Land Data Assimilation: Making sense of hydrological cycle observations
A proposal to ISSI for an "International Team"

W.A. Lahoz *et al.*

Abstract.

This proposal is for an "International Team" to bring together various ongoing efforts in hydrological cycle studies (data assimilation theory and application, NWP, models, observations) to spend 3 1-week periods at ISSI to discuss the state-of-the-art and bring together leading scientists, with focus on the upcoming ESA Earth Explorer mission SMOS (Soil Moisture and Ocean Salinity), scheduled for launch in 2009. An expected output is, in the first instance, enhanced collaboration and, ultimately, peer-reviewed papers.

1. Scientific rationale

1.1 Background. Data assimilation (DA) combines, in an objective way, information from observations with information from a model of the evolving system (*e.g.* atmosphere, land), taking account of errors in the observations and model (Nichols 2009). It is the cornerstone of Numerical Weather Prediction (NWP) where it has contributed to improved models, improved use of observations and significantly improved forecasts (Simmons and Hollingsworth 2002). Beyond NWP, DA has also played a key role in the organization and evaluation of observational and model information of various elements of the Earth System, and is increasingly being used to assess and prepare for future Earth Observation (EO) missions (see various chapters in Lahoz *et al.* 2009).

Accurate knowledge of spatial and temporal land surface storages and fluxes are essential for addressing a wide range of important, socially relevant science, education, application and management issues. Improved estimates of land surface conditions are directly applicable to agriculture, ecology, civil engineering, water resources management, rainfall-runoff prediction, atmospheric process studies, climate and weather prediction, and disaster management (Houser *et al*. 2009).

During 2009, the ESA Earth Explorer mission SMOS (Surface Moisture and Ocean Salinity; Kerr *et al.* 2000) will be launched. A focus of this International Team will be making sense of SMOS soil moisture (SM) observations. SM is one the key geophysical variables for understanding the Earth's hydrological cycle, as it is the key variable describing land surface hydrology. Table 1 (adapted from http://www.cp34-smos.icm.csic.es/index.htm) identifies key benefits from SMOS SM measurements.

The only practical way to observe the land surface on continental to global scales is by satellite remote sensing. However, this cannot provide information on the entire system, and measurements only represent a snapshot in time. Land surface models (LSMs) can predict spatial/temporal land system variations, but these predictions are often poor, due to model initialization, parameter and forcing errors and inadequate model physics and/or resolution.

Land DA (Houser *et al.* 2009) provides an objective way to make sense of hydrological cycle observations of SM, snow cover and land surface temperature, LST (or geophysical variables related to them). Assimilation of SMOS SM measurements into a LSM will provide a best estimate of the state of the land surface, and will help evaluate SMOS measurements and LSMs. This information will improve our understanding of the land surface and its interaction with the atmosphere. Land DA will also provide an objective way of monitoring SM. Integration of SM information from SMOS, and other measurements of SM, LST and snow cover (LST and snow cover, Aqua/Terra MODIS; LST, AATSR and MSG SEVIRI; SM, AMSR-E) into an LSM using DA, will provide a comprehensive picture of the state and variability of the land surface. In this way land DA will help fulfil the potential of SMOS.

Table 1: Key benefits from SMOS soil moisture measurements.

AREA	PRODUCTS	COMMENT
Meteorology	NWP models	Soil moisture (SM) plays a fundamental role in the transfer of water and energy between the surface and the atmosphere. Introduction of this variable in current NWP models will allow improving predictions, especially important under adverse meteorological conditions.
Climatology	Models	Variability of the SM series with a long integration period may provide relevant information for the study of climate change.

Risk Management	Flooding Risk Map	The soil's risk of flooding is significantly conditioned by the amount of water stored in the vadose zone. The generation of this type of products will require the inclusion of SM data in hydrological and NWP models (precipitation predictions)
	Fire Risk Map	The risk of fire is determined by several factors, including meteorological, geophysical and biophysical factors. The information on SM may be directly assimilated in drawing up fire risk maps as they provide direct information on evapo-transpiration, water content assimilated by vegetation, quality of vegetation,.
	Famine Risk Map	The merging of geopolitical, meteorological/ climatological information and data on the quality and estimates of agricultural and/or marine production (derived with the help of SM data) may be of great use in early prediction of famine episodes in areas of Earth where resources are scarce.
	Drought Risk Model	Analysing SM trends in large areas may serve to generate drought models, along with data from other sensors.
Agriculture	Agricultural Production Estimate	On the basis of SM data and by means of the application of hydrological models, it is possible to determine the amount of water assimilated by the vegetation, a value that is very useful for estimating agricultural production.
Hydrology	Models	The content of water stored in the soil is an important parameter to be taken into consideration in any hydrological model, as it is an indispensable variable in understanding the water cycle.

