1. Introduction to the System 3 Ocean Analysis

1.1 Ocean initial conditions, ocean reanalysis and data assimilation

The main purpose of the ocean analysis at ECMWF is to provide initial conditions for the extended range forecasts (seasonal and monthly). The ECMWF seasonal and monthly forecasting systems are based on a coupled ocean-atmosphere general circulation model that predicts both the lower boundary conditions (namely SSTs) and their impact on the atmospheric circulation. The quality of monthly and seasonal forecasts is determined by the various components of the system (the ocean initialization, the coupled model, the ensemble generation and the calibration strategy), which are closely interrelated:

- **Ocean initial conditions** are the main source of predictability at seasonal time scales, and their importance is critical in the prediction of ENSO (El Niño Southern Oscillation) and its impact on the climate system. The correct initialization of the upper ocean thermal structure is considered instrumental in the prediction of the tropical SST at seasonal timescales with dynamical models. At the monthly time scales, the prediction of phenomena such as the MJO (Madden Julian Oscillation) requires the correct representation of the ocean-atmosphere interactions.

- **A historical ocean reanalysis** is required to provide initial conditions for the calibration of the seasonal forecasts. The a-posteriori calibration of model output requires an estimate of the model climatology, which is obtained by performing a series of coupled hindcasts during some historical period (typically 10-20 years). A historical record of hindcasts is also needed for skill assessment. The interannual variability represented by ocean reanalysis will have an impact on both the calibration and on the assessment of the skill.

- **An ensemble of ocean analyses** (five in total) is performed to sample the uncertainty in the ocean initial conditions. The ensemble of ocean initial conditions provided by the five analyses contributes to the creation of the ensemble of forecasts for the probabilistic predictions at monthly and seasonal ranges.

Ocean initial conditions for the global ocean could in principle be estimated by forcing an ocean model with atmospheric fluxes of heat, momentum and fresh water. However both ocean models and atmospheric fluxes are far from perfect, and the estimation thus obtained (first guess) can have substantial uncertainty. In order to improve the estimation of the state of the ocean, this first guess is combined with ocean observations via a data assimilation procedure. At ECMWF, the ocean model is HOPE (Hamburg Ocean Primitive Equation) and the assimilation system is based on an OI (Optimal interpolation) scheme.

Data assimilation has a large impact on the mean state of the first guess, and consistently reduces the bias. The impact is more especially noticeable in the tropics, where both the mean state and the interannual variability is improved of the ocean analysis is improved by data assimilation.

Data assimilation has a favourable impact on the skill of seasonal forecasts of SST, especially in the western Pacific, where the forecast skill is improved, especially in the first 3 months, is improved by using data assimilation in the initialization of the ocean (Alves et al 2003, Balmaseda et al 2007).

1.2 The System 3 ocean analysis system

The new operational ECMWF ocean analysis system (system 3 or S3) consists of two analysis streams:

- A historical reanalysis from 1959 which is 09.03.2007e (BRT). It is used to initialize seasonal forecasts.
- An early delivery ocean analysis, produced daily in near real time (NRT), used to initialize the monthly forecasts.

Although the historical reanalysis goes back to 1959, only the period 1981-2005 is used to initialize the calibrating hindcasts of the S3 seasonal forecasting system (Anderson et al., 2007). The earlier period of S3 ocean analysis will be used to initialize seasonal and decadal predictions within the ENSEMBLES project. As well as providing initial conditions for coupled model forecasts, the S3 ocean re-analysis, based on the synthesis of surface and subsurface ocean observations, surface fluxes from atmospheric analyses and reanalyses, and a general circulation ocean model, constitutes an important resource for climate variability studies.

The S3 ocean analysis has several innovative features, including an on-line bias correction algorithm, the assimilation of salinity data on T-surfaces and assimilation of global seas level trends. Two main criteria have been considered in the design of the assimilation algorithm: making optimal use of the observation information at the same time as avoiding spurious climate variability in the resulting ocean reanalysis09.03.200709.03.2007is given in Balmaseda et al. 2007.

