International Conference
on Ensemble Methods in Geophysical Sciences
Toulouse, France, 12-16 November 2012

Oral Programme
(jam-08/11/2012)

(Invited presentations: 30 minutes - contributed presentations: 20 minutes)

Monday 12 November

9.00 Registration. Coffee

9.50 Opening

Session 1. ***Particle Filters and Applications***

Chair: Gérald Desroziers

10.15 Arnaud Doucet (invited, Department of Statistics, University of Oxford, UK)

Block Sampling Proposals for Particle Filters

Whenever using particle filters it is crucial to select carefully the proposal distribution so as to obtain Monte Carlo estimates with acceptable variance. Even the use of the so-called "optimal" proposal does not guarantee this will happen. We propose an original block sampling strategy where, at a given time index, we not only sample the current state variable but allow ourselves to sample again "recent" past state variables in light of the new observation available. For a fixed computational complexity, we show that such a sampling strategy can lead to much lower variance estimates.

10.45 Chris Snyder (invited, National Center for Atmospheric Research, Boulder, USA)

Performance bounds for particle filters using the "optimal" proposal density

Particle filters may be cast in the form of sequential importance sampling and thus allow the choice of a proposal density. This choice is known to be crucial to the performance of the algorithm. We show in the special case that the system in linear and Gaussian that the asymptotic arguments of Snyder et al. (2008) apply directly to the "optimal" proposal density, which minimizes the variance of the particle weights over different realizations of particles drawn from the proposal density. The asymptotic results indicate that the ensemble size needed to avoid degeneracy of the particle-filter update (i.e., the situation in which a single particle receives a weight very close to 1) may be dramatically smaller for the optimal proposal that for the standard proposal, in which new particles are drawn by evolving particles from the previous time under the system dynamics. The degree to which a smaller ensemble is feasible with the optimal proposal depends crucially on the magnitude of the system noise. At the same time, the asymptotic results also demonstrate that the required ensemble size for the optimal proposal will grow exponentially with an appropriately defined measure of the problem size, albeit with a smaller exponent than for the standard proposal.

11.15 Matthias Morzfeld (Lawrence Berkeley National Laboratory, Berkeley, USA), Alexandre Chorin, Robert Miller, Yvette Spitz and **Brad Weir** (Brad Weir: Oregon State University, Corvallis, USA)

Implicit Sampling for Data Assimilation in Geophysics

Realistic models for geophysical processes are often large dimensional and highly non linear. Approximate ensemble techniques for data assimilation, e.g. the ensemble Kalman filter (EnKF), can be implemented for such large models, but can be problematic in cases of strong non linearity. Standard Monte Carlo (MC) methods construct an empirical estimate of the posterior probability density function (pdf) without making linearity or Gaussianity assumptions, however often suffer from sample impoverishment when the state dimension is large and therefore can be inefficient. Much research effort in ensemble-based data assimilation thus focuses on making MC methods more efficient. We present results of the implicit sampling technique, in which we construct an importance density that is informed by the observations. The construction allows us to concentrate the MC sampling on the high-probability regions of the posterior pdf. Because the sampling (and thus the computations) is concentrated on the high-probability region of the posterior pdf, sample impoverishment is ameliorated, so that fewer particles are required and the computational effort remains manageable.

We show that this concentration of particles can be achieved by performing a minimization to identify the regions of high probability of the posterior pdf and that this minimization is similar to the minimizations performed in variational data assimilation. Once the high-probability regions of the state space are identified, samples can be obtained in these regions by choosing a solution of an under-determined set of algebraic equations with a random right-hand-side. The solutions to the algebraic equation define the importance function implicitly. This importance function, and thus each particle path, is directly influenced by the observations. We present efficient implementations of the implicit sampling method where we make use of standard minimization methods.

To demonstrate the efficiency and applicability of our approach, we consider data assimilation for a set of stochastic pde's that is important in geomagnetic data assimilation (this problem has 600 state variables), as well as a model of near-shore flow that has 10,000 state variables. In both cases, implementation of implicit sampling is straightforward and sample impoverishment is significantly reduced. Our method requires on the order of 10 particles when standard Monte Carlo techniques or EnKF require hundreds to thousands for comparable accuracy.

11.35 Emmanuel Cosme (Université Joseph Fourier, Grenoble, France) and Chris Snyder

A deterministic, fully non Gaussian analysis scheme for ensemble filters: the Multivariate Rank Histogram Filter

A deterministic, fully non Gaussian analysis scheme designed for ensemble filters is introduced. It is deterministic, in the sense that no random number generation is necessary, making any observational update reproducible. It is fully non Gaussian because any shape of probability densities can be dealt with. The filter analysis implements a sequential realization procedure (Tarantola, 2005) based on the Rank Histogram Filter approach (Anderson, Monthly Weath. Rev., 138, 4186-4198, 2010). For each direct observation of the system, the observed variable is first updated. Unobserved variables are then updated recursively. The method will be presented and compared with the Ensemble Kalman Filter, the Rank Histogram Filter, and the Particle Filter, through numerical experiments with Lorenz 63 and 95 models in highly nonlinear and non Gaussian settings. Preliminary experiments suggests that the Multivariate Rank Histogram Filter outperforms the other methods, in terms of estimation, with only 100 particles.

11.55 Christoph Bergemann (Deutsches Zentrum für Luft- und Raumfahrt (DLR), Cologne, Germany)

Particle filter based data assimilation into an air quality model

Air pollution is one of the main environmental hazards within many industrialized regions. Accurate analysis and forecast of atmospheric composition is therefore an important challenge for geophysical modelling. The most commonly used operational air pollution models are offline models providing no feedback to meteorology. These models differ from most geophysical models in that their equations are far less dependent on the initial values. In fact trajectories of the same model with different initial conditions converge with time. As a consequence, data assimilation should be used not only to gain insight into the current atmospheric composition but also increase our knowledge of the driving processes. Large uncertainty within the models is caused by insufficient knowledge about the emission situation, which is usually only approximated using static inventories. Data assimilation can be used to gain insight into emitted amount of pollutants. In this study we investigate the use of a particle filter based assimilation system for the regional air quality model POLYPHEMUS/DLR. Particle filters seem particularly well-suited for this task as they are able to deal with the non linearities inside the chemical model components. Furthermore the convergent nature of air pollution models helps prevent ensemble degeneration. Based on the model set up developed for the region of southern Germany within the FP7 project PASODOBLE, we construct an ensemble by modifying the model parameters like emissions and meteorology. Using a particle filter based system we can in turn evaluate the emissions into the model using in-situ observations provided by the local authorities. We will show current results and discuss options for development of an operational system.

12.15 Svetlana Dubinkina (Earth and Life Institute, Université catholique de Louvain, Louvain-La-Neuve, Belgium), Hugues Goosse, Violette Zunz and Yoann Sallaz-Damaz

Reconstructions of the climate states over last centuries using particle filtering

In contrast to meteorology, data assimilation in palaeoclimatology is relatively new, but the interest in it is growing as it gives a more reliable state estimation of the past climate changes. We conduct experiments with the three-dimensional Earth system model of intermediate complexity LOVECLIM using particle filtering to reconstruct past climate states over the last 150 years. We assimilate surface temperature from a twin experiment over southern hemisphere with assimilation period of three months.

We consider two particle filtering methods: sequential importance re sampling and an extremely efficient particle filter (P.J. van Leeuwen "Non linear data assimilation in geosciences: an extremely efficient particle filter", Q.J.R. Meteorol. Soc. 2010). In sequential importance re sampling, a set of particles, where a particle is a realization of the model with random perturbation of initial conditions, is integrated over a season. Then, when observations of sea surface temperature become available, the set is re sampled according to an importance weight of each particle. This importance weight comes from computation of the likelihood of the state obtained by each particle. The extremely efficient particle filter used here is based on sequential importance re sampling and nudging. In addition to the assimilation of surface temperature each three months, sea surface temperature is nudged every day through the fluxes coming from the atmosphere to the ocean such that it is slightly pulled towards the seasonal/monthly mean calculated from observations.

We compare results of the simulations obtained using LOVECLIM only, LOVECLIM constrained by sequential importance re sampling, and LOVECLIM constrained by the extremely efficient particle filter.

12.35 Rihab Mechri (1), Catherine Ottlé (1), Olivier Pannekoucke (2), Abdelaziz Kallel (3) and Ahmed Ben Hamida (4)

(1) Laboratoire des Sciences du Climat et l'Environnement, Institut Pierre Simon Laplace des sciences de l'environnement, Orme des Merisiers, 91191, Gif-sur-Yvette, France

(2) CNRM, Météo-France, 31057, Toulouse, France

(3) High Institute of Electronic and Communication of Sfax, Tunisia

(4) National Engineering school of Sfax, Tunisia

Sub-pixel temperatures estimation based on the assimilation of coarse resolution. Thermal Infrared data using particle filtering

Thermal infra-red (TIR) data are very useful for surface fluxes estimation giving the possibility to assess energy budgets through surface temperature. However, an accurate knowledge of such data at high spatial/temporal resolution is not possible considering the present instruments on board satellites. In fact, available instruments allow either the high spatial resolution versus a low temporal one (e.g. ASTER: repeat cycle of 16 days/spatial resolution of 15m to 90 m) or the high temporal resolution with a coarse spatial one (e.g. SEVIRI: repeat cycle of 15 min/spatial resolution of 3km). Then, it is necessary to develop methodologies to combine these multi-scale and multi-temporal data to better monitor fluxes at appropriate scale.

Our approach consists in the development of a new downscaling method based on particle filtering to extract sub-pixel variables from large scale data measurements. First step is to use a land surface model to generate ensemble of end-members temperatures (first guess temperature of each land cover class present in the TIR pixel) (the SEtHyS SVAT model is used for that purpose, Coudert et al. 2006). Second step is to aggregate the temperature for each ensemble member given a high resolution land cover to have a new ensemble containing random coarse spatial resolution temperatures. Last step consists in the selection of the optimal large scale temperature estimations that fit the observed temperatures using the particle filter. In this way, we can get for each end-member present in the TIR pixel, the best estimates of sub-pixel temperatures.

First, a sensitivity analysis has been performed to extract for each land cover class the most sensible model parameters on the surface temperature. Second the new downscaling approach will be presented and its performances will be analysed in terms of errors on the model and on the observations. Finally the performances and robustness of our approach will be discussed.

Keywords: Downscaling, Thermal Infra-red, Particle Filters

12.55 Lunch break

Session 1. *Particle Filters and Applications* (continued)

14.30 Melanie Ades (invited, University of Reading, Reading, UK) and Peter Jan van Leeuwen

Particle filters in high-dimensional systems

The majority of data assimilation schemes rely on linearity assumptions. However as the resolution and complexity of both the numerical models and observations increases, these linearity assumptions become less appropriate. A need is arising for data assimilation schemes, such as Particle filters, which are fully non linear. Unfortunately standard particle filters fail to represent the full posterior probability density function (pdf) of the model state given the observations when the number of independent observations is high. The equivalent weights particle filter is an adaptation to the standard particle filter designed to ensure a representation of the true posterior pdf, regardless of the number of observations. It uses proposal densities to ensure that particles are both close to the observations at analysis time and have equivalent significance when estimating the posterior.

The success of the equivalent weights particle filter has been shown in both the low dimensional Lorenz (1963) model and for a barotropic vorticity equation with over 65,000 variables. The encouraging results in such a high-dimensional setting show the potential for the scheme to be used operationally. However before such an implementation can be explored, one major question remains to be answered. In order to ensure that the majority of particles have equal significance in the posterior, the equivalent weights particle filter applies a movement to the individual particles. This movement has the potential to introduce unbalanced model states. It is well known that the background matrix in 4D-Var can be formulated to ensure that the analysis state is balanced. The equivalent weights particle filter does not use covariances of the state directly and so there is the potential for spurious gravity waves to be introduced by these unbalanced states.

We will investigate the severity of this problem using a one layer primitive equation model. Not only is this model high-dimensional but it also incorporates gravity waves, allowing us to explore in detail the effect of the equivalent weights movement. Potential solutions to this problem will be examined within this setting, since if this issue can be resolved then the equivalent weights particle starts to become a viable alternative for operational forecasting.

15.00 Rafal Wojcik (Massachusetts Institute of Technology, Cambridge, USA), Dennis McLaughlin, Hamed Alemohammad and Dara Entekhabi

Ensemble based fusion of noisy images

In this talk we use noisy measurements to characterize geometric features that are difficult to detect or observe. Relevant geophysical Examples include subsurface geological formations, algae blooms, forest fires, and rain storms. Our methodology accounts for uncertainty in the location and structure of hidden features by generating a conditional ensemble of images that depict the range of likely variability for a given set of observations. This conditional ensemble is derived using a non-parametric

Bayesian importance sampling approach that combines an unconditional (prior) ensemble with observations. The prior ensemble is generated with a non-stationary, multi-point stochastic procedure. The importance sampling ranks the prior images according to their probability (likelihood). The likelihood function is constructed from a set of paired historical noisy and ground truth measurements. Several aspects of the presented methodology make it flexible, statistically robust, and computationally efficient. These include versatile non-parametric representations of prior uncertainty and measurement error and a distance-preserving dimensionality reduction procedure that enables the importance sampling procedure to be carried out efficiently in a low-dimensional attribute space. The attributes that define this space are not specified in advance but are, rather, inferred from prior samples and observations. We illustrate the concepts above with a meteorological example that identifies rainy areas on the Earth's surface from noisy remote sensing measurements displayed as images. Conditional ensembles are presented for both single measurement and multiple measurement sources.

