Assimilation of Remotely Sensed Soil Moisture in the MESH Model

by

Xiaoyong Xu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Geography

Waterloo, Ontario, Canada, 2015

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Soil moisture information is critically important to weather, climate, and hydrology forecasts since the wetness of the land strongly affects the partitioning of energy and water at the land surface. Spatially distributed soil moisture information, especially at regional, continental, and global scales, is difficult to obtain from ground-based (in situ) measurements, which are typically based upon sparse point sources in practice. Satellite microwave remote sensing can provide large-scale monitoring of surface soil moisture because microwave measurements respond to changes in the surface soil's dielectric properties, which are strongly controlled by soil water content. With recent advances in satellite microwave soil moisture estimation, in particular the launch of the Soil Moisture and Ocean Salinity (SMOS) satellite and the Soil Moisture Active Passive (SMAP) mission, there is an increased demand for exploiting the potential of satellite microwave soil moisture observations to improve the predictive capability of hydrologic and land surface models.

In this work, an Ensemble Kalman Filter (EnKF) scheme is designed for assimilating satellite soil moisture into a land surface-hydrological model, Environment Canada's standalone MESH to improve simulations of soil moisture. After validating the established assimilation scheme through an observing system simulation experiment (synthetic experiment), this study explores for the first time the assimilation of soil moisture retrievals, derived from SMOS, the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) and the Advanced Microwave Scanning Radiometer 2 (AMSR2), in the MESH model over the Great Lakes basin. A priori

rescaling on satellite retrievals (separately for each sensor) is performed by matching their cumulative distribution function (CDF) to the model surface soil moisture's CDF, in order to reduce the satellite-model bias (systematic error) in the assimilation system that is based upon the hypothesis of unbiased errors in model and observation. The satellite retrievals, the open-loop model soil moisture (no assimilation) and the assimilation estimates are, respectively, validated against point-scale in situ soil moisture measurements in terms of the daily-spaced time series correlation coefficient (skill R).

Results show that assimilating either L-band retrievals (SMOS) or X-band retrievals (AMSR-E/AMSR2) can favorably influence the model soil moisture skill for both surface and root zone soil layers except for the cases with a small observation (retrieval) skill and a large open-loop skill. The skill improvement $\Delta R^{A\text{-}M}$, defined as the skill for the assimilation soil moisture product minus the skill for the open-loop estimates, typically increases with the retrieval skill and decreases with increased open-loop skill, showing a strong dependence upon $\Delta R^{S\text{-}M}$, defined as the retrieval skill minus the model (open-loop) surface soil moisture skill. The SMOS assimilation reveals that the cropped areas typically experience large $\Delta R^{A\text{-}M}$, consistent with a high satellite observation skill and a low open-loop skill, while $\Delta R^{A\text{-}M}$ is usually weak or even negative for the forest-dominated grids due to the presence of a low retrieval skill and a high open-loop skill. The assimilation of L-band retrievals (SMOS) typically results in greater $\Delta R^{A\text{-}M}$ than the assimilation of X-band products (AMSR-E/AMSR2), although the sensitivity of the assimilation to the satellite retrieval capability may become progressively weaker as the

open-loop skill increases. The joint assimilation of L-band and X-band retrievals does not necessarily yield the best skill improvement.

As compared to previous studies, the primary contributions of this thesis are as follows. (i) This work examined the potential of latest satellite soil moisture products (SMOS and AMSR2), through data assimilation, to improve soil moisture model estimates. (ii) This work, by taking advantage of the ability of SMOS to estimate surface soil moisture underneath different vegetation types, revealed the vegetation cover modulation of satellite soil moisture assimilation. (iii) The assimilation of L-band retrievals (SMOS) was compared with the assimilation of X-band retrievals (AMSR-E/AMSR2), providing new insight into the dependence of the assimilation upon satellite retrieval capability. (iv) The influence of satellite-model skill difference $\Delta R^{S\text{-}M}$ on skill improvement $\Delta R^{A\text{-}M}$ was consistently demonstrated through assimilating soil moisture retrievals derived from radiometers operating at different microwave frequencies, different vegetation cover types, and different retrieval algorithms.

Acknowledgements

First and foremost, I would like to express my special thanks to my advisors, Dr. Jonathan Li and Dr. Bryan Tolson. Their invaluable guidance is gratefully acknowledged here. It was only through their support and encouragement that I have been able to accomplish my research goals proposed for this study.

I would also like to thank my supervisory committee members, Dr. Jonathan Price, Dr. William Quinton and Dr. James Craig, for their helpful comments as well as their time and support throughout my study over the years. Sincere thanks are given to the external examiner, Dr. Jing M. Chen, for his time and his helpful comments on the original version of the thesis. My appreciation also goes on to Alan Anthony, Lori McConnell, and Susie Castela for their help during my PhD study. They always answered my questions with a quick response.

I am very grateful to the European Space Agency (ESA) and the ESA Earth Observation Missions Helpdesk Team for providing the SMOS soil moisture product, to the National Snow and Ice Data Center (NSIDC) and the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) for providing access to the AMSR-E soil moisture data, to the Japan Aerospace Exploration Agency (JAXA) and the GCOM-W1 Data Providing Service for providing access to the AMSR2 soil moisture product, and to the Michigan Automated Weather Network (MAWN) and the Natural Resources Conservation Service (NRCS) for their in situ soil moisture data used in this study.

Sincere thanks also go to Dr. Ralf Staebler for providing me with in situ data at Borden, to Dr. Bruce Davison and Dr. Frank Seglenieks for providing access to the MESH model code and the forcing data, to Dr. Amin Haghnegahdar for his calibration results.

This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca) and Compute/Calcul Canada.

I am so thankful for the financial support through an NSERC Alexander Graham Bell Canada Graduate Scholarship-Doctoral (CGSD), a Meteorological Service of Canada Graduate Supplement Scholarship, a Research Studentship provided by Dr. Bryan Tolson, the University of Waterloo President's Graduate Scholarship, the Dr. T.E. Unny Memorial Award, and the University of Waterloo Environmental Studies Graduate Experience Award.

Finally, I dedicate this work to my wife, my daughter, and my son for their love, patience and endless support. They are an inspiration to me every day.

Xiaoyong Xu

September 2015

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Background

Changes in the spatial and temporal distribution of water resources are expected to play a major role in driving the impacts of climate and global change on human settlements and infrastructure (Bates et al., 2008). The monitoring and prediction of water resources under climate change typically rely on in situ and remote sensing observations, and reliable numerical modeling systems. In situ observations of hydrological conditions (e.g. precipitation, snow, soil moisture and evapotranspiration) are generally based upon uneven point sources, and have limited and sparse spatial coverage except in developed areas or well-designed field experiments. Satellite remote sensing offers better geographical coverage and holds the capability to provide large-scale spatially distributed measurements. A variety of hydrology-related variables, such as precipitation, snow cover, snow water equivalent, surface soil moisture, land surface temperature, leaf area index, and evapotranspiration, can be estimated using remote sensing (see the review papers by Rango, 1994; Tang et al., 2009; Zheng and Moskal, 2009; Li et al., 2009; Wang and Qu, 2009; Dietz et al., 2012). Overall, however, the temporal and/or spatial coverage of remote sensing measurements is still not sufficient for many practical applications because remote sensing provides only instantaneous values of the object within the sampled area (instantaneous field of view, IFOV) at the observing time. Additionally, satellite remote sensing cannot measure the information (e.g. soil water content) below a thin surface layer.

On the other hand, land surface and hydrological model simulations, in particular for physically-based distributed models, allow for the estimation and prediction of hydrologic conditions at desired spatial and temporal scales. In practice, however, land surface and hydrologic modeling is often difficult because we have neither a perfect forecast model nor perfect forcing data. The accuracy of state estimation suffers from uncertainties in forcing fields and deficiencies in model physics and/or parameters. To improve the model simulations, one may constrain the model forecasted state in time with observations. A simplistic method is a direct insertion, which uses observations to directly replace the corresponding model predictions at measurement times. Nevertheless, observation errors, which always exist but are ignored in the direct insertion method, could degrade the state estimation (Additionally, a direct insertion is applicable only when the model variable is directly connected to the observed variable). To circumvent these problems, observations should be integrated into the model dynamical framework by taking into account both the model forecast and observation errors. The effects on the state estimation (analysis) of the model and observations will be controlled by their respective error statistics. This allows a state estimation superior to either the model forecast or the observation alone to be produced. Meanwhile, the observed information can, by means of consistency constraints based upon the time evolution and physical properties of the system, spread to times and locations that are not directly observed. This is the basic concept of data assimilation.

Soil moisture information is critical to the monitoring and modeling of climate and global changes (e.g. Zhang and Frederiksen, 2003). As a reservoir for evapotranspiration, soil moisture has an important controlling on the partitioning of energy fluxes between

latent and sensible forms at the land surface. In the presence of extensive soil moisture anomalies, land surface fluxes may modulate the large-scale atmospheric circulation during the summer (e.g. Wolfson et al., 1987; Fischer et al. 2007). Soil moisture affects precipitation across a range of spatial and temporal scales (e.g. Talyor et al., 2012; Collow et al., 2014). Soil moisture regulates the partitioning of rainfall into runoff (surface discharge) and infiltration on land surfaces, and therefore has significant impacts upon streamflow forecasting in rainfall-runoff models (e.g., Maurer and Lettenmaier, 2003; Berg and Mulroy, 2006). Traditionally, soil moisture can be in situ measured using a gravimetric method or ground-based sensors (probe, time domain reflectometry, ground penetrating radar, etc.). In situ measurements typically serve as the "ground truth", but spatially distributed soil moisture information, especially at regional, continental, or global scales, is difficult to estimate from in situ measurements that are typically based upon sparse point sources in practice. Satellite microwave remote sensing (e.g. Jackson, 1997; Bindlish et al., 2003; Njoku et al., 2003; Owe et al., 2008; Kerr et al., 2012; Entekhabi, et al., 2010a) holds the ability to provide a large-scale monitoring of surface soil moisture because microwave measurements respond to changes in the surface soil's dielectric properties, which are strongly controlled by soil water content. Over the past decade, satellite microwave soil moisture retrievals have shown great potential to improve the predictive skill of land surface and hydrologic models, especially through data assimilation techniques (see section 1.2.3). In a data assimilation system, near-surface soil moisture information derived from satellite microwave measurements can spread to deeper soil layers that cannot be directly measured by satellite microwave sensors. Furthermore, soil moisture observations from

different satellite platforms can, through data assimilation, be merged within the same model framework to yield a single optimal soil moisture estimation.

With recent advances in satellite microwave soil moisture estimation, in particular the launches of the Soil Moisture and Ocean Salinity (SMOS) satellite and the Soil Moisture Active Passive (SMAP) mission, there is an increased demand for exploiting the potential of satellite microwave soil moisture observations to improve the predictive capability of hydrologic and land surface models. This doctoral study aims to assimilate satellite microwave soil moisture observations into a distributed land surface-hydrological model and to demonstrate the contribution of the assimilation to the model soil moisture estimates for both surface and root zone soil layers. The improved soil moisture estimates resulting from the assimilation would benefit weather and climate forecast initializations. In the long term, outcomes of this study can improve the monitoring and prediction of water recourses under climate change, thus providing better guidance for water resource related applications and management.

1.2 Literature review

1.2.1 Data assimilation methods

In essence, data assimilation aims to estimate a posterior probability density function (PDF) of the model state given observations from one or more sources (Bayes' theorem). A great number of data assimilation methods have been developed for land and hydrologic

applications (Table A1). In most cases, we assume that the model and observation errors are Gaussian, and thus the analysis problem can be resolved with either a maximum-likelihood estimator (e.g. variational assimilation methods) or a variance minimizing estimator (e.g. Kalman filter, KF; extended Kalman filter, EKF; or ensemble Kalman filter, EnKF). A variational assimilation method (e.g. three-dimensional variational assimilation, 3DVAR; or four-dimensional variational assimilation, 4DVAR) seeks the state with the maximum likelihood by minimizing a cost function that contains, at least, a background term, a measure of the misfit between the model state x (unknown) and the forecast background (a priori sate), and an observation term, a measure of the misfit between x and observations. In contrast, the Kalman Filter (KF) and its variants (EKF, EnKF) directly compute the Kalman gain matrix and derive the analysis state based upon the analysis equation, which is supposed to ensure minimum analysis error variances. In the KF and the EKF, an additional error covariance equation is utilized to propagate the model forecast error information while the EnKF uses a Monte Carlo sampling (an ensemble of model states and an ensemble of perturbed observations) to estimate error statistics and the evolution of the model forecast error information is implemented by integrating the ensemble of model states forward in time. As compared to the variational methods, a variance minimizing estimator is relatively easy to implement since an adjoint version of the forecast model is not required. These assimilation techniques are further detailed as below.

a. Variational data assimilation

A variational method does not directly compute the analysis state. Instead, it seeks an equivalent solution to the analysis problem by minimizing a predefined cost function, given by

$$J(x) = (x - x^b)^T B^{-1}\left(x - x^b\right) + [Y - H(x)]^T R^{-1}[Y - H(x)] \quad (1.1)$$

where x^b represents the background (a priori) model state, Y is the observation, B and R denote the respective error covariances of x^b and Y. H is a linear or linearized observation operator, which relates the model state variable to the observed variable. The superscript T denotes the transpose of the matrix. The first right-hand-side term of (1.1) is called the background term, which is an objective measure of the misfit between the state x (unknown) and the background model state x^b. The second right-hand-side term of (1.1) is the observation term, which quantifies the misfit between x and the observation Y. If multiple observation types (assuming the observation errors are uncorrelated) are to be assimilated, the observation term in (1.1) can be broken down into multiple terms, each observation type having its own observation term. Additionally, to impose additional weak constraints, more terms (e.g. the penalty term) can be added to the right-hand-side of (1.1). In practice, a suitable descent algorithm is needed to iteratively search for an approximate solution to the minimization of $J(x)$. At each iteration step, a new estimation of x is made to produce as possible as great reduction in $J(x)$. The search direction (descent direction) is determined based upon the local slope, i.e. the gradient of the cost function. When the minimum of J(x) is found, the corresponding x is the optimal analysis state. In a realistic

application, only a small number of iterations are performed to ease the computational burden. The cost function in (1.1) can be employed to different spatial dimensions, such as one-dimensional variational assimilation 1DVAR (e.g. assimilation of satellite sounding data) and three-dimensional variational assimilation 3DVAR (e.g. a global assimilation analysis of the 3D meteorological fields).

Further, if the minimization is extended to the time domain, it is called the four-dimensional variational analysis (4DVAR). The corresponding cost function is defined as

$$J(x_{t_0}) = (x_{t_0} - x^b)^T B^{-1}(x_{t_0} - x^b) + \sum_{i=0}^{n}[Y_{t_i} - H(x_{t_i})]^T R_{t_i}{}^{-1}[Y_{t_i} - H(x_{t_i})] \quad (1.2)$$

In (1.2), the observation term of the cost function contains the differences between the state x and observation Y over a time interval (t_0 to t_n), while the background term is defined only at initial time t_0. The basic idea of 4DVAR is to seek an optimal state x at t_0 (i.e. initial condition x_{t_0}) to yield (through the forward integration of the model assuming the model is perfect) the sequence of optimal states x_{t_i} (t_i represent the observation time, i = 0, …, n), which will lead to the minimum of the cost function defined in (1.2). The minimization procedure of the 4DVAR cost function can be summarized as follows. (i) The forecast model is integrated forward (from t_0 to t_n) with a first guess of the initial condition to produce the forecast state at each observation time t_i over the time interval, and the cost function (1.2) is calculated. (ii) The adjoint model of the forecast model (a conjugate transpose of the tangent linear model of the forecast model) is integrated backward to the beginning of the time window (from t_n to t_0), and the gradient of the cost function with respect to the initial condition is computed. (iii) Check whether the minimization

convergence criterion is met or not; if the convergence criterion is not met, the initial condition of the forward model is adjusted based upon the descent direction, which is estimated using the gradient of the cost function as calculated in (ii). (iv) Repeat the steps (i) to (iii) until the convergence criterion is met and the optimal forecast trajectory is determined. To perform 4DVAR, the construction of the adjoint model of the forecast model is required, and could be difficult if the forecast model is highly nonlinear and complex. In practice, some approximations have to be adopted when deriving the adjoint model (e.g. ignoring moisture physical processes in the adjoint atmospheric model) to limit possible numerical instability arising from nonlinear processes. The readers are referred to the relevant references (e.g., Talagrand and Courtier, 1987) for more details on the minimization of the cost function, the adjoint equations, and the descent algorithms

Note that 4DVAR using the cost function defined in (1.2) deals with only the uncertainty in the model initial condition, and ignores deficiencies in the model physics and parameters (i.e. assuming that the model is perfect). This is the so-called strong-constraint, i.e. the sequence of model states over the time interval must completely comply with the forecast equations. To impose external weak constraints to deals with other errors such as the deficiencies in the model physics and noises in the forcing fields, additional terms can be placed to the right-hand-side of (1.2) (e.g. Reichle et al. 2001a). In 4DVAR, all the observations distributed in this interval are assimilated simultaneously, and the observational information is propagated not only from the past into the future but also from the future into the past. That is to say, the state estimation over the assimilation interval

(time window) is influenced by all the observations distributed in this interval. Therefore, 4DVAR is a typical representative of the smoothing (or batch) algorithm.

b. Kalman Filter and its variants

The Kalman Filter (KF) and its various variants (extended Kalman Filter, EKF; ensemble Kalman Filter, EnKF) are typical 'filtering' (or sequential) assimilation techniques. As compared to the 'smoothing' algorithm 4DVAR, the implementation of sequential 'filtering' technologies is relatively easy since an adjoint version of the forecast model is not required, and therefore makes themselves more attractive for land surface and hydrologic data assimilation. In the traditional KF, each assimilation cycle consists of two steps: a forecast step and an analysis step. In the forecast step, the forecast model is integrated forward in time (from an initial or analysis state) with an additional error covariance equation to propagate error information:

$$x_k^f = M(x_{k-1}^a, u_k) \tag{1.3}$$

$$P_k^f = MP_{k-1}^a M^T + Q \tag{1.4}$$

where x and P denote the model state and the associated error covariance matrix, respectively; the subscript k denotes the time index; the superscripts f and a represent forecast and analysis, respectively. u denotes uncertainties in the model (errors in forcing data and/or deficiencies in model parameters/physics). M denotes the model operator and Q stands for the model error.

In the analysis step (time index *k* omitted) the new observation is used to adjust the current forecast estimation. The analysis equation of the KF is given by,

$$x^a = x^f + P^f H^T \left(HP^f H^T + R\right)^{-1}[Y - H(x^f)] \tag{1.5}$$

Meanwhile, the error covariance matrix can be updated using

$$P^a = P^f - P^f H^T \left(HP^f H^T + R\right)^{-1} HP^f \tag{1.6}$$

Starting from the updated state x^a and error estimation P^a, equations (1.3) and (1.4) are then integrated forward to produce the forecasted X^f and P^f for the next observation. As such, the observational information is accumulated into the model state in a sequential manner.

The KF is valid only for linear systems. Its variants have been developed to solve the optimal estimation problem for nonlinear systems. The EKF still uses equations (1.3)-(1.6) but with *M* in equation (1.3) being a nonlinear operator and *M* in (1.4) being its linearized version. Equation (1.4) indicates that the linear KF and its nonlinear variant, the Extended Kalman filter (EKF), explicitly compute and propagate the error statistics. In practice, the full error covariances are difficult or impossible to be directly estimated due to an expensive computational cost and insufficient error information, especially for large-scale applications. Additionally, the use of a linearized and approximate error covariance equation may cause the EKF to fail to track the state space in a strongly nonlinear system since higher-order components are ignored. To this end, Evensen (1994) proposed the Ensemble Kalman filter (EnKF) where the Fokker-Planck equation defining the time evolution of the model state's probability density was solved using a Monte Carlo method. The probability density function (PDF) of the model state is represented using an ensemble where the mean is the best estimate (Gaussian assumption) and the ensemble spread defines

the error variance. The measurement errors are represented using another ensemble with the mean equal to zero. The evolution of the forecast error statistics is implicit in ensemble forecasts. In contrast to the EKF, the error evolution is fully nonlinear in the EnKF but with lower rank (finite ensemble size).

Similar to the KF and EKF, the EnKF sequentially conducts a forecast step and an analysis step. In the forecast step, the ensemble of model states, generated by a Monte Carlo method, are integrated forward in time, expressed as

$$x_{j,k}^{f} = M\left(x_{j,k-1}^{a}, u_{j,k}\right) \quad j = 1, \dots, N \tag{1.7}$$

where M denotes the model operator, x denotes the model state, u_j denotes uncertainties in the model (perturbations to the forcing data or deficiencies in model parameters/physics), the superscripts f and a represent forecast and analyzed state, respectively; the subscript k denotes the time index, and j is the ensemble member index, counting from 1 to the number of model state ensemble N.

In the analysis step (time index k omitted), the Kalman gain K is estimated from the forecast and measurement error covariances and each forecast ensemble member is then updated according to the Kalman analysis equation.

$$K = \sigma\left(X^{f}, \hat{Y}\right)[\sigma\left(\hat{Y}, \hat{Y}\right) + \sigma(\varepsilon, \varepsilon)]^{-1} \tag{1.8}$$

$$x_{j}^{a} = x_{j}^{f} + K\left[y_j - \hat{y}_j\right] \quad j = 1, \dots, N \tag{1.9}$$

where σ denotes the covariance between two vectors, x_j^f and x_j^a denote the forecast and analysis model state of the jth ensemble member. y_j and $\hat{y}_j$ represent the perturbed

observation and the corresponding model prediction. X^f, $\hat{Y}$ and ε represents the ensembles of $\{x_j^f\}$, $\{\hat{y}_j\}$, and observation errors, respectively. The best estimation is represented by the analysis ensemble mean. With an infinite ensemble size the EnKF will yield exactly the same analysis as the EKF.

1.2.2 Microwave remote sensing for surface soil moisture estimation

Soil moisture is an important variable for numerical weather, climate, and hydrologic forecasts. This is because soil moisture plays a crucial role in hydrological cycle by controlling the partitioning of water and energy fluxes at the land surface and the moisture exchanges at the soil-vegetation-atmosphere interface. Surface soil moisture can be estimated using various remote sensing instruments including microwave, optical, and thermal infrared sensors (see a review by Wang and Qu, 2009). Microwave techniques are of particular value for surface soil moisture estimation because microwave measurements are sensitive to changes in the soil dielectric properties, which are strongly controlled by soil water content. Liquid water has a very high dielectric constant (~80-90 at 0-20ºC) while the dielectric constant is very low (only ~ 4) for dry soil. Such a high contrast between the dielectric constants of wet and dry soils forms the basis for deriving soil moisture information from microwave remote sensing measurements. Over the past decades, both active and passive microwave technologies have been developed for surface soil moisture estimation (Jackson, 2005). Passive microwave sensors (radiometers) measure the natural thermal radiation emitted from the soil (brightness temperature), while active microwave sensors (radars) emit energy to the land surface and observe the radiation backscattered by

the soil (backscattering coefficient). The mixed dielectric constant of the soil's constituents can be retrieved based upon the acquired brightness temperatures or backscattering coefficients. Ultimately, soil water content can be estimated by means of a soil dielectric mixing model.

Tables A2 and A3 summarize the representative microwave sensors, respectively, for satellite active and passive soil moisture measurements. For spaceborne active measurements, the ESA Remote Sensing Satellite (ERS) Synthetic Aperture Radar (SAR) and Scatterometer (SCAT), the Canadian RADARSAT series (e.g. Merzouki et al., 2011), and the Advanced Scatterometer (ASCAT) onboard the Meteorological Operational (Metop) satellite (e.g., Bartalis et al., 2007; Albergel et al., 2009), successor of the SCAT, have played important roles within the past decades; while passive microwave observations typically relied upon the Special Sensor Microwave Imager (SSM/I) (e.g. Jackson, 1997), the Scanning Multichannel Microwave Radiometer (SMMR) (e.g. Reichle and Koster, 2005; Owe et al., 2008), the Tropical Rainfall Measuring Mission Microwave Imager (TMI) (e.g. Bindlish et al., 2003), the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) (e.g. Njoku et al., 2003; Njoku and Chan, 2006), the Advanced Microwave Scanning Radiometer 2 (AMSR2), or the Microwave Imaging Radiometer with Aperture Synthesis (MIRAS) onboard the Soil Moisture and Ocean Salinity (SMOS) satellite (Kerr et al., 2001; 2010; 2012). In particular, the SMOS mission and the newly launched (January 2015) Soil Moisture Active Passive (SMAP) mission (Entekhabi et al., 2010a) were designed exclusively for soil moisture monitoring. The L-band (1.3 or 1.4 GHz) sensors carried by the two satellites have stronger penetration of soil

and vegetation than those operating at higher frequencies (e.g. X or C-band), and thus greatly enhance our capability to map large-scale surface soil moisture.

There is usually an inverse relationship between a sensor's temporal frequency and spatial resolution. The active SAR technology is able to scan the land at a high spatial resolution, but the revisit time is very long. Passive microwave sensors onboard polar-orbiting satellites offer a higher time resolution (revisit per 1-3 days) due to their wide swaths, but generally result in relatively coarse spatial samplings. Overall, soil moisture retrieval is challenging for active microwave sensors because radar signal is highly sensitive to local features of the soil surface (surface roughness, topography, vegetation, etc.), while passive microwave soil moisture products are usually more reliable due to higher signal-to-noise ratio and mature retrieval algorithms. Microwave sensors measure only the soil moisture within a near-surface layer. The soil thickness measured increases with the wavelength (approximately several tenths of the wavelength). For bare soil, the penetration depth is about 3-5 cm for L-band (1-2 GHz) sensors (e.g. SMOS), and only ~1-1.5 cm for C (4-8 GHz) or X (8-12 GHz) band measurements (e.g. AMSR-E). Soil moisture estimation using microwave sensors is subject to vegetation effects. Where there is a vegetation cover, the radiation emitted or backscattered from the soil will be attenuated owing to the scattering and absorption by vegetation canopy. The magnitude of the vegetation attenuation increases with the sensor frequency and the vegetation density. Hence soil moisture retrieval at high microwave frequencies (> 5-6 GHz) is valid only for bare soil or sparely vegetated regions. Vegetation cover impacts upon sensors operating at low frequencies are less pronounced because the latter can penetrate moderately dense

canopies. For example, L-band sensors (e.g. SMOS) can provide reliable measurements over a wide range of vegetation cover (biomass $\leq$ 5 kg/m^2).

1.2.3 Assimilation of microwave remotely-sensed soil moisture in land surface and hydrological models

Thanks to the development of data assimilation technologies and the launches of various passive and active microwave sensor systems, there has been an intensive global research effort to assimilate microwave remote sensing soil moisture information into land surface or hydrological models within the past few decades (Table A4). The studies generally can be divided into three major categories as follows.

(i) *A direct assimilation of microwave brightness temperatures in land surface models (LSM)*. A series of synthetic assimilation experiments based upon the 1997 Southern Great Plains hydrology experiment demonstrated that a direct assimilation of microwave brightness temperature data in LSMs could provide reliable soil moisture estimates (e.g. Reichle et al., 2001a, 2001b, 2002b). In practical application, Margulis et al. (2002) used the EnKF method to assimilate airborne Electronically Steered Thinned Array Radiometer (ESTAR) 1.4 GHz surface brightness temperature measurements during SGP97 into Noah LSM. Crow and Wood (2003) conducted similar assimilation experiments with the TOPMODEL-Based Land Surface-Atmosphere Transfer Scheme (TOPLATS) model. Their results showed that the assimilation of ESTAR brightness temperature measurements led to good soil moisture estimates not only in the surface layer but also for root zone.

Mattia et al. (2009) demonstrated that high spatial resolution surface soil moisture could be estimated through an integration of SAR data and hydrologic modeling with a constrained minimization technique. In their SAR retrieval algorithm, the hydrological model provided the background information about soil moisture at coarse spatial resolution based on the Antecedent Precipitation Index (API) approach.

(ii) *Assimilation of microwave soil moisture retrievals in LSM.* As one of the pioneer studies, for instance, Houser et al. (1998) incorporated soil moisture derived from the NASA L-band (1-2 GHz) push broom microwave radiometer (PBMR) mounted on a NASA C-130 aircraft into the hydrologic model TOPLATS with several assimilation schemes. Results showed that all the assimilation schemes could produce substantial improvements in the model surface soil moisture simulations. Reichle and Koster (2005) assimilated the surface soil moisture retrievals from SMMR into the NASA Catchment Land Surface Model (CLSM) with the EnKF method. Reichle et al. (2007) assimilated both SMMR and AMSR-E soil moisture retrievals into CLSM. The assimilation led to overall improvements, relative to either the model estimates or satellite retrievals alone, in terms of soil moisture anomaly time series correlation with in situ measurements. In Draper et al. (2009a), the Extended Kalman Filter (EKF) method was used to assimilate the surface soil moisture derived from AMSR-E C-band brightness temperature measurements with the Land Parameter Retrieval Model (LPRM) into the Interactions between Surface, Biosphere, and Atmosphere (ISBA) land model. The introduction of AMSR-E soil moisture did yield substantial analysis increments (changes in the model estimate between before and after the implementation of the analysis equation) for both surface and root-

zone soil moisture, although the assimilation estimates were not validated against real in situ observations. Liu et al. (2011) showed that the assimilation of AMSR-E soil moisture was as efficient as the precipitation corrections for enhancing the model skill for soil moisture estimation (anomaly time series correlation coefficient with in situ measurements). The study assessed the contributions of two AMSR-E soil moisture products (June 2002 to July 2009), the NASA standard algorithm product archived at the National Snow and Ice Data Center (NSIDC) and the LPRM-derived AMSR-E soil moisture. The assimilation of LPRM product led to larger soil moisture skill improvement than the NSIDC product. Draper et al. (2012) suggested that the CLSM model soil moisture skill could be improved through the assimilation of either AMSR-E or ASCAT soil moisture retrievals. A joint assimilation of the two sensor products could produce the best soil moisture skill. Due to the bias (systematic error) between satellite retrievals and the model estimates, a priori rescaling on satellite retrievals (the cumulative distribution function (CDF) matching) was applied during the aforementioned efforts. Li et al. (2012) assimilated AMSR-E soil moisture retrievals (derived from the X-band brightness temperatures using single-channel algorithm), without a priori scaling, into the Noah land surface model. Their work was motivated by the assumption that the mean value of satellite retrievals have potential contribution to improving the model mean values of soil moisture. Although the study observed the improved soil moisture estimates (as indicated by reduced bias and root-mean-square-error against in situ measurements), especially for the mass conservation scheme, the analysis typically made systematic corrections to the model soil moisture estimation (a clear symptom of bias in the assimilation). This means that a

satellite-model bias removal is an indispensable part in a bias-blind assimilation system (i.e. correcting random errors only).

(iii) *Assimilation of microwave soil moisture observations in runoff models*. During a flooding event, the affected areas are usually characterized by wet pre-storm soil moisture conditions. This provides an opportunity to improve streamflow forecasts by identifying antecedent soil moisture conditions with microwave remote sensing observations. Pauwels et al. (2002) improved the simulated hydrographs in TOPLATS through the assimilation of ERS SAR soil moisture estimates using a statistical correction approach. Francois et al. (2003) applied the EKF method to assimilating soil moisture retrieved from ERS-1 SAR into a lumped rainfall-runoff model. The study demonstrated that the sequential assimilation of SAR data had the potential to improve the runoff simulations. Jacobs et al. (2003) introduced the surface soil moisture observed by ESTAR during the SGP97 hydrology experiment into a lumped rainfall-runoff model. The ESTAR soil moisture retrievals were used to represent antecedent soil moisture conditions and to update the curve numbers and the runoff predictions based upon a strong correlation between the curve number and soil moisture in the Soil Conservation Service (SCS) curve number method. Results showed an enhancement in runoff forecasts for the watersheds at different spatial scales. Regarding the earlier studies of assimilating ERS SAR soil moisture retrievals into hydrological models, we refer the readers to a review paper by Loumagne et al. (2001). Crow et al. (2005) suggested that the predictive capability of land surface's response to precipitation was enhanced when TMI-derived surface soil moisture was sequentially assimilated into an API model. Brocca et al. (2010) explored the impact on

flood forecasting of assimilating ASCAT-based soil wetness index in a rainfall-runoff model. Results revealed that the assimilation of the ASCAT soil moisture estimates via a simple nudging scheme led to an enhancement in runoff predictions. Crow and Ryu (2009) proposed a new assimilation scheme in which remotely-sensed surface soil moisture measurements were employed to simultaneously adjust both antecedent soil moisture and rainfall accumulations during hydrological modeling. Their work was motivated by the additional capability of soil moisture data to filter errors contained in satellite-based rainfall products (Crow and Zhan, 2007; Crow et al., 2009). Preliminary results indicated that the new approach outperformed those schemes which considering only the calibration of antecedent wetness conditions.

1.2.4 Research gaps and recommendations

The encouraging results warrant further research efforts in this field. In particular, with the launches of SMOS and SMAP, our capability to map large-scale soil moisture by microwave remote sensing is progressively enhanced, which will surely trigger more research efforts in satellite soil moisture assimilation over the next decade. The upcoming research efforts should cover the following aspects: (i) assimilation of new L-band soil moisture products (mainly from SMOS and SMAP). Until recently satellite soil moisture measurements at X (8-12 GHz) and C (4-8 GHz) bands (e.g. SMMR, AMSR-E, TMI, ASCAT, or SAR) have been dominant in land/hydrologic data assimilation applications. The launches of SMOS and SMAP have opened up new opportunities for L-band soil moisture assimilation. Assimilation of SMOS and SMAP soil moisture is more attractive

because the ability of SMOS and SMAP to measure surface soil moisture for a wide range of vegetation covers is clearly of advantage for assessing the vegetation modulation of the assimilation. (ii) Comparison between L-band and X-band (or C-band) soil moisture measurements for land data assimilation. In land/hydrologic data assimilation applications, to date, there has been a lack of comparative studies of multi-sensor soil moisture products, especially from different microwave bands, which represent different retrieval capabilities. At the moment, satellite soil moisture estimates for the same location and within the same time period can be available for both L-band sensors (e.g. SMOS, SMAP) and X-band (or C-band) measurements (e.g. AMSR-E/AMSR2, ASCAT). Their comparison for land data assimilation and even their joint assimilation will offer further insight into the dependence of the assimilation upon the satellite observation skill. (iii) Application of nonlinear/non-Gaussian assimilation approaches. Current methods (e.g. KF, EKF, EnKF, and 4DVAR) for assimilating satellite soil moisture are typically based upon the assumption of Gaussian error statistics. For a nonlinear system with the error statistics far from Gaussian, to take into account the influence of skewed PDF, general nonlinear/non-Gaussian filters (e.g. the Particle Filter) should be applied.

1.3 Research objectives

The overall goal of this research is to exploit the potential of satellite microwave soil moisture observations to improve the predictive capability of hydrologic and land surface models. The limited research objectives are set as follows: 1) to develop nonlinear filter algorithms for assimilation of satellite soil moisture in a distributed land surface-

hydrological model; 2) to explore the impact of satellite microwave soil moisture, through a data assimilation scheme, upon the model soil moisture estimates; and 3) to advance our understanding of the satellite observation skill and vegetation cover impacts on the assimilation results. In the long term, this study aims to improve the monitoring and prediction of water recourses under climate change, thus providing better guidance for water resource related applications and management.

As part of an operational forecasting system developed at Environment Canada, the modelling system called Modélisation Environmentale-Surface et Hydrologie (MESH) provides a flexible framework for coupling land surface schemes, a hydrological model, and an atmospheric model (Pietroniro et al., 2007). In this thesis, the ensemble Kalman filter (EnKF) is utilized to assimilate SMOS, AMSR-E and AMSR2 soil moisture retrievals into the standalone version of MESH in which CLASS, the Canadian Land Surface Scheme, is coupled with a distributed hydrological model over the Great Lakes basin. This is the first study to assimilate these satellite retrievals in the MESH model. This work will address the following questions. Fundamentally, how will satellite retrievals of surface soil moisture, through data assimilation, impact the MESH model soil moisture estimates? Next, how does the vegetation cover modulate the assimilation performance? Finally, how important is the satellite observation skill to the assimilation estimates? To this end, the main contents of this thesis include: 1) to equip the MESH model with the EnKF scheme for assimilation of satellite soil moisture retrievals; 2) to validate the performance of the established assimilation system through observing system simulation experiments (synthetic experiments) and assimilation of real satellite retrievals; 3) to reveal the

vegetation modulation of assimilation by taking advantage of the ability of L-band senor (SMOS) to measure surface soil moisture for a wide range of vegetation covers; 4) to compare L-band retrievals (SMOS) and X-band retrievals (AMSRE/AMSR2) in the same assimilation system.

In this thesis, the EnKF method is used because (i) as compared to the variational methods, the EnKF is relatively easy to implement since an adjoint version (a conjugate transpose of the tangent linear model) of the forecast model is not required; (ii) estimation of full error covariances (as used in the EKF) is not required in the EnKF, and the model and measurement error variances are defined by the ensemble spreading (although the input error parameters may need to be specified based upon empirical statistics or to be adaptively tuned); and (iii) the EnKF is an ensemble-based method and thus can be easily merged into the existing ensemble forecasting system in use with Environment Canada.

