Joint IMD-WMO Group Fellowship Training

on

Numerical Weather Prediction

managed by

India Meteorological Department (IMD), Pune

04 Oct – 10 Nov 2021

Lecture Notes on "*Postprocessing of NWP Forecasts*"

by

Kondapalli Niranjan Kumar

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

I. Model Output Postprocessing

A typical operational forecast model may produce of the order of a hundred or so different forecast fields – usually referred to as diagnostics. However, there is often a much smaller set of diagnostic fields that a forecaster or forecasting team will make regular use of. Obvious things that forecasters might want to look at are maps of precipitation intensity, wind speed and temperature through the forecast period, but these fields by themselves are often not enough. If an ensemble of forecasts is run even more information can be produced. Nevertheless, a numerical model will not produce one forecast of a particular variable at each grid point of the model.

Further, the forecasts are associated with the errors from multiple sources which have detrimental effect on the model skill. The primary source of these errors includes 1) sensitivity to initial conditions; 2) boundary condition errors and 3) model structural errors. Model structural errors include a missing or poor representation of subgrid dynamical and physical processes and inaccuracies associated with the numerical scheme. All these numerical weather predictions (NWP) model deficiencies induce errors that are rapidly amplified in time due to the chaotic nature of the model dynamics, and in turn affect the course of the forecasts by inducing what are often called **systematic and random errors**. The first two sources of errors can be partly removed by the nudging the basic state of a model towards the reanalysis while the third source of error can be reduced using the statistical postprocessing techniques.

The basic definition of postprocessing can be understood like this, let's say the model grid-resolution is around 25–50 km square. A single value of minimum temperature or precipitation or any other field variable in this grid box may not represent the local features and mask the local variability. This is particularly true if that grid box contains a land region with appreciable local topography and maybe also contains contrasts between rural and urban areas. Even in a high-resolution regional model with grid boxes perhaps 5 km square, a single value of some variables may still mask quite a lot of local variation. The obvious answer to this problem is to increase the resolution of the model so that the local variations in terrain and land surface type which lead to the variations are themselves resolved. This is a somewhat impractical solution as most NWP centres are already running their NWP systems at the highest resolution that is allowed by the constraints of the available computing power. A more pragmatic solution is to put the output of the NWP model through some sort of further processing in order to incorporate some local variability. This approach is usually referred to as 'post-processing'.

The post processing of model output should be viewed as an integral part of the modeling system, at least for operational applications. This processing can take many forms, but one involves the use of methods for reducing the systematic error in the forecast products. The resulting better correspondence with observations will result in improved predictability for the entire system. The following subsections review various methods for statistically correcting NWP model forecasts in order to reduce the systematic error. The static methods require the use of a lengthy period of model reforecasts in order to define the statistical corrections based on relationships between past model output and past observations.

Note: Only systematic error is reduced by these methods except Kalman Filter (KF) based bias-correction approach. The random errors that result from numerically induced phase errors in the propagation of features, the smoothing of small-scale propagating features associated with insufficient model resolution, and other causes, will remain in the solution and cannot be statistically removed. However, the systematic error can represent a significant fraction of the total error, especially near the ground, so removing it through the use of post-processing methods can be very beneficial.

Operational dynamical forecasting began in 1956 in the U.S. and dissemination to the public of products derived from statistically postprocessed dynamical forecasts was initiated in 1968. These early forecasts were based on a technique known as "**perfect prog**" (e.g., Wilks, 2011), which required no training data from the dynamical model. Shortly thereafter, the preferred **model output statistics** (MOS, Glahn & Lowry, 1972) method began to be used when sufficient dynamical-model training data became available.

A. Model Output Statistics (MOS)

This method, when applied under ideal circumstances, is the gold standard of NWP model output post-processing (Glahn and Lowry, 1972). MOS is essentially multiple linear regression, where the predictors h_{nj} are model forecast variables (e.g., temperature, humidity, or wind at any grid point, either near the surface or in the upper levels), and may also include other astronomical or geographical parameters (such as latitude, longitude and time of the year) valid at time t_n. The predictand y_n is a station weather observation (e.g., maximum temperature or wind speed) valid at the same time as the forecast. Here, as in any statistical regression, the quality of the results improves with the quality and length of the training data set used to determine the regression coefficients b_j .

The dependent data set used for determining the regression coefficients is

$$\left.\begin{array}{ll} y_n = y(t_n) & n = 1, \ldots, N \\ h_{nj} = h_j(t_n) & n = 1, \ldots, N; j = 1, \ldots, J \end{array}\right\}$$

where we consider one predictand $y_n$as a function of time t_n and J predictors h_{nj}.

