NILU Norsk institutt for luftforskning
Norwegian Institute for Air Research

Development of a land data assimilation system at NILU

W.A. Lahoz, wal@nilu.no

10th June 2009, 2nd Workshop on Remote Sensing and Modeling of Surface Properties, Toulouse, France

Thanks to S.-E. Walker (NILU), Met.no & Meteo-France

Contents

- Motivation
- Data assimilation set up
  - SURFEX model
  - Observations
  - Data assimilation algorithm
  - Plans
- Outlook

Motivation

Build Land DA capability at NILU:

Match expertise between NILU and other groups (Met.no, Meteo-France):

Strong national/international collaboration on:

Theory; Application to land surface; Operational implementation

- Theory: NERSC (Norway), LMD (France)

- NWP: Met.no, Meteo-France; HIRLAM/ALADIN

- Earth System: Land surface (GMU, USA); biosphere (U. Jena, Germany); interaction with atmosphere (several groups)

- Observations: SMOS (Y. Kerr); GlobSNOW (led by Finland); land surface satellite (several groups)

Land DA goals:

(1) Build DA algorithms for the land surface: e.g. hybrid, based on ensemble Kalman filter/particle filter, EnKF/PF (complementary methods) - focus is land forecasting;

(1) Assimilate variables, focus on EnKF (but other algorithms will be tested/ developed, e.g., EKF, extended Kalman filter): land soil temperature (LST, or associated variable), soil moisture, snow (or associated variable);

(1) Use DA to produce analyses, evaluate observations & models (e.g. SURFEX land surface parameters if appropriate) & DA algorithms

Norsk institutt for luftforskning
Norwegian Institute for Air Research

Issues in land DA:

•Observed quantities often non-linearly related to model variables

Abilities of different DA assimilation algorithms to handle non-linear observation operators need to be compared/evaluated;

•Most observations for land DA concerned with surface/near-surface conditions – important model variables represent more deep soil conditions (e.g. deep soil moisture)

Abilities of different DA methods to solve inverse problems must be compared/evaluated;

Issues in land DA (continued):

•Satellite data becoming more important for land DA. These data often have complicated observation error structures (biases, spatially correlated errors)

See ECMWF developments, OI->EKF (Drusch et al., GRL, 2009)

Abilities of different DA methods to handle complicated observation error structures must be compared/evaluated. Interaction between (systematic) model errors & (systematic) observation errors needs to be handled properly;

•Atmospheric forcing for land DA may come from models or observations (e.g. precipitation)

Specification of model error characteristics crucial.

N.B., issues EnKF/PF: non-linearity, non-Gaussianity, error characterization

Approach:

i. Theoretical development: compare EnKF/PF (NERSC/Met.no/NILU) + novel developments (hybrid DA system for land forecasting)

i. Work with NILU DA system (EnKF) + land model (SURFEX - MF) from HARMONIE/Met No + satellite observations + novel developments (errors, biases) – incorporate at NILU

i. Compare with EKF from MF (received from JF Mahfouf & compiled Met.No, ongoing NILU)

i. Consultation with Met.No & MF in future developments with land DA

Norsk institutt for luftforskning
Norwegian Institute for Air Research

Outcomes:

i. Improved land DA systems (e.g. hybrid EnKF/PF)

i. Evaluation of land DA systems (EnKF, EKF,...)

i. Understanding land/atmos interactions (feedbacks)

i. Improved use of EO (satellite data; error characterization)

i. Improved forecast & modelling capability (Better use of EO; better error characterization, model + observations: NWP)

i. Studies of land/atmosphere system (analyses)

