Joint IMD-WMO group fellowship Training

On

Numerical Weather Prediction

Through

Distance learning mode

Managed by

Meteorological Training Institute, India Meteorological Department (IMD), Pune

During

04 Oct to 10 Nov 2021

Lecture Notes

By

Somenath Dutta, IMD, Pune, India

Objective analysis: We know that in NWP the governing equations, which are essentially nonlinear partial differential equations, are integrated forward in time to obtain the future values of the field variables.

Since the governing equations are nonlinear partial differential equations and as there is no method to integrate analytically such equation, these equations are integrated numerically. To integrate the equations numerically the analytical time and spatial domain are discretised into finite number of time steps and spatial grid points. As the solving of these nonlinear PDE are BVP (boundary value problem) and IVP, (Initial value problem) we require the values of the field variable at all grid points at each level at initial time. But the observed field variables are not necessarily at the grid points. Thus, our first task is to prepare the values of the field variable at grid points from the observing points. This is known as objective analysis.

Different types of objective analysis scheme.

1. Polynomial fitting
2. Crossman's scheme
3. Optimum interpolation scheme

Polynomial fitting: In this method spatial variation of any field variable over a limited region at a given vertical level and at given time is expressed as a polynomial in x (longitude) and y (latitude); the degree of which is determined by the number of observing points over the region.

Let us consider an arbitrary field variable,$'Z'$, which is to be objectively analyzed. Obviously at any instant at any level, it is a function of (x, y).

Let there are $'(n+1)'$ number of observing points with coordinates $\{(x_i, y_i); i = 1, .., (n+1)\}$.

Let $Z(x_i, y_i) = Z_i, \forall i = 1, n$.

In this method, $Z(x, y)$ is expressed as a polynomial in x, y with degree $'n'$, like, $Z(x, y) = \sum_{i=1}^{n+1} a_{n-i+1} x^{n-i+1} y^{i-1} \ldots \ldots (1)$

Thus, we have following set of (n+1) linear equations in a_i for $\forall i = 0, n$

$$Z_k = \sum_{i=1}^{n+1} a_{n-i+1} x_k^{n-i+1} y_k^{i-1} \ldots \ldots (k = 1, .., n+1)$$

Above equations can be solved by inverting the (n+1) x (n+1) coefficient matrix.

Once, the polynomial is constructed using above computed values of a_i for $\forall i = 0, n$; values of the field variables Z can be computed at any arbitrary point.

Although this method is very robust mathematically, but it is very difficult and time & memory consuming in inverting a very large order matrix.

Cressman's successive correction method: Let us consider an arbitrary field, 'Z' to be analyzed at an arbitrary grid point 'G'. Let us consider also an observing point 'O' near the grid point.

This method starts with an initial guess for the true value of a field 'Z' at the 'G' & 'O' both.

Let $Z_G^{(0)}$ & $Z_O^{(0)}$ be the initial guess value of 'Z' at 'G' & 'O' both.

IfZ_G^T & Z_O^T are the true values of 'Z' at 'G' & 'O' respectively, then error in initial guess at these points are respectively $\{Z_G^T -\text{-}\, Z_G^{(0)}\}$ & $\{Z_O^T -\text{-}\, Z_O^{(0)}\}$.

In this method it is assumed that both the errors are same, i.e., $Z_G^T - Z_G^{(0)}$=$Z_O^T -\text{-}\, Z_O^{(0)}$. This gives the 1st improved guess for Z, denoted by $Z_G^{(1)} = Z_G^{(0)}$+$[Z_O^T - Z_O^{(0)}]$. Similarly, at the observing point 'O' also.

Thus 1st improved guess is obtained by adding a correction to the initial guess.

Correction is made successively, such that (m+1)th improved guess of 'Z' at the grid point 'G' is given by, $Z_G^{(m+1)} = Z_G^{(m)} + [Z_O^T - Z_O^{(m)}]$.

