沙漠研究 **22**-1, 183 -186 (2012)
Journal of Arid Land Studies

"ICAL 1 / DT X"
－Refereed Paper－

Effect of Forgetting Factor on Gray Box Auto Regressive Exogenous Algorithm for Short Term Diurnal Temperature Forecast in Remote Area Regions

Saurabh KUMAR[1)], Satyendra Narayan SINGH[1)], Umesh Chandra RAY[2)] and Sanjay KUMAR*[3)]

Abstract: Remote arid areas are devoid of weather data which may be called fairly accurate. They have to rely upon predictions in nearby cities, sometimes more than 300 kms away and that too in completely different demographic and physical structure. It is important to establish alternative method of weather prediction. Model based weather prediction is presented in this paper for such application. It is based upon the principle of dynamic modelling and parameter estimation. These parameters have no physical meaning but a change in their behaviour predicts the change in climatic condition. The selection of model based upon forgetting factor is preferred after evaluating Akaike information criterion, Bayesian information criterion and, Deviance information criterion. It can be interpreted accordingly to predict the climate at local scale. The methodology is based upon smaller data accumulation but provides fairly accurate results.

Key Words: Arid area, Forgetting factor, Gray box modelling, Weather

1. Introduction

The processes and systems in climatological prediction get more complex and difficult with our understanding of non-linear processes involved. Here, however, high reliability is not required. Short term weather predictions in remote areas are extrapolated from weather data collected at nearby cities, sometimes more than 300 km away with completely different geographical (physical) and demographic conditions. Remote sensing data are not cost-effective and reliable unless supported by ground observations, for such predictions. It is neither practical nor economically feasible to establish and maintain weather station at remote area locations unless it is strategically important. Forecast for short term period ranging from 3 to 7 days is likely to have decisive impact on sustainable developmental process and planning, especially when Very Large Scale Photovoltaic Systems and renewable energy based systems are being planned in remote areas and deserts. Over the last decades, there has been considerable research activity in the field of model based reasoning especially for information poor processes. Model based analysis (MBA) is based upon behavior descriptions and interconnections of the parameters affecting the whole system. It can be understood as the interaction of observed and predicted values. The basic idea of dynamic modelling can be described as a simple comparison between measured and modeled quantities. A straight forward model structure can be obtained if the characteristics of each component in the system are described by equations derived from the basic laws of physics. The number of parameters influencing climate is large and requires supercomputer for computations, which renders it enviable due to cost-effectiveness of requirement for remote areas.

Recursive Auto Regressive Exogenous (RARX) system identification methodology with forgetting factor is developed for such applications. The model rectifies itself on the basis of noises in the system and can be trained with small data sets. The parameters of the model have no physical significance but their deviation can be used to predict the changes in the climatic condition. They are easier to set up and require much less detailed information about the system to be modeled. However, for the identification phase, models may need rather rich data sets for a correct estimation of parameters, especially if it is a dynamic model. Three criteria is used to evaluate the model and its number of parameters, Akaike information criterion, Bayesian information criterion and, Deviance information criterion. The RARX model with fairly small no. of parameters with forgetting factor of 0.997 is found quite efficient and useful in the present case. Forgetting factor allows discounting past information and reducing computational time and capacity for prediction of temperature on diurnal basis.

* Corresponding Author: prof.kumars@gmail.com
P.O. Box - 5, Muzaffarpur-842001, Bihar, India
1) Centre for Renewable Energy and Environmental Research
2) Dept. of Physics, Ranchi University, Jharkhand, India
3) National Institute of Technology, Patna, Bihar, India

2. Model Formulation

A typical recursive identification algorithm is given by,

$$\theta'_n = \theta'(t-1) + K(t)((y(t) - y'(t)) \quad (1)$$

here, $y'(t)$ is the parameter estimate at time t, and $y(t)$ is the observed output at time t. Moreover, $y'(t)$ is a prediction of the value $y(t)$ based on the observations up to time $t-1$ and also based on the current model. The gain $K(t)$ determines in what way the current prediction error $y(t)$-$y'(t)$ affects the update of the parameter estimate. It is typically chosen as,

$$K(t) = Q(t)\varphi(t) \quad (2)$$

where, $\varphi(t)$ is (an approximation of) the gradient with respect to y of $y'(t \mid \theta)$. The latter symbol is the prediction of $y(t)$ according to the model described by $y(t)$. $Q(t)$ is the matrix which determines the adaptation gain and the way parameters are updated. The model structure like AR and ARX that correspond to linear regression can be written as,

$$y(t) = \varphi(t)^T \theta(t) + e(t) \quad (3)$$

where, the regression vector $\varphi(t)$. contains old values of observed inputs and outputs, and $e(t)$ is the noise source. The most logical approach to the adaptation problem is to assume a certain model for how "true" parameter $y'(t)$ changes. A typical choice is to describe these parameters as a random walk,

$$\theta'(t) = \theta'(t-1) + \omega(t) \quad (4)$$

here, $\omega(t)$ is assumed to be white Gaussian noise with covariance matrix,

$$E[\omega(t)\,\omega^{\mathrm{T}}(t)] = R_1 \quad (5)$$

where, R_1 is variation of the variance. Considering, underlying description of the observation, a linear regression, an optimal choice of Q_n in "eq. 1 and 2" can be computed from the Kalman filter. The above method is modified to discount old measurements so that the model adopts the changing situation dynamically. An observation that is Γ samples old carries a weight that is R_2^{τ} of the weight of the most recent observation. Now, the model can be constructed with several parameters. To decide the number of parameters, three major criterian are put forth in recent times. These can be used for model selection.

