Multivariate Hydrologic Data Assimilation for Model Structural Learning and Process-Diagnostics

Grey S. Nearing (PI)
University of Alabama at Tuscaloosa
Department of Geological Sciences
gsnearing@ua.edu

Christa D. Peters-Lidard (Co-PI)
NASA Goddard Space Flight Center
Earth Science Division
christa.d.peters-lidard@nasa.gov

Soni Yatheendradas (Co-I)
NASA Goddard Space Flight Center
Hydrological Sciences Laboratory
soni.yatheendradas-1@nasa.gov

Wade T. Crow (Collaborator)
USDA Agricultural Research Service
Hydrology and Remote Sensing Lab
wade.crow@ars.usda.gov

Table of Contents

Scientific/Technical/Management

Objectives and Expected Significance

It has often been suggested that data assimilation (DA) might be useful for helping scientists to improve land surface and other dynamical Earth systems models (e.g., van den Hurk et al., 2011, Liu and Gupta, 2007). This seems feasible because DA is quite successful at extracting information from remote sensing data to improve *simulations* from imperfect models (*i.e.,* to update model states and/or model parameters), and therefore it is reasonable to conclude that observation data contain information that might allow us to improve the models themselves. So far, this capability has not been realized in any systematic way, and the land modeling community lacks a general methodology for using DA to help diagnose or improve model structures.

This project will provide the essential foundation (theory, methods, code base, and examples) for using data assimilation directly for model structural learning and diagnostics. First, a dynamic correction factor is developed through data assimilation, and this correction factor is demonstrated to result in improved model simulations. Second, a set of novel diagnostic metrics is applied to this statistical correction factor to identify specific compensating error structures between individual processes and couplings in the model.

This project will focus on identifying compensating error structures in model representations of processes like infiltration, turbulent exchange, root water uptake, and soil moisture controls on stomatal resistance and photosynthesis. This will be done using a structural learning and process diagnostic approach to assimilate both in situ and remote sensing data related to (i) soil moisture, (ii) evapotranspiration, (iii) photosynthesis (fluorescence), and (iv) surface carbon fluxes. Including vegetation-related observation data along with hydrology-related data is essential because most modern land models simulate vegetation dynamically. Including dynamic vegetation in terrestrial hydrology models has been shown to add significant skill to predicting various components of the terrestrial water cycle (e.g., Yang et al., 2011). Our objective is to use directly improve the structure and dynamics of terrestrial hydrology models in a way that leads to more realistic simulations of coupled ecohydrological systems.

In addition, the approach that we propose toward generalized data assimilation will allow us to extract more information from satellite retrievals. As an example, our recent NASA THP-funded exploratory project found that an ensemble Kalman filter (EnKF) was able to extract less than half of the information content of AMSR-E soil moisture retrievals, due to nonlinear and non-Gaussian biases and statistical error structures in the retrievals. The DA method that we propose handles arbitrary (nonparametric) bias and error in models and retrievals.

<u>The primary project objective is to develop a general DA strategy that can:</u>

1. Systematically use remote sensing retrievals to directly enable structural improvements and process diagnostics in land models.
2. Account for nonlinear, non-Gaussian, and time/space-dependent biases and error structures in assimilated retrievals to avoid information loss during DA.

Technical Approach and Methodology

Background & Challenges

Theory and methods for using data assimilation to facilitate model structural learning exist, and have been applied to assimilate streamflow data into rainfall/runoff models (Nearing and Gupta, 2015, Wilkinson et al., 2011, Bulygina and Gupta, 2011). This project will extend those methods to land surface models for multivariate data assimilation.

We define data assimilation here as the application of Bayes' theorem to estimate the time-dependent state of a dynamic system (*e.g.,* Wikle and Berliner, 2007). The general form of a DA smoother is:

$$p_a(X_{1:t}|Y_{1:t}) \propto p_h(Y_{1:t}|X_{1:t})p_m(X_{1:t}). \qquad [1.1]$$

$X_t \in \mathbb{R}^{d_x}$ and $Y_t \in \mathbb{R}^{d_y}$ are the model state and assimilated observations respectively at time t. The Bayesian prior distribution p_m represents uncertainty related to a simulation model $\mathscr{m}$, and the Bayesian likelihood distribution p_h represents uncertainty related to an observation or retrieval model $\mathscr{h}$. The analysis posterior p_a is the best probabilistic estimate of the state of the system given the combined information from model and observations.

In hydrology, DA *filters* are more common than DA *smoothers*. A DA filter relies on certain Markov approximations to reduce the dimension of the analysis posterior. The general form of a filter is:

$$p_a(X_t|Y_{1:t}) \propto p_h(Y_t|X_t)p_m(X_t|X_{t-1}, Y_{1:t-1}). \qquad [1.2]$$

Our proposed work will use partial DA smoothers to avoid discontinuities in the analysis time series – this requirement is discussed in detail by Bulygina & Gupta (2009).

The two main challenges to implementing equations [1.1] and [1.2] are:

1. To identify the model and retrieval error distributions p_m and p_h, and
2. To either sample (e.g., ensemble DA methods) or maximize (e.g., variational DA methods) the product of these distributions in a manner that is both computationally and statistically efficient.

In current land DA systems, the sampling or variational (maximization) problem is often solved analytically by assuming a particular parametric form of statistical error. For example, the EnKF assumes that errors in both the model and retrievals are all zero-mean Gaussian with covariances that are either known (*e.g.,* Kumar et al., 2014, Vrugt et al., 2006) or estimated (*e.g.,* Crow and Reichle, 2008, Reichle et al., 2008). In cases where the retrieval errors are not actually zero-mean, cumulative density function (CDF) matching is used to shift the retrieval distribution (Kumar et al., 2012), but can remove valuable information about systematic model errors from the retrievals. There are various partial generalizations of this type of approach (e.g., Zupanski, 2005, Feraud et al., 2011), but all rely fundamentally on parametric approximations of the model and retrieval error structures.

Both a priori bias correction (e.g., CDF-matching) and different types of mis-specified statistical error distributions can result in substantial information loss during land data assimilation. We recently quantified information loss in soil moisture DA for an exploratory project sponsored by the NASA Terrestrial Hydrology Program, and found that the EnKF loses more than half of the available information from AMSR-E soil moisture retrievals. This is in line our other recent work on quantifying information loss for soil moisture DA (Nearing et al., 2013a, Nearing et al., 2013b, Crow and Van Loon, 2006).

In some cases, disagreements between the model and retrieval uncertainty distributions (p_m and p_h) are so large that DA gives spurious results. This is a notable problem when assimilating certain types of flux observations into dynamic vegetation components of terrestrial hydrology models. **Figure 1** shows example results from assimilating FluxNet observations of net ecosystem exchange (NEE) into the Noah-MP land model (Niu et al., 2011) using the EnKF to update water and carbon storage states in the soil and canopy. The two subplots show two different FluxNet sites – at one site NEE DA caused statistically significant improvements to model-estimated NEE (shown) and model-estimated latent and sensible heat fluxes (not shown), and at another site NEE DA degraded these same variables. The poor results at the forested site are due to highly non-Gaussian NEE observation errors, and also to large errors in the modeled process couplings between water and carbon variables (this particular model structural problem is illustrated in **Figure 4**, which is discussed later in this proposal).

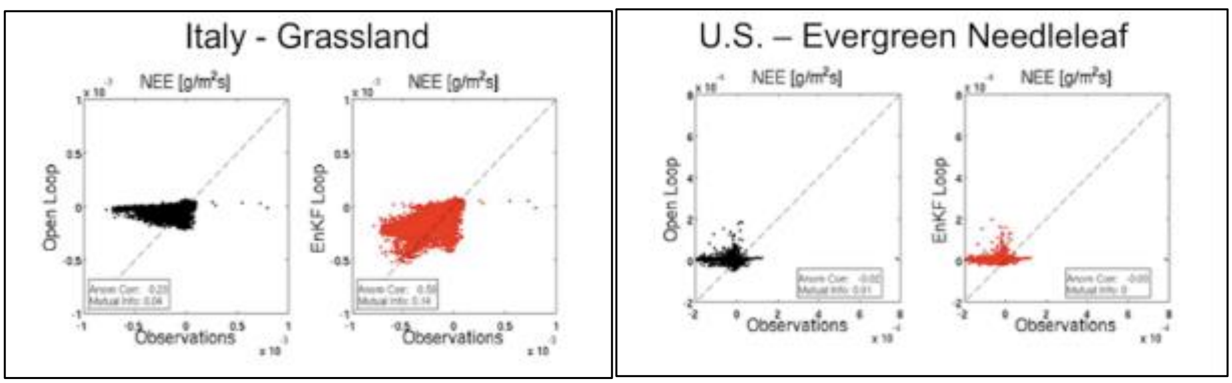

***Figure 1:** Examples of assimilating NEE data from two FluxNet sites into the Noah-MP land model to update soil and canopy water and carbon states. DA improved accuracy of modeled fluxes at the Grassland site, but degraded accuracy at the forest site. This inconsistency is due to differences in process-level model error at the two sites.*

This figure is representative of a larger body of results from several similar DA experiments at many global FluxNet sites in different eco-climates. These experiments were performed by the PI as part of an ongoing NASA-funded project to use quantum computing to facilitate computationally efficient carbon flux DA in terrestrial hydrology models.

In general, the couplings between dynamic vegetation and water cycle process components in terrestrial hydrology models show strongly nonlinear and non-stationary error patterns that are difficult to capture in the types of statistical error distributions that are required for typical DA applications. The proposed project will address this problem directly by using DA to both identify the underlying process-level error structures in the model, and also by estimating the resulting non-parametric and nonstationary structures of both model and retrieval uncertainty distributions.

To summarize, we plan to attack three general issues that are present in many types of land DA problems:

- First, a priori bias correction, and other types of **imperfect representations of model and retrieval errors result in (sometimes significant) information loss during DA.** This is especially undesirable when the observation data come from remote sensing platforms that are expensive to develop, launch, and maintain.
- Second, **diverse and unknown model error structures can result in spurious assimilation updates, actually degrading model performance**. This can be a significant problem when working directly with the dynamic vegetation components of modern terrestrial hydrology models like Noah-MP due to large and site-specific model errors related to the couplings between vegetation and water states.
- Third, DA currently doesn't tell us what is wrong with our models. **There is no systematic methodology to translate the information extracted from observation data by DA into improved understanding of modeled processes** or model deficiencies.

Particle filters can work with non-Gaussian model and retrieval error distributions, but there are two reasons why using particle filters does not address the underlying problem. First, particle filters require us to specify both the model and retrieval uncertainty distributions ($p_{\mathcal{m}}$ and $p_{\hbar}$), and hence require a priori knowledge of corresponding bias and statistical error distributions. Second, particle filters suffer from dimensionality problems resulting in computationally and statistically inefficient sampling (Snyder et al., 2008). Sampling inefficiency can be mitigated using proposal distributions (*e.g.,* van Leeuwen, 2010), but this does not address the fundamental source of disagreement between incorrect or over-simplified model and retrieval uncertainty distributions[1]. That is, **particle filters can *deal with* arbitrary error structures, but they cannot not help us *learn* about non-Gaussian model or retrieval errors, or to improve model structures using information extracted from DA**.