Above iteration method converges when following condition is satisfied, $\left|, Z_G^{(m+1)} - Z_G^{(m)}\right| < \epsilon$, for a pre assigned very small positive number ϵ. Then either of $Z_G^{(m+1)}$& $Z_G^{(m)}$ is taken as an estimated analyzed value, say, Z_O^{AG} based on the value at the observing point 'O'.

If there are a number of observing points, say, $O_1, O_2, O_3, \ldots etc.$; around a specific grid point 'G' and if $Z_{O_k}^{AG}$ be the estimated analyzed value of 'Z' based on observed value at 'O_k', for k=1,2,3…etc.

Then analyzed value of 'Z' at 'G' is given $Z_G^A = \frac{\sum_k w_k Z_{O_k}^{AG}}{\sum_k w_k}$,

Where $w_k = \frac{R^2 - {r_k}^2}{R^2 + {r_k}^2},\ if\ r_k < R$ and $w_k = 0,\ if\ r_k \geq R$

Where, R is a positive integral multiple of grid length, decided based on the scale of the weather, for which analysis is done and r_k is the distance between 'G' and the observing point O_k.

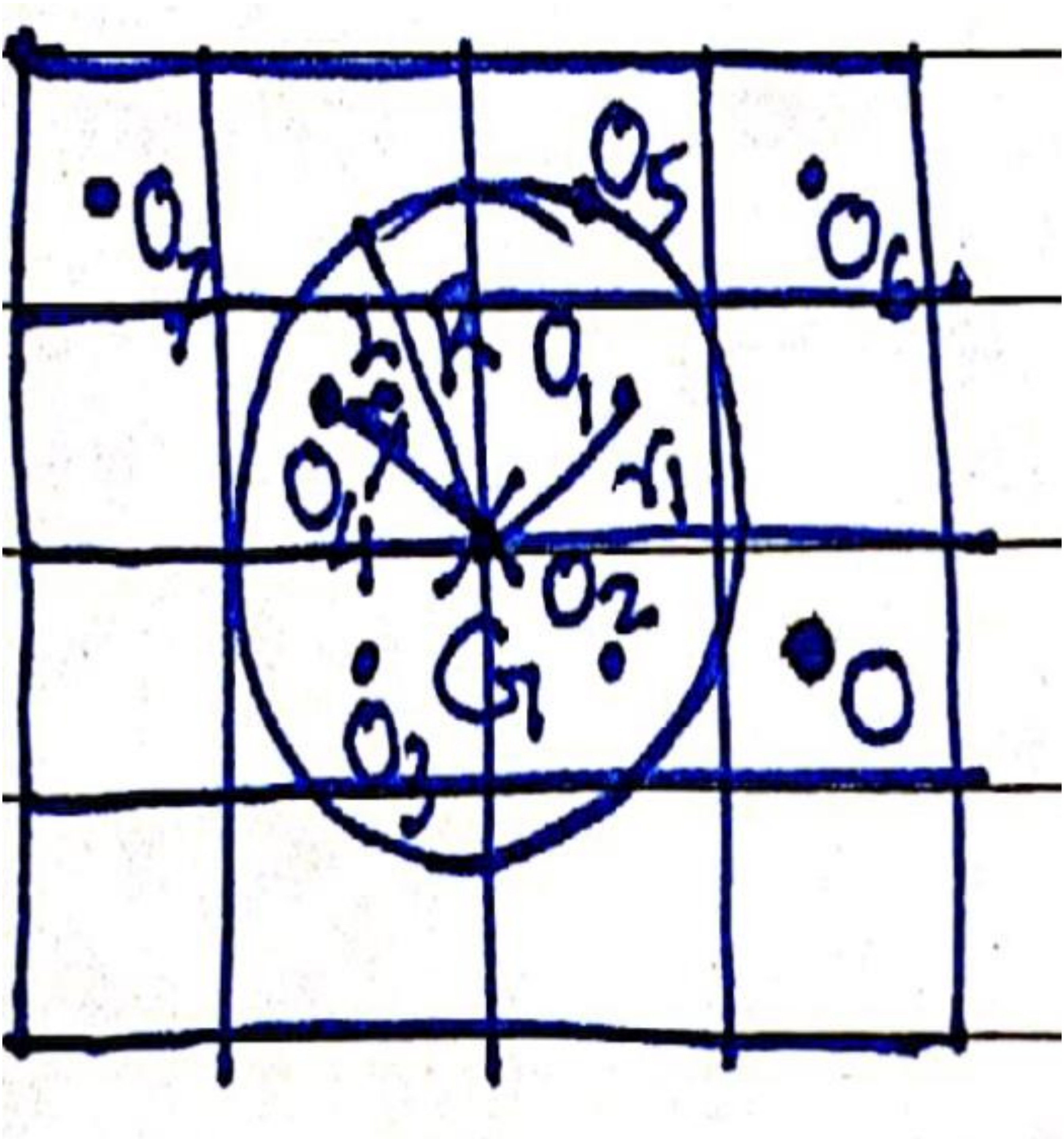
For the scale of the given weather, R is fixed.

