Joint IMD-WMO Group Fellowship Training

on

Numerical Weather Prediction

Managed by

India Meteorological Department (IMD), Pune,

04 Oct - 10 Nov 2021

Lecture Note on Ensemble data Assimilation technique, Hybrid data assimilation tech

by

Sumit Kumar

National Centre for Medium Range Weather Forecasting
Ministry of Earth Sciences
A-50, Sector-62 Noida-201309, UP, India

Outline

The purpose of this lecture is to introduce the reader to promising experimental methods for atmospheric data assimilation involving the use of ensemble forecasts by introducing the operational systems at NCMRWF, INDIA

- Understanding the natural linkage between data assimilation and ensemble forecasting: ensemble forecasts are designed to estimate the flow-dependent uncertainty of the forecast; data assimilation techniques require accurate estimates of forecast uncertainty in order to optimally blend the prior forecast(s) with new observations.
- Intent here is to demonstrate how these two endeavors are united and how and why this union may improve the quality of both initial conditions and ensemble forecasts.

Introduction

Data assimilation can broadly be defined as the mathematical discipline dedicated to the optimal reconciliation of a theoretical model with observed data: a primary application being state estimation of an imperfectly known dynamical system for example earth's atmosphere. It is often desired to make forecasts using a theoretical model that must be initialized using information about the current state of a dynamical system. We assume that the dynamical system under investigation may be probed via measurement; however, measurements may be sparse, due to cost and technological limitations, and are subject to error. To make accurate predictions, one must utilize all available information from current and past measurements, incorporate understanding of the dynamical system via a theoretical model, and account for all errors present in the forecasting process. At major operational weather forecasting centers around the world, hybrid data assimilation methods that combine a climatological (time-averaged) estimate of forecast errors with a dynamic (time-varying) estimate have been adopted as the primary approach for initializing numerical weather prediction (NWP) models. Hybrid systems have become the state of the art among data assimilation methods. These systems combine the benefits of two other systems that are traditionally used in operational weather forecasting: an ensemble-based system and a variational system.

1. NCMRWF Ensemble data Assimilation system

Initial approaches of Numerical Weather Prediction (NWP) were deterministic. Eady (1951) was first to express his concern about the strictly deterministic approach in NWP and advocated for probabilistic approach. The practical implementation of the approach that combines probability with determinism is called Ensemble Prediction.

An ensemble prediction system usually includes a control forecast and a good number of perturbed forecasts. The control forecast is one that starts from the best estimated state (based on available observations) of the atmosphere (analysis) prepared by the data assimilation system. Initial conditions for other ensemble members are generated by adding perturbations (or errors) to the analysis. During the early stage of the forecast, error grows more or less linearly with time and the deterministic forecast shows good skill. During this period the small error in the initial condition remains small and trajectories of the model forecast and the "truth" are close to each other in phase space. Beyond this range of linear error growth, deterministic forecast loses its skill but ensemble mean (or average) can be treated as a single forecast representing the best available estimate of the future atmosphere. By calculating the ensemble average the unpredictable components of the forecast are filtered out and those are retained that show agreement between the ensemble members. Spread in the forecast is a measure of disagreement between the ensemble members. Another important aspect of ensemble prediction is that it provides a quantitative basis for probabilistic forecasting.

Ensemble forecasting methods in different operational centres around the world mostly differ by the way in which initial condition perturbations are generated. NCMRWF uses Ensemble Transform Kalman Filter (ETKF) (Bishop et al., 2001) in its Global Ensemble Prediction System to generate initial perturbations. The objective of ETKF is to provide initial conditions for NCMRWF Ensemble Prediction System (NEPS) forecasts. It generates global perturbations for wind, temperature, humidity and pressure fields for the 22 ensemble members. In NEPS system, the perturbations generated by ETKF are combined with the operational 4D-VAR analysis so that a full Ensemble Kalman Filter (EnKF) analysis is not required. Implementation of EnKF is computationally expensive whereas calculation of transformation matrix in ETKF, which updates only the initial perturbation instead of updating the analysis, is much cheaper. The perturbations are added to the reconfigured analysis using the Incremental Analysis Update (IAU) scheme (Clayton, 2012) within the UM. The control forecast does not need any input perturbation from ETKF. It uses only the reconfigured deterministic analysis as its initial condition.