Several DA filtering techniques proposed in the hydrology literature can learn statistical error distributions "on-line" (i.e., during DA), but these so far have been restricted to stationary, linear, and Gaussian (Crow and Reichle, 2008). **Simply learning the covariance structures of the stationary component of model and/or retrieval error tells us little about the compensating error structures of process-based models.** For this we need a more general DA learning approach that explicitly represents the time-dynamics, process-dynamics, and nonlinearities in the modeled system.

[1] Sampling inefficiency in a Bayesian inference problem like DA occurs when the majority of samples of X from $p_{\mathcal{m}}$ return low probability on the actual retrieval Y according to $p_{\hbar}$. This type of disagreement is due either to incorrectly specified model and/or retrieval error distributions or simply to error in the dynamic model $\mathcal{m}$ and/or retrieval model $\hbar$ themselves.

Proposed Approach

Proposed Approach – Data Assimilation Theory

The DA theory developed by Ghahramani & Roweis (1999) is able to learn model error structures that are arbitrary, unknown, and heterogeneous and/or heteroscedastic. This DA theory has been widely adapted, extrapolated, and applied to many different types of model learning problems in many areas of science and engineering in general (e.g., Wang et al., 2008, Schön et al., 2011, Turner et al., 2009), including streamflow forecasting (Nearing and Gupta, 2015, Wilkinson et al., 2011, Bulygina and Gupta, 2011, Bulygina and Gupta, 2010, Bulygina and Gupta, 2009). However, aside from a recent small exploratory project funded by the NASA Earth Science Technology Office (ESTO), this theory has not – to our knowledge – been applied to either atmosphere or land models, or to assimilate any remote sensing retrievals. Certain results from the NASA ESTO feasibility study are presented in the *Proposed Approach – Data Assimilation Examples* subsection below.

The theory is built around expectation-maximization (EM). A nonparametric correction factor is added to the state transition function of the dynamical simulation model, and/or to the retrieval operator. Typically, the state transition function for DA takes a form like:

$$X_t = \mathcal{m}(X_{t-1}, U_t, \theta_m) + \varepsilon_t, \qquad [2]$$

where U_t are time-dependent boundary conditions or forcing data, θ_m are model parameters, and $\varepsilon_t \in \mathbb{R}^{d_x}$ is drawn from a (possibly time-dependent) error distribution. Instead, the form of our proposed state transition function is:

$$X_t = \mathcal{m}(X_{t-1}, U_t, \theta_m) + \mathcal{g}_m(X_{t-1}, U_t, \theta_m), \qquad [3.1]$$

where $\mathcal{g}_m \in \mathbb{R}^{d_x}$ is a dynamic correction function that takes as inputs all of the information that we currently have about the modeled system. Unlike a purely statistical model error term, this correction function acts on the same inputs as the model itself, and therefore contributes to the information flow between individual processes and couplings within the model.

Similarly, a correction function may be added to the retrieval operator as:

$$Y_t = \mathcal{h}(X_t, A_t, \theta_h) + \mathcal{g}_h(X_t, A_t, \theta_h), \qquad [3.2]$$

where A_t and θ_h are vectors of time-dependent ancillary data and time-independent retrieval parameters, respectively, and $\mathcal{g}_h \in \mathbb{R}^{d_y}$ is a retrieval correction function. Unlike in equation [3.1], equation [3.2] has no time dynamics, however it still has the ability to represent retrieval bias conditional on all available information about the current state of the system.

In our expectation-maximization data assimilation (EM-DA) strategy, the arbitrary correction functions are expanded using basis functions:

$$\mathcal{g}_m(X_{t-1}, U_t, \theta_m) = \sum_{b=1}^{N_b} \alpha_b^m g(X_{t-1}, U_t, \theta_m, \mathcal{H}_b^m)\text{, and} \qquad [4.1]$$

$$\mathcal{g}_h(X_t, A_t, \theta_h) = \sum_{b=1}^{N_b} \alpha_b^h g\left(X_t, A_t, \theta_h, \mathcal{H}_b^h\right). \qquad [4.2]$$

$g(.)$ is some as-yet-unspecified basis function and $\{\alpha_b^m, \mathcal{H}_b^m\}, \{\alpha_b^{\hbar}, \mathcal{H}_b^{\hbar}\}$ are sets of hyperparameters related to $\mathcal{g}_m$ and $\mathcal{g}_{\hbar}$ respectively. The state variable X in equation [1] is treated as a hidden variable, and the hyperparameters are optimized over this hidden variable by EM. The E-step takes the expectation of model state X given a current guess for the hyperparameters and the M-step then maximizes the likelihood of the hyperparameters with respect to the E-step values of *X*. In general, this iterative EM approach is guaranteed to converge to correct values of hyperparameters and hidden model states (Dempster et al., 1977). Notice that the E-step and M-step are analogous to the sampling and uncertainty estimation parts of the general DA problem outlined in the preceding subsection. Our EM strategy addresses these two parts of the DA problem jointly through this iterative procedure.

Equations [4.1] and [4.2] can be employed with more or less fidelity by choosing the (finite) number of basis functions N_b. A larger number of basis functions generally requires more assimilation data for robust learning, but provides more resolution to the nonlinear model and retrieval error structures. This tradeoff is also application-dependent, and assessing problem-specific relationships between length of terrestrial remote sensing data records and achievable fidelity of correction factors for land models will be a major thrust of the experimental work in this project.

The central goal of EM-DA is to learn the structure(s) of $\mathcal{g}_m$ and/or $\mathcal{g}_{\hbar}$, and thus to improve hydrologic simulations by improving hydrologic model itself, rather than simply by updating the state of the simulation. In the latter – which is what standard state-updating DA does – the model reverts back to the same patterns of simulation error during forecast periods, or when there are no observations to assimilate. This is also different than using DA to update model parameters. The problem with parameters estimation (by DA or otherwise) is that if the model structure is wrong, then the resulting parameterized model is kludged (Clark, 1987) – meaning that it does not necessarily *get the right answers for the right reasons*, and therefore cannot reasonably be expected to provide realistic simulation data under nonstationary conditions (Kirchner, 2006).

Our EM-DA approach differs from standard DA, where the primary goal is to estimate the state X_t or analysis distribution p_a. **The goal in EM-DA is that after a period of learning, the resulting dynamical correction factors can be used for (1) model diagnostics and to guide model/retrieval improvements, and (2) to understand model and retrieval uncertainty and bias.** The following subsections illustrate how this works:

- The *Proposed Approach – Data Assimilation Examples* subsection provides two hydrology-related examples of how this strategy leads directly to **improved model predictions, even in long-term simulation or forecast periods with no additional observation data**.
- The *Proposed Approach – Diagnostic Metrics* subsection outlines a quantitative strategy for obtaining process-level diagnostic information from the learned $\mathcal{g}_m$ and $\mathcal{g}_{\hbar}$ correction functions.

Proposed Approach – Data Assimilation Examples

As noted above, EM-DA has been applied several times to rainfall-runoff models to improve streamflow forecasts. **Figure 2** shows example results from experiments by Nearing & Gupta (2015), and qualitatively similar results were given by Bulygina & Gupta (2011) and Wilkinson et al, (2011).

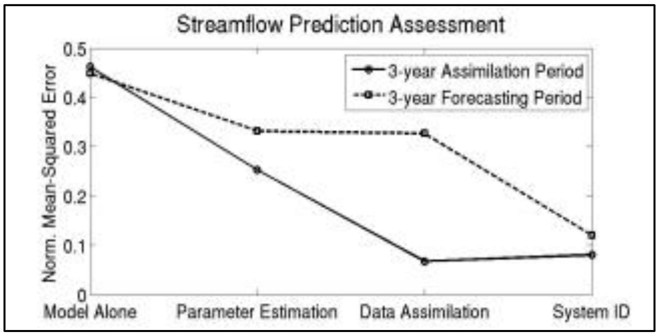

***Figure 2**: Mean squared errors of streamflow estimates made by a rainfall-runoff model at a watershed in Arkansas, USA. DA alone does not provide long-term forecast improvement because it does not address underlying causes of model error or bias. Expectation-maximization DA (labeled 'System ID' on the x-axis) translates into long-term improvement, even during periods where no data are assimilated.*

In this figure, mean squared errors (MSE) are compared between an un-calibrated conceptual rainfall-runoff model, a calibrated version of the same model, standard EnKF streamflow data assimilation, and EM-DA (labeled *system ID* in the figure). Daily streamflow data were assimilated over a 3-year observation period, and MSE statistics were estimated for a different 3-year period with no DA and no parameter estimation. The EM-DA procedure is seen to produce improved simulations with better streamflow predictions than even the calibrated model, even during a long (3-year) period where no observations were assimilated. **The difference between this and standard DA is that the information gained from assimilated observations was stored in the dynamics of the predictive model, and was therefore useful outside of simply updating initial forecast states. Importantly, this means that EM-DA has changed the internal time-dependent information flows within the simulation model, and quantifying such changes will give us insight into process-level error structures in the original simulation model.**

In the afore-mentioned exploratory project funded by NASA ESTO, we applied EM-DA to assimilate in situ observations of surface (5 cm) soil moisture measured with time-domain reflectometry probes to update root-zone soil moisture states in the Noah-MP land model. **Figure 3** here shows the root mean squared error (RMSE) of the Noah-MP model at estimating root-zone soil moisture (loosely defined as integrated over a 0-50 cm depth) first using default parameters from the North American Land Data Assimilation System (NLDAS; Xia et al., 2011), then with parameter estimation, and then with DA and EM-DA. EM-DA not only resulted in improved soil moisture states as compared to regular DA by the EnKF in both the calibrated and un-calibrated versions of the Noah-MP model, but this improvement persisted through an independent 1-year evaluation period where no new observations were assimilated. Here EM-DA was able to 'store' information extracted from the soil moisture data by updating the structure of the internal dynamics of the predictive model.

EM-DA is significantly more computationally expensive than the EnKF, but makes better use of the observation information. When this information comes from expensive remote sensing platforms, it is at least feasible that the tradeoff between increased computational expense and decreased information loss is justifiable – at least for some applications.

