

The NILU SURFEX-EnKF land data assimilation system

W.A. Lahoz, S.-E. Walker and D. Dammann



Abstract

In this report we describe the development of the NILU (Norwegian Institute for Air Research) SURFEX-EnKF (SURFEX-Ensemble Kalman filter) land data assimilation system. The system is being developed on a NILU computing platform in collaboration with the Norwegian Meteorological Institute (Met.no) and Météo-France. It is based on the SURFEX land surface model from Météo-France, and uses various variants of the Ensemble Kalman filter and Particle filter data assimilation method, the main focus being on the former. It also uses the Extended Kalman filter method currently in place at Météo-France. The system is currently used to assimilate land surface temperature and soil moisture, and will be extended to snow variables at a later stage. The observations assimilated are from in situ platforms in the first instance; at a later stage, observations from satellite platforms, including soil moisture from the ESA SMOS (Soil Moisture and Ocean Salinity) mission, launched in November 2009, will be assimilated. The main objective for developing this land data assimilation system is to evaluate observations and models; as part of this exercise, analyses of land surface variables will be produced. Results from the effort at NILU will be of benefit to Numerical Weather Prediction agencies, space agencies and the broad scientific community.

Contents

	Page
Abstract	1
1 Introduction	5
2 Land data assimilation system	6
2.1 Data assimilation scheme	7
2.2 Land surface scheme	14
2.3 Observations.....	17
2.4 Observation operator	19
2.5 Error characteristics.....	19
2.6 Evaluation and data assimilation.....	20
3 Results	22
3.1 Surface temperature (TG1 control variable)	25
3.2 Mean surface temperature (TG2 control variable).....	27
3.3 Superficial volumetric water content (WG1 control variable).....	30
3.4 Mean volumetric water content (WG2 control variable)	32
3.5 TG1 increments	35
3.6 TG2 increments	36
3.7 WG1 increments.....	38
3.8 WG2 increments.....	39
3.9 Discussion	41
4 Plans for future developments	44
4.1 Immediate future	44
4.2 Long-term future	45
4.3 Collaboration.....	45
5 Conclusions	46
6 Acknowledgments.....	46
7 References	47
Appendix A Practical documentation of the land surface analysis scheme based on an Ensemble Kalman filter within SURFEX	53
Appendix B User's Guide for compiling and running SURFEX-EKF and SURFEX-EnKF.....	65
Appendix C EKF results from Météo-France	75

The NILU SURFEX-EnKF land data assimilation system

1 Introduction

Variables such as land surface temperature (LST) and soil moisture are important for understanding interactions between the land and the atmosphere, involving energy, carbon and water cycles (Houser, 2003; Houser et al., 2010). This, in turn, benefits Numerical Weather Prediction (NWP) and climate prediction by allowing a better use of Earth Observation data; better simulation of land/atmosphere processes; and improved initial states for prediction at various temporal scales. Monitoring of the land surface is also becoming increasingly important for addressing climate change issues.

Land surface temperature and wetness conditions affect and are affected by numerous climatological, meteorological, ecological and geophysical phenomena. Therefore, accurate, high resolution estimates of terrestrial water and energy storages are valuable for predicting climate change, weather, biological and agricultural productivity, and flooding, and for performing a wide range of studies in the broader biogeosciences. In particular, terrestrial stores of energy and water modulate fluxes between the land and the atmosphere, and exhibit persistence on diurnal, seasonal and interannual time scales. Furthermore, because soil moisture, temperature and snow are integrated states, errors in land surface forcing and parametrization accumulate in the representations of these variables in operational NWP models, which lead to incorrect surface water and energy partitioning. Therefore, accurate initialization of water and energy state variables in these models is crucial.

Data assimilation (DA) is a technique involving the synthesis of a range of observations from many sources, and with a variety of errors, into a numerical model of an evolving system, e.g., the atmosphere or the land (Kalnay, 2003). It brings together information from observations and the model in an objective way. The key element of the DA method is to confront models with observations. The use of the DA method is already established in NWP; Simmons and Hollingsworth (2002) discuss how DA has helped improve the European Centre for Medium-Range Weather Forecasts (ECMWF) short-term forecasts over the last twenty years.

Assimilation of land surface observations is at an earlier stage than, e.g., assimilation of atmospheric observations. However, during the past decade, land DA has been a very active field of research. Assimilated observations include satellite retrievals of land surface temperature (LST), soil moisture, snow water equivalent (SWE) and snow cover area (e.g. van den Hurk et al., 2002; Andreadis and Lettenmaier, 2005; Slater and Clark, 2006; Bosilovich et al., 2007; Dong et al., 2007; Drusch, 2007; Reichle et al., 2008; Houser et al., 2010).

Land DA has been mainly used to, first, test the methodology and, second, produce land surface analyses to study specific phenomena (Houser et al., 2010). A further use of land DA is testing model and observational information. For example, many innovative land surface observations are becoming available that may provide additional information necessary to refine and constrain the physical parametrizations and initialization of land surface states critical for seasonal-to-interannual prediction. These constraints can be imposed in three ways: (i) by forcing the land surface primarily by observations (such as precipitation and radiation) the often severe atmospheric NWP land surface forcing biases can be avoided; (ii) by employing innovative land surface DA techniques, observations of land surface storages (such as soil temperature and moisture) can be used to constrain unrealistic simulated storages; and (iii) the land surface physical parametrizations themselves can be improved through the use of observed parameters, and through the DA process where model states are constantly being evaluated against observations.

The main objective of the land DA effort at NILU is to make use of ideas successfully applied to NWP, and now starting to be applied to chemistry models and climate models (see, e.g., Lahoz et al., 2007a; Palmer et al., 2008; Trenberth, 2008), to evaluate model and observational information for the land surface. As part of this, the land DA system will be used to test various state-of-the-art DA schemes and produce analyses of land variables.

In Section 2 we discuss the elements of the NILU land DA system; in Section 3 we present preliminary results from initial tests of the system; in Section 4 we discuss plans for further developments; Section 5 provides conclusions. Appendix A includes the SURFEX-EnKF (SURFEX-Ensemble Kalman filter) documentation; Appendix B includes the SURFEX EnKF user guide; Appendix C contains results from Météo-France against which we validate the NILU land DA system.

2 Land data assimilation system

The elements of the NILU SURFEX-EnKF land DA system (Lahoz, 2008) are: (i) DA scheme (mainly variants of the EnKF, but also variants of the Particle filter, and the Extended Kalman filter, EKF); (ii) land surface scheme, or model – this is the SURFEX model developed at Météo-France (Le Moigne, 2009); (iii) observations; (iv) the observation operator; and (v) error characteristics for the model and the observations. Figure 1 provides a schematic of the land DA set up, illustrated with the Ensemble Kalman filter.

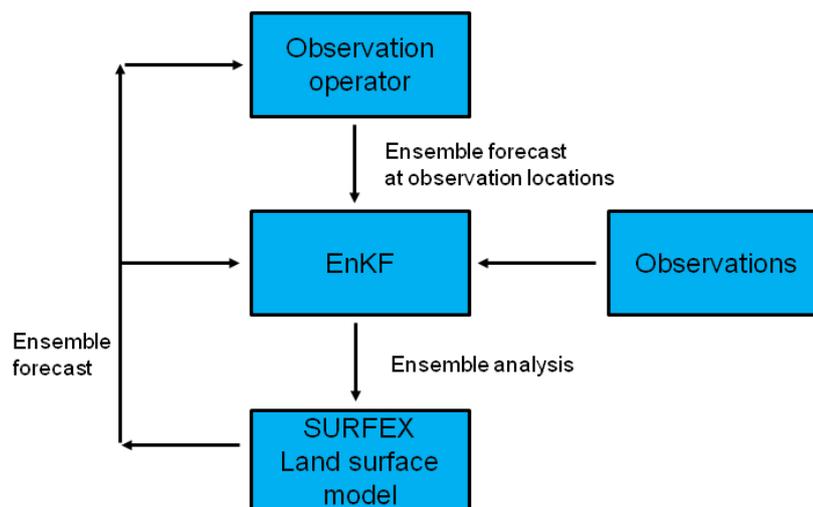


Figure 1: Schematic of the NILU SURFEX-EnKF land DA system methodology. Observations and the model have errors that will be characterized (see Section 2.5).

2.1 Data assimilation scheme

Most DA algorithms, such as the Kalman filter and variational methods are built on statistical linear estimation. Bouttier and Courtier (1999) provide details of these algorithms.

Statistical linear estimation achieves Bayesian estimation when the system is linear and the errors are Gaussian. In particular, statistical linear estimation provides a way of estimating the Best Linear Unbiased Estimate, BLUE (Talagrand, 2003a, 2010a). Independently of the notion of statistical estimation, there exist two broad classes of numerical algorithms for data assimilation: variational and sequential. In the context of statistical linear estimation, these algorithms take respectively the form of the 4-D variational method (4D-Var), or the Kalman filter. (If the time dimension is omitted, the 4D-Var method becomes the 3-d variational method, 3D-Var.) These are two different algorithms for determining the BLUE, and they are equivalent under the only condition of linearity.

Ensemble assimilation is a form of Monte Carlo approximation which attempts to estimate probability distribution functions (PDFs) from the spread of the ensemble. It is not based on statistical linear estimation. In present applications (e.g. the Ensemble Kalman filter; Evensen, 2003), the size of the analysed ensembles typically lies between a few tens to a few hundreds of model states.

We discuss and contrast below the assimilation schemes used in the NILU SURFEX-EnKF land DA system, viz., the Extended Kalman filter (EKF), and variants of the Ensemble Kalman filter (EnKF) and Particle filter (PF). As the focus of this land DA system is the development of EnKF and PF DA schemes, we only provide a brief description of the EKF. For a discussion of DA methods as applied to the land surface, see Houser et al. (2010).

Sequential methods

To discuss the EKF, we first introduce sequential methods. In the Kalman filter (KF), a recursive sequential algorithm is applied to evolve a forecast, \mathbf{x}^f , and an analysis, \mathbf{x}^a , as well as their respective error covariance matrices, \mathbf{P}^f and \mathbf{P}^a . The KF equations are (subscripts denote the time step; \mathbf{M} and \mathbf{H} are linear model and observation operators, respectively):

$$\mathbf{x}_n^f = \mathbf{M}_{n-1} \mathbf{x}_{n-1}^a; \quad (1a)$$

$$\mathbf{P}_n^f = \mathbf{M}_{n-1} \mathbf{P}_{n-1}^a \mathbf{M}_{n-1}^T + \mathbf{Q}_{n-1}; \quad (1b)$$

$$\mathbf{x}_n^a = \mathbf{x}_n^f + \mathbf{K}_n [\mathbf{y}_n - \mathbf{H}_n \mathbf{x}_n^f]; \quad (1c)$$

$$\mathbf{K}_n = \mathbf{P}_n^f \mathbf{H}_n^T [\mathbf{R}_n + \mathbf{H}_n \mathbf{P}_n^f \mathbf{H}_n^T]^{-1}; \quad (1d)$$

$$\mathbf{P}_n^a = [\mathbf{I} - \mathbf{K}_n \mathbf{H}_n] \mathbf{P}_n^f. \quad (1e)$$

Equation (1a) represents the forecast of the model fields from time-step $n-1$ to n , while Eq. (1b) calculates the forecast error covariance from the analysis error covariance \mathbf{P}^a and the model error covariance \mathbf{Q} . Equations (1c) and (1e) are the analysis steps, using the Kalman gain defined in Eq. (1d). \mathbf{Q} and \mathbf{P}^a are assumed to be uncorrelated (Bouttier and Courtier, 1999). For optimality, all errors must be uncorrelated in time.

The KF can be generalized to non-linear H and M operators, although in this case neither the optimality of the analysis nor the equivalence with 4D-Var holds. The resulting equations are known as the Extended Kalman filter, EKF (Bouttier and Courtier, 1999).

The EKF is used for land DA at Météo-France (Mahfouf et al., 2009; see also Draper et al., 2009) and there are plans to use it for land DA at ECMWF (Drusch et al., 2009). The land DA work at NILU will use the EKF scheme described in Mahfouf et al. (2009). One aim of this work at NILU will be to compare the EKF formulation due to Mahfouf et al. with variants of the EnKF and the PF, described below. Section 3 provides results (analyses, increments) from a first comparison between the EKF scheme described in Mahfouf et al. and two EnKF variants. Prior to this comparison, we ascertain that the land DA set-up at NILU involving the EKF gives the same results as that at Météo-France.

Ensemble methods

1. Ensemble Kalman Filter. The EnKF represents a state space formulation of the KF that uses an ensemble of model states to store and propagate the system state and the state error covariance. The ensemble representation is particularly beneficial for large-scale systems, when explicit storage and manipulation of the covariance matrix is impossible or not feasible. It also makes possible an efficient and numerically robust propagation of the covariance with a non-linear model by integrating the ensemble of model states. Another attractive feature of EnKF methods is their conceptual simplicity. Compared to variational methods, the EnKF delivers similar performance (e.g. Kalnay et al., 2007), but requires no tangent linear or adjoint models.

These qualities make the EnKF a highly attractive option as a basis for building operational forecast systems, as well as an increasingly popular alternative to variational methods. As a result, there is a strong growth of the number of publications on the EnKF in recent years, with a doubling time of about two years from 2000 (8 publications in ISI Web of Knowledge) to 2008 (153 publications). These studies investigate both fundamental and applied aspects of EnKF, such as non-linearity; localization; non-Gaussianity; forecasting and reanalysis applications; hybrid systems; comparison with 3D-Var and 4D-Var.

In the traditional EnKF (Evensen, 1994; Burgers et al., 1998) the state error covariance matrix is represented by the ensemble statistically, using a covariance estimator for a sample of random states. In Ensemble square root filters (ESRFs), e.g. Tippett et al. (2003), the ensemble of anomalies is regarded as a factorization of the state error covariance. Other EnKF schemes include the Maximum likelihood ensemble filter (MLEF; Zupanski et al., 2008), and the deterministic EnKF (e.g. Sakov and Oke, 2008b). The ensemble Kalman smoother (Evensen and van Leeuwen, 2000) provides a simple and efficient framework for assimilating asynchronous observations, making it equivalent to traditional (“strong constraint”) 4D-Var (Sakov et al., 2010).

Two new state-of-the-art versions of the EnKF recently proposed by Sakov and Oke have been incorporated into the land DA system developed at NILU: (i) an Ensemble square root filter (ESRF), Sakov and Oke (2008a); and (ii) a Deterministic ensemble Kalman filter (DEnKF), Sakov and Oke (2008b).

The ESRF of Sakov and Oke uses a product of a symmetric mean-preserving transform matrix with an optional mean-preserving orthonormal random rotation. This ensures that the KF equations are accurately solved for the ensemble mean and the variance-covariance matrix. The mean-preserving random rotations also prevent the build-up of ensemble outliers in ESRF-based DA systems; in particular for non-linear models (this is relevant for land DA). No perturbations of observations are necessary in this version of the filter. Sakov and Oke (2008a) also show that the new ESRF method is superior to other non-mean preserving ESRFs and the more traditional EnKF, using perturbed observations.

The DEnKF of Sakov and Oke is a simple modification of the traditional EnKF; it uses a linear approximation to the ESRF update matrix. This linear approximation has the property that it automatically increases the spread of the ensemble in cases where the observation-based analysis correction of the ensemble is large; conversely, the increase in spread is much smaller when the analysis correction is small. This prevents the collapse of the ensemble. As for the ESRF filter of Sakov and Oke, no perturbation of the observations is necessary. In Sakov and Oke (2008b) it is shown that this DEnKF performs almost as well as the ESRF filter and significantly better than the traditional EnKF with perturbed observations. It thus combines the simplicity and flexibility of the traditional EnKF method, with the robustness and superior performance of the ESRF. The DEnKF also readily permits the use of Schur product-based covariance localization schemes.

2. *Applications of the Ensemble Kalman filter.* The EnKF has a strong record of successful application to a wide range of problems. These include oil reservoir

modelling (Evensen et al., 2006); oceanography (Keppenne and Rienecker, 2002); meteorology (Houtekamer and Mitchell, 2001); air quality (Eben et al., 2005); and land modelling (Reichle et al., 2002; Clark et al., 2008). A variant of the EnKF is being developed at NILU, in collaboration with Met.no and Météo-France, for land data assimilation (Lahoz, 2008). Due to the relatively recent development of the EnKF and since operating a large-scale EnKF system requires substantial computational resources there are currently fewer EnKF-based large-scale forecasting applications than those based on variational methods. However, the situation is gradually changing, and several operational EnKF-based systems have been described in the last few years (Houtekamer et al., 2005; Whitaker et al., 2008; Szunyogh et al., 2008). EnKF methods are considered for assimilation of snow data and Leaf Area Index (LAI) in the future ECMWF system (<http://www.ecmwf.int/about/programmatic/2006/topical/6.pdf>).

In oceanography, the TOPAZ system (Evensen, 2007) developed at the Norway-based NERSC (Nansen Environmental and Remote Sensing Centre), remains the only large-scale operational EnKF-based system. TOPAZ runs an ensemble of 100 model states and provides forecasts of ocean currents and sea ice conditions in the North Atlantic and Arctic Oceans. It has been built as a Norwegian contribution to the Global Ocean Data Assimilation Experiment (GODAE), which currently is a part of the European MyOcean project that includes a number of major oceanographic centres. The currents from TOPAZ are used by the ECMWF as input to their operational wave model.

3. Particle filter. Particle filters constitute a broad class of Sequential Monte Carlo (SMC) methods based on statistical simulations of model error evolution using various Monte Carlo random draw techniques (Doucet et al., 2001, 2008; Chen, 2003). For illustration of the PF, we focus on the Sequential importance sampling and re-sampling (SIR) method, which recently has received some attention (van Leeuwen, 2003).

The SIR is an ensemble-based DA method, founded on a general Bayesian statistical framework (Gelman et al., 2004). It makes no assumptions of linearity in the model equations, or that model and observational errors are Gaussian. The SIR is thus, in principle, well suited to deal with systems such as the land surface where model evolution can be highly non-linear, and where model and observational errors can be non-Gaussian.

In the SIR, the ensemble of model states at any given timestep is used as an approximation for the prior and posterior density distribution of the true model state. The ensemble is evolved forward in time using the model operator as in the EnKF, but the analysis step is different. Instead of using the Kalman filter (KF) update equations (Eq. (1) above) to obtain the new analysed model states, importance sampling in the form of re-weighting of the ensemble is done; each forecast ensemble member gets a new weight proportional to the likelihood of the state given the observations. Ensemble members receiving high weights are kept in the ensemble; those receiving low weights are likely to be removed. At the beginning of the next timestep, the new ensemble members are re-sampled and given equal weighting again, and the process is repeated.

The re-sampling step is an essential element of the SIR; without it the method degenerates very quickly, and only a few ensemble members with high weights remain. The analysed model state is calculated as the arithmetic mean of the ensemble members. It can be shown that this model state represents a variance-minimizing estimate of the true model state, even for highly non-linear models. As for other ensemble-based methods, ensemble members provide an estimate of the analysed model state uncertainty.

A major drawback of the EnKF is the underlying assumption that the model states have a Gaussian distribution, making the PF attractive. Furthermore, the PF does not require a specific form for the state distribution, but the distribution of particle weights quickly may become highly skewed (this is known as “particle impoverishment”), and in this case a re-sampling algorithm needs to be applied (see, e.g., Weerts and El Serafy, 2006).

At NILU three PF algorithms are implemented (or will be implemented) as part of the SURFEX-EnKF DA system. These are the SIR PF described above and, in addition, a Regularized PF (RPF) and an Auxiliary PF (ASIR) (Doucet et al., 2001, 2008; Ristic et al, 2004). In the RPF, in addition to a standard re-sampling step, a regularization step is also performed by randomly drawing new particles using a Gaussian kernel. This increases the spread of the particles and is designed to avoid the particle impoverishment mentioned above. In the auxiliary PF, a more advanced re-sampling scheme is employed, which places higher weights on particles more likely to survive when considering states one time step ahead in the dynamic system. A review of the current status of various PF methods is given in Doucet et al. (2008).

4. Applications of the Particle filter. The ease of use and applicability of the PF to a wide range of non-linear and non-Gaussian systems makes the method potentially attractive for land surface data assimilation. The SIR, and other PF methods described above, has been applied with good results in different fields such as control theory, robotics, tracking and perception, where model non-linearity is important. It has also been used with success in several oceanographic studies involving relatively large multi-dimensional non-linear models (van Leeuwen, 2003).

The PF has been applied in hydrology (where assumptions of linearity and Gaussianity do not hold) to estimate model parameters and state variables (Moradkhani et al., 2005). Moradkhani et al. comment that the PF is easily and directly applicable to more complex models (e.g. land surface models, LSMs), but that prospective improvement in its application to hydrometeorological models is an open issue. For a conceptual rainfall-runoff model, Weerts and El Serafy (2006) have compared the SIR to the Residual-resampling (RR) filter and an EnKF that can handle dynamic non-linear/non-Gaussian models. Their results show that the EnKF performs best with a low number of ensemble members, and the RR filter performs best at intermediate and high number of particles, although differences are small. Weerts and El Serafy comment that these methods are feasible and easy to implement in real flood forecasting systems, and recommend further research on assumptions on model and measurement uncertainties.

5. *Intercomparison between the EKF, EnKF and PF.* The EKF is an example of a sequential data assimilation method; the PF and EnKF are examples of ensemble data assimilation methods. Each of them has advantages and limitations:

- The EKF is capable of handling some departure from Gaussian distributions of model errors and non-linearity of the model operator. However, if the model becomes too non-linear or the errors become highly skewed or non-Gaussian, the trajectories computed by the EKF will become inaccurate. The EKF needs to run the physical model (SURFEX) n times per period (time step), where n is the number of control variables. Currently, only at most $n = 4$ control variables are used in the land DA system making use of SURFEX, so the extra computational burden of running the physical model several times is modest;
- The EnKFs are capable of handling a larger degree of non-linearity in the model operator than the traditional EKF, but for its optimality it still relies upon the errors being approximately Gaussian. The EnKFs need to run the physical model (SURFEX) N times per period (time step), where N is the number of ensemble members. In the case one uses far more ensemble members (N) than control variables (n , as described for the EKF immediately above), i.e., $N \gg n$, the EnKF would generally be more computational demanding than the EKF. This could be especially true for current land DA systems based on SURFEX, since calculations have shown that for these systems running of the SURFEX model takes up a significant fraction of the CPU time compared to the assimilation part (approximately 95% of the time). As shown in Section 3 below, tests with the NILU land DA system suggest that the number of ensemble members required for the EnKF is essentially the same as the number of control variables, i.e., $N \sim n$;
- In contrast to the EKF and the EnKFs, PFs are able to handle arbitrary probability distributions and non-linearity. However, this is commonly done by operating on a quite large ensemble of particles (ensemble members). As for the EnKFs, PFs need to run the physical model (SURFEX) N times per period (time step) where N is the number of particles (ensemble members). Thus, a PF may require large resources for high-dimensional problems (Snyder et al., 2008). For such large numbers of particles, we thus may need to use parallel computing facilities. This could be particularly true for the SURFEX-EnKF DA system set up as a 3-D problem; currently it is set up at NILU as a 1-D problem, with independent columns in the soil (see Section 2.2). By 3-D we mean inclusion of horizontal information from neighbouring grid cells in the DA system.

The EKF is implemented (not necessarily operationally) at operational centres such as Météo-France and ECMWF; the EnKF and PF have started to be successfully applied for DA in land applications (see references above). However, there are only a few comparison studies between the EKF, EnKF and PF for these land applications, typically characterized by a combination of features that do not fit well any particular method: strong non-linearity, non-Gaussianity, and high dimensionality. Representative examples are provided below:

- Reichle et al. (2002) compare the EKF and the EnKF for soil moisture assimilation and conclude that the EnKF outperforms the EKF for ten or more ensemble members. This is ascribed to the EnKF's flexibility in representing model errors;
- Zhou et al. (2006) argued that the traditional ("perturbed observations") EnKF (Burgers et al. 1998) is able to provide a good approximation for non-linear and non-Gaussian land surface problems, despite its dependence on Gaussian assumptions. Weerts and El Serafy (2006) found that both the EnKF and PF are feasible and easy to implement in a realistic flood forecasting system.

Comparison between the EKF, EnKF and PF will help to understand the strengths and weaknesses of these DA methods, and address the weaknesses. This information will help improve the methods and will benefit, among others, the NWP community. Progress in operational application of these data assimilation methods requires an understanding of their strengths and weaknesses, with focus on several issues identified as crucial by the land DA community. These include:

- Observed quantities are often non-linearly related to the model variables, Thus, the abilities of the different DA assimilation algorithms to handle non-linear observation operators need to be compared and evaluated;
- Most observations for land DA are concerned with surface or near-surface conditions, while important model variables represent more deep soil conditions. Deep soil moisture is one example. The abilities of different DA methods to solve inverse problems must be compared and evaluated;
- Remote sensing data, from, e.g., satellites, are becoming more and more important for land DA (see, e.g., Drusch et al., 2009). These remote sensing data often have complicated observation error structures, for example biases and spatially correlated errors. The abilities of different DA methods to handle such complicated observation error structures must be compared and evaluated. The interaction between (systematic) model errors and (systematic) observation errors is another issue that must be handled in a proper way;
- The atmospheric forcing for the land DA may come from models or from observations, with precipitation as an important example. Here, again, specification of model error characteristics is crucial.