15.20 End of session 1

Session 2. *Miscellaneous*

Chair: Tim Palmer

15.20 Ashwanth Srinivasan (University of Miami, Coral Gables, USA), Mohamed Iskandarani, William C. Thacker and Omar M. Knio

Data assimilation using a polynomial chaos based ensemble

We explore the application of polynomial based uncertainty quantification methods for data assimilation and oceanic forecasts in the Gulf of Mexico. Polynomial Chaos methods are essentially spectral series in stochastic space that can be used very effectively to propagate input uncertainties through complex non-linear models with little restrictions on the model's differentiability or the uncertainties' statistics. The coefficients of that series can be effectively computed through an ensemble calculations which, once obtained, can be used in lieu of the model at a fraction of its cost. Subsequent model statistics can be obtained by mining that series. Here we explore the use of statistics derived from the series ocean for data assimilation. We illustrate the methodology through an application to the Gulf of Mexico, and discuss the pros and cons of the approach.

15.40 Heikki Järvinen (Finnish Meteorological Institute, Helsinki, Finland), Pirkka Ollinaho, Marko Laine, Antti Solonen and Heikki Haario

Ensemble prediction systems in parameter estimation and forecast skill optimization

We present an approach to utilize ensemble prediction systems to estimate numerical weather prediction (NWP) model parameters using optimal prediction skill as the target criterion. NWP models contain tunable parameters which appear in parametrisation schemes of sub-grid scale physical processes. Currently, the numerical parameter values are specified manually based on expert knowledge. The so-called EPPES method ("Ensemble prediction and parameter estimation system") utilizes ensemble prediction infra-structure for making statistical inference about the NWP model tunable parameters by (i) generating an ensemble of predictions so that each member uses different model parameter values, drawn from a proposal distribution, and (ii) updating the proposal distribution based on the likelihood of the forecasts and the corresponding parameter values as evaluated against verifying observations. In this presentation, we show how the EPPES method is used to optimize parameters related to cloud and precipitation formation, and significantly improve the medium-range forecast skill of the ECHAM5 atmospheric GCM. The reason for the improved skill is that EPPES method detects and corrects a tropical lower troposphere cold bias of the ECHAM5 model. Thus, the method is capable of relating the parameter variations to an optimal model response to correct the model bias in this complex modelling system. Finally, very promising results of the EPPES method in the context of the ECMWF forecasting system are presented.

16.00 Tea Break

16.30 Frédéric Chevallier (Laboratoire des Sciences du Climat et l'Environnement, Institut Pierre Simon Laplace des sciences de l'environnement, Gif-sur-Yvette, France)

On the parallelization of CO2-flux inversion schemes

CO2 concentrations in the atmosphere are functions of the CO2 surface fluxes, of the atmospheric transport fluxes and, to a smaller extent, of the photochemical CO2 production in the atmosphere. The combination of these three inputs makes inferring (or inverting) the surface fluxes from concentration measurements challenging. The Bayesian paradigm provides rigorous guidelines to build an inversion system in this context. However, an additional complication lies in the exceptionally long life time of CO2 that links any concentration measurement to global CO2 surface fluxes in the distant past: this long memory of the measurements combined with the high variability of atmospheric transport at short time-space scales hinders its implementation at reasonable computational cost. In practice, compromises have been chosen by the various scientific teams that apply Bayes' theorem. In the context of high-performance parallel computing, ensemble formulations of Bayes'

theorem have been increasingly popular but the relatively small sizes that can be currently afforded for the statistical ensembles imposes drastic simplifications of the inference problem. The variational formulation of Bayes' theorem is less limiting, but this approach is intrinsically sequential and implies fine-grained parallelism for the observation operator (i.e. a transport model) with much communication overhead. Moreover, transport models perform a lot of input operations (e.g., to read the transport mass fluxes), that are not well parallelizable. This problem is even bigger for their adjoint codes that read most of the information about the linearization point from the disk. Given the parallel structure of current supercomputers, finding higher levels of parallelism has become critical for the developers of the variational approach.

In this paper, we first discuss the size of the ensembles that would be needed for the implementation of a fully-fledged ensemble approach. We then introduce a physical parallelisation of the transport computation that yields a massive reduction in the computation wall clock time of a long inversion while still relying on a variational strategy. We test the performance of the algorithm on a 32-yr inversion (1979-2010).

16.50 Ralf Giering and Michael Vossbeck (FastOpt, Hamburg, Germany)

Performance gains by computing ensemble responses simultaneously

Computing the response for all members of an ensemble requires substantial computer resources. Traditionally, this task is split into several model runs performed sequentially on one processor and/or in parallel on several processors. Being straightforward, this approach, however, does not make use of the redundancy and symmetry present in such a computation. We present a new source-to-source code transformation which generates code to compute ensemble responses simultaneously on one processor.

Due to the increased memory locality of access patterns the new code speeds up the computation on processors using a cache hierarchy. On vector processors the vector length is increased within most loops in the transformed code. Similar, we could measure performance gains of the transformed code on Intel processors that support the most recent SIMD instruction set Advanced Vector Extensions (AVX).

Other sources for speed ups are computing so-called passive variables only once, performing I/O operations only once, and reducing communication cost when using MPI.

After explaining the transformation we will show a few code examples. We have implemented the transformation as a new mode of our Automatic Differentiation tool TAF. We will present and compare the run times to compute ensembles responses by the new and the classic method for a set of model codes, among them are the Planet Simulator (Plasim) and the general circulation model MITgcm. Results show that the speed up depends on the model code, the processor, the compiler and the number of ensemble members.

Finally, we will discuss code constructs that might inhibit speed up and show code modifications which allow generation of more efficient transformed code

17.10 Lewis Mitchell: University of Vermon – USA. Lewis.Mitchell@uvm.edu
Georg Gottwald (University of Sydney)

Ensemble data assimilation using stochastic homogenization in a slow-fast system with tipping points

A deterministic multi-scale toy model is studied in which a chaotic fast subsystem triggers rare transitions between slow regimes, akin to tipping points in weather or climate regimes. Using homogenization techniques from geometric singular perturbation theory, a reduced stochastic parameterization model is derived for the slow dynamics only.

We investigate whether such a stochastic climate forecast model can improve ensemble dat assimilation, in particular in the realistic setting when observations are only available for the slow variables. The main result is that the reduced stochastic model is far superior at detecting transitions between regimes the full deterministic forecast model, thus improving the analysis skill. We relate the forecast intervals over which skill improvements are obtained to the characteristic time scales of the system.

The stochastic climate model produces superior skill due to the finite ensemble size; ensembles obtained from the perfect deterministic forecast model lack sufficient spread even for moderate ensemble sizes. Stochastic climate models provide a natural way to provide sufficient ensemble spread to detect transitions between regimes. We corroborate these results with numerical simulations, and discuss the extension from a noise-induced tipping model to a model containing bifurcation-induced tipping.

17.30 End of Session

17.30 Welcome to the conference by Marc Pontaud (CNRM Scientific Deputy Director)

17.30 – 20.00 Cocktail

18.00 – 20.00 Poster Session 1 (authors in attendance 18.00 – 19.30)

20.00 Adjourn

Tuesday 13 November

Session 3. *Probabilistic Prediction*

Chair: Roberto Buizza

9.00 Tim Palmer (invited, University of Oxford, UK)

Towards the Probabilistic Weather and Climate Prediction Simulator

There is a constant tension in the development of weather and climate simulators. On the one hand, the higher the resolution is the simulator, the more accurately are we solving the underlying equations of motion – a key aspiration of our community is the production of a convectively resolved global model. On the other hand, the inevitable effects of truncation error must be represented adequately in our ensemble prediction systems if our customers are to be provided with statistically reliable probabilistic predictions. This means that computer time that might otherwise be used for increasing horizontal or vertical resolution must be allocated to ensure large enough ensemble sizes, that adequately span observation and model uncertainties, can be run. In this talk I want to present a new paradigm for weather and climate prediction which is consistent with the science of high-dimensional non linear dynamical systems, and in which both goals can be reached realistically within the coming years. In this paradigm the parametrised processes are represented stochastically, where the quantum mechanical noise of electrons flowing through wires replaces pseudo-random number generators, and where the high wave-number components of the resolved scales are integrated on chips with increased levels of speed and energy efficiency, but with decreased levels of bit-reproducible accuracy.

9.30 Luca Delle Monache (National Center for Atmospheric Research, NCAR, USA), F. Anthony Eckel (National Oceanic and Atmospheric Administration, NOAA, USA), Badrinath Nagarajan (NCAR), Daran Rife (GL Garrad Hassan, USA), Keith Searight (NCAR), Don Berchoff (NOAA), Martin Charron (Environment Canada, Canada), Ronal Frenette (Environment Canada), Jason Knievel (NCAR), Tim Mcclung (NOAA) and John Pace (Army Test and Evaluation Command, USA)

Probabilistic Weather Prediction with an Analog Ensemble

A new ensemble design based on a set of analog forecasts is proposed (analog ensemble, AnEn). The analog of a forecast for a given location and time is defined as the observation (or analysis grid point) that verified when a past prediction matching selected features of the current forecast was valid. The analogs are generated from the Environment Canada Global Environmental Multiscale Model (GEM) model deterministic run, and 10-m wind speed and 2-m temperature observations from 592 surface stations over the continental U.S., for a 15-month study period (1 May 2010 - 31 July 2011). Attributes of the 0-48 h probabilistic prediction of 10-m wind speed and 2-m temperature generated with AnEn are estimated and compared to a state-of-the-science operational system, the Environment Canada Regional Ensemble Prediction System (REPS).

AnEn exhibits better spread-error and statistical consistency, reliability, resolution, and value of the raw REPS system, while it exhibits a lower sharpness. Even after the REPS is calibrated, AnEn has superior or similar skill across the forecast lead times and for different events. The AnEn is based on a very simple approach that makes the whole procedure of generating probabilistic predictions extremely cost-effective (e.g., the real-time processing of AnEn is about three times cheaper than REPS), while producing a skill-level superior or similar to a calibrated state-of-the-science system, as REPS. The implication of these results will be discussed.

9.50 Michael Scheuerer (Heidelberg University, Germany)

Probabilistic Quantitative Precipitation Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation

Statistical post-processing of dynamical forecast ensembles is an essential component of weather forecasting. We present a post-processing method that generates full predictive probability distributions for precipitation based on ensemble model output statistics. Precipitation amounts are modelled by a left-censored generalized extreme value distribution which is fitted via minimum CRPS estimation. The conceptual simplicity of this approach permits an extension that also uses neighbourhood information contained in the dynamical forecasts in order to address displacement errors. The proposed method is illustrated with precipitation forecasts over Germany in 2011 by the COSMO-DE ensemble prediction system.

10.10 Thomas Hamill (NOAA Earth System Research Lab, Physical Sciences Division, Boulder, USA), Jeffrey S. Whitaker, Don Murray and Gary Bates

NOAA's multi-decadal global ensemble reforecast data set

NOAA recently developed a daily re-forecast (hindcast) for the 2012 version of its Global Ensemble Forecast Systems. 11-member daily re-forecasts are available to +16 days lead time every day from December 1984 to current. The forecasts were initialized at 00 UTC with the US NCEP Climate Forecast System Reanalysis (3D-Var) and with perturbed initial conditions using the ensemble transform with rescaling. Forecasts were based on the 2012 version of the NCEP Global Forecast System model. Model uncertainty was simulated with NCEP's stochastic tendency method. The forecast model resolution is T254 through day +8 (approximately 1/2 degree grid) and T190 thereafter (approximately 3/4 degree grid). Data was saved every 3 h to +72 h and every 6 h thereafter.

NOAA is maintaining a fast (hard-disk) archive of 99 of the most commonly requested variables, some at native resolution, and all at 1-degree resolution. Additionally, the full model state is archived on tape. Data is freely available to all through web interfaces and ftp.

The talk will provide more details on the configuration, the method for accessing the data, and several examples on how the re-forecasts can be used for statistical post-processing and diagnosis of model errors. Experimental products to de demonstrated may include improved probabilistic precipitation forecasts, renewable energy applications (wind, solar), improved tornado and hurricane forecasts, and diagnoses of systematic errors in forecasts of the Madden-Julian Oscillation.

10.30 Leonard Smith (London School of Economics, UK)

Guidance Information or Probability Forecast: Where do Ensembles Aim?

It is widely held that ensembles of simulations can provide a probability distribution of quantities of interest useful in decision support. This claim is challenged. It is suggested that while an ensemble of simulations provides information regarding the future, it is neither designed to nor best interpreted as providing a probability distributions reflecting future weather per se.

The seductive image of the output of an ensemble prediction system as a probability forecast, used to update a prior probability distribution (either from climatology or from yesterdays probability forecast) is inconsistent with actual practice, and arguably with the highest scoring probability forecasts.

In practice, alternative procedures are applied, procedures believed to yield both more skill and more value to the probabilistic forecast eventually produced. The ability of ensemble interpretations schemes to capture the information in the ensemble of simulations (contrasting Bayesian Model Averaging with kernel dressing) is explored, and sensible ways to use the ensemble forecast (probability updating vs blending) are contrasted. Each point holds implications for ensemble formation and resource allocation between observations, data assimilation and model complexity. The role of "sharpness" when we do not have "calibration" is clarified, and the question of whether or not post-processing ensemble prediction systems can ever yield sustainable odds (probabilities which could rationally be used as probabilities) is shown to impact the interpretation of ensemble systems.