The surface fluxes of heat, momentum and fresh water are an important component of the ocean analysis system. In S3, these are provided by ERA40 for the period 1959-2002 and by the operational system thereafter (ERA40/OPS). The five simultaneous ocean analyses are created by adding perturbations, commensurate with the estimated uncertainty, to the wind stress while the model is being integrated forward from one analysis time to the next. The wind perturbations have been revised in S3 to represent the perceived uncertainty in ERA40/OPS wind stress.

1.3 History of ocean analysis at ECMWF

ECMWF has produced operationally daily global ocean analyses to provide initial conditions for the seasonal forecasting system since 1997. There have been two versions of the ocean analysis, linked to the operational seasonal forecasting system. System 1 (S1) started in 1997 (Alves et al., 2003) and provided the initial conditions for the first ECMWF operational seasonal forecasting system (Stockdale et al., 1998). System 2 (S2) was introduced in 2001 (Balmaseda 2004), and has provided initial conditions for the ECMWF operational seasonal forecasts since 2002 (Anderson et al. 2003, Oldenborgh et al. 2005a,b, Vialard et al. 2005). A comparison between S2 and S1 ocean analyses is given in Balmaseda 2004. In 2004 an extension of S2 was introduced in order to initialize the monthly forecasting system (Balmaseda 2005, Vitart 2005). In summer 2006 the S3 ocean analysis was implemented operationally (Balmaseda et al 2007).
Although originally developed solely to provide ocean initial conditions for the seasonal forecast system, the scope of the ocean analysis at ECMWF has been slowly widening with time. Initially the length of the historical record of ocean initial conditions was not too long, since the hindcasts were mainly used to estimate the bias of the coupled model, which, although seasonally dependent, could be robustly estimated with a limited set of integrations. For instance, in the first seasonal forecasting system (S1), only 5 years of calibrating hindcasts were used. However, it was also necessary to provide an estimation of the forecast skill, and that required a larger historical sample, including as wide a range of climate conditions as possible. There was also the realization that calibrating not only the mean, but also the variance of the coupled model forecasts could lead to improved reliability of the seasonal forecasts. Besides, with the advent of multi-model activities it is clear that the robust bayesian combination and calibration of multi-model forecasts needs long records of realizations. All of these applications pointed towards the need for a long historical ocean reanalysis that could provide consistent initial conditions for the “calibrating” coupled hindcasts. And with a long record it is possible to start trying some decadal forecasts, as in the ENACT and ENSEMBLES projects, which try to assess the predictability at the decadal time scales.

At the other end of the spectrum, moving towards the shorter time scales, we have the monthly and medium range forecasting activities, where the demand for ocean initial conditions is increasing. Monthly forecasting activities have stirred quite some interest in the last few years, and the benefits of having an active ocean in the forecasting system has been demonstrated (Vitart et al. 2006, Woolnough et al. 2007). The monthly forecasting system uses the same ocean model as the seasonal forecasting system though there are some differences in the way the ocean initial conditions are produced. In the near future it is planned that the medium range EPS forecasts will also be performed with a coupled model, with the consequent need for real-time ocean initial conditions. The EPS will also have a need for the historical ocean reanalysis, since the reliable prediction of extreme events also requires hindcasts or re-forecasting activities just as is done now for monthly and seasonal forecasting.

In order to accommodate these different demands, the S3 ECMWF ocean analysis has been designed to deliver two kinds of analysis products: a delayed product 11 days Behind Real Time (BRT), used to initialize the seasonal forecasts, and an early delivery product, which runs in Near Real Time (NRT), to be used by the monthly forecasts, and which in the near future is envisaged to be used by the medium range forecasting system as well.

2. The Ocean Data Assimilation System

As for the previous systems, the ocean data assimilation system for S3 is based on the HOPE-OI scheme: The first guess is given by forcing the HOPE (Hamburg Ocean Primitive Equations) ocean model with daily fluxes of momentum, heat, and fresh water, while the observations are assimilated using an Optimal Interpolation (OI) scheme.

