Welcome!

Welcome to the AIMES Land DA Working Group's 4^{th} Annual Land DA Community virtual workshop on "Developments in Land Data Assimilation"! The key goals of this workshop are to continue building a global land DA community and to share knowledge on the technical challenges that we're all facing.

We look forward to seeing you on Monday, June 24th!

-- The AIMES Working Group on Land Data Assimilation

Table of Contents:

1.…………………………*Agenda*
2… *Asynchronous Participation*
3.………………*Abstract Booklet*
4.……………..*Participant Roster*

Updated June 20, 2024

MONDAY, JUNE 24, 2024		
EDT		
10:00 am	Welcome	AIMES Land DA Working Group
	SESSION 1: OUTLOOKS IN LAND DA	
10:05 am	INVITED SPEAKER: Developing a new JEDI-based Land Surface DA system for the Met Office	Samantha Pullen, Met Office, UK
10:20 am	Using 4DEnVar to estimate photosynthesis and stomatal conductance parameters in JULES.	Natalie Douglas, University of Reading, UK
10:30 am	Discussion	
	SESSION 2: LAND DA IN EARTH SYSTEM MODELS	
10:40 am	Advanced Arctic Terrestrial Data Assimilation: Integrating DVM-DOS-TEM with NCAR DART for Enhanced Ecological Modeling	Chu-Chun Chang, Woodwell Climate Research Center, USA
10:50 am	HGS-PDAF (version 1.0): a modular data assimilation framework for an integrated surface and subsurface hydrological model	Qi Tang, University of Neuchâtel, Switzerland
11:00 am	Harmonizing terrestrial carbon cycle observations over CONUS NEON sites: Assessing the information contributions of multiple data constraints	Dongchen Zhang, Boston University, USA
11:10 am	Discussion	
11:25 am	BREAK	
	SESSION 3: MACHINE LEARNING IN LAND DA RESEARCH AND APPLICATIONS (1st session)	
11:35 am	INVITED SPEAKER: SMAP-HydroBlocks: Challenges and Opportunities in Integrating Soil Moisture Data with Artificial Intelligence	Noemi Vergopolan, Rice University, USA
11:50 am	A deep learning observation operator to assimilate SIF in the LDAS-Monde system	Pierre Vanderbecken CNRM/Météo-France, France
12:00 pm	Discovering emergent soil carbon mechanisms with a biogeochemistry-informed neural network (BINN)	Haodi Xu, Cornell University, USA
12:10 pm	Discussion	

SESSION 4: COMMUNITY UPDATES		
12:25 pm	Parameter Data Assimilation and Machine Learning Review Paper Update	Natalie Douglas, Reading University, UK
	Terrestrial Carbon Community Assimilation System (TCCAS) Training Program	Tristan Quaife, Reading University, UK
	International Land Modeling Forum (ILMF) Update from the Parameter Estimation Working Group and CMIP7 Fast Track Activities	Natasha MacBean, Western University, Canada
	NEON Forecast Challenge	Michael Dietze, Boston University, USA
12:40 pm	Discussion	
1:00 pm	END	

TUESDAY, JUNE 25, 2024		
EDT		
10:00 am	Welcome	
SESSION 1: MACHINE LEARNING IN LAND DA RESEARCH AND APPLICATIONS (2nd session)		
10:05 am	Convergence in simulating global soil organic carbon by structurally different models after data assimilation	Feng Tao, Cornell University, USA
10:15 am	Exploring the Potential of History Matching for Land Surface Model Calibration	Simon Beylat, CNRS-LSCE/ISPL, France
10:25 am	Machine learning-based observation operators to assimilate microwave and SIF satellite observations into the ECMWF integrated forecast system	Sebastien Garrigues, ECMWF, UK
10:35 am	Discussion	
10:50 am	Breakout Groups	
11:20 am	Break	

SESSION 2: INCLUDING NOVEL OBSERVATIONS AND PROCESSES IN LAND DA SYSTEMS (1st session)		
11:25 am	Assimilating SIF and SWE observations into CLM using DART to improve GPP over the high mountains in the Western US	Xueli Huo, University of Utah, USA
11:35 am	Improving water cycling through satellite soil moisture data assimilation with DART	Daniel F.T. Hagan, Ghent University, Belgium
11:45 am	Terrestrial Carbon Community Assimilation System (TCCAS)	Thomas Kaminski, The Inversion Lab, Germany
11:55 am	Discussion	
SESSION 3: INCLUDING NOVEL OBSERVATIONS AND PROCESSES IN LAND DA SYSTEMS (2nd session)		
12:10 pm	Soil carbon assimilation effectively constrains carbon-cycle model forecasting	Qianyu Li, Independent, USA
12:20 pm	Contributions of irrigation modeling, soil moisture and snow data assimilation to the skill of high-resolution digital replicas of the Po basin water budget	Gabrielle De Lannoy KU Leuven, Belgium
12:30 pm	Diversifying stomatal parameters improves predictions of evapotranspiration and streamflow in the Catchment-CN4.5 land surface model	Trent W. Robinett, Stanford University, USA
12:40 pm	Discussion	
12:55 pm	Workshop Wrap Up	AIMES Land DA Working Group
1:00 pm	END	

AIMES
analysis, integration, and modeling of the earth system

Developments in Land Data Assimilation

June 24-25, 2024 · 10:00-13:00 EDT / 16:00-19:00 CEST / [Time Zone]
Virtual Workshop: Asynchronous Participation

At the end of each day, the videos from the talks will be posted (pending speaker approval), on the workshop's YouTube Playlist found here:

https://www.youtube.com/playlist?list=PLXPEVXGzl0mhx75RWiLMQUK2gTlw5_Poh

We also encourage you to use the Github discussion forum for this workshop to ask questions for the speakers, start a new discussion, or engage in existing discussions.:

https://github.com/land-da-community/2024-developments-in-land-DA-workshop/discussions

To participate, you can set up a free GitHub account here: https://github.com. The discussions are organized by the workshop's themes such as:

- Including novel observations and processes in land DA systems;
- Land data assimilation in Earth System Modeling;
- Machine learning in land DA research and applications;
- General Q&A;
- Community updates.

