J. Ind. Geophys. Union (2003)
Vol.7, No.2, pp.97-103

Modeling of foF_2 using Analysis of Variance Technique

T.Madhu, A.D.Sarma, R.Annapurna and S.N.N.Pandit
R&T unit for Navigational Electronics, Osmania University, Hyderabad-500 007.

ABSTRACT
One of the most important parameters in predicting the electron density profile is foF_2, the critical frequency of the F_2 layer. A statistical technique called "Analysis of Variance (ANOVA)" has been used to model the foF_2 values. The foF_2 data for 8 stations, spread over the Indian region (equatorial and low latitude regions) for a solar cycle, are used in this study. Coefficients generated using ANOVA can be used to predict the foF_2 values for any given location (in the Indian subcontinent), time and solar activity. The preliminary results show a significant improvement in the prediction of foF_2 values as compared to those due to IRI-90.

INTRODUCTION

The advent of GPS has revolutionized the field of navigation, worldwide. The accuracy of GPS is limited by the precision in measuring the time delays. For better accuracy, precise ionospheric and tropospheric time delay estimations are needed for position fixing, satellite navigation and geodesy (Hoffman, Lichtenegger & Collins 1992; Lieck 1990). As the troposphere is a non-dispersive medium for frequencies up to 15GHz, standard models are available to accurately measure the time delay. But as the ionosphere is an ever-changing medium, it is very difficult to model the Total Electronic Content (TEC).

The number of electrons in a vertical column with a cross sectional area of one square meter ($1m^2$) is called the TEC. TEC of the ionosphere introduces a time delay in trans-ionospheric propagation and is subjected to geographical, seasonal, diurnal and solar activity variations. Various electron density models such as Bent and IRI-90 have been developed to predict the TEC. However, these models have some limitation for application to the Indian latitudes. Further, Goodman & Aarons (1990) reported that the climatological behaviour of TEC is quite consistent with that of the F-region parameter foF_2, which has been determined from vertical incidence sounders. Aggarwal & Shastri (1991) also stated that the F_2-region exhibits marked variations with geographic latitude and longitude, local time, season of the year and solar activity. As there is no suitable TEC model available for Indian subcontinent spread over 6°-38°N latitude range, it is proposed to modify the IRI-90 model. Hence in this paper, prediction of foF_2 is carried out by using the Indian ionosonde data from several stations, which can be used for improving the prediction accuracy of IRI model over the Indian subcontinent.

NEED FOR MODIFICATION OF THE IRI-90 MODEL IN THE CONTEXT OF GPS APPLICATIONS

The ionosphere in both the polar and equatorial regions often causes problems for navigation, surveillance and communication systems. Sarma et al. (2002) suggested that as India comes under equatorial region, there is a necessity to thoroughly understand the ionospheric effects on GPS.

Ionospheric time delay models that yield good results at one place can prove to be inadequate at other places. The ionospheric time delay model developed by Klobuchar (1987), which is currently being used by most of the single frequency GPS users to compute the ionospheric errors, is based on Bent's electron density model. In Bent model only a few measurements were available for low latitude regions. As the TEC behaviour in many parts of the world has a diurnal maximum near 1400 hrs local time, Klobuchar made the peak amplitude of the cosine curve to coincide with 1400 hrs in his time delay model. But for Indian conditions it is observed that the peak amplitude occurs at different times. For example, Dabas, Bharadwaj & Lal (1979) reported that at Kurukshetra (India) the peak amplitude occurs at 1500 hours. More over, Klobuchar has taken the mean ionospheric height as 350Kms and used several approximations for calculating the time delay. In view of the above drawbacks, it may be concluded that the Klobuchar's time delay algorithm is not suitable for the Indian subcontinent.

International Reference Ionosphere 1990 (IRI-90) is a standard empirical ionospheric model established on behalf of international union of radio science (URSI) and committee on space research (COSPAR). A computer program was developed by CCIR (ITU) based on ionosonde data obtained at 100 or more stations for representation of electron density and ion composition. The IRI electron density profile presented in Fig.1 is given by Bilitza (1990).

