UNIVERSITY OF NOTTINGHAM

DEPARTMENT OF MINING ENGINEERING

A STATISTICAL ANALYSIS OF MONITORED DATA FOR METHANE PREDICTION

by

Darron William Dixon, BEng (Hons).

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

May 1992

For thought;

T. S. Elliot, Murder in the Cathedral.

However certain our expectation the moment forseen may be unexpected when it arrives.

A. A. Milne, House at Pooh Corner.

" Supposing a tree fell down, Pooh, when we were underneath it ? "
" Supposing it didn't ", said Pooh after careful thought.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF PLATES

ABSTRACT

This research describes an investigation into the application of a statistical method for the prediction of methane concentration in longwall coal districts. An important and necessary part of the research was the acquiring of representative mine environmental and coal production data and a number of shortcomings were identified in this area. The monitored data was used to build univariate time series models of general air body methane concentration, air velocity, barometric pressure, coal production and methane drainage variables of varying timescales according to the Box-Jenkins method of time series analysis. The univariate models were used to identify causal relationships between methane concentration and its explanatory variables. Coal production was found to be the dominant variable in the determination of the quantity of methane emitted and where appropriate, multivariate time series models were built in which expressions for methane concentration in terms of coal production were obtained. Forecasts of methane concentration values were generated from both univariate and multivariate models and a comparison was made of their forecasting capabilities. Finally, suggestions were made as to the potential use of time series models for application to mining process control.

CHAPTER ONE

INTRODUCTION

Currently the UK coal mining industry is undergoing a period of rapid change due to economic pressures. British Coal is striving to produce coal at a competitive price and quality to attract the recently privatised electricity generating companies and remain competitive in comparison to world coal prices. The realities of this situation are the possibility of more mine closures and the need for all mines to increase production and to lower costs. One way of increasing production would be to maximize production over a set period of time. To determine how this could be achieved involves the consideration of a number of factors which not only concern the ventilation engineer, but also need to encompass many other branches of mining engineering.

As coal faces produce larger amounts of coal the ventilation system is having to deal with an increased load of pollutants. These include dust, heat and methane. The importance of methane as a dangerous underground pollutant cannot be overstated. The recent tragic mining disaster at Incirharmani in Turkey in which over 250 people were killed is a reminder of the catastrophic potential of excessive methane concentrations [1]. It is explosive, asphyxiating and can halt coal production. There are essentially two methods of dealing with methane emission and these are generally used in conjunction with each other. The first is to dilute the gas that appears in the ventilating airstream to acceptable levels. The second and that which is now practised widely in the UK is methane drainage. The rate of methane emission is determined by the production cycle that results in a cyclical pattern of methane emission which to some extent is also influenced by changes in barometric pressure.

To combat the danger from excessive methane emission the underground environment is monitored. Underground environmental monitoring has developed steadily over the past 25 years and all collieries now have sophisticated environmental monitoring systems. The primary aim of such systems is ensure that underground conditions are safe and this is achieved by the reliable monitoring of environmental parameters at strategic points underground. The monitored information is relayed to the colliery control room where it is displayed usually as a spot value or as a simple graph showing the recent behaviour of a parameter. Beyond these limited applications the monitored environmental data is not

utilized.

Two key points have been mentioned so far. The first is an increased amount of methane emission due to higher levels of production and the second is an under-utilization of monitored information. This research attempts to link these two factors in order to provide a solution to the problem of increased methane emission and its production limiting potential by making use of existing environmental data. In short, the environmental data is used to predict future levels of methane emission. In this thesis research is described which attempts to predict methane emission statistically and whether such predictions could be of use in a coal production/methane emission control system to maximize coal production.

The benefits of such research are numerous. Safety cannot fail to be enhanced by an accurate knowledge of future methane levels. Any increase in the use of already monitored information will prove to management that such systems are indispensable, especially if it is shown that the information can be put to good use without the need for further expenditure. The mine managements concern is minimum cost coal which equates to optimized production. Thus, for a method of methane prediction to be effective it should be able to fulfil these criteria, i.e. require small expenditure and be an aid to optimize production.

Past methods of methane prediction have all needed some degree of physical knowledge of the prevailing mining conditions, such as the determination of strata gas contents and strata stress/permeability relationships. These factors are fundamental to many prediction techniques and are expensive and often difficult to determine. In this research a purely statistical method for methane prediction is proposed and investigated that differs from previous attempts in that it only requires information that is already in existence.

By analysing the monitored data it is hoped that the research will provide a clearer insight into the pattern of methane emission in longwall coal mining and indicate which are the most important parameters to consider. The statistical analysis takes the form of time series analysis and a time series technique is used to build models for methane emission with predictive capability. Previous work into methane prediction has been centred on the development of long-term prediction methods which provide the ventilation engineer

with an expected rate of methane emission. Such predictions are quite invaluable in deciding the ventilation quantities necessary to dilute methane to acceptable levels but have limited value from an immediate control viewpoint. The time series models have the ability to be used in a control strategy, whereby they can provide a simple warning of high levels of methane being encountered or, predict future levels according to changes in the rate of production. Thus, it is seen that a statistical technique may be of use in a fully automatic computer controlled coal face where the aim is to sustain coal production and keep methane concentration to acceptable levels.

Chapter 2 presents a concise account of the origin and occurrence of methane in mining while chapter 3 reviews previous research into methane prediction. The chapter highlights the reliance of present methods of methane prediction on a knowledge of physical parameters. Chapter 4 provides an introduction to the method of time series analysis and outlines the steps that are followed in order to build an appropriate time series model. A key feature of the time series approach is the ability to develop a model, a priori, which means that no initial knowledge of the data is required for a model to be built. Representative time series models can only be built from accurate data, the acquisition of which is of paramount importance. This is the subject of chapter 5 which details the efforts made to obtain data from Thoresby colliery and the difficulties incurred while doing so.

Chapter 6 describes the development of univariate time series models for methane concentration, indicator variables and methane drainage range variables. Models are built for different data time intervals and those for methane concentration are used for forecasting purposes. The univariate models are then analysed in chapter 7 to investigate the effects of the indicator variables on methane concentration, and where appropriate, multivariate models for methane concentration are built. Forecasts are also obtained from the multivariate models for methane concentration and these are compared to the univariate forecasts in chapter 8. Chapter 8 includes a discussion on the relative merits of the forecasting models and their suitability to mining process control.

The primary aim of the research is to investigate if representative time series models can be built from underground mine environmental and production data and if such models are capable of producing accurate predictions. In chapter 9, the main findings are summarized and recommendations are made to direct future research into this topic.

CHAPTER TWO

METHANE: ITS ORIGIN AND OCCURRENCE IN MINE WORKINGS

2.1 Introduction

The emission of methane from coal seams and surrounding strata into mine workings has long been of concern to the mining engineer. In modern times innovative mining techniques have been developed to enable greater quantities of coal to be mined from ever increasing depths and these have resulted in higher levels of methane emission. Methane is a dangerous gas that contributes a serious threat to safety and the maintaining of high levels of production. The development and application of a reliable and accurate method of prediction to ensure that these two criteria are met relies on a knowledge of the mechanisms of methane generation, storage and release and this chapter is intended as a general review of methane: its origin and occurrence in mine workings.

2.2 The Properties of Methane

Methane is a colourless, odourless and tasteless gas, with a relative density of 0.554 relative to air. At 0 °C and 750 mm Hg pressure, methane has a density of 0.716 kg/m^3 [2]. As a result of its low density it accumulates in the high places of mine workings where there is little or no air movement. Chemically, methane is fairly inert but the principal danger of methane in mining is its explosible character. It is explosive at concentrations between 5% and 15% in air, and at a concentration in the middle of this range (9.6%) the air/methane mixture reaches its maximum explosibility.

Methane will ignite or burn depending on the composition of air with which it is mixed. A low oxygen content or a high inert concentration will make ignition more difficult or even impossible. Investigations have shown that methane ceases to ignite at oxygen concentrations below 12% at standard temperatures and pressure [3]. Changes in atmospheric pressure can also influence the conditions required for flame propagation. The limits of flammability of mixtures of methane and air are illustrated by Figure 2.1

2.3 The Origin of Methane

Methane is a by-product of the diagenesis of carbonaceous material, originating mainly from vegetation, which eventually results in the formation of coal or oil. The process of conversion of vegetation into coal is termed 'coalification'. This process also produces water and carbon dioxide. The quantities of migratory products formed due to coalification are large. Typically, they can be in the order of 0.75 tonne of water, 600 m^3 of carbon dioxide and 600 m^3 methane, per tonne of matured coal at normal temperatures and pressures [4]. Most of these products, especially water which can contain gases in solution, are lost to the surrounding strata. The coalification process is a continuum starting from peat, passing through the lignite and brown coal stage into the high, medium then low-volatile bituminous region and culminating with anthracite.

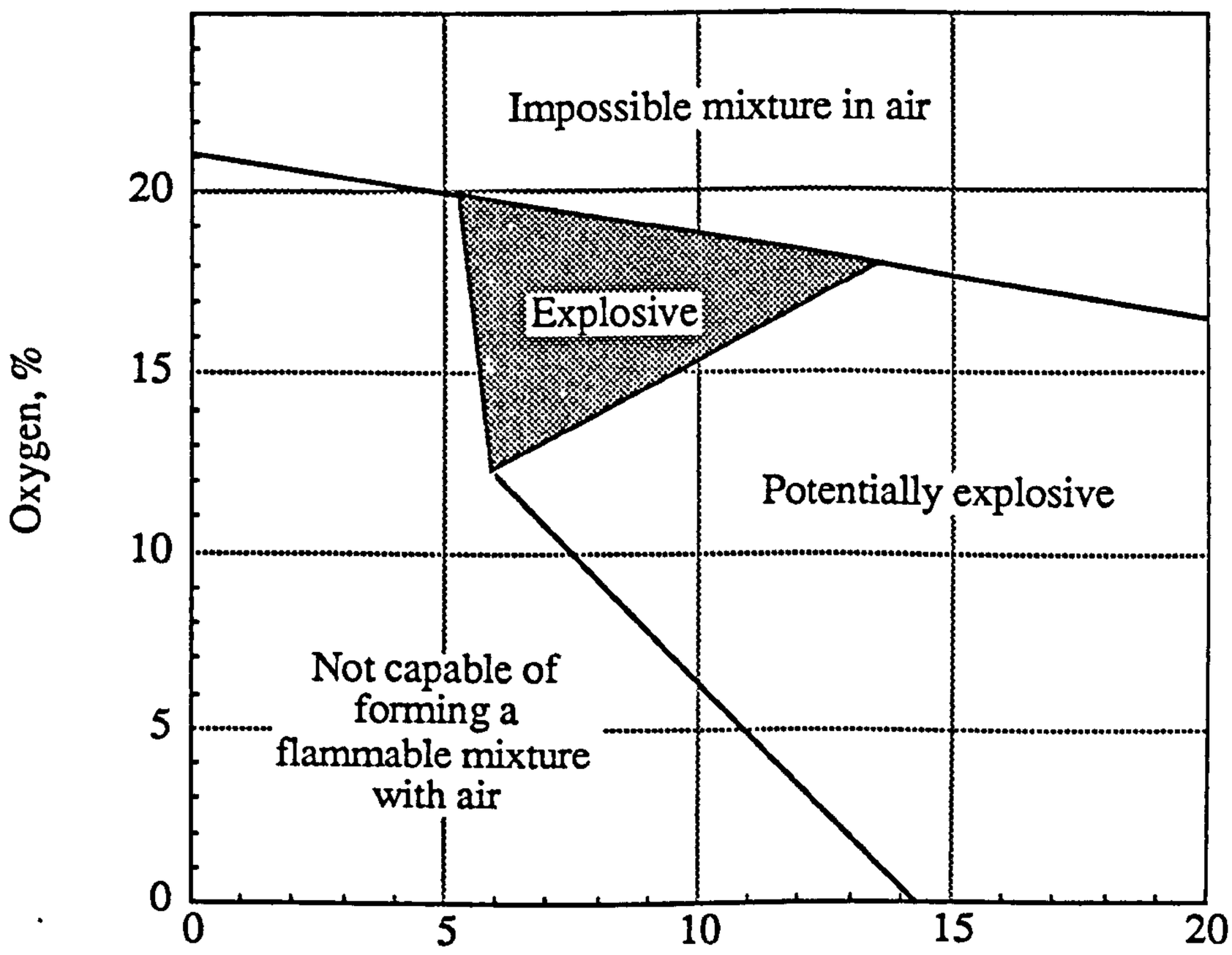


Figure 2.1 Explosibility Curve for Methane, (after Coward and Jones [3]).

Maturation of the coal-like material continues because of pressure from cover loads and geothermal heat . Any fluids within the coal seam are held at a pressure which is very

close to the hydrostatic pressure. Although carbon dioxide and methane are generated in similar proportions it is unusual to find a coal seam which contains comparable amounts of carbon dioxide to methane. Coal has a stronger affinity for carbon dioxide than methane but as coal matures, carbon dioxide is converted to methane.

Coal seam gas is refered to as 'firedamp' within the UK coal mining industry. Firedamp is a traditional name and is mostly composed of >90% methane with lesser amounts of higher hydrocarbons, carbon dioxide, nitrogen, oxygen, hydrogen and helium. Coal seams which contain a high percentage of higher hydrocarbons are thought to have acquired them from extraneous sources. These gases are associated with petroleum derivatives and were probably retained by coal while migrating.

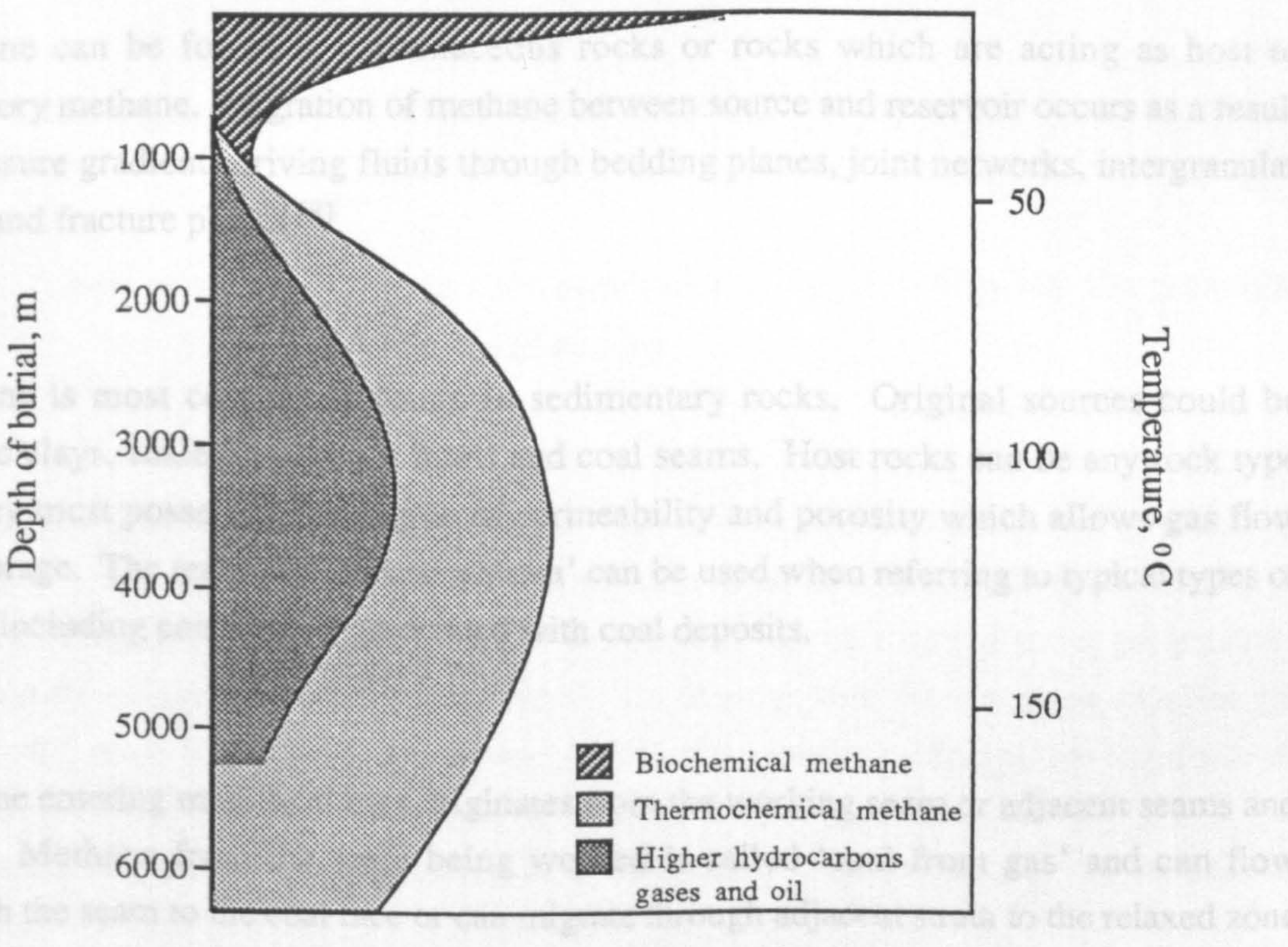


Figure 2.2 Relation Between Methane Generation, Depth of Burial and Temperature, (after Hedburg [7]).

Methane is first generated at shallow depths by biogenic processes. Such processes can only occur in reducing environments which are oxygen free and able to support the actions of anaerobic bacteria on carbonaceous material. It is estimated that biogenic methane accounts for 20% of the worlds gas reserves and lie at depths up to and possibly beyond 1000 m [5]. As burial and compaction progresses, a greater proportion of methane is generated by non-biogenic thermochemical degradation. At depths in excess of 1000 m biogenic processes cease to produce methane and thermogenic degradation becomes the principal methane generating mechanism [6]. A diagrammatic representation of the relationship between methane generation, depth of burial and temperature is shown in Figure 2.2.

2.4 Sources of Methane Emission

Methane can be found in carbonaceous rocks or rocks which are acting as host to migratory methane. Migration of methane between source and reservoir occurs as a result of pressure gradients driving fluids through bedding planes, joint networks, intergranular pores and fracture planes [8].

Methane is most commonly found in sedimentary rocks. Original sources could be organic clays, shales, carbonate muds and coal seams. Host rocks can be any rock type but they must possess some degree of permeability and porosity which allows gas flow and storage. The term 'coal measure strata' can be used when referring to typical types of strata (including coal seams) associated with coal deposits.

Methane entering mine workings originates from the working seam or adjacent seams and strata. Methane from the seam being worked is called 'coal front gas' and can flow through the seam to the coal face or can migrate through adjacent strata to the relaxed zone behind the moving face. Methane originating from other coal measure strata above and below mine workings migrates into the roadways and is termed 'strata gas' [9, 10]. Methane which is desorbed from the coal before it reaches the face may be released when the coal is cut. Any remaining gas will gradually desorb during transportation but may not have completely desorbed by the time the coal leaves the mine [11, 12].

The main contribution to methane emission is from the surrounding strata and studies have indicated that this can exceed 80% of the total gas emission [13]. For a comprehensive simulation and prediction of methane emission into mine workings, the contributions of coal front gas, strata gas and that from conveyed coal all need to be included.

2.5 The Retention of Methane in Coal

Methane is present in coal as adsorbed gas on the internal surface of the pores and as free gas occupying spaces within the internal structure of coal. The process of methane retention is called sorption and when the gas becomes free it is said to be desorbed [12, 14]. Sorption is sub-divided into two basic categories;

1. **Adsorption** : a reversible surface effect whereby one substance is physically held onto the surface of another,
2. **Absorption** : the uniform penetration of one substance into the molecular structure of another.

Most of the gas within coal is adsorbed onto the surface of coal pores in a mono-molecular layer. Absorption of methane is not considered to represent a significant role in the flow of methane from coal [15, 16]. It has been estimated that at a strata gas pressure of 20 bar the quantity of adsorbed methane is ten times greater than methane as a free gas in some US coals [17]. The high methane adsorption capacity of coal is due mainly to its very large internal surface area which could be as high as ~200 m^2/g [15]. A diagrammatic representation showing how methane is retained in coal can be seen in Figure 2.3

A number of models have been proposed to describe the process of adsorption onto the surface of coal. Langmuir [18] relates the quantity of gas adsorbed per unit mass of solid to the partial vapour pressure of the gas and describes the mono-molecular layer adsorption of gases with the following equation:

$$V = \frac{V_m\, b'\, P}{1 + b'P} \qquad [2.1]$$

where

V	=	volume of gas adsorbed,
V_m	=	maximum volume of gas adsorbable,
b'	=	desorption coefficient,
P	=	gas pressure.

Langmuir's theory gives the fraction of the adsorbent surface that is covered by molecules of adsorbed gas. If the maximum sorption capacity of the surface is known, then the volume of gas that can be adsorbed may be determined.

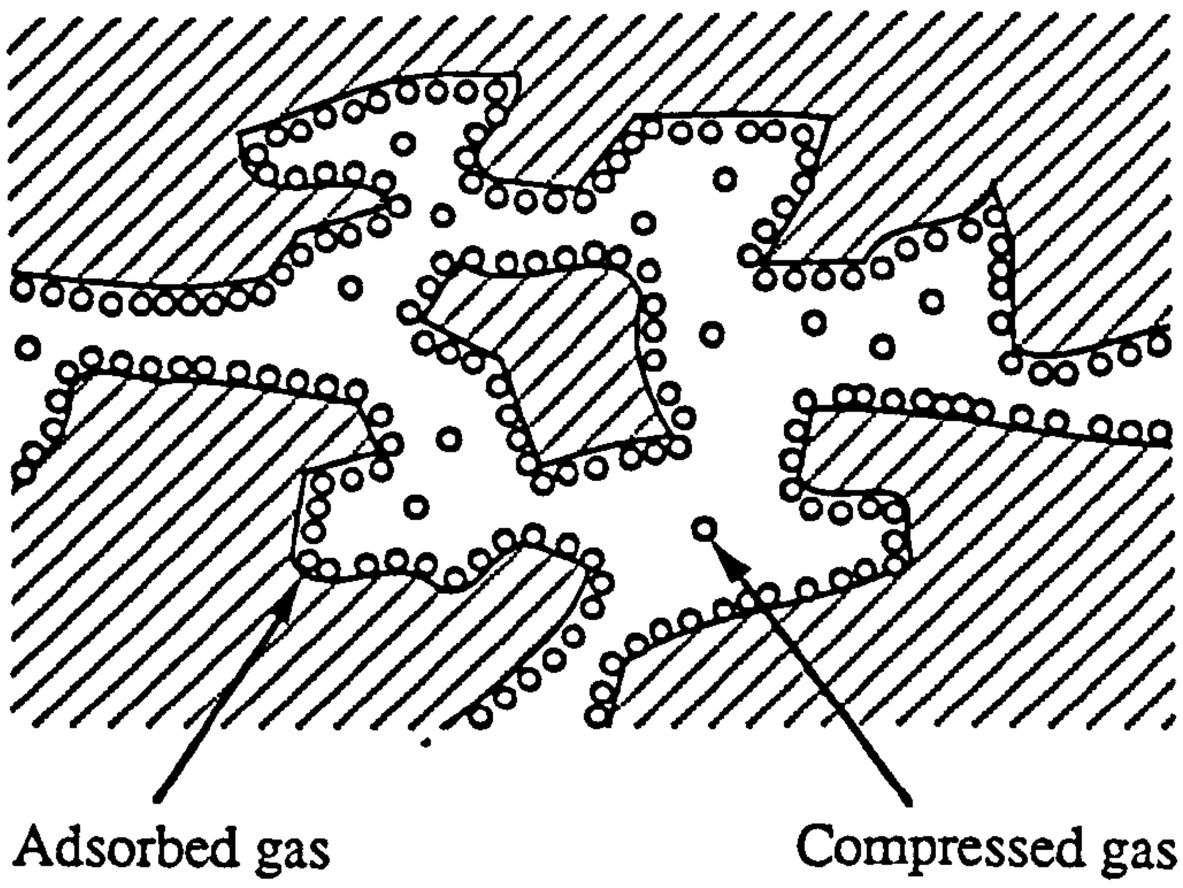


Figure 2.3 Adsorbed and Free Methane Within Coal.

Langmuir's theory only considers mono-molecular layer adsorption, and this was extended to consider multi-layer adsorption by Brunnauer, Emmett and Teller [19]. Keen [20], however, states that coal does not show a secondary layer of adsorption under normal gas pressures of less than 50 bar, and therefore, Langmuir's equation is considered to be sufficiently accurate to apply to the adsorption process.

As rank increases so does the adsorptive capacity and methane content of coal [21]. The US Bureau of Mines has found that for US coals the ratio of fixed carbon to volatile matter correlated well with the gas content of coal [22]. The critical parameter affecting the adsorption of methane by coal is the in-situ gas pressure. The greater the gas pressure the greater the adsorptive capacity of coal. This relationship is shown in Figure 2.4. Moisture content also has an effect on the adsorptive capacity of coal and is mainly related to the oxygen content of coal [23]. Strong interaction between polar water molecules and the surfaces of oxygen complexes hold water in pore spaces in an adsorbed state. Lower rank coals contain oxygen which is lost as the coalification process continues either by migration or in the form of carbon dioxide or water, resulting in decreased water adsorptive capacity.

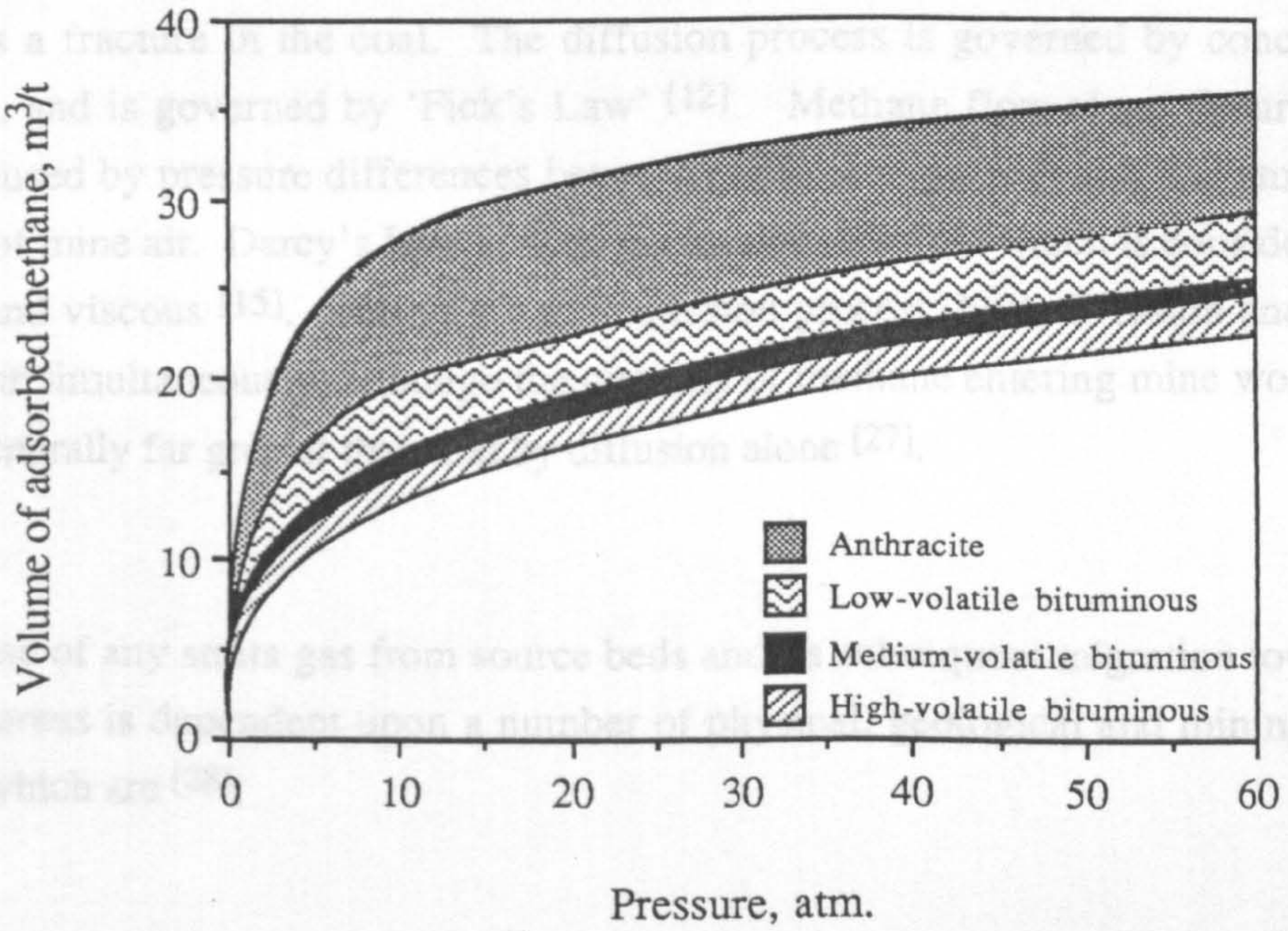


Figure 2.4 Variation of Methane Adsorption Isotherm with Coal Rank at 0 °C, (after Kim [22]).

2.6 The Release and Flow of Methane into Mine Workings

Methane present within coal and other source rocks will be at equilibrium with the local strata pressures. Underground mining activities cause disturbance to strata and upset this

equilibrium of adsorbed gas. These activities also cause relaxation of strata and the resultant fracturing provides flow paths for the gas to migrate into mine workings. Before strata is disturbed, methane is held in its source at high pressure. Mine workings, however, contains air at near atmospheric pressure and provide a 'pressure sink' into which methane flows from the zone of gas emission surrounding the working [24, 25]. The flow of methane is considered to be a two-step process [15, 26];

1. **diffusion** : through the micropore structure of the coal,

2. **flow** : along interconnected fissures in the coal bed.

After methane has desorbed it moves as a free gas by diffusion through solid coal until it intercepts a fracture in the coal. The diffusion process is governed by concentration gradients, and is governed by 'Fick's Law' [12]. Methane flow along fissures within coal is caused by pressure differences between the in-situ gas pressure and atmospheric pressure of mine air. Darcy's Law is used to describe this flow which is considered to be laminar and viscous [15]. During the gas emission process, both diffusion and laminar flow occur simultaneously, although the quantity of methane entering mine workings by flow is generally far greater than that by diffusion alone [27].

The release of any strata gas from source beds and its subsequent migration towards the working areas is dependent upon a number of physical, geological and mining factors, some of which are [28];

1. the gas content and the thickness of the coal seam,
2. the pressure at which the gas is held,
3. the permeability of the virgin coal seam and the surrounding strata,
4. the modifications of coal seam and strata permeabilities by stress changes caused by mining activities,
5. the subsidence of the overlying rock,
6. the method of mining and roof control,
7. the method of ventilation,
8. depth of working,
9. presence of other source beds in the vicinity of the seam worked,

10. barometric pressure,
11. rate of production.

Permeability is considered to be of primary importance in the emission of methane into mine workings. The release of methane from coal and its subsequent flow through strata towards mine workings is controlled mainly by the permeability of the formations concerned. Stress disturbances created by mining activities affect the permeability of both the seam being worked and that of adjacent strata and therefore, determine the nature of methane emission [29].

2.7 Conclusion

Methane is a by-product of the coalification process. It is found in coal and other rocks within which it is retained as both adsorbed and free gas, the predominat storage medium being through adsorption. Following mining operations, strata disturbances result in methane desorption and create breaks and fissures through which methane can flow into mine workings.

CHAPTER THREE

A REVIEW OF METHODS FOR METHANE PREDICTION

3.1 Introduction

The objective of methane prediction is to forecast future levels of methane emission or concentration. The availibility of accurate forecasts to the ventilation engineer can enable him to make important decisions concerning the ventilation strategy of the mine. The planning of a mine determines its viability, and a mine's efficiency can depend upon a good ventilation plan. A primary objective for the ventilation engineer is to ensure that air quantities are sufficient to meet statutory requirements for the dilution of methane in air. Mining Law in the UK requires that the methane concentration in the general air body must be less than 2% for men to work and must not exceed 1.25% where electrical power is in use [30].

The development of previous methane prediction techniques has required an understanding of the occurrence of methane in coal measures strata and mechanisms influencing its release and flow into mine workings. However, the emission of methane into mine workings is a very complex combination of processes which make accurate prediction difficult.

Methane can be emitted from a number of sources and the gas which appears in the ventilating air will normally be a mixture of strata gas, coal-front gas and gas from conveyed coal. According to the method of methane drainage employed, the gas captured will be a combination of strata gas, and seam gas. Of the total amount of gas released, strata gas can account for greater than 80%.while emission from the worked seam is normally never above 20% [13].

This chapter will discuss the various mathematical approaches that have been previously followed to develop a methane prediction technique and illustrate them with examples. It is not intended as an exhaustive review of the currently available techniques, but concentrates mainly on the approaches which have previously been used to develop them.

A methane prediction technique can be arrived at by following three main approaches, namely;

1. **empirical** : results in models based on simple mathematical representations of observable physical phenomenon,

2. **numerical** : using numerical solutions of either empirical or theoretical relationships to predict methane emission,

3. **statistical** : predicts methane emission purely by statistical analysis of observed data.

3.2 Empirical Approaches to Methane Prediction

Many researchers have undertaken studies into the area of methane prediction using empirical methods [12]. Such methods have mainly been developed in Belgium, France, FRG, Poland, UK, USA and the USSR. Although the prediction techniques vary quite considerably between countries they all consider the same basic parameters in their particular prediction method. These are;

1. the stratigraphy above and below the worked seam,
2. the desorbable gas content (m^3/t) of the worked coal seam and, if possible, of adjacent seams and strata,
3. the zone of gas emission in both the roof and floor strata,
4. the degree of gas emission from adjacent seams and strata.

3.2.1 Stratigraphy Above and Below the Worked Seam

Originally methane was formed by the coalification process together with the coal material and possibly contained within the coal seam. As geological time passed methane could migrate and be retained in neighbouring strata such as sandstones, shales, mudstones. To determine whether methane is present in rocks other than coal, core samples of all the

strata in the stratigraphic sequence need to be analysed.

Rocks with very low permeability and porosity, situated either above or below the working coal seam, can influence methane flow to mine workings. For example, a competent and elastic sandstone may cause less stress to be transmitted to methane sources. This prevents fracturing of strata and hence can affect the release and flow of methane. Low permeability rocks may also form a physical barrier within the gas emission zone through which methane cannot flow. Conversely, strata that is strong but brittle may suddenly fail resulting in a sudden 'outburst' of gas emission.

3.2.2 Desorbable Gas Content (m^3/t) of Strata in the Mining Zone

The desorbable gas content is the total quantity of methane present in a coal seam (or strata) before it is disturbed by the mining process. An accurate knowledge of the in-situ methane content of any potentially gas bearing rock is an important criterium for all of the present methods of methane prediction. The measurement of desorbable gas content can be achieved by 'direct' and 'indirect' methods. In the direct method, a coal sample is taken from underground and the gas content measured in the laboratory. The indirect methods determine the methane content from measurement of the in-situ gas pressure and a knowledge of the relevant adsorption isotherm of the coal.

3.2.2.1 Direct Methods of Determing Desorbable Gas Content

Direct methods of obtaining the desorbable gas content of coal usually consist of coring a sample, enclosing it as quickly as possible in a sealed vessel and then measuring the gas release over a period of time. Measurement of gas emission continues until the gas emitted from the sample is less than 0.05 cm^3/g per day for 5 consecutive days in the US Bureau of Mine's method [31]. At this point the sample will still contain some gas and this is found by crushing the sample at atmospheric pressure and measuring the gas release. A correction also has to be made for the gas lost prior to sealing the sample. In estimating the lost gas the time taken to transfer the sample from the drill case to the sealed container must be accurately known. The quantity of lost gas is then found by extrapolating the gas emission results for this time period.

The total gas content of the sample Q is obtained using the relationship:

$$Q = Q_1 + Q_2 + Q_3 \qquad m^3/t \qquad [3.1]$$

where

Q_1 = gas lost between drilling of the core and transfer of the core from the drill hole to the sealed container,

Q_2 = gas liberated from the core after placing it in the sealed container,

Q_3 = gas liberated when the coal sample is crushed.

The MRDE method differs on two points. Firstly, the ash content of the coal is taken into account. Secondly, Dunmore [12] observed that large blocks of coal can completely degas and that greater attention should be placed on the quantity of gas left after crushing at atmospheric pressure. Therefore, after crushing, the coal sample is allowed to desorb until it reaches zero methane partial pressure.

Table 3.1 shows the result of research undertaken by the USBM which compares gas content values determined by indirect and direct methods for a number of coal seams. The table shows how the value determined by either method can significantly differ and hence affect the results when used in a methane prediction technique. Neither method was found to produce gas content values which were consistently higher or lower.