Although focused on weather-like scenarios, where one has a large forecast-outcome archive and the model-lifetime is long compared to the forecast lead-time, these ideas also cast some light on the controversies regarding climate-like scenarios which do not have these properties. In particular, shortcoming in some of the criticisms of climate forecasts made by statisticians become clear when the aim and information content of ensembles is clarified.

The recognition that the best available initial condition was less useful than an ensemble of good initial conditions changed the nature of weather forecasting from point forecasting to probability forecasting. How might the nature of forecasting shift if model-based probability forecasts are recognised as a target we do not possess and arguably can never obtain.

10.50 Laurent Descamps (Météo-France/DIRIC, Toulouse, France) and Carole Labadie (Météo-France/GMAP/RECYF, Toulouse, France)

Ensemble post-processing methods for short-range ensemble forecasts

Calibration of ensemble prediction systems (E.P.S) aims at partly correcting deficiencies in estimating forecast errors. A number of studies have shown that post-processing methods could improve ensemble reliability. However, most of them have focused on medium range and some questions remained about the best approach for short-range E.P.S. Using the Météo-France short-range E.P.S, some of this key points will be addressed in the present study. Several post-processing methods, from simple approaches (such as bias correction) to more complex ones (such as Bayesian Model Averaging, analogs or CDF corrections) will be assessed. The need to use a reforecast data set as training data will also be investigated. Finally, the interest of post-processing methods for rare events will be illustrated on a heavy precipitation case that occured in the south-east of France in June 2010.

11.10 End of Session. Tea Break

Session 4. ***Kalman and Hybrid Filters. Theoretical Aspects***

Chair: Etienne Mémin

11.40 Marc Bocquet (invited, 1,2)
(1) Université Paris-Est, CEREA joint laboratory Ecole des Ponts ParisTech and EdF R&D, France.
(2) INRIA, Paris Rocquencourt research centre, France

Accounting for sampling errors in ensemble Kalman filtering

Because of sampling errors due to the finite size of the ensemble, the ensemble Kalman filter without corrections is useless for complex non linear models. The main solution used to fix the EnKF when applied to models with spatial extent is localisation. A subsidiary one is inflation which artificially inflates the spread of the ensemble to account for an underestimation of the errors.

In the context of an EnKF, I will show how to account for sampling errors traditionally dealt with by some ad hoc. inflation. This leads to the definition of the finite-size EnKF (EnKF-N) which accounts for sampling errors usually dealt with by inflation. One implementation of the EnKF-N is very close to the standard deterministic EnKF with the difference that it prescribes the presumably optimal inflation to be used to counter-act sampling errors.

This will be illustrated on 1D models (Lorenz '63, '95, Kuramato-Sivashinsky) and 2D models. In a perfect model context, the performance of the EnKF-N (without inflation) is usually as good as the corresponding optimally tuned EnKF.

The diagnostics of inflation by EnKF-N is local and has some resemblance with inflation adaptive schemes. A comparison will be made with adaptive inflation algorithms.

12.10 Shin'ya Nakano and Genta Ueno (The Institute of Statistical Mathematics, Tokyo, Japan)

Hybrid approach of ensemble transform and importance sampling for non-linear data assimilation

The ensemble-based approach is now recognized as a powerful tool for data assimilation in non-linear systems. In particular, the ensemble Kalman filter (EnKF) and its variants are widely used for various practical applications because of its computational efficiency and portability. However, since the EnKF is derived from a linear Gaussian observation model, it can give biased estimates in the case that the observation is non-linear or non-Gaussian. The particle filter (PF) is a ensemble-based algorithm which is applicable to problems with non-linear or non-Gaussian observations. However, the PF usually requires a prohibitively large ensemble size to achieve a good estimation. Accordingly, the PF tends to require huge computational cost.

The PF represents non-Gaussian probability density functions using the importance sampling method in order to allows us to use non-linear or non-Gaussian observations. In this paper, we propose an efficient hybrid algorithm which combines the ensemble transform Kalman filter (ETKF), which is one of the variants of the EnKF, with the importance sampling approach. In our algorithm, we first obtain a proposal distribution similar to the posterior distribution by using the ETKF. By exploiting a property of the ETKF, we can efficiently generate a large number of samples from this proposal distribution. Next, since the proposal distribution, which is obtained using the ETKF, is possibly biased due to non-linear or non-Gaussian observations, the samples from the proposal distribution are weighted so as to eliminate the bias according to the importance sampling theory. The weight for each sample in the importance sampling can also be calculated very efficiently by using the ETKF formula. Although this weighted ensemble provides a good approximation of the posterior distribution, its large ensemble size requires huge computational cost in the forecast step. In order to achieve high computational efficiency in the forecast step, we remake a new approximation of the posterior distribution with a small number of samples before the forecast step. There have been proposed several algorithms which takes advantage of both the EnKF and the PF. However, those existing algorithms do not appropriately avoid random errors introduced when a probability density function is approximated with a small number of samples. On the other hand, we generate an ensemble so as to represent the first and second order moments estimated by the importance sampling, which enables us to remarkably reduce the computational cost.

12.30 David Livings and Peter Jan van Leeuwen (University of Reading, Reading, UK)

Weighted Ensemble Square Root Filters for Non-linear, Non-Gaussian, Data Assimilation

In recent years the Ensemble Kalman Filter (EnKF) has become widely-used in both operational and research data assimilation systems. The particle filter is an alternative ensemble-based algorithm that offers the possibility of improved performance in non-linear and non-Gaussian problems. Papadakis et al (2010) introduced the Weighted Ensemble Kalman Filter (WEnKF) as a combination of the best features of the EnKF and the particle filter. Published work on the WEnKF has so far concentrated on the formulation of the EnKF in which observations are perturbed; no satisfactory general framework has been given for particle filters based on the alternative formulation of the EnKF known as the ensemble square root filter. This presentation will provide such a framework and show how several popular ensemble square root filters fit into it. No linear or Gaussian assumptions about the dynamical or observational models will be necessary. By examining the algorithms closely, short-cuts will be identified that increase both the simplicity and the efficiency of the resulting particle filter in comparison with a naive implementation. A procedure will be given for simply converting an existing ensemble square root filter into a particle filter. The procedure will not be limited to basic ensemble square root filters, but will be able to incorporate common variations such as covariance inflation without making any approximations.

12.50 François Le Gland (INRIA Rennes, France) and Valérie Monbet (Université de Rennes 1, France)

Large sample asymptotics for the ensemble Kalman filter

The ensemble Kalman filter (EnKF) has been proposed as a Monte Carlo, derivative-free, alternative to the extended Kalman filter, and is now widely used in sequential data assimilation, where state vectors of huge dimension (e.g. resulting from the discretization of pressure and velocity fields over a continent, as considered in meteorology) should be estimated from noisy measurements (e.g. collected at sparse in-situ stations). Even if the state and measurement equations are linear with additive Gaussian white noise, computing and storing the error covariance matrices involved in the Kalman filter is practically impossible, and it has been proposed to represent the filtering distribution with a sample (ensemble) of a few members and to think of the corresponding empirical covariance matrix as an approximation of the intractable error covariance matrix. Extensions to non linear state equations have also been proposed.

Surprisingly, very little is known about the asymptotic behaviour of the EnKF, whereas on the other hand, the asymptotic behaviour of many different versions of particle filters is well understood, as the number of particles goes to infinity. Our contribution is to fill this gap. Interpreting the ensemble members as a population of particles with mean-field interactions (and not merely as an instrumental device producing the ensemble mean value as an estimate of the hidden state), we prove the convergence of the EnKF, with the classical rate 1/sqrt(N), and we prove a central limit theorem, where the asymptotic variance satisfies a backward recursion, as the number N of ensemble members increases to infinity.

Notice that in the linear case, the limit of the empirical distribution of the ensemble members is the usual (Gaussian distribution associated with the) Kalman filter, as expected, but in the more general case of a non linear state equation with linear observations, this limit differs from the usual Bayesian filter.

Comparisons are provided with a popular version of a particle filter, in which particles are propagated according to an optimized Markov kernel, that takes observations into account. To illustrate our theoretical statements, numerical results are provided on a linear example, where all computations can be done explicitly.

13.10 End of session 1

13.10 Lunch break

Session 5. ***Kalman and Hybrid Filters. Continuation and Applications***

Chair: François Le Gland

14.30 Sébastien Beyou, Anne Cuzol and **Étienne Mémin** (INRIA Rennes, France)

Weighted Ensemble Transform Kalman Filter for Image Assimilation

We present a sequential data assimilation method based on the combination of two assimilation techniques: the ensemble Transform Kalman filter (ETKF) and the particle filter. Both techniques are based on Monte Carlo sampling allowing approximate solving of non linear stochastic filtering problems. However, while the ETKF is still based on a Gaussian assumption, the particle filter does not rely on such an assumption but is known to be less efficient when the number of available samples is small. In practice, both techniques are combined in the sense that the sampling step of the particle filter is based on the EnKF technique, followed by a weighting of samples using observations. The association of these two approaches is a step toward an efficient application of ensemble techniques to high-dimensional and non linear / non Gaussian problems, such as those encountered in meteorology or oceanography. We show the performances of this new approach on high-dimensional problems where the goal is to filter turbulent velocity fields from image observations. The assimilation technique associates a non linear stochastic dynamical model to linear observations extracted from the image sequences, or directly to the image data through a non linear observation operator. The method has been validated on a synthetic sequence, and applied to real oceanographic satellite image sequences of SST (sea surface temperature). This work corresponds to an extension of the preliminary study published in Tellus (N. Papadakis, E. Memin, A. Cuzol, N. Gengembre. Data assimilation with the Weighted Ensemble Kalman Filter. Tellus A, vol.62(5), p. 673-697, 2010).

In a second part, we present a way to improve the assimilation when the time step between observations (images) is very long. We make use of a conditional simulation technique in order to reduce dynamical discontinuities produced in that case by the sequential techniques.

14.50 Stefano Migliorini (University of Reading, Reading, UK)

Information-based data selection for ensemble data assimilation

A fundamental and well-known shortcoming of ensemble data assimilation techniques is that resulting analysis increments can only span a limited portion of the state space, in the case of practical importance when the size of the forecast ensemble is much smaller than the dimension of the state space. This consideration represents the main motivation for the use in ensemble data assimilation of localisation and inflation procedures that restrict analysis increments to a local portion of the state space and artificially increases the analysis variance, respectively. However, less attention is usually given to another consequence of the inherent rank deficiency of ensemble techniques, which affects the amount of observational information that is possible to exploit with these assimilation techniques. In this talk, an effective data selection methodology is presented, which allows the assimilation of only those observational components that provide information and that is compatible with existing localisation and inflation procedures. The use of this data selection method avoids unnecessary computations and round-off errors, minimises the risk of importing observational bias in the analysis and can potentially limit problems resulting from the use of too restrictive localisation domains that can perturb well-resolved balances of the flow. Results from the application of this data selection procedure are presented using a two-dimensional advection model both in the linear and non linear case. The possibility of using remote sounding observations with this technique is also demonstrated.

15.10 Georg Gottwald (University of Sydney, Australia), Lewis Mitchell and Sebastian Reich

Constraining overestimation of error covariances in ensemble Kalman filters

We investigate how to incorporate climatological information in ensemble data assimilation schemes. This can be done either on the level of providing additional observational, or on the level of parametrized forecast models.

In a first part, we consider the problem of an ensemble Kalman filter when only partial observations are available. For small ensemble sizes this leads to an overestimation of the error covariances. We show that by incorporating climatic information of the unobserved variables the variance can be controlled and superior analysis skill is obtained. We then employ this Variance Controlling Kalman Filter to control model error when the model is allowed to be void of stabilizing artificial numerical viscosity.

This is joint work with Lewis Mitchell and Sebastian Reich.

15.30 África Periáñez (Deutscher Wetterdienst, Offenbach am Main, Germany), Hendrik Reich and Roland Potthast

Adaptive Localization for Ensemble Methods in Data Assimilation

Ensemble-DA has high potential, but its theoretical justification is still under intensive development, both on the application side testing the development and operational setup of EnKF systems for different data types as well as the mathematical analysis looking into approximation properties of ensembles with ensemble-size small compared to the total number of degrees of freedom in your model.

It is very common to use space localization, in ensemble Kalman filter techniques, in order to reduce the effect of spurious long range correlations. It is the goal of our analysis to understand the basic properties of localization in this scheme. A decomposition study of the error sources is performed in order to determine its effect in the computation of the optimal localization size. The different sources of error are studied separately: the ensemble space, the localization and the observation measurements. We explore the performance of the analysis in the limits either when the ensemble size is increased or when the localization radius becomes small. This approach is analysed from the perspective of approximation theory and functional analysis, with basic insight in the form of convergence results. Secondly, it is addressed with numerical experimental results which demonstrate the validity of the theory.

15.50 Tuomo Kauranne, Idrissa Amour and Zubeda Mussa (Lappeenranta University of Technology (LUT), Lappeenranta, Finland)

Variational Ensemble Kalman filtering applied to shallow water equations

The Variational Ensemble Kalman Filter (VEnKF) is a hybrid Kalman filter that is based on a consistent variational approximation to the Extended Kalman Filter (EKF). It has been derived from the variational form of EKF, the Variational Kalman Filter (VKF), introduced by Auvinen et al in 2009. Like VKF, VEnKF re-samples its ensemble at every observation step. This feature allows VEnKF to overcome covariance leakage and maintain a robust ensemble approximation to the error covariance matrix all through the assimilation without resorting to covariance inflation. At observation steps, VEnKF reduces to 3DVAR with a dynamically computed snapshot of the state error covariance matrix. We present the VEnKF algorithm and its application to the assimilation of a real laboratory dam break experiment, modelled using the shallow water equations and observed with eight wave meters.