The linear regression (forecast)equation is

$$\hat{y}_n = b_0 + \sum_{j=1}^{J} b_j h_{nj} = \sum_{j=0}^{J} b_j h_{nj}$$

where for convenience the predictors associated with the constant term b_0 are defined as $h_{n0}\equiv$ 1. In linear regression the coefficients b_j are determined by minimizing the sum of squares of the forecast errors over the training period. The sum of squared errors (SSE) is given by:

$$SSE = \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 = \sum_{n=1}^{N} e_n^2$$

Taking the derivatives with respect to the coefficients b_j and setting them to zero we obtain:

$$\frac{\partial SSE}{\partial b_j} = 0 = \sum_{n=1}^{N} (y_n - \sum_{l=0}^{J} b_l h_{nl}) h_{nj} \qquad j = 0, 1, \ldots, J$$

or

$$\sum_{n=1}^{N} \left[h_{jn}^T y_n - h_{jn}^T \sum_{l=0}^{J} h_{nl} b_l \right] = 0 \qquad j = 0, \ldots, J$$

where $h_{jn}^T = h_{nj}$. The above Eqns. are the "normal" equations for multiple linear regression that determine the linear regression coefficients b_j , $j = 0, \ldots, J$. In matrix form, they can be written as

$$\mathbf{H}^T \mathbf{H} \mathbf{b} = \mathbf{H}^T \mathbf{y} \qquad \text{or} \qquad \mathbf{b} = \left(\mathbf{H}^T \mathbf{H}\right)^{-1} \mathbf{H}^T \mathbf{y}$$

where,

$$\mathbf{H} = \begin{bmatrix} 1 & h_{11} & \dots & h_{1J} \\ 1 & h_{21} & \dots & h_{2J} \\ \vdots & \vdots & h_{nj} & \vdots \\ 1 & h_{N1} & \dots & h_{NJ} \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_J \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

are, respectively, the dependent sample predictor matrix (model output variables, geographical and astronomical parameters, etc.), the vector of regression coefficients, and the vector of predictands in the dependent sample. $\hat{\boldsymbol{y}} = \boldsymbol{Hb}$, $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{Hb}$ are the linear predictions and the prediction error, respectively, in the dependent sample. The *dependent* estimate of the error variance of the prediction is $s_e^2 = \frac{SSE}{(N-J-1)}$ since the number of degrees of freedom is $N - J - 1$. This indicates that one should avoid overfitting the dependent sample by ensuring that $N \gg J$.

It is clear that for a MOS system to perform optimally, several conditions must be fulfilled:

a. The training period should be as long as possible (at least several years).

b. The model-based forecasting system should be kept unchanged to the maximum extent possible during the training period.

c. After training, the MOS system should be applied to future model forecasts that also use the same unchanged model system.

These conditions, while favorable for the MOS performance, are not favorable for the continued improvement of the NWP model, since they require “frozen” models.

In summary, the forecast statistical guidance (and in particular MOS)adds value to the direct NWP model output by objectively interpreting model output to remove systematic biases and quantifying uncertainty, predicting parameters that the model does not predict, and producing site-specific forecasts.

B. The “Perfect Prog” method

The earliest approach to statistical post processing is known as the Perfect-Prog (PP, perfect-prognosis) method (Klein et al. 1959). Perfect Prog is an approach similar to MOS, except that the regression equations are derived using as predictors, observations or analyses (rather than forecasts) valid at the prediction time, as if the forecasts were perfect. The regression

relationships are then applied to NWP-model forecasts of the predictors to produce forecasts of the predictands. Because the statistical relationships are not generated using model forecasts, they do not correct for model error. They simply statistically translate predicted variables into unpredicted or poorly predicted variables. In effect, it is assumed that the model prognosis is perfect. Because, as noted earlier, current models can explicitly predict many of the quantities that previously had to be statistically inferred through the PP method, this approach is less used operationally. A benefit of the PP method is that it is not dependent on the model to which it is applied, and thus the statistical relationships do not need to be recalculated when the model is modified.

Traditionally, most statistical downscaling methods applied in climate research have been of the PP type: the calibration of the statistical model is carried out completely in the real world, hence both predictor and predictand data are taken from observations (or observation proxies such as reanalyses). The calibrated model is then applied to predictors derived from model data. For this approach to be sensible, the predictors need to be "perfectly" simulated – thus the name. As a direct consequence of the calibration setup, the PP approach cannot correct model biases. Any bias in the simulated predictors will in general cause a biased downscaling.