3.2.2.2 Indirect Methods of Determing Desorbable Gas Content

Indirect methods allow the gas content to be determined from the relationship between strata gas pressure and gas content. There are several indirect methods for determining the gas content of strata. One of the simplest methods is that suggested by Kim [22]. Kim described the gas content of strata with the following relationship:

$$Q = \frac{(100 - \%\ \text{moisture} - \%\ \text{ash})}{100} \quad (0.75) \qquad [3.2]$$

$$x \quad [\ k\,(0.096h)^n - 0.14\,(\frac{1.8h}{100} + 1)\ ]$$

where

Q = gas content m^3/t,

h = depth m,

k, n = constants.

The value of k varies between 5.7 and 20 and n varies between 0.31 and 0.12. Figure 3.1 shows the relationship between k, n and the ratio of fixed carbon to the volatile matter of coal on an ash-free dry basis. Kim reports that the equation seems to produce adequate results for strata at deeper depths (higher pressures) but fails at shallower depths. Kim considers that the value of gas content determined from adsorption data is accurate to within ±30% of the direct method which is itself subject to possible errors of ±30%.

Location	Coalbed	Indirect method m^3/t (STP)	Direct method m^3/t (STP)
Vesta No. 5	Pittsburgh	3.4	2.6
Loveridge	Pittsburgh	5.2	5.8, 5.8
Howe	Hartshorne	10.5	11.1
Beatrice	Pocahontas No. 6	13.5	12.1
Inland	Illinois No. 6	2.7	1.7
Inland	Illinois No. 5	0.5	0.9
Kepler	Pocahontas No. 3	7.8	7.9
Price	Castlegate (subseam No. 3)	6.2, 5.0	4.2

Table 3.1 A Comparison of Direct and Indirect Methane Contents for Selected Coalbeds in the USA (after Kissel [31]).

Other indirect techniques consist of subjecting the coal sample to a constant pressure and measuring the gas uptake until equilibrium is reached. The gas content of the sample is considered to be equal to the amount of gas adsorbed. By varying the pressure and allowing equilibrium to be reached a complete gas adsorption curve can be obtained. The amount of gas adsorbed can be calculated from volumetric or gravimetric methods.

The strata pressure of methane has to be measured in-situ e.g. in order to make use of adsorption curves derived for a specific coal seam, the actual seam pressure must be known. The in-situ measurement of gas pressure presents a number of difficulties which are mainly related to the sealing of boreholes in order to obtain correct pressure build-up. Provided no leakage occurs, borehole gas pressure increases until an equilibrium pressure is reachead which will reflect the actual seam gas pressure. With present day borehole technology this is no longer a major problem [32].

Additionally, the laboratory measurements require the coal sample to have the same moisture content and temperature as the seam, which is difficult to obtain [31].

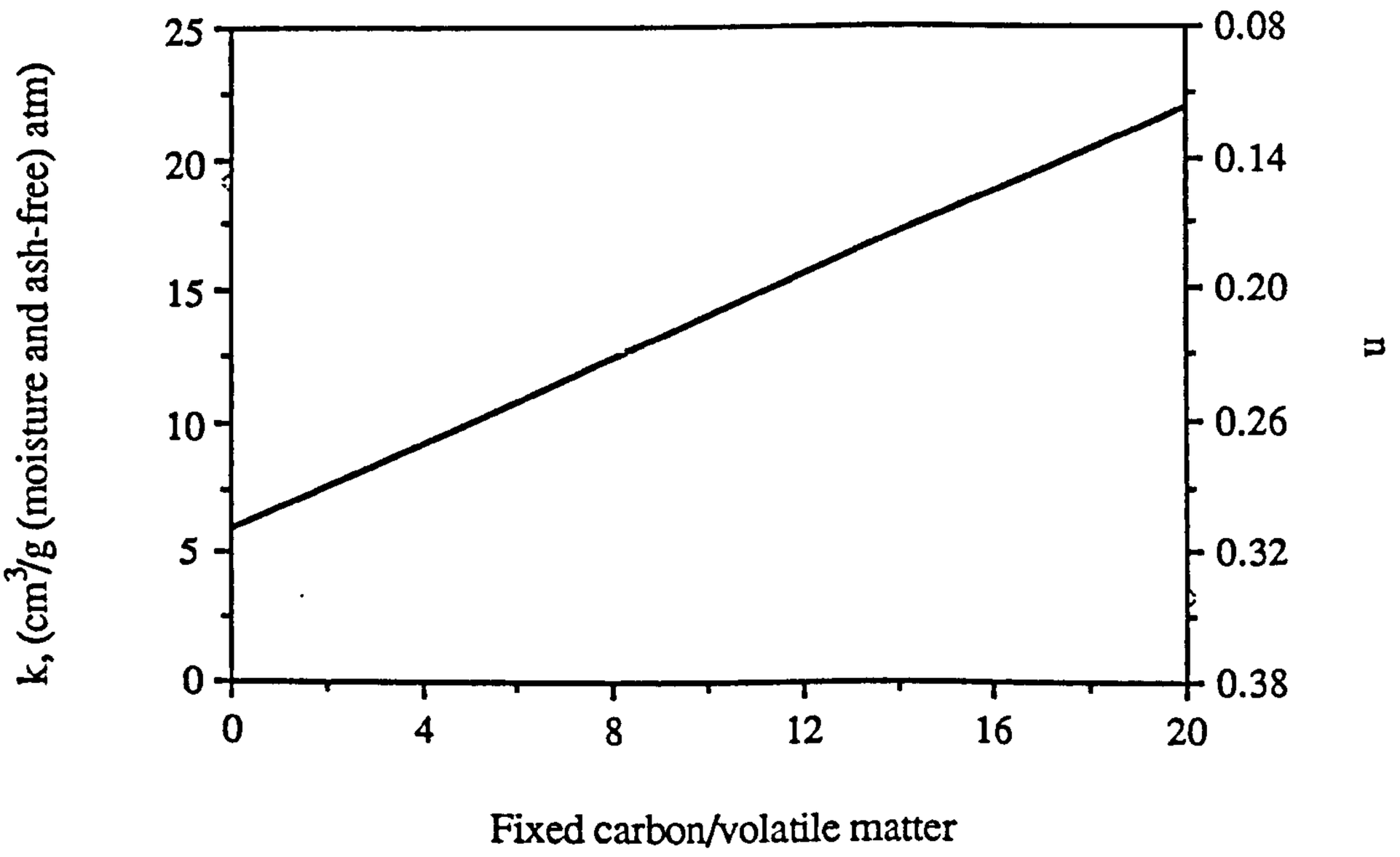


Figure 3.1 Value of Adsorption Constants k and n Versus the Ratio of Fixed Carbon to Volatile Matter, (after Kim [22]).

3.2.3 Zone of Gas Emission

Mining operations cause a modification of the stress patterns around a longwall area. Within this area, there is a zone in which the mechanical structure of strata is changed.

The extent of this affected area is called the 'zone of gas emission' and is usually considered to be smaller than the area in which the original stresses have changed [33]. The gas emission zone is further divided into the roof and floor zone of gas emission. The floor zone of gas emission is often much smaller than the roof zone, therefore, the contribution to total methane emission is usually higher from the roof zone. Within the gas emission zone, the formation of fissures leads to changes in strata permeability which increases gas flow to mine workings. Without changes in strata permeability by means of stress re-distribution, there would be very little gas emission into mine workings.

3.2.3.1 Stress-Permeability Profiles

McPherson [34] combined the theories of rock mechanics with the results of Mordecai's work [35] to produce a permeability profile for a longwall coal face (Figure 3.2). He suggested that the permeability of a coal seam would decrease in the stressed zone ahead of the face despite the fact that microfracturing would occur in this zone. Microfracturing would have the effect of causing partial sealing of the interconnected pores within the coal. This would cause a further decrease in the already low permeability. A localized increase in permeability occurs behind the face where the rock is relaxed. This increase of permeability would be of a few orders of magnitude due to the opening of microfractures, relaxation of normal cleavage, and planes of weakness between the beds. This induced permeability provides paths along which gas can flow. As the cover load is re-established, the permeability decreases, but to a greater level than its original value.

Research work by Durucan [36] produced a stress/permeability profile for a working longwall face, shown in Figure 3.3. Immediately in front of the face is the 'front abutment zone' where both principal stresses are thought to be compressive and increasing towards the face. The principal stress σ_1 is considered to reach its maximum value some 3 to 5 metres ahead of the face, whilst σ_3 decreases to become highly tensile resulting in fracturing of the coal seam. Permeability is seen to increase sharply here and this zone is referred to as the 'crushing zone'. Maximum permeability occurs in the 'stress relief zone' which is behind the face and reaches into the waste. In this zone, the state of stresses are very complex. Permeability begins to decrease as the cover load is re-established in the recompaction zone. Here the principal stresses are thought to be triaxial in form.

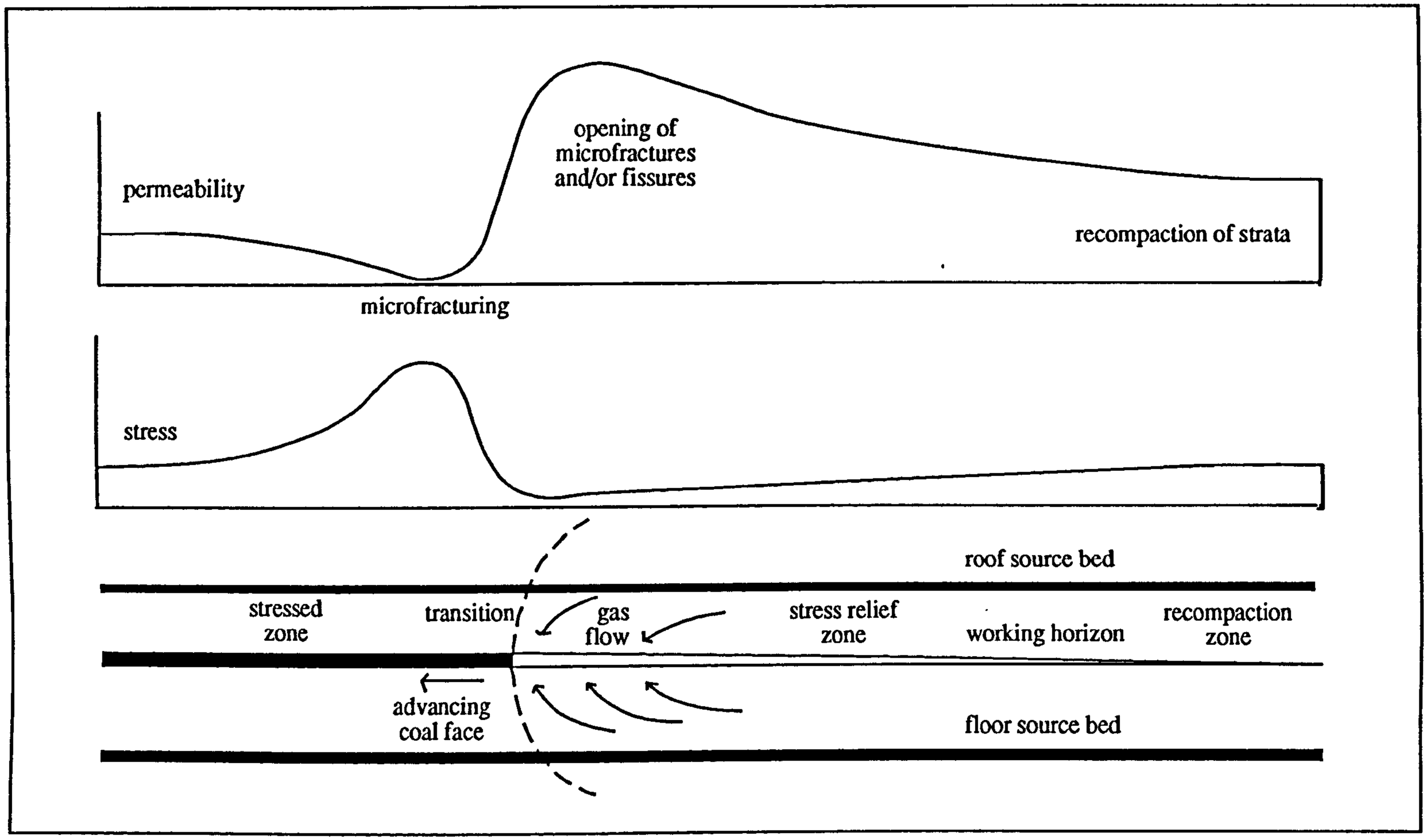


Figure 3.2 The Variation of Stress and Permeability in Strata Above and Below an Advancing Longwall Coal Face (after McPherson [34]).

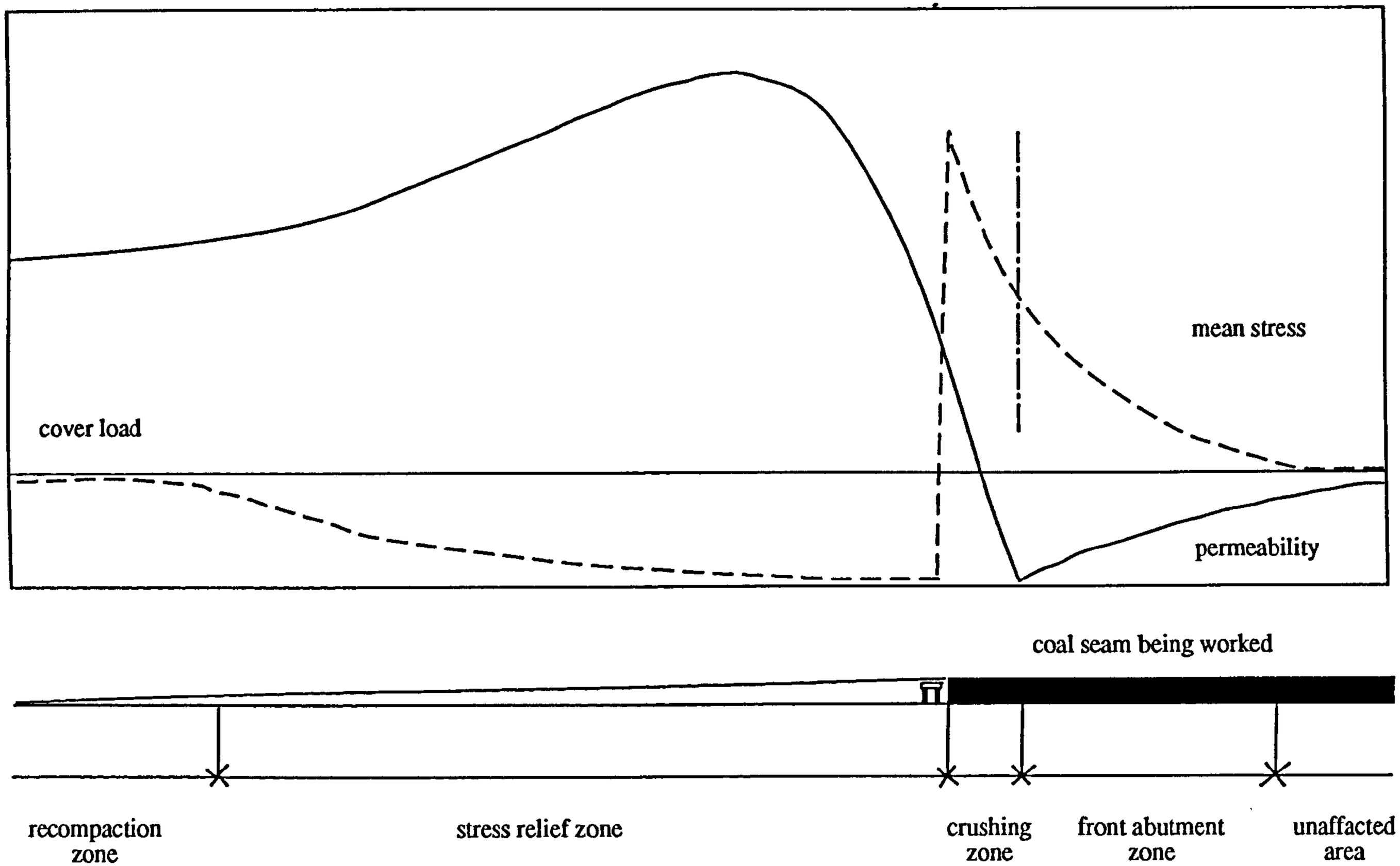


Figure 3.3 General Stress-Permeability Profile at the Roof Level of a Working Longwall Face (Not to Scale), (after Durucan [36]).

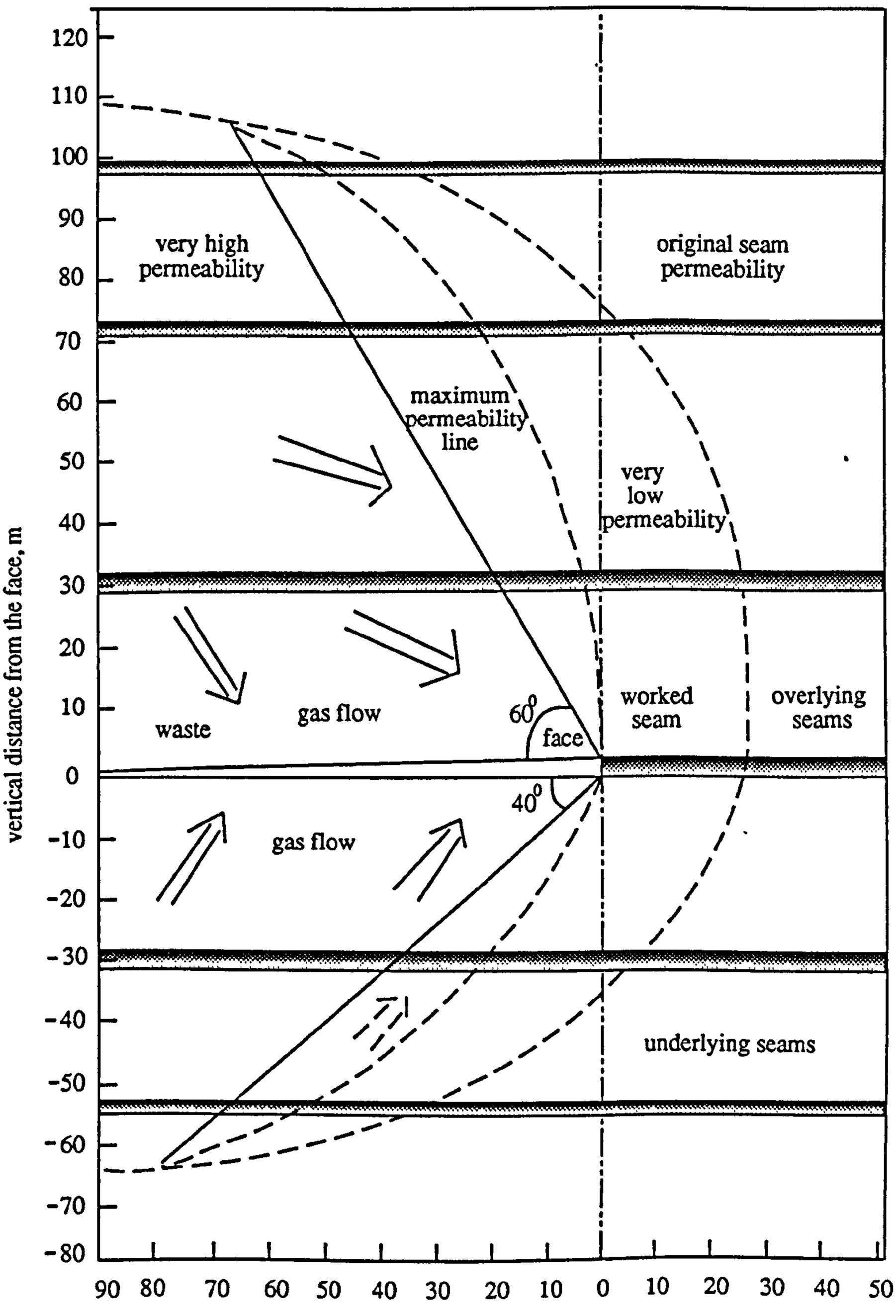


Figure 3.4 Different Permeability Zones and Suggested Flow Paths of Methane Around a Working Longwall Face (after Durucan [36]).

Changes of permeability within the zone lead to areas where gas flow will be highest. Figure 3.4 shows the different permeability zones and suggested flow paths of methane around a working longwall face which is assumed to be an unmined area [36]. In front of the face high abutment pressures give rise to very low strata permeabilities. The outer boundaries of this low permeability zone are defined by the parabola on the right hand side of the figure. In the crushing zone which lies between the maximum permeability line and the inner parabola, permeability starts to increase. Lines of maximum permeability will lie at angles of 60 degrees (in the roof) and 45 degrees (in the floor) to the working horizon, behind the face. The highest quantities of gas flow would originate from areas behind these lines, where permeability is very high. Durucan suggests that gas sources at distances more than 100 m above and 50 m below the working horizon are little affected by stress disturbances. Therefore, the permeabilities of these areas remain at their original value and contribute little to gas flow towards the workings.

3.2.4 Degree of Gas Emission

The degree of gas emission for a gas source is the percentage of the gas contained within the source which can desorb under certain conditions. The determination of the degree of gas emission is the main difference between the various empirical methods of methane prediction. Table 3.2 shows the factors taken into account by each method when determining the degree of gas emission [37].

3.2.5 MRDE Methane Prediction Method

The MRDE method of methane prediction, developed in the UK at the Mines Research and Development Establishment by a number of researchers, is based on the work done by Airey [38] to model the gas flow through broken coal.

The prediction of gas emission is calculated from information concerning the following factors [37, 39];

1. the thickness of the gas source relative to the thickness of the extracted coal seam; this converts the calculated emission quantity into cubic metres of gas per tonne of extracted coal from the mined seam,
2. the initial gas content of the source,
3. the degree of gas emission, or the percentage of the original gas content which is expected to be emitted into the mine working from the source,
4. the rate of coal production at the face.

It is in the calculation of the degree of gas emission which makes the MRDE prediction method different from those devised by others. In the MRDE method the degree of gas emission is a time-dependent function. The gas emission function used by MRDE was developed from Airey's theory of gas emission from broken coal. Airey considered disturbed coal seams to be composed of fragments of various size. His research resulted in an empirical equation which was established by laboratory measurements. This equation can be used to determine the release of gas with respect to time and partical size, represented by the following:

$$V(t) = A\left[1 - \exp\left(-\left(\frac{t}{t_0}\right)^n\right)\right] \qquad [3.3]$$

where

$V(t)$ = gas emitted up to time t after the coal starts to degas, the pressure at the coal surface being reduced to atmospheric at $t = 0$,

A = initial gas content of the coal,

t_0 = a time constant of the coal sample for 63% of the gas to desorb,

n = a factor describing the type of cracking within the lump (about 0.333 for bituminous coal).

This equation (3.3) is used in the MRDE method to determine the degree of gas emission. The emission function is evaluated for three main sources;

1. gas sources in the adjacent strata,
2. emission from the worked seam,
3. emission from broken coal on the conveyor.

No	Factor	Method							
		Stuffken NL	Schulz GFR	Flugge GFR	Winter GFR	Gunther F	Lidin USSR	Barbara PL	MRDE UK
1	Gas content of adjacent seams	√	√	√	√	√	√	√	√
2	Gas content of other strata		√	√	√	√			√
3	Thickness of adjacent strata	√	√	√	√	√	√	√	√
4	Distance of adjacent seams	√	√	√	√	√	√	√	√
5	Face length (determines emission zone)		√	√					
6	Intensity of firedamp drainage			√	√				
7	Dip of worked seam		*				√		
8	Worked seam thickness						√	√	
9	Caving or stowing						√	√	
10	Depth of working	*							√
11	Age of district								√
12	Advance rate of face								√
13	Emission from conveyor coal								√

* Factors used indirectly by Stuffken to calculate gas content values.

Table 3.2 Factors Taken into Account by the Various Prediction Methods (after Dunmore [37]).

For emission from sources in adjacent strata account is taken of the depth of working, the age of the district and the distance of the source relative to the working horizon. Sources at a height of 200 m and a depth of 100 m from the worked seam are believed not to contribute to the gas flow. The emission function is time dependent in that it varies according to the time which has passed since the face advanced. Emission from the worked seam is wholly dependent on the rate of advance and decreases with increasing face advance rate. Dunmore suggests that 50% of the gas content of the worked seam is emitted at a weekly advance rate of 17 m. His curve for the degree of gas emission from the worked seam is shown in Figure 3.5.

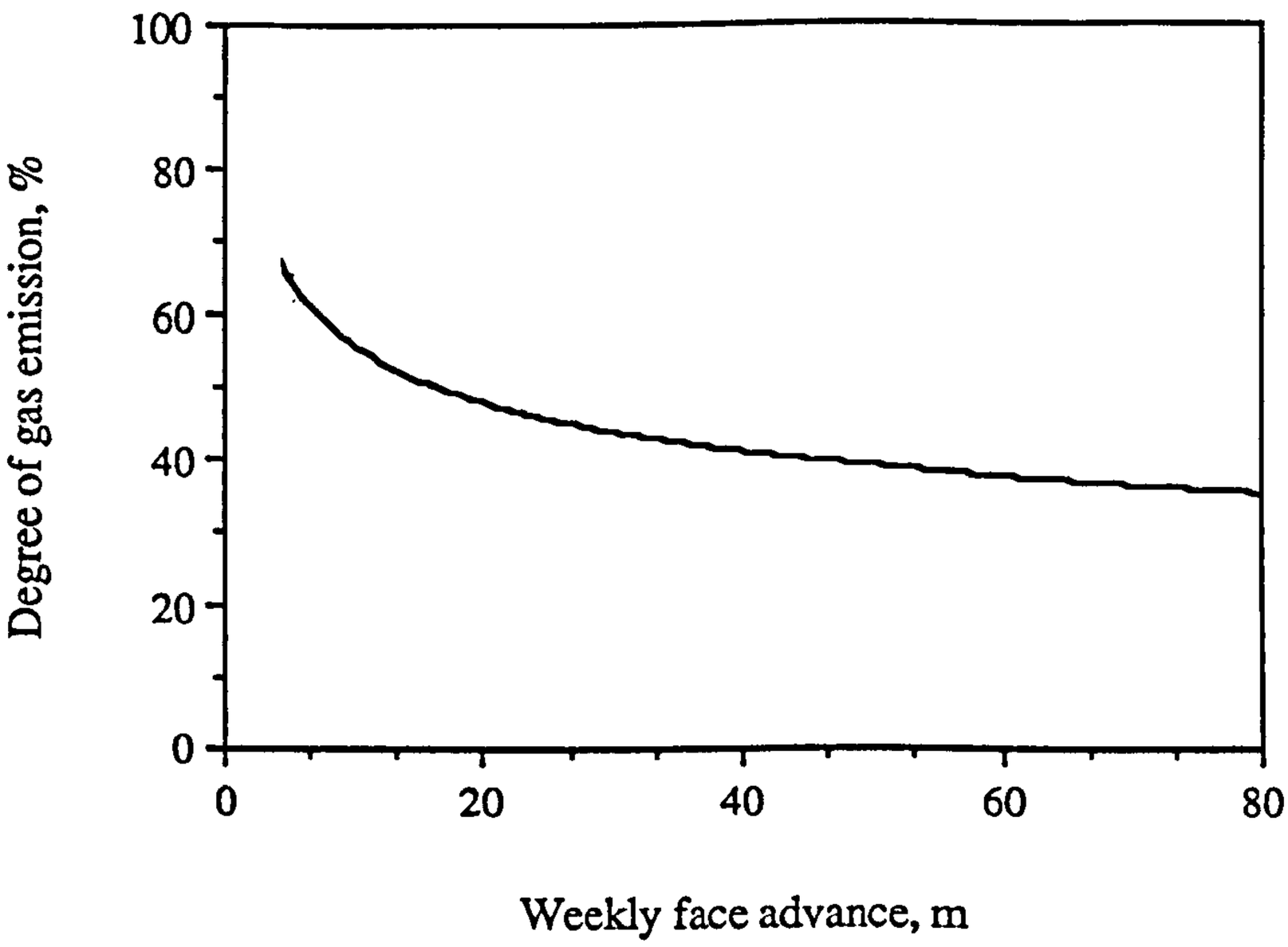


Figure 3.5 The MRDE Curve for the Degree of Gas Emission from the Worked Seam, (after Dunmore [40]).

Coal in clearance also contributes to the total methane emission and the quantity emitted is related to a number of factors. The most important is the length of time spent in the intake air and to obtain the best prediction results the emission curve should be determined for the prevailing situation. The gas emission curve for a coal of typical size is shown in Figure 3.6.

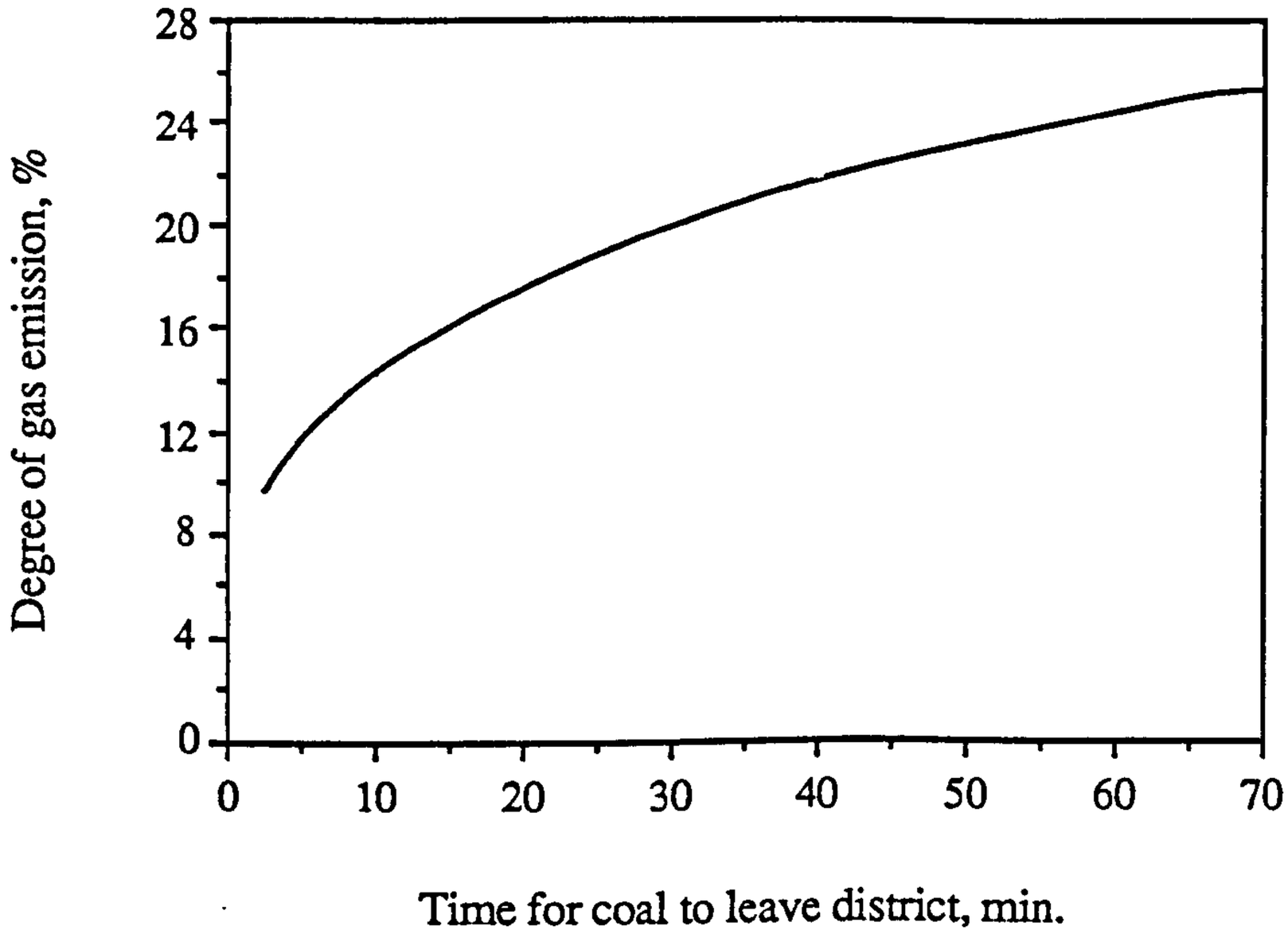


Figure 3.6 An Example Curve for the Degassing from a Bituminous Coal During Conveyor Clearance, (after Dunmore [40]).

To calculate the air required to dilute the emitted methane the program takes account of peak methane levels. MRDE have adopted the CERCHAR concept of a coefficient of irregularity [41]. This coefficient is obtained by finding the maximum and mean methane concentrations for a large number of observations. The coefficient of irregularity is then defined as the daily maximum value which is exceeded on 5% of the daily maximum values, divided by the mean of all the mean daily values over the observation period. Coefficient values are found to vary from site to site but typical values are 2.0 for the return end of the face airstream and 1.5 for the outbye end of a return [42]. The necessary airflow is found by multiplying the prediction amount by the appropriate coefficient value and deciding on the upper limit of methane concentration in the airstream.

The MRDE prediction method has been steadily developed into a powerful computer program which is capable of calculating methane emission levels resulting from variable face advance rates. This represents the total volume of methane flowing into the working district from all identified sources, and is expressed in m^3/t of mined coal at the face. Allowance can be made for gas captured by methane drainage where this is practised. The prediction results take the form of a set of tables and graphs.

Changes in the rate of face advance can be accommodated but the method assumes that the new rate will be sustained over a five day period. Today, the concept of a five day working week is historical since in many mines the economic climate has changed conventional working patterns.

The MRDE methane prediction program can be viewed as an aid to planning the airflow required to deal with the predicted level of methane emission. It details the sources of methane and the level of methane emission that can be expected at a specific rate of advance. It is not, however, capable of predictions for rapid and continual changes in face advance rate.

3.3 Numerical Approaches to Methane Prediction

Numerical methods of methane prediction are based on the principles of gas flow in porous permeable media. This flow is described by Darcy's Law and prediction techniques have mainly arisen from the numerical solutions by computers of gas flow equations mainly derived from Darcy's law [43, 44, 45].

The first researchers to consider numerical prediction techniques were Owili-Eger and Ramani [44] from the Pennsylvania State University. They developed a model which assumed steady-state methane flow with small variations in gas temperature. Further research by others found that this was only true for shallow depths and in deep mine situations larger temperature variations were found to occur which affects gas flow rates. Keen [20] and O'Shaughnessy [44] sought to apply numerical techniques to deep British longwall mining by developing transient solutions for methane flow. Their research made use of equlibrium sorption models which are based on the assumption that adsorbed gas is in a continuous state of equilibrium with the free gas pressure. Examples of equilibrium sorption based models are given in Figure 3.7.

The latest research into numerical methods of methane prediction has centred on the development of non-equilibrium sorption models. Such models take account of the kinetics of the diffusion/desorption process which are believed to be of importance in fractured coal seams. A full discussion on various modelling techniques and their

assumptions can be found in King and Ertekin's comprehensive survey of mathematical models related to methane production from coal seams [45]. It is important to note, however, that although a great deal of effort has been expendend, very few mathematical models have actually been sufficently developed to be able to predict methane emission into mine workings.