Norsk institutt for luftforskning
Norwegian Institute for Air Research

DA Set up

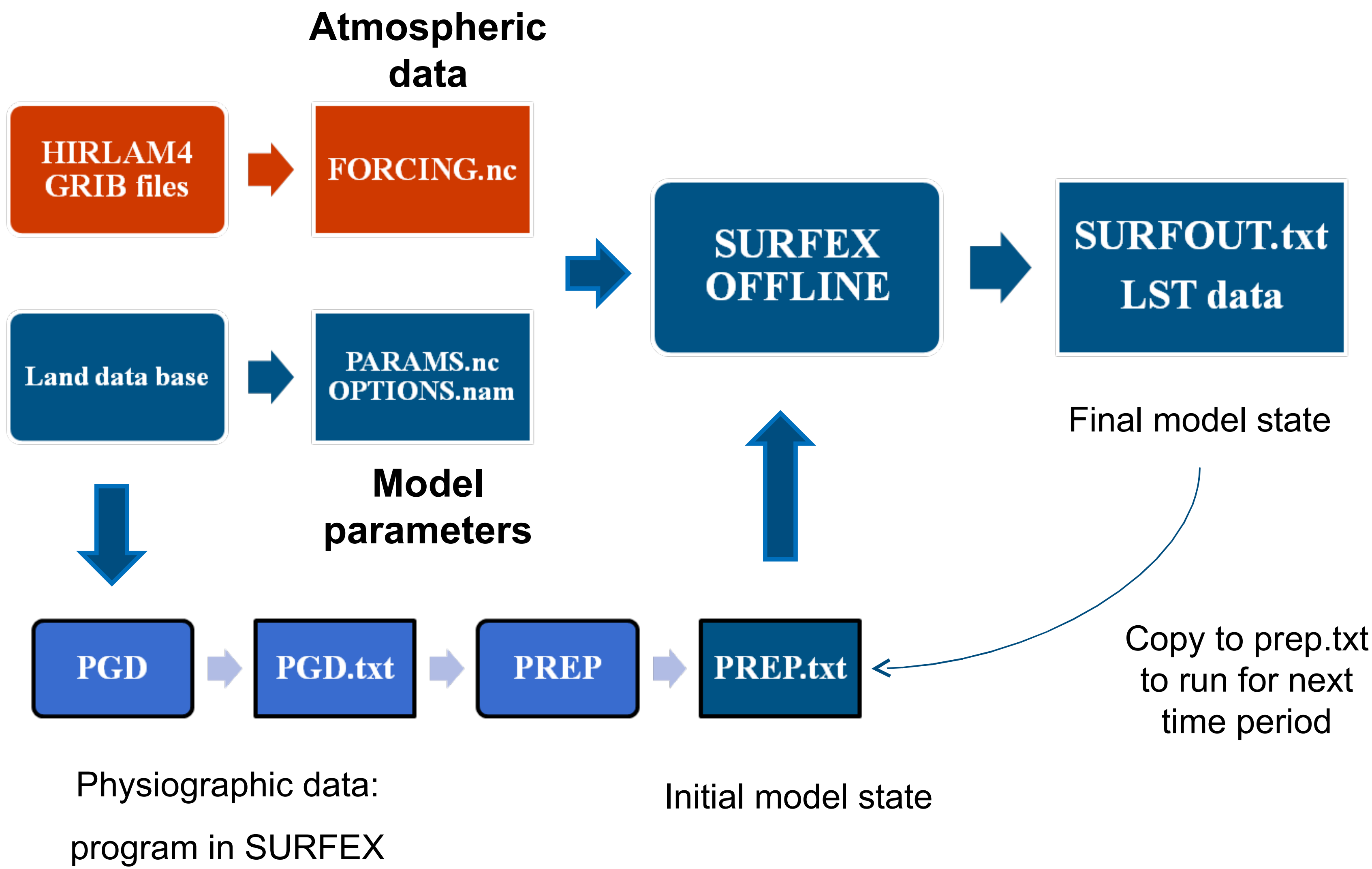
Land model:

SURFEX: Implementation at NILU

Collaboration with Met.No & Météo-France

- Using latest version (v4.8) off-line SURFEX model from Meteo-France (Giard and Bazile 2000; Le Moigne 2005;...)
- EKF from Meteo-France (Mahfouf et al. , JGR, 2009)

SURFEX model data flow (no DA) - schematic



Observations

- Plans for satellite observations of land surface temperature (LST, or related variable), snow cover (or related variable) & soil moisture (+ errors):

  - LST from EOS TERRA/MODIS and AQUA/MODIS satellites, 1 km spatial resolution: GENESI-DR application
  - Snow cover from same satellites with 500 m resolution
  - Soil moisture from EOS AQUA/AMSR-E with 25 km resolution
  - Focus on Scandinavia and specific periods (e.g. Aug 2007)

  - Soil moisture from SMOS
  - EUMETCAST (available at NILU) – SEVIRI, MODIS,...