The focus of this work is upon improvement in soil moisture estimates rather than in hydrologic forecasts (e.g. in the range of a few days). The soil moisture estimates (satellite, the model, and the assimilation) are quantitatively evaluated using the correlation *R* metric of skill, which is defined as the daily time series correlation with point-scale in situ soil moisture measurements. Note that satellite soil moisture retrievals are instantaneous observations (typically once every 1-3 days for this study). The satellite soil moisture skill *R* is computed only over the days with available satellite data, whereas the model and assimilation skill *R* values are obtained based upon the complete model simulation time series (except for snow covered or frozen soil periods) rather than using the soil moisture

estimates only at satellite observation (i.e. assimilation) times. The absolute error measure (e.g. the root-mean-square error (RMSE) metric) is not applicable here because (1) point soil moisture measurement typically contain substantial sampling errors in terms of absolute magnitude (Crow et al., 2012), while the temporal variability of soil moisture observed by point measurements used in this study is often spatially representative; and (2) the absolute magnitude of the soil moisture assimilation product is meaningless since the satellite retrievals are rescaled prior to the assimilation (Reichle et al., 2007). However, notice that through a percentile-based transformation (e.g., Entekhabi et al. 2010b) the time variations of soil moisture can be scaled to the soil moisture initial conditions of weather and climate models, while any bias (systematic error) in the soil moisture product can be scaled out (e.g. Zhang and Frederiksen, 2003). Therefore, the resulting soil moisture assimilation product can benefit weather and climate forecast initializations as long as the time variability of soil moisture is better captured.

The thesis is organized as follows. Chapter 2 describes the establishment of the EnKF assimilation scheme for the MESH model and examines the performance of the assimilation system under different conditions through the Observing System Simulation Experiment (synthetic experiments). Chapter 3 is focused upon the assimilation of SMOS soil moisture retrievals (2010-2013). Chapter 4 presents the assimilation of two AMSR-E soil moisture products (2003-2011) and their comparison with the assimilation of SMOS. Chapter 5 compares the assimilation of AMSR2 and SMOS retrievals for year 2013. The main conclusions and contributions of this thesis are presented in Chapter 6.

1.4 Statement of manuscript submission and author contributions

This doctoral thesis is comprised of the following published and submitted papers in refereed journals.

[1] Chapter 2: **Xu, X.**, B. A. Tolson, and J. Li (2015). Assimilation of satellite soil moisture data in the MESH model with the Ensemble Kalman filter: Synthetic experiments. *Journal of Water Resources Planning and Management,* under review (Manuscript number WRENG-2247).

[2] Chapter 3: **Xu, X.**, B. A. Tolson, J. Li, R. M. Staebler, F. Seglenieks, A. Haghnegahdar, and B. Davison (2015). Assimilation of SMOS soil moisture over the Great Lakes basin. *Remote Sensing of Environment*, 169: 163-175, doi:10.1016/j.rse.2015.08.017.

[3] Chapters 4 & 5: **Xu, X.**, B. A. Tolson, J. Li, B. Davison, and A. Haghnegahdar (2015). Comparison of X-band and L-band soil moisture retrievals for land data assimilation. *Remote Sensing of Environment*, under review (Manuscript number RSE-D-15-00783).

For all manuscripts, X. Xu conceived the study, performed the analysis and wrote the paper. B.A. Tolson and J. Li provided guidance and edited the manuscripts. R. M. Staebler provided in situ soil moisture data for paper [2]. B. Davison and F. Seglenieks provided access to the MESH model code and the forcing data. A. Haghnegahdar provided his MESH calibration results.

CHAPTER 2

Synthetic Assimilation Experiments

2.1 Introduction

Soil moisture plays a crucial role in the Earth's energy and water cycles. Specifically, (i) soil moisture is a key state variable linking the land surface and the atmosphere. Soil moisture can bring a significant component of memory into the soil-atmosphere system through an integration of precipitation and evaporation processes over time scales of days to weeks. The memory has an important influence on the surface saturation and the forecasting of flooding events; (ii) Soil moisture, as a reservoir for evapotranspiration, controls the partitioning of the incoming radiative energy between latent and sensible heat fluxes at the land surface. The soil water content greatly impacts the evaporative proportion from the land surface and the transpiration from the root zone, which in turn control the ground sensible heat fluxes; (iii) Soil moisture regulates the partitioning of rainfall into runoff (surface discharge) and infiltration at the land surface. In particular, accurate specification of antecedent (pre-storm) soil moisture conditions is crucial to flooding forecasts; (iv) Accurate initialization of soil moisture is critical to numerical weather and climate predictions. For example, in the presence of extensive soil moisture anomalies, land surface fluxes can modulate the large-scale atmospheric circulations during the summer, thus influencing short- and long-term precipitation and temperature patterns.

In situ measurements, which are typically based upon sparse point sources, cannot provide large-scale spatially distributed soil moisture information. Satellite microwave remote sensing holds the ability to map global surface soil moisture. This is attributed to a strong contrast between the dielectric constants of water (~80-90 at 0-20ºC) and dry soils (~ 4), which causes the sensitivity of microwave measurements to changes in the wetness of the soil. Both passive and active microwave remote sensing technologies have been developed for monitoring surface soil moisture (Jackson, 2005). Passive microwave sensors (radiometers) measure the natural thermal radiation emitted from the soil (brightness temperature), while active microwave sensors (radars) emit energy to the surface and observe the radiation backscattered by the soil (backscattering coefficient). The mixed dielectric constant of the soil's constituents can be retrieved based upon the acquired brightness temperatures or backscattering coefficients. Ultimately, soil water content can be estimated by means of a soil dielectric mixing model. Satellite microwave soil moisture measurements have been demonstrated to hold great potential to improve land and hydrologic modeling, especially through data assimilation (see a review by Xu et al., 2014). In land surface or hydrologic modeling, the state estimation is often subject to noises in forcing fields and deficiencies in model physical structure and/or parameters. Assimilation of satellite measurements can suppress stochastic noise in the model state estimation. Meanwhile, through data assimilation, satellite surface soil moisture information can propagate to the whole soil column.

With recent advances in satellite microwave soil moisture estimation, there is an increased demand for assimilating satellite soil moisture into land surface/hydrological

models, which will greatly exploit the potential of current satellite microwave observations in water resources related applications. Here, our focus is upon Environment Canada's MESH model (Pietroniro et al., 2007). This chapter describes a data assimilation scheme for assimilating satellite soil moisture in the standalone version of MESH in which CLASS, the Canadian Land Surface Scheme, is coupled with a distributed hydrological model. Prior to real applications, we conduct synthetic experiments, which are presented in this chapter. In a synthetic assimilation, satellite soil moisture data are simulated with the forecast model and the "true" states are known. This allows us to examine the performance of the established assimilation system in an idealized environment.

2.2 Forecast model

The forecast model used here is Environment Canada's standalone MESH, which is a coupled land-surface and hydrological model (Pietroniro et al., 2007). The model originates from the University of Waterloo's WATCLASS (Soulis et al., 2000). A primary innovation of MESH is that the model uses a Grouped Response Unit (GRU) approach (Kouwen et al., 1993) to resolve the heterogeneity in geophysical fields. A GRU is a grouping of subareas with similar soil and/or vegetation attributes (Fig. 2.1a). In the version of MESH used in this work, the identification of GRUs is based solely on the land cover types, i.e., each GRU corresponds to one land cover class (other soil characteristics are assumed to be same for the same GRU). Each model grid cell is represented by a limited number of distinct GRUs (tiles) weighted by their respective cell fractions. The land model is run on each GRU independently. The overall fluxes and prognostic variables of a grid cell are obtained by taking a weighted average of the results from GRUs. The soil column is

typically partitioned into three layers (0-10, 10-35, and 35-410 cm) to resolve water and temperature dynamics (Fig. 2.1b). At the moment, the land surface scheme considers only the vertical water movement between the soil layers, which is governed by the one dimensional Richard's equation (Soulis et al., 2000):

$$\frac{\partial \theta}{\partial t} = -\frac{\partial K(\theta)}{\partial z} + \frac{\partial}{\partial z}\left(K(\theta)\frac{\partial \psi(\theta)}{\partial z}\right) \tag{2.1}$$

where θ denotes the volumetric water content, t is time, z is the depth of soil, $K(\theta)$ is hydraulic conductivity, and $\psi(\theta)$ is pressure head (soil water suction). The lateral movement of water between grids/tiles is not taken into account. The resulting horizontal flows (overland flow, interflow, and base flow) at grid cells are ultimately be routed into the stream and river network systems.

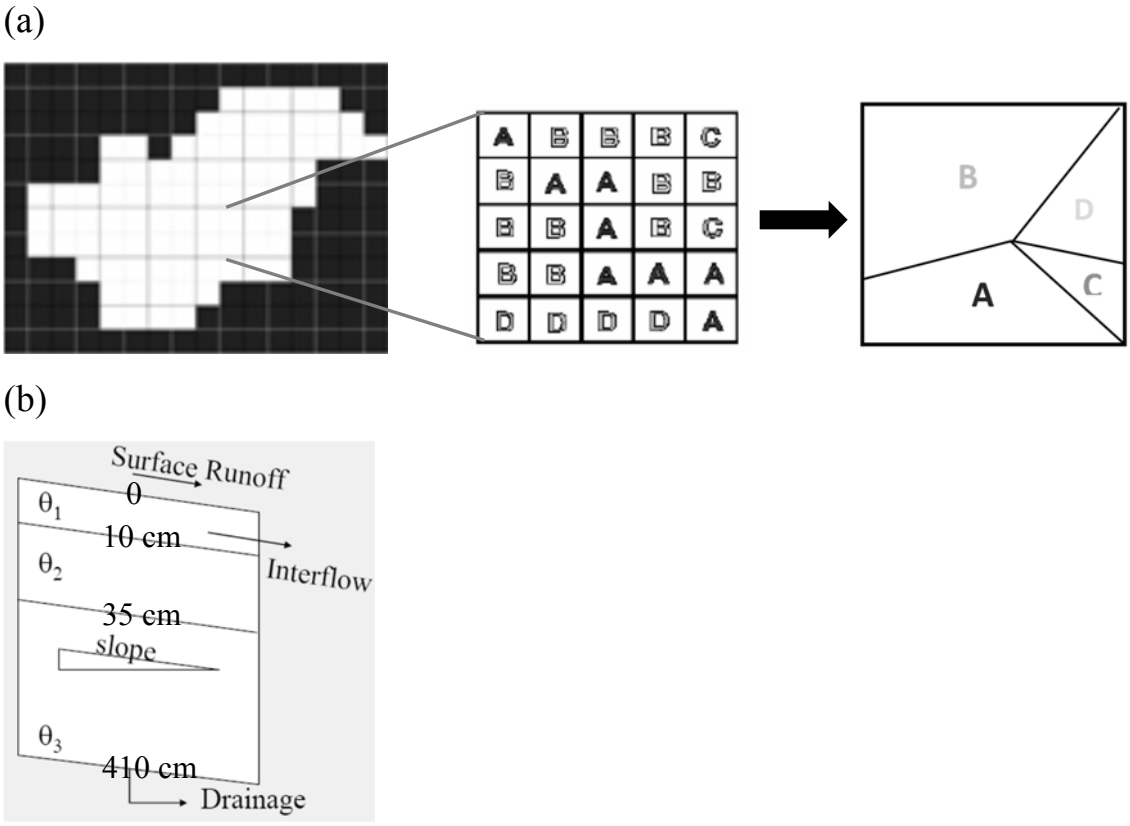

Fig. 2.1 (a) The Grouped Response Unit (GRU) approach to basin discretization and (b) Soil moisture and drainage representation in MESH (adapted from Kouwen et al., 1993; Soulis et al., 2000; 2005).

The experiment domain is the Great Lakes basin (Fig. 2.2). The model configurations are similar to those as used in Pietroniro et al. (2007). Seven GRU types are identified for the whole domain: crop, grass, deciduous forest, coniferous forest, mixed forest, water, and impervious. Each GRU class has a different model parameter set. The basin is gridded at 10 arcmins (~ 15 km × 15 km) in the modelling. Each model grid is a mosaic of the seven GRU classes weighted by their cell fractions.

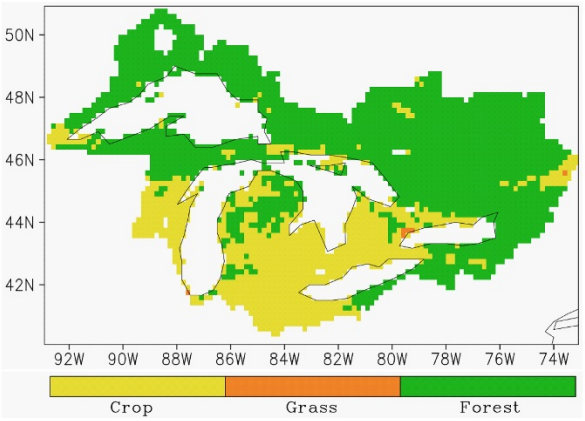

Fig. 2.2 The Great Lakes basin. Note that each grid cell (1/6th of a degree resolution) may consist of a maximum of seven land cover classes (crop, grass, deciduous forest, coniferous forest, mixed forest, water, and impervious). Only the dominant land cover class is displayed for each grid here (the forest cover represents the sum of the deciduous, coniferous, and mixed forest classes). Water and impervious surfaces are not labelled since soil moisture estimates are not considered for them.

2.3 Data assimilation scheme

In a data assimilation system, the observed information is integrated into the model framework by taking into account both the model forecast and observation error characteristics. This allows the model forecast and observation to be optimally merged while without violating the model physical constraints. Many technologies have been developed for land/hydrologic data assimilation (Table A1). In the present study, we use the ensemble Kalman Filter (EnKF) (Evensen, 1994, 2003) to assimilate synthetic satellite soil moisture in the MESH model. The EnKF, which was first introduced by Evensen (1994), uses a Monte Carlo approach to estimate the forecast and observation error statistics. An ensemble of model states is used to approximate the probability density of

the model state. The ensemble spread defines the forecast error variance and the ensemble mean is considered as the best estimate (Gaussian assumption). Thus, the error covariance equation (as used in the Kalman filer or the extended Kalman filter) for the evolution of the model forecast error information can be replaced by integrating the ensemble of model states forward in time, expressed as

$$x_{j,t}^{-} = M(x_{j,t-1}^{+}, u_{j,t}) \tag{2.2}$$

where *M* denotes the forecast model operator, $x_{j,t-1}^{+}$ represents a posterior (analysis) model state at measurement time *t*-1. $x_{j,t}^{-}$ is a priori (forecast) model state at measurement time *t*. $u_{j,t}$ denotes the uncertainties in the model (perturbations to the forcing data or deficiencies in model parameters/physics). The subscript *j* is the ensemble member index, counting from 1 to the size of the ensemble *N*. The observation is perturbed to generate an ensemble of perturbed observations with the ensemble mean equal to the actual observation and the spreading of the ensemble as the observation error variance, i.e.,

$$y_{j,t} = \bar{y}_t + \varepsilon_{j,t} \tag{2.3}$$

$$R_t = \frac{1}{N-1}\sum_{j=1}^{N} \varepsilon_{j,t}\, \varepsilon_{j,t}^{T} \tag{2.4}$$

where $\bar{y}_t$ and $y_{j,t}$ represent the actual observation and the perturbed observation at time *t*, respectively. $\varepsilon_{j,t}$ and R_t denote the observation error perturbation and the observation error covariance, respectively. The superscript *T* denotes the vector transpose. At measurement time *t*, each of the model forecast state ensemble members $x_{j,t}^{-}$ is updated to $x_{j,t}^{+}$ according to the Kalman analysis equation, given by

$$x_{j,t}^{+} = x_{j,t}^{-} + P_t^{-} H_t^{T} (H_t P_t^{-} H_t^{T} + R_t)^{-1} (y_{j,t} - H_t x_{j,t}^{-}) \tag{2.5}$$

where H_t is the measurement operator. P_t^- denotes the forecast error covariance. In the EnKF, P_t^- is only implicitly needed through

$$P_t^- H_t^T = \frac{1}{N-1}\sum_{j=1}^{N}(x_{j,t}^- - \frac{1}{N}\sum_{j=1}^{N}x_{j,t}^-)(H_t x_{j,t}^- - \frac{1}{N}\sum_{j=1}^{N}H_t x_{j,t}^-)^T \quad (2.6)$$

$$H_t P_t^- H_t^T = \frac{1}{N-1}\sum_{j=1}^{N}(H_t x_{j,t}^- - \frac{1}{N}\sum_{j=1}^{N}H_t x_{j,t}^-)(H_t x_{j,t}^- - \frac{1}{N}\sum_{j=1}^{N}H_t x_{j,t}^-)^T \quad (2.7)$$

Ultimately, through conducting in turn the forecast step (equation 2.2) and the update step (equation 2.5), the observational information is sequentially accumulated into the model state. Throughout this thesis, only the one-dimensional version of the Ensemble Kalman filter (1D-EnKF) is used where the horizontal correlations between the model grids are neglected and an observation influences only the model state at the observation location. The model state vector *x*, which has a dimension of 21 and is independent for each grid, is comprised of the volumetric liquid water content from the seven GRUs for the three soil layers. The observation y_j is the perturbed satellite retrievals of surface soil moisture, and the corresponding model prediction Hx_j^- denotes the volumetric liquid water content (a weighted sum of GRUs within the grid) in the model surface layer.

2.4 Experiment setup and results

The synthetic experiment is designed as follows. First, the standalone MESH is integrated for one year period (1 January to 31 December) using the meteorological forcings derived from the Global Environmental Multiscale (GEM) model (Mailhot et al., 2006) forecasts of year 2005. Each GRU class has its own model parameter set, which was based upon a

global calibration with streamflow observations (Haghnegahdar et al., 2014). The model integration is spun up by a repeated integration with the GEM forcings of year 2004. The simulated soil moisture serves as the reference solution ("truth"). The synthetic satellite soil moisture retrievals are generated, independently for each model grid, by applying random noise (the error standard deviation is set to 0.08 m^3/m^3) to the true surface soil moisture sequence. Next, an open loop model simulation (without data assimilation), which intentionally deviates from the true integration, is performed. To this end, the MESH model is integrated for a one-year period again but with the GEM forcings of year 2006 and using a different set of model parameters, which are generated by adding random noise with a standard deviation of 30% (of magnitude) to the model parameters used in the true integration. Finally, the synthetic soil moisture retrievals are assimilated into the open-loop integration, under different conditions, to examine the capability of the assimilation system to recover the true soil moisture. The assimilation experiments are listed in Table 2.1. More details on the experiments will be provided in the remainder of this section.

Table 2.1 List of synthetic experiments

Key	Description
A	Control experiment; Ensemble size N = 12; Assimilation interval is 24 hours
B1	Same as A, but with N = 6
B2	Same as A, but with N = 50
B3	Same as A, but with N = 100
C1	Same as A, but with assimilation interval of 12 hours
C2	Same as A, but with assimilation interval of 72 hours
D1	Same as A, except for a higher retrieval skill
D2	Same as A, except for a lower retrieval skill
E1-E25	Test the impact of the specified model and observation input error parameters

A. Control experiment

Experiment A is a control case. We use the 1D-EnKF with 12 ensemble members. To represent random errors in the forcing inputs, the cross-correlated forcing perturbation fields are generated (Table 2.2) following Reichle et al. (2007). Additionally, to account for the model uncertainties due to imperfect model parameters, temporally correlated error perturbations are added to the simulated volumetric liquid water content. Currently, the 0.001 m^3/m^3, 0.0005 m^3/m^3, and 0.00005 m^3/m^3 error standard deviations are applied to the three soil layers, respectively (Table 2.2). The model error parameters are specified largely based upon order-of-magnitude considerations and are not on-line (adaptively) tuned in our assimilation. Smaller error parameters are applied to deeper soil layers to avoid the bias between the ensemble mean (without assimilation) and the unperturbed open-loop integration. The model error correlation time is set to 1 day. A true observation error standard deviation of 0.08 m^3/m^3 is used for the synthetic retrievals (Recall that the synthetic retrievals are generated by adding white noise with standard deviation of 0.08 m^3/m^3 to the true surface soil moisture). The assimilation is performed at 24-hr intervals (i.e., we assume the observing frequency of once daily). The satellite retrieval-model discrepancies in climatological mean and scale are not present and thus not considered in our twin experiment.

Table 2.2 Error parameters for the selected forcing inputs and model variables

	Perturbation Method	Standard Deviation	Temporal Correlation	Cross Correlation
Forcing inputs[*]				
Precipitation (P)	Multiplicative	0.5	1 day	-0.8 with SW; 0.5 with LW
Incoming shortwave radiation (SW)	Multiplicative	0.2	1 day	-0.8 with P; -0.5 with LW
Incoming longwave radiation (LW)	Additive	40 W m^{-2}	1 day	0.5 with P; -0.5 with SW
Volumetric liquid water				
1st soil layer	Additive	0.001 m^3 m^{-3}	1 day	n/a
2nd soil layer	Additive	0.0005 m^3 m^{-3}	1 day	n/a
3rd soil layer	Additive	0.00005 m^3 m^{-3}	1 day	n/a

[*] Error parameters for the selected forcing inputs are adapted from Reichle et al. (2007).

Figure 2.3 shows the open-loop (without assimilation) and the assimilation surface (0-10 cm) soil moisture estimates across the study domain, in comparison with the 'true' state. The assimilation estimates show, relative to open loop, better overall agreement with the true fields in terms of the distribution and magnitude of soil moisture across the study domain. This demonstrates that the EnKF scheme installed for the MESH model is effective to improve the model surface soil moisture estimates. The counterpart of Fig. 2.3 for root-zone (0-35 cm) soil moisture is provided in Fig. 2.4. The root-zone soil moisture estimates are also improved through the assimilation of the surface soil moisture retrievals. The successful updating of root-zone soil moisture indicates that through data assimilation, the near-surface soil moisture information, which can be acquired by satellite microwave remote sensing, can spread to deeper soil layers that are not directly measured by satellite sensors. Note that an efficient constraint of satellite retrievals on root-zone soil moisture relies upon the model's accurate description of water movement in the soil column. In our twin experiment, the model physics is perfect since we used the same model (but with different model parameters) to generate the "true" state and to assimilate soil moisture

retrievals. In practice, it may be challenging to improve soil moisture estimates for root zone due to a lack of perfect forecast models and complete knowledge of the satellite observation errors.

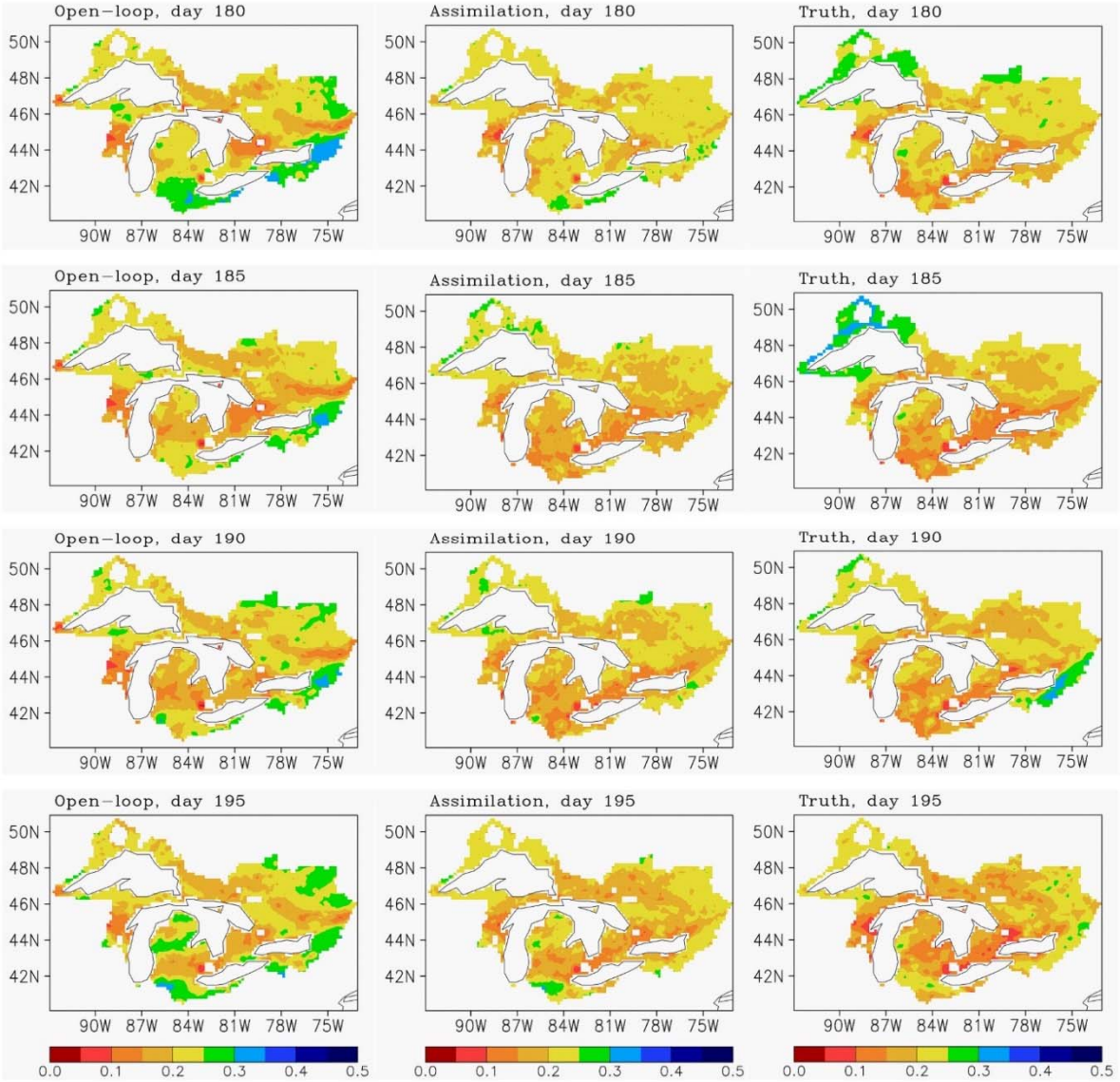

Fig. 2.3 Daily averaged surface (top 10 cm) soil moisture estimates from (left) the open-loop model, (middle) the assimilation, and (right) the true state for (top to bottom) days 180, 185, 190, and 195.

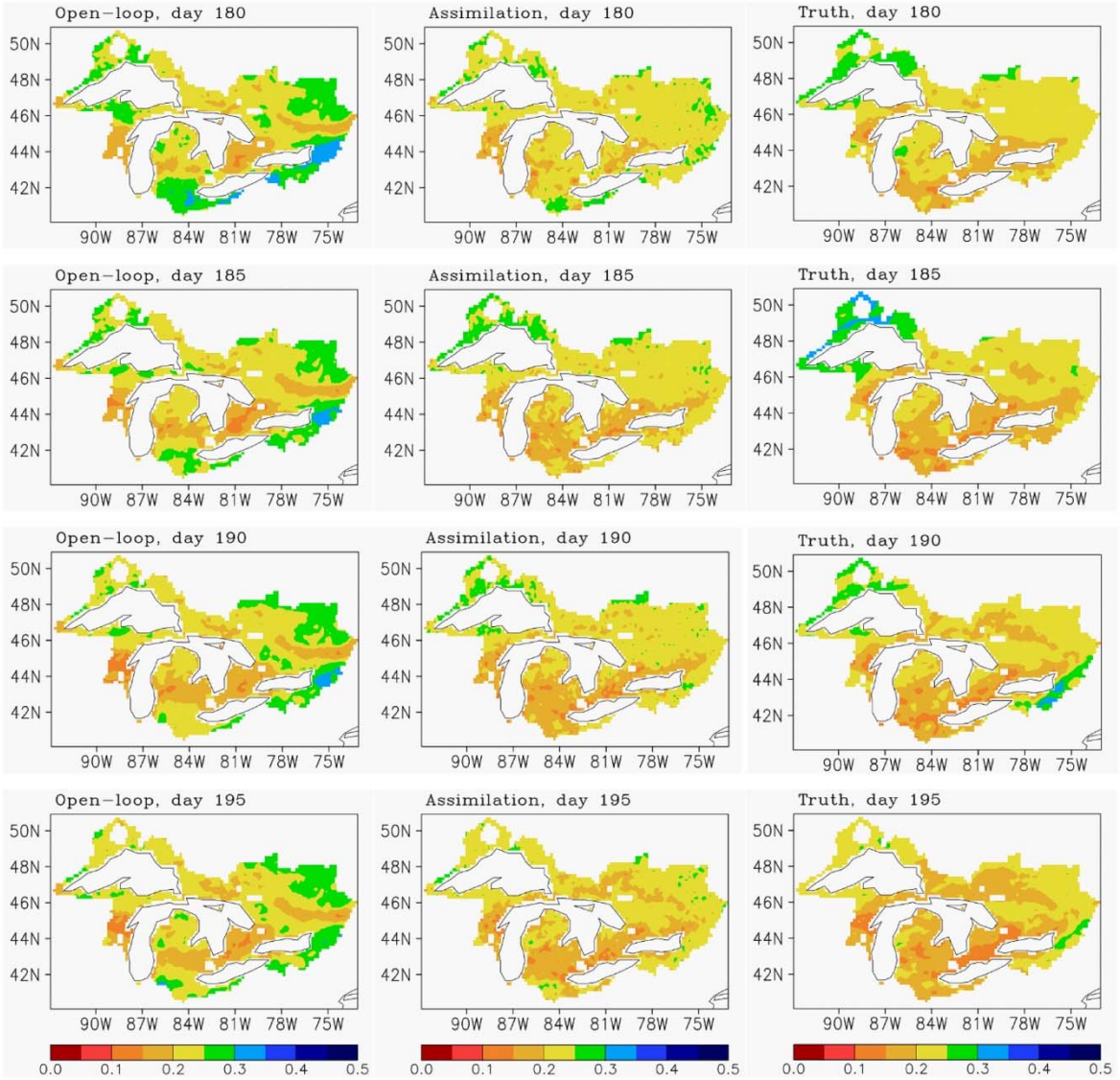

Fig. 2.4 Similar to Fig. 2.3, but for root-zone (top 35 cm) soil moisture, which is a depth-weighted average of soil moisture in the model's top two layers (0-10 and 10-35 cm).

The assimilation can be quantitatively validated against the truth using different performance criteria. Here we measure the performance of the EnKF with the soil moisture skill *R* metric, which is defined as the daily time series correlation of soil moisture estimates (open-loop or assimilation) with the truth state, to be consistent with real applications (Chapters 3-5). We compute the *R* values using data from 1 April to 31 September since

for our study domain, in practice, the effects of snow cover and frozen soils are minimal on satellite soil moisture estimates during this period.

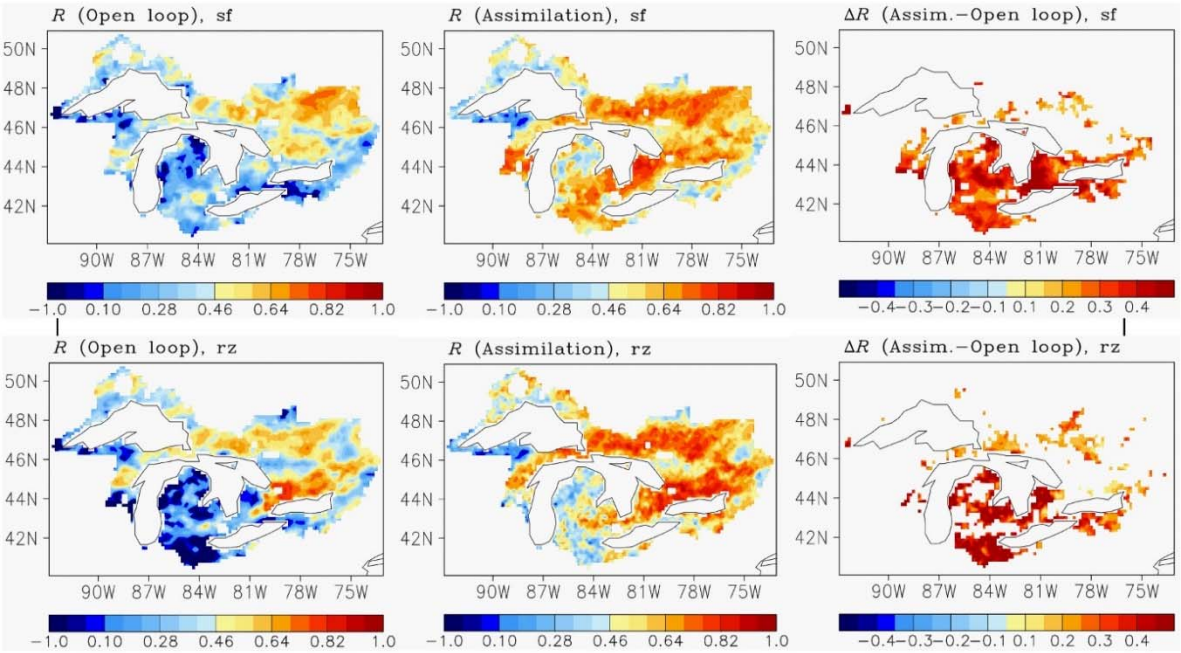

Fig. 2.5 Soil moisture skill *R* from (left) the open-loop model and (middle) the assimilation, and (right) the skill improvement Δ*R* (Assimilation minus Open-loop) for (top) the surface layer (0-10 cm) and (bottom) root-zone (0-35 cm). Δ*R* is displayed only when the open-loop *R* and the assimilation *R* are significantly (5% level) different from each other.

Fig. 2.5 compares the skill *R* values from the open-loop and the assimilation estimates for both surface and root zone soil moisture across the study domain. The Fisher *Z* transform method (Dunn and Clark, 1969; Meng et al., 1992) is used to test the significance of the skill improvement Δ*R*, defined as the skill for the assimilation product minus the skill for the open-loop estimates (Fig. 2.5, right). Overall, either the open-loop model or the assimilation provide similar spatial pattern of skill for surface (Fig. 2.5, top) and root-zone soil moisture (Fig. 2.5, bottom). The open-loop model (Fig. 2.5, left) typically provides

lower soil moisture skill R at the grids dominated by crop cover (mean of 0.29 (0.16) for surface (root-zone)) than for forest-dominated grids (mean of 0.41(0.42) for surface (root-zone)) (the vegetation cover distribution is shown in Fig. 2.2). Through data assimilation, almost all grids experience positive skill gains but with different magnitudes. In general, the skill improvement ΔR decreases with increased open-loop skill, coinciding with the finding by Reichle et al. (2008a). The stronger and statistically significant skill improvements are typically observed for crop-dominated grids. In real applications (Chapters 3-5), we will further assess the vegetation modulation of the assimilation and the impact of the open-loop skill and the satellite observation skill on the assimilation.

B. Impact of ensemble size

To perform the EnKF assimilation, the ensemble size N and the radius of influence for the observations r need to be appropriately configured. Throughout this work, r is zero since the 1D-EnKF is used (i.e. for a given observation, the analysis update (eq. (2)) is only applied to state variables at the observation location). In experiment A, an ensemble of 12 member are used. To assess the impact of ensemble size, we repeated experiment A with N = 6 (experiment B1), 50 (experiment B2), and 100 (experiment B3), respectively. The mean soil moisture skill results from these experiments are summarized in Fig. 2.6. The soil moisture skill values from experiment B1 (N=6) are slightly reduced from those obtained in experiment A (N = 12). On the other hand, increasing N to 50 or 100 has smaller (but opposite) effects. Such dependence on N is generally consistent with the synthetic EnKF results of Reichle et al. (2002a).

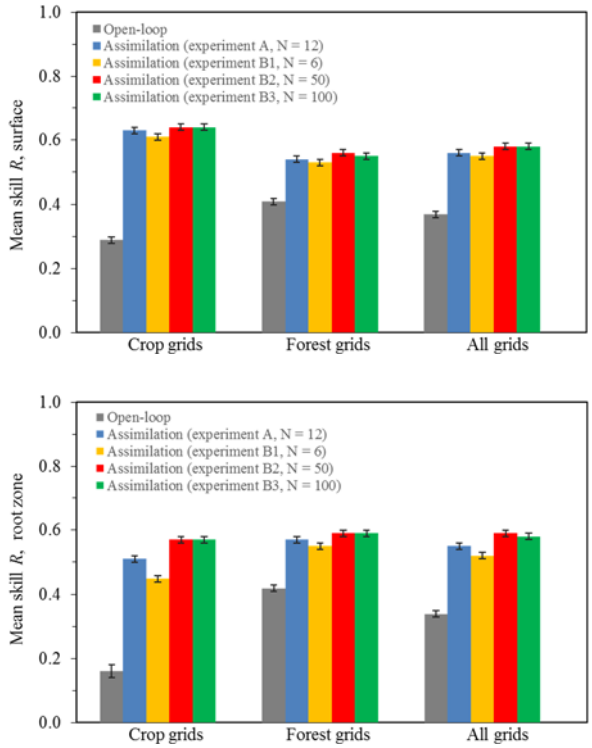

Fig. 2.6 Averaged soil moisture skill *R* for the open-loop model and the assimilation estimates from experiments A, B1, B2, and B3 (see Table 2.1 for key): (top) surface layer (top 10 cm) and (bottom) root zone (top 35 cm). The (area-averaged) *R* values are computed, respectively, for the crop-dominated grids, the forest-dominated grids, and all model grids (except for water and impervious surfaces) within the study domain. Error bars indicate 95% confidence intervals.