ETKF receives the forecast perturbations from the previous forecast cycle (T+6 state for the perturbed ensemble members) as input. The forecast perturbations valid at the new analysis time are mixed and scaled by ETKF to generate new set of mutually orthogonal analysis perturbations. The mixing of the evolved forecast perturbations is performed by the transformation matrix. The calculation of transformation matrix requires the model equivalent of each observation for each ensemble member, to provide the estimates of background uncertainty in observation space. These 'pseudo-observations' are calculated by the Observation Processing System (Figure 1) and provided to the ETKF (modelobs).

If the ensemble size is very small the background error covariance becomes large and the impact of observation is overestimated. This leads to unrealistically small analysis perturbations generated by ETKF. In order to counter this problem two methods are adopted: (1) horizontal localization and (2) covariance inflation.

In horizontal localization a number of equally spaced localization centres (currently 92) are defined around the globe. For each centre, a local transformation matrix is constructed by using the observations within a radius of 2000 km. Interpolation between the local transformation matrices for the nearest localization centres gives the final transformation matrix for each grid point. In this way longer range correlations in the error covariances are cut off at a specified distance. Rank of the analysis covariance estimate also gets improved by horizontal localization.

Further improvement in ensemble spread is made by multiplying the raw transformation matrix of each region by a region specific inflation factor. This inflation factor gets updated at each assimilation cycle. The inflation factor of the previous cycle is multiplied by the ratio of the root mean square (RMS) error of the ensemble mean with respect to observation to the RMS spread of the ensemble forecast. OPS provides the ensemble mean and spread through “modelobs” (observation equivalent from short-forecast) files and observation through varobs files processed against the control forecast (Figure 1).

The NEPS implemented in NCMRWF is fundamentally the same as the original MOGREPS developed at Met Office, UK. This global ensemble prediction system has a horizontal resolution of approximately 12 km and 70 vertical levels. A total of 23 ensemble members (22 perturbed forecasts and 1 control forecast) constitute this ensemble system. The 22 analysis perturbations for all the ensemble members are generated by ETKF system four times a day (at 00, 06, 12 and 18 UTC) from the previous 6 hr short forecast of the evolved perturbations for the variables u, v, θ, q and exner pressure on all levels. These analysis perturbations are added to the reconfigured analysis from the four–dimensional variational data assimilation system (4D-VAR) of Unified Model operational at NCMRWF (NCUM). A 10 day forecast of NEPS is routinely generated based on 00 UTC initial conditions which include a control forecast (Cntl) with the 4D-VAR analysis and 22 ensemble member forecasts with 22 perturbed initial conditions. The sequences of all the processes involved in NEPS operational at NCMRWF are represented by the flow diagram shown in Figure 1.