The figure depicts electron density(N) vs height(h) from 60-1000 kms. N_mF_2 and hmF_2 are the normalized F_2-peak density and the maximum height of F_2 layer respectively. Similarly N_mF_1 and hmF_1 are the normalized F_1-peak density and the maximum height of F_1 layer respectively. HZ is the upper boundary of the intermediate region and h_{VT} is the E-valley top. hmE and N_mE are the E-region peak height and density respectively. HA is the height at which the electron density profile starts. The whole profile is divided into six regions and are defined as the topside

($hmF_2 < h \le 1000$), the F_2 bottom side ($hmF_1 < h \le hmF_2$), the F_1-layer ($Hz < h \le hmF_1$), the intermediate region ($h_{VT} < h \le Hz$), the E-valley ($hmE < h \le h_{VT}$), and the E-bottom side and D-region ($h \le hmE$). The boundaries of these sub-regions are marked by several characteristic profile points including the F_2, F_1 and E-layer peaks. The critical frequencies foF_2, foF_1 and foE of different peaks have been monitored by the worldwide network of ionosondes since the forties. The IRI-90 programs are logically developed so that the parameters of the sub regions can be calculated independently.

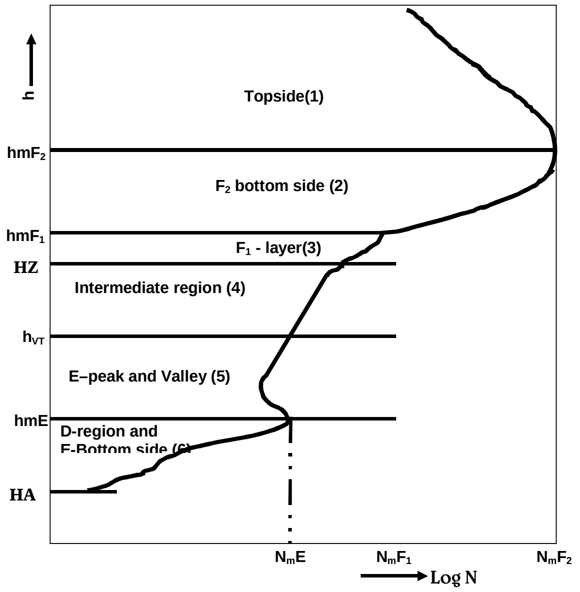

Figure 1. Buildup of IRI Electron Density Profile

In IRI-90 model, India is represented by a couple of stations. As data from only limited Indian stations is incorporated in the IRI electron density model, it is not expected to accurately estimate the ionospheric delay. foF_2 model developed with the Indian ionosonde data is to be plugged into the IRI-90 model so as to improve the accuracy of the model in predicting the ionospheric parameters relevant to the Indian conditions.

PREDICTION OF FOF_2 USING ANOVA METHOD

The ionospheric data was obtained from National Physical Laboratory (NPL), New Delhi over a period of one solar cycle from 1966-1976. The data was collected by NPL from vertical incidence stations in India viz: Delhi, Ahmedabad, Haringhatta, Bombay, Hyderabad, Madras, Tiruchirapalli, Kodaikanal and Thumba. A few data values were missing and for the purpose of analysis, they were assigned either with the interpolated values based on neighboring values or with values obtained from IRI-90 program. Since the data from Madras station were having very large number of missing values, that data were not used in our model construction. Hence, Indian ionosonde data of hourly median values of foF_2 for 8 stations, spread all over India, are modeled using ANOVA.

foF_2 data of one station corresponding to a selected month for all the 11 years is arranged in a 24x11 matrix. In this matrix, each row consists of foF_2 values for 11 years for a particular month and hour. The first row consists of the 1st hour values of foF_2 for all the 11 years. Similarly, the second row consists of the 2nd hour values of foF_2 for 11 years and so on.

