REQUEST FOR A SPECIAL PROJECT 2013–2015

MEMBER STATE: United Kingdom

Principal Investigator[1]: Tim Palmer

Affiliation: Oxford University

Address: Department of Physics
Atmospheric, Oceanic and Planetary Physics
Clarendon Lab
Parks Road
Oxford OX1 3PU

E-mail: T.N.Palmer@atm.ox.ac.uk

Other researchers:

Antje Weisheimer, Mirek Andrejczuk, Laure Zanna

Project Title:

Representing uncertainty in ocean observations and the ocean model for extended-range ensemble prediction

If this is a continuation of an existing project, please state the computer project account assigned previously.	**SPGBTPUO**	
Starting year: (Each project will have a well defined duration, up to a maximum of 3 years, agreed at the beginning of the project.)	2012	
Would you accept support for 1 year only, if necessary?	YES ☐	NO ☐

Computer resources required for 2013-2015: (The maximum project duration is 3 years, therefore a continuation project cannot request resources for 2015.)	**2013**	**2014**	**2015**
High Performance Computing Facility (units)	5,000,000	8,000,000	
Data storage capacity (total archive volume) (gigabytes)	6,000	10,000	

An electronic copy of this form ***must be sent*** *via e-mail to:* *special_projects@ecmwf.int*

Electronic copy of the form sent on (please specify date): 11/04/2012

Continue overleaf

[1] The Principal Investigator will act as contact person for this Special Project and, in particular, will be asked to register the project, provide an annual progress report of the project's activities, etc.

Principal Investigator: Tim Palmer

Project Title: Representing uncertainty in ocean observations and the ocean model for extended-range ensemble prediction

Extended abstract

It is expected that Special Projects requesting large amounts of computing resources (500,000 SBU or more) should provide a more detailed abstract/project description (3-5 pages) including a scientific plan, a justification of the computer resources requested and the technical characteristics of the code to be used. The Scientific Advisory Committee and the Technical Advisory Committee review the scientific and technical aspects of each Special Project application. The review process takes into account the resources available, the quality of the scientific and technical proposals, the use of ECMWF software and data infrastructure, and their relevance to ECMWF's objectives. - Descriptions of all accepted projects will be published on the ECMWF website.

Special project application - Project description:
"Representing uncertainty in weather and climate prediction"

1. Introduction

The basis of this proposal is to improve the reliability of extended-range forecasts of weather and climate. Broadly speaking, we define the notion of extended range as that range for which the use of coupled ocean-atmosphere models is crucial. Specifically, the proposal is in two halves. In the first half we describe the development of tools to improve the reliability of forecasts arising from a better representation of observation and model uncertainty in the ocean component of the coupled model system, both at data assimilation time, and during the forecast period itself. In the second part of the proposal we describe a series of proposed numerical experiments which will quantify the individual impacts of these different developments, on the reliability of the forecasts. These forecasts will primarily be on monthly and seasonal timescales.

To be more specific, the first half of the proposal describes an extension of the Ensemble-of-Data-Assimilations (EDA) methodology (*Isaksen et al, 2010*), recently implemented into the ECMWF operational medium range forcecast system, into the coupled forecast system. Specifically this will mean firstly implementing a "perturbed observation" methodology into the NEMOVAR ocean data assimilation system, and secondly implementing the "stochastically perturbed parametrisation tendency (SPPT)" scheme into the NEMO ocean model. This will have enormous benefit for the NEMO ocean modelling community. These developments will be discussed more in section 2-3 of this proposal. A series of numerical experiments using the ECMWF System 4 seasonal forecast system will be described in section 4, which will test the relative impact of these two individual implementations (observation uncertainty, model uncertainty) in monthly, seasonal and decadal forecast mode.

2. Ensembles of Data Assimilations

In this section we review some history in ensemble forecasting for medium range prediction, leading to the recent implementation of EDA. This is relevant in considering the specific proposals made in Sections 3 and 4 for the ocean component of a coupled model system.

It is commonly said that there are two sources of uncertainty for a weather forecast: initial uncertainty and model uncertainty. For longer-term climate forecast one has also to be concerned about greenhouse gas scenario uncertainty, but that will not concern us here. This delineation into initial and model uncertainty is extremely misleading as it implies that model uncertainty is somehow disjoint from initial uncertainty. Nothing could be further from the truth, as recent medium-range EDA ensemble prediction system (EPS) experimentation has demonstrated (M. Leutbecher personal communication). That is to say, based on the implementation of stochastic parametrisation (*Palmer et al 2009*) into the medium range EDA suite, a substantial part (c 15% of Z500 variance – M. Leutbecher, personal communication) of the initial uncertainty actually arises from model uncertainty in the data assimilation suite. On this basis, it makes sense to think of three sources of forecast uncertainty: observation uncertainty, the model-component of initial uncertainty, and model uncertainty during the forecast period.

