Forecasting **Products**

Research and Development

Forecasting Research
Technical Report No. 317

The Application of an ARIMA Process for Prediction of Crosswind Components at Birmingham Airport

by

S.A. Robson

January 2001

The Met.Office

Excelling *in weather services*

Forecasting Research
Technical Report No. 317

The Application of an ARIMA Process for Prediction of Crosswind Components at Birmingham Airport

By

S.A. Robson

January 2001

The Met Office
Forecasting Products
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

Abstract.

This paper describes attempts to apply an autoregressive, integrated, moving average (ARIMA) process to predict one-minute mean crosswind components at Birmingham airport. The motivation behind this work is the continuation of research into the prediction and avoidance of aircraft wake vortices. It is known that the use of persistence (assuming that the wind speed will remain constant from one minute to the next) gives a useful forecasting technique over such short periods. Previous work, using wind data from Memphis, has demonstrated that a straightforward ARIMA process can reduce the standard errors of a prediction using persistence by around 15% with the possibility of future enhancements increasing this figure to 25%. The prediction techniques used are purely statistical in nature although some meteorological knowledge is applied to some of the outcomes, i.e. results are not applied unless they are sensible in a meteorological context. The method used is that of an ARIMA process of which several models are compared.

The results of the study show that the most appropriate statistical predictor of crosswind components is an ARIMA (1,1,2) model that incorporates the lagged total wind speed. However, further studies should be undertaken to complete the understanding of the application of time series analysis to Birmingham data, as follows:

- Investigate the outliers in the model to assess the causes.
- Determine if the exclusion of outliers from the analysis improves the prediction still further, although some work would be required of the outliers before this should be accepted.
- Attempt to apply the model to predict forecasts of up to twenty minutes ahead.

THE APPLICATION OF AN ARIMA PROCESS FOR PREDICTION OF CROSSWIND COMPONENTS AT BIRMINGHAM AIRPORT.

Contents.

1 INTRODUCTION.

This paper describes the methods and results of attempting to predict one-minute mean crosswind components. It continues on from the work into the prediction and avoidance of aircraft wake vortices [reference 1], that demonstrated the applicability of using statistical prediction techniques in reducing errors in current forecasting techniques such as persistence. That study took a data set from Memphis Airport, the results of which were a reduction in forecasting errors of 15%, with foreseeable improvements of up to 25%. The study concluded by recommending the technique be applied to airport sites elsewhere and, in particular Birmingham, where a suitable data set was available.

Essentially, this work has been carried out to determine if such methods are successful using a data set from a site that is climatologically very different. Furthermore, the sample used is seasonally different. Analysis here concentrates on Birmingham winter data, whereas Memphis data were from summer. A comparison of the two sets of results in the context of error reduction only, has been made. The purpose of this paper is to attempt to demonstrate further that this method could reduce errors and, more importantly, provide a technique by which the errors can be quantified sufficiently well.

As discussed in the previous paper [reference 1], the prediction techniques used are purely statistical in nature. The same meteorological knowledge is applied to the data here to ensure results are not applied unless they are sensible in a meteorological context. Again the method employed is an autoregressive, integrated, moving average (ARIMA) process. No alternative methods are discussed or compared here (see [reference 2]).

2 THE DATA.

The data obtained from Birmingham consist of approximately a year (excluding months August and September) of consecutive one-minute mean wind vectors given as total wind speed in knots and direction in degrees. For the analysis that follows, the wind has been split into headwind and cross wind components.

Figure 1 shows the frequency distribution of the crosswind component at Birmingham for the period analysed. The line imposed on the histogram is a normal distribution having the same mean and variance as the Birmingham data. By inspection, it can be seen that the data have a distribution close to a normal distribution. It should also be noted that the mean of the distribution is not $0ms^{-1}$ but $1ms^{-1}$. Analysis of Memphis data found the mean to be at approximately $-0.5ms^{-1}$. The range of crosswind strength from the two sites is also notable. Values at Birmingham range between $-5ms^{-1}$ to as much as $11ms^{-1}$, compared with $-4ms^{-1}$ to $3ms^{-1}$ at Memphis. This is not important to the data analysis but again highlights the different climatology of the two sites.