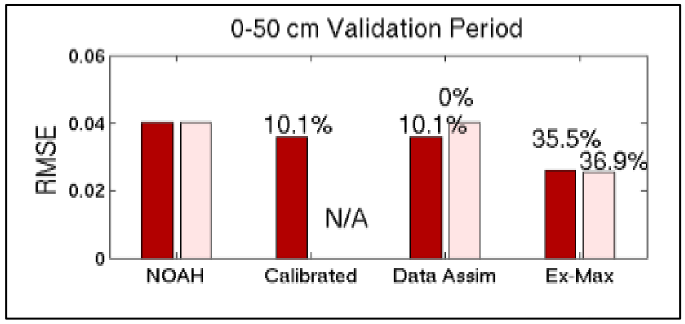

***Figure 3**: Anomaly root mean squared errors from calibrated (red) and un-calibrated (pink) Noah-MP land model simulations with DA by the EnKF and by expectation-maximization. Soil moisture observations at 5 cm depth were assimilated to update the soil water storage states over a total depth of 2 m during a three-year period. The statistics plotted here are from a subsequent 1-year period where no data were assimilated. Standard DA has no effect on model performance outside of the DA period, but expectation-maximization provides a ~35% improvement to anomaly RMSE as compared to the un-calibrated model.*

Proposed Approach – Diagnostic Metrics

The foundation of our approach to using EM-DA for process-level diagnostics is to quantify the way EM-DA changes patterns of information transfer within the terrestrial model. The basic tools for this are the dynamic process networks (DPN) that were proposed by Ruddell & Kumar (2009). A DPN represents a dynamic system as a Bayesian network, whereby each node in the network represents a system variable, and each edge in the network represents influence that one variable exerts on another variable at a certain timescale.

Influence between variables is quantified by measuring information transfers. To calculate the influence that one variable x_i (which could, for example, be a component of the state vector X) has on another variable x_j at a particular timescale τ, we integrate over the expected effect of probabilistically conditioning x_j at time $t+\tau$ on the value of x_i at time t given all of the variables in the model other than x_i at time t. Note that even in deterministic models each variable is probabilistic *conditional on only a subset of other variables*. Schreiber (2000) proposed a feasible Markov approximation of this integration that results in the following metric, called *transfer entropy*:

$$I(x_{j,t+\tau}; x_{i,t}|x_{j,t}) = \int p(x_{j,t}) \int \int p(x_{j,t+\tau}, x_{i,t}|x_{j,t}) \ln\left(\frac{p(x_{j,t+\tau}|x_{i,t}, x_{j,t})}{p\left((x_{j,t+\tau}|x_{j,t})\right)}\right) dx_{j,t+\tau}\, dx_{i,t}\, d\,x_{j,t}. \quad [5]$$

Two separate DPNs are constructed to represent all prognostic and all diagnostic variables in model m before and after EM-DA, with each directed edge between every pair of variables in the DPN quantified using equation [5]. Then, a third DPN is constructed to represent the differences between the prior and posterior model DPNs. This third difference network represents the (positive and/or negative) differences in information transfers among internal model variables that result from EM-DA.

All probability distributions in equation [5] are therefore derived empirically from modeled data, and this metric can be applied at any spatiotemporal scale by changing τ (which can represent spatial as well as temporal lags) and by changing the limits of integration. We will develop transfer entropy difference networks at many different spatiotemporal scales, to assess changes in how fixed-timestep simulation models represent larger patterns of behavior before vs. after EM-DA.

A simplified example of this type of analysis is given in **Figure 4**[2]. This figure shows differences between transfer entropy metrics calculated over Noah-MP simulations vs. similar metrics calculated directly from FluxNet observation data. This is slightly different than the type of analysis that will be used in this project – specifically, **Figure 4** shows differences between observed vs. modeled DPNs, whereas we will look at differences between modeled DPNs before vs. after structural updating by EM-DA. The main value of EM-DA for this type of process network analysis is that EM-DA allow us to account for observation uncertainty. Additionally, EM-DA allows us to assess the value of assimilating sparse and/or incomplete observation data sets into a model – it does not require that we directly observe every variable in the DPN. Using EM-DA to create difference networks like in Figure 4, allows us to ask the question *how much can we learn about internal process couplings given whatever partial data we have available for data assimilation*. Without EM-DA it would be difficult to construct DPNs directly from remote sensing data. Despite these differences, **Figure 4** is representative of one of the main types of illustrations that will be produced in this project.

Figure 4 also shows that Noah-MP exhibits a clear disconnect between hydrology and vegetation at the seasonal timescale at this particular FluxNet site, which is the US Evergreen site from **Figure 1**. **First-order improvements to the model at this site would target the role of vegetation on mediating Q_e (*e.g.,* the soil moisture controls on stomatal resistance) and soil moisture (*e.g.,* the root water uptake functions).**

One of the things that we will specifically evaluate in this project is whether this type of result (strong decoupling between vegetation and soil hydrology) is consistent across different eco-climates and model configurations, and whether the response surfaces of the EM-DA correction functions related to these couplings is generally consistent across different eco-climates. If this is the case, it will be indicative of general systematic error in the structure of the coupled ecohydrology model. On the other hand, if the EM-DA correction functions related to these decouplings have difference response surfaces at different sites, then this will indicate that the model structure is not sufficiently flexible to represent diversity in ecohydrological couplings.

[2] This figure comes from a presentation at the American Meteorological Society conference in 2017. A recording of this talk is available here: *ams.confex.com/ams/97Annual/webprogram/Paper312228.html*

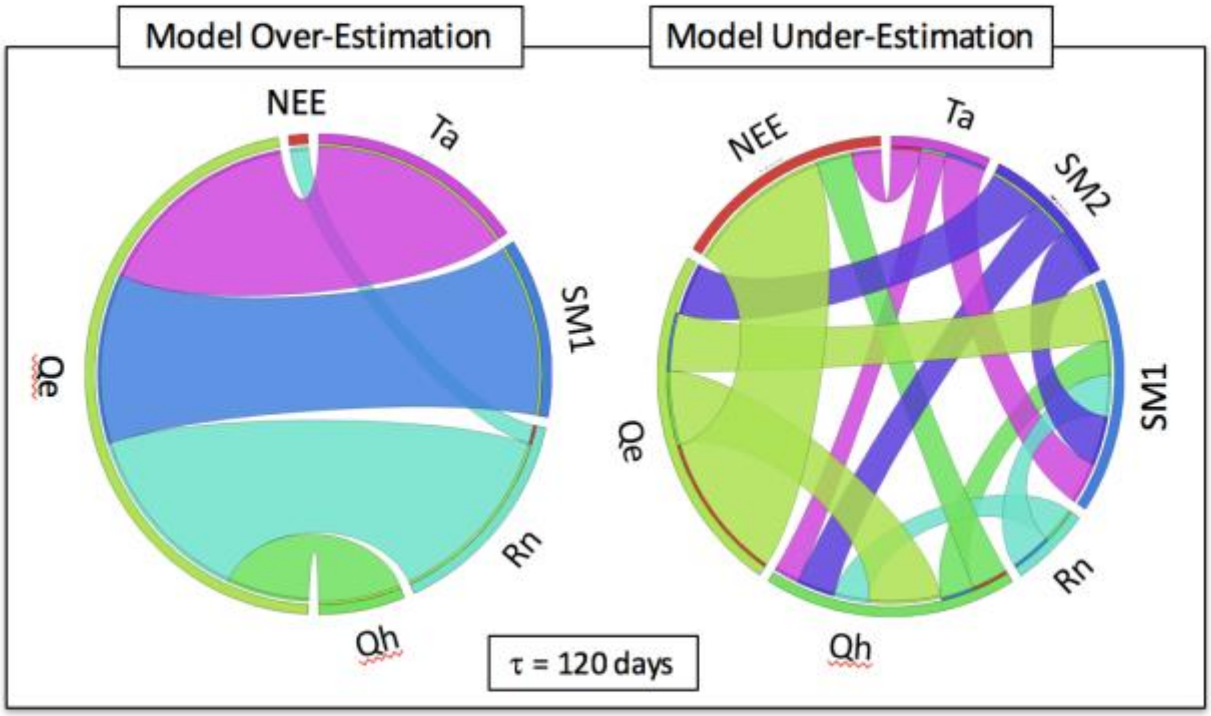

Figure 4: *Transfer entropy differences between DPNs estimated from Noah-MP simulations vs. those estimated directly from FluxNet data at a seasonal timescale. One thing that we see here is that the model over-estimates connections from energy to evapotranspiration, but under-estimates connections between water and carbon.*

R_n = net radiation
T_a = 2 m air temperature
SM1 = surface soil moisture
SM2 = sub-surface soil moisture
NEE = net ecosystem exch.
Qh = sensible heat flux
Qe = latent heat flux

Project Workflow

Project Workflow – Infrastructure Development

The first part of our work will be to develop the multivariate EM-DA system around the Structure for Unifying Multiple Modeling Alternatives (SUMMA; Clark et al., 2015b, Clark et al., 2015c, Clark et al., 2015a). SUMMA is an ideal land model for EM-DA because of its ability to isolate individual biogeophysical processes (and hence assess process sensitivity). That is, SUMMA includes options for different equations that represent various individual biogeophysical processes like albedo, turbulent exchange, stomatal resistance, root water uptake, snowmelt, etc. – including all of the ecohydrological processes mentioned in the previous subsections. In total there are ~30 individual configuration options for SUMMA, resulting in $O(10^9)$ possible land models, each with different configurations of individual process representations. As an example, one of the ~30 process options in SUMMA allows the user to choose from a Ball-Berry or Jarvis style stomatal resistance function.

Our EM-DA land data assimilation system (EM-LDAS) built around SUMMA will be capable of assimilating multiple different types of remote sensing and in situ observations to update the plant and soil states, to update SUMMA state-transition functions. SUMMA was designed from the ground up to interact easily with DA systems – specifically, the prognostic state vectors for each process parameterization are collected and isolated at key points in the code. The PI and Co-PI of this project were recently involved in a different NASA-funded project where SUMMA was ported into the NASA Land Information System.

Effective model structural learning must account for parameter uncertainty to avoid updating model structure when only parameter updates are necessary (in much the same way as before, this is necessary to avoid kludging). Thus, robust parameter estimation a necessary precursor to EM-DA. More specifically, it is necessary in EM-DA to marginalize over uncertain parameters, and so parameter *optimization* is not sufficient. Instead, it is necessary to estimate full posterior parameter distributions conditional on observation data. Thus, it will be necessary to first implement SUMMA into a Markov Chain Monte Carlo (MCMC) parameter estimation framework. We will implement the DiffeRential Evolution Adaptive Metropolis (DREAM; Vrugt, 2016) in a parallel computing environment. Members of this proposal team (PI and Co-PI) recently completed an extensive parameter sensitivity analysis of SUMMA as part of a NASA-funded project, and our team also has extensive previous experience implementing probabilistic parameter estimation in land models using remote sensing data (Nearing et al., 2010, Harrison et al., 2012). The project Co-I recently used $\text{DREAM}_{(ZS)}$ (ter Braak and Vrugt, 2008) to estimate parameter distributions in a slope-stability landslide model with remotely sensed data.