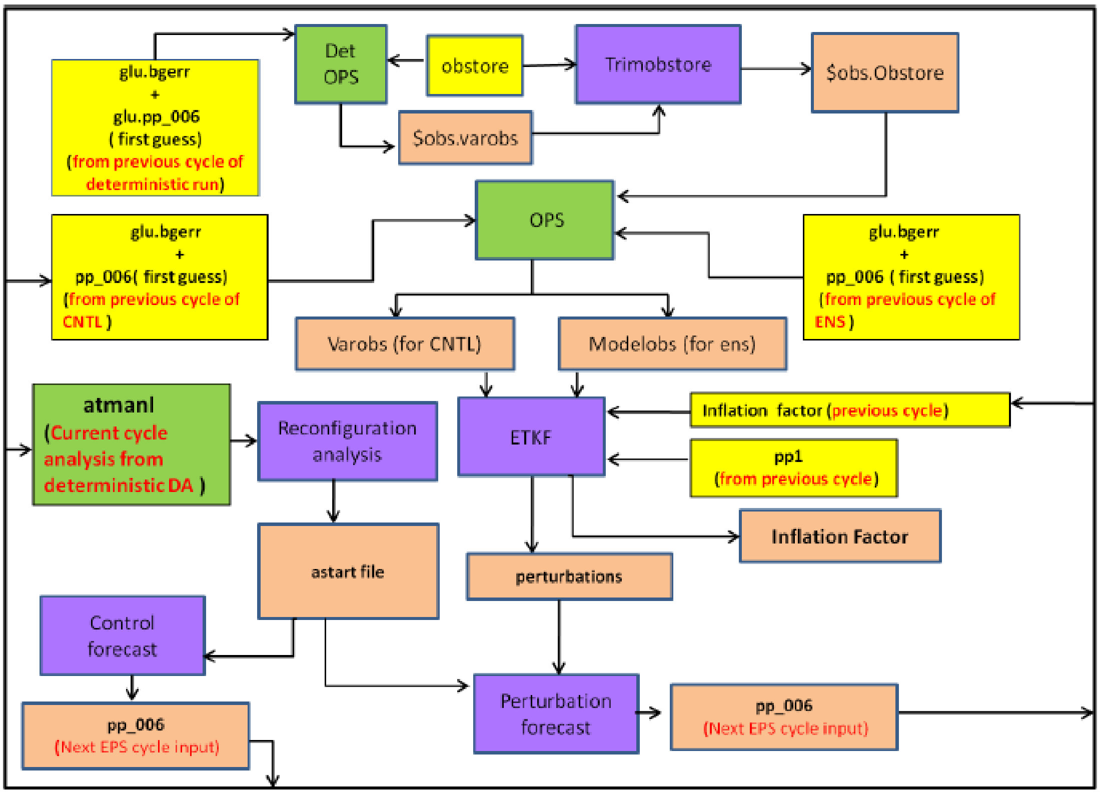

Figure 1 NEPS schematic

2. NCMRWF Hybrid Data Assimilation System at NCMRWF

NCMRWF is operationally using NCUM global NWP system for generating medium-range numerical forecast since 2012 (Sumit et al., 2020). NCUM system has been adapted from Unified Model (UM) seamless prediction system of "UM Partnership" and has been upgraded periodically to adapt scientific and technological advancements. The major components of the NCUM global NWP system include components for data processing, data assimilation and forecast model (George et al., 2016).

NCMRWF initially adapted UM's advanced 4D-Var Data Assimilation (DA) system for its global NCUM system in 2012. One of the weaknesses of this system was the use of a climatologically average background error. So in 2016, NCUM DA system has been upgraded to "Hybrid 4D-Var", which has the advantages of both traditional 4D-Var method and ensemble data assimilation approaches, providing flow-dependent forecast error covariances (Clayton et al., 2013). The term "hybrid" refers to the combination of climatological covariance and covariances calculated from an ensemble of forecasts, designed to sample the day to day varying uncertainty. The Hybrid 4D-Var system combines the advantages of traditional 4D-Var and the ensemble data assimilation. NCUM hybrid 4D-Var system uses the ensemble forecasts from the NCMRWF ensemble prediction system (NEPS) which uses ETKF (Ensemble Transform Kalman Filter) method for perturbation generation. The use of Hybrid 4D-Var method for data assimilation in NCMRWF is a major step towards the application of emerging paradigm of ensemble data assimilation in operational NWP in India.

Table 1 presents the details of currently operation DA-analysis system at NCMRWF. NCMRWF receives worldwide meteorological observations through Global Telecommunication System (GTS) via Regional Telecommunication Hub (RTH) at IMD, New Delhi. A large volume of satellite observations is being received through internet data services from various satellite data producers (NOAA/NESDIS, EUMETSAT, ISRO etc.) directly at NCMRWF. The NCUM Observation Pre-processing System (OPpS) (Figure 1 green boxes) packs all observations received at NCMRWF in a specific format known as "obstore", which can be read directly by the Observation Processing System (OPS) of NCUM. The Observation Processing

System (OPS) read the decoded observations packed by the OPpS. The OPS system has two components, the “extract” component of OPS retrieves the observations available in and calculates background values at the observation locations from the model background fields. The components of the OPS perform the quality control of observations and reformat them for their use in the Hybrid 4D-Var. It processes and packs observations for each 6 hourly window for the four data assimilation cycles in a day centered at 00, 06, 12 and 18 UTC.