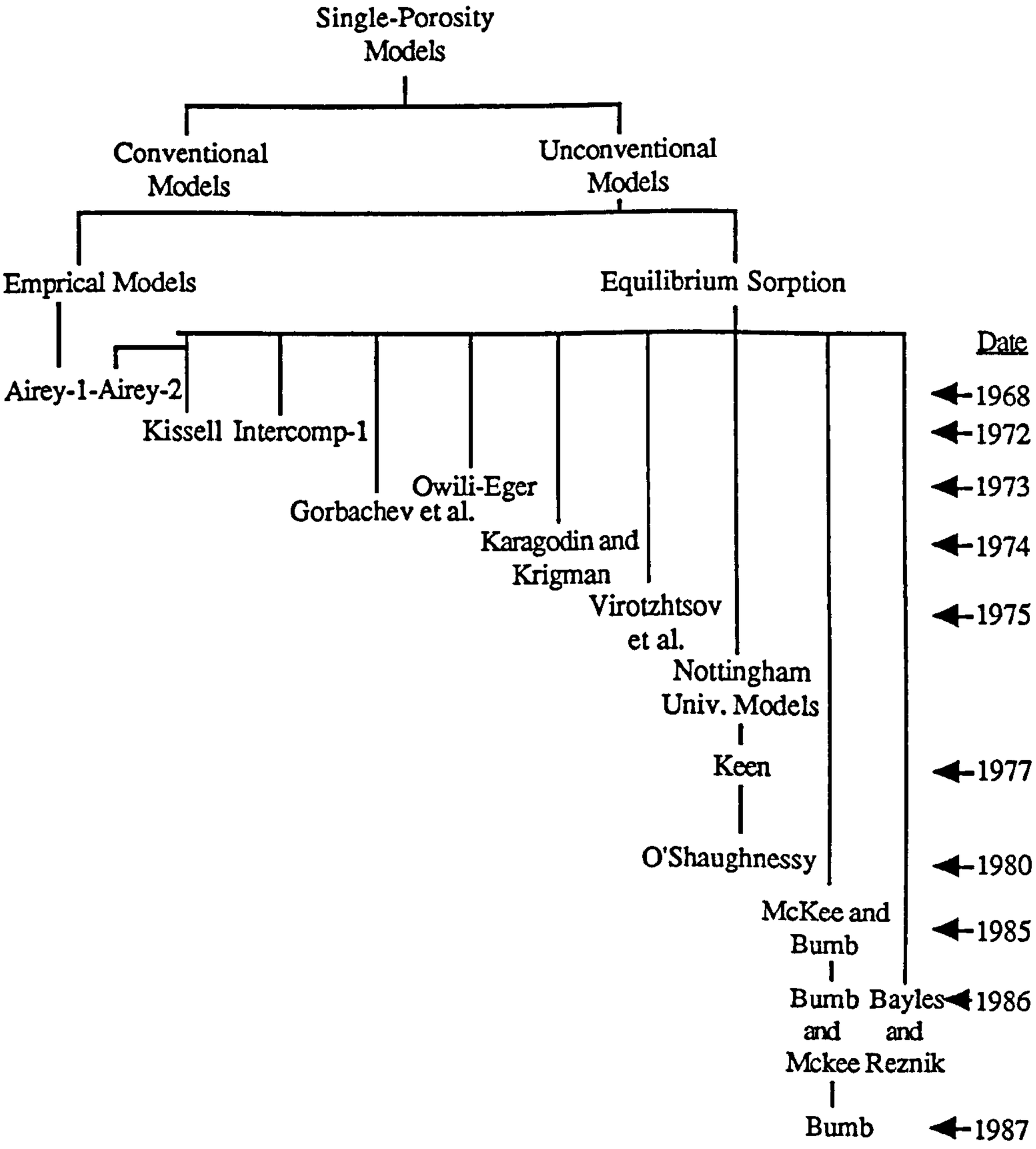


Figure 3.7 Relationships of Equilibrium Sorption Models (after King and Ertekin [45]).

The required input to computer programmes in mathematical prediction include parameters to define the model size, the initial and time dependent boundary conditions, properties of the coal seams and strata such as directional permeabilities and porosities

and the properties of the flowing gas such as viscosity [46]. The programmes terminate when the flow equation has been solved and the output gives the predicted gas pressure distribution.

Permeability is considered to be the most crucial factor in the solution of the gas flow equations. Therefore, recent studies have included elements of stress analysis and stress-permeability analysis in the simulation of methane flow using numerical techniques [47].

3.3.1 An Example of a Numerical Approach to Methane Prediction

Research carried out in the Department of Mining Engineering at the University of Nottingham resulted in the development of a time-dependent gas flow equation for variable anisotropic permeability derived from Darcy's law. This equation makes the following assumptions and is shown as equation 3.4 [46];

1. Flow is laminar,
2. Flow is single-phase,
3. Slip effects may be ignored,
4. Isothermal flow conditions exist,
5. Methane obeys the perfect gas law (shows no abnormal compressibility),
6. Darcy's law is valid,
7. The effect of adsorption may be ignored.

$$\frac{\partial p}{\partial t} = \frac{1}{2\mu\rho}\left\{ \frac{\partial}{\partial x_1}\left[k_1 \frac{\partial p^2}{\partial x_1}\right] + \frac{\partial}{\partial x_2}\left[k_2 \frac{\partial p^2}{\partial x_2}\right] + \frac{\partial}{\partial x_3}\left[k_3 \frac{\partial p^2}{\partial x_3}\right] \right\} \qquad [3.4]$$

where

μ = viscosity of the flowing gas,

ρ = porosity of the medium,

k_i = directional permeability of the medium,

p = gas pressure.

Finite element techniques are used to obtain approximate solutions to this equation to give time-dependent gas pressures. Ediz [47] gives a general time-dependent gas flow equation for anisotropic media with variable permeability derived from Darcy's Law and finds approximate solutions to the equation by using the finite element method. After obtaining the gas pressure distribution throughout the finite element mesh, he introduced a mass flow equation to calculate the methane flow rate into a roadway or waste area. The model also takes account of methane drainage and whether the boreholes are inclined or vertical. He achieved this by modifying and devising additions to a finite element program, originally written for the solution of heat flux calculations by PAFEC Ltd. An example of the predicted results are listed in Table 3.3. These were calculated for a typical gas boundary pressure of 10×10^5 N/m^2. The figures show how the application of methane drainage using inclined boreholes to a retreating face, changes the distribution of methane emission. The drainage boreholes were assumed to have been drilled from the return airway into the floor and roof strata.

Mining Type	Retreat Longwall	Advance Longwall
Methane Flow into Roadway from Roof Strata, l/s	26.58 20.04*	195.95 137.62*
Methane Flow into Roadway from Floor Strata, l/s	19.22 16.15*	119.56 83.65*
Methane Flow into Goaf from Roof Strata, l/s	324.38 139.92*	—
Methane Flow into Goaf from Floor Strata, l/s	110.31 59.72*	—
Total Return End Methane Flow, l/s	480.49 235.83*	315.51 221.27*

* with Drainage

Table 3.3 Results of Methane Flow Prediction for Retreat and Advance Models (Boundary Gas Pressure was Taken as 10×10^5 N/m^2), (after Ediz [47]).

Ediz suggests that his numerical gas flow simulation model is extremely versatile in the analysis of strata gas flow around a moving longwall coal face, although the model does not take account of methane emission from the worked seam. Ediz's program is able to allow for rapid variations in permeability and gas pressures with time. Ediz reports that the accuracy of the model depends upon the field data supplied, e.g. boundary gas pressures and strata permeabilities. Ediz concludes that there has been very little research carried out on the accurate determination of boundary gas pressures, strata permeabilities and on the post-failure stress permeabilities of strata measures. Further investigations by

Ediz included sensitivity tests to investigate the effects on methane flow rate of varying borehole length, pressure and spacing.

Numerical techniques for methane prediction have developed substantially over the past fifteen years. However, the usefulness of such numerical predictions to a ventilation engineer are as yet slight, at their current state of development due to the availability and the accuracy of the necessary input data. It is thought that there is some link between the rate of advance and initial values of boundary gas pressures but further research is needed before numerical techniques can accurately predict methane emission for varying face advance rates [48].

3.4 Statistical Approaches to Methane Prediction

Statistical approaches to methane prediction involve the use of various statistical techniques. These range from simple descriptive statistics such as histograms of data to analytical statistics such as regression analysis and others. Descriptive statistics are used to help format data into a useful form and usually play an important role in the development of empirical models. To some extent then, an empirical approach is analogous with a statistical one. However, an empirical model is a simple mathematical description of observable physical phenomenon which can be solved analytically or numerically. The use of analytical statistics to achieve a prediction model can be regarded as a separate approach because the final model, while still essentially empirical, will be purely due to statistical relationships deduced from the observed data. Such techniques allow inferences to be made concerning the mechanisms which generated the data.

Unlike empirical and numerical approaches, statistical ones make no direct attempt to model the physical processes influencing methane emission. This statement, however, is explained by noting a fundamental concept in the application of statistics to any task. Unless expert knowledge is available to explain and account for the results of statistical analysis, they cannot be interpreted and therefore, nothing is achieved. In other words, the observations obtained by statistical analysis are of limited value if there is no understanding, however limited, of the physical processes which generated the data.

Before analysis can be carried out great importance is placed on the acquisition of relevant data. If the data is not a true reflection of what has occurred there will be a danger of wrong interpretation. Fortunately, this has become easier over the past decade due to the extensive use of reliable and accurate electronic environmental monitoring instruments in use at collieries.

A number of analytical statistical techniques are available for use. These range from regression analysis which can be used in increasing levels of complexity to various forms of time series analysis. This thesis will concentrate on the application of time series analysis to methane prediction.

3.4.1 A Medium-Term Prediction Model

Statistical analysis can be defined as a scientific approach towards an understanding of information presenting itself in numerical form. Kaffanke [49] followed this approach to predict medium-term levels of methane emission. He uses 'medium-term' to describe a prediction period between one day to one month in length.

Kaffanke's method of making medium-term predictions of the methane content in the return air can be characterized as a point forecast using discrete multiple regression. The influencing variables chosen to describe the behaviour of methane emission were;

1. daily output in tonnes per day,
2. accumulated output in tonnes,
3. previous day's methane flow in m^3/s,
4. methane flow on previous Sunday, in m^3/s,
5. desorbable gas content in m^3/t,
6. number of preceding rest days in each case (only Saturdays, Sundays and Mondays).

The coal face studied produced coal Monday to Friday and no production took place at the weekend. Because of this, Kaffanke's data on methane concentration in the return air

showed distinct cyclical behaviour. This behaviour is shown in Figure 3.8. . The methane concentration is seen to increase throughout the working week and fall during the work-free weekend.

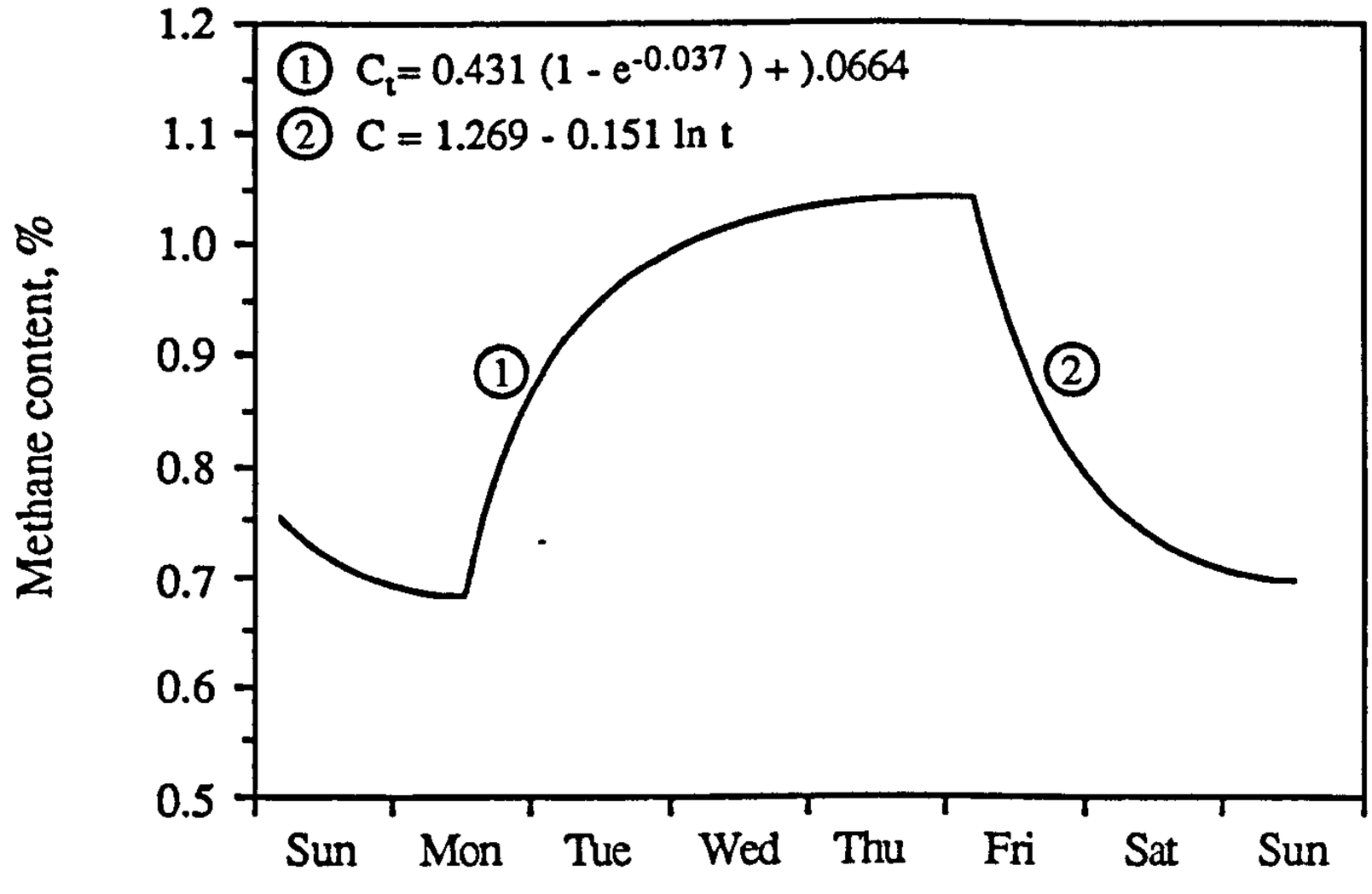


Figure 3.8 Typical Weekly Rhythm of the Methane Content in the Return Air Flow of a Working (hourly average values), (modified, after Kaffanke [49]).

The predictions were obtained by performing regression analysis in two steps,

1. the base data available right up to the current margin were subjected to regression analysis,
2. the structural equations thus obtained formed the basis for the next forecast.

The seven structural equations (one for each day of the week) obtained by Kaffanke were used to forecast values for methane flow and concentration. He was able to use the same structural equations throughout the modelling and forecasting period i.e. the structural equation for a Monday was used to forecast emission for all Mondays. This was because the pattern of production was constant for each particular day over the observation and forecasting period. The structural equations determined are shown in Table 3.4.

Day	Structural equation	MCF
Mon	$\dot{M}_{MO} = 1.324831\ M_{SU} + 0.049330$	0.804
Tue	$\dot{M}_{TU} = 1.22.291\ M_{SU} + 0.000049\ \Sigma\ m_{MO/TU} - 0.104899$	0.918
Wed	$\dot{M}_{WE} \doteq 0.000029 \Sigma\ m_{MO/WE} + 0.640939\ M_{SU} + 0.000057\ m_{WE} - 0.116830$	0.828
Thu	$\dot{M}_{TH} = 1.059347\ M_{WE} + 0.006453$	0.844
Fri	$\dot{M}_{FR} = 0.000034\ \Sigma\ m_{MO/FR} + 0.808406\ M_{SU} + 0.000001\ m_{FR} - 0.205325$	0.744
Sat	$\dot{M}_{SA} = 0.732386\ M_{FR} + 0.000007\ m_{SA} + 0.025258$	0.868
Sun	$\dot{M}_{SU} = 0.318048\ M_{FR} + 0.249384\ M_{SA} + 0.026625\ q_{SU} - 0.222538$	0.855

$\dot{M}$ = Methane flow in m^3/s m = Output in t/d

q = Gas content in m^3/t MCF = Multiple correlation factor

$\Sigma\ m_{MO/WE}$ = accumulated output in tonnes from Monday to Wednesday

Table 3.4 Structural Equations for Methane Flow, (after Kaffanke [49]).

The subsequent forecasts for methane concentration over the observation period are shown in Figure 3.9 . The backcasts from week 1 to week 14 coincide quite well with the actual methane concentrations recorded over this time period. This means that the structural equations determined over the historical part of the data generate forecasts which validate the accuracy of the statistical model. Further proof of the model accuracy is illustrated by the forecasts from week 14 onwards. The data recorded during this period was not used to obtain the structural equations yet the forecasts for this period continue to demonstrate the models accuracy.

By his research into the use of discrete linear multiple regression, Kaffanke has shown that methane emission could be predicted to acceptable accuracy. His research also revealed some interesting information. For example, his data demonstrated a cyclical pattern of methane emission and he showed how some influencing variables were more important than others. In his conclusion, he recommends further studies in the use of regression and other statistical techniques, such as time series analysis, linear filter theory and spectral analysis to methane prediction.

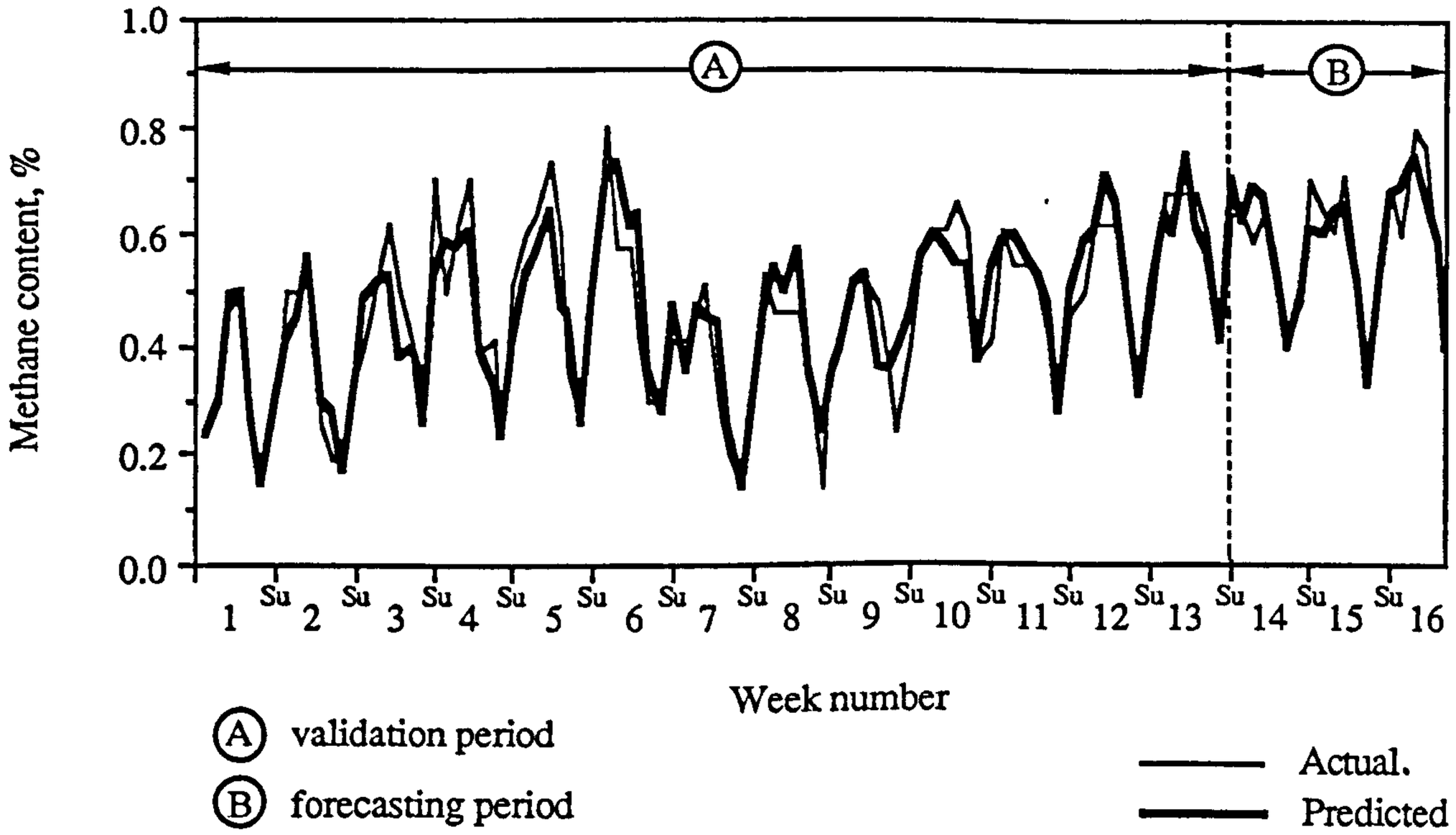


Figure 3.9 Graph of Methane Concentration Prediction, (modified, after Kaffanke [49]).

3.5 Conclusion

This chapter has reviewed the different approaches available to predict methane emission. By far, most effort has been channelled into the development of empirical techniques. Most empirical techniques, notably the MRDE method, have a mathematical basis but generally the techniques are of limited use outside of the specific geographical area which they were designed for. Although empirical methods are relatively simple they lack the theoretical base required for accurate prediction. Another major drawback of empirical techniques is the amount of information they use in order to obtain predictions, or rather their requirement of a good knowledge of the physical variables that are used in the predictive method. For example, all of the methods require strata gas content values and these are determined by drilling and laboratory analysis.

Numerical prediction techniques are becoming increasingly sophisticated but none are yet capable of determining methane emission in terms of face advance rate. Although this will undoubtedly be overcome in the future, it will not be before extensive research has

been undertaken on how to accurately determine the input parameters necessary to solve Darcy's gas flow equation. At the present time the accuracy of the predictions is dependent upon these factors and not upon the complexity of the mathematical model.

The application of analytical statistics to methane prediction is still in its infancy. Kaffanke's research demonstrated that methane forecasts could be made with sufficient accuracy. A statistical prediction method is of course empirical in nature and so only suited to adequately perform in the conditions for which it was built. A major benefit that a statistcal prediction method can offer is the ability to make accurate predictions without a thorough knowledge of physical conditions that limit the accuracy of numerical and empirical methods. For this reason this thesis carries out further investigations into the applicability of statistical methods, in the form of time series analysis, to predict methane emission.

The next chapter, chapter 4, provides an introduction to univariate time series analysis and is the basis of a method capable of predicting methane emission purely from a knowledge of monitored data.

CHAPTER FOUR

AN INTRODUCTION TO UNIVARIATE TIME SERIES ANALYSIS

4.1 Introduction

This chapter provides an introduction to the Box-Jenkins statistical method of time series analysis. In particular it is concerned with the principles that allow inference to be drawn from time series data and how, according to procedures advocated by Box and Jenkins, time series ARIMA models are built. The univariate ARIMA time series models can be used to obtain forecasts and also provide a basis for the building of multivariate models, which are topics of later chapters.

The primary objective of time series analysis is to develop a model which can be used to forecast future values of a variable or variables. However, this can be viewed as a narrow objective, especially when a great deal of knowledge can be obtained by performing the analysis and correctly interpreting results. Forecasting is widely used today by business managers and economists and is increasingly being used by scientists and engineers in technical applications. There are a number of techniques which can be used to build a mathematical model of a time series and by far the most popular is the Box-Jenkins approach. Since its conception in the 1960's it has steadily been developed into a sophisticated method, capable of producing models to many different forms of time series. As a consequence of its popularity and ability to produce accurate forecasts the basis of the Box-Jenkins approach has been incorporated into many statistical computer packages which can be used to build a time series model.

4.2 Definition of a Time Series

A time series is a time-ordered sequence of observations of a variable. The time interval between each observation must always be constant though it does not need to be the same interval in which the observations were taken. This means that a time series can be transformed into a suitable form for analysis so long as an appropriate time interval between observations is chosen. Examples of time series are: daily maximum air

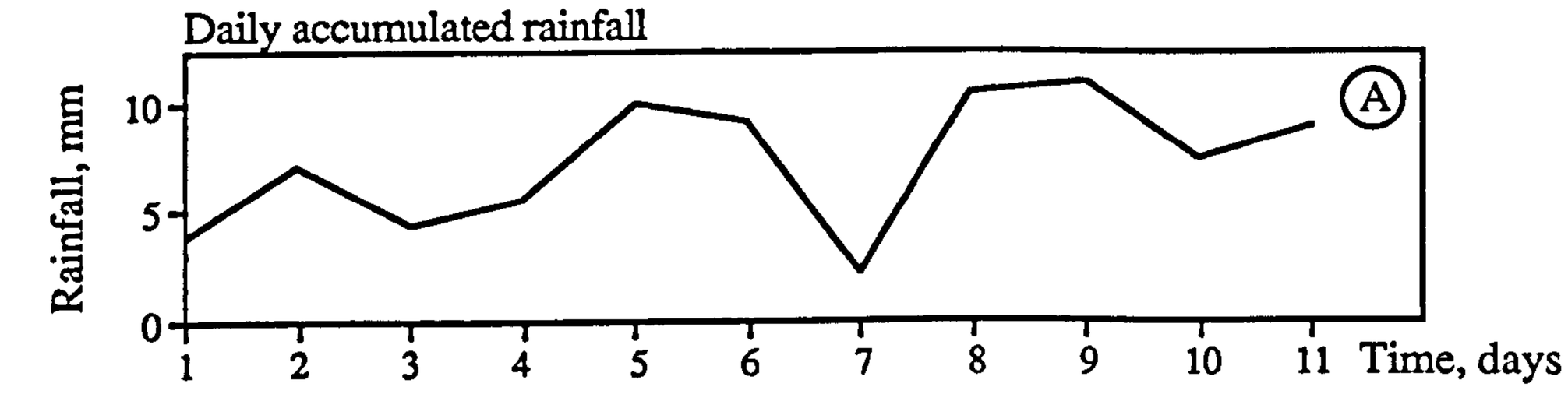


Figure 4.1 Examples of Time Series

temperature as a function of time in days, catalyst concentration in a chemical process as a function of hours, petrol consumption as a function of distance driven, ice cream sales as a function of temperature. A list of possible time series would be boundless.

Examples of some time series are shown in Figure 4.1. These series were chosen to demonstrate some important characteristics that time series can possess. These are discussed in the following sections.

4.2.1 Continuous or Discrete Time Series

A time series can be continuous or discrete in form, depending on how the observed data is presented. A continuous series (examples B and E) is a curve which is made up of an observed value of a variable, recorded at every moment in time. A discrete series (which are the most usual, examples A, C and D) consists of discrete observations of a variable at predetermined time points. The time intervals of discrete series may be yearly, monthly, weekly, etc. A time series that is considered to be continuous can also be a discrete series of points taken at certain time intervals. Whether a time series is continuous or discrete is of no real consequence, it is the length of the time interval which is important.

The selection of the time interval will depend on the behaviour of the variable being monitored. For example, if the variable is fluctuating rapidly in its continuous series and the monitoring instrument is unable to respond to these fluctuations then the recorded series will be essentially discrete in nature. Another type of discrete series is an accumulated series which arises when the variable being studied cannot be measured at every moment in time. Instead, the value of the variable is accumulated over equal intervals of time, such as total daily rainfall levels, shown by series A in Figure 4.1.

4.2.2 Stationary and Non-Stationary Time Series

A time series is referred to as being stationary or non-stationary depending on whether it is in statistical equilibrium, that is, its statistical properties do not evolve with time. A

stationary series has a constant mean and constant variance and is said to possess properties which are independent of any time difference within the period of observation. Series E in Figure 4.1 is stationary in that it appears to fluctuate about a level mean with similar amplitude of irregularity and thus constant variance. Series D has a changing amplitude of irregularity but the fluctuations are occurring about a fixed mean. Such a series is said to be stationary in the mean but non-stationary in the variance. Series A, B and C exhibit trend varying amplitude of irregularity and are non-stationary. Usually, it is necessary to decide whether the series is stationary or not and if not steps are taken to achieve statistical equilibrium.

4.2.3 Seasonal Time Series

A time series is said to be seasonal if it is known that it contains a component due to a regular cycle with a fixed period. If a series is suspected of possessing seasonality it is important that the seasonal cycle is determined from factors which are external to the data rather than by inspection of the data itself. For example, series C of monthly sun tan oil sales shows two annual peaks. The largest corresponds to sales during the summer months when people are sunbathing and the smaller to winter sales when such products are used by skiers. Thus, the series has two seasonal components which have a yearly cycle.

4.3 The Choice of the Box-Jenkins Approach to Time Series Analysis and Forecasting

There are a number of methods which can be used to analyse time series and the final choice of which method to use build a model is dependent on its possible application. In this thesis it is intended to develop models which can predict methane emission and have the possibility of application to a control system. The Box-Jenkins method requires a working knowledge of the procedure involved in identifying the correct model to fit the time series data and subsequently refining the model until it fits the data satisfactorily. This knowledge, while not difficult, requires patience and perseverance to acquire and is perhaps the greatest disadvantage in using the Box-Jenkins time series methodology. However, compared to other methods of analysing time series this difficulty is more than offset by the expectation of accurate forecasts. As well as the ability to model single or

univariate time series, Box-Jenkins methodology can also model causal relationships whereby the behaviour of a dependent variable is related to a number of independent variables. This is of importance when it becomes necessary to consider what dynamic relationships exist between methane emission and its influencing variables. Econometric methods and multiple regression are also capable of modelling causal relationships but it is generally acknowledged that they are more difficult to understand than univariate or multivariate Box-Jenkins models. Box-Jenkins models also have the advantage of application to a control system in which it is desired to maintain a variable as close as possible to a target value and once such a model is in operation it is possible to monitor its performance and determine if adjustments are necessary.

For a more detailed discussion on the main factors which govern the choice of a time series modelling and forecasting method refer to O'Donovan [50]. His findings are summarized in Table 4.1.

4.4 Models for Time Series Analysis and Forecasting

The models which are used within the Box-Jenkins approach are decided upon by an examination of the behaviour of the time series. In building a model, a mathematical analysis is carried out to determine whether the values of the series are in some way connected and whether this connection can be described mathematically.

As an example, consider the first 10 observations of a time series of methane concentration consisting of 430 observations and displayed in Figure 4.2. The ultimate aim is to have a time series model that adequately represents the historical data that it is based on and can be used to forecast future values of the series. To forecast the 11th value requires some inference to be drawn from the information contained in the first 10 observations. It seems that by an inspection of the plotted series the 11th value is influenced by the previous ones and, most importantly, the series does not appear to be totally random. A guess at a value for the 11th observation might be 0.78. This seems reasonable and on the basis of the previous observations it is highly unlikely to be say 0.70 or 0.90. In this case an intuitive guess gives a plausible answer which is close to the actual value of 0.79.

Factor	Forecasting Method						
	Simple exponential smoothing	Holt-Winters	De-composition	Box-Jenkins	Multiple Regression	Econometrics	Multivariate Box-Jenkins
Lead time							
Immediate	√	√	√	√	-	-	-
Short term	√	√	√	√	√	√	√
Medium term	-	-	-	-	√	√	√
Long term	-	-	-	-	√	√	√
Time to prepare forecast (1 - shortest ; 7 - longest)	1	2	3	5	4	7	6
Pattern of data							
Horizontal	√	-	√	√	-	-	√
Trend	-	√	√	√	√	√	√
Seasonal	-	√	√	√	√	√	√
Cyclical	-	-	√	-	√	√	-
Date requirements (S : period of seasonality)	10	15 2(S)	30 6(S)	30 6(S)	30 6(S)	Few 100	60 8(S)
Ease of understanding (1 - easiest ; 7 - hardest)	1	2	3	4	5	7	6
Cost (1 - least costly ; 7 - most costly)	1	2	3	5	4	7	6

Table 4.1 Comparison of Forecasting Methods (after O'Donovan [50]).

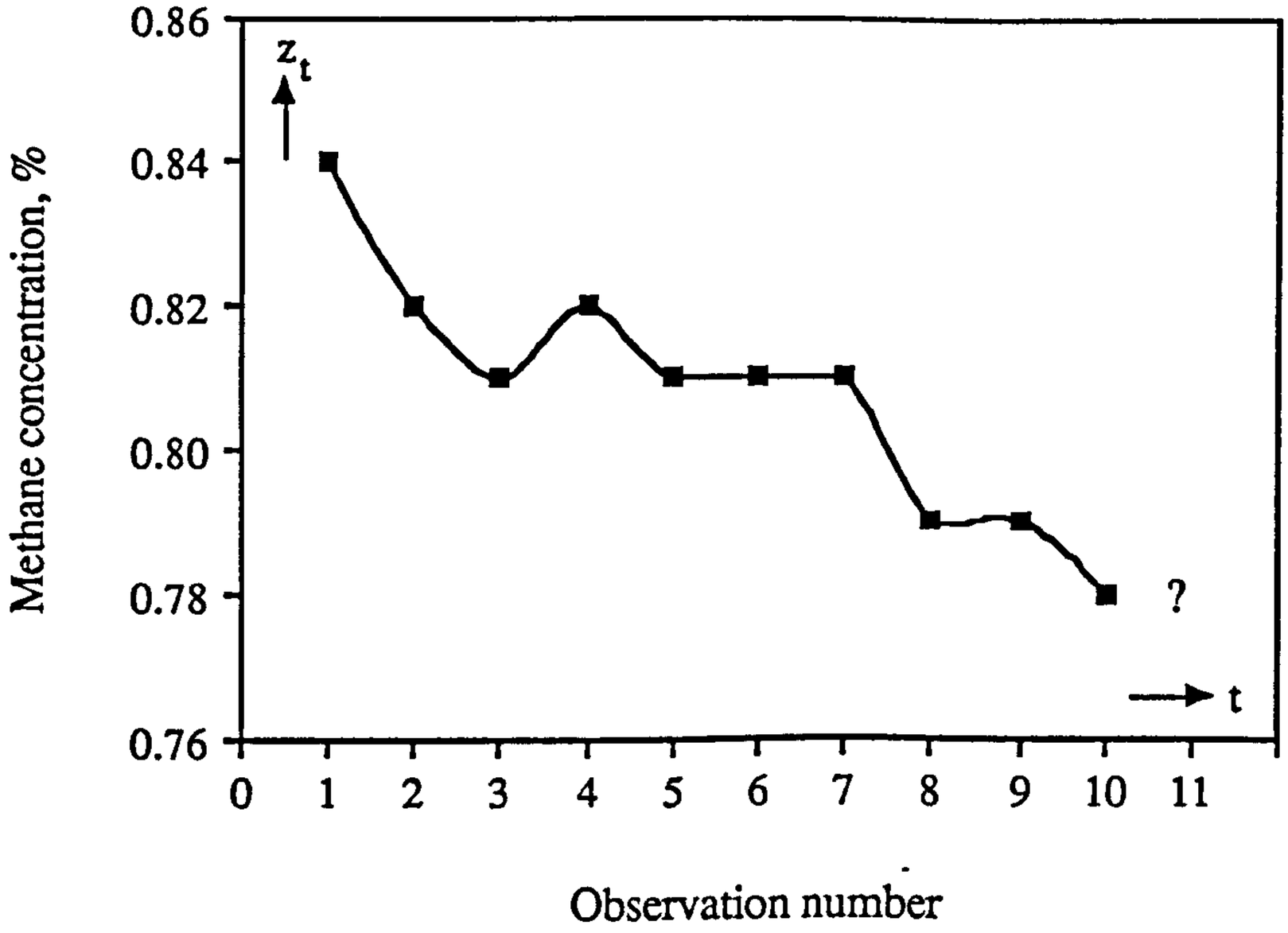


Figure 4.2 Plot of Methane Concentration Data

So far, no mathematics have been used, but to achieve a better forecast it is beneficial to make use of some techniques which enable a model to be built that can mathematically describe the historical series. The evidence that is needed is an indication of the correlation or memory between the observations. In predicting a value for the 11th value by inspection more attention was paid to the behaviour of the last few observations. The reason for this is that more significance is attached to these observations than the first few. However, the remainder of the series is not forgotten as these observations can indicate important information such as trend or seasonal patterns.

It is easy to express the 11th or nth observation as a function of the previous ones, the difficulty lies in finding the form of the function. Mathematically, the function is considered to consist of two components. The first component is thought of as being due to the process which generated the series and which is evident from the preceding observations. The second is a random element due to inexplicable causes in the process. The best model would be one that had a small random component. Thus, the 11th observation can be expressed mathematically as:

$$z_{11} = f(z_1, z_2, ..., z_{10}) + \varepsilon_{11} \qquad [4.1]$$

where

z_{11} = 11th observation, or the forecast,

$f(z_1, z_2, ..., z_{10})$ = a function of the first ten observations,

ε_{11} = a random variable.

The function, f and the random element, ε_t are termed a stochastic model representing z_t which in general terms is:

$$Z_t = f(z_{t-1}, z_{t-2}, ...) + \varepsilon_t \qquad [4.2]$$

The model expressed in equation 4.2 is the basis of a univariate stochastic model Z_t of the observed time series z_t.

The values of the observed time series, $z_1, z_2, ..., z_N$ are considered to be generated by some underlying process, which is currently unknown. If the nature of the time series does not allow a mathematical function to be used to exactly predict future values of the series then the underlying process is said to be stochastic. The word stochastic means random and indicates that the mechanism generating the values of the time series involves probability. In fact, the observed time series is considered to be a realization of a stochastic process. Since the generating mechanism involves probability, the observed time series $z_1, z_2, ..., z_N$ is regarded as a realization of the values of N random variables $Z_1, Z_2, ..., Z_N$ where z_1 is only one of a possible number of values for the variable Z_1 but is the value which happened to be observed at that point in time. The future values of the stochastic process are regarded as random variables and identifying the underlying process involves identifying the probability distribution of these random variables.

Each of the random variables Z_t has its own probability distribution with its own probability density function $f(z_t)$ depending on whether the variable being forecasted is discrete or continuous.

Each random variable Z_t has a mean,

$$\mu_t = E(Z_t) \qquad [4.3]$$

and a variance,

$$\sigma_t^2 = \sigma^2(Z_t) \qquad [4.4]$$

However, if the stochastic process is stationary then μ and σ_t^2 are equal to the mean μ and variance σ_z^2 of the stochastic process respectively.