Implementing both DREAM and EM-DA requires writing appropriate data readers. The specific forcing, remote sensing, and in situ data that we will use are detailed in the *Project Workflow – Application Experiments s*ubsection below. Many of the necessary data readers for this project will be adapted from previous efforts.

The infrastructure development portion of this project includes (1) Implementing SUMMA in DREAM, and (2) Implementing SUMMA in EM-DA.

As detailed in the *Plan of Work* subsection below, infrastructure development will require approximately 1.5 years of effort. This effort will result in an EM-LDAS capable of process-level sensitivity analyses, probabilistic parameter estimation, and EM-DA model structural learning.

We cannot propose to integrate our code into the NASA Land Information System directly as part of this project, because that would require extensive professional-level software development that we cannot afford on the project budget. Our purpose here is to offer fundamentally new DA capabilities that, once appropriately developed, may be added to integrated modeling systems like LIS. That being said, our proposal team has extensive experience with LIS development and integration, and all software produced in this project will be fundamentally aimed at future LIS integration.

Project Workflow – Application Experiments

The second half of this 3-year project will focus on using our EM-LDAS system in a series of in situ and remote sensing experiments to investigate the realism and quality of data products generated using the current generation of earth-orbiting satellites and terrestrial hydrology process models. Observation data will come from both in situ and remote sensing sources. For model diagnostics, it is essential to balance the spatial heterogeneity-based 'breadth' of remotely sensed observations and the relatively high accuracy/precision-based 'depth' of collocated observations at in situ sites (Gupta et al., 2013).

in Situ Experiments

For in situ data, we will use FluxNet. FluxNet measures precipitation, net radiation, air temperature, pressure, humidity, net ecosystem exchange, latent and sensible heat fluxes, and, in some cases, soil moisture – among a few other variables. FluxNet will be the primary source of "deep" data – observations that are relatively comprehensive, relatively dense in time (half-hourly), and relatively accurate. In total, several hundred years' worth of FluxNet data from many different eco-climates globally are publically available from the FluxNet 2015 repository (*http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/*).

Point-based FluxNet experiments are crucial for testing the EM-LDAS system, and also for conducing point-based model diagnostic experiments. These experiments will take approximately 1 year of total effort. We will perform these FluxNet EM-DA experiments in two different ways. First, we will use EM-DA to assimilate FluxNet site observations at each site of a number of diverse sites for a portion of the data record available at each site, and then assess the results using the complementary portions of the data record at the same site. In this case, the estimated $\mathcal{g}_m$ and $\mathcal{g}_h$ correction functions will vary between sites. Second, we will assimilate data from many FluxNet sites to learn <u>single</u> overall $\mathcal{g}_m$ correction function. We will test the added model skill from these lumped experiments using a leave-one-out approach on the FluxNet sites. Finally, we will compare both the response surfaces of the site-specific $\mathcal{g}_m$ functions vs. the site-general $\mathcal{g}_m$ functions, as well as the transfer entropy difference metrics derived from the site-specific and lumped experiments. **These comparisons will allow us to understand how much of process-level model error (relative to FluxNet data) is due to general site-independent errors in model structures across multiple eco-climates, vs. to the (in)ability of the model to perfectly represent site-specific process differences.**

Remote Sensing Experiments

Remote sensing experiments will focus on the CONUS domain – this is where we have the highest quality observation data for forcing and evaluation (which is important for testing an essentially new approach). Specifically, we will use parameters and forcing data from NLDAS. The Co-PI co-leads the NLDAS team that was recently funded to develop its own assimilation capabilities for soil moisture and evapotranspiration retrievals, and to implement new forcing uncertainty ensembles. Our proposed project will work closely with the NLDAS team, and will leverage both new forcing products, and also DA-related experience gained from those ongoing NLDAS efforts.

Remote sensing assimilation data will include level-3 soil moisture retrievals from SMAP, temperature-driven evapotranspiration retrievals from ALEXI (Anderson et al., 2011), fluorescence retrievals from GOME-2 (Joiner et al., 2014), and LAI retrievals from MODIS. These products will allow us to look at regional and continental-scale interactions between vegetation and water – and, specifically, to identify spatial and eco-climatic trends in model couplings between hydrology and vegetation stores and fluxes. By using distributed remote sensing data in conjunction with different SUMMA configurations, we will identify individual biogeophysical process representations that contribute significantly to disconnects and compensating error structures in these coupling relationships at different spatial and temporal scales. **This latter capability, facilitated by the marriage of multi-parameterization capable land models (in SUMMA) and EM-DA (developed here), will allow us to make concrete statements about priorities for future model development efforts.**

EM-DA is not inherently scale-dependent. It can be applied at any spatial scale, and can handle spatial autocorrelation in the models and retrievals. The transfer entropy metrics used in our DPN diagnostic analysis are scale dependent, both in the specification of the time lag, and also in the way samples are drawn for estimating the probability distributions and setting limits of integration in equation [5]. This means that we can develop DPNs at any spatiotemporal scale, and can therefore look at the ability of EM-DA to affect patterns in the internal connectivity between variables at different scales. We will apply DPN analysis to telescoping spatial and temporal scales, which will allow us to look at scale dependencies in the process-level information extracted by EM-DA from the different remote sensing data sources.

Techniques to assimilate most of these remote sensing variables – except fluorescence – are fairly mature. The PI of this project currently participates in a different NASA-funded project to develop traditional DA capabilities for fluorescence data from GOME-2, and this project will draw on that developing experience. Our project team has been involved in the community-wide efforts to develop and mature many of these existing DA techniques (Kumar et al., 2014, Nearing et al., 2012, Nearing et al., 2013a, Crow and Reichle, 2008, Crow and Van Loon, 2006). It is certainly the case that the community will continue to refine and improve techniques for assimilating soil moisture, evapotranspiration, leaf area observations, and many other types of remote sensing retrievals into land models, *and what we propose here is <u>not</u> a substitute for this kind of continued work on state-updating data assimilation*. Instead, what we propose here is a new type of DA that is designed for a different purpose – model structural learning and process-level model diagnostics. We will draw on the extensive existing experience by the community to guide our assimilation efforts (e.g., to choose appropriate observation operators, appropriate scaling functions, etc.).

Perceived Impact

Relevance to State of Knowledge in the Field

The development of model learning and model diagnostic capabilities using data fusion approaches like data assimilation is a longstanding, high-profile, and unsolved problem in the land modelling and terrestrial hydrology modelling communities. The need for the type of capability that we propose here has been explicitly recognized by several different science steering groups in the last decade. One example of this recognition is by the Global Land-Atmosphere Systems Study (GLASS). GLASS is the result of a thirty-year effort headed by the World Meteorological Organisation and the World Climate Research Program focused on developing and testing land models. In a decadal review of GLASS, van den Hurk et al. (2011) recognized a need for fundamentally new approaches to understanding model deficiencies, and the ability of land models to predict under nonstationarity.

Similarly, the US National Science Foundation recognized as one of their 'Grand Challenges for Environmental Modeling' (Challenge #7: System Identification) that *"radically novel procedures and algorithms are needed to rectify the chronic, historical deficit ... in engaging complex models systematically and successfully with field data for the purposes of learning and discovery."* They also recognized (Challenge #8: Predictive Science and Uncertainty) the need for "structural error/uncertainty and structural change in these models to be identified, quantified, rectified, and accounted for" (Beck et al., 2009).

This project represents a step-change in our ability to use data assimilation to address these types of questions and model-based learning objectives, and is the result of several years of dedicated effort to address these community-wide calls to action in a systematic and generalized way that can be extrapolated to many problems in land modelling.

Relevance to NASA Programs and Objectives

This proposal is directly relevant to THP program element 2.3 Multivariate Hydrological Simulation. In addition, this project directly extends several ongoing NASA efforts. This project is a direct continuation of a non-competitive exploratory project funded by the NASA THP program. It also directly extends work completed as part of a 2014 NASA ESTO Quick Response Submission project.

This project directly contributes to NASA objectives related maximizing value of space-based remote sensing products for earth systems modeling, including in terrestrial hydrology modeling. This project leverages work completed by a recent NASA ESTO funded project to develop terrestrial hydrology modeling for uncertainty quantification with SUMMA – specifically it adds a data assimilation component to that effort so that process diagnostics in SUMMA can directly leverage remote sensing data. SUMMA is now a part of the NASA Land Information System, and this proposal develops basic capabilities that are directly applicable to the DA objectives of LIS (Kumar et al., 2008) and the uncertainty-quantification and model-learning objectives of SUMMA (Clark et al., 2015b). Additionally, this project leverages and recent work related to uncertainty quantification in NLDAS (Nearing et al., 2016a); NLDAS is a major land modelling initiative shared jointly by NASA and NOAA.

The methodological and technological capabilities proposed here represent a fundamentally new way of deriving value from terrestrial remote sensing products. This project has the potential to increase the science value of several hydrology-related remote sensing data products.

Roles and Responsibilities

Work effort is divided between the University of Alabama (UA) and NASA GSFC. The first year and a half of the project will involve infrastructure development, and the second half of the project will involve experimental research.

The UA team is project PI (Dr. Nearing) and one dedicated graduate student (0.5 FTE). Dr. Nearing will be responsible for overall project administration and coordination between UA and NASA. The NASA team is project Co-PI (Dr. Peters-Lidard) and one project engineer (Dr.

Yatheendradas) at 0.30 FTE. Dr. Peters-Lidard will be responsible for administrating the NASA-side project development, including ensuring that the code and experiments from this project draw on and/or help to advance parallel NASA projects – specifically LIS, SUMMA, and NLDAS.

The UA team is responsible for implementing SUMMA into EM-DA, and will lead the FluxNet EM-DA experiments. PI Dr. Nearing has previous experience developing two hydrological EM-DA systems – as outlined in the *Proposed Approach – Data Assimilation Examples* subsection above. Dr. Nearing is the lead EM-DA technical developer, and is responsible for FluxNet experimental design. He will mentor one graduate student to assist with both of these tasks.

Dr. Yatheendradas is responsible for implementing SUMMA into DREAM, and for leading the remote sensing EM-DA experiments. Dr. Yatheendradas previous experience coupling a NASA landslide model with $\mathrm{DREAM}_{(ZS)}$. Dr. Yatheendradas and Dr. Nearing previously developed NLDAS, SMAP, and MODIS data readers for Noah-MP, and Dr. Yatheendradas will be responsible for adapting these data readers for SUMMA. He will also be responsible for developing GOME-2 and ALEXI data readers. The required DPN process diagnostic code was developed as part of a previous project jointly by Drs. Nearing and Yatheendradas.

Plan of Work

The specific project tasks listed in the table blow are detailed in the *Project Workflow* subsections above.

