Joint IMD-WMO Group Fellowship Training

on

Numerical Weather Prediction

04 Oct - 10 Nov 2021

Lecture Note

on

Gridpoint Statistical Interpolation (GSI) scheme & Concept of Observation Operator

by

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

Gridpoint Statistical Interpolation (GSI) Scheme

1. *Background:* Numerical Weather Prediction (NWP) is an initial value problem. A good NWP forecast requires an Analysis or Initial Conditions of high quality. The more accurate the estimation of the initial conditions, the better is the quality of forecasts. High-quality analysis requires a good Data Assimilation System, good observations to assimilate and a good NWP model providing the first guess or background information. *Data Assimilation* is the process that combines the observations of the atmospheric conditions with a short-range (usually 6 hours) model forecast (termed as first guess or background) to produce the best estimate of the current state of the atmosphere (Analysis) on the regular model grid. The real-world observations enter the numerical weather model's forecast cycles through data assimilation and safeguard against model error growth. The *first guess* contains background information retained by the model from previous analyses. It has footprints of past weather conditions and provides uniform information coverage over the assimilation domain. *Observational data* are considered over a range of time, called a time window, usually centred on the analysis time.

   Data Assimilation methods can be based on an empirical, constant statistical or adaptive statistical approach. *Successive Correction Method*, *Nudging* and, *Physical Initialization* and *Latent Heat Nudging* are types of empirical methods. The Adaptive Statistical approach comprises various forms of *Kalman Filter* (viz. *Extended Kalman Filter*, *Ensemble Kalman Filter*, etc.). The constant statistical method includes *Optimal Interpolation* (OI), *3-dimensional variational* (3DVar) and *4-dimensional variational* (4DVar) data assimilation techniques. At NCEP (National Centers for Environmental Prediction), USA, the first analysis system to be used was Optimal Interpolation. In the early '90s, NCEP operationalized its first variational analysis system called *Spectral Statistical Interpolation* (SSI) technique. It was based on the 3DVar approach.

Although it was a significant improvement over OI, it faced few lacunae. The spectral coefficients used in the spectral model are analyzed directly using the same basic equations as statistical (optimal) interpolation in the SSI analysis system. The analysis variables in SSI are spectral coefficients instead of grid point values. It uses all observations at once to solve a single global problem (Parrish and Derber 1992). In this formulation, the background error covariances are computed assuming that uncorrelated errors in spectral space lead to homogeneous, isotropic statistics in grid space. This is a significant weakness of the system. Thus, a new formulation needs to be developed to allow the variance and correlation length scale to vary nontrivially in all space directions. Also, in SSI, the analysis being formulated in spectral space made it difficult for atmospheric systems on a regional scale.

2. *GSI:* GSI analysis scheme is the evolutionary combination of the global SSI analysis system and the regional ETA 3D-VAR. It replaced spectral definition for background errors with grid point (physical space) version based on recursive filters. The 3D-Var version of this global analysis system in physical space is as effective as 3D-Var in spectral space with latitude-dependent structure functions and other error statistics. Diagonal background error covariance in spectral space (in SSI) allows little control over the spatial variation of the error statistics as the structure-function is limited to being geographically homogeneous and isotropic about its centre (Parrish and Derber 1992; Courtier et al. 1998). GSI allows greater flexibility in terms of inhomogeneity and anisotropy for background error statistics (Wu et al. 2002). Thus significant improvement of GSI over the SSI analysis scheme is its latitude-dependent structure functions and has more appropriate background errors in the tropics. The background error covariances are isotropic and homogenous in the zonal direction. Unlike SSI, the GSI system can easily be applied to the regional domain.

GSI system over the global domain is generally run in 6-hour intermittent assimilation cycles. A fresh analysis, i.e. a new estimate of the atmospheric state, is generated every 6-hour and is used as an initial condition for generating a 9-hour forecast. The background or the first guess used for the assimilation process is the 6-hour forecast from the previous cycle. The 9-hour forecast produced every cycle is necessary to interpolate all the asynoptic observations available within the assimilation time window (in general ± 3hour around the analysis time).