- Also interested in SYNOP observations (T_{2m}, RH_{2m})

Norsk institutt for luftforskning
Norwegian Institute for Air Research

From S. Briggs, ESA

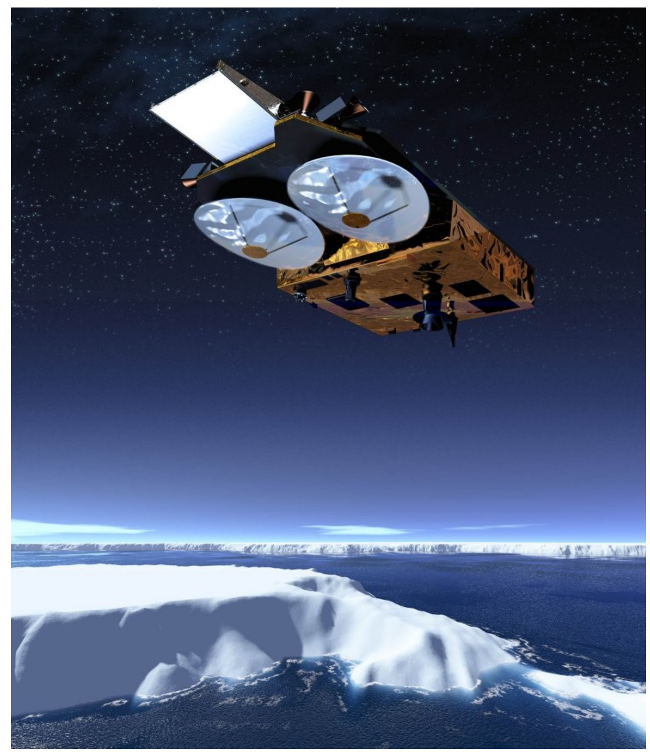
Two **more** Explorers set for launch in 2009

July

SMOS (Soil Moisture and Ocean Salinity), ESA's water mission

November

CryoSat, ESA's ice and snow mission



LAND SAF Products sent via Eumetcast

	Product	Product Coverage		Format	Eumetcast Availability	Application Category	status
LSA SAF	Surface Albedo	Europe, North Africa, South Africa, South America	3	HDF5	Europe Americas Africa	Land	MSG Based product Pre-op
LSA SAF	Land Surface Temperature	Europe, North Africa, South Africa, South America	2	HDF5	Europe Americas Africa	Land	MSG Based Product Pre-op.
LSA SAF	Downwelling Surface SW fluxes	Europe, North Africa, South Africa, South America	3	HDF5	Europe Americas Africa	Land	MSG Based product Op
LSA SAF	Downwelling Surface LW fluxes	Europe, North Africa, South Africa, South America	3	HDF5	Europe Americas Africa	Land	MSG Based Product Op
LSA SAF	Snow Cover	Europe, North Africa, South Africa, South America	3	HDF5	Europe Americas	Land	MSG Based product Pre-op

DA algorithm:

Two recent EnKF versions, implemented at NILU

- Sakov and Oke: MWR, 2008
  - An Ensemble Square Root filter (ESRF) using a symmetric Ensemble Transform Matrix (ETM): classical KF - ensemble mean. Prevents build up of ensemble outliers
- Sakov and Oke: Tellus, 2008
  - A Deterministic Ensemble Kalman Filter (DEnKF) using a linear approximation to the Ensemble Square Root Filter (ESRF) update matrix. Better for avoiding ensemble collapse. Use of localization schemes.

DA code:

- Build on EKF architecture/code (Mahfouf)
- Ensemble Square Root filter & Deterministic EnKF (Sakov & Oke)
- 1-D vertical DA initially for each grid cell (see EKF for MF)
  mainly 1-D processes in land
- 3-D assimilation later
- Fortran 90, BLAS (Basic Linear Algebra Subprograms) & LAPACK libraries
- Code general with matrix-free versions of observation operators

SURFEX EnKF CODE

- New run script: run_enkf.sh
- New main program: enkfassim.f90
- New makefile: Makefile.SURFEX.mk
- New namelist file: OPTIONS.nam