Important steps in Analysis of Variance (ANOVA)

1. For the selected station and month, arrange the foF_2 data of all the 11 years in a 24x11 matrix. i.e. $X=[x_{ij}]_{24x11}$ (1)
2. The row mean ($x_{i.}$) and the column mean ($x_{.j}$) are found. Row mean varies due to variations in time and column mean due to variations in solar sunspot number (ssn). The grand mean ($x_{..}$) is found using the row mean and column mean.
3. The total sum of squares is also calculated along with the sum of squares of rows and columns.

$$\text{Total sum of squares (tss)} = \sum_{i=1}^{24}\sum_{j=1}^{11}\left(x_{ij} - x_{..}\right)^2$$

where m=24 and n=11 (2)

$$\text{Sum of squares of rows (ssr)} = [\sum_{i=1}^{24}\left(x_{i.} - x_{..}\right)^2]\times 11 \quad (3)$$

$$\text{Sum of squares of columns (ssc)} = [\sum_{j=1}^{11}\left(x_{.j} - x_{..}\right)^2]\times 24 \quad (4)$$

4. The sum of squares of the residual is obtained by subtracting the sum of squares of rows and columns from the total sum of squares.

Sum of squares of the residual (ssres) = tss-ssr-ssc

5. Degree of freedom is found for rows, columns and the residual.
6. The mean square deviation between rows, columns and the residual is calculated by dividing the respective sum of squares by degree of freedom.

Mean square error is given by

$$MS = \frac{Sum\ of\ \text{squares}}{Degree\ of\ \ freedom} \quad (5)$$

msr=ssr/23;msc=ssc/10;msres=ssres/230;

This model can be considered as good enough if the mean square of residuals is small enough. We find that this is the case with our data, and hence we accept this as good enough model for representing the data. Afterwards, row effect, column effect and the effect of geographic latitude and longitude are estimated.

Row effect and column effect

Row effect (RE) is due to the variation in time.

RE = row mean- grand mean = $x_{i.}$- $x_{..}$ (6)

Similarly Column Effect (CE) =

column mean - grand mean = $x_{.j}$- $x_{..}$ (7)

As CE is a function of ssn (R), a 3rd order polynomial in R is used to get CE as

$$CE = c_0 + c_1R + c_2R^2 + c_3R^3 \quad (8)$$

Similary for all the remaining stations, analysis of variance is carried out and the above process of finding the coefficients for the CE and RE is done.

Effect of geographic Latitude and Longitude on CE and RE coefficients

For every station there are 28 coefficients (C_0 to C_{27}), i.e. 24 RE coefficients and 4 CE coefficients (Polynomial coefficients). The coefficients of all the stations are arranged in a (8x28) matrix. The variation in these coefficients from one station to another is due to the effect of geographic latitude (ϕ_g) and longitude (λ_g) and are modeled by using a bivariate quadratic fit.

$$C_i = b_0 + b_1f_g + b_2l_g + b_3f_g^2 + b_4l_g^2 + b_5l_gf_g \quad (9)$$

The coefficients b_0 to b_5 are generated using Least square technique for each of the 28 coefficients. The coefficients thus generated are of size (6x28).

Effect of geographic Latitude and Longitude on the grand mean

The variation in grand mean (x..) is due to the variation in λ_g and ϕ_g and a 2nd order polynomial in λ_g and ϕ_g is used in defining it.

$$x = d_0 + d_1\phi_g + d_2\lambda_g + d_3\phi_g^2 + d_4\lambda_g^2 + d_5\lambda_g\phi_g \quad (10)$$

The coefficients d_0 to d_5 are generated using least square method.

