MAUSAM, 51, 4 (October 2000) 319 - 328

551.509.316 : 551.506 (548.1)

On applicability of Frontier Regression to nowcast surface meteorological parameters at Chennai and Trichy airports in Tamilnadu

R. SURESH

Regional Meteorological Centre, Chennai-600 006, India

K. SURESH CHANDRA and N. SIVAGNANAM

University of Madras, Chennai-600 005, India

(Received 4 November 1999, Modified 30 March 2000)

सार — चेन्नई हवाई अड्डे पर विमानों के उड़ान की योजना बनाने के लिए मौसम विज्ञान के प्राचलों के अनुसार तत्काल पूर्वानुमान देने के लिए पूर्वानुमान क्षमता बढ़ाने हेतु एक उन्नत सांख्यिकीय पूर्वानुमान तकनीक अर्थात् अग्रिम समाश्रयण (एफ. आर.) का पता लगाया गया है। चूंकि समुद्री प्रभाव चेन्नई जैसे तटीय हवाई अड्डे के मौसम को अत्याधिक प्रभावित करता है अतः अन्तर्देशीय हवाई अड्डा अर्थात् तिरूचि हवाई अड्डे पर इस अध्ययन में बताए गए निदर्श का प्रयोग इसके परीक्षण के लिए तथा इसकी क्षमता का भी पता लगाने के लिए किया गया है।

ABSTRACT. An advanced statistical forecasting technique, *viz.*, Frontier regression (FR) has been explored to augment the forecasting capacity in nowcasting of meteorological parameters for aviation flight planning at Chennai airport. As maritime effects strongly influence weather over a coastal station like Chennai, the model contemplated in this study has been tried for an inland airport station, *viz.*, Trichy also to assess its efficacy.

Key words — Nowcasting, Statistical techniques, Frontier regression, Autoregression, Deterministic chaos, Persistence, Aviation take-off forecast..

1. Introduction

As the weather and climate affects the day-to-day human activities, there is an evergrowing demand for weather forecasts from various sections of the communities such as aviation and surface transportation agencies, farmers, planners and agencies concerned with outdoor activities like tourism, mountaineering etc. Short Range Forecast (SRF) has a period of validity upto three days. However, when the period of validity is six to twelve hours, the forecast is called Very Short Range Forecast (VSRF). Nowcasting is defined as a detailed description of the current weather and a forecast for few hours, specifically one to three hours [National Technical Information Service (NTIS), 1981]. In this context, it may be mentioned that aircraft 'take-off' forecast and 'landing' forecast fall under the category of nowcasting. Aircraft's take-off gross weight (TOGW) depends on surface temperature, pressure altitude (in turn surface pressure), surface wind at the time of take-off [India Meteorological Department (IMD), 1974]. An error beyond the permissible limit of forecast accuracy causes not only monetary losses to the airlines agencies but also has impact on passengers and cargo shipment indirectly. Hence the take-off forecast assumes socio-economic implications. International Civil Aviation Organization (ICAO) has set a target on operationally desirable accuracy of take-off forecast, *viz.*, in atleast 90% cases, the forecast absolute error (A.E) should be within 1°C, 1hPa, 30° in respect of surface temperature, pressure and wind direction respectively (ICAO, 1995).

The classical (statistical) methods are viable and useful for nowcasting since the numerical weather prediction (NWP) model output is not available with sufficient accuracy or promptness (Doswell, 1986; Wilks, 1995). For every short duration say 1 to 3 hrs. (IST), objective (statistical) prediction techniques using classical methods may work better than subjective (synoptic) methods (WMO, 1969; WMO, 1992). Till such time an efficient, quantitative, location specific LAM is operationally feasible with minimum computing requirement at field stations, it is desirable to develop classical methods to meet the operationally desirable forecast accuracy target set by ICAO.

Meteorological statistical forecasting generally requires the application of multivariate analyses with more

TABLE 1

Comparative performance of forecast of three hourly surface temperature within an absolute error 1°C and pressure within an absolute error 1 hPa at Chennai and Trichy airport, 1988

Season	Temperature				Pressure			
	Pers	AR	Win	MR	Pers	AR	Win	MR
				(a) Chennai airport				
Winter	78.2	79.8	65.2	79.8	70.3	87.2	81.9	85.9
Hot-weather	70.0	70.2	61.0	70.7	61.8	80.3	78.9	78.9
Monsoon	57.1	57.1	54.5	56.5	59.1	80.2	79.7	81.1
Post-monsoon	70.3	63.3	52.1	66.7	62.2	77.9	73.4	77.9
				(b) Trichy airport				
Winter	77.1	80.6	64.3	79.8	74.8	87.4	86.9	86.9
Hot-weather	57.7	59.6	56.4	56.4	65.7	79.6	74.5	78.3
Monsoon	60.8	66.7	55.4	64.3	64.4	83.7	83.7	81.1
Post-monsoon	74.8	76.9	62.3	76.9	68.8	81.1	76.9	80.0

Note: Pers - Persistency ; AR - Auto regression
Win - Winters' exponential smoothing; MR-Multiple Regression
Winter= January-February; Hot-weather = March-May;
Monsoon= June-September; Post-monsoon = October-December

TABLE 2

Results of dimensionality analyses of standardised and seasonal differenced temperature and pressure series, 1984-88