Task	Lead	2018	2019	2020
Infrastructure Development				
Implement DREAM around SUMMA	NASA			
Initial testing with FluxNet data	NASA			
Implement EM-DA around SUMMA	UA			
Initial testing with FluxNet data	UA			
Implement FluxNet data readers and obs. operators	UA			
Implement remote sensing and NLDAS data readers	NASA			
Technical Documentation	NASA/UA			
Experiments				
FluxNet site-specific experiments	UA			
FluxNet multi-site leave-one-out experiments	UA			
FluxNet process diagnostic analysis	UA			
FluxNet manuscript prep.	UA			
Remote sensing experiments	NASA			
Remote sensing diagnostic and cross-scale analysis	UA			
Remote sensing manuscript prep.	NASA			

References and Citations

Anderson, M. C., Kustas, W. P., Norman, J. M., Hain, C. R., Mecikalski, J. R., Schultz, L., González-Dugo, M. P., Cammalleri, C., d'Urso, G., Pimstein, A. and Gao, F. (2011) 'Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery', *Hydrol. Earth Syst. Sci.,* 15(1), pp. 223-239.

Beck, M. B., Gupta, H., Rastetter, E., Shoemaker, C., Tarboton, D., Butler, R., Edelson, D., Graber, H., Gross, L. and Harmon, T. (2009) 'Grand challenges of the future for environmental modeling', *White Paper, National Science Foundation, Arlington, Virginia*.

Bulygina, N. and Gupta, H. (2009) 'Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation', *Water Resources Research,* 45(12), pp. W00B13, doi: 10.1029/2007WR006749.

Bulygina, N. and Gupta, H. (2010) 'How Bayesian data assimilation can be used to estimate the mathematical structure of a model', *Stochastic Environmental Research and Risk Assessment,* 24(6), pp. 925, doi: http://dx.doi.org/10.1007/s00477-010-0387-y.

Bulygina, N. and Gupta, H. (2011) 'Correcting the mathematical structure of a hydrological model via Bayesian data assimilation', *Water Resources Research,* 47(5), pp. W05514, doi: 10.1029/2010WR009614.

Clark, A. (1987) 'The kludge in the machine', *Mind & Language,* 2(4), pp. 277-300.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W. and Brekke, L. D. (2015a) *The Structure for Unifying Multiple Modeling Alternatives (SUMMA), version 1: Technical description*: NCAR Technical Note NCAR/TN-514+ STR, 54 pp., National Center for Atmospheric Research, Boulder, Colo., doi: 10.5065/D6WQ01TD.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W. and Brekke, L. D. (2015b) 'A unified approach for process‐based hydrologic modeling: 1. Modeling concept', *Water Resources Research,* 51(4), pp. 2498-2514.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W. and Gochis, D. J. (2015c) 'A unified approach for process‐based hydrologic modeling: 2. Model implementation and case studies', *Water Resources Research,* 51(4), pp. 2515-2542.

Crow, W. T. and Reichle, R. H. (2008) 'Comparison of adaptive filtering techniques for land surface data assimilation', *Water Resources Research,* 44(8), pp. W08423, doi:10.1029/2008wr006883.

Crow, W. T. and Van Loon, E. (2006) 'Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture', *Journal of Hydrometeorology,* 7(3), pp. 421-432, doi: http://dx.doi.org/10.1175/JHM499.1.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) 'Maximum Likelihood from Incomplete Data via the EM Algorithm', *Journal of the Royal Statistical Society. Series B (Methodological),* 39(1), pp. 1-38.

Feraud, T., Chapuis, R., Aufrere, R. and Checchin, P. (2011) 'Improving results of rational non-linear observation functions using a Kalman filter correction', *2011 International Conference on Information Fusion (FUSION)*.

Ghahramani, Z. and Roweis, S. T. (1999) 'Learning Nonlinear Dynamical Systems Using an EM Algorithm', *Advances in Neural Information Processing Systems,* 11, pp. 431-437.
Gupta, H. V., Perrin, C., Kumar, R., Blöschl, G., Clark, M., Montanari, A. and Andréassian, V. (2013) 'Large-sample hydrology: a need to balance depth with breadth', *Hydrology and Earth System Sciences Discussions,* 10(7), pp. 9147-9189.
Harrison, K. W., Kumar, S. V., Peters-Lidard, C. D. and Santanello, J. A. (2012) 'Quantifying the change in soil moisture modeling uncertainty from remote sensing observations using Bayesian inference techniques', *Water Resources Research,* 48(11), pp. W11514.
Joiner, J., Yoshida, Y., Guanter, L., Lindstrot, R., Voigt, M., Jung, M., Vasilkov, A., Middleton, E., Huemmrich, K. F. and Tucker, C. J. 'New Measurements of Chlorophyll Fluorescence with Gome-2 and Comparisons with the Seasonal Cycle of GPP from Flux Towers'. *Proceedings of the 5th INT Ernational Workshop on Remote Sensing of Vegetation Fluoresce NCE, Paris, France*, 22-24.
Kirchner, J. W. (2006) 'Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology', *Water Resources Research,* 42(3).
Kumar, S. V., Peters-Lidard, C. D., Mocko, D., Reichle, R., Liu, Y., Arsenault, K. R., Xia, Y., Ek, M., Riggs, G. and Livneh, B. (2014) 'Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation', *Journal of Hydrometeorology,* 15(6), pp. 2446-2469.
Kumar, S. V., Reichle, R. H., Harrison, K. W., Peters‐Lidard, C. D., Yatheendradas, S. and Santanello, J. A. (2012) 'A comparison of methods for a priori bias correction in soil moisture data assimilation', *Water Resources Research,* 48(3).
Kumar, S. V., Reichle, R. H., Peters-Lidard, C. D., Koster, R. D., Zhan, X., Crow, W. T., Eylander, J. B. and Houser, P. R. (2008) 'A land surface data assimilation framework using the land information system: Description and applications', *Advances in Water Resources,* 31(11), pp. 1419-1432.
Liu, Y. Q. and Gupta, H. V. (2007) 'Uncertainty in hydrologic modeling: toward an integrated data assimilation framework', *Water Resources Research,* 43(7), pp. W07401, doi: 10.1029/2006WR005756.
Nearing, G. S., Crow, W. T., Thorp, K. R., Moran, M. S., Reichle, R. H. and Gupta, H. V. (2012) 'Assimilating remote sensing observations of leaf area index and soil moisture for wheat yield estimates: an observing system simulation experiment', *Water Resources Research,* 48(5), pp. W05525, doi: http://dx.doi.org/10.1029/2011WR011420.
Nearing, G. S. and Gupta, H. V. (2015) 'The quantity and quality of information in hydrologic models', *Water Resources Research,* 51(1), pp. 524-538.
Nearing, G. S., Gupta, H. V. and Crow, W. T. (2013a) 'Information loss in approximately bayesian estimation techniques: a comparison of generative and discriminative approaches to estimating agricultural productivity', *Journal of Hydrology,* 507(0), pp. 163-173.
Nearing, G. S., Gupta, H. V., Crow, W. T. and Gong, W. (2013b) 'An approach to quantifying the efficiency of a Bayesian filter', *Water Resources Research,* 49(4), pp. 2164-2173.
Nearing, G. S., Mocko, D. M., Peters-Lidard, C. D., Kumar, S. V. and Xia, Y. (2016a) 'Benchmarking NLDAS-2 soil moisture and evapotranspiration to separate uncertainty contributions', *Journal of Hydrometeorology*, pp. 745-759.

Nearing, G. S., Moran, M. S., Thorp, K. R., Collins, C. D. H. and Slack, D. C. (2010) 'Likelihood parameter estimation for calibrating a soil moisture model using radar bakscatter', *Remote Sensing of Environment,* 114(11), pp. 2564-2574.

Nearing, G. S., Ruddell, B., Clark, M. P. and Nijssen, B. 'The Role of Data Assimilation in Model Diagnostics (invited)'. *American Geophysical Union Fall Meeting*, San Francisco, CA USA.

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D. and Rosero, E. (2011) 'The community Noah land surface model with multiparameterization options (Noah‐MP): 1. Model description and evaluation with local‐scale measurements', *Journal of Geophysical Research: Atmospheres (1984–2012),* 116(D12).

Reichle, R. H., Crow, W. T. and Keppenne, C. L. (2008) 'An adaptive ensemble Kalman filter for soil moisture data assimilation', *Water Resources Research,* 44(3), pp. W03423, doi: 10.1029/2007WR006357.

Ruddell, B. L. and Kumar, P. (2009) 'Ecohydrologic process networks: 1. Identification', *Water Resources Research,* 45(3).

Schön, T. B., Wills, A. and Ninness, B. (2011) 'System identification of nonlinear state-space models', *Automatica,* 47(1), pp. 39-49.

Schreiber, T. (2000) 'Measuring information transfer', *Physical review letters,* 85(2), pp. 461.

Snyder, C., Bengtsson, T., Bickel, P. and Anderson, J. (2008) 'Obstacles to high-dimensional particle filtering', *Monthly Weather Review,* 136(12), pp. 4629-4640, doi: 10.1175/2008mwr2529.1.

ter Braak, C. J. and Vrugt, J. A. (2008) 'Differential evolution Markov chain with snooker updater and fewer chains', *Statistics and Computing,* 18(4), pp. 435-446.

Turner, R., M., D. and C., R. 'System identification in Gaussian process dynamical systems', *Nonparametric Bayes Workshop at NIPS*, Whistler, Canada.

van den Hurk, B., Best, M., Dirmeyer, P., Pitman, A., Polcher, J. and Santanello, J. (2011) 'Acceleration of Land Surface Model Development over a Decade of GLASS', *Bulletin of the American Meteorological Society,* 92(12), pp. 1593-1600.

van Leeuwen, P. J. (2010) 'Nonlinear data assimilation in geosciences: an extremely efficient particle filter', *Quarterly Journal of the Royal Meteorological Society,* 136(653), pp. 1991-1999; doi: http://dx.doi.org/10.1002/qj.699.

Vrugt, J. A. (2016) 'Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation', *Environmental Modelling & Software,* 75, pp. 273-316.

Vrugt, J. A., Gupta, H. V. and Nuallain, B. O. (2006) 'Real-time data assimilation for operational ensemble streamflow forecasting', *Journal of Hydrometeorology,* 7(3), pp. 548-565, doi: http://dx.doi.org/10.1175/jhm504.1.

Wang, J. M., Fleet, D. J. and Hertzmann, A. (2008) 'Gaussian process dynamical models for human motion', *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 30(2), pp. 283-298.

Wikle, C. K. and Berliner, L. M. (2007) 'A Bayesian tutorial for data assimilation', *Physica D-Nonlinear Phenomena,* 230(1-2), pp. 1-16, doi:http://dx.doi.org/10.1016/j.physd.2006.09.017.

Wilkinson, R. D., Vrettas, M., Cornford, D. and Oakley, J. E. (2011) 'Quantifying simulator discrepancy in discrete-time dynamical simulators', *Journal of agricultural, biological, and environmental statistics,* 16(4), pp. 554-570.