Overall, a small ensemble size is generally sufficient for the 1D-EnKF to perform well. This is primarily because the analysis increment calculations are conducted independently for the model grids (horizontal correlations are neglected) in the 1D-EnKF and the state vector is thus relatively small in dimension. An increased size of the ensemble would be required if horizontal correlations are taken into account (i.e., a three-dimensional EnKF) (e.g. Reichle et al. 2002b; Houtekamer and Mitchell, 1998). Additionally, note that a larger ensemble size would clearly be of advantage for suppressing statistical noise in error

covariance estimates of the state variables and error correlations between the states and the measured variable (e.g. Reichle et al. 2002a).

C. Effect of observing frequency

In reality, the satellite soil moisture retrievals derived from passive microwave sensors (e.g. AMSR-E, SMOS) or the Advanced Scatterometer (ASCAT) typically have a time resolution of 1-3 days (separately for the ascending and descending overpasses). Experiments C1 and C2 are designed to evaluate the impact of the frequency of the satellite observations on the assimilation estimates. Experiment C1 (C2) is the same as experiment A except for assimilating the synthetic retrievals at 12-hour (72-hour) intervals. Fig. 2.7 presents the mean soil moisture skill from experiments A, C1 and C2. Clearly the soil moisture estimates can be further enhanced (experiment C1) when the satellite observations are assimilated more frequently (i.e. a shorter assimilation interval). In contrast, a longer assimilation interval (experiment C2) decreased the skill improvement obtained by data assimilation. Inclusion of more observational information would clearly be of advantage. Therefore, in practice, to produce the best soil moisture estimates we should jointly assimilate the retrievals from both ascending and descending orbits or from different platforms. Draper et al. (2012) suggested that a joint assimilation of the AMSR-E and ASCAT soil moisture retrievals into the NASA Catchment land surface model led to higher soil moisture skill (anomaly R) than the alone assimilation of either of the two retrieval data sets.

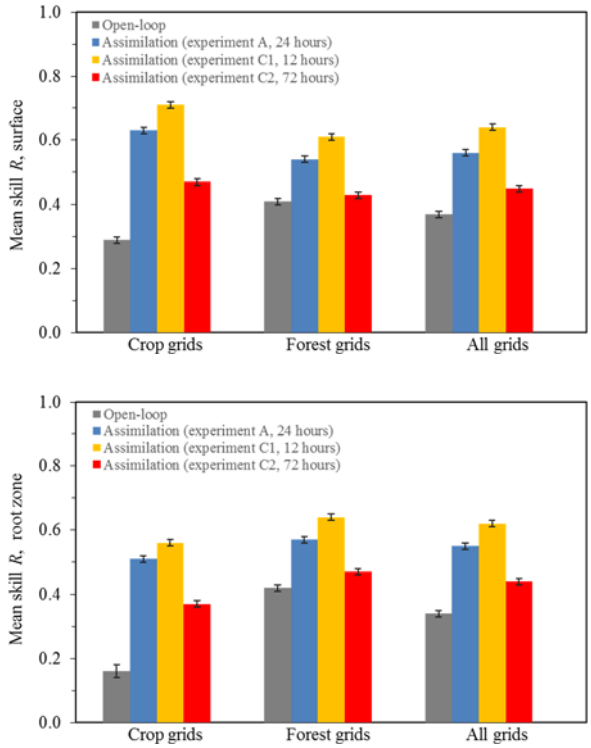

Fig. 2.7 Similar to Fig. 2.6, except that the assimilation estimates are from experiments A, C1, and C2 (see Table 2.1 for key), respectively.

In practice, however, the retrievals derived from different orbits (ascending or descending) or from different sensors may have different accuracies. For example, the descending AMSR-E soil moisture retrievals (1:30 A.M. LST) are usually expected to be more accurate than the ascending retrievals (1:30 P.M. LST) since the nighttime soil temperature and moisture profiles are more uniform. The L-band sensors (e.g. SMOS) are expected to provide better soil moisture estimates than the sensors operating at X or C bands (e.g. AMSR-E) since the latter (shorter bands) are more susceptible to vegetation effects. Such discrepancies could degrade the performance of a joint assimilation of different retrieval data sets, i.e., higher observing frequency (note that the best results from experiment C1 are based upon the synthetic retrievals with the same accuracy. In Draper et al. (2012), the

AMSR-E and ASCAT products also have similar retrieval skills and therefore their joint assimilation was promising). To demonstrate this point, we also performed an additional experiment in which we jointly assimilated two synthetic retrieval data sets with different accuracies. The first retrieval data set is same as that used in experiment A (the synthetic retrievals were obtained once daily from the truth soil moisture by adding white noise with a standard deviation of 0.08 m^3/m^3). The second retrieval data set is generated once daily (measurement time is shifted by 12 hours relative to the first data set) by applying the noise standard deviations of 0.12 m^3/m^3 to the truth soil moisture. The mean (basin-averaged) retrieval skill (correlation *R* between synthetic retrievals and the truth fields) for the first and second retrieval data sets, and their combination are 0.60, 0.44, and 0.54, respectively. Results (not shown here) suggest that the soil moisture skill *R* from the joint assimilation (12-hr intervals) is higher than that when assimilating the second retrieval data set alone (24-hr intervals), but is lower than that from experiment A (24-hr intervals). This confirms that a joint assimilation of different retrieval data sets does not necessarily yield the "best" estimation. Note that the combined retrieval set (twice daily) has a different retrieval skill from either the first or second retrieval set (once daily), although the influence of retrieval skill (see a comparison of experiments A, D1 and D2 presented below) is a consistent explanation for the observing frequency modulation on the assimilation skill.

D. Impact of retrieval skill

In experiment A, synthetic satellite retrievals (daily-spaced) were generated by applying white noise with a standard deviation of 0.08 m^3/m^3 to the truth soil moisture. The resulting retrieval data set has a mean (basin-averaged) retrieval skill of about 0.60. By applying the

noise standard deviations of 0.04 and 0.12 m^3/m^3, separately, to the truth soil moisture, two additional synthetic retrieval data sets (daily-spaced), with mean retrieval skill of 0.74 and 0.43, respectively, are produced, and are subsequently assimilated into the model (experiments D1 and D2). Figure 2.8 compares the mean soil moisture skill results from experiments A, D1 and D2. Clearly assimilating the three separate synthetic retrieval data sets produce three sets of assimilation estimates with different accuracies. Relative to experiment A, the mean soil moisture skill is further improved in experiment D1, but is degraded in experiment D2. This demonstrates that the skill improvement increases with increasing retrieval skill.

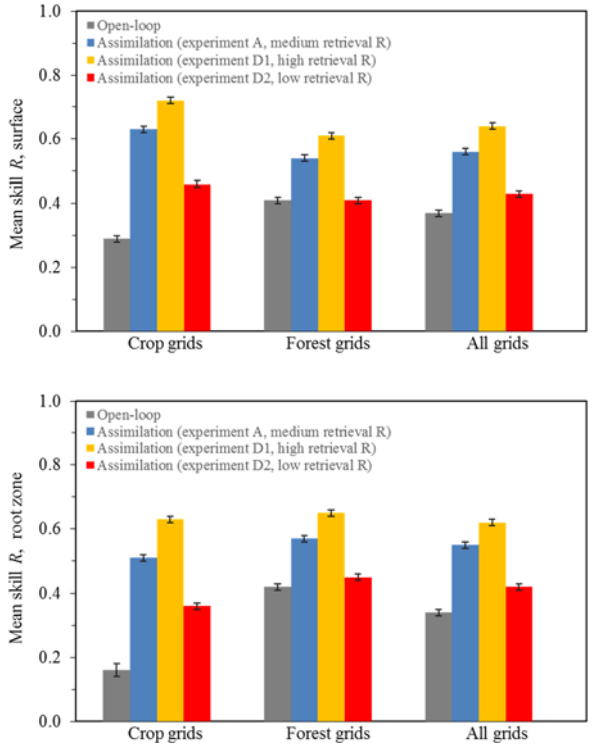

Fig. 2.8 Similar to Fig. 2.7, except that the assimilation estimates are from experiments A, D1, and D2 (see Table 2.1 for key), respectively.

E. Sensitivity to input error parameters

In experiment A, as mentioned earlier, the model error parameters, which were specified based upon order-of-magnitude considerations, were not tuned adaptively and a true observation error standard deviation was used for the synthetic retrievals. Now we test the impact of the specified model and observation error parameters through a group of new experiments. For the new experiments, the truth solution and synthetic satellite retrievals are different from those used in experiment A. Following Reichle et al. (2008b), the truth and synthetic satellite retrievals are generated as follows. First, we integrate the MESH model from 1 January to 31 December with the 2006 forcing data (the model is spun up by a 10-year repeated simulation using the 2005 forcings). The resulting soil moisture serves as the unperturbed 'open-loop' solution. Next, using the same forcing data and for the same period, we perform the MESH model ensemble simulations (12 members) with the error perturbation parameters listed in Table 2.2. We randomly select an ensemble member integration to serve as the synthetic ''truth'' (so that the perturbation parameters listed in Table 2.2 serve as the "truth'' model error inputs). For a given grid, the synthetic satellite soil moisture retrievals are generated by adding white noise with a standard deviations of 0.05 m^3/m^3 to the surface soil moisture estimates that are extracted from the 'true' fields at 24-hr intervals. Finally, the synthetic retrievals are assimilated into the MESH model with the 1D-EnKF. The assimilation will be repeatedly conducted using different sets of error perturbation parameters to explore the impact of input error parameters.

Table 2.3 Input error parameters (unit: m^3m^{-3}) for experiments E1-E25

Observation error stdev	Error stdev for the modeled volumetric liquid water in the 3 soil layers				
	2.E-04; 1.E-04; 2.E-06	5.E-04; 2.E-04; 1.E-05	1.E-03; 5.E-04; 5.E-05	2.E-03; 1.E-03; 1.E-04	4.E-03; 2.E-03; 2.E-04
0.005	E1	E2	E3	E4	E5
0.02	E6	E7	E8	E9	E10
0.05	E11	E12	E13*	E14	E15
0.08	E16	E17	E18	E19	E20
0.11	E21	E22	E23	E24	E25

* The reference (‘‘truth’’) model and observation error parameters are used.

Here we choose five sets of input model error parameters, which approximately represent five different forecast error stand deviations (stdev), and five values of observation error stdev (Table 2.3). Each of the five sets of input model error parameters and each observation error standard deviation are grouped together for use in the assimilation integrations. We do not change the forcing perturbations and the model error correlation time, which are taken from Table 2.2. Therefore, we can perform 25 assimilation experiments (E1 to E25). Note that one of the 25 experiments (E13 in Table 2.3) uses the reference (‘‘truth’’) model and observation error inputs.

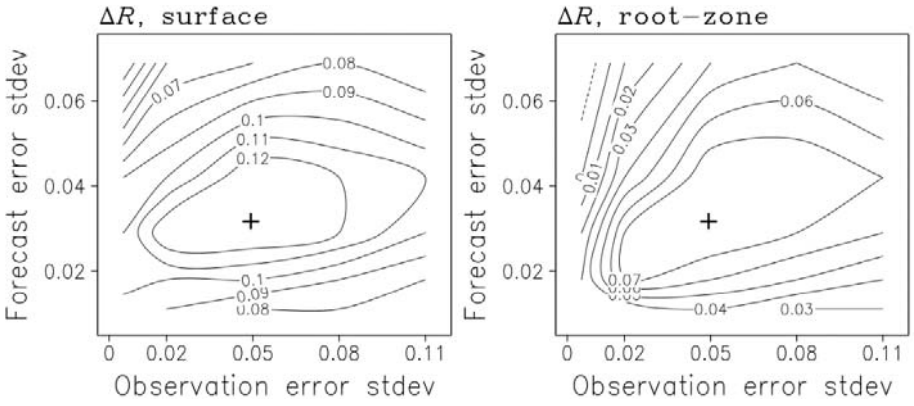

Fig. 2.9 Skill improvement Δ*R* (Assimilation - Open loop) for (left) surface and (right) root zone soil moisture as a function of the forecast and observation error standard deviations (stdev, units: m^3 m^{-3}). The plus sign denotes the assimilation experiment with the true model and observation error inputs.

Based upon the E1 to E25 results, we can obtain the assimilation performance, in terms of the (study domain-averaged) soil moisture skill improvement (Assimilation-Open loop), as a function of the (study domain and time averaged) forecast and observation error standard deviations (Fig. 2.9). As expected, the skill improvement Δ*R* in both surface and root zone soil moisture is strongest when the input error parameters are close to their true values (plus signs). Typically, when the input error parameters are wrongly specified (i.e. deviating from their true values) the EnKF filter assimilation still produce an increased skill (i.e. a positive Δ*R*), although the skill improvement Δ*R* will be decreased. In particular, if a severe underestimation of observation error occurs, the skill improvement Δ*R*, especially for root zone soil moisture, will be weak or even negative (i.e. the assimilation estimates are worse than the open loop). This illustrates that even without on-line (adaptive) tuning of the observation and model error parameters the EnKF filter will typically be able to improve soil moisture estimates as long as the observation errors are not severely underestimated.

Reichle et al. (2008b), which measured the assimilation performance using the root-mean-square error (RMSE) metric, came to the same general conclusions.

2.5 Summary and Discussion

In this chapter, we presented the implementation of the 1D-EnKF scheme to assimilate satellite soil moisture into the standalone version of MESH model. To examine the performance of the established assimilation scheme under different conditions, we have conducted a series of synthetic assimilation experiments in which satellite soil moisture observations and the reference ("truth") states were produced with the same forecast model. The experiments demonstrate the capability of the assimilation system to accurately approximate the "true" surface and root zone soil moisture states with satellite observations and the intentionally degraded model estimates. The soil moisture skill R metric is used to measure the performance of the EnKF filter, to be consistent with real assimilation applications (Chapters 3-5). Through assimilation of satellite soil moisture, almost all areas experience positive skill gains ΔR (assimilation-open loop) but typically with stronger and statistically significant ΔR for cropped grids, which generally exhibit low open-loop skill. A small ensemble size is generally sufficient for the 1D-EnKF to perform well because the analysis increment calculations are conducted independently for the model grids (horizontal correlations are neglected). An increased observing frequency (i.e. a shorter assimilation interval) typically can further enhance the assimilation estimates. Therefore, in any practical application, to produce the best estimates we should jointly assimilate the retrievals from all the available sources. However, note that a joint assimilation (i.e. an

increased observing frequency) of the retrieval sets with significantly different observation skills does not necessarily yield the "best" estimation. The skill improvement typically increases with increasing retrieval skill. Even without on-line (adaptive) tuning of the observation and model error parameters the skill of the assimilation product typically exceeds the skill of the open-loop model (i.e. a positive ΔR) except when the observation errors are severely underestimated. The findings provide an important reference for the practical application of the assimilation scheme (e.g. the choice of the ensemble size, the specification of input errors, etc.).

Note that in practice, the assimilation of satellite soil moisture will encounter a number of critical challenges, which were avoided in our synthetic experiments. They mainly include: (i) the model-satellite measurement scale discrepancy. Soil moisture derived from spaceborne passive microwave measurements typically have relatively coarse spatial resolution (> ~ 40 km; Table A3), whereas there is an increased demand for conducting land/hydrologic simulations at high spatial resolution. This raises a question: how to integrate coarse-scale satellite products and fine-scale land/hydrologic models. For the 1D-EnKF assimilation, a priori disaggregation scheme may be needed, i.e., coarse resolution observations are disaggregated and remapped onto the model grids prior to assimilation. For the 3D-EnKF filter (i.e. horizontal correlations between model grids are considered), we can conduct a direct assimilation of coarse-scale satellite products by upscaling the model forecast; (ii) Statistical bias between satellite and model soil moisture estimates. In practice, satellite-based soil moisture and model estimates typically exhibit different climatologies, which will violate the hypothesis of unbiased errors in model and

observation (for a bias-blind assimilation system). To reduce or remove the satellite-model bias, a priori observation rescaling by matching the cumulative distribution functions of the two data sets is often practical (e.g. Reichle and Koster, 2004). A prior calibration of the model with the climatology of satellite soil moisture also can be used to remove the bias (e.g. Kumar et al., 2012). (iii) Difficulty in quantifying satellite observation error. The synthetic experiments indicate that accurate specification of observation error covariance is crucial to the success of the analysis. Satellite soil moisture retrievals are typically subject to both instrumental errors and representativeness errors. The latter are caused primarily by the observation operator used in the retrieval algorithm and the misfit between the observation space and the model space. In reality, the errors in satellite retrievals, especially the representativeness errors are difficult or impossible to completely estimate since they often vary with time and space. Some approximations could be efficient. For example, the satellite climatology can approximately serve as the observation error inputs for the EnKF integrations (e.g. Liu et al., 2011; Draper et al., 2012); (iv) Sampling errors in point-scale ground measurements. In practice, in situ measurements typically serve as the "ground truth", but in most cases only sparse point measurements are available. Point-scale ground measurements typically contain large sampling errors (e.g. Crow et al., 2012), which poses an obstacle to the validation of assimilation, especially when using the RMSE metric. To this end, we need to upscale sparse point measurements (e.g. Crow et al., 2012) or we may measure the performance with the correlation *R* metric considering that the temporal variability of soil moisture observed by point measurement may be spatially representative (e.g. Loew and Mauser, 2008; Martinez-Fernandez and Ceballos, 2005). These problems will be appropriately addressed in our real assimilation experiments.

CHAPTER 3

Assimilation of SMOS Soil Moisture over the Great Lakes Basin

3.1 Introduction

Soil moisture, especially its anomaly information, is critical to weather and climate forecast initialization (e.g., Wolfson et al., 1987; Zhang and Frederiksen, 2003; Lau & Kim, 2012; Zeng et al., 2014). Microwave remote sensing technology offers an important approach for soil moisture estimation because changes in soil water content strongly affect the soil's dielectric properties. Satellite microwave remote sensing holds the ability to provide the large-scale spatially distributed near-surface soil moisture estimates, which, relative to point in situ measurements, are more compatible in space with land/hydrologic models, especially the distributed models. In the past decade, satellite microwave soil moisture observations have been intensively integrated into land surface and hydrologic models, in particular through advanced data assimilation that merges the observation and the model forecast based on estimates of their respective error characteristics (see a review paper by Xu et al., 2014). Data assimilation can spread and smooth the observed information in time and space (Reichle, 2008). Through data assimilation, the remotely-sensed near-surface soil moisture information can be propagated to the soil layers or the model variables that are not directly measured by satellites (e.g. Reichle & Koster, 2005). Additionally, in a data assimilation system satellite retrievals from different platforms can be merged into the same model framework to produce a single optimal state estimation of interest (e.g. Draper et al., 2012).

Until recently, the satellite soil moisture products were mainly based upon the X (8-12 GHz) or C (4-8 GHz) band measurements, such as the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E), the Scanning Multichannel Microwave Radiometer (SMMR), the Tropical Rainfall Measuring Mission Microwave Imager (TMI), the Advanced Scatterometer (ASCAT), and the RADARSAT series. A series of assimilation experiments based upon these products (e.g. Reichle & Koster, 2005; Reichle et al., 2007; Drusch, 2007; Liu et al., 2011; Draper et al., 2012; Crow et al., 2005; Brocca et al., 2010) have demonstrated the potential of satellite retrievals to improve the predictive capabilities of land surface and hydrologic models (e.g. soil moisture and runoff estimates) and provided insight into the main challenges in this field of research (e.g. the model-satellite scale discrepancy; the statistical biases between satellite product and model estimation). However, X and C band sensors are susceptible to vegetation cover and are sensitive to only the near-surface soil moisture (top 1 to 1.5 cm). The launch of European Space Agency's (ESA) Soil Moisture and Ocean Salinity (SMOS) satellite that carries an L-band (~1.4 GHz) Microwave Imaging Radiometer with Aperture Synthesis (MIRAS) (Kerr et al., 2001; 2010) has opened up the new opportunities for land data assimilation. The assimilation of SMOS soil moisture is more attractive because the L-band microwave has a stronger penetration of vegetation and soil (as opposed to those operating at X or C band), which can provide surface soil moisture estimates for a wide range of vegetation conditions and thus offer the new opportunities for assessing the vegetation modulation of the assimilation.

In recent years, there has been an intensive global research effort to assimilate SMOS soil moisture data in various models (e.g., Zhao et al., 2014; Ridler et al., 2014). Zhao et al. (2014) incorporated the SMOS soil moisture retrievals into a land surface model by minimizing the distance of the model solution from the SMOS observation and the background model estimate (by calibrating the model using the SMOS data first), which produced the improved surface soil moisture estimates. However, the study averaged the SMOS data across the entire domain (located in the central Tibetan Plateau) and the assimilation was performed at a coarser scale (~100 km) than the SMOS product scale (~ 15 km). A more recent study by Ridler et al. (2014) assimilated SMOS soil moisture in a bias-aware system (i.e., the observation bias is estimated jointly with the model state by state augmentation). The assimilation was conducted at a fine scale (by applying a vegetation-based disaggregation scheme to the SMOS observation bias) and led to superior soil moisture estimates (in terms of the square of the correlation), especially for the surface layer, although only one node retrievals were used.

In this chapter, an ensemble Kalman filter (EnKF) is utilized to assimilate SMOS soil moisture retrievals (Level 2) into a coupled land-surface and hydrological model MESH over the Great Lakes basin. Due to the bias between the retrievals and the model surface soil moisture, a priori rescaling on the SMOS retrievals is performed by matching their cumulative distribution function (CDF) to the model surface soil moisture's CDF. The retrievals, the open-loop model (no assimilation) soil moisture, and the assimilation estimates are validated against in situ soil moisture measurements from the Michigan Automated Weather Network, the Soil Climate Analysis Network, and the Fluxnet-Canada Research Network, in terms of the daily-spaced time series correlation coefficient (soil

moisture skill *R*). Our study differs from previous SMOS assimilation studies in three aspects: (1) the assimilation is conducted at a grid scale similar to the SMOS product scale (~ 15 km), and the assimilation estimates are validated at both the grid-scale and the subgrid-scale; (2) the Great Lakes basin was chosen as the study domain since it offers a range of vegetation conditions that favor the assessment of the vegetation impact on the assimilation; and (3) 4 years of SMOS data (2010-2013) are used, and the overall consistency between the years strongly demonstrates the robustness of our general conclusions. This chapter is organized as follows. In section 3.2, the data sets, the forecast model, and the assimilation scheme are described. Section 3.3 presents the skill for the SMOS soil moisture. Section 3.4 is focused upon the assimilation results. A summary and discussion is provided in section 3.5.

3.2 Data and methods

3.2.1 SMOS soil moisture retrievals

In this work, we use the SMOS Level 2 Soil Moisture User Data Product (MIR_SMUDP2) delivered by ESA. The product comprises the instantaneous soil moisture retrievals (rather than the daily composite as provided in the Level 3 product) and abundant reference information, such as geophysical features, retrieved standard deviation (RSTD), etc. The retrieved soil moisture is primarily based upon an iterative algorithm, which matches the modeled L-band emission of the surface to that observed by SMOS/MIRAS (Kerr et al., 2008; 2012). SMOS has a footprint of 43 km on average and a temporal resolution of 1-3

days for both ascending (6:00 am LST) and descending (6:00 pm LST) orbits. The MIR_SMUDP2 soil moisture retrievals are equally spaced at about 15 km (oversampled by a factor of nine). Four years (2010-2013) of SMOS retrievals from both ascending and descending overpasses are used in this study. Utilizing the attached reference information, a filtering is performed to exclude the retrievals with a large RSTD (> 0.08 m^3/m^3) and those contaminated by open water, frozen surface, snow, or rain, etc. To conduct the evaluation and assimilation, SMOS retrievals are resampled onto the hydrological forecast model grids (~15 km resolution) using a nearest neighbor approach. Whenever and wherever the model (combined with the rainfall forcing data) indicates the presence of precipitation, frozen soils, or snow cover, the satellite retrievals are also excluded from the evaluation and assimilation. Note that the processor version of the Level 2 product was changed over the four years with V501 (REPR data set) for 2010/2011 and V551 (OPER data set) for 2012/2013. Since different dielectric constant models are used in the two versions, there may be inconsistencies in the absolute magnitude of SMOS retrieval between 2010/2011 and 2012/2013.

3.2.2 Hydrological model and in situ measurements

The forecast model used here is Environment Canada's standalone MESH (Modélisation Environmentale-Surface et Hydrologie) model (Pietroniro et al., 2007), which originates from the coupling of the land surface scheme CLASS with the hydrological model WATFLOOD (Soulis et al., 2000). The primary feature of MESH is that the model uses a Grouped Response Unit (GRU) approach to resolve the subgrid-scale variability. A GRU

is a grouping of subareas with similar soil and vegetation attributes, and each model grid cell is represented by a limited number of distinct GRUs weighted by their respective cell fractions. In the version of MESH used in this work, the identification of GRUs is based solely on the land cover types, i.e., each GRU corresponds to one land cover class (other soil characteristics are assumed to be same for the same GRU). The soil column is partitioned into three layers (0-10, 10-35, and 35-410 cm) to resolve soil moisture and temperature dynamics. At the moment, the land surface scheme considers only the vertical water movement between the soil layers, which is governed by Richard's equation (Soulis et al., 2000). Within a grid cell, the fluxes and variables are computed independently for GRUs, ignoring the interactions between GRUs. The overall fluxes and prognostic variables of a grid cell are obtained by taking a weighted average of those from GRUs. The lateral movement of water between grid cells is not taken into account. The resulting horizontal flows (overland flow, interflow, and base flow) at grid cells are ultimately be routed into the stream and river network systems.

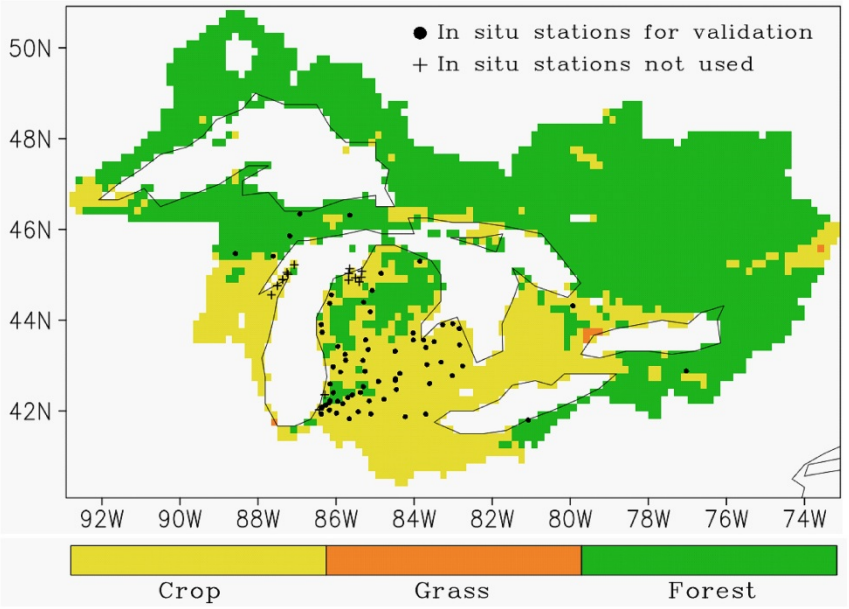

Fig. 3.1 Vegetation types (gridded at 1/6th of a degree resolution) over the Great Lakes basin and location of in situ stations for soil moisture measurements. In situ stations are from the Michigan Automated Weather Network (79 sites), the Soil Climate Analysis Network (3 sites), and the Fluxnet-Canada Research Network (1 site). Stations that are not used for validation are marked with plus signs (SMOS retrievals are not available or not considered over these stations due to the impact of open water).

The study domain for this work is the Great Lakes basin (Fig. 3.1). The basin, straddling the Canada-United States border, consists of the largest group of freshwater lakes on earth and the surrounding lands, with a drainage area of about 1000,000 km^2. The five primary fresh lakes are naturally interconnected and contain roughly one-fifth of the world's fresh surface water supply. The model configurations are similar to those used in Pietroniro et al. (2007) and Haghnegahdar et al. (2014). The model is run at a resolution of 1/6th of a degree (~15 km) using a time step of 30 minutes. Each model grid cell is divided into a mosaic of GRUs. Each GRU corresponds to one land cover type and is weighted by the

fraction of the land cover class within the cell. Seven GRU types are used for this domain, including crop, grass, deciduous forest, coniferous forest, mixed forest, water, and impervious. The land cover types were derived from a United States Geological Survey (USGS) climatological database. In this work, the model parameter sets are directly taken from a global calibration experiment where GRU specific parameters are calibrated basin-wide to streamflow observations (Haghnegahdar et al., 2014). Here MESH is forced using the gridded hourly precipitation data derived from the Canadian Precipitation Analysis (CaPA; Mahfouf et al., 2007); other meteorological forcing data (incoming shortwave and longwave radiations, surface air temperature, wind speed, pressure, and specific humidity) come from the Global Environmental Multiscale (GEM) model forecasts (Mailhot et al., 2006).

In this work, in situ soil moisture measurements (Fig. 3.1) from the Michigan Automated Weather Network (MAWN; http://www.agweather.geo.msu.edu/mawn/), the Soil Climate Analysis Network (SCAN; http://www.wcc.nrcs.usda.gov/scan/), and the Fluxnet-Canada Research Network (FCRN) are used to validate the SMOS retrievals, the model and the assimilation estimates. The specification of in situ stations and measurements is provided in Table A5. MAWN is comprised of about 79 stations. Each station uses two Campbell Scientific water content reflectometers (CS615 or CS616) to measure soil moisture. The two probes are horizontally inserted to provide hourly soil moisture measurements at depths of 10 and 25 cm (for 46 MAWN sites) or are vertically installed to measure soil moisture in the upper 60 cm profile (0-30 and 30-60 cm) (for 33 MAWN sites since about the middle of year 2008). Additionally, in situ data from three

SCAN sites (SCAN2003, 2011, and 2073) and one FCRN site (the Borden forest station) are included in this study. At SCAN sites, Stevens Hydra Probe sensors are horizontally inserted to provide hourly soil moisture measurements at 5, 10, 20, 50, and 100 cm below the surface, while at the Borden station (44.32°N, 79.93°W) 30 min-averaged soil moisture measurements are taken with CS615 probes at 2, 5, 10, 20, 50, and 100 cm below the surface at two locations. A filtering step is applied to all in situ data to ensure the reliability and effectiveness of the subsequent validations. In situ soil moisture observations are rejected if (1) they are beyond any realistic ranges (e.g., too high or too low to be explained by physical variability); (2) the time series contains sudden changes (significant "jump") that are impossibly attributed to any physical process; or (3) the soil is frozen.

3.2.3 The EnKF method

Data assimilation typically can be viewed as a process to optimally merge the model forecast and the observed information based upon some estimate of their error characteristics. A great number of methods have been developed for land/hydrologic data assimilation (Table A1). The reader is referred to the relevant articles for details on the properties of different algorithms. In the present study, the ensemble Kalman filter (EnKF) is used to assimilate SMOS soil moisture in the MESH model. The traditional Kalman Filter (KF) and its various variants (extended Kalman Filter, EKF; EnKF) are typical 'filtering' (or sequential) assimilation techniques. In the traditional KF, each assimilation cycle consists of a forecast step and an analysis step. In the forecast step, the forecast model is integrated forward in time (from an initial or analysis state) with an additional error covariance equation (linear model operator) to propagate the error information, while at

the analysis step the new observation is used to update the current forecast estimation. The KF is valid only for linear systems. Its nonlinear variant, the EKF, can be utilized to solve the nonlinear optimal estimation problem. The EKF still explicitly estimates and propagates the error information, but with a linearized and approximate error covariance equation. In practice, however, the full error covariances are difficult or impossible to directly estimate due to an expensive computational cost and insufficient error information, especially for large-scale applications. Additionally, the EKF may not be suitable for highly nonlinear systems since the high-order moments are ignored in its error covariance equation. To this end, Evensen (1994) proposed the EnKF scheme.

The primary innovation of the EnKF is that a Monte Carlo approach is used to estimate model and measurement error statistics. The probability density of the model states is represented using an ensemble where the mean is the best estimate (Gaussian assumption), and the ensemble spread defines the error variance. The model error statistics evolve by integrating the ensemble of model states forward in time. The measurement errors are represented using another ensemble with the mean equal to zero (Gaussian assumption) and the spreading of the ensemble consistent with the realistic or predefined observation error variance. The measurement errors are imposed onto the actual measurement to yield the ensemble of observations. At measurement times, a variance-minimizing analysis is applied to the ensemble of model forecast states, given by

$$x_j^+ = x_j^- + P^- H^T [H P^- H^T + R]^{-1} [y_j - H x_j^-] \quad , j = 1, \ldots, N \tag{3.1}$$

where j is the ensemble member index, counting from 1 to the ensemble size N. x_j^- and x_j^+ denote the a priori and posterior model state estimates, respectively. y_j represents the

perturbed observation. H is the measurement operator. P^- and R denote the error covariances for model forecast and observation, respectively. In contrast to the EKF, the error evolution is implicit and fully nonlinear in the EnKF but with a lower rank (finite ensemble size).

3.3 Skill for SMOS soil moisture retrievals

SMOS soil moisture products have been evaluated over different regions/scales with in-situ data from point (e.g. Al Bitar et al., 2012; Albergel et al., 2012) or network measurements (e.g. Gherboudj et al., 2012; Jackson et al. 2012; Zhao et al., 2014; Ridler et al., 2014). The validation studies have suggested that the SMOS retrievals typically exhibit an underestimation bias. The performance of the retrievals varies with the scale of the validation, typically showing a better accuracy for a large-scale average. Overall the desired accuracy of 0.04 m^3/m^3 for SMOS retrievals is met wherever the vegetation cover is light (nominal surfaces). However, the validation of coarse-scale satellite soil moisture unavoidably suffers from the disparity in spatial representativeness between satellite products and ground measurements (Jackson et al., 2010; Crow et al., 2012). Point-scale ground measurements, relative to the spatial averages, typically contain large uncertainties, which are strongly controlled by the precipitation type (e.g. convective or stratiform) and the local variability in geophysical fields (such as surface type, soil texture, and topography). Even for a soil moisture network, the spatial extent of ground observations may not always represent the satellite footprint area since the latter varies over time. These

factors pose an obstacle to validating satellite soil moisture products, especially when using the root-mean-square error (RMSE) metric.

Although point measurements are not readily converted to the spatial averages, the temporal variability of soil moisture observed by point measurement may be spatially representative (e.g. Brocca et al., 2009; Loew & Mauser, 2008; Martinez-Fernandez & Ceballos, 2005). Fig. 3.2 presents the soil moisture time sequences observed at four pairs of neighboring sites (all from MAWN). Each pair of sites may lie within the same SMOS footprint area. Although the absolute magnitudes of soil moisture are not necessarily matched, each pair of sites typically show good agreement for the temporal pattern of soil moisture. Likewise, at the Borden station soil moisture measurements taken at two locations are not always same in magnitude but showing consistent temporal dynamics for the period of record (not shown). Regarding the SCAN measurements, Liu et al. (2011) suggested that the SCAN point observations were highly correlated with the watershed average soil moisture obtained from network measurements and thus were suitable for evaluating the assimilation estimates with the correlation metric. Thus, overall the point-scale measurements (from MAWN, SCAN, and Borden) being used in this work are assumed to represent the areal average (satellite product scale or model grid cell) in terms of the temporal variability of soil moisture.