The workshop co-chairs will monitor the discussion.

Developments in Land Data Assimilation

June 24-25, 2024 · 10:00-13:00 EDT / 16:00-19:00 CEST / [Time Zone]
Virtual Workshop: Abstract Booklet

Exploring the Potential of History Matching for Land Surface Model Calibration

Simon Beylat[1], Nina Raoult[2]; Philippe Peylin[3]

[1] CNRS-LSCE/ISPL, France; [2] University of Exeter, UK; [3]CNRS-LSCE/ISPL, France

Land surface models are becoming increasingly complex in order to better represent the terrestrial part of the Earth System Models. Quantifying parameter uncertainty is therefore essential both for model development and for more accurate projections. In this study, we compare, via a twin experiment, 2 methods with different philosophies: History Matching (HM), a conservative emulator-based method, and Variational Data Assimilation (VarDA), commonly used for parameter estimation in land surface models. First, we evaluate the ability to recover known parameter values. The results obtained using VarDA correspond closely to the observations. However, the known parameter values are not always recovered, which highlights the equifinality of the parameter space and therefore the need to estimate a posterior distribution of the parameters. HM, on the other hand, provides a reasonably good fit and gives a good estimate of a posterior distribution of parameters more easily and under less stringent assumptions, which this time includes the true parameters. We then explore the ability of HM to simultaneously use different metrics targeting different model output behaviours. Our results show that using multiple metrics not only improves model-data fit, but also our understanding of model parameter relationships, highlighting its potential for future terrestrial data assimilation studies.

Advanced Arctic Terrestrial Data Assimilation: Integrating DVM-DOS-TEM with NCAR-DART for Enhanced Ecological Modeling

Chu-Chun Chang[1], Elchin Jafarov[1], Hélène Genet[2], Ruth Rutter[2], Doğukan Teber[1], Brendan Rogers[1]

[1]*Woodwell Climate Research Center, Falmouth, MA;* [2]*University of Alaska, Fairbanks, AK.*

Over the past 40 years, the Arctic has experienced rapid climate change, leading to extensive permafrost thaw and significant carbon release into the atmosphere. Current models estimating greenhouse gas (GHG) emissions from thawing permafrost display a wide range of uncertainties. To better constrain estimates of GHG emissions and enhance current predictions, we are developing an advanced data assimilation system.

To model ecosystem changes in the Arctic, we utilized the Dynamic Vegetation and Dynamic Organic Soil Terrestrial Ecosystem Model (DVM-DOS-TEM) model. This model encompasses the ecological and cryohydrological processes necessary to simulate the impact of thawing permafrost on soil, and vegetation carbon dynamics. We embedded the DVM-DOS-TEM model into the NCAR's Data Assimilation Research Testbed (DART) software package. TEM-DART can assimilate observational data to adjust model predictions to better match historical states and dynamics.

The TEM-DART is currently developed to effectively update model surface and subsurface components. We currently conduct tests to update the canopy leaf area index (LAI) and soil temperature. While leaf area index observations are observed at the canopy level, the model estimates LAI by plant functional types (PFT). We design a new way to treats LAI as an independent variable for each PFT. We use covariance between individual and total LAI to determine observation increments. This allows simultaneous updates for each PFT while constraining the total LAI within the gridcell, leveraging LAI observations for enhanced accuracy.

For soil temperature, our system updates the vertical profile by utilizing error correlations between layers, allowing adjustments throughout the soil column based on single-layer observations. Preliminary tests show promising results, although further DA tests are needed.

We will present preliminary results, challenges, and future developments using the new TEM-DART system to estimate vegetation productivity and permafrost dynamics at a long-term ecological monitoring site located in the boreal region of Alaska.

Contributions of irrigation modeling, soil moisture and snow data assimilation to the skill of high-resolution digital replicas of the Po basin water budget

Gabrielle De Lannoy[1], Michel Bechtold, Louise Busschaert, Zdenko Heyvaert, Sara Modanesi, Devon Dunmire, Hans Lievens, Augusto Getirana, Christian Massari

[1] KU Leuven, Research Institute for Geo-Hydrological Protection; [2]National Research Council, Perugia, Italy; [3]Department of Environment, Ghent University, Ghent, Belgium; [4]Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA

High-resolution water budget estimates benefit from modeling of human water management and satellite data assimilation (DA) in river basins with a large human footprint. Utilizing the Noah-MP land surface model, in combination with an irrigation module, Sentinel-1 backscatter and snow depth observations, we produce a set of 0.7-km^2 digital water budget replicas of the Po river basin (Italy) for 2015-2023.

The results demonstrate that irrigation modeling consistently improves the seasonal soil moisture variation and summer streamflow at all gauges in the valley after withdrawal of irrigation water from the streamflow (12% error reduction relative to observed low summer streamflow), even if the basin-wide irrigation amount is underestimated. Sentinel-1 backscatter DA for soil moisture updating strongly interacts with irrigation modeling: when both are activated, the soil moisture updates are limited, and the simulated irrigation amounts are reduced. Backscatter DA systematically reduces soil moisture in the spring, which improves downstream spring streamflow. Assimilating Sentinel-1 snow depth retrievals over the surrounding Alps and Apennines further improves spring streamflow in a complementary way (2% error reduction relative to observed high spring streamflow), via systematic local changes in the snow water equivalent.

