Monograph:
Redfern, E.J., Watson, S.M., Clark, S.D. et al. (1993) Modelling Outliers and Missing Values in traffic Count Data Using the ARIMA Model. Working Paper. Institute of Transport Studies, University of Leeds , Leeds, UK.

Working Paper 395

Published paper
Redfern, E.J., Watson, S.M., Clark, S.D., Tight, M.R., Payne, G.A. (1993) *Modelling Outliers and Missing Values in traffic Count Data Using the ARIMA Model.* Institute of Transport Studies, University of Leeds. Working Paper 395

UNIVERSITY OF LEEDS
Institute for Transport Studies

ITS Working Paper 395 *ISSN 0142-8942*

March 1993

MODELLING OUTLIERS AND MISSING VALUES IN TRAFFIC COUNT DATA USING THE ARIMA MODEL

EJ Redfern
SM Watson
SD Clark
MR Tight
GA Payne

Modelling Outliers and Missing values in traffic count data using the ARIMA model

Redfern E.J., Watson S.M., Clark S.D., Tight M.R. and Payne G.A.

Department of Statistics and Institute for Transport Studies
University of Leeds
Leeds LS2 9JT, England

Abstract. This paper considers the application of the methodology to traffic count time series in which both missing values and outliers are present. Intervention analysis and detection using large residuals are shown to be reasonably effective but possible problems that result from non-stationarity in the data are identified. It is shown that despite considerable variabilty in the types of series the model selected from the ARIMA family is surprisingly homogeneous.

Watson et al (1991) review the methods used for estimating missing values and identifying outliers in time series. These methods include techniques based on the estimate of the underlying autocorrelation structure and the use of a parametric model. Subsequent papers (Watson et al 1992a and 1992b) consider in more detail the use of methods based on the estimated autocorrelation function and their application to traffic count time series.

In this paper we consider the suitability of the Autoregressive Integrated Moving-Average (ARIMA) model and the use of the Box-Jenkins methodology (Box and Jenkins 1976) as a method for modelling traffic count series in the presence of missing values and outliers. This family of models has been used to model a variety of transport times series with varying degrees of success. The applications include rail and air passanger flows, journey times, public transport demand and the analysis of freeway traffic flow. (See Watson et al 1991 for a more detailed discussion.)

The family of models is represented by the equation

$$\phi(B)\Phi(B^s)\Delta\Delta_s X_t = \theta(B)\Theta(B^s)\epsilon_t$$

where ϕ and θ are polynomials in B of order p and q, Φ and Θ are polynomials in B^s of order P and Q respectively, B is the backward shift operator such that $BX_t = X_{(t-1)}$, $\Delta = (1-B)$ and $\Delta_s = (1-B^s)$, where s is the length of the seasonal period. ϵ_t is a noise process and is assumed to have a Gaussian distribution with zero mean and variance σ^2.

To model the outliers and missing values the model is extended by adding intervention series of the form

$$\zeta_t^{(k)} = \begin{cases} 1 & if\ t = k \\ 0 & elsewhere \end{cases}$$

where k is the time of the "event".

We therefore consider the extended model (Box and Tiao 1975)

$$X_t = \frac{\theta(B)\Theta(B^s)\epsilon_t}{\phi(B)\Phi(B^s)\Delta\Delta_s} + \sum_{k\in A} \frac{\omega(B)}{\delta(B)}\zeta_t^{(k)}$$

where $\omega(B)$ and $\delta(B)$ are polynomials in B which describe the "shape" of the event. Thus if $\omega(B) = \omega_0$ and $\delta(B) = 1$ we get a single spike at t = k equivalent to an outlier at that time, if $\omega(B) = \omega_0$ and $\delta(B) = 1 - \delta B$ we have a jump of size ω_0 which decays back to the original level at rate δ while if $\delta = 1$ we have a step of size ω_0.

Figures 1.1 to 1.5 show a selection of typical traffic count series. The data presented cover a range of road types and times of data over the period May to October. The series presented here cover Principle roads owned by a local authority (code P), trunk roads owned by the department of transport (T) and B class roads (B). They are subclassified into Built up roads with speeds greater than 40 (/B) and non-built up roads with speeds less than 40 (/N). The data covers the 153 observations between June 1st and October 31st 1991. Missing values were coded as zeros. From the figures we notice several features in the series.

(1) The weekly cycle varies from being very dominant (in rush hour traffic) to barely evident (in evening traffic).
(2) The amplitude and shape of the cycle varies according to the time and type of road.
(3) In some instances the cycle changes over the period of the data.
(4) The underlying volume of traffic also shows seasonal changes.

§2. Selection of Models

Figures 2.1 to 2.5 show the autocorrelation and partial autocorrelation functions for the seasonally differenced series of each of the series considered here after the identified outliers have been removed. The estimation of the two functions was done using a technique described by Marshall (1980).

Ignoring the seasonal factors at lags 7, 14 etc the **autocorrelation function** estimates for the T/N, B/B, T/B and P/N series all appear to decay to zero gradually over several values, while for the P/B series there is a single significant value at the first lag only. Considering the **partial autocorrelation function** estimates, the pattern is not quite as clear cut. For the T/N, B/B, T/B and P/N series the first is always significant and the second is often significant with occasional later ones (ignoring the seasonal lags for the time being). This suggests that the partial autocorrelation function cuts off before the auto-correlation implying that an autoregressive model of order 1 or 2 is appropriate for the short term non-seasonal correlation structure present in the data. For the P/B series the partial autocorrelation values, apart from those at 0800 hrs all have more than one significant term suggesting that a moving average model of order 1 is the most appropriate. At 0800 hrs only the first term is significant implying a pattern identical to the autocorrelation function. This implies that a mixed model may be needed but the problem of high correlation between the two parameters suggest that one of them can probably be ommitted.

Turning to the seasonal lags for T/N, B/B, T/B and P/N series there are no significant terms in the autocorrelation function while the partial autocorrelation function has significant terms at lag 7, usually at lag 14 and occasionally beyond. This suggests using a moving average model of order 1. For the P/B series there is no strong evidence of a seasonal component, however we include one in our models as an overfit to allow comparison with the other series.

Thus for most of the series a $(1,0,0)(0,1,1)_7$ model is appropriate with the possibility of an overfit by including a second autoregressive parameter in the non-seasonal part. In the case of the P/B series the $(0,0,1)(0,1,1)_7$ model appears to be prefered. To allow comparison over all the series we have fitted both these models to all the series, the results being summarized in tables 3.1 to 3.5.

§3. The fitted Models

To assess the models we present both the estimated variance and the value of Akaike's Information criterion (AIC) (Akaike 1974) which allows for the number of parameters in the models as well as the level of residual variance. The models were all fitted using the ARIMA procedure in the SAS package.

Tables 3.1 to 3.5 summarize the models fitted to most of our series using the large residuals to each fitted model to identify the possible outliers. These are then modelled by adding an intervention term to the series. In 32 of the 40 series analysed the Autoregressive model was the better fitting model, 5 of the discrepencies occuring in the case of the P/N series. The means (and standard errors) of the parameters for the five types of roads were

Road type	ϕ	Φ
T/B	0.539 (0.070)	0.491 (0.048)
P/B	0.373 (0.078)	0.492 (0.048)
B/B	0.283 (0.068)	0.614 (0.040)
T/N	0.386 (0.071)	0.664 (0.045)
P/N	0.403 (0.072)	0.713 (0.035)

Overall the mean of the ϕ parameter was 0.397 (0.033) while the Φ parameter had mean 0.597 (0.0233). Treating the parameter estimates as observations in an analysis of variance model, there was no significant difference between the estimated parameters at different times of the day nor between directions. Between the different roads there was no significant difference in the case of the non-seasonal parameter while the seasonal parameter was signifcant at p=0.002.

We observe therefore that there was considerable homogeneity in the resulting model over the data sets analysed, despite the large differences in the data as seen in figures 1.1 to 1.5.

Figure 3.1 illustrates the outliers detected in the various series. For the T/B series nearly all the events relate to Sunday peeks caused by day trippers. The patches of outliers in the T/N series about the beginning of August relate to problems with the counter while those at the end of September seem to be related to a similar event, the outliers surrounding a missing value. The cluster of outliers in the third P/N North series at the end of August are probably related to the clear change in the seasonality pattern that has occurred at that time. The only other significant event is the bank-holiday at the end of August which shows up on several of the series.

§4. Stability of the model.

To assess the stability of the model over time we took some longer sequences of the *P/B* traffic counts and fitted a global model and a series of models to successive quarters of the series. Tables 4.1 to 4.4 show the results for four of the extended series for 0800, 1200, 1700 and 2000 on the *P/B* road. The series and the divisions into segments are shown in figure 4.1.

We notice that the final model identified in each case was the same form as the global model. The parameter values do not change dramatically over time and the observation selected as outliers using the whole series are broadly consistent with those detected by using the model fitted to the local series. The estimated values for the missing values also show little difference between the global model and the partial model.

We also observe a consistency in the estimated noise variance between all the models and across the different times of observation.

§5. Some preliminary simulations

To evaluate the effectiveness of working with the largest residual as an indicator of possible outliers we did a preliminary study based on an actual traffic count series in which no outlier could be detected by any of the methods to which we added outliers following three patterns typical of the types of event that could occur in traffic counts. These were firstly one or more outliers at randomly selected points, secondly typical bankholiday patterns such as a high flow on a Friday low flow on a Monday, and finally a depressed flow throughout a week which may be typical of local events such as road works. Series 1-10 have a single outlier, 11-15 a bank holiday pattern,16-19 a random pattern of three outliers while 20-24 have a depressed or raised series of values over a block of working days. The standard deviation of the noise in the original series (σ) was 50 with fitted model

$$x_t - 0.57x_{t-1} = \epsilon_t - 0.73\epsilon_{t-1} \quad \text{where } x_t = y_t - y_{t-1}$$

In table 5.1 we show the errors in the replacement values suggested by using interventions at detected outliers, the success rate and the change in the parameters. The outliers identified here were solely on the basis of large residuals.

§6. Outliers and Influence

The outliers modelled above were all selected on the basis of large residuals. Such an approach is not guaranteed to identify the most influential points. In regression for example a point which has a large influence on the model can have a zero residual. To evaluate the direct translation of regression diagnostics in the time series environment, models were fitted to some of the DOT traffic count data using the regression procedure PROC REG in SAS. This supplies diagnostic values such as Cook's D (Cook 1977), the Hat matrix, DFFITS, COVRATIOs and DFBETAs (Belsey, Kuh and Welch 1984). These assess, for each observation in turn, the global effect on the parameters, the relative position in the X space, the prediction of the i^{th} value, the precision of the estimation and the effect of observations on the individual parameters in the model.

Tong (1989) suggested, for the AR(p) model, using nh_t where h_t is the t^{th} diagonal value of the hat matrix resulting from expressing the AR model in regression form. The autoregressive model is the one which translates directly into the regression environment, any moving average terms introducing a non-linearity into the estimation environment. Bearing this in mind we started by considering models in which X_t was regressed on X_{t-1} and X_{t-7}. This model was chosen since we have shown that the most common model to be used for the DOT data was an $(1,0,0).(0,1,1)_7$. Thus working with the data differenced at lag 7 but replacing the seasonal moving average term with an autoregressive term and omitting the lag 8 term that results from the multiplicative model we are left with the model

$$X_t = \beta_0 + \beta_1 X_{t-1} + \beta_7 X_{t-7} + \epsilon_t$$

We considered three alternatives for handling missing values. These were
(a) The values recorded as missing.
(b) The missing values treated as zeros
(c) The missing values modelled by interventions

Figure 6.1 shows the effect of this for the T/B west 1700 hrs series in which there are missing values at time points 27, 72 and 122.

Recording the missing values as missing (Case (a)) the regression diagnostics are the most promising as suggested by the first pair of graphs. The effect of a missing observation at time k is, as a result of the differencing, to produce an additional missing value at k+7. Ommitting all observations from the model that have missing values at some component also results in no information regarding Cook's D at times K+1, k+8 and k+15 while the hat statistic (since it only depends on the terms on the right hand side of the model) can be calculated at the time of the original missing value. A further consequence of the differencing and the inclusion of lagged variables as predictors is that an outlier will show up at k+7,k+8 and k+14 as well as k thereby making the interpretation difficult. The plots for this series suggest a strongly influential point at observation 36 (with significant values also occurring at 43,44 and 50). There is also a mildly influential point at time 85 (repeated at 92,93 and 99). This latter point is not shown up as having a large residual. The large hat value at 20 suggests that 13 may also be mildly influential. The residuals at these points do not show up as being large since they are swamped by the large residual at 36 and related lags. This agrees reasonably well with the analysis based on using residuals from a fitted ARIMA model in which observations 13 and 36 are identified but 85 is not. The latter approach detects a potential outlier at 40, a point where information is lost if we use the regression analysis ignoring missing values.

The effect of (b) was to produce large values of Cook's D and the Hat indicator at the time of the missing value and at lags 7 and 14 afterwards. It is obvious from the second pair of graphs that the observation missing at 72 is more critical than the others suggesting that not all missing values have equal effect on the modelling process. Using this approach the information on any other points is lost as a result of the dominance of the missing values.