Overview

Three methods are commonly used for land DA (Houser et al., 2010): variational (3D-/4D-var); sequential (Kalman filter and Extended Kalman filter, EKF); and ensemble (Ensemble Kalman filter, EnKF). The PF is being considered for land DA at, e.g., NILU. Because of the 1-D nature of most land surface processes, 1-D and 3-D versions of a DA scheme are often used (e.g. Reichle and Koster, 2003).

The EKF is capable of handling some departure from Gaussian distributions of model errors and non-linearity of the model operator. However, if the model becomes too non-linear or the errors becomes highly skewed or non-Gaussian the trajectories computed by the EKF will become inaccurate.

The EnKF is attractive as, e.g., it requires no derivation of a tangent linear operator or adjoint equations, and no integrations backward in time, as for 4D-Var (see Evensen, 2003). This can be problematic given non-linearities and on-off processes at the land surface. The EnKF also provides a cost-effective representation of the background error covariance matrix, \mathbf{B} (its analogue in the KF is commonly represented as \mathbf{P}^f – see Eq. (1)). Several issues need to be considered in, for example, developing the EnKF: (i) ensemble size; (ii) ensemble collapse; (iii) correlation model for \mathbf{B} ; and (iv) specification of model errors. However, the major drawback of the EnKF is the underlying assumption that the model states have a Gaussian distribution.

The 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 EnKF and PF are complementary. This complementarity makes a hybrid EnKF/PF version highly attractive for systems that can exhibit non-linear and non-Gaussian features, an example being the land surface. For example, the EnKF could be used as an efficient sampling tool to create an ensemble of particles with optimal characteristics with respect to observations. The PF methodology could then be applied on that ensemble afterwards to resolve non-linearity and non-Gaussianity in the system. This method is getting increased attention among DA experts (P. van Leeuwen and J. de Vries, pers. comm.). Developing such a hybrid system for the land is beyond the scope of current land DA efforts at NILU, but is being considered for the future (see Section 4).

2.2 Land surface scheme

Many land surface schemes (LSMs) have been developed and enhanced since the mid 1990s, with varying features, such as subgrid variability, community-wide input, advanced physical representations, and compatibility with atmospheric models (Houser et al., 2010).

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.

In collaboration with Met.no and Météo-France, NILU is developing a land DA system incorporating the SURFEX land surface model (Giard and Bazile, 2000; Le Moigne, 2009) used in the HIRLAM (High-Resolution Limited Area Model) and ALADIN (Aire Limitée Adaptation dynamique Développement International) NWP consortium. At NILU the SURFEX model will be used in the first instance as a stand-alone model, i.e., uncoupled from the atmospheric and ocean models.

The SURFEX model includes the following elements:

- **Soil and vegetation scheme** (ISBA, Interface Soil-Biosphere-Atmosphere, and ISBA-A-gs): This simulates the exchange of energy and water fluxes between the land surface and the atmosphere;
- **Water surface scheme** (COARE/ECUME – for the sea; FLAKE – for inland water): This simulates various features of the water surface: turbulent fluxes, temperature, salinity, heat budget, and mixed layer depth; and ice and snow cover for inland water;
- **Urban and artificial areas** (Town Energy Balance, TEB, model): The TEB model simulates the exchange of fluxes between a town/urban area and the atmosphere. A town/urban area is represented, e.g., by roofs, roads and facing walls;
- **Surface boundary layer (SBL) scheme**: This accounts for the way the vegetation canopy modifies the interaction between the land and the atmosphere. It incorporates the SBL equations into a surface scheme with implicit coupling to the atmosphere;
- **Chemistry and aerosols**: This takes account of the contribution of dust aerosols, sea salt emission, dry deposition of aerosols and gaseous species, and biogenic VOCs (volatile organic compounds) to: (i) surface fluxes (information from the land surface to the atmosphere), and/or (ii) atmospheric forcing (information from the atmosphere to the land surface);
- **Land use database (ECOCLIMAP)**: ECOCLIMAP is a global database of land surface parameters at 1 km horizontal resolution combining land cover maps with satellite information. It provides a detailed description of surface conditions such as vegetation types, sea/lake, and town;
- **Land surface analysis**: This uses a DA scheme (e.g. EKF at Météo-France) to update the SURFEX model state variables by assimilation of various in situ and satellite observations. The NILU effort extends the Météo-France land surface analysis to include variants of the EnKF and PF.

A brief description of how SURFEX works follows. During a model time step, each surface grid box receives from the atmosphere the following information: upper air temperature, specific humidity, horizontal wind components, pressure, total precipitation, long-wave radiation, short-wave direct and diffuse radiation, and, possibly, concentrations of chemical species and dust.

In return, SURFEX computes averaged fluxes of momentum, sensible and latent heat, and, possibly, chemical species and dust fluxes. These fluxes are then sent back to the atmosphere with the addition of radiative terms like surface temperature, surface direct and diffuse albedo, and surface emissivity.

The above information provides the lower boundary conditions for the radiation and turbulent schemes in an atmospheric model coupled to SURFEX, or forced by SURFEX output. In SURFEX, each grid box is made up of four adjacent surfaces: one for nature, one for urban areas, one for sea or ocean, and one for lake. The coverage of each of these surfaces is known through the global ECOCLIMAP database. The SURFEX fluxes are the average of the fluxes computed over nature,

town, sea/ocean or lake, weighted by their respective fraction. Figure 2 provides a summary of how SURFEX works.

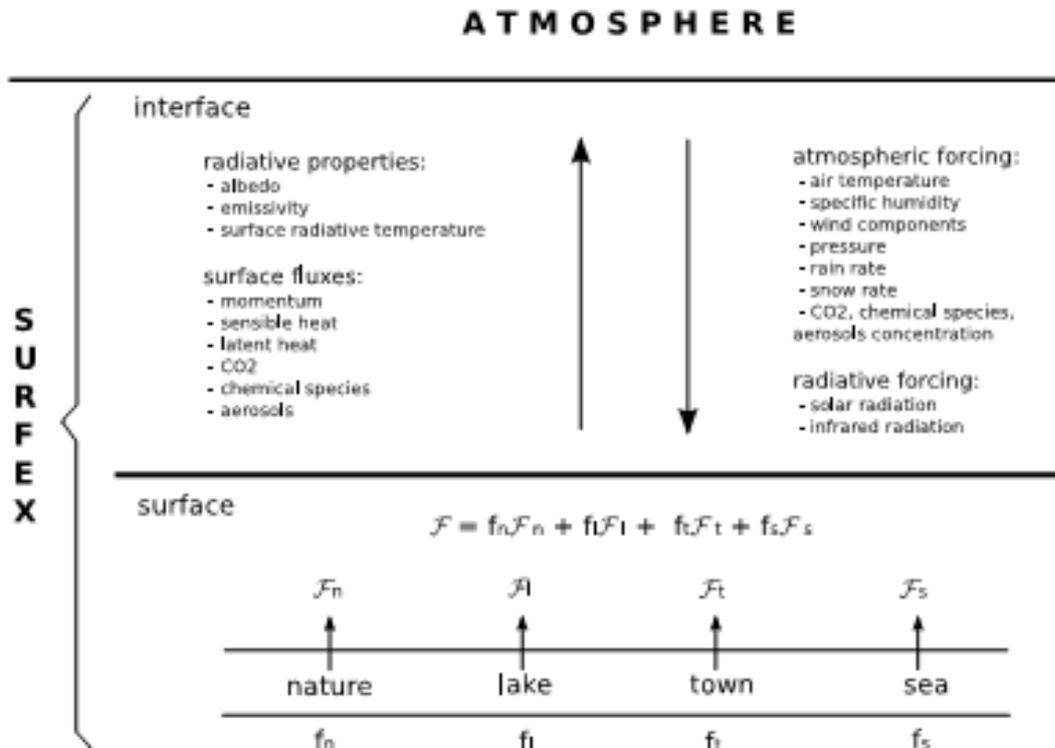


Figure 2: Exchanges between the atmosphere and land surface implemented in SURFEX. See text. (From Le Moigne (2009).)

At NILU we have successfully installed the SURFEX model on a local computing platform. The DA system implemented at NILU, incorporating SURFEX, is currently run as a 1-D problem, with independent soil columns. To set up this system, we have used the architecture of the Météo-France SURFEX-EKF land DA system, with the following changes:

- A new main program `enkfassim.f90` and module files `ensb_m.f90`, `rand_m.f90` and `util_m.f90` have been introduced incorporating the ensemble based data assimilation methods (EnKF and PF). This replaces the `varassim.f90` main program as used by the SURFEX-EKF system;
- Corresponding changes have been done in the file `Makefile.SURFEX.mk` in order to be able to compile the new SURFEX-EnKF system;
- The standard namelist file of SURFEX (`OPTIONS.nam`) has been extended with an extra namelist (`NAM.ENKF`) containing the type of Ensemble Kalman filter or Particle filter method to be used, and the number of ensemble members (N) to be used. For details see Appendix A of this report;
- A new run script (`run_enkf.sh`) incorporating the running of the SURFEX-EnKF system has been coded. A number of enhancements

have been made compared to the existing SURFEX-EKF script `run_ekf.sh`. For example, it is now possible to change a larger number of run parameters in the `OPTIONS.nam` file directly from the `run_enkf.sh` script. This includes the selection of the assimilation method, number of ensemble members, parameters governing the ensemble perturbations, and the observations to be included and their error variances (elements of the observational error covariance matrix, \mathbf{R});

- The documentation for the NILU SURFEX-EnKF land DA system is provided in Appendix A of this report. A manual for running the SURFEX-EnKF land DA system is provided in Appendix B of this report.

The SURFEX-EnKF land DA system has been compiled, run and tested at NILU for various land DA set-ups for 24 and 48 hours (four and eight 6-hour data assimilation cycles) in early July 2006, for which we have input data provided by Météo-France. The region over which the land data assimilation is performed is currently centred over France (see Section 3). These DA set-ups are: (i) no assimilation; (ii) EKF assimilation; (iii) EnKF assimilation, with 2, 3 and 5 ensemble members, and with two different EnKF methods (ESRF and DEnKF); and (iv) PF assimilation. Runs with 9 ensemble members have also been made.

2.3 Observations

Observations are commonly divided into conventional observations (e.g. in situ ground-based measurements such as screen-level temperature and relative humidity) and remotely-sensed satellite observations. Land DA considers both types of observations. Often, satellite land surface data are assimilated and the process validated using in situ measurements. Houser (2003) discusses the assimilation of land surface retrieved quantities and radiances.

The emphasis of land surface DA research is to assimilate remotely-sensed observations of the land surface that previous research suggests will provide memory to the land-atmosphere interaction. Remote observations of interest include: (i) temperature; (ii) soil moisture (surface moisture content, surface saturation, total water storage); (iii) other surface water bodies (lakes, wetlands, large rivers); and (iv) snow (areal extent, snow water equivalent – SWE). The NILU SURFEX-EnKF land DA system is designed initially to assimilate retrievals of land surface temperature and soil moisture (from in situ and/or satellite platforms). At a later stage, snow variables will be assimilated.

Assimilation of observation-based surface fields will continually steer the modelled states towards our understanding of reality. However, only a few variables can be observed directly, such as skin temperature. Other variables, such as subsurface soil temperature, will be affected by the assimilation, but their representation is dependent on imperfect model physical assumptions.

In the first instance, we assimilate at NILU in situ land surface temperature and soil moisture provided by the SYNOP/CANARI (Code d'Analyse Nécessaire à Arpege pour ses Rejets et son Initialisation; Taillefer, 2002) analysis. The data

assimilated in the current NILU set-up are: two-metre screen-level temperature (T_{2m}); and two-metre screen-level relative humidity (RH_{2m}). There is an option of assimilating superficial soil moisture content data from satellites, but this has not been implemented in the current set-up at NILU. The control variables are:

- Surface temperature (a surface quantity), units of K – TG1;
- Mean surface temperature (a volume quantity), units of K – TG2;
- Superficial volumetric water content (a surface quantity), units of m^3/m^3 – WG1;
- Mean volumetric water content of the root zone (a volume quantity), units of m^3/m^3 – WG2.

Results in Section 3 discuss the assimilation of T_{2m} and RH_{2m} observations using the above control variables. At a later stage, we will consider assimilation of satellite data. In particular, in the near future we propose to assimilate land surface temperature from at least one of Envisat-AATSR (Advanced Along Track Scanning Radiometer) and Aqua/Terra-MODIS (MODerate Resolution Imaging Spectroradiometer); and soil moisture from AMSR-E (The Advanced Microwave Scanning Radiometer - Earth Observing System; see, e.g., Draper et al., 2009) and SMOS (Soil Moisture and Ocean Salinity; Kerr et al., 2000). In the more distant future we propose to assimilate snow variables from, e.g., Aqua/Terra-MODIS. Brief details of the types of land surface remote (satellite) observations to be assimilated follow.

1. Land surface temperature. Remote sensing of surface temperature is a relatively mature technology. The land surface emits thermal infrared radiation at an intensity directly related to its emissivity and temperature. The absorption of this radiation by atmospheric constituents is lowest in the 3-5 μm and 8-14 μm wavelength ranges, making them the best atmospheric windows for sensing land surface temperature (LST). Some errors due to atmospheric absorption and improperly specified surface emissivity are possible, and the presence of clouds can obscure the signal. Generally, surface temperature remote sensing can be considered an operational technology, with many spaceborne sensors making regular observations (e.g. MODIS, Landsat Thematic Mapper, Advanced Very High Resolution Radiometer – AVHRR, and Advanced Spaceborne Thermal Emission and Reflection Radiometer – ASTER) - see Lillesand and Kiefer (1994) for details. The evolution of LST is linked to all other land surface processes through physical relationships. These land surface process interconnections can be exploited in a land DA framework to constrain all the predicted land surface states.

2. Soil moisture. Remote sensing of soil moisture content is a developing technology, although the theory and methods are well established (Eley, 1992). Long-wave passive microwave remote sensing is ideal for soil moisture observation, but there are technical challenges involved in correcting for the effects of vegetation and roughness. Houser et al. (2010) provides details of the use of soil moisture data from satellites.

3. Snow. Key snow variables of interest to land surface understanding include area coverage and snow water equivalent (SWE). While the estimation of SWE by

satellites is in many ways still a research activity, snow areal extent can be routinely monitored by many operational platforms. Houser et al. (2010) provides further details.

2.4 Observation operator

The non-linear observation operator H (see Section 2.1) transforms from the model space to the measurement space. It involves a mapping from geophysical inputs in model space (e.g. LST) to simulate an instrument measurement in measurement space (e.g. radiances), taking into account the physics of the measurement and the characteristics of the instrument. The role of the non-linear model operator M , or the forward model (see Section 2.1), depends on the data assimilation approach. In the sequential approach, the model operator maps the analysis forward in time to give a background state for a subsequent assimilation cycle (this applies to the EnKF and its variants if the evolution of the model state is treated as in the KF); in the variational approach, the model operator may be part of the analysis process. The observation operation must also take account of the different resolutions of the model and the measurement. Errors in mapping from the model to the measurement are accounted in the representativeness error (Cohn, 1997).

To represent the observation operator for a general land DA scheme computer code, we have made use of the approach described by Mandel (2007), which involves the use of the observation function, $h(x)$. This avoids explicit construction of the Jacobian matrix of H , which is often difficult.

2.5 Error characteristics

Representation of errors is fundamental to DA. One needs to consider errors in observations, background information, and model (see Eq. (1) for identification of the error covariance matrices mentioned in the following). \mathbf{R} , the observational error covariance matrix, is typically assumed to be diagonal, although this is not always justified. \mathbf{R} includes errors of the measurements themselves, \mathbf{E} , and errors of representativeness, \mathbf{F} ; $\mathbf{R}=\mathbf{E}+\mathbf{F}$. \mathbf{B} is the background error covariance matrix; its off-diagonal elements determine how information is spread spatially from observation locations. Estimating \mathbf{B} is a key part of the DA method. Estimating model error \mathbf{Q} is a research topic.

In the EnKF and also in the PF, the background (or forecast) errors are represented by the spread of the ensemble. This simplifies the computation of \mathbf{B} , and is an advantage of the EnKF; this can also be used to account for the model error \mathbf{Q} . For land DA, the model error \mathbf{Q} (associated with the temporal evolution of the model) is often tuned (e.g. Reichle et al., 2008). In land DA, an approach introduced by Desroziers et al. (2005) is also used to estimate the model and observation error covariances separately. Application of this approach to the heterogeneous land surface is discussed in Reichle et al. (2008).

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 variance, the solution of the DA problem becomes computationally practical. Note that the assumption of Gaussianity is often not justified in land DA applications.

Typically, there are biases between different observations, and between observations and model (see, e.g., Ménard, 2010). 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 that 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.

At NILU we will investigate the representation of observational errors; the development and implementation of bias correction methods; and the application of ideas from possibility theory.

2.6 Evaluation and data assimilation

General

A crucial element of DA is the evaluation of the quality of the observations, the models and the analyses, and the test of several assumptions built into DA algorithms, e.g. Gaussian errors; unbiased observations and models. Several diagnostics have been developed to do this (Talagrand, 2003b, 2010b). Broadly speaking, these consist of self-consistency tests and independent tests.

Self-consistency tests provide useful information for evaluating the quality of the data assimilation ingredients and the assumptions built into assimilation algorithms. Histograms of OmA (observation minus analysis) and OmF (observation minus forecast) differences are computed for a range of spatial and temporal scales to test whether the observations, forecast and analysis fields, and their errors, are consistent with each other. For example, the OmA histogram should be more peaked than that for OmF, as the analyses should be closer to the assimilated observations than the forecast. Furthermore, the OmF histogram should be Gaussian, if both the observation and forecast are assumed to have Gaussian errors. Time averages of the standard deviation of OmA can also be used to test whether the assimilation system is consistent with the concept of the Best Linear Unbiased Estimate, BLUE (Talagrand, 2003a, 2010a). Other tests check whether there are biases between observation and forecast, or between observation and analysis. Tests for Gaussian errors can also include tests of skewness and kurtosis.

Independent tests involve comparison of analyses with data that are independent from the analyses, i.e., data not assimilated to provide the analyses. Independent data can provide information on whether the analyses are realistic and can help attribute biases to observations, forecast and analysis; note that self-consistency

tests cannot be used to perform this attribution. Estimating the bias in the analyses by comparison against independent data is only possible when the error characteristics of the latter are well known.

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.

Evaluation of models

Data assimilation has been very successful at improving NWP models (Simmons and Hollingsworth, 2002). The desirability of applying DA ideas to improve climate models has long been recognized (Puri, 2002), but has not been fully implemented, mainly because climate models are not generally developed in conjunction with an appropriate DA scheme. However, recently, these ideas have been applied to the CAPT (Climate Change Prediction Program, Atmospheric Radiation Measurement Program, CCPP-ARM Parameterization Test-bed) initiative (see Phillips et al., 2004) with the aim to improve parametrizations in climate General Circulation Models. DA also has been used to evaluate ozone chemistry components in DA systems (Geer et al., 2006, 2007; Lahoz et al., 2007a, b). At a recent WCRP (World Climate Research Programme) Climate Summit at ECMWF, the use of observations, including DA ideas, to evaluate climate models was discussed (Trenberth, 2008).

Recently, Palmer et al. (2008) have discussed the concept of “seamless prediction” and its relevance for weather and climate forecasting (see also Rodwell and Palmer, 2007). In particular, Palmer et al. point out there are fundamental physical/dynamical processes common to both weather and seasonal forecasts on the one hand, and climate-change timescales on the other. Because of this, they propose that probabilistic validation of models at timescales where validation data exist (i.e., daily, seasonal and, to some extent, decadal timescales) can be used to calibrate climate-change probabilities at longer timescales. A key point made by Palmer et al. is that the need for calibration reflects a need for model improvement.

Data assimilation not only corrects weaknesses in models, but also identifies model deficiencies such as biases (e.g. between model and observations; between different observations), which as Rood (2005) states is likely the greatest challenge in DA.

Overview

At NILU we will use elements of the methodology described above to evaluate the fidelity of land surface observations, the SURFEX model, and various DA methods. In earlier work, Mahfouf (2007) compared the EKF and the EnKF methods in a land data assimilation context; we extend this work. Finally, to our knowledge, the data assimilation ideas discussed immediately above, have not been applied to evaluate land surface models.

3 Results

In this section we present first results for the four control variables in the NILU SURFEX-EnKF land data assimilation system: surface temperature – TG1; mean surface temperature – TG2; superficial volumetric water content – WG1; mean volumetric water content of the root zone – WG2 (see Section 2.3). Analysed fields are presented in Sections 3.1-3.4; analyses increments are presented in Sections 3.5-3.8; results are discussed in Section 3.9. Figure 3 provides a flowchart of the set-up for the SURFEX-EnKF (see also Appendix A). As part of the evaluation of the NILU SURFEX-EnKF land data assimilation system, in Section 3.9 we first compare the EKF results with those produced by the system at Météo-France (Appendix C; plots provided by J.-F. Mahfouf).

The NILU SURFEX-EnKF land data assimilation system assimilates two-metre temperature and relative humidity screen-level observations (T_{2m} and RH_{2m} , respectively) into the off-line SURFEX v4.8 model; Mahfouf et al. (2009) discusses the set up of the atmospheric forcing of the SURFEX model. The SURFEX model incorporates the two layer version of the ISBA (Interface Soil-Biosphere-Atmosphere; see Chapter 4 of Le Moigne, 2009) scheme. The T_{2m} and RH_{2m} observations come from SYNOP/CANARI analyses at 0000, 0600, 1200 and 1800 UTC. For the period analysed (see below) there are no superficial volumetric water content observations (termed WG1 and applicable to observations from a satellite platform, e.g., ASCAT – Advanced SCATterometer), and hence they are not assimilated. These three observation types (T_{2m} , RH_{2m} and WG1) are currently assimilated into the Météo-France land DA system, SURFEX-EKF (Mahfouf et al., 2009). An extension of this system to assimilate soil moisture data from AMSR-E is discussed in Draper et al. (2009). NILU have plans to extend the work of Draper et al. by performing the assimilation of AMSR-E soil moisture using variants of the EnKF.

Observational errors are 1 K for T_{2m} , 10% for RH_{2m} and 10% of the soil wetness index (SWI) for WG1. The observational background error covariance matrix, \mathbf{R} , is assumed to be diagonal. Model forecast errors in the EKF (error covariance matrix \mathbf{B}) are as in the SURFEX-EKF land DA system at Météo-France (see Section 2.3): 2 K for TG1 and TG2; 10% of the soil wetness index for WG1 and WG2. In the EKF, the model error matrix covariance \mathbf{Q} is specified to be 0.125 times the \mathbf{B} matrix. Only diagonal entries in \mathbf{B} and \mathbf{Q} are prescribed (see, e.g., Draper et al., 2009). The Jacobian calculations in the EKF use relative perturbations of 10^{-4} for the WG1 and WG2 humidity control variables, and 10^{-5} for the TG1 and TG2 temperature control variables.

Flowchart of SURFEX-EnKF

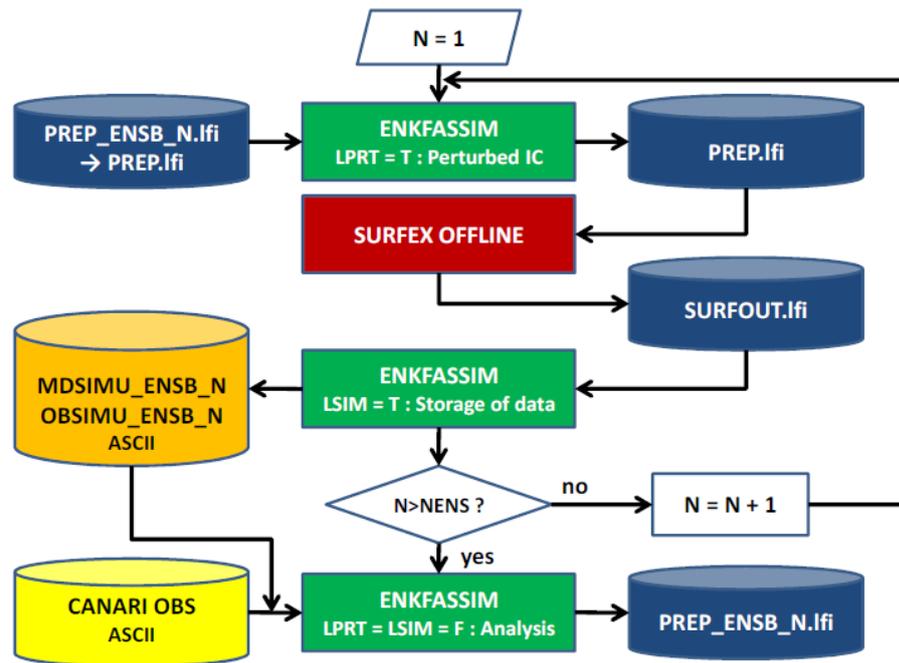


Figure 3: Flowchart of the NILU SURFEX-EnKF land DA assimilation system. This flowchart is similar to the one used for the EKF used at Météo-France (Mahfouf et al., 2009).

