ABSTRACT

Title of Dissertation: NON-GAUSSIAN ENSEMBLE FILTERING AND ADAPTIVE INFLATION FOR SOIL MOISTURE DATA ASSIMILATION

Emmanuel C. Dibia
Doctor of Philosophy, 2024

Dissertation Directed by: Professor Xin-Zhong Liang
Department of Atmospheric and Oceanic Science

The forecast error distribution in modern day land data assimilation systems is typically modeled as a Gaussian. The explicit tracking of only the first two moments can be problematic when trying to assimilate bounded quantities like soil moisture that are more accurately described using more general parameterizations. Given this issue, it is worthwhile to test how performance of land models is affected when the accompanying data assimilation system abides by a relatively more relaxed set of underlying assumptions.

To study this problem, we perform experiments using the ensemble Kalman filter (EnKF) and rank histogram filter (RHF) to assimilate surface soil moisture content observations into the NASA Catchment land surface model. The EnKF acts as the traditional (Gaussian) standard of comparison whereas the RHF represents the novel and more general data assimilation method. An additional parameter of our tests is the usage of an adaptive inflation scheme that is only applied to the ensemble prior. This is done in an attempt to mitigate the negative effects of systematic deficiencies not accounted for by either filter. The examinations were carried out at a number of globally-distributed test locations, deliberately coinciding with sites used to validate NASA SMAP soil moisture retrieval products.

Initial comparisons of the two filtering approaches in a perfect model context show both filters to provide significant benefits to the soil moisture modeling problem, with the RHF edging out the EnKF as the more performant filter. The relative performance gain of the RHF was most noticeable with respect to bias mitigation metrics and to the surface-level anomaly correlation scores, an interesting result given that neither filter is formulated to explicitly accommodate a systematic bias. When additionally applying adaptive inflation, both filters showed improvement in skill but such improvements were not significant. The use of synthetic observations and lack of a bias correction implementation may have led to exaggerated results.

To address this concern, the experiments were performed again but using real observations from SMAP soil moisture retrievals, with in situ validation data proxying as truth. A robust bias correction scheme was used as well to more closely approximate practices used in operational settings. The RHF continues to show better metrics than the EnKF, but no longer in a statistically significant sense. A similar result was noted with respect to inflation usage. The most likely reason for this outcome is the low observation count. The findings obtained from the data assimilation experiments in this dissertation offer insight on how best to focus development efforts in soil moisture modeling and land data assimilation.

NON-GAUSSIAN ENSEMBLE FILTERING AND ADAPTIVE INFLATION FOR SOIL MOISTURE DATA ASSIMILATION

by

Emmanuel Dibia

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2024

Advisory Committee:
Professor Xin-Zhong Liang, Chair/Advisor
Dr. Rolf Reichle
Dr. Jeffrey Anderson
Professor James Carton
Professor Jonathan Poterjoy
Professor Michael Evans, Dean's Representative

Acknowledgments

I'd like to offer a special thanks to Liang and Rolf. Looking back on everything, I'm in awe at the support these two have thrown my way over the years.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

catdef	Catchment deficit
prmc	Profile soil moisture content
rzexc	Root-zone excess
rzmc	Root-zone soil moisture content
sfmc	Surface soil moisture content
srfexc	Surface excess
ubRMSE	Unbiased root-mean-square error
ACC	Anomaly correlation coefficient
EI	Adaptively Inflated EnKF
EnKF	Ensemble Kalman filter
International Energy Agency	IEA
IPCC	Intergovernmental panel on climate change
LSM	Land surface model
MAB	Mean absolute bias
NASA	National Aeronautics and Space Administration
NCALM	National Center for Airborne Laser Mapping
NIC	Normalized information contribution
NMP	National Mesonet Program
NOAA	National Oceanic and Atmospheric Administration
OI	Optimal interpolation
OL	Open loop ensemble run
R^2	Pearson's correlation coefficient
RHF	Rank histogram filter
RI	Adaptively inflated RHF
RMSE	Root-mean-square error
SMAP	Soil Moisture Active-Passive
TLM	Tangent linear model

Chapter 1: Introduction

The majority of earth's atmosphere is composed of oxygen (O_2 - 21%), nitrogen (N_2 - 78%), and argon (Ar - 0.9%), all of which are mostly transparent to solar and terrestrial radiation. Compounds like water vapor (H_2O), carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O) are some of the more notable trace gases making up the remaining 0.1%; each member of this minority group is nearly opaque with respect to the infrared portion of the electromagnetic spectrum, and it is this opacity that is responsible for what's known as the greenhouse effect, whereby some of the outgoing heat from the earth is captured and emitted back toward the land surface. Using black-body radiation calculations, it can be readily determined that the earth would be cooler by roughly 30 ^{o}C (or about 60 ^{o}F) without this feedback present. In addition to shielding us from harmful ultraviolet rays and providing essential temperature regulation, the atmosphere exerts a pressure that allows for the persistence of liquid water on the earth's surface; the atmosphere also plays a major role in the development of the weather that we experience. Such functions make it clear how relevant the atmosphere is to our quality of life and justify why it should be studied. The goal of this chapter is to explain why focusing our attention on the land surface, specifically soil moisture, can be relevant to such research efforts.

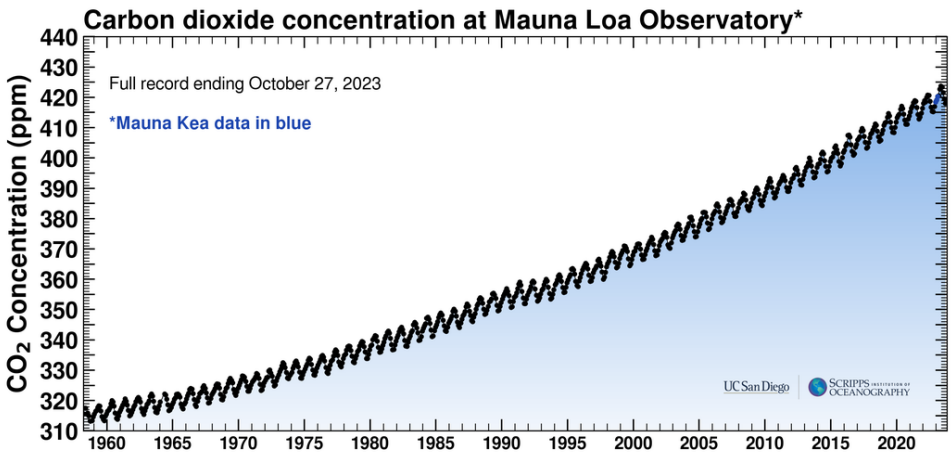

Figure 1.1: Keeling curve: Atmospheric concentration of CO_2 in parts per million (ppm) at Mauna Loa observatory in Hawaii. Superimposed on the general positive linear trend is the annual cycle, which is characterized by a Summer low and a Winter high. These relative differences can be partly explained by the fact that plants are most/least photosynthetically active during the Summer/Winter months. Figure is from Scripps Institution of Oceanography at University of California - San Diego.

1.1 Climate change

Pictured in Figure 1.1 is a recent realization of the Keeling Curve (Keeling and Keeling [1]). The Keeling Curve is one of the most important and well-known visualizations of the increase in atmospheric CO_2 concentrations over time. The sharp uptick observed in the accumulation rate coincides with the onset of modern-day human industrial activity, the effects of which can't be counteracted fast enough by the natural CO_2 sequestration mechanisms present in the carbon cycle (see Figure 1.2). The rising levels of CO_2 in the atmosphere contribute to the aforementioned greenhouse effect. This is a major driver of global climate change, with consequences including rising global temperatures, melting ice caps, sea-level rise, and shifts in weather patterns (Calvin et al. [2]).

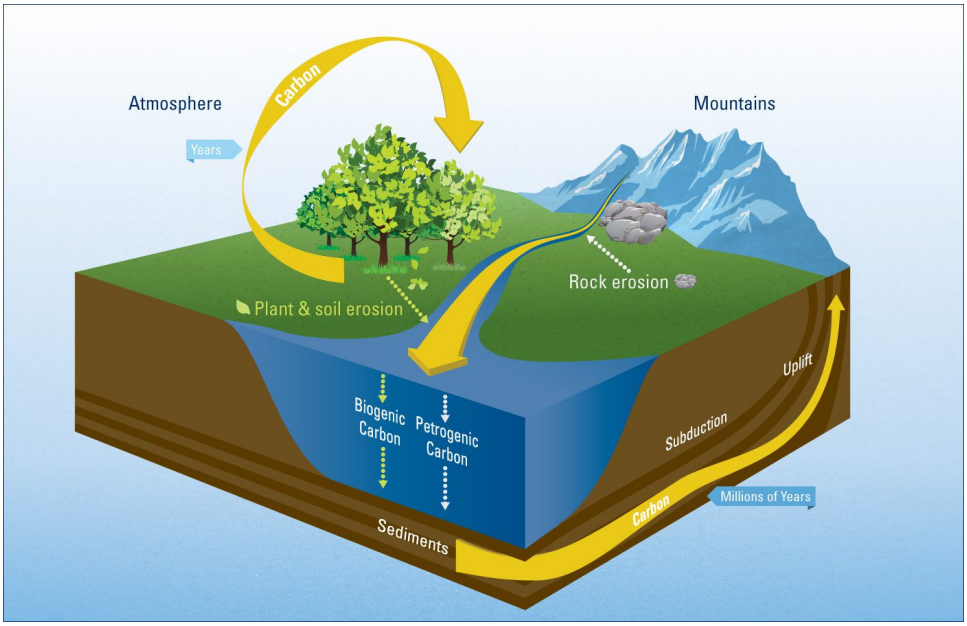

Figure 1.2: Carbon cycle: The carbon cycle is a dynamic process that helps regulate the Earth's climate by balancing the levels of carbon in different reservoirs. Human activities (not pictured), particularly the burning of fossil fuels and deforestation, have disrupted this natural cycle. Figure is from Woods Hole Oceanographic Institution.

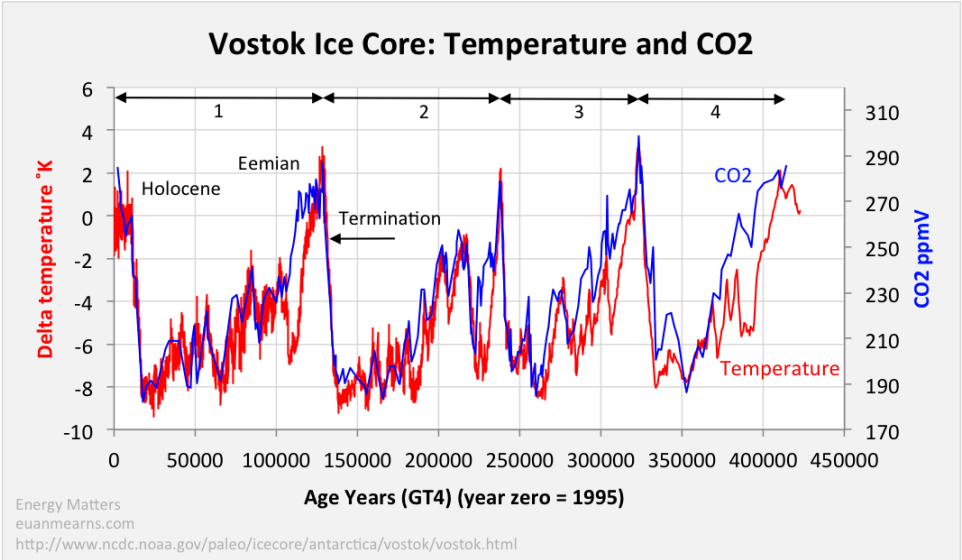

Figure 1.3: Relative temperature and CO_2 over time as derived from Vostok ice core data (Antarctica). Figure is from NOAA.

Direct measurements of CO_2 are limited to the relatively short time span of instrumental records. Paleoclimate researchers attempt to work around this limitation by using proxy data to produce reconstructions of not only CO_2, but also of different climatological variables of interest like sea level, temperature, and atmospheric composition (e.g., Grinsted et al. [3], Markwick [4]). The data leveraged to this effect includes tree rings, ice cores, sediment layers, and fossilized remains. One well-known example is the Vostok ice core record (Fischer et al. [5]), which shows that there is a strong correlation between global CO_2 levels and temperatures over geological timescales (see Figure 1.3). The key takeaway here is that paleoclimate datasets provide a baseline for understanding natural climate variability; this long-range perspective aids in being able to attribute observed changes in climate to human activities (Hansen and Sato [6]).

Historical data like the Keeling curve and Vostok ice core record are rather limited in terms of global spatiotemporal coverage. Despite this limitation, it is important that communities and

governments have access to climate information. Possessing such knowledge would better enable these groups to assess and manage risks associated with climate variability and change. The importance of adequate climate education is especially pressing for developing countries, for example, due to factors like limited economic resources that make it difficult to adapt to a changing climate. Climate education would ultimately lead to the development of a range of resilience-enhancing approaches. Such strategies could involve natural resource management practices that aim to preserve biodiversity of existing ecosystems[1]; the motivation for this point can be understood by noting that the biodiversity of our planet's ecosystems has direct implications on food security (Tscharntke et al. [10], Chappell and LaValle [11]) as well as on the regulation of water flow, quality, and availability (Olmstead [12], Srivastav et al. [13]). Another set of strategies to be incorporated for the sake of climate change adaptation might concern infrastructure, a component of which may focus on renewable energy, which is increasingly being adopted globally in its various forms in an effort to reduce the emission of harmful greenhouse gases (IEA; see Figure 1.4).

The preceding discussion highlights the usefulness of climate education and a number of coordinated data acquisition-related projects were also mentioned as evidence of this continued effort. Given the seriousness of climate change, it is natural to want to produce forecasts that attempt to quantify such changes at a future time. Climate models are crucial for integrating vast amounts of whatever observational data may be available into a single framework, testing different hypotheses, and exploring the potential impacts of various factors on the climate system, all of which can aid the climate science community in improving their understanding of the relevant

[1]There are a number of resources pledged to monitor biodiversity including the upcoming NASA Plankton, Aerosol, Cloud, Ocean Ecosystem mission (PACE; Werdell et al. [7]); additionally, other organizations like NCALM provide LiDAR data that can be used for various applications, biodiversity-focused ones included (OpenTopography [8], Catano and Stout [9])

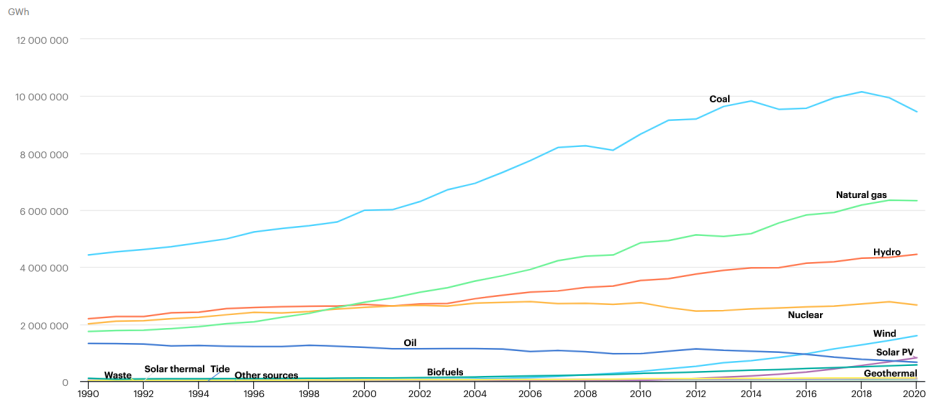

Figure 1.4: Global energy generation by source for renewables and nonrenewables from 1990-2020. Of note is the downward trend of global oil usage, especially in recent years. The variability in the coal line makes it difficult to speak convincingly of its specific longer-term trend direction. Generally, energy usage trends from renewable sources seem definitively positive. Figure is from International Energy Agency

processes. While climate modeling has provided invaluable insights into past and future climate scenarios, there are several challenges associated with it. For example, clouds and aerosols play crucial roles in the climate system, but they are among the most difficult components to model accurately. Similarly, climate models are increasingly incorporating biological and ecological processes, but these processes are still less well-understood compared to physical processes, so their inclusion introduces additional complexity and uncertainties (Ghan et al. [14]). The impacts of human behavior can also prove difficult to model, but such components are crucial to assessing the impacts of different mitigation and adaptation strategies that policymakers may be considering. In summary, the Earth's climate system is incredibly complex. Capturing all the relevant processes accurately in a model simulation is not straightforward, though continued effort in this regard may be helpful for improving our understanding of the problem.

Irrespective of limitations in data coverage and model quality, there is a widespread con-

sensus among climate scientists that our climate is indeed changing and that a significant part of this change is anthropogenic in nature. According to the 6th assessment report of the IPCC, modern-day human activity has already contributed to a global average warming of approximately 1.1 ^{o}C, with likely further increases projected over the coming decades if not enough is done to curb greenhouse gas output (Calvin et al. [2]). The bulk of anthropogenic CO_2 emissions come from the burning of fossil fuels, primarily for energy production (e.g., coal, natural gas, oil), transportation (e.g., gasoline and diesel fuels), and deforestation efforts (i.e., via release of stored carbon in vegetation and soils)[2].

1.2 Soil moisture as an environmental health indicator

The amount of moisture in the top-most portion of the soil column is a major player in deciding how incoming solar radiation is partitioned among the water and energy cycles. The nature of this partitioning has a significant impact on our day-to-day experience. Drier soils tend to have a lower albedo and less heat capacity; the relative lack of evapotranspiration means that the surface is less able to disperse heat, the result being a higher surface temperature. Such dry conditions could lead to the amplification of a heat wave instigated by a preexisting persistent high-pressure atmospheric system, for example; on the other end of the wetness spectrum, saturated soil conditions are more conducive to enhanced runoff and flooding events (Zhou et al. [15]).

Soil moisture also has direct implications on the ability of vegetation to grow and thrive. Changes in moisture availability can lead to a shift in vegetation type and distribution, potentially

[2]There are other greenhouse gases that are released by such activities, like CH_4 and N_2O, which tend to have much higher global warming potentials than CO_2, but are emitted in significantly smaller quantities.

impacting local climate patterns (Anderson and Goulden [16], Holsinger et al. [17]). One way to understand how plants achieve this is to recognize that they play a crucial role in carbon sequestration. Photosynthesis pulls CO_2 out of the atmosphere, which directly affects greenhouse gas concentrations (see Figure 1.2). [The water needed for photosynthesis also acts to couple the carbon and water cycles]. This carbon uptake occurs mostly during the warmer seasons, whereas being mostly inactive during the winter months due to frozen soils (see Figure 1.1). Climate change has been shown to result in longer growing seasons and less climatological support for the existence of permafrost soils (soil that remains continuously frozen at least 2 consecutive years) – these factors have competing implications for carbon sequestration levels [longer growing seasons suggest higher levels of temporally aggregated photosynthesis (higher net carbon sequestration), whereas if less carbon in dead organic matter is trapped in a frozen state, it is more readily available to act as fuel for erosion/weathering and wildfires (lower net carbon sequestration)].

In light of the preceding discussion on the anthropogenic origins of climate change, it is readily concluded that such feedbacks provide mechanisms for human activity to indirectly influence not only the atmosphere (e.g., atmospheric CO_2 volumes), but the soil moisture state as well (Seneviratne et al. [18]). But there are other more direct ways human beings can effect changes in soil moisture. Agriculture commonly relies on irrigation to supply water to crops. The water used for such efforts may be provided via groundwater extraction, leading to lowered water tables. Over-irrigation can lead to waterlogging and excess moisture, while under-irrigation can result in dry soil conditions. Upwards of one third of the global land surface is dedicated to agriculture (https://www.fao.org/sustainability/en/); that lower bound easily rises to well over 50% if only considering the United States (https://www.ers.usda.gov/). Urban areas that have

extensive pavement, concrete, and buildings have reduced natural infiltration of water into the soil. The result is increased surface runoff and potentially drier conditions in surrounding areas. Dams and reservoirs provide water storage and flood control, but this flow regulation also impacts the soil moisture state. These examples illustrate how human activities can directly influence soil moisture levels, which in turn can have cascading effects on ecosystems, agriculture, water resources, and other aspects of the environment in which we live. The arguments up to this point unequivocally communicate the usefulness of soil moisture as a sort of litmus test for different aspects of environmental health that encompass a number of the earth's systems, the atmosphere included. Naturally, given the information potential the soil moisture state offers, it is quite useful to understand how soil moisture is measured and modeled, and justified why it is worthwhile to reason about how these capabilities can be improved.