- Otherwise identical to the SURFEX-EKF system

NILU Norsk institutt for luftforskning
Norwegian Institute for Air Research

SURFEX EnKF CODE

5 options for DA: (0) no DA; (1) EnsRKF; (2) DEnsKF; (3) SIR; (4) RPF

(1)-(2): EnKF; (3)-(4): PF

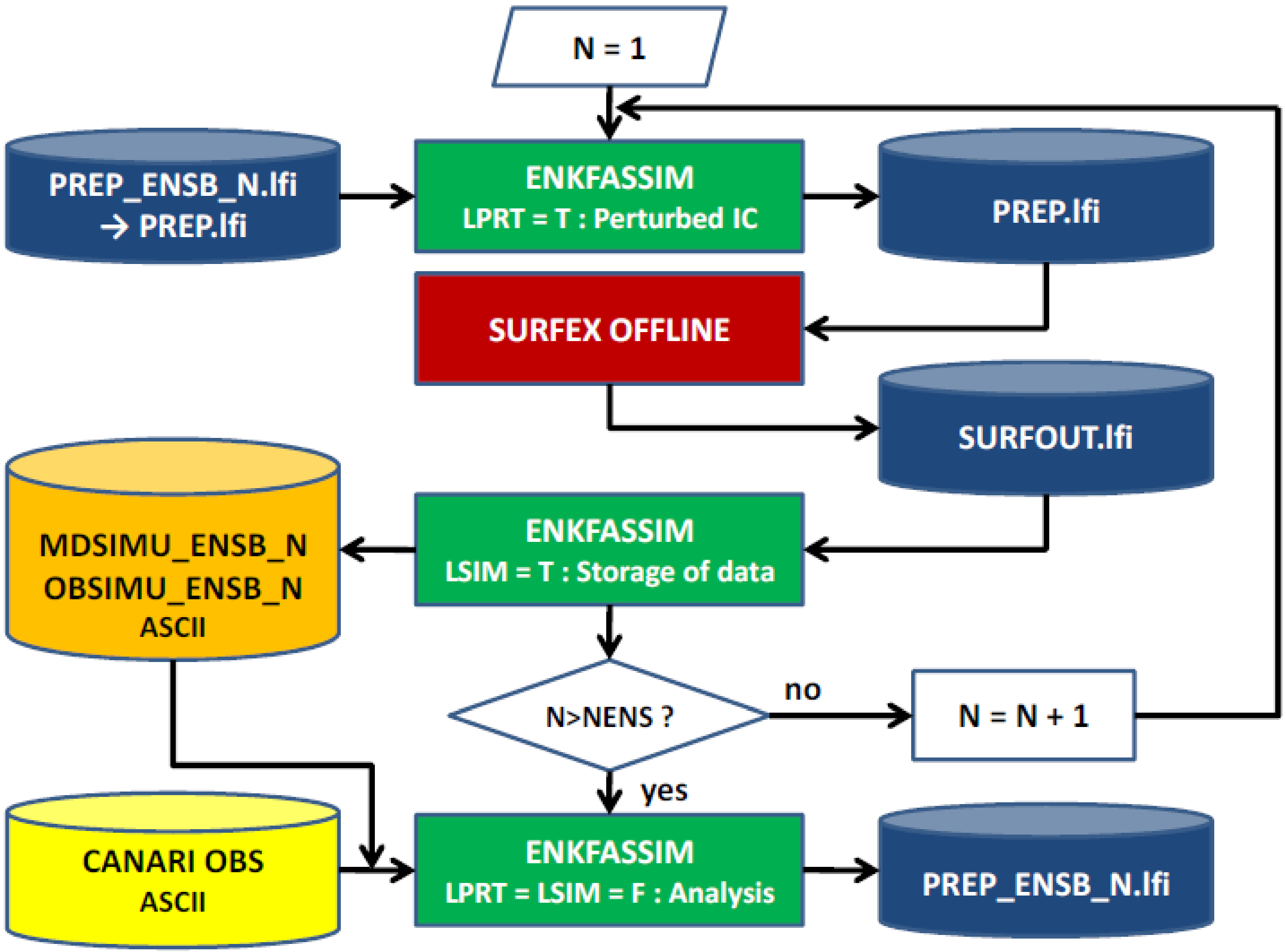
3 Observation types: screen level temp (T_{2m}); screen level RH (RH_{2m}); superficial soil moisture content (SWI) – initially use retrieved quantities – later, investigate use of radiances

Prognostic variables: TG1 (surface temp); TG2 (mean surface temp); WG1 (superficial volumetric water content); WG2 (mean volumetric water content in root-zone) – see EKF developments for SURFEX

- All can be control variables

Soil patches option: heterogeneous land surface

Flowchart of SURFEX-EnKF



PLANS:

- Compare EKF & EnKF (also various variants)
- Evaluate SURFEX model/observations using DA
- Land surface analyses – land-atmos interaction studies
- Compare/evaluate various variants of EnKF & PF: toward a hybrid system for land forecasting
- Later for DA system: Add features: e.g. more sophisticated observation errors (Verhoest,...); bias correction (Houser);...