The perturbation to create the ensemble members for the EnKF method is currently 1% of the logarithm of the SURFEX model value, treated as a random variable with mean value zero, this value being independent for each grid cell. This is done at the beginning of each assimilation time window to help avoid ensemble collapse. The 1% refers to the standard deviation of a Gaussian distribution about the SURFEX model value. This is done for all control variables, which are perturbed simultaneously. No extra factor is added to the error represented by the ensemble spread to take further account, e.g., of the model error (\mathbf{Q} in the traditional notation, applied to the EKF – see Section 2.5). This means that the errors in the initial conditions and model are assumed to be represented by the ensemble spread. Note that the observations are not perturbed, as this is not required in the variants of the EnKF we use (see Section 2.1).

The way the perturbation is currently performed is likely to account for the presence of random noise in the EnKF analysed fields (see Sections 3.1-3.4) and analyses increments (see Sections 3.5-3.8). To remedy this perceived weakness in the EnKF set-up, other methods of setting up the ensemble members are being developed and tested. For example, we are considering introducing a simple representation of the ensemble spread based on the diagonal (background error variance) of the \mathbf{B} matrix in the EKF set-up; later, we will consider a horizontal localization to account for correlations between physical variables over a finite spatial scale.

The assimilation period for the experiments described below is 24 hours and includes four 6-hour assimilation cycles (i.e., the assimilation time window is 6 hours). It starts on 0000 UTC, 1 July 2006, and ends on 0000 UTC, 2 July 2006.

There are a total of 38177 land observations of T_{2m} and RH_{2m} for this period. The spatial domain is a 289×289 grid over Europe, with horizontal resolution of 9.5 km in the North and East directions, and centred at 46.5° latitude and 2.6° longitude (see map in, e.g., Fig. 4) – at a later stage we will extend the experiments to other regions in Europe. In particular, the region of Norway/Sweden will be a focus of the assimilation of soil moisture from SMOS.

The forcing data and initial data are provided by Météo-France (Mahfouf et al., 2009). For this experiment (24 hours long), the forcing is not affected by the land data assimilation, which over time periods longer than 24 hours may affect the results as the atmospheric forcing may not be consistent with the land variables. At a later stage, we will take steps to develop a fully-coupled land DA system that can avoid this inconsistency. In particular, for long DA experiments (order one month) we will aim to provide atmospheric forcing that includes the effect of the soil variables. This could be accomplished by running in a free mode (i.e., no assimilation) an atmospheric model coupled to a soil model.

Wall-clock timings for 24-hour assimilation for the EKF and EnKF variants are given in Table 1 below. The machine on which the experiments were performed is a Linux Enterprise Server based computer system at NILU (“Ulven”). Note that the timings for the EKF and the EnKF variants for 5 ensemble members are very similar. The efficiency of the Ulven system is being studied, with a view to decreasing the wall-clock timings.

Table 1: Wall-clock timings.

DA scheme	Wall-clock time
EKF	1 hr 41 min
EnKF (square root, 2 ensemble members)	0 hr 53 min
EnKF (square root, 3 ensemble members)	1 hr 03 min
EnKF (square root, 5 ensemble members)	1 hr 44 min
EnKF (deterministic, 2 ensemble members)	0 hr 43 min
EnKF (deterministic, 3 ensemble members)	1 hr 02 min
EnKF (deterministic, 5 ensemble members)	1 hr 43 min

3.1 Surface temperature (TG1 control variable)

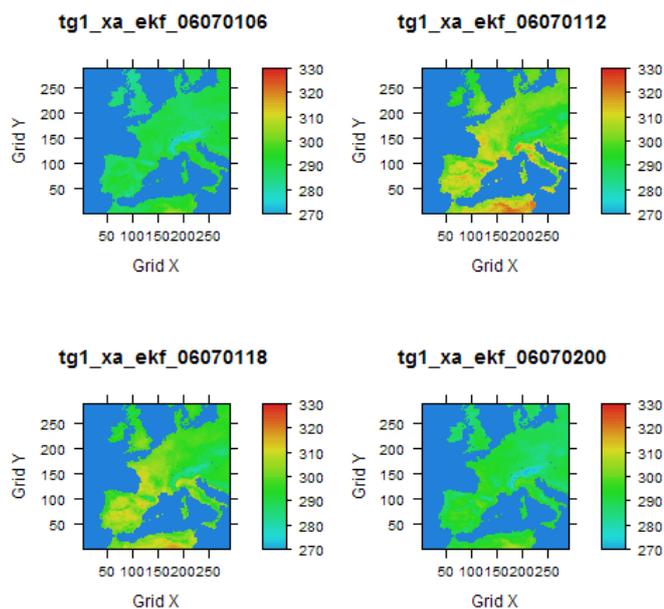


Figure 4: Surface temperature analysed field (K) at 4 times using the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

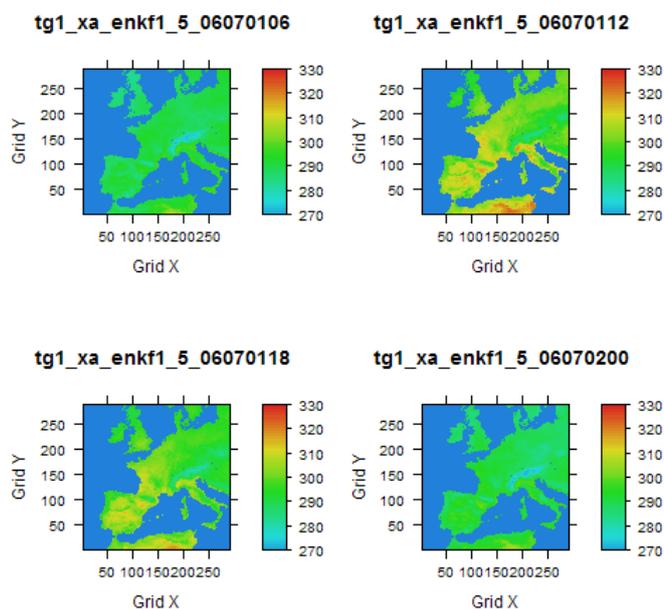


Figure 5: Surface temperature analysed field (K) at 4 times using the square root EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

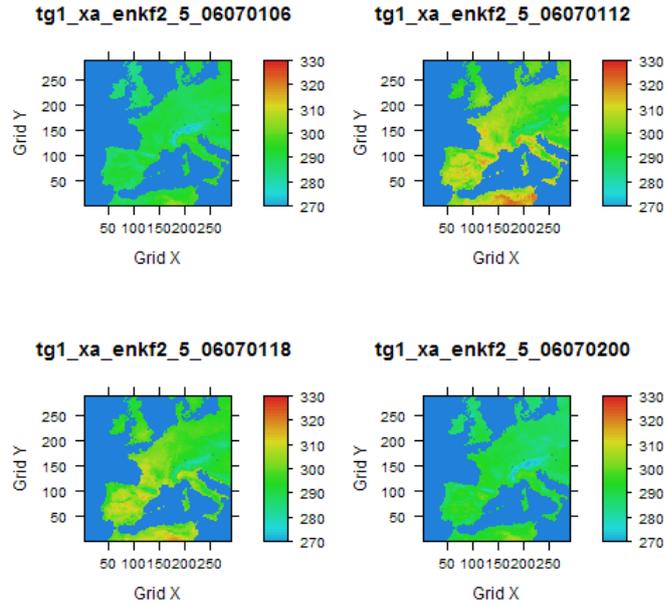


Figure 6: Surface temperature analysed field (K) at 4 times using the deterministic EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

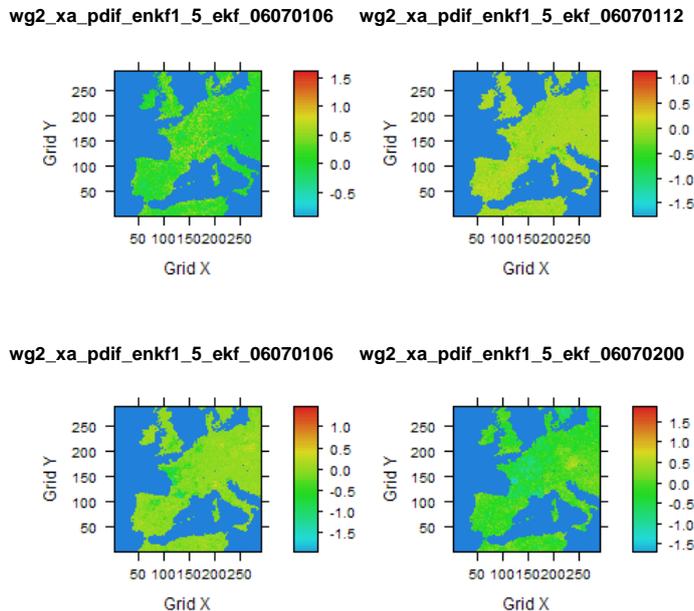


Figure 7: Surface temperature difference field (K) at 4 times, percentage difference between the square root EnKF, mean of 5 ensemble members, and the EKF, as a percentage of the EKF values. Positive values indicate the EnKF has higher values than the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

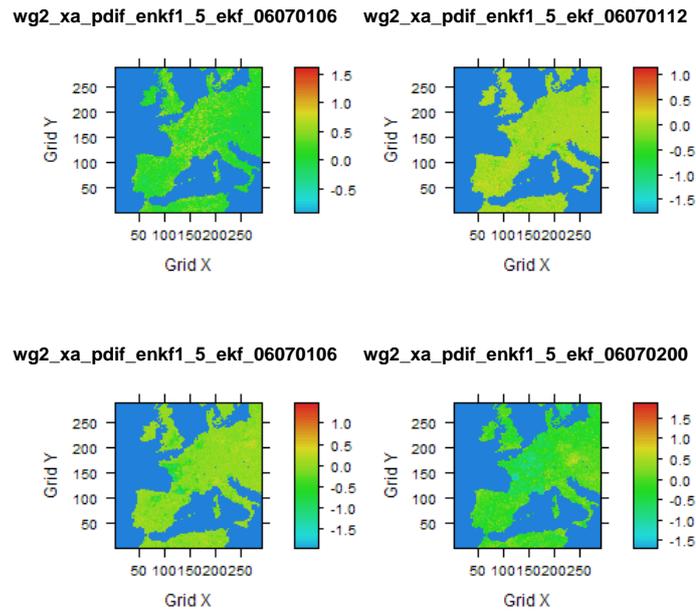


Figure 8: Surface temperature difference field (K) at 4 times, percentage difference between the deterministic EnKF, mean of 5 ensemble members, and the EKF, as a percentage of the EKF values. Positive values indicate the EnKF has higher values than the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

3.2 Mean surface temperature (TG2 control variable)

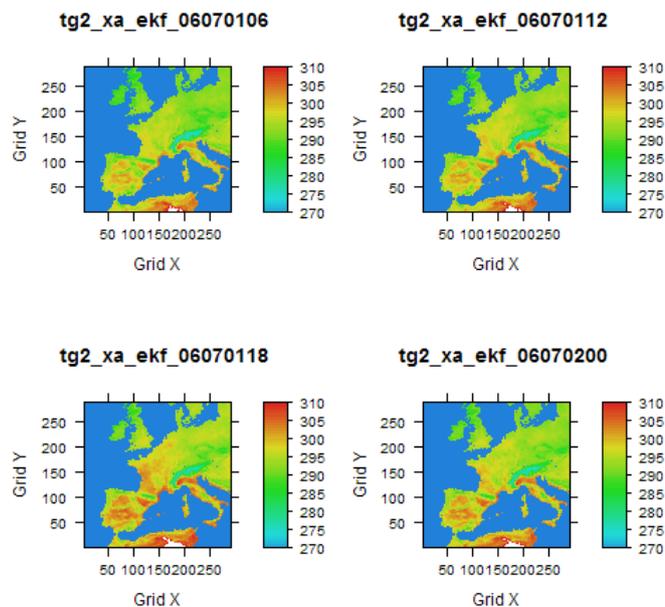


Figure 9: Mean surface temperature analysed field (K) at 4 times using the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

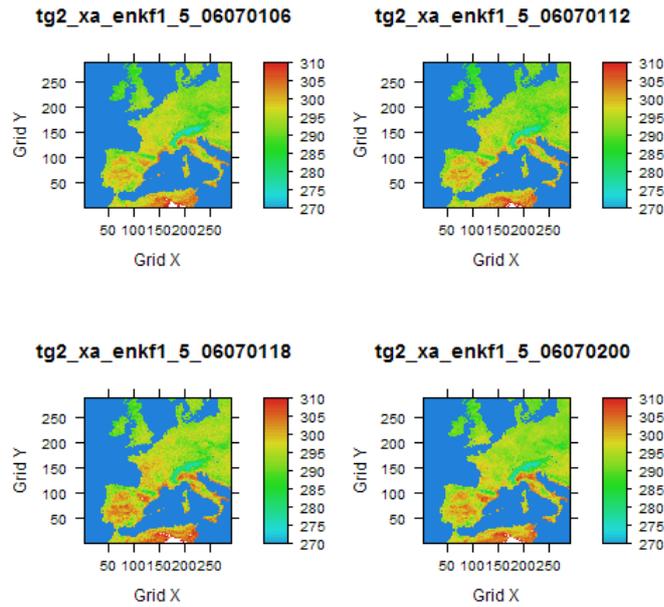


Figure 10: Mean surface temperature analysed field (K) at 4 times using the square root EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

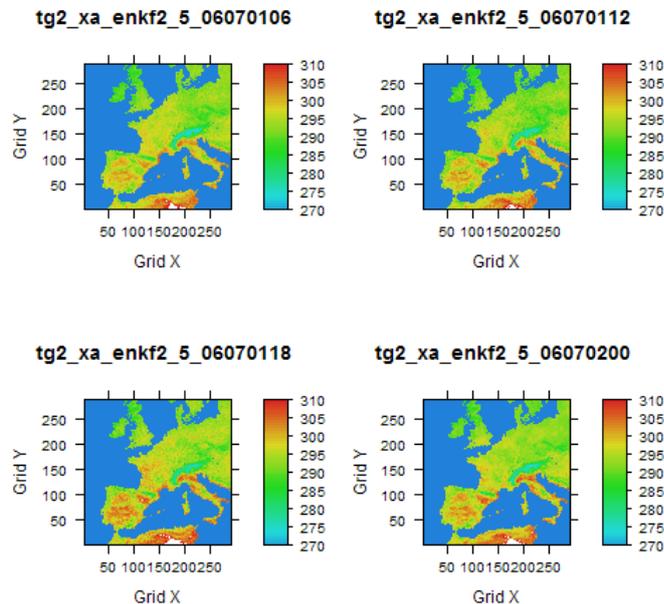


Figure 11: Mean surface temperature analysed field (K) at 4 times using the deterministic EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

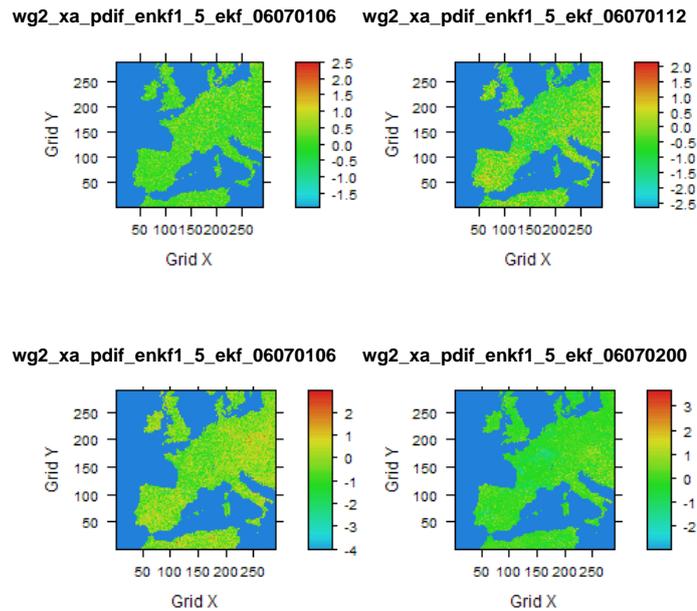


Figure 12: Mean surface temperature difference field (K) at 4 times, percentage difference between the square root EnKF, mean of 5 ensemble members, and the EKF, as a percentage of the EKF values. Positive values indicate the EnKF has higher values than the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

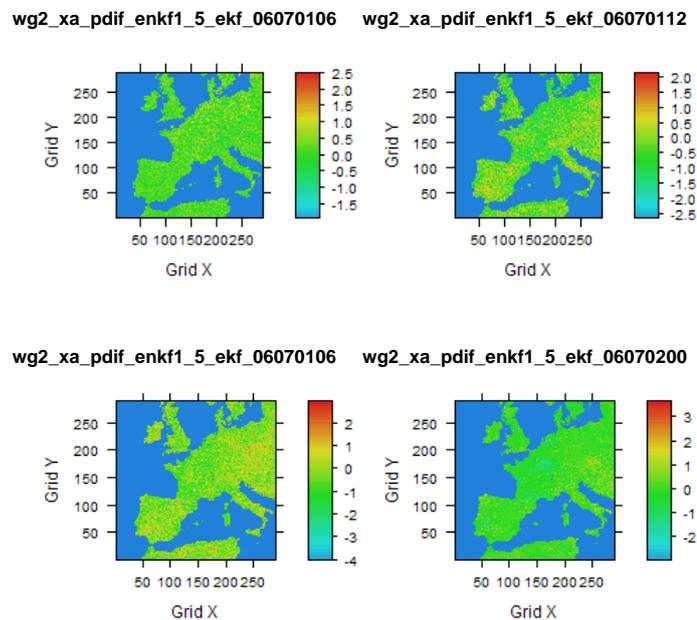


Figure 13: Mean surface temperature difference field (K) at 4 times, percentage difference between the deterministic EnKF, mean of 5 ensemble members, and the EKF, as a percentage of the EKF values. Positive values indicate the EnKF has higher values than the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

3.3 Superficial volumetric water content (WG1 control variable)

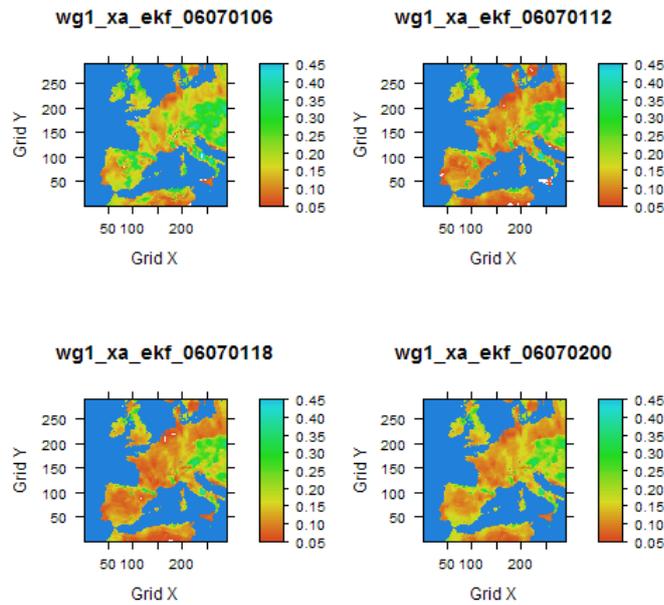


Figure 14: Superficial volumetric water content (m^3/m^3) analysed field at 4 times using the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

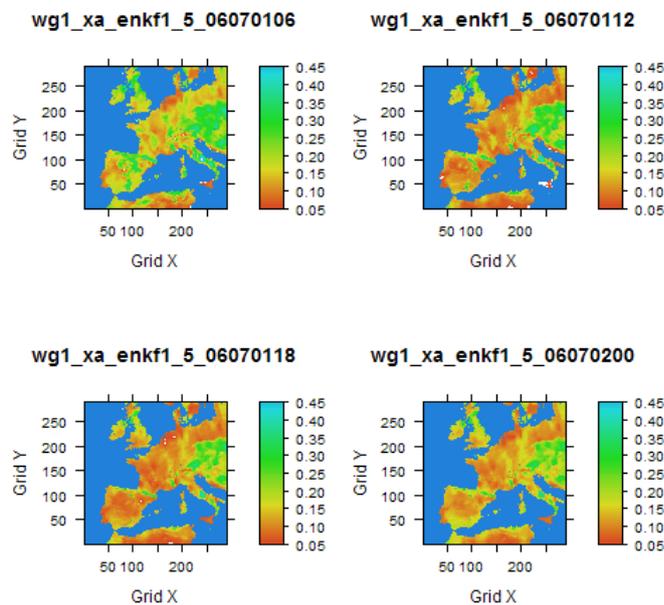


Figure 15: Superficial volumetric water content (m^3/m^3) analysed field at 4 times using the square root EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

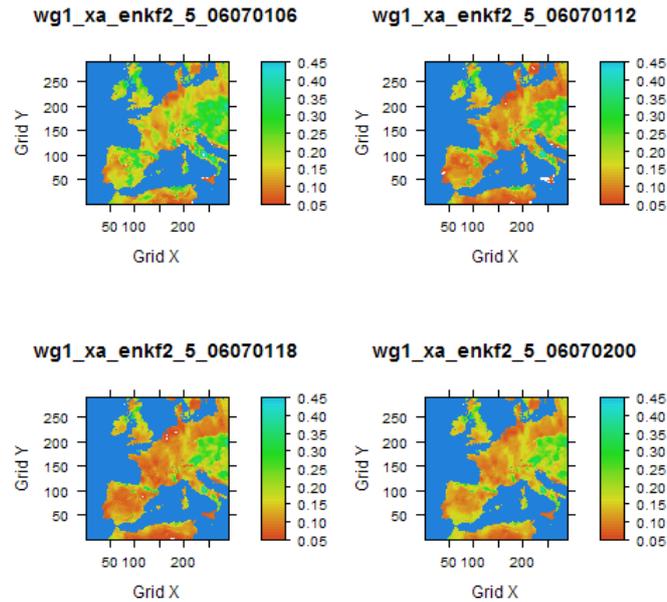


Figure 16: Superficial volumetric water content (m^3/m^3) analysed field at 4 times using the deterministic EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

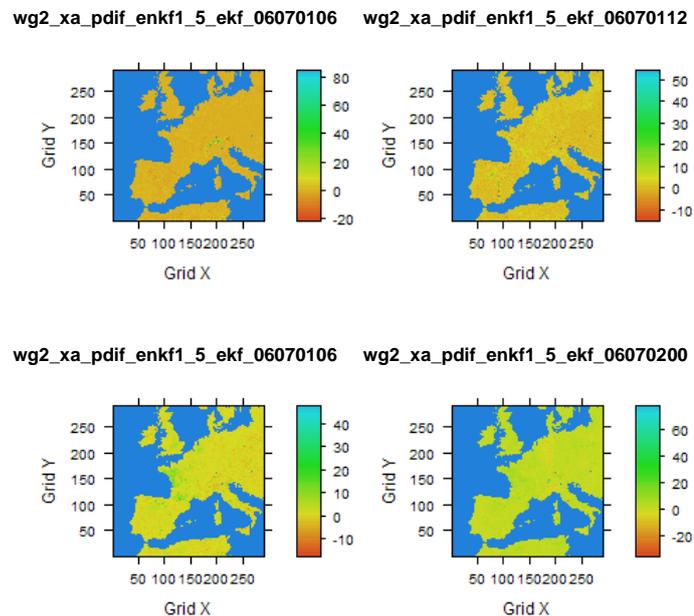


Figure 17: Superficial volumetric water content (m^3/m^3) difference field at 4 times, percentage difference between the square root EnKF, mean of 5 ensemble members, and the EKF, as a percentage of the EKF values. Positive values indicate the EnKF has higher values than the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