2.1 The analysis cycle

The ocean analysis is performed every 10 days. All the observations within a centered 10-days window are gathered and quality controlled. In S3 major upgrades have been introduced in the HOPE-OI system. In addition to subsurface temperature, the OI scheme now assimilates altimeter derived sea-level anomalies and salinity data. All the observations in the upper 2000m are assimilated (in S2 only the observations in the upper 400m were used). More information about the origin of the observation streams and processing can be found here. Before the observations are combined with the model background via the OI, the model background is bias corrected according to the procedure described in Balmaseda et al 2007. Then, the different data streams are assimilated sequentially as follows:

1. **Assimilation of sea level anomaly maps** (from altimeter): The detrended altimeter-derived sea level anomalies are combined with the bias-corrected model first-guess using the Cooper and Haines 1996 scheme to produce a first analysis.
2. **Assimilation of subsurface temperature** (from ARGO, XBTs, Moorings): The result of the previous step is then used as a first guess for a second assimilation step, where only subsurface temperature data are assimilated, and salinity is updated by imposing conservation of the model temperature/salinity (T/S) relationship (Troccoli et al. 2002), while the sea level and velocity field remain unchanged.
3. **Assimilation of Salinity** (from ARGO, Moorings): In a third assimilation step, the information provided by the salinity observations is used to modify the model T/S relationship. In this step, the T/S information is spread along isotherms following the scheme of Haines et al., 2006. Only salinity is modified in this step which results in the analysis. After this 3rd assimilation step, velocity updates are derived from the temperature and salinity increments imposing geostrophic balance (Burgers et al., 2002).
4. **Assimilation of global sea level trend** (from from altimeter): Finally, the trend in global (area averaged) sea level is assimilated. By combining the altimeter-derived trend in global sea level with the model trend in global dynamic height, it is possible to make the partition between changes in the global volume and changes in the total mass. By doing so, the global fresh water budget is closed and the global surface salinity and sea level adjusted accordingly.

Analysis increments in temperature, salinity and velocity are calculated using the OI scheme and balance relationships. To avoid exciting gravity waves, and to allow the model dynamics to adjust gradually to the changes in the density field, an Incremental Analysis Update (IAU) method (Bloem et al., 1996) is used: the increment is added slowly over the subsequent 10 days (IAU-10), after which a new background field is available, and the cycle repeated.

2.2 The ocean model

The HOPE ocean model (Wolff et al., 1997) uses an Arakawa E grid horizontal discretization. Several modifications took place over the years at ECMWF (Balmaseda 2004, Anderson and Balmaseda 2006). The horizontal resolution was increased to 1 x 1 degrees with equatorial refinement, i.e., the meridional resolution increases gradually towards the equator, where it is 0.3 degrees in the meridional direction. There are 29 levels in the vertical, with a typical vertical thickness of 10 meters in the upper ocean compared to 20 levels. The vertical mixing is based on Peters et al., 1998. The barotropic solver, originally implicit, was made explicit as described in Anderson and Balmaseda (2006).

2.3 Observations and fluxes data sets
In S3, the observations come from the quality controlled dataset prepared for the ENACT and ENSEMBLES projects until 2004 (Ingleby and Huddleston 2006), and from the GTS thereafter (ENACT/GTS). The OI scheme is now 3-dimensional, the analysis being performed at all levels simultaneously down to 2000m, where observations are taken independently and only to 400m. In addition, the decorrelation scales depend on the density gradient, which favours the propagation of information along isopycnals. A pictorial view of the various data sets used in S3 is given in fig 1. The analysis of SST is not produced using the OI-Scheme. Instead, the model SSTs are strongly relaxed to analyzed SST maps. The maps are daily interpolated values derived from the OIv2 SST product (Reynolds et al 2002) from 1982 onwards. Prior to that date, the same SST product as in the ERA40 reanalysis was used.

In S3, altimeter data are assimilated for the first time in the ECMWF operational ocean analysis. The altimeter information is given by maps of merged satellite product, provided by Ssalto/DUACS and distributed by AVISO, with support from CNES. Twice a week (on Wednesday and on Saturday mornings) (1/3x1/3 degrees) Maps of Sea Level Anomaly (MSLA) for a merged product combining all satellites (Envisat, Jason, Topex/Poseidon, ERS2, GFO) using optimal interpolation and accounting for long wavelength errors are produced (Le Traon et al., 1998, Ducet et al., 2000).

