

Improvement in Near Surface NWP Model Output using Kalman Filtering Technique: A Case Study for Trombay Site

Roopashree Shrivastava^{*,1}, Indumathi Srinivasan. Iyer¹ and Rajendrakumar Balkrishna Oza¹

⁽¹⁾ Radiation Safety Systems Division, Bhabha Atomic Research Centre, Mumbai – 400 085

Article history: received November 9, 2022; accepted August 8, 2023

Abstract

Numerical Weather Prediction (NWP) models exhibit systematic errors in the forecast of near surface atmospheric parameters due to various factors like grid resolution, parameterization schemes, treatment of sub-grid scale phenomena, data for initial and boundary conditions and interpolation techniques. One of the methods for reduction in model errors is the use of Kalman filter algorithm which recursively combines model output and observations such that the systematic errors are minimized. In the present study, the Kalman filter algorithm is utilized for correction of model output from The Air Pollution Model (TAPM) for the year 2013. The variables corrected are 2-m air temperature, 2-m relative humidity and zonal and meridional wind components at 10-m. Hourly observations of the same variables available at Trombay site are used in the study. In the present study, it is seen that, both wind speed and wind direction are better reproduced after Kalman filtering, in addition to near surface air temperature and relative humidity. Also, on an annual basis, biases in all the variables are eliminated. The standard statistical indices of model performance computed after Kalman filtering are superior to those computed using only model output. Time series plots of bias and RMSE in model after Kalman filtering indicate the advantage of Kalman filtering.

Keywords: Kalman filter; TAPM; Atmospheric model errors; Improvement; Statistical indices

1. Introduction

Numerical Weather Prediction (NWP) models have applications in various sectors like agriculture, renewable energy, aviation and disaster management during natural calamities like floods, heavy rainfall, cyclones, to name a few, as well as for routine weather updates. In the nuclear energy sector, the application of atmospheric models for weather forecasting is increasing because of the assistance it can provide in case of accidents. NWP models coupled with atmospheric dispersion models were applied to forecast the future trajectory of the plume during the Fukushima accident and also for estimating the source term. WRF-Chem model was employed to simulate the transport and dispersion of radioactive plume released from Japan after Fukushima Daiichi Nuclear Power Plant accident. By comparing model simulations with time series of measured concentration, it was possible to distinguish between

global and regional transport of radionuclides [Huh et al., 2013]. Katata et al. [2012] have used a combination of a non-hydrostatic atmospheric dynamical model (MM5) along with atmospheric dispersion model (GEARN) to estimate the release rates of ^{131}I and ^{137}Cs in the early phases of the accident at Fukushima. Several air quality applications like ensuring regulatory compliance of airborne releases require meteorological forecasts at local scale for prediction of associated impacts. In this regard, Thatcher and Hurley [2010] have developed a downscaling approach using the Conformal Cubic Atmospheric Model (CCAM) – The Air Pollution Model (TAPM) combined modeling system for forecasting local scale meteorology and air pollution in Melbourne, Australia. The combined modeling system was evaluated using measurements from various environmental monitoring stations in Melbourne, Australia during the year 2003. Results from this combined modeling system were found to be superior as compared to CCAM interpolated forecasts. As the demand for renewable energy like wind and solar energy increases across the globe, accurate and reliable predictions of atmospheric variables like solar radiation and wind speed extrapolated to wind power is required. Application of NWP models in atmospheric dispersion as well as in assessing renewable energy potential of a site requires realistic simulation of the related meteorological parameters.

However, NWP models exhibit errors due to several reasons like grid resolution, differences in station elevation used in the model and actual, model physics, numerical methods, data for initial and boundary conditions, interpolation at the required location of forecast and handling of sub-grid scale processes. Many times, the grid resolution of models is not sufficient to distinguish between the land surface and water bodies inside the model. In such cases, the model will use surface characteristics like albedo, surface roughness, land-use for the dominant surface type. In urban areas, meteorological observations may be carried out on the top of buildings due to lack of open spaces. This can also lead to differences between model results and observations. The presence of errors severely limits the application and utility of outputs from atmospheric models. Hence it is necessary to “post process” the output of NWP models using some technique to reduce the errors in the model output. Some of the techniques used to reduce errors in atmospheric models include Model Output Statistics (MOS), Glahn and Lowry [1972], Perfect Prog Method (PPM), Klein and Lewis [1970] and Kalman filter [Kalman, 1960; Kalman and Bucy, 1961]. MOS and PPM techniques are based on developing regression equations between the observed parameter to be predicted and a set of variables that are either forecasted or analyzed. Such kind of regression equations are generally developed on the basis of extensive and comprehensive datasets. However these techniques suffer from the drawback that if the NWP model is changed, then new regression equations need to be developed for the same variable.

One of the most convenient methods to reduce systematic errors in an atmospheric model is by Kalman filter. This is a technique in which model output is combined with observations recursively to reduce the systematic errors in the model output [Kalman, 1960; Kalman and Bucy, 1961]. A methodology for the correction of near surface forecasts of temperature and relative humidity based on a combination of Kalman filtering technique and empirical method with exponential smoothing was developed by Anadranistakis et al. [2004]. This methodology was tested in Northern Greece and substantial reduction in systematic errors for 2-3 days forecast of near surface temperature and humidity was seen. Galanis and Anadranistakis [2002] have implemented a one dimensional Kalman filter for correction of near surface temperature forecasts using Skiron model at two locations in Greece. They tested the Kalman filter algorithm for a period of seven days and were able to demonstrate that systematic errors almost vanish. Also the computation time was found to be small enough, making implementation easy even on a desktop computer. Crochet [2004] carried out Kalman filtering of 2-metre temperature and 10-metre wind speeds forecasts from Direct Model Output (DMO) at eleven synoptic stations in Iceland. The study duration was for two months during 1st January 2000 to 1st March 2000. Their results indicate that for both, wind and temperature forecasts, Kalman filter results are better than DMO. Likewise Cassola and Burlando [2012] have applied the Kalman filter technique for correction of wind speed forecast by the Bologna Limited-Area Model (BOLAM) model at two locations in Italy. Their study period extended for two years from 1st January 2007 to 31st December 2008. Results indicate that the Kalman filter method is able to provide significant forecast improvement as compared to direct model output as a result of minimization of systematic errors. Kalman filter can be implemented for a few hours i.e. now casting application or for medium range forecasting applications.

The Extended Kalman Filter (EKF) extends the concept of Kalman filter to a nonlinear system by linearization of the nonlinear system using first order Taylor approximation. The application of EKF requires calculation of the Jacobian matrix which is computationally intensive. An alternative to EKF is the Unscented Kalman Filter (UKF) which is an improvement over EKF and it generalizes the application of Kalman filters for both linear and nonlinear systems. UKF is also widely used for integration of Inertial Navigation System (INS) and Global Positioning

System (GPS). Hu et al. [2018, 2020] have combined inertial navigation equations with inertial measurement unit error equations for Inertial Navigation System/Global Navigation Satellite System (INS/GNSS) integration. Also, a Refined Strong Tracking Unscented Kalman Filter (RSTUKF) and Maximum Likelihood Adaptive Unscented Kalman Filter (MLAUKF) for application in INS/GNSS integration have been developed. The effectiveness and superiority of these techniques have been assessed through Monte Carlo simulations and practical experiments. Metia et al. [2016, 2018, 2020] have coupled the Extended Fractional Kalman filter with The Air Pollution Model and Chemical Transport model (TAPM-CTM) for obtaining improved estimates of emission inventories in urban and suburban areas. Using TAPM-CTM, concentrations of nitrogen oxide (NO), nitrogen dioxide (NO₂), and ozone (O₃) in Sydney and surrounding areas have been predicted. Thereafter EFKF is applied for obtaining improved emission inventory. Finally, model results are validated using satellite retrievals of O₃ in the region. Oktaviana et al. [2018] have applied a first order Fractional Kalman filter in TAPM-CTM to estimate the concentration of air pollution. Their simulations indicate that the Fractional Kalman filter has lesser errors as compared to the Kalman filter although at the cost of higher computational time. Voyant et al. [2017] have compared the performance of Kalman filter, autoregressive model and neural network based model for solar radiation forecast on hourly and shorter forecasting horizons and concluded that the performance of the various methods is a function of the forecast horizon. Gallego et al. [2020] have used UKF for estimating the metal-fluid temperature profiles and solar radiation at a solar cooling plant located in Seville. Simulations are carried out which compare the UKF with proportional integral derivative feed forward series controller and gain scheduling generalized predictive controller and the utility of UKF is demonstrated.

The present study is an application of Kalman filter for reduction of systematic errors in the output of TAPM. This study has been carried out at a tropical site in Mumbai, India. Yearlong data of both, model output and measurements at a single location have been utilized to evaluate the usefulness of the Kalman filtering Technique. The variables corrected are 2-m air temperature, 2-m relative humidity, and zonal and meridional wind components at 10-m height. Most of the studies published on the application of Kalman filter for improving NWP model output have focused on the improvement in prediction of maximum and minimum near surface air temperature, wind speed, solar radiation etc. However, in several applications, of equal importance is the accurate prediction of wind direction. In the present study, both wind speed and wind direction are better reproduced after Kalman filtering, in addition to near surface air temperature and relative humidity. Section 2 gives a brief description of TAPM. The Kalman filter algorithm is presented in Section 3. Section 4 describes the methodology of the Kalman filter and finally results are presented in Section 5.