1.3 Soil moisture: observations and modeling

Soil moisture observations could be obtained from in situ measurements. In the United States, for example, there is the NMP (https://nationalmesonet.us), whose goal is to maintain a number of (observing) mesonet networks in support of NOAAs Weather-Ready Nation initiative. Nearly all mesonet networks in the United States record soil moisture data. An added benefit is that these soil moisture measurements are colocated with other variables like precipitation and surface winds, allowing for more comprehensive analyses (see Figure 1.5 for an example monitoring station). Globally, similar campaigns are in place to provide in situ soil moisture observations, and ongoing effort is being made to aggregate such records into databases for the public to access (Dorigo et al. [19]). A limitation of these observations is that each is only

of point-scale resolution, but it isn't practical to support arbitrarily dense arrays of monitoring stations. With this in mind, there is usually no other choice but to assume that an observation station can be placed at a point that may be considered representative of the surrounding area. This assumption is less valid in regions where there are high levels of variability in the vertical soil structure or in the overlying vegetation.

It is possible to measure soil moisture remotely (e.g., via aircraft or satellite). Remote sensing doesn't observe soil moisture directly, but rather infers it via microwave measurements (this band of the electromagnetic spectrum is sensitive to soil moisture). Such measurements can be passive or active, the active class having relatively higher spatial resolution. The benefit of remotely sensed soil moisture observations over in situ measurements is that remote sensing allows for global coverage. This global coverage is obtained repeatedly via frequent revisit times. Satellite observations don't make in situ observations obsolete; this is in large part due to the limited sensing depth of satellite retrievals (just a few centimeters). Contrarily, in situ observations are able to obtain a much deeper assessment of the vertical soil moisture profile (see Figure 1.5).

Land surface models simulate the flow of water and energy between the land and atmosphere by keeping track of water storage variables such as soil moisture and snow. Forecast skill is critically dependent on the quality of the meteorological forcing data used (precipitation especially) to drive the land surface dynamics. In order to help correct the deficiencies present in the forcing data or model core, observations of land surface variables (either remotely sensed or in situ) can be integrated into the model via different data assimilation methods. A commonly used soil moisture assimilation method is the ensemble Kalman filter (EnKF; Evensen [21], Burgers et al. [22]), a Monte Carlo-based method that represents the forecast uncertainty via a sample of model runs. Perturbations are applied to a subset of the meteorological forc-

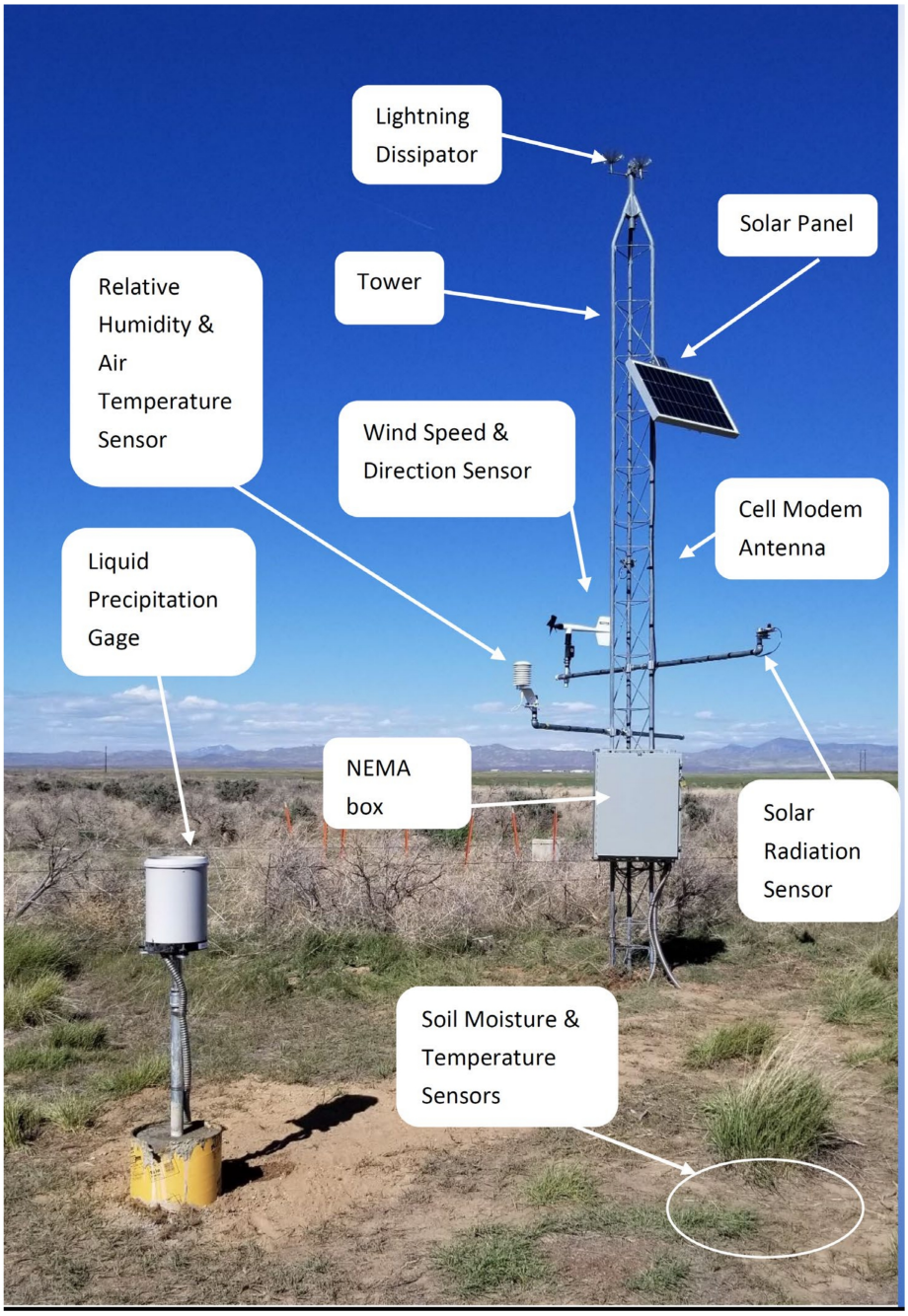

Figure 1.5: Example mesonet observation station maintained by the United States Department of Agriculture (USDA). Figure is from Caldwell et al. [20]

ing and state variables to generate the required ensemble. Some of the major advantages of the EnKF are its independence from model code, its flexibility with respect to model error treatment, and its appropriateness for nonlinear problems, like soil moisture dynamics (Pan and Wood [23], Andreadis and Lettenmaier [24], Durand and Margulis [25], Kumar et al. [26]). Despite these positive points, the EnKF is not completely general as it does indeed make limiting assumptions, namely that of Gaussian-distributed errors. The primary focus point for the remainder of this thesis concerns how the quality of soil moisture forecasts output by a land surface model depends on the different assumptions made by different data assimilation systems, specifically assumptions pertaining to how forecast errors ought to be parameterized (Gaussian versus non-Gaussian). The novel (non-Gaussian) data assimilation method to be tested on this point for comparison to the EnKF is the rank histogram filter (RHF; Anderson [27]).

1.4 Outline of thesis

Chapter 2 is an insert of Dibia et al. [28], which provides details of a perfect model experiment that is used to make an initial evaluation of different data assimilation methods, the EnKF and the RHF, when applied to the soil moisture modeling problem; the usefulness of adaptively inflating the ensemble prior will also be analyzed in this chapter. Chapter 3 performs the same evaluation as that done for Chapter 2, but instead using only real observations obtained from satellite retrieval data. Chapter 4 summarizes our findings and provides closing remarks and ideas for future work.

Chapter 2: Perfect model experiment: Improving soil moisture initialization with the RHF

2.1 Introduction

Subseasonal-to-seasonal forecasting at 2-week to 12-month lead times has become an increasingly relevant research area for the environmental modeling community due in large part to a stronger societal demand for reliable information about high-impact climate events (e.g., floods, droughts) further in advance of their onset. Several features in the Earth system have been investigated to determine to what extent they may be exploited as sources of predictability. One such source is the land surface, which directly modulates atmospheric flows, as indicated by the significant moisture and energy fluxes at the land-air boundary. Soil moisture variability can be regarded as a proxy for these interactions (Koster and Suarez [29], Seneviratne et al. [30]). Earlier studies have shown that satellite and in-situ soil moisture observations, when assimilated into a land surface model, improved the representation of soil moisture and other land surface states for short timescales (a few days) (Draper and Reichle [31]); within a coupled land- atmosphere modeling framework, improved land surface estimates resulted in more realistic evapotranspiration and precipitation (Koster et al. [32]). Earlier work has also demonstrated that enhanced soil moisture initialization has similar implications for the reproduction of large-scale atmospheric modes

with anomalies persisting on intradecadal and multidecadal timescales (e.g., Kuenzer et al. [33], Ford et al. [34]). Given a fixed set of soil moisture observations, we wish to examine how much useful information may be extracted by data assimilation systems having different levels of complexity.

Modern land surface assimilation has roots stemming from the Kalman filter (Kalman [35]), a state estimation method that takes into account the input uncertainties. The Kalman filter analysis is optimal if (i) the observation error and forecast error are mutually uncorrelated, unbiased, and white (temporally uncorrelated), (ii) the model state is linearly related to the observations, and (iii) the model dynamics are linear. Land surface models are driven by nonlinear processes, which suggests that an assimilation system would be more effective at producing improved state estimates if it did not rely so heavily on linearity assumptions. The extended Kalman filter (EKF; Evensen [36]) explicitly allows for the first-order treatment of weakly nonlinear model processes. The EKF requires the state-dependent linearization (or tangent linear model; TLM) of the model operator to propagate error estimates forward in time. The code in operational land models is not differentiable analytically (due to, for example, the presence of conditional statements), so the linearization must be determined numerically. This additional computational expense increases with the number of model variables that are included in the analysis state vector (Reichle et al. [37]). Deciding which variables to include and which to disregard when performing this computation requires a substantial amount of application-specific expertise to ensure that the resultant TLM approximation is useful. The EKF's use of closure approximations and its assumption of linear error growth may still lead to filter failure if nonlinearities are strong enough. One could simply increase the frequency at which the state update (or analysis) is performed or revise the TLM to include higher-order terms (e.g., Miller et al. [38]). Such engineering has been shown

to mitigate the threat of filter instabilities; however, the enhanced filter performance may not be enough to justify the extra computational burden.

The ensemble Kalman filter (EnKF; Burgers et al. [22]) entirely avoids the TLM and allows for the fully nonlinear propagation of the forecast errors. This is achievable because the complete probability distribution of the errors is approximated with a finite ensemble of model states at each update step. An added benefit of the EnKF is that it can more readily account for dynamic horizontal forecast error correlations (difficult for the EKF due to computational reasons) and permits the more careful treatment of model error. The EnKF has been shown to be capable in a broad range of applications (Houtekamer and Zhang [39]). In the context of soil moisture state estimation, it has been shown that the EnKF performs at least as well as the EKF when controlling for computational effort (Reichle et al. [37]). But even though the ensemble of the EnKF contains information about higher-order statistical moments that is advanced between analyses, the actual update equation explicitly considers only its mean and covariance, making the EnKF suboptimal for problems dominated by nonlinearity and/or non-Gaussianity.

Many other advanced data assimilation algorithms have been introduced that may be beneficial for approaching the nonlinear problems highlighted thus far. Such methods include reduced-rank filters (Verlaan and Heemink [40]), particle filters (Gordon et al. [41], Xiong et al. [42], Nakano et al. [43]), and the increasingly popular and broad suite of hybrid methods that consider input from both sequential and variational components when performing the analysis. Of particular interest here is the rank histogram filter (RHF) of Anderson [27], which is similar to the EnKF but does not assume that the ensemble prior and observation distributions are normal.

It has been demonstrated that the RHF can be a competitive alternative to the EnKF even for global numerical weather prediction problems (Anderson [27]). The models used in these

earlier examples have typically had dynamics meant to mimic the chaotic nature of a geophysical fluid such as the atmosphere or the ocean (Zhang et al. [44]). Land surface models, in contrast, are dissipative – any ensemble of initial perturbations will collapse toward a single common trajectory with time if something is not done to forcibly maintain its spread (Reichle et al. [45]). The dynamical regime of the model undoubtedly affects the performance characteristics of any ensemble data assimilation method. Therefore, a more comprehensive assessment of the RHF is needed that explores how the RHF behaves in a land surface application. The main objective of the present paper is to examine the performance differences between the EnKF and the RHF for soil moisture estimation within the context of a synthetic experiment using the NASA Catchment land surface model. A secondary objective is to assess the impact of adaptive error covariance inflation on the performance of the EnKF and RHF. We begin in Section 2.2 with a review of the two ensemble filter methods. Section 2.3 discusses the details of the land surface model and outlines the experimental design. Section 2.4 presents results, and Section 2.5 provides a conclusion.

2.2 Review of ensemble filtering methods

2.2.1 Ensemble Kalman filter (EnKF)

Given an ensemble of m model forecasts, $\mathbf{x}_i^f$, $i = 1, 2, ..., m$, each of dimension $n \times 1$, the EnKF updates the state, $\mathbf{x}_i$, of each ensemble member in the following manner:

$$\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t^f) \tag{2.1}$$

where the f and a superscripts are the forecast (prior) and analysis (posterior) quantities, respectively, corresponding to values before and after the update of the state estimate. Each member is given a unique $p \times 1$ observation vector, $\mathbf{y}_i$, obtained from a random draw from a normal distribution having a mean equal to the actual unperturbed observation vector, $\mathbf{y}$, and a covariance equal to the user-defined $p \times p$ observation error covariance matrix, $\mathbf{R}$. The $p \times n$ forward operator, $\mathbf{H}$, computes the observation-space equivalent of the model state. The innovation increment, $(\mathbf{y} - \mathbf{H}\mathbf{x}^f)$, is weighted by $\mathbf{K}$, the $n \times p$ Kalman gain (or influence matrix), which is given explicitly in Eq. (2.2).

$$\mathbf{K}_t = \mathbf{P}_t^f \mathbf{H}_t^T (\mathbf{H}_t \mathbf{P}_t^f \mathbf{H}_t^T + \mathbf{R}_t)^{-1} \tag{2.2}$$

where $\mathbf{P}^f$ is the $n \times n$ forecast error covariance matrix. The ensemble formulation of the analysis problem circumvents the need to ever represent $\mathbf{P}^f$ explicitly in computer memory as a matrix (Keppenne [46]). $\mathbf{P}^f$ can now be diagnosed as the sample covariance of the ensemble forecast as follows

$$\mathbf{P}^f = \mathbf{X}^f \mathbf{X}^{f,T} \tag{2.3}$$

where $\mathbf{X}^f$ is the normalized forecast perturbation matrix, the columns of which are formed by the normalized deviations of each ensemble member forecast from the ensemble mean:

$$\mathbf{X}^f = \frac{1}{\sqrt{m-1}} \left[(\mathbf{x}_1^f - \overline{\mathbf{x}}^f), \ \ldots \ , \ (\mathbf{x}_m^f - \overline{\mathbf{x}}^f) \right] \tag{2.4}$$

$$\overline{\mathbf{x}}^f = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}_i^f \tag{2.5}$$

The mean of the posterior ensemble, $\overline{\mathbf{x}}^a$, is the optimal state estimate, which, if we assume that all operators are linear, can also conveniently be written as follows in terms of the mean ensemble forecast and mean observation values.

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{K}(\overline{\mathbf{y}} - \mathbf{H}\overline{\mathbf{x}}^f) \tag{2.6}$$

Note that in Eq. (6), $\overline{\mathbf{y}}$ is the mean of the ensemble of $\mathbf{y}_i$, not the value of the actual observation, $\mathbf{y}$.

2.2.2 Rank Histogram filter (RHF)

The RHF is an ensemble filtering method that can accommodate any distribution for the ensemble prior or observation. Figures 6 and 7 in Anderson [27] provide useful schematics that show how the RHF works. Here, we will briefly describe its update procedure. To construct the ensemble prior, the RHF applies the forward operator, $\mathbf{H}$, to each member's state vector, $\mathbf{x}^f$, to compute the ensemble estimates of the observation. The values are then arranged in increasing order to partition the real number space into $m + 1$ intervals. The RHF assumes that the prior's cumulative density in each of these intervals is equal to $\frac{1}{m+1}$. Between members, the probability of the distribution is made uniform (constant); in the unbounded regions (the tails), the probability is set to match that of a Gaussian with variance equal to the ensemble variance and mean chosen such that the $\frac{1}{m+1}$ cumulative density condition is met. Here, the observation likelihood is specified as Gaussian with mean $\mathbf{y}$ and variance $\mathbf{R}$. To make the update computation less expensive, the observation likelihood is approximated as a piecewise linear function in between the values of each member in the prior distribution (the tails are left unchanged as partial Gaus-

sians). The prior and likelihood are then multiplied pointwise. After multiplication, the posterior ensemble is chosen from the resulting distribution by determining the points on the real number line that have cumulative densities satisfying $C_i = \frac{1}{m+1}$ for $i = 1, 2, ..., m$. Finally, each member's analysis increment (posterior minus prior) is regressed linearly onto all components of its corresponding analysis state vector. Available observations are assimilated serially, one at a time, which assumes that the observation errors are independent. Given the use of Gaussian tails, the range of possible state values is not constrained in the regions of the distributions outside of the ensemble. This is obviously not appropriate for bounded quantities. It is possible to force boundedness on the posterior ensemble as was mentioned by Anderson [27], but the subsequent regression may still result in physically impossible values for the unobserved quantities. (This problem is solved by the marginal adjustment RHF of Anderson [47], a recent modification of the RHF). As is discussed in Section 2.3, we impose extra-analysis physical consistency checks to reduce the harmful effects of unbounded posterior distributions. In summary, perhaps the most important feature of the RHF is that, unlike the EnKF, the RHF does not restrict the ensemble distribution of model states (prior or posterior) to be Gaussian. This novel aspect of the RHF is the focus of the present study.

2.2.3 Addressing systematic error with inflation

By design, the EnKF updates the magnitudes of the ensemble covariances during each analysis such that the norm of the error covariance matrix after the analysis is typically smaller and necessarily no larger than it was before the analysis; the user-defined observation error covariance matrix is usually kept static, meaning that, with successive analysis cycles, more and more

of the information provided by the observations will tend to be ignored and the data assimilation system will eventually stop altering the state estimate, a phenomenon known as filter divergence.

Filter divergence can often be avoided via appropriate use of a noise application scheme like the one employed during all of the land model integrations in this study (see Section 2.3). Such a strategy is an inflation method because it helps ensure that the ensemble spread does not get too small. Several different inflation formulations have been proposed in the literature, each being shown to generally improve ensemble representativeness and performance (e.g., Anderson and Anderson [48], Whitaker and Hamill [49]).

A common limitation of these inflation approaches is their use of static inflation parameters. A given parameter set may adequately inflate the ensemble for one time period but may over- or under-inflate the ensemble for another. Obtaining a set of manually tuned time-varying inflation parameters for large models is not feasible, prompting the introduction of adaptive techniques that leverage the real-time ensemble statistics to determine the appropriate inflation values. Such methods commonly rely on the ensemble's innovation (e.g., Anderson [50], Reichle et al. [51]). In particular, the inflation method of Anderson [52] computes the Bayesian update of a multivariate Gaussian inflation distribution after considering the impact of an observation on the expected innovation variance, which can be described as follows if considering the case of when a single observation is to be assimilated

$$E\left[(y - Hx^f)^2\right] = \lambda P^f + R \tag{2.7}$$

Here, P^f ,R , and λ, the inflation factor, are scalar quantities. It is further assumed that the observation errors and forecast errors are uncorrelated. Note that since λ is assumed to be a

normally-distributed variable, it can take on any real value so long as the expression above in Eq. (2.7) is satisfied; it could even be negative if the observation error is especially large (the restriction of to nonnegative values was externally imposed). Once an observation becomes available, the inflation factor in Eq. (2.7) is used to update the prior estimate of the inflation factor, the result of which is multiplicatively applied to the ensemble covariance. The adjusted ensemble then assimilates the observation in the usual manner. To prevent the auxiliary update of the inflation factor from diverging, it is damped toward unity after the analysis is performed

$$\lambda = 1 + f(\lambda - 1) \tag{2.8}$$

where λ is a damping factor.