16.10 Tea Break

16.40 Bertrand Bonan (Laboratoire Jean Kuntzmann, Institut national de recherche en informatique et en automatique, Université Joseph Fourier, INRIA/UJF, Grenoble, France), Maëlle Nodet (Laboratoire Jean Kuntzmann, INRIA/UJF, Grenoble, France) and Catherine Ritz (Laboratoire de Glaciologie et Géophysique de l'Environnement CNRS / UJF Grenoble, France)

Ensemble Methods for ice sheet model initialisation

A hot topic in ice sheet modelling is to run prognostic simulations over the next 100 years to investigate the impact of Antarctica and Greenland ice sheets on sea level change. Such simulations require an initial state of ice sheets which must be as close as possible to what is currently observed. Large scale ice sheet dynamical models are mostly governed by the following input parameters and variables: basal dragging coefficient, bedrock topography, surface elevation, temperature field. But we do not have satisfying initial states for simulations. Fortunately, some observations are available such as surface and (sparse) bedrock topography, surface velocities, surface elevation trend. The use of advanced inverse methods appears to be the adequate tool to produce satisfying initial states.

We develop ensemble methods based on Ensemble Kalman ﬁ lter to infer optimal initial states for ice sheet model initialisation thanks to available observations. As we ﬁrst want to assess the validity of the method we begin with twin experiments with a simple ﬂow-line large scale model, Winnie, as a ﬁrst step toward data assimilation for a full 3D ice sheet model, GRISLI. Despite its simplicity, Winnie flow-line model is strongly non-linear and is a good prototype to validate our methods. We also run several diagnostics to assess the quality of the recovered parameters.

17.00 Sébastien Barthélémy (Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique, CERFACS, Toulouse, France). Sophie Ricci (CERFACS, Toulouse, France), Olivier Pannekoucke (CNRM-GAME URA 1357, Toulouse, France), Olivier Thual (URA1875/CERFACS and INPT, CNRS, IMFT, Toulouse, France) and Pierre-Olivier Malaterre (UMR G-EAU, Irstea, Montpellier, France)

Data Assimilation On A Flood Wave Propagation Model: Emulation Of An Ensemble Kalman Filter Algorithm

This study describes the assimilation of synthetically-generated river water level observations in a flood wave propagation model. For this approach to be applied in the framework of real-time flood forecasting, the cost of the data assimilation procedure, mostly related to the estimation of the background error covariance matrix, should be reduced. An Ensemble Kalman Filter (EnKF) algorithm is applied, with a steady observation network, to demonstrate how the assimilation modifies the background correlation function at the observation point. It is shown that an initially Gaussian correlation function turns into an anisotropic function at the observation point, with a shorter correlation length-scale downstream of the observation point than upstream, and that the variance of the error in the water level state is significantly reduced downstream of the observation point.

Away from the observation point, an analytical expression describes the evolution of the error variance and the correlation length scale for the water level signal when the distance to the entrance of the domain increases: when the diffusion is small compared to the advection, the covariance function remains gaussian with an increasing correlation length-scale and a decreasing error variance. At the observation point, the reduction of the error variance and correlation length scale can be parametrized as a linear function of the observation error. This parametrization relies on the integration of the EnKF for a given observing network with given error statistics but can be used to fully describe the covariance function when additional observations are available with different error statistics.

The background error covariance matrix is thus fully characterized and can be modelled using a diffusion operator with an inhomogeneous diffusion coefficient that relates to the correlation length scale. The resulting covariance matrix is then used as an invariant background error covariance matrix for a series of successive Best Linear Unbiased Estimation (BLUE) algorithms which emulate an EnKF at a lower computational cost. This study shows how the background error covariance matrix can be computed off-line, with an advanced algorithm, and then used with a cheaper algorithm for real-time application.

17.20 Umer Altaf (1), Troy Butler (2), Xiaodong Luo (1), Clint Dawson (2), Talea Mayo (2) and Ibrahim Hoteit (1)
(1) King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
(2) Insitute for Computational Engineering and Sciences (ICES), University of Texas at Austin, Austin, Texas 78712, USA

Robust Ensemble filtering for efficient storm surge forecasting

We present a robust ensemble filtering approach for storm surge forecasting that is based on the H∞ filter implemented within the framework of the ensemble Kalman filters (EnKFs). By design, the H∞ filter is more robust than the standard Kalman filter in the sense that the estimation error in the H∞ filter has, in general, a finite growth rate with respect to the uncertainties in assimilation.

We first outline the general form of the ensemble H∞ filters (EnH∞F) and discuss some of their special cases. In particular, we show that an EnKF with certain covariance inflation is essentially an EnH∞F. In this sense, the EnH∞F provides a general framework for conducting (possibly adaptive) covariance inflation in the EnKFs.

The EnH∞F was implemented with the Advanced Circulation (ADCIRC) model for storm surge forecasting in the Gulf of Mexico. We use data obtained from Hurricanes Katrina and Ike as test cases. The ensemble filtering method is based on the Singular Evolutive Interpolated Kalman (SEIK) filter which is basically a square root EnKF with a random re sampling step. The experiments results suggest that the proposed H∞ SEIK filter provides more accurate forecasts of storm surge compared to the SEIK filter with constant inflation.

17.40 George Craig (Ludwig-Maximilians-Universität München, Germany), M. Würsch, H. Lange and M. Haslehner

Applying an ensemble Kalman filter to complex convective cloud fields in a hierarchy of models

Assimilation of high-resolution observations of cumulus convection such as radar reflectivities is complicated by the rapid evolution of the convective cloud field. The ensemble must maintain sufficient diversity to capture the initiation of new clouds, while responding closely enough to the observations to remove spurious convective cells – problems that are not captured in many idealised tests that focus on simulation of a single storm. The problem of complex convective fields has been explored using a hierarchy of models. Application of the Local Ensemble Transform Kalman Filter to a simple stochastic birth-death model shows that the filter is efficient in locating correct cloud locations, but poor in removing spurious cells, despite assimilation of "no rain" observations. These results are generalised using a modified shallow water model that allows coupling of cells through gravity waves, and also orographic influences on cell initialisation. Finally, results will be presented from the COSMO-KENDA system of the Deutscher Wetterdienst for idealised test cases (periodic boundary conditions and random initialisation of convective clouds).

18.00 End of Session

18.00 – 20.00 Poster Session 2 (authors in attendance 18.00-19.30)

20.00 Adjourn

Wednesday 14 November

Session 6. ***Ensemble Variational Assimilation***

Chair: Chris Snyder

9.00 Andrew Lorenc (invited, Met Office, Exeter, UK), Neill Bowler and Peter Jermey

4D-Ensemble-Var - a Development Path for the Met Office's Data Assimilation

The Met Office's global data assimilation system, a hybrid of four-dimensional variational data assimilation (4D-Var) with an ensemble generated by a localised ensemble transform Kalman filter (ETKF), works well. However maintenance and running costs are expected to be high, especially on future massively parallel computers.

As one possible development path, we have built a 4D-En-Var system. This uses the localised 4-dimensional trajectories from an ensemble of forecasts to fit the observations in the time-window, replacing the linear and adjoint models in 4D-Var. There are two goals in the current project:

1. A "deterministic" 4D-En-Var to match the performance of the operational 4D-Var system, at lower cost. To achieve this we have built the system into our existing VAR software, so that the years of detailed development of observation operators, variable transforms, etc. are not

lost.

2. An ensemble of 4D-En-Var to match or beat the performance of the localised ETKF in our ensemble system (MOGREPS). To achieve this we have copied the sophisticated adaptive inflation which gives MOGREPS the correct spread in short-period forecasts.

The design of a potential operational system will be presented, with preliminary results from the trials which are underway.

9.30 Gérald Desroziers, Jean-Thomas Camino and Loïk Berre (CNRM-GAME, Météo-France and CNRS, Toulouse, France)

A possible implementation of the 4D-Var based on a 4D-ensemble

Most operational assimilation systems rely on the four-dimensional variational (4D-Var) formalism. Such a formalism allows the background error covariances to evolve implicitly over the assimilation window. A limitation of the 4D-Var is that the background error covariances have classically to be modelled in 4D-Var and are partly homogeneous and static, whereas they should vary in space and time. The success of the Ensemble Kalman Filter (EnKF) is due to the fact that it allows to determine realistic flow dependent background error covariances. An alternative to the EnKF is the 4D-En-Var which may combine the respective advantages of both approaches. In this presentation, we propose a possible implementation of the 4D-En-Var. An application to a simplified assimilation problem, relying on the Burger's model, will also be shown.

9.50 Milija Zupanski (Colorado State University, USA)

Towards the development of hybrid variational-ensemble data assimilation: Minimization, Hessian preconditioning, and static error covariance model

There is a growing interest in combining the experience and methodological advantages of ensemble and variational data assimilation methods into a unifying methodology commonly referred to as hybrid variational-ensemble data assimilation. The most obvious combination of variational and ensemble methodologies is by forming an ensemble of three- or four-dimensional variational (3d-Var, 4d-Var) data assimilation algorithms, or by utilizing the error covariance from an ensemble algorithm in 3d- and 4d-Var. However, in addition to making the desired improvements, these approaches also inherit the undesirable features of the existing variational and ensemble algorithms alike. This points to a need to selectively use methodological advantages from ensemble and variational data assimilation systems and avoid incorporating sub-optimal features from the original algorithms.

With this in mind we attempt to make steps towards the development of a new class of selective hybrid variational-ensemble data assimilation methods that incorporate two main components: (1) iterative minimization coupled with superior Hessian preconditioning, and (2) flow-dependent, sufficient rank mixed variational-ensemble forecast error covariance.

In this presentation we will address the significance of minimization and preconditioning in such hybrid system, and implications to the design of the static error covariance model. We will also present a new development of static error covariance for use in hybrid system that explores the structure of circulant matrices, while maintaining the complexity of existing variational error covariance models.

10.10 Thomas Auligné (National Center for Atmospheric Research, Colorado, USA)

An Integrated Ensemble/Variational Hybrid Data Assimilation System

Ensemble/variational hybrid data assimilation systems are currently under a lot of scrutiny. They are very attractive since they show the potential to leverage the robustness and efficiency of variational systems with the flow dependency and uncertainty estimation from Ensemble Kalman Filters (EnKFs). One major drawback of this approach is the requirement to develop, interface and maintain two separate data assimilation systems, which need to be consistent with each other. We propose a new approach for updating the ensemble perturbations directly within the variational system and without the need for an external EnKF system. We will explain the implementation and show results for the Weather Research and Forecasting (WRF) model in full ensemble data assimilation mode. The evolution of the analysis and the ensemble spread will be studied and compared to a state-of-the-art EnKF system. We will also explain how this new approach can be used for a variational system in hybrid mode.

10.30 Thibaut Montmerle (CNRM-GAME, Météo-France and CNRS, Toulouse, France) and Pauline Martinet

Forecast errors in clouds and precipitation: diagnosis and modelling for the assimilation of cloudy radiances and radar data in the AROME model at convective scale

In NWP systems, the assimilation of observations linked to clouds and precipitation raises the issue of characterizing adequate forecast errors. In variational data assimilation, such forecast (or background) errors are represented in the so-called B matrix which allows to filter and to spread the information brought by the observations and to impose balance between the different initial fields.

At first, background error covariances at convective scale have been computed for precipitating cases using an ensemble assimilation of AROME forecasts coupled with the operational French global ensemble assimilation AEARP, starting from analyses that consider perturbed observations. 3h forecasts differences performed separately for precipitating and non-precipitating columns have then been used to compute forecast errors of all the model variables, including microphysical quantities. Multivariate relationships have been considered in this diabatic B matrix formulation in order to account for couplings between errors of the different variables. The results show strong discrepancies with the operational B matrix, which demonstrates the sub-optimal use of observations in precipitating areas in the current operational suite.

Two applications of these forecast errors will be shown: (i) assimilation of data from Doppler radars (radial velocities and reflectivities) in the AROME 3Dvar using an heterogeneous formulation allowing to use the diabatic B matrix specifically in rainy areas, and (ii) assimilation of cloudy radiances from IASI in a 1Dvar framework. For these particular applications, the importance of considering adequate balance relationships (e.g humidity and divergence, humidity and cloud contents) and proper spatial correlations in the background error covariances will be shown.

10.50 Tea Break

11.20 Stephen Pring (Met Office, Exeter, UK) and David Fairbairn

Comparing 4D-VAR and ensemble-VAR assimilation methods within toy models

Jointly with David Fairbairn, a PhD student at the Met Office, a number of variational and ensemble-variational methods have been compared within a toy model framework. The effect of the tangent-linear model used in 4D-VAR which approximates a non linear model is compared against the necessary localisation within ensemble-variational methods. In particular we show how time correlations of observations are better maintained by the tangent-linear model but if the model is highly non linear then the use of ensemble-variational methods, which do not require the use of a tangent-linear, have improved performance and localisation has a less significant effect. The analysis error for various ensemble sizes and different ensemble-variational methods will be illustrated.

11.40 Mohamed Jardak (Laboratoire de Météorologie Dynamique, ENS, Paris, France) and Olivier Talagrand

Bayesianity of Ensemble Variational Assimilation

A strong and a weak constraints ensemble-based variational data assimilation (Ens/4D-Var) have been objectively evaluated as an ensemble estimator. A comparison with the ensemble Kalman lters (EnKF), (ETKF) and particle lters (PF) has been conducted. Finally, in the absence of a general test of Bayesianity, the much weaker property of reliability has been put to use in order to have an objective evidence as to the impact of the non-linearity and non-Gaussianity on the Bayesian character of the estimation. Results of numerical tests are discussed. The conclusion is that weak model non-linearity significantly degrades the Bayesianity of the ensembles (while non-Gaussianity has no significant eect).