1.2 Questions to be addressed by this proposal. To fulfil the potential from SMOS and analogous satellite platforms this proposal addresses the following key questions:

"***What type of observations do we assimilate?"***. Partners will assimilate satellite observations (with a focus on retrieved quantities) and use *in situ* measurements to validate the results. Team members have access to these data. Issues such as data thinning, data level (retrievals, radiances) and geophysical representation of variables will be investigated.

"What are the observation errors?". Representation of errors is fundamental to DA. One needs to consider errors in observations, background information, and model. **R**, the observation error covariance matrix, is typically assumed to be diagonal, although this is not always justified. **R** includes errors of the measurements themselves, **E**, and errors of representativeness, **F**; **R**=**E**+**F**. **B** is the background error covariance matrix; its off-diagonal elements determine how information is spread spatially from observation locations. Estimating **B** is a key part of the DA method. Estimating model error **Q** is a research topic.

In general, in DA, errors are assumed to be Gaussian. The most fundamental justification for assuming Gaussian errors, which is entirely pragmatic, is the relative simplicity and ease of implementation of statistical linear estimation under these conditions. Because Gaussian probability distribution functions (PDFs) are fully determined by their mean and their variance, the solution of the DA problem becomes computationally practical.

Typically, there are biases between different observations, and between observations and model. These biases are spatially and temporally varying, and it is a major challenge to estimate and correct them. Despite this, and mainly for pragmatic reasons, in DA it is often assumed errors are unbiased. For NWP many assimilation schemes now incorporate a bias correction, and various techniques have been developed to correct observations to remove biases; these methods are now applied to land DA (DeLannoy *et al.* 2007a, b).

Possibility distributions (more general than PDFs) have been used in the retrieval of information from satellite imagery to account for incomplete information (Verhoest *et al.* 2007), and for non-Gaussian errors, which can occur with land surface variables.

Partners will investigate the representation of observational errors; the development and implementation of bias correction methods; and the application of ideas from possibility theory. SM measurements from SMOS will be a focus.

"What models do we use?". Many LSMs have been developed and enhanced since the mid 1990s, with varying features, such as sub-grid variability, community-wide input, advanced physical representations, and compatibility with atmospheric models (Houser *et al.* 2009). Partners will investigate assimilation of hydrological cycle observations into two state-of-the-art LSMs, JULES (Blyth *et al.* 2006) and SURFEX (Giard and Bazile 2000). JULES is used by the Met Office (UK); SURFEX is developed at Météo-France and used in the HIRLAM (HIgh-Resolution Limited Area Model) and ALADIN (Aire Limitée Adaptation dynamique Développement INternational) NWP consortium.

There are strong justifications for studying a LSM uncoupled from atmospheric and ocean models. In coupled models the atmospheric model can impose strong land surface forcing biases on the LSM. For example, biases in precipitation and radiation can overwhelm the behaviour of the LSM physics. By using an uncoupled LSM, we can better specify land surface forcing using observations, use less computational resources, and address most DA development questions. Partners will consider the assimilation of SMOS and other land surface observations into uncoupled versions of JULES and SURFEX.

"What assimilation algorithms do we use?" In developing a land DA capability, several issues must be considered: most processes are 1-D (vertical); heterogeneity of land surface; most observations are at the surface; presence of non-linearities and on-off (*i.e.*, threshold) processes; different scales involved (*e.g.* observational information, flux towers to satellites); and representation of observational and model errors, which for the former must take account of representativeness errors. Three methods are commonly used for land DA (Houser *et al.* 2009): variational (3-d/4-d var); sequential (Kalman filter, extended Kalman filter, EKF); and ensemble (ensemble Kalman filter, EnKF). Because of the 1-D nature of most land surface processes, 1-D and 3-D versions of a DA scheme are often used; by 3-D we mean inclusion of horizontal information from neighbouring grids in the DA system. The EnKF (*e.g.* Sakov and Oke 2008a, b) is attractive, as it avoids computing adjoints (as for 4d-var), which is problematic given non-linearities and on-off processes at the land surface; it also provides a cost-effective representation of **B**. Several issues need to be considered in the EnKF: (i) ensemble size; (ii) ensemble collapse; (iii) correlation model for **B**; and (iv) specification of model errors.

The major drawback of the EnKF is the underlying assumption that the model states have a Gaussian distribution. The Particle Filter (PF) does not require a specific form for the state distribution, but its major drawback is that distribution of particle weights quickly becomes skewed, and a re-sampling algorithm needs to be applied (Weerts and El Serafy 2006). The PF has been applied in hydrology (Moradkhani *et al.* 2005; Weerts and El Serafy 2006).

Partners will investigate 1-D and 3-D variants of the EnKF, and recent developments with Optimal Interpolation and the EKF at Météo-France (Mahfouf *et al.* 2009), as well as the PF.