Xia, Y., Ek, M., Wood, E., Sheffield, J., Luo, L. and Mo, K. 'North American Land Data Assimilation System (NLDAS) in support of theU.S. drought and flood analysis, monitoring, and prediction, including National Integrated Drought Information System (NIDIS)'. *International Union of Geodesy and Geophysics*, Melbourne, Australia.

Yang, Z. L., Niu, G. Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Longuevergne, L., Manning, K., Niyogi, D. and Tewari, M. (2011) 'The community Noah land surface model with multiparameterization options (Noah‐MP): 2. Evaluation over global river basins', *Journal of Geophysical Research: Atmospheres,* 116(D12).

Zupanski, M. (2005) 'Maximum likelihood ensemble filter: theoretical aspects', *Monthly Weather Review,* 133(6), pp. 1710-1726, doi: 10.1175/mwr2946.1.

Biographical Sketches

Dr. Grey S. Nearing (PI)
Assistant Professor
University of Alabama Department of Geological Sciences; Tuscaloosa, AL
gnearing@ua.edu

Education:

Ph.D.	Hydrology	University of Arizona, 2013
M.S.	Biosystems Engineering	University of Arizona, 2009
B.S.	Mathematics	Purdue University, 2006

Professional Experience:

Dr. Grey Nearing joined the Department of Geological Sciences at the University of Alabama at Tuscaloosa in August, 2017. Prior to that he was a Research Hydrologist in the Hydrological Sciences Laboratory at NASA Goddard Space Flight Center (50% time), and also a Support Scientist at the National Center for Atmospheric Research (50% time). Dr. Nearing's primary research focus is related to hydrological and agricultural data assimilation and land surface modeling. Since earning his PhD, Dr. Nearing has worked on evaluation and diagnostics for complex systems models using machine learning and system identification. He has worked with both modeling (Land Information System) and remote sensing (SMAP) groups at NASA, and with modeling (SUMMA) groups at NCAR. A brief list of projects follows:

- Project (Co-I): *Computational Technologies: Feasibility Studies of Quantum Enabled Annealing Algorithms for Estimating Terrestrial Carbon Fluxes from OCO-2 and the LIS Model* (ongoing)
  - Developed algorithms for data assimilation on quantum annealing computers.
- Project: *Development of Computational Infrastructure to Support High-Resolution Large-Ensemble Hydrology Simulations from Local-to-Continental Scales*
  - Developed multi-physics hydrology modeling platform for uncertainty analysis.
- Project: *SMAP Phase-E Early Adopter Program – Measuring the Value of SMAP Observations* (ongoing)
  - Developed metrics to quantify value of SMAP data for a variety of research and societal applications.
- Project (Science PI): *Next-Generation Data Assimilation Capabilities in a Mission Simulation Platform to Increase the Value of Terrestrial Remote Sensing Observations* (completed)
  - Developed a DA method to correct systematic biases in a soil column model.
- Project: *Data Assimilation in the Land Information System* (completed)
- Project: *Soil Moisture Data Assimilation in Crop Productivity Models*
  - Investigated the value of optical and microwave data assimilation for crop yield monitoring and forecasting

Awards and Grants:

- 2015: (PI) NASA Energy and Water Cycle Program; Exploratory Award
- 2015: NASA Hydrospheric and Biospheric Sciences Branch Award for Scientific and Technical Support
- 2014: NSF Geosciences; EAR Postdoctoral Fellowship
- 2014: (Co-I) NASA Earth Science Technology Office; Advanced Information Systems Technology
- 2014: (Science-PI) NASA Earth Science Technology Office; Quick Response Submission
- 2012: NSF East Asia and Pacific Summer Institute for US Graduate Students; Fellowship
- 2011 & 2012: Achievement Rewards for College Scientists (*two separate competitive awards*)

Selected Publications:

- G.S. Nearing, Y. Tian, H.V. Gupta, M.P. Clark, S.V. Weijs, K.W. Harrison; A philosophical basis for hydrologic uncertainty. *Hydrological Sciences Journal*, June, 2016.
- G.S. Nearing, D.M. Mocko, C.D. Peters-Lidard, S.V. Kumar, Y. Xia; Benchmarking NLDAS-2 soil moisture and evapotranspiration to separate uncertainty contributions. *Journal of Hydrometeorology*, January, 2016.
- G.S. Nearing, H.V. Gupta; The quantity and quality of information in hydrologic models. *Water Resources Research,* January, 2015.
- G.S. Nearing, H.V. Gupta, W.T. Crow; Information loss in approximately Bayesian estimation techniques: A comparison of generative and discriminative approaches to estimating agricultural productivity. *Journal of Hydrology*, December 2013.
- G.S. Nearing, H.V. Gupta, W.T. Crow, W. Gong; An approach to quantifying the efficiency of a Bayesian filter. *Water Resources Research,* April 2013.
- G.S. Nearing, W.T. Crow, K.R. Thorp, M.S. Moran, R.R. Reichle, H.V. Gupta; Assimilating remote sensing observations of leaf area index and soil moisture for wheat yield estimates: an observing system simulation experiment. *Water Resources Research*, May 2012.
- G.S. Nearing, M.S. Moran, R.L. Scott, G.E. Ponce-Campos; Coupling diffusion and maximum entropy models to estimate thermal inertia and soil moisture. *Remote Sensing of Environment*, January 2012.

Dr. Christa D. Peters-Lidard (Co-PI)
Deputy Director for Hydrosphere, Biosphere, and Geophysics, Earth Sciences Division
NASA/Goddard Space Flight Center, Greenbelt, MD 20771
Tel: 301/614-5811, *Fax:* 301/614-5808, *Email:* Christa.Peters@nasa.gov

Education:

Ph.D., Civil Engineering & Operations Research (Water Resources) 1997, Princeton University
M.A., Civil Engineering & Operations Research (Water Resources) 1993, Princeton University
B.S., Geophysics, Summa Cum Laude, 1991, Virginia Tech

Experience:

- ***Deputy Director for Hydrosphere, Biosphere and Geophysics, Earth Sciences Division,*** January 2015 – Present. Responsible for overall management of five laboratories with a budget of approximately $50M, including 50+ civil servant scientists, including organizational management, personnel, facilities and resources. Responsible for planning, leading, directing, implementing and assessing programs and activities dedicated to exploring and understanding the Earth's hydrosphere, biosphere, and solid earth.
- ***Physical Scientist, Hydrological Sciences Laboratory,*** May 2012 – January 2015. Serve as PI on research and applications projects including PMM science team, transitioning the award-winning Land Information System (LIS; http://lis.gsfc.nasa.gov) for research and applications. Lead NASA AIST- and MAP-funded projects including the NASA Unified WRF (NU-WRF).
- ***Lab Chief/Branch Head (Civil Servant).*** NASA/GSFC Hydrological Sciences Laboratory (formerly Branch), August 2005-May-2012. Supervise 10 scientists and over 30 contractors and research scientists. Responsible for $3M+ annual budget and coordination of NASA- and externally funded program of earth and space science research and applications.
- ***Assistant Professor.*** School of Civil and Environmental Engineering, Georgia Institute of Technology, 1997-2001. Taught undergraduate and graduate courses in Hydrology. Conducted independent research program on measurement and modeling of land-atmosphere interactions, including application of HPCC technologies.

Selected Professional Service:

- Journal of Hydrometeorology, Chief Editor, 2011-present; Editor, 2004-2007.
- Journal of Hydrology, Associate Editor, 2007-2011.
- Water Resources Research, Associate Editor, 2002-2003.
- American Meteorological Society: Councilor, 2012-2015; Executive Committee, 2013-2015; Member, Annual Meeting Oversight Committee, 2009-present; Committee on Hydrology, Chair 2002-2005, Member, 1998-2005.
- American Geophysical Union: Chair, Hydrology Early Career Award Committee, 2011-present, Member, Committee on Remote Sensing, 1997-present.

Selected Honors/Awards:

- Fellow, American Meteorological Society, 2012
- Arthur S. Flemming Award for Excellence in Federal Service, 2007
- NASA Software of the Year Award for the Land Information System, 2005
- Committee on Space Research (COSPAR) Commission A Zeldovich Medal, 2004
- Phi Beta Kappa, 1991

Selected Publications (Reverse Chronological):

- Nearing, G. S., D. M. Mocko, C. D. Peters-Lidard, S. V. Kumar, and Y. Xia. 2016. "Benchmarking NLDAS-2 soil moisture and evapotranspiration to separate uncertainty contributions." J. Hydrometeor. 17 (3): 745-759 [10.1175/jhm-d-15-0063.1]
- Tian, Y., G. S. Nearing, C. D. Peters-Lidard, K. W. Harrison, and L. Tang. 2016. "Performance Metrics, Error Modeling, and Uncertainty Quantification." Monthly Weather Review 144 (2): 607-613 [10.1175/mwr-d-15-0087.1
- Peters-Lidard, C.D., E. M. Kemp, T. Matsui, J. A. Santanello, Jr., S. V., Kumar, J. P. Jacob, T. Clune, W.-K. Tao, M. Chin, A. Hou, J. L. Case,, D. Kim, K.-M. Kim, W. Lau, Y. Liu, J.-J. Shi, D. Starr, Q. Tan,, Z. Tao, B. F. Zaitchik, B. Zavodsky, S. Q. Zhang, M. Zupanski, 2015: Integrated modeling of aerosol, cloud, precipitation and land processes at satellite-resolved scales, *Environmental Modelling & Software*, **67**, 149-159. doi:10.1016/j.envsoft.2015.01.007.
- Kumar, S. V., C. D. Peters-Lidard, K. R. Arsenault, A. Getirana, D. M. Mocko, and Y. Liu. 2015. Quantifying the Added Value of Snow Cover Area Observations in Passive Microwave Snow Depth Data Assimilation. *J. Hydrometeor*, 16(4,) 1736-1741 [10.1175/JHM-D-15-0021.1]
- Kumar, Sujay V., Christa D. Peters-Lidard, David Mocko, Rolf Reichle, Yuqiong Liu, Kristi R. Arsenault, Youlong Xia, Michael Ek, George Riggs, Ben Livneh, and Michael Cosh, 2014: Assimilation of Remotely Sensed Soil Moisture and Snow Depth Retrievals for Drought Estimation. *J. Hydrometeor*, **15**, 2446–2469. doi: 10.1175/JHM-D-13-0132.1.
- Xia, Y., M. B. Ek, C. D. Peters-Lidard, D. M. Mocko, M. Svoboda, J. Sheffield, and E. F. Wood. 2014. Application of USDM statistics in NLDAS-2: Optimal blended NLDAS drought index over the continental United States. J. Geophys. Res. Atmos. 119 (6): 2947–2965 [10.1002/2013JD020994]
- Kumar, S. V., C. D. Peters-Lidard, D. M. Mocko, R. H. Reichle, Y. Liu, K. R. Arsenault, Y. Xia, M. Ek, G. Riggs, B. Livneh, and M. Cosh. 2014. Assimilation of passive microwave-based soil moisture and snow depth retrievals for drought estimation. *J. Hydrometeor* 15 (6): 2446-2469 [10.1175/JHM-D-13-0132.1]
- Peters-Lidard, C. D., Kumar, S. V., Mocko, D. M. and Tian, Y., 2011: Estimating evapotranspiration with land data assimilation systems. *Hydrological Processes*, 25: 3979–3992. doi: 10.1002/hyp.8387.
- Peters-Lidard, C. D., P. R. Houser, Y. Tian, S. V. Kumar, et al., 2007: High Performance Earth System Modeling with NASA/GSFC's Land Information System. *Innovations in Systems and Software Engineering*, 3(3), 157-165, doi:10.1007/s11334-007-0028-x.