12.00 Yann Michel: CNRM-GAME, Météo-France and CNRS – FRANCE.

Applying tests of univariate Gaussianity on short range forecasts from an ensemble of variational assimilations.

Data assimilation schemes used in Numerical Weather Prediction have mainly relied on Gaussian Statistical Models. It is well known that non-linearities in the forecast model will produce non-Gaussian probability density functions, such that this assumption is questionable. Because the state and observation spaces have very large dimension, estimation of second order moments is already difficult, and estimating the full state probability density function or higher order moments is still an open area of research.

Measuring the deviation of Gaussianity produced by the non-linear model in a Gaussian-based assimilation cycle requires objectives measures or statistical tests. The goal of this presentation is to describe the application of simple statistical tests on an ensemble of variational assimilations. In such an ensemble, each member consists in a cycled, four dimensional data assimilation (4D-Var), with assimilation of perturbed observations. Such an ensemble is used to estimate background error covariances. Multiplicative inflation that insures that the short-range ensemble spread matches the forecast skill. The size of the ensemble has been extended to 90 members and the ensemble has been run over 15 days.

Sample values of analysis and forecast errors for all variables and grid points are given to standard tests of Gaussianity, such as D'Agostino's chi2 and Anderson-Darling tests. Small samples almost always pass a normality test. Normality tests have little power to tell whether or not a small sample of data comes from a Gaussian distribution. With large samples, minor deviations from normality may be flagged as statistically significant (physical bounds on variables...). Thus, a hundred members is a typical size where those tests can indicate likely deviation from Gaussianity with enough power. We will describe those local deviations from Gaussianity as a function of the variables considered, the vertical levels, and the analysis-forecast cycles.

12.20 Ricardo Todling: NASA/GMAO – USA.

A. El Akkraoui, R. M. Errico, J. Guo, J. Kim, D. Kliest, D. F. Parrish, M. Suarez, A. Trayanov, Yannick Tremolet, J. Whitaker and B. Zhang

Experimenting with the GMAO 4D Data Assimilation

The Global Modelling and Assimilation Office (GMAO) has been working to promote its prototype four-dimensional variational (4DVAR) system to a version that can be exercised at operationally desirable configurations. Beyond a general circulation model (GCM) and an analysis system, traditional 4DVAR requires availability of tangent linear (TL) and adjoint (AD) models of the corresponding GCM. The GMAO prototype 4DVAR uses the finite-volume-based GEOS GCM and the Grid-point Statistical Interpolation (GSI) system for the first two, and TL and AD models derived from an early version of the finite-volume hydrodynamics that is scientifically equivalent to the present GEOS non linear GCM but computationally rather outdated. Specifically, the TL and AD models hydrodynamics uses a simple (1-dimensional) latitudinal MPI domain decomposition, which has consequent low scalability and prevents the prototype 4DVAR from being used in realistic applications.

In the near future, GMAO will be upgrading its operational GEOS GCM (and assimilation system) to use a cubed-sphere-based hydrodynamics. This versions of the dynamics scales to thousands of processes and has led to a decision to re-derive the TL and AD models for this more modern dynamics, thus taking advantage of a two-dimensional MPI decomposition and improved scalability properties. With the aid of the Transformation of Algorithms in FORTRAN (TAF) automatic adjoint generation tool and some hand-coding, a version of the cubed-sphere-based TL and AD models, with a simplified vertical diffusion scheme, is now available, enabling multiple configurations of standard implementations of 4DVAR in GEOS.

Concurrent to this development, collaboration with the National Centers for Environmental Prediction (NCEP) and the Earth System Research Laboratory (ESRL) has allowed GMAO to implement a hybrid-ensemble capability within the GEOS data assimilation system. Both 3D- and 4D-ensemble capabilities are presently available thus allowing GMAO to now evaluate the performance and benefit of various ensemble and variational assimilation strategies. This presentation will cover the most recent developments taking place at GMAO and show results from various comparisons from traditional techniques to more recent ensemble-based ones.

12.40 Max Yaremchuk (Naval Research Laboratory, Stennis Space Center, USA) and Dmitri Nechaev

Localization of the ensemble covariances using the diffusion operator approach

Flow-dependent localization of the ensemble-generated covariances is essential for improving the performance of the data assimilation algorithms. Using numerical experiments simulating inhomogeneous Gaussian-shaped covariances B, two methods of retrieval the diffusion tensor field (DTF) from the simulated ensembles are tested. Although both methods rely on the analytic DTF approximations under the assumption of weak inhomogeneity, results of the experiments indicate that reconstruction of the DTF can be performed with a reasonable accuracy in realistically inhomogeneous settings. The diffusion-based localization (DL) methods are also compared with the traditional and adaptive ensemble covariance localization (AECL) techniques. Results of the experiments suggest that DL is capable of effectively approximating B with an accuracy of the AECL method when the ensemble size is lass than a hundred. With larger ensembles the accuracy of the DL approach is limited by the weak inhomogeneity assumption underlying the technique. Computationally, the cost of DL methods is comparable with the traditional technique if the ratio of the local decorrelation scale to the grid step is not too large.

13.00 End of Session. Adjourn

13.00 Lunch break (inside the Conference centre)

14.00 Bus to the town centre

16.00 Guided tour (meeting in front of the "Donjon")

20.00 Diner at "Chez Jules" restaurant

Thursday 15 November

Session 7. *(Quasi-)Operational Ensemble Prediction Systems*

Chair: Leonard Smith (to be confirmed)

9.00 Roberto Buizza (invited, ECMWF, Reading, UK) (ECMWF, www.ecmwf.int)

20 years of ensemble prediction at ECMWF

The first operational run of the ECMWF Ensemble Prediction System (EPS) was completed twenty years ago, on the 24th of November 1992. At that time, the EPS included a representation of initial uncertainties based on initial-time singular vectors, had 33 members and was run three-times a week at approximately a 200 km resolution with 19 vertical levels, for up to 10 days.

Mis en forme

Today, the system includes a more sophisticated simulations of initial perturbations based on singular vectors and an ensemble of data assimilations. Model uncertainties are also simulated using two stochastic schemes. Its size has been enlarged to 51 members. It is run twice a day for up to 15 days, coupled to a wave model, with a resolution of approximately 30 km up to forecast day 10 and 60 km afterwards. Twice a week the system is extended with a coupled dynamical ocean to 32 days. It includes a re-forecast suite used to compute the model climate and to generate calibrated forecast products such as the extreme forecast index.

During this talk, the EPS development is reviewed, recent changes are presented and plans for future improvements are discussed.

9.30 Carole Labadie (Météo-France, Toulouse, France) and Descamps, P. Cebron, Y. Michel, L. Raynaud, M. Boisserie, L. Berre, G. Desroziers and E. Bazil.e

Initialization with ensemble data assimilation and singular vectors in PEARP, the Météo-France ensemble prediction system

PEARP (Prévision d'Ensemble ARPEGE) is the operational Météo-France short-range Ensemble Prediction System based of the ARPEGE/IFS software. PEARP is a global system and the varying resolution of ARPEGE allows to produce with one single system high resolution probabilistic forecasts over Western Europe as well as ensemble products for tropical cyclone tracks over the Tropical Atlantic area and Indian Ocean. While initial uncertainty is taken into account by coupling a small Ensemble Data Assimilation (EDA) System with Singular Vectors (SV) technique, a multi-physics approach is used to account for model error. The respective influence of this two sources of initial perturbations (EDA and SV) will be shown.

9.50 Jing Chen (The Center of Numerical Weather Prediction, China Meteorological Administration, Beijing, China), TIAN Hua, DENG Guo, LI Xiaoli, GONG Jiandong, LI Yinglin, WANG Xiaocong and HU Jiangkai

The Introduction of Ensemble Prediction System at CMA and its application in monsoon season

The development of GEPS and REPS at CMA was launched in 2002 based on the T213 Global model and WMO B08RDP project sponsored by WWRP. From 2005-2010, We develop the GEPS based T213 global model and BGM initial perturbation method. From 2009 to 2010, we start to develop the REPS for China based on the B08RDP REPS system. This system has 15 members and runs twice a day at 00 and 12 UTC with 3-h of model output frequency.

The REPS system is based on WRF model with BGM as initial perturbation method. The multi-physics technique is deployed to present the model uncertainty. Eight types of combinations were conducted by use of two different cumulus convective (Betts-miller and Kain-Fritsch schemes), two planetary boundary layer schemes (YSU and MJY), and two land surface schemes (Noah and 5-layer thermal diffusion scheme), and two microphysics schemes (Lin and WSM6). This system has 15 members with 15km horizontal resolution covering China, and runs twice a day at 00 and 12 UTC with 3-h of model output frequency. The lateral boundary conditions of REPS are provided by CMA T213 model-based GEPS. The REPS also includes every 6-h 3D-VAR data assimilation and rescaling cycle for BGM. The conventional ensemble products and some specific tailored probabilistic products for severe weather forecast (such as convective risk index) are provided. The REPS has been operationally running since Jun 2010.

The evaluations of performance of GEPS and REPS for some typical heavy rainfall processes in monsoon season in China are performed in this study. The results show that compared to T213 model-based GEPS system the REPS has significant advantages for the forecasts of precipitation and high impact weather at short-range. Moreover, the Ensemble mean and probabilistic products for heavy rainfall events are able to provide more useful guidance information at the mesoscale. Finally the future plan for CMA REPS is discussed.

Key Words: GEPS and REPS at CMA, probabilistic products, heavy rainfall

10.10 Masashi Ujiie (Japan Meteorological Agency, Tokyo, Japan), Haruki Yamaguchi, Masakazu Higaki and Masayuki Kyouda

Current status and development of medium range EPS at JMA

The atmospheric predictability depends on flow and varies day to day. To quantify such uncertainty of the forecast in various time and scales, the Japan Meteorological Agency (JMA) has operated several ensemble forecast systems (EPSs), One-week, Typhoon, One-month and Seasonal EPSs. As a medium range forecast system, the One-week EPS targets on daily synoptic systems, heat and cold waves up to 9 days.

JMA started operation of the One-week EPS in 2001. For 11 years, the skill of the EPS has been continuously improved. Recently, the Anomaly Correlation Coefficients of 9day 500hPa geopotential height forecast for the extra tropical Northern Hemisphere has reached to 0.6, which means that it is worth while extending forecast period of the EPS. JMA also operates the Two-weeks EPS based on the One-month EPS whose target is weekly mean hot and cold temperature forecast. To make qualities of forecast consistent and develop the EPSs seamlessly, we plan to integrate the One-week and the Two-weeks EPSs. To construct the integrated EPS, we are reconsidering the specifications of the EPS such as initial perturbation generator, lower boundary condition, number of perturbed members and frequency of forecast. For initial perturbation generator, the One-week and the Two-weeks EPS currently employ different methods, the SV and the BGM method respectively. The SV method identifies perturbation related to synoptic systems over the extra tropics, while the method has difficulty in capturing growing modes over the tropics. On the other hand, the BGM method captures large scale growing modes by adjusting its definition of norm but the growth rate of the perturbations are still small relative to errors. The treatment of initial perturbations over the tropics will be one of the main issues for developing the integrated EPS. In the integrated EPS, we plan to produce a re-forecast dataset which is currently used in the Two-weeks EPSs to remove systematic errors. The gain of the systematic errors correction on the daily forecast is also our interest. Our preliminary results using the re-forecast dataset of the seasonal forecast system showed that correction of systematic errors which located stationary gave also positive impacts on the medium range forecast. The results showed the usefulness of re-forecast for the medium range forecast.

In the presentation we will introduce these activities on developing the integrated EPS.

10.30 Christoph Gebhardt (Deutscher Wetterdienst, Offenbach am Main, Germany), Susanne E. Theis, Zied Ben Bouallegue, Andreas Roepnack, Nina Schuhen and Michael Buchhold

COSMO-DE-EPS – an operational convection-permitting ensemble prediction system for the atmosphere

The convection-permitting ensemble COSMO-DE-EPS is in operational mode at DWD since 22nd May 2012. The currently 20 members are generated by applying variations to initial and boundary conditions based on dynamically downscaled forecasts of four global models. Additionally, the model physics of COSMO-DE are perturbed with a non-stochastic approach. The COSMO-DE-EPS has a three-hourly update cycle with forecasts up to 21 hours on a 2.8 km grid. Probabilistic products are provided for each forecast hour with focus on parameters which are relevant for weather warnings.The experiences made during more than one year of time-critical pre-operational and operational mode and general properties of the COSMO-DE-EPS will be presented based on forecasters' evaluation and objective probabilistic verification.

An overview of recent developments and future plans in member generation and applications of the COSMO-DE-EPS will be given.

