Weather prediction in a world of uncertainties: should ensembles simulate the effect of model approximations?

Roberto Buizza – ECMWF

This main topic of this workshop is the simulation of model uncertainties in ensembles designed to provide an estimate of the probability distribution function of analyses and forecast states. This is the context within which I will discuss the question posed above in this short communication.

Ensembles have proven, so far, to be the most effective way to provide a range of possible forecasts, thus complementing information about the most likely state with a confidence level. Ensembles, if accurate and reliable, provide more consistent (in time) and valuable information than single forecasts. To achieve greatest accuracy and reliability, the operational ensembles have been designed to simulate all the 'most relevant sources' of forecast error, which can be classified broadly as linked to initial condition (ICs) and to model uncertainties. The ICs' ones are due mainly to observations not being geographically uniform and being affected by measurement and representativeness errors, and to approximations and simplifications used in data assimilation. The model ones are linked to the fact that the equation of motion of the atmospheric flow are solved on a finite, discrete grid and include only an approximate description of the real physical processes.

In the early days on ensemble prediction (1980s and early 1990s), attention focused mainly on the simulation of ICs' uncertainties. In 1995, the Canadian global, medium-range ensemble was the first to include model uncertainties (*Houtekamer et al*, 1996, MWR 124). At ECMWF, the first stochastic scheme designed to simulate model uncertainties was implemented in 1999 (*Buizza et al*, 1999, QJRMS 125). Results from these two centres indicated that simulating model uncertainties was beneficial and improved accuracy and reliability. Following their examples, most of the operational ensembles have included model uncertainty schemes. Today, at ECMWF, two stochastic schemes are used to simulate model uncertainties: the Stochastically Perturbed Parameterized Tendencies (SPPT), an improved version of the original scheme with perturbations with up to 3 different scales, and the Stochastic Kinetic Energy Backscatter (SKEB) schemes (*Palmer et al*, 2007, ECMWF TM 540).

Following the Canadian example, since about 10 years ago ECMWF has been using ensembles also to estimate analyses' uncertainties, both for the atmosphere (say the wave-land-atmosphere) and the ocean. Considering the atmosphere, since 2008 an Ensemble of Data Assimilations (EDA; *Buizza et al,* 2008, QJRMS 134) has been used to give a measure of analysis' uncertainties, to provide flow-dependent background-error statistics to the ECMWF data assimilation systems, and to initialize the medium-range/monthly ensemble (ENS). Considering the ocean, an ensemble was used to produce the ocean analysis version 3, and is currently used to produce the operational Ocean Re-Analysis version 4 (ORAS4; *Balmaseda et al*, 2013: QJRMS, 139), which includes 5 members, generated perturbing the surface wind stresses. Also the ORAS4 ensemble members are used to initialize the ENS forecasts, since each of them is based on a coupled ocean-atmosphere model.

Ideally, the analysis and forecast ensembles should be consistent and have the same characteristics, to avoid initialization shocks and to initialize better all scales: the same number of members, the same, coupled model, with each forecast starting from one analysis, and with both ensembles using the same method to simulate model uncertainties. Full consistency would also allow diagnostics based on the analysis' ensemble to give us indications on how to improve the forecast ensemble, and vice-versa. We have not yet achieved full consistency, but we have been working hard to make two of these ensembles as consistent as possible (see Table A). Considering the atmosphere component, since March 2016 EDA and ENS use the same model version and horizontal resolution, albeit a different number of vertical levels. In terms of model uncertainties, the EDA uses a 1-time-scale version of the SPPT scheme, while ENS uses a 3-time-scale version of SPPT and SKEB. Furthermore, the EDA runs with a 12-hour delay and provides ENS only with a set of 25 perturbations (instead of 51 full model state), which are combined with the unperturbed high-resolution analysis and the singular vectors to generate the 51 ENS initial conditions. For the ocean component, both ORAS4 and ENS use the same version of the NEMO model with the same resolution. Ocean model uncertainties are not simulated in either ensemble. Finally, there are only 5 ocean analyses that are used to initialize the 51, coupled ENS forecasts.

Operational suites		Sources of uncertainty		
Type	**Hor. Resol. – Vert. levels – Fc length (days)**	**Obs**	**ICs**	**Model**
HRES	$T_{CO}1279$ (~9 km) - L137 – (0-10d)	--	--	--
H4DV	$T_{CO}1279$ (inner loops $T_{CO}255/319/399$) - L137	--	--	--
EDA	25 members: $T_{CO}639$ (~18km) - L137	δo	--	SPPT(1L)
ENS	51 members: $T_{CO}639$ (~18km) - L91 - (0-15d) $T_{CO}319$ (~36km) - L91 - (15-46d) - Ocean: NEMO ORCA100z42	-- --	$EDA^{25}+SVs^{50*Na}$ $ORAS4^{5}$	SPPT(3L) + SKEB --
S4	51 members: $T_{L}255$ (~80km) L91 - Ocean: NEMO ORCA100z42	-- --	SVs $ORAS4^{5}$	SPPT(3L) + SKEB --
ORAS4	5 members: NEMO ORCA 1 degree and 42 layers – Run with perturbed forcing fields			

Table A. Key characteristics of the ECMWF operational suites: the high-resolution forecast (HRES) and analysis (H4DV), the Ensemble of Data Assimilations (EDA), the medium-range/monthly (ENS) and the seasonal system-4 (S4) ensembles, and the ocean analysis ensemble (ORAS4). For the wave-land-atmosphere component (the Integrated Forecasting System, IFS), TcoNNN indicates a spectral-triangular truncation NNN with a cubic-octahedral grid; Lxx is the number of vertical levels (all suites have the top of the atmosphere at 0.01 hPa). Three sources of forecast error are simulated, linked to observations' errors (simulated in the EDA by perturbing the observations), initial-conditions (simulated both in ENS and S4 with two different methods) and model uncertainties. ORAS4, the ocean data assimilation, includes 5 members, which are used to initialize ENS and S4.

As part of this workshop, we will be discussing how to progress in the simulation of model uncertainties. It is worth recollecting few key recommendations that were made at three workshops held at ECMWF in 2005 on 'The representation of sub-grid scales', in 2007 on 'Ensemble Prediction', and in 2011 on 'Model uncertainty'. On diagnostic and evaluation, it was recommended to develop a methodology to diagnose the spectral energy transfer, to use coarse-graining strategies (with a factor of 10 difference in resolution) to determine the statistics that an effective stochastic scheme should generate, and to use initial tendencies and analysis increments to determine model error statistics. On the physical basis of model uncertainty simulation schemes, it was suggested to explore the physical basis of the stochastic schemes, to develop physical parameterisations that include explicitly model uncertainty estimations, and to apply 'falsification concepts' (does the model error scheme invalidate physical constraints?) in the scientific work.

I think that the recommendations listed above are still valid, and I would like to conclude by suggesting that we add 'consistency' between the analysis and forecast ensembles, as another goal to achieve. At ECMWF, last we have recently coupled the NEMO ocean model in ENS from day 0 because we have shown that this coupling improves the accuracy and reliability of our global, medium-range/monthly ensemble forecasts. Going back to the question that I posed at the start of this communication, my answer is affirmative: we should simulate all relevant sources of model uncertainties. Furthermore, I suggest that we aim to achieve full consistency and develop an Integrated Coupled Analysis and Forecast Ensemble (I-CAFÉ) that includes the same model uncertainty scheme(s) in both the coupled (ocean-wave-land-atmosphere to start with) analysis and the forecast elements, with forecasts' initial conditions given by the coupled ensemble of analyses.