Table 1 NCUM Global assimilation-forecast System(NCUM-G) up-grade

Model (NCUM-G:V7)	Atmospheric Data Assimilation	Surface analysis
Model: Unified Model; Version 11.2 **Domain:** Global **Horizontal Resolution:**12 km **Vertical levels:** 70 levels (model top at 80 km) **Time Step:** 5 minutes **Physical Parametrizations:** Based on GA7.2 **Dynamical Core:** ENDGame **Forecast length:** 10 days (based on 00 UTC and 12 UTC initial conditions)	**Method:** Hybrid incremental 4D-Var (Hybrid 4D-Var). Information on “errors of the day” is provided by NEPS (NCMRWF Ensemble Prediction System) forecast in all data assimilation cycles **Data Assimilation Cycles:** 4analyses per day at 00, 06, 12 and 18 UTC. Observations within +/- 3 hrs from the cycle time is assimilated in the respective DA cycle **Observations:** Observations received at NCMRWF from GTS(IMD) and various satellite data producers (NOAA/NESDIS, EUMETSAT, ISRO etc.) are used for assimilation. Observation Processing System does the quality control of observations. Variational bias correction is applied to satellite radiance.	**Soil Moisture analysis:** *Method:*SimplifiedExtended Kalman Filter *Analysis time:* 00, 06, 12 and 18 UTC *Observations assimilated:* ASCAT soil wetness observations, Screen Temperature and Humidity increments (pseudo observations from 3D-Var screen analysis) **SST**: Updated at 12 UTC DA cycle with OSTIA based SST and sea-ice analysis **Snow Analysis:** Satellite-derived snow analysis. Updated at 12 UTC DA cycle