Despite the seasonal improvements, irrigation modeling and Sentinel-1 DA also reveal limitations that need to be addressed to obtain optimal digital replicas. Both irrigation modeling and C-band backscatter DA cannot significantly improve short-term or interannual variations in soil moisture, and snow depth DA only impacts the deep snowpacks. Irrigation also results in a systematically prolonged high vegetation productivity. In short, irrigation modeling, Sentinel-1 backscatter and snow depth DA each have specific advantages and limitations and they act at different timescales in the production of digital water budget replicas for river basins.

Using 4DEnVar to estimate photosynthesis and stomatal conductance parameters in JULES.

Natalie Douglas[1], Tristan Quaife[1], Lina Mercado[2], Rebecca Oliver[3]

[1] University of Reading, University of Exeter[2], CEH[3]

The Joint UK Land Environment Simulator (JULES) has multiple options for modelling photosynthesis and stomatal conductance. The parameters specified in these processes are typically only known to be within a given range. The 4DEnVar data assimilation technique can be used against observational data to find improved estimates of these parameters. This method is particularly attractive in its speed, ease of implementation, lack of model code adaptation and in its avoidance of the construction of tangent linear/adjoint code through the use of a parameter ensemble. The aim of our study is to estimate the photosynthesis and stomatal conductance parameters and identify the optimal JULES option configuration using 4DEnVar against gross primary productivity, latent heat and sensible heat observations that are available at a subset of FluxNet sites. We present the progress made so far in this pursuit.

Machine Learning-Based Observation Operators to Assimilate Microwave and SIF Satellite Observations into the ECMWF Integrated Forecast System

Sebastien Garrigues[1]

[1]ECMWF

The CORSO EU-funded project aims at reducing the uncertainties in the land carbon sink estimates by leveraging Earth observation, data assimilation and machine learning. The objective is to consistently constraint both water and carbon fluxes over land simulated by the Integrated Forecast System (IFS) developed at ECMWF with the assimilation of microwave and solar-induced chlorophyll fluorescence (SIF) satellite observations. We present results on the developments of machine learning-based observation operators for (1) the normalized backscatter at 40 from ASCAT and (2) SIF from TROPOMI.

Several architectures of feedforward neural network (NN) and gradient boosted trees (XGBOOST) were tested to simulate the ASCAT backscatter and the SIF satellite measurements at global scale from the IFS model fields. For ASCAT (SIF), the ML model is trained over 2016-2017 (2019-2020), validated over 2018 (2021) and tested over 2019 (2022). For each training database, frozen soil, water bodies, snow area and orographic area were excluded. Features were selected using process-based knowledge and explainability methods.

For ASCAT backscatter, the set of predictors include soil moisture and soil temperature in the first 3 soil layers (up to 1m depth), Leaf Area Index (LAI) and latitude and longitude. A NN with 4 hidden layers, 60 neurons provides the most accurate predictions (r2=0.9 and rmse=0.87 db). The spatial distribution of the backscatter and its pattern as a function of soil moisture and LAI are accurately reproduced by the NN. SIF predictors include shortwave solar radiation, 2 m temperature and humidity, LAI, surface soil moisture, soil temperature, root-zone soil moisture and the week of the year. The SIF ML model shows good performances (r2=0.85) and accurately predict both the spatial distribution and seasonal cycle of SIF. However, lower correlation was found compared to ASCAT which is probably due to the limited information content on leaf physiology in the IFS.

Improving water cycling through satellite soil moisture data assimilation with DART

Daniel F.T. Hagan[1], Brett Raczka[2], Clara Draper[3], Jeffery Anderson[2], Mohamad El Gharamti[2], Guojie Wang[4]

[1] Ghent University; [2] National Center for Atmospheric Research; [3]CIRES/NOAA Earth System Research Laboratories; [4]Nanjing University of Information Science and Technology

In the last three decades, the Data Assimilation Research Testbed (DART) has been consistently developed to meet key challenges in assimilation of observations for earth system models. This study forms part of that endeavour to extend DART's added value to soil moisture (SM) DA. We assess the impact of some important features of DART for SM DA such as DART's adaptive inflation algorithm to help address the systematic bias between the prior estimate and observed soil moisture values. We also preprocess the observations with different methods (e.g. CDF matching) to help address systematic biases. Here, we assimilate SM from the ESA's CCI product into NCAR's CLM4.5 using DART over China, a domain which covers a wide range of climate scenarios. The results showed high consistencies (domain averaged R: 0.86) between the priors of the assimilation and SM observations relevant for drought and flood forecasts. Comparison of the posterior estimates with ERA5 SM showed high correlations over the humid, energy-limiting (R>0.8), and transition (R>0.6) regions where observations are mostly available across the year. We also found low ubRMSD in the energy-limiting(ubRMSD<0.03) although these were higher in transition and arid regions(ubRMSD<0.07). However, during the warm seasons, correlations increased(R>0.8) and ubRMSD decreased(<0.05) in transition and arid regions which characterize water-limiting areas of evaporation. These results were especially evident over the Tibetan Plateau, a challenging region for SM estimation. Validations with independent in-situ data (mostly in humid, East-China) across different soil depths showed improved(reduced) correlations(ubRMSD) in lower soil depths, although these improvements were minimal in stations close to openwater. This suggests the importance of applying vertical localization, to help dampen noise at deeper soil levels. Overall, we found improvements when the impact of the observations for the DA was prioritized over higher observation acceptance for both the priors and posterior SM estimates.