If we model the missing values using interventions with a pulse at the time of the missing values (Case (c)) we get diagnostics that suggest these points, and these points only, are the influential ones. i.e we observe the opposite of our desired objective - namely that of identifying outliers in the presence of missing values.

From this we see that the regression type diagnostics can be useful but in the presence of non-stationarity (which must be removed by an appropriate level of differencing) there is a problem of interpretation, events of interest showing up at times displaced by the amount of differencing involved as well as at the actual time. If the differencing is not done then the problem is avoided, however spurious outliers show up as a result of the non-stationarity present in the data.

§7. The Changing Seasonal Pattern

Earlier we identified a possible problem, due to the changes in seasonal pattern and we illustrate the problem below. Plotting X_t against X_{t-1} allows us to see the strength of the seasonal pattern in the series and any global shifts in it. Outlying periods will be shown by single cycles that do not conform to the general pattern. To illustrate any possible shifts a plot has been produced with a change of line pattern every 50 observations.

Figures 7.1 to 7.4 show the result using the P/B series from June 1st 1991 - October 31st 1991. The conclusions are summarized as

08.00 North: Gradual decay of the seasonal peak although the trough is fairly stable. One period of seven days during the middle set of 50 observations seems to standout with a changed shape while one in the first 50 observations shows a slight change at the trough.

08.00 South: The central block of the data shows a drop in the seasonal peak otherwise the pattern is fairly stable. One week in the middle part of the data does not conform to the general pattern. This is the week in which the bank holiday occurrs.

12.00 North: In this plot we see a different seasonal pattern which is not as strong as in the first two series. There also appears to be a drop in the level over the last third of the series. One week in the middle part also appears to be different from the rest. Finally the pattern appears to be more stable over the latter part of the series than the first part, there being considerably more variability in the envelope over the first third of the data.

12.00 South: The pattern here is not clear. There is some sort of seasonal structure to the data but it is far from stable. There is also some evidence to suggest a drop in the middle part of the data.

17.00 North: Here we see a strong seasonal structure with a definite drop in the peak in the final third of the data. One week in the middle does not conform to the general pattern having extra days at the trough. It also appears that this occurrs at the same time as the drop in the peak.

17.00 South: The seasonal structure here is not quite as strong as the North series showing more variability in the trough than at the peak. One week in the early part appears to have a deeper trough than the other weeks but there are no other obvious patterns.

20.00 North: The structure here is far from clear. The main feature is the reduction of variability over the successive thirds of the data and a drop in the final part. The first part of the data has greater variability and less constancy of pattern.

20.00 South: The structure is similar to that for the North series at this time. There is a definite shift in the final part of the data particularly in the trough while the first part shows a much greater variability.

The features illustrated here highlight potential problems that can occur when using the ARIMA model for traffic count series. Outliers identified at times of structural changes such as these may be as a result of the change and not genuine outliers. The smoothing effect present in the one step ahead updating means that the predicted values from the model rapidly adjust to any new level thus masking any long term evidence that might appear in the residuals.

§8. Concluding Remarks

The time series considered here have several common aspects despite the apparent differences in structure initially evident. Using the ARIMA family of models we need a seasonal differencing to model the non-stationarity in the seasonal component, evidenced by the variability in the positions of the peaks and troughs. There is no evidence of any non-stationarity in the overall level of the series suggesting a long-term stability in the level. The short-term correlation structure was usually adequately modelled by a single autoregressive parameter. The only departures to this suggested by the modelling process was to use a single moving average parameter or an extra auto-regressive parameter. Thus the flow on any one day is usually only related to the immediately preceeding value, and occasionally this could be extended by a couple of values. Any such extended dependence was never very strong. The seasonal part of the model always suggested a single moving average parameter was adequate a much stronger dependence on the observations from the same days of previous weeks. In every case the estimated parameters had positive values.

The large residuals to the fitted model present an adequate starting diagnostic for identifying extreme values but analysis based on the regression diagnostics suggest that other types of influential points need to be considered. These may show up as outliers if only the large residuals are used where as they are in fact indicative of structural changes in the series. Failure to consider such problems may also result in outliers being missed due to masking. The problem of changing structure is highlighted in the phase analysis of the seasonal structure which suggests that a global model may not be robust if it ignores these changes.

The preliminary study of 24 simulated series suggest that the technique based on large residuals may be adequate in the case of single or isolated outliers but starts to breakdown when patches of outliers occur. Other structural changes are also likely to be problematical. A more detailed study of event detection based on a large set of simulated time series will be presented in a subsequent paper (Redfern et al 1992).

We conclude therefore that the ARIMA family may be useful for identifying outliers in traffic counts but its application must be treated with caution. The $(1,0,0)(0,1,1)_7$ model appears to be fairly robust although the possibility of events such as structural changes require that the process must be monitored. One approach to online monitoring would be to use a Kalman filter to allow subsequent updating of the parameters and interpretation of the one-step ahead forecasting errors to flag potential errors.

References

[1] Akaike.H.(1974) ***A new look at the statistical model identification*** IEEE AC-19 716-723.

[2] Belsey Kuh and Welsch (1984)***Regression Diagnostics*** Wiley

[3] Box G.E.P and Jenkins G.M. (1976) ***Time series analysis : Forecasting and Control*** Holden Day New York.

[4] Box G.E.P. and Tiao G.C (1975) ***Intervention Analysis with application to Econometric and Environmental problems*** J.A.S.A 70 70-79

[5] Cooke D. (1977) ***Detection of Influential Observations in Linear Regression*** Technometrics 19 15-18.

[6] Marshall R. (1980) ***Autocorrelation estimation of time series with randomly missing observations*** Biometrika 67 567-570.

[7] Redfern E.J. Watson S.M., Clark S.D. and Tight M.R. (1993) ***Identifying Outliers and other events in seasonal time series*** ITS working paper University of Leeds (in preparation).

[8] Tong H. 1989 Discussion to Leave–k–out diagnostics Bruce and Martin J.Roy.Statist.Soc B 51 363-424

[9] Watson S.M. Clark S.D. Redfern E.J. and Tight M.R. (1991) ***Detection of Outliers in Time Series*** ITS working paper 362 University of Leeds.

[10] Watson S.M. Clark S.D. Redfern E.J. and Tight M.R. (1992a) ***An influence method for outlier detection applied to time series traffic data*** ITS working paper 365, University of Leeds.

[11] Watson S.M. Clark S.D. Redfern E.J. and Tight M.R. (1992b) ***Development on an influence function statistics for outlier detection with time series traffic series*** ITS working paper 366, University of Leeds.

Table 3.1 *Summary of models fitted to the T/B series*

dir	time	ϕ	θ	Θ	AIC	$\hat{\sigma}$	outliers
E	8	0.24		0.26	1309	20.71	9 65 88 95 128
			-0.21	0.26	1310	20.78	9 65 88 95 128
E	12	0.62		0.61	1611	60.65	8 9 57 88 89
			-0.64	0.57	1616	61.77	8 9 57 88 89
E	17	0.65		0.61	1656	68.08	45 52 66 88
			-0.43	0.41	1680	74.08	45 52 66 88
E	20	0.82		0.66	1431	32.39	52 64 66 67 87 88
			-0.59	0.55	1462	37.53	52 64 66 80 88
W	8	0.42		0.52	1543	46.40	51 65 96
			-0.31	0.55	1552	48.00	51 65 93
W	12	0.73		0.41	1705	81.46	24
			-0.59	0.32	1745	93.35	87
W	17	0.37		0.47	1580	52.74	13 36 40
			-0.32	0.46	1598	56.35	13 36
W	20	0.46		0.39	1385	27.86	22 85 86 92
			-0.34	0.31	1396	28.82	22 85 86 92 93

Table 3.2 *Summary of models fitted to the P/B series*

dir	time	ϕ	θ	Θ	AIC	$\hat{\sigma}$	outliers
N	8	0.51		0.42	1396	28.12	8 88
			-0.38	0.34	1407	29.19	8 88
N	12	0.45		0.48	1462	35.07	16 23 153
			-0.31	0.42	1469	35.81	16 23 88 153
N	17	0.63		0.35	1606	59.08	85 88 89 93 95 135
			-0.44	0.28	1595	61.17	85 88 89 90 91 95 135
N	20	0.45		0.52	1441	32.42	21 31 34 44 51
			-0.37	0.48	1447	33.13	21 31 34 44 51
S	8	0.56		0.37	1534	46.35	8 88 95 96
			-0.42	0.27	1546	48.41	8 88 95 96
S	12	0.01		0.41	1441	32.79	10 17
			-0.01	0.41	1441	32.79	10 17
S	17	0.12		0.69	1461	32.25	88
			-0.08	0.68	14.62	35.32	88
S	20	0.25		0.70	1416	29.95	31 34 42
			-0.17	0.65	1405	28.77	21 31 34 42

Table 3.3 *Summary of models fitted to the B/B series*

dir	time	ϕ	θ	Θ	AIC	$\hat{\sigma}$	outliers
E	8	0.60		0.44	1254	17.94	88 89
			0.02	0.86	1235	16.19	65 72 88 95 123
E	12	0.12		0.70	1272	18.64	57
			-0.14	0.49	1330	23.90	57 72 121 122 123
E	17	0.19		0.63	1311	21.23	14
			-0.35	0.84	1386	28.67	9 56 78 79 88 121 122 123
E	20	0.10		0.73	1291	19.92	
			-0.45	0.64	1404	28.87	32 53 55 56 88 122
W	8	0.56		0.52	1328	22.48	88 95 125
			-0.34	0.74	1355	26.23	5 13 51 54 55 62 65 72 79 88 121 123
W	12	0.27		0.73	1326	22.34	42 57
			-0.30	0.70	1439	34.50	42 63 65 67 72 121 122 123
W	17	0.21		0.50	1271	18.35	43 49 54 60
			-0.20	0.70	1308	21.96	54 61 64 65 67 121 122 123
W	20	0.21		0.66	1236	16.52	
			-0.25	0.66	1235	16.46	

Table 3.4 *Summary of models fitted to the T/N series*

dir	time	ϕ	θ	Θ	AIC	$\hat{\sigma}$	outliers
N	8	0.22		0.74	1172	13.01	65 88 121 123
			-0.17	0.75	1173	13.08	65 88 121 123
N	12	0.22		0.53	1350	23.91	57 121 123 128
			-0.21	0.52	1351	23.96	57 121 123 128
N	17	0.46		0.64	1423	30.79	38 88 121 123
			-0.38	0.63	1431	31.67	38 88 121 123
N	20	0.32		0.44	1283	20.50	4 8 45 53 54 55 56 88 89 116
			-0.21	0.41	1293	21.55	4 6 45 53 54 55 56 58
S	8	0.67		0.71	1427	33.27	64 65 66 88
			-0.17	0.80	1458	35.97	64 65 88
S	12	0.68		0.83	1358	27.59	52 63 64 65 66 67 121
			-0.32	0.58	1378	27.80	52 63 64 65 67 121 123
S	17	0.17		0.67	1289	20.46	61 63 64 65 67 121 123
			-0.17	0.67	1289	20.80	61 63 64 65 67 121 123
S	20	0.35		0.75	1248	17.89	63 64 65 67 121
			-0.26	0.73	1265	19.04	63 64 65 67

Table 3.5 ***Summary of models fitted to the P/N series***

dir	time	ϕ	θ	Θ	AIC	$\hat{\sigma}$	outliers
N	8	0.22		0.74	1172	13.01	65 88 121 123
			-0.26	0.55	1257	17.59	88
N	12	0.27		0.59	1368	25.30	27 57 88 123 128
			-0.38	0.48	1612	61.12	2 48 27 88 89
N	17	0.44		0.70	1448	33.56	38 86 88 123
			0.00	0.70	1335	22.80	38 86 93 135
N	20	0.34		0.61	1272	18.64	
			-0.27	0.55	1214	15.14	66 88
S	8	0.69		0.76	1474	36.68	64 66 88
			-0.07	0.73	1274	18.66	88
S	12	0.72		0.90	1384	30.37	52 63 64 65 66 67 86 121
			0.16	0.75	1282	19.06	86 105
S	17	0.18		0.66	1289	20.43	61 63 64 65 67 105 121
			-0.07	0.74	1281	19.01	51 88 105
S	20	0.36		0.74	1249	17.91	63 64 65 67 105 121
			-0.10	0.68	1194	14.23	105

Table 4.1 ***Summary of events handled modelling 2 years of data using a single model to cover the whole period and by splitting the data into 4 periods of about six months P/B series at 0800 hrs***