In MOS–type downscaling methods, the calibration links predictors from a numerical model to observed predictands. Potential model biases already enter the calibration procedure and are thus taken into account by construction. If a model is systematically, say, a degree too warm, the calibrated MOS will automatically subtract this bias

The differences between the two post-processing methods have been show below in Figure 1

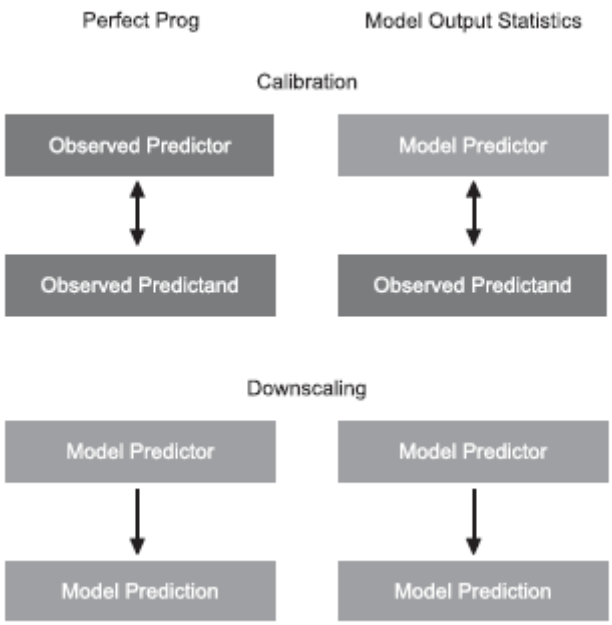

Figure 1: Perfect prognosis versus model output statistics. The top shows the origin of predictor and predictand data during calibration, the bottom the same for the actual downscaling.

C. Adaptive regression based on a simple Kalman filter approach

Adaptive regression based on Kalman filtering is another automatic post-processing method employs past observations and forecasts to estimate model bias in future forecasts. This method has been mainly used in data assimilation schemes to improve the accuracy of the initial conditions for both NWP. However, The KF has also been used for NWP model forecasts as a predictor bias correction method during postprocessing of short-term weather forecasts.

In the previous MOS and in other statistical prediction methods such as nonlinear regression or neural networks, the regression coefficients are computed from the dependent sample, and are not changed as new observations are collected until a new set of MOS equations are derived every 5 or 10 years. Because the regression coefficients are constant, the order of the observations is irrelevant in MOS, so that older data have as much influence as the newest observations used to derive the coefficients.

In adaptive regression, the Kalman filter equations are applied in a simple, sequential formulation to the multiple regression coefficients $b_k = b(t_k)$, whose values are updated every time step, rather than keeping them constant as in

$$\hat{y}_k = \sum_{j=0}^{J} b_j(t_k) h_{kj} = \begin{bmatrix} 1 & h_{k1} & \dots & h_{kJ} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \dots \\ b_J \end{bmatrix}_k = \mathbf{h}_k^T \mathbf{b}_k$$

Let's make the methodology clearer in the following section in slightly simpler way adopted from the Delle Monache et al. (2006)

In the KF approach, the information of recent past forecasts and observations is used to revise the estimate of the current raw forecast. Previous bias values are used as input to KF. The filter estimates the systematic component of the forecast errors, or bias. Once the future bias has been estimated, it can be removed from the forecast to produce an improved forecast. Such a corrected forecast should be statistically more accurate in a least-squares sense.

Methodology

The KF models the true (unknown) forecast bias x_t at time t, by the previous true bias plus a white noise η term [Bozic, 1994]:

$$x_{t|t-\Delta t} = x_{t-\Delta t|t-2\Delta t} + \eta_{t-\Delta t}$$

where $\eta_{t-\Delta t}$ is assumed uncorrelated in time and is normally distributed with zero-mean and variance σ_η^2, Δt is a time lag (flow figure below), and $t|t-\Delta t$ means that the value of the variable at time t depends on values at time $t-\Delta t$. Because of unresolved terrain features, numerical noise, lack of accuracy in the physical parameterizations, and errors in the observations themselves, the KF approach further assumed that the forecast error y_t (forecast minus observation at time t) is corrupted from truth by a random errorε_t:

$$y_t = x_t + \varepsilon_t = x_{t-\Delta t} + \eta_{t-\Delta t} + \varepsilon_t$$

where ε_t is assumed uncorrelated in time and normally correlated in time distributed with zero-mean and variance σ_ε^2. Thus, y_t includes both systematic biases plus random errors.

Kalman [1960] showed that the optimal recursive predictor of x_t (derived by minimizing the expected mean square error) can be written as combination of the previous bias estimate and the previous forecast error.

$$\hat{x}_{t+\Delta t|t} = \hat{x}_{t|t-\Delta t} + \beta_{t|t-\Delta t}(y_t - \hat{x}_{t|t-\Delta t})$$

where the hat (^) indicates the estimate

The weighting factor β, called Kalman gain, can be calculated from:

$$\beta_{t|t-\Delta t} = \frac{p_{t-\Delta t|t-2\Delta t} + \sigma_\eta^2}{(p_{t-\Delta t|t-2\Delta t} + \sigma_\eta^2 + \sigma_\varepsilon^2)}$$

where p is the expected mean square error, which can be computed as follows:

$$p_{t|t-\Delta t} = (p_{t-\Delta t|t-2\Delta t} + \sigma_\eta^2)(1 - \beta_{t|t-\Delta t})$$

Assume a time lag of Δt = 24 hours is used, today's forecast bias is estimated using yesterday's bias, which in turn was estimated using the day-before-yesterday's bias, and so on. Figure 2 shows the flow diagram of the KF algorithm.