4.4.1 The Sample Autocorrelation Function

In studying two variables it is often suspected that there is an association between them. One indication of the association between the two variables A and B (or the extent to which they tend to vary together) is the correlation coefficient $\rho(AB)$ defined as:

$$\rho(AB) = \frac{E(AB) - E(A)\,E(B)}{\sigma(A)\,\sigma(B)} \qquad [4.5]$$

where

$E(AB)$ = mean of the random variable AB, the product of A and B,

$E(A)$ and $E(B)$ = mean of A and B,

$\sigma(A)$ and $\sigma(B)$ = standard deviation of A and B.

It can be shown that the correlation coefficient has a number of properties that lend themselves to the identification of an appropriate time series model. Firstly, it is dimensionless, that is, its value does not depend on the scales of measurement of A and B, also:

$$-1 \leq \rho(AB) \leq 1$$

Recall that the observed time series values $z_1, z_2, ..., z_N$ are regarded as realizations of the random variables $Z_1, Z_2, ..., Z_N$. The distinctive feature of stochastic processes and what enables a time series model to be built, is that $Z_1, Z_2, ..., Z_N$ are dependent rather than independent variables. The correlation coefficient is used to identify the form of this dependence. Consider the correlation coefficient $\rho(AB)$ between the the variables Z_t and Z_{t+k}, which are separated by k intervals of time:

$$\rho(Z_t, Z_{t+k}) = \frac{E(Z_t Z_{t+k}) - E(Z_t)\, E(Z_{t+k})}{\sigma(Z_t)\, \sigma(Z_{t+k})} \qquad [4.6]$$

In general, the correlation coefficient $\rho(Z_t, Z_{t+k})$ depends on time t as well as on the separation of k intervals of time. If, however, the correlation coefficient between values of the stochastic process at two time points depends on only the interval between the time points and not on the time itself, then the stochastic process is said to be stationary. The concept of a stationary model is most important as statistical equilibrium needs to be achieved before selection of a suitable model can take place. For a stationary process with constant mean and variance, the correlation coefficient between Z_t and Z_{t+k} is called the theoretical autocorrelation at lag k, given by:

$$\rho_k = \frac{E(Z_t Z_{t+k}) - \mu^2}{\sigma_z^2} \qquad [4.7]$$

since $E(Z_t) = E(Z_{t+k}) = \mu$ and $\sigma(Z_t) = \sigma(Z_{t+k}) = \sigma_z$, and

$$\rho_0 = 1$$

since it can be shown that $E(Z_t^2) - \mu^2 = \sigma_z^2$.

The lag k is the difference between a time t in the time series and a time t+k in the same or a different series.

The theoretical autocorrelation has the same properties as the correlation coefficient. Thus, a time series in which the current value depends on past values is called an autocorrelated series. This dependence in the series between two points separated by k time units (the lag) can be described by the autocorrelation coefficient at lag k. It measures the extent to which a value of the series above or below the mean at time t tends to be followed by a value of the series above or below the mean k time units later.

A plot of ρ against the lag k for k = 1,2, ... is called the theoretical autocorrelation function of the time series. It describes how the correlation in the series dies out as the lag increases and can be used to decide if the series is stationary and, if so, to select an appropriate model from a class of stationary stochastic processes.

In practice, it is only possible to obtain estimates of the autocorrelations. A number of estimates of the autocorrelation function has been suggested by statisticians [51] and it has been concluded that the most satisfactory estimate of the kth lag theoretical autocorrelation r_k is:

$$r_k = \frac{c_k}{c_0}$$

where

$$c_k = \sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z}), \qquad k = 0, 1, 2, \ldots, K \qquad [4.8]$$

is the estimate of the autocovariance γ_k, given by,

$$\gamma_k = \rho_k \sigma_z^2$$

and $\bar{z}$ is the mean of the time series. Thus the sample autocorrelation function is given by:

$$r_k = \frac{\sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=1}^{N} (z_t - \bar{z})^2} \qquad [4.9]$$

This is known as the sample autocorrelation function, which when dealing with observed data can be referred to as simply the autocorrelation function or ACF. In practice, a minimum of 50 observations are needed to obtain a useful estimate of the autocorrelation function.

4.4.2 The Sample Partial Autocorrelation Function

The sample partial autocorrelation function is another tool used in the selection of an appropriate time series model. The autocorrelation function can reveal whether the original series is stationary and possesses characteristics of theoretical time series models but in itself is unable to identify these models. The theoretical partial autocorrelation function can be thought of as the autocorrelation between any two variables Z_t and Z_{t+k}, separated by a lag of k time units, with the effects of intervening variables Z_{t+1}, Z_{t+2}, ..., Z_{t+k-1} eliminated. It may be shown that $\rho_{11} = \rho_1$. The theoretical partial autocorrelation function is a listing or plot of ρ_{kk} for lags k = 1, 2, ...K.

The sample partial autocorrelation function r_{kk} is an estimate of ρ_{kk} calculated from observed time series values and are derived from the sample autocorrelations r_k by means of the following equations:

$$r_{11} = r_1$$

$$r_{22} = \frac{r_2 - r_1^2}{1 - r_1^2}$$

with more complicated equations for r_{33}, r_{44}, ...r_{nn}.

The sample partial autocorrelation function is referred to as the partial autocorrelation function or PACF when dealing with observed data.

As was the case for the sample autocorrelation, the sample partial autocorrelations are only estimates of the corresponding theoretical correlations. However, both the sample autocorrelation and partial autocorrelation functions tend to follow the same pattern as their theoretical counterparts, so they can be used to help identify an appropriate model for the underlying stochastic process.

4.5 The General Linear Process

A stochastic process can be represented as the output from a linear filter whose input is purely random [52]. Recall that the observed time series in which successive values are dependent, is regarded as a realization of a stochastic process, which is generated from a series of independent random shocks. These shocks are assumed to be normally and independently distributed with mean zero and variance σ^2.

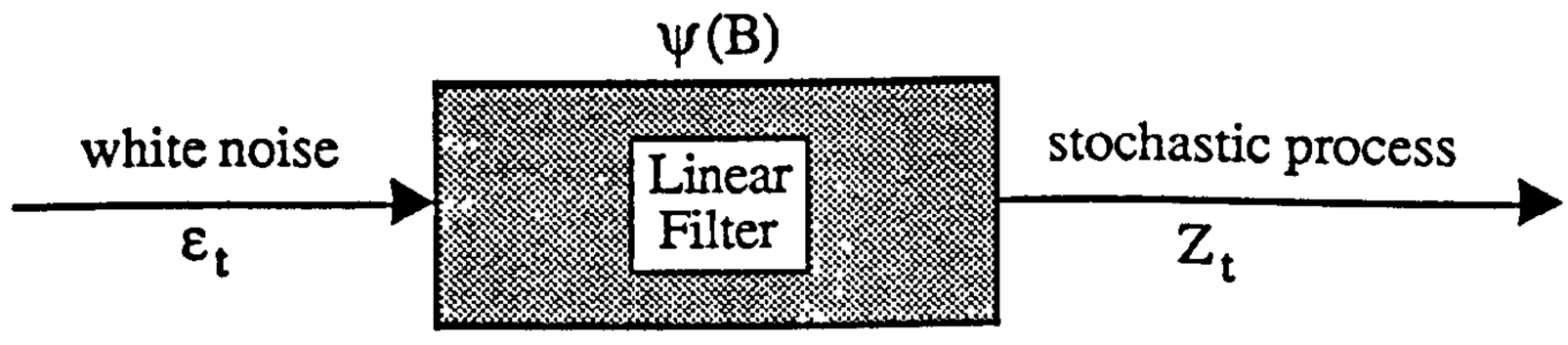


Figure 4.3 The Linear Filter.

The linear filter causes the independent random shocks ε_t to be transformed into the stochastic process Z_t. Thus the stochastic process can be written as follows:

$$\tilde{Z}_t = \varepsilon_t + \psi\varepsilon_{t-1} + \psi\varepsilon_{t-2} + \ldots \qquad [4.10]$$

$$= \varepsilon_t + \sum_{j=1}^{\infty} \psi\varepsilon_{t-j}$$

where

$\tilde{z}_t$ = $Z_t - \mu$, is the deviation of the process from some arbitrary origin, or from its mean, if the process is stationary and,

ψ = a weight coefficient.

This is known as the general linear process and allows $\tilde{Z}_t$ to be represented as a weighted sum of past and present values of a purely random process ε_t. It is conventional to refer to the purely random process as white noise.

4.6 The White-Noise Model

The simplest possible model for a stationary time series is one in which the variables Z_t are independent. This means that the series is completely random and the variables are independently and normally distributed with mean μ and variance σ_z^2, expressed as:

$$Z_t = \theta_0 + \varepsilon_t \qquad [4.11]$$

where

$\varepsilon_t \sim IN(0, \sigma_\varepsilon^2)$ and

$Z_t \sim IN(\mu, \sigma_Z^2)$.

Thus, the sequence consists of a series of random shocks ε_t with mean zero and constant variance and is known as a white-noise model. The parameter θ_0 is a constant term which is equal to the process mean μ when the variables Z_t are independent and σ_Z^2 (the variance of the variables Z_t) equals σ_ε^2.

The main characteristic property of a white-noise model is that the variables are completely random and this is reflected in the behaviour of the theoretical autocorrelations and partial autocorrelations which are zero for all lags. It is, however, unusual for a stochastic realization of an observed time series to possess sample autocorrelations and partial autocorrelations which are completely zero. The importance of the white-noise model becomes apparent when it is realized that if an observed series displays these characteristics it will not be possible to accurately forecast future values because the relationships between the variables are seen to be random. In such cases where this occurs and it was believed that some connection would be apparent then it may be possible to transform the format (i.e. the time interval between observations) of the observed series to obtain a degree of dependence.

The most important use of the white-noise model lies in its application to the remaining residual element after fitting an appropriate model to the observed time series. The principal objective in building a time series model is to fit a model to the series so that what is left unexplained by the model is a random series, or white-noise. If this is so, then the best possible forecasts will be achieved by the fitted model.

To illustrate the limited use of an observed time series which is found to be a realization of a white- noise process, consider the forecast of the value Z_{t+1}. Firstly, the white-noise model contains two unknown parameters θ_0 and σ_ε^2. Because of their relationship with the process mean and process variance of the variables it can be shown that:

$$\theta_0 = \mu$$

and

$$\sigma_\varepsilon^2 = \sigma_z^2$$

Estimates of θ_0 and σ_ε^2 are $\bar{z}$ and $\hat{\sigma}_z^2$, respectively, and are the mean and variance of the observed time series.

In a random series the forecast of the variable Z_{t+l} (written as $\hat{z}_t(l)$) is simply the ordinary expected values of Z_{t+l}, which is the mean $\mu = E(Z_{t+l})$. Therefore:

$$\hat{z}_t(l) \quad = \quad E(Z_{t+l}) \; = \; \mu \qquad [4.12]$$

The estimate, $\bar{z}$ of the process mean μ, is used as a point forecast of all future values of the time series which means that useful forecasts cannot be obtained from an observed time series which is identified as being random.

4.7 The Class of AutoRegressive Integrated Moving Average (ARIMA) Models

The models applied in this thesis to mine environmental data can be classified under the general heading of AutoRegressive Integrated Moving Average models, commonly referred to as ARIMA models. Such models are capable of describing both stationary and non-stationary time series. ARIMA models can be broken down into three different components. These are the autoregressive component (AR), the moving average component (MA), and the integrated component (I) which is introduced to induce a stationary condition. The general form of a non-seasonal ARIMA model of order, p, d, q, is given by:

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)\, \nabla^d Z_t \quad = \quad (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)\, \varepsilon_t \qquad [4.13]$$

where

ϕ_p = autoregressive parameters,
B = backward shift operator, i.e. $B^j Z_t = Z_{t-j}$,
∇^d = degree of differencing, d, to induce stationarity,
Z_t = observation at time t,
θ_q = moving average parameters,
ε_t = random element, ~ $IN(0, \sigma_A^2)$.

If a time series is stationary then it is possible to approximate it by either an AR(p), MA(q) or ARMA(p,q) model. Where a time series is found to be non-stationary it is necessary to make it stationary and subsequent models are referred to as ARI, IMA or ARIMA. In addition, ARIMA models can describe seasonal time series and special provision is made for series which exhibit seasonal behaviour. The necessary changes to deal with seasonal series are explained later in the chapter.

In this thesis the notation ARIMA (p,d,q) is used for models that contain only a non-seasonal component. The notation ARIMA ([a,b,c],d,[x,y,z]) is used when AR and MA parameters are fitted to specific lags. For example, the ARIMA ([1,3],1,2) consists of AR parameters fitted to lags 1 and 3, one degree of differencing and MA parameters fitted to lags 1 and 2 as normal.

4.8 The Autoregressive Model

An autoregressive model is one where the current value Z_t of the series is related to the variable Z_{t-p} where p = 1, 2, 3, ..., p, written in backshift operator form as:

$$Z_t = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p \qquad [4.14]$$

In an autoregressive process each value in a series is a linear function of the preceding value or values. The parameter ϕ is used to ascribe a weight to the importance of the Z_{t-p} variable and is an indication of how strongly each value depends on the preceding value. The process defined by this equation is called an autoregressive process of order p, or AR(p).

If p=1, then the model takes the form of an AR(1) or ARIMA(1,0,0). The AR(1) model is the simplest progression from the white noise model (ARIMA{0,0,0}) where the variable Z_t is regarded as a linear function of the preceding variable Z_{t-1} and a random element ε_t, given in difference form by:

$$Z_t = \theta_0 + \phi_1 Z_{t-1} + \varepsilon_t \qquad [4.15]$$

where

$$-1 < \phi > 1$$

This means that if the autoregressive model is to represent a stationary autoregressive time series then the p roots of $\phi_p Z_{t-p}$ must all be greater than $|1|$.

The characteristic pattern of an autoregressive process depends upon the roots and order p of the autoregressive term and the magnitude of the parameters ϕ_p. For example, in an AR(1) process the theoretical autocorrelations and partial autocorrelations are given by:

$$\rho_k = \phi_p Z_{t-p} \quad (k \geq 1)$$

and

$$\rho_{11} = \phi_1$$

$$\rho_{kk} = 0 \quad (k > 1)$$

If ϕ_1 is positive the autocorrelations decline smoothly towards zero while if ϕ_1 is negative the autocorrelations decline towards zero in an oscillatory manner, i.e. ρ_k will be positive for even lags and negative for odd lags. The partial autocorrelations for an AR(1) process cut off after the first lag and in general, are zero for all lags greater than p. Examples of theoretical autocorrelation and partial autocorrelation functions for autoregressive processes up to order p=2 are shown in Figure 4.4.

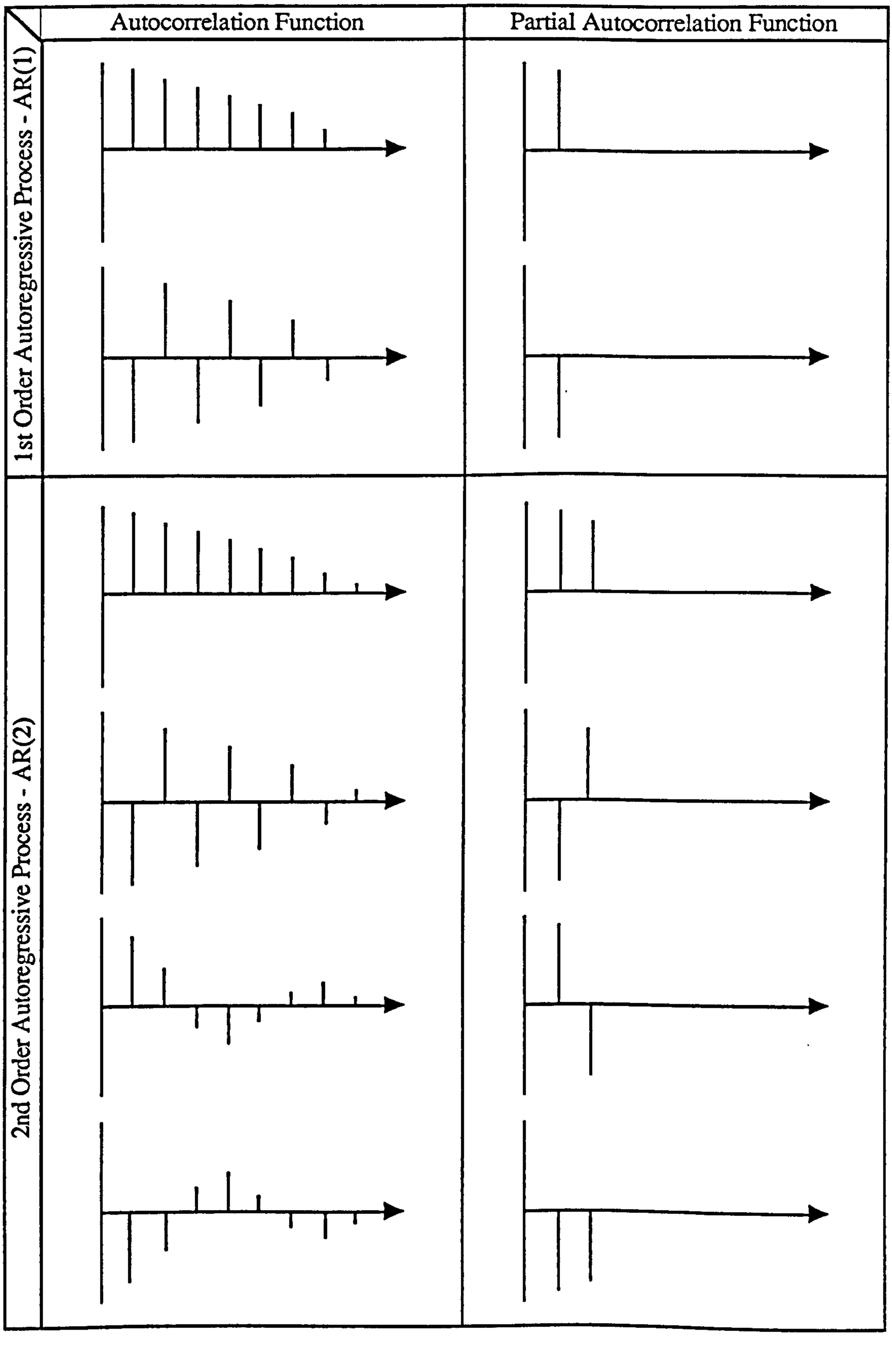


Figure 4.4 Theoretical Correlation Functions of Non-Seasonal Autoregressive Processes.

4.9 The Moving-Average Model

The moving average model is a method of representing a time series by a weighted average of its terms. A moving average process is one which is considered to be a moving average of successive terms of a white noise process, given in general form by:

$$Z_t = \theta_0 + \theta_1\varepsilon_{t-1} + \theta_q\varepsilon_{t-q} + \varepsilon_t \qquad [4.17]$$

The parameter θ_q is known as the moving average coefficient of order q. In a moving average process, each value is determined by the average of the current disturbance and one or more disturbances up to order q. The simplest case of a moving average model is one where q=1

$$Z_t = \theta_0 + \theta_1\varepsilon_{t-1} + \varepsilon_t \qquad [4.18]$$

In this model the variable Z_t is a constant plus a moving average of the current random element ε_t and the previous random element ε_{t-1}. This is called a moving average model of the first order or MA(1).

From equation 4.18 it follows that $\mu = E(Z_t) = \theta_0$ which means that the parameter θ_0 is the mean of the process. Also:

$$\sigma_z^2 = (1 + \theta_1^2)\,\sigma_k^2 \qquad [4.19]$$

From this equation it can be deduced that the moving average coefficient does not need to satisfy any conditions for stationarity and all moving average processes are stationary. However, it is necessary for the parameters in a moving average model to satisfy what is known as the invertibility condition if it is to produce efficient forecasts. The condition for invertibility is:

$$-1 < \theta < 1$$

and means that it is necessary for the q roots of an MA model to lie outside the unit circle [53].

As with AR models, the characteristic pattern of an MA process depends on the roots and order of the MA term and the magnitude of the parameters θ_q. For example in an MA(1) process the theoretical autocorrelations are given by:

$$\rho_1 = \frac{-\theta_1}{1 + \theta_1^2} \qquad [4.20]$$

This means that the autocorrelation function cuts off after the first lag. In general the autocorrelation function of a MA(q) model cuts off after lag q.

The partial autocorrelation function of an MA process gradually die down to zero, either in an oscillatory manner or as wholly positive or negative lags. Examples of theoretical autocorrelation and partial autocorrelation functions for moving average processes up to order q=2 are shown in Figure 4.5.

4.10 Mixed Processes

Commonly, when fitting a model to time series data it is found that the series is best represented by a combination of AR and MA terms. In such cases, the characteristics of the autocorrelation and partial autocorrelation functions exhibit lag structures which are common to both AR and MA processes but are more difficult to identify.

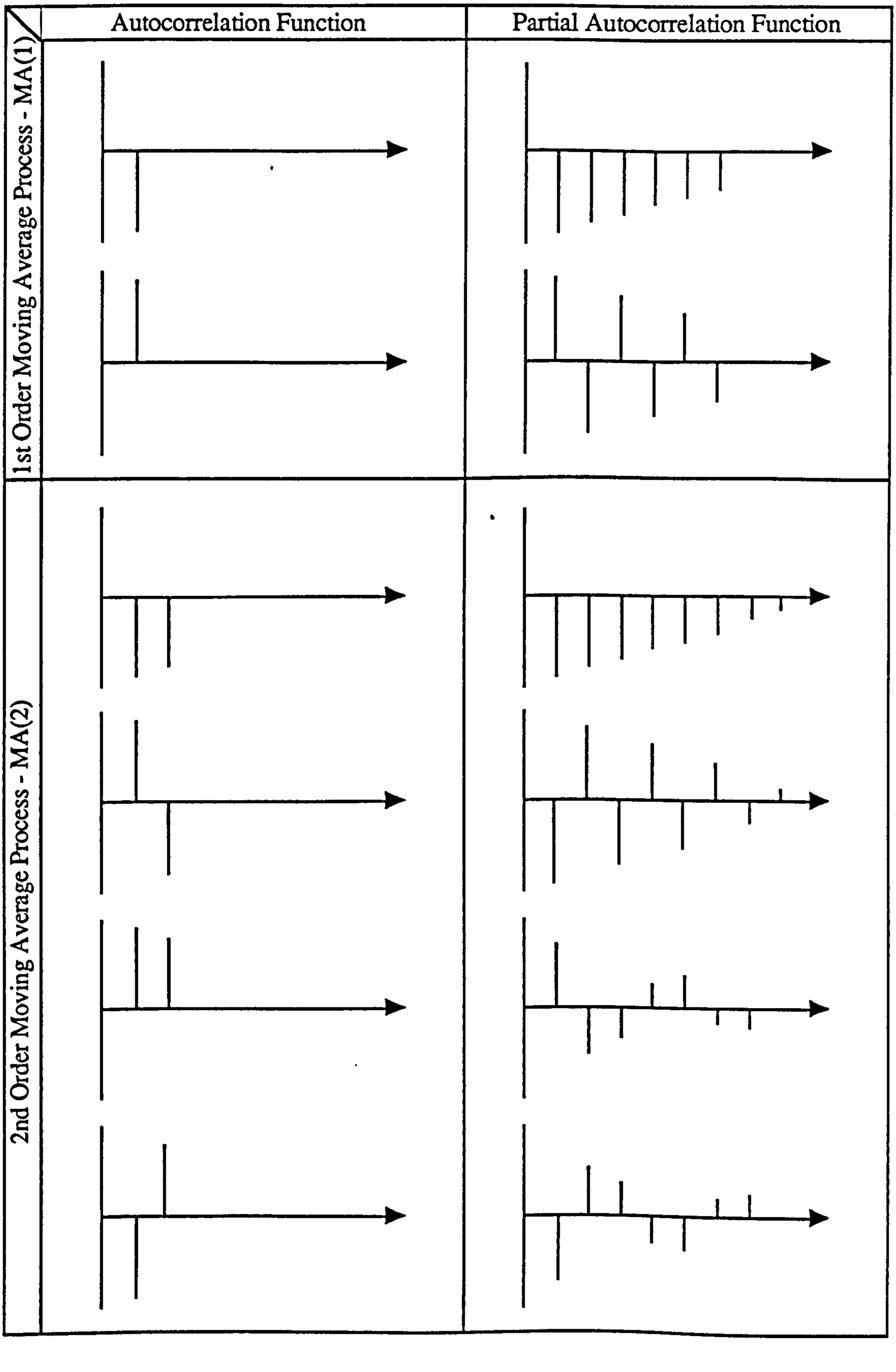


Figure 4.5 Theoretical Correlation Functions of Non-Seasonal Moving Average Processes.

4.11 Differencing a Time Series to Induce Stationarity

The concept of a time series exhibiting stationarity is very important. A time series needs to be stationary if it is to be modelled by the general class of ARMA models. Differencing is necessary because a time series often reflects the cumulative effect of some process. The stochastic process is not responsible for the observed level of the series but does cause the changes in the level. Thus, the observed time series may have to be differenced before a model can be fitted to it.

4.12 Box-Jenkins Univariate Methodology

There are many reasons for wanting to build a model of a real-life system. One of them might be to gain an understanding of what mechanisms are influencing the outputs of the system. However, an important objective is usually to use the time series model to forecast future values of a variable, either for general information or use in a control system.

The objectives will lead to the formulation of certain hypotheses about the time series data and to the need to collect data that can be used to verify them. For example, it is believed that changes in barometric pressure temporarily affects the rate of methane emission, but is this relationship evident from a time series model of the two variables, barometric pressure and methane emission? Thought must be given on what type of data is required and in what form it should be collected. To build a model from the data requires a model building methodology that is flexible. This allows the model to adapt to account for a hypothesis which does not explain the data and so be replaced by ones which do. The time series model is examined at every stage of its development to determine whether it fulfils the mathematics of the analysis. When a suitable model has been found, it is further tested to see how well it fulfils the original objectives and whether it gives accurate forecasts or requires further modification. All of the processes involved during the building of a time series model are designed to follow a logical path to produce the best possible model of the time series and these are illustrated by a flowchart in Figure 4.6. Some guide-lines for the building of models for forecasting, planning and control are [54, 55],

1. Understand the problem and the purpose of building the model.
2. Understand the decision-taking system which the model will serve.
3. Work out from the start how the model is to be implemented.
4. Structure the quantitative model by building a conceptual model of the appropriate environmental system, displaying the mechanisms involved.
5. Select the data carefully, understand its limitations and plot it in a variety of ways.
6. Aim for simple models, involving few variables first and then elaborate later, if necessary.
7. Proceed iteratively via,
   - Identification (Specification),
   - Estimation (Fitting),
   - Checking (Criticism).
8. Aim for parsimony in parameterization - avoid over parameterization.
9. Understand what the model has to say about the data.
10. Conduct experiments with the model (simulations) to understand its limitations.
11. Present the results of the model in simple terms to those that have to use it.

There are three stages in the building of a univariate time series model as advocated by Box and Jenkins, namely, identification, estimation and diagnostic checking and these are discussed in the following sections. The approach to building a Box-Jenkins time series model has been developed into a systematic and logical strategy.

4.13 Identification

At the identification stage the objective is to select an appropriate model which is thought to be representative of the observed time series. At this stage it is possible to identify a number of potential time series models which appear to explain the movements of the time series data and they can all be taken on to the estimation stage of model development. The actual identification of a potential model involves the use of the sample correlations (both auto and partial) of the observed time series. After these are generated they are compared to the theoretical correlations of various models and a model is selected on the basis of resemblance between a model's theoretical correlations and the time series sample correlations. Recall from 4.4 that the observed time series is assumed to be a realization of some underlying stochastic process and from 4.4.1 and 4.4.2 that only the sample

correlations of the observed time series can be calculated. This means that the sample auto and partial correlations will not behave exactly as their theoretical counterparts and allowances have to be made for this when identifying a possible model from the correlations. It is important to emphasize here that there is often difficulty in identifying a correct model and only the general behaviour of the correlations can be considered rather than a close scrutiny of detail. Fortunately, if a model is incorrectly chosen at this stage it will quickly become evident in later stages and can be discarded.

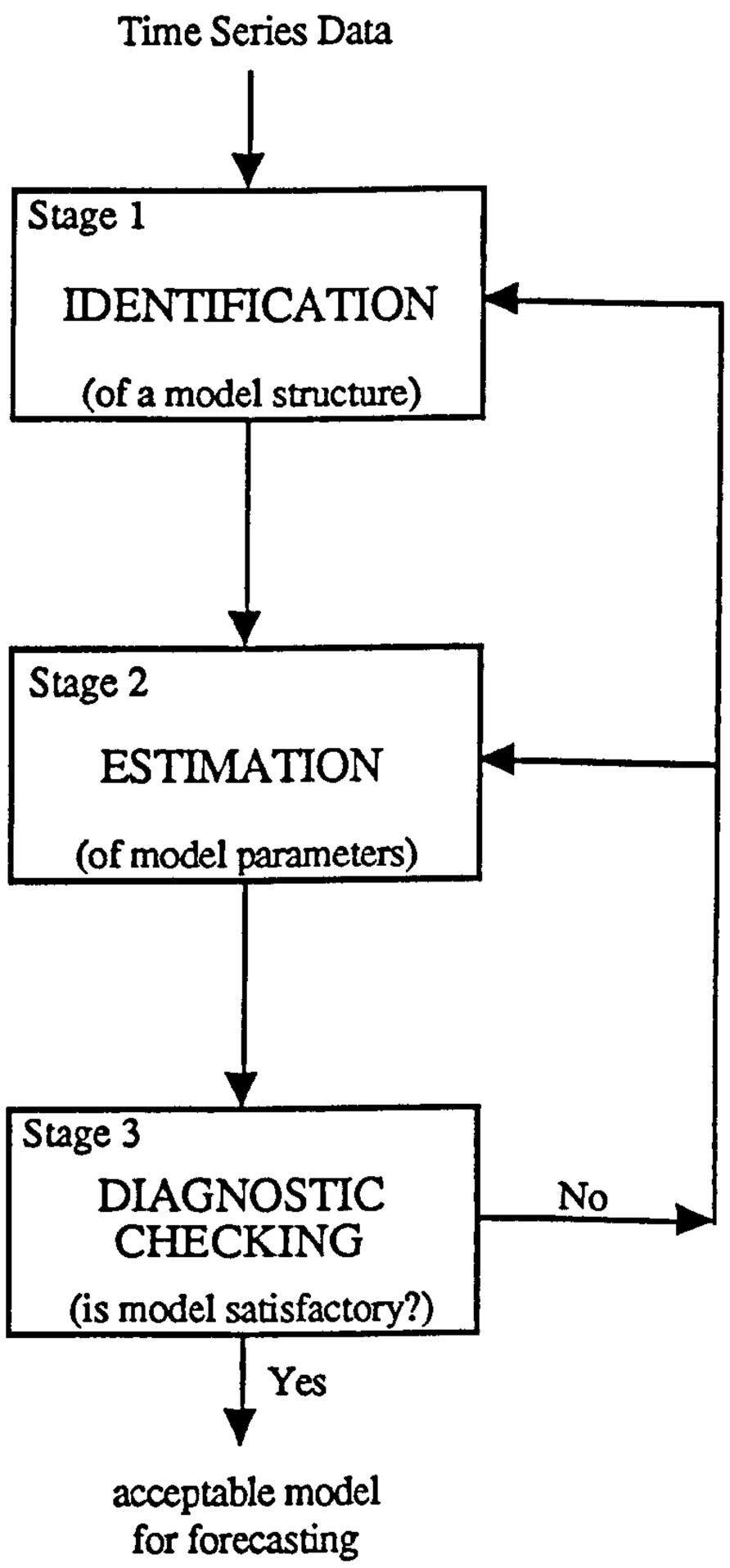


Figure 4.6 The Box-Jenkins Model Building Methodology.

If it is undecided whether the data is stationary an inspection of the correlations of the original time series data will usually indicate if this condition is true. A series that is non-stationary displays sample autocorrelations that do not die out quickly but continue for very high lags and fade away slowly. Differences of the data are taken until the sample autocorrelations display stationary characteristics. The characteristic properties of six stationary ARIMA models are shown in Table 4.2

Model	Theoretical ACF	Theoretical PACF
White Noise	all autocorrelations zero	all partial autocorrelations zero
AR(1)	tails off to zero	cuts off after lag 1
AR(2)	tails off to zero	cuts off after lag 2
MA(1)	cuts off after lag 1	tails off to zero
MA(2)	cuts off after lag 2	tails off to zero
ARMA(1,1)	tails off to zero	tails off to zero

Table 4.2 Characteristic Properties of Six ARIMA Models.

As an example consider the sample autocorrelations and partial autocorrelations of the methane concentration data, the first ten data points of which are illustrated by Figure 4.2. They are shown in Figure 4.7 and are those for a series of 430 data points with one degree of differencing. The plot of the autocorrelations for the raw series (not shown) indicated that the series was non-stationary and needed differencing. After differencing, and from an inspection of the plot of autocorrelations and partial autocorrelations the series is evidently stationary and appears to be neither a pure autoregressive or moving average process. A tentative identification might be that the series is best represented by an ARIMA (1,1,1) model. Indeed, for this example it is not easy to identify what the correct model to represent the series may be and in cases such as this it is best to consider all of the six common ARIMA models and eliminate inappropriate ones later.

In order to decide whether the correlations are significant in the identification of an appropriate model probability theory is used. A suitable guide-line for deciding on correlation significance is whether the value is greater than two standard deviations. This is known as the Quernouille [56] statistic and means that one standard deviation can be approximated by the reciprocal of the square root of the total number of observations and is equal to $1/\sqrt{N}$. Thus, any auto or partial correlation that is greater than $2/\sqrt{N}$ is regarded as significant and should be taken into account when identifying a suitable

```
Autocorrelations:   C4-Methane Concentration Data
 Transformations:  difference (1)
      Auto- Stand.

Lag  Corr.    Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1   Box-Ljung  Prob.
                   +----+----+----+----+----+----+----+----+----+
  1   .209    .048                    . |*.**                      18.970   .000
  2   .106    .048                    . |**                       23.818   .000
  3  -.072    .048                    .*| .                        26.075   .000
  4  -.102    .048                    **| .                        30.623   .000
  5   .078    .048                    . |**                       33.297   .000
  6   .024    .048                    . * .                       33.542   .000
  7   .095    .048                    . |**                       37.469   .000
  8  -.001    .048                    . * .                       37.470   .000
  9  -.035    .048                    .*| .                        38.023   .000
 10  -.107    .048                    **| .                        43.113   .000
 11  -.115    .047                    **| .                        49.020   .000
 12  -.027    .047                    .*| .                        49.342   .000
 13   .021    .047                    . * .                       49.533   .000
 14   .028    .047                    . |*.                       49.878   .000
 15  -.017    .047                    . * .                       50.014   .000
 16  -.053    .047                    .*| .                        51.292   .000
 17  -.005    .047                    . * .                       51.302   .000
 18   .035    .047                    . |*.                       51.853   .000
 19  -.038    .047                    .*| .                        52.515   .000
 20  -.006    .047                    . * .                       52.529   .000

 Plot Symbols:       Autocorrelations *     Two Standard Error Limits .
 Total cases:  431     Computable first lags after differencing:  429

Partial Autocorrelations:   C4-Methane Concentration Data
 Transformations:  difference (1)
    Pr-Aut- Stand.

Lag  Corr.    Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1
                   +----+----+----+----+----+----+----+----+----+
  1   .209    .048                    . |*.**
  2   .065    .048                    . |*.
  3  -.112    .048                    **| .
  4  -.078    .048                    **| .
  5   .140    .048                    . |*.*
  6  -.008    .048                    . * .
  7   .054    .048                    . |*.
  8  -.026    .048                    .*| .
  9  -.025    .048                    . * .
 10  -.097    .048                    **| .
 11  -.062    .048                    .*| .
 12   .008    .048                    . * .
 13   .028    .048                    . |*.
 14  -.015    .048                    . * .
 15  -.025    .048                    .*| .
 16  -.025    .048                    . * .
 17   .041    .048                    . |*.
 18   .044    .048                    . |*.
 19  -.084    .048                    **| .
 20  -.009    .048                    . * .

 Plot Symbols:       Autocorrelations *     Two Standard Error Limits .
 Total cases:  431     Computable first lags after differencing:  429
```

Figure 4.7 Autocorrelations and Partial autocorrelations of Differenced Methane Concentration Data.

model. However, this statistic is only a guide-line and although a correlation value below this is said to be small, there is a possibility that it could be significant.