Prediction of hourly foF_2 values

f_g, l_g along with b_0 to b_5 coefficients give the c_0 to c_{28} coefficients required to generate RE and CE (Eq. 9). The first 24 coefficients are used to get the effect of time (Eq. 6) and the remaining 4 coefficients are used to get the effect of ssn i.e. Year (Eq. 8). Using the generated RE and CE, Row Mean (RM) and Column Mean (CM) are estimated (Eqs. 6 &7). ϕ_g and λ_g together with the coefficients d_0 to d_5 are used to generate the Grand Mean (GM) for the selected station (Eq. 10). Now using the generated GM (xx), RM (xr) and CM (xc), foF_2 values can be estimated for any given place, time and solar activity.

$$\text{Predicted } foF_2 = RM + CM - GM \quad (11)$$

Following the above procedure, foF_2 values at Delhi station for January month for all the 11 years are predicted. The predicted foF_2 values are superimposed on the ionosonde foF_2 values as shown in Fig. 2.

The root mean square error between the ionosonde foF_2 and the predicted foF_2 is found to be 0.85. Similar procedure is carried out for all the other stations.

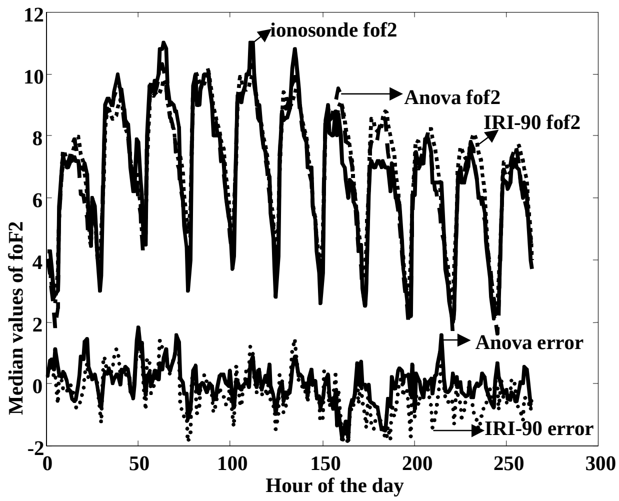

Figure 2. Comparison of modeled foF_2 and ionosonde foF_2 for Delhi (1966-1976).

Prediction of foF_2 for intermediate values of hour

For the intermediate values of hour, which are not integers, interpolation is used to find the corresponding RE value. The Lagrange interpolation formula reported by Chapra & Canale (1990) is used to estimate RE at that time. For this, the time of the day for which the foF_2 value has to be calculated is considered and the near by lowest integer is considered along with the next two higher integers. RE for each of these integers is calculated and is used to calculate the foF_2 value.

Suppose the time of day is ft, and let f1 be the previous integer to this value, then the fraction 'fa' is given as fa=ft-f1. Let 'f2' and 'f3' be the next two higher integers to ft and let y1, y2 and y3 be their respective row effects. Suppose 'y' is the row effect value at time ft, then

RE is given by $xr = y = y1 + fa\times(y2-y1) + fa\times(fa-1)\times\left(\frac{y3-2y2+y1}{2}\right) \quad (12)$

Similarly, CE and GM are calculated using Eqns. 8 and 10 and then foF_2 is calculated using Eq. 11.

EXAMPLE ESTIMATION:

Prediction of hourly values of foF_2

Suppose, the monthly median value of foF_2 at 11A.M(t=11) for January, 1969 (ssn=111) for Hyderabad station (ϕ_g = 17.35°; λ_g = 78.46°) is to be calculated. Using Eqns 6 to 10, RM (xr), CM (xc) and GM (xx) are predicted as xr= 11.279; xc=10.127; xx=8.4102

Hence, foF_2 = xr+xc-xx = 11.279+10.127-8.4102 = 12.996 MHz. The actual value measured at 11 A.M. is 12.6 MHz.