Observations	ϕ	Θ	event	error	Observations	ϕ	Θ	event	error
			61	116	1 – 176	0.47	0.77	61	118
			64	283				64	284
			65	123				65	125
			92	313	$\sigma = 23$			92	309
			120	235				120	277
					$AIC = 1547$			121	101
			131	401 (missing)				131	403 (missing)
			142	74				142	64
			211	265	177 – 352	0.58	0.60	211	263
			219	387 (missing)				219	385 (Missing)
								327	62
			330	403	$\sigma = 24$			330	406
			331	345				331	342
			332	263	$AIC = 1509$			332	260
			333	231				333	236
			334	212				334	224
			335					335	40
			337	327				337	334
			418	131	353 – 528	0.57	0.46	418	340
			421	314				421	309
			422	152				422	144
			451	448 (missing)	$\sigma = 29$			451	453 (missing)
			452	463 (missing)				452	471 (missing)
			453	455 (missing)	$AIC = 1623$			453	466 (missing)
			456	352 (missing)				456	353(missing)
			484	348				484	340
			485	116				485	122
			499	481 (missing)				499	486 (missing)
			500	460 (missing)				500	453 (missing)
					529 – 705	0.55	0.53	547	47
			575	359				575	361
			601	215 (missing)				601	220 (missing)
			694	412	$\sigma = 26$			694	427
			695	432				695	426
			696	287	$AIC = 1582$			696	285
			697	243				697	243
			698	226				698	243
								701	102
			705	332				705	329

Table 4.2 ***Summary of events handled modelling 2 years of data using a single model to cover the whole period and by splitting the data into 4 periods of about six months P/B series at 1200 hrs (m=missing)***

Observations	ϕ	Θ	event	error	Observations	ϕ	Θ	event	error
1-707	0.24	0.77	27	83	1 – 176	0.20	0.71	27	82
			42	77					
			97	108				97	107
$\sigma = 25$			131(m)	388	$\sigma = 27$			131(m)	392
			293	88	177 – 352	0.20	0.67	293	87
			335	84	$\sigma = 28$				
			336	93					
			451(m)	371	353 – 528	0.22	0.70	451(m)	371
			452(m)	385	$\sigma = 25$			452(m)	385
			461	171				461	172
			499(m)	373				499(m)	371
			500(m)	368				500(m)	366
			601(m)	550	529 – 705	0.29	0.54	601(m)	550
			699	129	$\sigma = 29$			699	127
			700	97				700	96
			706	157				706	154

Table 4.3 ***Summary of events handled modelling 2 years of data using a single model to cover the whole period and by splitting the data into 4 periods of about six months P/B series 1700 hrs 9 m = missing***

Observations	ϕ	Θ	event	error	Observations	ϕ	Θ	event	error
1-727	0.33	0.65	105	66	1 – 182	0.20	0.59	105	89
$\sigma = 24$			131(m)	338	$\sigma = 24$			131(m)	339
			337	55	183 – 364	0.49	0.62	337	53
					$\sigma = 28$				
			424	82	365 – 546	0.31	0.70	424	79
			451(m)	284	$\sigma = 27$			451(m)	283
			452(m)	328				452(m)	328
			461	57					
			462	56					
			464	46					
			499(m)	296				499(m)	320
			500	68					
					547-727	0.36	0.64	563	72
			601(m)	298	$\sigma =$			601(m)	298
			694	57				694	57
			705	45				705	44

Table 4.4 ***Summary of events handled modelling 2 years of data using a single model to cover the whole period and by splitting the data into 4 periods of about six months P/B series at 2000 hrs***

Observations	ϕ	Θ	event	error	Observations	ϕ	Θ	event	error
1– 727	0.33	0.65	105	66	1 – 182	0.20	0.59	105	69
$\sigma = 24$			92		$\sigma = 24$				
			131	338 (missing)				131	339 (missing)
			337	55	183 – 364	0.49	0.62	337	53
					$\sigma = 20$				
			418	78	365 – 546	0.31	0.70		
			424	82				424	79
			451	284 (missing)	$\sigma = 27$			451	283 (missing)
			452	328 (missing)				452	328 (missing)
			461	57					
			462	56					
			463	46					
			499	296 (missing)				499	320 (missing)
			500	68					
					547 – 727	0.36	0.64	563	72
			601	298 (missing)				601	298 (missing)
			694	57	$\sigma = 24$			694	57
			705	45				705	44

Table 5.1 *Summary of outliers detected in 24 simulated situations*

	Error size	initial Model ϕ	Φ	Final Model ϕ	Φ	Outliers found	Outliers missed	Error in replacement
1	4σ	0.53	0.69	0.56	0.72	44		21
2	4σ	0.52	0.79	0.55	0.72	45		5
3	6σ	0.42	0.80	0.56	0.73	24		-23
4	6σ	0.49	0.79	0.56	0.73	59		1
5	6σ	0.47	0.74	0.56	0.73	80		9
6	6σ	0.39	0.82	0.57	0.73	38		26
7	$> 6\sigma$	0.12	0.99	0.56	0.72	59		1
8	$> 6\sigma$	0.22	0.99	0.56	0.72	66		34
9	$> 6\sigma$	0.23	0.99	0.57	0.73	73		-56
10	$> 6\sigma$	0.23	0.99	0.56	0.73	24		-23
11	$> 6\sigma$	0.31	0.91	0.57	0.73	21 24		60 -22
12	$> 6\sigma$	0.13	0.98	0.55	0.73	56 59		52 1
13	$> 6\sigma$	0.20	0.89	0.56	0.73	77 80		13 -8
14	4σ	0.25	0.89	0.55	0.73	56 59		51 1
15	4σ	0.38	0.82	0.57	0.73	42 45		29 -18
16	8σ	0.31	0.84	0.60	0.65	27 41 71		128 -20 -24
17	6σ	0.42	0.95	0.57	0.72	34 56 82		-72 53 -2
18	4σ	0.38	0.86	0.56	0.74	30 45 87		-6 -20 -54
19	4σ	0.38	0.73	0.56	0.73	31 47 54		-50 -25 27
20	4σ	0.69	0.66	0.70	0.67	45	46 47 48 49	-158 * * * *
21	4σ	0.63	0.85	–	–	–	80 81 82 83 84	* * * * *
22	8σ	0.71	0.96	0.58	0.76	59 60 61 62 63		5 -4 17 -22 9
23	12σ	0.76	0.98	0.57	0.70	45 46 47 48 49		-58 -100 -74 -7 -29
24	6σ	0.53	0.89	0.57	0.70	45 46 47 49		-56 -98 -70 -10