The identification of an appropriate ARIMA process depends on the interpretation of the patterns shown by the sample autocorrelations and partial autocorrelations. AR(p) processes are identified mainly from the behaviour of the partial autocorrelations. These usually remain significant for the first p lags after which there is a cut-off point and the rest of the partial autocorrelations are below the $2/\sqrt{N}$ significance level. The autocorrelations of AR(q) processes do not lend themselves to model identification and display a pattern of a geometric decay curve. MA(q) processes are characterized by autocorrelations of a stationary series that are greater than two standard deviations up to and including lag q, after which there should be a cut-off point when the remaining sample autocorrelations should be small. The partial autocorrelations in an MA(q) process tend to decay beyond q with increasing k. Mixed processes, however, are considerably more difficult to identify. The sample autocorrelations generally have large values, with no definite pattern up to and including lag q after which they die out, similar to the autocorrelations of an AR(p). The sample partial autocorrelations tend to die out as the lag increases. Thus, mixed processes are characterized by auto and partial correlations that eventually die-out, instead of displaying an abrupt cut-off point.

4.14 Estimation

Once a tentative model has been identified, the next step in the Box-Jenkins model building process is the estimation of the model parameters. This procedure is fully automatic and is carried out by computer. To estimate the value of the coefficients for an AR(p) process, ordinary least squares regression methods are employed. For models that possess MA(q) terms, ordinary least squares estimation is inadequate and non-linear least squares is required. For example, consider an ARMA(1,1) process written as:

$$(1 - a_1B)\, Z_t = \quad (1 + b_1B)\, \varepsilon_t \qquad [4.21]$$

The coefficients a_1 and b_1 are the sample estimates of the autoregressive term ϕ_1 and the moving average term θ_1 respectively. Non-linear least squares performs iterative

calculations to determine the maximum-likelihood values of the model coefficients, or more specifically, aims to minimize the sum of the squares of the residuals $\sum\varepsilon_t^2$, estimated by $\sum e_t^2$, the sum of the squares of the observed residuals. Some computer programmes require starting values for a_1 and b_1 to be supplied before iterative calculations can begin. In such cases it is necessary to check that these preliminary estimates satisfy the stationarity and invertibility conditions required by the identified model. The minimization procedure continues over a range of values for a_1 and b_1 until no significant reduction in $\sum e_t^2$ occurs. Initial values for the model coefficients are found by considering the autocorrelations of the observed time series. For example, for an AR(1) model it can be shown that:

$$\rho_1 = \phi_1$$

and thus an initial estimate of the parameter ϕ is simply the sample autocorrelation r_1.

4.15 Diagnostic Checking

Diagnosing the suitability of the estimated model parameters is the final and most lengthy step in the Box-Jenkins ARIMA model building procedure. Ideally, one of the models tentatively identified in the first step will prove to be the right choice and so will be a satisfactory representation of the observed data. Others that are subsequently shown to be unsatisfactory are usually discarded but the nature of their inadequacies may suggest how and if they can be improved by respecification.

The whole model building philosophy aims to produce a parsimonious model. A parsimonious model is one that represents the data adequately with the minimum number of specified parameters. Thus, if it is found that more than one model fits the data sufficiently well the model with the fewest parameters is the one that should be chosen for use. Often, models that do not adhere to this rule possess badly estimated parameters and do not give rise to useful forecasts.

There are a number of checks that can be used to determine whether there is any evidence of inadequacy in a specified model. The first check normally performed is to see if the parameter estimates satisfy stationarity and invertibility requirements. If these have not been met, the tentative model is unacceptable and may mean that the observed time series has not been correctly differenced. Many checks are peculiar to the particular statistical computer package that is used to analyse the data. In this thesis the computer package SPSS-X Trends™ was used to perform the ARIMA analysis and provides for a comprehensive selection of diagnostic checks. After parameter estimation it reports several statistics describing how well the model fits the data [57]. Among the quality-of-fit statistics are two termed 'AIC' and 'SBC', known as the Aikaike Information Criterion and the Schwartz Bayesian Criterion, respectively. Of the models choosen to represent the series, these statistics can be used to choose between them and generally the model containing parameters with the lowest AIC or SBC is the best. Another important test on the significance of the model parameters is the t-test. If the t-statistic is greater than two, the parameter can be regarded as being significantly different from zero and should be included in the model. Parameters with a t-statistic of less than one are not thought of as significant and are removed from the model. When the t-statistic has a value between one and two it is difficult to decide if the parameter should be dropped or not but it is usually wise to experiment on the effect of removing the parameter. If a model is respecified on the basis of a parameters t-statistic it is necessary to repeat the estimation and diagnosis steps for the new specification. The t-test also provides a good indication of whether a model has been over-specified.

The checks described so far are essential preliminary tests on the estimated model parameters. Since the primary objective of time series analysis is to fit the observed data with a stochastic representation and result in a residual series that is random, i.e. with zero mean and constant variance, the main checks on model adequacy are applied to the residuals. By examining a plot of the residual series it is often seen that their behaviour appears to be random but further checks have to be made for proof. The autocorrelations and partial autocorrelations of the residuals should demonstrate white noise behavioural characteristics i.e. they should have correlations that are zero or nearly zero. If any of the correlation values exceed the $2/\sqrt{N}$ significance limits it is probable that the model has been under-specified and needs additional parameters. To help assess the significance of the correlations use is made of the Box-Ljung statistic [58]. The Box-Ljung statistic is also known as the modified Box-Pierce statistic, given by:

$$Q = N(N+2) \sum_{k=1}^{k} \frac{r_k^2}{N - k} \qquad [4.22]$$

where

Q = Box-Ljung statistic

N = number of residuals

r = sample autocorrelation at lag k.

SPSS-X Trends™ computes the Box-Ljung statistic and its significance at each lag. If Q is large it is because the sample autocorrelations are large and indicates that the model is inadequate. The statistic shows that if the correct ARIMA (p,d,q) model has been identified and estimated, Q has a chi-square distribution with k-p-q degrees of freedom. If the value of Q exceeds a critical value determined from chi-squared tables, the residuals are not white noise and the model needs to be respecified. The significance of Q can also be assessed by its associated probability value.

Another test of model adequacy is to overfit the initial identification. It is possible that by adding further parameters the fit of the original model can be improved, even though diagnostic checks did not detect anything untoward. Overfitting must be done with care otherwise redundant over-parameterization can easily result. Additional autoregressive and moving average parameters are introduced individually to determine their effect and are not added simultaneously. After each new parameter has been introduced, checks are made to assess whether the augmented model fits the data better than the original one.

A further test of model adequacy involves splitting the observed time series into two parts before the model building procedure begins. The greater portion of the series is used as normal to identify an appropriate model. Once an adequate model is available it is used to generate forecasts for the remaining 'out-of sample' portion of the series. The forecasts are compared with the actual values of the remaining series and the mean square forecasting errors are calculated. This form of diagnostic testing is known as out-of sample testing, and is a means of testing the forecasting performance of the estimated model. The model with the lowest mean square forecasting error is the best one. As a final check, the model parameters are again estimated by fitting them to the complete series and performing the various in-sample checks. If all is well, the final model will be

the best that can be achieved and it is this updated model which is used for forecasting purposes.

4.16 Seasonal Models

The methods described so far have dealt with the building of time series models that do not have any seasonal component within them. A time series contains a seasonal component if it exhibits a regular pattern that repeats itself after a certain number of basic time intervals. The time between each interval is known as the period, s. In the example of possible time series at the beginning of the chapter a seasonal example was the sales of sun-tan lotion on a monthly basis. This series contained a seasonal component of period 12 in which the sales in a particular month are related to the sales in that month from year to year. The presence of a seasonal component in time series from engineering is often much more difficult to determine but the basic method of model building is similar to that used for non-seasonal series.

The general form of seasonal ARIMA models is closely linked to that given in section 4.7 for non-seasonal ARIMA models. In the most general case the seasonal effect operates alongside the non-seasonal component and so the general model contains both non-seasonal and seasonal parameters. The most common type of seasonal model that is advocated by Box and Jenkins is the multiplicative seasonal ARIMA model given by:

$$(1 - \phi_p B^p)(1 - \phi_P B^P)\nabla^d \nabla^D Z_t = (1 - \theta_q B^q)(1 - \theta_Q B^Q)\varepsilon_t \qquad [4.23]$$

where

$(1 - \phi_p B^p)$ = non-seasonal AR component of order p,

$(1 - \phi_P B^P)$ = seasonal AR component of order P,

$(1 - \theta_q B^q)$ = non-seasonal MA component of order q,

$(1 - \theta_Q B^Q)$ = seasonal MA component of order Q,

∇^d = degree of non-seasonal differencing, d, to induce stationarity,

∇^D = degree of seasonal differencing, d, to induce stationarity,

Z_t = observation at time t,

ε_t = random element, ~ $IN(0, \sigma_A^2)$.

In this thesis the notation ARIMA $(p,d,q)(P,D,Q)^s$ is used for models that contain a seasonal component.

The identification of a multiplicative model is essentially the same as that used for non-seasonal models. The identication of the seasonal P, D, Q parameter orders can either be done before or after the correct order of non-seasonal p, d, q parameters are determined. Assuming that the non-seasonal component of the series has been found, the auto and partial autocorrelations of the residual series will show evidence of seasonal behaviour if it is present. Table 4.3 shows the characteristic properties of four stationary seasonal model components of period 12.

Once both the seasonal and non-seasonal parameters have been found it is necessary to re-estimate them simultaneously so that their correct values can be found. In some instances both non-seasonal and seasonal parameters can be found to be redundant once this is done. Model diagnostic checking is identical to the methods described in section 4.15

Model	Theoretical ACF	Theoretical PACF
SAR(1)	autocorrelations at lags 12, 24, 36 tail off towards zero	partial autocorrelations at lags 12, 24, 36 cut off after lag 12
SAR(2)	autocorrelations at lags 12, 24, 36 tail off towards zero	partial autocorrelations at lags 12, 24, 36 cut off after lag 24
SMA(1)	autocorrelations at lags 12, 24, 36 cut off after lag 12	partial autocorrelations at lags 12, 24, 36 tail off towards zero
SMA(2)	autocorrelations at lags 12, 24, 36 cut off after lag 24	partial autocorrelations at lags 12, 24, 36 tail off towards zero

Table 4.3 Characteristic Properties of Four Seasonal ARIMA Model Components.

4.17 Conclusion

This chapter has examined the principles behind the Box-Jenkins approach to time series analysis and has discussed the three stages involved in building a univariate seasonal or non-seasonal time series model. The basic methodology allows models to be constructed without any knowledge of the series itself and such a model is built from a consideration

of past values of itself only. This is beneficial in that a lack of practical knowledge does not prevent a time series from being modelled and any knowledge of the physical system from which the time series data originated can subsequently be used to improve the model.

Practically, the application of any theoretical method is usually much more difficult than the ideal presented in theory and the application of this method to mining environmental and production data proves this to be true. However, before any analysis can be presented the relevant data needs to be acquired and this is the subject of the next chapter.

CHAPTER FIVE

UNDERGROUND ENVIRONMENTAL MONITORING

5.1 Introduction

The success of time series analysis depends on the acquisition of accurate and representative data. This chapter details the environmental parameters that were monitored to enable time series models to be built and the instruments that were used to monitor them. All of the data was collected from Thoresby Colliery and the development of a data acquisition procedure at this colliery is also described.

5.2 Thoresby Colliery

Thoresby Colliery was first developed in 1925 by the Bolsover Colliery Company. Production began in the Top Hard seam in 1928 and was the only mined seam for more than 40 years. In 1972 production started in the High Hazles seam and in 1977 from the Parkgate seam from which all of Thoresby's coal is now obtained. Thoresby is Nottinghamshire's and British Coal's most profitable colliery and set a new record for coal production with an output of 2,459,557 tonnes in the financial year 1989/90. Thoresby is regarded as a long-life pit with estimated reserves to enable production to continue into the 21st Century. The mine aims to have three coal faces capable of producing coal. Two of these are usually retreat whilst the third is either advance or of the Z-type.

Thoresby does not suffer from any adverse environmental conditions other than the need to limit methane emission to ensure optimum production and also the Parkgate seam is considered to be a spontaneous combustion risk. Ventilation is provided by one of two identical 3.6 m diameter single stage axial flow fans with auto-variable pitch blades driven at 990 rpm by 2.5 MW motors. These were installed in 1986/87 to increase the airflow to the Parkgate seam.

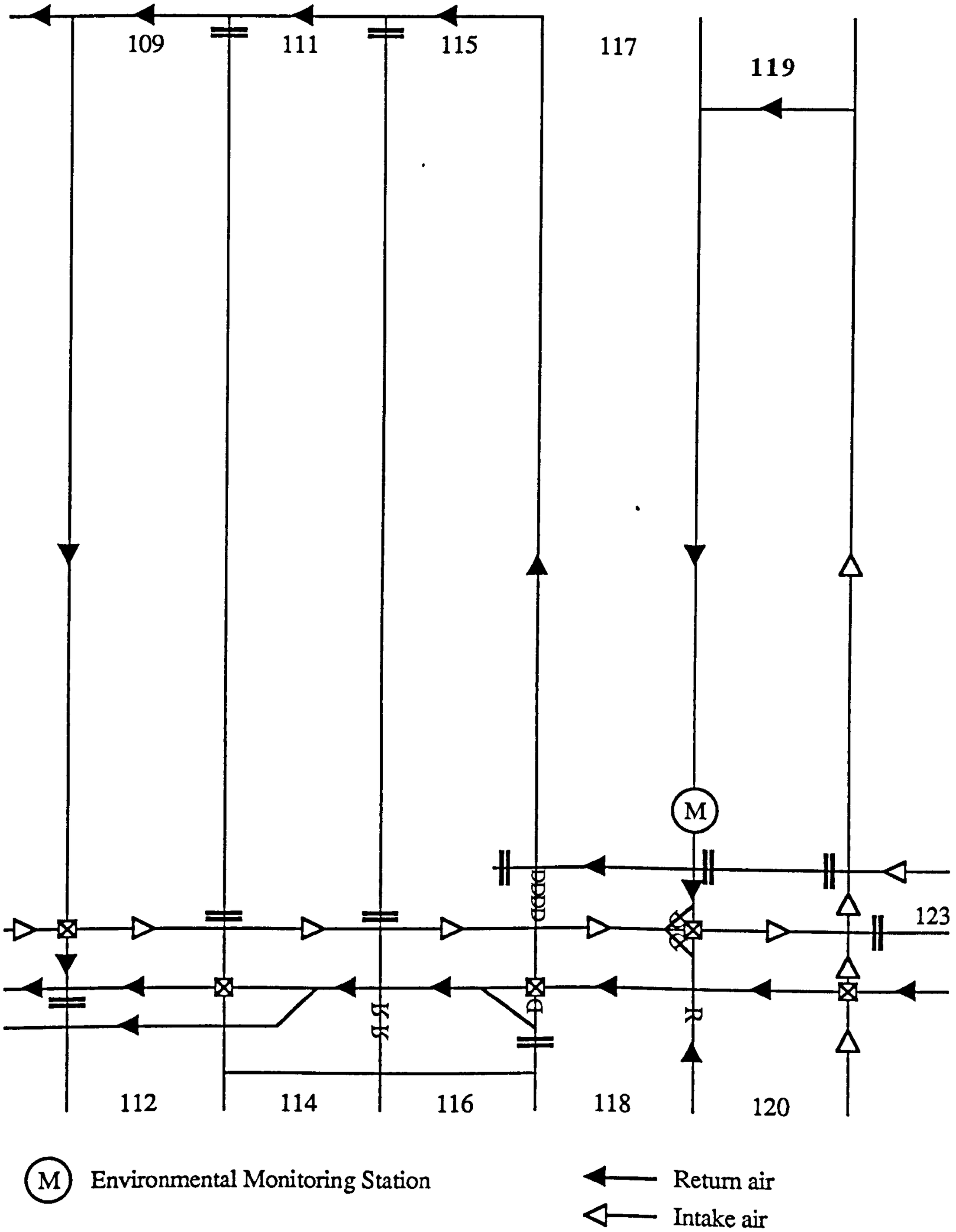


Figure 5.1 119's District Layout.

5.3 119's Retreating Face

119's district was a retreating face with a design face run of 1600 m. It commenced coal production in October 1990 and halted in September 1991 after completing its full design run. The length of the face was 255 m and the extraction height was 1.96 m. 119's was conceived as a heavy duty face and production was consistently in the region of 40,000 tonnes per week. Figure 5.1 shows its layout. The district layout is conventional U-type retreat with a split intake and a split return. The ventilation was antitropal in nature whereby the coal was conveyed in the intake gate road. Ventilation quantities were between 20 - 24 m^3/s. As a means of methane control an arrangement called the 'Sherwood Curtain' was used at the return end of the face. 119's did not suffer from excessive temperatures and was considered to be a medium spontaneous and frictional ignition risk.

5.3.1 The Sherwood Curtain

The Sherwood Curtain was developed in the Nottinghamshire Area of British Coal and is used by many Nottinghamshire collieries. It can be used on both advance and retreat faces. On an advancing face the ventilating air tends to keep the waste methane away from the men and machinery on the face side and so electrical equipment is kept free of gas. On a retreat face this does not happen and the methane fringe moves forward out of the waste at the return end of the face towards the supply gate machinery. This can cause problems if the methane concentration exceeds 1.25%. The Sherwood Curtain was developed (at Sherwood Colliery) to overcome this problem and is an arrangement where the gas fringe is held into the waste by directing the airflow at the face end [59]. The ideal arrangement of the Sherwood Curtain is illustrated by Figure 5.2.

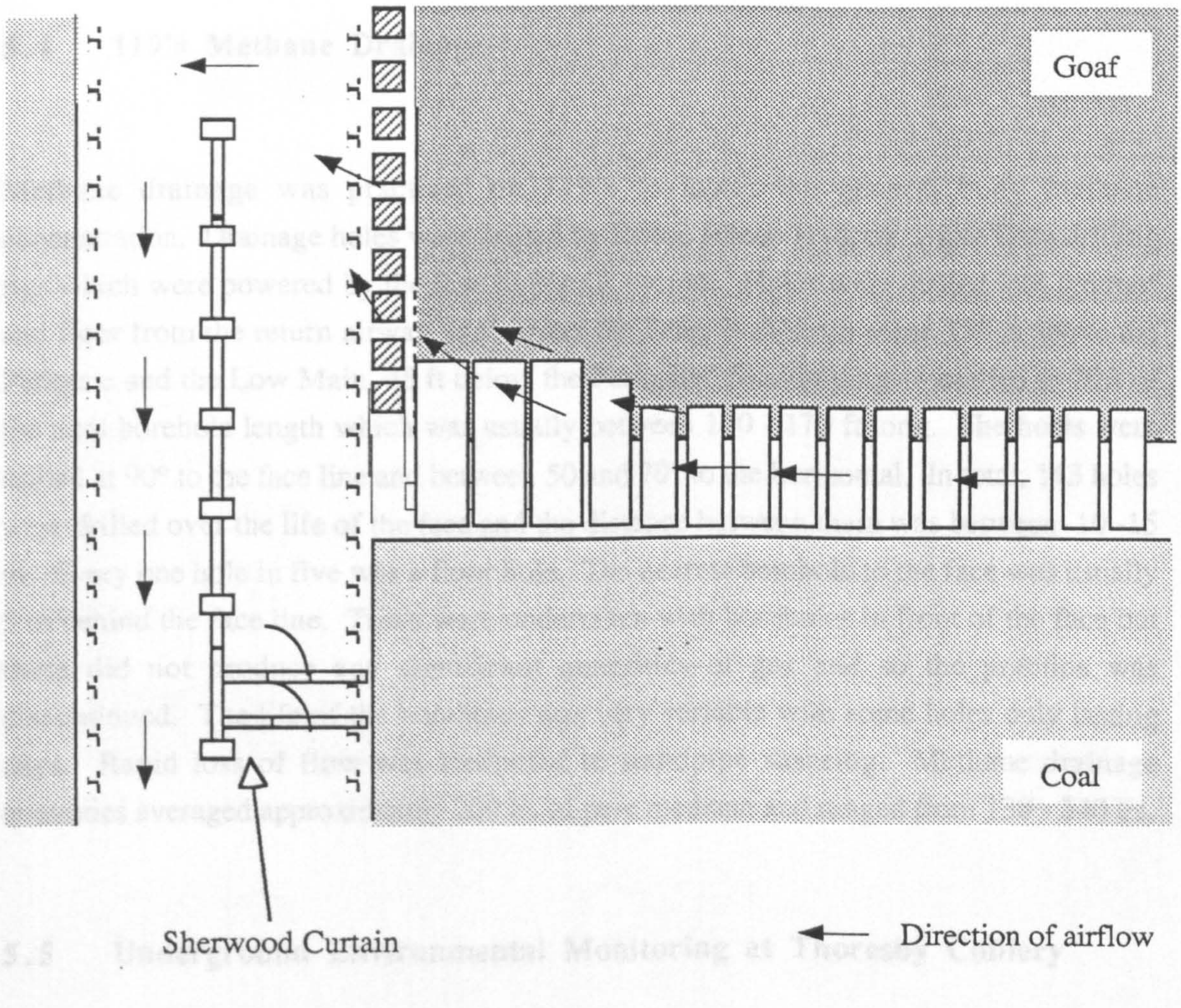


Figure 5.2 The Sherwood Curtain.

The ideal arrangement is where the pressure lock provided by the two sets of doors is intact and the air/methane mixture is forced to travel around the back of the Sherwood Curtain and so keep the gas fringe well into the waste. At the limit of the Sherwood Curtain where it is well behind the line of the goaf the author has recorded methane levels in excess of 5% (the limit of a portable methanometer) and using a high concentration device the methane concentration was found to be around 15%. In contrast, the supply gate methane readings were less than 0.5%, a sure sign of the usefulness and worth of the Sherwood Curtain. The arrangement and condition of the curtain is quite critical if it is to remain effective. The curtain does not interfere with the methane concentration or air velocity readings which are taken at the end of the supply gate. At the end of the face life when it was some 30 m away from the monitors, it was still thought to have no effect on these parameters as this distance was deemed to be sufficient for adequate mixing of the air/methane mixture.

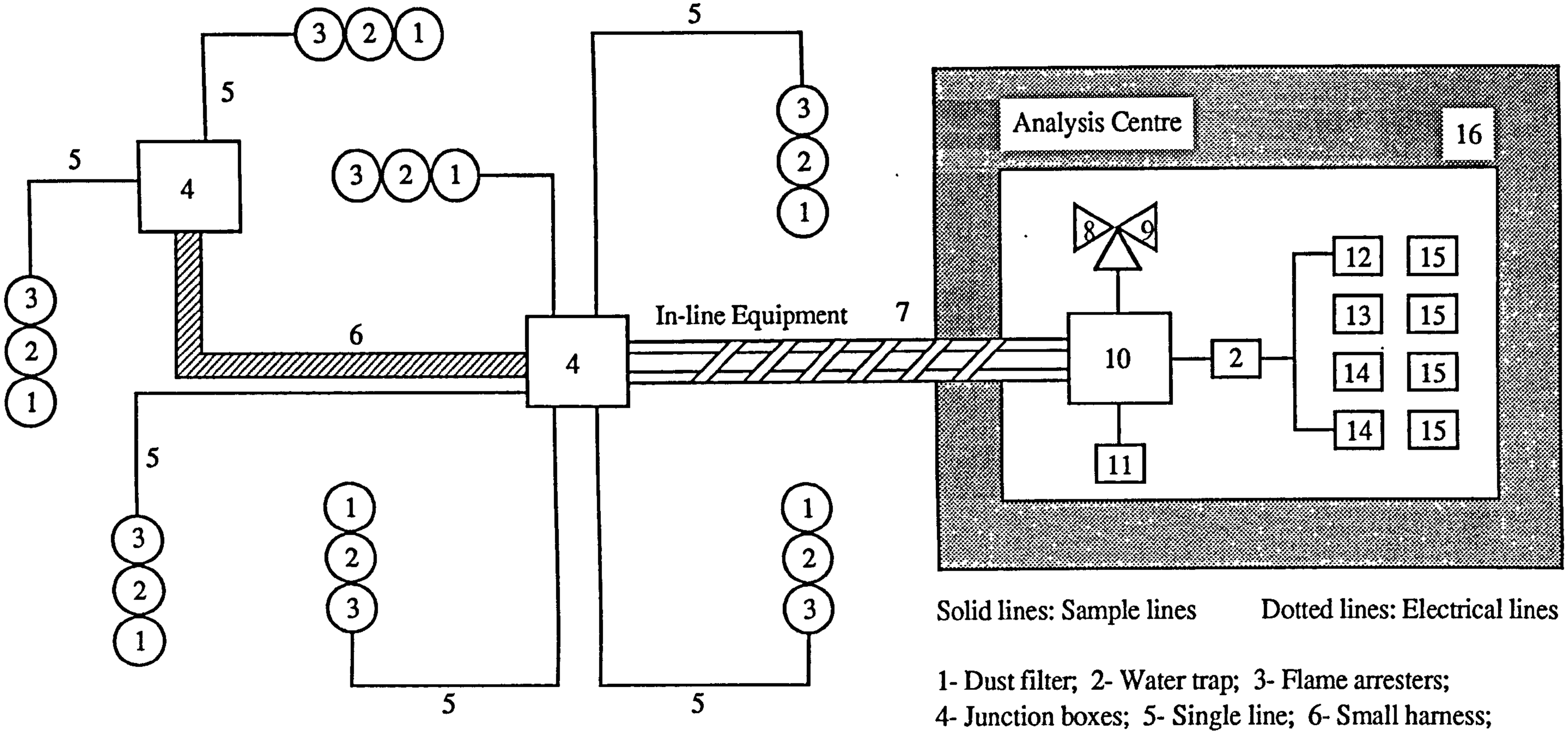
5.4 119's Methane Drainage

Methane drainage was practised on 119's to lower the general body methane concentration. Drainage holes were drilled by Edeco Minor-Hydrate Triple Ram drilling rigs which were powered by the face hydraulic system. Holes were drilled into the roof and floor from the return airway to intersect the Deep Soft seam some 130 ft above the Parkgate and the Low Main, 45 ft below the Parkgate. Standpiping amounted to 70ft of the total borehole length which was usually between 150 - 170 ft long. The holes were drilled at 90° to the face line and between 50 and 70° to the horizontal. In total, 143 holes were drilled over the life of the face and the distance between them was between 10 -15 m. Every one hole in five was a floor hole. The nearest borehole to the face was usually 5 m behind the face line. Trials were undertaken with boreholes in front of the face but these did not produce any significant quantities of gas and so the practice was discontinued. The life of the boreholes was very variable with some holes only lasting days. Rapid loss of flow was attributed to standpipe shearing. Methane drainage quantities averaged approximately 200 l/s of pure methane and ranged from 130 - 240 l/s.

5.5 Underground Environmental Monitoring at Thoresby Colliery

Thoresby Colliery has a comprehensive environmental monitoring system that was installed in the 1970's as part of British Coal's expanding environmental policy. The quality and state of the underground air must be tested for contaminants according to mining legislation which states that the quantities of noxious, asphyxiant or inflammable gases must be less than a specified level. Environmental monitoring is now a routine for all mines and is carried out not only to satisfy the legal requirement but to increase the safety of the mine in general.

Three forms of environmental monitoring are in use. The simplest and that which gives only a limited indication of environmental conditions is the use of portable detection instruments. These range from the flame safety lamp which can be used to provide an indication of the presence of methane or lack of oxygen to portable hand held detectors for a range of gases. The latest electronic instruments feature integrated capabilities that replace the flame safety lamp, spot-reading methanometers, aspirated chemical tubes, portable oxygen meters and the canary (for carbon monoxide testing) by a single unit.



Solid lines: Sample lines Dotted lines: Electrical lines

1- Dust filter; 2- Water trap; 3- Flame arresters;
4- Junction boxes; 5- Single line; 6- Small harness;
8- Standardization gases; 9- Air; 10- Control;
11- Timing control; 12- CO analyser; 13- Oxygen analyser;
14- Other analysers; 15- Recorders; 16- Display cabinet.

Figure 5.3 Schematic Diagram of a Tube-bundle System (after Fink and Adler, [60]).

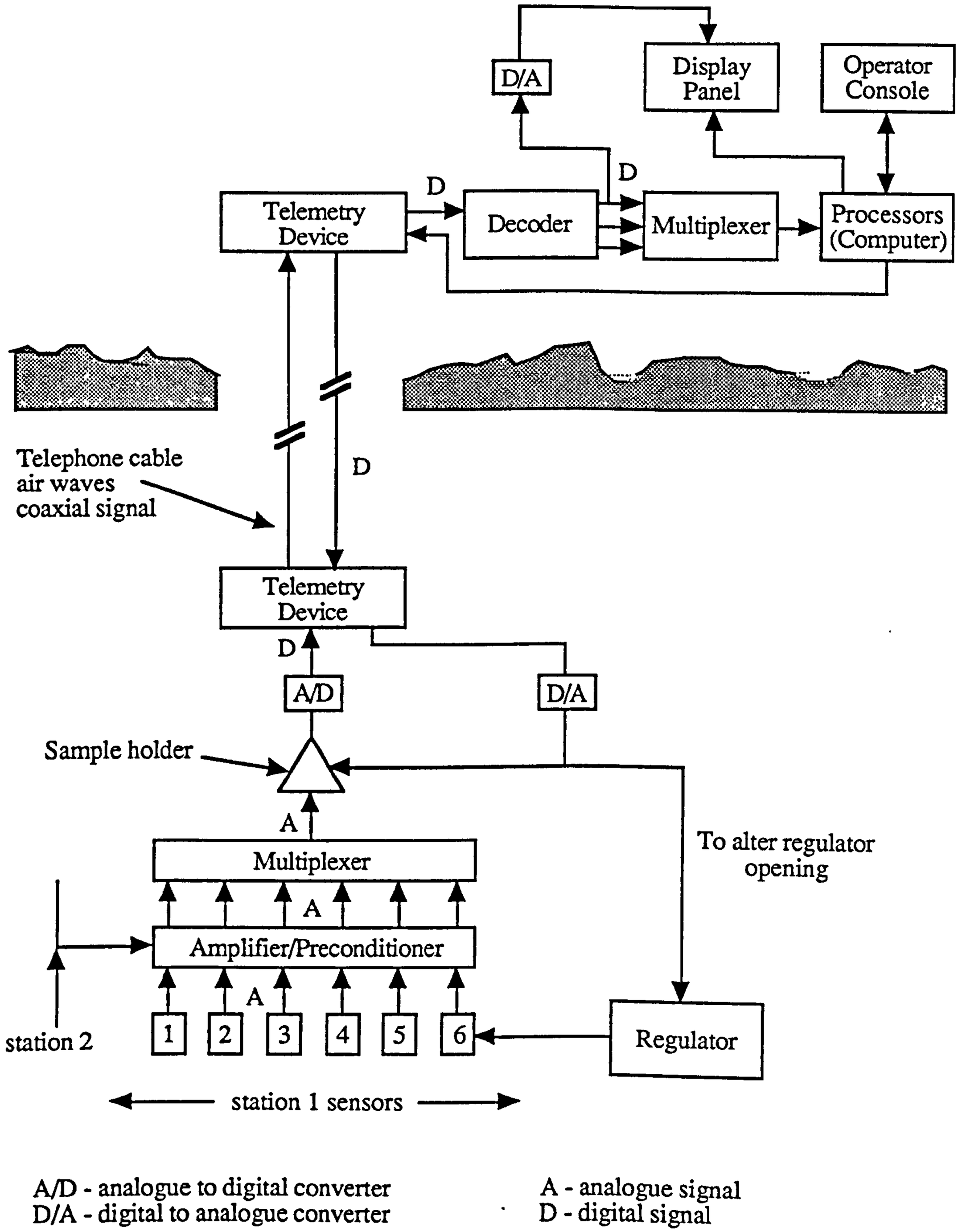


Figure 5.4 Diagram of a Real-Time Automated Monitoring and Control System (after Hormozdi [61]).

The second form of environmental monitoring is that provided by the tube-bundle system. In this system mine air is drawn through small diameter plastic tubes and analysed at the surface. A tube is taken to a particular sampling point and although each sample stays discrete a number of tubes are clustered as a bundle for easier sample transportation. The major drawback of this system is the delay associated with the travelling and sampling time of the gas. This can be up to 2 hours long and so the system cannot be regarded as providing real-time knowledge of the mine atmosphere. A typical arrangement of a tube-bundle system is shown in Figure 5.3.

The most important environmental information is that provided by a real-time system. These rely on the use of electronic instruments that provide information covering the whole spectrum of gaseous pollutants. By using transducers and electronic signal transmission the colliery control room knows instantly the condition of the mine atmosphere and there can be no doubt that such systems are invaluable to the safe and productive operation of any mine. A typical arrangement of a real-time monitoring system is shown in Figure 5.4.

5.6 Requirements and Components of a Real-Time Monitoring System

Some essential requirements of a real-time environmental monitoring system are,

1. reliable electronic transducers,
2. fast response time to changing environmental parameters,
3. instantaneous data transmission.

This list is not exhaustive and the final system should do all that is necessary to achieve information that is a true reflection of the underground conditions at an instant in time.

The basic components of a real-time environmental monitoring system are,

1. a detector or sensing head,

2. a data transmission system,
3. a computer system for data display, analyses, recording and storage.

Detector heads are used to test for a particular parameter e.g. methane concentration and provide an electric output that is in a known proportion to the quantity of contaminant in the sample. This signal is then sent by a data transmission system to the surface where it is received by a computer and stored for further use.

5.6.1 Data Transmission System

The signal or voltage produced by an environmental monitor as a response to the presence of a contaminant is transmitted to the surface by a medium known as multiplexing. Multiplexing is a method whereby a number of signals are combined and transmitted together along a single line. The analogue voltage from the monitor is converted into a digital signal. Two forms of multiplexing can be used and these are,

1. time division multiplexing (TDM),
2. frequency division multiplexing (FDM).

In time division multiplexing the signal from each monitor is transmitted digitally as a series of pulses along with an identity pulse. The number of signals that can be transmitted depends on the frequency of the pulses. Only information from one transducer is transmitted at any one time. Frequency division multiplexing is similar but varies the frequency of the transmitted signal. Each transducer signal has its own encoder with an associated frequency. Many signals can be transmitted at the same time in FDM. The FDM system is more expensive than TDM but is less susceptible to electrical interference. At Thoresby Colliery the TDM system is used.

5.7 Monitored Environmental Parameters

It was seen in chapter 2 that the release of strata gas from source beds and its subsequent migration towards the working areas is dependent upon a number of physical, geological and mining factors. A decision had to be made as to which of these factors should or could be monitored to contribute to the analysis and development of a methane prediction model. Some of the factors are very difficult to quantify and lie outside the scope of this investigation. The criteria used to select which parameters should be monitored was whether it was thought that they had any influence on methane emission (both to the general air body and methane drainage range) and could easily be monitored. On this basis, only two parameters could be used and these are barometric pressure and rate of production but they were not electronically monitored. Air velocity (hence air quantity) was also monitored as was known to have a definite effect on the methane concentration. It is possible that methane emission is constant while variations in airflow can cause appreciable changes in methane concentration, if only in the short-term. For completeness, it was decided to take account of the methane present in the drainage range and so drainage methane concentration, static pressure and differential pressure were also monitored. All of the monitors described were made by Seiger Limited and Status Scientific Controls Limited.

5.7.1 Instrument Calibration

Accurate calibration of the monitors beyond their normal level of maintenance was not considered to be appropriate. The monitors were checked to see that they were functioning correctly but were not removed beforehand for calibration. It was thought that the most representative data would be from instruments that were maintained as usual and not specially prepared or maintained. In fact, it was found that the monitors and their sensing heads were maintained extremely well as part of Colliery and British Coal policy. A routine of changing the sensing head of the Status methane monitor (general air body methane concentration) every two weeks was followed, with the head being checked at the mine and then sent to British Coal's Headquarters Technical Department at Bretby for further calibration checks. The sensing heads of the other instruments were changed and checked less frequently as past experience showed that they remained in calibration longer. Periodically the vortex-shedding head of the air velocity monitor was cleaned with a small brush to remove dust and dirt.

5.7.2 Methane Concentration

General air body methane concentration was monitored by a Status type CH4/03 methane system. The CH4/03 system is comprised of a regulator/display unit and a methane sensor head which can be placed up to 3 m distant from the regulator/display section. The regulator/display unit contains the circuitry to perform power supply regulation, and transmit its signal to an outstation. The CH4/03 monitors continuously and a relay circuit within it has the ability to shut-down electrical power in a district when methane levels reach a pre-set alarm point. The instrument range is 0 to 3% methane at an accuracy of ±0.1% methane by volume, or ±8% of the true value, whichever is the larger. The CH4/03 monitor utilizes the standard British Coal 0.4 - 2 V analogue output (to BS5754) for data transmission purposes. It also features an LCD display which can indicate the methane concentration, 'power', 'fault' and 'over-range' conditions. The CH4/03 can be powered by a number of external power sources. These features are common to the range environmental monitors described in this thesis. The arrangement of the whole system and its positioning in an underground roadway can be seen in Plate 5.1.