Prediction of foF_2 for intermediate values of hour

Suppose, the monthly median value of foF_2 at 11.30 A.M (t=11.30) for January, 1969 (ssn=111) for Hyderabad station

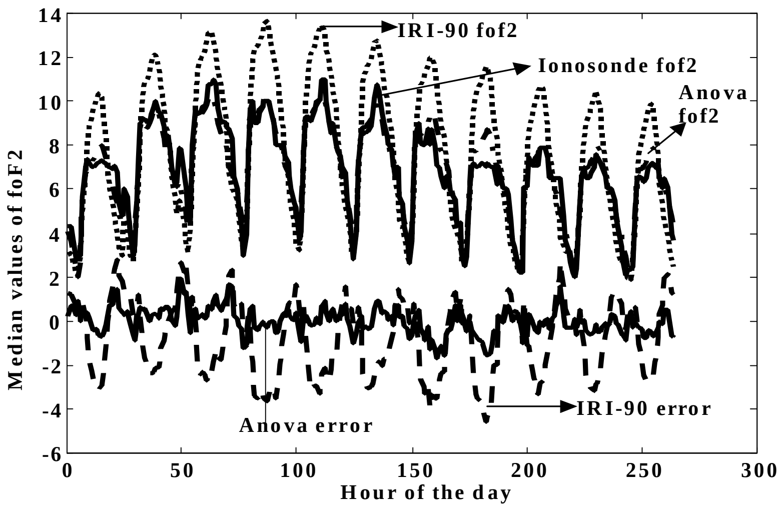

Figure 3. Comparison of ionosonde foF_2 values with ANOVA and IRI-90 for January month at Ahmedabad (1966-1976).

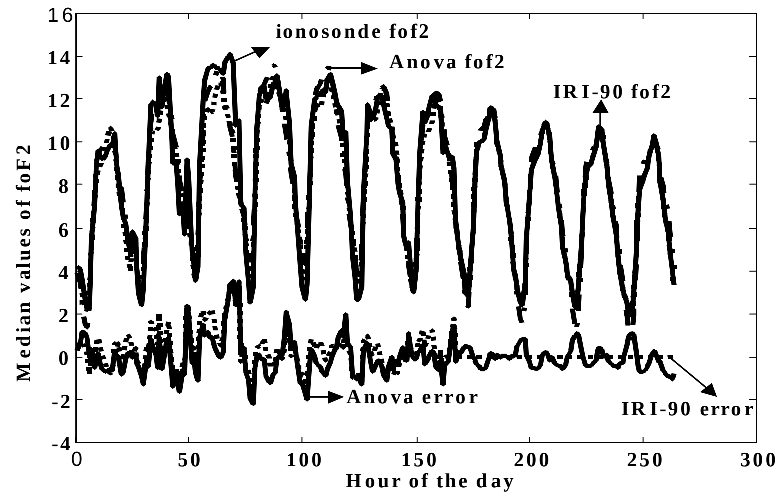

Figure 4. Comparison of ionosonde foF_2 values with ANOVA and IRI-90 for January month at Hyderabad(1966-1976).

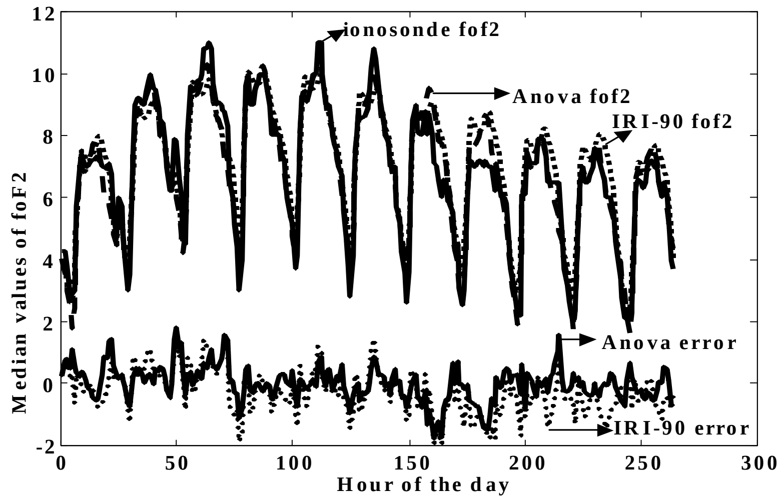

Figure 5. Comparison of ionosonde foF_2 values with ANOVA and IRI-90 for January month at Thumba (1966-1976).