2. The Air Pollution Model (TAPM)

TAPM is developed by Commonwealth and Industrial Research Organization (CSIRO), Australia [Hurley, 2005; Hurley et al. 2008]. This is a coupled meteorology and air pollution model. The meteorological component of TAPM is an incompressible, non-hydrostatic, primitive equation model with a terrain-following vertical coordinate for three-dimensional simulations. The model solves the momentum equations for horizontal wind components, the incompressible continuity equation for vertical velocity, scalar equations for potential virtual temperature and specific humidity of water vapour, cloud water/ice, rain water and snow. The turbulence terms in these equations are considered by solving separate equations for turbulence kinetic energy and eddy dissipation rate, and then using these values in representing the vertical fluxes by a gradient diffusion approach, including a counter-gradient term for heat flux. The model can be used to predict the flows important to local-scale air pollution, such as sea breezes and terrain induced flows, against a background of larger-scale meteorology provided by synoptic analyses. Meteorological parameterizations include explicit cloud microphysics, a vegetative canopy and soil scheme, turbulence parameterization as well as radiation at the surface. The soil parameterizations consist of the force – restore method. Soil characteristics are defined for three types of soil namely sand, sandy clay loam and clay. The surface temperature is calculated from the surface energy balance equation in which the individual short and long wave fluxes are modified for the presence of clouds. This model also includes urban parameterizations for the representation of urban areas. A particular region may be treated as urban area by the model on the basis of the land use category. For such simulations, the surface energy balance equation considers the anthropogenic heat flux. The air pollution module uses predicted meteorology and turbulence at each time step and represents pollutant dispersion through a combined Eulerian and Lagrangian approach. It includes plume rise, gas and

aqueous phase chemical reactions, wet and dry deposition. The air pollution algorithms in TAPM consist of the following modules namely Eulerian Grid Module (EGM), the Lagrangian Particle Module (LPM), the Plume Rise Module and the Building Wake Module. The EGM solves prognostic equations for the mean and variance of concentration, and for the cross – correlation of concentration and virtual potential temperature. The LPM uses the PARTPUFF [Hurley, 1994] approach in which the pollutant is modeled like a particle in the vertical and a puff in the horizontal. This module is used to represent near source dispersion. As the name suggests, the Plume Rise Module accounts for the plume rise due the momentum and buoyancy of the plume whereas the Building Wake Module incorporates the effects of structures like buildings in dispersion calculations. The model can also simulate gas-phase photochemical reactions based on the Generic Reaction Set, Azzi et al. [1992], gas and aqueous-phase chemical reactions for sulfur dioxide and particles, and a dust mode for total suspended particles. Wet and dry deposition processes can also be included. Several validation studies have been carried out using this model. Details can be accessed in Hurley et al. [2008].

3. Algorithm of the Kalman filter

A brief summary of the equations used in the Kalman filter are presented here [Kalman, 1960; Kalman and Bucy, 1961; Galanis and Anadrastakis, 2002]. Let x_t be a state vector denoting an unknown process at time t and y_t be an observation vector. The change of process x from time $(t - 1)$ to time (t) is given by the system equation

$$x_t = F_t x_{t-1} + w_t \quad (1)$$

The measurement equation relates the observation vector and the unknown process and is given by

$$y_t = H_t x_t + v_t \quad (2)$$

The coefficient matrices F_t and H_t are called system matrix and observation matrix respectively. The change in x_t with respect to time is assumed to be random and hence the system matrix i.e. $F_t = 1$. They have to be determined prior to use of the filter. Also, the covariance matrix W_t of the random vector w_t and V_t of v_t should be known before using the filter. The vectors w_t and v_t should be normally distributed with zero mean, be time independent and mutually uncorrelated.

Nomenclature used:

$t - 1|t - 1$ previous time step

$t|t$ current time step

$t|t - 1$ intermediate time step

$t + 1$ future time step

Kalman filter theory gives a method for recursive estimation of the unknown state vector x_t utilizing all observation values y up to time t . Based on the vector x_{t-1} and its covariance P_{t-1} at time $(t - 1)$, the optimal estimate for their values at time t is

$$x_{t|t-1} = F_t x_{t-1|t-1} (F_t = 1) \quad (3)$$

$$P_{t|t-1} = F_t P_{t-1|t-1} F_t^T + W_t \quad (4)$$

Improvement in Near Surface NWP Model Output

Equations (3) and (4) are referred to as prediction equations. As soon as the new observation y_t is available at time t , the new value of x can be calculated at time t denoted by x_t . For doing this, the Kalman gain at time t , K_t has to be calculated.

$$K_t = P_{t|t-1} H_t^T (H_t \cdot P_{t|t-1} \cdot H_t^T + V_t)^{-1} (H_t = 1) \quad (5)$$

$$x_t = x_{t|t-1} + K_t \cdot (y_t - H_t \cdot x_{t|t-1}) \quad (6)$$

The Kalman gain decides how easily the filter adapts to new observations. The new value of covariance matrix is

$$P_t = (I - K_t \cdot H_t) P_{t|t-1} \quad (7)$$

Equations (5)-(7) are known as update equations. W_t and V_t are the covariance matrices of the random vectors w_t and v_t respectively. The initial values x_0 and P_0 have to be defined before running the filter. Since convergence is achieved quickly, their individual values are less significant. However, the calculation of the covariance matrices W_t and V_t affects the final outcome in terms of the Kalman gain and the adaptability of the filter to new observations. Following Homleid [1995] and Simonsen [1991], constant values are used for W_t and V_t . The variances W_t and V_t need to be specified. They can be estimated using statistical estimation procedure or by tuning them to make the Kalman filter work as required. The ratio W_t/V_t decides how much weight is given to the new observation. In Homleid [1995] simulations are carried out for different values of the ratio W_t/V_t . Higher values of W_t/V_t give a filter that is more sensitive to the latest observation while a lower value of the ratio gives a more stable filter. The study concluded that the ratio $W_t/V_t = 0.06$ is an optimum value for use in the Kalman Filter algorithm. Hence the present study uses the same value of the ratio W_t/V_t . Hence as per Homleid [1995], $W_t/V_t = 0.06$ is used in the present study.

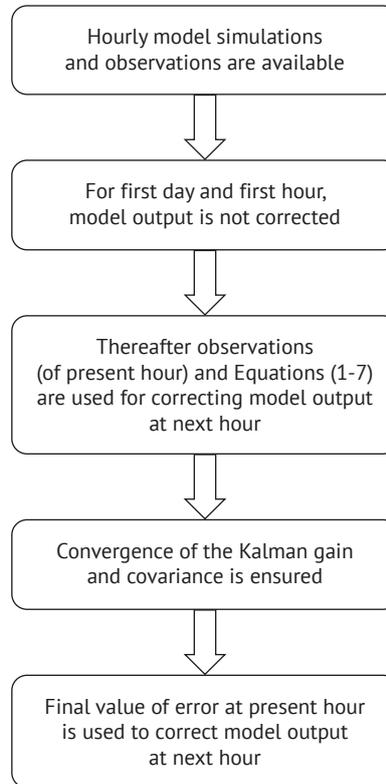


Figure 1. Flowchart of the Kalman filter.

Data assimilation is a process in which all available information is used to estimate as accurately as possible the state of a system [Parker, 2016]. In atmospheric data assimilation this information includes observations from various platforms like ground based stations, ships, airplanes, satellites and forecasts from numerical weather prediction (NWP) models. From observations which are scattered in space and time, the assimilation system delivers a complete gridded state estimate that provides initial conditions for all NWP models. Since the 1990s, data assimilation has also been used to construct long-term datasets for use in climate and other research, in a process known as retrospective analysis, or simply reanalysis. Reanalysis involves performing data assimilation for past periods, using a current NWP model and data assimilation method and some or all of the data that are available for those past periods. It produces a long sequence of comprehensive snapshots (analyses) of atmospheric conditions – a reanalysis dataset. The present study is an example of data assimilation using Kalman Filter method at a single point. Since the entire model output is not improved after Kalman filtering, it is not a reanalysis.

Following the procedure of Kalman Filter, it implies that the improved forecasted value at time $t + 1$ is the summation of model forecasted value at time $t + 1$ and the forecast error estimated at time t using the Kalman filter, and this has been used by other researchers [Anadranistakis et al. 2002; Anadranistakis et al. 2004; Galanis and Anadranistakis, 2002]. Moreover, in using the filter for any real application, error estimation at $t + 1$ is not available as measurement is not available at that time. Hence, even if the forecast error estimated at time t can improve the model forecast at $t + 1$, it is really useful from application point of view, and as mentioned earlier, this has been extensively used by other researchers. The entire procedure is described by a flowchart in Figure 1. Corrections are made only for a single grid point which is coincident with the observation site.

4. Material and Methods

Simulations are performed using TAPM model for 2013. The model is integrated in a nested domain configuration as shown in Figure 2. The grid resolution is 27 km for the outermost domain and 9 km, 3 km and 1 km respectively for the inner domains. Each domain has 40×40 grid points in the horizontal and 25 grid points in the vertical. With this configuration the study domain covers western India centered at the Trombay site ($19^\circ 1' 30''$ N, $72^\circ 56' 00''$ E). Model simulations are driven by using Conformal Cubic Atmospheric Model (CCAM) [Thatcher, 2008] which in turn uses synoptic scale meteorology from National Centre for Environmental Prediction (NCEP) at $1^\circ \times 1^\circ$ grid resolution. CCAM then provides TAPM compatible files which are used in TAPM. The terrain height, land use and soil type data used required for simulations with TAPM are obtained from United States Geological Survey (USGS) database. The initial and boundary conditions are obtained using Conformal Cubic Atmospheric Model (CCAM). Air temperature and relative humidity are measured on an hourly average basis using chart recorders (thermograph and hair hygograph respectively) inside a Stevenson screen. Wind data are measured using an ultrasonic anemometer at an interval of fifteen minutes on the terrace of a 20 m tall building. Wind data are further processed to produce hourly averages. All parameters are compared on average basis since model output corresponds to hourly average variables. In the output of TAPM, temperature and relative humidity are at 2-m from the surface and wind speed and direction are at 10 m from the surface. To facilitate comparison, the wind speed and wind direction data from model output are converted to zonal and meridional wind components using standard equations of vector calculus. The minor differences in the height of measured and modeled variables are less likely to have a significant impact in the final results. In the implementation of the filter, the observations at time t are used to correct the model results at time $(t + 1)$. Hence if observations are not available, those hours are excluded from the study.