The sensing head of the CH4/03 contains a pair of pellistors housed behind a pair of sintered metal filters between which is sandwiched an active carbon cloth filter. A filter is used as protection against substances which may poison the pellistors. Gas is allowed to reach the pellistor by natural diffusion. Electrical current flows through the catalytic elements (pellistors) which are arranged into a Wheatstone bridge. The methane present in the sensing chamber is oxidised and the heat from this reaction increases the temperature of the catalytic elements resulting in a change of their electrical resistance. One of the pellistors is an active catalytic detector whilst the other is non-active and provides temperature compensation. The methane concentration is proportional to the change in resistance and in the case of the CH4/03 is linear over the 0 - 3% measurement range. This type of sensing head is only suitable for methane concentrations of less than 5% in air. At concentrations in excess of 5% oxygen content can be decreased which may result in an incorrect reading. If the oxygen content is below 12%, perhaps due to very high levels of methane or another gas such as carbon dioxide, a false low reading will be obtained. Response time of the sensor head is approximately 10 seconds to reach 90% of a final reading. A dummy head is available to check the regulator unit output voltages, the calibrated offset zero and a zero to greater than full-range variable signal-voltage. Problems were experienced with early versions of methane sensing heads where it was found that sensor reliability deteriorated if they were subjected to methane

Plate 5.1

Photograph of Environmental Monitors

Left: Air Velocity

Centre: Carbon Monoxide

Right: Methane Concentration

concentrations >0.8% and high relative humidities for long periods of time [62]. This situation was resolved by the introduction of more robust pellistors.

Methane concentration values are very often characterized by rapid fluctuations in level. This can be a problem due to the inertia of the sensing head which can have a profound effect on the accuracy of readings if the time interval between observations is small and the methane level is fluctuating rapidly. Research work [63] has identified factors that influence a sensing heads response time. These can be classified as external influences such as the ventilation ram effect whereby the response time is reduced due to ventilation pressure and correct siting of the sensing head, and internal influences such as the construction of the sensing head. Careful consideration to a number of factors can reduce the response time markedly but for the purpose of obtaining methane concentration data no other instrument other than the CH4/03 was readily available. However, deliberately choosing an instrument that was not typical of those in general use was inappropriate. Thus, the response time of the Status CH4/03 was ignored and the values of methane concentration were regarded as correct at the time the observation took place. Ideally, a sensing head with negligible response time would be the best to use.

5.7.3 Air Velocity

The air velocity was measured by a Status type AV/02/030 velocity monitor. Similar to the CH4/03 it is comprised of a regulator/display unit and a detatchable sensing head. The monitor can measure velocity in a number of ranges from 0.5 to 30 m/s at an accuracy of ±10%. The sensing head operates on the vortex shedding principle and requires no moving parts. The vortex shedding principle is where air flowing past an obstruction results in the creation of vortices in a region downstream of the obstruction. The vortex frequency is proportional to air speed and by measuring this frequency it is possible to produce an electrical output signal that is proportional to air velocity. It is advantageous to use this type of sensing head rather than an anemometer type sensing head where damage to the vanes by dust or tampering can alter the velocity value. The sensor has been specially designed to cope with out of alignment variations and yaw angles of up to ±15° from normal do not adversely affect the velocity measurement. The control unit has a switched powered output for an externally intrinsically safe flameproof relay so that power can be cut-off if the measured air velocity falls below a pre-set level.

The sensing head of the monitor was placed in the middle to top right-hand corner of the roadway, as illustrated by Plate 5.1. The value of velocity measured in this part of the roadway is only a point value and would not be the truest possible. In fact the sensor is in the worst position for accurate measurement of velocity. In the area close to the rib or roof large changes in the true airflow may not be apparent in the values measured by the sensing head placed in a low-flow boundary layer. Ideally it should be placed close to the middle of the roadway but usually this is not possible. A method to overcome this difficulty is to assume that a constant relationship exists between the velocity measured by fixed sensing head and the average velocity at that airway cross-section [64]. If this is assumed then the velocity measured by the sensing head can be related to the true average velocity at the specific cross-section by a constant. This constant changes with velocity level. Fortunately, because time series analysis seeks only to relate the rate of change within a single parameter or between parameters matters can be simplified. It is not necessary to convert the velocity values to flowrates or apply constants. However, this only holds true if the values recorded by the velocity monitor represent the true changes in the average velocity of the airway. For 119's it was thought that placing the sensing head 1.15 m away from the ribside and 0.5 m from the roof was sufficient to safeguard this assumption.

5.7.4 Drainage Methane Concentration

The drainage methane concentration was measured by a Seiger BM2H high concentration reading methane unit. For drainage range differential pressures that are greater than 250 Pa a gas flow regulator needs to be used but at Thoresby 119's the drainage differential pressure was below 250 Pa and no regulation was needed. The BM2H reads from 0 - 100% methane to an accuracy of ±3% of the full range reading.

The catalytic method used to determine the methane concentration by the BM3 cannot be used by the BM2H. Such a method can only measure to approximately 5% methane and beyond this a non-catalytic method needs to be used. The BM2H uses a method based on the principle of thermal conductivity. Two thermal conductivity sensors are incorporated into the sensing head and arranged as a Wheatstone bridge. One is kept as a reference and is sealed in air while the other, the active sensor, is exposed to the air/methane mixture. Current is passed through both sensors and if methane is present the active sensor cools down. The cooling effect depends on the concentration of methane and

causes the current flowing through the active sensor to drop resulting in an imbalance in the bridge from which the methane concentration can be determined.

5.7.5 Drainage Static and Differential Pressure

The methane drainage range static and differential pressures were monitored by a Seiger BP1 pressure monitoring system. The BP1 is designed to monitor the differential pressure across ventilation doors, booster fans, stoppings and across an orifice plate placed in a pipe. It is also capable of monitoring static pressure if an appropriate static pressure transducer is used. By using different transducers a wide range of pressures can be monitored. Differential pressure can be monitored from 0 to 10 kPa and static pressure from 0 to 100 kPa. The accuracy of the pressure transducers are ±3% of the full range reading. The pressure transducers are of the diaphragm capacitance type and pressure connections are made by using rubber hoses and clips.

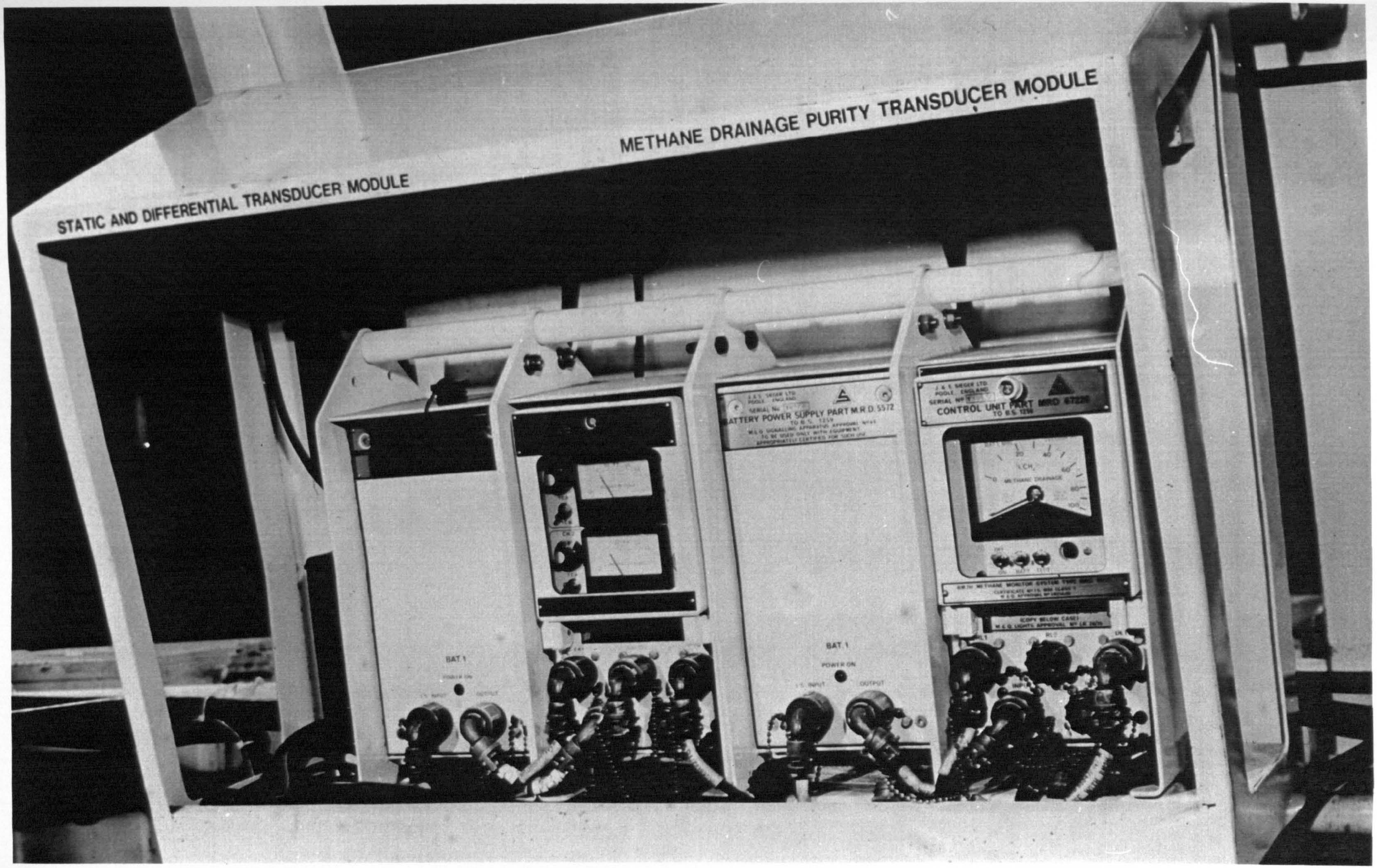
The drainage static pressure provides an indication of the suction available at the borehole. Exhausters are used to maintain a suction of 0.5 to 0.98 kPa to overcome the resistance of the pipeline to gas flow and improve production. The suction pressure created by these exhausters is not thought to be carried to the end of borehole because of pressure losses. Therefore, at the end of the borehole the borehole pressure is assumed to be slightly smaller than atmospheric [47]. The drainage differential pressure which is also known as the velocity pressure was measured by either side of an orifice plate and by using a simple formula the total flow of mixture from each borehole can be found. A typical arranagement of the methane drainage range monitors can be seen in Plate 5.2

5.7.6 Barometric Pressure

Unfortunately an electronic instrument for measuring absolute barometric pressure was not available. Instead, an ordinary, but well maintained barograph, was used. This instrument was sited in the colliery control room. The barometric pressure charts show a continuous weekly record from midday Monday to the following midday Monday with a small period of overlap. The barograph is scaled to measure barometric pressure from 950 to 1050 mb. The pressure values were entered into the computer by hand after

Plate 5.2

Photograph of Methane Drainage Range Monitors


METHANE DRAINAGE PURITY TRANSDUCER MODULE
STATIC AND DIFFERENTIAL TRANSDUCER MODULE
BATTERY POWER SUPPLY PART M.R.D. 5572
CONTROL UNIT PART MRD 67229
BAT. 1
POWER ON
BAT. 1

carefully reading the value from the chart. A prerequisite of time series analysis is that if parameters are to be cross-correlated they must all have the same time interval between observations. Since the chosen time interval for the electronically monitored parameters was 2 minutes this meant that 720 values per day had to be read from the chart. At times this proved difficult and resulted in a stepped appearance when the pressure was changing rapidly.

Barometric pressure was monitored with the intention of determining whether it could be used to account for changes in methane emission. To this effect it was necessary to only measure it on the surface. If a barograph had been installed underground the pressure variations due to ventilation system changes (static, velocity and NVP) would be superimposed onto the barograph chart.

5.7.7 Rate of Production

Coal production is perhaps the most important parameter that affects methane emission. If there was no production there would be no changes in strata stress and permeability, hence no gas flow would be possible. Information as to the quantity of coal production was taken from the colliery control room sheets. These are known as the control centre record sheet. The control centre record contains information on all of the mines producing coal faces, development progress, belt and skip delays and other items. They are completed for each shift by control room operators. An example of a control centre record sheet is given in Appendix 1.

The start and finish time for each face cut or strip is entered onto the sheet by the control room operator when he has been informed by a member of the face team. There was no method of knowing in the control room if the coal shearer was cutting unless instruction had been received from a member of the face team. Consequently, an investigation had to be carried out to determine the accuracy of the strip stop and start time.

The strip cutting behaviour of a complete afternoon shift was studied to find out what normally occurred. Production data is characterized by definite periods of coal cutting (during each shift) and periods when coal is not cut (rest time, maintenance and swap-

over time between shifts). Occasionally the cutting operation stopped and the control room was notified by telephone and the stop time noted on the control sheet. When the machine was ready to cut again this time was also telephoned to the control room and noted. For a complete and uninterrupted strip it was possible to calculate the cutting rate in metres per minute. It was found that this was not constant for each strip because the machine operator could control the speed of the machine to a small degree and stop the machine for a brief period of time for any reason. Generally, the machine operator slowed the machine as it approached the end of the strip. A typical complete strip took approximately 40 minutes. During a shift there were short periods between each strip when an operation termed the 'shuffle' was performed. The shuffle lasted an average of 20 minutes and its purpose was to remove coal that was left in the roof and floor at the face end. The decision to inform the control room of a machine halt was at the discretion of the machine operator and if it was felt that the stoppage was only for a matter of seconds or at most a few minutes the control room was often not informed. If the stoppage lasted longer than 5 minutes, however, the control room was informed of the reason and the correct stop time noted. For the purpose of analyses the investigation showed that the production start and stop times both for a complete or interrupted strip were accurate to within 5 minutes.

Production data was taken from the control room sheets and entered into a spreadsheet by hand. The base data format was whether the machine was cutting or not and if so at what average speed was maintained over the cutting period at intervals of one data point every 2 minutes.

5.8 Factors Influencing the Siting of Monitors

It is to be expected that the positioning of the sensing heads of monitors is an important criteria in achieving representative data. At Thoresby, the environmental siting policy is that they should be sited at the closest convenient point towards the end of the face return airway before it is split or joins another return for both advance and retreat faces. On 119's district, the instrument sensing heads (for the BM3 and BA4) were only placed a few metres from the environmental monitoring station.

Recommendations by the US Bureau of Mines advise that the methane sensing head be

placed at least 30cm from the roof and well away from the ribside [65]. This is because methane is lighter than air and the highest concentrations are likely to be found near the roof. When measuring air velocity the best place is the geometric centre of the airway cross-section. The sensing head should also be placed in an area free of obstructions both upstream and downstream.

5.9 Errors Due to Data Transmission

Errors in data transmission can be caused by interference from electromagnetic noise. This can be a problem in mining but can be minimized by careful cable shielding and the use of a system whereby once MINOS receives a signal it asks the transmitting device to verify the signal. Most British Coal data transmission systems subject incoming signals to a number of tests to decide whether it is valid data. However, it was found that most errors were caused not by the data transmission system but by the colliery information system computer. It may have been that actual values for a monitored parameter were erroneous, i.e. if methane concentration of 1.01% was observed but the CIS reported it as 1.11% and this sort of error cannot be detected without signal verification. This type of error occurs because the data contains transmission noise that can be miss-read as a different value on de-multiplexing. Tests carried out by MRDE at Bedlay Colliery to investigate data inaccuracies of the mines environmental management system identified two main sources of error [62]. The first was due to the fluctuating behaviour of some monitored parameters such as methane concentration and air velocity where additional signal noise could be enough to produce erroneous values. The second was caused by electrical noise and this was identified as the most significant problem. This problem was resolved by better earthing and cable shielding. Ultimately, for the purpose of data analysis, errors associated with data security during transmission are ignored unless it is obvious that a reading is vastly different to its preceding and succeeding values.

At Thoresby, the monitors send their information or signal to an outstation which waits for instructions from the control room computer to send data. The actual data transmission is by time division multiplexing. The outstation receives a signal from a monitor continuously but does not send a mean value, for the time interval that has passed since the last transmission. This means that the transmitted data are actually spot values. This is no real problem so long as the level of the monitored parameter has been fairly constant over the time interval between the data transmission. If a parameter is known to

be fluctuating then it would be better to use instruments that transmitted a mean value to smooth the fluctuations. Of the parameters monitored, it was found that methane concentration and air velocity often showed marked fluctuations in value between the monitored time interval of 2 minutes and there was no way of knowing how much variation in value occurred during the monitoring time interval. This was not ideal and could not be easily resolved and any problems that this may have caused were ignored.

5.10 Data Storage - Minos to CIS

The control room monitoring system includes facilities for comprehensive displays of environmental data [66]. Graphical displays of data showing trends and alarms are all possible, however, the scope to utilize the data beyond what is provided for in the control room system software is very limited.

The colliery information system or CIS (and also known as the mine information system - MIS) in use at Thoresby dates back to 1975. The environmental data is stored by the CIS on hard disks.

5.11 Data Transfer - CIS to PC

After the conception of the project, data was initially obtained for methane concentration only. This data was not in a digital format and was in the form of a printout. Thoresby Colliery generally prints the values for methane concentration daily with a sampling interval of 5 minutes. This data was kept for information purposes. It was necessary to gain approval to increase the sampling time to a frequency of every two minutes and the printout data was a carbon copy of the one kept for colliery information. This information was entered into a spreadsheet database by hand which soon proved to be a lengthy and unsatisfactory arrangement. Each day generated 720 data points. A method of more efficient data collection became a necessity and a number of methods were subsequently investigated to achieve this.

5.11.1 The Megalog MGL-21

The first idea was to consider using portable data logging equipment that was available within the Mining Department. The Megalog MGL-21 (manufactured by Technolog Ltd.) is an intrinsically safe device purposefully designed for underground use in coal mines. It is a solid state data recorder capable of receiving information from 15 analogue inputs and down-loading this data to an IBM compatible computer. As it records the unit has a facility to view a real-time graphical display of a selected channel. This facility is especially useful in determining, after the unit has been programmed for use, if it is functioning correctly.

The Megalog has a number of time-base options to select from. The fastest sampling rate possible is 2000 samples/sec, or a time-base of 0.5 msec, but only one channel can be monitored at this rate. For the expected purpose of the data logger this time-base was far too small and a time-base of one observation every two minutes was decided upon. At this recording rate the data logger was able to monitor 15 channels, which comfortably exceeded the number of channels to be monitored.

A disadvantage of the data logger is the need to specify an initial real start time so that the observation is actually recorded at the correct moment in time. For example, if the recording is to start at the arbitrary start time of 14:00:00 hours, the unit is required to be programmed with this start time and started when this time is reached. The difficulty lies with knowing the right point in time to start the data logger. However, this was not thought to be a major problem since a start timing accuracy of one or two seconds was not deemed critical and could be easily achieved by using an accurately set clockwork watch, or by using an underground telephone to consult the speaking clock if one was within reach. Connection to the environmental monitors was by an 18-way connector from a common supply, i.e. a junction box. A typical arrangement for underground use is shown in Figure 5.5.

The main problem anticipated by the use of the data logger was its ability to monitor the parameters continuously. This was absolutely necessary if time series modelling techniques were to be employed. There was no doubt as to the data logger's reliability and ruggedness for operating underground, rather its suitability was questioned according

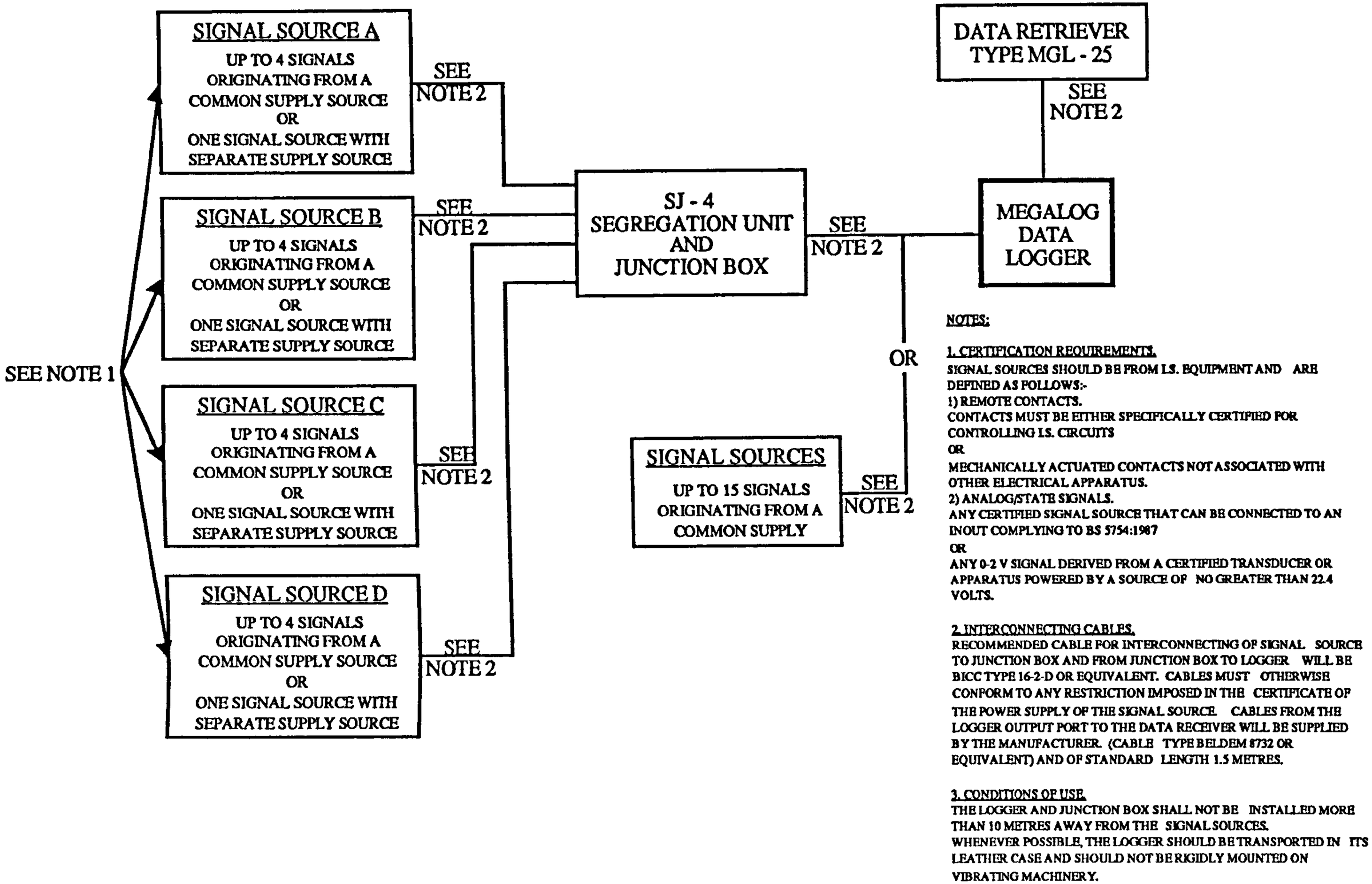


Figure 5.5 Underground Configuration Arrangement of the MGL-21.

to its battery and memory capacity. The recorder is powered by a maximum of two PP3 alkaline batteries. For continuous recording over a long period of time power consumption could be lowered by switching off the display. Recording data from 5 channels at a sampling rate of 1 observation every 2 minutes resulted in an estimated battery life of 26 days. The MGL-21 is supplied with approximately 64kB of memory. and at the same sampling rate the memory capacity of the data logger would have been consumed after approximately 14 days. For these reasons the data logger was rejected as being unsuitable.

5.11.2 Transfer from CIS to PC

The next idea was to consider how to make the maximum use of the existing colliery computer system facilities as possible. This was deemed to be a sensible approach since the data was already in existence and only a way to utilize it was to be found. Although simple in concept, actually getting the data into a usable form required much thought and perseverance. A computer communication package called 'Procomm Plus' was obtained to facilitate data transfer from the CIS to a PC [67]. This programme, after some early teething difficulties soon proved easy to use. By accessing the relevant data channel, data was streamed across to a PC. In fact, the communications programme allowed the PC to operate exactly as a CIS terminal. On running the programme it is necessary to select a CIS option which displays the transducer data in a report format (see Figure 5.6) on screen. After selecting the required channel the whole display is transferred and saved to the hard disk of the PC under a suitable filename. At first the data was not transferred continually and a small computer programme needed to be written to ensure continuous data streaming. The CIS data was displayed and transferred page by page with the user pressing the return key to transfer each page. This was time consuming and quickly became a very tedious task. It was possible to programme Procomm Plus to cause automatic scrolling of the data for transfer. The programme searches the incoming data for a key phrase, such as 'next page', and inserts a carriage return character into the output stream every time this phrase was found.

A problem was experienced with the storage of the data files. The selected 2 minute logging interval resulted in the generation of 5040 data points per week per channel. For the five monitored channels this meant a total of 25200 data points per week. The data transfer format was the same as the display format i.e. headers and footers at the top and

bottom of each page with the subsequent date/time/channel/analogue value/alarm and fail status in between which resulted in the transfer of many more characters than was necessary. This meant that a colossal amount of data needed to be stored onto numerous IBM formatted high density 3.5" disks, each capable of holding 1.4 Megabytes of data. It was soon apparent that such large amounts of data would take up too much disk space and this was the next problem to be overcome.

MON 19-AUG-91 TIME UNTIL VALUE
20'S/119'S 119'S MET

16:48		1.29 %	ALARM
16:50		1.19 %	ALARM
16:40	16:51	1.31 %	ALARM
16:51		1.22 %	
16:52		1.19 %	
16:54		1.23 %	
16:55		1.25 %	ALARM
16:56		1.27 %	ALARM
16:56		1.27 %	ALARM
16:58		1.24 %	ALARM
17:00		1.29 %	ALARM
17:02		1.22 %	ALARM
16:55	17:03	1.30 %	ALARM
17:03		1.18 %	
17:04		1.19 %	

* PAGE = 33

Figure 5.6 Transducer Report Format - for Methane Concentration Data.

Since the transferred data was in the format used by the CIS for report purposes and could not be altered either on the CIS itself or during transferring, it was necessary to reduce the size of the data files after transfer. Unfortunately, it was not possible to instruct the CIS to transfer only the selected channel time and value (the only two items of interest) and discard the headers and footers from each screen display page. After transference the data files were not in a suitable format for time series analysis and it proved necessary to write a short computer programme to remove the unwanted information. The data files were firstly copied from the floppy discs to a hard disk and

checked for missing or corrupted data. If data was discovered to be missing an attempt was made to see whether it could be recovered and at times such attempts did prove successful.

5.12 Missing Data

Missing data was usually found to be caused by faults within the computer system and not because of instrument failure. The instruments themselves proved to be extremely reliable and during the course of the study did not appear to malfunction at all. Missing data was noticed once the data had been transferred to PC and brought back to the University. Each file was checked to see if any data was missing, rather than data that had been lost due to instrument or transmission failure. It quickly became evident that if data was missing from one channel e.g. methane concentration it was also missing from the other channels as well. It was not possible to check that the transfer of data from the colliery CIS to an IBM PC was completely successful or that there was a complete record of data to transfer initially. The computer disks were collected regularly from the colliery and checked immediately for missing data in the hope that it could be possible to retrieve it. At times this was possible once the control room computer engineer knew the dates and times of missing data and if it was possible to access the archive disks for the information. It was Colliery practice to hold up to four weeks of back-up data on a recording archive disk [68]. When an archive disk was approaching capacity the procedure was to copy the back-up data to a new disk before the new disk was installed. The new disk was then prepared ready for use when the other disk became full. It was important to ensure that the archive disk was changed before it became completely full as data that was saved at the limit of disk capacity was often found to be corrupted.

5.13 Data Conversion

After the data had been reduced to four columns i.e. time, value, alarm and fail status it was transferred to a Macintosh SE 30 computer. The conversion from IBM MS-DOS to Macintosh format enabled a further reduction in data file size and after transfer, data files (containing two days worth of information) that were originally 60 kB in size were reduced to around 5 kB. Subsequent use of file compression software then reduced file size even further, resolving all file storage problems completely. A procedure was

developed to process the files as easily as possible since they had to be formatted by hand to enable it to be used for time series analysis. Each data file contained information recorded over a period of two days. Ideally, each file should have contained 1440 lines of data (after all page headers and footers had been removed) which was easily checked by instructing the computer to count the number of lines in the text document. Generally, it was rare for a file to contain the exact number of lines because of missing data and peculiarities caused by alarm and fail warnings. It was also found that the behaviour of alarm levels was different in each monitored parameter. For example, in the case of methane concentration data if an alarm level was reached the time at which this occurred was displayed. The alarm indication continued to be flagged until the methane concentration dropped below the alarm level at which time the start and end time of the alarm period was displayed. Any file that contained more than 1440 lines of data had to be manually checked and stripped of unnecessary information.

5.14 Data Transfer from IBM to VME Mainframe

After the data files had been manipulated into the correct time series format, it was then necessary to covert them back to MS-DOS files so that they could be transferred to the University's mainframe computer. Use was made of the extensive communications software available at the University to do this. Once the time series files had been stored on the VME system the next task was to subject them to time series analysis.

5.15 Conclusion

This chapter has described the variables that were monitored at Thoresby 119's and will subsequently be used to build univariate models. This chapter has also brought to attention the difficulty a researcher has in acquiring data that is being monitored at a colliery for research purposes. It is assumed that since the data had been monitored and stored it would be easily available for any other use, whereas in the course of the research this was not found to be true. Aspects of data monitoring with respect to the acquisition of accurate and representative data were also detailed. Finally the steps necessary to transform the data into a useable computer format were also described.

CHAPTER SIX

UNIVARIATE MODELS OF MINE ENVIRONMENTAL AND PRODUCTION DATA

6.1 Introduction

In this chapter the Box-Jenkins model building methodology outlined in chapter 4 is utilized to build univariate ARIMA models for environmental and production data. Four different models are built for each variable from data recorded over the month of April 1991. In addition to the original recorded time interval of one observation every 2-minutes the data was transformed to provide an average value over 10-minutes and an hourly average value. Lastly, models are presented that are built from only one day of monitoring at the original time interval. The main reasons for building the univariate models are to generate univariate forecasts of methane concentration and to use the univariate models for all of the variables to explore the possibility of building multivariate models for methane concentration that can also be used to forecast methane concentration.

The Box-Jenkins model building procedure is only part of a framework that is used to build a model in a systematic manner. Although this chapter is concerned with the building of univariate models, it is best to first consider them within an appropriate conceptual frame that presents a conceptual priori model of how methane concentration is thought to be explained by other variables. At this stage the objective is to determine if it is possible to build time series models of environmental and production data from differing time intervals. A complete example of the three stages of model building is given for only the first example for methane concentration. For the other models built, only items of interest or peculiarity are noted.

6.2 A Conceptual Model

It is intended to investigate whether it is possible to forecast methane emission from a consideration of possible explanatory variables, such as production and barometric pressure. Before this can be done univariate models for all of the variables need to be

built. A univariate model is built from the past history of a single variable and so forecasts from it are based on knowledge of its past history only. In a real-life situation such a model may be suspect where its forecasting ability is concerned. In some circumstances it may only be possible to generate forecasts from the past history of a single variable, either because no other variables relating to it can be found or the physical system that generated the series may be too complicated to model adequately.

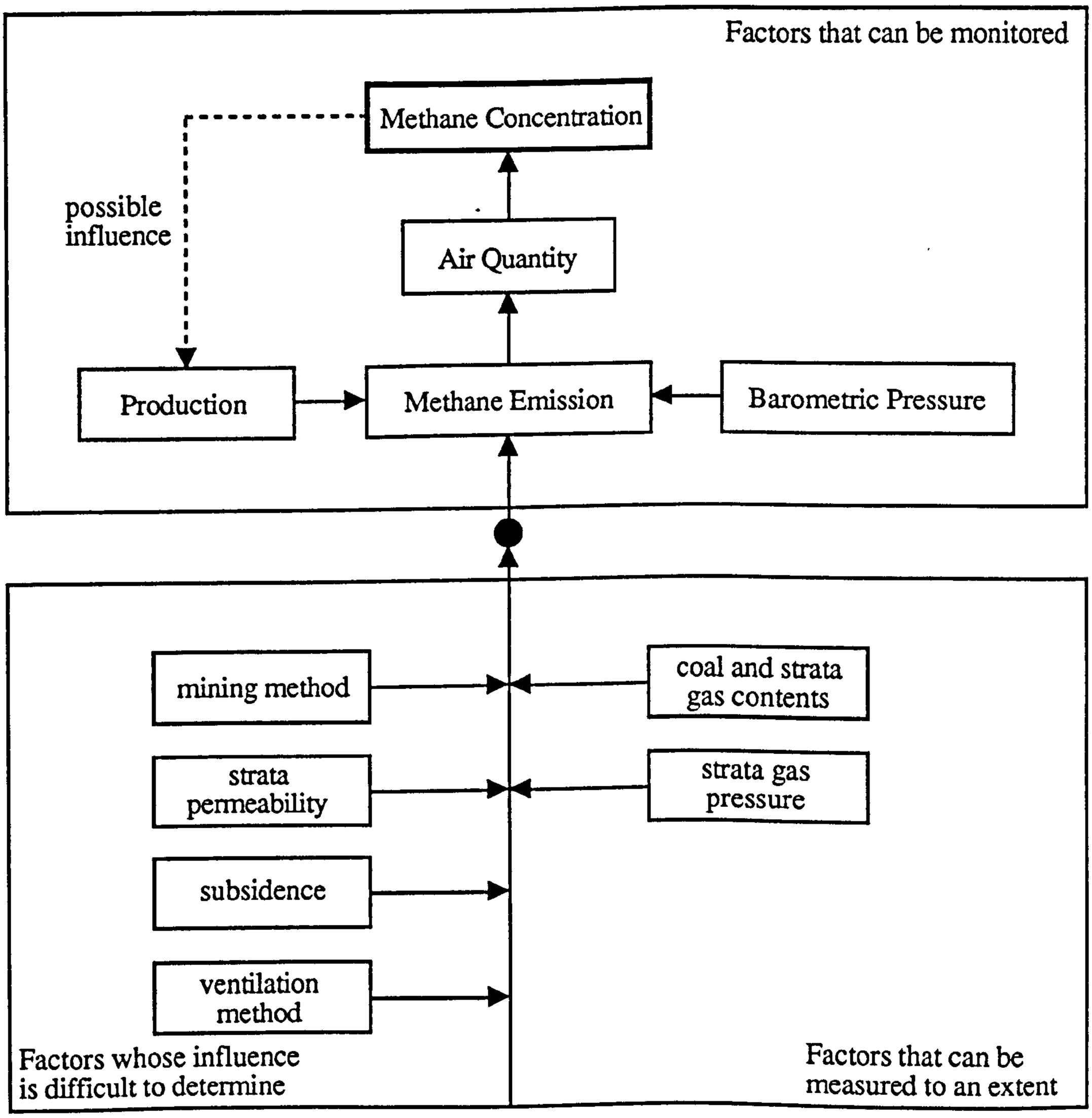


Figure 6.1 A Conceptual Model for Methane Emission.

Figure 6.1 presents a simple conceptual model for methane emission. In this model methane concentration is presented as the dependent variable and the others as independent variables. This means that there is a one-way causal relationship between methane concentration and the other variables. For example, barometric pressure may influence methane emission (hence concentration) but methane concentration can in no way cause changes in barometric pressure. A possible exception to the assumption of a one-way causal relationship between methane concentration and its explanatory variables occurs when methane concentration exceeds 1.25%, the statutory value at which electrical power must be switched off. At this event in time it would be expected that production is stopped but at Thoresby 119's this was not usually necessary. The methane concentration was monitored at the return end and although it often exceeded 1.25% the protection afforded by the Sherwood curtain ensured that the face end methane concentration was lower than 1.25% and production could continue. Other factors, mentioned in chapter 2 that can influence methane emission, although contained in the conceptual model are inappropriate for time series analysis. For example, strata gas contents can be determined but cannot be represented by a time series whilst strata gas pressures and permeabilities are extremely difficult to determine.