The notion that model uncertainty contributes substantially to initial uncertainty provides some justification for the use of singular vectors as a way to perturb initial conditions as part of an ensemble weather forecast. That is to say, techniques which were based on the premise that initial uncertainty was purely due to observational uncertainty produced initial-time ensembles that were massively inflated over the values associated with pure observation uncertainty. The fact that model uncertainty becomes more substantial as one approaches the truncation limit of a model was consistent with the fact that the singular vectors (eg based on an energy metric) had significant amplitude at sub-synoptic scales.

With the development and implementation of an EDA system at ECMWF which includes model uncertainty, the amplitude of singular vectors has been reduced by 50% operationally in the EPS. It has not been possible so far to completely eliminate them, though it is possible that they may not be needed in the future as the stochastic parametrisations improve, and the development of correlated observation error statistics are made. However, even if singular vectors are dropped operationally, they will continue to play a major diagnostic role for the development of the EPS, and for this reason singular vector play a role in this proposal too.

As we move from medium-range weather prediction to extended range prediction where coupled ocean atmosphere models are needed, then the delineation of sources of uncertainty can be further expanded to include:

a) Uncertainty in atmospheric (and land surface) observations
b) Contribution of atmospheric (and land surface) model uncertainty to initial conditions
c) Atmospheric (and land surface) forecast model uncertainty
d) Uncertainty in ocean observations
e) Contribution of ocean model uncertainty to initial conditions
f) Ocean forecast model uncertainty

Broadly-speaking, the current ECMWF EDA –EPS system contains the elements a)-c). It is proposed here to expand this to include d)-f). This will benefit not only ECMWF, but the whole climate forecast community, and in particular, any group using the NEMO ocean model.

3. Development of EDA to the Coupled System

In this Section we expand on the development aspects of the proposal.

a) Representation of uncertainty in ocean observations

It is proposed to develop a stochastic scheme to represent ocean observation uncertainty, in much the same way as atmospheric observation uncertainty is represented in the atmospheric EDA. As in the NEMOVAR system, we propose to assimilate temperature and salinity fields from ARGO floats, moorings, XBT and CTD profiles and sea level anomalies from altimetry. For each member of the ensemble, random perturbations will be added to the observations. The perturbations will be drawn from a Gaussian distribution with zero mean and standard deviation equal to the estimate of observation error standard deviation. To provide an accurate forecast, one needs to include the spatial correlation of the observation errors. Observations of temperature and salinity in the ocean are taken in a vertical column, apart from a small horizontal drift, therefore the vertical correlation of the observation errors will be taken into account when perturbing observations from ARGO, moorings, XBT or CTD profiles. For satellite altimetry data, the horizontal correlation of the observation errors will be considered. The covariance statistics of the ocean EDA will be used to determine the spatial correlation of the observation errors. A simple way to estimate spatially correlated uncertainties will be to consider small random perturbations to the position the observations are made.

b) Representation of model uncertainty

One of the major developments in weather and climate modelling over the last years, has been that of stochastic parametrisation (*Palmer, 2001*). There are broadly, two classes of stochastic parametrisation at ECMWF: the Stochastically Perturbed Parametrisation Tendency (SPPT) scheme, and the Spectral Stochastic Backscatter Scheme (SPBS). SPPT can be thought of as an attempt to represent uncertainty in existing parametrisations, and is essentially a set of AR-1 processes acting on individual spherical harmonic components (eg of temperature), which multiply the net parametrisation tendency from the sum of all sub-grid processes. By contrast, SPBS represents a specific physical process which, it is suggested, is missing in conventional deterministic models. See *Palmer et al (2009)* for a detailed technical description of SPPT and SPBS.

From a technical point of view, the SPPT scheme is relatively straightforward to implement in an ocean model such as NEMO. Hence it is proposed to so do, based on a multiplicative noise scheme using diabatic tendencies. In order to keep matters straightforward in the initial phase of implementation, it is proposed to return to the simple scheme developed *by Buizza et al (1999)* based on stochastic sampling from a uniform distribution over (say) a 10x10 degree box with specified temporal autocorrelation.