Assimilating SIF and SWE observations into CLM using DART to improve GPP over the high mountains in the Western US

Xueli Huo[1]

[1] University of Utah

High mountains over the Western US has high biomass and the potential for terrestrial carbon storage. We would like to investigate in which area and time GPP over the high mountains in the western US is limited by leaf nitrogen and soil moisture. To answer this question, we assimilated solar-induced fluorescence (SIF) and snow water equivalent (SWE) observations into CLM to constrain leaf nitrogen and soil moisture, respectively. We then compared the gross primary productivity (GPP) from these two data assimilation experiments with the default CLM run. In the SIF assimilation experiment, we used the CLM version with the SIF module developed by Li to model SIF. In the SWE assimilation experiment, we introduced a new scheme to add SWE in CLM when observations indicate there is snow on the ground, while CLM indicates that all snow has melted. It turns out SIF assimilation is able to improve GPP by adjusting leaf nitrogen, and assimilating snow water equivalent improves soil moisture and, consequently, enhances GPP to some extent.

Terrestrial Carbon Community Assimilation System (TCCAS)

Thomas Kaminski[1], Wolfgang Knorr[1], Michael Voßbeck[1], Mathew Williams[2], Timothy Green[2], Luke Smallman[2], Marko Scholze[3], Tristan Quaife[4], Tea Thum[5], Sönke Zaehle[6], Peter Rayner[1], Susan Steele-Dunne[7], Mariette Vreugdenhil[8], Mika Aurela[5], Alexandre Bouvet[9], Emanuel Bueechi[8], Wouter Dorigo[8], Tarek S. El-Madany[6], Marika Honkanen[5], Yann H. Kerr[9], Anna Kontu[5], Juha Lemmetyinen[5], Hannakaisa Lindqvist[5], Arnaud Mialon[9], Tuuli Miinalainen[5], Amanda Ojasalo[5], Shaun Quegan[10], Pablo Reyez Muñoz[11], Nemesio Rodriguez-Fernandez[9], Mike Schwank[12], Jochem Verrelst[11], Matthias Drusch[13], and Dirk Schüttemeyer[13]

[1] The Inversion Lab; [2]University of Edinburgh, UK; [3]University of Lund, Sweden; [4]University of Reading, UK; [5]FMI, Helsinki, Finland; [6]MPI BGC Jena, Germany; [7]TU Delft, The Netherlands; [8]TU Wien, Austria; [9]CESBIO Toulouse, France; [10]University of Sheffield, UK; [11]University of Valencia, Spain; [12]Swiss Federal Institute for Forest, Snow and Landscape Research, Switzerland; [13]ESA, ESTEC, The Netherlands

We report on the development of the Terrestrial Carbon Community Assimilation System (TCCAS), an activity funded by the European Space Agency within its Carbon Science Cluster. TCCAS is built around the newly developed D&B terrestrial biosphere model. D&B builds on the strengths of its two component models, DALEC and BETHY, in that it combines the dynamic simulation of the carbon pools and canopy phenology of DALEC with the dynamic simulation of water pools, and the canopy model of photosynthesis and energy balance of BETHY. A suite of observation operators allows the simulation of solar-induced fluorescence (SIF), fraction of absorbed photosynthetically active radiation, vegetation optical depth from passive microwave sensors, and surface layer soil moisture. The model is embedded into a variational assimilation system that adjusts a control vector to match the observational data streams. For this purpose TCCAS is provided with efficient tangent and adjoint code. The control vector consists of a combination of initial pool sizes and process parameters in the core model and in the observation operators. We show validation against independent data streams including GPP and above-ground biomass. We quantify the added value of each assimilated data stream by its capability to constrain the simulated terrestrial carbon uptake. We are about to release TCCAS as a community tool and are organising a hybrid training event on October 7 and 8 at ESRIN in Italy.

Soil carbon assimilation effectively constrains carbon-cycle model forecasting

Qianyu Li[1], Dongchen Zhang[2], Alexis Helgeson[3], Michael Dietze[2], Shawn P. Serbin[4]

[1] Independent; [2]Boston University; [3]Yale University; [4]NASA Goddard Space Flight Center

Accurate modeling and prediction of soil organic carbon (SOC) stocks are critical for realistic estimates of climate-carbon feedback and ecosystem carbon sequestration potential. Process-based models are widely used for simulating the dynamics of SOC stocks but often have considerable uncertainty. Exploring how process-based models might be better informed by observations is crucial for improving the ability to forecast SOC. In this study, we assimilated the SoilGrids SOC datasets into the Simplified Photosynthesis and Evapotranspiration (SIPNET) model for 39 National Ecological Observatory Network (NEON) terrestrial sites across the continental US using the Predictive Ecosystem Analyzer (PEcAn) platform. We explored several assimilation strategies to determine the best approach for assimilation. Our results show that assimilation results compare best with other SOC measurements when using a time-averaged and variance-inflated constraint from SoilGrids SOC data and when spatial correlation is considered. Indirect constraints on SOC from Moderate Resolution Imaging Spectroradiometer (MODIS) leaf area index, LandTrendr aboveground biomass, and Enhanced Soil Moisture Active Passive (SMAP) soil moisture data also help to reduce the uncertainty in SOC prediction. The effectiveness of data constraints is strongly determined by the relative uncertainty of both models and observations across sites, therefore careful characterization and propagation of uncertainties is important. Overall our results show that data assimilation represents a promising approach to providing more realistic SOC estimates for regional and global carbon budgets in the future.