6.3 Models For Methane Concentration

6.3.1 Original Series of Methane Concentration

The first model to be built was that for methane concentration recorded in the return gate of Thoresby 119's during April 1991. At a sampling interval of one observation every two minutes this generated a total series length of 21,600 observations. Each day consisted of 720 observations and the data was stored in daily intervals before being assimilated and transferred to the mainframe. It was necessary to examine each days data to see whether any spurious values were present and check for missing data. No data was thought to be incorrect and missing data was a rare event. Data that was missing was replaced by most likely values by using an appropriate SPSS-X Trends™ subroutine [57]. It proved impossible to produce a plot of the whole series both for computing and practical reasons but plotting a random selection of any manageable portion demonstrated the continually changing level of methane concentration. An inspection of Figure 6.2 reveals two types of trend. The first is a very definite long-term (over a few hours), upwards and downwards behaviour and the second is a seemingly rapid fluctuation

between observations over a short period of time i.e. over a few minutes. It is very difficult to explain definitively the reasons for such behaviour but the long-term trend may be accounted for by changes in the rate of production. The short-term changes may result from variations in airflow, perhaps themselves caused by the opening and closing of an air-door(s) somewhere in the ventilation circuit.

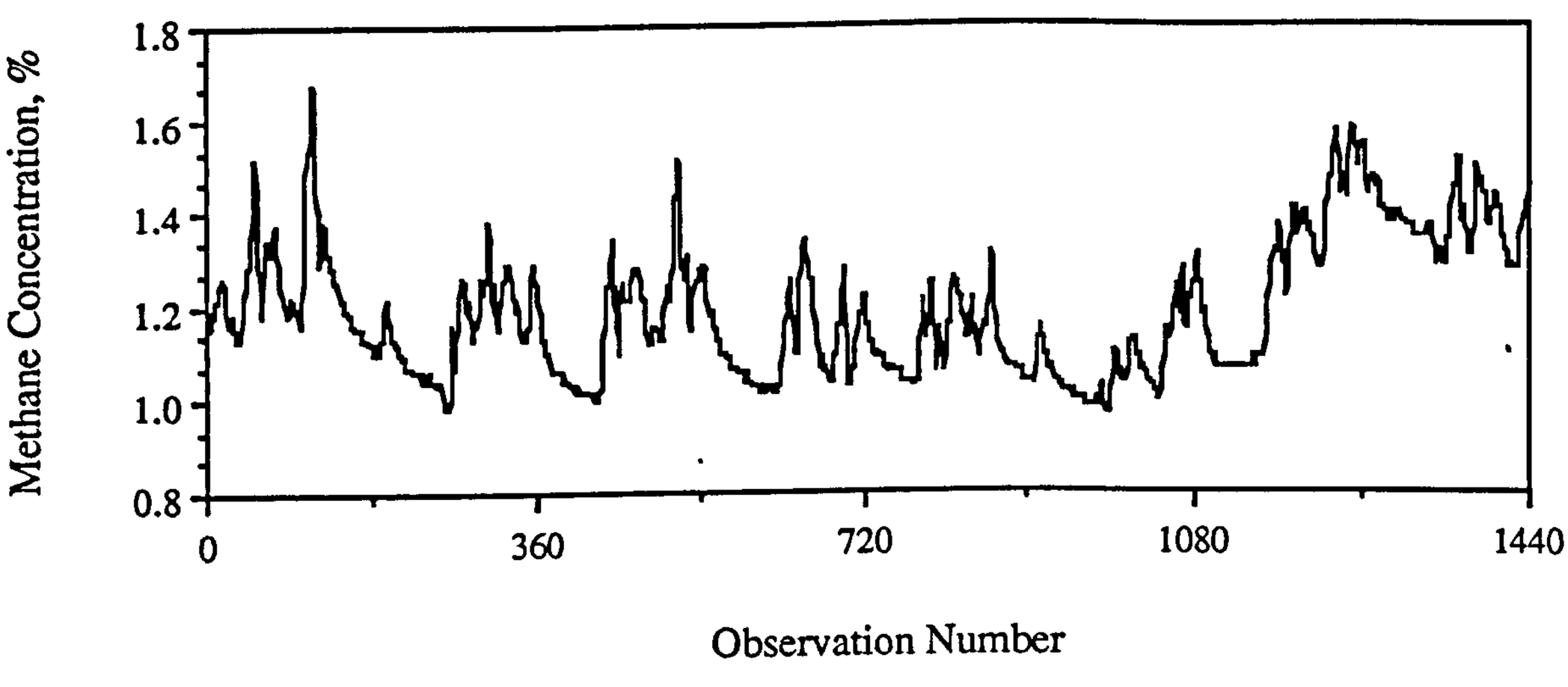


Figure 6.2 Plot of Methane Concentration Data.

From an inspection of this sample data it may be anticipated that many other portions of the data will exhibit similar trend behaviour. Such displays of trend mean that the series is non-stationary and the first step in the identification process of a model that fits the series is to make it stationary. For the model building process the series was split into a historical portion over which the parameter values were estimated. This portion was 16,000 observations in length. The rest of the series was kept as a validation portion so that the parameter fit of the most suitable model could be tested. The plot of the autocorrelations of the raw series, shown in Figure 6.3, die out very slowly and alongwith the value for the autocorrelation at lag 1 of 0.993 being very close to unity it is evident that the series is non-stationary. Taking a first difference was sufficient to obtain a stationary series and the correlations of the differenced series are shown in Figure 6.4.

Autocorrelations: METHAPRIL

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	.993	.008	.**********	15789.038	.000
2	.984	.008	.**********	31272.463	.000
3	.973	.008	.**********	46420.284	.000
4	.962	.008	.**********	61241.224	.000
5	.952	.008	.**********	75748.045	.000
6	.942	.008	.**********	89946.195	.000
7	.931	.008	.**********	103837.018	.000
8	.921	.008	.**********	117414.665	.000
9	.910	.008	.**********	130684.673	.000
10	.900	.008	.**********	143664.981	.000
11	.891	.008	.**********	156375.919	.000
12	.882	.008	.**********	168841.617	.000
13	.874	.008	.**********	181085.681	.000
14	.867	.008	.**********	193131.318	.000
15	.861	.008	.**********	205001.856	.000
16	.855	.008	.**********	216721.911	.000
17	.851	.008	.**********	228320.503	.000
18	.847	.008	.**********	239822.098	.000
19	.845	.008	.**********	251252.460	.000
20	.843	.008	.**********	262634.698	.000

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 16000 Computable first lags: 15999

Partial Autocorrelations: METHAPRIL

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
1	.993	.008	.**********
2	-.226	.008	*****.
3	-.036	.008	*.
4	.033	.008	.*
5	.010	.008	*
6	-.012	.008	*
7	-.012	.008	*
8	-.026	.008	*.
9	.011	.008	*
10	.032	.008	.*
11	.023	.008	*
12	.033	.008	.*
13	.032	.008	.*
14	.034	.008	.*
15	.037	.008	.*
16	.041	.008	.*
17	.052	.008	.*
18	.036	.008	.*
19	.047	.008	.*
20	.042	.008	.*

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 16000 Computable first lags: 15999

Figure 6.3 Correlations of Raw Methane Concentration Series.

Autocorrelations: METHAPRIL
Transformations: difference (1)

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	.224	.008	.****	800.944	.000
2	.079	.008	.**	901.636	.000
3	-.012	.008	*	904.069	.000
4	-.024	.008	*	913.121	.000
5	-.005	.008	*	913.559	.000
6	.006	.008	*	914.102	.000
7	.023	.008	*	922.853	.000
8	-.006	.008	*	923.434	.000
9	-.038	.008	*.	946.006	.000
10	-.045	.008	*.	978.556	.000
11	-.057	.008	*.	1030.542	.000
12	-.057	.008	*.	1083.188	.000
13	-.057	.008	*.	1135.838	.000
14	-.059	.008	*.	1191.765	.000
15	-.064	.008	*.	1257.208	.000
16	-.078	.008	**.	1355.073	.000
17	-.066	.008	*.	1425.603	.000
18	-.071	.008	*.	1506.867	.000
19	-.065	.008	*.	1574.468	.000
20	-.056	.008	*.	1624.607	.000

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 16000 Computable first lags after differencing: 15998

Partial Autocorrelations: METHAPRIL
Transformations: difference (1)

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
1	.224	.008	.****
2	.031	.008	.*
3	-.038	.008	*.
4	-.016	.008	*
5	.007	.008	*
6	.008	.008	*
7	.020	.008	*
8	-.018	.008	*
9	-.037	.008	*.
10	-.028	.008	*.
11	-.038	.008	*.
12	-.037	.008	*.
13	-.038	.008	*.
14	-.041	.008	*.
15	-.044	.008	*.
16	-.057	.008	*.
17	-.039	.008	*.
18	-.051	.008	*.
19	-.046	.008	*.
20	-.040	.008	*.

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 16000 Computable first lags after differencing: 15998

Figure 6.4 Correlations of Differenced Methane Concentration Series.

Although the series is now stationary or rather in statistical equilibrium, it will still be autocorrelated and additional features now need to be introduced to remove the autocorrelation and eventually result in a series of random residuals. This is the next step in the model building process and the objective is to examine the correlations and identify the number of p and q (autoregressive and moving average) parameters to estimate a tentative model as described in chapter 3. The autocorrelation values all appear to be significant, that is they are all in excess of the 2sd (2 standard deviations) mark. This is unique and is a function of the large number of observations from which the correlations were estimated. This complicates the identification of the appropriate number of model parameters since the correlation values are significant and likely to remain so at large lags. However, because this is caused by the large number of observations, significant correlations at large lags can be ignored and model identification proceeds as normal.

There are two large spikes at lags 1 and 2 in the autocorrelations and one large spike at lag 1 in the partial autocorrelation. It is most likely that the series is best fitted by a combination of autoregressive and moving average parameters and so is a mixed process. For this case, however, it is convenient to forgo the identification of the model form and proceed in a logical manner by specifying different numbers of parameters and seeing which models fits best. The first model to try is therefore an AR(1) or ARIMA(1,1,0). Since this is the first tentative model identification the next stage is to estimate a starting value for the AR(1) parameter.

SPSS-X Trends™ does not require to be supplied with initial parameter estimates. Instead it uses the value of the autocorrelation at lag 1 which is equal to 0.224. The estimation process proceeds iteratively until the optimum parameter value is reached and SPSS-X Trends™ offers a variety of estimation criteria. Although initial parameter values can be supplied it is better to use the value suggested by SPSS-X Trends™, otherwise the number of iterations may be large and exceed the default estimation criteria.

At this stage the series was split to provide a historical portion over which the parameter values were estimated and a validation period over which the fit of the parameter was tested. After estimation the final value of the AR(1) parameter was 0.2237 with a t-statistic of 29.03 which is highly significant. Examination of the autocorrelation residuals, shown in Figure 6.5, reveals that only the value at lag 1 is insignificant or is white noise, confirmed by its low Box-Ljung value and its associated high probability.

Autocorrelations: ERR_1 Error for METHAPRIL from ARIMA (1,1,0)

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	-.007	.008	*	.757	.685
2	.038	.008	.*	23.742	.000
3	-.027	.008	*.	35.159	.000
4	-.022	.008	*	43.013	.000
5	-.002	.008	*	43.052	.000
6	.002	.008	*	43.127	.000
7	.026	.008	.*	53.859	.000
8	-.003	.008	*	54.036	.000
9	-.029	.008	*.	67.937	.000
10	-.028	.008	*.	80.115	.000
11	-.039	.008	*.	104.298	.000
12	-.036	.008	*.	125.583	.000
13	-.036	.008	*.	146.292	.000
14	-.037	.008	*.	167.938	.000
15	-.038	.008	*.	191.473	.000
16	-.056	.008	*.	241.207	.000
17	-.038	.008	*.	264.541	.000
18	-.048	.008	*.	301.150	.000
19	-.042	.008	*.	329.217	.000
20	-.037	.008	*.	350.654	.000

Plot Symbols: Autocorrelations * Two Standard Error Limits .

Total cases: 16000 Computable first lags: 15998

Figure 6.5 Residual Autocorrelations from ARIMA (1,1,0).

All of the others are significant and this indicates that extra parameters needed to be added to the model. These were added one-by-one and by doing this it was found that this series could be represented acceptedly by a number of models.

Instead of fitting an AR(1) parameter to the series it was re-estimated with an MA(1) parameter. After estimation it was found that the Box-Ljung statistics and the residual error was larger than for the AR(1) fit but at this stage either model could be appropriate. Fitting extra parameters in turn reveals the difficulty in achieving random correlation of the residuals. After each new estimation it is found that the Box-Ljung values of the residuals are still significant at relatively low lags. Table 6.1 details model parameter and residual statistics. Some of the t-values indicate parameter redundancy and the best model could be one of seven (highlighted by bold text).

Lowering the Box-Ljung values would improve the fit of the models but this cannot be done by continually adding extra parameters. Too many parameters, even though they

may be significant, make the model unwieldy and difficult to interpret. In most circumstances the majority of time series can be adequately represented by low-order ARIMA models, that is models that are ARIMA's (3,2,3) or less. The exception occurs when very long series are modelled, as in this example. For very long series it is impossible to greatly reduce the correlation values beyond the lag that is equal to the largest sensible parameter order of the model. However, it will be demonstrated in chapter 8 that the forecasting performance of a model is not too critical of the number of parameters in the model.

Since seven models appear to contain parameters whose t-values are significant it is necessary to decide which is the best one. Table 6.1 shows the values of residual variance for all of the models and their standard error. It is seen that decreases in both the residual variance and standard error occur until model (2,1,2) is reached. Models (2,1,1) and (1,1,2) have higher residual variance and standard errors, so they are rejected. Models (2,1,3) and (3,1,2) have lower residual values than (2,1,2) but only by a very small amount. Model (2,1,3) is rejected because the t-value of the MA(3) parameter is insignificant. The most suitable models for this series is therefore the ARIMA (2,1,2).

Although the best fit model proved to be the ARIMA (2,1,2) it was decided to also choose the very simple ARIMA (1,1,0) as another model that could be used to represent the time series. The reduction in residual standard error after fitting extra parameters to acheive the more complex model was only 0.13% which is very small. Thus, from this simple diagnostic check there would appear to be little difference between them and in the interests of model parsimony the more complex model should be rejected. However, at this stage it is beneficial to consider both of the models as suitable and compare their forecasting performance so that a more informed decision can be made as to which is the most suitable one. This will be done in chapter 8. Further tests were carried out to decide if one of these two models was substantially better. The residual autocorrelations for the (2,1,2) model are very small up to lag 6, with correspondingly insignificant Box-Ljung values. The residual autocorrelations for the (1,1,0) model are larger, again favouring the acceptance of the (2,1,2) as the best. A final test is to examine the residual autocorrelations once the parameter values have been applied to the whole of the series. The parameters were estimated over the historical period and it is desirable to see how well they fit the complete series, i.e., the combination of the historical and validation portions.

Model	Parameter	B	SEB	T-value	Probability	Res. Var	Res. SE
(1,1,0)	AR(1)	0.2237	0.0077	29.0316	0.00000	0.00014882	0.01219901
(0,1,1)	MA(1)	-0.2000	0.0077	-25.8186	0.00000	0.00014964	0.01223262
(1,1,1)	AR(1)	0.3225	0.0330	9.7585	0.00000	0.00014872	0.01219517
	MA(1)	0.1037	0.0347	2.7987	0.00282		
(2,1,0)	AR(1)	0.2168	0.0079	27.4361	0.00000	0.00014868	0.01219360
	AR(2)	0.0308	0.0079	3.8985	0.00010		
(0,1,2)	MA(1)	-0.2178	0.0078	-27.6606	0.00000	0.00014850	0.01218602
	MA(2)	-0.0896	0.0078	-11.3792	0.00000		
(2,1,2)	AR(1)	0.5617	0.1485	3.7831	0.00016	0.00014842	0.01218269
	AR(2)	-0.2627	0.0537	-4.8877	0.00000		
	MA(1)	0.3448	0.1482	2.3261	0.02002		
	MA(2)	-0.2283	0.0339	-6.7273	0.00000		
(2,1,1)	AR(1)	-0.1940	0.1658	-1.1704	0.24183	0.00014860	0.01219029
	AR(2)	0.1276	0.0351	3.6343	0.00030		
	MA(1)	-0.4103	0.1665	-2.4641	0.01375		
(1,1,2)	AR(1)	-0.1708	0.0882	-0.1936	0.84648	0.00014851	0.01218639
	MA(1)	-0.2347	0.0878	-2.6729	0.00725		
	MA(2)	-0.0929	0.0190	-4.8810	0.00000		
(2,1,3)	AR(1)	0.5536	0.1687	3.2822	0.00103	0.00014841	0.01218257
	AR(2)	-0.3545	0.0752	-4.7148	0.00000		
	MA(1)	0.3366	0.1689	1.9927	0.04630		
	MA(2)	-0.3195	0.0821	-3.8921	0.00100		
	MA(3)	-0.0255	0.0259	-0.9858	0.32423		
(3,1,2)	AR(1)	0.7528	0.1029	7.3114	0.00000	0.00014840	0.01218188
	AR(2)	-0.6296	0.1117	-5.6366	0.00000		
	AR(3)	0.0765	0.0273	2.8030	0.00506		
	MA(1)	0.5350	0.1021	5.2371	0.00000		
	MA(2)	-0.5468	0.1006	-5.4341	0.00000		

Table 6.1 Model Statistics for Methane Concentration Data.

After fitting the parameter values estimated for each of the two models to the complete series it was found that the residual autocorrelation values were lower for the (2,1,2) model. This means that its parameters fit the whole series better than the (1,1,0) model does. However, the Box-Ljung values of the whole series for the (2,1,2) model do not compare favourably to the residual series of the augmented series, over which the parameter values were originally estimated. Although they are larger, they confirm that the model (2,1,2) is still suitable but will need checks to ensure the parameter values are optimum in an on-going monitoring and forecasting situation. Such situations are discussed in chapter 8.

Various diagnostic checks have shown that the ARIMA (2,1,2) fits better than the ARIMA (1,1,0). However, because their diagnostic statistics are very similar they can

both be retained and a final choice could be made on the basis of their forecasting performance. The two models are therefore:

$$\text{ARIMA (2,1,2):} \quad \underset{(3.9)\ (-5.0)}{(1 - 0.56B + 0.26B^2)}\nabla^1 M_t = \underset{(2.4)\ (6.8)}{(1 - 0.35B - 0.23B^2)}\,\varepsilon_t \qquad [6.1]$$

$$\text{ARIMA (1,1,0):} \quad \underset{(7.6)}{(1 - 0.22B)}\nabla^1 M_t = \varepsilon_t \qquad [6.2]$$

6.3.2 Hourly Average Series of Methane Concentration

The models presented in this section were fitted to a series of hourly average values of methane concentration for April 1991. That is, from the original series the thirty observations recorded each hour were averaged. This resulted in a total series length of 720 observations. The complete plot of the series is shown in Figure 6.6. The plot suggests cyclical behaviour with two periods. The most obvious is a weekly period and the possibility of a daily period as well. Although these are evident from an inspection of the plot only the behaviour of the correlations will indicate if there are true seasonal components within the series. As always the first step in the model building process is to plot the correlations of the series, and as expected the series is not stationary. Taking non-seasonal first differences is sufficient to render the series non-seasonally stationary and the autocorrelations are shown in Figure 6.7. Lags 1, 2 and 3 have correlation values that are less than the 2sd mark and significant spikes are seen at lags 4 and 5. Very large spikes occur at lags 24 and 48. This confirms that there is a seasonal component in the series, with period 24. That is to say, there is some degree of similarity between the value at time t and t+24. Since there is a seasonal component in the series it is necessary to find the correct order of seasonal parameters before the non-seasonal parameters are fitted.

A tentative seasonal model might be an SMA(1) or possibly an SAR(1). In either case it is necessary to take a seasonal difference of period equal to 24 before the parameter values are estimated. The purpose here is to remove the seasonal effect and if this is done no spikes at lags 24 and 48 will be evident in the residual correlations. The SMA(1) has a

value of 0.83 and a t-value of 30.876 while the SAR(1) parameter has a value of -0.18 and a t-value of -4.49. The SMA(1) model seems to be the best one and this is confirmed by its removal of the seasonal spikes from the correlations, whilst the SAR(1) parameter does not. So far the model is an ARIMA $(0,0,0)(0,1,1)^{24}$, i.e. it is a multiplicative seasonal model whose non-seasonal parameters have not yet been estimated and whose seasonal component is described by an SMA(1) parameter equal to 0.83. This value will alter once non-seasonal parameters are added to the model.

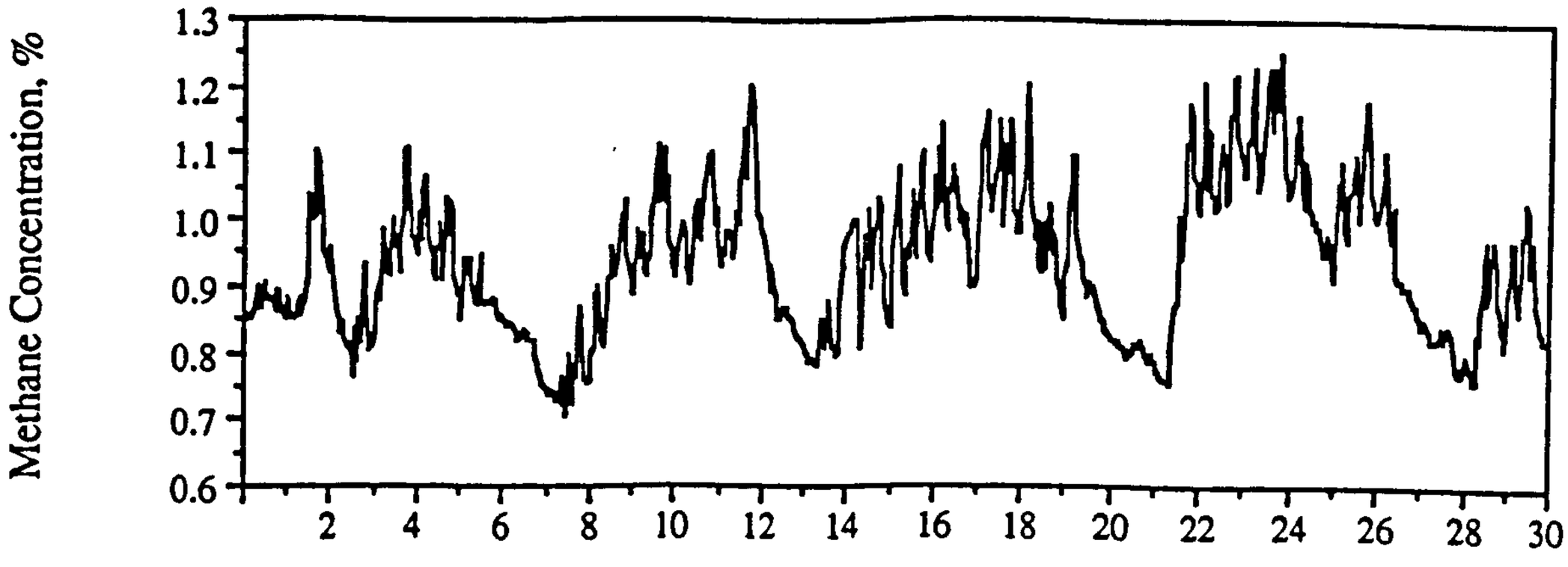


Figure 6.6 Plot of Hourly Average Methane Concentration Data.

The next step is to fit the non-seasonal parameters and model identification now proceeds as normal. After one degree of non-seasonal differencing the only significant correlations are at lags 5 and 6. A model could be estimated with parameters fitted to lags 5 and 6 only but it is most likely that lower-order parameters should be included as well. Since the seasonal component of the series is best represented by an SMA(1) model the first non-seasonal parameter to fit is an MA(1) and then re-estimate the model with an MA(5), fitted to lag 5 only. The autocorrelations of the MA(1) models residuals are all low apart from high values at lag 4, 5 and 6. Introducing the MA(5) parameter improves the fit of the model noticeably with none of the autocorrelations being significant. To test the fit of the model an AR(1) parameter was introduced. This had the effect of further reducing the Box-Ljung values but its t-value was 0.47. Furthermore, the t-values for the MA parameters were lower than for the purely MA model. This model was rejected because of the effect of the AR(1) parameter but the exercise was interesting as it seemed that

Autocorrelations: METHAPRIL HOURLY AVERAGE
Transformations: difference (1)

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	-.025	.039	.*| .	.419	.811
2	-.068	.039	.*| .	3.411	.332
3	-.012	.039	. * .	3.502	.478
4	-.039	.039	.*| .	4.674	.001
5	-.208	.039	**.*| .	48.190	.000
6	.065	.039	. |*.	50.978	.000
7	.022	.039	. * .	51.285	.000
8	.028	.039	. |*.	51.817	.000
9	.034	.039	. |*.	52.597	.000
10	.013	.039	. * .	52.706	.000
11	-.034	.039	.*| .	53.472	.000
12	.001	.039	. * .	53.472	.000
13	-.033	.039	.*| .	54.203	.000
14	.029	.039	. |*.	54.747	.000
15	.011	.039	. * .	54.821	.000
16	.090	.039	. |**	60.255	.000
17	.001	.039	. * .	60.255	.000
18	-.012	.039	. * .	60.360	.000
19	-.147	.039	*.*| .	74.898	.000
20	-.129	.039	*.*| .	86.100	.000
21	-.062	.039	.*| .	88.686	.000
22	-.003	.039	. * .	88.694	.000
23	.079	.038	. |**	92.900	.000
24	.294	.038	. |*.****	151.162	.000
25	.088	.038	. |**	156.412	.000
26	.008	.038	. * .	156.453	.000
27	-.053	.038	.*| .	158.333	.000
28	-.092	.038	**| .	164.068	.000
29	-.172	.038	*.*| .	184.195	.000
30	.028	.038	. |*.	184.736	.000
31	.002	.038	. * .	184.738	.000
32	.027	.038	. |*.	185.240	.000
33	-.020	.038	. * .	185.505	.000
34	.016	.038	. * .	185.671	.000
35	.037	.038	. |*.	186.632	.000
36	-.025	.038	.*| .	187.078	.000
37	-.055	.038	.*| .	189.133	.000
38	.052	.038	. |*.	191.024	.000
39	-.046	.038	.*| .	192.480	.000
40	.042	.038	. |*.	193.713	.000
41	.063	.038	. |*.	196.447	.000
42	.021	.038	. * .	196.749	.000
43	-.137	.038	*.*| .	209.877	.000
44	-.110	.038	**| .	218.281	.000
45	-.056	.038	.*| .	220.470	.000
46	.040	.038	. |*.	221.608	.000
47	.035	.038	. |*.	222.450	.000
48	.237	.038	. |*.***	261.853	.000
49	.065	.038	. |*.	264.801	.000
50	-.069	.038	.*| .	268.129	.000

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 650 Computable first lags after differencing: 648

Figure 6.7 Autocorrelation of Hourly Average Methane Concentration with 1 Degree of Non-Seasonal Differencing.

Autocorrelations: ERR_1 Error for DAT from ARIMA, MOD_2 NOCON

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	.003	.040	. * .	.006	.997
2	-.059	.040	.*| .	2.189	.534
3	-.005	.040	. * .	2.207	.698
4	-.072	.040	.*| .	5.436	.365
5	-.002	.040	. * .	5.439	.489
6	.064	.040	. |*.	8.051	.328
7	.007	.040	. * .	8.084	.425
8	.004	.040	. * .	8.095	.525
9	.030	.040	. |*.	8.677	.563
10	-.036	.040	.*| .	9.518	.574
11	-.032	.040	.*| .	10.163	.602
12	.000	.040	. * .	10.163	.681
13	-.040	.040	.*| .	11.177	.672
14	-.015	.039	. * .	11.320	.730
15	.016	.039	. * .	11.494	.778
16	.070	.039	. |*.	14.628	.622
17	-.028	.039	.*| .	15.123	.654
18	-.047	.039	.*| .	16.553	.620
19	-.018	.039	. * .	16.760	.668
20	-.029	.039	.*| .	17.289	.693
21	-.038	.039	.*| .	18.230	.692
22	.016	.039	. * .	18.392	.736
23	.014	.039	. * .	18.519	.777
24	.014	.039	. * .	18.648	.814

Total cases: 650 Computable first lags: 624

Partial Autocorrelations: ERR_1 Error for DAT from ARIMA, MOD_2 NOCON

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
1	.014	.038	. * .
2	-.053	.038	.*| .
3	-.008	.038	. * .
4	-.076	.038	**| .
5	-.002	.038	. * .
6	.058	.038	. |*.
7	.002	.038	. * .
8	.000	.038	. * .
9	.027	.038	. |*.
10	-.029	.038	.*| .
11	-.027	.038	.*| .
12	.003	.038	. * .
13	-.046	.038	.*| .
14	-.021	.037	.*.
15	.015	.037	.*.
16	.068	.037	.|*
17	-.016	.037	.*.
18	-.037	.037	*|.
19	-.025	.037	*|.
20	-.035	.037	*|.
21	-.034	.037	*|.
22	.016	.037	.*.
23	.022	.037	.*.
24	.021	.037	.*.

Total cases: 650 Computable first lags: 624

Figure 6.8 Residual Correlations of the ARIMA $(0,1,[1,5])(0,1,1)^{24}$ Model.

Model	Parameter	B	SEB	T-value	Probability	Res. Var	Res. SE
S(1,1,0)	SAR(1)	-0.1822	0.0405	-4.4972	0.00000	0.01440681	0.12002838
S(0,1,1)	SMA(1)	0.8295	0.0268	30.8673	0.00000	0.01232511	0.11101852
(0,1,1)	MA(1)	0.1450	0.0380	3.8106	0.00015	0.00136551	0.0369528
(0,1,0)	MA(1)	0.1450	0.0380	3.8106	0.00015	0.00136551	0.0369528
q=1,5	MA(5)	0.0969	0.0380	2.5518	0.01095		
	SMA(1)	0.9447	0.0373	25.287	0.00000		
(1,1,0)	AR(1)	0.1065	0.2238	0.4758	0.63437	0.00136739	0.03697828
q=1,5	MA(1)	0.2540	0.2168	1.1713	0.24191		
	MA(5)	0.0882	0.0375	2.3490	0.01913		
	SMA(1)	0.9447	0.0374	25.2616	0.00000		
(0,1,0)	AR(1)	-0.1373	0.0380	-3.6084	0.00033	0.00136656	0.03696701
p=1,5	AR(5)	-0.1047	0.0380	-2.7555	0.00603		
	SMA(1)	0.9436	0.0369	25.5801	0.00000		
(0,1,0)	MA(1)	0.1373	0.0383	3.5840	0.00036	0.00136297	0.03691840
q=(1,2,5)	MA(2)	0.0672	0.0383	1.7568	0.07945		
	MA(5)	0.0955	0.0378	2.5221	0.01191		
	SMA(1)	0.9438	0.0371	25.4654	0.00000		

Table 6.2 Model Statistics for Methane Concentration Data (Hourly Average Series).

introducing AR parameters into the model could lower the Box-Ljung values but result in dubious parameter statistics. To test this theory the model was re-estimated with only AR parameters at lags 1 and 5.

The residual Box-Ljung values for the AR(1,5) model were lower than for the MA(1,5) model at higher lags but its residual statistics were higher indicating that the MA(1,5) model was still the best fit so far. All that remained to be done was to add extra parameters until no improvement in the residual statistics occurred. The MA(1,5) model was fitted with an MA(2) parameter making it an MA(1,2,5) model. The addition of the MA parameter at lag 2 further reduced the residual Box-Ljung values and lowered the residual statistics slightly. However, its t-value was low and in the interest of model parsimony it was rejected. No further investigation was attempted and the final model for the series was the SARIMA(0,1,[1,5])$(0,1,1)^{24}$ given by:

$$\nabla^1\nabla^{24}M_t = \underset{(3.8)}{(1 - 0.15B + 0.10B^5)}\underset{(-2.6)}{}\underset{(25.5)}{(1 - 0.94B^{24})}\ \varepsilon_t \qquad [6.3]$$

Fitting this model to the whole of the series the residual autocorrelations compare favourably with those from the historical estimation period, shown in Figure 6.8, confirming that the model is suitable.

6.3.3 10-Minute Average Series of Methane Concentration

The original methane concentration data was averaged into 10-minute values to produce this third series. It had a total series length of 4320 values. The series was split into a historical portion of 3500 observations. The complete series was too long to plot onto one graph and Figure 6.9 shows the plot of the first 10 days worth of data. An inspection of this plot reveals that the series is most likely non-stationary and this proves to be the case once the correlations are calculated. After one degree of non-seasonal differencing the series is rendered stationary and the differenced correlations are shown in Figure 6.10. The pattern of the differenced auto and partial correlations leads to the tentative identification of a 2nd order autoregressive process but it is by no means clear what the generating process could be.

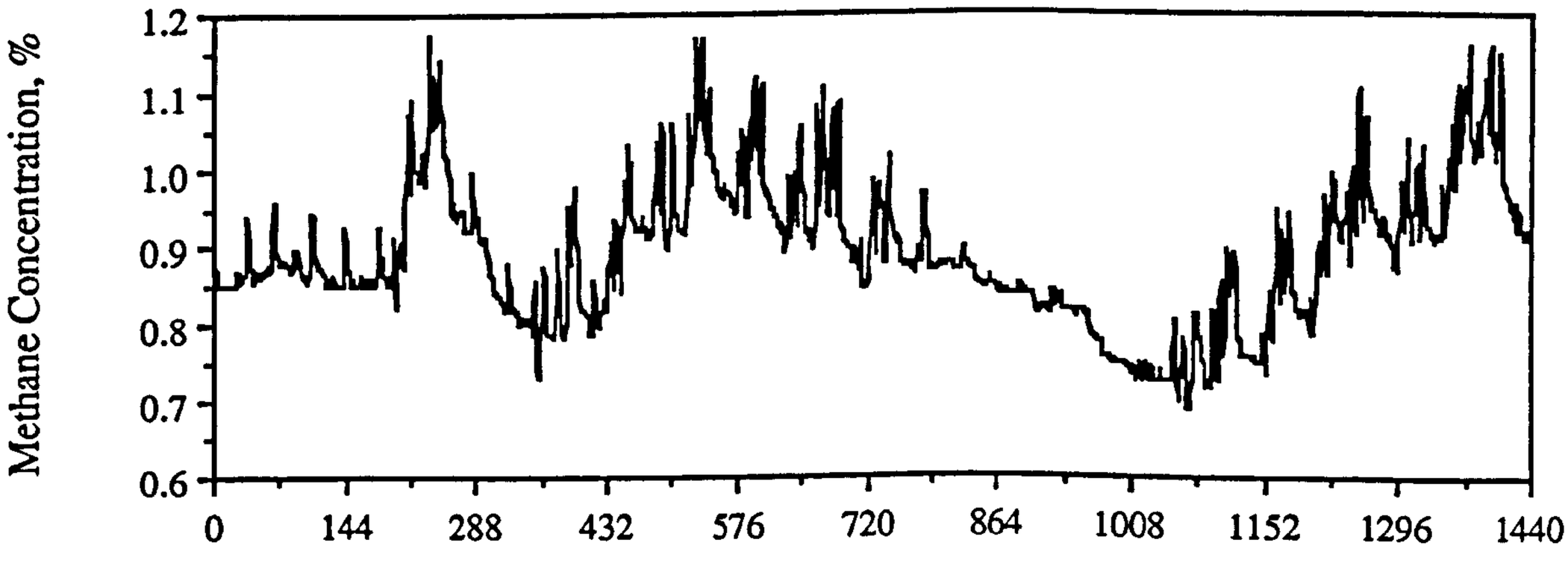


Figure 6.9 Plot of 10-Minute Averages of Methane Concentration.

```
Autocorrelations:   DAT
 Transformations:   difference (1)

Lag  Corr.    Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1    Box-Ljung  Prob.
                   +----+----+----+----+----+----+----+----+----+
  1   .304    .017                    .|.*****                      323.322   .000
  2  -.183    .017                 ***.|.                           440.863   .000
  3  -.329    .017                ******.|.                         821.214   .000
  4  -.258    .017                  ****.|.                        1053.865   .000
  5  -.064    .017                    *|.                          1068.064   .000
  6   .121    .017                    .|.*                         1119.446   .000
  7   .184    .017                    .|.***                       1238.081   .000
  8   .082    .017                    .|.*                         1261.399   .000
  9  -.058    .017                    *|.                          1273.060   .000
 10  -.120    .017                   *.|.                          1323.208   .000
 11  -.071    .017                    *|.                          1340.878   .000
 12  -.014    .017                    .*.                          1341.568   .000
 13   .015    .017                    .*.                          1342.361   .000
 14   .049    .017                    .|*                          1350.923   .000
 15   .046    .017                    .|*                          1358.309   .000
 16   .013    .017                    .*.                          1358.943   .000
 17  -.016    .017                    .*.                          1359.854   .000
 18  -.018    .017                    .*.                          1360.948   .000
 19  -.029    .017                    *|.                          1363.911   .000
 20   .003    .017                    .*.                          1363.943   .000

 Plot Symbols:      Autocorrelations *     Two Standard Error Limits .

 Total cases:  3500     Computable first lags after differencing:  3498

Partial Autocorrelations:   DAT
 Transformations:   difference (1)

Lag  Corr.    Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1
                   +----+----+----+----+----+----+----+----+----+
  1   .304    .017                    .|.*****
  2  -.304    .017                *****.|.
  3  -.200    .017                  ***.|.
  4  -.161    .017                   **.|.
  5  -.064    .017                    *|.
  6   .013    .017                    .*.
  7   .037    .017                    .|*
  8  -.018    .017                    .*.
  9  -.033    .017                    *|.
 10  -.029    .017                    *|.
 11   .000    .017                    .*.
 12  -.033    .017                    *|.
 13  -.041    .017                    *|.
 14   .004    .017                    .*.
 15   .004    .017                    .*.
 16   .005    .017                    .*.
 17   .002    .017                    .*.
 18   .010    .017                    .*.
 19  -.020    .017                    .*.
 20   .023    .017                    .*.

 Plot Symbols:      Autocorrelations *     Two Standard Error Limits .
 Total cases:  3500     Computable first lags after differencing:  3498
```

Figure 6.10 Correlations of 10-Minute Average Methane Concentration for 1 Degree of Differencing.