Outlook

- Test & build DA algorithms (theory)
- Confront models with observations (ideas from NWP)
- Exciting & important problem theoretically & scientifically

- Interest to DA theory
  - Innovative DA algorithms

- Interest to NWP (recall comments from F. Bouttier & S. English)
  - Better use of observations affected by the land
  - Better initial conditions for 2-week to seasonal forecasts

- Interest to climate modellers
  - Better land surface schemes
  - Better climate models
  - Better set up/design of experiments
  - Better understanding of performance of climate models

- General interest to land/atmosphere scientists: atmosphere-land processes

Extra slides

Visits to NILU/workshops

W'shops (June 2009, Meteo-France):

- 2nd workshop Remote sensing & modelling of surface properties
- 1st meeting Expert team surface processes

COST Action 0804: "Advancing the integrated monitoring of trace gas exchange between biosphere and atmosphere"; 2009-2013 – WAL in MC

Upscaling between local measurements & global measurements: heterogeneity

From Paul Houser:

Land Surface Data Assimilation: Progress and Realities

Current Status:

- Soil moisture, skin temperature, and snow assimilation have been demonstrated.
- Evapotranspiration, runoff, groundwater (gravity), and carbon assimilation are underway

Data Assimilation Tradeoffs:

- Tradeoff between using **complex data assimilation techniques**, the **ability to use all the available data** and **operational needs and realities** due to the large computational burdens.
- Tradeoff in dimensionality of data assimilation methods –need may depend on scale.
- Tradeoff between fine resolution and large area implementation.

Land Surface Data Assimilation Realities

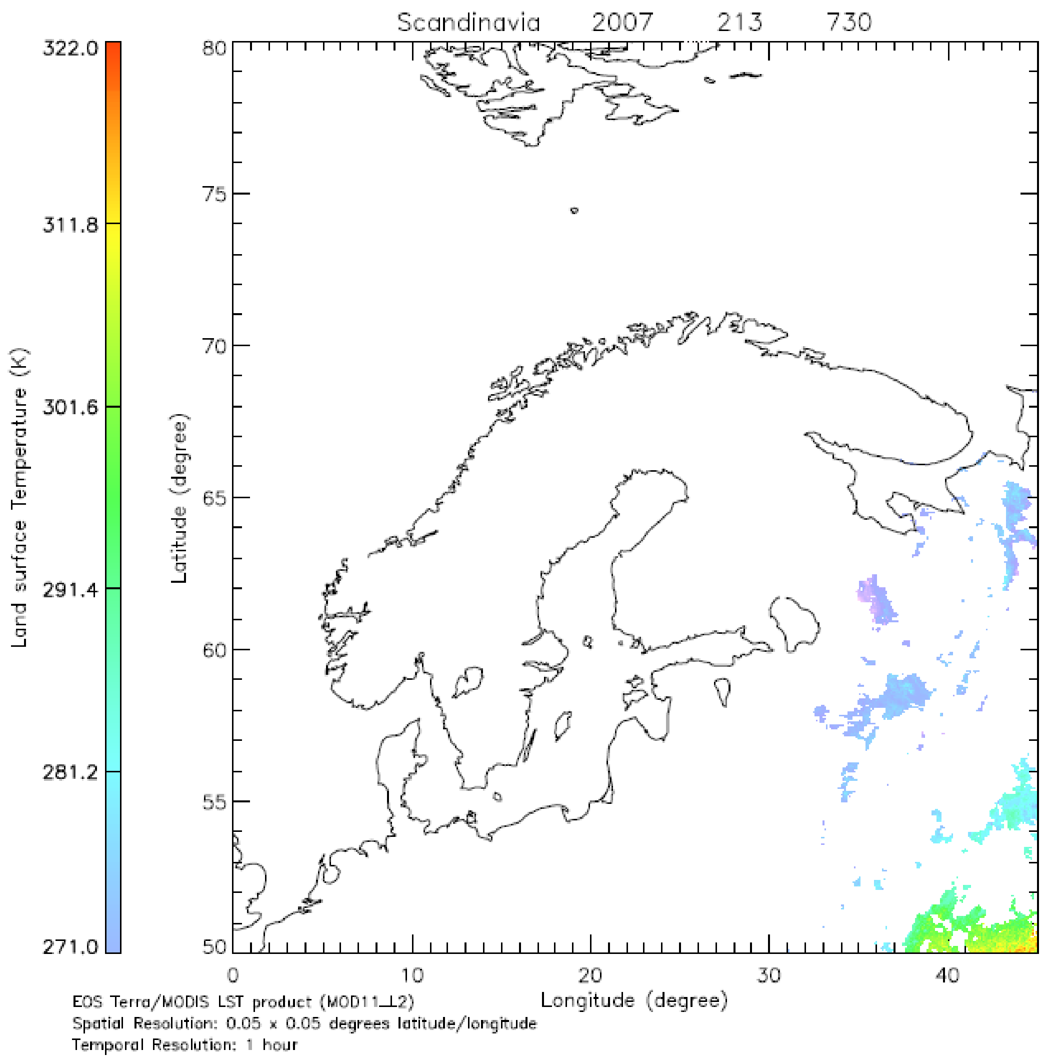
- Large-scale land data assimilation is severely limited by a lack of observations
- Observation and model *errors are not known* – educated guesses must be used.
- We need to pay attention to the *consequences of assimilation*, not just the optimum assimilation technique. i.e. does the model do silly things as a result of assimilation, as in snow assimilation example.
- Land model physics can be biased, leading to incorrect fluxes, given correct states.
- Most land observations are only available at the surface, meaning that biased differences in surface observations and predictions can be improperly propagated to depth.
- Assimilation does not always make everything in the model better. In the case of skin temperature assimilation into an uncoupled model, biased air temperatures caused unreasonable near surface gradients to occur using assimilation that lead to questionable surface fluxes.