Season	Time-shift	Embedding dimension considered	Dimension at which saturation occurred	Fractal dimension	No. of variables required for modelling	
					Minimum	Maximum
			Chennai airport - Temperature			
Winter	5	37	35	4.492	5	35
Hot-weather	5	68	66	3.836	4	66
Monsoon	7	38	36	7.230	8	36
Post-monsoon	5	40	38	5.934	6	38
			Chennai airport - Pressure			
Winter	8	48	46	1.719	2	46
Hot-weather	8	59	57	3.543	4	57
Monsoon	8	44	42	3.980	4	42
Post-monsoon	8	56	52	2.590	3	52
			Trichy airport- Temperature			
Winter	4	46	44	6.229	7	44
Hot-weather	7	66	64	6.448	7	64
Monsoon	7	44	40	5.724	6	40
Post-monsoon	5	46	43	7.654	8	43
			Trichy airport - Pressure			
Winter	8	49	45	2.233	3	45
Hot-weather	8	74	71	3.878	4	71
Monsoon	9	59	54	3.843	4	54
Post-monsoon	8	70	64	4.095	5	64

Note: Total number of data points used to estimate dimensionality are:
Winter-2336; Hot-weather & Post-monsoon-3640; Monsoon-4640.

number of variables. As such the application of multivariate analysis requires enormous computing power to handle large data matrices. Of course with the technological improvement in computing speed and data handling capacity from early 1960s, statistical applications to meteorology grew rapidly, classical forecasting methods may be useful for the first 6 to 12 hrs (IST) of forecast, as meteorological variables such as temperature, pressure etc. exhibit strong persistency (WMO, 1992) with short time span. An attempt has been made in this paper to explore the applicability of univariate, multivariate regression techniques to nowcast surface temperature and pressure at Chennai and Trichy airports in Tamilnadu.

2. Data used

The 3 hourly surface observations in respect of Chennai, Trichy, Madurai, Coimbatore, Thiruvananthapuram, Cuddalore, Tondi, Nagapattinam (locations shown in Fig. 1) for the period 1984 to 1988 have been obtained from the National Data Centre (NDC), India Meteorological Department (IMD), Pune. The surface data of Chennai and Trichy for the year 1995 have also been used for validating the models developed based on 1984-87 data.

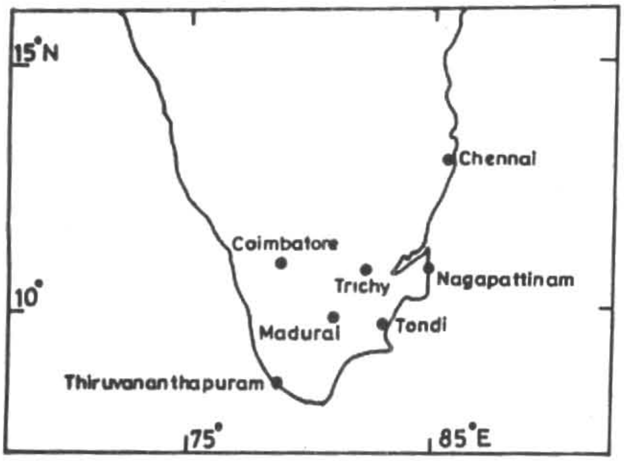


Fig. 1. Map showing locations of stations whose data have been considered in the study.

3. Method of analysis

3.1. *Univariate analysis*

As the meteorological variables exhibit high degree of persistency, univariate statistical schemes like correlogram analyses, auto regressive / moving average processes and exponential smoothing and filters have been attempted to nowcast surface temperature and pressure. Winters' three parameters linear smoothing (Winters, 1960) which accounts for trend, seasonality and randomness in the series have also been attempted. Analyses reveal very little improvement over the (operationally viable) method of climatology and persistency. ICAO's operationally desirable forecast accuracy could not be met through these methods. (Table 1).

3.2. *Multivariate analysis*

Linear multivariate regression analysis has also been carried out after carefully selecting the predictors based on the physical and statistical relationships they have with the predictands. For this purpose, data of the nearby stations have been considered and the method of forward screening has been employed. The results obtained through multiple regression (MR) has been tabulated in Table 1. However, in this case also, the desirable efficiency could not be achieved.

3.3. *Method of deterministic chaos*

As the efforts put on the univariate and multivariate analyses have not been commensurate with the benefits of such efforts, this complex dynamical system have been theorised into the theory of deterministic chaos. Dimensionality analysis have been conducted based on Grassberger and Procaccia (1983) algorithm to estimate the dimensions of attractors. In order to avoid spurious results, Ruelle's (1990) inequality on the minimum number of data points have also been taken care of. To ensure independency of the data points for the computation of correlation integral, the three hourly data have been standardised, seasonal differenced (length of seasonality 8). The results (Table 2) indicate that the surface temperature so also pressure system have low order fractal dimensions suggesting feasibility of modelling and the slow convergence suggests a large number of variables neded for modelling. Variability of minimum and maximum number of parameters needed for modelling between a coastal and an inland station have also been brought out from this computation explaining mesoscale variability of weather parameters. As the time tested conventional univariate and multivariate analyses did not meet our requirements, the recent regression model, *viz.*, frontier regression has been explored.