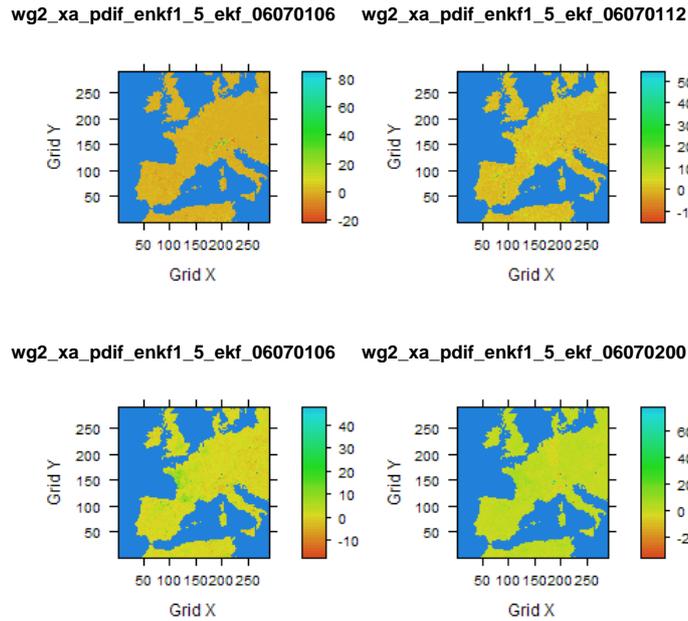


Figure 18: Superficial volumetric water content (m^3/m^3) difference field at 4 times, percentage difference between the deterministic EnKF, mean of 5 ensemble members, and the EKF, as a percentage of the EKF values. Positive values indicate the EnKF has higher values than the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

3.4 Mean volumetric water content (WG2 control variable)

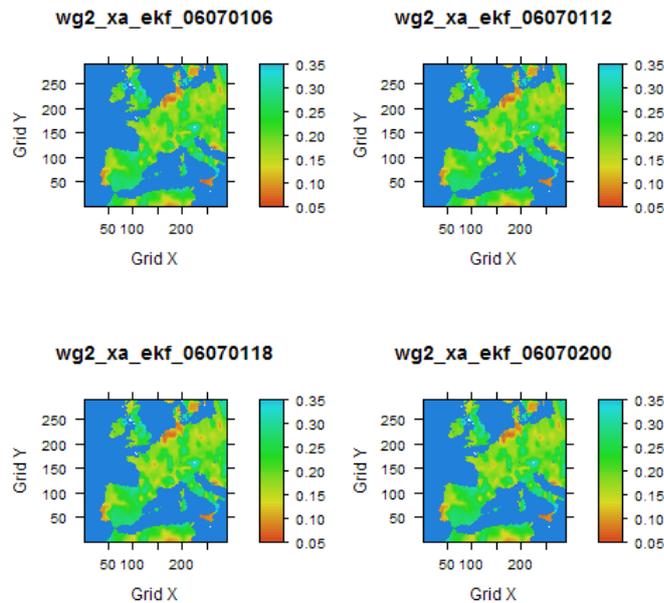


Figure 19: Mean volumetric water content (m^3/m^3) analysed field at 4 times using the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

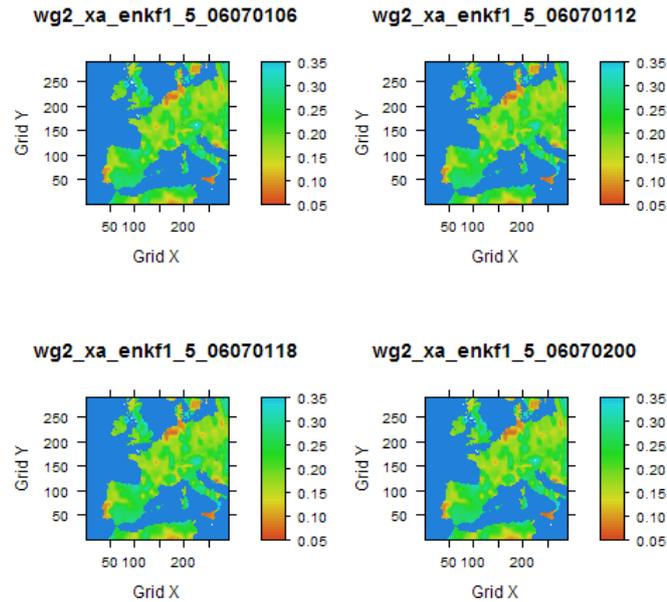


Figure 20: Mean volumetric water content (m^3/m^3) analysed field at 4 times using the square root EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

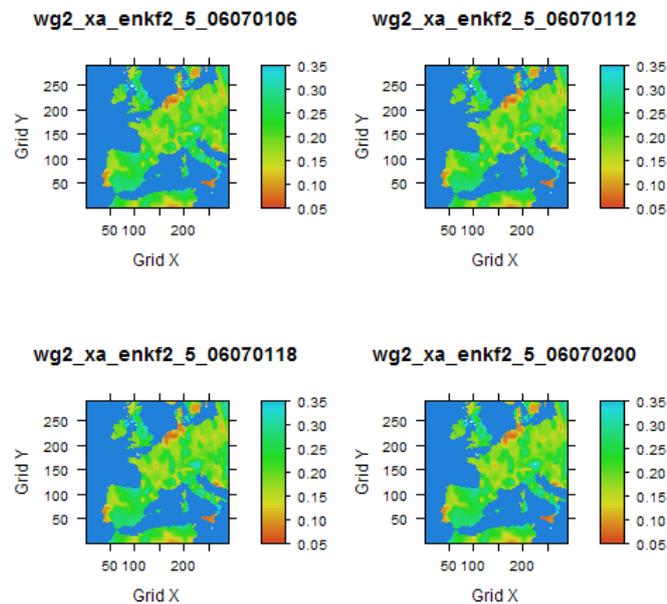


Figure 21: Volumetric water content (m^3/m^3) analysed field at 4 times using the deterministic EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

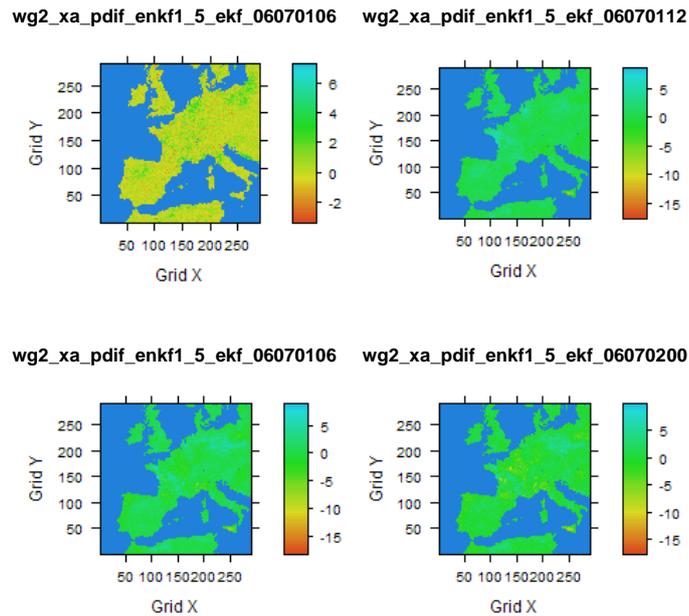


Figure 22: Volumetric water content (m^3/m^3) difference field at 4 times, percentage difference between the square root EnKF, mean of 5 ensemble members, and the EKF, as a percentage of the EKF values. Positive values indicate the EnKF has higher values than the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

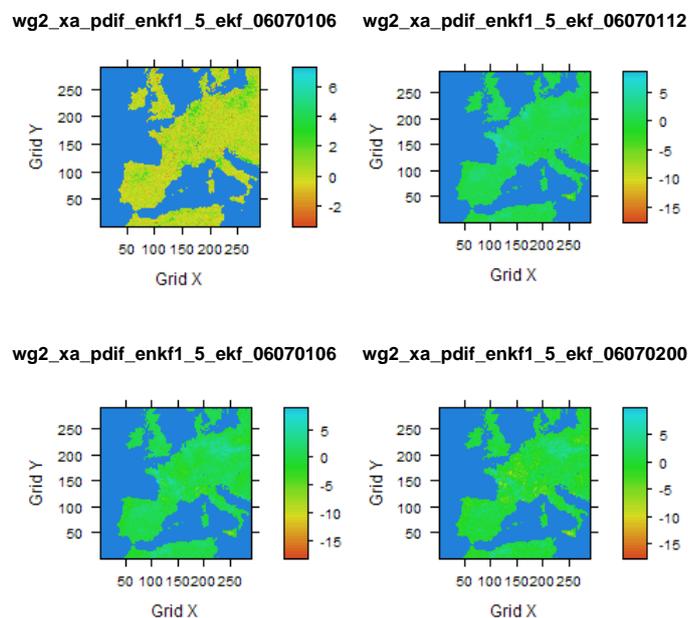


Figure 23: Volumetric water content (m^3/m^3) difference field at 4 times, percentage difference between the deterministic EnKF, mean of 5 ensemble members, and the EKF, as a percentage of the EKF values. Positive values indicate the EnKF has higher values than the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

3.5 TG1 increments

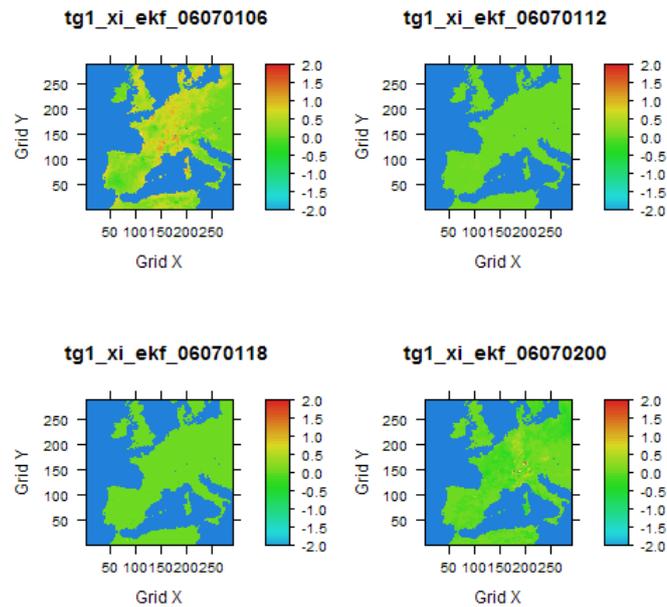


Figure 24: Surface temperature analyses increments (K) at 4 times using the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

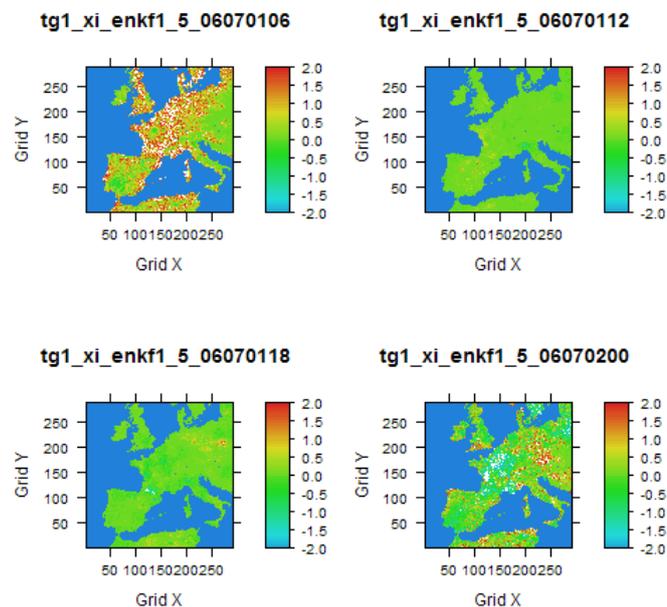


Figure 25: Surface temperature analyses increments (K) at 4 times using the square root EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

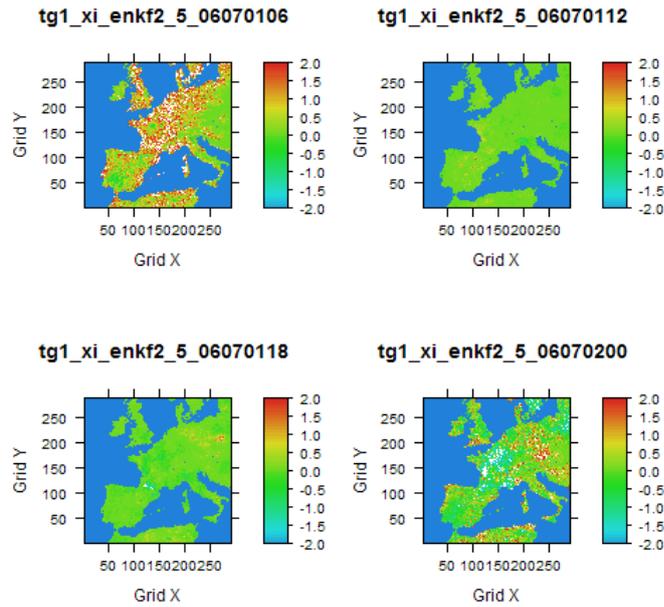


Figure 26: Surface temperature analyses increments (K) at 4 times using the deterministic EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

3.6 TG2 increments

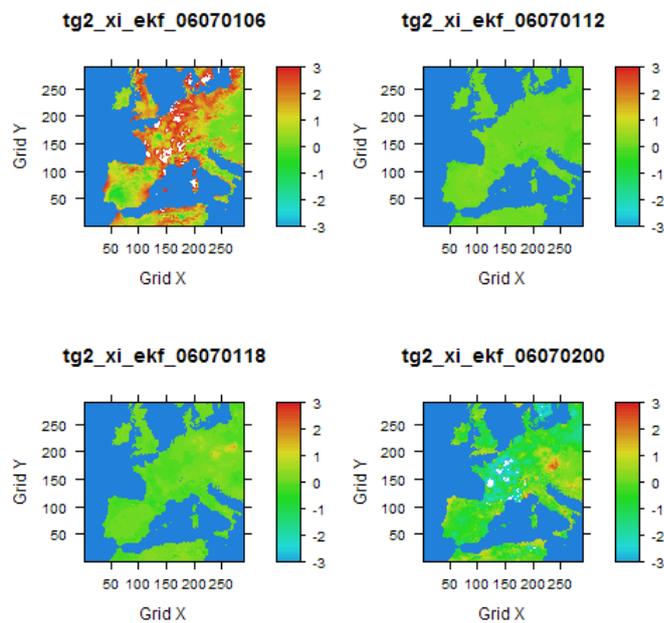


Figure 27: Surface temperature analyses increments (K) at 4 times using the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

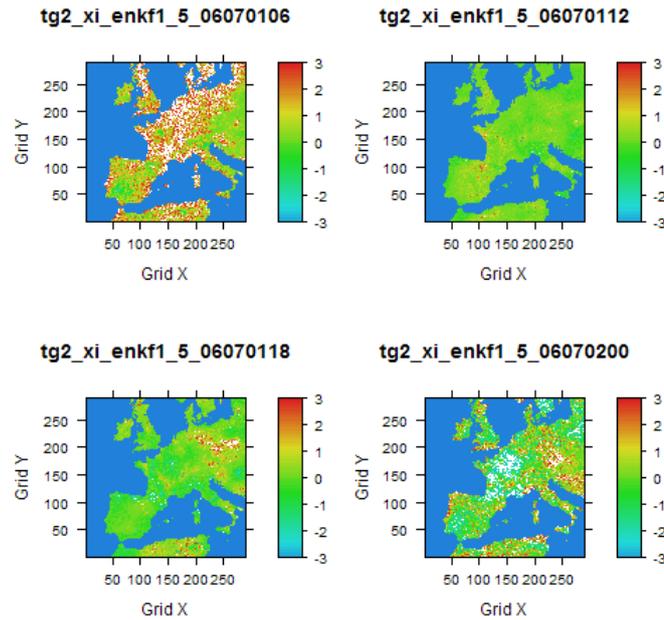


Figure 28: Surface temperature analyses increments (K) at 4 times using the square root EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

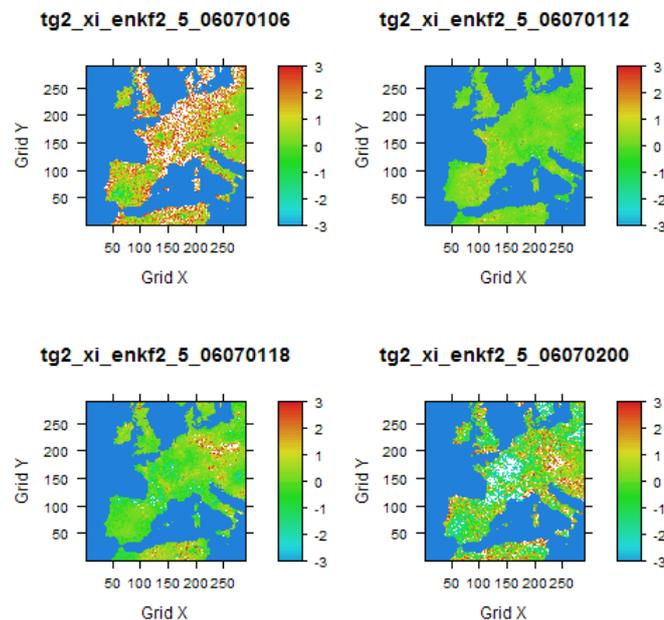


Figure 29: Surface temperature analyses increments (K) at 4 times using the deterministic EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

3.7 WG1 increments

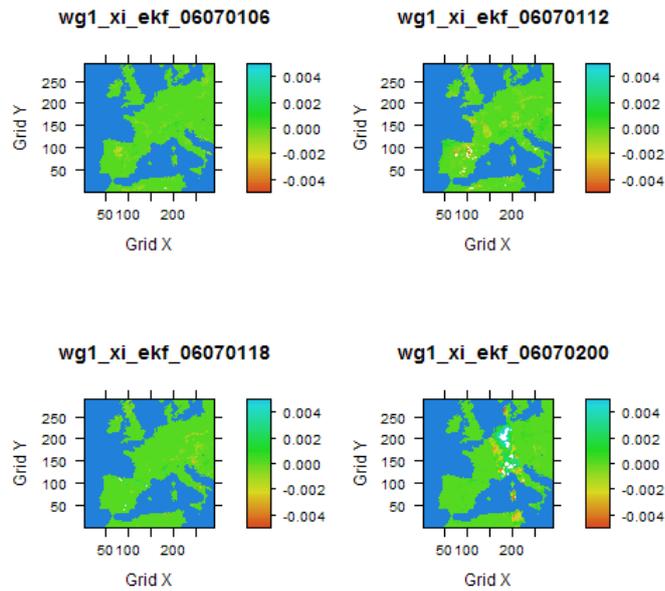


Figure 30: Superficial volumetric water content (m^3/m^3) analyses increments at 4 times using the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

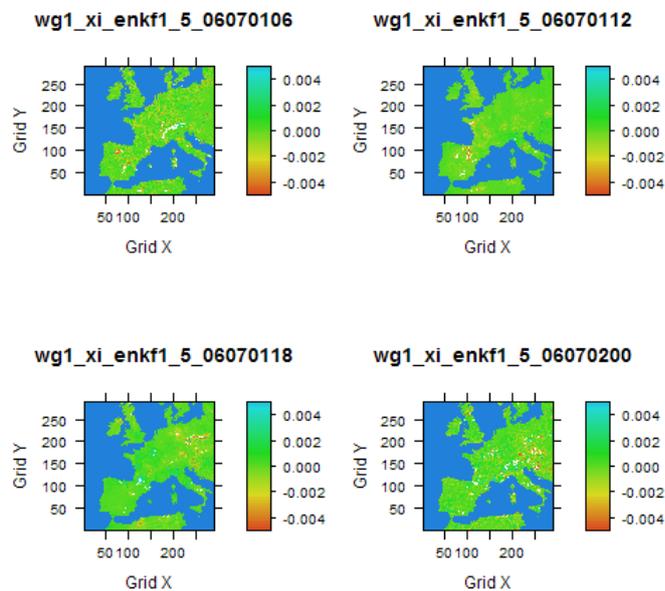


Figure 31: Superficial volumetric water content (m^3/m^3) analyses increments at 4 times using the square root EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

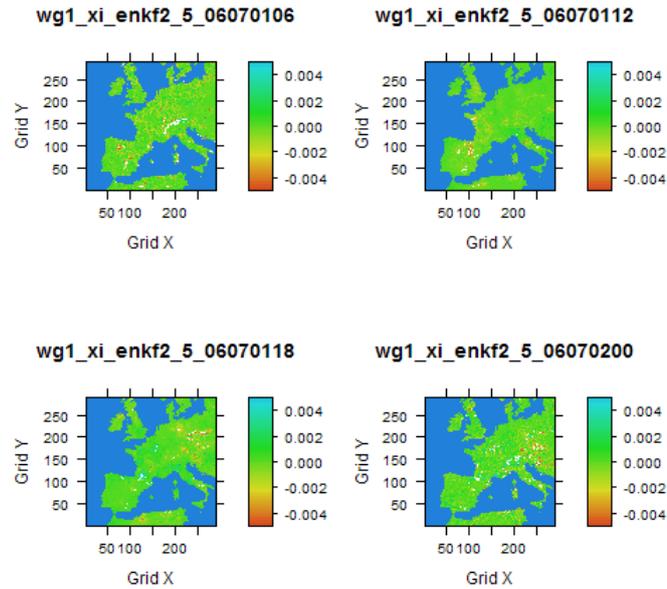


Figure 32: Superficial volumetric water content (m^3/m^3) analyses increments at 4 times using the deterministic EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

3.8 WG2 increments

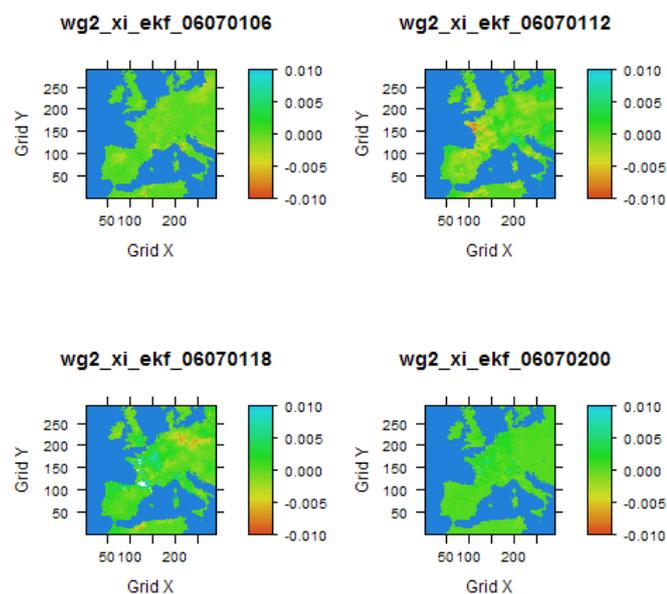


Figure 33: Volumetric water content (m^3/m^3) analyses increments at 4 times using the EKF. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

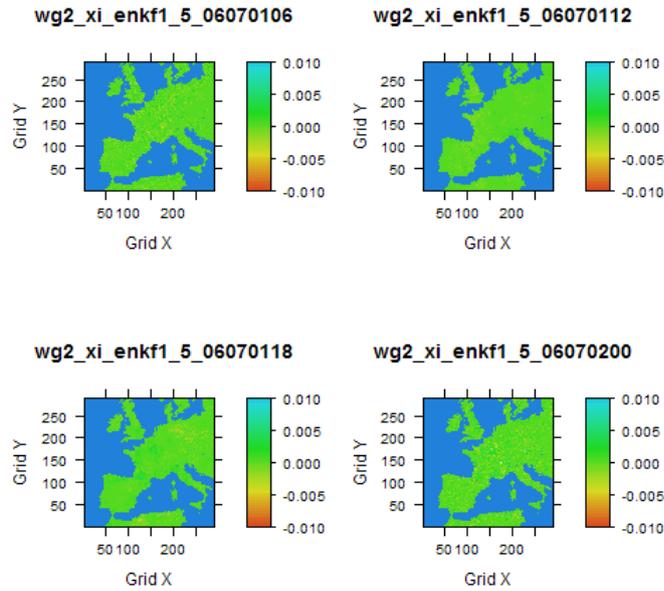


Figure 34: Volumetric water content (m^3/m^3) analyses increments at 4 times using the square root EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

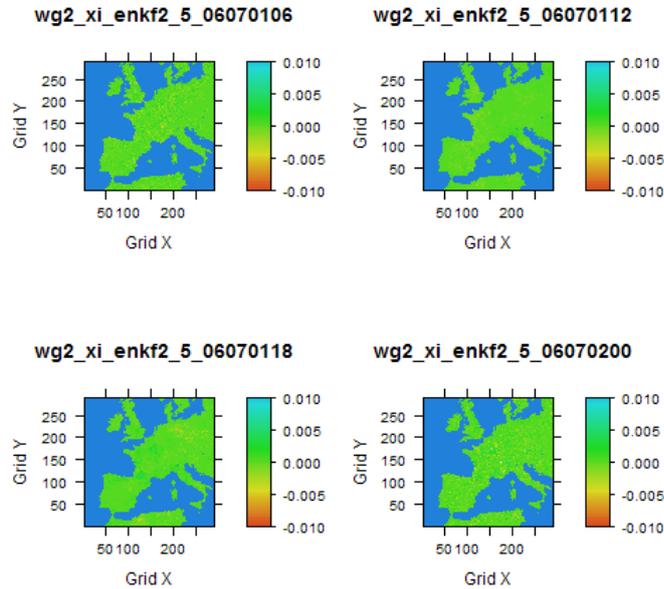


Figure 35: Volumetric water content (m^3/m^3) analyses increments at 4 times using the deterministic EnKF, mean of 5 ensemble members. Top left panel: 0600 UTC, 1 July 2006; top right panel: 1200 UTC, 1 July 2006; bottom left panel: 1800 UTC, 1 July 2006; bottom right panel: 0000 UTC, 2 July 2006.