GSI Analysis scheme was initially built as a 3D-Var system. In recent years, it has been under continuous development and has evolved into an Ensemble-Variational Hybrid system. Over the course of development, multiple options of the GSI analysis scheme are available. It can be used as 3DEnVar, Hybrid 3DEnVar, 4DEnVar, Hybrid 4DEnVar, or 4DVar (when coupled with an adjoint model supported by GSI) system. For the present course, we will concentrate on the 3D-Var configuration. In a variational assimilation system, the analysis is computed by minimizing a cost function. In the 3D-Var approach, one defines a cost function proportional to the square of the distance between the analysis and both the background and the observations (Sasaki 1970; Kalnay 2003). The cost function is minimized directly to obtain the analysis. In 3D-Var, the cost function can be formulated as:

$$J_{var}(x) = J_b + J_o + J_c \tag{1}$$

where,

J_{var} ⇒ The cost function

J_b ⇒ Fit to background: $\frac{1}{2}(x - x_b)^T B^{-1} (x - x_b)$

J_o ⇒ Fit to observation: $\frac{1}{2}\left(y - H(x)\right)^T R^{-1} \left(y - H(x)\right)$

J_c ⇒ Constraint terms include the penalties for negative humidity constraint, excess moisture constraint, negative visibility constraint, negative gust constraint, negative PBL constraint, and conservation of global dry mass.

Equation-1 can be represented as:

$$J_{var}(x) = \frac{1}{2}(x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2}\left(y - H(x)\right)^T R^{-1} \left(y - H(x)\right) + J_c \tag{2}$$

where,

x ⇒ analysis/control variable

x_b ⇒ background vector

$B \Rightarrow$ background error covariance matrix

$H \Rightarrow$ observation operator

$R \Rightarrow$ observation error covariance matrix (instrument error + representative error)

$y \Rightarrow$ observation vector

The analysis/control variables in GSI upon which the background errors are computed includes:

- Streamfunction (ψ)
- Unbalanced velocity potential (χ)
- Unbalanced temperature (T)
- Unbalanced surface pressure (Ps)
- Pseudo relative humidity or normalized relative humidity
- Satellite bias correction coefficients
- Ozone (global GSI only)
- Cloud condensate mixing ratio
- Trace gases / aerosols / chemistry (for chemical DA)
- Gust and visibility (for RTMA)

In the 3D-Var analysis system, the observations within the assimilation time window are all considered valid at the same time. The minimum or optimal value of the cost function is obtained for $x = x_a$ $(the\ analysis)$ by minimizing equation-2:

$$\nabla J_{var}(x) = 0 \qquad (3)$$

A detailed explanation of the variational equation is provided in the lecture notes on *'Introduction to cost function for 3-DVAR as well as for 4- DVAR data assimilation and its minimization, giving rise to analyzed field'*.

Minimization of the cost function in GSI involves an <u>*outer*</u> and an <u>*inner iteration*</u>. The outer iteration covers the Quality Control of observation data and running of Non-linear Observation Forward Operator. During inner iteration, GSI performs the Minimization procedure, Variational Quality Control, runs Simple Forward Operator, and provides a solution to start the next outer iteration.

In most cases for running GSI, three types of input data are required viz; background / first guess file, observations and fixed files. The fixed files are comprised of CRTM (Community Radiative Transfer Model) coefficient files (used for satellite radiance data assimilation), configuration files, statistic files, and bias correction files. GSI, when run with ensemble/hybrid option, will, in addition, require ensemble forecasts. The GSI system, when used for the global domain, constitutes the following steps:

i. *Observation Processing*: Before ingesting into the model, the observations require pre-processing. They are decoded, grouped and encoded into single BUFR ((Binary Universal Form for the Representation of Meteorological data) files containing observations that fall within the assimilation time window. The synoptic observations generally constitute a single file. In contrast, the asynoptic or non-conventional data (especially the observations from satellites) are used as separate files based on their type and sensors used for recording them. Observation processing is being done in real to near-real-time at operational centres. For archived runs, the observation processing is already completed, and the observation files are available to be directly used for assimilation.