3.4. *Frontier regression (FR)*

3.4.1. *Background of stochastic frontier functions*

The classical regression equation provides the expected value of the dependent variable (Y) for given values of the explanatory variables [$X = (x_1, x_2, \ldots, x_k)$], included in the equation. Frontier regressions are basically meant for estimating the maximum or minimum value of Y for given X. The literature on frontier analysis, generally indicates two types of approaches for identifying extremal value of Y for given X. One, the non-parametric approach, wherein no distributional assumptions are made in the identification of a frontier on which the extremal is located. Data envelopment analysis is one such method. The second is a parametric approach, wherein the equation to be fitted is assumed to have a negatively skewed error term. To be specific, the general model for frontier regression, under this approach is given by

$$Y_i = f(x_{1i}, x_{2i}, x_{3i} \ldots, x_{ki}) + \eta_i \quad (1)$$

where Y_i is the predictand at i^{th} time period (i=1,2,3.., N) and η_i has two components, one being a symmetric random variable and the other a non-negative random variable. In practical situations, one assumes

$$\eta_i = v_i - u_i \quad (2)$$

where $\{v_i\}$ are independent and identically distributed (i, i, d) random variables with mean zero and variance σv^2 and $\{u\}$ are i, i, d (independent of v's) half normal with common density

$$g(u, \sigma_u) = \frac{2}{\sqrt{2\pi}\sigma_u} \exp(-u^2/2\sigma_u^2) \quad (3)$$

so that η is negatively skewed. In Economics analysis the interest is in the measurement of the departure of Y from its maximum of [$f(x_{1i}, , \ldots, x_{ki})$] in order to provide an index of inefficiency of the i^{th} unit. In fact the index of efficiency of the i^{th} unit is given by

$$TE\,(i) = Y_i / f\,(x_{1i}, \ldots, x_{ki}) \qquad (4)$$

3.4.2. *Linear stochastic frontier equation*

The stochastic frontier function has been proposed for cross- sectional (multilevel) data. In the linear specification of f in (1), one has

$$y_{it} = \beta_0 + x_{it}\beta + (v_{it} - u_{it});\ (i = 1,2 \ldots N);\ (t = 1,2, \ldots, T) \qquad (5)$$

where y_{it} is the i^{th} value of dependent variable at t^{th} time period. x_{it} is $(k \times 1)$ vector of predictors at t^{th} time period.

β is a vector of unknown parameters.

v_{it} are random variables which are assumed to be i, *i,i,d*

$N\,(0, \sigma_v^2)$and independent of the u_{it} which are non-negative random variables an assumed to account for inefficiency in the i^{th} dependent value at time t and $u_{it} = [u_i \exp\{-\eta\,(t-T)\}]$ and η is to be estimated.

The problem of estimation of parameters of the model in (5) has been excellently described in Coelli (1993) and Greene (1993), Coelli (1996) provides a computer programme, which can be freely downloaded from Internet, that would provide maximum likelihood estimates (MLE) of the parameters of (5), under many other options, such as allowing truncated normal/half-normal distribution for u_i. In order to calculate the MLEs, parameterization of σ_u^2 and σ_v^2 has been attempted by $\sigma^2 = \sigma_u^2 + \sigma_v^2$ and $\nu = \sigma_u^2 / (\sigma_u^2 + \sigma_v^2)$. The parameter ν lies between 0 and 1 and thus this closed interval [0,1] can be searched to provide a good starting value for use in an interactive maximization process using Davidon-Fletcher-Powell Quasi - Newton algorithm (Battesse and Coelli, 1992). The computer program gives a three step procedure in estimating the unknown β parameters of the stochastic frontier functions, *viz.*, (*i*) ordinary least square (OLS) estimates, (*ii*) a two -phase grid search of β is conducted with parameters (except β_0) set to OLS values and the β_0 & σ^2 parameters adjusted according to the corrected OLS formula. Grid search is conducted across the parameter space of ν. Values of β are considered from 0.1 to 0.9 in steps of 0.05. The grid search aims at extreme value of predict and for each set of predictors at a given point of time. (*iii*) The value selected through grid search are used as *starting* values in the Quasi-Newton iterative procedure to obtain final MLEs. Detailed procedure for computing the MLEs have been described in Himmelblau (1972). The computer programme provides the covariance matrix of explanatory variables, final MLEs of β's, and technical efficiencies for each unit besides significance of regression coefficients.

3.4.3. *Frontier regression in forecasting weather parameters*

It is important to note that frontier regression are primarily meant for identifying extremal values of Y for a given value of $(x_1, x_2, \ldots x_k)$. Since the classical regression analysis did not provide forecasts with desired accuracy, our interest was to try other estimating devices, which can perform better than the classical regression analysis. It is that curiosity that set us on to FR model. The three hourly synoptic data has been transfered as cross-sectional data proposed in FR model by taking t=0000, 0300, 0600, ...2100 UTC and i=1,2,3,.....N (number of days in each season). One can estimate the expected value of Y given X, using

$$E(Y_i : X = x_i)\ f(x_i) - E\,(u_i : X = x_i) \qquad (6)$$

since $E\,(v_i : x)$ is assumed to be zero. Under the half normal assumption on u, it can be verified that

$$E\,(u_i : X = x_i) = E\,(u_i) = \sigma_u \left(\sqrt{\frac{2}{\pi}}\right) \qquad (7)$$

Using the maximum likelihood estimates provided by the software, we propose the predicted value of Y for given X, as given by (7) on substituting the MLEs of the parameters. It may be noted that in the classical regression analysis $E\,(Y : X = x_i)$ is linear in Y, while estimator through (7) is non-linear in Y (since σ_u is involved on its right hand side). Further with the inclusion of σ_u in the predicted value, the degree of skewness in the data has also been taken into account.

4. Selection of predictors

4.1. *Identification*

Based on the physical relationships between the temperature (T) and pressure (P), which in turn are related with the other meteorological variables like water vapour (moisture), wind etc through equation of state, equation of continuity and conservation of mass etc, the following potential predictors have been identified. Temperature related predictors have been described here explicitly and similar definitions will hold good for pressure related predictors also, simply by interchanging 'T' by 'P'. It may be remembered that each data point is separated by a time period of 3 hours (IST) (*i.e*). Lag 1 refers to the data observed three hours before, Lag 8 means the data observed 24 hours (IST) before.