Kalman filter algorithm is developed using equations described above and utilized for correction of model output and observations for one year namely 1st January 2013 to 31st December 2013. Variables corrected include wind speed, wind direction at 10-m height in terms of zonal and meridional wind components, 2-m air temperature and 2-m relative humidity. Following Galanis and Anadranistakis [2002], the initial value x_0 of the systematic error is 0.0 and initial variance P_0 is 4.0. Twenty five iterations are used at the beginning of the Kalman filter run to make sure that the initial conditions and model gain matrices will settle to converged stable values. Once this is done, the filter does not need to be iterated twenty five times again. When the next observed value is obtained, the filter just use the previous values (of Kalman Gain, and covariance) to obtain the updated value. Li et al. [2018] discuss the stability criteria for different Kalman filters. For the linear Kalman filter used in the present study, stability and boundedness of the error covariance is ensured as per this paper. The value of P_t in the previous iteration becomes the value of $P_{t|t-1}$ in the present iteration.

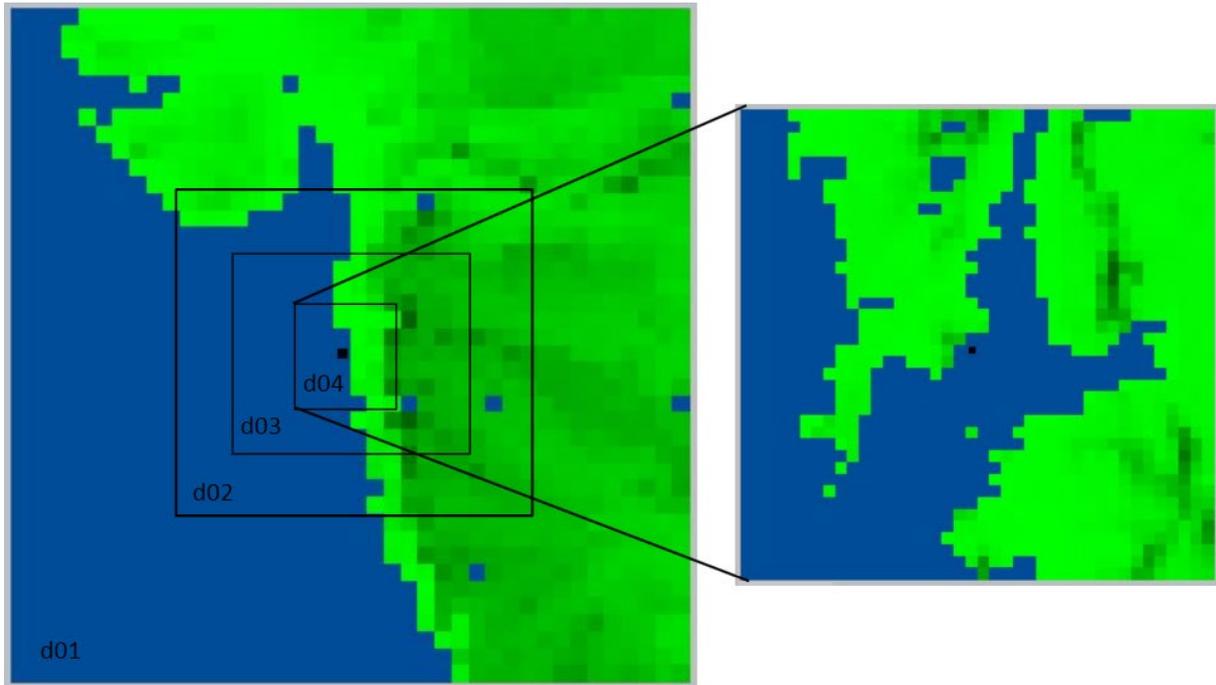


Figure 2. Configuration of the domains used for simulation in TAPM.

5. Results and Discussion

The Kalman filter algorithm is applied sequentially from 1st January 2013 0 IST to 31st December 2013 23 IST. For the first day and first hour when observations are not available, the model output is used without any correction factor. Thereafter for each subsequent hour, observations from the previous hour are used to obtain the correction factors which are applied to model data at the current hour. For better visualization, and to provide a snapshot of the results for the entire year, figures of monthly average hourly values of 2-m relative humidity, 2-m air temperature and 10-m zonal and meridional wind components from observations (green curve), TAPM (red curve) and TAPM corrected with Kalman filter (blue curve) are given in Figures 3-6 respectively. For example at 08.00 am in any month, monthly average curves are averages over the entire month. Referring to Figures-3 and 4, it is seen that after Kalman filtering the model biases present in the calculation of 2-m relative humidity and 2-m air temperature from the NWP model are eliminated at all hours. Also, eliminated completely are the differences in the times at which model and observations show minimum relative humidity and/or maximum air temperature in a given month. Here it is to be noted that, the minima/maxima refers to the monthly average value. The magnitude of the minima/maxima as well as the diurnal variation has also improved after Kalman filtering. From both the figures, it is seen that the blue and green curves are almost overlapping indicating the utility of the filter.

Figures 5 and 6 describe the variation of zonal and meridional wind components at 10-m height on monthly average hourly basis. These wind components are utilized to estimate the wind speed and wind direction. A common observation from both the figures is that the over prediction in wind speed is minimized because of the Kalman filter. This is a common feature across almost all months of the year. Also, for most of the months the blue and green curves are overlapping indicating the advantage of Kalman filtering. The utility of the Kalman filter is apparent in the estimation of wind speed as well as wind direction. In general, it can be concluded that the diurnal pattern of wind has also improved due to Kalman filtering.

To evaluate the skill of the filter in a better way, time series of monthly average hourly bias values in 2-m relative humidity, 2-m air temperature and 10-m zonal and meridional wind components from TAPM (red curve) and TAPM corrected with Kalman filter (blue curve) are given in Figures 7-10. Referring to Figure-7 which shows the bias in 2-m relative humidity before and after filtering, it is seen that TAPM has a bias varying between -45% to 30% throughout the year except the monsoon period during June to September. In the monsoon season, the bias in TAPM is around -15%. It should be noted that the reduced bias in the monsoon season could be due to the fact

that relative humidity is high during the monsoon season and also diurnal variation is less. It is also seen that most of the months are characterized by over and under prediction of relative humidity (i.e. positive and negative bias) occurring on a time scale of a day. Bias in TAPM corrected with Kalman filter is within $\pm 15\%$ throughout the year with reduced diurnal variation as compared to TAPM.

Likewise from Figure 8 which shows the bias in 2-m air temperature before and after filtering, it is seen that TAPM has a bias varying between $\pm 5^\circ\text{C}$ throughout the year except the monsoon period during June to September. Here too, most of the months are characterized by over and under prediction of air temperature (i.e. positive and negative bias) occurring on a time scale of a day. In the monsoon season, the bias in TAPM is between $\pm 2^\circ\text{C}$. As in case of relative humidity, the bias in the monsoon season is less because of the small diurnal variation in air temperature. The same in TAPM corrected with Kalman filter is within $\pm 2^\circ\text{C}$ throughout the year with reduced diurnal variation.

Finally, the time series of biases in 10-m zonal and meridional wind components are shown in Figures 9 and 10 respectively. The advantage of Kalman filtering is clearly evident from these figures. Except the monsoon season, for TAPM, the bias in 10-m zonal wind (shown in Figure 9) varies between $\pm 4\text{ m/s}$. The same during monsoon season is 2 m/s . The bias is reduced during the monsoon season, owing to nearly constant wind direction in this season. After Kalman filtering the biases are reduced to within $\pm 1\text{ m/s}$ throughout the year and near zero in the monsoon season. In case of 10-m meridional wind, bias in TAPM is seen to vary between -3 m/s to $+2\text{ m/s}$ throughout the year with reduced bias in the monsoon season for reason already described. Like 10-m zonal wind, when TAPM results are corrected with Kalman filter, the bias in 10-m meridional wind is reduced to within $\pm 1\text{ m/s}$ throughout the year and near zero in the monsoon season.

While bias is used to understand the over and under prediction of a model, Root Mean Square Error (RMSE) gives an idea of the effect of large errors. Hence time series of monthly average hourly RMSE values in 2-m relative humidity, 2-m air temperature and 10-m zonal and meridional wind components from TAPM (red curve) and TAPM corrected with Kalman filter (blue curve) are given in Figures 11-14. Except the monsoon season, RMSE in TAPM output of 2-m relative humidity is seen to vary between 7% to as high as 45% with large diurnal variation occurring in a day. The same in TAPM corrected with Kalman filter is $\sim 10\%$ with reduced diurnal variation. The discussion for 2-m air temperature follows that of 2-m relative humidity with TAPM showing large RMSE as well as large diurnal variation in RMSE as compared to TAPM corrected with Kalman filter.

Figures 13 and 14 show the RMSE in 10-m wind components, and the reduction in RMSE after Kalman filtering is clearly seen in the two figures. The reduction in biases and RMSE in wind components is evident in a better prediction of wind direction as well. From the above discussion, it can be concluded that the reduction in model bias and RMSE in all the variables after Kalman filtering leads to an improvement in model performance at the location of comparison.

While time series plots are useful in studying diurnal variation, they have less utility in estimating the predominant wind direction and wind speed for a site. Both these are important in atmospheric dispersion studies. Hence, in Figure 15, wind roses are plotted on annual basis using the same wind speed classes for TAPM data, observed data and TAPM data corrected with Kalman filter. In Figure 15, the extreme left wind rose is plotted using TAPM data, centre one using observed data and extreme right using TAPM data corrected with Kalman filter. Referring to the wind rose plotted using TAPM data alone, the over prediction in wind speed is apparent from the color of the petals in the wind rose (blue green color indicates wind speed of 3-5 m/s).