3.9 Discussion

The analyses and increments from the 24-hour experiment (four 6-hour assimilation cycles) described in the preamble to Section 3 are some of the first results produced with the NILU SURFEX-EnKF system and require further investigation. Extension to a longer assimilation period (order weeks) is needed to properly assess and compare the behaviour of the EKF and EnKF land DA schemes. However, although a 24-hour assimilation experiment is very short in the context of the timescale of soil variables, it is still possible to discern some general features in the behaviour of the DA system and DA schemes, which we identify below.

Comparison of NILU EKF results with Météo-France EKF results

We first discuss comparison of the EKF results shown in Sections 3.1-3.8 against results produced by Météo-France and kindly provided to us by J.-F. Mahfouf (Appendix C). This comparison shows that the NILU EKF results are essentially the same as those of from Météo-France. This suggests that NILU have successfully implemented the SURFEX and EKF set-up at Météo-France.

Analyses

TG1 and TG2. The diurnal cycle in temperature is captured by all data assimilation schemes, with the cycle being less evident in the mean surface temperature (TG2; a volumetric quantity) than for the surface temperature (TG1; a superficial quantity), as expected. The maximum in TG1 should be at 1200 UTC, and the maximum in TG2 should be at 1800 UTC (lagged from TG1). These features are seen for the EKF and the EnKF (both versions give similar results). Note that for TG2, and both the EKF and the EnKF, while the values at 1800 UTC tend to be higher than those at the other times, the values at the other three times (0000 UTC, 0600 UTC and 1200 UTC) are less different between themselves than for TG1.

The percentage difference plots show that the EnKF analysed temperature fields (TG1 and TG2) are very close to the EKF analysed temperature fields, differences being generally around 0%, with maximum values being generally less than 1% in magnitude. This would be expected given the short time period (24 hours) of the data assimilation experiment.

There are a few highly localized outliers in the percentage difference field, but these difference values are generally less than 1% of the EKF analysed values. The difference fields for the TG1 and TG2 forecasts tend to be noisier than for the analyses, especially for TG2 (not shown). These noisy features arise likely due to the way the ensemble members are produced (see preamble to Section 3), and appear to concern the evolution of the SURFEX model ensemble members rather than the way the observations are assimilated.

WG1 and WG2. The general patterns (which reflect the precipitation forcing) in WG1 (a superficial quantity) are similar for the EKF and EnKF, with generally dry areas over most of Western Europe, and generally wet areas over Eastern Europe (except Poland). These patterns do not vary much throughout the day (i.e., the four assimilation cycles). The general patterns for WG2 (a volumetric quantity) are also similar for the EKF and EnKF, and do not vary much

throughout the day; the WG2 patterns have somewhat smaller spatial scales than for WG1.

The percentage difference plots show that very often the differences between the EKF and EnKF (results are similar for both variants of the EnKF) WG1 and WG2 analysed fields are around zero, and small in absolute value. This would be expected given the short time period (24 hours) of the data assimilation experiment.

However, there are some exceptions where the differences are not close to zero, and occur over areas bigger than patches, and thus less likely to be attributable to noise. For WG1, the EnKF is slightly wetter than EKF at 1800 UTC over regions that are relatively very dry (e.g. Southern England, Southwest France) – confirmed by the absolute difference plots (not shown). For WG2, the EnKF is slightly drier than EKF at 1800 UTC over regions that are neither very dry nor very wet (e.g. parts of Southwest France) – confirmed by the absolute difference plots (not shown).

Note that because the analysed values of WG1/WG2 concerned are very small, occasionally the percentage difference can be large in magnitude. There is a suggestion that the percentage difference field for WG1/WG2 is noisier than for TG1/TG2. As for TG1/TG2, the way the ensemble members are set up is likely to contribute to this noisy feature, although the smaller values for WG1/WG2 are also likely to contribute. Although results are not clear cut, there is a suggestion that the difference fields for the WG1 and WG2 forecasts are occasionally slightly noisier than for the analyses (not shown). This suggests that, as well as the way the ensemble is set up (see above), the smaller values for WG1/WG2 could also play a role in producing these noisy difference fields.

Summary. The results presented in Sections 3.1-3.4 show that the EKF and the two versions of the EnKF (for 5 ensemble members) represent qualitatively similar analysed fields with, for example, relatively warm regions coinciding with relatively dry regions (e.g. low lying areas in Spain), and relatively cold regions coinciding with relatively wet regions (e.g. mountain areas such as the Alps and the Pyrenees). The performance of the EnKF schemes for 2 and 3 ensemble members is similar to that for 5 ensemble members shown in the figures in Sections 3.1-3.4 (this is also the case for the increments, Sections 3.5-3.8 – see below).

In comparison to the EKF scheme, the EnKF variants perform similarly for the temperature control variables. The performance for the humidity control variables (WG1/WG2) is slightly different for the EnKF variants (with respect to the EKF), with some differences occurring over larger areas than for TG1/TG2, although there is no clear cut evidence of a bias in any of the schemes.

Results with 9 ensemble members (not shown) show EnKF analysed fields that are slightly closer to the EKF analysed fields, but show no decrease in the noisy features present in the EnKF analyses and increments. This supports the notion that the main reason for the noisy features in the EnKF is the way the ensemble members are set up.

Increments

General comments. The increments strongly vary according to the diurnal cycle, with soil temperature increments being larger during nighttime (0000 and 0600 UTC), and soil moisture increments being larger during daytime (1200 and 1800 UTC). At their maximum, value, the increments should have the same magnitude as the background errors (J.-F. Mahfouf pers. comm.), i.e., order 1 K for soil temperature and order $0.01 \text{ m}^3/\text{m}^3$ for soil moisture. This is seen to be generally the case. This subsection has benefited from discussions from J.-F. Mahfouf, which have helped clarify the behaviour of the TG1/TG2 and WG1/WG2 increments.

TG1 and TG2. For both the EKF, and the EnKF variants, the largest increments (in magnitude) for TG1/TG2 occur during nighttime (0000 UTC and 0600 UTC), as expected. The pattern in the TG2 increments is similar to that in the TG1 increments (e.g. positive increments over France at 0000 UTC), with the TG2 increments being larger in magnitude by a factor of about 3-4. This is expected, as the surface layer (with which TG1 is associated) has very low thermal inertia, i.e., it has short-term memory. At 1800 UTC the largest values in the TG2 increments are in the Eastern part of the domain (e.g. Poland). This is to be expected as this region experiences nighttime first during the 24-hour assimilation experiment.

The increments for TG1/TG2 for the EKF and EnKF variants have the same spatial patterns. However, for the EnKF, the increments are slightly larger in magnitude and have small-scale noise. This is likely due to the way the ensemble members are set up (see the above discussion on the analyses), which would affect the representation of the background errors (as discussed above, at their maximum, the increments should have the same magnitude as the background errors).

WG1 and WG2. For both the EKF, and the EnKF variants, the increments are generally small, especially for WG1. WG1 increments should be small as this reservoir is very shallow, as it has no memory, i.e., no sensitivity to initial conditions. The increments for WG2 should be bigger in magnitude than those for WG1, as WG2 is the most sensitive variable to surface fluxes. An example occurs for the EKF over Northwest France, where there are strong negative increments at 1200 UTC, and strong positive increments at 1800 UTC. For the EnKF, and over the same region, the range of the colour palettes suggests this is also the case, although the effect is less clear than for the EKF. Overall, the largest increments (in magnitude) for WG1/WG2 occur during daytime (1200 UTC and 1800 UTC), as expected.

As for TG1/TG2, the increments for WG1/WG2 for the EKF, and EnKF variants, have the same spatial patterns. There are cases of slightly more negative increments for the EnKF for WG1 (e.g. 0600 UTC over Spain, 1800 UTC over Eastern Europe), and slightly more positive increments for the EnKF for WG2 (e.g. 1200 UTC over Northwest France), but as for the analyses, there is no clear cut evidence of a bias in any of the schemes. As for TG1/TG2, a difference between the EKF and EnKF is the presence of small-scale noise in the increments for the EnKF variants, likely due to the way the ensemble members are set up (see the above discussion on the analyses).

4 Plans for future developments

4.1 Immediate future

Technical

For technical reasons, the current focus of the NILU SURFEX-EnKF land DA system is France and its surrounding area for a period in early July 2006 (see results in Section 3). At a later stage, with the collaboration of Met.no, NILU will shift its focus to Scandinavia and, in particular, the Norway/Sweden area, and other time periods will be considered. NILU will also consider other forcings of the land DA system, including at a later stage the possibility of an on-line system where the atmosphere and land are coupled, and other observations (in situ and satellite).

As seen in section 3.9, the EnKF analyses and increment fields have noisy features that are unlikely to be realistic, and are likely to arise due to the way the ensemble members are set up. This makes it important that a better way to set up the ensemble members be found. To remedy this perceived weakness in the EnKF set-up, other methods of setting up the ensemble members are being developed and tested. For example, we are considering introducing a simple representation of the ensemble spread based on the diagonal (background error variance) of the **B** matrix in the EKF set-up; later, we will consider a horizontal localization to account for correlations between physical variables over a finite spatial scale.

Finally, the computer code will be optimized to work with ten or more ensemble members.

Data assimilation set up

As part of the development of the land DA system, the schemes incorporated into the system (EKF; variants of the EnKF and PF) will be tested and compared for various spatial and temporal scenarios. Standard tests for evaluating the self-consistency of the land DA system, and tests against independent data to evaluate the quality of the analyses, will be implemented (see Section 2.6 for details).

Measures of non-linearity (e.g. Lyapunov exponents) may be considered to assess the non-linear behaviour (or otherwise) of the various DA schemes.

Assimilation of satellite data

Soil moisture from AMSR-E will be assimilated into the NILU SURFEX-EnKF system for the month of July 2006. This work will extend that described in Draper et al. (2009), and compare the performance of the EKF against that of the EnKF variants implemented in the NILU SURFEX-EnKF system. As of January 2010, the data assimilation experiment (satellite data, forcing data) is being set up at NILU.

NILU lead the ISSI (International Space Science Institute) International Team “Land Data Assimilation: Making sense of hydrological cycle observations”, which focuses on assimilation of soil moisture from SMOS. The International Team has funding for 3 meetings to be held at ISSI in Bern, Switzerland. The first meeting will take place February 2010. International Team members include leading experts in the disciplinary areas in land data assimilation, and land surface

Earth Observation and modelling in Europe and the US. They come from institutions in Norway (NILU, Met.no and NERSC), France (Météo-France, LMD and U. Toulouse), Belgium (U. Ghent), The Netherlands (KNMI), USA (George Mason University), UK (U. Exeter), Spain (U. Valencia) and Finland (FMI). The PI of SMOS, Y. Kerr (based at U. Toulouse) is a member. The ISSI International Team will be used to develop assimilation of soil moisture data from SMOS, and strengthen collaboration in the hydrological sciences.

4.2 Long-term future

In the long-term future, NILU will move from a 1-D approach to a 3-D approach, to avoid potential inconsistencies between the atmospheric forcing data and the land variables. This will be relevant for assimilation experiments longer than a few days.

NILU will consider various refinements to the treatment of errors. In particular, the suitability of the Desroziers approach to estimate model and observation error covariances separately will be tested; bias correction schemes to correct biases in the observations and/or the model will be tested; and the potential of possibility theory ideas will be tested for characterizing observational errors.

Development of a hybrid land DA system incorporating the advantages of the EnKF and the PF will also be considered.

Availability of system to SURFEX-modelling community

Once the 1-D NILU SURFEX-EnKF land DA system has been tested further and its functionality extended to include more general time periods and geographical areas, it will be made available to the SURFEX-modelling community. This could be done with the SURFEX model being off-line. At a future time, an off-line 3-D version, and an on-line version (1-D or 3-D) could be made available.

4.3 Collaboration

The NILU land DA system owes a lot to collaboration between NILU, Météo-France and Met.no, and the invaluable help provided by the latter two to NILU. The broad areas of collaboration for NILU are currently expected to be:

- DA theory: NERSC and Met.no, Norway; Météo-France, LMD (France);
- Comparison of land DA schemes: NERSC and Met.no, Norway; Météo-France, LMD (France); Paul Houser and colleagues, USA;
- Bias correction: Paul Houser and colleagues, USA;
- Possibility theory: Niko Verhoest and colleagues, Belgium;
- Hybrid land DA system: NERSC and Met.no, Norway;
- Land surface observations, including temperature, soil moisture (including SMOS) and snow: members of the ISSI International Team “Making sense of hydrological cycle observations” – this includes the SMOS PI, Yann Kerr; members of the GlobSNOW consortium (<http://globsnow.fmi.fi>);
- Assimilation of soil moisture data from AMSR-E: Météo-France (France).

5 Conclusions

In collaboration with Météo-France and Met.no, NILU is developing the SURFEX-EnKF land data assimilation (DA) system. This is based on the SURFEX land surface model developed at Météo-France, and used in the HIRLAM/ALADIN NWP consortium. The main motivation is to develop a land DA system to test various state-of-the-art DA schemes (EKF; variants of the EnKF and PF), produce analyses of land variables, and test land surface observations and models. This information will allow better use of Earth Observation data by the NWP community, will help improve short- and medium-term NWP forecasts, and will help improve land surface models and climate models. These outcomes will also benefit the scientific community interested in the variability and evolution of the land component of the Earth System, and its interactions with other elements of the Earth System.

This report presents very preliminary results of the SURFEX-EnKF land DA system being developed at NILU, which show a flavour of what it can do. These results suggest that the NILU system for the EKF has been set up properly, and that the NILU system produces qualitatively and quantitatively reasonable results with both the EKF and the EnKF schemes. One clear difference between the DA schemes is the presence of small-scale noise in the EnKF analyses and increments, likely due to the way the ensemble members are set up in the EnKF. To remedy this perceived weakness in the EnKF set-up, other methods of setting up the ensemble members are being developed and tested. For example, we are considering introducing a simple representation of the ensemble spread based on the diagonal (background error variance) of the \mathbf{B} matrix in the EKF set-up; later, we will consider a horizontal localization to account for correlations between physical variables over a finite spatial scale.

It is still early days and many things remain to be done, but NILU now have the capability (in collaboration with colleagues at Met.no and Météo-France) to develop a powerful tool to study the land component of the Earth System.

6 Acknowledgments

We thank Météo-France (Jean-François Mahfouf, Eric Martin) and Met.no (Trygve Aspelien, Jelena Bojarova, Nils Gustafsson, Mariken Homleid, Viel Ødegaard) for providing the SURFEX model and the EKF land DA system, and for invaluable help and advice. We thank other colleagues (Paul Houser, Pavel Sakov and Olivier Talagrand) for useful discussions which have helped develop our ideas of how to perform land data assimilation. This effort is supported via an internal project at NILU, and various external projects (NILU Data Assimilation SIP, EU-GENESI, ESA harmonization, ISSI International Team, and COST Action “Advancing the integrated monitoring of trace gas exchange between biosphere and atmosphere”).

7 References

- Andreadis, K. and Lettenmaier, D. (2005) Assimilating remotely sensed snow observations into a macroscale hydrology model. *Adv. Water Resour.*, 29, 872-886.
- Bosilovich, M., Radakovich, J.D., Silva, A. da, Todling, R. and Verter, F. (2007) Skin temperature analysis and bias correction in a coupled land-atmosphere data assimilation system. *J. Meteorol. Soc. Jpn.*, 85A, 205-228.
- Bouttier, F. and Courtier, P. (1999) Data assimilation concepts and methods, March 1999. Meteorological training course lecture notes. London, ECMWF. Available from <http://www.ecmwf.int>.
- Burgers, G., van Leeuwen, P.J. and Evensen, G. (1998) Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.*, 126, 1719-1724.
- Chen, Z. (2003) Bayesian Filtering: From Kalman Filters to Particle Filters, and Beyond. Technical Report. Hamilton, ON, Canada, Adaptive Syst. Lab. McMaster University.
- Clark, M.P., Rupp, D.E., Woods, R.A., Zheng, X., Ibbitt, R.P., Slater, A.G., Schmidt, J. and Uddstrom, M.J. (2008) Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Adv. Water Resour.*, 31, 1309-1324.
- Cohn, S.E. (1997) An introduction to estimation theory. *J. Meteorol. Soc. Jpn.*, 75, 257-288.
- Dee, D.P. (2005) Bias and data assimilation. *Quart. J. Roy. Meteorol. Soc.*, 131, 3323-3343.
- Dee, D. and da Silva, A. (1998) Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteorol. Soc.*, 124, 269-295.
- De Lannoy, G.J.M., Reichle, R.H., Houser, P.R., Pauwels, V.R.N. and Verhoest, N.E.C. (2007a) Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. *Water Resour. Res.*, 43, W09410.
- De Lannoy, G.J.M., Houser, P.R., Pauwels, V.R.N. and Verhoest, N.E.C. (2007b) State and bias estimation for soil moisture profiles by an ensemble Kalman filter. *Water Resour. Res.*, 43, W06401.
- Desroziers, G., Berre, L., Chapnik, B. and Poli, P. (2005) Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteorol. Soc.*, 131, 3385-3396.
- Dong, J., Walker, J. P., Houser, P. R. and Sun, C. (2007) Scanning multichannel microwave radiometer snow equivalent assimilation. *J. Geophys. Res.*, 112, D07108, doi:10.1029/2006JD007209.
- Doucet, A. and Johansen, A.M. (2008) A tutorial on particle filtering and smoothing: Fifteen years later. In: *Handbook of nonlinear filtering*. Ed. by D. Crisan and B. Rozovsky. Oxford University Press (in press). Available from http://www.cs.ubc.ca/~arnaud/doucet_johansen_tutorialPF.pdf
- Doucet, A., de Freitas, N. and Gordon, N. (eds.) (2001) Sequential Monte Carlo methods in practice. New York, Springer Verlag.

- Draper, C.S., Mahfouf, J.-F. and Walker, J.P. (2009) An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme. *J. Geophys. Res.*, *114*, D20104, doi: 10.1029/2008JD011650.
- Drusch, M. (2007) Initializing numerical weather prediction models with satellite surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set. *J. Geophys. Res.*, *112*, doi: 10.1029/2006JD007478.
- Drusch, M., Scipal, K., de Rosnay, P., Balsamo, G., Andersson, E., Bougeault, P. and Viterbo, P. (2009) Towards a Kalman Filter based soil moisture analysis system for the operational ECMWF Integrated Forecast System. *Geophys. Res. Lett.*, *36*, L10401, doi:10.1029/2009GL037716.
- Eben, K., Jurus, P., Resler, J., Belda, M., Pelikán, E., Krüger, B. C. and Keder, J. (2005) An ensemble Kalman filter for short-term forecasting of tropospheric ozone concentrations. *Quart. J. Roy. Meteorol. Soc.*, *131*, 3313–3322.
- Eley, J. (1992) Summary of Workshop, Soil Moisture Modeling. Proceedings of the NHRC Workshop held March 9-10, 1992. (NHRI Symposium Proceedings, 9).
- Evensen, G. (2003) The Ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.*, *53*, 343–367.
- Evensen, G. (2007) Data assimilation: the ensemble Kalman filter. Berlin, Springer.
- Evensen, G. and van Leeuwen, P.J. (2000) An ensemble Kalman smoother for nonlinear dynamics. *Mon. Weather Rev.*, *128*, 1852–1867.
- Evensen, G., Hove, J., Meisingset, H.C., Reiso, E., Seim, K.S. and Espelid, O. (2006) Using the EnKF for assisted history matching of a North Sea reservoir mode. Society of Petroleum Engineers (SPE 106184). URL: <http://enkf.nersc.no/Publications/eve07a.pdf>
- Geer, A.J., Lahoz, W.A., Bekki, S., Bormann, N., Errera, Q., Eskes, H. J., Fonteyn, D., Jackson, D. R., Juckes, M. N., Massart, S., Peuch, V.-H., Rharmili, S., and Segers, A. (2006) The ASSET intercomparison of ozone analyses: method and first results. *Atmos. Chem. Phys.*, *6*, 5445–5474.
- Geer, A.J., Lahoz, W.A., Jackson, D.R., Cariolle, D. and McCormack, J. P. (2007) Evaluation of linear ozone photochemistry parametrizations in a stratosphere-troposphere data assimilation system. *Atmos, Chem. Phys.*, *7*, 939–959.
- Gelman A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004) Bayesian data analysis, 2nd ed. Boca Raton, Fla., Chapman & Hall/CRC.
- Giard, D. and Bazile, E. (2000) Implementation of a new assimilation scheme for soil and surface variables in a global NWP model. *Mon. Weather Rev.*, *128*, 997-1015.
- Houser, P.R. (2003) Land data assimilation systems. In: *Data Assimilation for the Earth System*. Ed. by R. Swinbank, V. Shutyaev and W.A. Lahoz, Dordrecht, Kluwer. (NATO science series: IV: Earth and Environmental Sciences, 26). pp. 345-360.

- Houser, P.R., et al. (2010) Land surface data assimilation. In: *Data Assimilation: Making sense of observations*. Eds, W.A. Lahoz, B. Khatatov and R. Ménard. Dordrecht, Springer, due March 2010.
- Houtekamer, P.L. and Mitchell, H.L. (2001) A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.*, 129, 123–137.
- Houtekamer, P.L., Mitchell, H.L., Pellerin, G., Buehner, M., Charron, M., Spacek, L. and Hansen, B. (2005) Atmospheric data assimilation with an ensemble Kalman filter: Results with real observation. *Mon. Weather Rev.*, 133, 604–620.
- Kalnay, E. (2003) Atmospheric modeling, data assimilation and predictability. Cambridge, Cambridge University Press.
- Kalnay, E., Li, H., Miyoshi, T, Yang, S.-C. and Ballabrera-Poy, J. (2007) 4-D-Var or ensemble Kalman filter? *Tellus*, 59A, 758–773.
- Keppenne, C.L. and Rienecker, M.M. (2002) Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. *Mon. Weather Rev.*, 130, 2951–2965.
- Lahoz, W.A. (2008) Status and Plans for cooperation with SURFEX between NILU and Met.no, presentation made to HIRLAM Management Group, November 2008. Available from www.nilu.no.
- Lahoz, W.A., Errera, Q., Swinbank, R. and Fonteyn, D. (2007a) Data assimilation of stratospheric constituents: A review. *Atmos. Chem. Phys.*, 7, 5745–5773, 2007.
- Lahoz, W.A., Geer, A. J., Bekki, S., Bormann, N., Ceccherini, S., Elbern, H., Errera, Q., Eskes, H. J., Fonteyn, D., Jackson, D. R., Khatatov, B., Marchand, M., Massart, S., Peuch, V.-H., Rharmili, S., Ridolfi, M., Segers, A., Talagrand, O., Thornton, H. E., Vik, A. F. and von Clarmann, T. (2007b) The Assimilation of Envisat data (ASSET) project. *Atmos. Chem. Phys.*, 7, 1773-1796.
- Le Moigne, P. (2009) SURFEX Scientific Documentation. Note de Centre No. 87, Groupe de Meteorologie a Moyenne Echelle, Centre National de Recherches Météorologique, Météo-France (SURFEX, v. 5 - Issue 1). URL: http://hirlam.org/index.php?option=com_docman&task=doc_download&gid=605&Itemid=70.
- Lillesand, T.M. and Kiefer, R.W. (1994) Remote sensing and photo interpretation. 3rd ed. New York, John Wiley & Sons.
- Mahfouf, J.-F. (2007) L'Analyse dans le sol a Meteo-France. Partie 1: Evaluation et perspectives a l'échelle locale. Note de travail du Groupe de Meteorologie a Moyenne Echelle, No. 84. Available from Météo-France.
- Mahfouf, J.-F., Bergaoui, K., Draper, C., Bouyssel, F. Taillefer, F. and Taseva, L. (2009) A comparison of two off-line soil analysis schemes for assimilation of screen-level observations. *J. Geophys. Res.*, 114, D08105, doi: 10.1029/20008JD011077.
- Mandel, J. (2007) A brief tutorial on the Ensemble Kalman Filter. University of Colorado at Denver and Health Sciences Center. (UCDHSC/CCM Report no. 242).