Histogram of Crosswind Strength

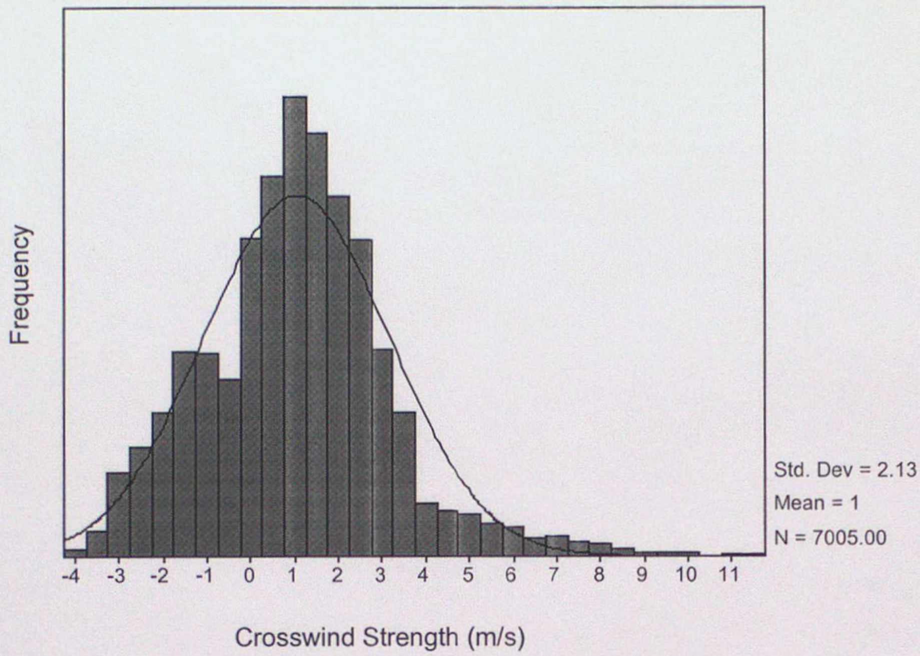

Figure 1: Frequency distribution of crosswind components at Birmingham with a line showing a normal distribution carrying the same mean and variance.

3 TIME SERIES ANALYSIS.

3.1 AN OVERVIEW OF ARIMA MODELS.

The ARIMA process is discussed in reference 1, with a more complete explanation in reference 2. This section provides only a brief summary of the process.

ARIMA is the acronym for AutoRegressive Integrated Moving Average, the three components that are the ARIMA method. In essence, the method uses a combination of weighted previous observations and errors at predicted values, to determine the next value.

The general model is written as ARIMA (p,d,q), where p is the order of autoregression, d is the degree of differencing and q is the order of moving average.

The ARIMA procedure comprises three steps – identification, estimation, and diagnosis. The process is iterative and is complete when the model is satisfactory, or meets its requirements. There is no algorithm to determine the perfect model.

The remainder of this paper is concerned with applying an ARIMA method using the three model building stages and analysing the results.

3.2 USING ARIMA WITH BIRMINGHAM DATA.

3.2.1 IDENTIFICATION.

The method begins by analysis of a plot of the data to determine whether or not it is stationary. This is a necessary condition that needs to be satisfied in order to determine values for the autoregressive and moving average components, p and q respectively.

A stationary series has the same mean and variance throughout.

Figure 2 shows a sample plot of 48 hours of consecutive observations of one-minute average crosswind values taken from Birmingham.