The same in the wind roses plotted using observed wind data and model data corrected with Kalman filter indicates wind speed of 1-2 m/s. Also, the percentage of calm hours (wind speed less than 0.5 m/s) in the model results is only 0.64%. The same in the wind roses plotted using observed wind data and TAPM model data corrected with Kalman filter indicates calm frequency of $\sim 20\%$ and 18% respectively. Of equal importance in atmospheric dispersion studies, is the prediction of wind direction by a NWP model. Again referring to the wind rose plotted using TAPM data alone, it is seen that the NWP model has predominantly indicated a south westerly to north westerly flow (wind direction varying from 225° to 315°) for most of the year. Some north easterly flow is also seen. From observed data it is seen that only few wind sectors namely South South West (SSW), North North East (NNE) and South (S) assume predominance on annual basis for this site. Among the three, the SSW wind sector has higher frequency of occurrence. Apart from these three, other wind sectors have negligible frequency of occurrence. When the model data are corrected using the Kalman filter, the SSW wind sector is correctly shown to be the predominant wind sector. It is to be noted that even after Kalman filtering, there are differences with respect to the observed wind rose in terms of wind direction persistence and frequency of calm hours. However, as compared to the wind rose from NWP model data alone, there is a significant improvement after Kalman filtering.

Improvement in Near Surface NWP Model Output

Finally, to quantify the utility of Kalman filter, some standard statistical indices of model performance like average of model, average of observation, standard deviation of model, standard deviation of observation, bias, mean absolute error, root mean square error, index of agreement, and correlation coefficient described in Table 1 are computed using model output before and after Kalman filtering. Equations 8 to 12 are used for the computation of statistical indices. Data availability is more than 99% for 2-m air temperature and 2-m relative humidity and ~94% for wind components at 10-m. As evident from Table 1, model biases in the variables used for comparison are eliminated after Kalman filtering. The mean absolute error (MAE) and RMSE are reduced by more than a factor of two after Kalman filtering. Also, there is an improvement in the index of agreement and correlation coefficient due to Kalman filtering.

$$BIAS = \frac{1}{N} \sum_{i=1}^N (P_i - O_i) \quad (8)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_i - O_i| \quad (9)$$

$$RMSE = \frac{1}{N} \sum_{i=1}^N \sqrt{(P_i - O_i)^2} \quad (10)$$

$$IOA = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (|P_i - O_{mean}| + |O_i - O_{mean}|)^2} \quad (11)$$

$$CORR = \frac{N(\sum_{i=1}^N O_i P_i) - (\sum_{i=1}^N O_i)(\sum_{i=1}^N P_i)}{\sqrt{[N(\sum_{i=1}^N O_i^2) - (\sum_{i=1}^N O_i)^2][N(\sum_{i=1}^N P_i^2) - (\sum_{i=1}^N P_i)^2]}} \quad (12)$$

Statistical Indices	TAPM				TAPM + KF			
	2-m TT	2-m RH	10-m u	10-m v	2-m TT	2-m RH	10-m u	10-m v
NO. OF HRS	8721	8736	8269	8269	8721	8736	8269	8269
% AVAILABILITY	99.55	99.72	94.39	94.39	99.55	99.72	94.39	94.39
AVG_MOD	25.71 °C	71.84%	0.99 m/s	-0.23 m/s	26.68 °C	76.18%	0.32 m/s	0.21 m/s
AVG_OBS	26.68 °C	76.18%	0.32 m/s	0.21 m/s	26.68 °C	76.18%	0.32 m/s	0.21 m/s
STD_DEV_MOD	3.55 °C	18.05%	2.07 m/s	2.36 m/s	4.06 °C	20.35%	0.72 m/s	1.46 m/s
STD_DEV_OBS	4.31 °C	21.65%	0.69 m/s	1.46 m/s	4.31 °C	21.65%	0.69 m/s	1.46 m/s
BIAS	-0.98 °C	-4.34%	0.67 m/s	-0.45 m/s	0.0 °C	0.0%	0.0 m/s	0.0 m/s
MAE	1.77 °C	12.55%	1.73 m/s	1.48 m/s	0.65 °C	4.43%	0.45 m/s	0.51 m/s
RMSE	2.26 °C	16.29%	2.00 m/s	1.78 m/s	0.92 °C	6.48%	0.63 m/s	0.68 m/s
IOA	0.92	0.82	0.45	0.76	0.98	0.98	0.77	0.94
CORR	0.88	0.70	0.41	0.68	0.98	0.95	0.60	0.89

AVG_MOD – Average of model, AVG_OBS – Average of observation, STD_DEV_MOD – Standard Deviation of model, STD_DEV_OBS – Standard Deviation of observation, BIAS – bias, MAE – Mean Absolute Error, RMSE – Root Mean Square Error, IOA – Index of Agreement, CORR – Correlation Coefficient.

Table 1. Model Performance Statistics.

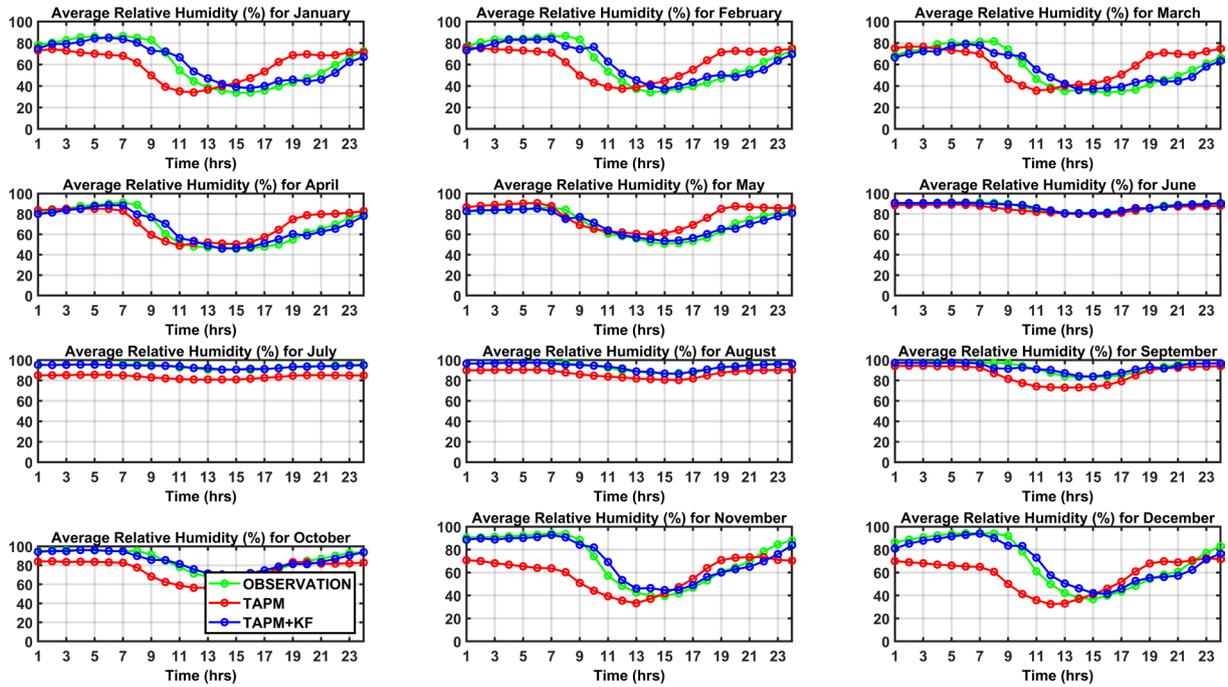


Figure 3. Monthly average hourly time series of 2-m relative humidity from observed data (green curve), TAPM data (red curve) and TAPM + KF data (blue curve).

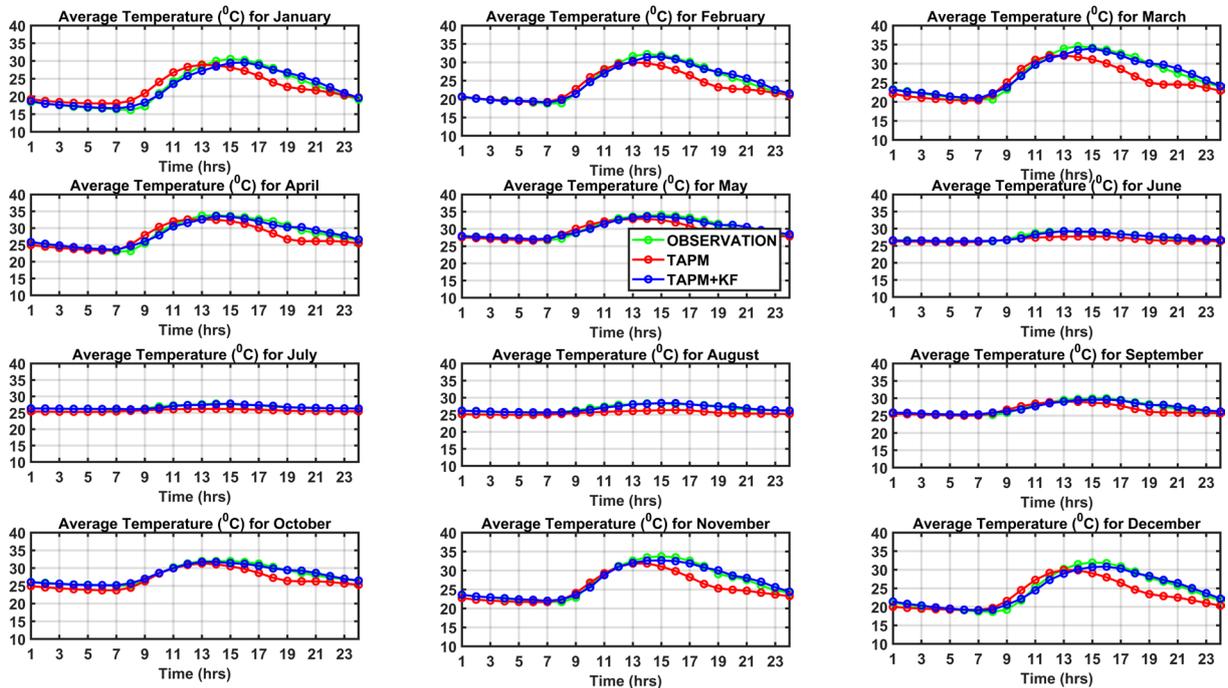


Figure 4. Monthly average hourly time series of 2-m air temperature from observed data (green curve), TAPM data (red curve) and TAPM + KF data (blue curve).