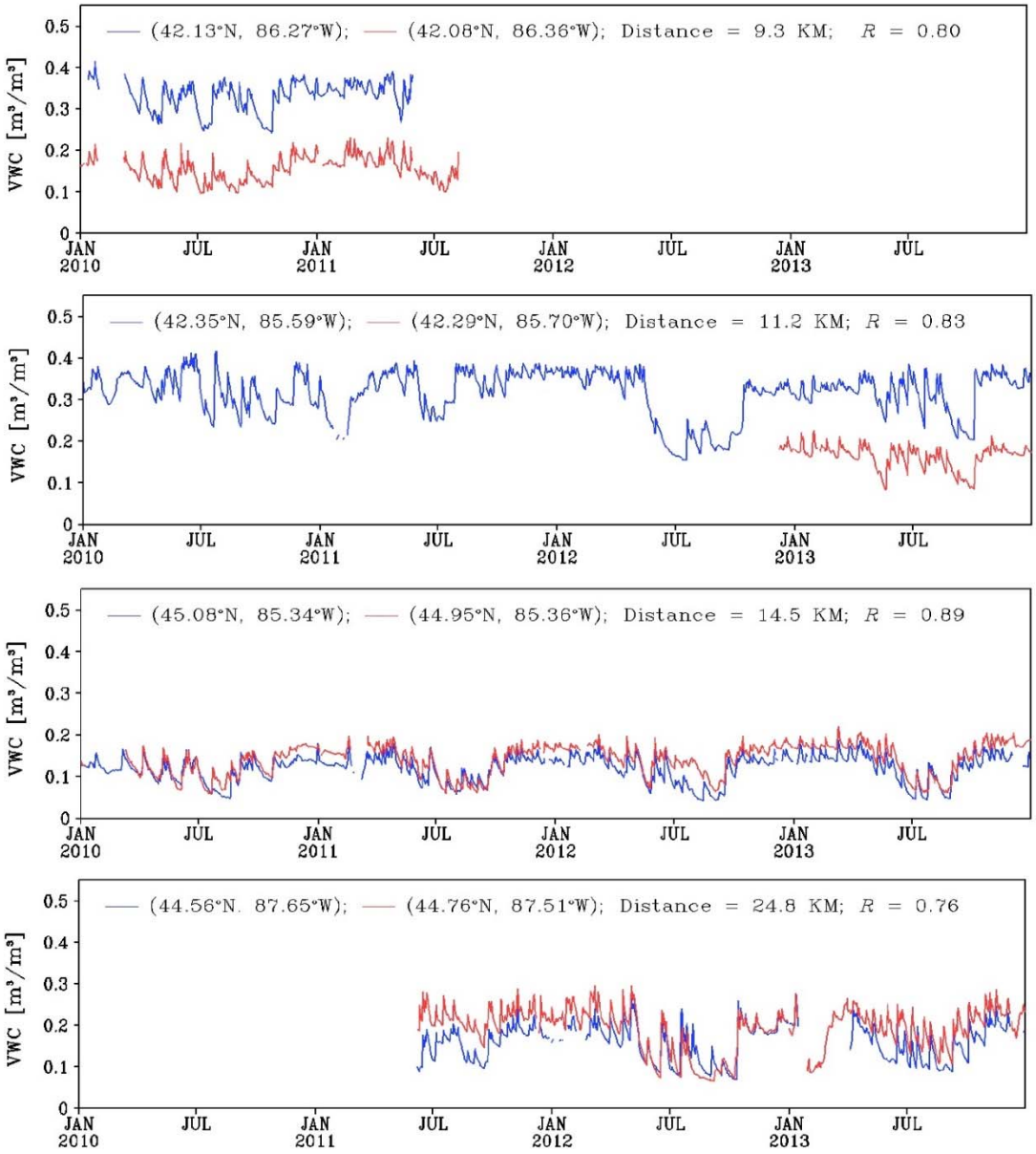

Fig. 3.2 Comparison of volumetric water content (VWC) daily time sequences for four pairs of MAWN sites. For each panel, location of the two sites and their distance are shown, and *R* denotes the correlation coefficient between the two soil moisture sequences. The labels on the *x*-axis denote the first day of each month.

Since the absolute magnitude of soil moisture for the areal average (corresponding to the satellite footprint scale) is difficult to estimate based upon point-source observations,

the SMOS retrievals are not validated with the RMSE metric in this study. Instead, we only assess the SMOS soil moisture skill *R*, which is defined as the daily time series correlation of SMOS retrievals with point measurements. SMOS measures only the water content within the top ~ 5-6 cm soil layer. Although the 5 cm depth matches well with the average soil penetration of SMOS, here the SMOS soil moisture skill is computed using in situ measurements taken at 10 cm depth or in the top 30 cm profile (for those sites with the vertically installed probes), to be consistent with the subsequent assessment of the model surface soil moisture skill (sections 3.4.2 and 3.4.3). Overall the use of 10 cm-depth and 0-30 cm measurements is acceptable in this study since typically the time patterns of soil moisture between in situ measurements taken at 5 cm, 10 cm, and 20/25 cm are highly correlated.

To be consistent with the subsequent 1D-EnKF (section 3.4), the SMOS retrievals (from both ascending and descending orbits) are mapped onto the MESH model grid cells (at a 1/6th degree resolution) using a nearest neighbor approach. Given a model grid, the SMOS skill (daily time series correlation *R* with in situ data) is assessed using in situ measurements falling within the grid cell. Typically only one in situ site is available per model grid cell. We do not compute the *R* values when any of the following occurs: (1) the effective length of SMOS soil moisture daily time series is less than 60 days per year; (2) in situ soil moisture (unfrozen) time series are shorter than 100 days per year; (3) the time series standard deviation of in situ soil moisture is less than 0.02 m^3/m^3 (since the measurement noise may significantly impact the *R* values when the time series standard deviation is too small); or (4) linear or quadratic trends in the SMOS or in situ time series

significantly contribute to the correlation (by examining if a linear regression or a polynomial of the 2nd degree gives statistically significant trends).

Fig. 3.3 shows the SMOS soil moisture skill. To be consistent with the subsequent validation of the assimilation estimates, we classify the model grid cells into four types: (1) sCmC: the SMOS soil moisture has a nominal (low vegetation) surface type (the retrieval case value is 12 in MIR_SMUDP2; in this study, for the grids of interest, a nominal surface is typically a crop surface) and the crop cover is also dominant (> 50%) within the model grid square; (2) sCmF: the SMOS soil moisture is from a crop surface node, but the fraction of forest cover (the sum of the deciduous, coniferous, and mixed forest classes) within the model grid cell exceeds 50% (note that since a model grid square and the SMOS node mapped onto the grid are not exactly matched in space their surface types may be not always the same); (3) sFmC: the SMOS retrieval mapped onto a model grid is from a forest surface node (the retrieval case values equals 11 in MIR_SMUDP2), but the model grid is dominated by crop cover; and (4) sFmF: the SMOS retrieval case is a forest surface and the model grid is also covered dominantly by forest. Table 3.1 provides the median and mean skill *R* for each grid type.

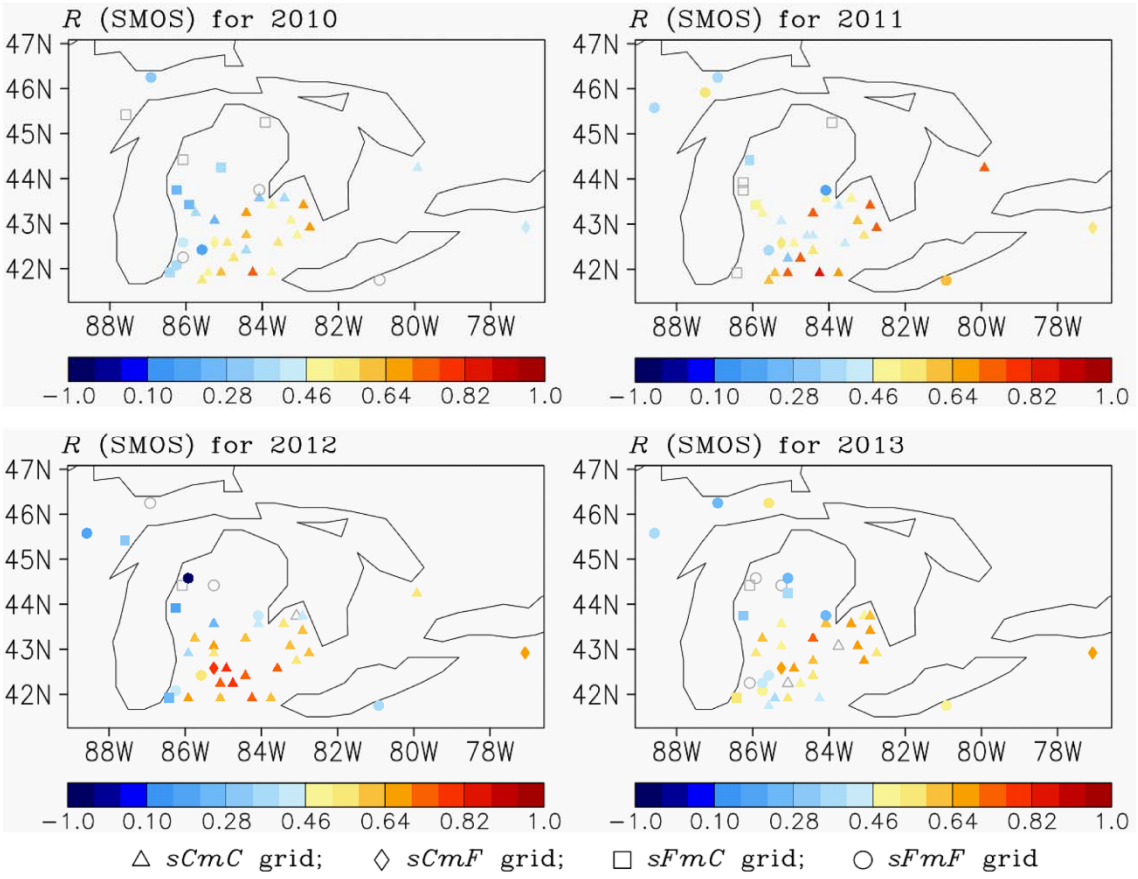

Fig. 3.3 SMOS soil moisture skill, which is defined as the correlation coefficient *R* of daily averaged SMOS retrievals with in situ measurements, over four individual years. *R* is computed after the SMOS retrievals are mapped onto the model grid coordinate system. Symbols indicate the model grid types as defined in the text: (triangles) sCmC, (diamonds) sCmF, (squares) sFmC, and (circles) sFmF. *R* values that are not significantly (5% level) different from zero are indicated by open symbols in grey.

The SMOS retrievals from crop surfaces, i.e., at the sCmC and sCmF grids (triangles and diamonds in Fig. 3.3), typically show modest to high skill *R* (median of 0.55 for sCmC and 0.64 for sCmF), which means that the time variation of SMOS soil moisture at these grids agrees well with the temporal pattern of in situ measurements. In contrast, the SMOS

observation skill is usually low at the sFmC and sFmF grids (squares and circles) where the retrievals come from forest cover-dominated surfaces (with a median of 0.23 for sFmC and 0.32 for sFmF). The identified SMOS skill disparity between forested and cropped surfaces is consistent with the fact that the satellite retrieval capabilities decrease with increased canopy density. Additionally, the forest grids with low SMOS skill are typically located near the lakes. The corresponding SMOS retrievals may also be impacted by the presence of open water and a low quality of the reconstructed brightness temperatures caused by the Gibbs effect (Gibbs, 1899) over the coast. Al Bitar et al. (2012) suggested that the temporal dynamics of soil moisture between SMOS and SCAN/SNOTEL point stations were typically well matched, but negatively affected by the increasing forest and/or water fractions within the satellite node. Note that such a vegetation modulation of the SMOS observation skill can strongly impact the model soil moisture skill gain through data assimilation (sections 3.4.2 and 3.4.3).

Table 3.1 Median and mean skill *R* within each grid type for soil moisture from SMOS, the open-loop model, and the assimilation, respectively

Soil layer	Grid type	*N*	Median *R*			Mean *R* with 95% confidence intervals		
			SMOS	Open-loop	Assim.	SMOS	Open-loop	Assim.
0-10 cm	sCmC	91	0.55	0.39	0.64	0.55±0.01	0.39±0.01	0.64±0.01
	sCmF	8	0.64	0.60	0.74	0.62±0.04	0.61±0.03	0.73±0.02
	sFmC	21	0.23	0.40	0.52	0.23±0.04	0.42±0.02	0.50±0.02
	sFmF	33	0.32	0.62	0.60	0.29±0.03	0.60±0.02	0.61±0.02
0-35 cm	sCmC	89	-	0.51	0.72	-	0.47±0.01	0.71±0.01
	sCmF	8	-	0.65	0.80	-	0.67±0.03	0.77±0.02
	sFmC	20	-	0.49	0.54	-	0.48±0.02	0.53±0.02
	sFmF	32	-	0.67	0.65	-	0.64±0.02	0.62±0.02

Grid types are defined in the text. *N* denotes the combined number of grid-based *R* values for 2010-2013.

3.4 Assimilation of SMOS soil moisture

A 1D-EnKF (i.e., the analysis increment computation is performed independently for the model grids) with 12 ensemble members is applied to assimilating SMOS retrievals into the MESH model. Given a model grid, in the EnKF analysis equation (3.1) the model state vector x_j (dimension is 21) is comprised of the volumetric liquid water content from all the seven GRUs within the grid cell and all the three soil layers modelled in MESH. The observation y_j is the perturbed SMOS soil moisture and the corresponding model prediction Hx_j^- denotes the model estimates of the grid-averaged volumetric liquid water content (a weighted sum of GRU values) in the model surface layer (0-10 cm). The assimilation period is from 1 January 2010 through 31 December 2013. The model is spun up for a 8-year period with the 2002-2009 forcing data.

In the EnKF, the estimates of the model forecast errors are derived from an ensemble of model integrations. To represent random errors in the forcing inputs, cross-correlated forcing perturbation fields are generated (Table 2.2), following Reichle et al. (2007). The selected perturbation parameters are largely based upon order-of-magnitude considerations (Reichle et al., 2002a). To account for the model forecast errors due to deficiency in model physics and/or parameters, temporally correlated error perturbations are applied to soil moisture (volumetric liquid water content) estimates in the model. The following equation is used to yield the time evolution of error perturbations.

$$q_k = \sigma \left[(1 - k/\tau)\, w_0 + \sqrt{1 - (1 - k/\tau)^2}\, w_k \right] \tag{3.2}$$

where q is the error perturbation ensemble, w is white noise ensemble with mean of 0 and variance of 1, τ is the correlation time length (unit: the model time step), k denotes the time index ($0 \leq k < \tau$), and σ represents the specified model error standard deviation. Currently, the 0.001 m^3/m^3, 0.0005 m^3/m^3, and 0.00005 m^3/m^3 error standard deviations are applied to the model's three layers (0-10, 10-35, and 35-410 cm), respectively. The model error correlation time is set to 1 day, which is the approximate frequency for the SMOS observations (1 or 2 observations every 3 days for both ascending and descending passes). In the EnKF, the measurement errors are represented using another ensemble with the mean equal to zero and the variance equal to the observation error variance. In this study, a uniform error standard deviation of 0.08 m^3/m^3 (derived from the SMOS climatology) is assumed for the SMOS retrievals.

3.4.1. Bias detection and reduction

If we directly assimilate the unscaled SMOS soil moisture product, the analysis (updating the model forecast with a SMOS observation) typically makes systematic corrections to the model estimate. Negative mean increments (change in the model estimate between after and before the updating) are pronounced across the study region for both the surface layer and the root zone (not shown here). This provides clear evidence of the presence of bias in the system. If the system is bias-free (i.e., no systematic errors in either the model or the SMOS observation), mean analysis increments should be close to zero. This bias problem was also indicated by non-zero mean innovations and non-zero difference between climatology of satellite retrievals and that of their model equivalents.

Data assimilation systems are usually designed to produce an optimal estimate based upon the hypothesis of unbiased (and uncorrelated) errors in model and observation (i.e., a bias-blind system). In practice, biases in model forecast or observation (including observation operator) would contribute to the error variances, resulting in a suboptimal analysis. Observation biases, if present and known, should be removed prior to the assimilation. Provided that we can attribute the systematic errors to proper sources, and they also can be represented, by design, using appropriate parameters, the biases can be estimated jointly with the model state by adding the designed parameters to the state vector (i.e., a bias-aware system). However, this is extremely complicated to achieve considering limited reference data and thus beyond the scope of this work.

Following previous studies (e.g., Reichle and Koster, 2004; Reichle et al., 2007; Liu et al., 2011; Draper et al., 2012) we utilize a bias reduction scheme that matches the cumulative distribution function (CDF) of SMOS retrievals to the MESH model surface soil moisture's CDF by scaling the retrievals. The CDF matching scheme can effectively remove the climatological difference (mean and standard deviation) between satellite retrievals and model data, with little impact on the SMOS soil moisture skill. The skill for the rescaled SMOS retrievals is almost identical to the skill of unscaled SMOS (Fig. 3.3). However, notice that since the absolute magnitude of SMOS soil moisture is changed the assimilation products are meaningful only in terms of the time variability of soil moisture, which is consistent with the advantage of point measurements (section 3.3). In the present study, the model CDF is based on the 4-year (2010-2013) model surface soil moisture, while the SMOS soil moisture CDF (and the scaling of SMOS) is calculated separately for

2010/2011 and 2012/2013 since there are non-negligible inconsistencies in SMOS retrievals between the two periods (due to the change of the dielectric constant model in the retrieval algorithms). Correspondingly, the SMOS observation error standard deviation (0.08 m^3/m^3) is rescaled by multiplying it with the ratio between the scaled SMOS time series standard deviation (very close to the model soil moisture standard deviation) and the unscaled SMOS time series standard deviation. The rescaling of the SMOS retrievals and their error standard deviations is conducted locally. In addition, we also matched the satellite and model CDFs separately for the two model periods (2010-2011 and 2012-2013) and independently for each season. Results indicated that the rescaling parameters depended only weakly upon the model period and the season for this study.

3.4.2. Skill improvement over open-loop

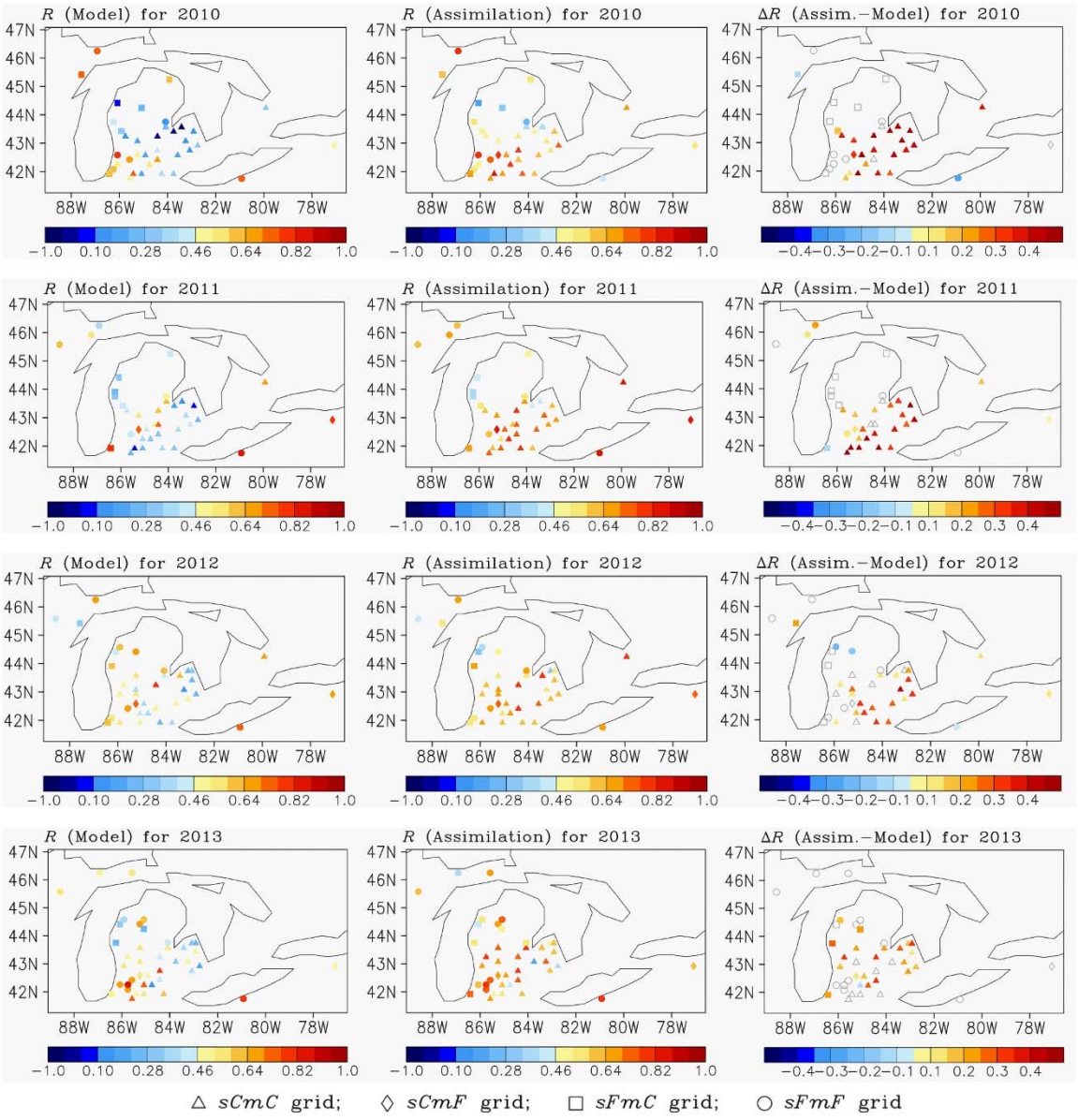

Fig. 3.4 Skill for surface soil moisture (0-10 cm) from (left) the open-loop model and (middle) the assimilation, and (right) the skill improvement $\Delta R^{\text{A-M}}$ (Assimilation minus Open-loop) over four individual years (top to bottom: 2010, 2011, 2012, and 2013). In the right column, $\Delta R^{\text{A-M}}$ is denoted by an open symbol in grey if the open-loop R and the assimilation R are not significantly (5% level) different from each other. Symbols denote the model grid types, same as in Fig. 3.3.

Fig. 3.4 compares the surface soil moisture skills from the open-loop model (single integration without assimilation) and the assimilation estimates based upon the scaled SMOS retrievals. Here the surface soil moisture skill refers to the correlation *R* (daily time series) between the grid-averaged soil moisture from the model surface layer (0-10 cm) and in situ measurements taken at 10 cm depth or in the 0-30 cm profile (the probe is vertically installed for some sites). *R* values are not computed if the length of SMOS and/or in situ soil moisture time series is short or when the correlation is strongly affected by the in situ measurement noise or the trends (section 3.3). Consistent with the assessment of the SMOS skill, the model grids are categorized as the sCmC, sCmF, sFmC, and sFmF types (section 3.3). Table 3.1 summarizes the median and mean skill *R* within each grid type for each soil moisture product.

To test the significance of the difference between skills for the three soil moisture products (SMOS, the open-loop, and the assimilation), the Fisher *Z* transform method is used. Assuming that two correlations R_1 and R_2 are independent, the *Z*-score for the difference between the two correlations can be expressed as (Dunn & Clark, 1969; Meng et al., 1992)

$$z = \frac{\left(0.5\ ln\frac{1+R_1}{1-R_1}\right)-\left(0.5\ ln\frac{1+R_2}{1-R_2}\right)}{\sqrt{\frac{1}{N_1-3}+\frac{1}{N_2-3}}} \tag{3.3}$$

where N_1 and N_2 are the sample sizes for R_1 and R_2. Given a significance level, the two correlations are statistically different from each other if the absolute *Z*-score exceeds the corresponding critical value. In practice, the assumption that the correlations (skills) are independent is not strictly valid for the three soil moisture products. To this end, the

significance was estimated using a Monte Carlo approach for a limited number of grids (due to computational burden). This preliminary test confirmed the results assuming independence very closely approximate the Monte Carlo-based results. Thus, all statistical tests for the skill difference reported in the work utilize the independence assumption and are not Monte Carlo based.

The open-loop model (Fig. 3.4, left column) typically provides higher surface soil moisture skill R at the sFmF and sCmF grids (median/mean of about 0.61), which are covered dominantly by forest, than at the sCmC and sFmC grids (median/mean of about 0.40) that are dominated by crop cover. Through the assimilation, the four grid types experience different skill gains $\Delta R^{\text{A-M}}$, defined as the skill for the assimilation soil moisture product minus the skill for the open-loop estimates (Fig. 3.4, right). Overall the sCmC grids (triangles) have the largest improvement $\Delta R^{\text{A-M}}$, and the sFmF grids (circles) show the weakest or even negative $\Delta R^{\text{A-M}}$; while soil moisture from the sCmF and sFmC grids (diamond and square signs) typically shows low to modest increase in skill. The skill gain $\Delta R^{\text{A-M}}$ is typically statistically significant for the sCmC grids. After the assimilation (Fig. 3.4, middle), the surface soil moisture skill R for the sCmC grids (median/mean of about 0.64) are typically closer to or even larger than R for the forest-dominated grids (sCmF and sFmF). Similarly, Draper et al. (2012) revealed larger skill (anomaly R) improvements for the cropland than for the mixed cover class (10-60% trees or woody plants) when assimilating the AMSR-E and ASCAT retrievals in the Catchment Land Surface Model (CLSM).

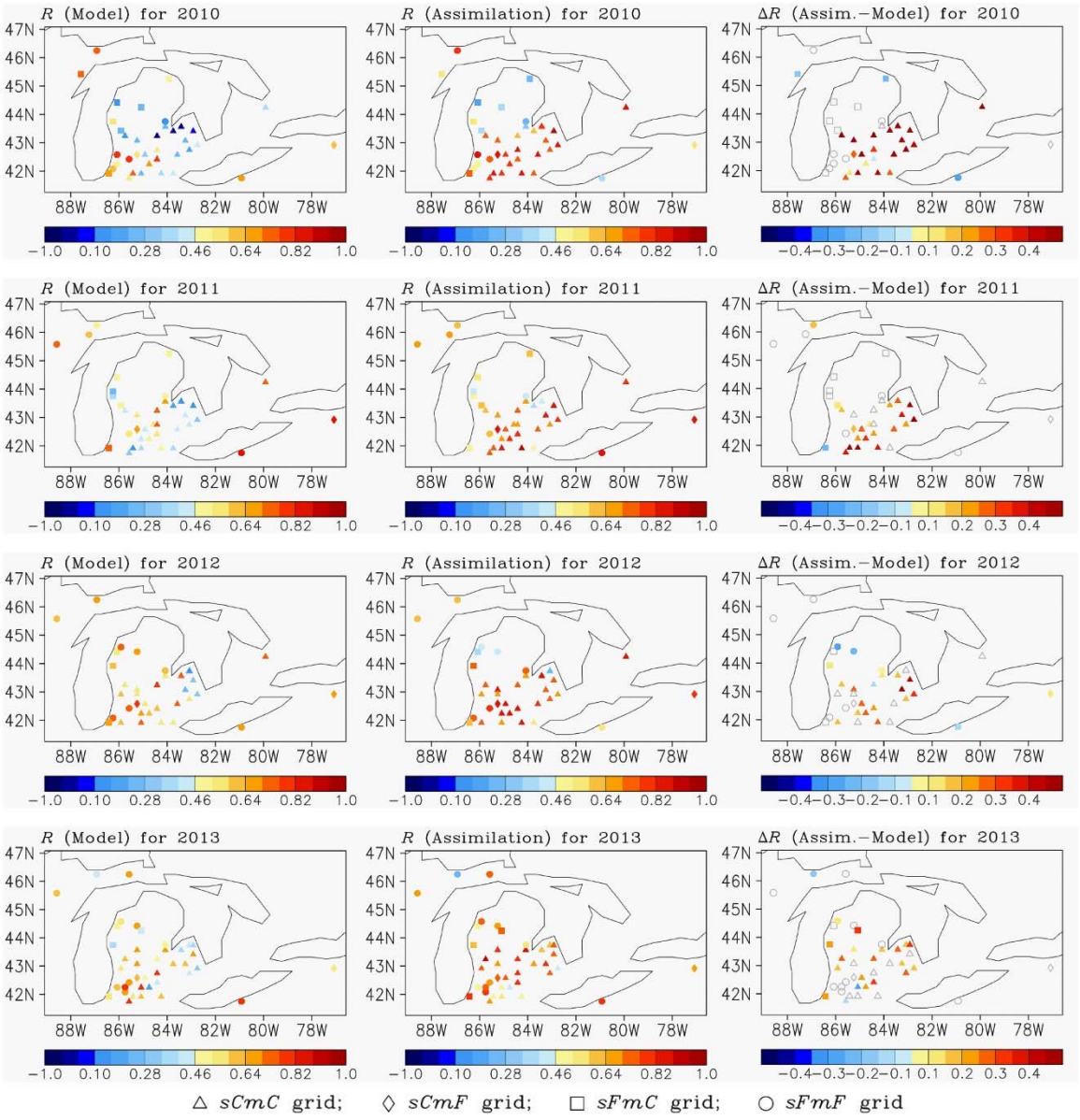

Fig. 3.5 Similar to Fig. 3.4, but for root-zone (top 35 cm) soil moisture.

The counterpart of Fig. 3.4 for root-zone soil moisture (0-35 cm) is provided in Fig.3.5. The root-zone soil moisture skill is derived using a depth-weighted average of soil moisture estimates in the model's top two layers (0-10 and 10-35 cm) against the arithmetic mean of in situ measurements at 10 and 25 cm depths or the 0-30 cm profiles measured by vertically installed sensors. The variations with the grid types of the open-loop skill and

the skill gain ΔR^{A-M} for root-zone soil moisture are quite similar to those observed for the surface soil moisture. Overall the open-loop skill for root-zone soil moisture (Fig. 3.5, left column) is higher at forest-dominated grids (sFmF and sCmF) than at crop cover-dominated grids (sCmC and sFmC). The strongest skill improvement ΔR^{A-M} for root-zone soil moisture are also observed for the sCmC grids (triangles in Fig.3.5, right). This clearly indicates that the surface soil moisture information measured by SMOS, through the EnKF assimilation, can be propagated to the soil layers that are not directly measured. For a given grid type, on average, the skill for root-zone soil moisture is slightly higher than the surface soil moisture skill (for either the open-loop or the assimilation product) (Table 3.1).

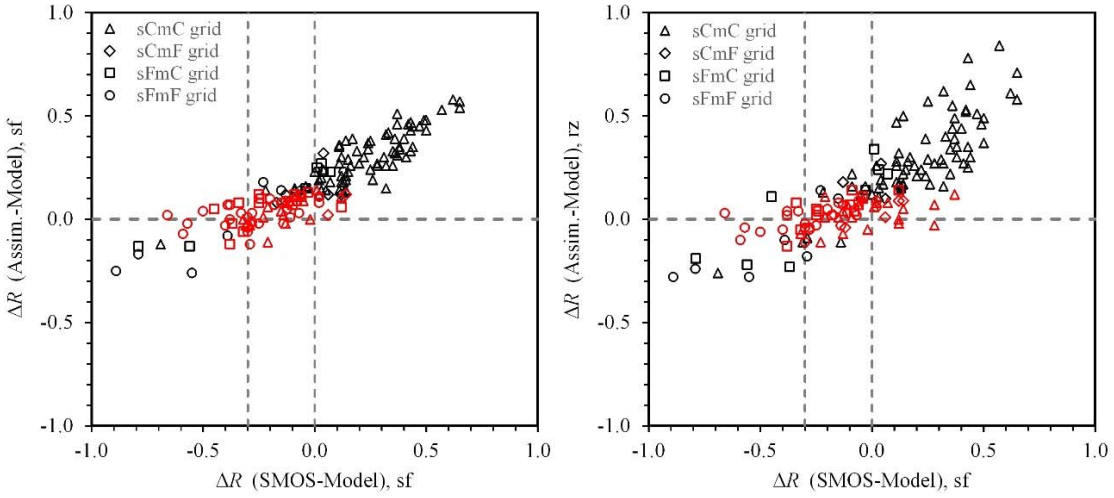

Fig. 3.6 Skill improvement ΔR^{A-M} (skill for the assimilation minus the open-loop skill, ordinate) for (left) surface and (right) root-zone soil moisture against ΔR^{S-M} (skill for the SMOS observation minus skill for the open-loop surface soil moisture, abscissa). Symbols indicate the model grid types as defined in the text: (triangles) sCmC, (diamonds) sCmF, (squares) sFmC, and (circles) sFmF. Symbols in red mean that ΔR^{A-M} are not statistically significant at the 5% level. The horizontal dashed line denotes $\Delta R^{A-M} = 0$. The two vertical dashed lines denote $\Delta R^{S-M} = -0.3$ and $\Delta R^{S-M} = 0$, respectively.

The skill improvement $\Delta R^{A\text{-}M}$ is controlled not only by the satellite observation skill but also by the skill for the open-loop estimates. In general, the skill improvement $\Delta R^{A\text{-}M}$ increases as the satellite observation skill, but decreases with increased open-loop skill (Reichle et al., 2008a). Therefore, when the satellite observation skill is high and the model (open-loop) skill is low, the largest skill improvement $\Delta R^{A\text{-}M}$ is expected, which typically corresponds to the sCmC case. On the contrary, if the satellite observation skill is low and the open-loop model skill is high, we usually expect weak $\Delta R^{A\text{-}M}$, as observed for the sFmF grids. When the satellite skill and the open-loop skill are either both high (e.g. sCmF grids) or both low (e.g. sFmC grids), $\Delta R^{A\text{-}M}$ are typically low to modest.

The skill improvement $\Delta R^{A\text{-}M}$ (the assimilation skill minus the open-loop skill) against $\Delta R^{S\text{-}M}$, defined as the SMOS observation skill minus the skill for the open-loop surface soil moisture, is provided in Fig. 3.6. Overall the skill improvement $\Delta R^{A\text{-}M}$ for both surface and root-zone soil moisture (the ordinate) is strongly related to $\Delta R^{S\text{-}M}$ (the abscissa). Every time the SMOS skill is greater than or equal to the open-loop surface soil moisture skill, the assimilation is typically able to significantly improve the skill of the model estimates. Such is the case with most of the sCmC grids (triangles). When the satellite observation skill is about 0~0.3 lower than the open-loop model (i.e., $\Delta R^{S\text{-}M}$ along the abscissa is between -0.3 and 0), the open-loop skill was still improved by the assimilation for most cases (85% for surface soil moisture and 80% for root-zone soil moisture), but the improvements are not always statistically significant. If the skill for SMOS retrievals is more than about 0.3 below the open-loop skill (i.e., $\Delta R^{S\text{-}M}$ is less than -0.3), the assimilation is not helpful and even

negatively affects the open-loop skill. The results are fairly consistent with Draper et al. (2012). The study showed that the assimilation of AMSR-E and ASCAT retrievals in CLSM typically generated an improved skill (in terms of anomaly R) for both the surface and root zone soil moisture as long as the satellite observation skill is no more than about 0.2 lower than the open-loop skill.

For the retrievals with very low or even negative skill ($\Delta R^{S\text{-}M}$ is thus small in Fig.3.6), which generally reflect poor satellite observations, their real errors could be severely underestimated by the input error parameters, thus causing negative $\Delta R^{A\text{-}M}$. Overall, negative $\Delta R^{A\text{-}M}$ is more severe in root zone than for the surface layer (Fig. 3.6). This is generally consistent with the finding that poorly specified observation errors have a fiercer impact on the assimilation estimates of root zone soil moisture than on surface soil moisture estimates (Reichle et al., 2008b). The on-line quality control routines (e.g., Reichle, 2008) and on-line tuning of error covariances (Reichle et al., 2008b) may be helpful for controlling the occurrence of negative $\Delta R^{A\text{-}M}$. Note that although the assimilation skill does not necessarily exceed the skill of the open-loop model for individual grids, the assimilation product always outperforms or at least match the open-loop counterpart in terms of the averaged skill for each grid type (Table 3.1), coinciding with the finding based on synthetic assimilation experiments (Reichle et al., 2008a). Additionally, as shown in Fig. 3.6, overall the surface soil moisture $\Delta R^{A\text{-}M}$, relative to root-zone $\Delta R^{A\text{-}M}$, exhibits a better linear relationship with $\Delta R^{S\text{-}M}$. For a given $\Delta R^{S\text{-}M}$, the skill improvement $\Delta R^{A\text{-}M}$ is usually more variable (along the ordinate) for root-zone soil moisture than for surface soil moisture. This may be due to the fact that during the assimilation the updating of root-zone soil moisture

is subject to the accurate information exchanges between the surface soil and the deeper layers, which, in turn, are controlled by many factors (e.g. the model dynamics and the input error parameters). However, notice that a perfect linear relationship between $\Delta R^{A\text{-}M}$ and $\Delta R^{S\text{-}M}$ is not expected since the sensitivity of $\Delta R^{A\text{-}M}$ to $\Delta R^{S\text{-}M}$ is additionally affected by the magnitude of open loop skill.

3.4.3. Skill improvement over SMOS

In theory, the assimilation seeks to produce superior estimates, relative to both the open-loop model and the observation product alone. In this section, we investigate the skill improvement, relative to the SMOS observation skill, by the assimilation. Fig. 3.7 shows $\Delta R^{A\text{-}S}$, defined as the skill for the surface soil moisture assimilation product minus the SMOS observation skill. It is expected that $\Delta R^{A\text{-}S}$, as opposed to $\Delta R^{A\text{-}M}$, increases as the open-loop skill (since the assimilation product skill typically increases with the open-loop skill for the same observation skill), but decreases with increased satellite observation skill. As expected, overall the variation of $\Delta R^{A\text{-}S}$ with the grid type (Fig. 3.7) is opposite to that for $\Delta R^{A\text{-}M}$ (Fig. 3.4, right column). At the sFmF and sFmC grids (circles and squares in Fig. 3.7), the surface soil moisture skill for the assimilation typically significantly exceeds the skill of SMOS product alone (but the corresponding $\Delta R^{A\text{-}M}$ is typically small or even negative, as discussed above). This is mainly because that for the two grid types the open-loop skill is typically much higher than the satellite skill (e.g. Table 3.1). In contrast, smaller $\Delta R^{A\text{-}S}$ are usually observed for the sCmC grids (triangles in Fig. 3.7; the corresponding $\Delta R^{A\text{-}M}$ is typically the strongest).