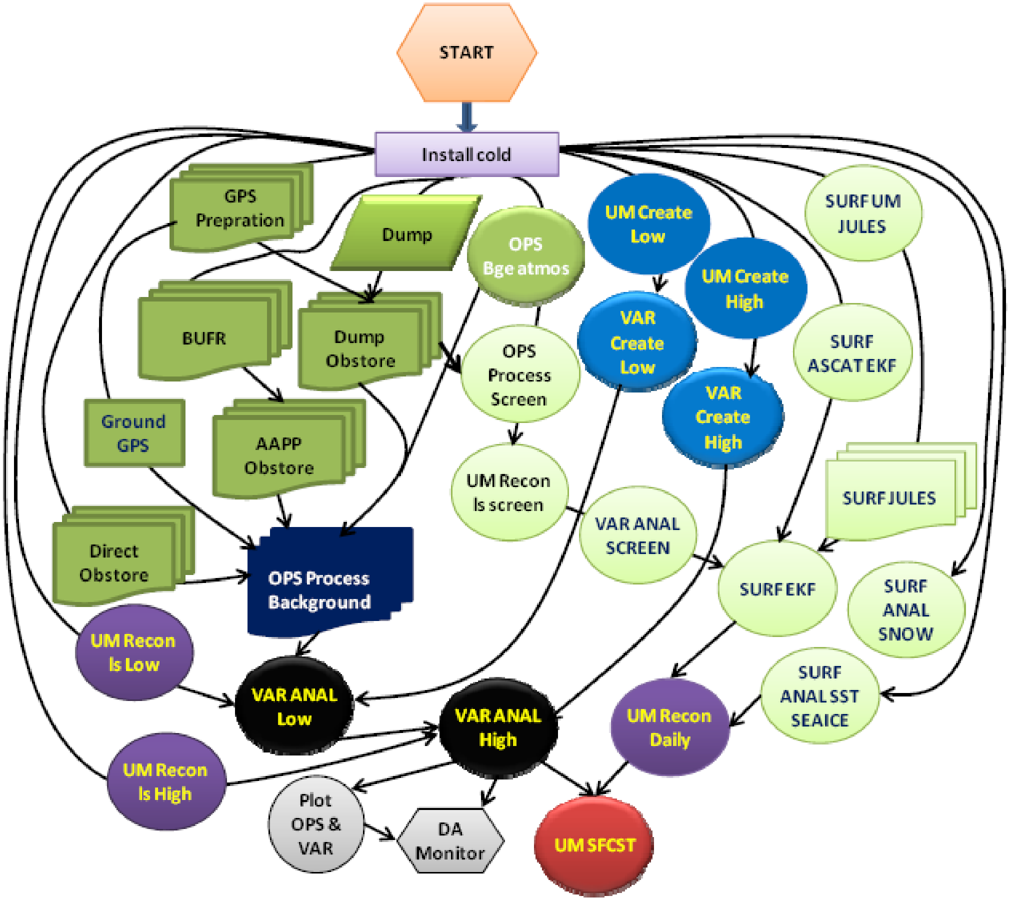

Figure 2 Schematic of NCMRWF's operational Hybrid DA-analysis system

In Figure 1 the light blue section computes the "Flow-dependent" error information by making use of the 22 member ensemble forecast from "NCMRWF Ensemble Prediction System (NEPS). The hybrid variational-ensemble DA approach (Light blue and black circles) merges the two sources of covariance information, climatological and flow-dependent, into the 4D-Var algorithm. The ETKF prepares ensemble perturbations for the NEPS at 00, 06, 12 and 18 UTC and the short forecast is prepared based on the Hybrid 4D-Var analysis with ensemble perturbations. Hybrid 4D-Var uses these ensemble forecasts for the flow dependent background errors. The linearization states for the perturbation forecast model in VAR is taken from the short forecast of the high resolution UM (Figure 1: purple colour). The high resolution NCUM forecasts are interpolated to (reconfiguration) to the VAR resolution. Global atmospheric analysis valid for 00, 06, 12 and 18 UTC is produced every day.

References

Bishop, C. H., Etherton, B. J. and Majumdar, S. J., 2001: Adaptive samplingwith the ensemble transform kalman filter. Part 1: theoretical aspects Mon. Wea. Rev. 129, 420–436.

Clayton, A. M., , A. C. Lorenc, , and D. M. Barker, 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. Quart. J. Roy. Meteor.Soc., 139, 1445–1461, doi:10.1002/qj.2054.

Clayton, A., 2012: Incremental Analysis Update (IAU) Scheme. Unified Model Documentation Paper No. 31

Eady, E., 1951: The quantitative theory of cyclone development. Compendium of Meteorology, T. Malone, Ed., Amer. Meteor. Soc., 464–469.

George J P., S. Indira Rani, A. Jayakumar, SajiMohandas,SwapanMallick, A. Lodh, R. Rakhi, M. N. R. Sreevathsa and E. N. Rajagopal (2016): NCUM Data Assimilation System, NMRF/TR/01/2016, 20p

Sumit Kumar, M. T. Bushair, BuddhiPrakash J., AbhishekLodh, Priti Sharma, Gibies George, S. Indira Rani, John P. George, A. Jayakumar, Saji Mohandas, Sushant Kumar, Mohana S. Thota, RaghavendraAshrit, and E. N. Rajagopal (2020), NCUM Global NWP System: Version 6 (NCUM-G:V6), NCMRWF Technical Report, NMRF/TR/06/2020

Section 1 NCMRWF Ensemble data Assimilation system: based on

https://www.ncmrwf.gov.in/Reports-php/Implementation%20of%20Unified%20Model%20based%20Ensemble%20Prediction%20System%20at%20NCMRWF%20(NEPS).php

Section 2: NCMRWF Hybrid Data Assimilation System at NCMRWF is based on:

https://www.ncmrwf.gov.in/Reports-php/NCUM-Global-NWP-System-Version-6.php