(*i*) As the weather parameters normally have persistence, the corresponding value of temperature at the same time period (t), say 0000,0300,0600,....2100 UTC of the previous day (*i.e.* with a lag of 8 periods) has been identified as one of the potential predictors and it is defined here as TLAG8 which can be expressed as $T(t$-8).

TABLE 3

Performance of frontier regression in forecasting surface temperature and pressure within an absolute error of 1°C and 1hPa at Chennai airport, 1988

Time (UTC)	Winter			Hot weather			Monsoon			Post-monsoon		
	S. D.	Eff	RMSE	S. D.	Eff	RMSE	S. D.	Eff	RMSE	S. D.	Eff	RMSE
(a) Surface temperature												
0000	1.51	70.7	2.14	2.17	72.2	1.19	1.66	62.5	1.66	1.77	66.7	1.11
0300	1.39	82.8	1.38	2.20	71.1	1.16	1.92	62.5	1.32	1.89	58.9	1.17
0600	1.51	70.7	1.28	2.51	72.2	1.78	2.35	52.5	1.85	2.18	61.1	1.55
0900	1.51	81.0	0.93	2.78	55.6	2.20	2.70	50.0	2.12	2.39	63.3	1.70
1200	1.07	94.8	0.50	2.09	72.2	1.46	2.48	43.3	2.48	1.78	75.6	1.16
1500	1.06	93.1	0.55	1.83	81.1	0.90	1.79	68.3	1.24	1.73	66.7	1.10
1800	1.38	86.2	0.73	2.02	87.8	0.69	1.74	68.3	1.19	1.75	81.1	1.04
2100	1.38	70.7	1.01	2.13	67.8	1.28	1.74	65.0	1.37	1.76	71.1	1.10
(b) Surface pressure												
0000	2.32	70.7	0.99	3.23	56.7	1.30	2.46	49.2	1.30	2.99	63.3	1.14
0300	2.26	94.8	0.46	3.09	83.3	1.10	2.41	87.5	0.67	2.64	93.3	0.62
0600	2.19	94.8	0.55	3.37	94.4	0.63	1.76	87.5	0.63	2.92	94.4	0.70
0900	2.31	98.3	0.44	3.39	92.2	0.56	2.19	86.7	0.71	2.92	94.4	0.66
1200	2.48	100.0	0.45	3.24	85.6	0.65	2.39	83.3	0.79	2.84	96.7	0.61
1500	2.41	98.3	0.44	3.45	78.9	0.96	2.13	85.8	0.71	2.67	94.4	1.01
1800	2.23	91.4	0.55	3.32	88.9	0.68	1.96	90.8	0.62	2.50	93.3	1.02
2100	2.24	69.0	0.92	3.10	67.8	1.06	2.06	60.0	1.25	2.58	68.9	0.61

S.D. - Standard deviation , Eff - Efficiency, RMSE- Root mean squared error

TABLE 4

Performance of frontier regression in forecasting surface temperature and pressure within an absolute error of 1°C and 1hPa at Trichy airport, 1988

Time (UTC)	Winter			Hot weather			Monsoon			Post-monsoon		
	S.D.	Eff	RMSE	S.D.	Eff	RMSE	S.D.	Eff	RMSE	S.D.	Eff	RMSE
(a) Surface temperature												
0000	1.45	77.6	0.93	1.95	61.1	1.35	1.21	70.8	0.97	1.78	68.9	1.11
0300	1.17	81.0	0.75	1.95	74.4	0.83	1.44	74.2	0.84	1.67	76.7	1.10
0600	1.47	72.4	0.86	2.29	60.0	1.70	1.95	63.3	1.23	1.94	65.6	1.35
0900	2.10	84.5	0.71	2.83	63.3	1.91	2.23	59.2	1.18	2.39	70.0	1.53
1200	2.34	79.3	0.82	2.92	50.0	2.15	2.70	57.5	2.34	2.33	64.4	1.58
1500	1.55	79.3	0.77	2.83	65.6	1.92	2.20	55.8	2.19	1.82	78.9	1.13
1800	1.28	87.3	0.69	2.00	71.1	1.16	1.77	75.8	1.30	1.75	88.9	0.84
2100	1.27	72.4	0.87	2.05	63.3	1.14	1.44	71.7	1.22	1.77	74.4	1.08
(b) Surface pressure												
0000	2.25	70.7	0.90	2.86	66.7	0.99	2.07	64.2	1.15	2.41	72.2	1.02
0300	2.35	93.1	1.11	2.81	85.6	0.89	1.76	85.8	0.72	2.42	96.7	0.50
0600	2.11	96.5	0.44	2.90	87.8	0.78	1.65	89.2	1.15	2.44	93.3	0.58
0900	2.05	91.4	0.54	2.86	92.2	0.56	1.28	90.8	0.74	2.39	96.7	0.57
1200	2.28	93.1	0.56	2.79	80.0	0.84	1.34	87.5	0.87	2.59	96.7	0.52
1500	2.37	93.1	0.70	2.56	84.4	0.69	1.46	82.5	0.85	2.40	93.3	0.59
1800	2.05	93.1	0.61	2.75	90.0	0.61	1.69	88.3	0.67	2.33	96.7	0.63
2100	2.12	74.1	1.22	2.66	71.1	1.03	1.28	60.8	1.18	2.68	77.7	1.10

S. D. - Standard deviation, Eff - Efficiency, RMSE- Root mean squared error

(*ii*) In order to study as to how the difference in the temperature of the previous time period from its corresponding value of a day before influences the change in temperature value for the time period being studied, 'TLAG1D' has been constructed. Mathematically, TLAG1D = $T(t\text{-}1) - T(t\text{-}9)$.