Fig 1.1 T/B Series June 1st 1991 - October 31st 1991

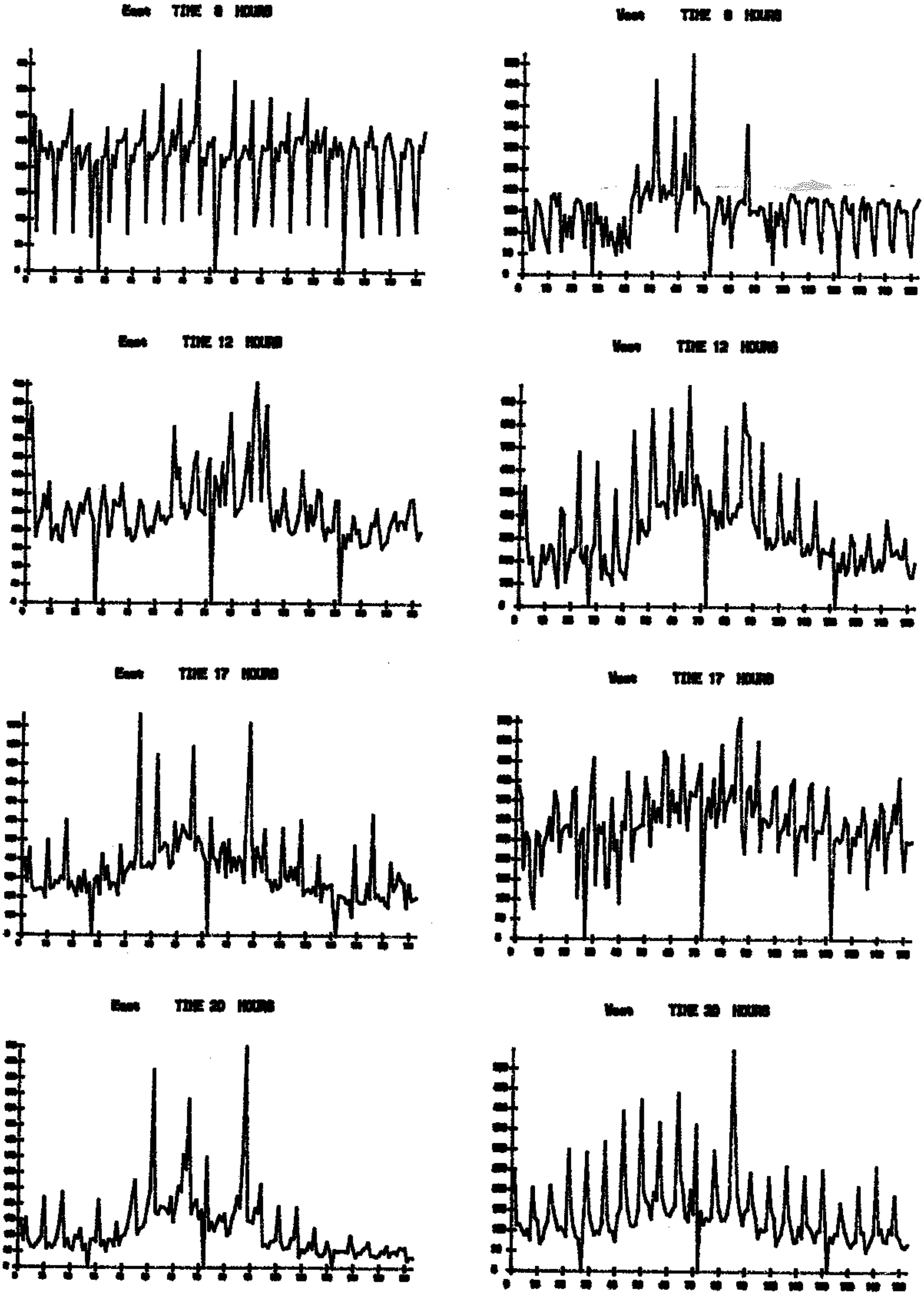


Fig 1.2 P/B Series June 1st 1991 - October 31st 1991

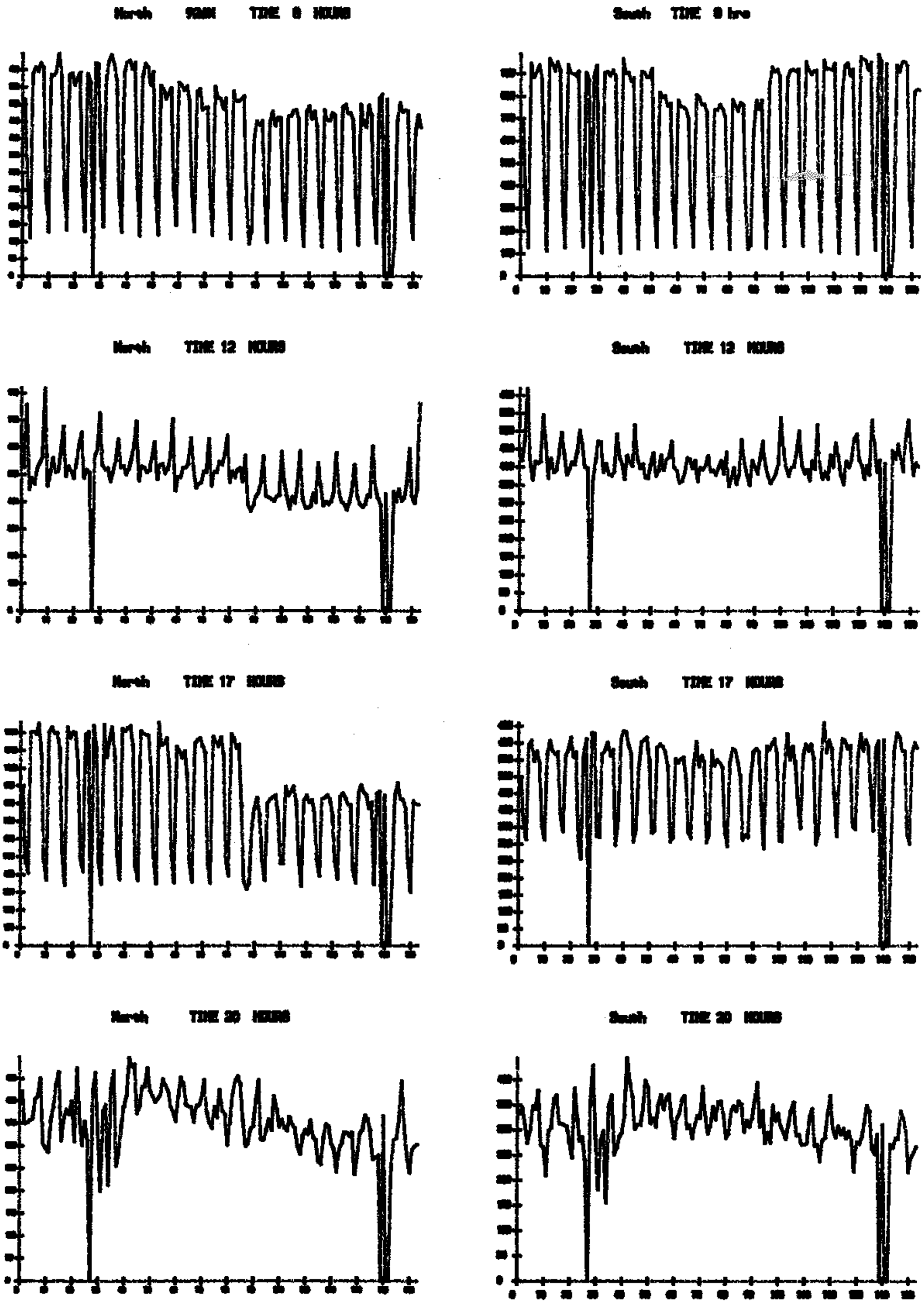


Fig 1.3 B/B Series June 1st 1991 - October 31st 1991

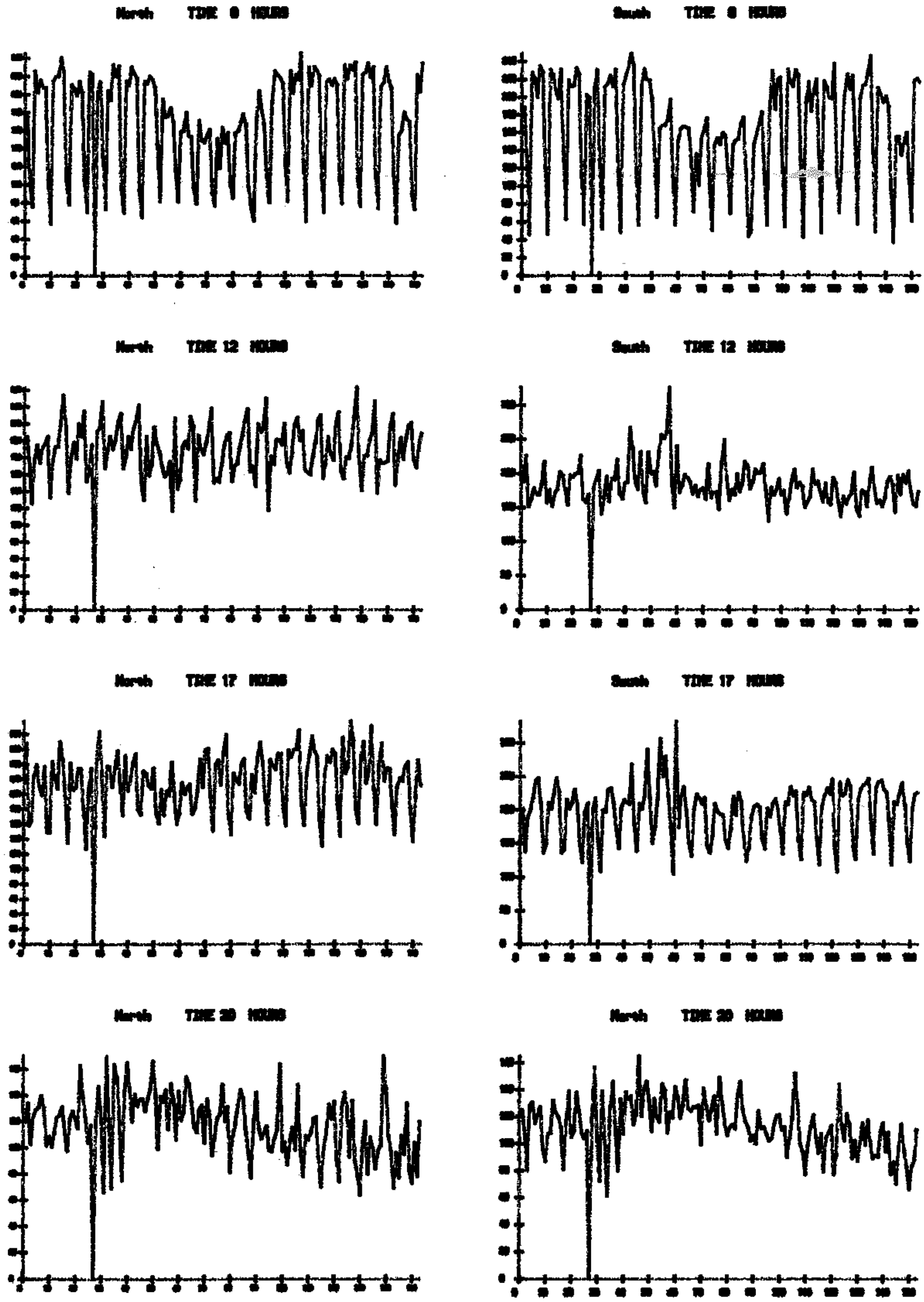


Fig 1.4 T/N Series June 1st 1991 - October 31st 1991

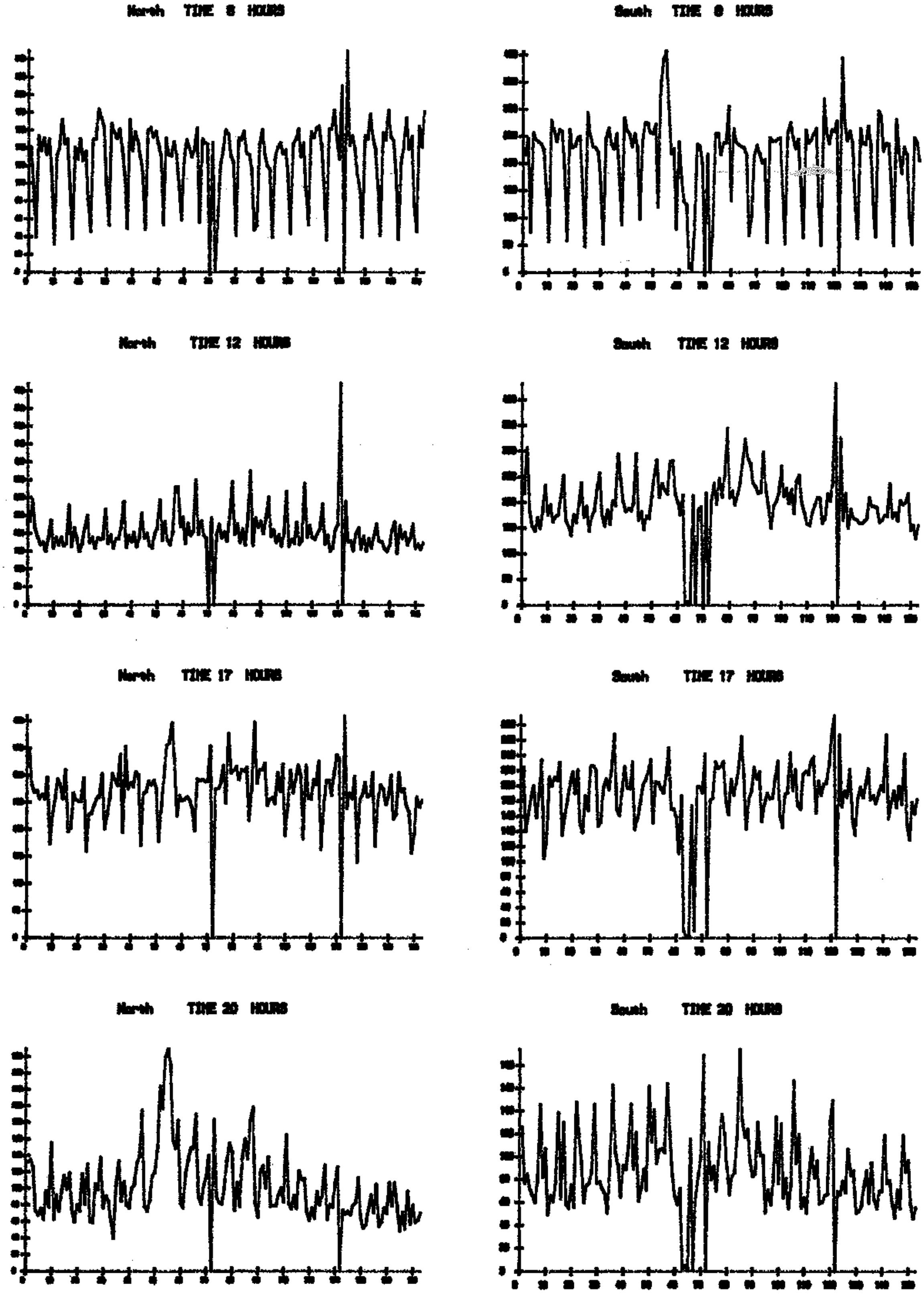


Fig 1.5 P/N Series June 1st 1991 - October 31st 1991

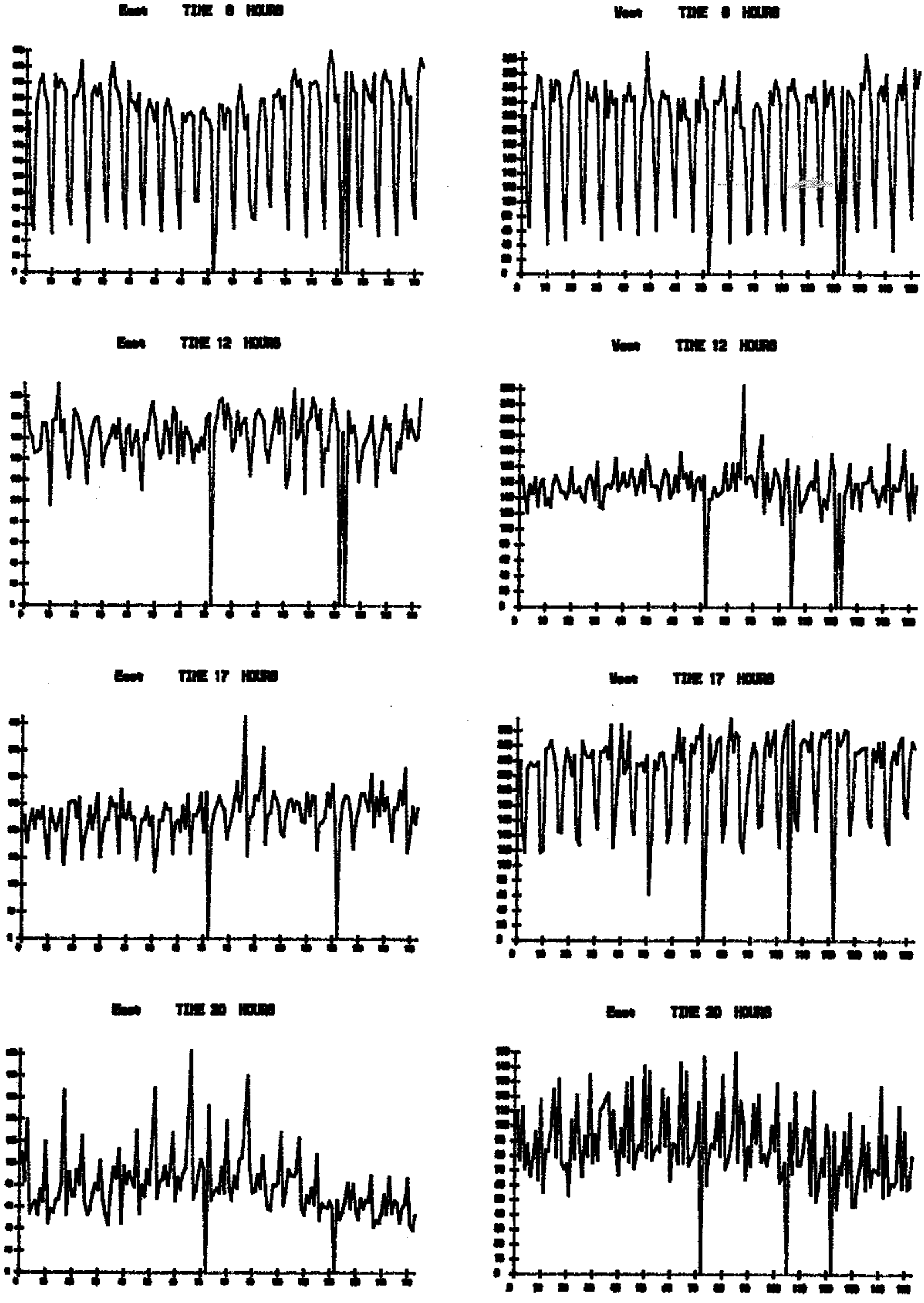