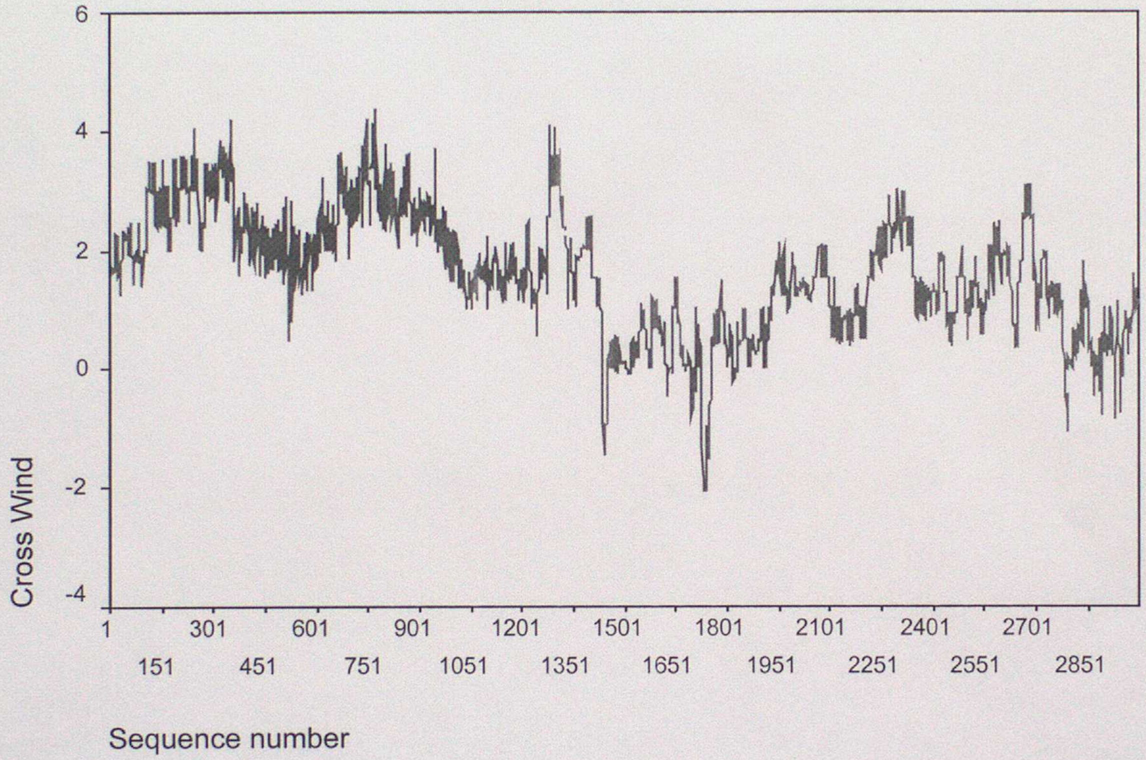

Figure 2: A sample of 48 hours of consecutive crosswind speeds taken from the Birmingham data.

From the data in Figure 2 it is clear that the data are by no means stationary as both the mean and the variance vary quite markedly through the time period shown.

When a series is not stationary, it is common practice in time series analysis to study the differences between consecutive observations rather than the values themselves. Figure 3 shows data taken from the same time period as Figure 2 but plotting the differences between consecutive observations.

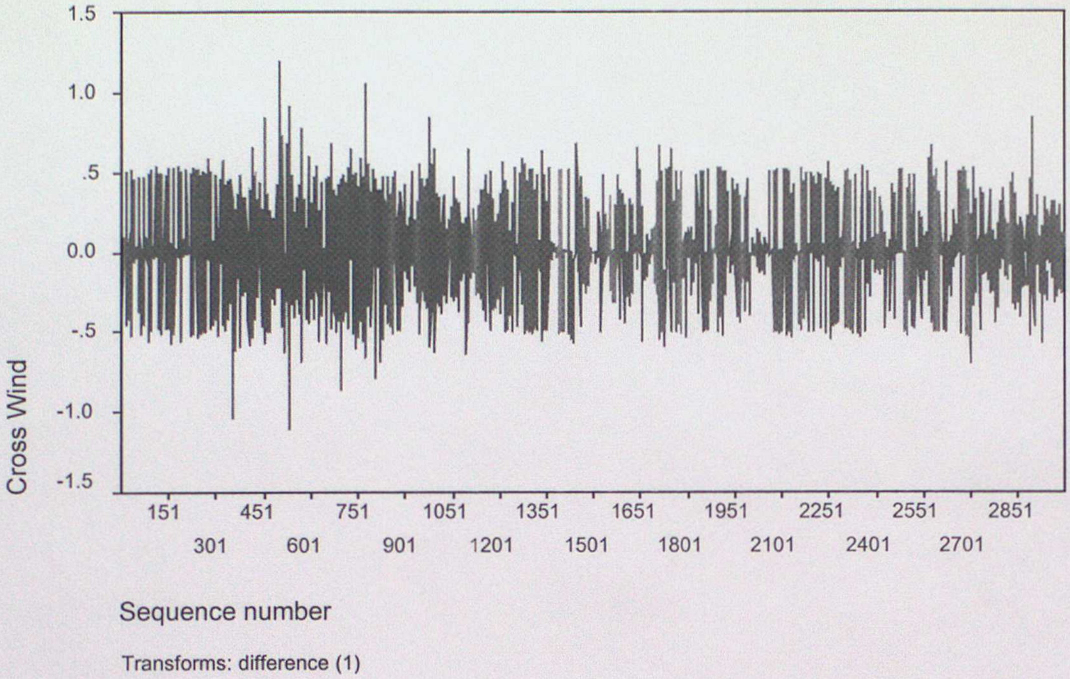

Figure 3: Differences between 48 hours of consecutive crosswind observations taken from Birmingham.