Improvement in Near Surface NWP Model Output

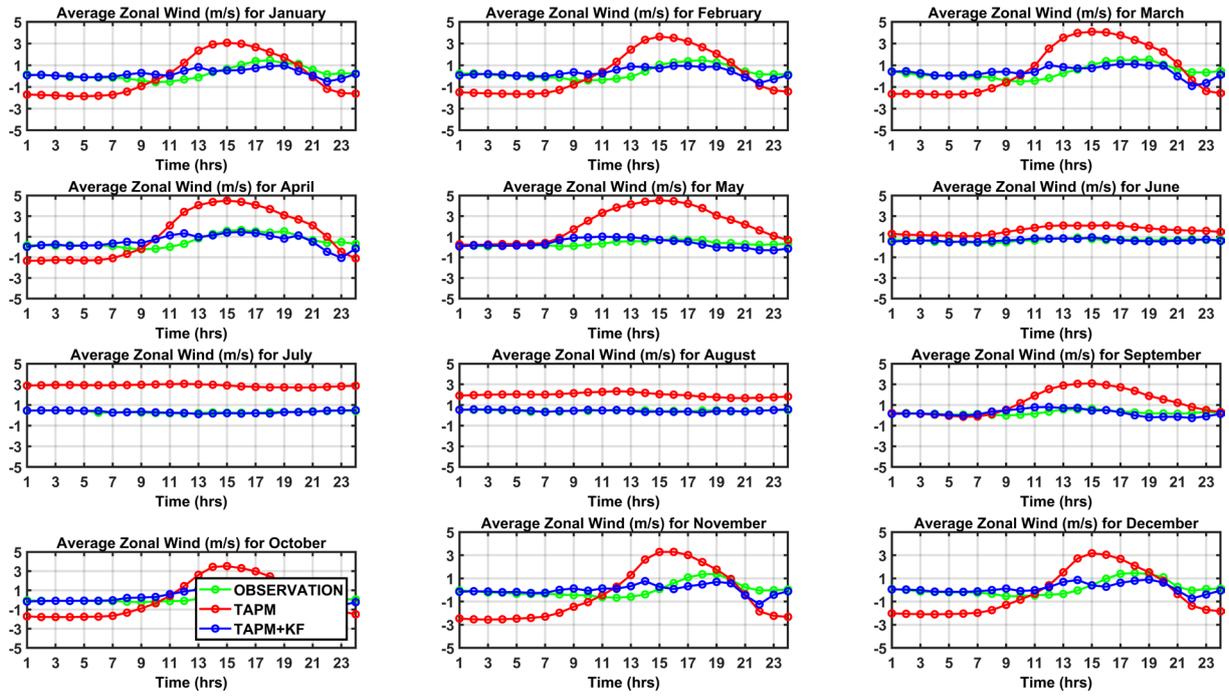


Figure 5. Monthly average hourly time series of 10-m zonal wind from observed data (green curve), TAPM data (red curve) and TAPM + KF data (blue curve).

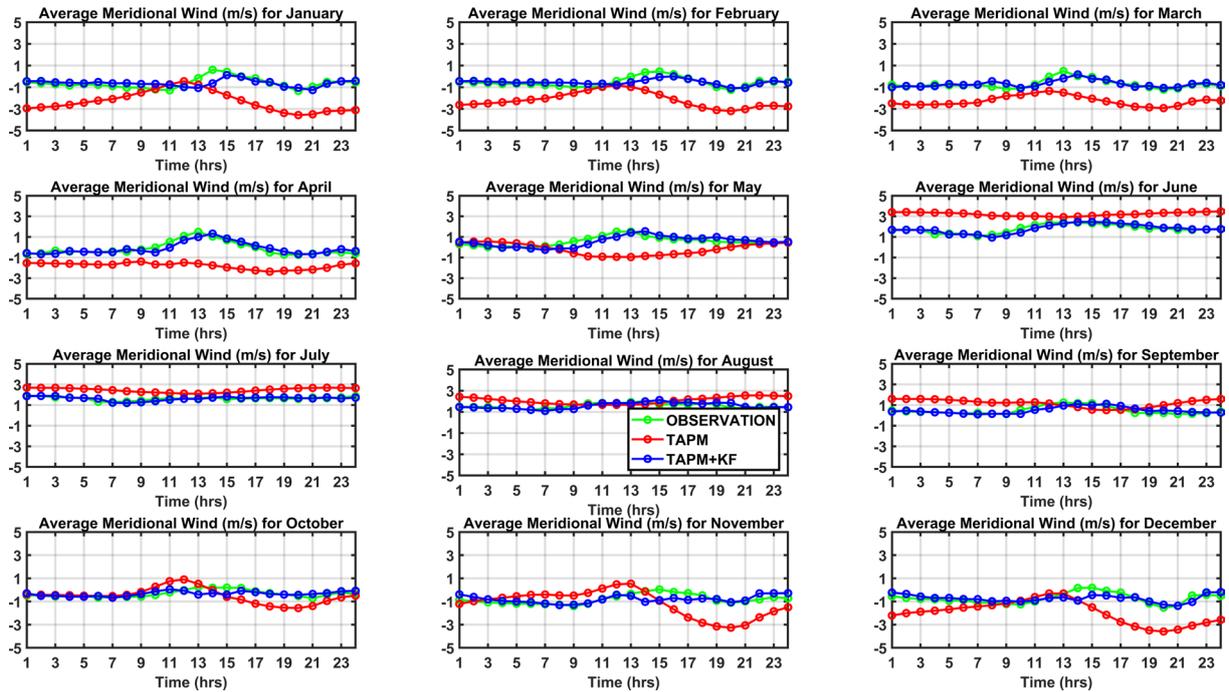


Figure 6. Monthly average hourly time series of 10-m meridional wind from observed data (green curve), TAPM data (red curve) and TAPM + KF data (blue curve).

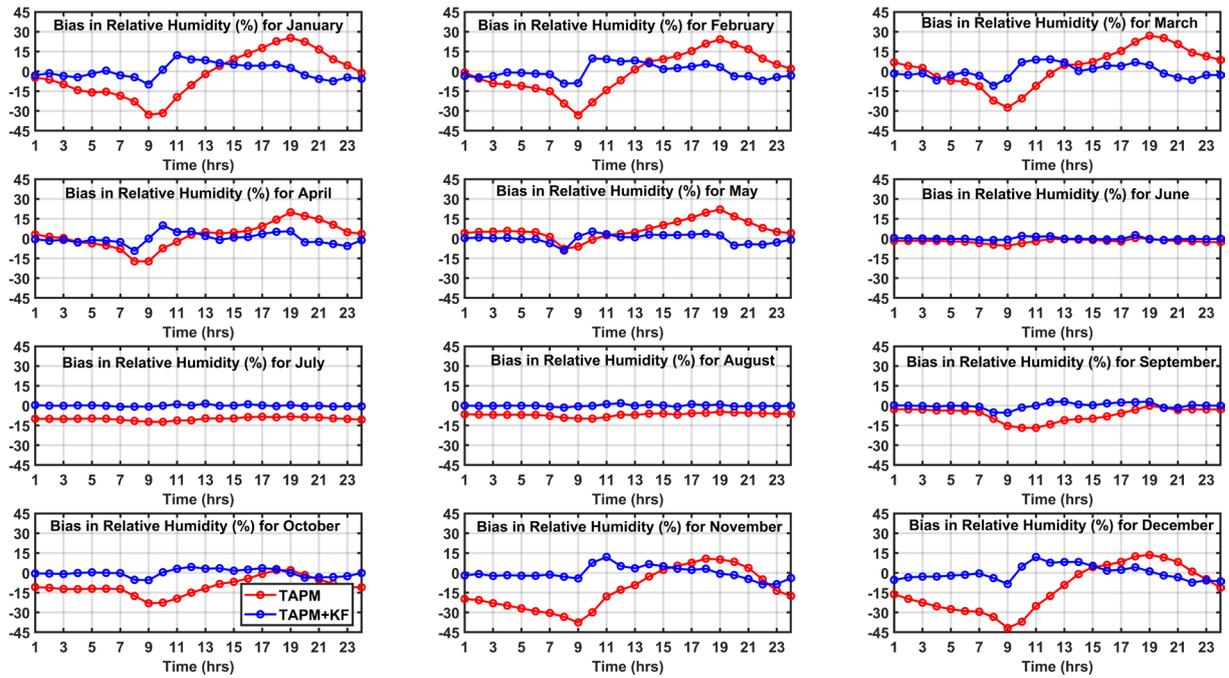


Figure 7. Monthly average hourly time series of bias in 2-m relative humidity from TAPM data (red curve) and TAPM + KF data (blue curve).