Fig 2.1 T/B Series June 1st 1991 - October 31st 1991

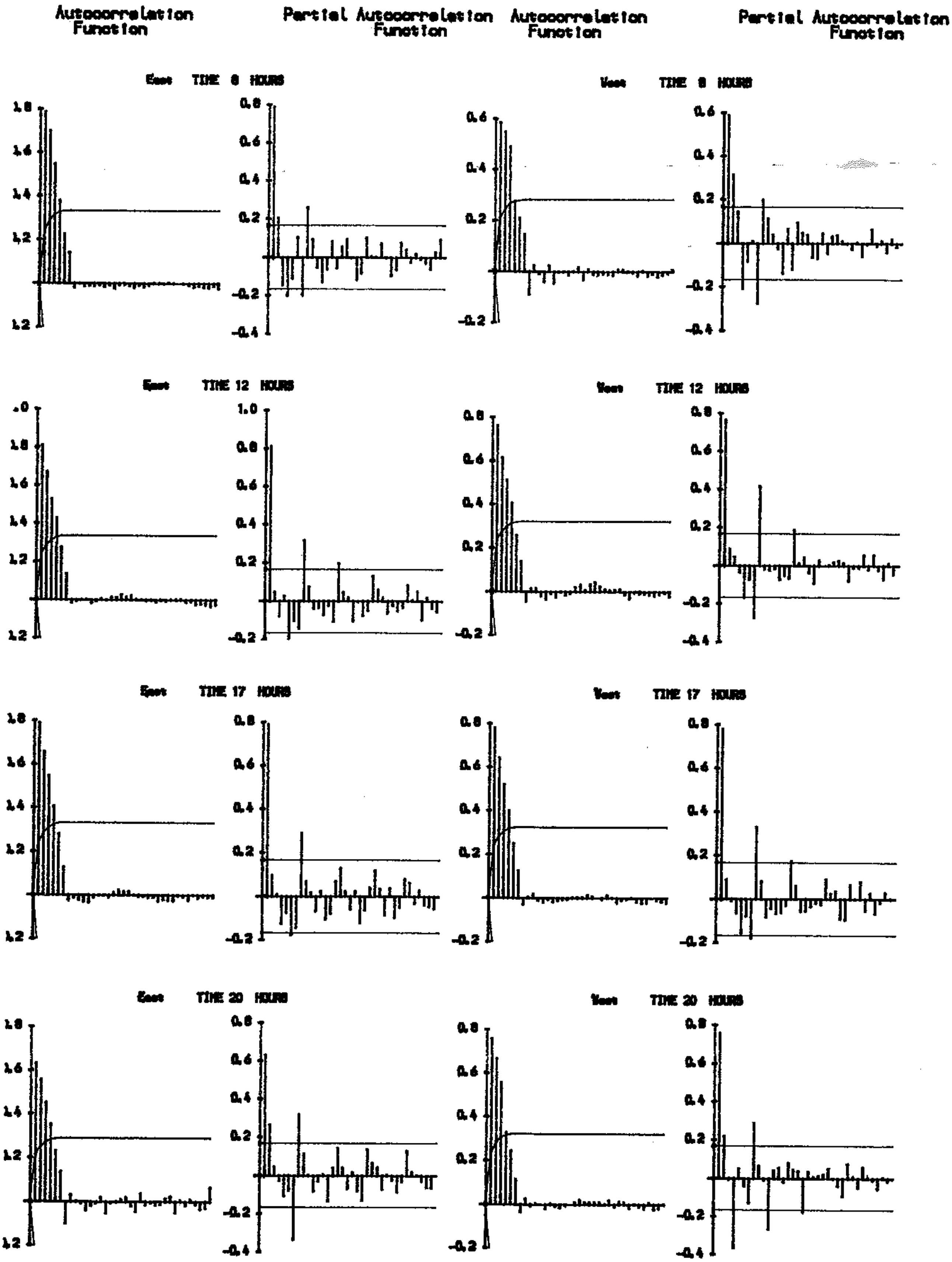


Fig 2.2 P/B Series June 1st 1991 - October 31st 1991

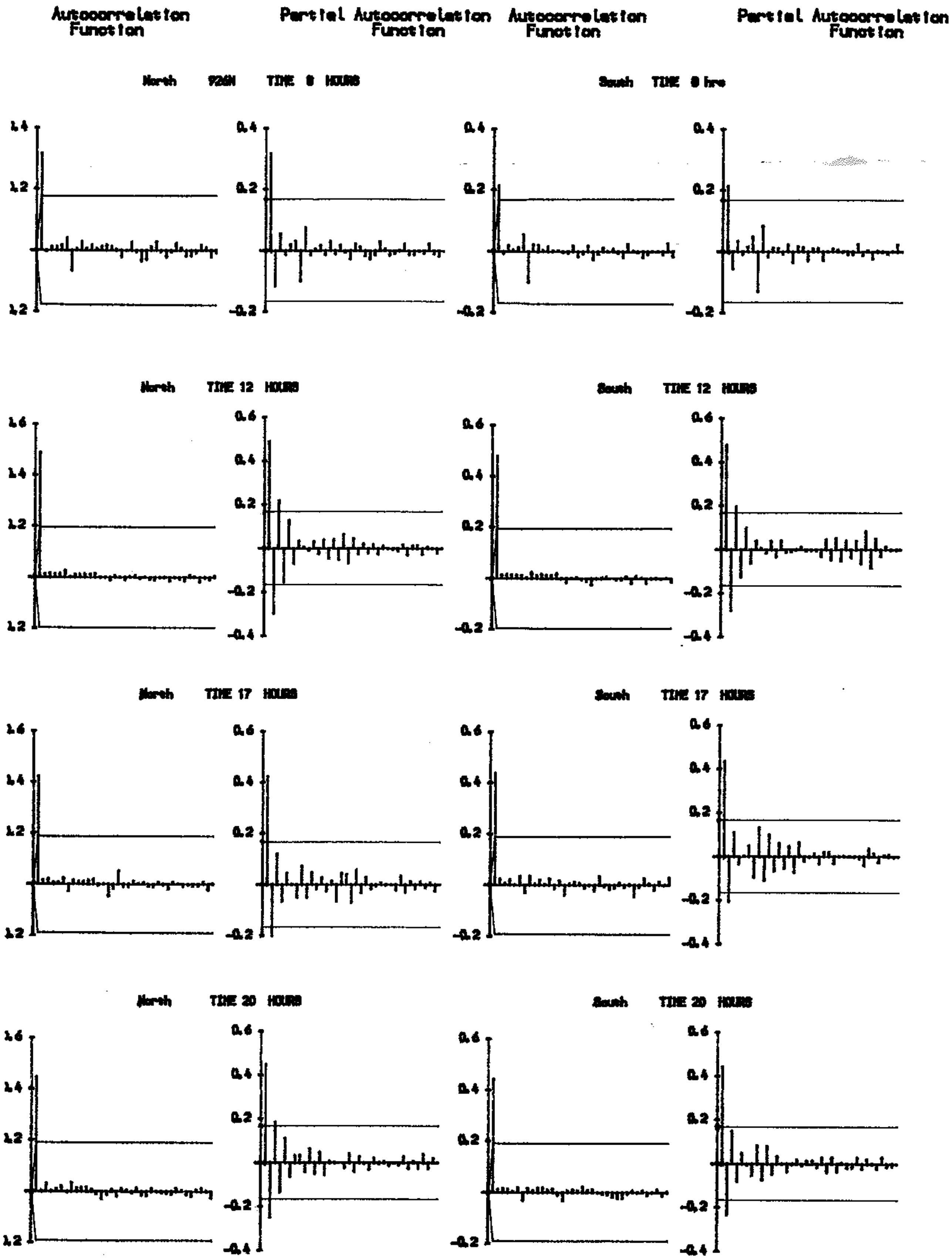


Fig 2.3 B/B Series June 1st 1991 - October 31st 1991

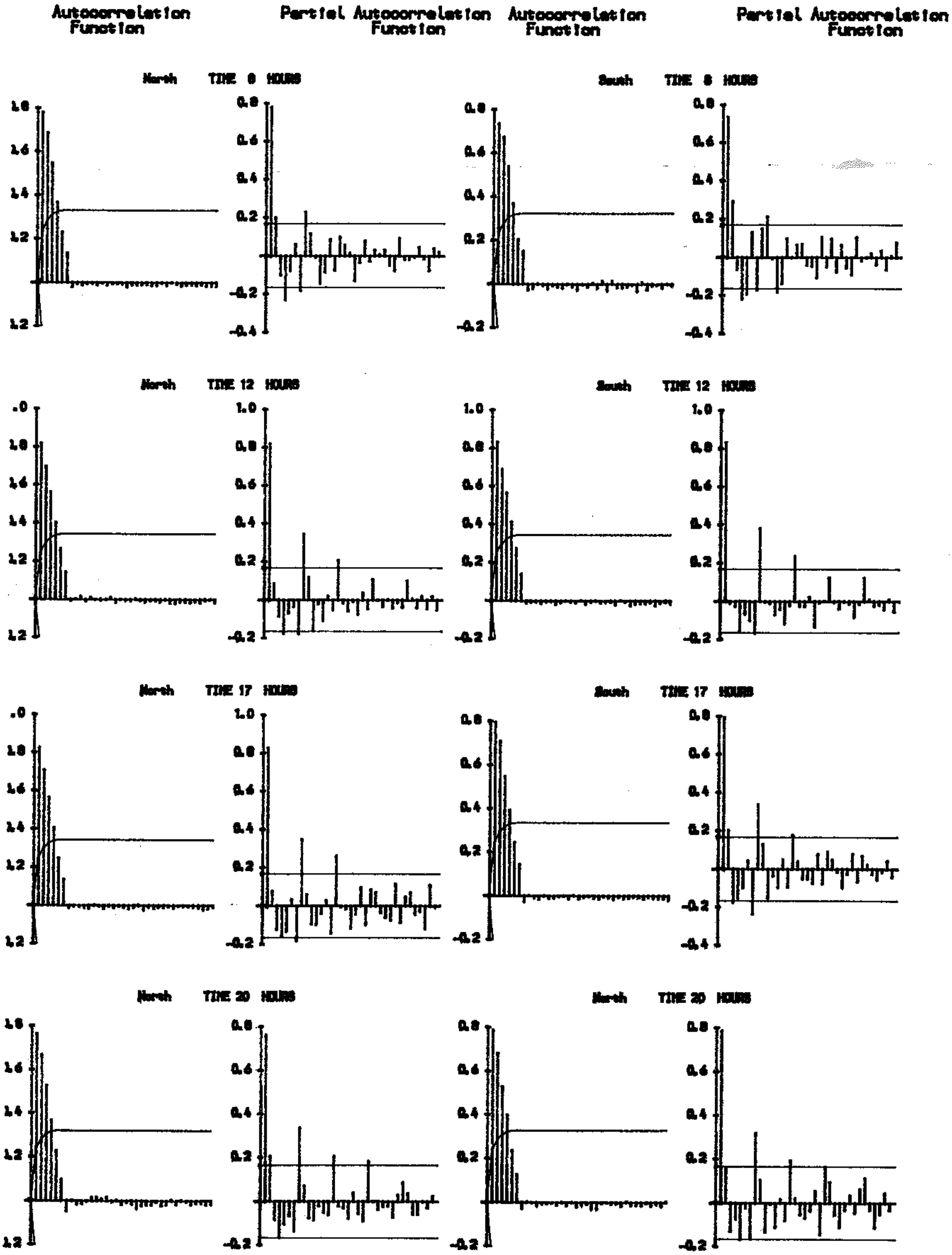


Fig 2.4 T/N Series June 1st 1991 - October 31st 1991

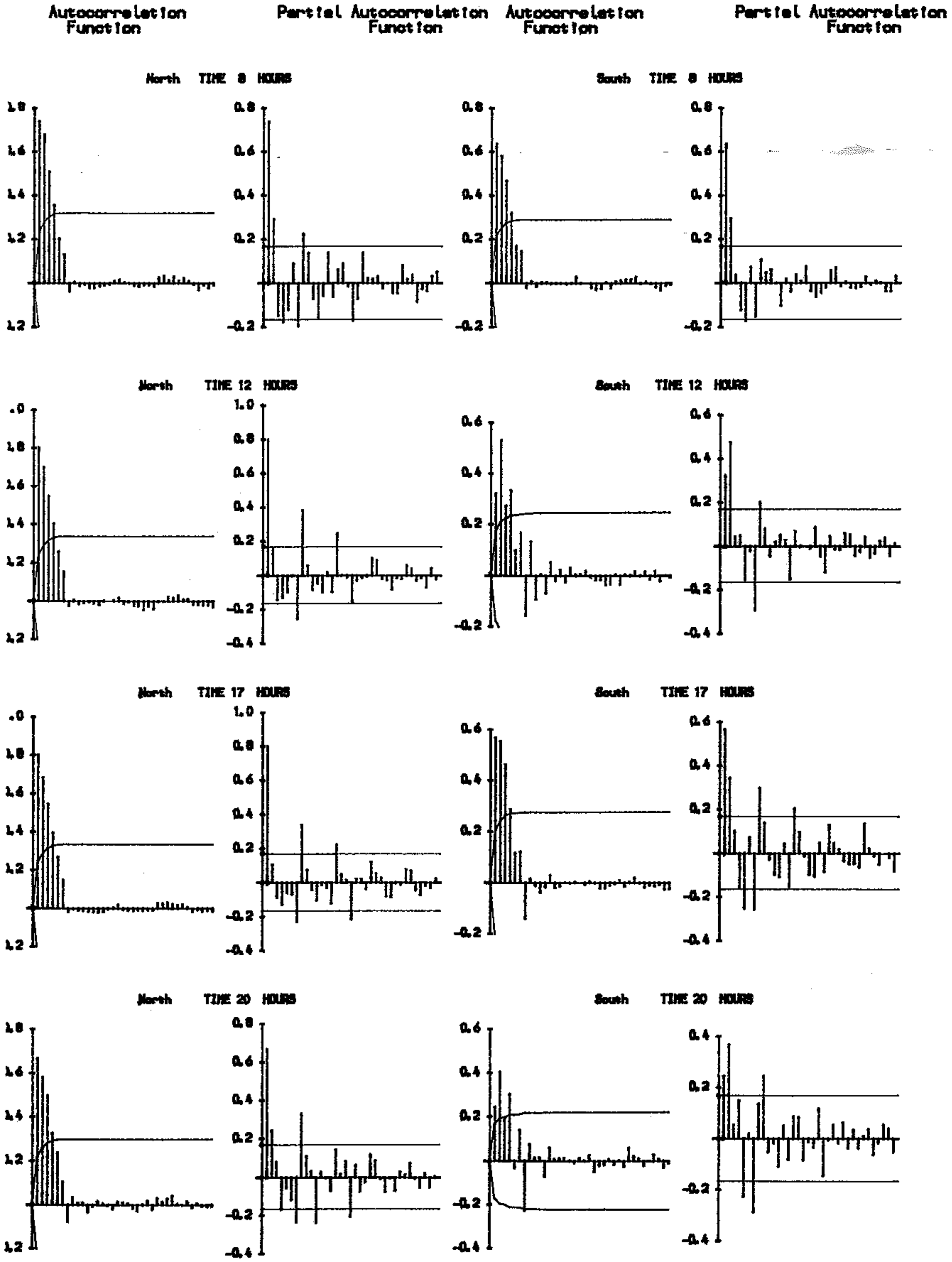


Fig 2.5 P/N Series June 1st 1991 - October 31st 1991