CREW

Paul R. Houser, 14 March 2007, Page 34

NILU
Norsk institutt for luftforsk
Norwegian Institute for Air

Example: TERRA-MODIS LST: 1 August 2007, 7.30 am UTC



Focus on Scandinavia

Equations

Sakov and Oke: MWR, 2008

- An Ensemble Square Root filter (ESRF) using a symmetric Ensemble Transform Matrix (ETM): classical KF - ensemble mean

$$\mathbf{x}^{f} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_{i}^{f}$$

Forecast ensemble & mean

$$\mathbf{A}^{f} = \left[\mathbf{A}_{1}^{f},\ldots,\mathbf{A}_{N}^{f}\right] = \left[\mathbf{X}_{1}^{f} - \mathbf{x}^{f},\ldots,\mathbf{X}_{N}^{f} - \mathbf{x}^{f}\right]$$

Forecast anomalies

$$\mathbf{P}^{f} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{X}_{i}^{f} - \mathbf{x}^{f})(\mathbf{X}_{i}^{f} - \mathbf{x}^{f})^{T} = \frac{1}{N-1}\mathbf{A}^{f}\mathbf{A}^{fT}$$

Forecast error

Norsk institutt for luftforskning
Norwegian Institute for Air Research

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

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

$$\mathbf{P}^{a} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^{f}$$

Analysis step in KF eqns

for ensemble mean

$$\mathbf{A}^{a} = \mathbf{A}^{f}\mathbf{T}$$

Update: analysis anomalies

$$\mathbf{T} = \left[\mathbf{I} + \frac{1}{N-1}\left(\mathbf{H}\mathbf{A}^{f}\right)^{T}\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{A}^{f}\right)\right]^{-1/2}$$

Transformation matrix

(several choices)

$$\mathbf{I} + \frac{1}{N-1}\left(\mathbf{H}\mathbf{A}^{f}\right)^{T}\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{A}^{f}\right) = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{T}$$

$$\mathbf{T} = \mathbf{V}\boldsymbol{\Lambda}^{-1/2}\mathbf{V}^{T}$$

Singular value decomposition with V orthonormal and Λ diagonal with eigenvalues

$$\mathbf{X}^a = \left[\mathbf{X}_1^a, \ldots, \mathbf{X}_N^a\right] = \left[\mathbf{A}_1^a + \mathbf{x}^a, \ldots, \mathbf{A}_N^a + \mathbf{x}^a\right]$$ Analysis ensemble

$$\sum_{i=1}^{N} \mathbf{A}_i^a = \mathbf{0}$$ Mean preserving T $$\rightarrow \mathbf{x}^a = \frac{1}{N}\sum_{i=1}^{N} \mathbf{X}_i^a$$ Analysis ensemble mean

$$\mathbf{P}^a = \frac{1}{N-1}\sum_{i=1}^{N} (\mathbf{X}_i^a - \mathbf{x}^a)(\mathbf{X}_i^a - \mathbf{x}^a)^T = \frac{1}{N-1}\mathbf{A}^a \mathbf{A}^{aT}$$ Analysis error

Kalman filter equations exactly solved for the ensemble mean
Note: No perturbations of observations