The first-guess is obtained from integrating the HOPE ocean model from one analysis time to the next, forced by ERA40/OPS fluxes (ERA40 fluxes from the period January 1959 to June 2002 and NWP operational analysis thereafter). In S2 the fluxes were from ERA15/OPS, but the wind stresses were not directly used: instead, the wind stress was derived from the analyzed winds using an off-line bulk formula. The representation of the upper ocean interannual variability is improved when using the ERA40 wind stress (Uppala et al., 2006), although the stresses are biased weak in the equatorial Pacific. The fresh water flux from ERA-40 (Precipitation - Evaporation, denoted P-E) is known to be inaccurate. S3 uses a better but by no means perfect estimate, obtained by 'correcting' the ERA-40 precipitation values (Troccoli and Kallberg 2004).

3. The Ocean Observations and Quality Control

3.1 Subsurface data

Subsurface temperature observations that are available in near-real-time are currently provided by the TAO/TRITON and PIRATA arrays in the equatorial region (McPhaden 1998, Servain et al 1998) and the global Volunteer Observing Ship (VOS) programme which provides XBT (eXpandable BathyThermographs) measurements mainly along merchant shipping routes. More recently, observations are provided by the ARGO network of drifting profilers. Salinity data is available from ARGO and from the TRITON moorings.

Prior to 2004, the temperature and salinity profiles come from the quality controlled ENACT/ENSEMBLES data set (Ingleby and Huddleston 2006). From January 2005, the observations come from the GTS (Global Telecommunication System). An automatic quality control procedure (an extension of Smith et al., 1991) is then performed in several stages:

1. **Daily averaging.** First of all, daily averages are created if applicable: if some site reports more frequently than once per day, daily averages are created (this is the case for the TRITON moorings, which report hourly).
2. **Black list of coastal observations.** Data in the vicinity of the coast are rejected, as a way of accounting for representativeness error.
3. **Background check.** A level-by-level check between the distance between model values and observations in relation to the error statistics.
4. **Buddy check.** A consistency test between observations is performed.
5. **Super-obbing.** Profiles which are close in space and time are superobbed, following the same criteria as in Smith et al., 1991.

6. **Completeness of profiles.** in S3 there is an additional check for completeness of the profiles: a profile is considered incomplete, and therefore rejected, if the sparsity of the remaining observations in the vertical is judged insufficient to resolve the vertical temperature gradients. (An observation profile will be rejected if the temperature difference between consecutive levels is larger than 5 deg C or if it contains a vertical temperature gradient larger than 0.1 deg C/m).

The observation coverage and the quality control decisions for the different assimilation cycles can be seen here. The figures show that thanks to the ARGO system the coverage of salinity is now comparable to that of temperature, and it is almost global. In recent times the TRITON moorings in the West Pacific and Indian ocean also provide salinity in real time. The PIRATA and TAO moorings report only temperature in real time, even when the sensors are also able to measure salinity. Most of the data from XBTs and Mooring are superobbed, whilst the ARGO profiles are often partially rejected by our QC system.

### 3.2 Sea level data

Information about sea level is provided by the altimetric data. The altimeter information is given by maps of merged satellite product, provided by Ssalto /DUACS and distributed by AVISO, Twice a week (on Wednesday and on Saturday mornings) Maps of Sea Level Anomaly (MSLA) for a merged product combining all satellites (Envisat, Jason, Topex/Poseidon, ERS2, GFO). Prior to assimilation, these maps are smoothed to remove unrepresented features and interpolated onto the model grid. These are then interpolated in time to produce daily maps. Only the map corresponding to the centre of the assimilation window is assimilated (09.03.2007 referred to a 7 year mean (1993-2000). To enable comparison with the background field a reference mean sea level is required.

In S3, this is the 7-year mean sea level from an ocean analysis spanning the period 1993-2000. The possibility of using a reference mean sea level derived from the GRACE gravity mission has also been explored, but it was found that sensitivities to the reference mean sea level are large, and could potentially introduce abrupt jumps in the analysis.

The trend in global sea level is quite substantial. If this trend is produced by thermal expansion due to global warming, it can not be represented by the ocean model: in absence of fresh water fluxes, most ocean models used for climate are volume preserving, since they make use of the Boussinesq approximation. Therefore, if not treated correctly, the trend in sea level can be a problem when assimilating altimeter observations. In S3, the global sea level trend is removed from the altimeter sea level anomalies before they are assimilated via the GH96 scheme. In S3, the global sea level trend is later assimilated as described in Balmaseda et al 2007.