10.50 Tea Break

11.20 Jonathan Flowerdew (Met Office, Exeter, UK)

MOGREPS ensemble forecasting

This presentation briefly highlights three topics related to ensemble forecasting at the Met Office:

- The operational ensemble system (MOGREPS) will be described, with particular emphasis on the initial condition perturbations and the adaptive inflation scheme which is used to calibrate their spread.
- Raw ensemble forecasts can be subject to deficiencies arising from systematic errors in the underlying model, data assimilation and perturbations. These can be mitigated by combining output from multiple ensemble systems with different systematic errors, and by statistically transforming the forecasts based on information about past performance. These techniques will be illustrated for the case of medium-range precipitation over the UK. The multimodel ensemble is competitive with or superior to the best single model, for both the reliability and resolution components of the Brier Skill Score. Forecast probabilities were calibrated using a method which preserves spatial structure. This improves reliability in most cases, and brings further improvements to resolution for light rain.
- One key role of ensemble forecasts is to predict the risk of severe events that may endanger

life or property. Under contract to the Environment Agency, the Met Office produces an ensemble forecast of coastal storm surges by driving a storm-surge model from each MOGREPS ensemble member. These forecasts have recently been extended from two to seven days using data from the medium-range component of MOGREPS. Both case studies and statistical verification indicate the potential for useful forecasts over this lead time range, superior to that which can be obtained with simple dressing of a deterministic forecast.

11.40 Piers Buchanan (Met Office, Exeter, UK) and Ken Mylne

The Use of Met Office Ensemble Model data to Generate Severe Weather Forecasts

The Met Office Global and Regional Ensemble Prediction System (MOGREPS) was made operational in 2008 following 3 years of trials. The model output from this model has been used in a variety of ways by both internal and external customers. This presentation focuses on the use of global MOGREPS data to generate 3 different products for forecasting severe weather:

1. A year's trial has shown potential benefit in using ensemble data to generate clear air turbulence (CAT) forecasts for civil aviation. Objective verification has allowed comparison of the performance in using 24 members against just using a single control member.
2. Some preliminary work to use global MOGREPS data to generate products to aid forecasting propensity for lightning, hail and tornados will be shown. Several case studies will be presented.
3. Global 15 day tropical cyclone forecast products using MOGREPS-15 data will be presented. Some results verifying forecast track error, combining MOGREPS, ECMWF EPS and NCEP GEFS data will also be shown.

12.00 Christopher Cunningham: Center for Weather Forecast and Climate Studies (CPTEC/INPE) Brazil. christopher.cunningham@cptec.inpe.br
Jose Paulo Bonatti, Pedro Leite da Silva Dias

Improvements of the eigenvector method to produce a ensemble of perturbed initial conditions for the CPTEC Ensemble Prediction System

In Brazil, the Center for Weather Forecast and Climate Studies (CPTEC) produces operational ensemble forecasts since 2001. Currently, CPTEC is member of the THORPEX Interactive Grand Global Ensemble (TIGGE), a World Weather Research Programme to accelerate the improvements in the accuracy of 1-day to 2 week high-impact weather forecasts for the benefit of humanity. The method used to create perturbed initial conditions is based on an eigenvector/eigenvalue problem. This work aims to present the basic aspects of this method and recent efforts made to improve the skill of the CPTEC ensemble prediction system (CPTEC-EPS). Modifications that were introduced in the originally proposed method, which is currently the operational version, include perturbing four additionally target regions and two new variables (CPTEC EPS-MB2009). The effect on forecast skill of this modifications was measured trough anomaly correlations, spread among members, root mean squared error and continuous ranked probability score (CRPS). The variables evaluated were absolute temperature at 850 hPa and sea level pressure, averaged over Northern and Southern Hemisphere, and Tropics. A system that compares systematically and in real-time the operational and improved CPTEC-EPS was developed. Results showed that spread was enhanced and root mean squared error was diminished, mainly over the extra tropics; hence, improving the error growing representation by the ensemble system. The deterministic aspect of the ensemble prediction system was measured through the anomaly correlation pattern. This score indicated larger improvements over the northern hemisphere and for temperature at 850 hPa. The CRPS results also show improvements. This score, which is negative oriented, diminished over all regions, indicating that the proposed modifications improve the probabilistic predictive characteristic. Hence, the proposed modified method resulted in an overall improvement of the skill of the CPTEC EPS. A bias correction of the CPTEC E! PS-MB200 9 outputs resulted in a further improvement in the performance of the EPS. Best progresses were achieved over the tropics and southern hemisphere.

12.20 François Bouttier (Météo-France CNRM-GAME 42 Av Coriolis 31057 Toulouse, France), Olivier Nuissier, Laure Raynaud and Benoît Vié

Regional short-range ensemble prediction with the AROME atmospheric system

The pre-operational AROME ensemble prediction system aims to predict atmospheric event probabilities at fine scale (dx=2.5km) and short range (3-36 hours). The main challenges lies in the small affordable ensemble size and in peculiarities of the phenomena of interest, such as strong rain, which tend to be relatively rare, intermittent and highly non-Gaussian. The presentation will describe ongoing research in the ensemble perturbation techniques (initial conditions by ensemble of 3DVar assimilations, boundary conditions from a larger-scale ensemble, stochastic model perturbations) and performance and verification issues. The user point of view will be explained, with a focus on heavy Mediterranean precipitation, on aviation users, and on risk management in a real-time operational setting. There are complex interactions with advanced data assimilation techniques, too.

12.40 End of Session. Lunch Break

Session 8. ***(Quasi-)Operational Assimilation Systems***

Chair: Tom Hamill

14.30 Loïk Berre (invited, CNRM-GAME, Météo-France and CNRS, Toulouse, France), Gérald Desroziers, Hubert Varella, Laure Raynaud, Carole Labadie and Laurent Descamps

Variational ensemble data assimilation at Météo-France for covariance modelling and ensemble prediction

Since July 2008, a variational ensemble data assimilation system (EnVar) has been operational at Météo-France. It is used to calculate flow-dependent background error covariances for optimizing data assimilation. It provides also perturbed initial states for the Météo-France ensemble prediction system.

Prominent features and developments of this EnVar will be presented. The ensemble approach has been chosen to be consistent with an error simulation technique of the deterministic 4D-Var data assimilation cycle. Spectral filtering techniques based on objective signal and noise estimates are applied for modelling background error standard-deviations, and wavelet filtering techniques are considered for correlations.

The link between ensemble approaches and data assimilation techniques has been emphasized in this system, with for instance a combination of observation- and ensemble-based information to calibrate and validate this EnVar system.

15.00 Anna Shlyaeva (Hydrometeorological Research Center of Russia, Moscow, Russia), Vasily Mizyak and Mikhail Tolstykh

Local Ensemble Transform Kalman Filter assimilation scheme for the global atmospheric model SL-AV

Local Ensemble Transform Kalman Filter assimilation scheme is being implemented in the Hydrometeorological Research Center of Russia. The LETKF scheme is developed for the operational global NWP model SL-AV. SL-AV is a semi-Lagrangian global atmospheric model that uses semi-implicit finite-difference block of own development for solving dynamics and parameterizations from ALADIN/LACE model. The first experiments on the ensemble assimilation are carried out with the 1.4x1.1 degrees, 28 vertical levels model version.

The observations currently assimilated are synops (surface pressure, temperature and wind components), radiosondes (temperature and wind components on the pressure levels) and satellite retrievals (wind components on the pressure levels).

Localization is carried out in the observation space via Schur product. We currently use Gaspari-Cohn localization function with localization distances varying for different vertical levels, and latitudes. The distances are user defined at several vantage points (in vertical and latitude) and the localization distances in between are calculated by the linear interpolation. Vertical localization is made in terms of pressure logarithm.

Multiplicative as well as additive inflation is used to avoid the filter divergence and to account for the model error. The experiments on using stochastic physics for the accounting for the model error that were carried during our tests on the shallow water model have shown the probable positive outcome of the approach. We are now incorporating simple stochastic physics scheme for the SL-AV model, perturbing some parameters of the physical parametrizations during the forecast step. The experiments with the several months full assimilation cycle using LETKF scheme and SL-AV model have shown the stability of the ensemble scheme.

We are now investigating the use of the AMV observations taking into consideration their spatial auto covariances. We plan to estimate the impact of different strategies for dealing with this issue in the assimilation scheme: thinning, super observations, using non-diagonal observation error covariance matrices, variable transformation.

15.20 Amal El Akkraoui (NASA/Global Modelling and Assimilation Office/Science Systems and Applications, Inc., Greenbelt, USA) and Ricardo Todling

The GMAO 3DVAR-Hybrid Data Assimilation System

The Global Modelling and Assimilation Office (GMAO) has collaborated with the National Centres for Environmental Prediction (NCEP) in the development of the Grid-point Statistical Interpolation (GSI) system. Originally 3DVAR-capable only, this collaboration has enabled the Grid-point Statistical Interpolation analysis with a 4D-capability. Concurrently, collaboration of Earth System Research Laboratory (ESRL) and NCEP has introduced to GSI an ensemble-hybrid capability. [see Daryl (2012) for a comprehensive study of various configurations of ensemble-variational hybrid strategies for the NCEP Global Forecasting System]. GMAO is presently investigating traditional 4DVAR and 3DVAR-hybrid, as possible intermediate steps toward a fully 4D-hybrid data assimilation for the Goddard Earth Observing System (GEOS). This presentation summarizes results from the implementation of a hybrid-3DVAR component for GEOS DAS.

Together with the GEOS general circulation model (GCM) and GSI, the ensemble Kalman filter of Whitaker et al. (2008) is used to provide analysis updates for GEOS GCM ensemble forecasts that serve as background to the 3D-hybrid strategy. Experiments with different configurations of a 32-member ensemble 3DVAR-hybrid are compared with results from the traditional 3DVAR used thus far to generate GMAO products. An update on the progress toward replacing the GMAO operational system with a 3DVAR-hybrid scheme will also be given.

Kleist, D. T, 2012: An evaluation of hybrid variational-ensemble data assimilation for NCEP-GFS. Ph. D. Thesis, University of Maryland

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. Mon. Wea. Rev., 136, 463-482.

15.40 Yonghong Yin (The Centre for Australian Weather and Climate Research, and the Bureau of Meteorology, Melbourne, Australia), Oscar Alves and Debra Hudson

Ensemble data assimilation/initialisation for intra seasonal to seasonal prediction using POAMA at the Bureau of Meteorology

POAMA (http://poama.bom.gov.au/) is a state-of-the-art seasonal forecast system based on a coupled ocean and atmosphere general circulation model and data assimilation systems. The latest version, POAMA-2, is designed to deliver both seasonal and intra seasonal forecasts. The data assimilation systems for the seasonal prediction version of POAMA-2 include an ensemble ocean data assimilation system (PEODAS; Yin et al. 2011) and an atmosphere/land initialisation (ALI; Hudson et al. 2011) system. For developing the intra seasonal forecasting capability, we have recently developed a coupled ensemble initialisation (CEI) scheme using a coupled-model breeding method to generate an ensemble of perturbed ocean and atmosphere states that can be used to initialise the forecasts.

Using the unperturbed analyses from PEODAS and ALI, the CEI system generates coupled bred vectors (BVs) of both the ocean and atmosphere based on the one-sided breeding method. Considering the different characteristics of the time scales of variability in the global ocean and atmosphere, we choose different rescaling norms for each component. Zonally-averaged 10 m zonal wind and three dimensional oceanic temperatures have been selected as atmospheric and oceanic rescaling norms, respectively. The perturbations produced by the CEI system display significant state-dependence for both atmosphere and ocean components. The BVs show that they can be sampled as the growing errors for intra seasonal to seasonal time scales for improving the forecasting skills and can also be used for estimation of error covariances in an ensemble data assimilation system.

Currently a preliminary version of a coupled ensemble data assimilation system for POAMA-2 (called PECDAS) has been developed. This is based on the CEI system but the unperturbed atmosphere and ocean analyses is done sequentially using modified methods of the ALI and the PEODAS respectively and using the background states from the coupled model run. The ocean error covariances are estimated from the ensemble perturbations. The characteristics of the coupled perturbations from the coupled system and the ocean reanalysis from PECDAS are compared to those from PEODAS.

References:
Hudson, D., Alves, O., Hendon, H.H. and Wang, G., 2011: The impact of atmospheric initialisation on seasonal prediction of tropical Pacific SST. Climate Dynamics. 36, 1155-1171.
Yin, Y., O. Alves, and P. R. Oke, 2011: An ensemble ocean data assimilation system for seasonal prediction. Mon. Wea. Rev., 139, 786-808.

16.00 Tea Break

16.30 Pau Escribà (ECMWF, Reading, UK, and Agencia Estatal de Meteorología, Barcelona, Spain), Massimo Bonavita, Mats Hamrud, Lars Isaksen and Paul Poli

LETKF compared to 4DVAR for assimilation of surface pressure observations in IFS (ECMWF)

European Centre for Medium-Range Weather Forecasts (ECMWF) runs operationally a 4DVAR assimilation system using a 12 hours window in its Integrated Forecasting System (IFS) model. Flow-dependency in background error estimation is taken into account by running a lower resolution Ensemble of Data Assimilations.

ECMWF has codified the Local Ensemble Transform Kalman Filter (LETKF) with IFS for experimental purposes. Several methods for additive and multiplicative inflations and for localization have been implemented. Because of the assumption about linearity in the LETKF equations, a six hours assimilation window is used to ensure that the background ensemble propagation is reasonably linear.

In the first part of the present work four of the analysis systems available at ECMWF are compared assimilating only one type of observations, surface pressure, as a first sanity check. The four methods are LETKF, 4DVAR with static background covariance, hybrid 4DVAR-EDA and hybrid 4DVAR-LETKF. The comparison looks at the performances of the different analysis states, taking as the truth the operational 4DVAR analysis that runs at ECMWF with the full observing system and includes some diagnostics of the nature of the results.

In the second part results for the Kalman Smoother LETKF extension are presented. This technique recomputes the analysis at one assimilation cycle by using the weights computed in the next cycle. Using Kalman Smoother LETKF assimilation window is extended from 6 hours to 12 hours in a way that the assumption about linearity is respected.