- Ménard, R. (2010) Bias estimation. In: *Data Assimilation: Making sense of observations*. Ed. by W.A. Lahoz, B. Khattatov and R. Ménard. Springer, due March 2010.
- Moradkhani, H., Hsu, K.-L., Gupta, H. and Sorooshian, S. (2005) Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resour. Res.*, *41*, W05012, doi:10.1029/2004WR003604.
- Palmer, T.N., Doblas-Reyes, F. J., Weisheimer, A. and Rodwell, M. J. (2008) Toward seamless prediction. calibration of climate change projections using seasonal forecasts. *Bull. Amer. Meteorol. Soc.*, *89*, 459-470.
- Phillips, T.J., Potter, G.L., Williamson, D.L., Cederwall, R.T., Boyle, J.S., Fiorino, M., Hnilo, J.J., Olson, J.G., Xie, S. and Yi, J.J. (2004) Evaluating parameterizations in general circulation models: Climate simulation meets weather prediction. *Bull. Amer. Meteorol. Soc.*, *85*, 1903–1915.
- Puri, K., ed. (2002) Activities of the CAS/JSC Working Group on Numerical Experimentation, WGNE. In: *Research activities in atmospheric and oceanic modelling*. Ed. by H. Ritchie. Geneve (WMO/TD-No. 1105, Report 32).
- Reichle, R.H. and Koster, R.D. (2003) Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. *J. Hydrometeorol.*, *4*, 1229-1242.
- Reichle, R.H., Walker, J.P., Koster, R.D. and Houser, P.R. (2002) Extended versus ensemble Kalman filtering for land data assimilation. *J. Hydrometeorol.*, *3*, 728-740.
- Reichle, R.H., Crow, W.T. and Keppenne, C.L. (2008) An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resour. Res.*, *44*, W03423 , doi:10.1029/2007WR006357.
- Ristic, B., Arulampalam, S. And Gordon N. (2004) Beyond the Kalman Filter: Particle filters for tracking applications. Boston, Artech House Publishers.
- Robinson, D., Bevins, R. E. and Rowbotham, G. (1993) The characterization of mafic phyllosilicates in low-grade metabasalts from eastern North Greenland. *Am. Mineral.*, *78*, 377-390.
- Rodgers, C.D. (2000) Inverse methods for atmospheric sounding: Theory and practice. Singapore, World Scientific Publishing Co.
- Rodwell, M.J. and Palmer, T.N. (2007) Using numerical weather prediction to assess climate models. *Quart. J. Roy. Meteorol. Soc.*, *133*, 129-146.
- Rood, R.B. (2005) Assimilation of stratospheric meteorological and constituent observations: A Review. *SPARC Newsletter*, No. 25, July 2005.
- Sakov, P. and Oke, P.R. (2008a) Implications of the form of the ensemble transformation in the ensemble square root filters. *Mon. Weather Rev.*, *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.

- Sakov, P., Evensen, G. and Bertino, L. (2010) Asynchronous data assimilation with the EnKF. *Tellus*, 62A, 24-29.
- Simmons, A.J. and Hollingsworth, A. (2002) Some aspects of the improvement in skill in numerical weather prediction. *Quart. J. Roy. Meteorol. Soc.*, 128, 547-677.
- Slater, A. and Clark, M. (2006) Snow data assimilation via an ensemble Kalman filter. *J. Hydrometeorol.*, 7, 478-493.
- Snyder, C., Bengtsson, T., Bickel, P. and Anderson, J. (2008) Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.*, 136, 4629–4640.
- Szunyogh, I., Kostelich, E.J., Gyarmati, G., Kalnay, E., Hunt, B.R., Ott, E., Satterfield, E. and Yorke, J.A. (2008) A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus*, 60A, 113–130.
- Taillefer, F. (2002) CANARI (Code for the Analysis Necessary for Arpege for its Rejects and its Initialization): Technical Documentation. Internal CNRM/GMAP report.
URL: <http://www.cnrm.meteo.fr/gmapdoc/spip.php?article3>.
- Talagrand, O. (2003a) Bayesian estimation. Optimal interpolation. Statistical linear estimation. In: *Data assimilation for the earth system*. Ed. by R. Swinbank, V. Shutyaev and W.A. Lahoz, Dordrecht, Kluwer. (NATO science series: IV: Earth and Environmental Sciences, 26). pp. 21-35.
- Talagrand, O. (2003b) A posteriori validation of assimilation algorithms. In: *Data assimilation for the earth system*. Ed. by R. Swinbank, V. Shutyaev and W.A. Lahoz, Dordrecht, Kluwer. (NATO science series: IV: Earth and Environmental Sciences, 26). pp. 85-96.
- Talagrand, O. (2010a) Variational assimilation. In: *Data assimilation: Making sense of observations*. Eds, W.A. Lahoz, B. Khattatov and R. Ménard. Springer, due March 2010.
- Talagrand, O. (2010b) Evaluation of assimilation algorithms. In: *Data assimilation: Making sense of observations*. Eds, W.A. Lahoz, B. Khattatov and R. Ménard. Springer, due March 2010.
- Tippett, M.K., Anderson, J.L., Bishop, C.H., Hamill, T.M. and Whitaker, J.S. (2003) Ensemble square root filters. *Mon. Weather Rev.*, 131, 1485–1490.
- Trenberth, K.E. (2008) Exploiting and evaluating models with observations. WCRP Modelling Summit, ECMWF, UK, May 2008. URL: <http://www.ecmwf.int/newsevents/meetings/workshops/2008/ModellingSummit/presentations/Trenberth.pdf>
- van den Hurk, B.J.J., Jia, L., Jacobs, C., Menenti, M., and Li, Z. –L. (2002) Assimilation of land surface temperature data from ATSR in an NWP environment – a case study. *Int. J. Rem. Sens.*, 23, 5193-5209.
- van Leeuwen, P.J. (2003) A variance-minimizing filter for large-scale applications. *Mon. Weather Rev.*, 131, 2071–2084.

- Verhoest, N.E.C., De Baets, B., Mattia, F., Satalino, G., Lucau, C. and Defourny, P. (2007) A possibilistic approach to soil moisture retrieval from ERS synthetic aperture radar backscattering under soil roughness uncertainty. *Water Resour. Res.*, 43, WR005295, doi:10.1029/2006WR005295.
- Weerts, A.H. and El Serafy, G.Y.H. (2006) Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resour. Res.*, 42, W09403, doi:10.1029/2005WR004093.
- Whitaker, J.S., Hamill, T.M., Wei, X., Song, Y. and Toth, Z. (2008) Ensemble data assimilation with the NCEP global forecast system. *Mon. Weather Rev.*, 136, 463–482.
- Zupanski, M., Navon, I.M. and Zupanski, D. (2008) The Maximum Likelihood Ensemble Filter as a non-differentiable minimization algorithm. *Quart. J. Roy. Meteorol. Soc.*, 134, 1039–1050.

Appendix A

Practical documentation of the land surface analysis scheme based on an Ensemble Kalman filter within SURFEX

Practical documentation of the land surface analysis scheme based on an Ensemble Kalman filter within SURFEX

Sam-Erik Walker
Norwegian Institute for Air Research - NILU
20 October 2009

1 Introduction

The present description is based on an off-line version of SURFEX v4.8 that runs on a PC. One assumes that this version is currently running on your computer, if not, the first step is to install such a version before trying to use the SURFEX Ensemble Kalman filter Land Data Assimilation System, LDAS (SURFEX-EnKF LDAS).

2 Source code - creation of the binary

To get the source code, contact Sam-Erik Walker at NILU (sew@nilu.no). The source code will be provided to you in a tar file `SURFEX-ENKF-SRC.tar`. You should untar the directories `ENKFASSIM` and `MYSRC` under the directory `SURFEX_EXPORT/src`. Once this is done you will have the following files in the directory `ENKFASSIM`:

- `enkfassim.f90`: Main program that performs the various steps of the assimilation: definition of ensemble members by initial perturbed states, reading of fields from SURFEX outputs, writing out of fields necessary for the analysis and, finally, the surface analysis;
- `ensb_m.f90`: Module containing the Ensemble Kalman filter and Particle filter subroutines used by the main program;
- `rand_m.f90`: Module containing various subroutines handling the pseudo random number generation necessary to set up the ensemble;
- `util_m.f90`: Module containing other utility routines used by the system;
- `mkl_vsl.f90`: Interface file to the Intel Math Kernel Library (MKL) Vector Statistical Library (VSL);
- `trans_chaine.f90`: Transformation of an integer into a character;
- `get_file_name.f90`: Gets the name of files for the current assimilation window.

In order to compile these routines and to get an executable `ENKFASSIM`, the file `Makefile.SURFEX.mk` (provided in the tar file) contains the following sequence of instructions:

```
Source ENKFASSIM
DIR_ENKFASSIM += ENKFASSIM
ifdef DIR_ENKFASSIM
DIR_SOURCE += DIR_ENKFASSIM
endif
```

In the variable `PROG_LIST` defining the various main programs to be generated, `ENKFASSIM` has been added. You can then type "make" in order to generate the LDAS

executable, to be located in the directory `SURFEX_EXPORT/src/exe` (which is also where the other executables: `PGD`, `PREP` and `OFFLINE` are).

The additional files in the directory `MYSRC` correspond to additional features: one is an interpolation of the atmospheric forcing that conserves the accumulated fluxes; another concerns the diagnostic of relative humidity for each individual patch (only required if you use the ISBA version with tiles); the last one is a modified definition of an array in order to read the forcing data files in binary format.

3 The EnKF scheme

The tar file `SURFEX-ENKF.tar.gz` contains a sample of all the required data and scripts to run the SURFEX-EnKF LDAS. First, you need to have all the required data to run a “standard” SURFEX integration: a file of *initial conditions* (e.g. `PREP.lfi` if you work with the LFI format) as well as a set of *forcing data* (e.g. `Forc_TA_YYYYMMDD_r12.txt` and `Params_config_YYYYMMDD_r12.txt` if you work with an ASCII format). If you want to run the LDAS over a long period of time, the forcing should be split according to the length of your assimilation window and not according to the actual duration of the time period. Therefore, if you have already run an off-line integration without data assimilation (called an “open loop” run), you should redo this exercise by splitting the forcing data set into a number of files corresponding to the duration of your integration divided by the length of the assimilation window (with the same units for time in all cases). You should set the logical `LRESTART` to `TRUE` and copy the output file `SURFOUT.lfi` from a given SURFEX integration to define the input file `PREP.lfi` for the next integration (see the example for the script `run_enkf.sh`).

In addition to the initial conditions and forcing files, you need *observation files*. Currently, these files are written in ASCII and observations have been interpolated from the raw data onto the model grid. There is one file per assimilation window that contains all types of observations that are located around the analysis time (i.e., the end of the assimilation window). Note that for asynoptic data there is a mismatch between model and observation times. The generation of this single file needs some pre-processing (this strategy could be revised in the future both in terms of data format and content). When an observation is missing at a given model grid point, it is set to 999.0 (this is the default for undefined values within the SURFEX system).

4 The namelist

The standard namelist of SURFEX (`OPTIONS.nam`) has to be complemented by options related to the EnKF LDAS. The extra namelist options related to the EnKF LDAS are as follows (as set for you in the example provided):

```
&NAM_IO_ENKFASSIM
LPRT = F,
LPRF = F,
LSIM = F
/
&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
/
&NAM_VAR
IVAR       = 1,
NVAR       = 1,
XVAR_M(1)  = 'WG2',
XVAR_M(2)  = 'WG1',
XVAR_M(3)  = 'TG2',
XVAR_M(4)  = 'TG1',
PREFIX_M(1) = 'X_Y_WG2 (m3/m3)
',
PREFIX_M(2) = 'X_Y_WG1 (m3/m3)
',
PREFIX_M(3) = 'X_Y_TG2 (m3/m3)
',
PREFIX_M(4) = 'X_Y_TG1 (m3/m3)
',
XSIGMA_M(1) = 0.1,
XSIGMA_M(2) = 0.1,
XSIGMA_M(3) = 2.0,
XSIGMA_M(4) = 2.0,
TPRT_M(1)  = 0.01,
TPRT_M(2)  = 0.01,
TPRT_M(3)  = 0.01,
TPRT_M(4)  = 0.01,
INCV(1)    = 1,
INCV(2)    = 0,
INCV(3)    = 0,
INCV(4)    = 0
/
&NAM_ENKF
IENS = 1,
NENS = 5,
ENKFM = 1
/

```

Table 1 below gives an overview and description of all the different namelist parameters.

Table 1. Description of variables in the namelist `OPTIONS.nam` for the blocks relative to the SURFEX-ENKF LDAS. The elements with a star (*) should be kept at their value in bold – note their actual values are defined by the script `run_enkf.sh`.

Namelist block	Variable	Type	Description
NAM_IO_ENKFASSIM	LPRT*	F T	Perform analysis Perturb control variables for ensemble
	LPRF*	F T	Perform analysis Perturb forcing data for ensemble (currently not used)
	LSIM*	F T	Perform analysis Write state vectors and simulated observations

NAM_OBS	NOBSTYPE	Integer	Number of possible observation types. <i>This value must be consistent with the obs. file</i>
	YERROROBS(1)	Real	Observation error for T_{2m} in K
	YERROROBS(2)	Real	Observation error for RH_{2m} (no unit)
	YERROROBS(3)	Real	Observation error for w_g (fraction of SWI – soil wetness index)
	INCO(i)	Integer	1 if observation type included 0 if observation type excluded
NAM_VAR	IVAR*	1	Control variable of interest
	NVAR*	1	Number of control variables
	XVAR_M(i)	Character	Control variable identifier in PREP file
	PREFIX_M(i)	Character	Control variable prefix in PREP.txt file
	XSIGMA_M(1)	Real	(Initial) BG (background) error for w_2 (fraction of SWI) (currently not used)
	XSIGMA_M(2)	Real	(Initial) BG error for w_g (fraction of SWI) (currently not used)
	XSIGMA_M(3)	Real	(Initial) BG error for T_2 (K) (currently not used)
	XSIGMA_M(4)	Real	(Initial) BG error for T_s (K) (currently not used)
	TPRT_M(1)	Real	SDev (standard deviation) of perturbation of $\log w_2$ for ensemble
	TPRT_M(2)	Real	SDev of perturbation of $\log w_g$ for ensemble
	TPRT_M(3)	Real	SDev of perturbation of $\log T_2$ for ensemble
	TPRT_M(4)	Real	SDev of perturbation of $\log T_s$ for ensemble
	INCV(i)	Integer	1 if control variable number i included 0 if control variable number i excluded
NAM_ENKF	IENS	Integer	Ensemble member of interest (currently not used)
	NENS	Integer	Number of ensemble members
	ENKFM	Integer	Type of Ensemble Kalman filter method 0 = No assimilation performed 1 = Ensemble square-root Kalman filter, ENSRKF (Sakov and Oke 2008a) 2 = Deterministic ensemble Kalman filter, DENSKF (Sakov and Oke 2008b) 3 = Importance sampling and re-sampling particle filter SIR 4 = Regularized particle filter RPF 5 = Auxiliary PF, ASIR (currently not used)

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: (i) The

surface temperature, T_g (TG1); (ii) the mean surface temperature, T_2 (TG2); (iii) the superficial volumetric water content, w_g (WG1); and (iv) the mean volumetric water content in the root-zone, w_2 (WG2). The choice of the control variables is done by setting the corresponding element of the array `INCV` to one. For the chosen variables, the corresponding value `TPRT` must also be set. This value denotes the standard deviation of a random Gaussian variable used to perturb the logarithm of the initial value of this prognostic variable. The mean value of these random perturbations is assumed to be zero. The EnKF could also be run with the activation of the patches, which means that in such circumstances the analysis of the prognostic variables will be done separately for each patch (see Section 9 in Appendix A).

Regarding the observations, three observation types are considered: (i) screen level temperature, (ii) relative humidity and (iii) superficial soil moisture content. As for the control variables, the elements of the array `INCO` control which type of observation one wants to assimilate.

The number of ensemble members to be used in the Ensemble Kalman filter is selected by choosing a value for `NENS`. For example by choosing `NENS = 100` you will run the system with 100 ensemble members. The type of Ensemble Kalman filter method is selected by setting the parameter `ENKFM` according to the possibilities as listed in Table 1. In addition to the two main Ensemble Kalman filter methods, two Particle filters have also been added to the list of ensemble methods (a third is planned). Currently the value `IENS` is not used. It may be used in the future in order to perform an operation on a given ensemble member.

5 Run script

You have a script `run_enkf.sh` (in `SURFEX-ENKF/rundir` from the tar file `SURFEX-ENKF.tar.gz`) that allows you to run the EnKF over a specified time period. This script is the main driver of the assimilation. It does the looping over assimilation windows, gets the required data, stores outputs, creates temporary files, cleans directories, etc. It operates in several steps (see the flowchart in Fig. A1):

- **Step 0:** Call `ENKFASSIM` in order to create perturbed initial conditions. This option is triggered by the logical `LPRT = T` in the namelist `&NAM_IO_VARASSIM`. A new ensemble member, i.e., a perturbed file of initial conditions (`PREP_ENSB_N.lfi`), is created, where `N` is the ensemble member number, and where `N = 1, ..., NENS`, with `NENS` the number of ensemble members;
- **Step 1:** Run `SURFEX` (name `OFFLINE`) with the current ensemble member (perturbed initial conditions);
- **Step 2:** Call `ENKFASSIM` in order to store the ensemble member related data, i.e., the perturbed evolved prognostic variables and simulated observations in temporary ASCII files `MDSIMU` and `OBSIMU`. These values are read from the output file generated during the previous step. This option is triggered by the logical `LSIM = T` in the namelist `&NAM_IO_VARASSIM`;
- **Step 3:** Redo steps 0 to 2 for each ensemble member in turn;
- **Step 4:** Calls `ENKFASSIM` in order to perform the soil analysis: the corresponding switches are `LPRT = F` and `LSIM = F`. Store the analysis for the model state for the next assimilation cycle. Go to step 0 until the maximum number of assimilation cycles is reached. During this step, the following instructions are performed:
 - Read observations and perform a bias correction if required;

- Read model state and simulated observations from each ensemble member run;
- Compute the covariance matrix R of observation errors;
- Perform the analysis and store the result in `PREP_ENSB_N.lfi` files (for the next cycle).

In Fig. A.1 a flowchart of the SURFEX-ENKF LDAS is shown, corresponding to the various steps of the script `run_enkf.sh`.

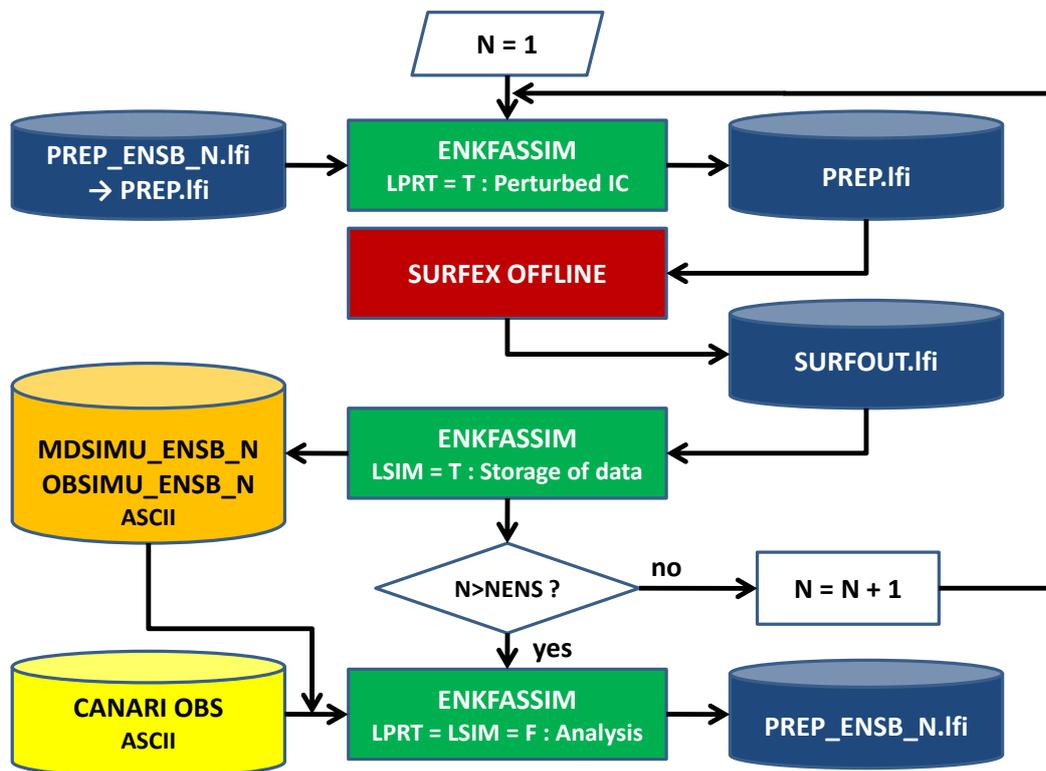


Figure A1. Flowchart of the EnKF-SURFEX LDAS (corresponding to the various steps of the script `run_enkf.sh`). `NENS` corresponds to the number of ensemble members.

6 Management of dates

The dates defined as `YYYYMMDDHH` are evolved in time using the command `smsdate` which is a script that uses an executable `decdate` generated from the C program `decdate.c` using the command:

```
gcc -o decdate decdate.c
```

The script and the C program are available in the directory `UTILITY` of the tar file `SURFEX-ENKF-SRC.tar`. If other tools are available in your computing environment you can use them accordingly.

7 Directory structure

A number of directories should be created and provided:

- **repforcing**: Directory where the forcing data are stored (sample for one day provided in ASCII);
- **repreults**: Directory where the results will be stored;
- **reprun**: Defines the working directory in the script (script `run_enkf.sh` provided);
- **repobs**: Directory where the observations are stored (sample for one day provided in ASCII);
- **repname1**: Directory where the namelist is located (namelist `OPTIONS.nam` provided);
- **repanalyse**: Directory where the initial conditions are stored (`PREP.lfi`) – currently provided for the ALADIN-France domain on 01 July 2006 at 0000 UTC;
- **repbin**: Directory where the binary files to execute SURFEX and the EnKF are located.

This structure has been created for you in the example provided in `SURFEX-ENKF.tar.gz`. You will find in this tar file the script `run_enkf.sh` and a namelist `OPTIONS.nam`. Once you have created the executables `OFFLINE` and `ENKFASSIM`, the content of `SURFEX-ENKF.tar.gz` should allow you to run one day of EnKF assimilation of screen-level parameters every 6 hours over the ALADIN-France domain (see above).

8 Acknowledgments

The author is indebted to Jean-François Mahfouf for providing the off-line version of SURFEX v4.8 and the SURFEX-EKF (Extended Kalman filter) code, scripts and documentation on which the current SURFEX-EnKF code, scripts and documentation are built. We also gratefully acknowledge the help from Met.no in Oslo, Norway, for providing us with computing facilities and help in setting up, compiling and running the SURFEX-EKF and SURFEX-EnKF packages on their Linux system. Thanks to Viel Ødegaard, Mariken Homleid and, last but not least, Trygve Aspelien for their efforts. Recently, the system has also been installed at NILU. Many thanks to Dyre Dammann for his efforts in getting the SURFEX-EnKF compiled and run locally at NILU and sorting out many difficulties during his work at NILU in June and July 2009.

9 Appendix: On the use of the EnKF with the SURFEX “patch” option

9.1 Introduction

An EnKF surface analysis scheme has been coded within SURFEX. The first version was not designed for the “patch” approach of the ISBA scheme. It has been recently extended to include such an option, which is compulsory when considering the ISBA-Ags scheme. The “patch” approach is similar to the “mosaic” land surface model of Koster and Suarez (1992) where the `NATURE` tile within a grid box is divided into a number of independent patches, each having its own set of prognostic variables, and surface energy and water balances. On the other hand, the forcing level for the fluxes and the meteorological variables is assumed to be identical for each individual patch. When

aggregated values are needed (in particular to be passed to the atmospheric model) a simple weighted average of each tile parameter is computed as:

$$\bar{x} = \sum_{k=1}^M \alpha^k x^k$$

Where α^k is the fraction occupied by patch k within the NATURE tile, and x^k is the value of the parameter computed over this specific patch. Currently, the number of patches, M , is set to 12.

9.2 Ensemble Kalman filter without patches

With only one patch, the dimension of the control vector x is equal to the number of prognostic variables to be initialized N_x (the analysis problem is solved independently for each individual model grid point). The (non-linear) observation operator H projects this vector onto the observation space vector y :

$$y = H(x)$$

This vector is then compared to the actual observation vector y_o to produce the innovation vector $y_o - y$. The dimension of the observation vector is equal to the number of independent observations to be assimilated, N_y . The observations are interpolated onto the model grid before analysis, which means that the observation operator does not include any spatial interpolation; this part is done in an independent pre-processing of the data.

The two Ensemble Kalman filter methods implemented here use an observation matrix free method to solve the Kalman filter equations. Thus, the computation of the Kalman gain and the new analysed control vector is done without the need to calculate the Jacobian matrix of the observation operator H . This also applies to the Particle filter methods available.