### 3.3 Sea Surface temperature

An accurate initial SST is of great importance for the accuracy of the forecasts. In the present system, we ... The analysis is linearly interpolated to daily values before use. Because the SST analysis is produced only once a week (on Mondays, with a valid central time of the previous Wednesday), and because we need the following Wednesday to produce the interpolated value for a Thursday, this implies a delay of up to 11 days in the availability of the SST field. Daily SST fields are available in near real time from the NWP system, but these are less stable, and there are significant inconsistencies in the data from different periods. We prefer to use the higher quality data, even if it is slightly delayed. Despite using what we consider to be the best SST analysis available, there are still significant errors, estimated to be of the order of 0.3 K.

The arrival of SST information is one of the major factors for introducing a delay in the analysis. The real-time analysis does not wait for the second map of SST to be available. Instead, a daily SST product is created by adding the latest SST anomaly to the daily climatology.

### 4. The Reanalysis and Real Time Streams

When in 2004 the monthly forecasting system became operational at ECMWF, it was necessary to produce near-real-time ocean initial conditions, since the typical 11 days delay of the reanalysis product was not adequate for the monthly system. To this end, an early delivery ocean analysis system was introduced in May 2004. Details of the timing of the near real time S2 ocean analysis are given in Balmaseda 2005. In what follows, we denote the Near Real Time stream as NRT, and for BRT we refer to the reanalysis, which runs 11 days Behind Real Time. The BRT and NRT analysis systems are almost identical: they only differ in the details of the assimilation cycle and in the earlier cut-off of the assimilation windows. The BRT product is a continuation of the historical ocean reanalysis, while the NRT only exists for the recent period.

In S3, the NRT ocean analysis is an integral part of the operational ocean analysis system. To avoid degradation of the NRT product, the NRT analysis always starts from the most recent BRT analysis and is then brought forward to real time every day. In principle, the design of the S3 ocean analysis system allows the paradigm of a seamless prediction system, since the forecasts for the different time scales (medium range, monthly and seasonal) could all be initialized using the NRT product. Equally, the BRT product would provide initial conditions for the hindcasts needed for the calibration (or re-forecasting) of the seamless products.

The figure below shows schematically the schedule followed in the production of the NRT and BRT ocean analyses in S3 (which is slightly different from that described in Balmaseda 2005). Every day, the BRT ocean analysis is advanced by 1 day, starting from the BRT analysis from the previous day (D0-12 in fig) to produce the BRT analysis at day D0-11. If the 10-day assimilation cycle is due on that day (D0-12), the OI analysis is performed using a 10-day centered window,
Assimilation cycle in the daily production of the behind-real-time (BRT) and near-real-time (NRT) ocean analyses in S3.

Similarly, the NRT ocean analysis is also produced daily. Starting from BRT (D0-12), the model is integrated forward 12 days, so bringing it up to real time. All the available observations during that period are used. During the 12 days there are always two assimilation cycles. The first assimilation takes place at (D0-12), using a 10-day window centred at D0-12 (W1), and the assimilation increment using the incremental update method is computed. This increment is only applied for the next 7 days, at which time the second assimilation W2 is performed (i.e. at D0-5), using observations from the 10-day window as shown in fig 5. The corresponding increment is then computed, and applied during the next 5 days of the integration. Effectively, W2 only has 9 days worth of data, since the daily averages for day D0 are not available at the time of the operational production. The assimilation windows W1 and W2 overlap for 3 days. The observations within this overlapping period, although used in two different analysis cycles are not given extra weight, since the increment of the first assimilation cycle is not applied to completion.

5. Ocean Analysis Products

The ECMWF ocean analysis system produces daily analysis of the global ocean for temperature, salinity, velocity, sea level and other derived variables. The reanalysis stream spans the period 1959 up to 11 days behind real time. The real-time stream started in August 2006.