It can be seen from Figure 3 that the sequence of differences does have a constant mean. Furthermore, the mean of the series is very close to zero for this sample. The variance however is not constant and methods to rectify this are discussed later. This result mirrors that of the Memphis data.

Having differenced the series, we now have the second ARIMA model parameter, d. It represents the number of times the series was differenced.

- In order to ascertain the most appropriate combination of p and q two functions of a series, known as the autocorrelation function (ACF) and the partial autocorrelation function (PACF), usually reveal the correct values. The definitions of these complex functions are expressed in simple terms in reference 1.

The ACF and PACF of the time series derived from the crosswinds at Birmingham are given in Figure 4 and Figure 5 respectively. Two important theoretical results are stated below (see reference 2 for details).

- ARIMA (0, 0, n) processes have a slowly decreasing PACF with many significant lags and precisely n significant lags in the ACF.
- ARIMA (n, 0, 0) processes have a slowly decreasing ACF with many significant lags and precisely n significant lags in the PACF.
- Mixed Autoregressive and Moving Average models have more complex ACF and PACF patterns. Suitable model identification requires many iterations of the ARIMA cycle

Figures 4 and 5 show the appropriate functions for the differenced series. Upon first inspection, making a comparison with plots of theoretical ACF and PACF functions for standard ARIMA models, it appears that the most appropriate function to employ in this case is an ARIMA (0,1,1), or perhaps an ARIMA (0,1,2). As in the analysis of Memphis data, from the ACF and the PACF graphs, it seems reasonable to assume that it is a moving average process of small order q that is likely to be the dominant feature of any process used.

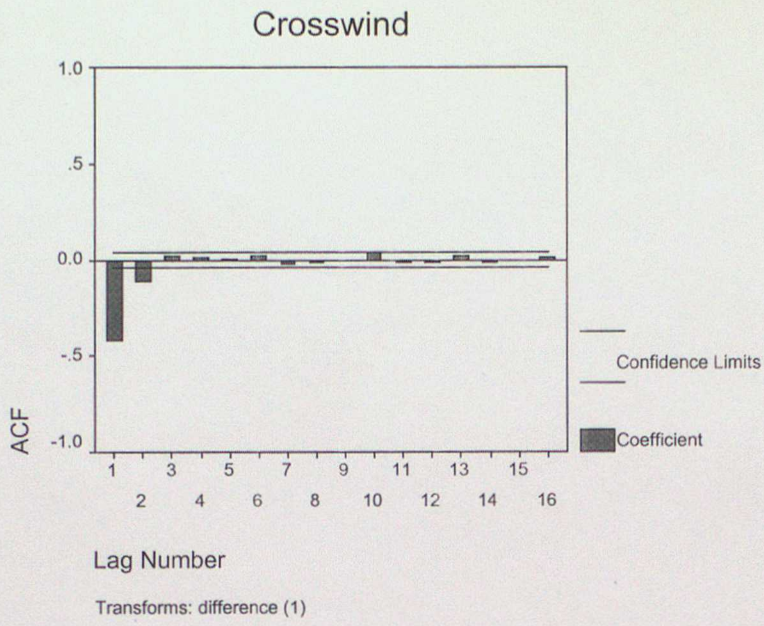

Figure 4:The autocorrelation function of the differenced crosswind readings from Birmingham.

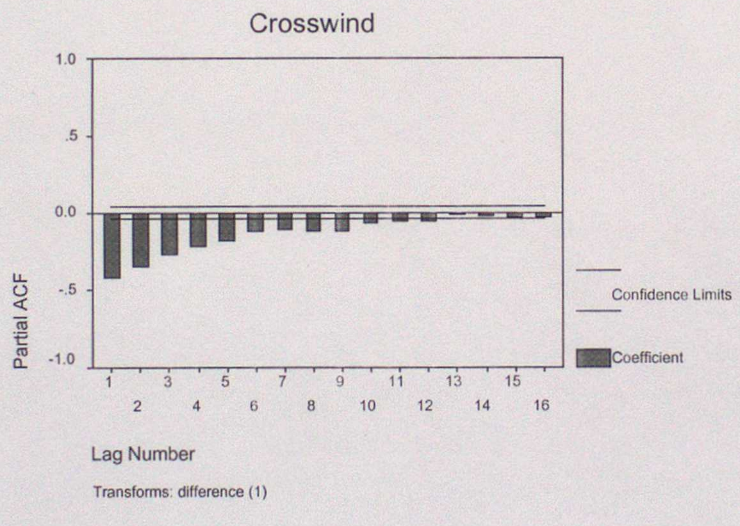

Figure 5:The partial autocorrelation function of the differenced crosswind readings from Birmingham.