Dr. Soni Yatheendradas (Co-I)
Assistant Research Engineer
NASA Goddard Space Flight Center; Greenbelt, MD
University of Maryland Earth Systems Science Interdisciplinary Center; College Park, MD
soni.yatheendradas-1@nasa.gov

Education:

Ph.D.	Hydrology	University of Arizona, 2007
M.S.	Hydrology	University of Arizona, 2003
B.Tech	Civil Engineering	India Institute of Technology Bombay, 2000

Research and Professional Experience:

- **Assistant Research Engineer**, UMD & NASA/GSFC Hydrological Sciences Lab (December 2012 - present): Landslide hazard and exposure indicators for climate assessment. Scalable Geophysical Big Data Analysis. Diagnosing the efficiency of Data Assimilation using information theory. Sensitivity and interaction of landslide binary outputs to parameters and rain source.
- **Research Associate**, UMD and NASA/GSFC Hydrological Sciences Lab (September 2008 - December 2012): FEWS NET early warning of agricultural drought; Satellite snow area assimilation for streamflow improvement towards NWS Decision Support
- **Post-Doctoral Research Associate**, EES Dept., New Mexico Institute of Mining and Technology (May 2007 - August 2008)

Professional Services and Awards

- Co-Investigator on NASA ROSES-2014 NCA and ROSES-2009 Terra/Aqua proposals.
- 2015 NASA Hydrospheric and Biospheric Sciences Award for Scientific/Technical Support.
- Salt River Research Fellowship (University of Arizona, 2004-2005 & 2003-2004)
- The World Laboratory Fellowship (University of Arizona, 2000-2001).

Selected Publications:

- Nearing, G. S., S. Yatheendradas, W. T. Crow, et al (2017), Nonparametric triple collocation, *Water Resour. Res.*, 53, doi:10.1002/2017WR020359.
- Yatheendradas, S., et al. 2014. Comment on ‘Shang S. 2012. Calculating actual crop evapotranspiration under soil water stress conditions with appropriate numerical methods and time step. *Hydrol. Process.* 26: 3338-3343. DOI: 10.1002/hyp.8405’. *Hydrol. Process.*, 28 (11): 3833-3840 [10.1002/hyp.10138]
- Kumar, S., et al. 2012. "A comparison of methods for a priori bias correction in soil moisture data assimilation." *Water Resour. Res.*, 48 (3): W03515
- Rosero, E., et al. (2010), Quantifying parameter sensitivity, interaction and transferability in hydrologically enhanced versions of the Noah-LSM over transition zones during the warm season, *J. Geophys. Res.*, *115*, D03106.
- Yatheendradas, S., et al. 2012. "Distributed assimilation of satellite-based snow extent for improving simulated streamflow in mountainous, dense forests: An example over the DMIP2 western basins." *Water Resour. Res.*, 48 (9): W09557

Dr. Wade T. Crow (Collaborator)
Research Physical Scientist
USDA ARS Hydrology and Remote Sensing Lab; Beltsville, MD
wade. crow@ars.usda.gov

Education:

2001 Princeton University, Civil and Environmental Engineering, Ph.D.
1998 Princeton University, Civil and Environmental Engineering, M.S.E.
1995 Carleton College, Physics, B.A.

Positions:

2001 to 2002: Post-doctoral research associate at Princeton University.
2002 to present: Research Physical Scientist, USDA ARS, Beltsville, MD.

Research and Professional Activities:

Dr. Crow received his Ph.D. in 2001 from Princeton University and is currently a Research Physical Scientist and Project Lead Scientist at the USDA ARS Hydrology and Remote Sensing Laboratory in Beltsville, MD. His research focuses on the development of hydrologic and agricultural applications for remote sensing data and the implementation of appropriate data assimilation approaches to facilitate this goal - with a special emphasis on techniques that fuse information from various disparate remote sensing sources. He has served (or currently serves) on the science teams for the NASA Precipitation Measurement Mission program, the NASA Hydros mission, the NASA SMAP mission, and the NASA AirMOSS mission. He recently completed a 3-year term as editor of the American Meteorological Society's *Journal of Hydrometeorology.*

Selected Peer-Reviewed Publications:

- **Crow, W.T.**, E. Han, D. Ryu, C.R. Hain, and M.C. Anderson, M.C. and Ryu, "D. Estimating annual water storage variations in medium-scale (10^3 to 10^4 km^2) basins using microwave-based soil moisture retrievals," *Hydrologic and Earth System Sciences*. 2017.
- **Crow, W.T.**, F. Chen, R.H. Reichle, and Q. Liu, "L band microwave remote sensing and land data assimilation improve the representation of prestorm soil moisture conditions for hydrologic forecasting," *Geophysical Research Letters,* 44. 2017.
- Gruber, A., C.-H. Su, **W.T. Crow**, S. Zwieback, W. A. Dorigo and W. Wagner, "Estimating error cross-correlations in soil moisture data sets using extended collocation analysis," *Journal of Geophysical Research*, 121. 2016.
- Pan, M., C.K. Fisher, N.W. Chaney, W. Zhan, **W.T. Crow**, F. Aires, D. Entekhabi and E.F. Wood, "Triple collocation: Beyond three estimates and separation of structural/non-structural errors," *Remote Sensing of Environment*, 171. 2015.
- **Crow, W.T.**, F. Lei, C. Hain, M.C. Anderson, R.L. Scott, D. Billesbach and T. Arkebauer, "Robust estimates of soil moisture and latent heat flux coupling strength obtained from triple collocation," *Geophysical Research Letters*, 42. 2015.

Table of Personnel and Work Effort

The following table reflects the level of support required of all personnel necessary to perform the proposed investigation, regardless of whether these individuals require funding from this proposal. The proposed work level is appropriate to perform the investigation on the basis of previous investigation experience.

Name and/or Position Title	Role	Institution	PY 1 FTEs	PY 2 FTEs	PY 3 FTEs	Total
NASA-Funded Work Effort						
Grey Nearing	PI	Geological Sciences Department, University of Alabama	0.08	0.08	0.08	0.24
Christa Peters-Lidard	Co-I	Hydrospheric and Biospheric Sciences, NASA Goddard Space Flight Center	0.08	0.08	0.08	0.24
Soni Yatheendradas	Co-I	University of Maryland, Earth Systems Science Interdisciplinary Center & Hydrospheric and Biospheric Sciences, NASA Goddard Space Flight Center	0.30	0.30	0.30	0.90
Graduate Student[a]	Support	Geological Sciences Department; University of Alabama	0.50	0.50	0.50	1.5
Subtotal:			0.96	0.96	0.96	2.88
Work Effort for Which No Funding is Requested						
Grey Nearing[b]	PI	Geological Sciences Department, University of Alabama	0.13	0.12	0.12	0.375
Wade Crow	Collaborator	Hydrology and Remote Sensing Lab, USDA-ARS	de minimis	de minimis	de minimis	de minimis
Subtotal:			0.13	0.12	0.12	0.37
Total:			1.09	1.08	1.08	3.25

[a]In addition to 1 month (0.08 FTE) of NASA-funded effort, PI Dr. Nearing will dedicate approximately one-third of his University-funded research appointment (a 50% appointment over 9 months) for project administration, student advising, and technical work related directly to this project. Dr. Nearing's 1 month of NASA-funded summer salary will be dedicated solely to technical work, including code development and experimental analysis.

Current and Pending Support

Dr. Grey S. Nearing (PI)

Current Support: None.

Pending Support: None.

Dr. Christa Peters-Lidard (Co-PI)

Current Support:

Project Title: Predicting Middle Eastern and African Seasonal Water Deficits using NASA Data and Models
Principal Investigator: Christa D. Peters-Lidard
Supporting Agency: NASA
Program: Earth Science Applications: Water Resources
Point of Contact: Bradley D. Doorn, (202) 358-2187, Bradley.Doorn@nasa.gov
Commitment (person-months): 2.4 per year
Performance Period: October 2014 – September 2018

Project Title: An integrated hydrological modeling system for high-resolution coastal applications
Principal Investigator: Teddy Holt (NRL)
Supporting Agency: Office of Naval Research
Program: NOPP Funding Announcement - Integrating the Hydrological Cycle for Improved Coastal and Global Forecasting
Point of Contact: Dan Eleuterio, Daniel.eleuterio@navy.mil
Commitment (person-months): 1.2 per year
Performance Period: May 2016 – September 2018

Project Title: Dynamic Emissivity Estimates to Support Physical Precipitation Retrievals for GPM
Principal Investigator: Christa Peters-Lidard
Supporting Agency: NASA
Program: Precipitation Measurement Missions Science Team (PMM)
Point of Contact: Dr. Ramesh Kakar, (202) 358-0240, ramesh.k.kakar@nasa.gov
Commitment (person-months): 1.6 per year
Performance Period: January 1, 2016 – December 31, 2018

Project Title: FY16-18 Land Information System (LIS) Mobility Support Development Tasks
Principal Investigator: Christa D. Peters-Lidard, NASA/GSFC
Supporting Agency: Air Force, Life Cycle Management Center
Program: 557th Weather Wing, 2nd Weather Group, 16th Weather Squadron
Point of Contact: Dan Rozema, daniel.rozema@us.af.mil
Commitment (person-months): 1.2 per year
Performance Period: August 2016 – July 2019

Award or Project Title: Climate risks in the water sector: Advancing the technical readiness of emerging technologies in climate downscaling and hydrologic modeling
Principal Investigator: Martyn Clark, UCAR
Government Agency: NASA
Program: Advanced Information Systems Technology (AIST)
Point of Contact: Michael Little, (757) 864-6837, m.m.little@nasa.gov
Commitment (person-months): 1.2 per year
Performance Period: September 1, 2017 – August 31, 2019