The SMOS observation skill could even exceed the assimilation skill at a few of the sCmC grids (Fig. 3.7). Reichle et al. (2008a), based upon synthetic experiments (Fig. 2a therein), also found that the surface soil moisture skill from the assimilation was not always above the satellite observation skill (anomaly R was used therein), especially in the presence of a poor open-loop model skill and a high satellite skill (such is the case with our sCmC grids showing negative ΔR^{A-S}). As they pointed out, the reasons for the occurrence of negative ΔR^{A-S} may include the effects from the nonlinearity of the system, a small ensemble size, and the imperfect input error parameters, etc. However, note that overall the surface soil moisture assimilation skill (median/mean of 0.64) is still significantly better than the SMOS product skill (median/mean of 0.55) for the sCmC-type grid (Table 3.1).

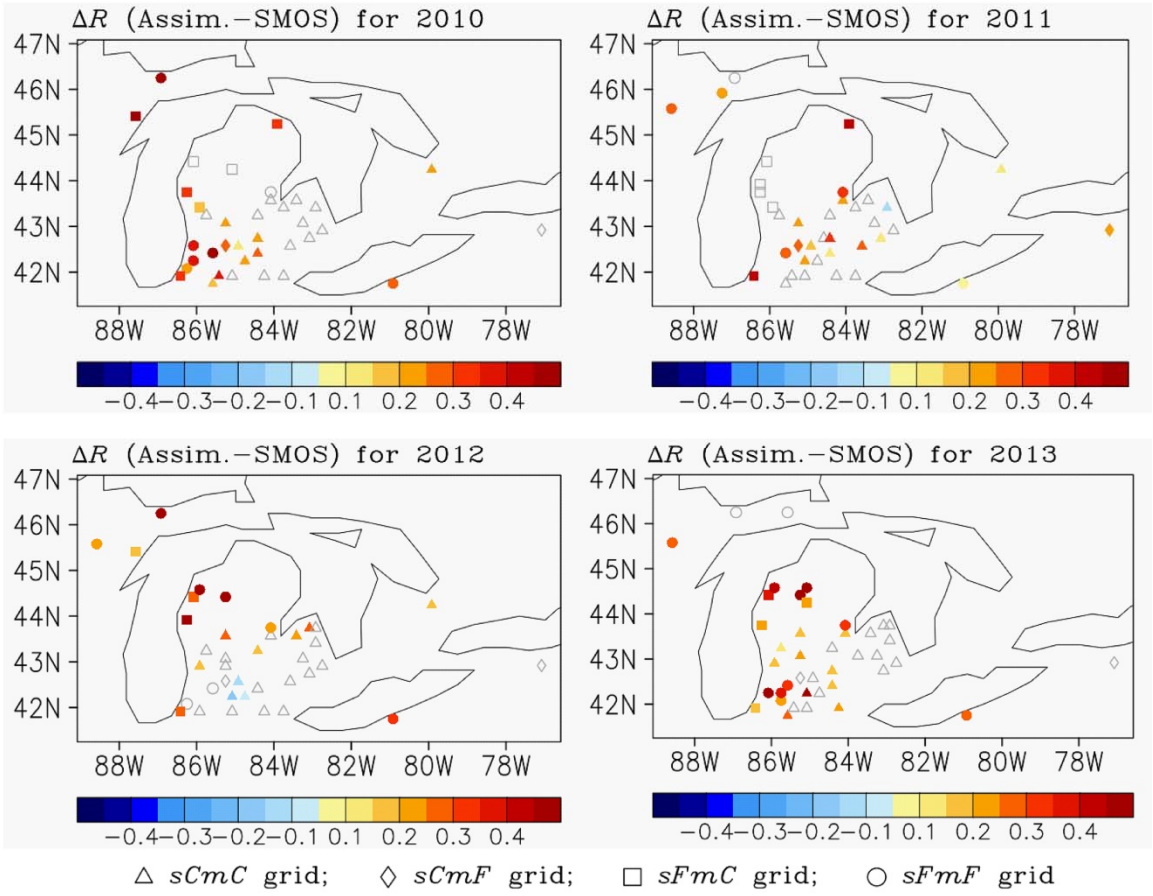

Fig. 3.7 Skill improvement $\Delta R^{A\text{-}S}$, defined as the skill for the surface soil moisture assimilation product minus the SMOS observation skill. $\Delta R^{A\text{-}S}$ in grey open symbol means that the assimilation skill and the SMOS skill are not significantly (5% level) different from each other. Symbols denote the model grid types, same as in Fig. 3.4.

3.4.4. Subgrid-scale (GRU) soil moisture skill

In the above, point in situ measurements are used to assess the skill for the grid-scale soil moisture. It is acknowledged that there could be a mismatch in vegetation or soil characteristics between the two products with different spatial scales. A model grid square typically represents a mixture of multiple land cover and soil attributes, while a point

station corresponds to only a specific vegetation and/or soil type. In this study, however, this factor is expected to have negligible effects on the skill evaluation above since the land cover type for in situ station is typically consistent with the dominant land cover class for the grid-scale soil moisture.

We also computed the subgrid-scale soil moisture skill, i.e., point measurements are compared with the model soil moisture from a subgrid area that has the same vegetation or soil characteristics as the point site. In the MESH model, the subgrid-scale variability is resolved using the GRU approach (section 3.2.2). Each model grid cell is a mosaic of up to seven GRUs. Each GRU corresponds to one land cover class (other soil characteristics are assumed to be the same for the same GRU type) and is weighted by the fraction of the land cover class within the grid cell. Hence, for a given grid location, the soil moisture skill for a specific GRU, which corresponds to the land cover class for the in situ station, is assessed. Overall the subgrid-scale (GRU) soil moisture (not shown) and the grid-averaged soil moisture reveal a consistent vegetation modulation of skill for both the open-loop and the assimilation. The open-loop model usually provides strong soil moisture skill for forest GRUs and weaker skill for crop GRUs. A crop GRU, if the SMOS soil moisture sampled from a crop surface node is assimilated, typically experiences a large skill improvement $\Delta R^{\text{A-M}}$. When the assimilated SMOS retrievals come from a forest-type surface, the skill improvement $\Delta R^{\text{A-M}}$ for the crop GRU soil moisture is relatively weak. The assimilation typically leads to smaller or even negative $\Delta R^{\text{A-M}}$ for forest-GRUs, even when the assimilated SMOS soil moisture is from a crop surface node. To further improve

the assessment of the soil moisture skill, dense in situ observations would clearly be of advantage, although such data are not available for this study.

3.5 Summary and Discussion

Since the launch of SMOS satellite mission, the validation and assimilation of SMOS soil moisture has been an active research area. In this chapter, the 1D-EnKF is used to assimilate SMOS soil moisture retrievals into the MESH model over the Great Lakes basin. The satellite retrievals, the open-loop soil moisture, and the assimilation estimates are validated against point-scale in situ soil moisture measurements from MAWN, SCAN and FCRN, in terms of the daily time series correlation coefficient (soil moisture skill R). Due to the bias between the SMOS retrievals and the model soil moisture estimates, a priori rescaling on the retrievals is performed using the CDF matching. Our focus in this work is thus on the assimilation of the scaled SMOS retrievals. The main results from this study are as follows.

(1) The observation skill is typically low for the SMOS retrievals from forest surfaces, but becomes high for those from crop surfaces, consistent with the effect of canopy density on the satellite retrieval capabilities. On the other hand, the open-loop model typically provides higher soil moisture skill R over forests than over crops.

(2) Overall the assimilation can favorably influence the model soil moisture skill for both the surface layer and the root zone except for the cases with a small SMOS observation skill and a large open-loop skill. The skill improvement $\Delta R^{A\text{-}M}$, defined as the skill for the assimilation soil moisture product minus the skill for the open-loop estimates, for

both surface and root-zone soil moisture typically increases as the SMOS observation skill and decreases with increased open-loop skill, showing a strong dependence upon $\Delta R^{S\text{-}M}$, defined as the SMOS observation skill minus the open-loop surface soil moisture skill. When the SMOS skill is greater than or equal to the open-loop surface soil moisture skill, the assimilation is typically able to significantly increase the open-loop soil moisture skill.

(3) The crop-dominated grids typically experience the largest $\Delta R^{A\text{-}M}$ if the assimilated SMOS retrievals also come from crop surfaces, consistent with a high satellite observation skill and a low open-loop skill, while $\Delta R^{A\text{-}M}$ is usually the weakest for the forest-dominated grids when the SMOS retrievals from forested surfaces are assimilated, due to a low observation skill and a high open-loop skill.

(4) On average, the skill for the surface soil moisture assimilation product is significantly better than the skill for the SMOS product alone, although the dependence of $\Delta R^{A\text{-}S}$ (skill for the surface soil moisture assimilation product minus the SMOS observation skill) upon the open-loop skill and the satellite observation skill is opposite to that for $\Delta R^{A\text{-}M}$. The forest-dominated grids, if the assimilated SMOS retrievals also come from forest surfaces, typically have large $\Delta R^{A\text{-}S}$ because the corresponding open-loop skill is generally higher than the satellite skill. In contrast, smaller $\Delta R^{A\text{-}S}$ are typically observed when the assimilated SMOS retrievals are from cropped surfaces since the corresponding SMOS observation skill is high.

(5) We also investigated the subgrid-scale (GRU) soil moisture skill by comparing point measurements with the GRU soil moisture (a GRU and an in situ site lie within the same grid cell and have the same land cover class). Overall the GRU soil moisture skill and the grid-scale soil moisture skill show a consistent vegetation modulation for both the open-loop and assimilation estimates. This confirms a negligible impact of point measurements (in situ data) on the skill assessment for the grid-scale soil moisture (model and SMOS) due to the possible disparity in vegetation characteristics between them.

Unlike previous assimilation studies of SMOS soil moisture (e.g. Zhao et al., 2014; Ridler et al., 2014), this work assimilated 4 years of SMOS retrievals (2010-2013) at a grid scale of ~15 km. The overall agreement within the same grid type and the overall consistency between the years are observed for each of the three soil moisture products (SMOS, the open-loop, and the assimilation), which demonstrates the robustness of our results. This study also suggests that the ability of SMOS/MIRAS to measure surface soil moisture for a wide range of vegetation covers is clearly of advantage for assessing the vegetation modulation of the assimilation. The results offer further insight into the dependence of the assimilation upon the open-loop skill and the satellite observation skill.

In this work, only the correlation *R* metric of skill is used to assess the three data sets (SMOS alone, the open-loop model, and the assimilation estimates) because (1) the temporal variability of soil moisture (rather than the absolute magnitude) observed by point measurements is spatially representative; and (2) the absolute magnitude of the soil

moisture assimilation product is meaningless since the satellite retrievals are rescaled prior to the assimilation (Reichle et al., 2007). Note that through a percentile-based transformation (e.g., Entekhabi et al. 2010b) the time variations of soil moisture can be scaled to the soil moisture initial conditions of weather and climate models, while any bias (systematic error) in the soil moisture product can be scaled out (e.g. Zhang & Frederiksen, 2003). Therefore, the resulting soil moisture assimilation product can benefit weather and climate forecast initializations as long as the time variability of soil moisture is captured accurately. The skill *R* values presented in this work are derived based upon the raw soil moisture time series that include the seasonal cycle. To assess the impact of soil moisture seasonality on the skill *R* evaluation, we also analyzed the anomaly *R*. The soil moisture anomalies are defined as departures of daily soil moisture from the seasonal (monthly mean) climatology (e.g., Reichle et al., 2007). Note that the anomaly *R* computation is restricted to only a portion of the validation grids presented above since (i) at least three years of complete estimates, for each soil moisture product, are required for extracting the soil moisture seasonal climatology, and (ii) for more sites (relative to the raw time series) the variability of the anomaly time series is lower than the measurement noise level, which results in meaningless correlations. Overall, the raw *R* metric and the anomaly *R* results (not shown here) are qualitatively similar and consistently lead to the general conclusions presented above.

In the present work, overall the open loop soil moisture skill for 2010/2011 is lower than that for 2012/2013 (Figs. 3.4 and 3.5). The difference may be caused by two sources: (i) the meteorological forcing data (notably rainfall) used for 2010/2011 may be of relatively

low quality; and (ii) the model parameters (related to physiography, vegetation, and soil characteristics), which were based upon a calibration with the 2004-2005 streamflow observations (Haghnegahdar et al., 2014), may be not the "best" for 2010/2011. If the improved forcing data and/or calibrated model parameters are applied, the 2010/2011 open-loop skill could be increased and the corresponding skill improvement through the assimilation is expected to decrease (as shown for 2012/2013). However, our general conclusions remain valid.

CHAPTER 4

Assimilation of AMSR-E Soil Moisture in the MESH Model

4.1 Introduction

Soil moisture information is critically important to the monitoring and modelling of climate and global changes since the wetness of the land strongly affects the energy, water and biogeochemical cycles of the climate system (e.g. Seneviratne et al. 2010). In practice, spatially distributed soil moisture information, especially at regional, continental, and global scales, is difficult to estimate using ground-based (in situ) measurements since in situ stations routinely acquiring soil moisture are sparse globally. Moreover, in situ soil moisture measurements are typically point-scale observations and thus have difficulty representing spatial averages, especially in terms of the absolute magnitude of soil moisture. Space-based remote sensing, especially microwave measurements that are sensitive to changes in soil water content, holds the ability to measure large-scale spatially distributed surface soil moisture. So far, spaceborne microwave soil moisture observations typically come from passive sensors such as the Special Sensor Microwave Imager (e.g. Jackson, 1997), the Scanning Multichannel Microwave Radiometer (e.g. Reichle and Koster, 2005), the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (e.g. Bindlish et al., 2003), the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) (e.g. Njoku et al., 2003), and the Advanced Microwave Scanning Radiometer 2 (AMSR2), as well as active systems such as the European Space Agency Remote Sensing Satellite (ERS) Synthetic Aperture Radar (SAR), the Canadian

RADARSAT series (e.g. Merzouki et al., 2011), the Scatterometer (SCAT), and the Advanced Scatterometer (ASCAT) (e.g., Bartalis et al., 2007; Albergel et al., 2009). The Microwave Imaging Radiometer with Aperture Synthesis (MIRAS) onboard the Soil Moisture and Ocean Salinity (SMOS) satellite (Kerr et al., 2012) and the newly launched Soil Moisture Active Passive (SMAP) mission, which were designed exclusively for soil moisture measurement, hold enhanced capabilities to estimate surface soil moisture. Spaceborne microwave sensors measure only soil moisture within a near-surface layer. The soil thickness measured increases with the wavelength (several tenths of the wavelength). For bare soil, the penetration depth is about 3-5 cm for L-band (1-2 GHz) measurements (e.g. SMOS), and only ~1-1.5 cm for C- (4-8 GHz) or X- (8-12 GHz) band measurements (e.g. AMSR-E, ASCAT). Where there is a vegetation cover, the radiation emitted or backscattered from the soil would be attenuated owing to the scattering and absorption by vegetation canopy. The magnitude of the vegetation attenuation increases with the sensor's operating frequency and the vegetation density. Hence soil moisture retrievals at high microwave frequencies (> 5-6 GHz) are typically valid only for bare soil and lightly or moderately vegetated regions.

Over the past decades, satellite microwave soil moisture retrievals, which are, relative to point measurements, more compatible in space with distributed numerical models, have shown great potential to improve the predictive skill of land surface and hydrologic models, especially through advanced data assimilation (e.g. Xu et al., 2014). In practice, land surface and hydrological modelling is typically subject to uncertainties in forcing fields and deficiencies in model physical structure and parameters, while satellite microwave soil

moisture retrievals are only available for the surface layer and typically contain instrumental and representativeness errors. Advanced data assimilation offers a framework to merge model forecast and satellite retrievals based upon some estimate of their error characteristics. The merging is achieved, most commonly, either by maximum-likelihood estimators (e.g. variational methods) or by variance minimizing estimators (e.g. the Kalman filter, KF; the extended Kalman filter, EKF; and the ensemble Kalman filter, EnKF). The former assumes that the error statistics of the background (a priori state) and the observation are known and seeks the state with the maximum likelihood by minimizing a cost function that measures the distance of the model state (unknown) to the observations and to the background. In contrast, a variance minimizing estimator directly derives the analysis state based upon an analysis equation, which can ensure that the total analysis error variances are minimum over the whole space and time domain. Through data assimilation near-surface soil moisture information derived from satellite microwave measurements can spread to deeper soil layers that cannot be directly measured by satellite microwave remote sensing. Furthermore, in a data assimilation system, soil moisture retrievals from different satellite platforms can, by means of consistency constraints in time evolution and physical properties, be interpolated and extrapolated to yield a single optimal soil moisture estimation.

Ten years (2002-2011) of operations of the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) had acquired important near-surface soil moisture data for land and hydrologic data assimilation. Reichle et al. (2007) assimilated the Level-2B AMSR-E soil moisture product, based upon the NASA standard algorithm,

into the NASA Catchment Land Surface Model (CLSM) using the EnKF method. The assimilation led to an overall improvement relative to either the model estimates or satellite retrievals alone, in terms of soil moisture anomaly time series correlation with in situ measurements. In Draper et al. (2009a), the Extended Kalman Filter (EKF) method was utilized to assimilate the surface soil moisture derived from AMSR-E C-band brightness temperature measurements with the Land Parameter Retrieval Model (LPRM) into the Interactions between Surface, Biosphere, and Atmosphere (ISBA) land model. The introduction of AMSR-E soil moisture did yield substantial analysis increments (changes in the model estimate between before and after the implementation of the analysis equation) for both surface and root-zone soil moisture, although the assimilation estimates were not validated against real in situ observations. Liu et al. (2011) showed that the assimilation of AMSR-E soil moisture was as efficient as the precipitation corrections for enhancing the model skill for soil moisture estimation (anomaly time series correlation coefficient with in situ measurements). The study assessed the contributions of two AMSR-E soil moisture products (June 2002 to July 2009), the NASA standard algorithm product archived at the National Snow and Ice Data Center (NSIDC) and the LPRM-derived AMSR-E soil moisture. The assimilation of LPRM product led to larger soil moisture skill improvement than the NSIDC product. More recently, Draper et al. (2012) suggested that the CLSM soil moisture skill could be improved through the assimilation of either AMSR-E or ASCAT soil moisture products. A joint assimilation of the two sensor products could produce the best soil moisture skill. Due to the bias (systematic error) between satellite retrievals and model soil moisture estimates, a priori rescaling on AMSR-E retrievals (the cumulative distribution function (CDF) matching) was applied for the aforementioned efforts. Li et al.

(2012) assimilated AMSR-E soil moisture retrievals (derived from the X-band brightness temperatures using single-channel algorithm), without a priori scaling, into the Noah land surface model. Their work was motivated by the assumption that the mean value of satellite retrievals have the potential to improve the model mean values of soil moisture. Although the study observed the improved soil moisture estimates (as indicated by reduced bias and root-mean-square-error against in situ measurements), especially for the mass conservation scheme, the analysis typically made systematic corrections to the model soil moisture estimation (a clear symptom of bias in the assimilation). This means that a satellite-model bias removal is an indispensable part in a bias-blind assimilation system (i.e. correcting random errors only).

In this chapter, the 1D-EnKF is applied to assimilating AMSR-E soil moisture retrievals, based upon the NASA standard and LPRM algorithms, respectively, into the standalone version of MESH in which CLASS, the Canadian Land Surface Scheme, is coupled with a distributed hydrological model. The assimilation estimates are validated over the Great Lakes basin using in situ soil moisture measurements from the Michigan Automated Weather Network. Our goal is to investigate how the assimilation of AMSR-E soil moisture affects the MESH model soil moisture estimates and how well our results are compared with the AMSR-E assimilation in other models and for other regions. Relative to AMSR-E, newer satellite instruments operating at L-band and designed exclusively for soil moisture measurement, such as SMOS and SMAP, are expected to hold an enhanced capability to estimate surface soil moisture due to their stronger penetration of vegetation and soils. The assimilation of AMSR-E could be used a benchmark for evaluating the

assimilation of the SMOS and SMAP soil moisture, in particular for assessing the impact of the retrieval skill on the assimilation of remotely-sensed soil moisture. The present study differs from other AMSR-E assimilation studies in two aspects: (1) this work is the first time to assimilate AMSR-E soil moisture in the MESH model and for the Great Lakes basin; and (2) the AMSR-E assimilation is compared with the assimilation of SMOS retrievals, providing further insight into the dependence of the assimilation upon the satellite retrieval skill.

4.2 Data and Methods

4.2.1 Standalone MESH model

Environment Canada's standalone MESH is a coupled land-surface and hydrological model (Pietroniro et al., 2007). The model involves both the vertical water transfers and the horizontal flows (overland flow, interflow, base flow). The soil column is partitioned into three or more layers to resolve water and temperature dynamics. The vertical water movement between the soil layers is governed by Richard's equation. The model uses a Grouped Response Unit (GRU) approach to resolve the subgrid-scale variability. In the version of MESH used in this study, each model grid cell is divided into a limited number of distinct GRUs based on the distribution of land cover classes within the cell. Within a grid cell, the fluxes and variables are computed independently for GRUs, ignoring the interactions between GRUs. The overall fluxes and prognostic variables of a grid cell are

obtained by taking a weighted average of those from GRUs. The lateral movement of soil water between grid cells is not taken into account.

The study domain for this work is the Great Lakes basin (Fig. 3.1). The model configurations are similar to those used in Pietroniro et al. (2007) and Haghnegahdar et al. (2014). Seven GRU types based upon land cover class are used here, including crop, grass, deciduous forest, coniferous forest, mixed forest, water, and impervious. Each grid cell is a mosaic of the seven GRUs, weighted by their cell fractions. In this work, the model parameter sets for all GRU types are directly taken from a global calibration strategy using streamflow observations (Haghnegahdar et al., 2014). The forecast model is forced using the gridded hourly precipitation data derived from the Canadian Precipitation Analysis (CaPA; Mahfouf et al., 2007); other meteorological forcing data (incoming shortwave and longwave radiations, surface air temperature, wind speed, pressure, and specific humidity) come from the Global Environmental Multiscale (GEM) model forecasts (Mailhot et al., 2006). A three-layer soil structure is used: 0-10, 10-35, and 35-410 cm. The simulations are performed at a resolution of 1/6th of a degree (~15 km) with a time step of 30 minutes.

4.2.2 AMSR-E soil moisture products and data assimilation scheme

A number of algorithms had been developed to extract soil moisture from AMSR-E measurements. In this work, we assimilate two AMSR-E soil moisture products: (i) the AMSR-E/Aqua Level-2B land surface product (Njoku 2004), based upon the NASA standard algorithm, archived at NSIDC (data version V09), and (ii) the LPRM algorithm-based AMSR-E Level 2 soil moisture product (Owe et al. 2008) archived at the NASA

Goddard Earth Sciences Data and Information Services Center (GES DISC). The two products have been widely used in various validation and assimilation studies. In the remainder of this thesis, the two AMSR-E products are referred to as NSIDC and LPRM products, respectively.

The NSIDC product is delivered at a 25 km Equal-Area Scalable Earth Grid (EASE-Grid) cell spacing. The soil moisture retrievals were derived from the X-band (10.7 GHz) brightness temperature measurements using the Polarization Ratios (PR) approach (modified version from Njoku et al., 2003; Njoku and Chan, 2006). The use of normalized PRs (brightness temperature difference between the vertical and horizontal polarizations at a given frequency normalized by their sum) can effectively remove the surface temperature dependence. PRs at 10.7 GHz and 18.7 GHz are used to derive the vegetation/roughness parameter based upon empirical relationships. Soil moisture is then estimated based upon departures of PR at 10.7 GHz from local monthly minima, which is used as a baseline value. Except for surface soil moisture and vegetation/roughness parameter, the NSIDC product also contains useful ancillary data, such as surface type and quality control flags. Utilizing the ancillary information, a filtering is performed to exclude the soil moisture retrievals that are contaminated by dense vegetation, open water, frozen surface, snow cover, radio-frequency interference (RFI), or rainfall, etc.

The LPRM algorithm uses a forward radiative transfer model to retrieve surface soil moisture and vegetation optical depth through a nonlinear iterative procedure (Owe et al., 2001; 2008). The LPRM product includes soil moisture retrievals and vegetation optical

depths derived from both the AMSR-E's X-band (10.7 GHz) and C-band (6.9 GHz) brightness temperature measurements and the land surface temperature that is separately derived from the vertical polarization brightness temperatures at 36.5 GHz. In order to reduce the RFI effects, here we use only the X-band LPRM retrievals. The retrievals are not used whenever (1) the land surface is frozen, (2) the vegetation optical depth exceeds 0.8, or (3) the uncertainty of soil moisture (simultaneously provided) is greater than 8%.

Temporal coverage for the AMSR-E soil moisture products is from 18 June 2002 through 4 October 2011, with a resolution of 1-2 days for both ascending (01:30 pm LST) and descending (01:30 am LST) orbits. Prior to the assimilation, the AMSR-E retrievals are resampled onto the MESH model grids (~15 km resolution) using a nearest neighbor approach. Whenever and wherever the model (combined with the rainfall forcing data) indicates the presence of precipitation, frozen soils, or snow cover, the satellite retrievals are also excluded from the evaluation and assimilation.

Here we use a 1D-EnKF (i.e. the analysis increment computation is performed independently for the model grids) with 12 ensemble members to assimilate soil moisture retrievals into the MESH model. The primary innovation of the EnKF is that a Monte Carlo approach is used to estimate model and measurement error statistics. The probability density of the model states is represented using an ensemble where the mean is the best estimate (Gaussian assumption), and the ensemble spread defines the error variance. The model error statistics evolve by integrating the ensemble of model states forward in time. The measurement errors are represented using another ensemble with the mean equal to

zero (Gaussian assumption) and the spreading of the ensemble consistent with the realistic or predefined observation error variance. The measurement errors are imposed onto the actual measurement to yield the ensemble of observations. At measurement times, a variance-minimizing analysis is applied to the ensemble of model forecast states, given by

$$x_j^+ = x_j^- + \sigma\left(X^-, \hat{Y}\right)[\sigma\left(\hat{Y}, \hat{Y}\right) + \sigma(\varepsilon, \varepsilon)]^{-1}\,[y_j - \hat{y}_j] \quad , j = 1, \ldots, N \tag{4.1}$$

where σ denotes the covariance between the two vectors, j is the ensemble member index, counting from 1 to the ensemble size N. x_j^- and x_j^+ denote the a priori and posterior model state estimates, respectively. y_j and $\hat{y}_j$ represent the perturbed observation and the corresponding model prediction. X^-, $\hat{Y}$ and ε denote the ensembles of $\{x_j^-\}$, $\{\hat{y}_j\}$, and observation errors, respectively. In the 1D-EnKF, x_j, for a given grid, is comprised of the volumetric liquid water content from all GRUs within the grid cell and all the three soil layers. The observation y_j is the perturbed satellite soil moisture and the corresponding model prediction $\hat{y}_j$ denotes the model estimates of the grid-averaged volumetric liquid water content (a weighted sum of GRU values) in the model surface layer (0-10 cm).

In the EnKF, the estimates of the model forecast errors are derived from an ensemble of model integrations. To represent random errors in the forcing inputs, cross-correlated forcing perturbation fields are generated following Reichle et al. (2007). To account for the model forecast errors due to deficiency in model physics and/or parameters, temporally correlated error perturbations are applied to the forecasted soil moisture (volumetric liquid water content). Currently, the 0.001 m^3/m^3, 0.0005 m^3/m^3, and 0.00005 m^3/m^3 error standard deviations are applied to the model's three layers (0-10, 10-35, and 35-410 cm),

respectively. The model error correlation time is set to 1 day. The observation errors are represented using another ensemble with the mean equal to zero and the variance equal to the observation error variance. In this study, the error standard deviations of 0.02 m^3/m^3 and 0.08 m^3/m^3 (derived from the climatology) are assumed for NSIDC product and LPRM product, respectively.

Since the satellite retrievals and model surface soil moisture exhibit different climatologies, which would impede an optimal merging of the two data sets, a priori rescaling (bias reduction) is applied to AMSR-E retrievals and the observation error standard deviations (stdev). The retrievals are rescaled by matching their cumulative distribution function (CDF) to the MESH surface soil moisture's CDF. The observation error stdev is rescaled by multiplying it with the ratio between the time series stdev of the scaled retrievals (almost identical to the model soil moisture stdev) and that of the unscaled retrievals. The rescaling of the retrievals and their error stdev is conducted independently for each grid cell. The CDF matching scheme can effectively remove the climatological difference (mean and standard deviation) between satellite retrievals and model data, with little impact on the retrieval skill (correlation with in situ measurements). However, note that since the absolute magnitude of satellite soil moisture is changed, the assimilation product has merit only in the time variability of soil moisture, which is consistent with the advantage of point measurements. The assimilation period is from 1 January 2003 through 04 October 2011.

4.2.3 In situ soil moisture observations and skill metric

In this work, in situ soil moisture measurements from the Michigan Automated Weather Network (MAWN; http://www.agweather.geo.msu.edu/mawn/) are used to validate the satellite retrievals, the model and the assimilation estimates (Fig. 3.1). MAWN is comprised of 79 stations. Each station uses two Campbell Scientific water content reflectometers (CS615 or CS616) to measure soil moisture. The two probes are horizontally inserted to provide hourly soil moisture measurements at depths of 10 and 25 cm (for 46 MAWN sites) or are vertically installed to measure soil moisture in the upper 60 cm profile (0-30 and 30-60 cm) (for 33 MAWN sites since about the middle of year 2008). In this study, only 30 stations, which generally correspond to the "sCmC" grids (i.e. a crop cover-dominated grid with satellite soil moisture also from a cropped surface) defined in Chapter 3, are used for the validation because the AMSR-E retrievals are not available or not considered for the rest stations due to the impact of open water and/or dense vegetation. A filtering step is applied to all in situ data to ensure the reliability and effectiveness of the subsequent validations. In situ soil moisture observations are rejected if (1) they are beyond any realistic ranges (e.g., too high or too low to be explained by physical variability); (2) the time series contains sudden changes (significant "jump") that are impossibly attributed to physical process; or (3) the surface soil is frozen.

Typically only one in situ site is available per model grid cell. Since the absolute magnitude of soil moisture for the areal average (satellite product scale or model grid cell) is difficult to estimate based upon point-source observations, the root-mean-square error (RMSE) metric is not used for the validation in this study. Instead, we use only the skill

metric *R*, which is defined as the daily time series correlation with point measurements. This is because the temporal variability of soil moisture observed by point measurement may be spatially representative (e.g. Brocca et al., 2009; Loew and Mauser, 2008; Martinez-Fernandez and Ceballos, 2005), although point measurements are not readily converted to the spatial averages. The neighboring MAWN sites typically show good agreement for the temporal pattern of soil moisture (Chapter 3), indicating that overall the point ground measurements being used in this work can represent the areal average (satellite product scale or model grid cell) in terms of the temporal variability of soil moisture.

4.3 Results

4.3.1 Soil moisture skill comparison between NSIDC and LPRM products

As mentioned earlier, only point ground measurements are available for the validation in this study. Since the absolute magnitude of soil moisture for the areal average (corresponding to the satellite footprint scale or the model grid cell) is difficult to estimate based upon point-source observations, we use the skill metric *R*, which is defined as the daily time series correlation with point measurements. AMSR-E measures only the water content within the top ~ 1 cm soil layer. The retrievals (from both ascending and descending orbits) are mapped onto the model grid cells using a nearest neighbor approach. After using the CDF matching approach, the satellite retrievals correspond to the model surface layer (0-10 cm). Note that the CDF matching scheme has no impact on the retrieval

skill. For a given model grid, the skill for the retrievals, the open-loop model, and the assimilation is assessed using in situ measurements falling within the grid cell. Typically only one in situ site is available per model grid cell. In this study, the surface soil moisture skill (satellite retrievals, open-loop and assimilation) are computed using in situ measurements taken at 10 cm depth or in the top 30 cm profile (for those sites where the probes are vertically installed). The root-zone soil moisture skill (open-loop and assimilation) is derived using a depth-weighted average of soil moisture estimates in the model's top two layers (0-10 and 10-35 cm) against the arithmetic mean of in situ measurements at 10 and 25 cm depths or the 0-30 cm profiles measured by vertically installed sensors. The skill *R*, for a given grid cell, is computed separately for each year (unfrozen period). We do not compute the *R* values when any of the following occurs: (1) the effective length of satellite retrieval daily time series is less than 60 days per year; (2) in situ soil moisture (unfrozen) time series are shorter than 100 days per year; (3) the time series standard deviation of in situ soil moisture is less than 0.02 m^3/m^3 (since the measurement noise may significantly impact the *R* values when the time series standard deviation is too small); or (4) linear or quadratic trends in the soil moisture time series significantly contribute to the correlations. Fig. 4.1 shows the spatial distribution of the retrieval skill for the two AMSR-E products. The comparison between the skill of the open-loop model (single integration without assimilation) and the skill for the assimilation estimates are provided in Fig.4.2 for surface soil moisture and in Fig. 4.3 for root-zone soil moisture. The years 2009, 2010 and 2011, which have the most in situ data available, are chosen for shown.

The two AMSR-E soil moisture products (NSIDC and LPRM) have been evaluated using in situ point or network measurements over different regions such as the United States (e.g., Jackson et al. 2010; Crow and Zhan, 2007), Canada (e.g. Champagne et al., 2010), Europe (e.g., Brocca et al., 2011; Wagner et al., 2007), and Australia (e.g. Draper et al. 2009b). Each product performed differently in different studies. It is generally accepted that the LPRM product has better correlations with in situ data than the NSIDC retrievals. In the present work (Fig. 4.1), the two retrieval products typically have low to modest skill *R* with a median of 0.31 for the NSIDC product and a median of 0.42 for the LPRM product. Each product performed differently at different locations. Overall, the retrievals from LPRM (mean $R = 0.39$) showed higher skill than the NSIDC product (mean $R = 0.31$), which is fairly consistent with the evaluations over other regions (e.g., Jackson et al. 2010; Brocca et al., 2011; Draper et al. 2009b).

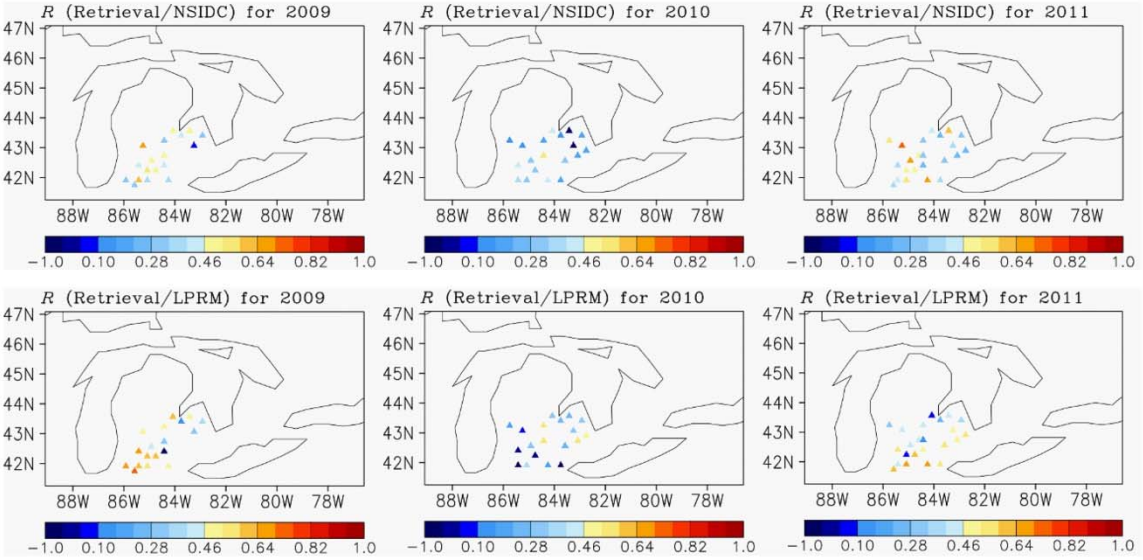

Fig. 4.1 Retrieval skill for (top) AMSR-E NSIDC product and (bottom) AMSR-E LPRM product over years (left) 2009, (middle) 2010, and (right) 2011, respectively.

The open-loop model (Fig. 4.2, top row) typically provides low to modest surface soil moisture skill (a median of 0.38). After the assimilation (Fig. 4.2, the 2nd and 3rd rows), the model surface soil moisture skill was increased, especially for the LPRM assimilation (to a median of 0.59). The statistically significant skill gain $\Delta R^{A\text{-}M}$, defined as the skill for the assimilation minus the skill for the open-loop, are observed less frequently for the NSIDC assimilation (Fig. 4.2, the 4th row) than through assimilating the LPRM product (Fig. 4.2, bottom row). Overall, skill improvement $\Delta R^{A\text{-}M}$ is larger for the LPRM assimilation, consistent with higher retrieval skill for the LPRM product.