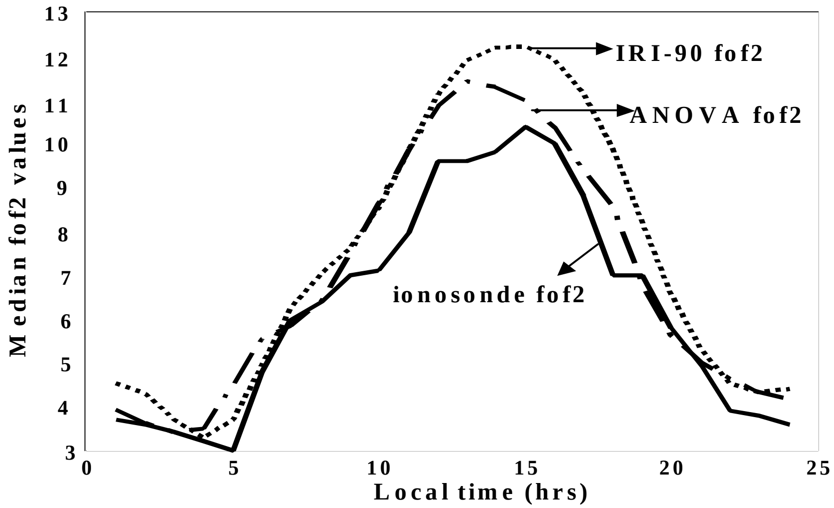

Figure 6. Comparison of ionosonde foF_2 values with ANOVA and IRI-90 for Ahmedabad in May, 1997.

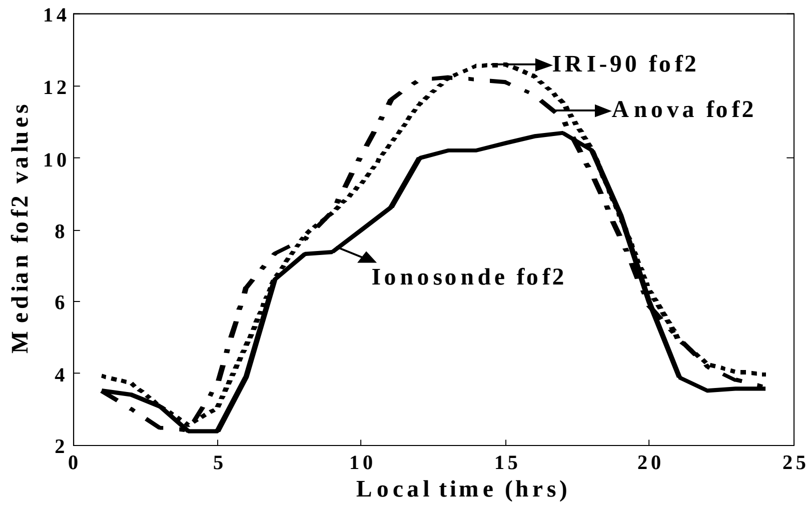

Figure 7. Comparison of ionosonde foF_2 values with ANOVA and IRI-90 for Ahmedabad in September, 1997.

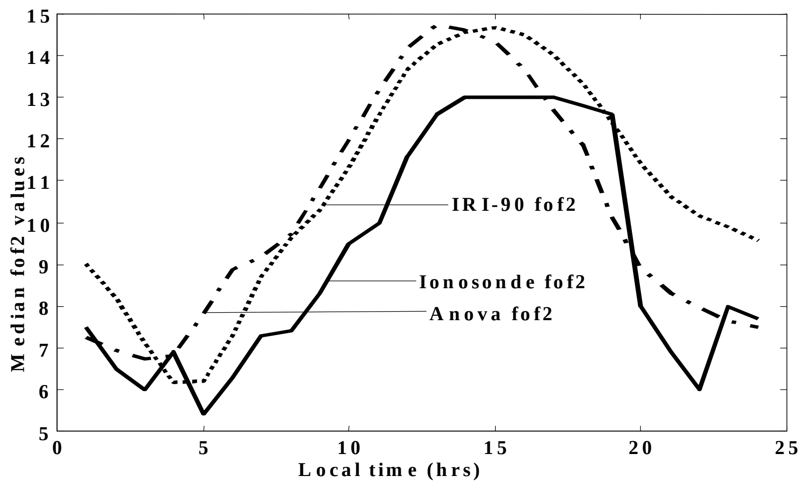

Figure 8. Comparison of ionosonde foF_2 values with ANOVA and IRI-90 for Ahmedabad in May, 1992.