9.3 Ensemble Kalman filter with patches

With M patches, the dimension of the control vector x is extended to $N_x \cdot M$. On the other hand, the number of observations is still equal to N_y . The model counterpart of the observation y_o is assumed to be the average of the corresponding value y^k for each patch k :

$$\bar{y} = \sum_{k=1}^M y^k$$

Therefore, the innovation vector becomes $y_o - \bar{y}$.

For the computation of the Kalman gain, the dimension of the background error covariance matrix \mathbf{B} has to be increased to the size $N_x M \times N_x M$, whereas the observation error covariance matrix \mathbf{R} keeps the same size $N_y \times N_y$. For each patch k there is an observation operator H providing the simulated observation y^k from the control vector x^k :

$$y^k = H(x^k)$$

This relation states that the simulated observation over the patch k only depends upon the control vector over the same patch. The actual observation operator combines the above relation with the spatial averaging over the patches:

$$\bar{y} = \sum_{k=1}^M \alpha^k y^k = \sum_{k=1}^M \alpha^k H(x^k)$$

Again, since the Ensemble Kalman filter methods implemented here use an observation matrix free method to solve the Kalman filter equations, the computation of the Kalman gain and new analysed control vector is done without the need to calculate the Jacobian matrix of the observation operator H . This also applies to the Particle filter methods available.

9.4 Conclusion

In Section 9 of Appendix A we have shown that it is possible to extend the EnKF coded within SURFEX for one patch to a set of M patches. The analysis equation and methodology for integrating the individual ensemble members are kept unchanged. In particular, the number of perturbed integrations to be performed remains equal to the number of ensemble members N_e , and not $N_e \cdot M$. This comes from the fact that the simulated observations are a linear combination of independent results from each patch, therefore, they can be perturbed simultaneously. In practice, the control-vector needs to be enlarged from N_x to $N_x \cdot M$ (and, accordingly, the \mathbf{B} matrices to $N_x M \times N_x M$). The simulated observations need to be computed from the weighted contribution of each patch. Since an observation matrix free method is used to solve the Kalman filter analysis equations, no calculations of Jacobian matrices are necessary. This also applies to the Particle filter methods. The changes have been coded in the most recent version of the SURFEX-EnKF and are available from the author.

References

- Koster, R. And M. Suarez, 1992: Modeling the Land Surface Boundary in Climate Models as a Composite of Independent Vegetation Stands, *J. Geophys. Res.*, 97(D3), 2697-2715.
- Sakov, P. and Oke, P.R., 2008a: Implications of the form of the ensemble transformation in the ensemble square root filters, *Monthly Weather Review*, 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.

Appendix B

User's Guide for compiling and running SURFEX-EKF and SURFEX-EnKF

User's Guide for compiling and running

SURFEX-EKF and SURFEX-EnKF

Sam-Erik Walker and Dyre Dammann
Norwegian Institute for Air Research - NILU
20 October 2009

1 Introduction

This document describes how to compile and run the SURFEX-EKF and SURFEX-EnKF land data assimilation systems currently implemented at NILU. It may also be read as a guide for others who wish to set up and run these systems using their own local computing facilities.

2 Computing environment

The SURFEX-EKF and SURFEX-EnKF land data assimilation systems are currently implemented on a Linux based computer at NILU with host name "ulven". In order to access this computer from other Windows based PCs at NILU, open an MS-DOS command prompt window from Windows and type

```
>> telnet fenris
```

and log in with your user name and password. After you have successfully logged in to the computer "fenris", you can log in to "ulven" by issuing the command

```
>> ssh ulven
```

and provide the same password as before.

The SURFEX-EKF and SURFEX-EnKF systems are located in the directories

```
/nilu/inby/b106015/williaml/SURFEX-EKF
```

and

```
nilu/inby/b106015/williaml/SURFEX-EnKF
```

respectively.

In order to obtain faster access to these two folders you may wish to include the following two aliases in your C-shell and BASH-shell system files `.cshrc` and `.bashrc`, typically located in your home directory:

```
alias goto_ekf 'cd /nilu/inby/b106015/williaml/SURFEX-  
EKF/EXPORT_v4.8'
```

```
alias goto_enkf 'cd /nilu/inby/b106015/williaml/SURFEX-  
EnKF/EXPORT_v4.8'
```

We may then proceed directly to the two specific directories by issuing the commands

```
>> goto_ekf
```

or

```
>> goto_enkf
```

Typically you will only compile or run one of these two systems at any given time.

3 SURFEX-EKF

3.1 How to compile SURFEX-EKF

3.1.1 How to set up the computer for compiling

Add the following line to the file `.bashrc`, in order to set the environmental variables associated with the Intel Math Kernel Library (Intel MKL):

```
./opt/intel/Compiler/11.0/081/mkl/tools/environment/mklvars32.sh
```

The Intel MKL is only used by the SURFEX-EnKF system.

The following three lines must also be added in order to define the SURFEX-version and to set the SURFEX environmental variables:

```
export SURFEX_EXPORT=/nilu/inby/b106015/williaml/SURFEX-
EKF/EXPORT_v4.8
```

```
source $SURFEX_EXPORT/conf/profile_surfex
```

```
export
PATH=$PATH:$SURFEX_EXPORT/src/LIB/netcdf3.6.1/bin/:$SURFEX_EXPORT/src/exe/
```

3.1.2 How to compile

In order to compile, we first need to move to the directory program source directory. Type:

```
>> goto_ekf
>> cd src
```

Then we need to be in a BASH shell. Type:

```
>> bash
```

Then we compile by typing:

```
: make
```

The model will then compile using the file `Makefile.SURFEX.mk` in the current directory.

If the compilation was successful the following main program executables should exist in the sub directory `exe`:

```
PGD PREP OFFLINE SXPOST OI_MAIN VARASSIM
```

The make command will generally only compile changes made to the source code of the system since the last compilation. In order to compile all of the source code of the system completely from scratch, type:

```
: make clean
```

and then

```
: make
```

3.2 How to run SURFEX-EKF

3.2.1 How to set up the computer for running

Before running SURFEX-EKF we need to make a path to the program `smsdate`. We created a path to the directory `UTILITY`, where `smsdate` is located, in the file:

```
\inby\b106015\williaml\SURFEX-EKF\EXPORT_v4.8\conf\profile_SURFEX
```

The following line was added to this file:

```
PATH
$SURFEX_EXPORT/bin/:::SPATH/SURFEX_EXPORT/:UTIL:$surfex_export/src/UTILIT
Y
```

3.2.2 How to set up the program for running

The script file to run SURFEX-EKF is called `run_ekf.sh` and is located in:

```
N:\inby\b106015\williaml\SURFEX-EKF\EXPORT_v4.8\SURFEX_EKF\rundir
```

Open the script `run_ekf.sh` in a text editor.

In this file you have the options of setting the following parameters:

- `expid` – the current run experiment identifier text string;
- `vm` - the number of control variables;
- `l` - the number of time periods.

If you edit this file in a text editor like `WORDPAD`, remember to use the command `dos2unix` on the file in order to convert it to a Linux (UNIX) ASCII format again. For more information, please see Section 5.1 below.

N.B.! The above script file automatically edits the contents of the `OPTIONS.nam` file associated with the SURFEX-EKF system regarding the above three parameters.

You may also need to edit the `OPTIONS.nam` file, in order to set the following parameters:

- `tprt_m` – amount of perturbation for each control variable;
- `nobs` – number of observation types;
- `yerrorobs` – amount of observational error for each type of observation;
- `inco` – include observation type logical variable (0=false,1=true).

After editing the file, please remember to use the command `dos2unix` on the file in order to convert it to a Linux (UNIX) ASCII format again. For more information, please see Section 5.1 below.

A result catalogue needs to be created in the directory:

```
N:\inby\b106015\williaml\SURFEX-
EKF\EXPORT_v4.8\SURFEX_EKF\result
```

with the same name as the `expid` text string as defined in the run script.

3.2.3 How to run the program

The file to run SURFEX-EKF is called `run_ekf.sh` and is located in the directory:

```
N:\inby\b106015\williaml\SURFEX-EKF\EXPORT_v4.8\SURFEX_EKF\rundir
```

Please make sure first that you have started the BASH-shell. You do this by writing

```
>> bash
```

and you will receive a `:` prompt in the terminal window.

To run the model, move to the above directory, and type:

```
: run_ekf.sh
```

3.3 SURFEX-EKF results

The following directory will contain all the results:

```
N:\inby\b106015\williaml\SURFEX-
EKF\EXPORT_v4.8\SURFEX_EKF\result\

```

where `<expid>` is the `expid` text string defined in the run script.

4 SURFEX-EnKF

4.1 Differences between EnKF and EKF

The SURFEX-EnKF directory structure is generally a copy of the SURFEX-EKF directory structure with identical directories, subdirectories and filenames, except for the following four files:

```
SURFEX-EKF/EXPORT_v4.8/SURFEX-EKF/rundir/run_ekf.sh
SURFEX-EKF/EXPORT_v4.8/src/VARASSIM/varassim.f90
SURFEX-EKF/EXPORT_v4.8/SURFEX_EKF/namelist/OPTIONS.nam
SURFEX-EKF/EXPORT_v4.8/src/Makefile.SURFEX.mk
```

which for the SURFEX-EnKF system are replaced by the following modified files:

```
SURFEX-EnKF/EXPORT_v4.8/SURFEX-EnKF/rundir/run_enkf.sh
SURFEX-EnKF/EXPORT_v4.8/src/ENKFASSIM/enkfassim.f90
SURFEX-EnKF/EXPORT_v4.8/SURFEX_EnKF/namelist/OPTIONS.nam
SURFEX-EnKF/EXPORT_v4.8/src/Makefile.SURFEX.mk
```

The changes that have been made to `run_enkf.sh`, `enkfassim.f90` and `OPTIONS.nam` are extensive and will not be described here. However, regarding the file `Makefile.SURFEX.mk` the following changes have been made:

The code block:

```
DIR_VARASSIM += VARASSIM
#CPPFLAGS_VARASSIM=
#
ifdef DIR_VARASSIM
DIR_SOURCE += $(DIR_VARASSIM)
CPPFLAGS += $(CPPFLAGS_VARASSIM)
```

has been replaced by:

```
DIR_ENKFASSIM += ENKFASSIM
DIR_MKL += $MKLPATH
CPPFLAGS_ENKFASSIM=-Dmkl -Dlinux
#
ifdef DIR_ENKFASSIM
DIR_SOURCE += $(DIR_ENKFASSIM)
CPPFLAGS += $(CPPFLAGS_ENKFASSIM)
INC += $(DIR_MKL)/include
LIBS += $(DIR_MKL)/lib/libmkl_intel.a
LIBS += $(DIR_MKL)/lib/libmkl_sequential.a
LIBS += $(DIR_MKL)/lib/libmkl_core.a
LIBS += -lpthread
Endif
```

Also, the line:

```
PROG_LIST = PGD PREP OFFLINE SXPOST OI_MAIN VARASSIM
```

has been replaced by

```
PROG_LIST = PGD PREP OFFLINE SXPOST OI_MAIN ENKFASSIM
```

4.2 How to compile SURFEX-EnKF

4.2.1 How to set up the computer for compiling

Add the following line to the file `.bashrc`, in order to set the environmental variables associated with the Intel Math Kernel Library (Intel MKL):

```
./opt/intel/Compiler/11.0/081/mkl/tools/environment/mklvars32.sh
```

The following three lines must also be added in order to define the SURFEX-version and to set the SURFEX environmental variables:

```
export SURFEX_EXPORT=/nilu/inby/b106015/williaml/SURFEX-
EnKF/EXPORT_v4.8

source $SURFEX_EXPORT/conf/profile_surfex

export
PATH=$PATH:$SURFEX_EXPORT/src/LIB/netcdf3.6.1/bin/:$SURFEX_EXPORT/src/exe
/
```

Note that if you have already made the above changes to the `.bashrc` file for the SURFEX-EKF system (Section 3.1.1 in Appendix B), the only change you need to do here is to replace the text string `"SURFEX-EKF"` with `"SURFEX-EnKF"`. You may, alternatively, issue the command

```
>> export SURFEX_EXPORT=/nilu/inby/b106015/williaml/SURFEX-EnKF/EXPORT_v4.8
```

directly in the terminal window, if you want to avoid changing the `.bashrc` file.

4.2.2 How to compile

In order to compile, we first need to move to the program source directory. Type:

```
>> goto_enkf
>> cd src
```

Then we need to be in a BASH shell. Type:

```
>> bash
```

Then we compile by typing:

```
: make
```

The model will then compile using the file `Makefile.SURFEX.mk` in the current directory.

If the compilation was successful the following main program executables should exist in the sub directory `exe`:

```
PGD PREP OFFLINE SXPOST OI_MAIN ENKFASSIM
```

The `make` command will generally only compile changes made to the source code of the system since the last compilation. In order to compile all of the source code of the system completely from scratch, type:

```
: make clean
```

and then

```
: make
```

4.3 How to run SURFEX-EnKF

4.3.1 How to set up the computer for running

Before running SURFEX-EKF we need to make a path to the program `smsdate`. We created a path to the directory `UTILITY`, where `smsdate` is located, in the file:

```
\inby\b106015\williaml\SURFEX-EnKF\EXPORT_v4.8\conf\profile_SURFEX
```

The following line was added to this file:

```
PATH =
$SURFEX_EXPORT/bin/::SPATH/SURFEX_EXPORT/:UTIL:$surfex_export/src/UTILITY
```

4.3.2 How to set up the program for running

The script file to run SURFEX-EnKF is called `run_enkf.sh` and is located in:

```
N:\inby\b106015\williaml\SURFEX-
EnKF\EXPORT_v4.8\SURFEX_EnKF\rundir
```

Open the script `run_enkf.sh` in a text editor.

In this file you have the options of setting the following parameters:

- `expid` – the current run experiment identifier text string;
- `vm` – the number of control variables;
- `l` – the number of time periods;
- `nens` – the number of ensemble members in the Ensemble Kalman filter and Particle filter;
- `enkfm` – the type of ensemble-based data assimilation method;
- `tprt_m` – amount of perturbation as relative standard deviation for each control variable ;
- `nobs` – number of observation types;
- `yerrorobs` – amount of observational error for each type of observation;
- `inco` – include observation type logical variable (0=false,1=true).

If you edit this file in a text editor like WORDPAD, remember to use the command `dos2unix` on the file in order to convert it to a Linux (UNIX) ASCII file again. For more information, please see Section 5.1 in Appendix B.

N.B.! The above script file automatically edits the contents of the `OPTIONS.nam` file associated with the SURFEX-EnKF system regarding the above parameters. However, if for some reason you should wish to edit this file, remember to use the command `dos2unix` on the file in order to convert it to a Linux (UNIX) ASCII file again. For more information, please see Chapter 5.1 in Appendix B.

A result catalogue needs to be created in the directory:

```
N:\inby\b106015\williaml\SURFEX-
EnKF\EXPORT_v4.8\SURFEX_EnKF\result
```

It must be given the same name as the `expid` text string set in the script.

4.3.3 How to run the program

The file to run SURFEX-EnKF is called `run_enkf.sh` and is located in the directory:

```
N:\inby\b106015\williaml\SURFEX-
EnKF\EXPORT_v4.8\SURFEX_EnKF\rundir
```

Please make sure first that you have started the BASH-shell. You do this by writing

```
>> bash
```

and you will receive a `:` prompt in the terminal window.

To run the model, move to the above directory, and type:

```
: run_enkf.sh
```

4.4 SURFEX-EnKF results

4.4.1 Result directory

The following directory will contain all the results:

```
N:\inby\b106015\williaml\SURFEX-
EnKF\EXPORT_v4.8\SURFEX_EnKF\result\

```

where `<expid>` is the `expid` text string defined in the run script.

4.4.2 Missing data

If observation data are missing (i.e., have the value 999.0) they will be assigned the modelled value at the same grid point.

5 Other issues to be aware of

5.1 DOS ASCII and Linux (Unix) ASCII

If the text editor used to make changes to files use DOS ASCII format (for example WORDPAD) you need to convert the files to UNIX ASCII format. You do this by typing:

```
>> dos2unix filename
```

It is highly recommended you use a text editor which preserves the Linux (UNIX) ASCII type of the file, when editing Linux (UNIX) ASCII files from Windows. For example, the EditPad Lite text editor for Windows is one such text editor.

5.2 Sed-command

Avoid changing the number of spaces in the file `OPTIONS.nam`. It can cause problems for the `sed`-command if the number of spaces between values, equal signs (“=”), and values are different in this file and in the script.

5.3 Changing the .bashrc file

If you want to run the SURFEX-EnKF version after having run the SURFEX-EKF version, or vice versa, you must remember to repeat the commands in Section 4.3.1, or the reverse of Section 4.3.1, respectively.

5.4 Empty work catalogue

If an error message shows up right away when running the scripts `run_enkf.sh` or `run_ekf.sh`, saying “can’t erase ‘*’- no such file or directory”, create a file in the work directory and try again.

5.5 Permission denied

If an error message shows up right away when running the script `run_ekf.sh` or `run_enkf.sh`, e.g., “permission denied”, try to use the command `dos2unix` on the file (see also Section 5.1 in Appendix B).

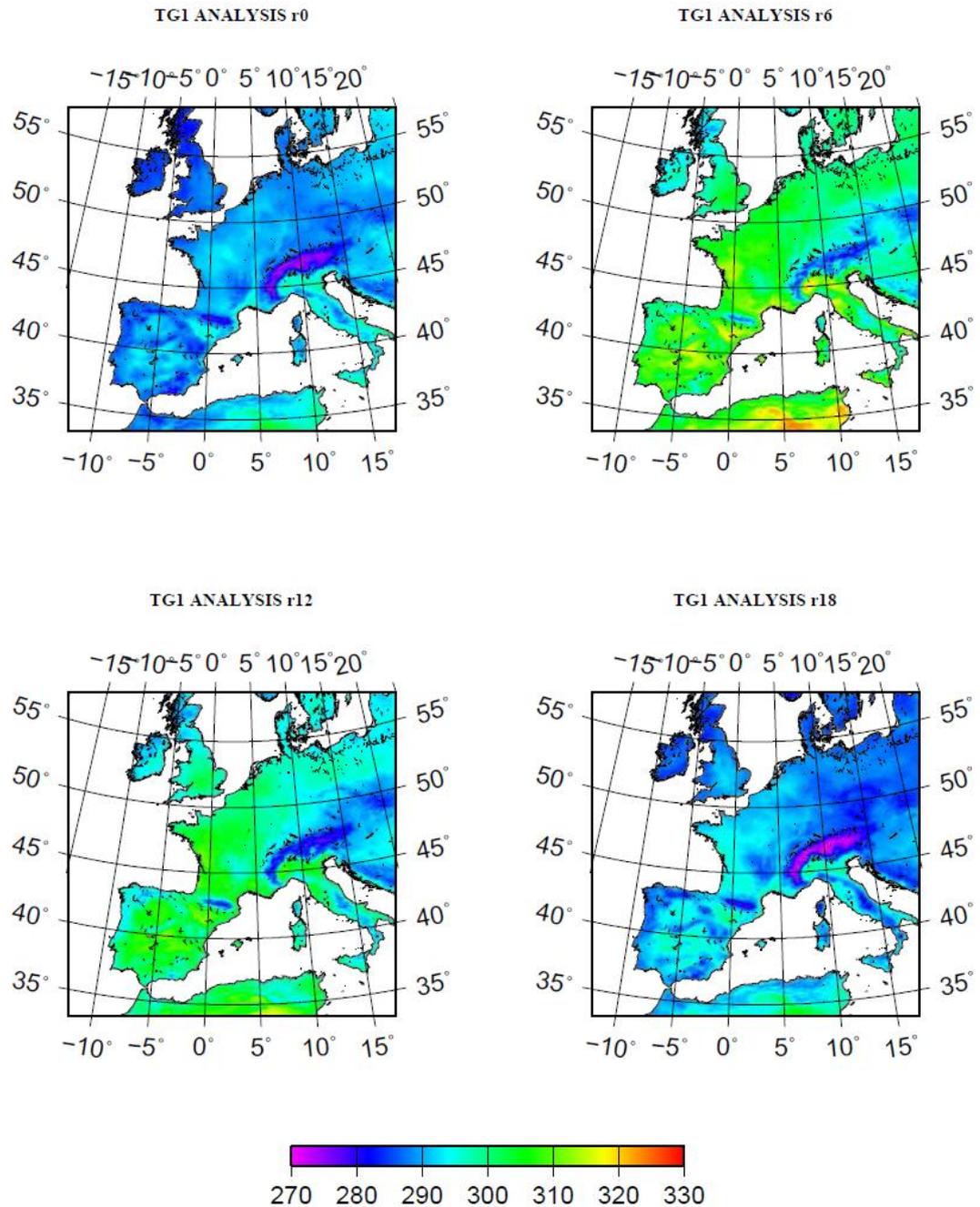
Appendix C

EKF results from Météo-France

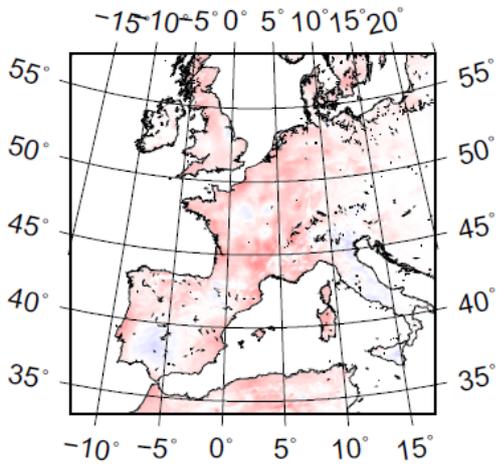
EKF results from Météo-France

Note: units, geographical area and time stamps are as in Sections 3.1-3.8

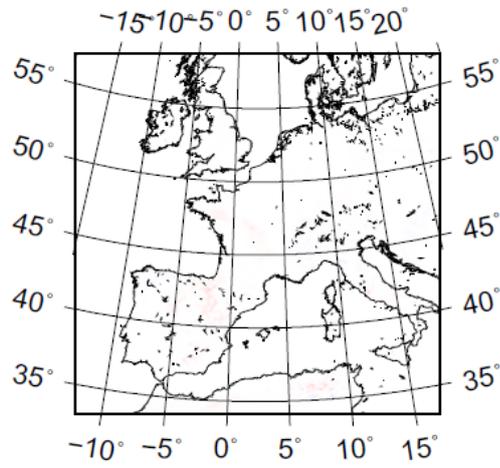
TGI.