(*iii*) To account for the effects of relative humidity (RH) and vapour pressure (VP) at a previous time period (just three hours before) over the temepratute at any specific time, the RH, VP parameters have been considered. RH and VP are defined as RH = RH (t-1); VP = VP (t-1).

(*iv*) The surface wind is resolved into zonal component (denoted usually by 'U') and meridional component (denoted by ('V'). As the warm/cold air advection will modify the weather parameters, the effects of U and V components of wind at a previous time period are also considered as ULAG1 and VLAG1 which are defined as ULAG1 = $U(t\text{-}1)$; VLAG1=$V(t\text{-}1)$.

(*v*) As the weather parameters of neighbouring stations may affect weather over a place, correlation matrices have been constructed for temperature and pressure of Chennai with the related weather parameters of Trichy, Tondi, Nagapattinam, Madurai and Trivandrum (which will be discussed in section 4.2). However, the signal was found good only from Trichy, presumably due to its nearness and favourable atmospheric flow pattern in most of the seasons. The teleconnections between the weather parameters of Chennai and Trichy are also considered and the variables have been described almost in the same style defined above but by suffixing 'TR' to denote these parameters pertaining to Trichy, (*i.e*)

$$\text{LAG1DTR} = T(t\text{-}1) - T(t\text{-}9) \text{ of Trichy}$$

TABLE 5

Comparative performance of frontier regression with the method of persistent in forecasting surface temperature and pressure within an absolute error of 1°C and 1hPa at Chennai airport, 1995

Time (UTC)	Winter		Hot weather		Monsoon		Post-monsoon	
	Pers	FR Eff	Pers	FR Eff	Pers	FR Eff	Pers	FR Eff
(a) Surface temperature								
0000	67.2	67.8	86.7	85.6	61.5	56.7	75.0	68.9
0300	72.4	85.7	78.3	75.6	56.6	61.7	60.9	55.6
0600	70.7	68.4	67.4	63.3	63.9	74.2	67.4	65.6
0900	75.9	76.7	66.3	60.0	54.1	57.5	64.1	65.6
1200	84.5	87.5	79.4	78.9	50.8	53.3	64.1	75.6
1500	75.9	91.1	90.2	92.2	53.3	54.2	63.0	70.0
1800	65.5	78.9	82.6	86.7	47.5	54.2	66.3	82.2
2100	70.7	63.2	80.4	76.7	51.6	54.2	72.8	74.4
(b) Surface pressure								
0000	87.9	77.2	64.1	64.4	61.5	60.8	68.5	68.9
0300	82.8	98.2	67.4	97.8	66.4	90.8	71.7	93.3
0600	86.2	100.0	60.9	93.3	65.6	93.3	67.4	97.8
0900	86.2	96.5	53.3	92.2	64.8	93.3	69.6	92.2
1200	79.3	96.5	59.8	94.4	66.4	85.8	70.7	93.3
1500	72.4	77.2	59.8	91.1	68.9	84.2	70.7	94.4
1800	81.0	96.1	55.4	86.7	66.4	81.2	72.8	92.2
2100	79.3	78.9	63.0	62.2	65.6	65.8	63.0	64.4

Pers - Persistency ; FR Eff- Frontier Regression efficiency

TABLE 6

Comparative performance of frontier regression with the method of persistent in forecasting surface temperature and pressure within an absolute error of 1°C and 1hPa at Trichy airport, 1995

Time (UTC)	Winter		Hot weather		Monsoon		Post-monsoon	
	Pers	FR Eff	Pers	FR Eff	Pers	FR Eff	Pers	FR Eff
(a) Surface temperature								
0000	65.5	67.8	64.1	62.2	66.4	82.5	64.1	61.1
0300	67.2	66.1	65.2	68.9	77.1	81.7	67.4	67.8
0600	70.7	71.4	65.2	67.8	63.9	72.5	66.3	57.8
0900	67.2	67.8	55.4	66.7	54.9	70.0	64.1	64.4
1200	67.2	75.0	57.6	51.1	44.3	60.0	67.4	63.3
1500	70.7	64.3	53.3	53.3	41.8	65.8	63.0	65.6
1800	60.3	71.4	62.0	76.7	47.5	96.7	56.5	74.4
2100	69.0	66.1	70.7	68.9	59.0	75.0	75.0	71.1
(b) Surface pressure								
0000	91.4	86.0	77.2	70.0	64.8	65.0	79.4	74.4
0300	86.3	100.0	75.0	87.8	69.7	90.0	75.0	98.9
0600	86.2	94.7	70.7	96.7	72.1	90.0	75.0	91.1
0900	84.5	100.0	66.3	94.4	70.5	90.0	78.3	88.9
1200	75.9	98.3	58.7	85.6	65.6	92.5	71.7	93.3
1500	81.0	91.2	64.1	86.7	63.9	90.8	76.1	86.7
1800	77.6	94.7	64.1	86.7	73.0	89.2	75.0	94.4
2100	87.9	82.5	71.7	68.9	71.3	66.7	70.7	70.0

Pers - Persistency ; FR Eff - Frontier regression efficiency

UTR = $U(t\text{-}1)$; VTR = $V(t\text{-}1)$ of Trichy

RHTR = RH $(t\text{-}1)$; VPTR = VP$(t\text{-}1)$ of Trichy.

(*vi*) The departure of temperature from its corresponding value at the same time period a day before has also been considered. It is denoted as TANOM and defined by

TANOM = $T(t)\text{-}T(t\text{-}8)$. However this parameter has been used as predictand as will be explained in section 4.2.

Few other predictors like previous 3/6/12/24 hrs (IST) trend of the same station as well as that other stations have also been considered. But the results are not discussed here since their efficacies are either below and/or on par with that obtained through the predictors described above.