Then as shown in adjoining figure, draw a circle centered at 'G' with radius 'R'.

Consider those observing points only, which lie inside the circle.

Then assign weight $\frac{R^2-{r_k}^2}{R^2+{r_k}^2}$ for such observing point O_k, with more weight to the nearer points and decreasing as going away from G, as shown in the figure.

Major limitation is fixing weight in an empirical method; thus, robustness is missing.



Optimum interpolation method: Let us consider an arbitrary field, ‘Z’ to be analyzed at an arbitrary grid point ‘G’. Let, Z_1&$Z_2$are two unbiased estimates of the of the true value Z_T of ‘Z', obtained from two independent sources. If ε_1& ε_2 are the errors in the above two estimates, then $\varepsilon_1 = Z_T - Z_1$& $\varepsilon_2 = Z_T - Z_2$ are two random variables with mean zero, i.e., $\bar{\varepsilon_1} = 0$ & $\bar{\varepsilon_2} = 0$.

Now another one estimate for Z is constructed as $Z_A = aZ_1 + (1-a)Z_2$.

If ε be the error in this estimate, then

$\varepsilon = Z_T$-$[aZ_1 + (1-a)Z_2]$

$$= (a + 1 - a)Z_T - [aZ_1 + (1-a)Z_2]$$

$= a(Z_T$-$Z_1) + (1-a)(Z_T - Z_2)$

$$= a\varepsilon_1 + (1-a)\varepsilon_2$$

$$\Rightarrow \bar{\varepsilon} = a\bar{\varepsilon_1} + (1-a)\bar{\varepsilon_2} \Rightarrow \bar{\varepsilon} = 0$$

Thus, $Z_A = aZ_1 + (1-a)Z_2$ is also an unbiased estimate of Z.

Now the coefficient ‘a’ is determined in such a way that the variance of the error ε is minimum.

We know that for two random variables, X & Y, if $T = mX + nY,\ then\ Var(T) = m^2Var(X) + 2mnCov(X,Y) + n^2Var(Y)$.

Hence, $Var(\varepsilon) = a^2Var(\varepsilon_1) + 2a(1-a)Cov(\varepsilon_1, \varepsilon_2) + (1-a)^2Var(\varepsilon_2)$.

Let, $\sigma_1 = +\sqrt{Var(\varepsilon_1)}$, $\sigma_2 = +\sqrt{Var(\varepsilon_2)}$ and $\sigma = +\sqrt{Var(\varepsilon)}$

As, Z_1&Z_2 are independent, hence $\varepsilon_1, \varepsilon_2$ are also so. Hence, $Cov(\varepsilon_1, \varepsilon_2) = 0$. Hence, $\sigma^2 = a^2{\sigma_1}^2 + (1-a)^2{\sigma_2}^2$.