As stated previously, Anderson [52] models the inflation factor using a normal distribution, where the forecasted inflation distribution is the most recent (damped) posterior inflation distribution. This study uses the adaptive inflation algorithm of Gharamti [53], which improves on that of Anderson [52] by instead modeling the inflation factor using an inverse- gamma distribution, yielding a more stable implementation that completely avoids negative values, and better combats overdispersion.

2.3 Land model and experiment setup

2.3.1 Land surface model

The land surface model used here is the Catchment model of the Goddard Earth Observing System (GEOS) modeling and assimilation framework (Koster et al. [54]). The catchment-based

approach considers the effects of subgrid heterogeneity on the horizontal structure of hydrological land surface processes. The model has been shown to be viable when used standalone (Reichle et al. [55]) as well as when coupled to a general circulation model (Gelaro et al. [56]). The model prognostic variables include soil temperature and heat content, snow water equivalent, and soil water. The latter are water excess and deficit variables that measure the total amount of water stored in the catchment and the departure of a layer's water content from the equilibrium vertical profile. The excess and deficit variables include the surface excess, root zone excess, and catchment deficit, which correspond to three nested layers of 0-5 cm, 0-100 cm, and 0 cm to the (spatially varying) bedrock depth, respectively. These soil water prognostic variables are used to diagnose the corresponding volumetric soil moisture values – surface moisture content (sfmc), root zone moisture content (rzmc), and profile moisture content (prmc). The distribution of rzmc is used to spatially partition a catchment fractionally into three different hydrological regimes: (i) saturated - ground surface completely saturated, (ii) transpiration- sufficient – surface not saturated but transpiration proceeds without stress or severe water limitation (moisture-stress evapotranspiration is not supported), and (iii) wilting - soil moisture is below the wilting point and therefore too dry for any transpiration to occur. Different parameterization schemes for the surface energy balance and soil water transfer are used in the different hydrological regimes, a distinction that is ignored in more conventional, layer-based land surface models.

2.3.2 Ensemble spread and stochastic perturbations

To maintain the ensemble spread, each member is treated during the model integration with regular noise applications (Reichle et al. [45], Reichle and Koster [57]). The noise varies

according to an autoregressive model of order 1 [AR(1)], which can be described as follows

$$u_i^t = a u_i^{t-1} + b\epsilon_i^t \tag{2.9}$$

$$a = \exp(-\Delta t / t_c) \tag{2.10}$$

$$b = \sqrt{1 - a^2} \tag{2.11}$$

ϵ is a random noise sample from a standard-normal distribution, Δt is the interval between noise applications, and t_c is the correlation timescale for the component of the state that will ultimately be perturbed. The formulation in Eq. (2.9) – Eq. (2.11) explicitly enforces a temporally correlated error. Once u_i^t is obtained, the state, x^t, is modified according to the following logic:

If x_i^t must be nonnegative, then

$$x_i^t = x_i^t \exp \frac{(S^2 + 1)}{2} + u_i^t \sqrt{(\ln(S^2 + 1))} \tag{2.12}$$

otherwise

$$x_i^t = x_i^t + S u_i^t \tag{2.13}$$

The expression in Eq. (2.12) uses the inverse transform method, which makes it possible to multiplicatively apply u_i^t after converting it to a sample of a lognormal distribution with mean 1 and standard deviation S. Multiplicative noise application is necessary for variables that cannot be negative (e.g., downwelling shortwave radiation, precipitation). S is a fixed parameter used to scale the variance of the unit-variance noise of Eq. (2.9) to a level consistent with the expected errors of the perturbed state variable. After the perturbations are applied, the updated quantities

are passed through a physical consistency check procedure, and, if necessary, are adjusted to remain within the physical limitations encoded in the model; this same check is applied at every call to the model core, regardless of whether a perturbation scheme is being used.

2.3.3 Experiment design

The domain for the synthetic experiment consists of 18 globally-distributed locations mainly clustered in North America and Europe (Reichle et al. [55]). The locations were selected because each is also a validation site for the NASA SMAP soil moisture products. As a group, the sites have a wide array of soil and vegetation properties that provide challenges to real-world land-assimilation efforts. Site selection was made in anticipation of future work that will rely on the in-situ measurements for validation of real-data assimilation experiments (please see Appendix A for validation site details). The synthetic experiment is carried out at each site independently from the others. MERRA-2 data (Gelaro et al. [56]) is used for the surface meteorological forcing inputs. The Catchment model parameters and boundary conditions are those used in Version 4 of the SMAP L4_SM system (Reichle et al. [55]).

Conducting the experiment in a synthetic environment allows for the fine-grained control of every aspect of the experiment. The potential usefulness of the results obtained in a real-world context is made more apparent when the parameters of the experiment are set in a sensible way. An important consideration is how to simulate model error, which can be difficult to characterize given that it is unknown in the real world. Studies focusing on land data assimilation have tried to account for model error in a synthetic framework by using distinct meteorological forcing datasets and/or model parameters for the truth and data assimilation runs (e.g., Reichle and

Koster [57]). Here, we follow a similar rationale and prescribe model error by requiring that the truth be generated using meteorological forcing for the years 2000-2009 while the data assimilation experiments are run with forcing corresponding to the years 2010-2019; additionally, the data assimilation ensembles were initialized with the 1 January 2010 model states taken from the end of the experiment period; stated differently, we are trying to estimate the truth given intentionally incorrect (or error-prone) meteorological forcing and initial conditions. A schematic of the approach is provided in Figure 2.1 and is explained in more detail below.

The Catchment model initial conditions were obtained as follows. Firstly, the model state was spun-up from a cold start by running the model four times through the 20-year period from 1 January 2000 through 31 December 2019. After this spinup, the model was run from 1 January 2000 through 31 December 2009. This 10-year period serves as the truth for the assimilation experiments (Figure 2.1). Daily sfmc observations of the truth were generated by perturbing the true value with a random sample of normally distributed noise with mean zero and standard deviation equal to 0.03 m^3 m^{-3} . The standard deviation was chosen to be roughly equal to typical errors in SMAP soil moisture retrievals (Colliander et al. [58]). Once the observations and truth states had been recorded, the model was reset to 1 July 2009 at which point a 100-member ensemble was initialized, each with an identical copy of the truth state at that time (Figure 2.1). To spin up the ensemble, each member was run for another six months through 31 December 2009 with a unique set of AR(1) perturbations applied to a subset of the meteorological forcing and prognostic variables. The perturbations were samples from either lognormal or normal distributions with mean 1 and 0, respectively. The timescale of the AR(1) correlation for each perturbation time series was chosen to provide a physically meaningful error timescale; there was no prescription of cross-correlated errors (see Table B1 in the appendix). The ensemble was then

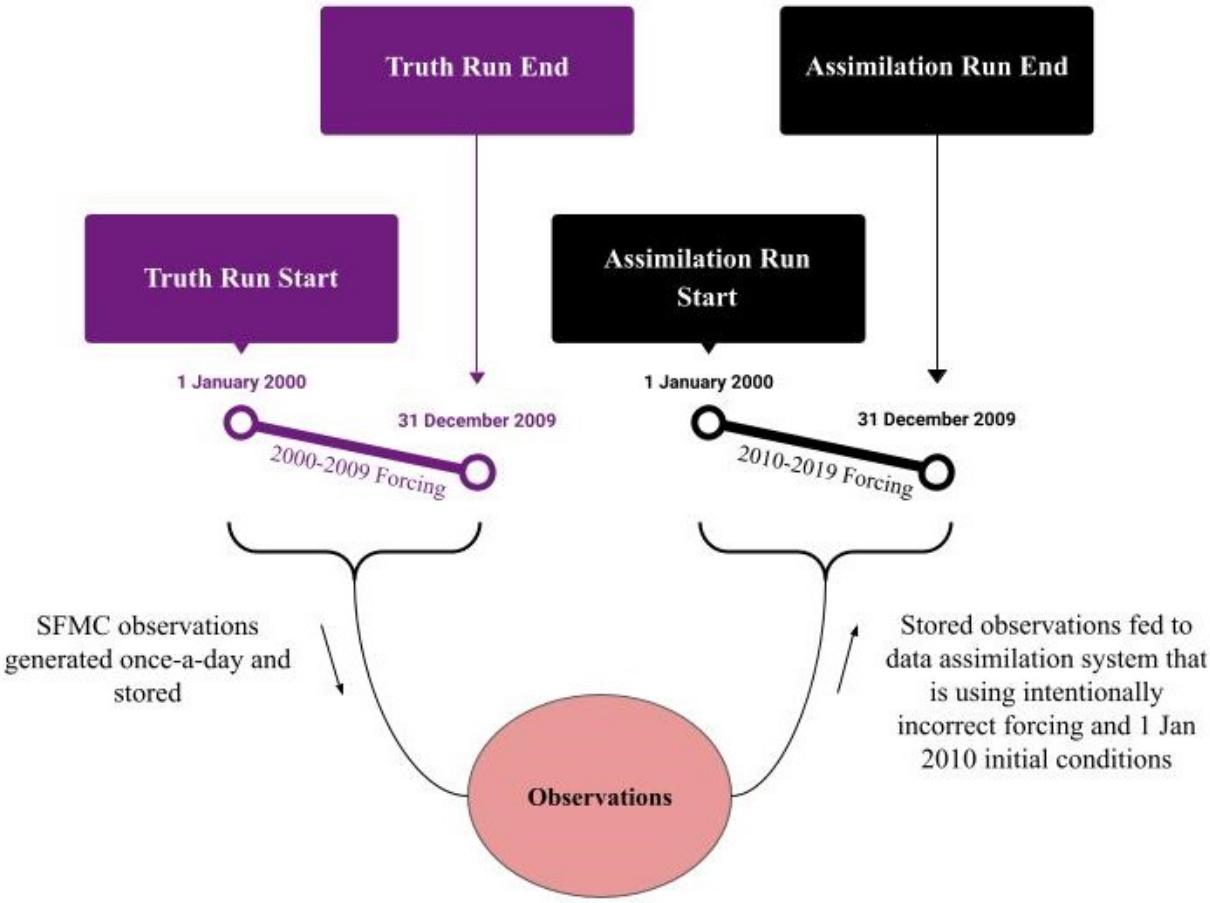

Figure 2.1: Schematic of the synthetic experiment showing the 10-yr offset of the meteorological forcing used to drive the truth run (single-member) and that used to drive the assimilation run (100-member ensemble).

run from 1 January 2000 through 31 December 2009, using the (intentionally wrong) 2010-2019 meteorological forcing data and 1 January 2010 ensemble of initial conditions. The ensemble trajectory during this ten-year period was recorded as the open loop simulation. The open-loop establishes the base-level performance of the modeling system without the benefits of soil moisture data assimilation. Finally, the data assimilation experiments consist of ensemble simulations that are identical to the OL but with the assimilation of the synthetic daily observations of surface soil moisture. To mimic the quality controls in place for real-world satellite soil moisture retrieval algorithms, soil moisture observations were not assimilated if any one of the ensemble members indicated that snow was on the ground or that the soil was frozen. After an analysis was performed, the updated quantities were passed through a physical consistency check procedure, and, if necessary, were adjusted to remain within the physical limitations encoded in the model.

There were four assimilation experiments performed, each one highlighting a different filter/inflation combination – (i) EnKF without adaptive inflation, (ii) RHF without adaptive inflation, (iii) EnKF with adaptive inflation, (iv) RHF with adaptive inflation. (Note that from here onward, 'inflation' is used to refer to the adaptive inflation algorithm of Gharamti [53] described in Section 2.2.3, and not to the general application of model prognostics and forcing perturbations outlined in Section 2.3.2). Each case used the exact same Catchment model, meteorological forcing, initial ensemble, and member-specific temporal sequence of AR1 perturbations and had access to the same set of sfmc observations.

The update of the state and of the inflation parameters is handled wholly by NCARs Data Assimilation Research Testbed (DART; Anderson et al. [59]), which allows for any discretized computer model to be interfaced with a suite of sequential ensemble data assimilation methods and common inflation techniques. In these experiments, the prior state vector provided to DART

is not the entire model state vector but rather a subset consisting of only the surface excess and root zone excess model prognostic variables, which matches the EnKF state vector of the SMAP L4_SM system (Reichle et al. [55]; their section 2.2.2). The adaptive inflation algorithm was set to inflate only the ensemble prior; the mean and standard deviation of the inflation distribution were initialized to 1.0 and 0.6, respectively, and were not allowed to be negative; before the inflation factor was updated and applied to the ensemble, the difference between the inflation factor and 1 was multiplied by a damping constant (equal to 0.9), the result of which was then added to 1 (see Eq. (2.8)).

2.3.4 Validation

2.3.4.1 Performance metrics and normalized information contribution

The impact on model skill of assimilating soil moisture observations was evaluated using several performance metrics, namely the anomaly correlation coefficient (ACC), the mean absolute bias (MAB), and the unbiased root mean square error (ubRMSE). Each metric was applied to the three diagnostic soil moisture content variables output by the model (sfmc, rzmc, prmc). The computation for the ACC is identical to that of the regular Pearson correlation coefficient but with the extra step of first removing the mean seasonal cycle from each of the two time series. The mean seasonal cycle was determined using the mean trajectory of the open loop (no meaningful difference in performance was seen when the climatologies of the respective experiments were used instead of the open loop mean). On a given day, the average value across all available years was taken from the mean open-loop run; the resulting time series (of 365 daily values) was then smoothed with an equal-weighted, 15-day average centered on the day of interest; leap-day

quantities (February 29th) were discarded in all input time series prior to the calculation of the anomaly correlation coefficient. The MAB was determined by taking the absolute value of the difference between two time series, then taking the average value of the result. The ubRMSE computes the root mean square error of two time series after having first subtracted the respective mean values from each.

Confidence intervals for the ACC were computed using a Fisher-Z transform of the anomaly correlation values. As in Draper et al. [60], the effective sample size of the input time series was reduced to account for inherent autocorrelations in the data. This consideration leads to progressively larger uncertainties for moisture variables with more temporal memory. The confidence intervals for the MAB and ubRMSE were computed using a two-tailed t-test with 10,000 replicates. The open loop estimates provide a baseline skill level without assimilation. For each performance metric, ν, the skill improvement gained from the assimilation relative to the open loop baseline can be measured as a normalized information contribution (NIC_ν; Kumar et al. [61])

$$NIC_\nu = \frac{\nu_{analysis} - \nu_{openloop}}{\nu_{target} - \nu_{openloop}} \tag{2.14}$$

Here, ν_{target} is the value corresponding to perfect performance for the given ν. For example, if looking at NIC_{ACC} , then ν_{target} would be equal to 1. The NIC_ν measures how well the analysis does when compared to the open loop run, where positive/negative values imply that data assimilation helped/hurt model performance relative to the open loop. A NIC_ν equal to 1 (the highest possible value, regardless of the particular ν) means that the analysis exactly matched the truth over the entire experiment, while a NIC_ν equal to 0 means that the analysis performed

just as well as the open loop (i.e., assimilating observations had no effect). A negative-valued NIC_ν indicates that assimilating observations led to a decline in skill. Here, the $\nu_{analysis}$ and $\nu_{openloop}$ quantities are regarded as being independent, meaning that an expression for the error in NIC_ν estimates can be obtained by summing the uncertainty components of Eq. (2.14) in quadrature.

$$\delta NIC_\nu = \sqrt{\left(\frac{\delta\nu_{analysis}}{\nu_{target} - \nu_{openloop}}\right)^2 + \left[\frac{\delta\nu_{analysis}\left(\nu_{analysis} - \nu_{target}\right)}{\left(\nu_{target} - \nu_{openloop}\right)^2}\right]^2} \tag{2.15}$$

The confidence intervals for the NIC scores are computed as the uncertainty in Eq. (2.15), δNIC_ν, multiplied by 2. The difference in any two values is considered statistically insignificant if the respective confidence intervals overlap.

2.3.4.2 Normalized Innovations

Innovation statistics will also be used to help characterize filter performance. The EnKF assumes that the observations and model estimates are unbiased and uncorrelated, meaning that a histogram of the innovations, after having been scaled with the inverse of the expected error standard deviation ($\mathbf{R} + \mathbf{HP}^f\mathbf{H}^T$), ought to closely approximate a standard normal distribution (zero mean and standard deviation of one). Given the synthetic nature of the experiment, the observation error $\mathbf{R}$ is known exactly (set as a scalar quantity equal to the square of 0.03 m^3 m^{-3}). So a normalized innovation distribution for the EnKF with standard deviation not equal to 1 suggests that the ensemble's simulated errors of the observation ($\mathbf{HP}^f\mathbf{H}^T$) are incorrect, which can only be due to sampling error, nonlinearity and/or a violation of the EnKF's basic assumptions. The RHF does not make bias or correlation assumptions, so there is not a known

target distribution for its corresponding normalized innovation histogram, but it will regardless be presented alongside that of the EnKF to further highlight any departure of the soil moisture assimilation problem from the EnKF's underlying assumptions.

2.4 Results and discussion

2.4.1 Soil moisture

Improved accuracy of the soil moisture state has the most potential to improve near-surface atmospheric flows in regions with strong land-atmosphere coupling (Koster and Suarez [29]). One such location is the Great Plains of the United States, where the Little Washita test site is positioned. For portions of the analysis that follow, we focus our attention on the Little Washita site, given that it is representative of where we would expect to see maximal impact of assimilating soil moisture observations. In some cases, we also illustrate results for the sites at Carman, St. Josephs, and Tonzi Ranch to highlight interesting aspects. We did not detect any meaningful general dependence of the assimilation performance on site-specific environmental features (soil properties, land cover type, climate) or the ensemble itself (e.g., innovations statistics). If such a signal is to be detected, a much larger number of sites is likely needed, which is beyond the scope of the present study.