3.3 ESTIMATION.

Using the software package SPSS, and in particular the Trends Arima procedure, the model parameters identified are used to estimate the coefficients of the ARIMA model. The model output, which includes the predicted series and the error (residual), is used to complete the final step of the modelling process – diagnosis.

3.4 DIAGNOSIS.

In order to identify the most appropriate model a check was made by calculating the standard errors of all ARIMA (p, 1, q) for $p, d, q \leq 2$. It is logical that the model that gives rise to the lowest standard error will be the most appropriate model for this sequence of data. Furthermore, the ACF and PACF of the error series should not be significantly different from 0 and the residuals should be without pattern. That is, they should be *white noise.* Section 3.6 addresses the analysis of the residuals.

Little difference was found in the standard errors (see section 4, table 1) of the models as described above although the inclusion of an autoregressive term does reduce the standard error by around 0.7%. Consequently, an ARIMA (1, 1, 2) model was considered to be the most appropriate model to use for this data sequence.

3.5 VARIANCE OF THE SEQUENCE.

ARIMA processes assume constant variance throughout a time series. As this is not the case with the Birmingham crosswind data, as in the case of the Memphis data, there is a need to find methods that alleviate the problem. Several potential solutions were presented in reference 1. The results of that study have been applied here. In summary, it was found that the most appropriate application was to re-calculate the appropriate parameters of the model on an hourly basis and to include the lagged total wind speed as a further regressed variable. This application was deemed as being the most appropriate as it produced the lowest standard error in the residual distribution.

3.6 RESIDUALS.

Two aspects of the residual distribution need to be checked in all applications of ARIMA models. Firstly, it is assumed that the residuals follow a normal distribution. Secondly, the ACF and PACF of the residual distribution should demonstrate no significant values, otherwise there would be an indication of some unknown factor that should be included in the model.

Figure 7 shows the residual error distribution taken for an ARIMA (1, 1, 2) model for forecasts one minute ahead. Again, the results mirror that found in reference 1. It can be seen by inspection that this distribution does not match the normal distribution of the same mean and variance that is shown as a line in the figure. However, the actual residual distribution does demonstrate properties that are beneficial. In particular there is a larger than expected proportion of errors that lie in the very low category shown by the peak in the figure. When applying consideration to the appropriateness of the model it is the tails of the distribution that become important as the frequency of large errors is critical to the practical application in question.

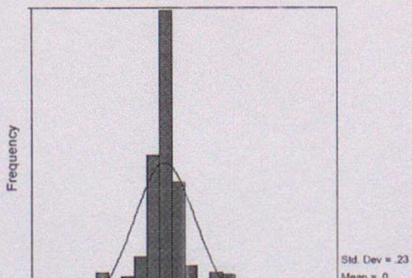

Figure 7: Residual error distribution of an ARIMA (1, 1, 2) model applied to the Birmingham data.

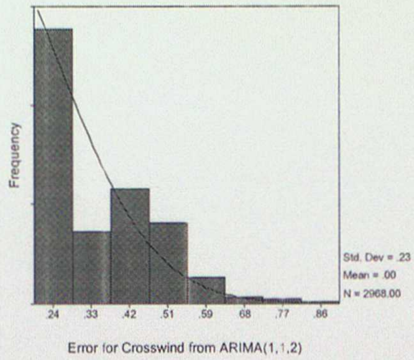

Figure 8:The tail of the residual error distribution shown in figure 7.

Looking in more detail at the tail of the same distribution, in Figure 8, some error values do in fact carry a slightly higher frequency than would be expected by a normal distribution. As in the previous study, it is felt this should be balanced by noting the peak already mentioned around the mean at 0 m/s.