Developing a new JEDI-based Land Surface DA system for the Met Office

Samantha Pullen[1], Cristina Charlton-Perez[1], Brett Candy[1], Lorenzo Milazzo[1]

[1] Met Office, UK, Met Office, UK

The Met Office is undertaking a massive programme of work reformulating and redesigning our complete weather and climate research and operational/production systems, including oceans and the environment, to allow us to fully exploit future generations of supercomputer for the benefits of society. For our observation processing and data assimilation capabilities we have chosen to adopt the JEDI (Joint Effort for Data assimilation Integration) framework and work is already well underway developing JEDI-based systems for observation processing and data assimilation for the atmosphere. Work is now starting on planning and development of a new JEDI-based land surface data assimilation (LSDA) system to bring land surface and atmosphere DA under a common framework. This will allow more efficient sharing of expertise and of development and maintenance effort while opening up opportunities for increased coupling between the land and atmosphere. This presentation will discuss our motivations and aims for the new JEDI-based LSDA system, describe the progress so far, and show how we plan to transition towards implementation in the Met Office operational systems.

Diversifying stomatal parameters improves predictions of evapotranspiration and streamflow in the Catchment-CN4.5 land surface model

Trent W. Robinett[1], Jana Kolassa[2,3.], Meng Zhao[4], Yanlan Liu[5], Alexey Shiklomanov[2], Alexandra G. Konings[1]

[1] Department of Earth System Science, Stanford University, Stanford, CA, United States; [2]Science Systems and Applications Inc., Lanham, MA, United States; [3]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, United States; [4]Department of Earth and Spatial Sciences, University of Idaho, Moscow, ID, United States; [5]The Ohio State University, School of Earth Sciences, Columbus, Ohio, United States

Land surface models (LSMs) struggle to predict the response of vegetation to drought, reducing model prediction accuracy for many hydrologic variables. We hypothesize that this shortcoming arises in part from poor representation of the spatial variability of the LSM parameters that characterize vegetation hydrologic processes. For example, stomatal marginal water use efficiency (g1, the slope of the Medlyn stomatal conductance equation) varies strongly with vegetation type and structure. However, LSMs typically parameterize g1 via plant functional type, which is ineffective at capturing the true spatial of variability g1. Here, we develop an environmental filtering (EF) scheme that generates pixel-specific g1 within the Catchment-CN4.5 LSM by predicting g1 as a function of local precipitation and canopy height. Although the functional form of this relationship is prescribed, the exact shape is updated within Catchment-CN4.5 itself to minimize compensating errors with other preset parameters. This is achieved by constraining EF-parameterized Catchment-CN4.5 outputs with GLEAM evapotranspiration and CAMELS streamflow. Catchment-CN4.5 is then run across the continental United States with the location-specific g1 values predicted by the EF scheme. We found that parameterizing g1 via EF consistently improves Catchment-CN4.5 predictions of streamflow, decreasing model error by as much as 80% in some basins and by 4% on average across all basins. For ET, we found that EF-based g1 parameterization reduced model error by 0.5% across all model pixels, but individual pixel responses were mixed. However, plant functional type-based and EF-based parameterizations led to very different responses of ET to drought in drought-sensitive pixels, further informing the importance of correct vegetation parameterization for capturing the response of ET to drought in LSMs.

HGS-PDAF (version 1.0): a modular data assimilation framework for an integrated surface and subsurface hydrological model

Qi Tang[1], Hugo Delottier[1], Wolfgang Kurtz[3], Lars Nerger[4], Oliver S. Schilling[1,2,5], and Philip Brunner[1]

[1]Centre for Hydrogeology and Geothermics (CHYN), University of Neuchâtel, Neuchatel, 2000, Switzerland; [2]Hydrogeology, Department of Environmental Sciences, University of Basel, Basel, 4056, Switzerland; [3]German Meteorological Service, Centre for Agrometeorological Research, Branch Office Weihenstephan, Freising, 85354, Germany; [4]Alfred-Wegener-Institut, Helmholtz Zentrum für Polar- und Meeresforschung, Bremerhaven, 27570, Germany; [5]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, 8600, Switzerland

This work describes a modular ensemble-based data assimilation (DA) system which is developed for an integrated surface–subsurface hydrological model. The software environment for DA is the Parallel Data Assimilation Framework (PDAF), which provides various assimilation algorithms like the ensemble Kalman filters, non-linear filters, 3D-Var and combinations among them. The integrated surface–subsurface hydrological model is HydroGeoSphere (HGS), a physically based modelling software for the simulation of surface and variably saturated subsurface flow, as well as heat and mass transport. The coupling and capabilities of the modular DA system are described and demonstrated using an idealised model of a geologically heterogeneous alluvial river–aquifer system with drinking water production via riverbank filtration. To demonstrate its modularity and adaptability, both single and multivariate assimilations of hydraulic head and soil moisture observations are demonstrated in combination with individual and joint updating of multiple simulated states (i.e. hydraulic heads and water saturation) and model parameters (i.e. hydraulic conductivity). With the integrated model and this modular DA framework, we have essentially developed the hydrologically and DA-wise robust toolbox for developing the basic model for operational management of coupled surface water–groundwater resources.