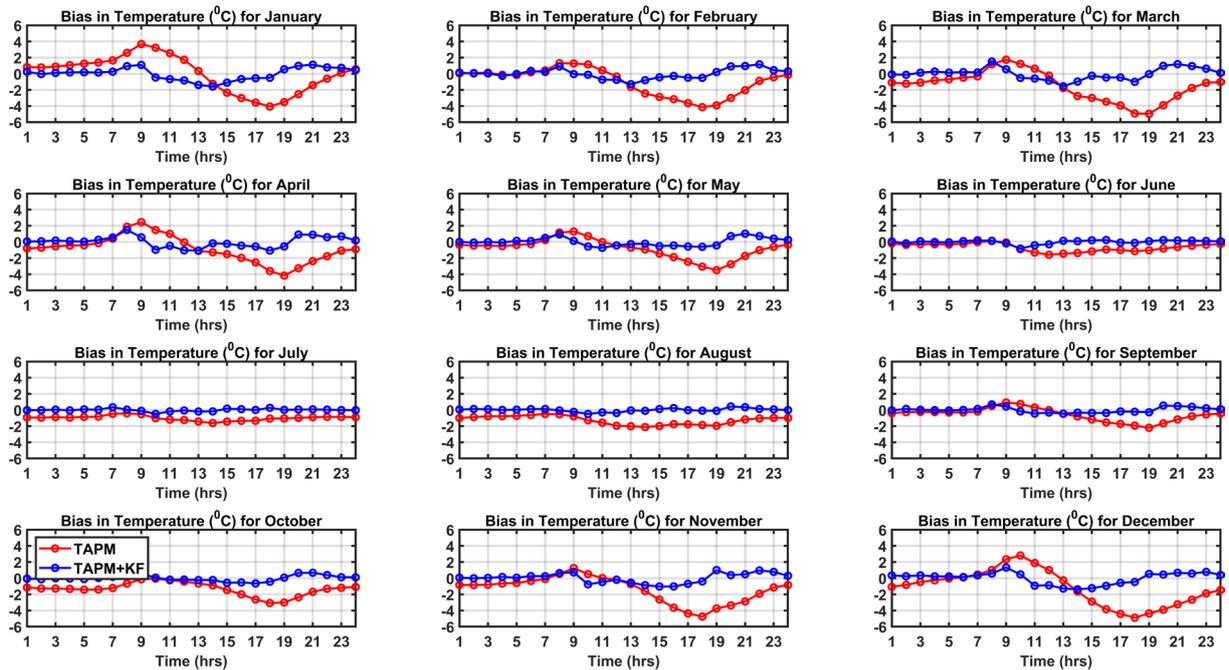


Figure 8. Monthly average hourly time series of bias in 2-m air temperature from TAPM data (red curve) and TAPM + KF data (blue curve).

Improvement in Near Surface NWP Model Output

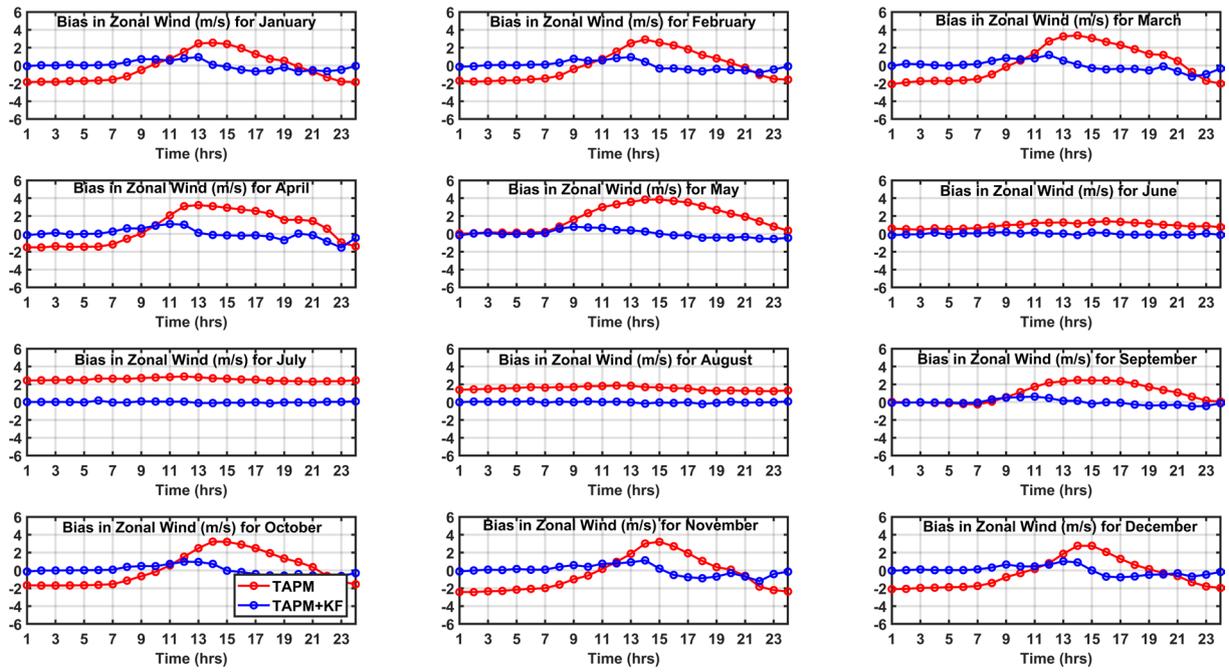


Figure 9. Monthly average hourly time series of bias in 10-m zonal wind from TAPM data (red curve) and TAPM + KF data (blue curve).

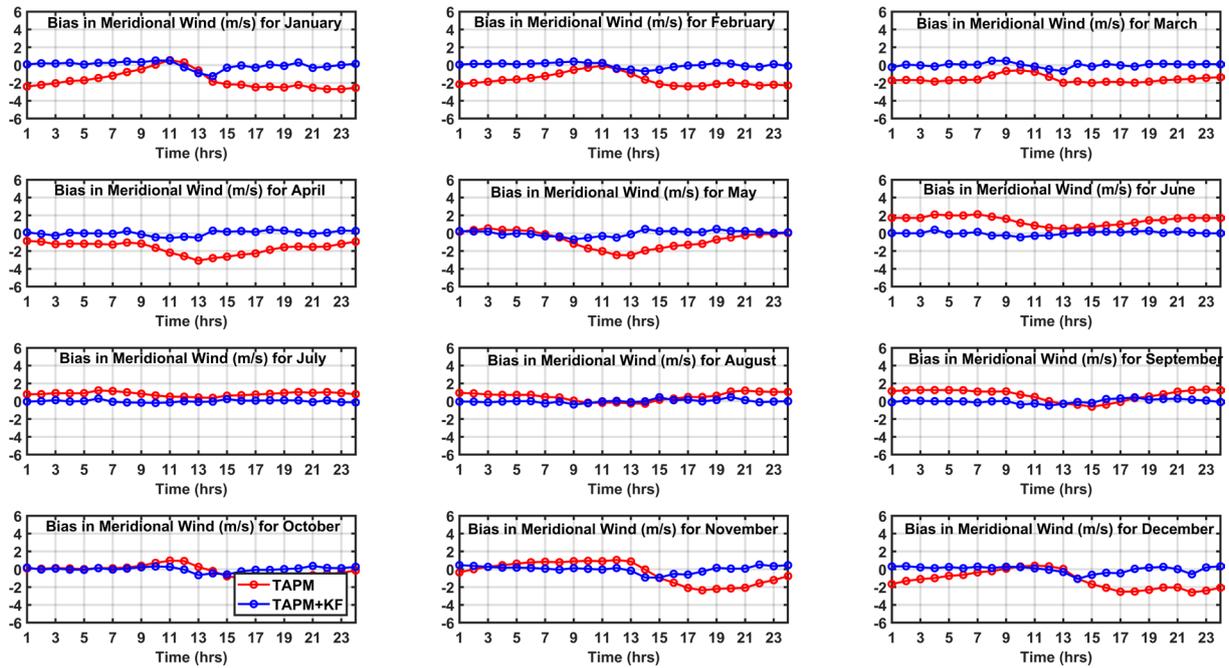


Figure 10. Monthly average hourly time series of bias in 10-m meridional wind from TAPM data (red curve) and TAPM + KF data (blue curve).

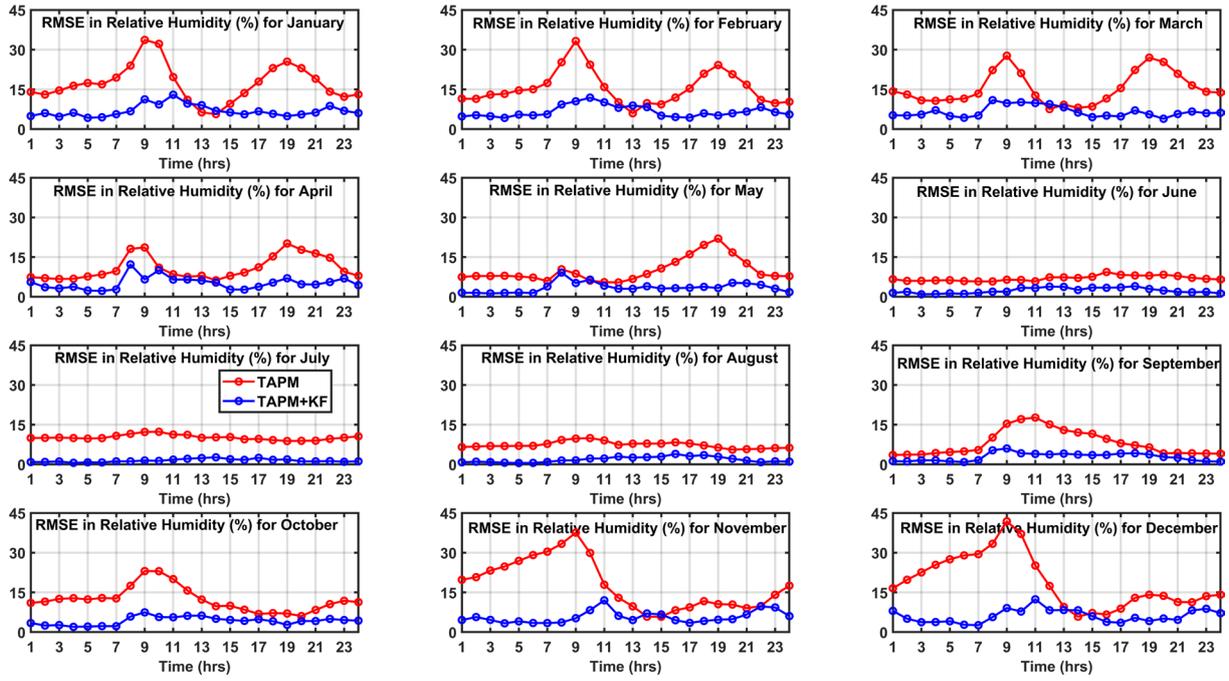


Figure 11. Monthly average hourly time series of RMSE in 2-m relative humidity from TAPM data (red curve) and TAPM + KF data (blue curve).