4.2. *Selection*

Matrices of correlation coefficients (CC) between temperature (so also for pressure) of each sub-period and their potential predictors for all seasons in respect of Chennai and Trichy have been constructed. Those predictors which have physical and statistical relationship (taking care of multicollinearity) with temperature and pressure have been chosen for each sub-period and for each season for modelling. In order to lessen the degree of multi-collinearity, the predictand is chosen as TANOM rather than $T(t)$. Under this arrangement, we achieve not only parsimony in regression variables but also explore a new horizon to look at the predictand as deviation from its corresponding value of the previous day and not the predictand and specific time. Once TANOM [i.e. $T(t)$- $T(t\text{-}8)$]is predicted, $T(t)$ can be easily calculated by simply adding the known value $T(t\text{-}8)$ which is the temperature of the previous day at the forecast hour. For predicting weather parameters of Chennai, all predictors listed in para 4.1 have been considered. In a similar way, for predicting temperature and pressure at Trichy, in addition to predictors based on past data of its own, lagged variables of temperature, pressure, RH, VP, *U* and *V* components of Chennai have been considered. As such, we have developed 8 model equations of varying number of predictors separately for temperature and pressure for each season corresponding to each sub-period (0000,0300, . . . , 2100 UTC) for each station.

5. Results of FR models

The computer program developed by Coelli (1996) has been used for computations. The data pertaining to the period 1984-87 have been used for developing model coefficients and that of 1988 have been used for verification. The model coefficients are obtained through OLS, GRID search and MLE methods. The MLEs are estimated through iterative procedures taking the values obtained through GRID search as their initial (starting) values. The regression coefficients are tested for their significance through '*t*' test. As stated already, since we want to forecast the weather parameter as close as possible to the actual and are not interested in maximizing the weather parameter, the estimated value of *u*, *viz.*, $\sqrt{\frac{2}{\pi}}\,\sigma_u$ has been deducted from the output obtained (to be specific, from the constant term of the regression equation) through MLE method. The FR models estimate the regression coefficients through entirely new concepts on estimation theory which can not be ordinarily obtained by the method of OLS. Nevertheless, the root mean squared errors (RMSE) of MLE methods are lower than that of OLS and GRID methods. It is this feature that attracted us to propose a new estimator based on FR model.

The model efficiency in predicting temperature and pressure during 1988 at Chennai and Trichy within an absolute error of 1°C and 1 hPa respectively, RMSE for all subperiods of all seasons are furnished in Tables 3 and 4 respectively. The model RMSE is very much comparable with that of the standard deviation (SD) indicating the suitability of the model. Nowcasting surface pressure at Chennai and Trichy through FR models has met the ICAO standard of take-off forecast in many sub-periods. However in forecasting surface temperature, FR model efficiencies are lagging behind the desirable accuracy standards in many sub-periods. It has also been seen that the efficiency of the FR model in predicting temperature and pressure within the prescribed error limit is better than the method of persistency, Winters' exponential smoothing and AR processes. However the desired standards of accuracies set by ICAO have not been fully achieved even through this model for different time periods of various seasons. Performance of the model in comparison with the method of persistency for the year 1995 has been summarised in Tables 5 and 6 and pictorially depicted in Fig. 2. The model has performed well in many sub-periods of different seasons in comparison to the method of persistency which is normally used by operational meteorologists.

TABLE 7

Empirical temperature correction factors for low clouds

Clouds oktas 3 hours before	Correction factor (°C)
0	0.0
≤ 2	0.1
≤ 4	0.2
≥ 5	0.3
≥ 7	0.4

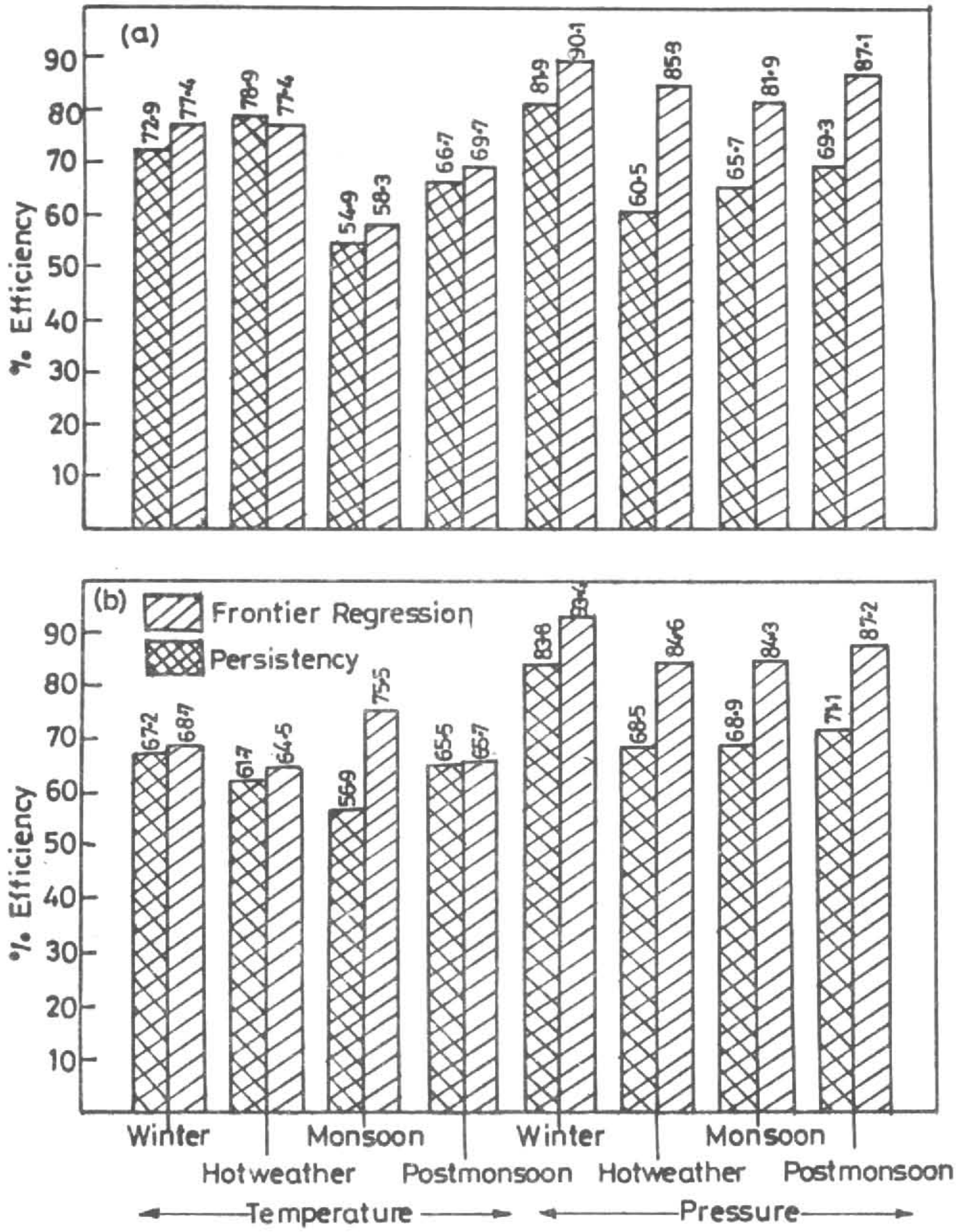
5.1. *Incorporating stability criteria in FR models*