Convergence in simulating global soil organic carbon by structurally different models after data assimilation

Feng Tao[1]

[1]Cornell University

Current biogeochemical models produce carbon–climate feedback projections with large uncertainties, often attributed to their structural differences when simulating soil organic carbon (SOC) dynamics worldwide. However, choices of model parameter values that quantify the strength and represent properties of different soil carbon cycle processes could also contribute to model simulation uncertainties. Here, we demonstrate the critical role of using common observational data in reducing model uncertainty in estimates of global SOC storage. Two structurally different models featuring distinctive carbon pools, decomposition kinetics, and carbon transfer pathways simulate opposite global SOC distributions with their customary parameter values yet converge to similar results after being informed by the same global SOC database using a data assimilation approach. The converged spatial SOC simulations result from similar simulations in key model components such as carbon transfer efficiency, baseline decomposition rate, and environmental effects on carbon fluxes by these two models after data assimilation. Moreover, data assimilation results suggest equally effective simulations of SOC using models following either first-order or Michaelis–Menten kinetics at the global scale. Nevertheless, a wider range of data with high-quality control and assurance are needed to further constrain SOC dynamics simulations and reduce unconstrained parameters. New sets of data, such as microbial genomics-function relationships, may also suggest novel structures to account for in future model development. Overall, our results highlight the importance of observational data in informing model development and constraining model predictions.

A deep learning observation operator to assimilate SIF in the LDAS-Monde system

Pierre Vanderbecken[1], Vural Jasmin[1], Rojas-Muñoz Oscar[1], Bonan Bertrand[1], Calvet Jean-Christophe[1]

[1] CNRM/Météo-France, France

The Solar-Induced Fluorescence (SIF) of the vegetation is a good indicator of the photosynthesis activity. The recent level 2 SIF product from the TROPOMI instrument onboard SENTINEL-5P provides an all-sky, daily estimate of the SIF all around the globe. Due to the large swath width and the heliosynchronous orbit of the satellite, we have an almost global coverage of SIF every two days. Yet, one major issue is the noisiness of the data. Hence, the questions we address here are: (1) can the SIF from TROPOMI be assimilated in a land surface model, and (2) will it improve our vegetation and carbon fluxes monitoring in the LDAS-Monde system?

A deep learning approach is used to design the observation operator, adopting a feed-forward neural network architecture. Such a method was assessed to work with other remote sensing products like ASCAT backscatter. The predictors (aka features) used as inputs of the operator are either geographical parameters or observed Leaf Area Index (LAI). Using observed LAI avoids learning LDAS-Monde model bias. The observation-driven operator is trained on a global and continental scale for a whole year of observations and evaluated for another year. The forward operator prediction exhibits a good match with the observations but still poorly reproduces the high SIF values from TROPOMI.

The operator was used in different assimilation experiments over the Ebro Valley, a region difficult to model due to irrigated croplands. The assimilation of SIF improves our vegetation monitoring, yet it is preferable to co-assimilate SIF with LAI observations due to the noise of the SIF estimate.

SMAP-HydroBlocks: Challenges and Opportunities in Integrating Soil Moisture Data with Artificial Intelligence

Noemi Vergopolan[1]
[1] Rice University

This presentation introduces challenges and opportunities in integrating artificial intelligence (AI) and land data assimilation (LDA) exemplified through the development of SMAP-HydroBlocks, the first 30m resolution satellite-based surface soil moisture dataset in the United States. By integrating satellite observations and land surface modeling with AI trained on in-situ data, this research aimed to overcome some of the challenges of traditional LDA, including the computational expense of high-resolution ensembles, the need for unbiased models and observations, and spatial scale mismatches between processes, observations, and models, which challenge calibration and validation in heterogeneous environments.

This integration enhances our ability to model processes at fine spatial resolution and combine multi-source and multi-scale observations, thereby improving the representativeness and accuracy of these processes. Specifically, AI aids in identifying and mitigating biases, uncovering the drivers of errors, quantifying system uncertainties, and quantifying the added value of satellite observations. The class of recurrent neural networks, in particular, offers opportunities to capture process memory, further refining our models. Despite these advancements, challenges remain with respect to benchmarking and transferability to sparsely monitored regions, highlighting the need for reliable spatiotemporal coverage of in-situ data and adequate training/testing sampling strategies. The SMAP-HydroBlocks dataset exemplifies the successful application of these approaches, providing detailed soil moisture information that supports a range of applications, from hydroclimate extremes to predicting vegetation carbon storage.

We discuss future opportunities, emphasizing the integration of AI data assimilation models trained online within physics models' runtime—similar to how covariance error matrices are updated online in traditional DA. By bridging data gaps with innovative AI approaches, this research aims to achieve unprecedented resolution and accuracy in soil moisture data. These advancements will support more effective environmental management and decision-making in a changing climate, ultimately enhancing our ability to respond to and mitigate climate-related challenges.

Discovering Emergent Soil Carbon Mechanisms with a Biogeochemistry-Informed Neural Network (BINN)

<u>Haodi Xu</u>[1] Feng Tao[1], Joshua Fan[1], Carla P. Gomes[1], Yiqi Luo[1]

[1]Cornell University

The advent of big data and ever-developing artificial intelligence (AI) provides unprecedented opportunities to enhance our understanding of the global carbon cycle and other biogeochemical processes. However, retrieving knowledge of processes from big data remains a grand challenge. Here, we developed a Biogeochemistry-Informed Neural Network (BINN) that integrates a vectorized process-based soil carbon cycle model (i.e., Community Land Model version 5, CLM5) into a neural network (NN) structure to gain emerging knowledge of soil organic carbon (SOC) storage from big data. BINN demonstrates a high accuracy in retrieving the mechanistic causes underlying observational data. We use BINN with a real-world observational SOC database to retrieve six major processes in soil carbon cycle across the US continent and compare the results with another Bayesian-inference approach. The high agreement on the spatial patterns of processes between these two approaches further confirms BINN's ability in retrieving mechanistic knowledge from big data. In addition, the seamless integration of NN and vectorized CLM5 supports an improved optimization efficiency by >30-fold compared with other state-of-the-art data assimilation approaches. A transformative tool harnessing the power of both AI and process-based modeling in interpreting big data is thus established for new scientific discoveries in biogeochemistry.