Now, minimalization of $Var(\varepsilon)$ requires, $\frac{\partial[\sigma^2]}{\partial a} = 0 \Rightarrow a = \frac{{\sigma_2}^2}{{\sigma_1}^2 + {\sigma_2}^2}$

Hence optimally interpolated estimate of ‘Z’ is an unbiased and minimum error variance estimate and is given by

$$\left(\frac{{\sigma_2}^2}{{\sigma_1}^2 + {\sigma_2}^2}\right)Z_1 + \left(\frac{{\sigma_1}^2}{{\sigma_1}^2 + {\sigma_2}^2}\right)Z_2$$

Maximum likelihood estimate of a variable: Let us consider an arbitrary field, 'Z' , the maximum likelihood estimate of which is to be analyzed at an arbitrary grid point 'G'. Let, $Z_1 \& Z_2$are two unbiased estimates of the of the true value Z_T of 'Z', obtained from two independent sources.

If $\varepsilon_1 \& \ \varepsilon_2$ are the errors in the above two estimates, then $\varepsilon_1 = Z_T - Z_1 \& \ \varepsilon_2 = Z_T - Z_2$ are two random variables with mean zero, i.e., $\bar{\varepsilon_1} = 0 \ \& \ \bar{\varepsilon_2} = 0$. Now, both $\varepsilon_1, \varepsilon_2$ follow Gaussian distribution with mean zero and SDs $\sigma_1 \& \sigma_2$ respectively. So, $f_{\varepsilon_1}(\varepsilon_1)$=$\frac{1}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{{\varepsilon_1}^2}{{\sigma_1}^2}\right)$ & $f_{\varepsilon_2}(\varepsilon_2)$=$\frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{{\varepsilon_2}^2}{{\sigma_2}^2}\right)$; where, $\sigma_1 = +\sqrt{Var(\varepsilon_1)}$, & $\sigma_2 = +\sqrt{Var(\varepsilon_2)}$. Thus Joint probability density function of the bivariate random variable $(\varepsilon_1, \varepsilon_2)$ is $f_{\varepsilon_1,\varepsilon_2}(\varepsilon_1, \varepsilon_2) = f_{\varepsilon_1}(\varepsilon_1) f_{\varepsilon_2}(\varepsilon_2)$; because $\varepsilon_1, \varepsilon_2$ are independent. So, $f_{\varepsilon_1,\varepsilon_2}(\varepsilon_1, \varepsilon_2)$=$\frac{1}{2\pi\sigma_1\sigma_2} exp\left(-\frac{{\varepsilon_1}^2}{{\sigma_1}^2} - \frac{{\varepsilon_2}^2}{{\sigma_2}^2}\right)$.

Maximum likelihood function is defined as the natural logarithm of $f_{\varepsilon_1,\varepsilon_2}(\varepsilon_1, \varepsilon_2)$. Thus, $MLF = \ln\left[\, f_{\varepsilon_1,\varepsilon_2}(\varepsilon_1, \varepsilon_2)\right] = -\ln(2\pi) - \ln(\sigma_1\sigma_2) - \left(\frac{{\varepsilon_1}^2}{{\sigma_1}^2} + \frac{{\varepsilon_2}^2}{{\sigma_2}^2}\right)$.

Maximum Likelihood Estimate of 'Z' is that one which maximizes the maximum likely hood function (MLF). Thus MLF will be maximum, when $J = \left(\frac{{\varepsilon_1}^2}{{\sigma_1}^2} + \frac{{\varepsilon_2}^2}{{\sigma_2}^2}\right) = \left[\frac{(Z_T - Z_1)^2}{{\sigma_1}^2} + \frac{(Z_T - Z_2)^2}{{\sigma_1}^2}\right]$ is minimum. J is known as cost function.

Now, minimization of J requires, $\frac{\partial J}{\partial Z_T} = 0$,

$$\Rightarrow Z = \left(\frac{{\sigma_2}^2}{{\sigma_1}^2 + {\sigma_2}^2}\right) Z_1 + \left(\frac{{\sigma_1}^2}{{\sigma_1}^2 + {\sigma_2}^2}\right) Z_2$$

$$\Rightarrow Optimally\ Interpolated\ Estimate = Maximum\ Likelihood\ Estimate$$

$\Rightarrow Optimally\ Interpolated\ Estimate$ is the best estimate. It is known as best linear unbiased estimate (BLUE).