Pasquill's stability classes as modified by Turner (1964) have been worked out with a view to analyse the cause (s) for the low efficiencies in predicting temperature between 0600 and 1200 UTC from March to December. It has been observed by Suresh (1998) that the between month unstable category frequencies have large variation during morning hours of post monsoon season due to varying nature of early morning precipitation and dissipation of clouds. Moreover, the time of onset of sea breeze over Chennai during hot weather and monsoon seasons is quite uncertain and it has the modifying effect of temperature between 2 and 4°C (Atkinson, 1981; Bhaskar Rao *et al.* 1984) which causes inaccuracy in prediction of temperature at 0600 and 0900 UTC. The computed stability classes, after being assigned with some empirical values, were also used as predictors to ascertain the improvement in forecast. But this approach did not yield satisfactory results.

5.2. *Inclusion of cloud amounts and heights*

As clouds scatter the visible (VIS) radiation and absorbs the infrared (IR) radiation, the day and night temperature are considerably modified by the cloudy weather conditions. The scattering of solar radiation (of the form VIS) by the clouds modify the albedo (which is defined as the ratio of reflected solar radiation to the incoming solar radiation). The absorption of terrestrial radiation (of the form IR) by clouds enhances the greenhouse effect. It is assumed that the presence of low clouds at time *t* will modify the temperature at time (*t*+3) where *t*=0000,0300, 0600,...,2100 UTC. The modification due to the presence of clouds, specifically low clouds, is positive (negative) during night (day) where of radiation from earth surface is of the form IR. Such corrections factors (based on the data of 1984 to 1987) for low clouds of varying base heights have been worked out empirically. The FR forecast for the next sub- period was recast for hot weather season over Chennai and Trichy. The correction factor was subtracted from the FR model outputs for the forecast period 0600,0900 and 1200 UTC and added in rest cases. The correction factors for low clouds whose base heights are less than or equal to 999 m above surface have been given in Table 7.

The FR temperature forecast during hot weather season is furnished in Table 8. Though an improvement in efficiencies was of the order, a maximum of 4.4% in respect of Chennai and 7.8% in respect of Trichy, reduction in effi-



Figs. 2 (a&b). Forecasting surface temperature and pressure at (a) Chennai and (b) Trichy airport during 1985

ciency to the same extent, as that of improvement due to incorporation of correction factors, also could be seen in some cases. Hence, it has been decided not to take this empirical correction factor into consideration as the inclusion of this parameter in this statistical model does not improve the efficiency.

6. Summary

The FR models are normally designed for estimating the maximum output for a given set of input and the regression coefficients are interatively estimated through MLE approach by following certain assumptions on the 'error term' of the ordinary regression model. However, in this paper a new estimate for the constant term of the regression equation has been proposed by subtracting estimated value of one of the error terms considered in the model from the value of the constant term of the model. The salient features of FR model in forecasting surface temperature and pressure at Chennai and Trichy airports are summarised below.

(*i*) The efficacies of the FR models are far better than that could be obtained from the method of persistency, auto regression and Winters' three parameter exponential smoothing, except for a very few sub-periods.