Harmonizing terrestrial carbon cycle observations over CONUS NEON sites: Assessing the information contributions of multiple data constraints

Dongchen Zhang[1], Qianyu Li[2], Alexis Helgeson[3], Shawn P. Serbin[4], Michael Dietze[1]

[1] Boston University; Brookhaven National Laboratory[2]; Yale University[3]; NASA Goddard Space Flight Center[4]

Accurate inventories of terrestrial carbon pools and fluxes are crucial for understanding ecosystem processes, tracking climate change impacts, and meeting the monitoring, reporting, and verification (MRV) requirements in international treaties and voluntary carbon markets. In meeting this need, the fusion of process-based modeling, field data, and remote sensing observations has the potential to provide more accurate and precise estimates than each alone. However, as the number of data constraints on a system increases, different sources of information can interact with each other in complex ways across space, time, and processes. In this study, we undertake a value-of-information analysis to assess the contribution of different observations to reducing carbon cycle uncertainties across pools, fluxes, and spatial domains within the PEcAn carbon cycle data assimilation system. We used a novel block-based Tobit Gamma Ensemble Filter to assimilate four synergistic data constraints, MODIS leaf area index, Landtrendr aboveground biomass, SMAP soil moisture, and SoilGrids soil organic C, into a process-based ecosystem model (SIPNET) at 39 National Ecological Observatory Network sites across the contiguous U.S. from 2012 to 2021. Results showed that not only did we greatly reduce uncertainty among the directly constrained pools but many observations were able to share information across variables and space. These indirect constraints helped identify synergies and conflicts among data streams and across space, which provides the insights for further constraining carbon inventories. Overall, soil carbon remains the largest source of uncertainty in the overall carbon budget due to both its large size and limited observational constraints.

Developments in Land Data Assimilation

June 24-25, 2024 · 10:00-13:00 EDT / 16:00-19:00 CEST / [Time Zone]
Virtual Workshop: Participant Roster

Name	Affiliation	Bio
Cédric Bacour	LSCE (CEA/IPSL), France	Research scientist at LSCE with expertise in remote sensing and inverse modeling of land surface properties. Current research focuses on: 1) data assimilation techniques to enhance simulations of carbon, water, and energy fluxes using the ORCHIDEE land surface model, and 2) algorithmic development for estimating XCO2 from satellite data.
Simon Beylat	CNRS-LSCE/ISPL, France	
Asher Samuel Bhatti	Bacha Khan University, Pakistan	I am more interested in causality and information flow
Camille Birman	Météo France, France	I work on surface data assimilation for NWP at Météo France
Yasmin L Bohak	Weizmann Institute of Science, Israel	
Bertrand Bonan	CNRM / Météo-France, France	I am a researcher at CNRM, Météo-France's research laboratory. My work is dedicated to maintaining and developing our land DA system called LDAS-Monde (scientifically and technically). I also aim to transfer LDAS-Monde into operational contexts.
Jean-Christophe Calvet	Meteo-France, CNRM, France	
Chu-Chun Chang	Woodwell Climate Research Center, USA	
Amie Corbin	Leiden University, Netherlands	
Gabrielle De Lannoy	KU Leuven, Belgium	
Jan De Pue	Royal Meteorological Institute of Belgium	Background in soil physics, currently working with ISBA-CTRIP on hydrology-vegetation interactions, and the development of a regional LDAS for applications in Belgium.
Michael Dietze	Boston University, USA	
Jiarui Dong	Lynker at NOAA/NCEP/EMC, USA	
Natalie Douglas	University of Reading, UK	
Istem Fer	Finnish Meteorological Institute, Finland	
Craig Ferguson	NASA, USA	Background in land and regional modeling, remote sensing of the water cycle, LIS-DA of soil moisture
Nina Fischer	University of Edinburgh, Scotland	
Andrew Fox	GMAO, NASA GSFC & GESTAR-II, Morgan State University, USA	

Name	Affiliation	Bio
Zheng Fu	Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences, China	
Yanjun Gan	University of Texas at Arlington, USA	
Sebastien Garrigues	ECMWF, UK	
Yang Gu	Boston University, USA	
Daniel F.T. Hagan	Ghent University, Belgium	
Dalei Hao	Pacific Northwest National Laboratory, USA	
Farhad Hassani	University of Texas at Arlington, USA	
Pin-hsin Hu	Max Planck Institute for Biogeochemistry, Germany	
Min Huang	University of Maryland, USA	
Xueli Huo	University of Utah, USA	
Akihiko Ito	University of Tokyo, Japan	
Thomas Kaminski	The Inversion Lab, Germany	Thomas has been working on assimilation of observations into terrestrial biosphere models for decades. He likes to set up new assimilation systems and to improve the existing ones.
Rajsekhar Kandala	Indian Institute of Science, Bangalore, India	I am Rajsekhar, PhD research scholar at Indian Institute of Science, Bangalore. My research focuses on the estimation of root zone soil moisture and soil hydraulic parameters using data assimilation techniques, like the Ensemble Kalman Filter.
Aleya Kaushik	CU CIRES/NOAA GML, USA	I am interested in using DA and ML for more efficient parameter optimization for land carbon processes.
Jana Kolassa	NASA GSFC, USA	
Alexandra Konings	Stanford University, USA	
Gerbrand Koren	Utrecht University, Netherlands	
Prashant Kumar	Space Applications Centre, ISRO, India	Working on satellite data assimilation, also developed a long term rainfall product over India.
Jiameng Lai	Cornell University, USA	I am interested in using DART to assimilate LAI and soil moisture
Martin Lange	Deutscher Wetterdienst, Germany	Land Surface DA for NWP application. Analysis of snow, T2m, soil temperature, and soil moisture.
Fangni Lei	University of Connecticut, USA	