Figure 2.2 shows the sfmc and rzmc time series at the Little Washita validation site during June, July, and August (JJA) 2009, the final year of the experiment period. The same plot was created for different years (not pictured) to see if there was a trend in the difference between the experimental configurations that were tested (e.g., EnKF versus the RHF), but no such trend was found; hence the year 2009 was chosen arbitrarily for presentation purposes. Pictured are

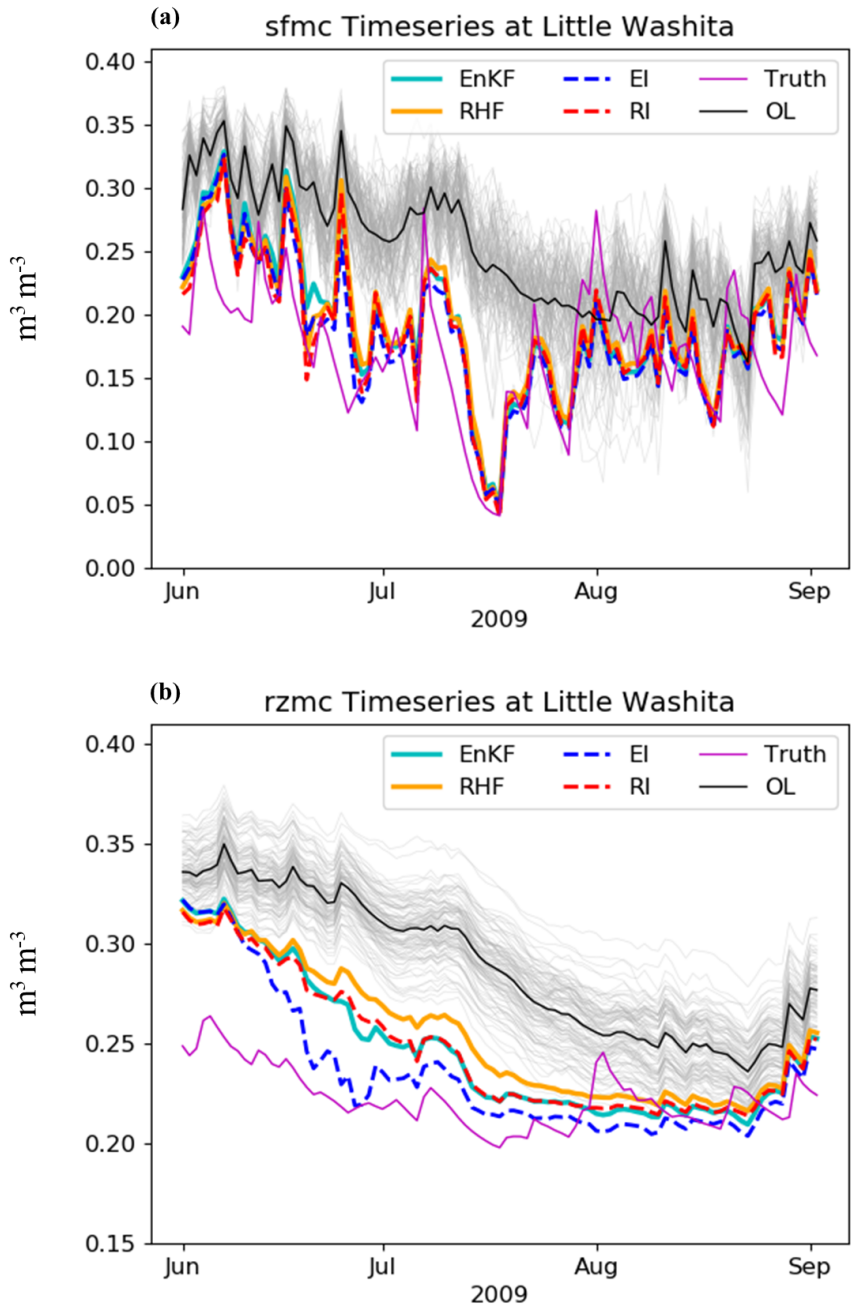

Figure 2.2: Time series of volumetric soil moisture content ($m^3\ m^{-3}$) for (a) sfmc and (b) rzmc at Little Washita during JJA 2009 (OL: open loop, EnKF: ensemble Kalman filter, RHF: rank histogram filter, EI: EnKF with adaptive inflation enabled, RI: RHF with adaptive inflation enabled).

the truth trajectory and the ensemble mean values of the open loop and data assimilation runs; also shown are the 100 individual members of the open loop ensemble (grey lines). The peaks in the soil moisture trajectories coincide with precipitation events and are typically followed by relatively slower dry-down periods, which are marked by little to no precipitation. Note that the open loop ensemble spread is small during precipitation events and increases during the dry- down periods. For both sfmc and rzmc, the mean of the open loop at this site is generally drier than the truth at the Little Washita site, meaning that the assimilation of sfmc observations should add moisture to correct the systematic dry bias coming from the model error specification (the incorrect meteorological forcing). Qualitatively, this was achieved for each of the four data assimilation runs as indicated by the smaller differences between their ensemble mean values and the truth (relative to the difference between the open loop and the truth). Also noticeable are the modest improvements in correlation with the truth. None of the filters responded particularly well to surface soil moisture observations following relatively strong precipitation events during this time period. There seems to be rapid peaking at the onset of the wetting episodes but not of a high enough magnitude. This last point can be explained by recalling that the analysis is effectively a weighted average of the model simulation (ensemble mean) and the assimilated observation, the weights being inversely proportional to the variance of the respective errors. The observation error was intentionally held fixed for the entirety of the experiment period whereas the ensemble estimate of the model forecast error (the ensemble spread) was changing over time. During heavy precipitation events, the model ensemble was increasingly confident that the soil was nearing saturation as evidenced by relatively small values of ensemble spread. As a result, the observation information is not given as much relative weight in the averaging process, sometimes leading to insufficient soil wetting. The severity of this effect is mitigated as the precipitation ceases

and the soil dries, allowing the ensemble spread to increase to more climatologically normal values (boosting responsiveness of the ensemble to information introduced by the assimilated observations).

Figure 2.3 shows the average NIC_{ACC} (Figure 2.3a), NIC_{MAB} (Figure 2.3b), and NIC_{ubRMSE} (Figure 2.3c) metrics for sfmc, rzmc, and prmc, along with their corresponding 95% confidence intervals. The EnKF had a positive NIC_{ACC} for each of the 3 soil moisture variables, indicating a systematic improvement over the open loop. The NIC_{ACC} scores are 0.51 for sfmc, 0.41 for rzmc, and 0.42 for prmc. The NIC_{MAB} for the EnKF was positive for sfmc and rzmc with respective values of 0.16 and 0.03 but was near- zero for prmc ($\tilde{0}$). The assimilation of soil moisture observations with the EnKF improved model skill in terms of the bias for sfmc but seems to have led to no real performance benefit for the deeper soil moisture reservoirs of the catchment (rzmc and prmc) relative to the open loop. The EnKF NIC_{ubRMSE} scores are all positive with a value of 0.32 for sfmc and a more modest 0.25 for both rzmc and prmc. Generally, the EnKF provides a clear net benefit over the open loop for model estimates of soil moisture.

The RHF had a NIC_{ACC} of 0.52 for sfmc, 0.41 for both rzmc and prmc, performing significantly better than the EnKF for sfmc and just as well for rzmc and prmc. For NIC_{MAB}, the RHF showed scores of 0.23, 0.12, and 0.08 for sfmc, rzmc, and prmc, respectively, performing significantly better than the EnKF in each case. Each of the NIC_{MAB} values for the RHF were positive, meaning that the RHF had a positive effect on model performance in terms of model bias relative to the open loop. The RHF had NIC_{ubRMSE} values of 0.33 for sfmc and 0.25 for both rzmc and prmc, effectively identical to those for the EnKF. In all cases, NIC_{ν} values for the RHF were positive and never significantly worse than those for the EnKF. The most notable difference between the two filters is seen for the bias, where the RHF shows itself to be the significantly

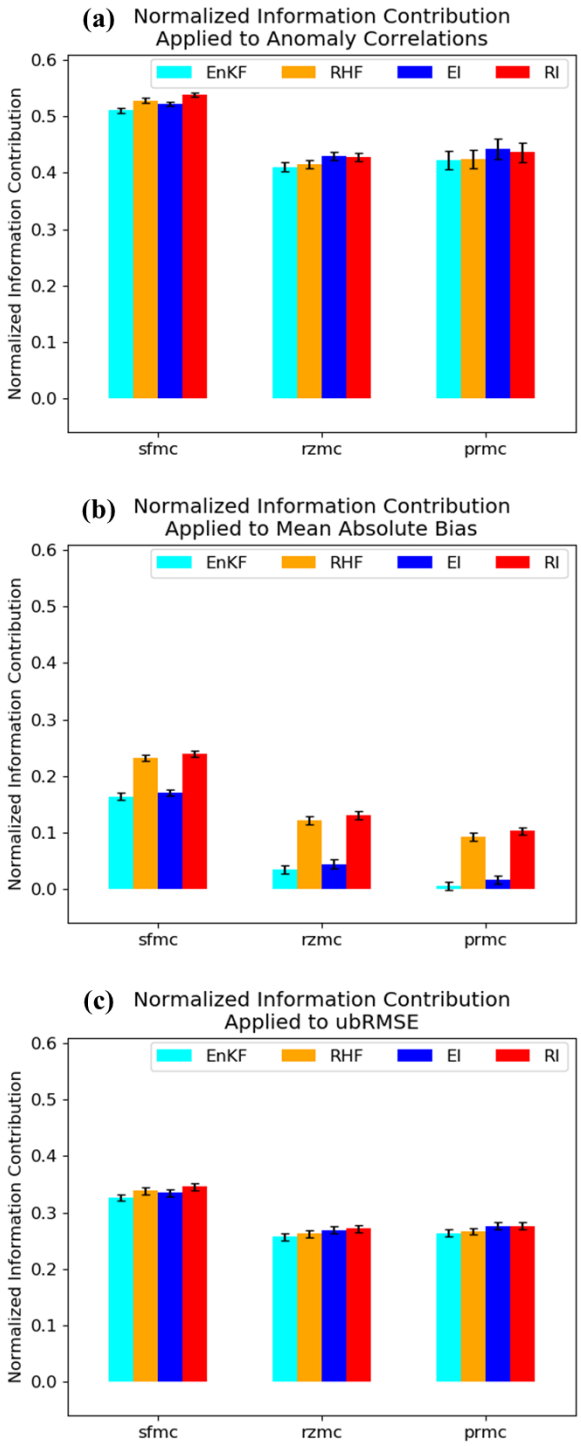

Figure 2.3: Normalized information contribution metrics for (a) anom- aly correlation coefficients, (b) mean absolute bias, and (c) unbiased root-mean-square error (ubRMSE). Metrics are averaged across all 18 validation sites. The entire 10-yr experiment period was consid- ered. Confidence intervals pictured are 95%.

better filter, increasing the NIC_{MAB} for each of the three soil moisture variables by 0.05 over the EnKF.

With the addition of adaptive inflation, the RHF achieved NIC_{ν} values at least as good as its non- inflated variant, but in only one case was that improvement statistically significant (NIC_{ACC} for sfmc). Similarly, the use of adaptive inflation with the EnKF always improved performance but significantly so in only two cases (NIC_{ACC} for sfmc and rzmc). In terms of NIC_{ACC}, the inflated RHF performed significantly better than all of the other filter experiments for sfmc, but such a difference did not translate to the unobserved soil moisture variables (rzmc and prmc), for which no one filter configuration showed itself to be best. The NIC_{MAB} and NIC_{ubRMSE} subpanels illustrate that inflation generally does improve the MAB and ubRMSE values but never in a statistically significant sense. The NIC_{ν} value magnitudes across all experiments and metrics indicate that the impact of soil moisture assimilation on skill was always positive and never negative. The largest improvements over the open loop were observed for the anomaly correlations (NIC_{ACC} ranging from 0.40 to 0.53) and the smallest skill improvements were for the bias (NIC_{MAB} ranging from 0 to 0.24). The metrics show that the RHF performed somewhat better than the EnKF and that adaptive inflation mostly results in minor skill improvements for both filters compared to the respective noninflated filter variants.

2.4.2 Ensemble representativeness

The rank histograms for the open loop and for each of the four assimilation experiments are shown in Figure 2.4 for a representative set of locations, including St Josephs, Carman, Tonzi Ranch, and Little Washita. The rank histograms were created by considering every prior ensem-

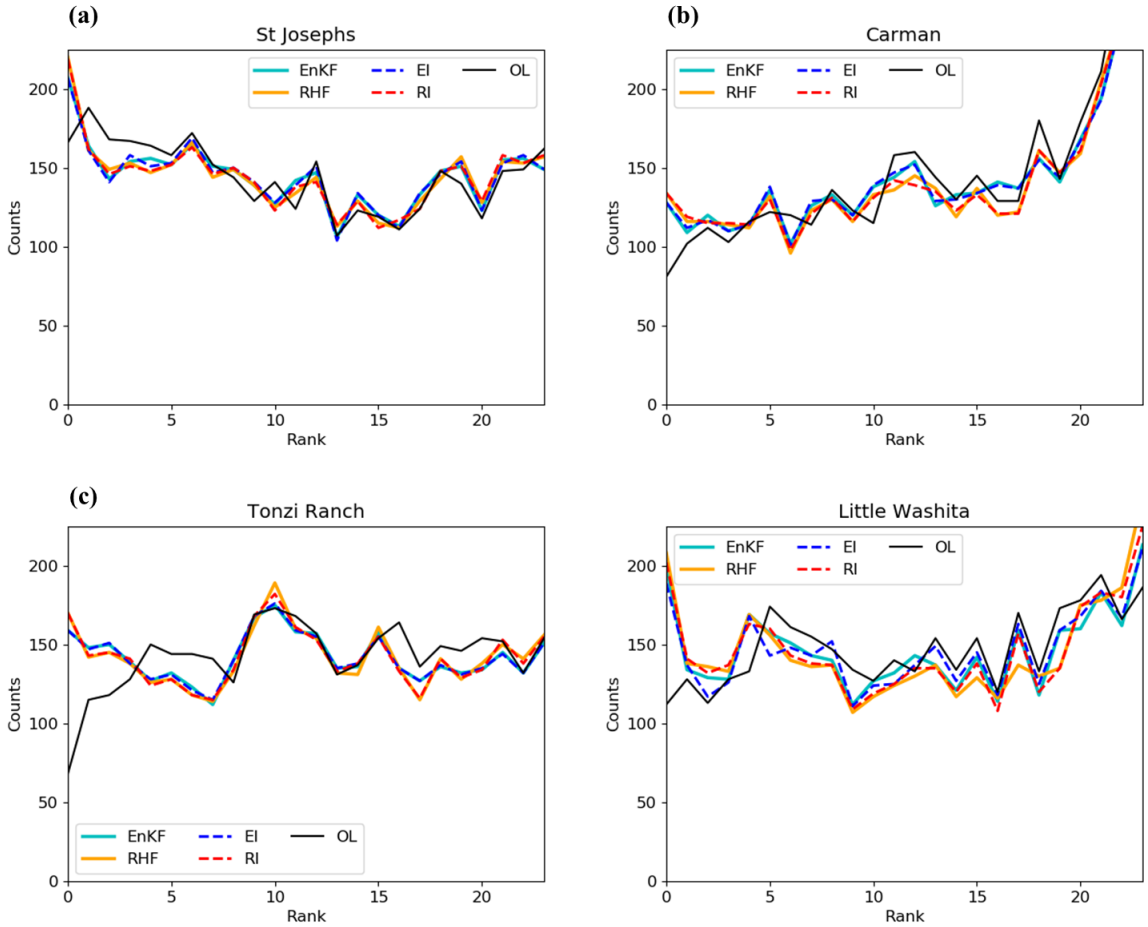

Figure 2.4: Rank histograms of SFMC observations for each assimilation experiment at (a) St. Josephs, (b) Carman, (c) Tonzi Ranch, and (d) Little Washita. All 10 years of available data were considered.

ble used over the experiment period, then ordering (or ranking) all the members of this augmented set from low to high according to the surface soil moisture state value; these bins are then populated using all the available observations (over the 10 years). The bins corresponding to each experiment are not identical, but the observations used to populate the bins are the same. To reduce noise in the illustrations, the counts of adjacent bins were averaged (the smoothing has no impact on the major features of the histograms). It is important to note that a rank histogram says nothing about whether the ensemble or the observations are representative of the actual system variable being estimated (soil moisture in this case). It simply provides a way to gauge if an ensemble is representative of the space spanned by the observations. One-to-one correspondence between the two would be evidenced by a flat line for a given experiment. None of the ensembles in Figure 2.4 has a histogram with such a horizontal profile. At St Josephs (Figure 2.4a), the rank histogram for the open loop shows some underdispersion (as evidenced by the gentle 'U' shape) and a forecast ensemble that is biased wet compared to the observations. Data assimilation does not seem to correct for either issue in any meaningful way as evidenced by the close tracking of the filter and open loop histogram profiles; in fact, data assimilation seems to exaggerate the bias (note the higher left edge of the histograms of the assimilation experiments compared to that of the open loop).

All the rank histograms at Carman (Figure 2.4b) also show biased and underdispersive ensembles but instead with forecasts that tend to be too dry with respect to the observations; there is no clear distinction between the open loop and the assimilation experiments at this location, except, again, at the left-most edge where the assimilation experiments show slightly wetter ensembles than for the open loop. The open loop rank histogram at Tonzi Ranch (Figure 2.4c) does not show an obvious global dispersion problem, but it does exhibit a forecast ensemble that is too

dry compared to the observations. The assimilation experiments reduce the severity of the bias (given by the relatively flatter outer left edge). At Little Washita (Figure 2.4d), the rank histogram profile for the open loop indicates a forecast ensemble biased wet compared to the observations; but the filter experiments show more general underdispersion with little to no bias (as evidenced by their roughly symmetrical 'U' shapes). The results imply that data assimilation leads to small changes in the ensemble's representativeness of the observation space; most of these changes are concentrated on the left side of the histogram figures, where assimilating the soil moisture observations clearly leads to generally wetter ensembles. The main aspects of the rank histograms do not vary coherently with filter type or whether inflation was used, nor do they seem to depend on climatology (dry versus wet).

Figure 2.5 shows time series of the open loop and ensemble prior standard deviations of surface and root zone soil moisture at the same four locations during JJA 2009. At St Josephs (Figure 2.5a), the standard deviation for the open loop ensemble had a JJA 2009 mean value of 0.026 $m^3\ m^{-3}$, ranging from 0.01 to 0.042 $m^3\ m^{-3}$ over the 3-month time period. The mean standard deviations were somewhat lower for the RHF (0.024 $m^3\ m^{-3}$) and higher for the EnKF (0.29 $m^3\ m^{-3}$), both having roughly similar ranges of 0.01 to 0.045 $m^3\ m^{-3}$. The time series for the inflated cases were identical to those of their respective noninflated counterparts. The variability of the time series does not show a clear dependence on experimental configuration. At Carman (Figure 2.5b), the time series for the open loop standard deviations shows a mean value of approximately 0.031 $m^3\ m^{-3}$ and a 3-month range of 0.012 to 0.045 $m^3\ m^{-3}$. The RHF standard deviation mean is 0.024 $m^3\ m^{-3}$ while that for the EnKF is 0.026 $m^3\ m^{-3}$. Inflation does not impact the mean for the EnKF but does lead to a small reduction for the RHF to a value of 0.023 $m^3\ m^{-3}$. Here, the time series corresponding to the data assimilation experiments do

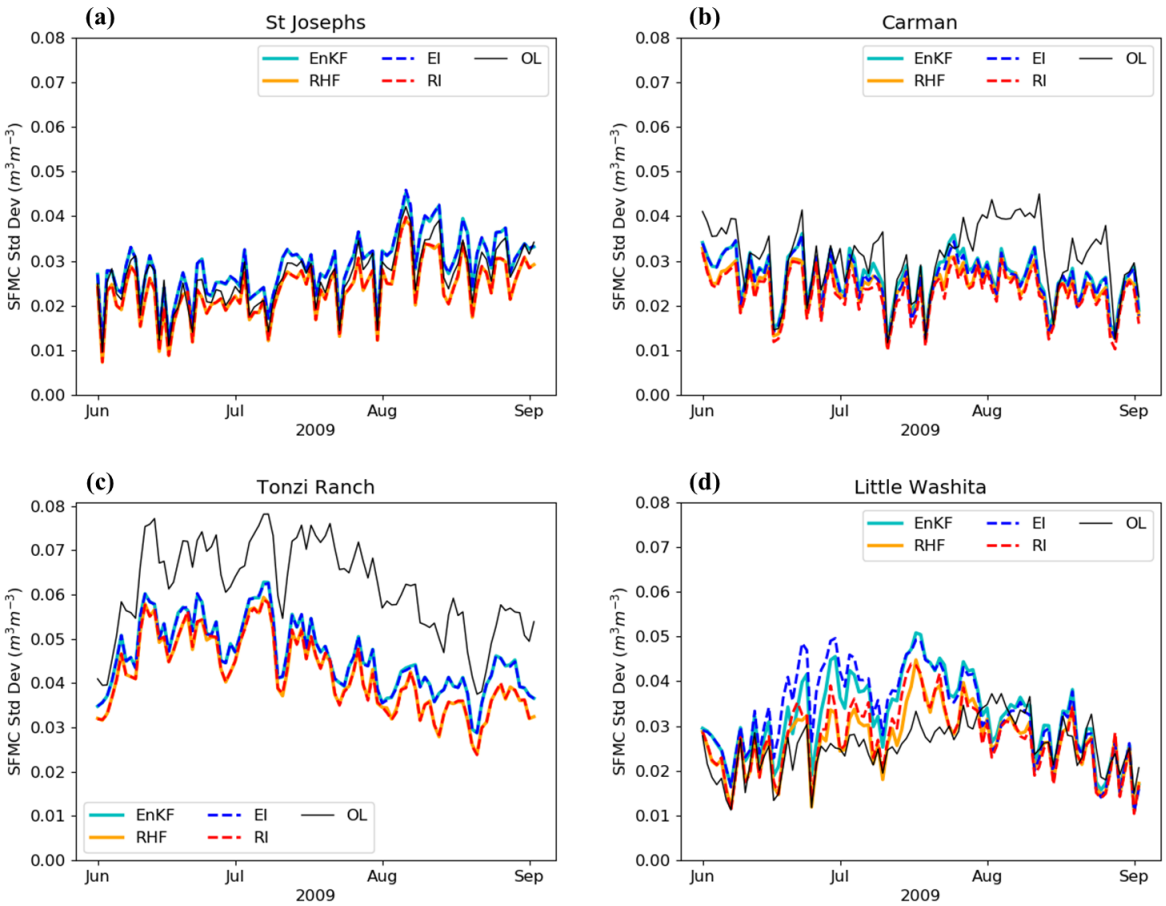

Figure 2.5: Time series examples of volumetric soil moisture content (m3 m23) standard deviation of data assimilation experiment ensembles (priors) and of the open loop ensemble for JJA 2009 at (a) St. Josephs, (b) Carman, (c) Tonzi Ranch, and (d) Little Washita.

not track the open loop case as closely (e.g., in early August) as was seen for St Josephs. The open loop at Tonzi Ranch (Figure 2.5c) had a mean standard deviation of 0.062 m^3 m^{-3} , with a range of 0.037 to 0.078 m^3 m^{-3} . The spreads in the ensemble priors are significantly lower than that for the open loop, with mean values of 0.046 m^3 m^{-3} for the EnKF and 0.042 m^3 m^{-3} for the RHF (the values were the same for the respective inflated variants of each filter). At Little Washita (Figure 2.5d), the ensemble standard deviations for the open loop showed a mean value of 0.026 m^3 m^{-3} (ranging from 0.010 to 0.038 m^3 m^{-3}). The EnKF variants are relatively less dispersed, having mean values of 0.032 m^3 m^{-3} and 0.033 m^3 m^{-3} for the non-inflated and inflated forms, respectively. The RHF variants track each other closely, with means of 0.027 m^3 m^{-3} and a common range of 0.011 to 0.046 m^3 m^{-3} .