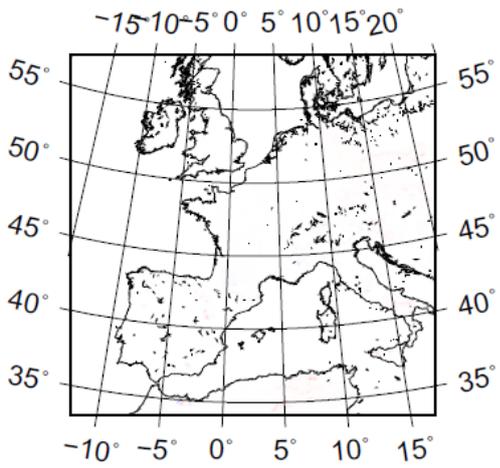
TGI INCREMENTS r0



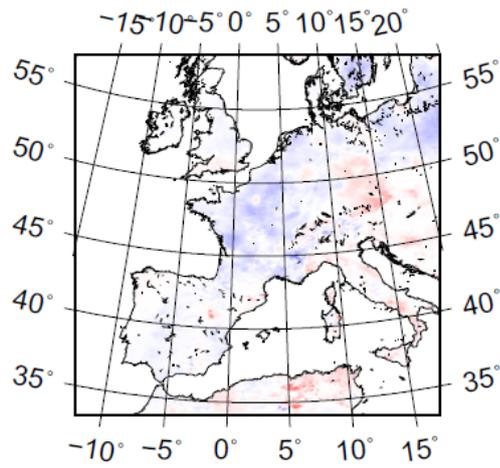
TGI INCREMENTS r6

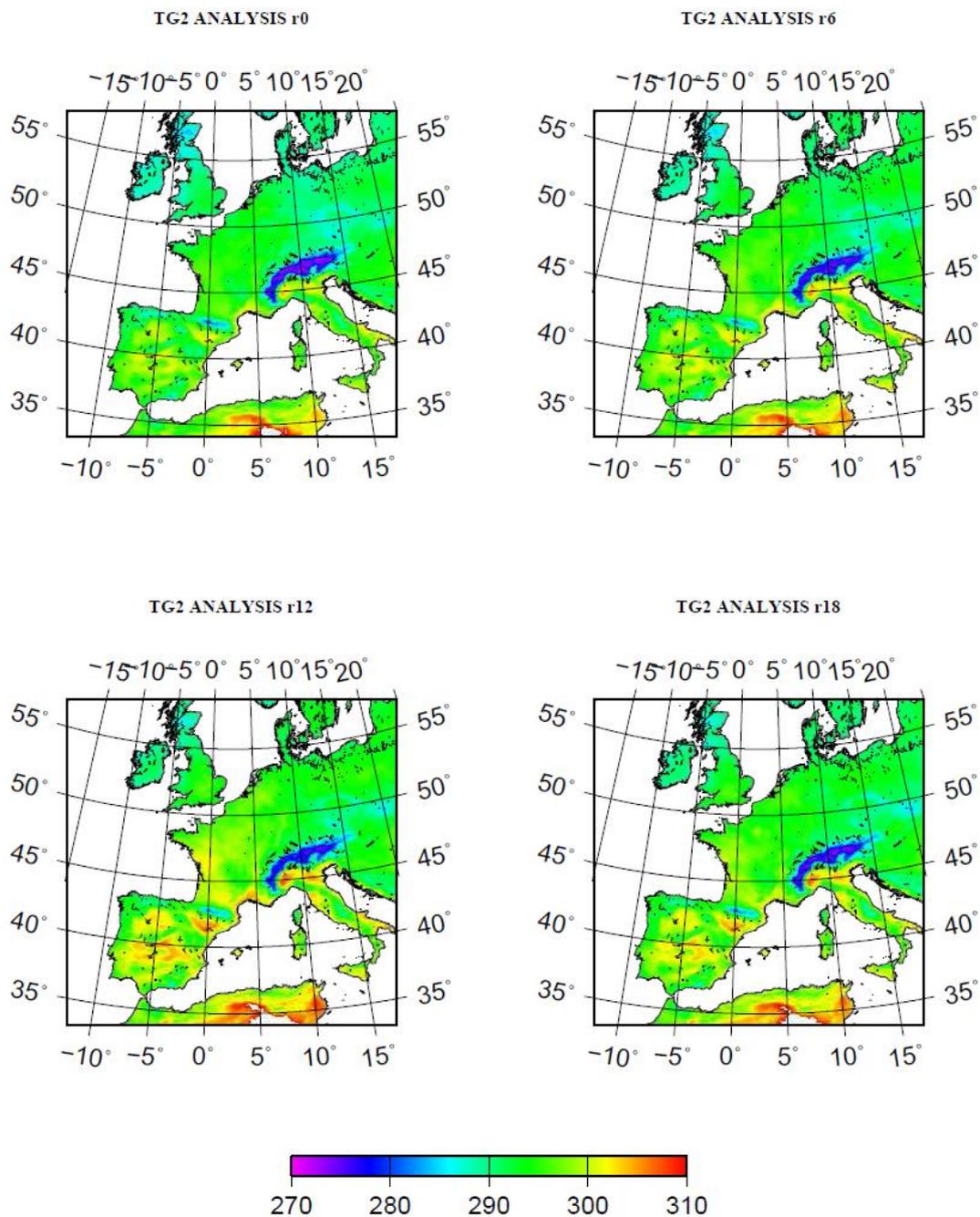


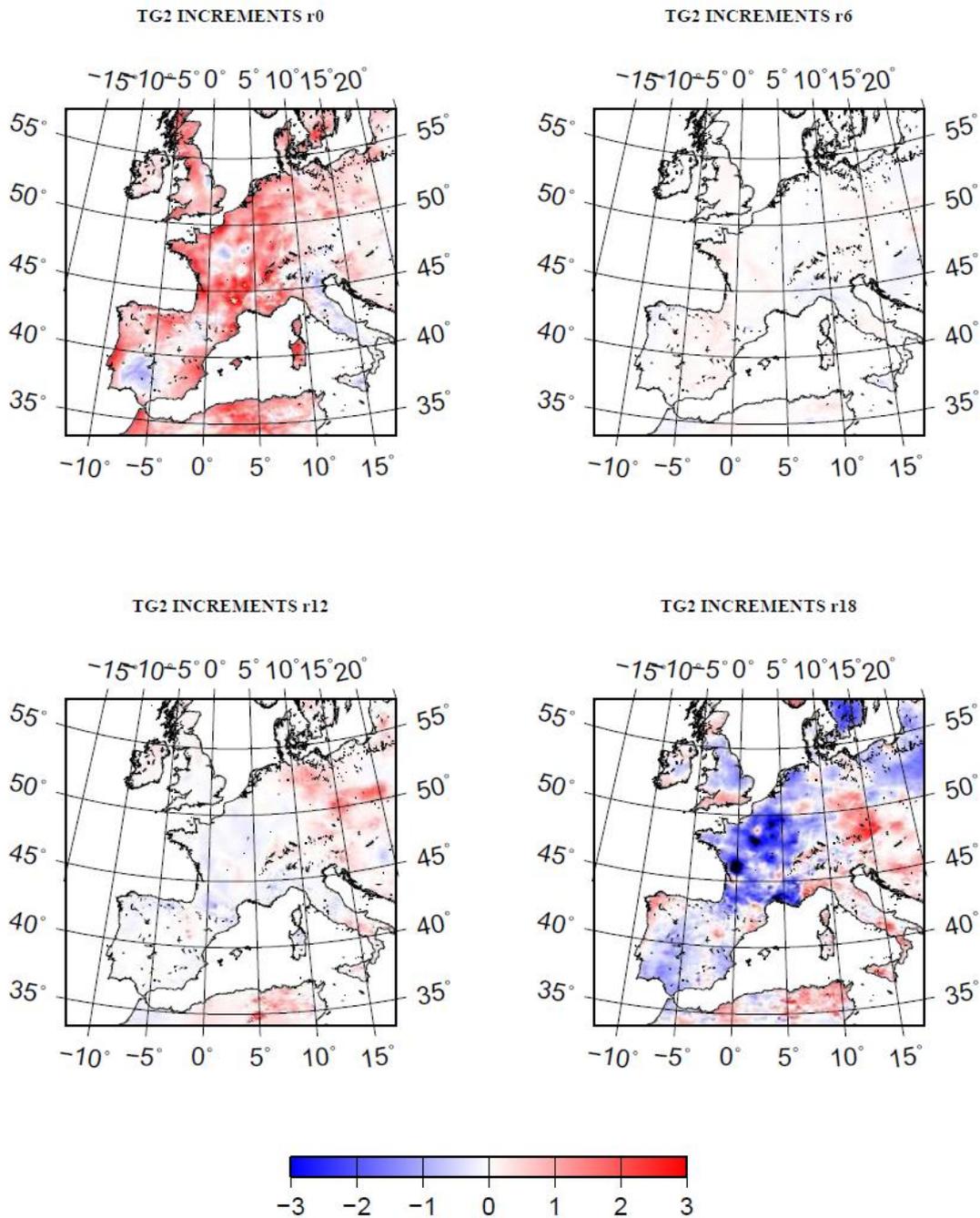
TGI INCREMENTS r12

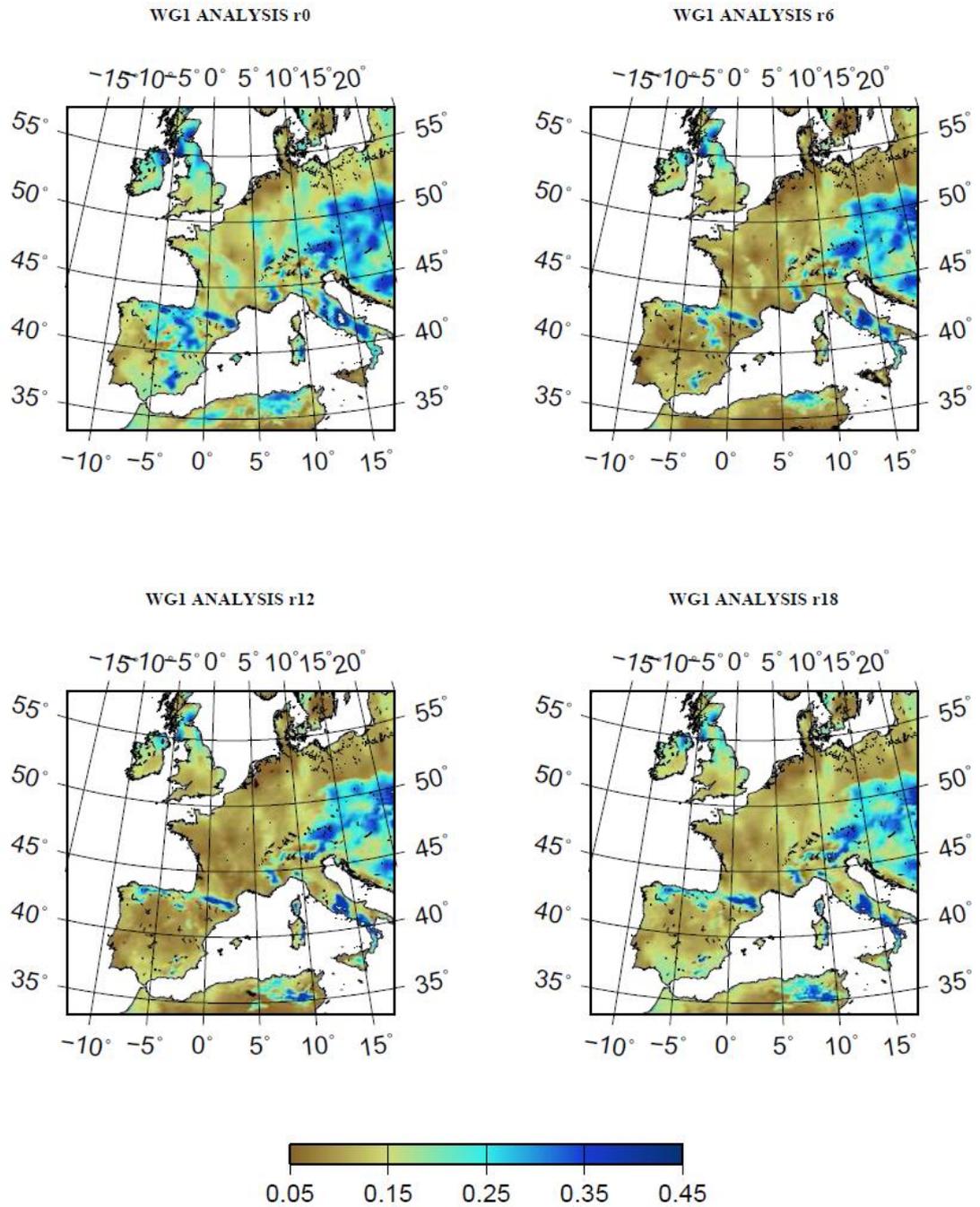


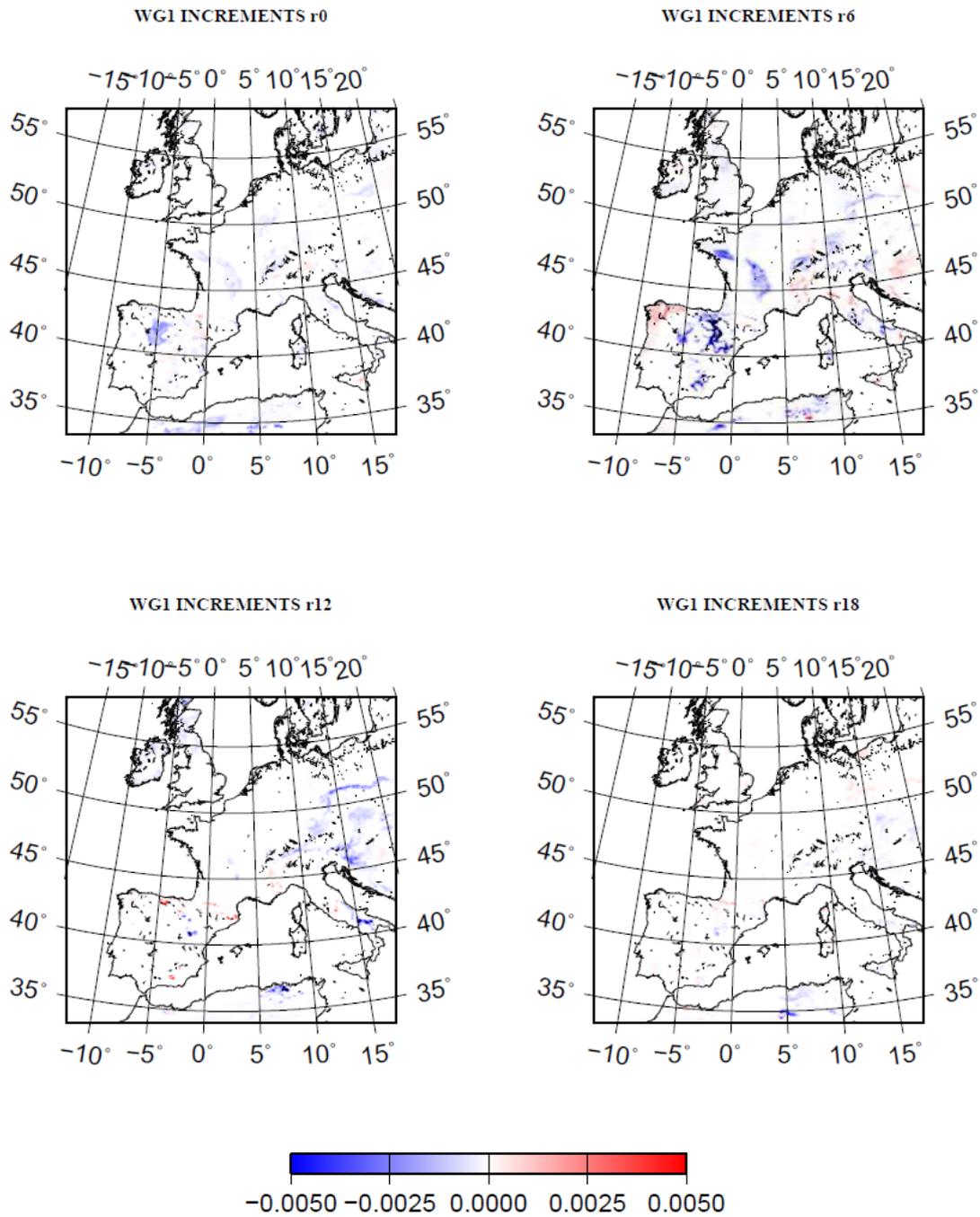
TGI INCREMENTS r18



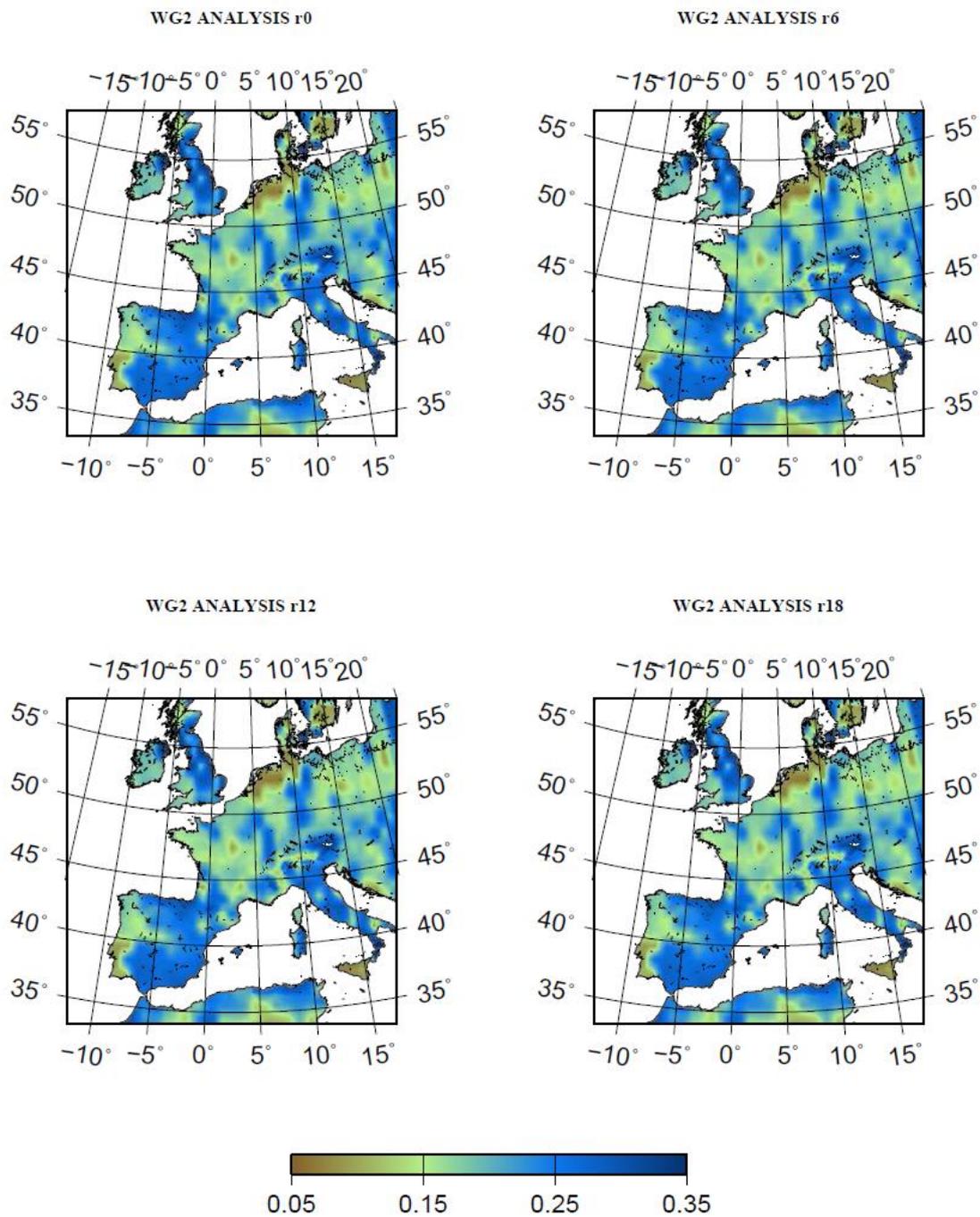
TG2.

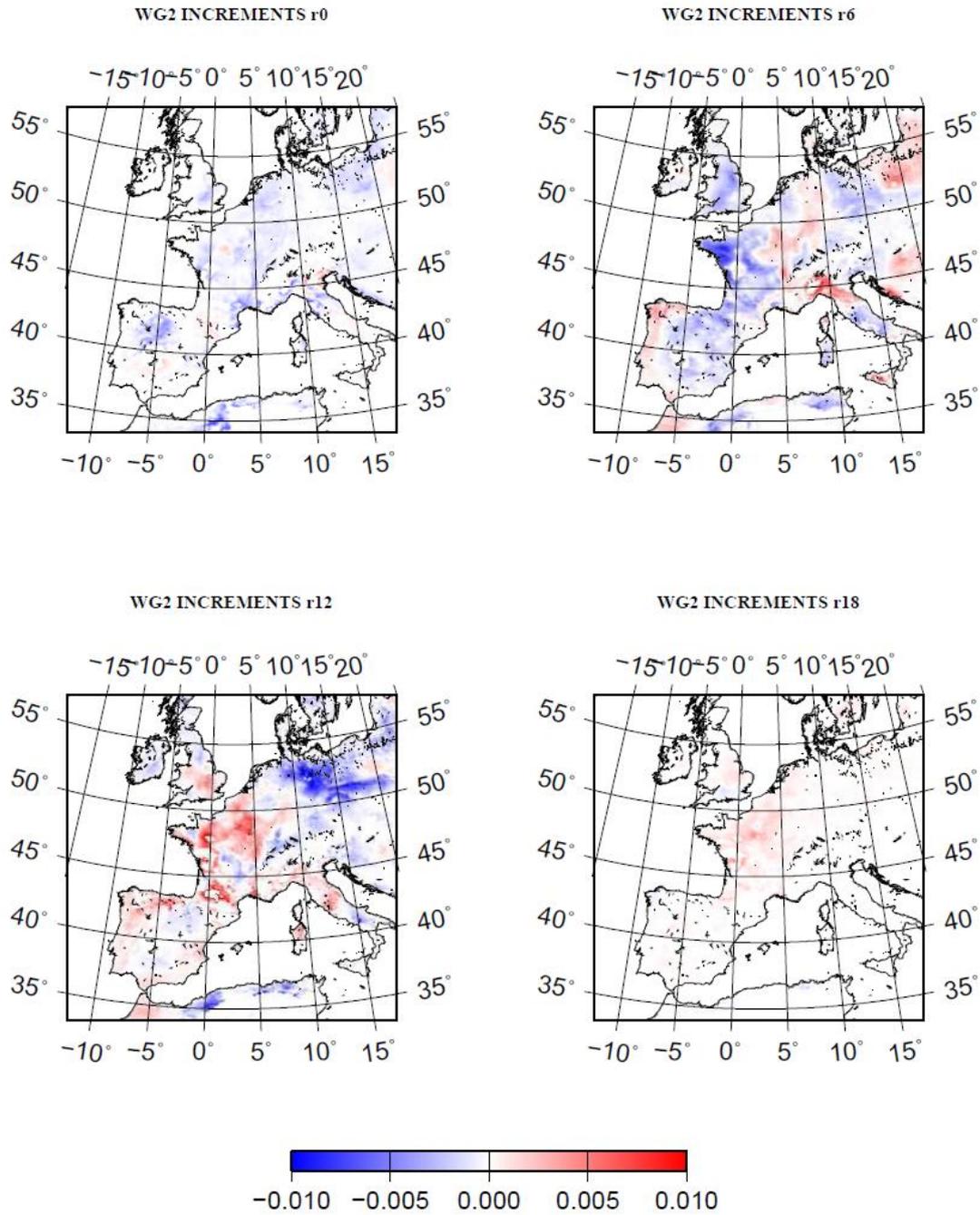


WGI.



WG2.







**Norwegian Institute
for Air Research**

Norwegian Institute for Air Research
P.O. Box 100, N-2027 Kjeller, Norway
*Associated with CIENS and the
Environmental Research Alliance of Norway
ISO certified according to NS-EN ISO 9001*

REPORT SERIES TECHNICAL REPORT	REPORT NO. TR 2/2010	ISBN: 978-82-425-2174-3 (printed) 978-82-425-2175-0 (electronic) ISSN: 0807-7185	
DATE	SIGN.	NO. OF PAGES 84	PRICE NOK 150.-
TITLE The NILU SURFEX-EnKF land data assimilation system		PROJECT LEADER William A. Lahoz	
		NILU PROJECT NO. E-107123	
AUTHOR(S) W.A. Lahoz, S.-E. Walker and D. Dammann		CLASSIFICATION * A	
		CONTRACT REF.	
REPORT PREPARED FOR Norwegian Institute for Air Research			
ABSTRACT Development of the NILU SURFEX-EnKF land data assimilation system is described. The system is being developed in collaboration with the Norwegian Meteorological Institute and Météo-France. It is based on the Météo-France SURFEX land surface model, and uses various variants of the Ensemble Kalman filter and Particle filter method. It also uses the Extended Kalman filter method from Météo-France. The system assimilates land surface temperature and soil moisture; it will be extended to snow variables later. Observations assimilated are from in situ platforms; later, observations from satellite platforms will be assimilated. The main objective for developing this assimilation system is to evaluate observations and models; analyses of land surface variables will be produced. Results will benefit Numerical Weather Prediction agencies, space agencies and the broad scientific community.			
NORWEGIAN TITLE NILUs SURFEX-EnKF land data assimilasjonssystem			
KEYWORDS Data assimilation	Land surface	Earth Observation	
ABSTRACT (in Norwegian) Utvikling av NILU's SURFEX-EnKF land data assimilasjonssystem blir beskrevet. Systemet utvikles i samarbeid med Meteorologiske Institutt (met.no) og Météo-France. Det er basert på SURFEX land overflate modellen fra Météo-France, og bruker ulike varianter av Ensemble Kalman filter og partikkel filter metoder. Det inneholder også Extended Kalman filter metoden fra Météo-France. Systemet assimilerer land overflate temperatur og jord fuktighet, og vil senere utvides til å omfatte snø parametre. Observasjoner som assimileres er fra bakkebaserte målinger, senere vil systemet også omfatte satellittbaserte observasjoner. Hovedhensikten med å utvikle assimilasjonssystemet er for å evaluere observasjoner og modeller, med tilhørende analyse av overflate variabler. Resultatene vil være til nytte i forbindelse med numerisk værvarsling, jordobservasjoner med mer.			

* Classification
A Unclassified (can be ordered from NILU)
B Restricted distribution
C Classified (not to be distributed)

REFERENCE: E-107123
DATE: JANUARY 2010
ISBN: 978-82-425-2174-3 (printed)
978-82-425-2175-0 (electronic)

NILU is an independent, nonprofit institution established in 1969. Through its research NILU increases the understanding of climate change, of the composition of the atmosphere, of air quality and of hazardous substances. Based on its research, NILU markets integrated services and products within analyzing, monitoring and consulting. NILU is concerned with increasing public awareness about climate change and environmental pollution.



Norsk institutt for luftforskning
Norwegian Institute for Air Research