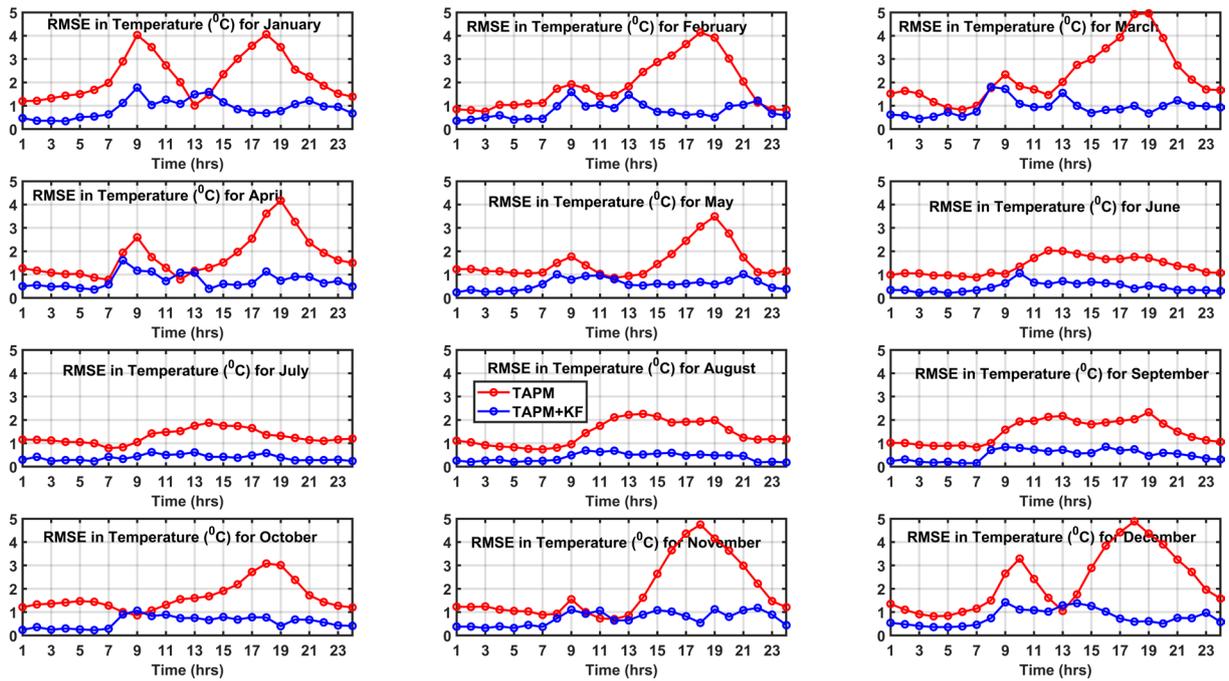


Figure 12. Monthly average hourly time series of RMSE in 2-m air temperature from TAPM data (red curve) and TAPM + KF data (blue curve).

Improvement in Near Surface NWP Model Output

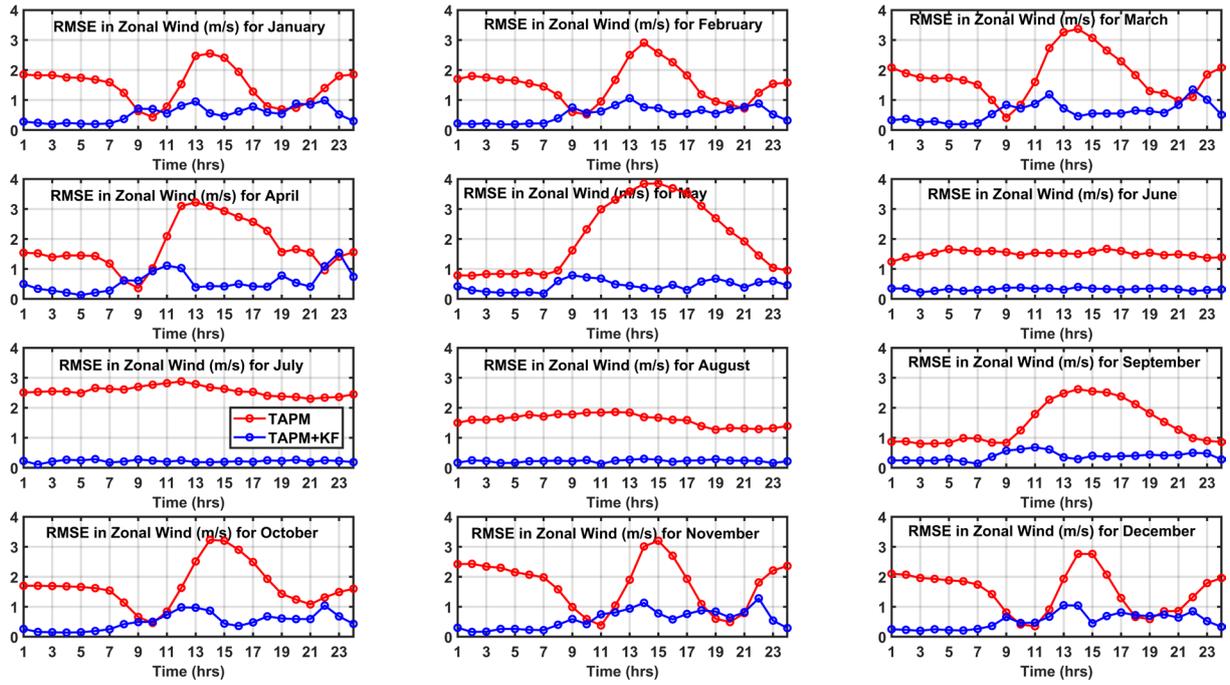


Figure 13. Monthly average hourly time series of RMSE in 10-m zonal wind from TAPM data (red curve) and TAPM + KF data (blue curve).

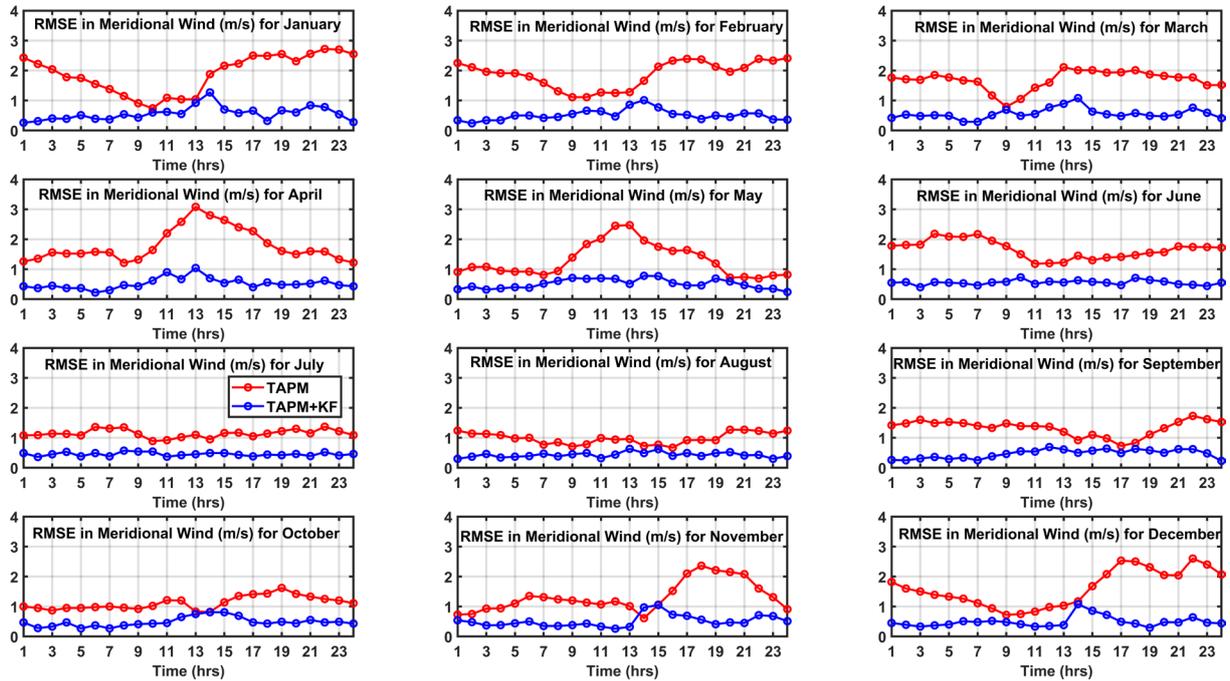


Figure 14. Monthly average hourly time series of RMSE in 10-m meridional wind from TAPM data (red curve) and TAPM + KF data (blue curve).