Generally, the RHF tends to have smaller spreads than the EnKF (the same is true for the relative differences between the inflated variants); the EnKF variants, as we have previously shown, exhibit positive performance gains over the open loop, but the large standard deviations relative to the RHF can be interpreted as evidence that the EnKF successfully compensates for the slight inappropriateness of its assumptions by boosting the ensemble spread.

The ensemble spread in the inflated EnKF closely tracks that of the non-inflated EnKF, and the same is true for the RHF variants. In fact, at St. Josephs and Tonzi Ranch there is no discernible difference in ensemble spread between the inflated and non-inflated filter versions across the entire JJA 2009 period (Figure 2.5a,c), and only very small differences are seen at Carman and Little Washita (Figure 2.5b,d). It remains unclear why the inflation does not lead to larger differences in ensemble spread during JJA 2009.

Of note is that the standard deviations of the open loop are not necessarily always greater than or equal to those of the noninflated filter configurations, which seems nonsensical if one

neglects the boundedness of the sfmc problem being considered here; when an ensemble is especially wet or dry, its spread tends to collapse as the mean approaches the extremes. One example of this behavior can be seen clearly at Little Washita by comparing Figures 2.2a and Figure 2.5d, where the open loop shows significantly smaller standard deviations than the filters during time periods for which the filter ensemble sfmc values are moderate and the open loop ensemble has a relatively much wetter soil column. Across sites, it was also observed that inflation values tended to be closer to one during periods when the mean of the open loop was closer to the truth trajectory.

2.4.3 Measuring non-Gaussianity

Figure 2.6 illustrates the normalized innovation distributions of the assimilation experiments at the same representative locations as in Figure 2.4 and over the entire 10-year experiment period. Distributions for the EnKF and RHF are shown in the left column, and distributions for the filter variants with adaptive inflation are shown in the right column. For reference, the dotted line pictures the unit normal distribution (mean = skewness = 0, variance = 1), which is the distribution of the normalized innovations of the EnKF when all its assumptions are satisfied (see Section 2.1). Each of the distributions in Figure 2.6 is biased and skewed, clearly showing the systematic departure of the soil moisture assimilation problem considered here from those assumptions.

At St Josephs (Figures 2.6a and 2.6b), the dry biases for the non-inflated filters are the same at - 0.17; the standard deviations are greater than one (1.40 and 1.46 for the EnKF and RHF, respectively), suggesting that the ensembles are underestimating the expected background

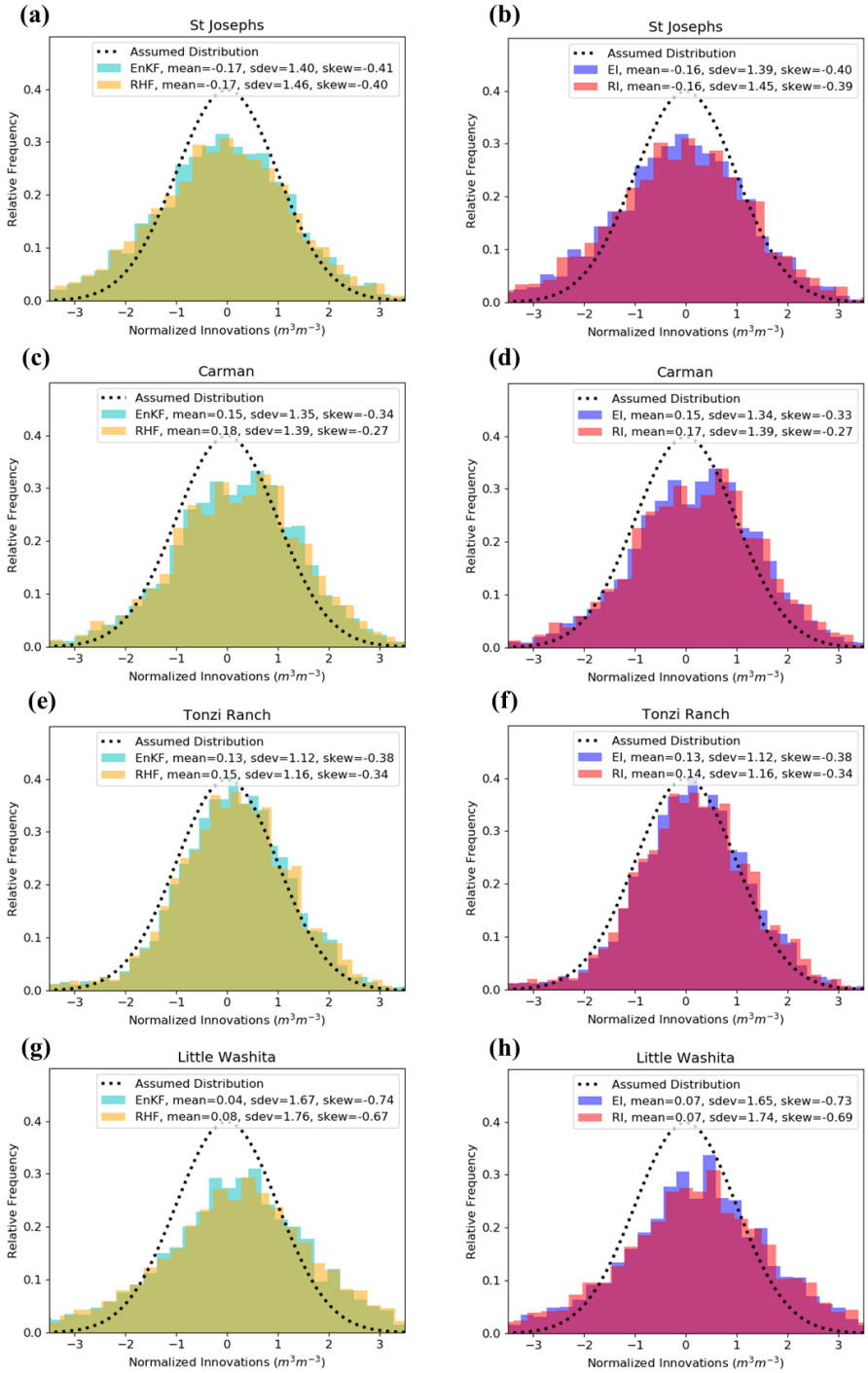

Figure 2.6: Representative set of normalized innovations distributions at (a),(b) St. Josephs, (c),(d) Carman, (e),(f) Tonzi Ranch, and (g),(h) Little Washita for the (left) noninflated and (right) inflated filter variants. Histograms were computed over the entire assimilation period (1 Jan 2000–31 Dec 2009). Quantities listed in key show the mean, standard deviation (sdev), and skewness of each distribution.

error covariance; the skewness values for the EnKF and the RHF are nearly identical at -0.41 and - 0.40, respectively. The negative skewness values imply that the probability density tends to be somewhat more concentrated at the higher end of the distribution when compared to a Gaussian (the relationship is reversed for positive skewness). The use of inflation (Figure 2.6b) didn't lead to any meaningful change in the mean, standard deviation, or skewness for either filter; whatever difference in statistics was observed seems to depend more on filter type (RHF vs EnKF) than it does on use of inflation.

Carman (Figures 2.6c and 2.6d) instead shows wet biases (0.15 and 0.18 for the EnKF and RHF, respectively); the standard deviations are 1.35 for the EnKF and 1.39 for the RHF; also shown are negative skewness values with -0.34 and -0.27 for the EnKF and RHF. As at St Josephs, inflation leads to negligible change in the distribution statistics considered here. At Tonzi Ranch (Figures 2.6e and 2.6f), the innovations indicate a systematic wet bias relative to expectation with values of 0.13 and 0.15 for the EnKF and RHF; the respective standard deviations are 1.12 and 1.16, with skewness values of -0.38 for the EnKF and -0.34 for the RHF. The normalized innovation statistics for the corresponding inflated cases are nearly identical. As at Carman and St Josephs, the differences in skewness are more dependent on filter type than on whether inflation was used.

The bias of the normalized innovations at Little Washita was 0.04 for the EnKF and 0.08 for the RHF (observations tended to be wetter than the forecasts); the EnKF had a standard deviation and skewness of 1.67 and -0.74, respectively, with the RHF having similar values (standard deviation of 1.76 and skewness of -0.67). Consideration of inflation (Fig. 2.6h) shows little to no change for the RHF, which had a reduction of 0.01 in the mean and one of 0.02 in both the standard deviation and skewness; somewhat larger differences are present between the EnKF variants

where now the inflated case shows an increase of 0.03 in the bias and a 0.02 reduction in the standard deviation of the normalized innovations distribution.

When looking at similar plots for all 18 locations (not pictured), there is no apparent relationship between the mean, bias, and skewness. Also of note is that, for each filter, inflation always led to a skewness magnitude that was less than or equal to that for the corresponding noninflated case. The RHF, inflated or not, tended to lead to skewness values closer to zero than the EnKF. These results imply that adaptive inflation was more likely to generate ensembles with relatively dry outlier members; these drier outliers were also more common for the EnKF than for the RHF. The results also show that the standard deviations of the normalized innovations distributions were generally larger for the RHF than for the EnKF, irrespective of inflation usage, a finding that is in line with the tendency of the RHF to have smaller ensemble spreads than the EnKF (as previously discussed).

To better quantify and understand the effects of non-Gaussianity on filter performance, we use the Kolmogorov-Smirnov (KS) statistic to test if a sample comes from a normal distribution. This is achieved by comparing the cumulative distribution functions of the sample and of a continuous normal distribution having the same mean and variance. The KS statistic is defined as the maximum distance between the two functions across the range of the random variable. Assuming that the sample is sufficiently large, higher KS statistics generally communicate more confidence that the sample distribution does not sample from the reference distribution. Here, we test every ensemble prior that was used as input for an analysis at each location and use a significance level of 5%. We take the null hypothesis to be that the ensemble prior is indeed Gaussian. This analysis was carried out for each of the 18 locations for every ensemble prior during JJA 2009, which is representative of the 10-year experiment period.

The KS values across all locations ranged from 0.5-0.65, which are all quite large. Every corresponding p-value was orders of magnitude smaller than the target significance level. This suggests that the evidence allows for us to reject the null hypothesis at each of the analysis times in JJA 2009; in other words, we can be confident that none of the prior distributions were Gaussian. When comparing the average KS value for each filter, the RHF had slightly higher values than the EnKF ensemble at all 18 sites. The difference does not depend on whether inflation was used (this distinction is present when making the same comparisons for the posterior ensembles). Also noted was that the mean KS values for the open loop were usually significantly higher than those of the data assimilation experiments. This result highlights how the update procedures used by the two filters lead to the destruction of non-Gaussian information in the ensemble priors, with the RHF maintaining more non-Gaussianity in the ensemble than the EnKF. In Figure 2.7, Little Washita is used as a representative example of the relative relationships of the KS time series for the open loop and for the ensemble priors corresponding to the four different data assimilation experiments. Here, the KS values for the RHF tend to be somewhat higher than those for the EnKF. The inflated EnKF shows the lowest KS readings. Note the relatively higher KS values for the open loop compared to all the assimilation cases.

2.5 Summary and conclusion

In this chapter, we presented results highlighting the differences between the EnKF and RHF when surface soil moisture observations were assimilated in a synthetic experiment. There were four cases: each filter was used with and without an adaptive inflation technique. The synthetic framework allows for more control of the experimental details, namely the better con-

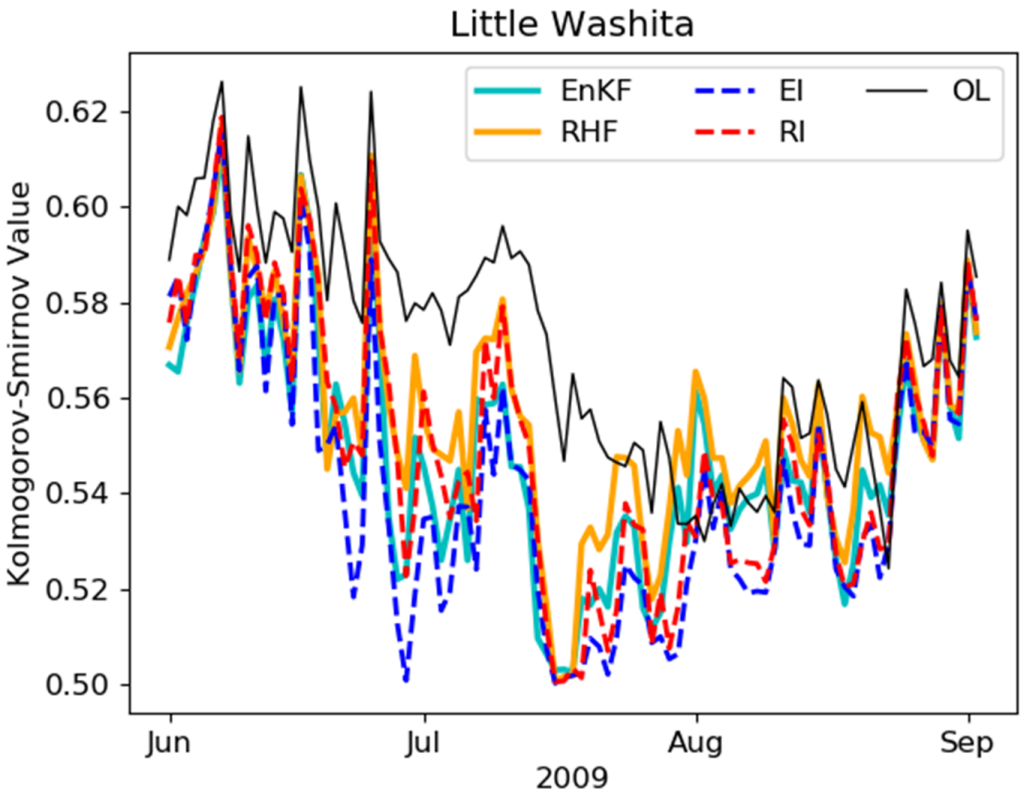

Figure 2.7: Time series of Kolmogorov–Smirnov values for the ensemble prior at the Little Washita site during JJA 2009.

textualization of model error, and the complete avoidance of representativity issues associated with using complex forward operators. In all of the metrics observed, the RHF performs at least as well as the EnKF but usually slightly better. This slight but consistent performance difference suggests that the non-Gaussianity present in the problem was significant enough to be exploited by the RHF. Additionally, the adaptive inflation implementation proved to be beneficial for both filters, suggesting that, given the experimental configuration, the deficiencies in ensemble spread were indeed relevant to performance and therefore requiring treatment.

The RHF allows for the non-Gaussian representation of the forecast and observation distributions. In our experiments, the observation error distributions were deliberately kept Gaussian to isolate the effects of the forecast-error parameterization. The RHF performed better than the open loop for all the NIC_{ν} metrics examined here. This was not the case for the EnKF, which showed near-zero NIC_{MAB} values for prmc. Evaluation of the KS statistics suggests that the RHF, inflated or not, allows for improved conservation of the ensemble's non-Gaussian information. Given the results here, we propose that it is worthwhile to further examine the RHF as a potential alternative to the EnKF for more complex land assimilation problems (including using bias corrected filters, modifying a filter for spatially distributed application, or assimilating multiple types of observations).

Chapter 3: Real observation experiment

3.1 Introduction

The results in the previous chapter were obtained using a perfect model experiment, in which the RHF showed itself to be the significantly better filter compared to the EnKF when it came to assimilating synthetically generated surface moisture observations. It was also shown that adaptively inflating the ensemble prior was somewhat beneficial to soil moisture quality, a result that applies to both the RHF and the EnKF. It is desirable to understand to what extent the perfect model results carry over to a testing environment that more closely resembles an operational setup, namely one that uses real (not synthetic) observations and implements a practical bias correction. Such is the primary concern of this chapter.

3.2 SMAP satellite observations

The NASA SMAP mission is a satellite-based Earth observation mission designed to provide global measurements of soil moisture and freeze-thaw state (Entekhabi et al. [62]). Launched on 31 January 2015, SMAP represents a collaborative effort between NASA and several international partners to improve our understanding of the Earth's water cycle and its influence on weather, climate, and agricultural productivity.

The data obtained from the SMAP mission have widespread applications across various fields. They contribute to improved weather and climate forecasts, hydrological modeling, drought monitoring, flood prediction, agricultural management, and water resource planning (Crow et al. [63], Zhang et al. [64], Reichle et al. [65]). Additionally, SMAP data are openly available to the scientific community and general public alike, encouraging further research and innovation. [For the experiments to be outlined, all SMAP data was obtained using AppEEARS (Application for Extracting and Exploring Analysis Ready Samples; https://appeears.earthdatacloud.nasa.gov/), which is a web-based data extraction and visualization tool developed by NASA designed to facilitate easy access and processing of large Earth observation datasets].

3.3 Experiment design

The domain for the real-observation experiments is the same set of 18 SMAP validation sites that were studied during the perfect model portion of this dissertation. Also unchanged are the land model (NASA Catchment model) and the meteorological forcing dataset (MERRA2). The perturbation settings are different as noted in Table 3.1, with an increase in the standard deviation of the catdef perturbations to 0.24 kg m $^{-2}$ h^{-1}.

The Catchment model initial conditions were obtained in the following manner. The model state was spun up from a cold start by running the model several times over the period from 1 Jan 2000 through 31 Dec 2021. On the last cycle, the model was stopped on 1 January 2015 and then replicated until a total of 100 identical copies were at-hand. To spin up this ensemble, each member was then forecast from 1 Jan 2015 through 31 Mar 2015 using a unique set of meteorological forcing perturbations, in much the same manner as in the perfect-model case but with minor

Table 3.1: Perturbed meteorological forcing variables in real observation experiments.

Perturbed Variable	Surface excess (additive)	Catchment deficit (additive)	Precipitation (multiplicative)	Shortwave radiation (multiplicative)	Longwave radiation (downwelling) (additive)
Standard deviation (S)	0.16 mm	0.24 mm	0.5	0.3	20 W m^{-2}
Time scale [hr]	3	24	72	72	72

changes of the perturbation magnitudes (see Table 3.1). To obtain the open loop, the ensemble was then run from 1 Apr 2015 through 31 Mar 2021, the new experimental time period. The same experimental configurations will be used here (i.e., EnKF, RHF, EI, and RI). The observations to be assimilated are obtained from version 5 of the SMAP L3_SM_P_E product (ONeill et al. [66]). Quality control was enforced via filtering on the appropriate quality flags that are provided in the observation metadata. A simple bias correction approach was implemented before assimilating any observations. The approach used involved independently computing the z-scores of the observation and forecast ensemble prior to computing the analysis. For illustration, as follows is the formula for computing the observation z-score, z_{y_i}:

$$z_{y_i} = \frac{y_i - \mu_y}{\sigma_y} \tag{3.1}$$

Here, μ_y and σ_y are the sample mean and standard deviations of the SMAP observations over the entire 6-yr experiment period – these values are static. y_i is the single raw SMAP observation available at the analysis step.

The mean and standard deviation values used to normalize the filter ensemble were taken from the daily-averaged open loop ensemble mean and standard deviation over the 6-yr experiment period. Statistics were only evaluated for the subset of days during the experiment period on which there was a valid observation.

The performance metrics of Chapter 2 were likewise computed for the real-observation experiments. Of relevance here is that the in-situ core data used to validate the SMAP soil moisture retrievals will proxy for the truth.

3.4 Results and discussion

Figure 3.1 shows the plain (not NIC) ACC, MAB, and ubRMSE values for the 6-yr experiment period after having been averaged across all 18 test sites.