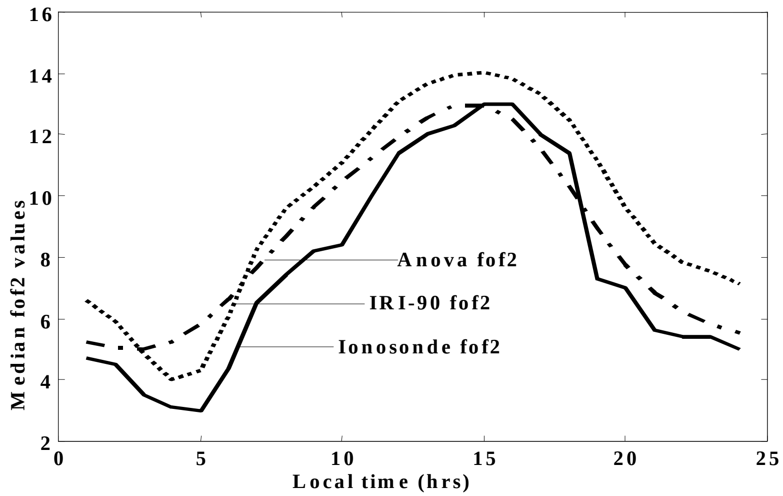

Figure 9. Comparison of ionosonde foF_2 values with ANOVA and IRI-90 for Ahmedabad in September, 1992.

(ϕ_g =17.35^0; λ_g =78.46^0) is to be calculated. The foF_2 value is calculated using quadratic interpolation when the time of the day is not an integer. For t=11.30, Ft=11.5, since 30/60=0.5. and fa=ft-f1=0.5.

Now, f1=11, f2=12 and f3=13.

For each of these hours foF_2 values are calculated. Let y1, y2 and y3 be the row effects at hours f1, f2 and f3 respectively. y1=2.869; y2=3.1942; y3=3.3239

By using Eq. 12, the row effect value at ft is obtained as y = RE = 2.987.

Now, Row Mean(RM) = y + GM = 2.987 + 8.4102 = 11.3972. In this example, the column mean and grand mean are the same as in Example 1.

Therefore, foF_2=xr+xc-xx =11.3972+10.127-8.4102=13.1140 MHz.

Comparison between ionosonde foF_2 with ANOVA and IRI-90

The modeled (ANOVA) foF_2 values are compared with the values due to IRI-90 and ionosonde for Ahmedabad, Hyderabad and Thumba stations (see figs. 3-5). The x-axis corresponds to monthly median values of 1-24 hours for all the 11 years. For example, 97th hour corresponds to 1st hour of January, 1970 and 250th hour corresponds to 10th hour of January, 1976.

It is obvious from Figs. 3 and 5 that the values set by the current model are better fitting to the ionosonde data than the IRI-90 values do for Ahmedabad and Thumba stations. At Hyderabad station (see Fig. 4) the results are not tallying well with the ionosonde foF_2 due to lack of continuous measured data. Also, the residual (i.e. the error) is seen to be not random but shows some trend to periodicity. Hence, an attempt is being made to model this residual itself. This model can then be plugged into the ANOVA model proposed in the present paper, and hence can improve considerably the predictability of our model. Table 1 gives the root mean square error between the modeled, experimental and IRI-90 foF_2 values for Ahmedabad, Hyderabad and Thumba stations.