The following products are displayed in the web:

- Horizontal maps of SST, Sea Level, Depth of 20 deg Isotherm, Surface Salinity, Zonal Wind Stress and Averaged Temperature and Salinity over the upper 300m.
- Zonal sections along the equator of temperature.
- Meridional temperature sections at 165E, 140W and 30W.
- Time-Longitude sections along the equator of SST, Sea Level, Depth of 20 degrees Isotherm, Zonal Wind Stress.
- Observation coverage maps and quality control decisions.

Both the full fields and the anomalies respect the 1981-2005 are displayed. The maps represent daily fields from the real-time analysis, and weekly and monthly from the reanalysis stream. The monthly fields go back to 1959. The weekly fields are only displayed since January 2007. For the real time, only the last 30 daily fields are displayed.

6. Ocean Archive

The operational ocean analysis is archived in MARS under stream=OCEA, expver=0001 and system=3.

As for all the ocean data archived in MARS, the leveltype=depth (dp).

The ocean reanalysis is archived as type=OR

The real-time ocean analysis is archived as type=AN

Time and date attributes

For the ocean analysis and reanalysis products, time=0,step=0, and date=YYYYMMDD as the verifying date.
Parameters

The ocean data is archived in GRIB using the ECMWF local table 2 version 151 and GRIB extension in definition 4. The most common parameters are:

- 129 for potential temperature
- 130 for salinity
- 131 for zonal velocity
- 132 for meridional velocity
- 133 for vertical velocity
- 145 for sea level
- 163 for depth of 20 degree isotherm
- 148 for mixed layer depth
- 164 for averaged temperature in upper 300m
- 175 for averaged salinity in upper 300m
- 153 for zonal wind stress
- 154 for meridional wind stress

Time products

There are different time products:

- PROD=INST for instantaneous fields.
- PROD=TACC for accumulated fields (only for OR).
- PROD=TIMS for timeseries (only for Reanalysis).
- PROD=TAVG for time averaged fields

Spatial Sections

There are different spatial sections:

- SEC=H for horizontal fields (i.e. x-y sections)
- SEC=Z for zonal sections (x-z or z-time)
- SEC=M for meridional sections (y-z or y-time)
- SEC=V for vertical sections (z-time)

Other Attributes

- method=1 (for data assimilation initialization)
- number=0 for centered unperturbed analysis. (n=1/2/3/4 for the perturbed analysis)
- system=3 (system 3 operational analysis)
- levelist: needed for H sections
- latitude: needed for Z sections
- longitude: needed for M sections
- range: time range (in hours) needed for TIMS and TAVG products.

Examples

- Mars retrieval of Sea Surface Temperature and Surface Salinity from the real time analysis, for the 2nd of March 2007. Only the centered analysis (number=0):

  ```
  retrieve,
  number=0,
  system=3,
  time=00:00:00,
  date=2007-03-02,
  section=h,
  levtype=dp,
  method=1,
  param=129.151/130.151,
  levelist=5.0,
  stream=ocea,
  expver=1,
  type=sn,
  product=inst,
  class=od
  ```

- Mars retrieval of accumulated sea level and depth of 20 deg isotherm from the ocean reanalysis for the 1st November 1982:
Mars retrieval of an instantaneous zonal section of temperature along the equator, from the reanalysis, all 5 ensemble members:

```
retrieve,
number=0/1/2/3/4,
system=3,
time=00:00:00,
date=1982-11-01,
section=z,
latitude=0,
levtype=dp,
expver=1,
method=1,
param=129.151
```

Mars retrieval of an instantaneous meridional section of zonal velocity at 140W, from the real-time analysis. The user needs to indicate the closer model longitude:

```
retrieve,
longitude=220.078,
number=0,
system=3,
time=00:00:00,
date=1982-11-07,
section=m,
levtype=dp,
expver=1,
method=1,
type=or,
product=inst,
class=od,
param=131.151
```

Mars retrieval of timeseries of averaged temperature in the upper 300m along the equator from the reanalysis, for the period 19870101-19870201:
References and further literature

Ocean Analysis References:


ECMWF Newsletter No. 105 - Autumn 2005.


Levitus S NODC (Levitus) World Ocean Atlas 1998 data provided by the NOAA-CIRES Climate Diagno09.03.2007 at http://www.cdc.noaa.gov/