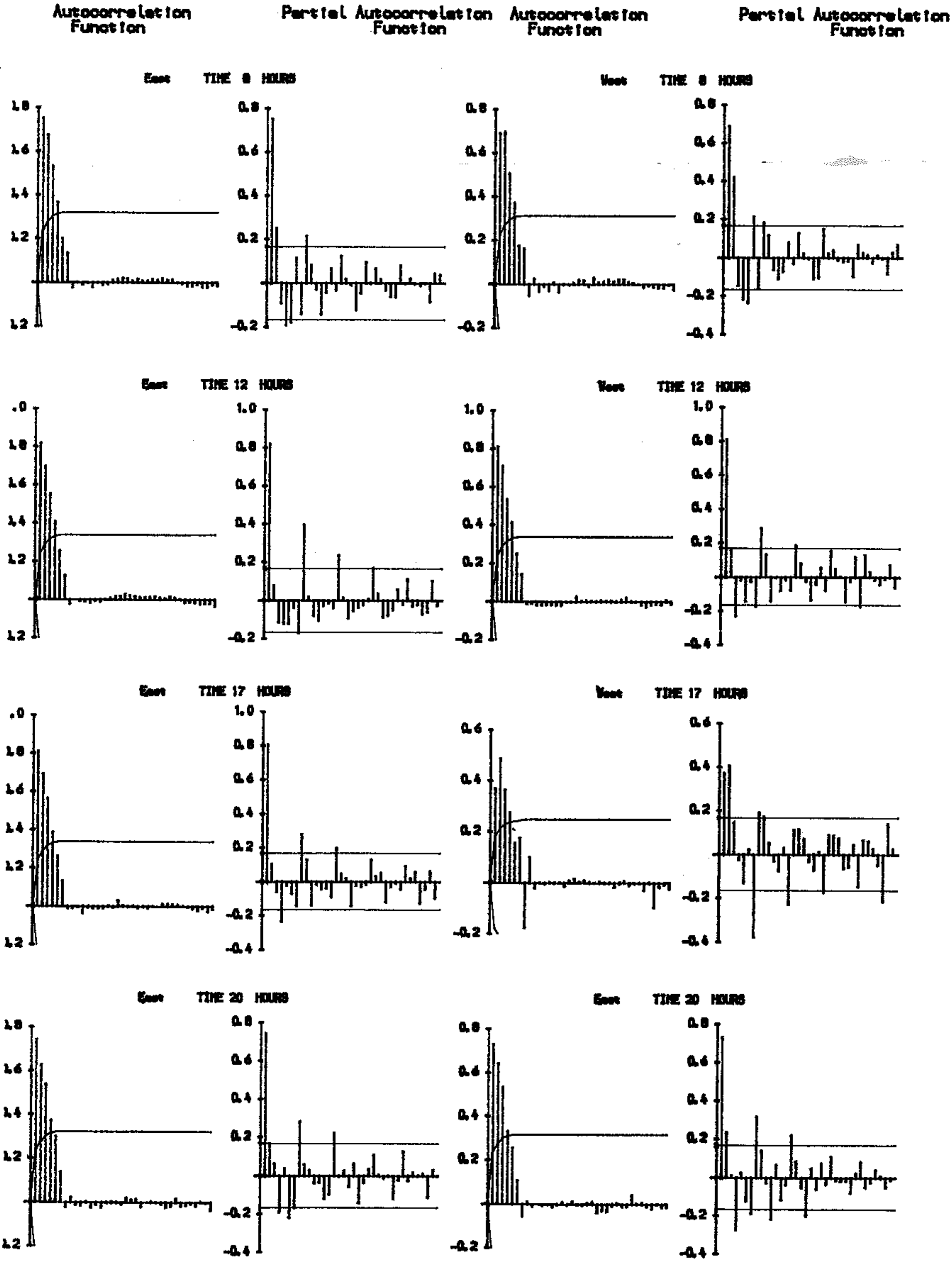


Fig 3.1. Outliers detected in the 40 traffic count series
in the period June 1st to October 31st
Vertical lines indicate the Sundays

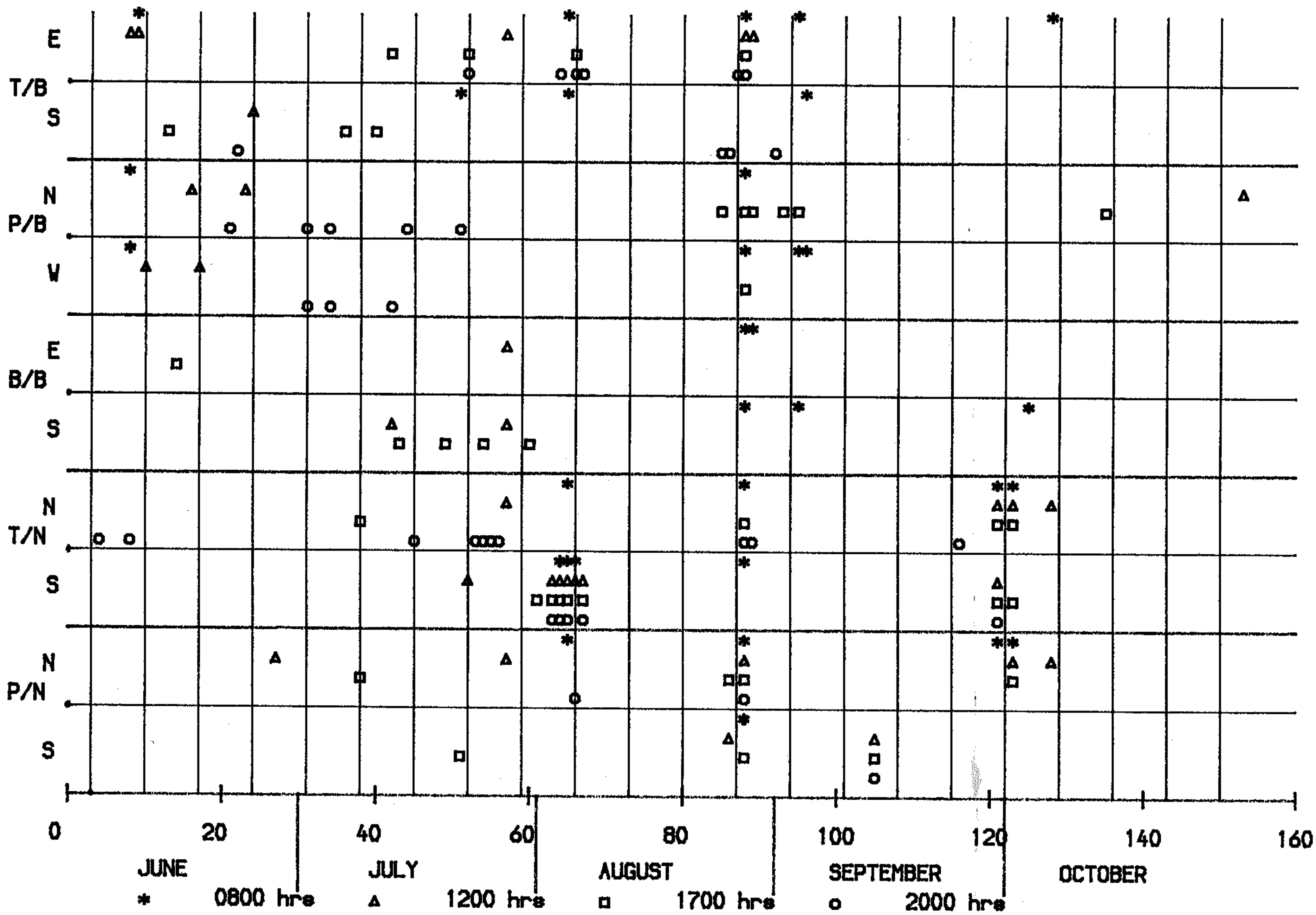
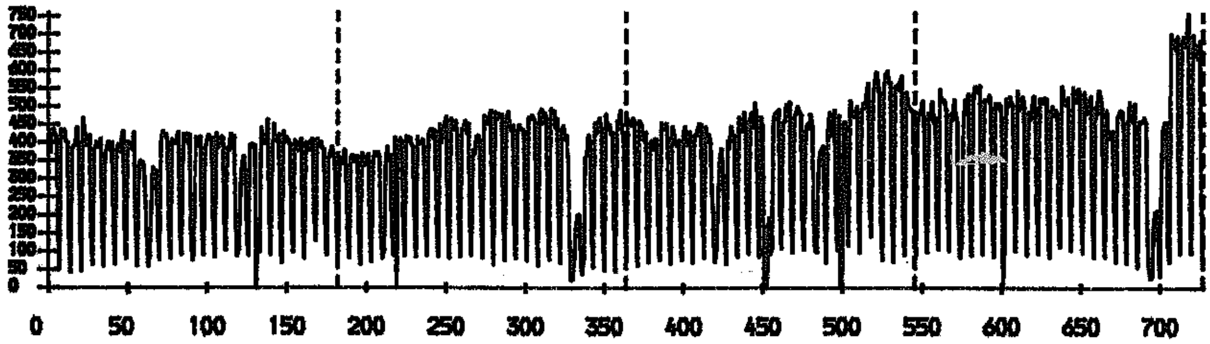
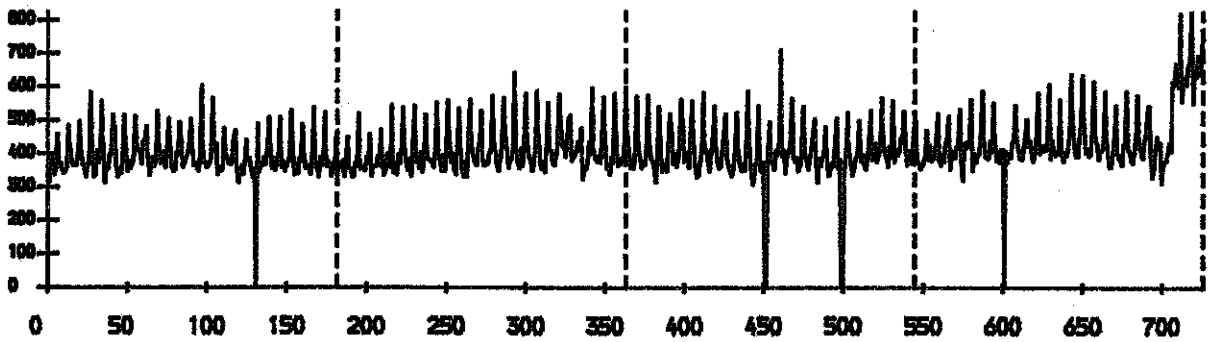


Fig 4.1 Four series no P/B class road for the period
February 1st 1988 - January 28th 1991
showing the split into divisions for assessing
the stability of the model