Name	Affiliation	Bio
Qianyu Li	Independent, USA	
Hannah Liddy	Columbia University, USA	
Yiqi Luo	Cornell University, USA	
Zhenqi Luo	Cornell University, USA	
Natasha MacBean	Western University, Canada	
Fabienne MAIGNAN	LSCE, France	
Sanjay N C	Indian Institute of Technology Bombay, India	I am working on hyper resolution land surface model and data assimilation implementation into that
Devavat Chiru Naik	Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India	Improving the Land Data Assimilation System for Drought Monitoring
Sudarsan Natarajan	Monash University, Australia	I currently work on Land surface models for groundwater estimation and in future I am also planning to improve the land surface variables through DA techniques.
Isaac Kwesi Nooni	Wuxi University, China	
Biljana Orescanin	NASA/GMAO/SSAI, USA	
SHIVAJI PATEL	National Centre for Medium Range Weather Forecasting (NCMRWF), NOIDA, India	I am working as a senior research fellow in a data assimilation group. My main activity related to extreme weather events and it assimilation in regional model (NCUM-R, UK Met office).
Cristina Charlton-Perez	Met Office, UK	I'm working on the next generation JEDI land surface DA system at the UK Met Office
Richa Prajapati	IITB-Monash Research Academy, India	
Evan Pu	Zhejiang University, China	Integrating machine learning into a data assimilation system for global soil organic carbon simulation
Samantha Pullen	Met Office, UK	
Tristan Quaife	University of Reading, UK	
Brett Raczka	NSF NCAR, USA	I support the assimilation of land surface observations through the use of ensemble DA tools (NCAR DART). I am interested in improving the representation of carbon, water and energy cycling processes within land surface models with the help of novel/challenging observations like snow, soil moisture and SIF.

Name	Affiliation	Bio
Rolf Reichle	NASA GSFC, USA	Dr. Rolf Reichle's research interests are primarily in data assimilation of satellite land surface observations, satellite remote sensing of the land surface, the global terrestrial water, energy, and carbon cycles, large-scale land-atmosphere interactions, and applications to numerical weather prediction, short-term climate forecasting, and retrospective climate analysis. Dr. Reichle received the M.S. degree ("Diplom") in Physics from the University of Heidelberg, Germany, in 1996 and the Ph.D. degree in Environmental Engineering from the Massachusetts Institute of Technology in 2000.
Trent Robinett	Stanford University, USA	Trent is a 3rd year Ph.D. student in the Department of Earth System Science at Stanford University. Trent's research looks to update the representation of vegetation trait variability in land surface models to improve predictions of the hydrologic cycle.
Vanya Romanova	Deutscher Wetterdienst (German Weather Services), Germany	
Dibyandu Roy	Postdoctoral Fellow, IIT BOMBAY, India	I am a Hillslope hydrologist. Interested to assimilate atmospheric data with existing soil moisture data on land surface model.
Jyoti Singh	Rutgers University, USA	I am working on enhancing processes in CLMcrop to improve crop models. I am keen to listen to and collaborate with researchers focusing on data assimilation for advancing crop modeling.
Wu Sun	Carnegie Institution for Science, USA	Dr. Wu Sun is a Research Scientist in the Department of Global Ecology at the Carnegie Institution for Science. His research explores the interactions between the terrestrial carbon cycle and climate across scales by combining field-scale, atmospheric, and remote sensing observations with terrestrial biosphere models and atmospheric inverse models. He currently leads a project funded by NASA's Arctic–Boreal Vulnerability Experiment program to quantify responses of photosynthesis and respiration to climate in Arctic and boreal biomes.
Yinghao Sun	Boston University, USA	
Qi Tang	University of Basel, University of Neuchatel, Switzerland	
Feng Tao	Cornell University, USA	

Name	Affiliation	Bio
Svetlana Travova	Hydrometcenter of Russia, Russia	My research interests include land surface models and data assimilation for such models. Presently I study ML-techniques and hope to implement some of them in the LSM and/or Land DA.
Muhammad Umair	Universite de Montreal, Canada	
Pierre Vanderbecken	CNRM/Météo-France, France	
Noemi Vergopolan	Rice University, USA	
Toni Viskari	JRC, Finland	
Jasmin Vural	Météo-France, France	
Pete Weston	ECMWF, UK	
Mikki Wilburn	UW - Madison, USA	
Youlong Xia	SAIC @NCEP/EMC, USA	I am a support scientist and working on land observation processing and DA task at EMC. Major task is to work on (1) land ioda converters development and tests, (2) testing land offline DA workflow, and (3) producing land initial conditions use the snow DA to provide them in support the coupled experiments at EMC.
He-Ming Xiao	National Taiwan University, Taiwan	
Haodi Xu	Cornell University, USA	
Yuan Xue	Lynker at NOAA, USA	
Sönke Zaehle	MPI BGC, Germany	Interest in land surface modelling and model data integration to improve climate-carbon feedback simulations. Using DA and not DA methods :)
Dongchen Zhang	Boston University, USA	
Yu Zhou	Cornell University, USA	
Yu Zhu	Max Planck Institute for Biogeochemistry, Germany	
Dan Zhu	Peking University, China	