   Observations operational in the Data Assimilation System at National Center for Medium Range Weather Forecasting (NCMRWF) is summarized in table-1. The observations have quality markers assigned to them by the data processing centres. The initial filtering of the observations data is based on the associated quality markers. In addition, before the minimization procedure, the observations are passed through several quality control checks to ensure good data assimilation. Based on the analysis resolution, high-density data are thinned or superobbed. Thinning also alleviates observation error correlation and error of representativeness.

ii. *Storm Relocation*: In the presence of cyclonic systems, this step is run to adjust the first guess. Storm relocation adjusts the cyclonic system in the first guess before assimilation, close to observed storm intensity, location and structure, and ensures better-analyzed storm fields.

iii. *Prep*: This step prepares the data to be used in the analysis system.

iv. *Analysis*: This step runs the data assimilation and generates analysis/initial conditions for the subsequent forecast.

v. *Forecast*: Using the analysis fields generated from the previous step, the forecast model runs up to the desired number of hours (based on the assimilation cycle) and produces the first guess/background for the next assimilation cycle.

vi. *Post*: This step converts the analysis and forecast fields generated from the previous step into WMO (World Meteorological Organization) GRIB (Gridded Binary) format. The resulting post-processed files can be used for diagnostic purposes and also can be used by other NWP model systems.

The community version of the GSI system is being maintained and supported by the Developmental Testbed Center (DTC):

https://dtcenter.org/community-code/gridpoint-statistical-interpolation-gsi

The GSI system in the form of community model code is freely available in the public domain for use by the research community:

https://dtcenter.org/community-code/gridpoint-statistical-interpolation-gsi/download

DTC currently focuses on testing and evaluating the GSI system on a regional scale coupled with the WRF (Weather Research and Forecasting) model system.

Concept of Observation Operator

Observations used for data assimilation are from locations different from the analysis grid points. Assimilation of these data requires horizontal and vertical interpolation of the first guess values from model space to observation space. It is more straightforward when the observations are the direct model variables (i.e., temperature, wind, moisture and, surface pressure). Complexity arises when dealing with observations which are measurements influenced by the direct model variables. These observations, viz; radiances, bending angle, refractivities, reflectivities, doppler shifts, horizontal line of sight winds, zenith total delays, etc., are measured by remote sensing instruments. For assimilating such data, an operator is required. Using the values from the model first guess, the operator will simulate the first guess of the observation or can be termed as the model equivalent of the observation. These operators in NWP, known as '*Observation Operator*' (denoted by 'H' in the previous section), provides the link between the NWP model variables and the observations (Lorenc 1986; Pailleux 1990). The simulation of the model equivalent of observations by the observation operator allows the correct comparison of forecast variables with the observations for assimilation.

The operator 'H' shown in the previous section represents the ensemble of all the observation operators, which will transform the control variable (x) into a quantity at the observation location equivalent to the parameter (y) measured by the instruments. The actual and simulated parameters are then compared, and their difference $\left(y - H(x)\right)$ known as the '*observational increments*' or innovations, are computed. The model variables are transformed as per the physical laws. For example, the *Radiance Operator* uses vertical profiles of temperature and moisture fields from the model first guess and computes the first guess values equivalent to observed satellite radiances. The observation operator also performs spatial interpolations (or transformation from spectral to physical space) from the model values to the observation location.

The observation operator depends on the level of prepossessing (Huang et al. 2002). For example, to assimilate the refractivity profiles, they can be transformed into profiles of temperature versus pressure. Instead of incorporating the refractivity profiles, these temperatures vs pressure profiles could also be assimilated. But it has been proven by multiple studies that it is more beneficial to assimilate the observations directly in as raw form as possible.

Given below are examples of two observation operators (Huang et al. 2002):

For zenith total delays (ZTD, which are measured by ground-based Global Positioning System (GPS) receivers), the observation operator includes

$$ZTD = p_a\, f(\theta, h) + \frac{1}{g(\theta)} \sum_{i=1}^{N} q_i \left(p_{i+1/2} - p_{i-1/2}\right)$$

where,

p_a ⇒ the pressure at the GPS antenna

$p_{i\pm 1/2}$ ⇒ the pressures at the model half levels

q ⇒ specific humidity

g ⇒ the gravitational acceleration

θ ⇒ the latitude

h ⇒ the geometric height

f ⇒ a function depending on the geographical location of the site.