The open loop ensemble had an ACC of 0.6. The EnKF and RHF each showed modest improvements over the open loop with respective ACC values of 0.63 and 0.64. Inflation does not result in any appreciable change in ACC for either filter. No one experimental configuration separates itself as the singly superior choice with respect to the ACC metric in a statistically significant sense. Also of note is that no experiment does significantly better than the open loop, implying that data assimilation falls short in providing a statistically meaningful benefit to base-level model skill. Such findings collectively represent a noticeable departure from the results collected from the perfect-model experiments in Chapter 2, for which the RHF displayed significantly better performance than the EnKF for sfmc, and the performance benefits introduced by inflation usage were also significant for each filter; recall also that it was possible to select the RI experiment as the unequivocal (statistically significant) best choice (for sfmc ACC). One

obvious potential explanation for the different outcomes may indeed be the existence of a bias correction implementation. Another contributing factor is likely the difference in how frequently observations are being assimilated, which is undoubtedly much lower for the real-observation case, limiting whatever ability the filters may have in being able to modify the model state. It is important to keep in mind that such factors do indeed complicate the evaluation of ACC (and all other metrics), though, in fairness, these are exactly the type of constraints practitioners actually deal with, so these points are not particularly novel.

For the location-averaged ubRMSE, the open loop ensemble shows a reading of 0.041 m^3 m^{-3}. The EnKF and RHF have somewhat better results at 0.039 m^3 m^{-3} and 0.037 m^3 m^{-3}, respectively. Adaptively inflating the prior ensembles does not impact these metrics. As was the case for ACC, statistical noise makes it impossible to prescribe any filter over another with respect to the ubRMSE metric. For the perfect-model runs, it was possible to determine that the RI approach performed significantly better than the EnKF, though not better than the RHF or EI runs – for the real-observation analogue, this performance gap effectively vanishes.

The open loop shows an average MAB of 0.035 m^3 m^{-3}. The EnKF and RHF both have marginally better scores with a common value of 0.034 m^3 m^{-3}. When adding in inflation, the performance improves such that the EI and RI experiments show respective MAB values of 0.033 m^3 m^{-3} and 0.032 m^3 m^{-3}, where the use of adaptive inflation seems to benefit the RHF more but not such that there is a statistically meaningful distinction between any two filters. As for ACC and ubRMSE, data assimilation always leads to a positive impact over the base-level model skill (OL) in terms of MAB. It is worth mentioning that, by design, bias correcting the observations ought to make it such that the OL and assimilation experiments have more or less the same climatology (though there may still be a bias present with respect to the in situ observation used

for verification). It is therefore not expected that assimilating soil moisture observations would lead to any real change in the MAB metric (positive or negative). Any departure from zero must be attributed to noise or residual systematic errors. The MAB results obtained here are in stark contrast to what was noted for the synthetic observation case, where the RHF and RI experiments showed themselves as being the best choices in terms of bias mitigation. It might be concluded that a simple bias correction would have produced results in Chapter 2 more in line with those seen here for the real observation experiments. Technically, the lack of a bias correction for the perfect model experiments should have negatively impacted the performance of both filters, since neither is formulated to account for biases, but the findings support the notion that the RHF's flexibility with respect to non-Gaussian likelihoods may make it better suited for more local bias mitigation approaches. Exploring this potential point of contrast between the two filters is worth future effort as any observed differences in behavior may have more relevant implications on the forecasting of extreme weather events. Given all the results shown in Figure 3.1, it is not feasible to say that a single filter performs significantly better than any other when across all 3 metrics (ACC, ubRMSE, MAB).

3.4.1 Surface moisture content

The sfmc time series for JJA 2018 at a subset of the 18 validation sites is provided in Figure 3.2. Included are the means of the open loop and filter ensembles, as well as the assimilated SMAP observations and the in situ sfmc observations that are used as the proxy for truth.

Generally, all the time series at Tonzi Ranch (Figure 3.2a) show a gentle, seasonal drying trend over the 3-month span. The open loop ensemble is much wetter than the in situ observations.

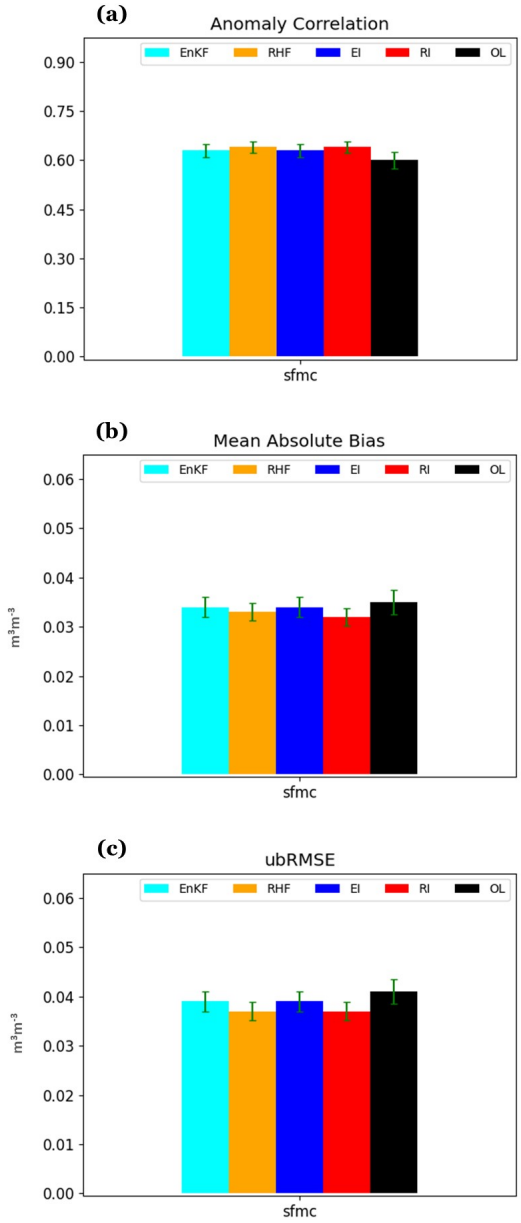

Figure 3.1: NIC values averaged over entire 6-yr experiment period for real-observation evaluation. Averaging was done over all locations.

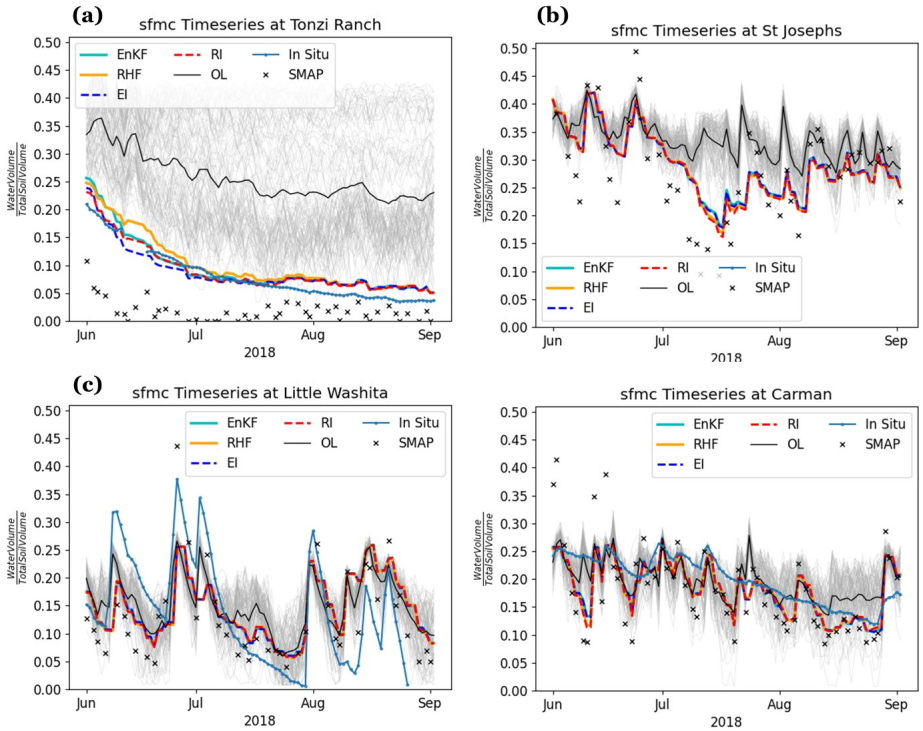

Figure 3.2: Time series of volumetric soil moisture content (m^3 m^{-3}) for sfmc at representative set of locations during JJA 2018. SMAP observations are scaled according to Eq. (3.1). [Note that St Josephs has no in situ data available during this time period.]

The observations, in contrast, are much drier (very near zero) over most of the same period. The filter ensembles do very well to track the in situ values throughout. There are small differences in performance between the different experimental configurations. Here, the EnKF seems to do a better job of approximating the in situ reference than the RHF. Inflation usage seems to provide additional performance benefits to both filters but not such that the relative performance ranking is changed (i.e., the EI ensemble shows more skill than that of the RI experiment). Generally, the open loop ensemble spread is quite large, suggesting a high sensitivity to data assimilation, which is supported by how well the filters are getting rid of moisture. A likely reason for why the filters do not seem to take advantage of the dispersion in the open loop more may be stemming from the perturbations used for ensemble generation; when at the dry end of the wetness spectrum, perturbations can only add moisture into the soil column, limiting how representative a forecast ensemble can be when such dry observations are being assimilated. The separation between the different time series is quite glaring. As noted earlier, the bias correction implementation made use of data across the entire 6-yr period for computing a single pair of mean and standard deviation values. It is expected that there still be some difference when using a more granular temporal window for analysis. Such differences were noted for other time slices as well.

The time series at St Josephs (Figure 3.2b) shows several differences from what was just observed for the same time slice at Tonzi Ranch. The difference between the open loop and the filter ensembles is relatively more mild than what was noted at Tonzi Ranch. Here, the SMAP observations also tend to be more randomly dispersed about the model ensembles (both positive and negative, relatively speaking). Of note is that there are no in situ validation observations during this time period. The complete lack of temporal coverage for extended periods is quite common (at all 18 locations - not just St Josephs) and does not allow for gauging the filter

performance during this 3-month span. What can be gleaned from the time series is that the filter ensembles all track each other faithfully with any differences that occur being essentially negligible. The filter ensembles also seem to be reasonably responsive to the observations, but the observations do seem to imply the presence of what may be unrealistic soil moisture dynamics. This is suspected because of the high-magnitude wetting events indicated by rapid peaking of the SMAP data. The dry-downs seem more reasonable. An alternative view could be that the scaling of the observations into the model space may have been less correct for this specific location.

At Little Washita (Figure 3.2c), the open loop ensemble mean does a good job of approximating the in situ observations, with modest differences in terms of temporal variability or 3-month average. Assimilating the sfmc retrievals does lead to readily noted changes, whereby the filter configurations are clearly syncing with the observation time series, but it is not clear that such changes are actually helpful on average in light of the verification data. Regardless, whatever negative effects are present are marginalized by the high base-level skill of the model.

For Carman (Figure 3.2d), the model ensemble and observation time series show very high levels of variability compared to the validation set. Overall, the general drying trend with time is also indicated by all of the time series. The variability gap notwithstanding, the open loop ensemble seems to approximate the in situ values reasonably well. Data assimilation actually seems to make things somewhat worse during this time period as the high sensitivity to the observations pulls the filters further away from the in situ time series.

3.4.2 Ensemble representativeness

The ensemble histograms for the real-observation experiments are provided in Figure 3.3 for the same four representative locations as in (Figure 3.2). The rank histograms shown are formulated the same way as in the perfect-model portion of this dissertation, albeit with a considerably smaller set of populating observations and for a different experiment period (1 Apr 2015 - 1 Apr 2021).

At Tonzi Ranch, the open loop shows a mostly flat rank histogram profile, indicating reasonable correspondence between the space spanned by the observations and that spanned by the model ensemble. The assimilation of sfmc observations via the different filter configurations seems to introduce more moisture in the ensembles, meaning much wetter model estimates relative to the observations. This outcome is somewhat more exaggerated for the EnKF and EI set than it is for either the RHF or RI experiment. The findings just reported are not impacted by inflation usage. A particularly interesting feature for the EnKF and EI experiments is the steep drop in ensemble representativeness for very dry soil moisture values. This implies that the EnKF and EI ensembles tend to be more relatively more spread out than the set of RHF and RI ensembles. Recalling the rank histograms at Tonzi Ranch for the perfect-model case, data assimilation did indeed result in a wetter ensemble compared to observations but no such representativeness gap was noted. A potential explanation for this is that the perturbation scheme is too aggressive for this location, leading to the systematic addition of too much moisture in the model ensembles.

The open loop and filter configurations at St Josephs all show similar rank histograms indicating ensembles that are slightly underdispersive and wet when compared to the observations. Data assimilation seems to marginally exacerbate the oversaturation present in the open loop. The

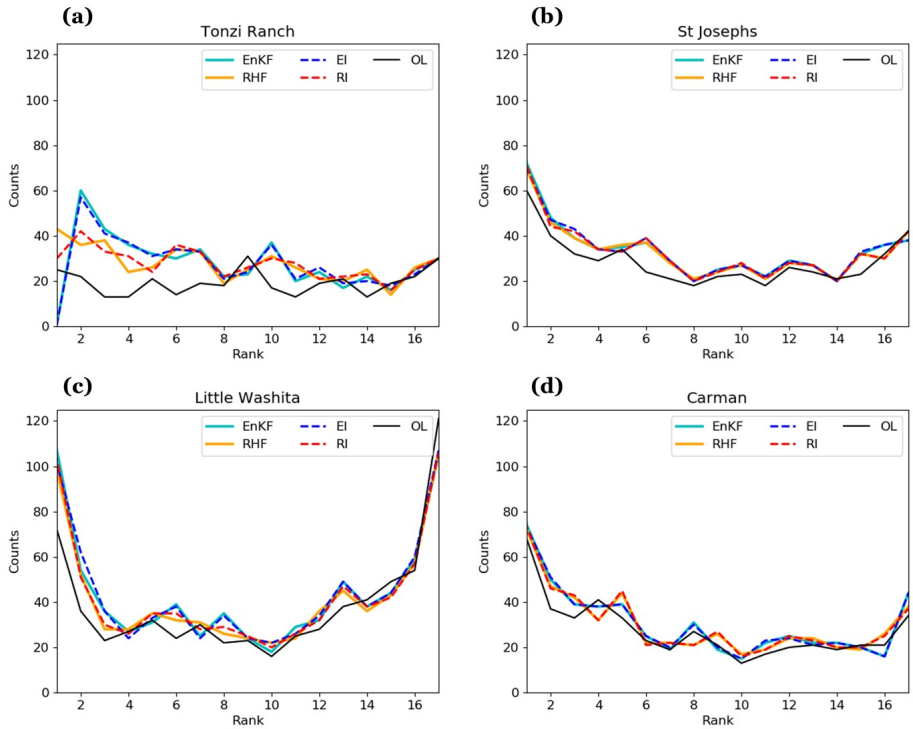

Figure 3.3: Rank histograms for sfmc at a representative set of the testing locations.

general rank histogram features and the relative differences between the filter and open loop profiles are actually quite close to those noted for the synthetic observation case, with no discrepancy in major features.

The rank histograms at Little Washita all show rather extreme underdispersion. Adaptive inflation does not seem to affect this result at all. The model ensembles do appear to be too dry with respect to the observations. The findings here just about match those seen at Little Washita for the synthetic observation case, with the only departure being that the open loop in the perfect-model scenario was significantly less underdispersive.

Lastly, the open loop ensemble at Carman tends to be underdispersive and wet when compared to the observations. Data assimilation does not seem to impact this result as indicated by the negligible separation between the rank histogram for the open loop and those for the filter ensembles. As at the previously mentioned locations, inflation usage leads to no meaningful change in the rank histogram profile. The perfect-model experiment rank histograms at Carman are perfectly aligned with the major features noted here.

As mentioned briefly for the rank histogram discussion for Tonzi Ranch, it is possible that the general biases present in all of the 4 histograms just overviewed may have origins stemming from the meteorological perturbations that were applied for ensemble generation. One potential avenue for future work is exploring the efficacy of a spatially-varying set of perturbation parameters with regard to addressing such biases. The scheme would undoubtedly have implications for mitigating the severity of underdispersion features as well, which were present to some degree for all four locations. On another note, the rank histograms were remade but using the in situ data as the observations (not shown); it was readily observed that the general features of the histograms in Figure 3.2 were maintained, which is the expected result.

3.4.3 Standard deviations

The standard deviation time series for the same four locations and JJA 2018 time period are provided in Figure 3.4.

At Tonzi (Figure 3.4a), the standard deviation for the open loop shows a 3-month mean of 0.089 m^3 m^{-3}. This average is dominated mainly by the latter two months , with the spread increasing quickly during the first month and dwarfing the standard deviations of the filter ensembles. The average values over this period for the EnKF and RHF are 0.068 m^3 m^{-3} and 0.050 m^3 m^{-3}, respectively, with the EnKF showing a much higher spread than the RHF. Inflation leads to modest changes in both filters, with the EI and RI experiments showing respective time-average standard deviations of 0.069 m^3 m^{-3} and 0.049 m^3 m^{-3}. The really large standard deviation for the open loop can also be seen in terms of the sfmc ensemble members in Figure 3.4. Also noted are how the filter ensembles are much closer to the bottom end of saturation, being quite dry relative to the open loop. As noted for the perfect-model experiments, the RHF and RI set continue the observed trend of exhibiting significantly lower spreads than the EnKF and EI ensembles on average. [This relative difference is exhibited for the remaining three locations as well].

The open loop at St Josephs (Figure 3.4b) shows an average standard deviation of roughly 0.023 m^3 m^{-3}. The EnKF and RHF have respective mean standard deviations of 0.038 m^3 m^{-3} and 0.033 m^3 m^{-3}. Assimilating observations did not result in a noteworthy impact on spreads for this period. The EnKF and EI experiments come in with higher spreads than the RHF and RI set, and the open loop has significantly lower spreads than the other model ensembles. A quick consultation of the accompanying sfmc plot shows that the open loop is much nearer saturation. The compression of the open loop ensemble against the upper bound of physically allowable soil

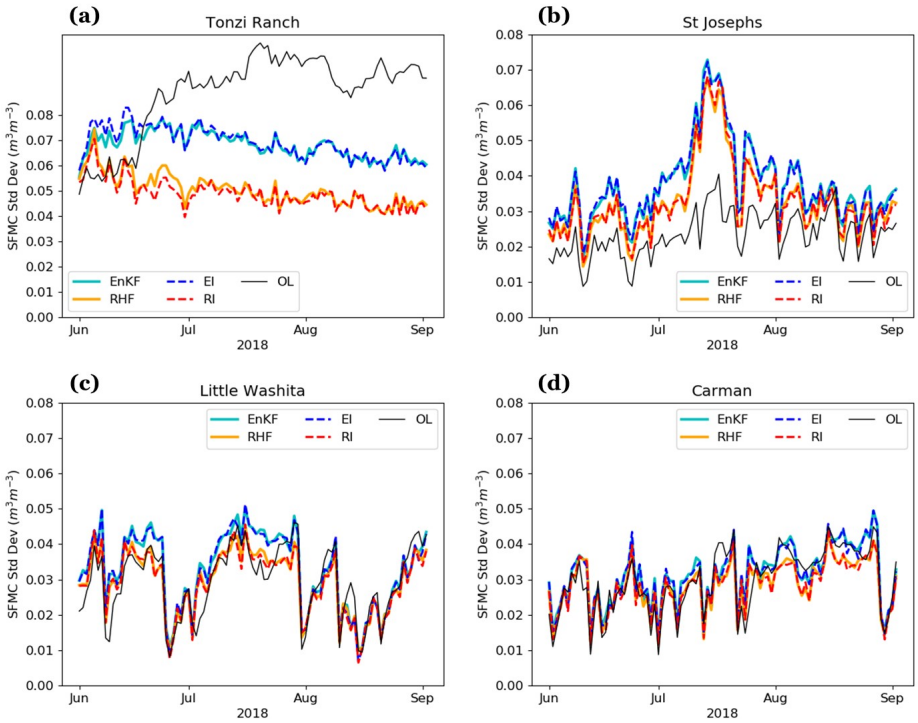

Figure 3.4: Time series of ensemble standard deviations at a representative set of test locations.

moisture explains the relatively lower spreads that are observed here.