3. Model Selection

3.1. Akaike information criterion

This criteria is used when relative measure of the information has lost while describing the real situation (Akaike, 1974). It can be said to describe the tradeoff between bias and variance in model construction. In other words, it can be used for a trade off between accuracy and complexity of the model. AIC also provides a means for comparison among models for selecting the better structure. However, it can tell nothing about how well a model fits the data in an absolute sense. If all the models are bad, this criteria fails. It is defined as,

$$AIC = 2k - 2\ln(L) \quad (6)$$

here, k = the number of parameters in the model, and L = the maximized value of the likelihood function for the estimated model. The preferred model is the one with the minimum AIC value. AIC not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters. This penalty discourages overfitting. Increasing the number of free parameters in the model improves the goodness of the fit. AICc is AIC with a correction for finite sample sizes,

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \quad (7)$$

where, k denotes the number of model parameters. Thus, AICc is AIC with a greater penalty for extra parameters (Burnham and Anderson, 2002). If n is small or k is large it gives better rsult. AICc converges to AIC as n gets large. AIC increases the probability of selecting models that have too many parameters, i.e., of overfitting when n is not many times larger than k^2. The time series models provides better result when AICc criteria is used (Brockwell and Davis, 2009).

The underlying errors should be normally distributed and independent. This leads to χ^2 model fitting,

$$L = \prod_{i=1}^{n}\left(\frac{1}{2\pi\sigma_i^2}\right)^{1/2} \exp\left(-\sum_{i=1}^{n}\frac{(y_i - f(x))^2}{2\sigma_i^2}\right) \quad (8)$$

It reduces to,

$$\ln(L) = C - \chi^2/2 \quad (9)$$

where, C is a constant independent of the model used, and depends only on the use of particular data points. i.e. it does not change if the data do not change. The AIC is therefore given by,

$$AIC = 2k - 2C + \chi^2 \quad (10)$$

here, C can be ignored since only differences in AIC are meaningful.

3.2. Bayesian information criterion

Bayesian information criterion (BIC) is also used for model selection among a class with different numbers of parameters. To optimize BIC is a form of regularization. When estimating model parameters using maximum likelihood estimation, it is possible to increase the likelihood by adding parameters, which

may result in overfitting. The BIC resolves this problem by introducing a penalty term for the number of parameters in the model. This penalty is larger in the BIC than in the related AIC (McQuarrie and Tsai, 1998).

Let, x = the observed data; n = the number of data points in x, the number of observations, or equivalently, the sample size; k is the number of free parameters to be estimated, $p(x|k)$ is the probability of the observed data given the number of parameters, L is the maximized value of the likelihood function for the estimated model. Then, the BIC can be written as,

$$-2\ln p(x|k) \approx BIC = -2\ln L + k \cdot \ln(n) \qquad (11)$$

here, it is assumed that the model errors or disturbances are independent and identically distributed according to a normal distribution and that the boundary condition that the derivative of the log likelihood in respect to the true variance is zero, this becomes,

$$BIC = 2n \cdot \ln(\sigma^2) + k\ln(n) \qquad (12)$$

where σ^2 is the error variance. The error variance in this case is defined as,

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (13)$$

Under the assumption of normality the following,

$$BIC = \chi^2 + k \cdot \ln(n) \qquad (14)$$

In any estimated model, the model with the lower value of BIC is more accurate. The BIC is an increasing function of σ^2 and an increasing function of k. This implies unexplained variation in the dependent variable and the number of explanatory variables increases the value of BIC. Hence, lower BIC implies either fewer explanatory variables, better fit, or both. The BIC generally penalizes free parameters more strongly than does the Akaike information criterion, though it depends on the size of n and relative magnitude of n and k.

3.3. Deviance information criterion (DIC)

The deviance information criterion (DIC) is a hierarchical modeling generalization of the AIC and BIC. It is particularly useful in model selections where the posterior distributions of the models have been obtained by Markov chain Monte Carlo (MCMC) simulation. It is only valid when the posterior distribution is approximately multivariate normal (Claeskens and Hjort, 2008). Let the deviance is defined as,

$$D(\theta) = -2\log(p(y|\theta)) + C \qquad (15)$$

where y are the data, θ are the unknown parameters of the model and $p(y \mid \theta)$ is the likelihood function. C is a constant which cancels out in all calculations that compare different models. The expectation,

$$\overline{D} = E^{\theta}\left[D(\theta)\right] \qquad (16)$$

is a measure of how well the model fits the data. A larger value means worse fit. The effective number of parameters of the model is computed as,

$$p_D = \overline{D} - D(\theta') \qquad (17)$$

where θ' is the expectation of θ. Its larger value makes it easy to model the data. The deviance information criterion is calculated as,

$$DIC = p_D + \overline{D} \qquad (19)$$

The models with smaller DIC should be preferred to models with larger DIC. Models are penalized both by the value of D, which favors a good fit, but also by the effective number of parameters p_D. Since D will decrease as the number of parameters in a model increases, the p_D term compensates for this effect by favoring models with a smaller number of parameters. The advantage of DIC over other criteria in the case of Bayesian model selection is that the DIC is easily calculated from the samples generated by Monte Carlo simulation.