Table 1. Error comparison table

rms error between:	Ahmedabad	Hyderabad	Thumba
Ionosonde and predicted (ANOVA)	0.6055	0.8536	0.5724
ionosonde and IRI-90	1.9152	0.8259	0.7558

Further, the generated coefficients for Ahmedabad station are used to predict the foF_2 values for ssn=18.3 (May, 1997), 28.3(September, 1997), 100.3(May, 1992) and 79.5(September, 1992) respectively. The predicted foF_2 values are compared with the values due to IRI-90 and ionosonde (see figs.6-9). Table 2 depicts the root mean square error between the modeled, experimental and IRI-90 foF_2 values for Ahmedabad station for May and September in 1992 and 1997.

Table 2. Error comparison table for Ahmedabad in 1992 and 1997.

rms error between:	May, 1997 (ssn=18.3)	September, 1997 (ssn=28.3)	May, 1992 (ssn=100.3)	September 1992 (ssn=79.5)
ionosonde and ANOVA	0.9235	1.3441	1.7662	1.9644
ionosonde and IRI-90	1.3775	1.1163	1.9987	1.9884

From the above figures, it is evident that the Anova program is predicting the foF_2 values better than IRI-90, except at some instants of time. However, incorporation of more data would enable the ANOVA program to predict the foF_2 values more accurately for any given place, time and solar activity within the Indian subcontinent.

CONCLUSIONS

In this paper, the hourly median values of foF_2 are modeled using statistical techniques. Coefficients are generated using analysis of variance method. Later, these coefficients are used to predict the foF_2 values at any given place, time and solar activity. For time values that are not integer, interpolation method is used. The foF_2 values predicted using the proposed model are compared with the foF_2 values due to IRI-90. It is observed that the modeled foF_2 values in general compare better with the experimental foF_2 values than those due to IRI-90. For better accuracy, more data from several stations within the Indian continent are to be modeled using ANOVA program.

ACKNOWLEDGEMENTS

The work in this paper is carried out under the project sponsored by Department of Science and Technology, New Delhi, vide sanction letter No. ESS/075/035/98, dated May 26, 1999. Thanks are due to Dr.D.R.Lakshmi, Emeritus Scientist, for her valuable suggestions and Ms.Paramita Karanjai for her help in developing some of the programs.

REFERENCES

Aggarwal, S. & Shastri, S., 1991. Atlas of Ionospheric communication Parameters over the Indian subcontinent, pp.2, National Physical Laboratory, New Delhi.

Bilitza, D., 1990. International Reference Ionosphere, pp.51, NASA report No: NSSdc/ WDC-A-R&S 90-22, National Space Science data Center, Science applications Research, Lanham, Maryland, 20706, USA.

Chapra, S. C. & Canale, R. P., 1990. Numerical Methods for Engineers, pp.379-385, Mc GRAW-HILL.

Dabas, R.S., Bharadwaj, R.M. & Lal, J.B., 1979. A study of Total Electron Content & Slab Thickness Variations at

Kurukshetra, Ind. J. of Radio & Space Phy.,, 8, 15-19.
Goodman, J. M. & Aarons, J., 1990. Ionospheric Effects on Modern Electronic Systems, Proc. of the IEEE, 78, 512-528.
Hoffman B., Lichtenegger H. & Collins, J., 1992. GPS theory and practice, pp. 88-106, Springer, N.Y.
Klobuchar, J.A., 1987. Ionospheric time delay algorithm for single frequency GPS users, IEEE trans. Aerospace and Electronic Systems, 23, .325-331.
Lieck. A., 1990. GPS Satellite Surveying, pp. 218-226, John Wiley, N.Y.
Sarma, A.D., Reddy, B.M., Madhu, T., Niranjan Prasad, Ravindra, K. & Lakshmi, D.R., 2002. Ionospheric Constraints and time delay statistics for Indian WAAS, 10th International Ionospheric Effects Symp. (IES-2002), pp. 179-183, 7-9, May, Alexandria, Virginia, USA.

(Accepted 2003 January 22. Received 2003 January 13; in original form 2002 November 11)