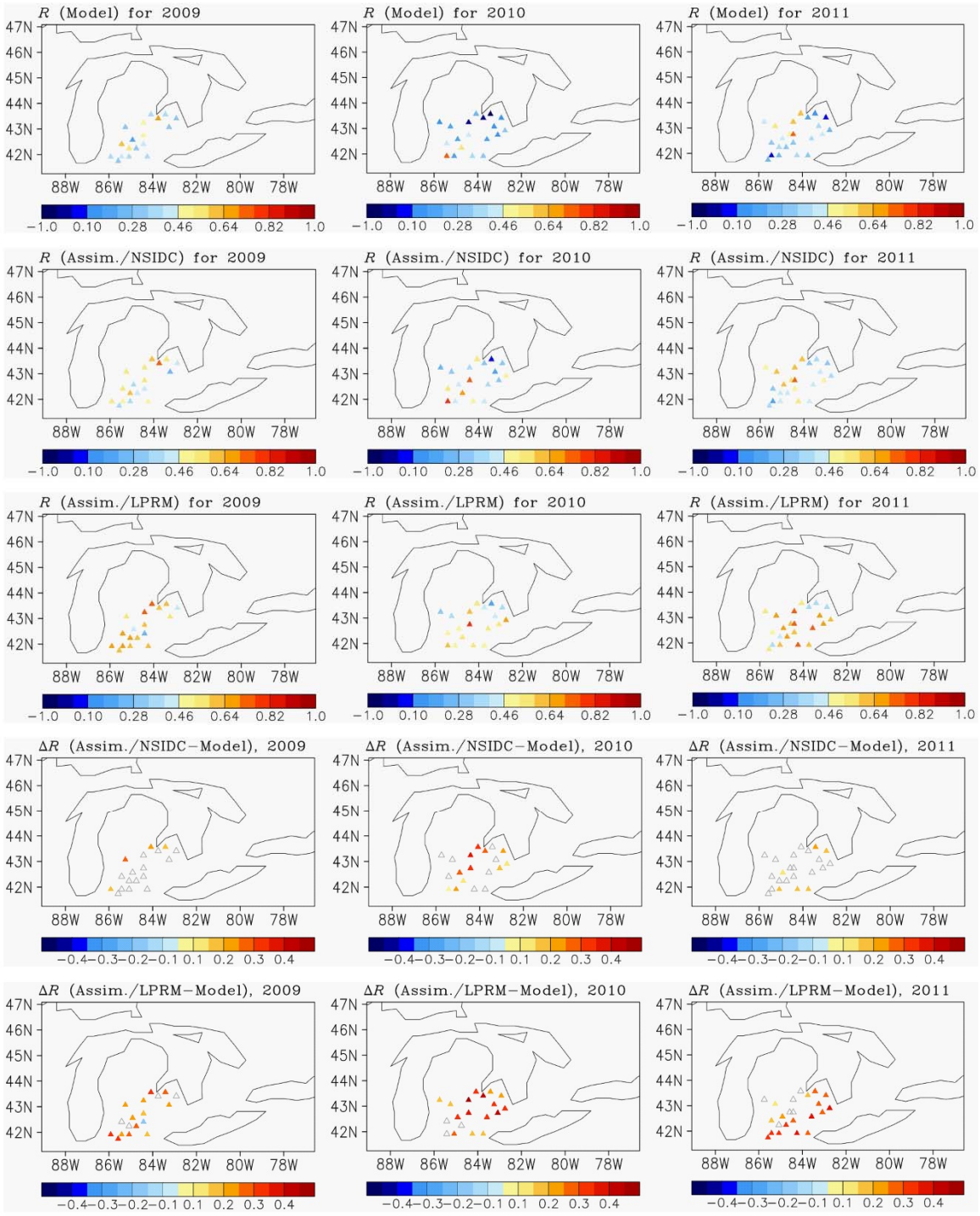

Fig. 4.2 Skill for surface soil moisture from (top row) the open-loop model, (2nd row) the assimilation of NSIDC product, (3rd row) the assimilation of LPRM, (4th row) the skill improvement $\Delta R^{\text{A-M}}$ (Assimilation minus Open-loop) from the NSIDC assimilation, and (bottom) $\Delta R^{\text{A-M}}$ from the LPRM assimilation, over three individual years (left to right: 2009, 2010, and

2011). In the bottom two rows, $\Delta R^{A\text{-}M}$ is denoted by an open symbol in grey if the open-loop R and the assimilation R are not significantly (5% level) different from each other. Symbols denote locations of the validation grids.

The open-loop skill and its response to the assimilation for root-zone soil moisture (Fig. 4.3) are similar to those observed for the surface soil moisture (Fig. 4.2). Overall, the root-zone soil moisture skill is slightly higher than the surface soil moisture skill for both the open-loop and the assimilation estimates. This reflects the increase of the assimilation product skill with the open loop skill. The stronger skill improvement $\Delta R^{A\text{-}M}$ for the LPRM assimilation than for the NSIDC assimilation are also observed for root-zone soil moisture. The skill gain in root zone soil moisture indicates that the surface soil moisture information acquired by AMSR-E, through the EnKF assimilation, can propagate to the soil layers that are not directly measured. Note that the single soil moisture skill R (for single grid and single year) from the assimilation estimates was not always above that from the open loop model. Negative $\Delta R^{A\text{-}M}$ may be observed (blue triangles in the last two rows in Figs. 4.2 and 4.3) for the grids where the retrieval skill is much lower than the open loop skill (see section 4.3.2 for details).

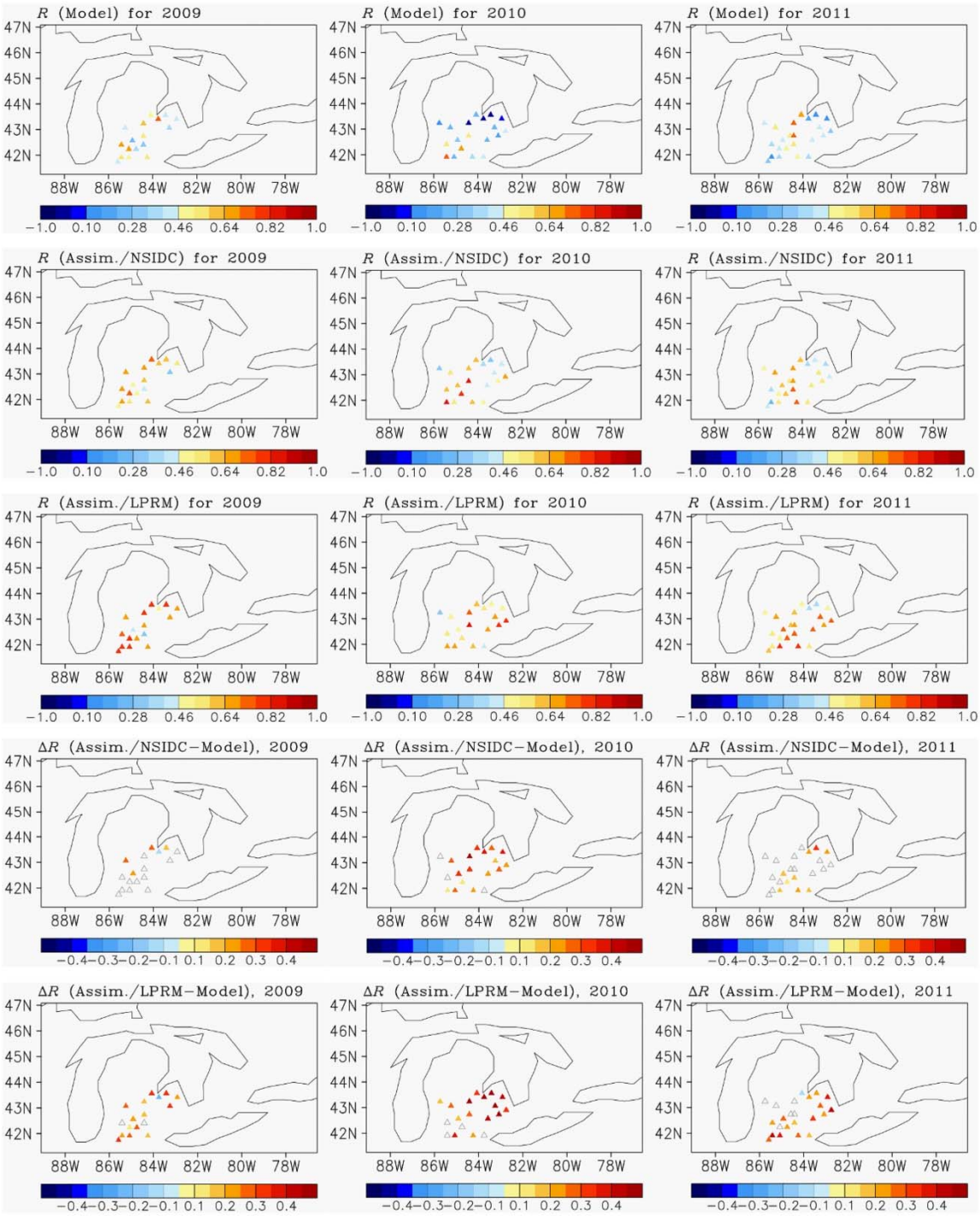

Fig. 4.3 Similar to Fig. 4.2, but for root zone soil moisture.

Mean soil moisture skill R values (over all sites and all years) with 95% confidence intervals are summarized in Figure 4.4 for the retrievals, the open-loop, and the assimilation estimates, respectively. The confidence interval for the mean R is estimated using $[\bar{R} - t\frac{S}{\sqrt{N}},\ \bar{R} + t\frac{S}{\sqrt{N}}]$, where $\bar{R}$ is the sample mean of all the single R values (single site and single year) for a given soil moisture product, N is the sample size, which is the summation over available validation sites and years in this study, and S is the sample standard deviation. The value of t, which depends upon the degrees of freedom (i.e. N-1) and the level of confidence, can be determined from the known t table. The mean retrieval skill is significantly higher for AMSR-E/LPRM (0.39 (mean R) ± 0.03 (95% confidence level)) than for AMSR-E/NSIDC (0.31 ± 0.03). The mean R for the model open-loop is 0.37 for the surface soil moisture, and is 0.46 for the root zone soil moisture. As expected, the improvement in the model skill R through assimilation increases with increasing retrieval skill. After the assimilation of NSIDC product, the mean model skill R is increased by about 0.11 for both the surface and root zone soil moisture. After assimilating the LPRM product, the mean skill improvement $\Delta R^{\text{A-M}}$ is about 0.19 for both the surface and root zone soil moisture. All the mean skill improvements are statistically significant. The skill difference between the two assimilation estimates (about 0.08) for either surface or root zone soil moisture is approximately equal to that between the two retrieval products (0.31 for NSIDC and 0.39 for LPRM). However, note that the increase in assimilation skill, for a given open loop model skill, is typically slower (smaller) than the increasing in retrieval skill (Reichle et al., 2008a).

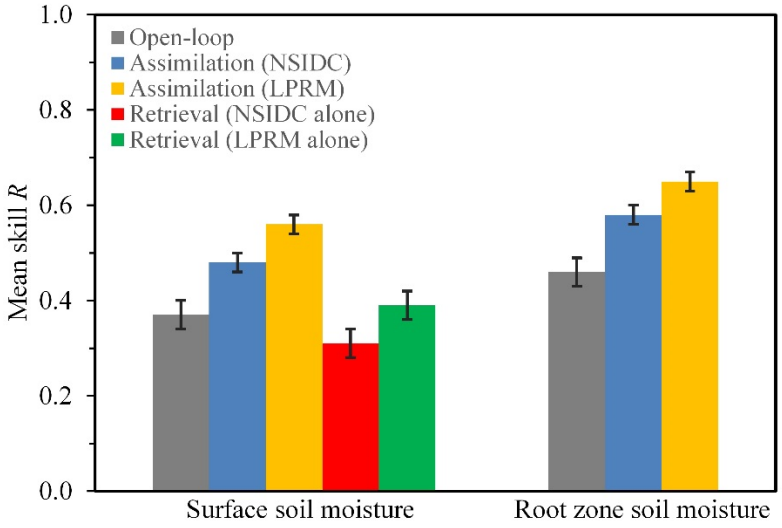

Fig. 4.4 Mean soil moisture skill *R* for the AMSR-E retrievals, the open-loop model, and the assimilation estimates. The retrievals and the assimilation estimates are based upon the AMSR-E/NSIDC product alone and the AMSR-E/LPRM product alone, respectively. Error bars indicate 95% confidence intervals

Fig. 4.4 shows that the mean skill for the assimilation estimates always exceeds that of the open-loop model, even when the retrieval skill (e.g. NSIDC) is lower than that of the open-loop model. Through synthetic assimilation experiments, Reichle et al. (2008a) pointed out that if the open-loop model skill was low to modest even the retrievals of low skill could contribute to the assimilation skill. The synthetic study also indicated that the surface soil moisture skill from the assimilation estimates was typically above the satellite observation skill, except for the presence of a poor open-loop model skill and a high satellite skill. Similarly, Fig. 4.5 shows that the mean $\Delta R^{A\text{-}S}$, defined as the skill for the surface soil moisture assimilation product minus the observation skill is about 0.17 for the assimilation of either NSIDC or LPRM product. This evidently demonstrates that the assimilation

produced superior soil moisture estimates, relative to both the open-loop model and the observation product alone.

Liu et al. (2011) assessed the contributions of both the NSIDC and LPRM products (June 2002 to July 2009), through the EnKF assimilation, to the CLSM model soil moisture skill using in situ measurements from the continental United States Soil Climate Analysis Network (SCAN). We can compare their skill levels with our results. Note that differences between the two studies are expected since the model, forcing data, and in situ measurements used in Liu et al. (2011) are different from those used in our study. However, the two studies showed similar modulation of the two AMSR-E products on the model soil moisture skill. Liu et al. (2011) showed that the retrieval skill was higher for LPRM than for NSIDC (Fig. 3 therein). Accordingly, the LPRM retrievals resulted in greater skill in assimilation product than the NSIDC product. For the CLSM model forced with precipitation from the NASA Modern Era Retrospective analysis for Research and Applications (MERRA), which has the mean open-loop skill (0.43 for surface soil moisture and 0.47 for root zone) similar to ours (Fig. 4.4), the skill improvement from the assimilation (for both surface and root zone) is about 0.05 for NSIDC and about 0.11 for LPRM. Their skill improvements are smaller than those obtain in the present study (0.11 for NSIDC and 0.19 for LPRM), which may be due to the use of anomaly *R* metric in their study. Draper et al. (2012) demonstrated that through assimilating 3.5 years (January 2007 to May 2010) of LPRM retrievals in CLSM the mean skill (anomaly *R*) improvement is about 0.09 (the open loop skill is about 0.45) over the United States SCAN/SNOTEL network and the Murrumbidgee Soil Moisture Monitoring Network in southeast Australia.

Overall our raw *R* metric of skill and the anomaly *R* metric (e.g. Liu et al., 2011; Draper et al., 2012) supports the same general conclusions, especially for the assimilation dependence on the satellite retrieval skill.

4.3.2 Dependence of the assimilation upon the retrieval-model skill difference

The skill improvement ΔR^{A-M} is controlled not only by the satellite observation (retrieval) skill but also by the skill for the open-loop estimates. In general, the skill improvement ΔR^{A-M} increases as the satellite observation skill, but decreases with increased open-loop skill (Reichle, et al. 2008a). To further investigate the impact of the open-loop skill and the retrieval skill on the assimilation, Fig. 4.5 provides the skill improvement ΔR^{A-M} (the assimilation skill minus the open-loop skill) against ΔR^{S-M}, defined as the retrieval skill minus the skill for the open-loop surface soil moisture. The skill improvement ΔR^{A-M} approximately increases linearly with ΔR^{S-M} when assimilating either AMSR-E/NSIDC or AMSR-E/LPRM retrievals. As long as ΔR^{S-M} exceeds -0.2 (i.e., assimilating retrievals with a skill no more than 0.2 below the open-loop skill), the assimilation is typically able to increase the model skill (i.e. a positive ΔR^{A-M} is expected). If the retrieval skill is greater than or equal to the open-loop surface soil moisture skill (i.e. $\Delta R^{S-M} \geq 0.$), the skill improvement ΔR^{A-M} are often statistically significant. When the skill for the retrievals is more than 0.2 below the open-loop skill (i.e., $\Delta R^{S-M} < -0.2$), the chances for positive ΔR^{A-M} become slim. The results are fairly consistent with Draper et al. (2012). The study showed that the assimilation of AMSR-E and ASCAT retrievals in CLSM typically generated an improved skill (in terms of anomaly R) for both the surface and root zone soil moisture as

long as the satellite observation skill is no more than about 0.2 lower than the open-loop skill. Chapter 3 also suggested that the assimilation of SMOS soil moisture may be not helpful and even negatively affect the open-loop skill if the skill for SMOS retrievals is more than about 0.3 below the open-loop skill.

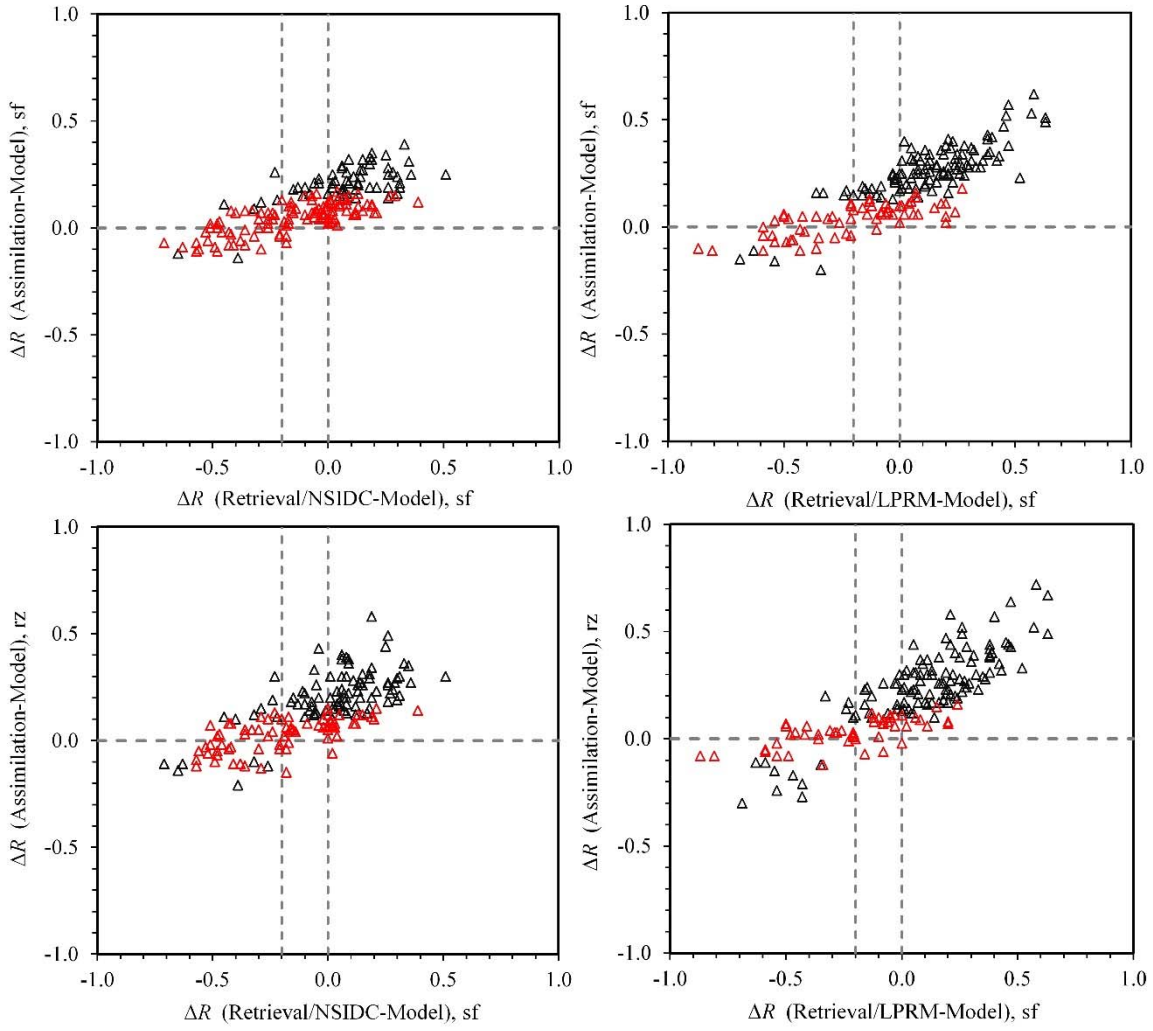

Fig. 4.5 Skill improvement $\Delta R^{A\text{-}M}$ (skill for the assimilation minus the open-loop skill, ordinate) for (top) surface and (bottom) root-zone soil moisture against $\Delta R^{S\text{-}M}$ (skill for the satellite retrievals minus skill for the open-loop surface soil moisture, abscissa), derived from (left) the

assimilation of AMSR-E/NSIDC and (right) the assimilation of AMSR-E/LPRM. Symbols in red mean that ΔR^{A-M} are not statistically significant at the 5% level. The horizontal dashed line denotes $\Delta R^{A-M} = 0$. The two vertical dashed lines denote $\Delta R^{S-M} = -0.2$ and $\Delta R^{S-M} = 0$, respectively.

As shown in Fig. 4.5, overall the surface soil moisture ΔR^{A-M}, relative to root-zone ΔR^{A-M}, exhibits a better linear relationship with ΔR^{S-M}. For a given ΔR^{S-M}, the skill improvement ΔR^{A-M} is usually more variable (along the ordinate) for root-zone soil moisture than for surface soil moisture. This may be associated with the fact that during the assimilation the update of root-zone soil moisture is subject to the accurate information exchanges between the surface soil and deeper layers, which, in turn, are controlled by many factors such as model dynamics and input error parameters. However, note that the linear association between ΔR^{A-M} and ΔR^{S-M} is not expected to be perfect since the sensitivity of ΔR^{A-M} to ΔR^{S-M} is additionally affected by the magnitude of open-loop skill R. Reichle et al. (2008a) showed that along the axis of retrieval skill the contour lines of the skill improvement ΔR^{A-M} are denser at low to modest open-loop R than at higher open-loop R (Figure 2c and 2d therein), i.e., the skill improvement ΔR^{A-M} is more sensitive to the increase in the retrieval skill when the open-loop R is low to modest than when the open-loop R is high. Therefore, the same ΔR^{S-M} typically leads to larger ΔR^{A-M} at low open-loop R than at high open-loop R. From their Figure 2c, for example, for the retrieval with skill R = 0.4 and an open-loop surface soil moisture with skill $R = 0.3$, ΔR^{S-M} was 0.1 and ΔR^{A-M} was about 0.13, while for the retrieval with $R = 0.5$ and an open-loop skill $R = 0.4$, ΔR^{S-M} was still 0.1 but ΔR^{A-M} was only about 0.08.

4.3.3 Comparison with SMOS assimilation

The launch of SMOS satellite that carries an L-band (~1.4 GHz) Microwave Imaging Radiometer with Aperture Synthesis (MIRAS) has opened up new opportunities for land data assimilation. Chapter 3 reported the assimilation of SMOS soil moisture retrievals (2010-2013) in the MESH model over the Great Lakes region. The study revealed the impact upon the assimilation of the open-loop skill and the satellite observation skill. The crop-dominated grids typically experienced substantial skill improvement $\Delta R^{A\text{-}M}$ when the assimilated SMOS retrievals also came from crop surfaces, due to the presence of a high satellite observation skill and a low open-loop skill. Here we perform the comparison between the assimilation results from SMOS and AMSR-E, which may offer further insight into the dependence of the assimilation upon the satellite retrieval skill. As a reminder, the AMSR-E products covered June 2002 to October 2011 while the SMOS retrievals are available from January 2010 to present. First, we compare the summarized results from the 2003-2011 AMSR-E assimilation (Fig. 4.4) to those based upon the 2010-2013 SMOS assimilation (Table 3.1). For the SMOS assimilation, we only take into account the validation results over the "sCmC" grids, to be consistent with the AMSR-E assimilation. The open-loop model skill for 2010-2013 (R = 0.39 for surface and R = 0.47 for root zone soil moisture) is similar to that for 2002-2011(R = 0.37 for surface and R = 0.46 for root zone soil moisture). The mean retrieval skill is higher for SMOS (R = 0.55) than for the AMSR-E products (R=0.31 for NSIDC and R=0.39 for LPRM). Accordingly, the assimilation of SMOS retrievals resulted in greater skill improvement (mean $\Delta R^{A\text{-}M} = 0.25$)

than the assimilation of AMSR-E products (mean $\Delta R^{A\text{-}M} = 0.11$ for NSIDC and 0.19 for LPRM).

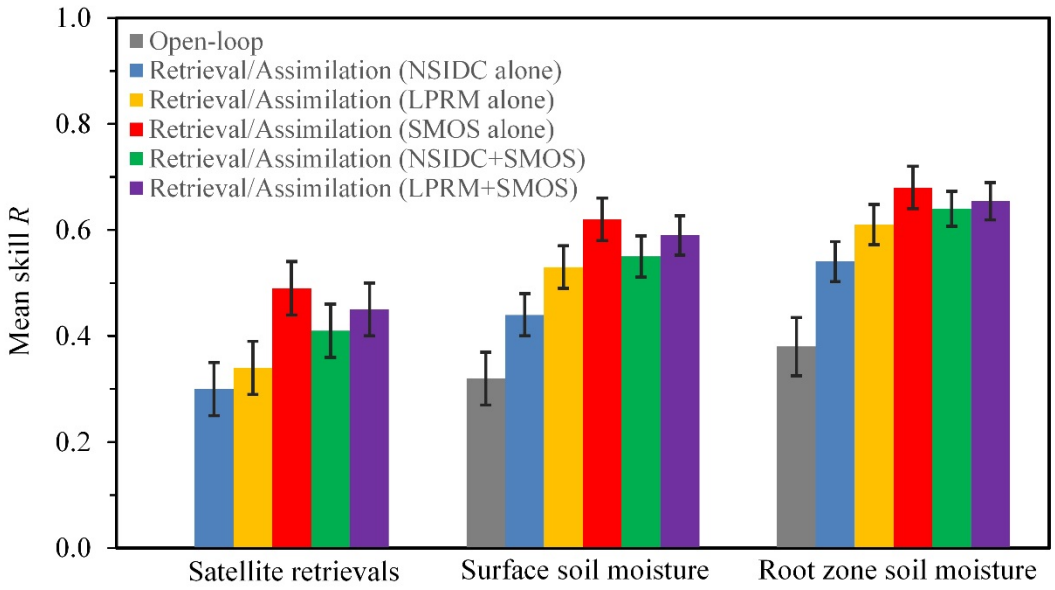

Fig. 4.6 Mean soil moisture skill *R* for the satellite retrievals, the open-loop model, and the assimilation estimates, based on two years (2010-2011) of results. The retrievals and the assimilation estimates are based upon the AMSR-E/NSIDC product alone, the AMSR-E/LPRM product alone, the SMOS product alone, the combination of NSIDC and SMOS, and the combination of LPRM and SMOS, respectively. Error bars indicate 95% confidence intervals

Next, we compute the mean skill based upon only the years (i.e. 2010 and 2011) for which both SMOS and AMSR-E retrievals are available. Furthermore, since the two sensors have different overpassing time (01:30 am/pm LST for AMSR-E; 06:00 am/pm LST for SMOS), we also perform the combined assimilation of instantaneous soil moisture retrievals from the two instruments (AMSR-E/NSIDC + SMOS and AMSR-E/LPRM + SMOS). The SMOS soil moisture product used here is same as that used in Chapter 3. Prior to the joint

assimilation, all the retrieval products (separately) are rescaled using the CDF matching where the model CDF is based on the 10-year (2002-2011) MESH model surface soil moisture. Figure 4.6 summarizes the mean skill for the retrievals, open-loop model, and the assimilation estimates, using a space-time mask for AMSR-E and SMOS measurements. The results further reveals that the skill for the assimilation estimates for both surface and root zone soil moisture increases as the satellite retrieval skill. For the retrieval skill, SMOS soil moisture (R = 0.49) is significantly higher than the LPRM product (R = 0.34) and the NSIDC product (R=0.30). The mean open-loop skill R is 0.32 for surface soil moisture. After assimilating the three products separately, the mean skill R values for surface soil moisture are increased (statistically significant at 5% level) to 0.44 (NSIDC), 0.53 (LPRM), and 0.62 (SMOS), respectively. For the root zone soil moisture, the NSIDC, LPRM, and SMOS products lead to the gains (statistically significant at 5% level) of 0.16, 0.23, and 0.30 in the mean skill, respectively.

The retrieval skill for the combined AMSR-E and SMOS (R=0.41 for NSIDC + SMOS; R = 0.45 for LPRM+SMOS) is significantly higher than the AMSR-E product alone (R=0.30 for NSIDC and R=0.34 for LPRM), but is lower than the SMOS product alone (R=0.49) (Fig. 4.6). Consequently, the assimilation skill from the joint assimilation of AMSR-E and SMOS is greater than that from the assimilation of AMSR-E alone (especially for the NSIDC product), but is not superior to that from the SMOS alone assimilation. Draper et al. (2012) showed that the joint assimilation of AMSR-E and ASCAT soil moisture could produce slightly better skill improvement (not statistically significant) than assimilating either of them. This could be because that AMSR-E and

ASCAT operate at similar microwave frequencies and have similar retrieval skills. Our results indicate that the combined assimilation of X-band (AMSR-E) and L-band (SMOS) products does not necessarily yield the best skill improvement.

4.4 Discussion

In our EnKF assimilation, the model input error parameters are not on-line tuned. Although the assimilation estimates are subject to the specified input error parameters, Reichle et al. (2008b) demonstrates that a non-adaptive EnKF typically leads to improved soil moisture estimates (over the open loop), even when the input error parameters moderately deviate from their true values. However, when the error estimates for the model and/or the retrievals are far from the realistic conditions, the assimilation estimates may be even worse than the open-loop model (Reichle et al., 2008b). This could be used to explain the occurrence of negative $\Delta R^{A\text{-}M}$ when the retrieval skill is very low or even negative (the corresponding $\Delta R^{S\text{-}M}$ is thus small) (Fig. 4.5). For the retrievals with very low or even negative skill, which generally reflect poor satellite observations, their real errors could be severely underestimated by the input error parameters, thus causing negative $\Delta R^{A\text{-}M}$. Overall, negative $\Delta R^{A\text{-}M}$ is severer in root zone than for the surface layer (Fig. 4.5). This is because poorly specified observation error variances have a stronger impact on the assimilation estimates of root zone soil moisture than on surface soil moisture estimates (Reichle et al., 2008b). To avoid this problem, on-line quality control routines (e.g. cross-comparisons, Reichle, 2008) and on-line adjusting of input errors parameters (Reichle et al., 2008b) need to be added to the assimilation system.

In the present work, the *R* value is derived based upon the raw time series rather than using the anomaly time series (e.g. Reichle et al., 2007; Liu et al., 2011; Draper et al., 2012) considering that the extraction of anomaly soil moisture for this study is hampered by a number of factors. First, the soil moisture anomalies, defined as departures of daily soil moisture from the climatological annual cycle, are sensitive to the computation of the climatology. Different climatology calculations, such as monthly mean (e.g. Reichle et al., 2007; Liu et al., 2011); 31-day window running average (e.g. Draper et al., 2012), and normalized 5-week moving average (e.g. Brocca et al. 2011) typically lead to different anomalies. Second, the in situ record lengths are different across the validation sites, and the time coverage is less than 3 years for a substantial portion of stations. Third, depths of in situ soil moisture were changed over the years. Some sites have measured soil moisture in the upper 60 cm profile (0-30 and 30-60 cm) since the middle of year 2008, as compared to previous observations at horizontal depths of 10 and 25 cm. However, note that overall our raw *R* metric of skill and the anomaly R metric (e.g. Liu et al., 2011; Draper et al., 2012) qualitatively provide the same general conclusions, especially for the assimilation dependence on the satellite observation skill.

4.5 Summary and Conclusion

The assimilation of AMSR-E soil moisture has been an active research area. In this chapter, the 1D-EnKF is adopted to assimilate the two AMSR-E retrieval products (2003-2011), NSIDC and LPRM, into the MESH model over the Great Lakes basin. The satellite

retrievals, the open-loop model, and the assimilation estimates are validated against point-scale in situ soil moisture measurements, in terms of the daily time series correlation coefficient (soil moisture skill R). Due to the bias (systematic error) between the satellite retrievals and the model soil moisture estimates, a priori rescaling on the retrievals is performed (the CDF matching). Additionally, the AMSR-E sensor-based assimilation results are compared with the assimilation of SMOS retrievals (only for lightly or moderately vegetated areas). The following conclusions can be drawn from this work.

- Overall, the assimilation of either NSIDC or LPRM product led to superior skill, relative to either the open-loop model skill or the retrieval skill. The improvement in the model skill R through assimilation increases with increasing retrieval skill. On average, the LPRM assimilation yielded larger skill improvement $\Delta R^{A\text{-}M}$, defined as the skill for the assimilation soil moisture product minus the skill for the open-loop estimates, for both surface and root-zone soil moisture than the NSIDC assimilation.

- The skill improvement $\Delta R^{A\text{-}M}$ is strongly controlled by $\Delta R^{S\text{-}M}$, defined as the retrieval skill minus the open-loop surface soil moisture skill. For a single site and a single year, as long as $\Delta R^{S\text{-}M}$ exceeds -0.2 (i.e., assimilating retrievals with a skill less than 0.2 below the open-loop skill), the assimilation is typically able to produce increased skill (i.e. a positive $\Delta R^{A\text{-}M}$).

- The assimilation of SMOS retrievals resulted in greater skill improvement $\Delta R^{A\text{-}M}$ than the assimilation of AMSR-E products. The skill improvement $\Delta R^{A\text{-}M}$ from the joint

assimilation of AMSR-E and SMOS is greater than that from the assimilation of AMSR-E alone, but is lower than that from the assimilation of SMOS alone.

CHAPTER 5

Comparison of AMSR2 and SMOS Soil Moisture Retrievals for Land Data Assimilation

5.1 Introduction

Data assimilation aims to optimally merge model estimates and observations by quantifying their respective error characteristics. For land surface and hydrologic modeling, data assimilation techniques handle the optimization problem in the presence of random noise in forcing fields and uncertainties in model physics and/or parameters. Assimilation of satellite microwave soil moisture in land surface and hydrologic models has received considerable attention within the past decades. In particular, ten years (2002-2011) of operations of the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) provided key data sources for advances in land data assimilation (e.g. Reichle et al., 2007; Draper et al., 2009, 2012; Liu et al., 2011; Li et al., 2012). In Chapter 4, we presented the assimilation of AMSR-E soil moisture in the MESH model as well as the comparison between the assimilation of AMSR-E and SMOS. Since October 2011, the AMSR-E soil moisture data have been no longer available due to a technical problem with the instrument's antenna. As the successor of AMSR-E, the Advanced Microwave Scanning Radiometer 2 (AMSR2) onboard the first generation of the Global Change Observation Mission-Water (GCOM-W1) satellite, launched by the Japan Aerospace Exploration Agency (JAXA) in May 2012, is currently in operation. The coexistence of AMSR2 (May 2012-present), SMOS (November 2009-present), and SMAP (January 2015-

present) will open up the new opportunities for filling the research gap regarding assimilation of satellite soil moisture measurements obtained at different microwave frequencies (namely, X/C band vs. L band). In this chapter, we conduct a comparative study of assimilating AMSR2 and SMOS soil moisture in the MESH model. This study is intended to complement the AMSR-E and SMOS comparison, as presented in Chapter 4.

5.2 Methodology

We will use the 1D-EnKF to assimilate AMSR2 and SMOS soil moisture retrievals, separately and jointly. The forecast model, the assimilation filter, the study domain and the SMOS retrievals used here are same as those described in Chapter 3. Here we describe only the AMSR2 soil moisture retrievals. AMSR2 is generally same as the AMSR-E instrument (Table A3). AMSR2 acquires microwave emission from the Earth's surface and atmosphere with a temporal resolution of 1-2 days for both ascending (1:30 pm LST) and descending (1:30 am LST) overpasses. In this work, we assimilate the AMSR2 Level 2 Soil Moisture Content (SMC) product released by JAXA. The product version is Ver. 1.1 (1.110.100) (during the preparation of this work, the Ver. 2.0 (2.220.2.00) of AMSR2 products was released). The inversion of soil moisture is based upon a lookup table method, which was also adopted to produce the JAXA AMSR-E soil moisture (Fujii et al., 2009). The lookup table, which was derived from theoretical calculations using a forward radiative transfer model, describes the relation of soil water content and vegetation water content (as well as the fractional vegetation cover) with two indices, the normalized polarization difference at 10.7 GHz (i.e. brightness temperature difference between the vertical and

horizontal polarizations normalized by their average) and the normalized frequency difference between 36.5 GHz and 10.7 GHz horizontal polarizations (i.e. difference between brightness temperatures obtained at the two frequencies normalized by their average). By looking up the table, soil moisture and vegetation water content can be estimated based upon the observed polarization difference and frequency difference (as well as the observed fractional vegetation cover). Currently only volumetric soil moisture data are stored in the AMSR2 Level 2 SMC product.