The ACF and PACF of the residual errors were calculated for the chosen model using several sets of different data from Birmingham to ensure no pattern was demonstrated by these functions. Figure 9 shows the autocorrelation function for the ARIMA (1,1,2) residuals. The plot confirms that the residuals are *white noise* as required. Therefore, in general the model appears appropriate but again there are a number of outliers to the model (a discussion regarding the treatment of outliers is given in reference 1 and will not be considered further here).

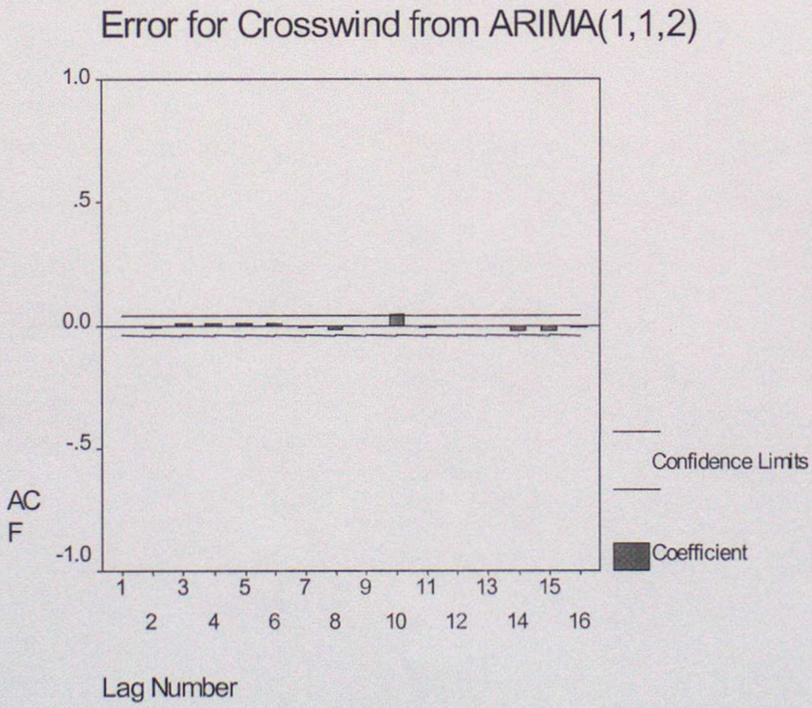

Figure 9:The autocorrelation function of the residuals

In conclusion, the model is considered to be valid for the Birmingham data.

4 RESULTS.

Table 1 below shows the results of reducing sample size and including lagged wind on the Standard error. The results are compared against persistence. The fits are not very good compared with predictions like stock values, which tend to be much more easily followed. In other words there is a far greater level of randomness in wind flow which surely would be as expected. It can be seen that by reducing the sample size we improve the accuracy, this is obvious, as the degree of randomness will be reduced over a shorter period. Beyond this the differences are small. The best solution, i.e. the one with the lowest error, indicates the ARIMA (1,1,2) most suitable.

Model	(0,1,1)	(0,1,2)	(1,1,2)	(1,1,1)	Persistence
Standard error	.23638939	.23370171	.23217337	.2359516	0.23585
Standard error (Reduced Sample + lagged wind)	.22464548	.22578751	.21944986	.22658085	

Table 1:A comparison of Standard Errors for ARIMA models and Persistence

5 CONCLUSIONS.

The most appropriate statistical predictor of crosswind components at minute intervals was found to be an ARIMA (1, 1, 2) model that incorporates the lagged total wind speed and a regressed variable. In addition, assessment of the suitability of using standard error as the judge is also valid as the extremes suggested by use of the normal distribution turn out to be slightly pessimistic.

The standard errors for this model were around 90% of the errors produced by a persistence forecast.

6 FURTHER WORK.

In addition to the points highlighted in reference 1, the following studies should be undertaken in order to complete the understanding of the application of time series analysis to Birmingham data.

- Investigate the outliers in the model to assess the causes.
- Determine if the exclusion of outliers from the analysis improves the prediction still further, although some work would be required of the outliers before this should be accepted.
- Attempt to apply the model to predict forecasts of up to twenty minutes ahead (see reference 1).

7 REFERENCES.

1 Halsey, N.G.J., 1998: The prediction of crosswind component over very short periods in the context of wake vortex avoidance.

FR Tech Report 244 (Available from Met Office Library.)

2 Box, G.E.P. and Jenkins, G.M. 1976: Time series analysis, forecasting and control.

Holden Day.