(*ii*) As the likelihood ratio test and '*t*' test of significance of the regression coefficients indicate the suitability

TABLE 8

Efficacy of frontier regression in forecasting surface temperature within + 1°C error at Chennai and Trichy airport during hot weather season, 1988 by incorporating correction for low clouds

Time (UTC)	Frontier regression model efficiency			Efficiency after incorporating corrections for low clouds		
	OLS	GRID	MLE	OLS	GRID	MLE
			Chennai airport			
0000	72.2	63.3	72.2	70.0	57.8	71.1
0300	71.1	62.2	71.1	72.2	64.4	71.1
0600	70.0	55.6	72.2	71.1	60.0	73.3
0900	57.8	33.3	55.6	55.6	36.7	55.6
1200	73.3	62.2	72.2	73.3	64.4	71.1
1500	81.1	70.0	81.1	80.0	73.3	82.2
1800	85.6	85.6	87.8	87.8	87.8	87.8
2100	67.8	62.2	67.8	70.0	60.0	68.9
			Trichy airport			
0000	61.1	46.7	61.1	62.2	44.4	62.2
0300	74.4	70.0	74.4	77.8	66.7	76.7
0600	60.0	56.7	60.0	61.1	55.6	61.1
0900	64.4	40.0	63.3	62.2	46.7	66.7
1200	51.1	26.7	50.0	50.0	33.3	48.9
1500	65.6	47.8	65.6	66.7	48.9	66.7
1800	67.8	71.1	60.0	70.0	70.0	67.8
2100	63.3	57.8	63.3	66.7	57.8	67.8

OLS - Ordinary least square method : GRID - Grid search method ; MLE - Maximum likelihood estimator method.

of the model coefficients, it is plausible to select the best model for forecasting the weather parameters.

(*iii*) Incorporation of stability classes and cloud characteristics in the FR model did not improve the forecast efficiency significantly.

(*iv*) The efficiency in forecasting surface pressure through this model is almost close to the target set by ICAO. However in respect of the surface temperature forecast, the efficiency varies from 43.3 to 94.8% (51.1 to 88.9%) during different time period of various seasons at Chennai (Trichy).

7. Conclusions, limitations and future work

The models proposed in this study are simple, parsimonious and operationally feasible. The models for predicting surface pressure have almost met the target set by ICAO on operationally desirable accuracy of forecasts. However for nowcasting surface temperature, desired efficiency is yet to be achieved even through FR. As such, till such time operationally viable NWP models are developed to nowcast surface parameters, models contemplated in this study can be used. Further evaluation of the models may given greater satisfaction for their operational utility. Hence, real time evaluation of the models proposed in this study may have to be done with utmost care in view of the aircract safety implications. None of the models attempted could forecast the surface wind direction within the permissible error limits prescribed by ICAO albeit the wind speech could be forecast by few models within the error limits. Further analysis on vector wind is being attempted.

Acknowledgements

The authors thank the Additional Director General of Meteorology, IMD, Pune for sparing the data and the Deputy Director General of Meteorology, Regional Meteorological Centre, Chennai for extending all facilities to conduct this study. Referee's constructive comments are appreciated with thanks. The authors acknowledge the services rendered by S/Shri M. Bharathiar and N. Selvam in tracing the figures.

References

Atkinson, B.W., 1981, "Meso-scale atmospheric circulations", Queen Mary College, University of London, 495p.

Battesse, G.E., and Coelli, T.J., 1992, "Frontier Production Functions, Technical Efficiency and Panel Data: With Application of Paddy Farmers in India", *Journal of Productivity Analysis*, **3**, 153-169.

Bhaskar Rao, N.S., William, S.D., Mary Chandy and Usha Devi, 1984, "Study of sea breeze at Madras", *Mausam*, **35**, 4, 537-538.

Coelli, T.J., 1993, "Finite sample properties of stochastic frontier estimators and associated test statistics", Working papers in econometrics, University of New England, Armidale.

Coelli, T.J., 1996, "A guide to FRONTIER Version 4.1: A computer Program for Stochastic Frontier Production and Cost Function Estimation", CEPA Working Paper 96/07, University of New England, 32p.

Doswell, C.A., 1986, "Short-Range Forecasting", in : Mesoscale Meteorology and Forecasting, ed. by Peter S. Ray, *Amer. Met. Soc., Boston*, 793p.

Grasberger, P., and Procaccia, I., 1983, "Characterization of strange attractors", *Phy. Rev. Lett.*, **50**, 5, 346-349.

Greene, W.H., 1993, "The econometric approach to efficiency analysis", in: The measurement of productivity efficiency, ed. H.O. Fried, C.A.K. Lovell and S.S. Schmidt, Oxford University Press, New York, 68-119.

Himmelblau, D.M., 1972, "Applied non-linear programming", McGraw Hill, New York.

India Meteorological Department, 1974, "Forecasting Manual - Part V - Aviation meteorology", prepared by N.S. Bhaskar Rao and V. Surya Rao, IMD, Pune, India, 130p.

International Civil Aviation Organization, 1995, "Meteorological service for international air navigation", Annex 3, 12th ed., ICAO, Montreal, Canada, 89p.

National Technical Information Service, 1981, "Nowcasting. Mesoscale observations and short-range prediction", Proc. Internat. IAMAP symp., 25-28 August 1981, Humburg, Germany, NTIS, US Dept. of Commerce, 415p.

Ruelle, D., 1990, "Deterministic chaos: the science and the fiction" (The Claude Bernard lecture), *Proc. R. Soc. London (Math. Phys. Sci.)*, **A427**,, 241-248.

Suresh, R., 1998, "On some aspects of boundary layer over Madras airport", *Mausam* **49**, 4, 528-530.

Turner, D.B., 1964, "A diffusion model for an urban area", *J. Appl. Met.*, **3**, 83-91.

Wilks, D.S., 1995, "Statistical methods in the atmospheric sciences", Academic press Inc., San Diego, California 467p.

Winters, P.R., 1960, "Forecasting sales by Exponentially weighted Moving Averages", *Management Science*, **6**, 324-342.

World Meteorological Organization, 1969, "Aeronautical meteorology", Proc of WMO scientific and technical conf. on aeronautical meteorology, London March 1968, WMO technical note No. 95, WMO No. 227. TP. 121, WMO, Geneva, Switzerland, 493p.

World Meteorological Organization, 1992, "Methods of interpretating numerical weather prediction output for aeronautical meteorology", WMO No. 770, WMO, Geneva, Switzerland, 89p.