Using only surface pressure in the assimilation along with the fact that the NWP based model is the same for all the analysis schemes (IFS), ensures a quite clean comparison between LETKF and the variational techniques without having the problem of localisation for satellite radiances which is still and issue to investigate for improving LETKF analyses.

Finally some results for the different assimilation schemes with the full observing system are presented as well as a comparison of computing efficiency.

16.50 Nils Gustafsson (Swedish Meteorological and Hydrological Institute, Norrköping, Sweden) and Jelena Bojarova (Norwegian Meteorological Institute, Oslo, Norway)

Sensitivity to the ensemble generation approach in hybrid 4D-Variational ensemble data assimilation scheme in HIRLAM

The Hybrid Variational ensemble data assimilation scheme has been implemented into the HIRLAM forecasting system. The structures of the static forecast error covariance are merged with the flow-dependent structures sampled by the ensemble of HIRLAM perturbations. These flow-dependent structures are incorporated into the existing HIRLAM 3D- and 4D-Variational data assimilation schemes by the augmented control state method. A number of real observation experiments (conventional and satellite radiance observations) and observing system simulation experiments (single observation experiments) were conducted in order to understand properties of the hybrid scheme. We investigate sensitivity of the scheme to the different ensemble generation methods (EnsDA, ETKF, downscaling of global EPS) and to the various ensemble generation strategies, with regards to variance inflation, ensemble size, forecast length and forecast time lagging of ensemble perturbations. Sensitivity to other tunable parameters such as assumptions on smoothness and variability of the local linear weights is investigated as well.

17.10 Christoph Schraff (Deutscher Wetterdienst, Offenbach am Main, Germany), Hendrik Reich and Andreas Rhodin

Development of a LETKF for Km-scale Ensemble Data Assimilation for the COSMO model

A novel ensemble data assimilation system for convection-permitting numerical weather prediction using mesh widths typically less than 3 km is being developed in the framework of a COSMO Priority Project called KENDA (Km-scale ENsemble-based Data Assimilation). The system is based on the LETKF (Local Ensemble Transform Kalman Filter) method. The main objective of this system is to provide suitable perturbed initial conditions for an ensemble prediction system, and to use a situation-dependent description of the background errors for combining short-range forecast information with new observations. To provide high-quality perturbed lateral boundary conditions, a hybrid 3DVAR-LETKF system is also being developed for the global model at DWD. In the presentation, we will focus on the system implementation and outline current work.

17.30 End of session

17.30 Session 9. ***Identification and Representation of Model Errors***
Chair: Marc Bocquet

17.30 Carla Cardinali (ECMWF, Reading, UK), Nedjeljka Zagar, Gabor Radnoti and Roberto Buizza

Representing model uncertainties in the ensemble data assimilation

The ECMWF operational EDA simulates the effect of both data and model uncertainties. The data errors are represented by adding perturbations with statistical characteristics implied by the observation error covariance matrix whilst the model uncertainties are represented by adding stochastic perturbations to the physical tendencies to simulate the effect of random errors in the physical parameterizations (ST-method). In this work an alternative method (XB-method) is proposed to simulate model uncertainties, based on adding to the model background field, perturbations with statistical characteristics defined by the model background error covariance matrix. The perturbation amplitude varies in accordance with the amplitude of the lack of spread in the ensemble analysis when the ensemble spread is compared with the innovation vector spread for a certain period. The perturbation amplitude is a function of model levels, latitude band and parameter. A full diagnostic assessment on the different ensembles analysis generated by the different model error representation methodologies used is presented. Also, the different ensembles analysis have been used to generate the initial perturbations of the ECMWF Ensemble Prediction System (EPS). The XB-method induces the largest EPS spread in the medium-range.

17.50 Adjourn

Friday 16 November

Session 9. ***Identification and Representation of Model Errors (continued)***

Chair: Marc Bocquet

9.00 Renaud Marty: Chaire de recherche EDS en prévisions et actions hydrologiques, Département de génie civil et de génie des eaux, Université Laval, Canada. renaud.marty.1@ulaval.ca
Vincent Fortin: Recherche en prévision numérique environnementale, Environnement Canada, Dorval, Canada
Heri Kuswanto: Department of Statistics, Faculty of Mathematics and Natural Science, Institute Technology of Sepuluh Nopember (ITS), Surabaya 60111, East Java, Indonesia
Anne-Catherine Favre: Grenoble-INP/ENSE3, Laboratoire d'Etude des Transferts en Hydrologie et Environnement (LTHE), UMR 5564, Grenoble, F-38041, France
Eric Parent: Dpt. of Statistics, UMR518 AgroParis- Tech/INRA, 16 rue C Bernard, 75005 Paris, France

Bayesian Processor of Ensemble Members: combining the Bayesian Processor of Output with Bayesian Model Averaging for reliable weather forecasting

Meteorological forecasts are provided by numerical weather prediction systems and their associated uncertainty (initial conditions, meteorological modeling) is generally estimated through an ensemble approach. Current ensemble prediction systems typically underestimate this uncertainty, partly because not all sources of uncertainty are being taken into account, partly because there exists a discrepancy between the temporal and spatial scales at which the variables are predicted and the scales at which they are needed for specific applications, including hydrological ensemble prediction systems.

Post-processing of ensemble forecasts is therefore necessary in order to obtain a reliable assessment of the forecast uncertainty. Different statistical approaches have been recently proposed to build reliable probabilistic forecasts from raw ensemble forecasts, including:

- 2005 the Bayesian Model Averaging (BMA) representing the predictive distribution by a weighted combination of distribution centred on bias-corrected ensemble members;
- 2006 the Bayesian Processor of Output (BPO) summarizing the probabilistic forecasts by ensemble means and building a likelihood function from the joint record of ensemble means and observations, and combining it with a prior distribution (climatology) to obtain the predictive distribution.

The proposed Bayesian Processor of Ensemble Members (BPEM) is based on a hierarchical model assuming that ensemble prediction systems are reliable but according to an unobserved latent variable and considering that a linear relationship exists between this latent variable and the predictand. Parameters of the mixture model are estimated on the basis of the BMA framework assuming that ensemble members are exchangeable with the latent variable. Conditional on the latent variable, the predictive distribution of the quantity of interest is then obtained using the BPO framework. This statistical approach can be applied to calibrate meteorological ensemble forecasts.

It is illustrated with temperature forecasts in Québec City in summer 2008. The proposed framework successfully generates reliable forecasts and outperforms slightly both the BMA and BPO approaches as well as a climatological forecast. The comparison with BMA and BPO reveals that the introduction of the climatological information in the predictive distribution and the application of calibration on each ensemble members through the latent variable explain mainly this good behaviour.

9.20 So-Young Ha (National Center for Atmospheric Research, Boulder, USA), Chris Snyder and Judith Berner

Model error representation in mesoscale WRF-DART cycling

Mesoscale forecasts are strongly influenced by physical processes, from turbulence and mixing in the planetary boundary layer to moist convection and microphysics, that are either poorly resolved or must be parameterized in numerical models.

Due to the model errors, mesoscale ensemble systems generally suffer from under dispersiveness which often leads to poor forecast skills. In an ensemble Kalman filter data assimilation, insufficient ensemble spread leads to poor filter performance in the analysis cycle.

To alleviate the underestimate of ensemble spread, we compare two approaches to this issue: a multi-physics ensemble, in which each member's forecast is based on a distinct suite of physical parameterizations, and stochastic backscatter, in which small noise terms are included in the model equations for momentum and potential temperature.

We perform our experiments in a domain over the continental U.S. using the WRF/DART system, which employs the Weather Research and Forecasting model for ensemble forecasts and the Data Assimilation Research Testbed for the ensemble Kalman filter.

Verification against independent observations for a one-month summer period shows that including model-error techniques improves not only an ensemble of analyses, but also short-range forecasts started from these analyses. The stochastic backscatter scheme outperforms the multi-physics ensemble near the surface throughout the whole cycling period.

9.40 Alison Rudd (University of Reading, Reading, UK), Laura Baker, Stefano Migliorini and Ross Bannister

Representation of model error in convective-scale ensembles

An experimental 1.5-km resolution convection-permitting version of the Met Office's 24-member Global and Regional Ensemble Prediction System (MOGREPS) has recently been developed (the 1.5km-EPS). The 1.5km-EPS accounts for model error from unresolved processes through the use of a stochastic perturbation technique known as the random parameters (RP) scheme. We assess the ability of the RP scheme to increase ensemble spread and make comparisons with observations. We determine the sensitivity of diagnostics, such as variances and correlation length scales, to the size and structure of the model errors. By switching on and off the different sources of forecast error (initial condition error, lateral boundary condition error and model error) we attempt to disentangle the effect of model error from initial condition and boundary condition errors. A series of forecasts performed for different lead times are used to study the impact of model error on the evolution of the forecast error covariance. A comparison is also made between the ensemble-derived and observation-derived forecast error statistics.

10.00_Kirstin Kober (Meteorologisches Institut Ludwig-Maximilians-Universität München, Germany), George C. Craig, Tobias Selz and Annette Förster

Stochastic parameterisations in multi-scale ensemble systems

A multi-scale ensemble forecasting system enables to examine the relation between the predictability of high impact weather and uncertainty coming from both small and large scales. Within the project Pandowae (Predictability ANd Dynamics Of Weather Systems in the Atlantic-European Sector) such a multi-scale ensemble system is under development focusing on variability caused by convection. The large scale variability is provided by a selection of 10 members of the global IFS ensemble prediction system (EPS) of ECMWF. Within each of these 10 members, ten COSMO (Consortium for Small-scale Modelling) model runs with 7 km horizontal resolution are nested. The 100 high resolution forecasts use the Plant-Craig stochastic convection parameterization representing convective variability. The skill of this existing part of the ensemble system (100 members) is investigated and compared with observations with focus on precipitation forecasts. Neighbourhood verification techniques are applied to compare the skill of the stochastic realizations with the standard Tiedtke convection scheme. First results show that for high thresholds the forecasts with the Plant-Craig scheme are superior to the Tiedtke forecasts. The probabilistic skill of the ensemble is also investigated.

The last component of the multi-scale EPS representing small scale variability is currently under development. COSMO experiments with 2.8 km horizontal resolution are nested into the 100 7 km members with perturbations to the convective boundary layer. The amplitude of the variability is determined based on physical processes affecting the initiation of convection in the boundary layer like surface heating, subgrid scale orography or cold pools. Perturbations based on these processes are added to the tendencies and analysed.

10.20 Alfons Callado (Spanish Meteorological Agency, Barcelona, Spain), G. Shutts, J--J. Morcrette and R. Forbes

Stochastic Perturbation of Physical Parameterization Tendencies and Model Uncertainty

Stochastic parameterizations provide a methodology for representing model uncertainty in ensemble forecast systems (EPS), and have the ability to improve the EPSs performance. The operationally forecast ensembles at European Centre for Medium-Range Weather Forecasts (ECMWF) have included an explicit stochastic perturbation to the net effect of parameterized physical processes, called Stochastically Perturbed Parameterization Tendencies scheme (SPPT), in their ensemble prediction system since October 1998 (Buizza et al. 1999; Palmer et al. 2009). SPPT is based on a random multiplicative noise drawn from a Gaussian distribution and applied with spatial and temporal correlations. And it accounts for a cloud of sub-grid possibilities scattered around a mean parameterized value compatible with resolved grid constraints, so exploring nearby regions of phase space that more restricted deterministic parameterizations do not reach. Hence SPPT generalises the output of existing sub-grid parameterizations as probability distributions.

The present study aims to assess the strengths and limitations of SPPT scheme based on the current implementation at ECMWF and also explore new approaches with the potential to improve SPPT. It is also intended to gain knowledge about model errors and/or uncertainties in Integrated Forecast System (IFS) from ECWMF in order to devise and calibrate SPPT using the coarse-graining methodology (Shutts and Palmer, 2007).

A number of different settings and variants of SPPT in T159 ECMWF EPS have been experimented like:1) Applying SPPT independently to each parameterization; 2) Using distinct spatial and temporal pattern generators to each parameterization; 3) Not perturbing clear-sky in radiation parameterization; or 4) Trying another perturbation functions. In spite of some different benefits and/or weakness between these approaches, it is concluded there seems not to be a single clear optimum SPPT implementation, but rather there exists a range of them producing quite similar results. It is discussed that where 'theoretical' NWP model uncertainty is expected to be does not always correspond to 'practical' findings. Finally like first Buizza's implementation it was found that SPPT is very sensitive and so dependant to space- and time-scales of stochastic forcing playing a main role the generator patterns.

To infer some statistical information on the component of model error that comes from NWP model resolution, ECMWF IFS forecasts at T1279 resolution have been used as 'truth' relative to matching T159 forecasts. So both tendency contributions coming from dynamics, and parameterized radiation, convection, large-scale condensation, turbulent mixing, non-orographic gravity wave drag and their total are coarse-grained to common spatial and temporal scales to derive the error of lower resolution model error. Results suggest that variance of the model error is proportional to the mean tendency rather than the standard deviation assumed by SPPT scheme. Moreover the associated PDF resembles more to Poisson distribution for which the variance equals the mean than to a Gaussian one (Shutts and Callado, 2011).