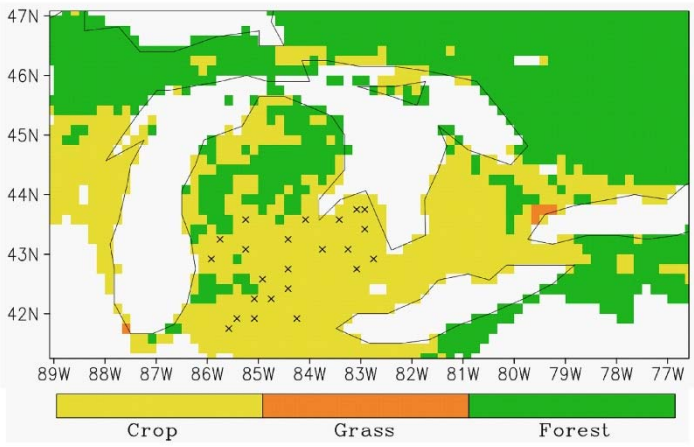

Fig. 5.1 Location of validation sites/grids (23). In situ soil moisture measurements are taken from the Michigan Automated Weather Network.

Prior to the assimilation, both AMSR2 and SMOS retrievals are resampled onto the MESH model grids (~15 km resolution) using a nearest neighbor approach. Whenever and wherever the model (combined with the precipitation data) indicates the presence of precipitation, frozen soils, or snow cover, the retrievals are excluded from the evaluation and assimilation. We also exclude the forested grids (the sum of deciduous, coniferous, and

mixed forest classes within the grid cell exceeds 50%) for analysis considering that AMSR2 soil moisture is typically valid only for low or moderate vegetation cover.

Since the satellite retrievals and model surface soil moisture exhibit different climatologies, a priori rescaling is applied to the retrievals and the observation error standard deviations (stdev). The retrievals are rescaled by matching their cumulative distribution function (CDF) to the model surface soil moisture's CDF. The observation error stdev (0.05 m^3/m^3 and 0.08 m^3/m^3, derived from the satellite climatology, are assumed for the AMSR2 product and the SMOS product, respectively) is rescaled by multiplying it with the ratio between the time series stdev of the scaled retrievals and that of the unscaled retrievals. The rescaling of the retrievals and their error stdev is conducted independently for each model grid, respectively, for AMSR2 and SMOS. Only one year (2013) of soil moisture retrievals are used in this study. The assimilation period is from 1 January through 31 December 2013. The model (ensemble integration without assimilation) is spun up for a 3-year period with the 2010-2012 forcing data. The study domain is still the Great Lakes basin (Fig. 5.1). The satellite retrievals, the open-loop model, and the assimilation estimates are validated against point-scale in situ soil moisture measurements from the Michigan Automated Weather Network (MAWN), in terms of the soil moisture skill *R* (see previous chapters for details). After the quality control step, *R* values are computed only for 23 model grids (Fig. 5.1) where both AMSR2 and SMOS products are available.

5.3 Results and discussion

5.3.1 AMSR2 vs. SMOS scatterplot

Fig. 5.2 presents the SMOS versus AMSR2 retrieval skill for the 23 validation sites (grid cells). The retrieval skill for SMOS soil moisture always exceeds or matches that of the AMSR2 product (The symbols of triangle are typically located above the one-to-one line in Fig. 5.2, left). This is consistent with the dependence of satellite soil moisture retrieval capabilities upon the microwave frequency. The L-band measurements (SMOS) are more sensitive to changes in soil water content than the X-band measurements (AMSR2). Although the two instruments are not in agreement in term of the magnitude of the retrieval skill, the correlation between the two sets of retrieval skill is as high as 0.81 (not listed in Fig. 5.2). This means that the spatial patterns are similar for the two sets of retrieval skill. Note that the retrieval skill (daily time series correlation between retrievals and in situ measurements) is calculated using only the days with available retrievals. Since there are different revisit times (separately for the descending and ascending orbits) for the two sets of satellite measurements with 1-3 days for SMOS and 1-2 days for AMSR2, the number of days with AMSR2 retrievals (between 150 to 180 after the quality control step) is always greater than the daily sequence length of SMOS retrievals, ranging from 120 to 140, in this study (Fig. 5.2, right). Therefore, the SMOS and AMSR2 retrieval skill values are obtained based upon different data sequence lengths.

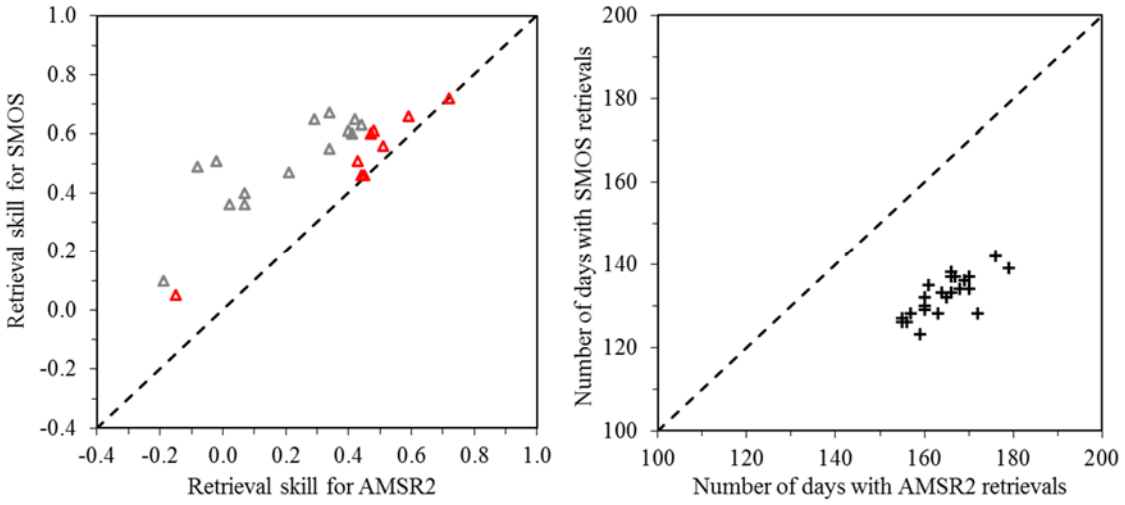

Fig. 5.2 Scatterplot of SMOS vs AMSR2 for (left) the retrieval skill and (right) the number of days with available retrieval data (after quality control). The dashed line is the one-to-one line. In the left plot, a grey symbol of triangle means that the retrieval skill values for SMOS and AMSR2 are significantly (5% level) different from each other.

Fig. 5.3 shows the SMOS versus AMSR2 comparison, in terms of the skill for the assimilation estimates and the skill improvement $\Delta R^{\text{A-M}}$ (defined as the skill for the assimilation soil moisture product minus the skill for the open-loop estimates), respectively. In contrast with their evident difference in retrieval skill (Fig. 5.2, left), the assimilation soil moisture product skill (and thus the skill improvement $\Delta R^{\text{A-M}}$) obtained with the two sets of retrievals are in good overall agreement. In general, the contribution of satellite retrievals to the assimilation (i.e. the skill improvement $\Delta R^{\text{A-M}}$) is expected to increase as the retrieval skill increases (e.g. Chapter 2; Chapter 4; Reichle et al., 2008). Here, while the SMOS retrieval skill is typically higher than the AMSR2 retrieval skill, the skill gain $\Delta R^{\text{A-M}}$ caused by assimilating SMOS does not necessarily exceed the corresponding $\Delta R^{\text{A-M}}$ by assimilating AMSR2. For either surface or root zone soil moisture,

the skill improvements $\Delta R^{A\text{-}M}$ from assimilating SMOS is within 0.05 of the corresponding $\Delta R^{A\text{-}M}$ obtained through assimilating AMSR2 for 18 out of 23 validation sites (i.e. squares scattered near the one-to-one line in Fig. 5.3). This means that the SMOS and AMSR2 retrievals, although with statistically different observation skill, made similar contributions to the model soil moisture skill through data assimilation. Considering that an increased assimilation frequency (i.e. a shorter assimilation interval) typically can favorably influences the assimilation skill (Chapter 2), a deficit in retrieval skill for the AMSR2 assimilation may be partly offset by a higher frequency of the measurements (Fig. 5.2, right).

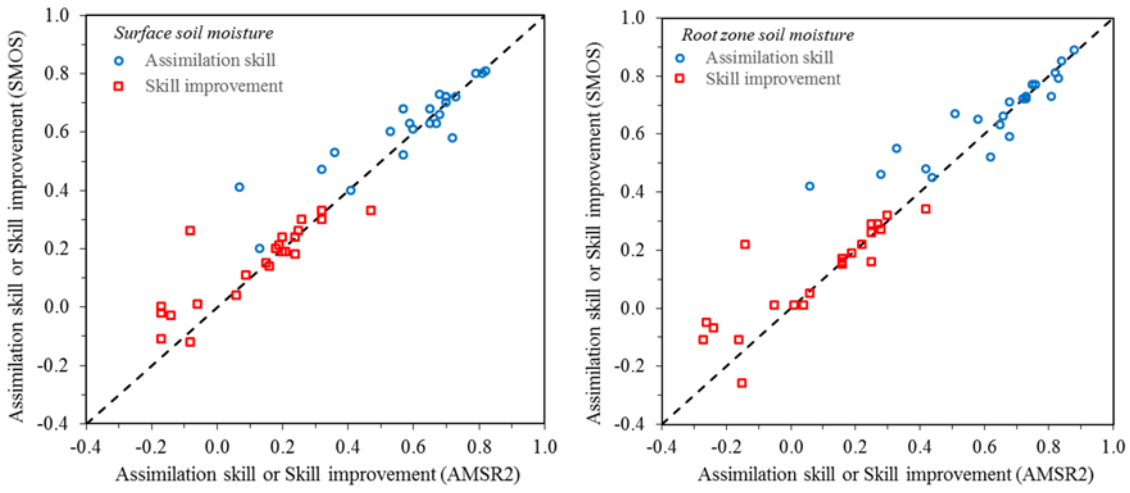

Fig. 5.3 (left) Surface and (right) root zone soil moisture scatterplot of SMOS vs AMSR2 for (circles) the assimilation skill and (squares) the skill improvement (Assimilation-Open loop). The dashed line is the one-to-one line.

Although good overall agreement is seen between SMOS and AMSR2 in terms of contribution to the assimilation estimates, the assimilation product skill, for an individual

grid (validation site), does not always exceed the open loop skill. Fig. 5.4 illustrates the dependence of the skill improvement $\Delta R^{A\text{-}M}$ for both surface and root zone soil moisture upon the retrieval skill and the open loop skill. Overall the SMOS and AMSR2 assimilation experiments provide consistent information. When the satellite observation skill (retrieval skill) is modest to high ($\geq \sim 0.3$) and the open-loop model skill is low to modest ($\leq \sim 0.5$), the skill gain $\Delta R^{A\text{-}M}$ is typically strong (i.e. the symbols appearing in the lower-right portion of the plot in Fig. 5.4). In contrast, if the satellite retrieval skill is low or even negative ($\leq \sim 0.2$) or if the open-loop model skill is high ($\geq \sim 0.6$), $\Delta R^{A\text{-}M}$ is often weak or even negative (i.e. the assimilation is worse than the open-loop). A positive skill improvement (but not necessarily significant in statistics) is typically expected as long as the retrievals skill is less than about 0.2 below the open-loop skill (i.e., located at the lower-right side of the dashed line in Fig. 5.4).

The skill for the assimilation estimates should, in theory, always exceed or match the open-loop skill if the model and observation errors are accurately quantified in the assimilation system. Here negative $\Delta R^{A\text{-}M}$ values, which occurred primarily for very low or negative retrieval skill, especially for the AMSR2 assimilation (Fig. 5.4), are due to the lack of on-line quality control routines (Reichle, 2008) and on-line adjusting of input errors parameters (Reichle et al., 2008b). When the poor quality of satellite retrievals (the corresponding retrieval skill is very small or negative) are assimilated, their real errors could be severely underestimated by the specified input error parameters. Under this situation, the assimilation estimates may be worse than the open-loop model (Reichle et al., 2008b), thus causing the occurrence of negative $\Delta R^{A\text{-}M}$.

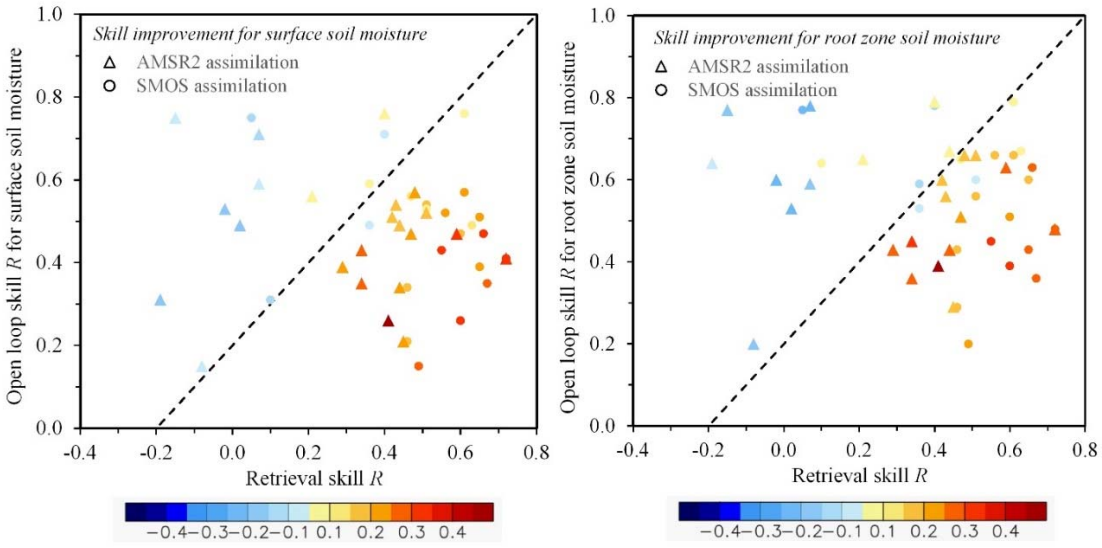

Fig. 5.4 Skill improvement $\Delta R^{A\text{-}M}$ (skill for the assimilation minus the open-loop skill, symbols) for (left) surface and (right) root-zone soil moisture as a function of the retrieval skill (abscissa) and open loop skill (ordinate), from assimilating either (triangles) AMSR2 or (circles) SMOS. The dashed line denotes $\Delta R^{S\text{-}M}$ (skill for the retrievals minus skill for the open-loop) = -0.2.

5.3.2 Mean soil moisture skill

We also compute the mean soil moisture skill over the 23 validate sites for the retrievals, the open-loop, and the assimilation estimates, respectively (Fig. 5.5). The confidence interval for the mean R is estimated using $[\bar{R} - t\frac{S}{\sqrt{N}},\ \bar{R} + t\frac{S}{\sqrt{N}}]$, where $\bar{R}$ is the sample mean of all the single R values (single validation site) for a given soil moisture product, N is the sample size, which is the summation over available validation sites in this study, and S is the sample standard deviation. The value of t, which depends upon the degrees of freedom (i.e. N-1) and the level of confidence, can be determined from the known t table. For either the retrieval skill or the assimilation skill, the confidence interval is quite large for AMSR2 (Fig. 5.5), which should be related to the negative AMSR2 retrieval skill

appeared at several validation grids (Fig. 5.4). As expected, the mean retrieval skill for SMOS soil moisture (R = 0.51) is significantly higher than that of AMSR2 product (R = 0.29).

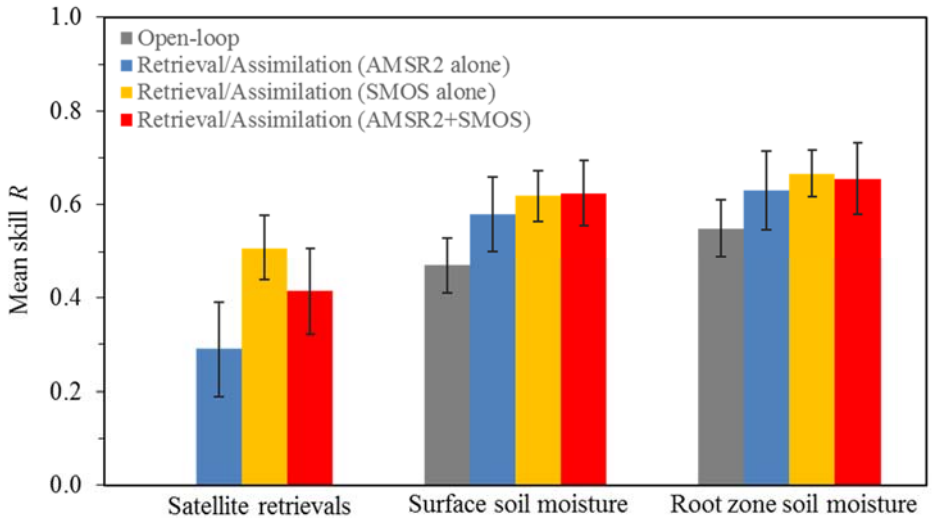

Fig. 5.5 Mean soil moisture skill R for the satellite retrievals, the open-loop model, and the assimilation estimates. The retrievals and the assimilation estimates are based upon the AMSR2 product alone, the SMOS product alone, and the combination of AMSR2 and SMOS, respectively. Error bars indicate 95% confidence intervals.

The mean open-loop skill R are 0.47 for surface soil moisture and 0.55 for root zone soil moisture. After assimilating the AMSR2 and SMOS products, separately, the mean skill R are increased to 0.58 (AMSR2) and 0.62 (SMOS) for surface soil moisture, and to 0.63 (AMSR2) and 0.67 (SMOS) for root zone soil moisture. On average, the assimilation skill (and thus the skill improvement $\Delta R^{A\text{-}M}$) is only marginally sensitive to the increase in the retrieval skill. This is attributed largely to a relatively high open-loop skill for this year (a higher measurement frequency for AMSR2 may also play a role, as discussed in the above

section). Reichle et al. (2008b) showed that along the axis of retrieval skill the contour lines for assimilation product skill become progressively flatter as the open-loop skill *R* increases (Figures 2a and 2b therein). This means that the sensitivity of the assimilation to the retrieval skill becomes weaker as the open-loop *R* increases, i.e., the same increase in the retrieval skill typically leads to a weaker improvement in the assimilation skill for a high open-loop *R* than for a lower open-loop *R*. In Fig. 5.5, however, note that although the two sets of assimilation skill are similar (not significantly different), the confidence levels for their respective assimilation improvement (relative to the open-loop skill) are different. The confidence level is greater (lower) than 95% for the skill improvement in either surface or root zone soil moisture through the assimilation of SMOS (AMSR2).

Since AMSR2 and SMOS have different overpassing time (01:30 am/pm LST for AMSR2; 06:00 am/pm LST for SMOS), we can also perform the combined assimilation of instantaneous soil moisture retrievals from the two instruments. Recall that prior to the joint assimilation, the two sets of retrieval products are (independently) rescaled using the CDF matching approach. The mean retrieval skill for the combined AMSR2 and SMOS is 0.41, which does not significantly (5% level) differ from that for either the AMSR2 product alone or the SMOS product alone (Fig. 5.5). The joint assimilation of AMSR2 and SMOS increases the mean model skill from 0.47 to 0.62 for surface soil moisture, which is statistically significant (5% level), and from 0.55 to 0.66 for root zone soil moisture, which does not reach the 5% level of significance. Overall the combined assimilation of two sensor products does not, relative to the assimilation of either product alone, further significantly improve the assimilation skill.

5.4 Summary

As a complementary work to the AMSR-E vs. SMOS comparison for land data assimilation (section 4.3.3), this chapter compares the contributions of AMSR2 and SMOS retrievals, through data assimilation, to the MESH model soil moisture estimates. The AMSR2 and SMOS retrievals (year 2013) are assimilated into the MESH model over the Great Lakes basin with the 1D-EnKF. A priori rescaling on the retrievals, separately for the two instruments, is performed by matching their cumulative distribution function (CDF) to the model surface soil moisture's CDF. Across the 23 validation grids, the retrieval skill for L-band measurements (SMOS product) typically exceeds that of X-band measurements (AMSR2), consistent with the impact of the microwave frequency upon satellite retrieval capabilities. For most of the validation sites, however, the two sets of retrievals made similar contributions to soil moisture skill in the assimilation system. The possible explanations include (i) the relatively high open-loop skill (since the sensitivity of the assimilation to the retrieval skill becomes weaker as the open-loop R increases) and (ii) the higher measurement frequency for AMSR2 (since a shorter assimilation interval typically can favorably influences the assimilation skill). In terms of mean soil moisture skill, although the two sets of assimilation skill are not significantly different, the confidence level for the assimilation improvement (relative to the open-loop skill) is higher for assimilating SMOS than for assimilation of AMSR2. The combined assimilation of two sensor products does not further significantly improve the assimilation skill.

Overall, the SMOS and AMSR2 assimilation experiments provide consistent information regarding the dependence of the skill improvement $\Delta R^{A\text{-}M}$ upon the retrieval skill and the open loop skill. When the satellite observation skill (retrieval skill) is modest to high and the open-loop model skill is low to modest, the skill gain $\Delta R^{A\text{-}M}$ is typically strong. A positive skill improvement (but not necessarily significant in statistics) is typically expected as long as the retrievals skill is less than about 0.2 below the open-loop skill. To suppress the occurrence of negative $\Delta R^{A\text{-}M}$ values, on-line quality control routines and adaptive adjusting of input error parameters should be useful.

CHAPTER 6

Conclusions and Contributions

6.1 Chapter synthesis and conclusions

As indicated in Chapter 1, this work was motivated by several questions: how will satellite retrievals of surface soil moisture, through data assimilation, impact the MESH model's soil moisture? how does the vegetation cover modulate the assimilation performance? And how important is the satellite observation skill to the assimilation estimates? To address these questions, the one-dimensional version of the Ensemble Kalman filter (1D-EnKF) scheme was designed for the assimilation of satellite soil moisture in the standalone version of MESH model. After validating the established assimilation scheme through the Observing System Simulation Experiment (Chapter 2), this study explored the assimilation of soil moisture retrievals, derived from SMOS (Chapter 3), AMSR-E (Chapter 4), and AMSR2 (Chapter 5), in the MESH model over the Great Lakes basin. A priori rescaling on satellite retrievals (separately for each sensor) was performed by matching their cumulative distribution function (CDF) to the model surface soil moisture's CDF, in order to reduce the satellite-model bias (systematic error) in the assimilation system that was based upon the hypothesis of unbiased errors in model and observation. The satellite retrievals, the open-loop soil moisture (no assimilation) and the assimilation estimates were, respectively, validated against point-scale in situ measurements, in terms of the daily-spaced time series correlation coefficient (skill R). The main findings and conclusions are as follows.

- The skill for satellite soil moisture retrievals typically decreases with increased canopy density. The retrieval skill for L-band measurements (SMOS) is unsurprisingly higher than that for X-band measurements (AMSR-E/AMSR2).

- Assimilating either L-band retrievals (SMOS) or X-band retrievals (AMSR-E/AMSR2) can favorably influence the model soil moisture skill for both the surface layer and root zone except for the cases with a small observation (retrieval) skill and a large open-loop skill. The skill improvement $\Delta R^{A\text{-}M}$, defined as the skill for the assimilation estimates minus the skill for the open-loop soil moisture, is strongly controlled by $\Delta R^{S\text{-}M}$, defined as the retrieval skill minus the open-loop surface soil moisture skill.

- The ability of SMOS to measure surface soil moisture for a wide range of vegetation covers is clearly of advantage for assessing the vegetation modulation of the assimilation. The crop-dominated grids typically experience strong $\Delta R^{A\text{-}M}$ if the assimilated SMOS retrievals also come from crop surfaces (note that a model grid cell and the SMOS node mapped onto the grid are not exactly matched in space), consistent with a high satellite observation skill and a low open-loop skill, while $\Delta R^{A\text{-}M}$ is usually weak or even negative for the forest-dominated grids when the SMOS retrievals also from forested surfaces are assimilated, due to the presence of a low observation skill and a high open-loop skill.

- The assimilation of L-band retrievals (SMOS) typically resulted in greater skill improvements $\Delta R^{A\text{-}M}$ than the assimilation of X-band retrievals (AMSR-E/AMSR2), although the sensitivity of the assimilation to the retrieval skill becomes weaker as the open-loop skill increases. The joint assimilation of L-band and X-band retrievals, relative to assimilation of either band product alone, does not necessarily yield better skill improvement.

6.2 Originality and contributions

This thesis explored for the first time the assimilation of SMOS, AMSR-E and AMSR2 soil moisture in the MESH model. As compared to previous studies, the present work has provided new results to demonstrate the impact upon soil moisture estimates of assimilating remotely sensed soil moisture in a hydrological model. The original contributions are as follows. (1) This work examined the potential of latest satellite soil moisture products (SMOS and AMSR2), through data assimilation, to improve soil moisture model estimates. (2) This work revealed the vegetation modulation of satellite soil moisture assimilation. (3) The assimilation of L-band retrievals (SMOS) was compared with the assimilation of X-band retrievals (AMSR-E/AMSR2), providing further insight into the dependence of the assimilation upon satellite soil moisture retrieval capability. (4) The influence of satellite-model skill difference $\Delta R^{S\text{-}M}$ on skill improvement $\Delta R^{A\text{-}M}$ was consistently demonstrated through the assimilation of soil moisture retrievals derived from radiometers operating at different microwave frequencies, different vegetation cover types, and different retrieval algorithms.

6.3 Future work

Overall, assimilation of satellite soil moisture in land surface/hydrological models is still in its infancy. There are some important issues that were outside the scope of this thesis and need to be considered in the future.

(i) *On-line quality control of satellite soil moisture*. In theory, the assimilation results should be always better than or at least match the modeling without assimilation (open-loop). In practice, however, negative skill improvement $\Delta R^{A\text{-}M}$ may occur when observation errors for very poor quality of satellite retrievals (the corresponding retrieval skill is typically very small or negative) are severely underestimated in the assimilation system. If on-line quality control routines (e.g. Reichle, 2008) are applied to the assimilation system, those inconsistent satellite observations, through cross-comparison between multiple observational sources and models, can be excluded for the assimilation, thus controlling the occurrence of negative $\Delta R^{A\text{-}M}$ and improving the assimilation performance.

(ii) *On-line (adaptive) tuning of model and observation input error parameters*. The assimilation estimates are subject to model and observation error variances, which are strongly controlled by the specified input error parameters in this work (although a non-adaptive EnKF typically performs well for soil moisture estimates, even when the input error parameters moderately deviate from their true values). The input error parameters can be on-line (adaptively) tuned based upon the variance information of the normalized

innovations, which is defined as the innovations (satellite observation minus model background residuals) divided by the square root of the sum of the forecast error covariance in observation space and the measurement error covariance (Reichle et al., 2007). For instance, if the variance of the normalized innovations is greater than 1, it means that the input error covariances are typically smaller than their optimal values that should lead to the expected value of 1 for the normalized innovations variance. Thus, by increasing the strength of the model error perturbations, our normalized innovations variance could draw close to its expected value of 1, which should, in theory, lead to better assimilation products. Reichle et al. (2008b), based on synthetic assimilation experiments, demonstrated the potential of an adaptive tuning of input errors to improve the assimilation estimates, although a long time period is typically required for steady estimates of model and measurement error variances.

(iii) *Establishment of a bias-aware assimilation system*. The satellite-model bias removal is an indispensable part in a bias-blind assimilation system (i.e. correct random errors only) since satellite retrievals and model surface soil moisture typically exhibit different climatologies, which impedes an optimal merging of the two data sets. As used in this study, a priori rescaling on the retrievals by matching their CDF to the model surface soil moisture's CDF has the ability to eliminate the satellite-model bias (in terms of climatological mean and standard deviation). To deal with the bias problem, another type of idea is to establish a bias-aware system. Provided that we can attribute the systematic errors to proper sources, and they also can be represented, by design, using appropriate

parameters, the bias can be estimated jointly with the model state by adding the designed parameters to the state vector (i.e., a bias-aware system).

(iv) *Application of nonlinear/non-Gaussian filters*. The Kalman Filters (KF, EKF, and EnKF) directly estimate the posterior state using the Kalman analysis equation that is derived on the basis of the assumption of Gaussian statistics. Realistic applications have suggested that the Gaussian assumption-based methods work efficiently for the systems with PDFs being or close to Gaussian, but are nonetheless subject to the influence of the PDF skewness (e.g., Bengtsson et al., 2003; Reichle et al., 2002b). A general nonlinear filter (nonlinear/non-Gaussian) is expected to be efficient when used for nonlinear dynamical models with considerably skewed PDF for the predicted errors (e.g. multi-modal structure) (Miller et al., 1999; Han and Li, 2008). Nonlinear/non-Gaussian filters can be implemented through particle interpretations (e.g. Pham, 2001) or kernel approximation (Miller et al., 1999). For example, the Particle Filter (PF) uses a set of particles (sampled from a proposal distribution) with associated importance weights to approximate the posterior PDF. In theory, the PF is suitable for all types of systems (linear or nonlinear) and PDFs (Gaussian or non-Gaussian) due to the ability to track the full state space (Arulampalam et al., 2002). Some researchers have conducted satellite snow assimilation with the PF (e.g. Dechant and Moradkhani, 2011) and the preliminary results showed promising.

(v) *Impact of assimilating satellite soil moisture upon streamflow prediction*. This topic may be more attractive to the hydrological community. However, note that the contribution

of satellite soil moisture assimilation is expected to be weaker to runoff estimates than to soil moisture simulations (e.g. Reichle et al., 2008b). If the non-adaptive EnKF method is used, the estimation of streamflow is susceptible to the specified input error parameters. Wrongly specified input error parameters could severely degrade the flow estimates. Reichle et al. (2008b) demonstrated the potential of an adaptive tuning of input error parameters to improve streamflow estimates. Additionally, the model runoff estimates, for a given watershed, are influenced by all upstream grids that make up the watershed, while satellite soil moisture measurements are not always available, especially for the presence of dense vegetation cover. Therefore, the 1D-EnKF assimilation may not be conducted for all upstream grids, thus impacting the improvement in streamflow estimates.

(vi) *Assimilation of SMAP soil moisture*. The SMAP mission, which was launched during the preparation of this thesis (January 2015), represents the current state-of-the-art in satellite microwave soil moisture estimation. In particular, the simultaneous acquisition of an L-band radiometer (1.41 GHz) and an L-band radar (1.26 GHz) onboard SMAP will provide soil moisture products at different scales (radiometer-only, 36 km resolution; radar-only, 3 km and the fusion of radiometer and radar, 9 km). It is worthwhile to explore how SMAP retrievals of surface soil moisture, in comparison with the SMOS and AMSR2 products, through data assimilation, impacts the hydrological model soil moisture simulations, which can provide further insight into the dependence of the assimilation upon the satellite observation skill.

REFERENCES

Al Bitar, A., Leroux, D., Kerr, Y. H., Merlin, O., Richaume, P., Sahoo, A., et al. (2012). Evaluation of SMOS soil moisture products over continental U.S. using the SCAN/SNOTEL network. *IEEE Transactions on Geoscience and Remote Sensing*, 50, 1572-1586.

Albergel C., Rudiger C., Carrer D., Calvet J-C., Fritz N., Naeimi V., Bartalis Z., and Hasenauer S. (2009). An evaluation of ASCAT surface soil moisture products with in-situ observations in Southwestern France. *Hydrology and Earth System Sciences*, 13, 115–124.

Albergel, C., de Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., et al. (2012). Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. *Remote Sensing of Environment*, 118, 215-226.

Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2), 174−188.

Bartalis Z., Wagner W., Naeimi V., Hasenauer S., Scipal K., Bonekamp H., Figa J., and Anderson C. (2007). Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). *Geophysical Research Letters*, 34, L20401, doi:10.1029/2007GL031088.

Bates B. C., Kundzewicz Z.W., Wu S, and Palutikof J.P. (eds) (2008). *Climate Change and Water*. Technical Paper of the Intergovernmental Panel on Climate Change, IPCC Secretariat, Geneva, 210 pp.

Bengtsson, T., Snyder, C., and Nychka, D. (2003), Toward a nonlinear ensemble filter for high-dimensional systems. *Journal of Geophysical Research*, 108 (D24), 8775−8785.

Berg, A. A., and K. Mulroy (2006). Streamflow predictability given macro-scale estimates of the initial soil moisture status. *Hydrological Sciences Journal,* 51, 642–654, doi:10.1623/hysj.51.4.642.

Bindlish R., Jackson T. J., Wood E., Gao H., Starks P., Bosch D., and Lakshmi V. (2003). Soil moisture estimates from TRMM microwave imager observations over the Southern United States. *Remote Sensing of Environment*, 85: 507–515.

Brocca, L., Melone, F., Moramarco, T., and Morbidelli, R. (2009). Soil moisture temporal stability over experimental areas of Central Italy. *Geoderma*, 148(3-4), 364-374, doi:10.1016/j.geoderma.2008.11.004.

Brocca L., Melone F., Moramarco T., Wagner W., Naeimi V., Bartalis Z., et al. (2010). Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrology and Earth System Sciences* 14: 1881–1893.

Brocca L., S. Hasenauer, T. Lacava, F. Melone, T. Moramarco,W. Wagner, W. Dorigo, P. Matgen, J. Martínez-Fernández, P. Llorens, J. Latron, C. Martin, M. Bittelli (2011). Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe, *Remote Sensing of Environment* 115, 3390-3408.

Champagne, C., Berg, A., Belanger, J., McNairn, H., and De Jeu, R. (2010). Evaluation of soil moisture derived from passive microwave remote sensing over agricultural sites

in Canada using ground-based soil moisture monitoring networks. *International Journal of Remote Sensing,* 31, 3669-3690

Collow, T. W., A. Robock, and W. Wu (2014), Influences of soil moisture and vegetation on convective precipitation forecasts over the United States Great Plains, *J. Geophys. Res.*, 119, 9338–9358, doi:10.1002/2014JD021454.

Crosson W.L., Laymon C.A., Inguva R., and Schamschula M.P. (2002). Assimilating remote sensing data in a surface flux-soil moisture model. *Hydrological Processes*, 16: 1645-1662.

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

Crow W.T., Bindlish R., and Jackson T.J. (2005). The added value of spaceborne passive microwave soil moisture retrievals for forecasting rainfall-runoff ratio partitioning, *Geophysical Research Letters* 32, L18401, doi:10.1029/2005GL023543.

Crow W. and Zhan X. (2007). Continental-scale evaluation of remotely sensed soil moisture products. *IEEE Geoscience and Remote Sensing Letters* 4: 451-455.

Crow W.T. and Ryu. D. (2009). A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. *Hydrology and Earth System Sciences* 13:1–16.

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., de Rosnay, P., Ryu, D. & Walker, J. P. (2012). Upscaling sparse ground-based soil moisture

observations for the validation of coarse-resolution satellite soil moisture products. *Reviews of Geophysics*, 50, RG2002, doi:10.1029/2011RG000372.

Dechant C. and Moradkhani H. (2011). Radiance data assimilation for operational snow and streamflow forecasting. *Advances in Water Resources* 34: 351–364.

De Lannoy G.J.M., Reichle R.H., Houser P.R., Arsenault K.R., Verhoest N.E.C., and Pauwels V.R.N. (2010). Satellite-scale snow water equivalent assimilation into a high-resolution land surface model. *Journal of Hydrometeorology* 11: 352-369.

De Lannoy, G.J.M, Reichle R.H., Arsenault K.R., Houser P.R., Kumar S., Verhoest N.E.C., and Pauwels V.R.N. (2012). Multiscale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado. *Water Resources Research* 48, W01522, doi:10.1029/2011WR010588.

Dietz A.J., Kuenzer C., Gessner U., and Dech S. (2012). Remote sensing of snow - a review of available methods. *International Journal of Remote Sensing* 33:13, 4094-4134.

Draper C. S., Mahfouf, J-F, & Walker, J. P. (2009a). An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme. *Journal Geophysical Research* 114, D20104, doi:10.1029/2008JD011650.

Draper, C., Walker, J. P., Steinle, P., De Jeu, R. A. M., & Holmes, T. R. H. (2009b). An evaluation of AMSR-E derived soil moisture over Australia. *Remote Sensing of Environment*, 113(4), 703–710, doi:10.1016/j.rse.2008.11.011.

Draper, C. S., R. H. Reichle, G. J. M. De Lannoy, and Q. Liu (2012), Assimilation of passive and active microwave soil moisture retrievals, *Geophys. Res. Lett.*, 39, L04401, doi:10.1029/2011GL050655.

Drusch, M. (2007). Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set. *Journal of Geophysical Research,* 112, D03102, doi:10.1029/2006JD007478.

Dunn, O. J., & Clark, V. A. (1969). Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association*, 64, 366-377.

Entekhabi, D., et al. (2010a). The Soil Moisture Active and Passive (SMAP) mission, *Proc. IEEE*, 98(5), 704–716, doi:10.1109/JPROC.2010.2043918.