For refractivity, the observation operator includes the part

$$N = \frac{p}{T} \frac{1}{1 + q(1/\epsilon - 1)} \left(k_1 + \frac{q}{\epsilon} \left(k_2 - k_1/\epsilon + \frac{k_3}{T} \right) \right)$$

where,

p ⇒ pressure at the observation location

T ⇒ temperature at the observation location

q ⇒ humidity at the observation location

$k_{1,2,3}$ & ϵ ⇒ constants

The observation operator includes a complicated integral for radiance data depending on the radiative transfer up through the atmosphere.

References

Courtier, P. and Coauthors, 1998: The ECMWF implementation of the three-dimensional variational assimilation (3D-Var). I: Formulation; Quart. J. Roy. Meteorol. Soc. 124, 1783–1807.

Dutta, S. K. and V. S. Prasad, 2011: Impact of gridpoint statistical interpolation scheme over Indian region; J. Earth Syst. Sci., 120, 6, 1095-1112.

Huang, Xiang-Yu & Henrik Vedel, 2002: An Introduction To Data Assimilation.

Kalnay, E., 2003: Atmospheric modeling, data assimilation and predictability; Cambridge University Press

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction; Q. J. R. Meteorol. Soc., 112, 1177–1194.

Pailleux, J., 1990: A global variational assimilation scheme and its application for using TOVS radiances; In Proc. WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography, 325–328, Clermont-Ferrand, France.

Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System; Mon. Weather Rev., 120, 1747–1763.

Rajagopal, E. N., M. Das Gupta, S. Mohandas, V. S. Prasad, J. P. George, G. R. Iyengar and Preveen Kumar D, 2007: Implementation of T254L64 global forecast system at NCMRWF; NCMRWF Technical Report, 1–42.

Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis; Mon. Wea. Rev., 98, 875-883.

Wu W-S, R. J. Purser and D. F. Parrish, 2002: Three dimensional variational analysis with spatially inhomogenous covariances; Mon. Weather Rev., 130, 2905–2916.

Table 1: Observations operational in Data Assimilation system at National Center for Medium Range Weather Forecasting (NCMRWF)

Conventional Observations	**Satellite Observations**						
	Satellite Winds (AMV)		***Scatterometer Winds***	***Satellite Radiances***			***GPSRO***
	GEO	***LEO***		***GEO***	***LEO***		***LEO***
				IR	**IR (HyS)**	**MW**	***Bending Angle***
Surface: Land SYNOP (TAC & BUFR) SHIP (TAC & BUFR) BUOY, TC BOGUS *Profiles:* PILOT, TEMP (RS/RW- Both TAC & BUFR) Wind Profiler, Drop Sonde DWR VAD Winds *Aircraft:* AMDAR, AIREP	INSAT-3D	NOAA-15	ASCAT (MetOp-A)	INSAT-3D Imager	IASI (MetOp-A)	AMSU-A (MetOp-A)	COSMIC-2
	Meteosat-8	NOAA-18	ASCAT (MetOp-B)	SEVIRI (Meteosat-8)	IASI (MetOp-B)	AMSU-A (MetOp-B)	METOP A, B & C
	Meteosat-11	NOAA-19	Scatsat	SEVIRI (Meteosat-11)	AIRS (AQUA)	AMSU-A (NOAA-18)	TanDEM-X
	HIMAWAR I	MetOp-A	Windsat (Coriolis)	AHI (HIMAWARI-8)	CrIS (SNPP)	AMSU-A (NOAA-19)	TerraSAR-X
	GOES-16	MetOp-B		INSAT-3D/3DR Sounder	CrIS (NOAA-20)	MHS (MetOp-A)	KOMPSAT-5
	GOES-17	AQUA				MHS (MetOp-B)	FY-3C
		TERRA				MHS (NOAA-19)	
		SNPP				MT-SAPHIR	
		NOAA-20				ATMS (SNPP)	
						SSMIS (DMSP-F17)	
						AMSR-2 (GCOM-W1)	
						FY-3C	
						GMI (GPM)	
						ATMS (NOAA-20)	