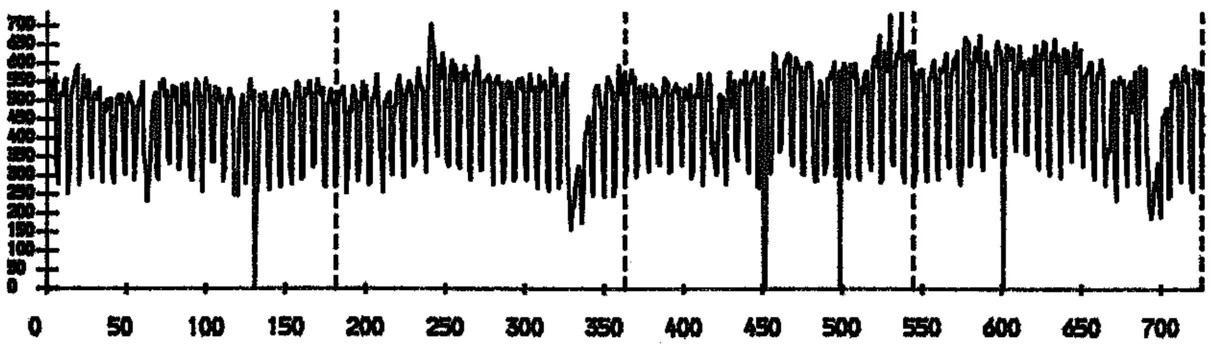
North 0800



North 1200



South 1700



South 2000

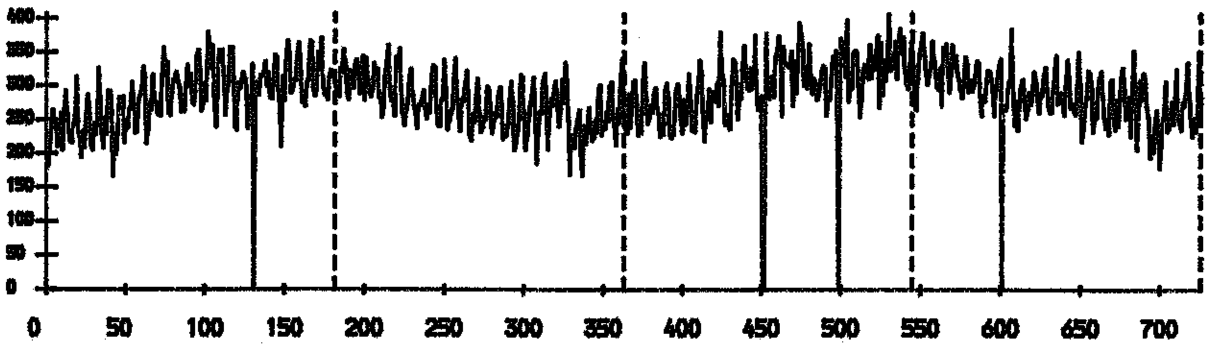
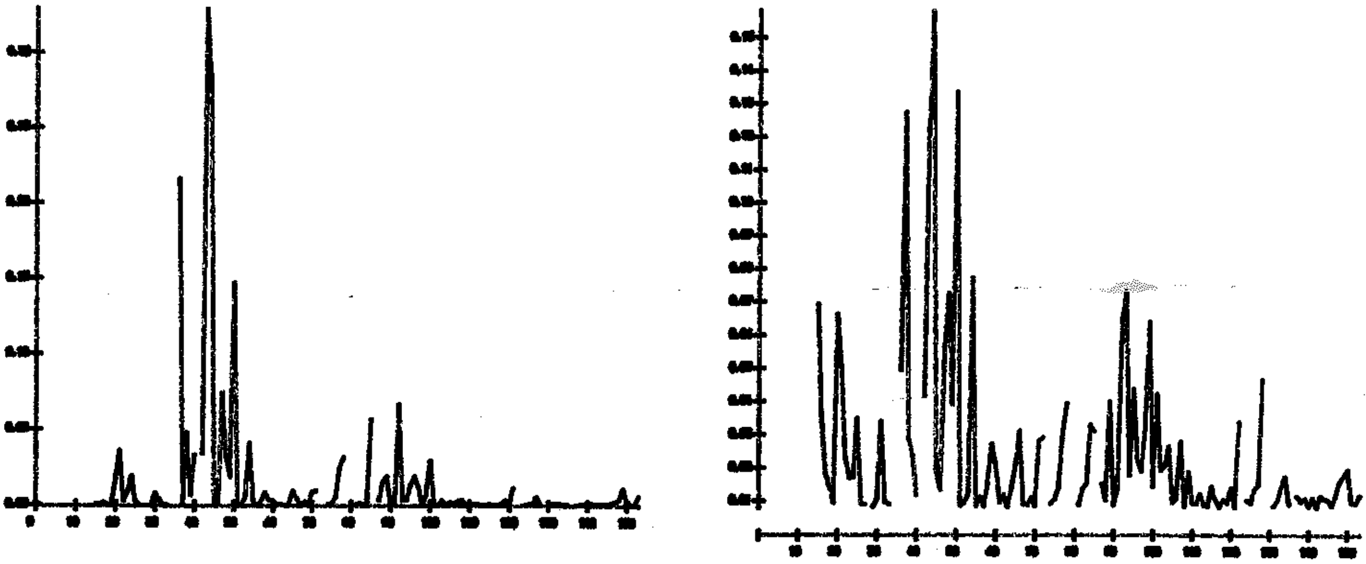


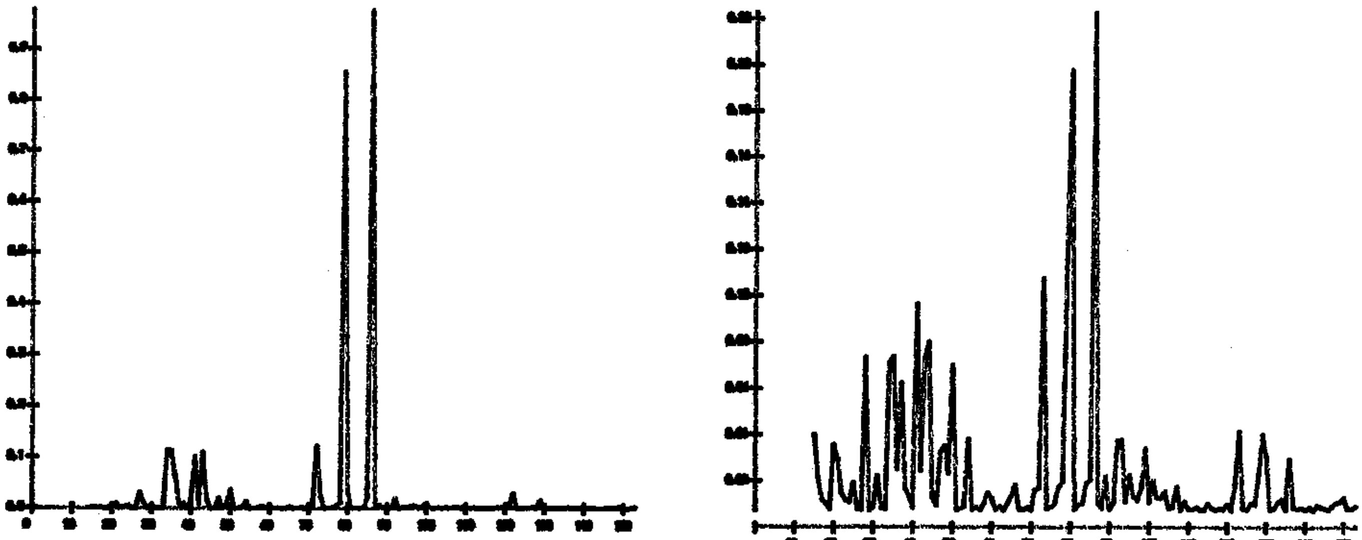
Fig 6.1 Regression Diagnostics applied to T/B West

Cook's D Statistic | Hat Statistic

Case a) Missing values ommitted



Case b) Missing values treated as zeros



Case c) Missing values modelled by interventions

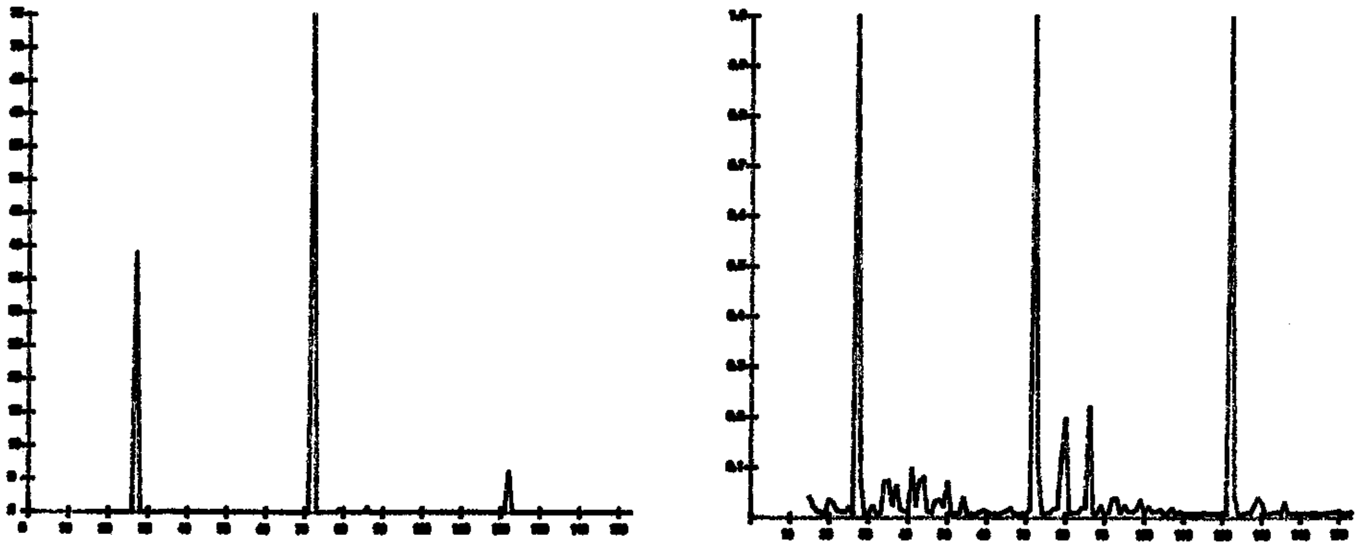
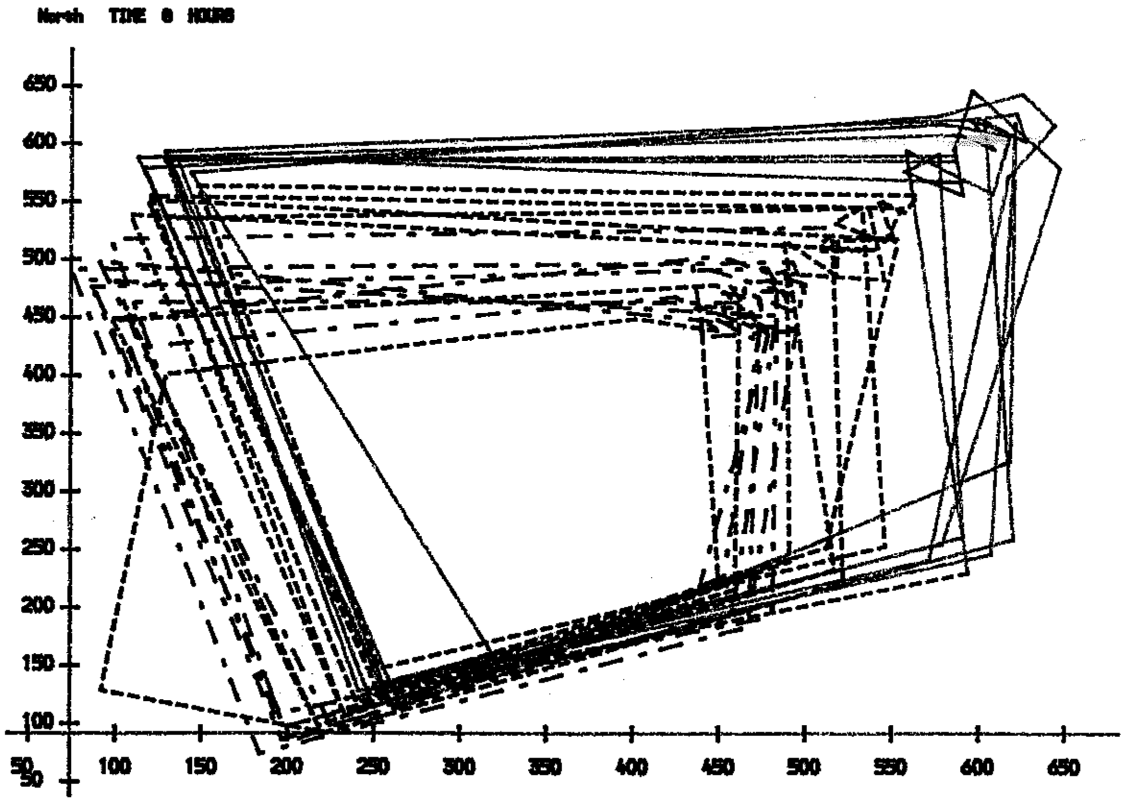