Importance of mean-preserving feature: improved performance

-> prevents build-up of ensemble outliers

Equations

Sakov and Oke: Tellus, 2008

- A Deterministic Ensemble Kalman Filter (DEnKF) using a linear approximation to the Ensemble Square Root Filter (ESRF) update matrix. Better for avoiding ensemble collapse. Use of localization schemes

$$\mathbf{x}^{f} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_{i}^{f}$$

Forecast ensemble & mean

$$\mathbf{A}^{f} = \left[\mathbf{A}_{1}^{f},\ldots,\mathbf{A}_{N}^{f}\right] = \left[\mathbf{X}_{1}^{f} - \mathbf{x}^{f},\ldots,\mathbf{X}_{N}^{f} - \mathbf{x}^{f}\right]$$

Forecast anomalies

$$\mathbf{P}^{f} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{X}_{i}^{f} - \mathbf{x}^{f})(\mathbf{X}_{i}^{f} - \mathbf{x}^{f})^{T} = \frac{1}{N-1}\mathbf{A}^{f}\mathbf{A}^{fT}$$

Forecast error

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

KF analysis

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

KF gain

$$\mathbf{A}^a = \mathbf{A}^f - \frac{1}{2}\mathbf{K}\mathbf{H}\mathbf{A}^f = \mathbf{A}^f\mathbf{T}$$

Update: analysis anomalies

KH <<1 ("small")

$$\mathbf{T} = \left[\mathbf{I} - \frac{1}{2}\frac{\left(\mathbf{H}\mathbf{A}^f\right)^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{H}\mathbf{A}^f\right)}{N - 1}\right]$$

Transformation matrix

(several choices)

$$\mathbf{M} = \mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$$

$$\mathbf{M}^{-1} = (\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^T$$

Singular value decomposition with $\mathbf{V}$ orthonormal and Λ diagonal with eigenvalues

$$\mathbf{X}^a = \left[\mathbf{X}_1^a, \ldots, \mathbf{X}_N^a\right] = \left[\mathbf{A}_1^a + \mathbf{x}^a, \ldots, \mathbf{A}_N^a + \mathbf{x}^a\right]$$ Analysis ensemble

$$\sum_{i=1}^{N} \mathbf{A}_i^a = \mathbf{0} \qquad \mathbf{x}^a = \frac{1}{N}\sum_{i=1}^{N} \mathbf{X}_i^a$$ Analysis mean

$$\mathbf{P}^a = \frac{1}{N-1}\sum_{i=1}^{N} (\mathbf{X}_i^a - \mathbf{x}^a)(\mathbf{X}_i^a - \mathbf{x}^a)^T = \frac{1}{N-1}\mathbf{A}^a \mathbf{A}^{aT}$$ Analysis errors

See eqn to update forecast anomalies

$$\mathbf{P}^{a} = (\mathbf{I} - \mathbf{KH})\mathbf{P}^{f} + \frac{1}{4}\mathbf{KHP}^{f}\mathbf{H}^{T}\mathbf{K}^{T}$$

Increased ensemble spread

$$\mathbf{T} = \left[\mathbf{I} - \frac{1}{2}\frac{\left(\mathbf{HA}^{f}\right)^{T}\mathbf{M}^{-1}\left(\mathbf{HA}^{f}\right)}{N-1}\right] \approx \left[\mathbf{I} - \frac{\left(\mathbf{HA}^{f}\right)^{T}\mathbf{M}^{-1}\left(\mathbf{HA}^{f}\right)}{N-1}\right]^{1/2}$$

First two terms in Taylor approximation of this

N.B.

$$\mathbf{M} = \mathbf{HP}^{f}\mathbf{H}^{T} + \mathbf{R} = \mathbf{V\Lambda V}^{T}$$

$$\mathbf{M}^{-1} = (\mathbf{HP}^{f}\mathbf{H}^{T} + \mathbf{R})^{-1} = \mathbf{V\Lambda}^{-1}\mathbf{V}^{T}$$

Singular value decomposition with V orthonormal and Λ diagonal with eigenvalues

Advantages of EnKF vs EKF (1)

- Some major problems associated with using the EKF in connection with (larger) nonlinear models:
  - Inaccuracy in the evolution of the model error covariance matrix and huge computational requirements associated with the storage and forward integration of this matrix
  - Use of the central forecast as the estimate of the state. For non-linear dynamics the central forecast is not equal to the mean or expected value

- The EnKF was designed to resolve the points above. It has gained in popularity due to its simple conceptual framework and relative ease of implementation
  - No derivation of a tangent linear operator (Cf 4D-Var)
  - Model error covariance implicitly defined through maintaining a set of model states in the form of an ensemble
  - The mean of the ensemble representing the estimated state

Advantages of EnKF vs EKF (2)