For Little Washita (Figure 3.4c), the open loop ensemble shows a mean standard deviation of 0.029 m^3 m^{-3} over JJA 2018. The EnKF and EI experiments share a common average of roughly 0.033 m^3 m^{-3}, with the RHF and RI experiments characterized by notably less spread, sharing a 3-month mean of 0.030. The results indicate that spread has little dependence on data assimilation, as indicated by how closely the each of the individual time series track each other.

Carman (Figure 3.4d) shows an open loop 3-month average standard deviation of about 0.030m^3 m^{-3}. The EnKF ensemble tends to be somewhat more dispersive with an average value of 0.031 m^3 m^{-3} and the RHF being somewhat less dispersive showing a mean value of 0.028. Adaptively inflating the ensemble prior did not lead to any meaningful changes in the spreads.

3.4.4 Measuring non-Gaussianity

In an attempt to quantify (detect) ensemble departures from Gaussianity, the KS values will be analyzed for a subset of the validation sites during JJA 2018.

The KS statistic of the open loop ensemble at Tonzi Ranch during JJA 2018 showed an average value of roughly 0.539 over the 3-month period. The EnKF and RHF share a common 3-month average of 0.503, coming in noticeably lower than that of the OL. Adaptively inflating the ensemble prior actually leads to increases as evidenced by the comparatively higher KS values for the EI and RI experiments at 0.509 and 0.507, respectively. Of note is that the KS time series over JJA 2018 declines sharply over June and the early part of July before plateauing somewhat for the remainder of the evaluation period. The relatively lower values of the filter ensembles imply that data assimilation is forcing the ensemble closer to Gaussianity. The destruction of

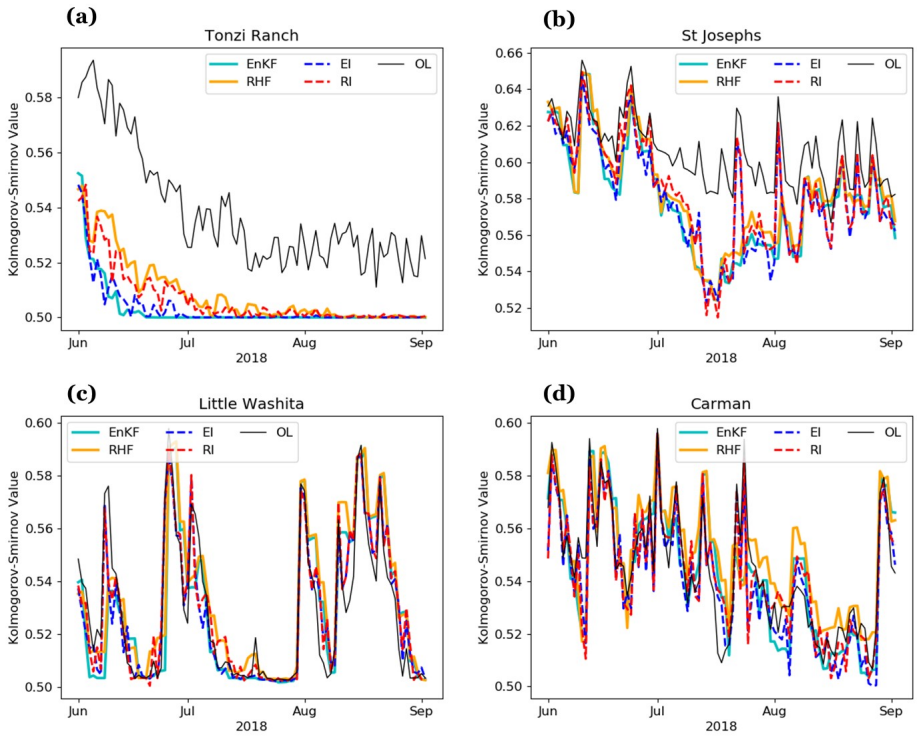

Figure 3.5: Kolmogorov-Smirnov values of sfmc for representative set of locations.

non-Gaussian information caused by data assimilation was also seen in the KS discussion of Chapter 2. There is a notable flattening of the KS values at the 0.5 mark. Also of note is that the general outline and relationships between the individual KS time series actually seem to match the corresponding sfmc time series to some extent. This correspondence between KS and sfmc time series was noted for the other 3 locations as well.

At St Josephs, the open loop ensemble has a mean KS value of 0.606. The EnKF and RHF have relatively smaller KS metrics with respective averages of 0.579 and 0.584. Inflation usage does not lead to the creation of a meaningful distinction in the mean values, though, via an inspection of the KS time series, there does appear to be some noticeable noise (small scale variability) that is introduced. Overall, the RHF and RI ensembles tend to contain a little more non-Gaussian information than the EnKF and EI ensembles.

Little Washita has an open loop JJA 2018 mean KS of 0.528. The EnKF and EI set have roughly the same mean, and the RHF and RI experiments have respective averages of 0.533 and 0.529. Unlike in the perfect model case, the KS values for the filter configurations at Little Washita are at least as high as that for the OL. It is not at all clear how to explain the relative differences between the configurations. Based on the results obtained in this dissertation, if inflation has any effect on KS values, it seems to be such that a reduction is realized. This same pattern is on display at Carman. The open loop at Carman shows an average KS for the open loop of 0.544. The EnKF and EI experiments have means of 0.545 and 0.541, respectively, while the RHF and RI experiments have higher respective means of 0.551 and 0.543. Unlike at Tonzi Ranch and St Josephs, Little Washita and Carman show examples of the filter experiments having higher KS values than that of the OL. In each of these latter two locations, the RHF tended to have a more extreme departure relative to the EnKF.

3.5 Summary and conclusion

In this chapter, results were obtained for the real observation analogue to the perfect model experiments of Chapter 2. Performance of the EnKF and RHF were compared at the same 18 test locations, but this time over a shorter time period (6 years) to accommodate the limited observation data, for which SMAP soil moisture retrievals were used. The in situ data meant to validate SMAP soil moisture retrieval products stood for truth in our experiments. Unlike in the perfect model case, there was no temporal offset imposed between the meteorological forcing and the initial conditions of the model. Additionally, a bias correction scheme was implemented whereby the first two moments of the observation and model forecast records over the 6-yr period were removed from the respective raw observation and model forecast being assimilated each day. The adaptive inflation algorithm of Chapter 2 was also tested for the real observation case.

Based on the aggregate NIC_ν metrics, every one of the filter configurations tended toward improved performance over the base-level model skill of the OL, but in no case was this improvement statistically significant. The lack of statistical significance was also present for the differences introduced by inflation usage. Such results are not totally unexpected given the relative dearth of assimilated observations. Based on the KS values at Little Washita in the perfect model case, it was concluded that the use of data assimilation resulted in the removal of non-Gaussian information in the ensemble, with the RHF being relatively less destructive than the EnKF in this regard. The real observation case tells a somewhat different story, mainly pertaining to the relative differences in KS value between the filter configurations and the OL, with the OL not always having a higher reading than the filters. It is suspected that the general dispersion of the ensemble plays a role here, but such must be the focus of more extensive future work. Similarly,

when looking at the rank histograms for both the real observation and perfect model cases, the idea of using a meteorological perturbation scheme having spatially dependent parameters might be worth exploring in future to address some of the persistent bias and underdispersion problems observed for some of the locations. Improvements in this regard may be more fruitful than trying to leverage more general data assimilation methods. This latter point is mainly predicated, again, on the data requirement necessary to realize significant performance differences between current and more novel approaches being too high at this time.

Chapter 4: Concluding remarks

The primary goal of this dissertation was to explore the effects of non-Gaussian error parameterizations of the prior on the quality of analyses output by a land data assimilation system. Of secondary importance was the desire to better understand how performance might further be affected when an adaptive inflation scheme is applied to counteract the negative effects of systematic deficiencies. To these ends, the relative differences in skill between the EnKF and the RHF were evaluated for the surface moisture assimilation problem at a number of testing locations, both filters being examined with and without the assistance of an auxiliary inflation system acting on the ensemble prior. The comparison was first carried out in a synthetic context, in which it was possible to maintain tight control over the experimental details, namely how to contextualize model error. The synthetic environment also allowed for the avoidance of representativeness issues associated with using complex forward operators.

- It was noted in the results that the assimilation of sfmc observations led to unequivocal improvements in the ability of the model ensemble to better approximate the truth trajectory, irrespective of filter type or whether inflation was used.

This success is largely in agreement with other examples in the literature that also highlight the wide-ranging benefits of surface moisture assimilation seen not only for the observed variable but also for the unobserved portions of the land model via adequate information spread.

- The RHF showed itself to be at least as good as the EnKF for every metric and was significantly better than the EnKF for a number of metrics in the perfect model experiments.
- Inflation usage also provided modestly significant improvements for both filters, with no sufficient evidence of a net negative impact anywhere in the evaluation.

One note about the perfect model experiments is that the same observation-error standard deviation was used at every site; using a site-specific error may be the better solution moving forward.

Even if the RHF is indeed the better filter, it is relevant to note that this performance gap was realized in an idealized setting, with artificial prescriptions of model error and with an abundance of high-fidelity observations available for assimilation. Would this skill gap continue to be as apparent if the evaluation was to be carried out in an environment more consistent with an operational setting? This was the driving question behind the second set of experiments, which assimilated SMAP surface soil moisture retrievals after a climatological bias correction application; an independent set of in situ observations stood in as proxy for the true sfmc values.

The results show that the average performance differences between the two filters are largely diminished in this more realistic setting, such that that performance gains of the RHF relative to the EnKF dip just below the threshold of statistical significance (95%), which is not surprising. A similar finding was noted for the adaptive inflation usage when transitioning to the real observation case. One result noted for both the perfect-model and real-observation portions was that the inflation algorithm yielded an unexpected lack of impact. In offline testing, the damping toward one was turned off completely; no meaningful resulting change in the inflation runs was observed after having made this change. The conclusion from this study is not that inflation is unnecessary for this problem but rather that a different algorithm ought to be ex-

plored. With respect to the topic of bias, as noted in Chapter 3, neither filter formulation implies a relative advantage in handling systematic bias – relatively speaking, the RHF simply has more flexibility in modeling non-Gaussian likelihoods. But still the fact remains that, in the presence of a bias (Chapter 2), the RHF managed to perform significantly better than the EnKF in terms of mitigating this problem (i.e., MAB). On another note, it was suggested in this work that a spatially-varying (location-specific) set of meteorological forcing and model state perturbations may be helpful in addressing the both the bias and underdispersion features present in the rank histograms; exploring how best to determine this may be a relevant focus point for future experiments. In conclusion, the performance gains of the RHF over the EnKF noted here are quite small, likely not warranting the overhaul of land data assimilation systems at more established centers when considering the expense that would undoubtedly be required to make such a transition. With that said, the incremental benefits of the RHF outlined in this dissertation clearly speak to the importance of further study/consideration of non-Gaussian and nonlinear features in environmental modeling problems. Treatment of such features ought to be increasingly relevant to land model performance (and earth climate study) as the robustness of observation networks continues to improve.

Appendix A: SMAP L4_SM Validation Sites

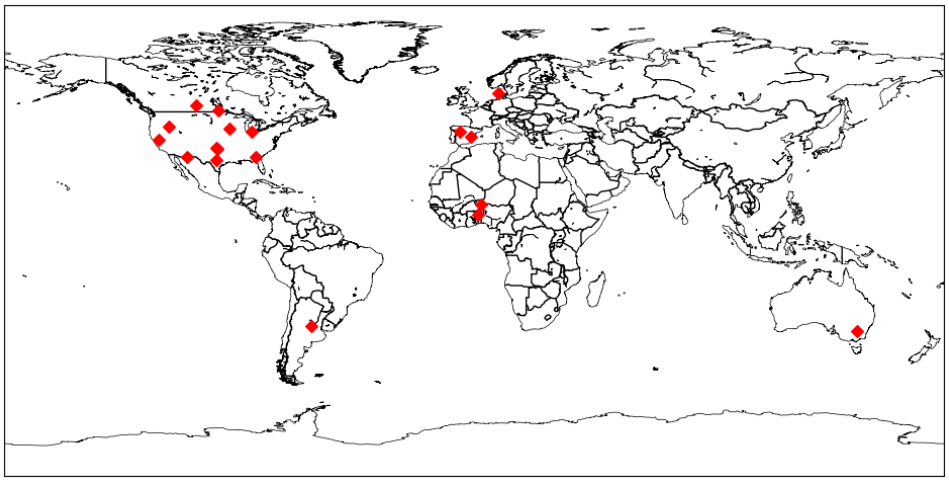

Figure A.1: Map of SMAP validation sites. The experiments outlined in this dissertation were carried out exlusively at these locations.

Table A.1: Latitude and longitude of SMAP L4_SM validation sites, along with corresponding climate regime and MODIS IGBP land cover type.

Site name	Latitude (°)	Longitude (°)	Climate regime	MODIS IGBP land cover
1.Remedhus	41.42	-5.37	Temperate	Croplands
2.Reynolds Creek	43.19	-116.72	Arid	Grassland
3.Yanco	-34.72	146.13	Semiarid	Croplands (crop–grass mix)
4.Carman	49.67	-97.98	Cold	Croplands
5.Walnut Gulch	31.72	-110.09	Arid	Shrub open
6.Little Washita	34.92	-98.23	Temperate	Grasslands
7.Fort Cobb	35.38	-98.57	Temperate	Grasslands (crop–grass mix)
8.Little River	31.72	-83.73	Temperate	Cropland/natural mosaic
9.St Josephs	41.45	-84.97	Cold	Croplands
10.South Fork	42.42	-93.53	Cold	Croplands
11.Monte Buey	-33.01	-62.49	Arid	Croplands
12.Tonzi Ranch	38.43	-120.95	Temperate	Savannas woody
13.Kenaston	51.39	-106.51	Cold	Croplands
14.Valencia	39.57	-1.26	Arid	Savannas woody
15.Niger	13.57	2.66	Arid	Grasslands
16.Benin	9.77	1.68	Arid	Savannas
17.TxSON	30.43	-98.61	Temperate	Grasslands
18.HOBE	55.97	9.10	Temperate	Croplands (crop–natural mix)

Bibliography

[1] Ralph F Keeling and Charles D. Keeling. Atmospheric monthly in situ co2 data - mauna loa observatory, hawaii. in scripps co2 program data, 2017. URL https://library.ucsd.edu/dc/object/bb3859642r.

[2] Katherine Calvin, Dipak Dasgupta, Gerhard Krinner, Aditi Mukherji, Peter W. Thorne, Christopher Trisos, José Romero, Paulina Aldunce, Ko Barrett, Gabriel Blanco, William W.L. Cheung, Sarah Connors, Fatima Denton, Aïda Diongue-Niang, David Dodman, Matthias Garschagen, Oliver Geden, Bronwyn Hayward, Christopher Jones, Frank Jotzo, Thelma Krug, Rodel Lasco, Yune-Yi Lee, Valérie Masson-Delmotte, Malte Meinshausen, Katja Mintenbeck, Abdalah Mokssit, Friederike E.L. Otto, Minal Pathak, Anna Pirani, Elvira Poloczanska, Hans-Otto Pörtner, Aromar Revi, Debra C. Roberts, Joyashree Roy, Alex C. Ruane, Jim Skea, Priyadarshi R. Shukla, Raphael Slade, Aimée Slangen, Youba Sokona, Anna A. Sörensson, Melinda Tignor, Detlef van Vuuren, Yi-Ming Wei, Harald Winkler, Panmao Zhai, Zinta Zommers, Jean-Charles Hourcade, Francis X. Johnson, Shonali Pachauri, Nicholas P. Simpson, Chandni Singh, Adelle Thomas, Edmond Totin, Andrés Alegría, Kyle Armour, Birgit Bednar-Friedl, Kornelis Blok, Guéladio Cissé, Frank Dentener, Siri Eriksen, Erich Fischer, Gregory Garner, Céline Guivarch, Marjolijn Haasnoot, Gerrit Hansen, Mathias Hauser, Ed Hawkins, Tim Hermans, Robert Kopp, Noëmie Leprince-Ringuet, Jared Lewis, Debora Ley, Chloé Ludden, Leila Niamir, Zebedee Nicholls, Shreya Some, Sophie Szopa, Blair Trewin, Kaj-Ivar van der Wijst, Gundula Winter, Maximilian Witting, Arlene Birt, and Meeyoung Ha. *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.* July 2023. doi: 10.59327/ipcc/ar6-9789291691647. URL http://dx.doi.org/10.59327/IPCC/AR6-9789291691647.

[3] Aslak Grinsted, J. C. Moore, and S. Jevrejeva. Reconstructing sea level from paleo and projected temperatures 200 to 2100 ad. *Climate Dynamics*, 34(4):461–472, January 2009. doi: 10.1007/s00382-008-0507-2. URL https://doi.org/10.1007/s00382-008-0507-2.

[4] Paul J Markwick. Fossil crocodilians as indicators of late cretaceous and cenozoic climates: implications for using palaeontological data in reconstructing palaeoclimate. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 137(3-4):205–271, March 1998.

[5] H. Fischer, M. Wahlen, J. Smith, D. Mastroianni, and B. Deck. Noaa/wds paleoclimatology

- vostok ice core co2 data, 1105-2856m, 1999. URL https://www.ncei.noaa.gov/metadata/geoportal/rest/metadata/item/noaa-icecore-2443/html.

[6] James E. Hansen and Makiko Sato. Paleoclimate implications for human-made climate change. 2011. doi: 10.48550/ARXIV.1105.0968. URL https://arxiv.org/abs/1105.0968.

[7] P. Jeremy Werdell, Michael J. Behrenfeld, Paula S. Bontempi, Emmanuel Boss, Brian Cairns, Gary T. Davis, Bryan A. Franz, Ulrik B. Gliese, Eric T. Gorman, Otto Hasekamp, Kirk D. Knobelspiesse, Antonio Mannino, J. Vanderlei Martins, Charles R. McClain, Gerhard Meister, and Lorraine A. Remer. The plankton, aerosol, cloud, ocean ecosystem mission: Status, science, advances. *Bulletin of the American Meteorological Society*, 100(9): 1775–1794, September 2019. ISSN 1520-0477. doi: 10.1175/bams-d-18-0056.1. URL http://dx.doi.org/10.1175/BAMS-D-18-0056.1.

[8] OpenTopography. Napa watershed, ca, 2010. URL https://opentopography.org/meta/OT.052010.26910.1.

[9] Christopher P. Catano and I. Jack Stout. Functional relationships reveal keystone effects of the gopher tortoise on vertebrate diversity in a longleaf pine savanna. *Biodiversity and Conservation*, 24(8):1957–1974, April 2015. ISSN 1572-9710. doi: 10.1007/s10531-015-0920-x. URL http://dx.doi.org/10.1007/s10531-015-0920-x.

[10] Teja Tscharntke, Yann Clough, Thomas C. Wanger, Louise Jackson, Iris Motzke, Ivette Perfecto, John Vandermeer, and Anthony Whitbread. Global food security, biodiversity conservation and the future of agricultural intensification. *Biological Conservation*, 151 (1):53–59, July 2012. ISSN 0006-3207. doi: 10.1016/j.biocon.2012.01.068. URL http://dx.doi.org/10.1016/j.biocon.2012.01.068.

[11] Michael Jahi Chappell and Liliana A. LaValle. Food security and biodiversity: can we have both? an agroecological analysis. *Agriculture and Human Values*, 28(1):3–26, November 2009. ISSN 1572-8366. doi: 10.1007/s10460-009-9251-4. URL http://dx.doi.org/10.1007/s10460-009-9251-4.

[12] Sheila M. Olmstead. Climate change adaptation and water resource management: A review of the literature. *Energy Economics*, 46:500–509, November 2014. ISSN 0140-9883. doi: 10.1016/j.eneco.2013.09.005. URL http://dx.doi.org/10.1016/j.eneco.2013.09.005.

[13] Arun Lal Srivastav, Rajni Dhyani, Manish Ranjan, Sughosh Madhav, and Mika Sillanpää. Climate-resilient strategies for sustainable management of water resources and agriculture. *Environmental Science and Pollution Research*, 28(31):41576–41595, June 2021. ISSN 1614-7499. doi: 10.1007/s11356-021-14332-4. URL http://dx.doi.org/10.1007/s11356-021-14332-4.

[14] S. J. Ghan, X. Liu, R. C. Easter, R. Zaveri, P. J. Rasch, J.-H. Yoon, and B. Eaton. Toward a minimal representation of aerosols in climate models: Comparative decomposition of aerosol direct, semidirect, and indirect radiative forcing. *Journal of Climate*, 25 (19):6461–6476, April 2012. ISSN 1520-0442. doi: 10.1175/jcli-d-11-00650.1. URL http://dx.doi.org/10.1175/JCLI-D-11-00650.1.