Fig 7.1 Phase Plot of X(t) v X(t-1) for P/B Series



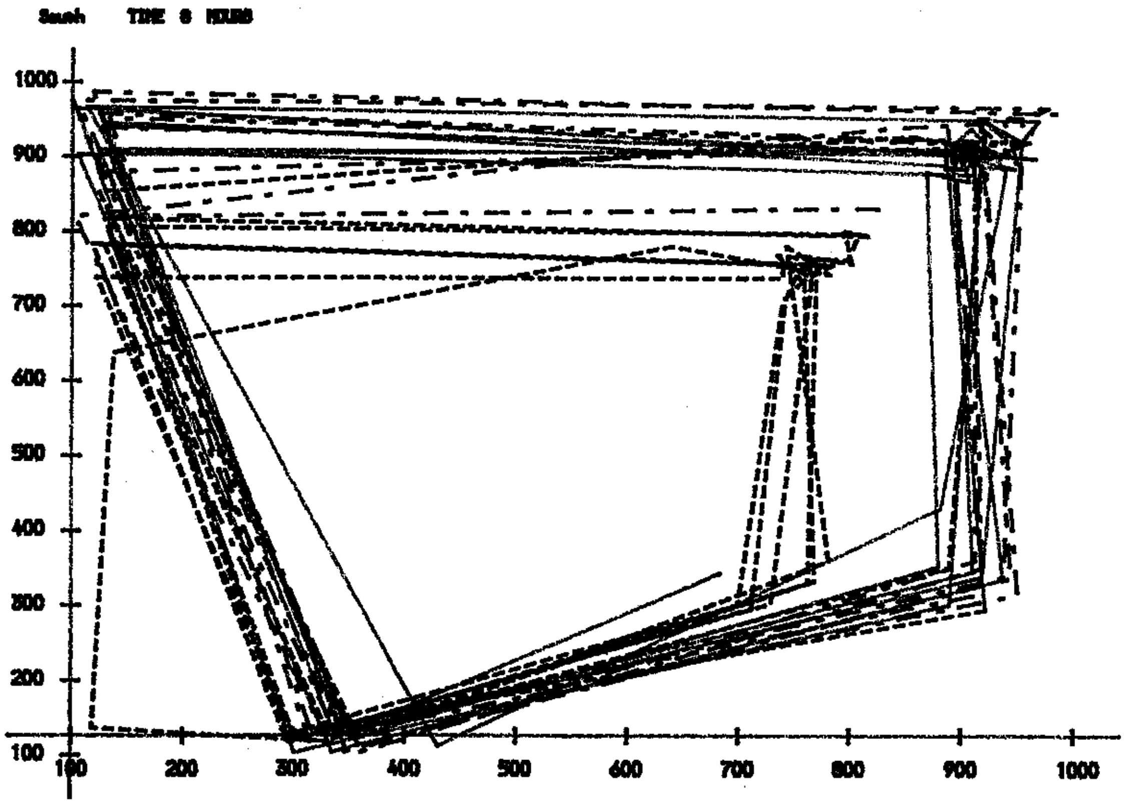


Fig 7.2 Phase Plot of X(t) v X(t-1) for P/B Series

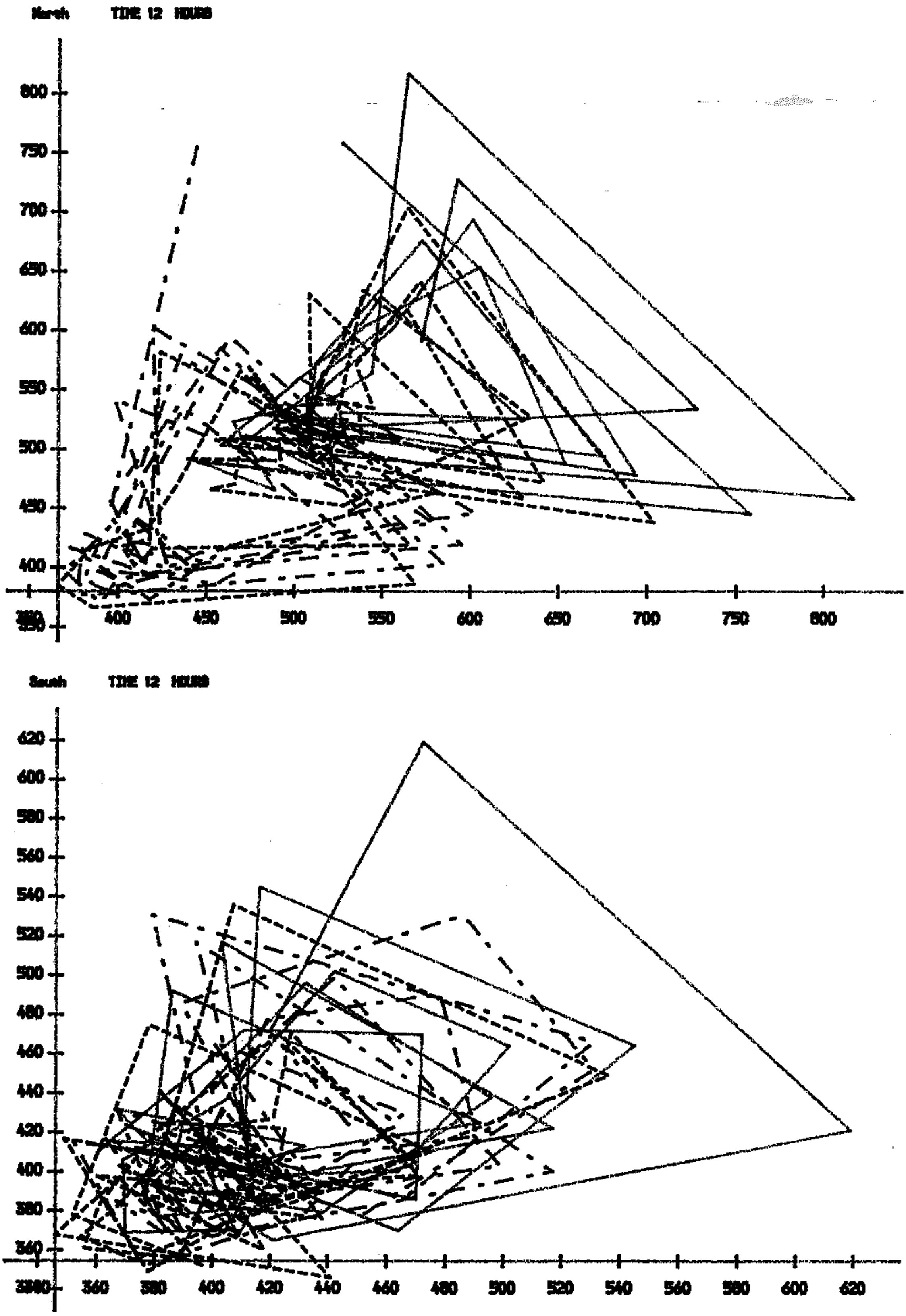
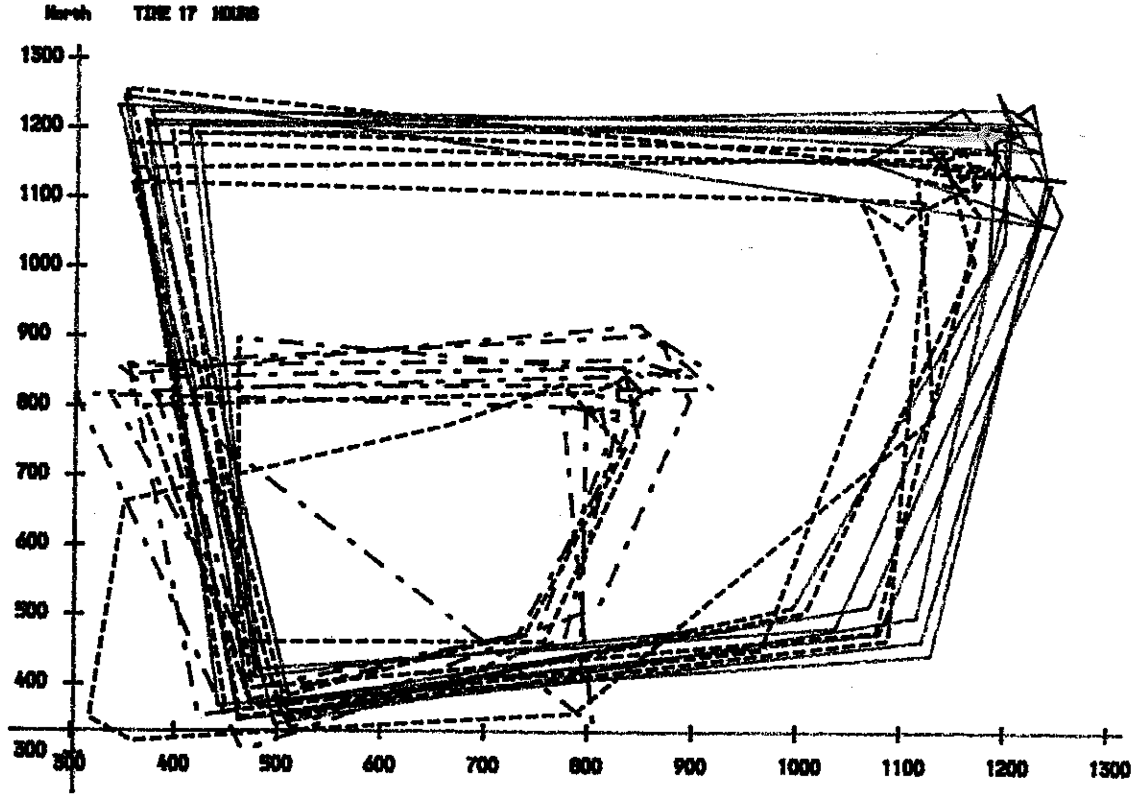


Fig 7.3 Phase Plot of X(t) v X(t-1) for P/B Series



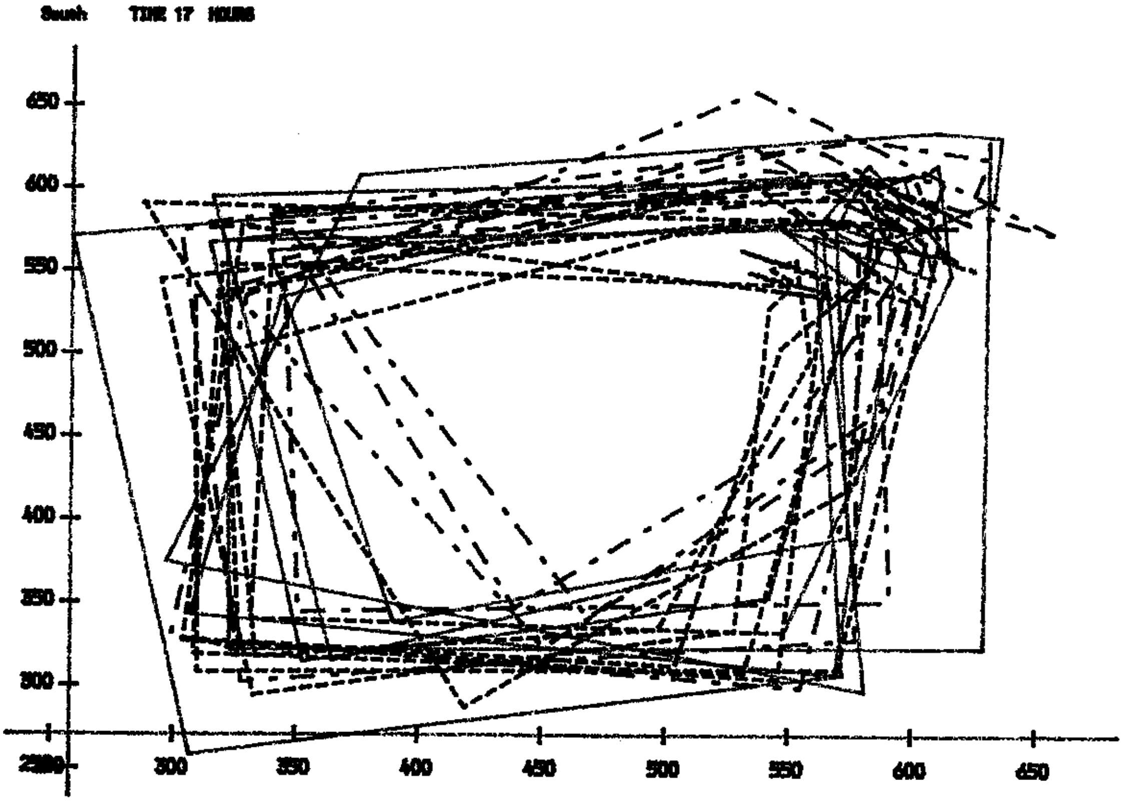


Fig 7.4 Phase Plot of X(t) v X(t-1) for P/B Series