See later

- In EnKFs (and particle filters) each ensemble member is run forward in time through the model
- Uncertainty (or spread in the ensemble) is introduced by stochastic model dynamics (stochastic physics) when integrating each ensemble member forward in time
- In the EKF uncertainty in the estimated state is introduced in the update of the B-matrix (background error covariance) and in the added Q-matrix (Model error)
- However, both are optimal and correct strictly speaking only when the underlying PDFs (prior and posterior to the observations) are Gaussian

Particle filter versions at NILU

- Importance Sampling and Resampling Particle Filter (SIR)
- Auxiliary SIR Particle Filter (ASIR)
- Regularized Particle Filter (RPF)
  - Ristic et al.: "*Beyond the Kalman Filter. Particle filters for tracking applications*", Artech House, Boston, 2004

These are included in the same Fortran module ensemble_m.f90 as the two ensemble Kalman filters (and copied into the enkfassim.f90 code).

Advantages of particle filters

- A particle filter can handle PDFs which can be quite different from Gaussian, e.g. skewed, heavy-tailed, bi-modal, multi-modal etc.

- However, in order to take advantage of this often a *very large* ensemble is needed ; also need resampling to avoid skewed particle distribution

- But another nice feature of particle filters is that they do not alter model physics, only reassigns the probabilities of model states

NOTE:

- In all ensemble based methods, EnKFs, particle filters, the ensemble should capture the true state well, ideally through the mean, and should also describe well the uncertainty of the true state via the spread of the ensemble

SUMMARY:

PF/EnKF represent 2 main groups of ensemble DA methods. Each of them has advantages & limitations.

PF can handle arbitrary probability distributions & non-linearity, but may require excessive resources for high-dimensional problems.

EnKF is more suitable for high-dimensional problems, but should not be used in strongly non-linear/non-Gaussian systems.

Complementarity of EnKF & PF makes a hybrid version highly attractive for non-linear/non-Gaussian systems, e.g., land surface.

Collaboration between NERSC/Met.no/NILU

NILU Norsk insti Norwegian

SURFEX EnKF: OPTIONS.nam

```
&NAM_IO_ENKFASSIM
LPRT =  F,
LPRF =  F,
LSIM =  F
/
```

LPRT associated with perturbation (random draw) of each control variable in order to create new ensemble members.

LPRF associated with perturbation of forcing data in order to create new ensemble members. Currently this option is not used.

LSIM is used to read and write model variables and simulated observations as in the SURFEX-EKF version.

```
&NAM_VAR
IVAR          = 1,
NVAR          = 1,
XVAR_M(1)  = 'WG2',
XVAR_M(2)  = 'WG1',
XVAR_M(3)  = 'TG2',
XVAR_M(4)  = 'TG1',
...
TPRT_M(1)  = 0.1,
TPRT_M(2)  = 0.1,
TPRT_M(3)  = 0.1,
TPRT_M(4)  = 0.1,
INCV(1)       = 1,
INCV(2)       = 1,
INCV(3)       = 0,
INCV(4)       = 0
/
```

Currently the EnKF runs with the two-layer version of the ISBA scheme. It means that the control variables can be the four main prognostic variables of this scheme:

The surface temperature T_s (TG1), the mean surface temperature T_2 (TG2), the superficial volumetric water content w_g (WG1), and the mean volumetric water content in the root-zone w_2 (WG2).

TPRT is the standard deviation of Gaussian random variables for perturbation of the logarithm of each control variable (mean values assumed to be zero).

```
/
&NAM_OBS
NOBSTYPE    = 3,
YERROBS(1) = 1.0,
YERROBS(2) = 0.1,
YERROBS(3) = 0.4,
INCO(1)       = 1,
INCO(2)       = 1,
INCO(3)       = 1
/
```

Regarding the observations, three observation types are considered:

Screen level temperature, relative humidity and superficial soil moisture content.

Like for the control variables, the elements of the array INCO control which type of observation one wants to assimilate.

```
&NAM_ENKF
IENS      = 1,
NENS     = 1,
ENKFM = 1
/
```

IENS is currently not used. It may be used in the future to operate on a single ensemble member

NENS denotes the number of ensemble members

ENKFM denotes the type of ensemble method:

1: Ensemble square-root Kalman filter: ENSRKF
2: Deterministic ensemble Kalman filter: DENSKF
3: Importance sampling and resampling particle filter: SIR
4: Regularized particle filter: RPF

All ensemble data assimilation methods are based on BLAS, Sparse BLAS and LAPACK subroutine libraries. Generation of pseudo random numbers based on subroutines using the Mersenne Twister RNG.