Project Title: Evaluating and Advancing the Representation of Lake-Atmosphere Interactions and Resulting Heavy Lake-Effect Snowstorms Across the Laurentian Great Lakes Basin Within the NASA-Unified Weather Research and Forecasting Model
Principal Investigator: Michael Notaro (University of Wisconsin-Madison)
Supporting Agency: NASA
Program: Modeling, Analysis, and Prediction (MAP)
Point of Contact: David Considine, (202) 358-2277, david.b.considine@nasa.gov
Commitment (person-months): 1.2 per year
Performance Period: June 2017 – May 2021

Pending Support:

Project Title: FEWS NET's NASA PAPA Operational Seasonal Water Availability Measures and Forecasts
Principal Investigator: C. Peters-Lidard
Supporting Agency: USAID
Program: FEWS NET
Point of Contact: Romaine Williams, rowilliams@usaid.gov
Commitment (person-months): 1.2 per year
Performance Period: October 1, 2017 – September 30, 2022

Dr. Soni Yatheendradas (Co-I)

Current Support:

Project Title: Landslide Hazard and Exposure Indicators for the National Climate Assessment
Principal Investigator: Dalia B. Kirschbaum (NASA/GSFC)
Supporting Agency: NASA
Program: NASA ROSES 2014 A. 29 solicitation *Climate Indicators and Data Products for Future National Climate Assessments*
Point of Contact: Lucia Tsaoussi: Lucia.S.Tsaoussi@nasa.gov (202) 358-4471
Commitment (person-months): 4 months/year
Performance Period: March 1, 2016 - Feb 28, 2019

Project Title: Optimal Data Layout for Scalable Geophysical Analysis in a Data-intensive Environment
Principal Investigator: Hongfeng Yu (University of Nebraska-Lincoln)
Supporting Agency: NASA
Program: NSF EarthCube IA Collaborative Proposal
Point of Contact: Eva Zanzerkia: ezanzerk@nsf.gov , (703) 292-4734
Commitment (person-months): 6 months/year
Performance Period: Sep 1, 2015 to Aug 31, 2017

Pending Support: None.

Letters of Support

College of Arts and Sciences
Department of Geological Sciences

June 26, 2017

THE UNIVERSITY OF ALABAMA
ARTS & SCIENCES

Dr. Jared Entin
ROSES Terrestrial Hydrology Program
National Aeronautics and Space Administration

Dear Dr. Entin:

On behalf of the Department of Geological Sciences (DGS) at the University of Alabama, I write as Department Chair to strongly support Dr. Grey Nearing's application to your program for research funding. Dr. Nearing was recently hired as a tenure-track assistant professor in the DGS. As part of our program to encourage research productivity in recently-hired faculty, the Department and other sources on campus will provide Dr. Nearing at least $3000 a year in travel funding in his first two years of employment, and at least $1500 each year subsequently. Additionally, the Department assists all tenure track faculty in meeting publication costs as the situation arises. As such, Dr. Nearing did not include travel and publication costs in his proposed budget.

In summary, I support Dr. Nearing's application without reservation and can clearly affirm he has the full and unqualified support of the entire Department of Geological Sciences.

Sincerely,

C. Fred T. Andrus

C. Fred T. Andrus
Associate Professor and Chair
Department of Geological Sciences
University of Alabama
2003 Bevill Building
Tuscaloosa, AL 35487
Voice: 205-348-5177
FAX: 205-348-0818

202 Bevill Building
Box 870338
Tuscaloosa, AL 35487-0338
(205) 348-5095
FAX (205) 348-0818

Office of
Information Technology

August 1st, 2017

To Whom It May Concern:

This letter is to certify that The University of Alabama, Office of Information Technology, will provide access to the OIT managed high performance computing cluster (UA HPC) to Grey S. Nearing for the duration of his grant award.

Announcement ID: NNH17ZDA001N-THP: Terrestrial Hydrology

Proposal Title: Multivariate Hydrologic Data Assimilation for Model Structural Learning and Process-Diagnostics

Sincerely,

John P. McGowan, Ph.D.

Vice Provost & CIO
Office of Information Technology
The University of Alabama
A313 Gordon Palmer Annex
Box 870346
Tuscaloosa, AL 35487
Office 205-348-5610 | Mobile 210-287-7555
john.mcgowan@ua.edu

Associate Vice Chancellor for IT
The University of Alabama System
500 University Blvd.
Tuscaloosa AL 35401
Office 205-348-5205 | Mobile 210-287-7555
john.mcgowan@ua.edu

Budget Justification

Budget Justification – Narrative

University of Alabama Funding by Program Year

	PY1	PY2	PY3	Total
University of Alabama	10,780	11,319	11,885	33,984
NASA	5,204	5,416	8,131	18,750
Total	**15,984**	**16,735**	**20,016**	**52,735**

**Note that these numbers do not include salary and benefits, as per guidelines in Section IV(b)(iii) of the NASA ROSES NRA (2017).*

Personnel Roles and Cost Basis:

PI Dr. Grey Nearing will provide overall project management, including overall organization and coordination between UA and NASA for code development and experimental design/analysis. He will also provide graduate student mentorship and supervision, and will contribute directly to technical tasks related to both code development and performing and analyzing scientific experiments. In particular, he will lead the EM-DA code development. Dr. Nearing is funded for one month (0.08 FTE) under the proposed budget, but will contribute approximately 0.20 FTE overall during the three years on this project.

Dr. Nearing will supervise one graduate student who will participate in software development – especially on testing the software developed by Drs. Nearing and Yatheendradas. The graduate student will be involved in experimental design, execution, and analysis with the completed code base.

Other Direct Costs: There are no other direct costs for this proposal. Dr. Nearing's travel and publications are covered as a new hire by departmental funds for the duration of this project. This includes $3K in travel for the first two years of the project and $1.5K for the third year, as well as publication fees as necessary.

Other Applicable Costs – GSFC: The GSFC portion of the budget will be sent directly to GSFC from NASA HQ and will not be managed by UA. The basis of estimate and detailed budget for GSFC are provided in the Budget Details section below. Dr. Christa Peters-Lidard (Co-PI) is the deputy director of Hydrosphere, Biosphere and Geophysics in the Earth Sciences Division at NASA GSFC and has been the PI or Co-PI on four AIST projects, with extensive experience in administrating and supervising NASA-side project development. Dr. Soni Yatheendradas (Co-I) recently submitted a paper that measured soil moisture retrieval information content and quantified the major challenges related to extracting that information during DA. The proposed staff from GSFC are uniquely qualified to support the proposal work in these areas.

Facilities and Equipment

Sufficient computing resources were requested form NASA NCCS at GSFC. The eBooks request number is HEC-SMD-17-1285 – the NASA eBooks pre-request is attached to this proposal. All necessary software, including Fortran and c++ compilers, as well as Matlab and Python scripting packages, are available on the NCCS Discover machine. No other specialized equipment is required for this project. Additional computing resources for this project will be available at the University of Alabama UA-HPC center.

Budget Justification – Details

ORGANIZATION

The University of Alabama

PRINCIPAL INVESTIGATOR/PROJECT DIRECTOR

Nearing NASA 08032017

		Year 1 Research	Year 2 Research	Year 3 Research	IRG1 Total
A.	SENIOR PERSONNEL: PI/PD, Co-PIs, Faculty and Other Senior Associates				
	List each separately with title, (A.7. show number in brackets)				
1	Nearing (1 summer mo)	[redacted]	[redacted]	[redacted]	[redacted]
6	( ) OTHERS (LIST INDIVIDUALLY ON BUDGET EXPLANATION PAGE)	0	0	0	0
7	**( ) TOTAL SENIOR PERSONNEL (1-6)**	[redacted]	[redacted]	[redacted]	[redacted]
B.	OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)				
1	( ) ASSISTANT RESEARCH SCIENTISTS/Technicians	0	0	0	0
2	( ) POST DOC	0	0	0	0
3	( ) GRADUATE STUDENTS	[redacted]	[redacted]	[redacted]	[redacted]
4	( ) UNDERGRADUATE STUDENTS	0	0	0	0
5	( ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)	0	0	0	0
6	( ) OTHER	0	0	0	0
	TOTAL SALARIES AND WAGES (A+B)	[redacted]	[redacted]	[redacted]	[redacted]
C.	FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				
	SENIOR PERSONNEL (32%)	[redacted]	[redacted]	[redacted]	[redacted]
	ASST RESEARCH SCIENTISTS /TECHNICIANS (32%)	[redacted]	[redacted]	[redacted]	[redacted]
	POST DOC (17.38%)	[redacted]	[redacted]	[redacted]	[redacted]
	GRADUATE STUDENTS (SALARY/4 x 7.7%)	[redacted]	[redacted]	[redacted]	[redacted]
	GRAD STUDENTS HEALTH INS ($104/MO x WKG Mos x GRAs)	[redacted]	[redacted]	[redacted]	[redacted]
	TOTAL FRINGE BENEFITS	[redacted]	[redacted]	[redacted]	[redacted]
	TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A+B+C)	[redacted]	[redacted]	[redacted]	[redacted]
D.	EQUIPMENT				
	TOTAL EQUIPMENT	0	0	0	0
E.	TRAVEL				
	1. DOMESTIC (INCL. CANADA MEXICO AND U.S. POSSESSIONS)	0	0	0	0
	2. FOREIGN	0	0	0	0
	TOTAL TRAVEL	0	0	0	0
F.	PARTICIPANT SUPPORT (1 REU)				
	1. STIPENDS	0	0	0	0
	2. TRAVEL	0	0	0	0

	3. SUBSISTENCE	0	0	0	0
	4. OTHER (Tuition)	0	0	0	0
	(1) TOTAL PARTICIPANT COSTS	0	0	0	0
G.	OTHER DIRECT COSTS				
	1. MATERIALS AND SUPPLIES	0	0	0	0
	2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION	0	0	0	0
	3. CONSULTANT SERVICES	0	0	0	0
	4. RENT (see comment to determine if IC is charged on rent)	0	0	0	0
	5. SUBAWARDS	0	0	0	0
	6. ( )TUITION ($10,470 FY16/17)	10,780	11,319	11,885	33,984
	7. OTHER	0	0	0	0
	TOTAL OTHER DIRECT COSTS	10,780	11,319	11,885	33,984
H.	TOTAL DIRECT COSTS (A THROUGH G)	[redacted]			
I.	INDIRECT COSTS (SPECIFY RATE AND BASE)				
	Base 1 (Subtract Subaward Amt > 25K)	[redacted]			
	Rate 1 (ie: 49%, 26%, 55%)	0.490	0.490	0.490	
	TOTAL INDIRECT COSTS	[redacted]			
J.	TOTAL DIRECT AND INDIRECT COSTS (H + I)				
K.	RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECT SEE GPG II.D.7j.)	0	0	0	0
L.	AMOUNT OF THIS REQUEST (J) OR (J MINUS K)	[redacted]			
M.	COST-SHARING: PROPOSED LEVEL $	0	0	0	0