Entekhabi, D., Reichle, R. H., Koster, R. D., & Crow, W. T. (2010b). Performance metrics for soil moisture retrievals and application requirements. *Journal of Hydrometeorology*, 11, 832–840.

Evensen, G. (1994). Sequential data assimilation with a non-linear quasigeostrophic model using Monte-Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5), 10143-10162.

Evensen, G. (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343-367.

Fischer, E. M. et al. (2007), Soil moisture-atmosphere interactions during the 2003 European summer heat wave. *J. Clim*. 20, 5081-5099.

Francois C., Quesney A., and Ottle C. (2003). Sequential assimilation of ERS-1 SAR data into a coupled land surface–hydrological model using an extended Kalman filter. *Journal of Hydrometeorology* 4: 473-487.

Fujii,H., Koike,T., Imaoka,K. (2009). Improvement of the AMSR-E Algorithm for Soil Moisture Estimation by Introducing a Fractional Vegetation Coverage Dataset Derived from MODIS Data, *Journal of the remote sensing society of Japan*, 29: 282-292.

Gherboudj, I., Magagi, R., Goita, K., Berg, A. A., Toth, B., & Walker, A. (2012). Validation of SMOS data over agricultural and boreal forest areas in Canada. *IEEE Transactions on Geoscience and Remote Sensing*, 50, 1623–1635.

Gibbs, J. W. (1899). Fourier's series. *Nature*, 59, 606, doi:10.1038/059606a0.

Haghnegahdar, A., Tolson, B. A., Davison, B., Seglenieks, F. R., Klyszejko, E., Soulis, E. D., Fortin, V., and Matott, L., S. (2014). Calibrating Environment Canada's MESH Modelling System over the Great Lakes Basin. *Atmosphere-Ocean*, 52, 281-293, doi: 10.1080/07055900.2014.939131

Han, X., and Li, X. (2008). An evaluation of the nonlinear/non-Gaussian filters for the sequential data assimilation, *Remote Sensing of Environment*, 112:1434-1449.

Houser P.R., Shuttleworth W.J., Famiglietti J.S., Gupta H.V., Syed K., and Goodrich D.C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research* 34: 3405–3420.

Houtekamer, P. L., and H. L. Mitchell (1998). Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, 126, 796-811.

Jackson T.J. (1997). Soil moisture estimation using special satellite microwave/imager satellite data over a grassland region. *Water Resources Research* 33: 1475–1484.

Jackson T.J. (2005). Estimation of surface soil moisture using microwave sensors. *Encyclopedia of Hydrology* 54: 799-809.

Jackson, T. J., Cosh, M. H., Bindlish, R., Starks, P. J., Bosch, D. D., Seyfried,M., Goodrich, D. C., Moran, M. S., & Du, J. (2010). Validation of advanced microwave scanning radiometer soil moisture products. *IEEE Transactions on Geoscience and Remote Sensing*, 48(12), 4256–4272, doi:10.1109/TGRS.2010.2051035.

Jackson, T. J., Bindlish, R., Cosh, M. H., Tianjie, Z., Starks, P. J., Bosch, D.D., et al. (2012). Validation of soil moisture and ocean salinity (SMOS) soil moisture over watershed networks in the U.S. *IEEE Transactions on Geoscience and Remote Sensing*, 50, 1530–1543.

Jacobs J.M., Meyers D.A., and Whitfield B.M. (2003). Improved rainfall/runoff estimates using remotely-sensed soil moisture. *Journal of the American Water Resources Association* 39: 313–324.

Kerr, Y., Waldteufel, P., Richaume, P., Davenport, I., Ferrazzoli, P., & Wigneron, J. -P. (2008). SMOS Level 2 Processor Soil Moisture ATBD. Toulouse, France: CESBIO.

Kerr Y.H., Waldteufel P., Richaume P., Mahmoodi A., Wigneron J.P., et al. (2012). The SMOS soil moisture retrieval algorithm. *IEEE Transactions on Geoscience and Remote Sensing* 50: 1384-1403.

Kerr, Y. H., Waldteufel, P., Wigneron, J. -P., Martinuzzi, J., Font, J., & Berger, M. (2001). Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 1729–1735.

Kolberg S, Rue H, and Gottschalk L (2006) A Bayesian spatial assimilation scheme for snow coverage observations in a gridded snow model. *Hydrology and Earth System Sciences* 10:369–381.

Kouwen, N., E.D. Soulis, A. Pietroniro, J. Donald and R.A. Harrington. (1993). Grouped Response Units for Distributed Hydrologic Modelling. *Journal of Water Resources Planning & Management. ASCE*, 119(3): 289-305.

Kumar, S. V., Reichle, R. H., Harrison, K. W., Peters-Lidard, C. D., Yatheendradas, S., & Santanello J. A. (2012). A comparison of methods for a priori bias correction in soil moisture data assimilation. *Water Resources Research*, 48, W03515, doi:10.1029/2010WR010261.

Lau, W. K. M. & Kim, K.-M. (2012). The 2010 Pakistan flood and Russian heat wave: Teleconnection of hydrometeorological extremes. *Journal of Hydrometeorology*, 13, 392-403, 2012.

Li, B., Toll, D., Zhan, X., and Cosgrove, B. (2012), Improving estimated soil moisture fields through assimilation of AMSR-E soil moisture retrievals with an ensemble Kalman filter and a mass conservation constraint, *Hydrol. Earth Syst. Sci.,* 16, 105-119, doi:10.5194/hess-16-105-2012.

Li Z, Tang R, Wan Z, et al. (2009). A review of current methodologies for regional evapotranspiration estimation from remotely sensed data. *Sensors* 9: 3801–3853.

Liston G. and Hiemstra C.A. (2008). A simple data assimilation system for complex snow distributions (SnowAssim). *Journal of Hydrometeorology* 9: 989-1004.

Liu, Q., Reichle, R. H., Bindlish, R., Cosh, M. H., Crow, W. T., de Jeu, R. A. M., De Lannoy, G. J. M., Huffman, G. J., & Jackson, T. J. (2011). The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system. *Journal of Hydrometeorology*, 12, 750-765, doi:10.1175/JHM-D-10-05000.1.

Loew, A., and Mauser, W. (2008). On the disaggregation of passive microwave soil moisture data using a priori knowledge of temporally persistent soil moisture fields. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 819–834, doi: 10.1109/TGRS.2007.914800.

Loumagne C., Normand M., Riffard M., Weisse A., et al. (2001). Integration of remote sensing data into hydrological models for reservoir management. *Hydrological Sciences Journal (Journal des Sciences Hydrologiques)* 46: 89-102.

Mahfouf, J., Brasnett, B., & Gagnon, S. (2007). A Canadian precipitation analysis (CaPA) project: Description and preliminary results. *Atmosphere-Ocean*, 45, 1-17.

Mailhot, J., Bélair, S., Lefaivre, L., Bilodeau, B., Desgagné, M., Girard, et al. (2006). The 15-km version of the Canadian regional forecast system. *Atmosphere-Ocean*, 44, 133-149. doi:10.3137/ao.440202

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

Martinez-Fernandez, J., & Ceballos, A. (2005). Mean soil moisture estimation using temporal stability analysis. *Journal of Hydrology*, 312, 28-38, doi:10.1016/j. jhydrol.2005.02.007.

Mattia F., Satalino G., Pauwels V.R.N., and Loew A. (2009). Soil moisture retrieval through a merging of multi-temporal L-band SAR data and hydrologic modeling. *Hydrology and Earth System Sciences* 13: 343-356.

Maurer, E. P., and Lettenmaier, D. P. (2003) Predictability of seasonal runoff in the Mississippi River basin. *J. Geophys. Res.* 108(D16), 8607, doi:10.1029/2002JD002555.

Meng, X., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111, 172-175.

Merlin O, Chehbouni A., Boulet G., and Kerr Y. (2006). Assimilation of disaggregated microwave soil moisture into a hydrologic model using coarse-scale meteorological data. *Journal of Hydrometeorology* 7: 1308–1322.

Merzouki A., McNairn H., and Pacheco A. (2011). Mapping soil moisture using RADARSAT-2 data and local autocorrelation statistics. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4: 128-137.

Miller R.N., Carter E.F., and Blue S.T. (1999) Data assimilation into nonlinear stochastic models. *Tellus. Series A, Dynamic Meteorology and Oceanography* 51: 167−194.

Moradkhani H., DeChant C.M., and Sorooshian. S. (2012. Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method. *Water Resources Research* 48, W12520, doi:10.1029/2012WR012144.

Ni-Meister, W., Houser, P. R., and Walker, J. P. (2006), Soil moisture initialization for climate prediction: Assimilation of scanning multifrequency microwave radiometer soil moisture data into a land surface model, *J. Geophys. Res.,* 111, D20102.doi:10.1029/2006JD007190.

Njoku E.G., Jackson T.J., Lakshmi V., Chan T.K., and Nghiem S.V. (2003). Soil moisture retrieval from AMSR-E. *IEEE Transactions on Geoscience and Remote Sensing* 41: 215-229.

Njoku, E. G. (2004). AMSR-E/Aqua L2B Surface Soil Moisture, Ancillary Parms, & QC EASE-Grids. Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center. doi:10.5067/AMSR-E/AE_LAND.002

Njoku, E. G.,& Chan, S.K. (2006). Vegetation and surface roughness effects on AMSR-E land observations. *Remote Sensing of Environment*, 100(2), 190–199, doi:10.1016/j.rse. 2005.10.017.

Owe, M., De Jeu, R. A. M., & Walker, J. (2001). A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. IEEE Transactions on Geoscience and Remote Sensing, 39(8), 1643–1654, doi:10.1109/36.942542.

Owe, M., de Jeu, R. & Holmes, T. (2008), Multisensor historical climatology of satellite derived global land surface moisture. *J. Geophys. Res.*, 113, F01002.

Pauwels R. N., Hoeben R., Verhoest N.E.C., De Troch F..P, and Troch PA (2002). Improvements of TOPLATS-based discharge predictions through assimilation of ERS-based remotely-sensed soil moisture values. *Hydrological Processes* 16: 995–1013.

Pham D. T. (2001). Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon Weather Rev.*, 129:1194–1207.

Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., Verseghy, D., Soulis, E. D., Caldwell, R., Evora, N., and Pellerin, P. (2007), Development of the

MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale*, Hydrol. Earth Syst. Sci.,* 11, 1279–1294.

Rango A. (1994). Application of remote sensing methods to hydrology and water resources. *Hydrological Sciences Journal* 39: 309-320.

Reichle R.H., McLaughlin D.B., and Entekhabi D. (2001a). Variational data assimilation of microwave radio brightness observations for land surface hydrology applications. *IEEE Transactions on Geoscience and Remote Sensing* 39: 1708-1718.

Reichle R.H., Entekhabi D., and McLaughlin D.B. (2001b). Downscaling of radio brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach. *Water Resources Research* 37: 2353-2364.

Reichle, R. H., Walker, J. P., Koster, R. D., & Houser, P. R. (2002a). Extended versus ensemble Kalman filtering for land data assimilation. *Journal of Hydrometeorology*, 3, 728-740.

Reichle R, McLaughlin D.B., and Entekhabi D. (2002b). Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review* 130: 103–114.

Reichle, R. H., & Koster, R. D. (2004). Bias reduction in short records of satellite soil moisture. *Geophysical Research Letters*, 31, L19501, doi:10.1029/2004GL020938.

Reichle, R. H., and R. D. Koster (2005). Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model. *Geophys. Res. Lett.*, 32, L02404, doi:10.1029/2004GL021700

Reichle, R., R. Koster, P. Liu, S. Mahanama, E. 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. (2008). Data assimilation methods in the Earth sciences. *Advances in Water Resources*, 31, 1411-1418.

Reichle, R. H., Crow, W. T., Koster, R. D., Sharif, H. O., & Mahanama S. P. P. (2008a). Contribution of soil moisture retrievals to land data assimilation products. *Geophysical Research Letters*, 35, L01404, doi:10.1029/2007GL031986.

Reichle, R. H., Crow, W. T., & Keppenne, C. L. (2008b). An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resources Research,* 44, W03423, doi:10.1029/2007WR006357.

Ridler, M.-E., Madsen, H., Stisen, S., Bircher, S., & Fensholt, R. (2014). Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetation atmosphere transfer model in Western Denmark. *Water Resources Research,* 50, 8962–8981, doi:10.1002/2014WR015392.

Rodell M. and Houser P.R. (2004). Updating a land surface model with MODIS-derived snow cover. *Journal of Hydrometeorology* 5: 1064-1075.

Rodríguez E., Navascués B., Ayuso J.J., and Järvenoja S (2003). Analysis of surface variables and parameterization of surface processes in HIRLAM. Part I: Approach and verification by parallel runs. *HIRLAM Technical Report* No. 58 pp. 1-52.

Seneviratne, S. I., T. Cortia, E. L. Davina,M. Hirschia, E. B. Jaegera, I. Lehnera, B. Orlowskya, and A. J. Teuling (2010), Investigating soil moisture–climate interactions in a changing climate: A review, *Earth Sci. Rev.,* 99(3–4), 125–161, doi:10.1016/j.earscirev.2010.02.004.

Seo D., Koren V., and Cajina N. (2003a). Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. *Journal of Hydrometeorology* 4: 627-641.

Seo D., Koren V., and Cajina N. (2003b). Real-time assimilation of radar-based precipitation data and streamflow observations into a distributed hydrological model. In: *Weather Radar Information and Distributed Hydrological Modeling (Proceedings of Symposium HS03 held during IUGG2003).* Sapporo, Japan, July 2003, IAHS Publ. No. 282:138-142.

Soulis, E. D., Snelgrove, K., Kouwen, N., Seglenieks, F., and Verseghy, D. L. (2000). Towards closing the vertical water balance in Canadian atmospheric models: Coupling of the land surface scheme CLASS with the distributed hydrological model WATFLOOD. *Atmosphere-Ocean*, 30, 251-269.

Soulis, E. D., Kouwen, N., Pietroniro, A., Seglenieks, F. R., Snelgrove, K. R., Pellerin, P., Shaw, D.W. and Martz, L.W. (2005), A framework for hydrological modelling in MAGS, in: Prediction in Ungauged Basins: Approaches for Canada's Cold Regions, edited by: Spence, C., Pomeroy, J.W. and Pietroniro, A., Canadian Water Resources Association, 119–138.

Tang Q., Gao H., Lu H., and Lettenmaier D.P. (2009). Remote sensing: hydrology. *Progress in Physical Geography* 33: 490-509.

Tang Q. and Lettenmaier D.P. (2010). Use of satellite snow-cover data for streamflow prediction in the Feather River Basin, California. *International Journal of Remote Sensing* 31: 3745-3762.

Talagrand, O., and P. Courtier (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation - Part I. Theory. *Quart. J. Roy. Meteor. Soc.*, 113, 1311–1328.

Taylor, C.M., R.A.M. de Jeu, F. Guichard, P.P. Harris, and W.A. Dorigo (2012). Afternoon rain more likely over drier soils. *Nature* 489 (7416), 423-426, doi:10.1038/nature11377.

Wagner, W., Naeimi, V., Scipal, K., De Jeu, R. A. M., & Martinez-Fernandez, J. (2007). Soil moisture from operational meteorological satellites. *Hydrogeology Journal*, 15, 121–131, doi:10.1007/s10040-006-0104-6

Walker J.P. and Houser P.R. (2001). A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations. *Journal of Geophysical Research* 106 (D11): 11,761–11,774.

Wang L. and Qu J.J. (2009). Satellite remote sensing applications for surface soil moisture monitoring: A review. *Frontiers of Earth Science in China* 3: 237–247.

Wolfson, N., Atlas, R., and Sud, Y. C. (1987). Numerical experiments related to the summer 1980 US heat wave, *Mon. Weather Rev*, 115, 1345–1357.

Xu, X., Li, J., & Tolson, B. A. (2014). Progress in integrating remote sensing data and hydrologic modeling. *Progress in Physical Geography*, 38, 464-498, doi: 10.1177/0309133314536583.

Zhang H., and Frederiksen, C. S. (2003). Local and nonlocal impacts of soil moisture initialization on AGCM seasonal forecasts: A model sensitivity study. *J. Climate.* 16, 2117-2137.

Zheng G. and Moskal L.M. (2009). Retrieving Leaf Area Index (LAI) using remote sensing: Theories, Methods and Sensors. *Sensors* 9: 2719-2745.

Zeng, X., Wang, B., Zhang, Y., Song, S., Huang, X., Zheng, Y., Chen, C., & Wang, G. (2014). Sensitivity of high-temperature weather to initial soil moisture: a case study using the WRF model. *Atmos. Chem. Phys.*, 14, 9623-9639.

Zhao, L., Yang, K., Qin, J., Chen, Y., Tang, W., Lu, H., & Yang, Z. (2014). The scale-dependence of SMOS soil moisture accuracy and its improvement through land data assimilation in the central Tibetan Plateau. *Remote Sensing of Environment*, 152, 345-355.

Appendixes

Table A1 Summary of data assimilation methods for land/hydrologic applications

Methods	Specifications	Examples
Direct insertion	Directly replace the model forecast with an observation	Rodell and Houser (2004); Tang and Lettenmaier (2010)
Statistical correction	Statistical characteristics (e.g. mean, standard deviation) of the modeled variables are adjusted to match those observed	Houser et al. (1998) Pauwels et al. (2002)
Successive corrections	Update the modeled variables at each grid based on the surrounding observations by multiple passes. The weight of an observation depends upon its distance to the model grid	Rodríguez et al. (2003)
Nudging	Add a nudging term, which is proportional to the model/observation difference, to the prognostic equations. The nudging term will force the integration of prognostic equations towards observations	Houser et al. (1998) Brocca et al. (2010)
Optimal interpolation	Observations are weighted according to known or estimated errors. Determine the optimum weight (gain) matrix using least squares so that the total analysis error is minimum	Houser et al. (1998) Liston and Hiemstra (2008)
Three dimensional variational (3DVAR)	Seek a state with the maximum likelihood by iteratively minimizing a cost function, which measures the misfit between the model simulations and observations. Dynamical constraints are included when minimizing the cost function	Seo et al. (2003a, 2003b)
Four dimensional variational (4DVAR)	Extension of 3DVAR to the time dimension. Being a smoothing algorithm; Seek an optimal fit of the model forecast to observations over an assimilation interval; The state estimation is affected by all the observations within the assimilation interval	Reichle et al. (2001a, 2001b)
Kalman Filter (KF)	A variance minimizing analysis in the framework of a sequential assimilation; Each assimilation cycle consists of two steps: the forecast step (the model state is integrated forward in time) and the analysis step (the model prediction is updated with observations); Explicit error covariance propagation; Valid only for linear systems	Walker and Houser (2001) Crow et al. (2005) Crow and Zhan (2007)
Extended Kalman Filter (EKF)	A nonlinear counterpart of the KF; A linearized and approximate error covariance is used; Being able to deal with some non-linear/Gaussian processes	Francois et al. (2003); Draper et al. (2009a)

(*continued*)

Table A1 (continued)

Methods	Specifications	Examples
Ensemble Kalman Filter (EnKF)	A Monte Carlo variant of the KF; The error statistics are represented by an ensemble of model states and the ensemble spread defines the error variance; The ensemble mean is the best estimate (assumption of Gaussian statistics)	Reichle et al. (2002b; 2007); Crow and Wood (2003); De Lannoy et al. (2010; 2012); Draper et al. (2012)
Ensemble Kalman Smoother (EnKS)	Similar to EnKF except that the time dimension is included; Being a smoothing variance-minimizing estimator	Crow and Ryu (2009)
Bayesian Filter	Seeks a posterior probability density function (PDF) at a current time given all the observations up to the current time based on Bayes theorem	Kolberg et al. (2006)
Particle Filter (PF)	A Monte Carlo importance sampling is used, and the posterior PDF of the model state is represented by a weighted sum of the particles that are sampled from a proposal distribution. Update the importance weights at the analysis step; Being a nonlinear/non-Gaussian filter	Moradkhani et al. (2012); Dechant and Moradkhani (2011)

Table A2 Summary of spaceborne active microwave sensors for soil moisture estimation

Sensor/Satellite	Period of Operation	Frequency	Polarization	Spatial Resolution	Repeat Cycle
SAR/ERS-1	1991-1999	5.3 GHz	VV	30 m	35 days
SAR/ERS-2	1995-2011	5.3 GHz	VV	25 m	35 days
ASAR/Envisat	2002-2012	5.3 GHz	VV/HH,HV/HH, VH/VV	30 - 1000 m	35 days
SAR/TerraSAR-X	2007-present	9.6 GHz	HH,VV,HV,VH	1 - 18 m	11 days
SAR/RADARSAT-1	1995-present	5.3 GHz	HH	8 - 100 m	24 days
SAR/RADARSAT-2	2007-present	5.4 GHz	HH,VV,HV,VH	3 - 100 m	24 days
SAR/JERS-1	1992-1998	1.3 GHz	HH	18 m	44 days
PALSAR/ALOS	2006-present	1.3 GHz	HH,VV,HV,VH	7 - 100 m	46 days
SCAT/ERS-1&2	1991-2011	5.3 GHz	VV	25 km / 50 km	3-4 days
ASCAT/Metop	2006-present	5.3 GHz	VV	25 km / 50 km	1-2 days
Radar/SMAP	Jan 2015-present	1.26 GHz	VV,HH,HV	3 km	2-3 days

Abbreviations: ERS, European Remote Sensing Satellite; SAR, Synthetic Aperture Radar; ASAR, Advanced Synthetic Aperture Radar; Envisat, Environmental Satellite; JERS, Japanese Earth Resources Satellite; PALSAR, Phased Array type L-band Synthetic Aperture Radar; ALOS, Advanced Land Observing Satellite; SCAT, Scatterometer; ASCAT, the Advanced Scatterometer; SMAP, the Soil Moisture Active Passive mission.

Table A3 Summary of spaceborne passive microwave sensors for soil moisture estimation

Sensor (Satellite)	Period of Operation	Frequency/Footprint Size (along track × cross track)	Polarization	Data Acquisition
SMMR (Nimbus-7)	1978-1987	6.6 GHz / 148 km × 95 km 10.7 GHz / 91 km × 59 km 18.0 GHz / 55 km × 41 km 21.0 GHz / 50 km × 38 km 37.0 GHz / 27 km × 18 km	H & V	Every other day
SSM/I (DMSP)	1987-present	19.3 GHz / 69 km × 43 km 22.0 GHz / 60 km × 40 km 37.0 GHz / 37 km × 29 km 85.5 GHz / 15 km × 13 km	H & V	Daily
TMI (TRMM)	1997-present	10.7 GHz / 63 km × 39 km 19.4 GHz / 30 km × 18 km 21.3 GHz / 28 km × 28 km 37.0 GHz / 16 km × 10 km 85.5 GHz / 7 km × 5.1 km	H & V	Daily
AMSR-E (Aqua)	2002-2011	6.9 GHz / 74 km × 43 km 10.7 GHz / 51 km × 30 km 18.7 GHz / 27 km × 16 km 23.8 GHz / 31 km × 18 km 36.5 GHz / 14 km × 8 km 89.0 GHz / 6 km × 4 km	H & V	1-2 days
AMSR2 (GCOM-W1)	2012-present	6.9/7.3 GHz / 62 km × 35 km 10.7 GHz / 42 km × 24 km 18.7 GHz / 22 km × 14 km 23.8 GHz / 26 km × 15 km 36.5 GHz / 12 km × 7 km 89.0 GHz / 5 km × 3 km	H & V	1-2 days
MIRAS (SMOS)	2010-present	1.4 GHz / ~ 43 km × 43 km	H & V	Every 1-3 days
Radiometer (SMAP)	Jan 2015-present	1.41GHz/ 47 km × 39 km	H, V, 3rd & 4th Stokes	Every 2 -3 days

Abbreviations: SMMR, Scanning Multichannel Microwave Radiometer; SSM/I, Special Sensor Microwave/Imager; DMSP, Defense Meteorological Satellite Program; TMI, Tropical Rainfall Measuring Mission (TRMM) Microwave Imager; AMSR-E, Advanced Microwave Scanning Radiometer for EOS; AMSR2, Advanced Microwave Scanning Radiometer 2; GCOM-W1, Global Change Observation Mission 1st - Water "SHIZUKU"; MIRAS, Microwave Imaging Radiometer with Aperture Synthesis; SMOS, Soil Moisture and Ocean Salinity satellite. SMAP, the Soil Moisture Active Passive mission.

Table A4 Summary of efforts to assimilate microwave soil moisture data into land/hydrologic models

Authors	Microwave data	Data period	Land/hydrologic models	Assimilation methods	Study region
Reichle et al.(2001a, b; 2002b)	Simulated L-band brightness temperatures based upon SGP97	June to July, 1997	A soil-vegetation-atmosphere transfer scheme (SVAT)	4D-VAR; EnKF	Central Oklahoma
Crosson et al. (2002)	Brightness temperatures by a ground-based SLMR during SGP97	23 June-16 July 1997	A land surface flux–soil moisture model SHEELS	KF	Central Oklahoma
Margulis et al. (2002); Crow and Wood (2003)	Brightness temperatures by Airborne ESTAR during SGP97 (resolution: 800 m)	16 of the 30 days (June 18-July 17, 1997)	Noah LSM; TOPLATS model	EnKF (updating with both point-and footprint-scale observations)	Central Oklahoma
Merlin et al. (2006)	Synthetic SMOS-type soil moisture data		A distributed SVAT model	EnKF	The Walnut Gulch Watershed, Arizona
Pauwels et al. (2002)	Soil moisture retrievals from SAR on ERS-1 and 2 (resolution: 25-30 m)	13 measurements between October 1995-February 2000	TOPLATS	Statistical correction	The Zwalm catchment, Belgium
Francois et al.(2003)	Soil moisture retrievals from SAR on ERS-1 (spatial resolution: 30 m)	25 measurements between 1995-1997	The two-layer conceptual hydrological model GRKAL	EKF	The Orgeval agricultural river basin, France
Reichle and Koster (2005); Ni-Meister et al. (2006)	SMMR (C-band) surface soil moisture retrievals (resolution: ~120 km)	January 1979-August 1987	Catchment LSM	EnKF	Global; Eurasian catchments
Reichle et al. (2007)	Soil moisture retrievals from AMSR-E (X-band) and SMMR (C-band)	June 2002-May 2006 (AMSR-E); January 1979-August 1987 (SMMR)	Catchment LSM	EnKF	Global

(*continued*)

Table A4 (continued)

Authors	Microwave data	Data period	Land/hydrologic models	Assimilation methods	Study region
Draper et al. (2009a)	Surface soil moisture retrievals from AMSR-E C-band observations (resolution: ~60 km)	Over 2006	The Interactions between Surface, Biosphere, and Atmosphere (ISBA) LSM	EKF	European domain
Li et al. (2012)	AMSR-E (X-band) soil moisture product (resampled spatial resolution: ~ 25 km)	2006-2007	Noah LSM	EnKF	The Little Washita watershed, Oklahoma
Brocca et al. (2010)	Soil wetness index (SWI) product from ASCAT (spatial resolution: 25 km)	January 2007-June 2009	A rainfall-runoff model MISDc	Update the modeled saturation degree using a nudging scheme	5 subcatchments of the Upper Tiber River in central Italy
Draper et al. (2012)	Surface soil moisture retrievals from AMSR-E (X-band) and ASCAT	January 2007 to May 2010	Catchment LSM	EnKF	CONUS and southeast Australia
Crow et al. (2005)	Soil moisture retrievals from TMI X-band observations	December 1997 - September 2002	An antecedent precipitation index (API) model	Updating API based upon TMI soil moisture with the KF	26 basins in the U. S. Southern Great Plains
Crow and Zhan (2007)	Soil moisture retrievals from dual- and single-polarization AMSR-E X-band brightness temperatures, ERS-1 and -2 SCAT backscattering coefficients, and GOES thermal observations, respectively	July 1, 2002 to December 31, 2005 (AMSR-E); January 1, 1997 to December 31, 2005 (SCAT); 2002–2004 growing seasons (GOES)	API model	Rainfall correction based upon remotely sensed soil moisture with the KF	CONUS
Crow et al. (2009)	AMSR-E (X-band) soil moisture retrievals	1 July 2002 to 31 December 2006	API model	Rainfall correction with the KF	CONUS
Crow and Ryu (2009)	Synthetic remotely sensed soil moisture retrievals by the SAC model		The Sacramento (SAC) hydrologic model	Simultaneously adjust the model state (EnKF or EnKS) and rainfall accumulations (KF)	MOPEX basins, USA

Abbreviations: KF, the Kalman Filter; EKF, the Extended Kalman Filter; EnKF, the Ensemble Kalman Filter; EnKS, the Ensemble Kalman Smoother; SLMR, S- and L-band Microwave Radiometer; TOPLATS, TOPMODEL-based land atmosphere transfer scheme; MOPEX, the Model Parameterization Experiment.

Table A5. List of in situ soil moisture measurements

Network	Site ID	Location	Latitude (°)	Longitude (°)	Vegetation cover	Soil depth
MAWN	ALB	Albion, MI	42.26	-84.77	Crop	10, 25cm
MAWN	ARL	Manton,MI	44.40	-85.29	Tree	10, 25cm
MAWN	BBC	Watervliet, MI	42.13	-86.27	Crop	0-30, 30-60 cm
MAWN	BAT	Laingsburg, MI	42.83	-84.37	Crop	0-30, 30-60 cm
MAWN	BLK	Bear Lake, MI	44.37	-86.16	Crop	0-30, 30-60 cm
MAWN	BEL	Belding, MI	43.11	-85.31	Crop	0-30, 30-60 cm
MAWN	SWM	Benton Harbor, MI	42.08	-86.36	Crop	0-30, 30-60 cm
MAWN	BNZ	Benzonia, MI	44.56	-86.12	Tree	10, 25cm
MAWN	BRS	Berrien Springs, MI	41.94	-86.38	Crop	0-30, 30-60 cm
MAWN	CAS	Cassopolis, MI	41.95	-85.99	Crop	10, 25cm
MAWN	CER	Ceresco, MI	42.22	-85.15	Crop	0-30, 30-60 cm
MAWN	CLT	Charlotte, MI	42.65	-84.92	Crop	10, 25cm
MAWN	CTH	Chatham, MI	46.34	-86.93	Tree	10, 25cm
MAWN	CLR	Clarksville, MI	42.87	-85.26	Crop	0-30, 30-60 cm
MAWN	CLD	Coldwater, MI	41.93	-85.11	Crop	0-30, 30-60 cm
MAWN	CMC	Commerce Twp., MI	42.60	-83.60	Crop	0-30, 30-60 cm
MAWN	CNT	Constantine, MI	41.83	-85.66	Crop	0-30, 30-60 cm
MAWN	DOW	Dowagiac, MI	42.02	-86.17	Tree	10, 25 cm
MAWN	MSU	East Lansing, MI	42.67	-84.49	Crop	0-30, 30-60 cm
MAWN	HTC	East Lansing, MI	42.71	-84.48	Crop	0-30, 30-60 cm
MAWN	ELD	Suttons Bay, MI	45.03	-85.67	Tree	10, 25 cm
MAWN	EPR	Eastport, MI	45.08	-85.34	Crop	10, 25 cm
MAWN	ELK	Elk Rapids, MI	44.84	-85.41	Tree	10, 25 cm
MAWN	EMT	Emmett, MI	42.99	-82.76	Crop	0-30, 30-60 cm
MAWN	ENT	Lakeview, MI	43.35	-85.18	Crop	0-30, 30-60 cm
MAWN	ESC	Escanaba, MI	45.86	-87.18	Tree	10, 25cm
MAWN	FGV	Fairgrove, MI	43.53	-83.49	Crop	0-30, 30-60 cm
MAWN	FEV	Fennville, MI	42.60	-86.16	Tree	10, 25 cm
MAWN	FLT	Flint, MI	43.02	-83.67	Crop	10, 25cm
MAWN	FRL	Freeland, MI	43.56	-84.02	Crop	0-30, 30-60 cm
MAWN	FRM	Fremont, MI	43.42	-85.96	Crop	0-30, 30-60 cm
MAWN	GAY	Gaylord, MI	45.03	-84.85	Crop	10, 25 cm
MAWN	GRJ	Grand Junction, MI	42.40	-86.07	Tree	10, 25 cm
MAWN	HRT	Hart, MI	43.74	-86.36	Crop	0-30, 30-60 cm
MAWN	HFD	Hartford, MI	42.23	-86.16	Crop	0-30, 30-60 cm
MAWN	PCC	Hastings, MI	42.54	-85.30	Tree	10, 25 cm
MAWN	HAW	Hawks, MI	45.30	-83.85	Crop	10, 25 cm
MAWN	KBS	Hickory Corners, MI	42.41	-85.37	Crop	5, 20 cm
MAWN	HDN	Clayton, MI	41.87	-84.23	Crop	10, 25 cm
MAWN	HVL	Hudsonville, MI	42.86	-85.89	Crop	0-30, 30-60 cm

MAWN	ITH	Ithaca, MI	43.32	-84.49	Crop	0-30, 30-60 cm
MAWN	KZO	Kalamazoo, MI	42.35	-85.59	Tree	10, 25 cm
MAWN	KAL	Kalkaska, MI	44.66	-85.07	Crop	10, 25 cm
MAWN	KLR	Hartford, MI	42.18	-86.18	Tree	0-30, 30-60 cm
MAWN	KCT	Kent City, MI	43.25	-85.77	Crop	10, 25 cm
MAWN	KWD	Kewadin, MI	44.95	-85.36	Crop	10, 25 cm
MAWN	KND	Kinde, MI	43.92	-83.01	Crop	10, 25 cm
MAWN	LPR	Lapeer, MI	43.08	-83.31	Crop	0-30, 30-60 cm
MAWN	TPD	Paw Paw, MI	42.21	-85.96	Tree	10, 25 cm
MAWN	LAW	Lawton, MI	42.16	-85.83	Crop	10, 25 cm
MAWN	LES	Leslie, MI	42.47	-84.46	Crop	10, 25 cm
MAWN	LIN	Kawkawlin, MI	43.72	-84.03	Crop	0-30, 30-60 cm
MAWN	LDT	Ludington, MI	43.90	-86.38	Crop	0-30, 30-60 cm
MAWN	MCB	McBain, MI	44.19	-85.13	Crop	10, 25 cm
MAWN	MCT	Mecosta, MI	43.57	-85.24	Crop	10, 25 cm
MAWN	MEN	Mendon, MI	41.98	-85.43	Crop	0-30, 30-60 cm
MAWN	MML	McMillan, MI	46.31	-85.65	Crop	10, 25 cm
MAWN	MGR	Munger, MI	43.56	-83.76	Crop	0-30, 30-60 cm
MAWN	NTH	Northport, MI	45.14	-85.65	Tree	10, 25 cm
MAWN	OLD	Old Mission, MI	44.93	-85.50	Crop	10, 25 cm
MAWN	OST	Kalamazoo, MI	42.30	-85.70	Tree	10, 25 cm
MAWN	PTR	Dundee, MI	41.93	-83.70	Crop	10, 25 cm
MAWN	PIG	Pigeon, MI	43.90	-83.27	Crop	0-30, 30-60 cm
MAWN	RVL	Frankenmuth, MI	43.40	-83.70	Crop	10, 25 cm
MAWN	ROM	Romeo, MI	42.78	-83.02	Crop	0-30, 30-60 cm
MAWN	SDK	Sandusky, MI	43.46	-82.84	Crop	10, 25 cm
MAWN	SCD	St. Joseph, MI	42.02	-86.43	Crop	0-30, 30-60 cm
MAWN	SHV	South Haven, MI	42.36	-86.29	Tree	10, 25 cm
MAWN	SPO	Sparta, MI	43.12	-85.76	Crop	0-30, 30-60 cm
MAWN	STV	Stephenson, MI	45.41	-87.61	Crop	10, 25 cm
MAWN	NWM	Traverse City, MI	44.88	-85.68	Tree	10, 25 cm
MAWN	VER	Verona, MI	43.81	-82.85	Crop	10, 25 cm
MAWN	WEO	West Olive, MI	42.97	-86.08	Crop	0-30, 30-60 cm
MAWN	LUX	Casco, WI	44.56	-87.65	Tree	10, 25cm
MAWN	EHB	Egg Harbor, WI	45.05	-87.26	Tree	10, 25cm
MAWN	NSW	Nasewaupee, WI	44.76	-87.51	Tree	10, 25 cm
MAWN	SRB	Sister Bay, WI	45.22	-87.07	Tree	10, 25 cm
MAWN	STB	Sturgeon Bay, WI	44.89	-87.37	Tree	10, 25 cm
MAWN	WJP	West Jacksonport, WI	45.01	-87.23	Tree	10, 25 cm
SCAN	2003	Wabeno, WI	45.47	-88.58	Tree	5, 10, 20, 50, 100 cm
SCAN	2011	Geneva, NY	42.88	-77.03	Tree	5, 10, 20, 50, 100 cm
SCAN	2073	Sunleaf Nursery, OH	41.80	-81.08	Tree	5, 10, 20, 50, 100 cm

FCRN	Borden, ON	44.32	-79.93	Tree	2, 5, 10, 20, 50, 100 cm