4. Formulation and Forgetting Factor

To show the effect of forgetting factor and selection of parameter order in metereological parameter prediction, a model is formed to predict temperature profile in a typical summer day (April 21, 2010) at Muaffarpur, India. System identification modelling are based upon identification of important input parameter affecting the output to be predicted. Present model is formulated with two input parameters, humidity and air flow and one single output parameter, temperature profile. The model is trained with daily data collected of the three parameters over one month. These data sets are used ten times with several forgetting factors.

Forgetting factor allows the model to remember its previous data and its effect on its prediction. The model accordingly assigns weightage to the data remembered. The forgetting factor R_2 is also called as variance of the innovations $e(t)$. A typical choice of R_2 is in the range of 0.97 to 0.997, which amounts to approximately remembering 33 to 333 last observations, respectively. In the present case, a forgetting factor of 0..995 was selected. The model, hence, remembered last 200 data in predicting the output.

MATLAB software is used to calculate the predicted value of the temperature profile. The time frame used for prediction was one day, two days and three days. Since the error function reintroduces error in prediction and changes

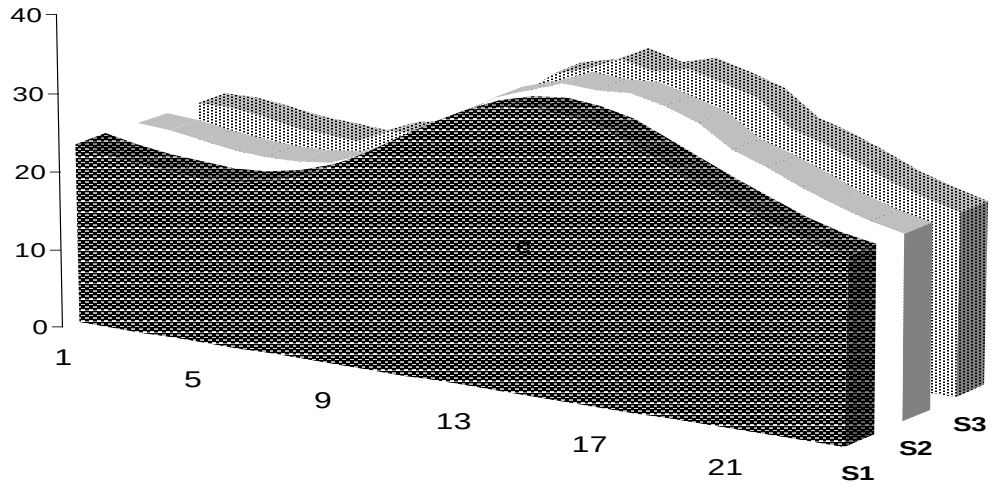

Fig. 1. Variation of temperature - (I) actual (S_1), (II) predicted one day in advance (S_2), (III) predicted two days in advance (S_3).

dynamically the model itself, large error causes instability. Hence, as we move from one day to two days and three days, the model becomes more and more unstable and the error in prediction also increases. In **Figure 1** S_1 shows the actual value of temperature profile. S_2 and S_3 shows predicted values for one day in advance and two days in advance. It can be easily observed from the figure that S_3 has more instability as compared to actual data (S_1) and one day prediction data (S_2).

Result indicates that temperature can be predicted with 90% accuracy for up to two days in advance. It becomes unstable with deviation in forgetting factor. The predicted data also becomes slowly unstable for more than two days. Nevertheless, sudden changes in climatological parameters need to be studied further as these are indicated by large changes in parameters. Such changes can be interpreted in terms of impending dust storm or other harsh climatological changes to issue warning. These predictions need to be quantified with large data accumulation and experience. Mostly climate in arid areas do not change suddenly except in case of storms. Hence, at least five years would be preferred choice even from nearby meteorological station for setting up the initial model. With experience and accumulation of further data, the model will refine itself continuously for better prediction.

References

Akaike H. (1974): A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**(6): 716–723.

Burnham K.P., Anderson D.R. (2002): *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd ed.* Springer-Verlag.

Brockwell P.J., Davis R.A. (2009): *Time Series: Theory and Methods*. 2nd ed. Springer.

McQuarrie A.D.R., Tsai C.L. (1998): *Regression and Time Series Model Selection*. World Scientific.

Claeskens G., Hjort N.L. (2008): *Model Selection and Model Averaging*. Cambridge.