi (2005), An ensemble-based reanalysis approach to land data assimilation, *Water Resour. Res.*, *41*, W02013, doi:10.1029/2004WR003449.
- Famiglietti, J. S., and M. Rodell (2013), Water in the balance, *Science*, *340*(6138), 1300–1301, doi:10.1126/science.1236460.
- Forman, B. A., and S. A. Margulis (2010), Assimilation of multiresolution radiation products into a downwelling surface radiation model. 1: Prior ensemble implementation, *J. Geophys. Res.*, *115*, D22115, doi:10.1029/2010JD013920.
- Forman, B. A., R. H. Reichle, and M. Rodell (2012), Assimilation of terrestrial water storage from GRACE in a snow-dominated basin, *Water Resour. Res.*, *48*, W01507, doi:10.1029/2011WR011239.
- Gaspari, G., and S. Cohn (1999), Construction of correlation functions in two and three dimensions, *Q. J. R. Meteorol. Soc.*, *125*, 723–757.
- Houborg, R., M. Rodell, B. Li, R. Reichle, and B. F. Zaitchik (2012), Drought indicators based on model assimilated GRACE terrestrial water storage observations, *Water Resour. Res.*, *48*, W07525, doi:10.1029/2011WR011291.
- Koster, R. D., M. J. Suarez, A. Ducharme, 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.
- Landerer, F. W., and S. C. Swenson (2012), Accuracy of scaled GRACE terrestrial water storage estimates, *Water Resour. Res.*, *48*, W04531, doi:10.1029/2011WR011453.
- Li, B., M. Rodell, B. F. Zaitchik, R. H. Reichle, R. D. Koster, and T. M. van Dam (2012), Assimilation of GRACE terrestrial water storage into a land surface model: Evaluation and potential value for drought monitoring in western and central Europe, *J. Hydrol.*, *446*, 103–114.
- Mela, K., and J. N. Louie (2001), Correlation length and fractal dimension interpretation from seismic data using variograms and power spectra, *Geophysics*, *66*(5), 1372–1378.
- Reichle, R. H., R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe (2007), Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR), *J. Geophys. Res.*, *112*, D09108, doi:10.1029/2006JD008033.
- Reichle, R. H., W. T. Crow, and C. L. Keppenne (2008), An adaptive ensemble Kalman filter for soil moisture data assimilation, *Water Resour. Res.*, *44*, W03423, doi:10.1029/2007WR006357.
- Rienecker, M. M., et al. (2011), MERRA—NASA's modern-era retrospective analysis for research and applications, *J. Clim.*, *24*, 3624–3648.
- Rodell, M., et al. (2004), The global land data assimilation system, *Bull. Am. Meteorol. Soc.*, *83*, 381–394.
- Rodell, M., I. Velicogna, and J. S. Famiglietti (2009), Satellite-based estimates of groundwater depletion in India, *Nature*, *460*, 999–1002.
- Rowlands, D. D., S. B. Luthcke, S. M. Klosko, F. G. R. Lemoine, D. S. Chinn, J. J. McCarthy, C. M. Cox, and O. B. Anderson (2005), Resolving mass flux at high spatial and temporal resolution using GRACE inter-satellite measurements, *Geophys. Res. Lett.*, *32*, L04310, doi:10.1029/2004GL021908.
- Steiger, J. H. (1980), Tests for comparing elements of a correlation matrix, *Psychol. Bull.*, *87*, 245–251.
- Stieglitz, M., A. Ducharme, R. D. Koster, and M. Suarez (2001), The impact of detailed snow physics on the simulation of snow cover and subsurface thermodynamics at continental scales, *J. Hydrometeorol.*, *2*, 228–242.
- Strassberg, G., B. R. Scanlon, and D. P. Chambers (2009), Evaluation of groundwater storage monitoring with the GRACE satellite: Case study of the High Plains Aquifer, Central United States, *Water Resour. Res.*, *45*, W05410, doi:10.1029/2008WR006892.
- Sturm, M., J. Holmgren, and G. E. Liston (1995), A seasonal snow cover classification system for local to global applications, *J. Clim.*, *8*, 1261–1283.
- Su, H., Z.-L. Yang, R. E. Dickinson, C. R. Wilson, and G.-Y. Niu (2010), Multisensor snow data assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment terrestrial water storage information, *J. Geophys. Res.*, *115*, D10104, doi:10.1029/2009JD013035.
- Swenson, S., and J. Wahr (2006), Post-processing removal of correlated errors in GRACE data, *Geophys. Res. Lett.*, *33*, L08420, doi:10.1029/2005GL025285.
- Swenson, S., P. J.-F. Yeh, J. Wahr, and J. S. Famiglietti (2006), A comparison of terrestrial water storage variations from GRACE with in situ measurements from Illinois, *Geophys. Res. Lett.*, *33*, L16401, doi:10.1029/2006GL026962.
- Syed, T. H., J. S. Famiglietti, and D. P. Chambers (2009), GRACE-based estimates of terrestrial freshwater discharge from basin to continental scales, *J. Hydrometeorol.*, *10*, 22–40.
- Tang, Q., H. Gao, P. Yeh, T. Oki, F. Su, and D. P. Lettenmaier (2010), Dynamics of terrestrial water storage change from satellite and surface observations and modeling, *J. Hydrometeorol.*, *11*, 156–170.
- Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins (2004), GRACE measurements of mass variability in the Earth system, *Science*, *305*, 503–505.
- Wahr, J., S. Swenson, V. Zlotnicki, and I. Velicogna (2004), Time-variable gravity from GRACE: First results, *Geophys. Res. Lett.*, *31*, L11501, doi:10.1029/2004GL019779.
- Wang, X., C. de Linage, J. S. Famiglietti, and C. S. Zender (2011), Gravity Recovery and Climate Experiment (GRACE) detection of water storage changes in the Three Gorges Reservoir of China and comparison with in situ measurements, *Water Resour. Res.*, *47*, W12502, doi:10.1029/2011WR010534.
- Yeh, P. J.-F., S. C. Swenson, J. S. Famiglietti, and M. Rodell (2006), Remote sensing of groundwater storage changes in Illinois using the Gravity Recovery and Climate Experiment (GRACE), *Water Resour. Res.*, *42*, W12203, doi:10.1029/2006WR005374.
- Zaitchik, B. F., M. Rodell, and R. H. Reichle (2008), Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi river basin, *J. Hydrometeorol.*, *9*, 535–548, doi:10.1175/2007JHM951.1.