Whilst it is likely that an equivalent form of SPBS is appropriate for the ocean, this lies outside the scope of the current proposal, and would require a separate study. By way of justification, it can be noted that at the time of writing, the ECMWF EDA system only contains SPPT and not SPBS (though it is expected that SPBS will be implemented into EDA in 2011 or 2012).

4. Experimentation

We now come to the heart of this project: to examine the extent to which more reliable extended-range predictions can be made from coupled ocean models using the EDA methodology, but a) with explicit representations of observation and model uncertainty in the ocean initial conditions, and b) with explicit representation of ocean model uncertainty in the forecast itself.
We propose a set of experiments with the ECMWF System 4 coupled model system to perform the following experiments.

a) Ensemble hindcasts using atmospheric EDA only (with SPPT in the atmospheric model)
b) As a) but additionally with ocean observation uncertainty
c) As a) but additionally with SPPT in the ocean
d) As a) but additionally with ocean observation uncertainty and SPPT in the ocean

The length of the forecasts, the resolution, and the ensemble size, remain to be determined, and strongly depend on available resources. It is likely that the initial phase of experimentation will be

based on relatively small ensemble sizes (c. 10) run at T159 resolution to a year ahead. Depending on progress, longer, higher resolution integrations may be possible.
Certainly a key measure of performance here will be a simple spread/skill measure. That is to say, the spread of a well balanced EPS about its ensemble mean, should equal the rms error of the ensemble mean forecast. Using analysis of variance, a key output of our experimentation will be an estimate of how ocean observation uncertainty and ocean model uncertainty at initial time, respectively, contribute to the balance between ensemble spread and error at forecast time (eg 3,6,9,13 months) for various atmospheric and oceanic variables. We will also quantify the extent to which oceanic SPPT in the forecast model contributes to a well balanced ensemble system.
Returning to the opening paragraph of this proposal, we will study from these experiments how representation of uncertainty in ocean observations and representation of ocean model uncertainty in the data assimilation process, improves the reliability of forecasts (eg of atmospheric variables). Reliability is a standard probabilistic metric of performance. Essentially a probability forecast is reliable, if for forecasts of a binary event E, which lie in a probability bin labelled by probability P, then the corresponding frequency of occurrence of E equals P. Hence, for example, in a sample of forecasts where the forecast probability of E was 60%, E should have occurred on 60% of occasions.

5. Technical requirements

The estimate for the computer resources requirements is based on the costs of running the ECMWF seasonal forecasting system of the IFS CY36R4 coupled to NEMO ocean model, similar to the envisaged System 4. For the (low) horizontal resolution of T95 and 62 vertical levels in the atmosphere coupled to a 1 degree version of NEMO, a single integration over 1 month costs approximately 450 System Billing Units (SBU) on C1A or C1B and the total archive volume are approximately 220 MBytes. The corresponding numbers for the horizontal resolution T159 also with 62 vertical levels and coupled to a 1 degree NEMO version are approximately 700 SBU and 800MBytes, respectively.

Given that a typical seasonal forecast experiment would run for at least two start dates (one for the summer and one for the winter season), approximately 20 years of hindcasts over a length of 4 to 7 months with an ensemble of ~10 members, the total costs for one such experiment would be around 1,000,000 SBU for T95 and up to 2,000,000 SBU for 7 month long hindcasts in T159 with data storage requirements of 500,000 Mbytes and 2,300,000 Mbytes, respectively.

References

Buizza, R., M.J. Miller and T.N. Palmer, 1999: Stochastic simulation of model uncertainties in the ECMWF Ensemble Prediction System. *Q.J.R.Meteorol. Soc*., **125**, 2887-2908.

Isaksen, L., J. Haseler, R.Buizza and M.Leutbecher, 2010: The New Ensemble of Data Assimilations, ECMWF Newsletter, **Spring 2010**, 17-21

Isaksen, L., M. Fisher & J. Berner, 2007: Use of analysis ensembles in estimating flow-dependent background error variance. In Proc. ECMWF Workshop on Flow-dependent Aspects of Data Assimilation, 11–13 June 2007, 65–86.

Palmer, T.N., 2001: A nonlinear dynamical perspective on model error: a proposal for nonlocal stochastic-dynamic parametrisation in weather and climate prediction models. *Q.J.R.Meteorol.Soc*., **127**, 279-304.

Palmer, T.N, R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G.J. Shutts, M. Steinheimer and A Weisheimer, 2009: Stochastic parametrization and model uncertainty. *ECMWF Technical Memorandum,* **598**.