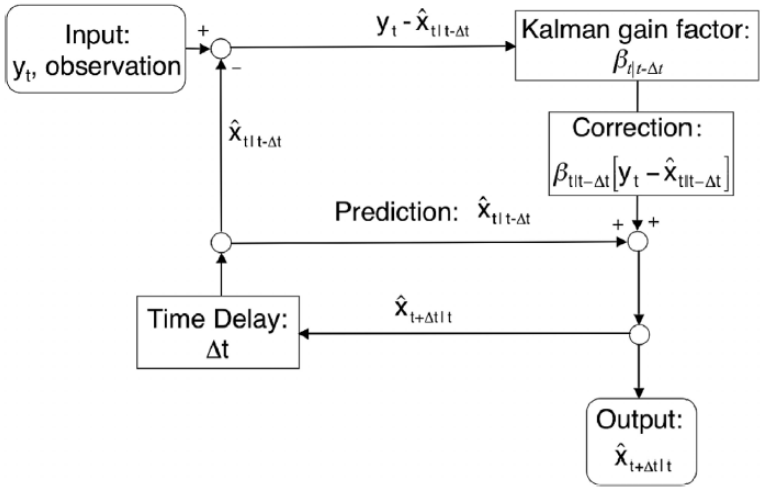

Figure 2: *Flow diagram of the Kalman filter bias estimator. It uses a predictor corrector approach, starting with the previous estimate of the bias ($\hat{x}_{t|t-\Delta t}$) and correcting it by a fraction (β) of difference between the previous bias estimate and previous observed forecast error (y_t) to estimate the future bias ($\hat{x}_{t+\Delta t|t}$).*

D. Weather generators

Even though model time steps may be relatively short - perhaps tens of minutes – much of the short-time-scale variability associated with some phenomena is not represented in the model solution. For example, precipitation rates in nature can be highly variable, as rain bands, or other small convective features that are in various stages of their life cycle, pass across a location. Variability on these time scales is not represented in most operational models. This is especially true for large AOGCM grid boxes for which the time series of variables are

smoother than those that apply to single points, simply because of the averaging that is implied over the large area. Unfortunately, high-frequency rain rates are needed for many hydrological applications, where the rate determines the partitioning of the rainwater between runoff and infiltration. Another example of high-frequency variability that is not represented in NWP models is wind gustiness, which is needed in models of dust elevation and transport, ocean waves, etc. To address such needs for high-frequency information from NWP and climate models, *synthetic high-resolution time series* can be generated with what are called stochastic weather generators. These methods essentially postprocess the model-generated time series, adding realistic higher-frequency variability. For NWP model simulations, the weather generators can add high-frequency spatial and temporal variations in the precipitation rate.

SUGGESTED GENERAL REFERENCES FOR FURTHER READING

1. Wilks, D. S., 2011: Statistical Methods in the Atmospheric Sciences. 3rd ed. Elsevier, 676 pp.
2. T. Warner, Numerical Weather and Climate Prediction (Cambridge Press, Cambridge, UK, 2011)
3. P. Inness, S. Dorling, Operational Weather Forecasting, DOI:10.1002/9781118447659 (John Wiley & Sons, Ltd)
4. Haraldur Ólafsson, Jian-Wen Bao (Eds), Uncertainties in Numerical Weather Prediction, Elsevier, 2021, https://doi.org/10.1016/B978-0-12-815491-5.09991-2.
5. Kalnay, E., 2003: Atmospheric modeling, data assimilation and predictability, Cambridge University Press, USA
6. Glahn, H.R., Lowry, D.A., 1972. The use of Model Output Statistics (MOS) in objective weather forecasting. J. Appl. Meteorol.11, 1203–1211.
7. Klein, W. H., B. M. Lewis, and I. Enger, 1959: Objective prediction of 5-day mean temperature during winter. J. Meteor. 16, 672-682.
8. Maraun, D., & Widmann, M. (2018). Statistical Downscaling and Bias Correction for Climate Research. Cambridge: Cambridge University Press. doi:10.1017/9781107588783
9. Delle Monache, L., T. Nipen, X. Deng, and Y. Zhou (2006). Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction. J. Geophys. Res., 111, D05308, doi: 10.1029/ 2005JD006311.
10. Bozic, S. M. (1994), Digital and Kalman Filtering, 2nd ed., 160 pp., John Wiley, Hoboken, N. J.
11. Website: https://dtcenter.org/community-code/model-evaluation-tools-met