"What are the model and analysis errors?". A crucial element of DA is evaluation of observations, models and analyses, and testing of several assumptions, *e.g.* Gaussian errors; unbiased observations and models. Several diagnostics have been developed to do this (Talagrand 2009). These consist of self-consistency tests and independent tests. In general, comparison against independent data is much more significant than comparison against the assimilated observations. Thus, independent data are the ultimate arbiter of the quality of analyses. Independent data for the land can be either *in situ* or remote.

Partners will investigate these methods to evaluate the DA algorithms, including the suitability of *in situ* data (relatively high resolution, local scale) to evaluate analyses provided at relatively low resolution and regional to global scales (*i.e.*, upscaling of information). A further method investigated will be the approach of Desroziers *et al.* (2005) to estimate model/observation error covariances separately (see Reichle *et al.* 2008 for land DA).

2. Timeliness

The proposed "International Team" will provide a strong example of the multi-disciplinary nature of land DA, involving: theory and application; LSMs; EO; NWP agencies and consortia; space agencies. All participants cover these areas and will be brought together by the International Team. The International Team will be relevant to ISSI's new focus on EO. Results will be of great interest to all actors in the EO community, including:

- ***NWP agencies:*** Results will help guide the implementation of algorithms to assimilate land surface parameters;
- ***Space agencies:*** Results will provide information on the quality of SMOS data (and other EO data), and will help guide plans for future EO missions;
- ***EO data users:*** Results will provide information on the quality of EO data, and will guide scientific use of this EO data;
- ***Instrument teams:*** Results will provide information on SMOS instrument errors (as well as for other instruments), and will help guide efforts to remedy shortcomings;

- ***Modellers:*** Results will provide information on LSMs and will help guide efforts to remedy model shortcomings;
- ***Data assimilators:*** Results will provide information on land DA techniques, and will help guide efforts to improve them;
- ***Public:*** Results will help improve models, observations and algorithms used in NWP; results will also help improve climate models.

3. Objectives of the project

- **The objective of the "International Team" is to provide a multi-disciplinary platform to make sense of hydrological cycle observations, with a focus on SMOS soil moisture data.**

4. Role of ISSI

ISSI and the "International Team" will be very important to European-wide activities in land DA, and the exploitation of SMOS soil moisture observations. First, it will allow partners in the proposal to hold three week-long meetings in a focused environment, something that the partners do not have funds to do. Second, the International Team will provide a focused platform to discuss SMOS soil moisture observations. Third, the International Team will provide a focused environment for developing and strengthening collaborations, and disseminating the results of these collaborations via publications in the peer-reviewed literature. Results will also be disseminated at community-wide workshops on land DA, and SMOS. The experience of the successful "ASSET International Team", led by the proposer W.A. Lahoz, is a testament of the benefits of ISSI to the scientific community.

ISSI qualifies as a preferred implementation site for the following reasons:

Location: ISSI is centrally located in Europe;

Facilities: ISSI has excellent facilities for presenting and interpreting results;

Set-up: The focused nature of the ISSI "International Team" provides an ideal environment for bringing together the actors in a multi-disciplinary exercise. The low administrative burden associated with ISSI means that more time is dedicated to scientific activities;

Financial support: ISSI funds allow partners to do something they are not funded to do.

5. Expected output from the proposal

- "International Team". The membership will include leading experts in the disciplinary areas in land DA, and land surface EO and modelling in Europe and US;
- Reports and peer-reviewed papers.

6. Schedule

All dates in the schedule are relative to the first meeting date (Month 0; M0).

- Preparation: M-3 -> M0 (all participants);
- Meetings: 3 meetings for 1 week each, Monday-Friday (M0, M6, M12);
- Anticipated dates for the "International Team" at ISSI (October 2009; M0);
- Disseminating results at international workshops: from M2 onwards;
- Writing reports and papers: M2 -> M12 and beyond. The experience of the ASSET International Team suggests that submission of the first peer-reviewed papers would occur from M12 onwards;
- Report to ISSI: M12.

7. Workshop participation

Leading European and US actors in the disciplinary areas involved in land DA will participate in this International Team. The proposal invitation list (15, plus 1 observer from HIRLAM to participate in the first meeting, and 1 ISSI liaison) for the "International Team" is as follows:

- ***Team:*** William Lahoz - Team leader, and Sam-Erik Walker (NILU, Norway); Olivier Talagrand (LMD, France); Niko Verhoest and Valentijn Pauwels (U. Ghent, Belgium); Pavel Sakov (NERSC, Norway); Paul Houser (GMU, USA); Gabriëlle De Lannoy (GMU, USA & U.Ghent, Belgium); Jean-François Mahfouf (Météo-France, MF, France); Peter Cox (U. Exeter, UK); Nils Gustafsson (SMHI, Sweden & Met.No, Norway); Chris Schmullius (U. Jena, Germany); Yann Kerr (SMOS PI, CESBIO Toulouse, France); César Coll (U. Valencia, Spain); Jouni Pulliainien (FMI, Finland).