10.40 End of Session. Tea Break

Session 10. ***Further Developments***

Chair: Martin Leutbecher

11.10 Hailiang Du, Leonard A. Smith and **Emma Suckling** (The London School of Economics and Political Science, UK)

Pseudo-orbit Gradient Descent Ensemble Method

Berliner (1991) pointed out a number of difficulties in applying the Bayesian paradigm to state estimation in chaotic systems. Even in the perfect model scenario, likelihood methods have difficulty in providing good estimate of the initial condition (or the model parameters). In large part, the difficulty lies in the failure to skilfully meld information in the dynamics of the non-linear system itself with that from the observations. Data assimilation via Pseudo-orbit Gradient Descent provides an attractive approach to data assimilation in non-linear systems. In the perfect model scenario, noise observations prevent estimating the current state precisely. In this setting, Ensemble Kalman Filtering approaches are hampered by foundational assumptions of dynamical linearity and its failure to respect dynamical manifolds, while variational approaches suffer due to local minima in the cost function. The Pseudo-orbit Gradient Descent approach addresses the nowcast problem by enhancing the balance between extracting information from the dynamic equations and information in the observations. Outside the perfect model scenario, most traditional methods fail to produce consistent estimates of the model state due to structural model error. The Pseudo-orbit Gradient Descent approach produces more consistent estimates of the model state, and estimates the model error simultaneously, but requires a stopping criteria. The Pseudo-orbit Gradient Descent approach is shown to outperform the Ensemble Kalman Filter method and the variational method in the context of 18-dimensional Lorenz96 flow and 2-dimensional Ikeda map in both the perfect and some imperfect model scenarios.

11.30 Vivien Mallet (INRIA & CEREA, France), Sergiy Zhuk (IBM Research, Dublin, Ireland) and Alexander Nakonechniy (Taras Shevchenko National University of Kiev, Kiev, The Ukraine)

Ensemble forecast of analyses with uncertainty estimation

Ensemble forecast of analyses (EFA) couples a classical data assimilation method with sequential aggregation of ensemble forecasts. The assimilation method produces analyses whenever new observations become available. At the same time, an ensemble of given simulations is generated in order to forecast future time steps. The objective of EFA is to forecast the upcoming analyses (to be generated with the future observations) with a linear combination of the ensemble of forecasts. The weights of the linear combination depend on the ensemble simulation, on time and on the state component. Hence, for any forecast variable and at any location in space, EFA linearly combines the ensemble of forecasts to produce one single forecast of the upcoming analysis for the given variable at the given location.

In previous work, aggregation was carried out by machine learning algorithms that are well adapted to operational forecasting as they enjoy robustness properties. The approach is indeed applied on the French national air quality forecasting platform, Prév'air. Nevertheless, the method does not provide uncertainty estimations. Instead of machine learning, we propose to apply a minimax filter on the aggregation weights, which allows us to preserve the forecast performance while introducing uncertainty estimation along with the forecasts. The approach forecasts the upcoming analyses, which are the best a posteriori representation of the real state in some sense, significantly better than any model. A Kalman filter can also be applied on the aggregation weights; we will discuss how it compares to the minimax filter in this context.

The method will be illustrated for the forecast of peak ozone concentration fields over Europe, based on an ensemble of chemistry-transport models.

11.50 C. Santos: Agencia Estatal de Meteorología (AEMET) – Spain. csantosb@aemet.es
A. Amo, E. Abellán, A. Callado, P. Escribà, J. Sancho, J. Simarro

AEMET-SREPS: past, present and future

The Spanish meteorological Agency (AEMET) Short-Range Ensemble Prediciton System (AEMET-SREPS), was an European pioneer last decade as a multi-model multi-boundaries LAMEPS. Deterministic models are currently going down through convective scales, and ensembles should follow. However, new predictability issues arise at these scales and the design of an EPS on the gamma-mesoscale needs further insights. Current research lines include Stochastic Perturbed Parameterization Tendency (SPPT) for model error sampling, Local Ensemble Transform Kalman Filter (LETKF) for IC uncertainties, perturbation of LBCs testing different GCM or global ensemble combinations, and also new verification methods able to cope with spatial patterns, such as the Method for Object-Based Diagnostic Evaluation (MODE). Here it is presented a historical review of the system, as well as an outlook to the main research lines to build a feasible AEMET-gamma-SREPS.

12.10 Ernest Koffi (Laboratoire des Sciences du Climat et l'Environnement, Institut Pierre Simon Laplace des sciences de l'environnement, Gif-sur-Yvette, France), S. Kuppel, **P. Peylin**, F. Chevallier and D. Santaren

Variational vs. ensemble strategies to optimize the uncertain parameters of a terrestrial biosphere model

Terrestrial ecosystem models can provide major insights into the responses of Earth's ecosystems to environmental changes and rising levels of atmospheric CO2. To achieve this goal, biosphere models need mechanistic formulations of the processes that drive the ecosystem functioning from diurnal to decennial time-scales. However, the subsequent complexity of model equations is associated with unknown or poorly calibrated parameters which limit the accuracy of long-term simulations of carbon or water fluxes. We have developed a data assimilation framework to constrain the parameters of a state of the art mechanistic vegetation model (ORCHIDEE) which is involved in the AR5-IPCC report. In this study, we assimilate eddy-covariance observations of CO2 and latent heat fluxes and investigate the relative benefits of two different approaches to minimize the Bayesian cost function. The first one belongs to the family of the gradient-based methods and relies on the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm to solve the non-linear optimization problem with bounds imposed on some of the variables. In this case, the local gradients are computed accurately with the tangent linear model of ORCHIDEE. The main drawback of this approach is that it may end up on local minima instead of locating the global minimum. Our second method to minimize the cost function is based on an ensemble approach. We use a genetic algorithm that operates a stochastic search over the entire parameter space. It manipulates an ensemble of realisations of the state vector and makes them evolve towards the optimal solution. This approach is simpler to implement than the first one and less sensitive to local minima but potentially much more expensive in computational terms. In this study, we compare the performances of the two approaches in terms of model-data fit and estimated parameters (value and uncertainty).

12.30 Ehouarn Simon: NERSC - Norway.

Annette Samuelsen, Laurent Bertino, Dany Dumont

Estimation of positive sum-to-one constrained parameters with ensemble-based Kalman filters: application to an ocean ecosystem model

Ocean biogeochemistry models are sensitive to numerous and poorly known parameters. In addition, the strong non-linearity of the equations, the constraints on the state variables and parameters, and the regional variations are challenging for classical optimization methods. This motivates the use and further development of ensemble data assimilation methods for combined state-parameter estimation in coupled physical-biogeochemical systems. We consider the estimation of positive sum-to-one constrained parameters. However, the sum-to-one constraint cannot be guaranteed by ensemble-based Kalman filters when parameters are bounded. Therefore, we investigate two types of variables transformations for the estimation of positive sum-to-one constrained parameters that lead to the estimation of new set of parameters with normal or bounded distributions. These transformations are illustrated and discussed with the estimation of the grazing preferences parameters of zooplankton in ocean ecosystem models. These parameters are introduced to model the relative diet composition of zooplankton species. Experiments are performed with the 1D coupled model GOTM-NORWECOM with Gaussian anamorphosis extensions of the deterministic ensemble Kalman filter.

12.50 End of Session. Lunch Break

Session 11. ***Evaluation and Validation***

Chair: Andrew Lorenc

14.30 Mark Buehner (invited, Environment Canada, Canada), Josée Morneau and Cécilien Charrette

A comparison of Ensemble-Variational data assimilation and 4D-Var for global deterministic weather prediction

A new approach, called Ensemble-Variational data assimilation (En-Var), is tested and compared with the Canadian operational global atmospheric 4D-Var. The En-Var approach relies on the four-dimensional ensemble-based background error covariances produced by the ensemble Kalman filter (EnKF) and a variational minimization to produce a single deterministic analysis. It is well-suited for producing the high-resolution analyses required to initialize global deterministic medium-range forecasts because: 1) it can efficiently assimilate the very large volume of observations currently available, 2) it is more computationally efficient than 4D-Var, and 3) it is more flexible with respect to modelling the background error covariances than traditional EnKF approaches. Verification results from using En-Var with covariances obtained by averaging the EnKF-based and traditional NMC-method-based covariances will be shown. These results will be compared with those obtained from the standard 4D-Var approach.

15.00 Simon Lang and Martin Leutbecher (ECMWF, Reading, UK)

On the reliability of ensemble spread generated by singular vectors

Perturbations of the initial conditions based on the leading singular vectors of the forecast model propagator are used to represent initial uncertainties in the ensemble prediction systems of the European Centre for Medium-Range Weather Forecasts. The singular vectors are calculated operationally with T42 spatial resolution. For the extra-tropics of each hemisphere the initial perturbations are generated from the leading 50 singular vectors.

To better understand the properties of ensemble forecasts initialised by singular vectors, the ensemble variance is compared with the variance of the ensemble mean error in subspaces defined by the leading evolved singular vectors. Starting from ensemble configurations that are perturbed only with singular vector initial perturbations, we discuss the statistical consistency of the ensemble forecasts and how this depends on the number of singular vectors used to generate the initial condition perturbations. Furthermore, it is investigated how the statistical consistency depends on the spatial resolution of the singular vector computations.

15.20 Matthias Grzeschik (WESS-Water & Earth System Science Competence Center c/o University of Tübingen, Germany), Thomas Schwitalla, Hans-Stefan Bauer and Volker Wulfmeyer

Capability of WRF 3D-Var and DART EnKF to represent covariance structures

Three dimensional variational data assimilation (3D-Var) and Ensemble Kalman Filter (EnKF) are two typical methods for data assimilation in the field of weather forecast. For global models, EnKF performed better than or at least comparable to 3D-Var as shown in recent publications. A few studies furthermore showed some advantages of the EnKF compared to 3D-Var for limited area models. The advantage of the EnKF is that the background error covariance matrix (B), which is not explicitly calculated for EnKF algorithm but implicitly represented by the ensemble, is sequentially updated during an assimilation cycle. For 3D-Var, however, a stationary (constant) B matrix is used instead.

The two methods use also different approximations in the representation of the B matrix. The ensemble representation is based on the assumption that a few ensemble members are enough to describe the uncertainties and correlations. The 3D-Var B matrix, however, needs to be inverted, which requires coordinate transformations and a-priori assumptions.

As a case study, the two methods were compared for an intensive observation period (IOP) during the COPS campaign. Beside using the same model (Weather Research & Forecast Modelling System, WRF), the same modelling domain, and the same types of observations, the initial background error covariance matrix for the 3D-Var is created from the same ensemble as the

EnKF is applying to perform a fair comparison.

An ensemble driven by a global ensemble taken from the ECMWF was simulated over a spin-up period of 24h to generate realistic correlations for the WRF model domain. Based on this ensemble, mean and B matrix were calculated for the WRF 3D-Var. The comparison between 3D-Var and Ensemble Kalman Filter (EnKF), implemented in the data assimilation research test-bed (DART), was performed for a 24h assimilation cycle with 3h analysis intervals. The choice of observations to be assimilated were determined by DART and therefore limited to surface and upper-air observations.

15.40 Giovanni Leoncini (Met Office, Exeter, UK) and **Nigel. Roberts**

Convective scale ensemble at the Met Office Part II: new tools for assessment and evaluation

Many weather services nowadays provide model based ensemble forecasts of the weather with lead times ranging from a few hours to a few days. Such models in essence solve, numerically, a discretised version of the equations of motion on a gridded set of points. Most ensemble forecasts make use of models whose grid has characteristic spacings of the order of ten kilometres or larger (i.e. large scale). However increasing computing power has made high resolution ensemble forecast feasible with grid spacings of the order of 1km (i.e. the convective scale). As a result a few weather services run or are planning to run convective scale ensemble forecasts. Among these is the Met Office. This talk builds on the ideas of Roberts and Leoncini (Part 1) to show that convective scale ensembles have much larger variances and this renders ineffective many of the tools and diagnostics that have proven very useful for the large scale ensembles. The second part of the talk proposes new scale selective diagnostics and tools to evaluate and assess convective scale ensembles. Such diagnostics are based on the Fraction Skill Score and aim to assess the spread-skill relationship for the convective scale. This study is based on six cases encompassing different weather regimes and uses the Unified Model of the Met Office, but the conclusions are very likely to be model independent.

16.00 Tilmann Gneiting (Heidelberg University, Germany), Roman Schefzik and Thordis Thorarinsdottir

Ensemble copula coupling: Towards physically consistent, calibrated probabilistic forecasts of spatio-temporal weather trajectories

Ensemble forecasts call for statistical post-processing, in that model biases and dispersion errors need to be addressed. Methods such as Bayesian model averaging (BMA) and ensemble model output statistics

(EMOS) achieve this for single weather variables at single locations and single look-ahead times, but fail to recognize spatial, temporal and inter-variable dependence structures.

We describe a three-stage procedure, tentatively called ensemble copula coupling (ECC), that addresses this challenge, by generating post-processed ensemble forecasts of spatio-temporal weather trajectories:

1. Apply state of the art statistical post-processing techniques to obtain calibrated and sharp predictive distributions for each weather variable, location and look-ahead time individually.
2. Draw a discrete sample from each univariate, post-processed predictive distribution.
3. Generate a post-processed ECC ensemble of spatio-temporal weather trajectories by using the (discrete) ensemble copula.

The key idea is that the ECC post-processed ensemble inherits the multivariate rank dependence structure from the raw ensemble, thereby capturing the flow dependence.

We study various variants of the ECC approach, all of which are straightforward computationally, and describe relations to the Schaake shuffle and other extant techniques. An application to the ECMWF ensemble is given in a companion talk by Roman Schefzik.

16.20 End of Session. Closure of Conference

16.30 End of Conference