[15] Sha Zhou, A. Park Williams, Alexis M. Berg, Benjamin I. Cook, Yao Zhang, Stefan Hagemann, Ruth Lorenz, Sonia I. Seneviratne, and Pierre Gentine. Land–atmosphere feedbacks exacerbate concurrent soil drought and atmospheric aridity. *Proceedings of the National Academy of Sciences*, 116(38):18848–18853, September 2019. ISSN 1091-6490. doi: 10.1073/pnas.1904955116. URL http://dx.doi.org/10.1073/pnas.1904955116.

[16] Ray G. Anderson and Michael L. Goulden. Relationships between climate, vegetation, and energy exchange across a montane gradient. *Journal of Geophysical Research*, 116(G1), March 2011. ISSN 0148-0227. doi: 10.1029/2010jg001476. URL http://dx.doi.org/10.1029/2010JG001476.

[17] Lisa Holsinger, Sean A. Parks, Marc-André Parisien, Carol Miller, Enric Batllori, and Max A. Moritz. Climate change likely to reshape vegetation in north america's largest protected areas. *Conservation Science and Practice*, 1(7), May 2019. ISSN 2578-4854. doi: 10.1111/csp2.50. URL http://dx.doi.org/10.1111/csp2.50.

[18] Sonia I. Seneviratne, Thierry Corti, Edouard L. Davin, Martin Hirschi, Eric B. Jaeger, Irene Lehner, Boris Orlowsky, and Adriaan J. Teuling. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, 99(3):125–161, 2010. ISSN 0012-8252. doi: https://doi.org/10.1016/j.earscirev.2010.02.004. URL https://www.sciencedirect.com/science/article/pii/S0012825210000139.

[19] W. A. Dorigo, W. Wagner, R. Hohensinn, S. Hahn, C. Paulik, A. Xaver, A. Gruber, M. Drusch, S. Mecklenburg, P. van Oevelen, A. Robock, and T. Jackson. The international soil moisture network: a data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences*, 15(5):1675–1698, May 2011. ISSN 1607-7938. doi: 10.5194/hess-15-1675-2011. URL http://dx.doi.org/10.5194/hess-15-1675-2011.

[20] Todd G. Caldwell, Michael H. Cosh, Steven R. Evett, Nathan Edwards, Heather Hofman, Bradley G. Illston, Tilden Meyers, Marina Skumanich, and Kent Sutcliffe. Soil moisture sensors in undisturbed soils. *Journal of Visualized Experiments*, (189), November 2022. ISSN 1940-087X. doi: 10.3791/64498. URL http://dx.doi.org/10.3791/64498.

[21] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5):10143, 1994. doi: 10.1029/94jc00572. URL https://doi.org/10.1029/94jc00572.

[22] Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble kalman filter. *Monthly Weather Review*, 126(6):1719–1724, June 1998. doi: 10.1175/1520-0493(1998)126⟨1719:asitek⟩2.0.co;2. URL https://doi.org/10.1175/1520-0493(1998)126<1719:asitek>2.0.co;2.

[23] Ming Pan and Eric F. Wood. Data assimilation for estimating the terrestrial water budget using a constrained ensemble kalman filter. *Journal of Hydrometeorology*, 7(3):534–547, June 2006. ISSN 1525-755X. doi: 10.1175/jhm495.1. URL http://dx.doi.org/10.1175/JHM495.1.

[24] Konstantinos M. Andreadis and Dennis P. Lettenmaier. Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, 29(6): 872–886, June 2006. ISSN 0309-1708. doi: 10.1016/j.advwatres.2005.08.004. URL http://dx.doi.org/10.1016/j.advwatres.2005.08.004.

[25] Michael Durand and Steven A. Margulis. Effects of uncertainty magnitude and accuracy on assimilation of multiscale measurements for snowpack characterization. *Journal of Geophysical Research: Atmospheres*, 113(D2), January 2008. ISSN 0148-0227. doi: 10.1029/2007jd008662. URL http://dx.doi.org/10.1029/2007JD008662.

[26] Sujay Kumar, Christa Peters-Lidard, Yudong Tian, Rolf Reichle, James Geiger, Charles Alonge, John Eylander, and Paul Houser. An integrated hydrologic modeling and data assimilation framework. *Computer*, 41(12):52–59, December 2008. ISSN 1558-0814. doi: 10.1109/mc.2008.475. URL http://dx.doi.org/10.1109/MC.2008.475.

[27] Jeffrey L. Anderson. A non-gaussian ensemble filter update for data assimilation. *Monthly Weather Review*, 138(11):4186–4198, November 2010. doi: 10.1175/2010mwr3253.1. URL https://doi.org/10.1175/2010mwr3253.1.

[28] Emmanuel C. Dibia, Rolf H. Reichle, Jeffrey L. Anderson, and Xin-Zhong Liang. Non-gaussian ensemble filtering and adaptive inflation for soil moisture data assimilation. *Journal of Hydrometeorology*, 24(6):1039–1053, June 2023. doi: 10.1175/jhm-d-22-0046.1. URL https://doi.org/10.1175/jhm-d-22-0046.1.

[29] Randal D. Koster and Max J. Suarez. Impact of land surface initialization on seasonal precipitation and temperature prediction. *Journal of Hydrometeorology*, 4(2):408–423, April 2003. doi: 10.1175/1525-7541(2003)4⟨408:iolsio⟩2.0.co;2. URL https://doi.org/10.1175/1525-7541(2003)4<408:iolsio>2.0.co;2.

[30] Sonia I. Seneviratne, Thierry Corti, Edouard L. Davin, Martin Hirschi, Eric B. Jaeger, Irene Lehner, Boris Orlowsky, and Adriaan J. Teuling. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, 99(3-4):125–161, May 2010. doi: 10.1016/j.earscirev.2010.02.004. URL https://doi.org/10.1016/j.earscirev.2010.02.004.

[31] C. Draper and R. Reichle. The impact of near-surface soil moisture assimilation at subseasonal, seasonal, and inter-annual timescales. *Hydrology and Earth System Sciences*,

19(12):4831–4844, December 2015. doi: 10.5194/hess-19-4831-2015. URL https://doi.org/10.5194/hess-19-4831-2015.

[32] Randal D. Koster, Paul A. Dirmeyer, Andrea N. Hahmann, Ruben Ijpelaar, Lori Tyahla, Peter Cox, and Max J. Suarez. Comparing the degree of land–atmosphere interaction in four atmospheric general circulation models. *Journal of Hydrometeorology*, 3(3):363–375, June 2002. doi: 10.1175/1525-7541(2002)003⟨0363:ctdola⟩2.0.co;2. URL https://doi.org/10.1175/1525-7541(2002)003<0363:ctdola>2.0.co;2.

[33] C. Kuenzer, D. Zhao, K. Scipal, D. Sabel, V. Naeimi, Z. Bartalis, S. Hasenauer, H. Mehl, S. Dech, and W. Wagner. El niño southern oscillation influences represented in ERS scatterometer-derived soil moisture data. *Applied Geography*, 29(4):463–477, December 2009. doi: 10.1016/j.apgeog.2009.04.004. URL https://doi.org/10.1016/j.apgeog.2009.04.004.

[34] Trent W. Ford, Steven M. Quiring, and Oliver W. Frauenfeld. Multi-decadal variability of soil moisture–temperature coupling over the contiguous united states modulated by pacific and atlantic sea surface temperatures. *International Journal of Climatology*, 37(3):1400–1415, June 2016. doi: 10.1002/joc.4785. URL https://doi.org/10.1002/joc.4785.

[35] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, March 1960. doi: 10.1115/1.3662552. URL https://doi.org/10.1115/1.3662552.

[36] Geir Evensen. Using the extended kalman filter with a multilayer quasi-geostrophic ocean model. *Journal of Geophysical Research: Oceans*, 97(C11):17905–17924, November 1992. doi: 10.1029/92jc01972. URL https://doi.org/10.1029/92jc01972.

[37] Rolf H. Reichle, Dennis B. McLaughlin, and Dara Entekhabi. Hydrologic data assimilation with the ensemble kalman filter. *Monthly Weather Review*, 130(1):103–114, January 2002. doi: 10.1175/1520-0493(2002)130⟨0103:hdawte⟩2.0.co;2. URL https://doi.org/10.1175/1520-0493(2002)130<0103:hdawte>2.0.co;2.

[38] Robert N. Miller, Michael Ghil, and François Gauthiez. Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of the Atmospheric Sciences*, 51(8):1037–1056, April 1994. doi: 10.1175/1520-0469(1994)051⟨1037:adaisn⟩2.0.co;2. URL https://doi.org/10.1175/1520-0469(1994)051<1037:adaisn>2.0.co;2.

[39] P. L. Houtekamer and Fuqing Zhang. Review of the ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144(12):4489–4532, November 2016. doi: 10.1175/mwr-d-15-0440.1. URL https://doi.org/10.1175/mwr-d-15-0440.1.

[40] M. Verlaan and A. W. Heemink. Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrology and Hydraulics*, 11(5):349–368, October 1997. doi: 10.1007/bf02427924. URL https://doi.org/10.1007/bf02427924.

[41] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F Radar and Signal Processing*, 140(2):107, 1993. doi: 10.1049/ip-f-2.1993.0015. URL https://doi.org/10.1049/ip-f-2.1993.0015.

[42] X. Xiong, I. M. Navon, and B. Uzunoglu. A note on the particle filter with posterior gaussian resampling. *Tellus A: Dynamic Meteorology and Oceanography*, 58(4):456, January 2006. doi: 10.1111/j.1600-0870.2006.00185.x. URL https://doi.org/10.1111/j.1600-0870.2006.00185.x.

[43] S. Nakano, G. Ueno, and T. Higuchi. Merging particle filter for sequential data assimilation. *Nonlinear Processes in Geophysics*, 14(4):395–408, July 2007. doi: 10.5194/npg-14-395-2007. URL https://doi.org/10.5194/npg-14-395-2007.

[44] Fuqing Zhang, Meng Zhang, and James A. Hansen. Coupling ensemble kalman filter with four-dimensional variational data assimilation. *Advances in Atmospheric Sciences*, 26(1):1–8, January 2009. doi: 10.1007/s00376-009-0001-8. URL https://doi.org/10.1007/s00376-009-0001-8.

[45] Rolf H. Reichle, Jeffrey P. Walker, Randal D. Koster, and Paul R. Houser. Extended versus ensemble kalman filtering for land data assimilation. *Journal of Hydrometeorology*, 3(6):728–740, December 2002. doi: 10.1175/1525-7541(2002)003⟨0728:evekff⟩2.0.co;2. URL https://doi.org/10.1175/1525-7541(2002)003<0728:evekff>2.0.co;2.

[46] Christian L. Keppenne. Data assimilation into a primitive-equation model with a parallel ensemble kalman filter. *Monthly Weather Review*, 128(6):1971–1981, June 2000. doi: 10.1175/1520-0493(2000)128⟨1971:daiape⟩2.0.co;2. URL https://doi.org/10.1175/1520-0493(2000)128<1971:daiape>2.0.co;2.

[47] Jeffrey L. Anderson. A marginal adjustment rank histogram filter for non-gaussian ensemble data assimilation. *Monthly Weather Review*, 148(8):3361–3378, August 2020. doi: 10.1175/mwr-d-19-0307.1. URL https://doi.org/10.1175/mwr-d-19-0307.1.

[48] Jeffrey L. Anderson and Stephen L. Anderson. A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12):2741–2758, December 1999. doi: 10.1175/1520-0493(1999)127⟨2741:amciot⟩2.0.co;2. URL https://doi.org/10.1175/1520-0493(1999)127<2741:amciot>2.0.co;2.

[49] Jeffrey S. Whitaker and Thomas M. Hamill. Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, 140(9):3078–3089, September 2012. doi: 10.1175/mwr-d-11-00276.1. URL https://doi.org/10.1175/mwr-d-11-00276.1.

[50] Jeffrey L. Anderson. Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena*, 230(1-2):99–111,

June 2007. doi: 10.1016/j.physd.2006.02.011. URL https://doi.org/10.1016/j.physd.2006.02.011.

[51] Rolf H. Reichle, Wade T. Crow, and Christian L. Keppenne. An adaptive ensemble kalman filter for soil moisture data assimilation. *Water Resources Research*, 44(3), March 2008. doi: 10.1029/2007wr006357. URL https://doi.org/10.1029/2007wr006357.

[52] Jeffrey L. Anderson. Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A*, 61(1):72–83, January 2009. doi: 10.1111/j.1600-0870.2008.00361.x. URL https://doi.org/10.1111/j.1600-0870.2008.00361.x.

[53] Mohamad El Gharamti. Enhanced adaptive inflation algorithm for ensemble filters. *Monthly Weather Review*, 146(2):623–640, February 2018. doi: 10.1175/mwr-d-17-0187.1. URL https://doi.org/10.1175/mwr-d-17-0187.1.

[54] Randal D. Koster, Max J. Suarez, Agnès Ducharne, Marc Stieglitz, and Praveen Kumar. A catchment-based approach to modeling land surface processes in a general circulation model: 1. model structure. *Journal of Geophysical Research: Atmospheres*, 105(D20): 24809–24822, October 2000. doi: 10.1029/2000jd900327. URL https://doi.org/10.1029/2000jd900327.

[55] Rolf H. Reichle, Qing Liu, Randal D. Koster, Wade T. Crow, Gabrielle J. M. De Lannoy, John S. Kimball, Joseph V. Ardizzone, David Bosch, Andreas Colliander, Michael Cosh, Jana Kolassa, Sarith P. Mahanama, John Prueger, Patrick Starks, and Jeffrey P. Walker. Version 4 of the SMAP level-4 soil moisture algorithm and data product. *Journal of Advances in Modeling Earth Systems*, 11(10):3106–3130, October 2019. doi: 10.1029/2019ms001729. URL https://doi.org/10.1029/2019ms001729.

[56] Ronald Gelaro, Will McCarty, Max J. Suárez, Ricardo Todling, Andrea Molod, Lawrence Takacs, Cynthia A. Randles, Anton Darmenov, Michael G. Bosilovich, Rolf Reichle, Krzysztof Wargan, Lawrence Coy, Richard Cullather, Clara Draper, Santha Akella, Virginie Buchard, Austin Conaty, Arlindo M. da Silva, Wei Gu, Gi-Kong Kim, Randal Koster, Robert Lucchesi, Dagmar Merkova, Jon Eric Nielsen, Gary Partyka, Steven Pawson, William Putman, Michele Rienecker, Siegfried D. Schubert, Meta Sienkiewicz, and Bin Zhao. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, 30(14):5419–5454, July 2017. doi: 10.1175/jcli-d-16-0758.1. URL https://doi.org/10.1175/jcli-d-16-0758.1.

[57] Rolf H. Reichle and Randal D. Koster. Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. *Journal of Hydrometeorology*, 4(6):1229–1242, December 2003. doi: 10.1175/1525-7541(2003)004⟨1229:atiohe⟩2.0.co;2. URL https://doi.org/10.1175/1525-7541(2003)004<1229:atiohe>2.0.co;2.

[58] Andreas Colliander, Rolf H. Reichle, Wade T. Crow, Michael H. Cosh, Fan Chen, Steven Chan, Narendra Narayan Das, Rajat Bindlish, Julian Chaubell, Seungbum Kim, Qing Liu, Peggy E. O'Neill, R. Scott Dunbar, Land B. Dang, John S. Kimball, Thomas J. Jackson,

Hala Khalid Al-Jassar, Jun Asanuma, Bimal K. Bhattacharya, Aaron A. Berg, David D. Bosch, Laura Bourgeau-Chavez, Todd Caldwell, Jean-Christophe Calvet, Chandra Holifield Collins, Karsten H. Jensen, Stan Livingston, Ernesto Lopez-Baeza, Jose Martinez-Fernandez, Heather McNairn, Mahta Moghaddam, Carsten Montzka, Claudia Notarnicola, Thierry Pellarin, Isabella Greimeister-Pfeil, Jouni Pulliainen, Judith Gpe. Ramos, Mark Seyfried, Patrick J. Starks, Zhongbo Su, R. van der Velde, Yijian Zeng, Marc Thibeault, Mariette Vreugdenhil, Jeffrey P. Walker, Mehrez Zribi, Dara Entekhabi, and Simon H. Yueh. Validation of soil moisture data products from the NASA SMAP mission. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:364–392, 2022. doi: 10.1109/jstars.2021.3124743. URL https://doi.org/10.1109/jstars.2021.3124743.

[59] Jeffrey Anderson, Tim Hoar, Kevin Raeder, Hui Liu, Nancy Collins, Ryan Torn, and Avelino Avellano. The data assimilation research testbed: A community facility. *Bulletin of the American Meteorological Society*, 90(9):1283–1296, September 2009. doi: 10.1175/2009bams2618.1. URL https://doi.org/10.1175/2009bams2618.1.

[60] C. S. Draper, R. H. Reichle, G. J. M. De Lannoy, and Q. Liu. Assimilation of passive and active microwave soil moisture retrievals. *Geophysical Research Letters*, 39(4):n/a–n/a, February 2012. doi: 10.1029/2011gl050655. URL https://doi.org/10.1029/2011gl050655.

[61] Sujay V. Kumar, Rolf H. Reichle, Randal D. Koster, Wade T. Crow, and Christa D. Peters-Lidard. Role of subsurface physics in the assimilation of surface soil moisture observations. *Journal of Hydrometeorology*, 10(6):1534–1547, December 2009. doi: 10.1175/2009jhm1134.1. URL https://doi.org/10.1175/2009jhm1134.1.

[62] Dara Entekhabi, Eni G. Njoku, Peggy E. O'Neill, Kent H. Kellogg, Wade T. Crow, Wendy N. Edelstein, Jared K. Entin, Shawn D. Goodman, Thomas J. Jackson, Joel Johnson, John Kimball, Jeffrey R. Piepmeier, Randal D. Koster, Neil Martin, Kyle C. McDonald, Mahta Moghaddam, Susan Moran, Rolf Reichle, J. C. Shi, Michael W. Spencer, Samuel W. Thurman, Leung Tsang, and Jakob Van Zyl. The soil moisture active passive (smap) mission. *Proceedings of the IEEE*, 98(5):704–716, May 2010. ISSN 1558-2256. doi: 10.1109/jproc.2010.2043918. URL http://dx.doi.org/10.1109/JPROC.2010.2043918.

[63] Wade T. Crow, Jianzhi Dong, and Rolf H. Reichle. Leveraging pre-storm soil moisture estimates for enhanced land surface model calibration in ungauged hydrologic basins. *Water Resources Research*, 58(8), August 2022. ISSN 1944-7973. doi: 10.1029/2021wr031565. URL http://dx.doi.org/10.1029/2021WR031565.

[64] Chen Zhang, Zhengwei Yang, Haoteng Zhao, Ziheng Sun, Liping Di, Rajat Bindlish, Pang-Wei Liu, Andreas Colliander, Rick Mueller, Wade Crow, Rolf H. Reichle, John Bolten, and Simon H. Yueh. Crop-casma: A web geoprocessing and map service based architecture and implementation for serving soil moisture and crop vegetation condition data over u.s. cropland. *International Journal of Applied Earth Observation and Geoinformation*, 112:

102902, August 2022. ISSN 1569-8432. doi: 10.1016/j.jag.2022.102902. URL http://dx.doi.org/10.1016/j.jag.2022.102902.

[65] Rolf H. Reichle, Sara Q. Zhang, Jana Kolassa, Qing Liu, and Ricardo Todling. A ¡scp¿weakly coupled¡/scp¿ land surface analysis with ¡scp¿smap¡/scp¿ radiance assimilation improves ¡scp¿geos¡/scp¿ medium-range forecasts of near-surface air temperature and humidity. *Quarterly Journal of the Royal Meteorological Society*, 149(754):1867–1889, June 2023. ISSN 1477-870X. doi: 10.1002/qj.4486. URL http://dx.doi.org/10.1002/qj.4486.

[66] Peggy E. ONeill, Steven Chan, Eni G Njoku, Tom Jackson, Rajat Bindlish, M. Julian Chaubell, and Andreas Colliander. Smap enhanced l3 radiometer global and polar grid daily 9 km ease-grid soil moisture, version 5, 2021. URL https://nsidc.org/data/spl3smp_e/versions/5.