- ***HIRLAM Observer:*** Jeanette Onvlee (KNMI, The Netherlands);
- ***ISSI liaison:*** Lennart Bengtsson

Areas covered by participants:

Data assimilation theory: O.Talagrand, N. Verhoest, V. Pauwels, P. Sakov;
Land data assimilation developments: P. Houser, G. De Lannoy, W. Lahoz, S.-E. Walker;
Land surface models: J.-F. Mahfouf (SURFEX), P. Cox (JULES);
Numerical Weather Prediction: N. Gustafsson (SMHI/Met.No), J.-F. Mahfouf (MF);
Biosphere: C. Schmullius;
Observations: Y. Kerr (Soil moisture, SMOS), C. Coll (LST), J. Pulliainien (GlobSNOW);
HIRLAM: J. Onvlee (only 1st meeting).

8. Facilities required

One meeting room for a maximum of 20 participants: overhead projector; facilities for powerpoint presentations; small room for breakout meetings (5-10 people); wireless network; internet connections; 2-3 PCs for those people who do not bring a lap-top; coffee machine.

9. Financial requirements

- Team leader: travel expenses from Norway to Bern (1 return trip).
- "International Team" meeting: accommodation and living expenses for up to 16 participants (including Team leader).

10. External financial support

Partners have funds to support their travel to Bern. The HIRLAM project will support partners involved in the International Team (*e.g.* J. Onvlee, N. Gustafsson).

References

Blyth, E., *et al.*, 2006. JULES: a new community land surface model. *Glob. Change News.*, No. 66, October 2006.

De Lannoy, G.J.M., *et al.*, 2007a. Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. *WRR*, **43**, W09410.

De Lannoy, G.J.M., *et al.*, 2007b: State and bias estimation for soil moisture profiles by an ensemble Kalman filter. *WRR*, **43**, W06401.

Desroziers, G., *et al.*, 2005. A simple method to diagnose and adapt observation and background errors. *QJRMS*, **131**, 3385–3396.

Giard D. and E. Bazile 2000. Implementation of a new assimilation scheme for soil and surface variables in a global NWP model. *MWR*, **128**, 997-1015.

Houser, P.R., *et al.*, 2009. Land surface data assimilation. In *Data Assimilation: Making sense of observations*, Eds. W.A. Lahoz, B. Khattatov and R. Ménard, Springer, due April 2009.

Kerr, Y.H., *et al.*, 2000. The Soil Moisture and Ocean Salinity Mission: SMOS. *Earth Obs. Quarterly*, pp 18-26.

Lahoz, W.A., *et al.,* 2009. *Data Assimilation: Making sense of observations*, Eds. W.A. Lahoz, B. Khattatov and R. Ménard, Springer, due April 2009.

Mahfouf, J.-F., *et al.*, 2009. A comparison of two off-line soil analysis schemes for assimilation of screen-level observations. *JGR*, in press.

Moradkhani, H., *et al.*, 2005. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *WRR*, **41**, W05012, doi:10.1029/2004WR003604.

Nichols, N.K., 2009. Mathematical concepts of data assimilation. In *Data Assimilation: Making sense of observations*, Eds. W.A. Lahoz, B. Khattatov and R. Ménard, Springer, due April 2009.

Reichle, R.H., *et al.*, 2008. An adaptive ensemble Kalman filter for soil moisture data assimilation. *WRR*, **44**, W03423.

Sakov, P. and Oke, P.R., 2008a. Implications of the form of the ensemble transformation in the ensemble square root filters. *MWR*, **136**, 1042-1053.

Sakov P. and Oke, P.R., 2008b. A deterministic formulation of the ensemble Kalman filter: An alternative to ensemble square root filters. *Tellus*, **60A**, 361-371.

Simmons, A.J. and Hollingsworth, A., 2002: Some aspects of the improvement in skill in numerical weather prediction. *QJRMS*, **128**, 547-677.

Talagrand, O., 2009. Evaluation of assimilation algorithms. In *Data Assimilation: Making sense of observations*, Eds. W.A. Lahoz, B. Khattatov and R. Ménard, Springer, due April 2009.

Verhoest, N.E.C., *et al.*, 2007. A possibilistic approach to soil moisture retrieval from ERS synthetic aperture radar backscattering under soil roughness uncertainty. *WRR*, **43**, WR005295.

Weerts, A.H. and G.Y.H. El Serafy, 2006. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models.*WRR*, **32**, W09403, doi:10.1029/2005WR004093.