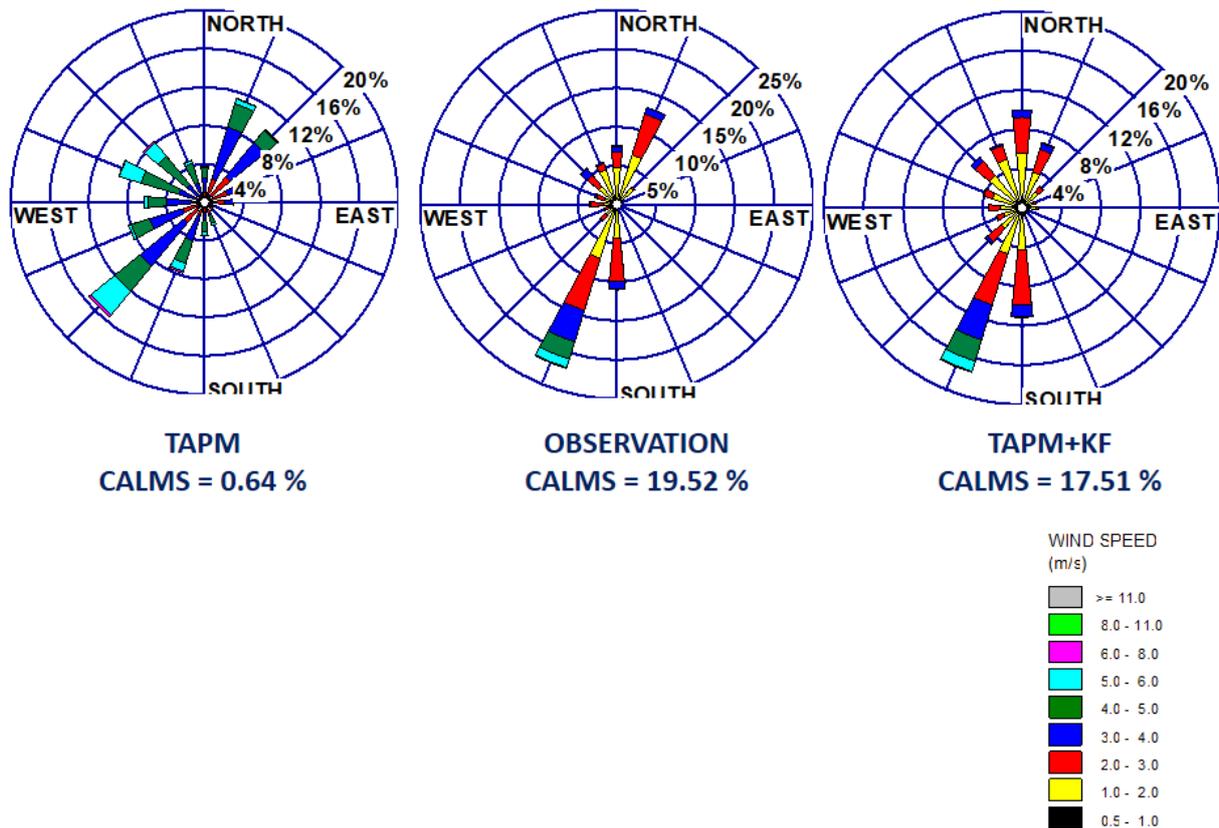


Figure 15. Annual wind roses from TAPM data (extreme left), observed data (centre) and TAPM + KF data (extreme right) (Calms refers to the percentage of hours when the wind speed was less than or equal to 0.5 m/s).

6. Conclusions

In the present study, a one dimensional Kalman filter framework for improvement of results from NWP model has been utilized. The algorithm has been applied in hind cast mode for correction of output from The Air Pollution Model (TAPM) for 2013. After Kalman filtering, the variables namely 2-m air temperature, 2-m relative humidity, zonal and meridional wind components at 10-m compare better with observations. Specific advantages of the Kalman filter are that on an annual basis, biases in all the variables are eliminated, over prediction in wind speed is minimized and wind direction is better reproduced. Also, all standard statistical indices of model performance computed after Kalman filtering are superior to those computed using only model output. Time series plots of bias and RMSE in model after Kalman filtering indicate the advantage of Kalman filtering. Given the ease of implementation and minimal run time of the algorithm even for annual simulations, it is a simple and cost effective tool to reduce the errors of NWP models. To conclude, it can be said that Kalman filter algorithm is a value addition to a NWP model.

Acknowledgements. The authors express their gratitude to Dr. D. K. Aswal, Director, Health, Safety and Environment Group, BARC and Shri Probal Chaudhury, Head Radiation Safety Systems Division, BARC for constant encouragement and fruitful discussions during the period of study. The authors also thank Shri B. B. Ghorpade of Radiation Safety Systems Division for assistance in measurements of meteorological variables.

Data Availability. The Final Analyses (FNL) data for this study are from the Research Data Archive (RDA) which is maintained by the Computational and Information Systems Laboratory (CISL) at the National Center for Atmospheric Research (NCAR). NCAR is sponsored by the National Science Foundation (NSF). The original data are available from the RDA (<http://rda.ucar.edu>) in dataset number ds083.2.

References

- Anadranistakis, M., V. Kotroni, K. Lagouvardos and K. Skouras (2002). Combination of Kalman filter and an empirical method for correction of near-surface temperature forecasts: Application over Greece, *Geophys. Res. Lett.*, 29, 16, 23-1-23-4.
- Anadranistakis, M., K. Lagouvardos, V. Kotroni and H. Eleftheriadis (2004). Correcting temperature and humidity forecasts using Kalman filtering: potential for agricultural protection in Northern Greece, *Atmos. Res.*, 71, 115-25.
- Azzi, M., G.M. Johnson and M. Cope (1992). An introduction to the generic reaction set photochemical smog mechanism, In Proceedings of the 11th International clean Air and Environment Conference, Brisbane, Clean Air Society of Australia and New Zealand.
- Crochet, P. (2004). Adaptive Kalman filtering of 2-metre temperature and 10-metre wind-speed forecasts in Iceland, *Meteor. Appl.*, 11, 173-187.
- Cassola, F. and M. Burlando (2012) Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output, *Appl. Energy*, 99, 154-166.
- Galanis, G. and M. Anadranistakis (2002). A one dimensional Kalman filter for the correction of near surface temperature forecasts, *Meteor. Appl.*, 9, 437-441.
- Gallego, A.J., A. J. Sanchez, M. Berenguel and E. F. Camacho (2020). Adaptive UKF-based model predictive control of a Fresnel collector field, *J. Process. Control.*, 85, 76-90.
- Glahn, H.R. and D. A. Lowry (1972). The use of model output statistics (MOS) in objective weather forecasting, *J. Appl. Meteor.*, 11, 1203-1211.
- Homleid, M. (1995). Diurnal corrections of short-term surface temperature forecasts using the Kalman filter, *Weather Forecast.*, 10, 689-707.
- Hu, G., W. Wang, Y. Zhong, B. Gao and C. Gu (2018). A new direct filtering approach to INS/GNSS integration, *Aerosp. Sci. Tech.*, 77, 755-764.
- Hu, G., B. Gao, Y. Zhong, C. Gu (2020). Unscented Kalman filter with process noise covariance estimation for vehicular ins/gps integration system, *Inf. Fusion.*, 64, 194-204.
- Huh, C.A., C.Y. Lin and S.C. Hsu (2013). Regional Dispersal of Fukushima-Derived Fission Nuclides by East-Asian Monsoon: A Synthesis and Review, *Aerosol. Air. Qual. Res.*, 13, 537-544.
- Hurley, P.J. (1994). PARTPUFF - A Lagrangian particle/puff approach for plume dispersion modeling applications, *J. Appl. Meteor.*, 33, 285-294.
- Hurley, P.J. (2005). TAPMV4. Part 1: Technical Description. CSIRO Marine and Atmospheric Research Paper 25, 59.
- Hurley, P.J., Physick, W.L. and A. K. Luhar (2005). TAPM: A practical approach to prognostic meteorological and air pollution modeling, *Environ. Modell. Software.*, 20, 737-752.
- Hurley, P.J., M. Edwards and A. K. Luhar (2008). TAPMV4. Part 2: Summary of some verification studies. CSIRO Marine and Atmospheric Research Paper, 26, 31.
- Kalman, R.E. (1960). A new approach to linear filtering and prediction problems, *Trans. ASME.*, 82, 35-45.
- Kalman, R.E. and R. S. Bucy (1961). New results in linear filtering and prediction problems, *Trans. ASME.*, 83, 95-108.
- Katata, G., M. Ota, H. Terada, M. Chino and H. Nagai (2012). Atmospheric Discharge and Dispersion of Radionuclides during the Fukushima Dai-ichi Nuclear Power Plant accident. Part I: Source Term Estimation and Local-scale Atmospheric Dispersion in Early Phase of the Accident, *J. Environ. Radioact.*, 109, 103-113.
- Klein, W.H. and F. Lewis (1970). Computer Forecasts of maximum and minimum temperature, *J. Appl. Meteor.*, 9, 350-359.
- Li, W., G. Wei, D. Ding, Y. Liu, and F.E. Alsaadi (2018). A New Look at Boundedness of Error Covariance of Kalman filtering, *IEEE. Trans. Syst. Man. Cybern.*, 48, 2, 309-314.
- Metia, S., S.D. Oduro, H.N. Duc and Q. Ha (2016). Inverse Air-Pollutant Emission and Prediction Using Extended Fractional Kalman filtering, *IEEE J. Sel. Top. Appl. Earth. Obs. Remote. Sens.*, 9, 5, 2051-2063.
- Metia, S., Q.P. Ha, H.N. Duc, M. Azzi (2018). Estimation of Power Plant Emissions with Unscented Kalman filter, *IEEE J. Sel. Top. Appl. Earth. Obs. Remote. Sens.*, 11, 8, 2763-2772.
- Metia S, Q.P. Ha, H.N. Duc, Y. Scorgie (2020). Urban air pollution estimation using unscented Kalman filtered inverse modeling with scaled monitoring data, *Sustain Cities Soc.*, 54:101970, <https://doi.org/10.1016/j.scs.2019.101970>.
- Oktaviana, Y.V., E. Apriliani and D. K. Arif (2018). Fractional Kalman filter to estimate the concentration of air pollution, *J. Phys. Conf. Series.*, 1008, 012008.

- Parker, W. (2016). Reanalyses and Observations: What's the Difference?, *Bull. Amer. Meteor. Soc.*, 69, 1565-1572.
- Simonsen, C. (1991). Self adaptive model output statistics based on Kalman filtering. Lectures and Papers Presented at the WMO Training Workshop on the Interpretation of NWP Products in terms of local weather phenomena and their verification, Wageningen the Netherlands; WMO, XX-33-XX-37.
- Thatcher, M. (2008). CCAM GUI instruction manual for TAPM users (v905t). In *CSIRO Marine and Atmospheric Research*, 30.
- Thatcher, M. and P.J. Hurley (2010). A customizable downscaling approach for local-scale meteorological and air pollution forecasting: Performance evaluation for a year of urban meteorological forecasts, *Env. Modell. Software*, 25, 82-92.
- Voyant, C., F. Motte, A. Fouilloy, G. Notton, G., C. Paoli and M. Nivet (2017). Kalman filtering and classical time series tools for global radiation protection. In 4th International Conference on Energy, Sustainability and Climate Change, IRCC, Greece.

***CORRESPONDING AUTHOR: Roopashree SHRIVASTAVA,**

Radiation Safety Systems Division, Bhabha Atomic Research Centre, Mumbai – 400 085

e-mail: roopa@barc.gov.in