Model	Parameter	B	SEB	T-value	Probability	Res. Var	Res. SE
(1,1,0)	AR(1)	0.3037	0.0161	18.8581	0.00000	0.00068410	0.02615525
(0,1,1)	MA(1)	-0.3994	0.0155	-25.7643	0.00000	0.00065417	0.02557678
(1,1,1)	AR(1)	-0.0427	0.0425	-1.0037	0.31558	0.00065411	0.02557565
	MA(1)	-0.4328	0.0384	-11.2827	0.00000		
(2,1,0)	AR(1)	0.3960	0.0161	24.5760	0.00000	0.00062125	0.02492486
	AR(2)	-0.3033	0.0161	-18.8292	0.00000		
(0,1,2)	MA(1)	-0.3669	0.0169	-21.7270	0.00000	0.00065362	0.02556608
	MA(2)	-0.0526	0.0169	3.1168	0.00184		
(2,1,2)	AR(1)	0.9844	0.0286	34.3965	0.00000	0.00058139	0.02411197
	AR(2)	-0.5797	0.0286	-20.2812	0.00000		
	MA(1)	0.6953	0.0348	19.9809	0.00000		
	MA(2)	-0.1359	0.0348	-3.9086	0.00009		
(2,1,1)	AR(1)	0.9246	0.0266	34.6898	0.00000	0.00058283	0.02414182
	AR(2)	-0.4762	0.0149	-32.0170	0.00000		
	MA(1)	0.6160	0.0285	21.6385	0.00000		
(1,1,2)	AR(1)	0.6644	0.0291	22.7952	0.00000	0.00061163	0.02473106
	MA(1)	0.3976	0.0284	13.9994	0.00000		
	MA(2)	0.4169	0.0154	27.0844	0.00000		
(2,1,3)	AR(1)	1.0787	0.0431	24.9875	0.00000	0.00058016	0.02408646
	AR(2)	-0.6601	0.0333	-19.7941	0.00000		
	MA(1)	0.7813	0.0469	16.6765	0.00000		
	MA(2)	-0.1701	0.0339	-5.0241	0.00000		
	MA(3)	-0.0769	0.0278	-2.7657	0.00571		
(3,1,2)	AR(1)	-0.0131	0.0252	-0.5216	0.60198	0.00058445	0.02417538
	AR(2)	0.5116	0.0241	21.2631	0.00000		
	AR(3)	-0.4736	0.0152	-31.1136	0.00000		
	MA(1)	-0.3154	0.0256	-12.3054	0.00000		
	MA(2)	0.6841	0.0253	27.0554	0.00000		

Table 6.3 Model Statistics for 10-Minute Average Model.

Estimating an ARIMA (2,1,0) results in the AR1 parameter having a value of 0.39 (t-value = 24.57) and the AR2 parameter having a value of -0.30 (t-value = -18.83). Both of these parameters are significant but an examination of the residual autocorrelations reveals that they are still highly correlated. Therefore, the original identification is inadequate and it is necessary to fit additional parameters. This could be done on a trial and error basis but it was more efficient to fit a range of models to the series automatically, as done for the modelling of the original series. Such a procedure, where it is evident that a relatively simple model is sufficient to model a series, allows model parameters to be rejected if their t-values are too low. The residuals of the resulting models are then compared to see which is the best fit. Table 6.3 details the various model statistics, including the residual standard error and variance.

Autocorrelations: ERR_17 Error for DAT from ARIMA, MOD_35 NOCON

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	-.003	.017	.*.	.023	.989
2	-.009	.017	.*.	.295	.961
3	.018	.017	.*.	1.442	.837
4	-.011	.017	.*.	1.848	.870
5	-.031	.017	*|.	5.149	.525
6	-.020	.017	.*.	6.538	.478
7	.012	.017	.*.	7.061	.530
8	-.023	.017	.*.	8.990	.438
9	-.047	.017	*|.	16.800	.079
10	-.052	.017	*|.	26.161	.006
11	.002	.017	.*.	26.173	.010
12	.007	.017	.*.	26.355	.015
13	-.021	.017	.*.	27.972	.014
14	.011	.017	.*.	28.396	.019
15	.016	.017	.*.	29.273	.022
16	.008	.017	.*.	29.503	.030
17	-.005	.017	.*.	29.595	.042
18	.018	.017	.*.	30.731	.043
19	-.024	.017	.*.	32.807	.035
20	.023	.017	.*.	34.745	.030

Plot Symbols: Autocorrelations * Two Standard Error Limits .

Total cases: 3500 Computable first lags: 3498

Partial Autocorrelations: ERR_17 Error for DAT from ARIMA, MOD_35 NOCON

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
1	.003	.015	.*.
2	.005	.015	.*.
3	.010	.015	.*.
4	-.011	.015	.*.
5	-.023	.015	.*.
6	-.010	.015	.*.
7	.012	.015	.*.
8	-.021	.015	.*.
9	-.049	.015	*|.
10	-.057	.015	*|.
11	-.009	.015	.*.
12	.000	.015	.*.
13	-.013	.015	.*.
14	.020	.015	.*.
15	.010	.015	.*.
16	.000	.015	.*.
17	-.006	.015	.*.
18	.013	.015	.*.
19	-.019	.015	.*.
20	.018	.015	.*.

Plot Symbols: Autocorrelations * Two Standard Error Limits .

Total cases: 3500 Computable first lags: 3498

Figure 6.11 Correlations of the ARIMA (2,1,3) Fit.

Autocorrelations: ERR_18 Error for DAT from ARIMA, MOD_36 NOCON

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	-.002	.015					.*.					.021	.989
2	-.010	.015					.*.					.422	.936
3	.013	.015					.*.					1.181	.881
4	-.008	.015					.*.					1.478	.916
5	-.030	.015					*|.					5.471	.485
6	-.024	.015					.*.					7.905	.341
7	.002	.015					.*.					7.917	.442
8	-.022	.015					.*.					10.097	.343
9	-.043	.015					*|.					18.093	.053
10	-.048	.015					*|.					28.011	.003
11	-.002	.015					.*.					28.025	.005
12	.002	.015					.*.					28.037	.009
13	-.017	.015					.*.					29.259	.010
14	.016	.015					.*.					30.373	.011
15	.009	.015					.*.					30.687	.015
16	.001	.015					.*.					30.695	.022
17	-.003	.015					.*.					30.737	.031
18	.015	.015					.*.					31.709	.034
19	-.018	.015					.*.					33.140	.033
20	.018	.015					.*.					34.496	.032

Plot Symbols: Autocorrelations * Two Standard Error Limits .

Total cases: 4320 Computable first lags: 4318

Partial Autocorrelations: ERR_18 Error for DAT from ARIMA, MOD_36 NOCON

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-.002	.015					.*.				
2	-.010	.015					.*.				
3	.013	.015					.*.				
4	-.008	.015					.*.				
5	-.030	.015					*|.				
6	-.024	.015					.*.				
7	.001	.015					.*.				
8	-.022	.015					.*.				
9	-.043	.015					*|.				
10	-.050	.015					*|.				
11	-.004	.015					.*.				
12	.001	.015					.*.				
13	-.018	.015					.*.				
14	.011	.015					.*.				
15	.003	.015					.*.				
16	-.001	.015					.*.				
17	-.006	.015					.*.				
18	.010	.015					.*.				
19	-.022	.015					.*.				
20	.016	.015					.*.				

Plot Symbols: Autocorrelations * Two Standard Error Limits .

Total cases: 4320 Computable first lags: 4318

Figure 6.11A Correlations of the Complete ARIMA (2,1,3) Fit.

From Table 6.3 it is seen that 9 models contain parameters that constitute a suitable model. Of these models the ARIMA (2,1,3) has the best fit as seen by a comparison of its residual statistics shown in Table 6.3. The next best model is the ARIMA (2,1,2) and the ARIMA (2,1,3) is the result of an acceptable overfitting of this model. Adding an AR3 parameter to the ARIMA (2,1,2) produces an AR1 parameter with a very low t-value and a worsening of the residual statistics. Thus the ARIMA (3,1,2) is rejected. A slight improvement of the residual statistics is achieved by fitting an MA3 parameter. This parameter is small and indicates that there will be nothing to be gained by the addition of any further parameters.

The residual autocorrelations of the ARIMA(2,1,3) are shown in Figure 6.11. The correlation values are low up to lag 6 after which they are somewhat larger. However, this is of no concern and is due to the large number of observations used to estimate the model. The final test of the model competence is to fit the parameters to the complete series, i.e. the historical and validation portions. The autocorrelations for the complete series show that the parameters fit well and are shown in Figure 6.11A.

The best model to represent this series is:

$$\text{ARIMA (2,1,3):} \quad \underset{(25.09)\ (-19.8)}{(1 - 1.08B + 0.66B^2)}\nabla^1 M_t = \underset{(16.7)\ (-5.0)\ (-2.8)}{(1 - 0.78B + 0.17B^2 + 0.08B^3)}\varepsilon_t \quad [6.4]$$

6.3.4 One Day Series of Methane Concentration

The last model of this section was built from one whole days data of methane concentration taken from 10th April 1991. The series is 720 observations in length and is seen in Figure 6.12. During the model building stages the series was split into a historical portion of 650 observations and a validation portion of 70 observations. The plot shows several rise and fall cycles over the day but there is no evidence of any real periodic behaviour.

Taking first differences of the raw series was sufficient to render it stationary. Both the autocorrelation and partial autocorrelation plots have significant values at lag 1 while there

is also a significant value at lag 11 in the autocorrelations. This could be ignored since the 95% confidence limits generated for the correlations expect that 1 in 20 values will exceed the 2sd mark by chance alone. Thus, the identification of a tentative model is relatively easy as the correlations indicate a possible ARIMA(1,1,1) process. However, the estimation of this model indicates overfitting as the MA parameter has an insignificant t-value of 0.25. Also, the Box-Ljung values increased markedly from lag 11 onwards as seen in the residual autocorrelations, shown in Figure 6.13. Re-estimating with just an AR(1) parameter produced residuals with low Box-Ljung values up to lag 10, after which they also increased markedly. This was unusual as the original and differenced correlations did not indicate any possible reasons to explain this behaviour.

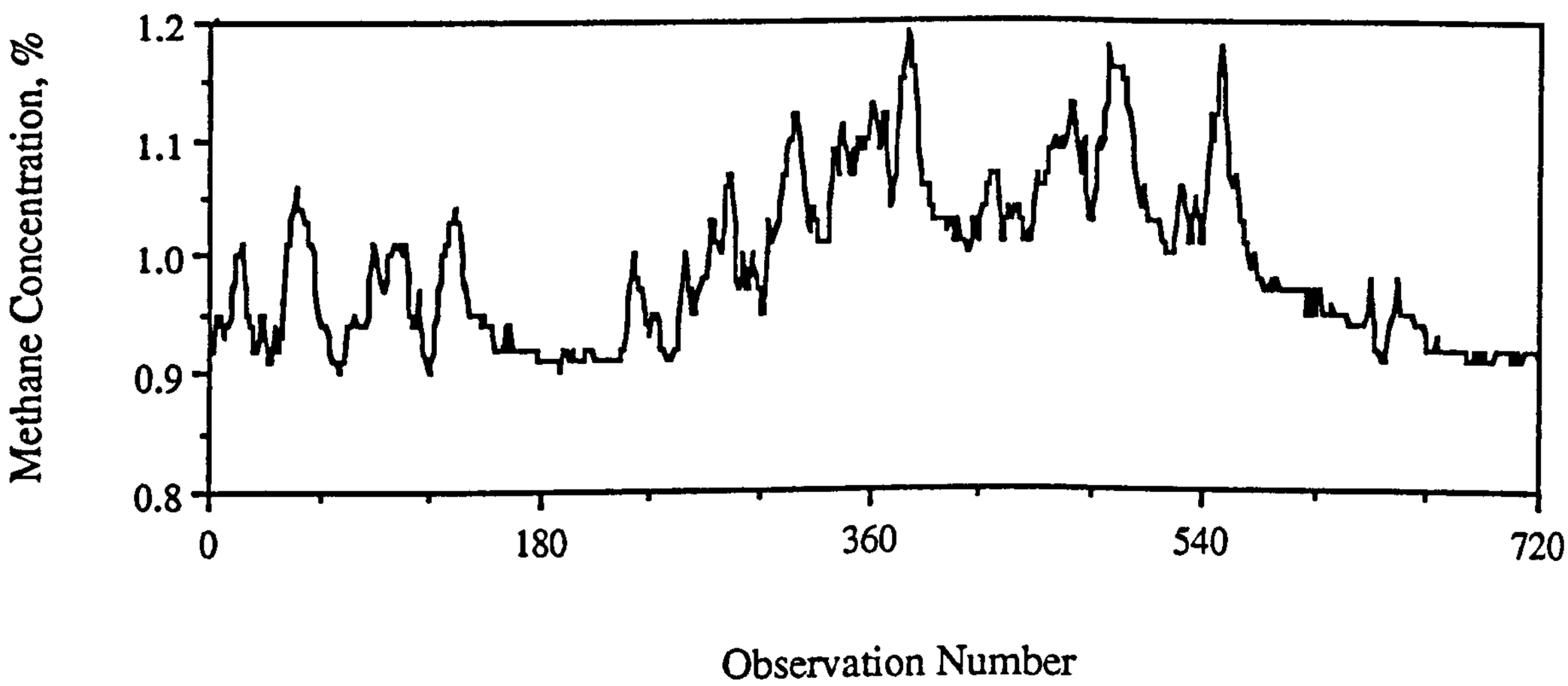


Figure 6.12 Plot of One Days Methane Concentration.

It was possible that the series could contain a seasonal component with a period of 10 but from an a priori consideration of the raw series there was no reason to justify this. This hypothesis was investigated by assigning a seasonal period of 10 to the data but the corresponding changes in the autocorrelation behaviour confirmed that no seasonality was present. Another attempt was made to lower the Box-Ljung values from lag 11 onwards by fitting an AR parameter at lag 11. This parameter proved to be significant with a value of -0.12 (t-value = -3.28) and an examination of the residual correlations revealed that its addition was appropriate. The autocorrelations of the residual series shown in Figure 6.14 are an improvement on those from the ARIMA (1,1,0) but the autocorrelation Box-Ljung values still increase markedly from lag 10 onwards. This indicates that many more

Autocorrelations: ERR_1 Error for DAT from ARIMA, MOD_1 NOCON

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	.000	.039	. * .	.000	1.000
2	.008	.039	. * .	.039	.998
3	-.005	.039	. * .	.057	1.000
4	-.057	.039	.*| .	2.173	.825
5	-.004	.039	. * .	2.182	.902
6	.007	.039	. * .	2.216	.947
7	.020	.039	. * .	2.484	.962
8	-.031	.039	.*| .	3.116	.959
9	.033	.039	. |*.	3.841	.954
10	-.042	.039	.*| .	8.249	.691
11	-.117	.039	**| .	17.306	.138
12	-.053	.039	.*| .	19.186	.117
13	.014	.039	. * .	19.323	.153
14	.029	.039	. |*.	19.901	.176
15	.021	.039	. * .	20.188	.212
16	-.069	.039	.*| .	23.406	.136
17	-.070	.039	.*| .	26.637	.086
18	-.058	.039	.*| .	28.866	.068
19	-.060	.039	.*| .	31.278	.052
20	-.026	.039	.*| .	31.727	.062

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 650 Computable first lags: 648

Figure 6.13 Residual Autocorrelations of the ARIMA (1,1,1) Model.

Autocorrelations: ERR_3 Error for DAT from ARIMA, MOD_5 NOCON

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	-.005	.039	. * .	.018	.991
2	.024	.039	. * .	.383	.944
3	-.003	.039	. * .	.387	.984
4	-.056	.039	.*| .	2.476	.780
5	-.015	.039	. * .	2.615	.855
6	-.005	.039	. * .	2.631	.917
7	.004	.039	. * .	2.640	.955
8	-.044	.039	.*| .	3.943	.915
9	.027	.039	. |*.	4.428	.926
10	-.041	.039	.*| .	10.387	.496
11	.006	.039	. * .	10.409	.580
12	-.021	.039	. * .	10.694	.636
13	.030	.039	. |*.	11.291	.663
14	.031	.039	. |*.	11.915	.685
15	.010	.039	. * .	11.988	.745
16	-.071	.039	.*| .	15.345	.571
17	-.081	.039	**| .	19.783	.345
18	-.065	.039	.*| .	22.627	.254
19	-.066	.039	.*| .	25.558	.181
20	-.033	.039	.*| .	26.306	.195

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 650 Computable first lags: 648

Figure 6.14 Residual Autocorrelations of the ARIMA ([1,11],1,0) Model.

additional parameters could be fitted to the model to reduce the residual autocorrelations at higher lags. The residual variance for the model with the additional AR11 parameter is lowered by 1.5% as compared to the ARIMA (1,1,0). This improvement is small and demonstrates that little is gained by increasing the number of model parameters, especially at higher lags. Thus, the most suitable model for this series is a very simple ARIMA (1,1,0) given by:

$$\text{ARIMA (1,1,0):} \qquad \underset{(114.7)}{(1 - 0.26B)}\nabla^1 M_t = \varepsilon_t \qquad [6.5]$$

A final test to the models suitability is how well the AR parameter fits the complete series. In this case the residuals after the model has been fitted to the complete series do not increase greatly nor is their any appreciable change in the Box-Ljung values. This confirms that the model is suitable to represent the original series.

Model	Parameter	B	SEB	T-value	Probability	Res. Var	Res. SE
(1,1,1)	AR(1)	0.2979	0.1420	2.0977	0.03632	0.00016431	0.01281835
	MA(1)	0.0366	0.1487	0.2459	0.80582		
(1,1,0)	MA(1)	0.2638	0.0379	6.9633	0.00000	0.00016408	0.01280918
(0,1,0)	MA(1)	0.2503	0.0378	6.6169	0.00000	0.00016168	0.01271543
p=1,11	MA(11)	-0.1230	0.0381	-3.2281	0.00131		

Table 6.4 Parameter Statistics for One Days Methane Concentration Data.

6.4.1 Original Series of Air Velocity

As mentioned earlier, only brief descriptions of the model building process are now presented for all of the univariate models in this chapter. Air velocity was chosen as it is a variable that effects methane concentration in the short-term. That is, even if methane emission is constant, the air velocity is usually continuously varying and hence so is methane concentration. It was not thought necessary to transform the original air velocity series into air quantity values since this involves a constant scalar that is redundant. As usual the series was split into a historical portion of 16,000 observations and a validation portion of 5500 observations. The plot of a portion of the series is shown in Figure 6.15.

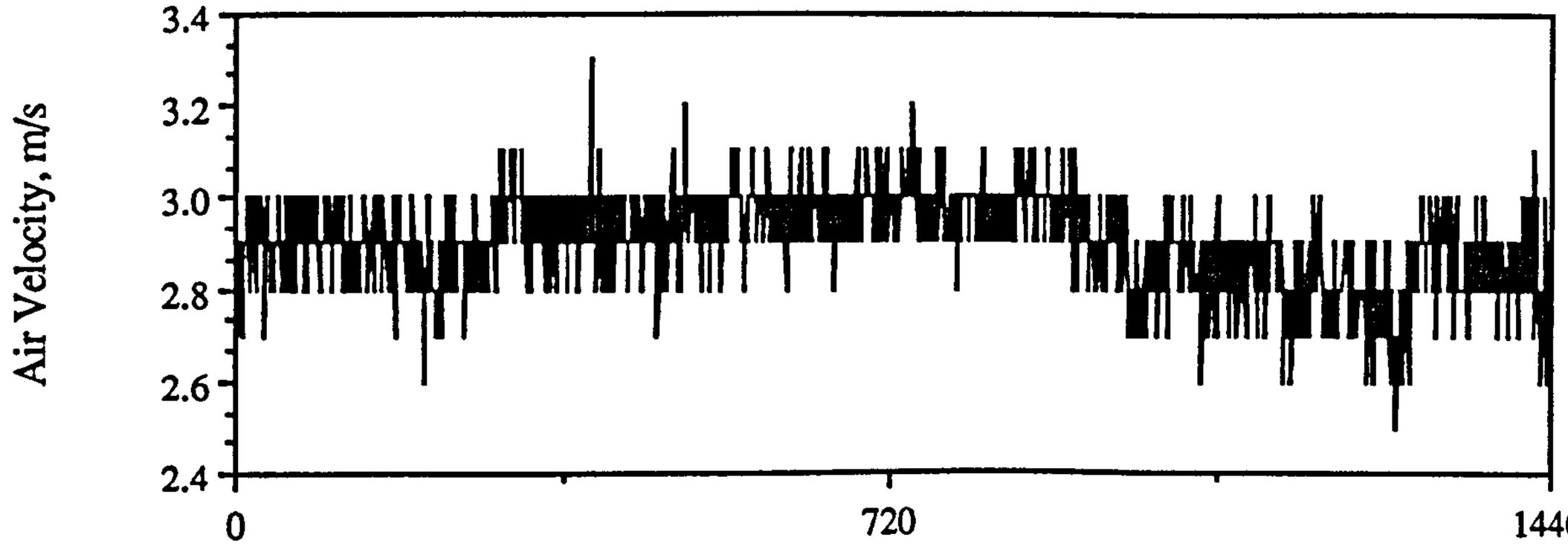


Figure 6.15 Plot of Air Velocity.

The air velocity values fluctuate over an average range of 0.2 m/s, with occasional large variations that could be due to a number of reasons. A large number of observations were lost due to transmission failure and these were replaced by instructing SPSS-X Trends™ to assign likely values. The vortex head of the velocity meter is sited adjacent to the sensing head of the methane monitor (see Plate 5.1, chapter 5). Thus, it is likely that the variation in the methane concentration (not methane emission) value is determined by the air velocity (hence quantity).

The series was rendered stationary by one degree of differencing and proceeding through the structured modelling process, the best standard model was an ARIMA (1,1,2). The t-value of the MA1 parameter was insignificant (-1.35) and re-estimating as an ARIMA (1,1,[2]) resulted in an improved fit. The residual Box-Ljung values were low until lag 13 after which they increased dramatically. An MA parameter was added to the model at lag 13 and this reduced the residual variance by 2.3%. This was small and the inclusion of the AR parameter at lag 13 was not thought worthwhile. Thus the model for the original air velocity series was:

$$\text{ARIMA (1,1,[2])}: \quad \underset{(-114.7)}{(1 + 0.75B)}\nabla^1 A_t = \underset{(73.1)}{(1 - 0.58B^2)}\varepsilon_t \qquad [6.5]$$

6.4.2 Hourly Average Series of Air Velocity

The plot of air velocity hourly averages for the complete month of April 1991 is shown in Figure 6.16. There are three items of interest on this plot. The first two are periods over which no actual data was available and are represented by the two very steady portions of the plot. The third is a very large drop in air velocity over the period 19 to 20th April, after which the velocity recovers to its usual level. On investigation, no circumstances were apparent to explain this behaviour. Both of these factors are detrimental to the building of a representative model but their effects are attenuated due to the large number of observations used.

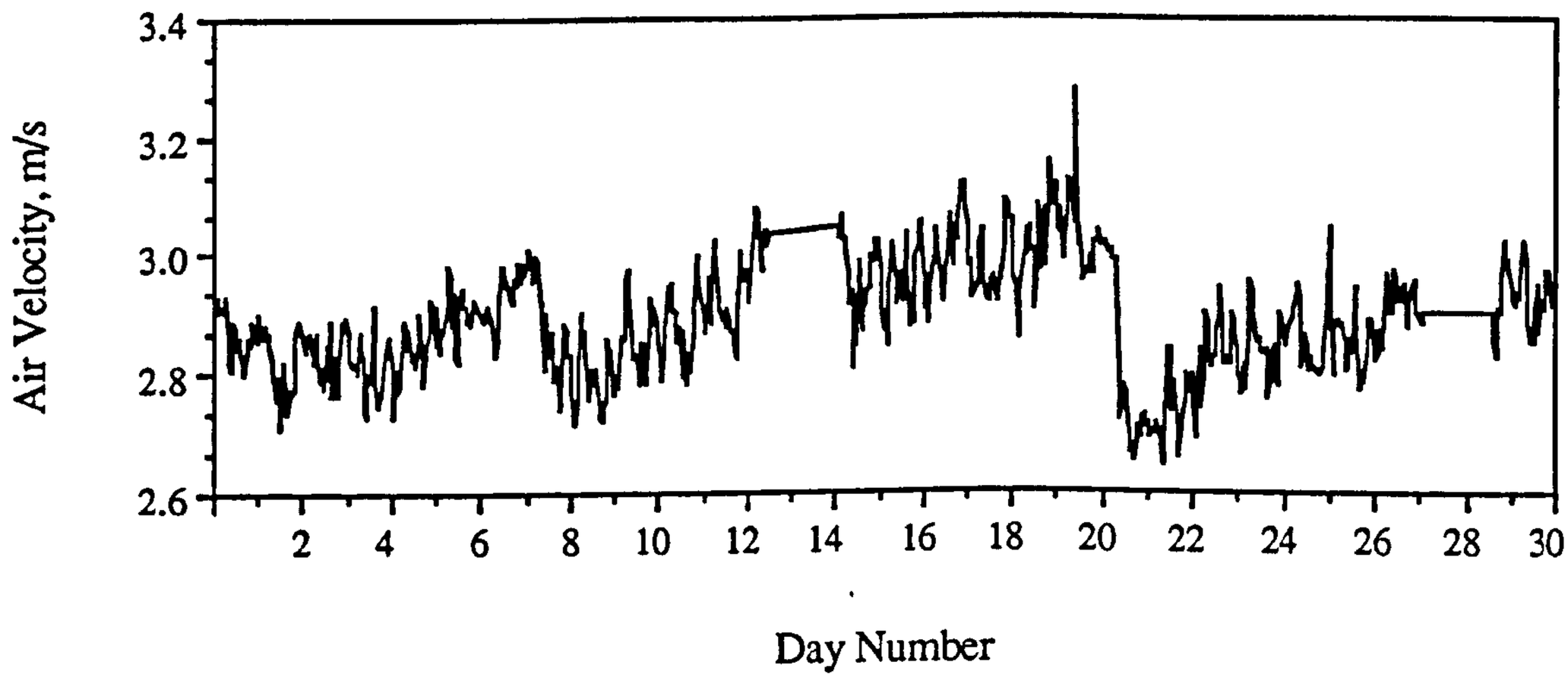


Figure 6.16 Plot of Air Velocity Hourly Averages.

The differenced correlations demonstrated weak seasonality with a period of 24, identical to the seasonality of the hourly average methane concentration series. This was surprising since there was no reason to explain this behaviour. To test whether this assumption was correct, models were estimated with and without seasonal components and revealed that models with a seasonal SMA(1) component fitted best. The best, simple model was an ARIMA $(1,1,1)(0,1,1)^{24}$ and this was improved by fitting an extra AR parameter to lag 7 in an attempt to lower the residual Box-Ljung values. However, the addition of the AR7 parameter only improved the residual statistics by a small amount. Thus, the best model to represent the series was the ARIMA $(1,1,1)(0,1,1)^{24}$ given by:

$$(1 - 0.46B)\nabla^1\nabla^{24}A_t = (1 - 0.47B)(1 - 0.80B^{24})\,\varepsilon_t \qquad [6.7]$$

(6.5) (15.4) (29.5)

Fitting the model to the complete series actually resulted in a lowering of the residual autocorrelations Box-Ljung values, indicating a good model fit.

6.4.3 10-Minute Average Series of Air Velocity

The 10-minute average series of air velocity did not show any seasonal behaviour and it was relatively straightforward to build a time series model for it. The plot of the first ten days worth of data is shown in Figure 6.17. The plot displays no notable characteristics apart from indicating that the series is non-stationary both in mean and variance. One degree of non-seasonal differencing was sufficient to render it stationary.

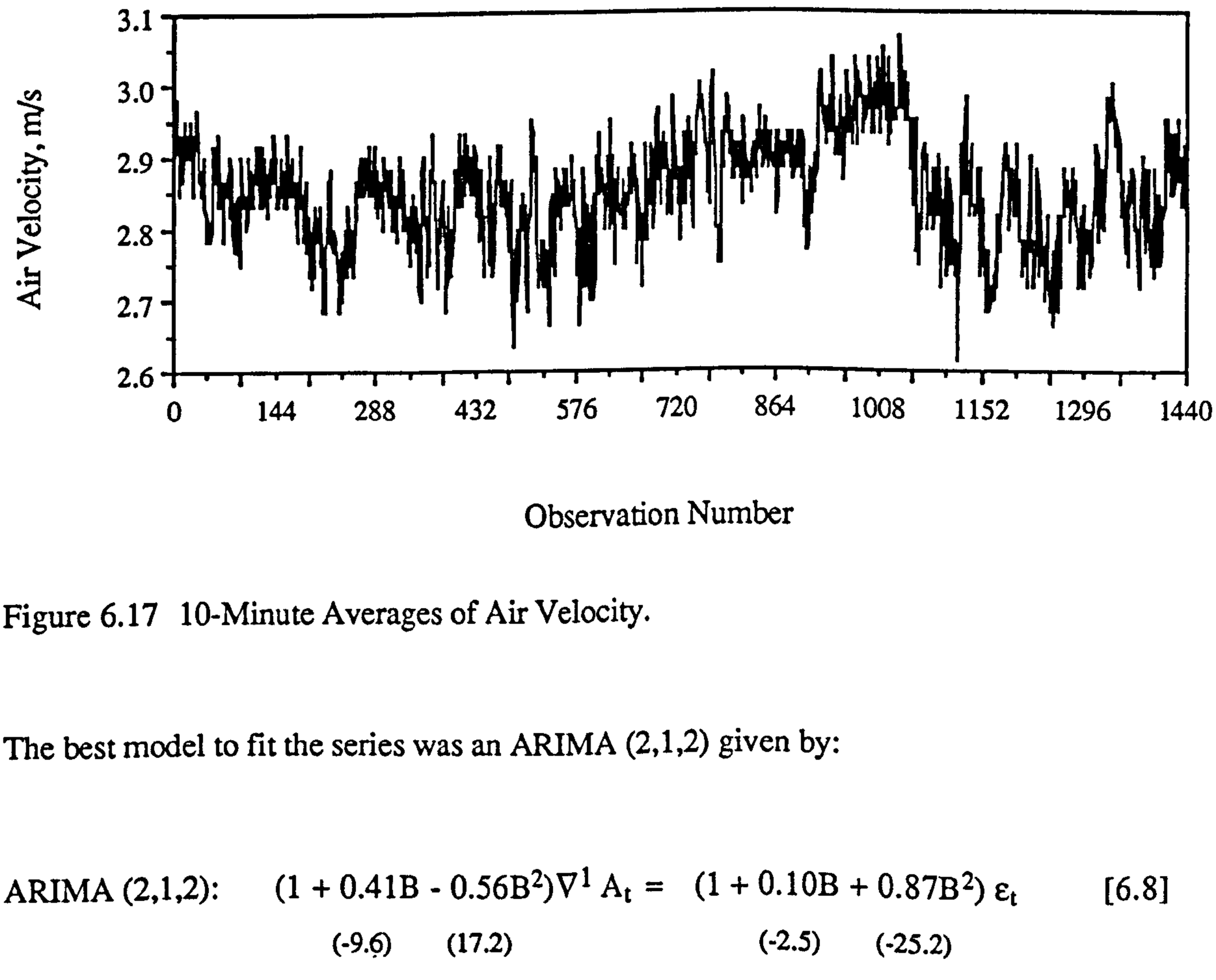


Figure 6.17 10-Minute Averages of Air Velocity.

The best model to fit the series was an ARIMA (2,1,2) given by:

$$\text{ARIMA (2,1,2):} \quad (1 + 0.41B - 0.56B^2)\nabla^1 A_t = (1 + 0.10B + 0.87B^2)\,\varepsilon_t \qquad [6.8]$$

(-9.6) (17.2) (-2.5) (-25.2)

Although the value of the MA1 parameter was small its t-value of -2.5 warranted its retention. The model fitted the series well with none of the residual autocorrelations having significant Box-Ljung values up to lag 20. The final test was to apply the model to the complete series and it was found that the residual autocorrelation Box-Ljung values only altered slightly, indicating the suitability of the model parameters.

6.4.4 One Day Series of Air Velocity

The plot of the series for one days worth of air velocity recorded on the 10th April 1991 can be seen in Figure 6.18. It was found that a very simple ARIMA (0,1,1) model fitted the series very well, but on testing it by gross overfitting it was found that an ARIMA (2,1,3) was also suitable. Both of the models are:

ARIMA (0,1,1): $\nabla^1 A_t = (1 - 0.86B)\,\varepsilon_t$ [6.9]

(42.5)

ARIMA (2,1,3):

$$(1 + 1.91B + 0.94B^2)\nabla^1 A_t = (1 + 1.06B - 0.70B^2 - 0.82B^3)\varepsilon_t \quad [6.10]$$

(-52.2) (-27.3) (-24.8) (17.4) (22.3)

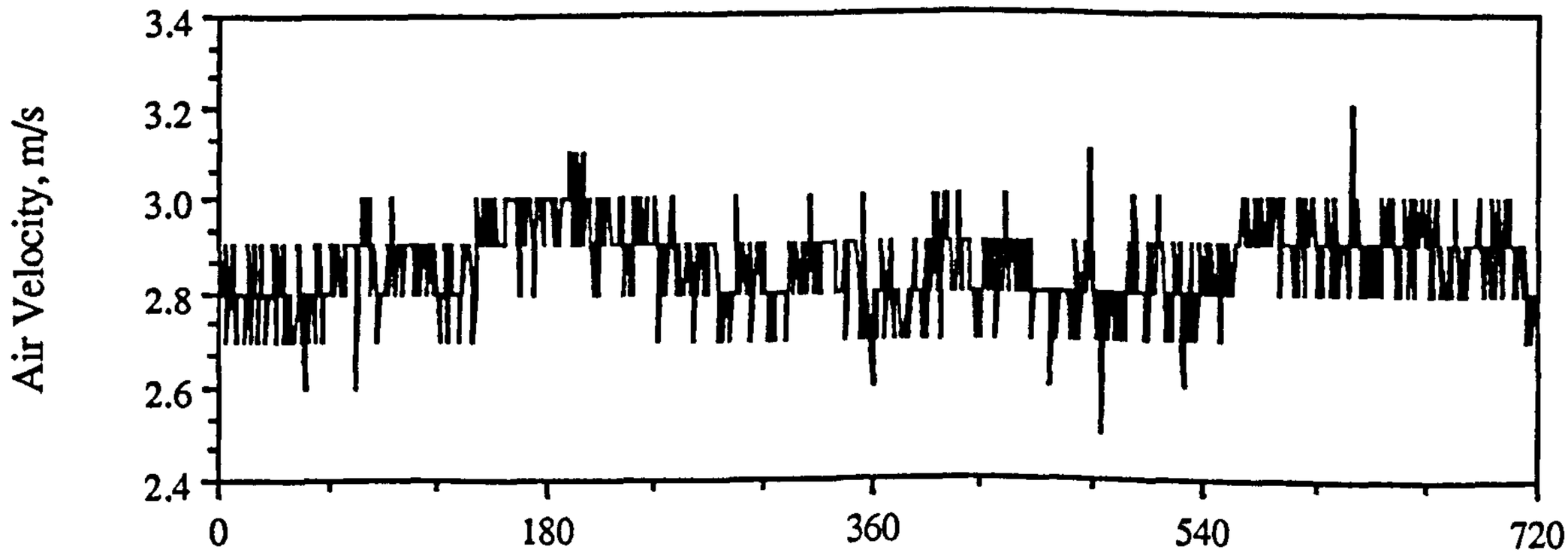


Figure 6.18 Plot of One Days Air Velocity.

The ARIMA (2,1,3) has lower Box-Ljung values for the residual autocorrelations but the fitting of the extra parameters to the ARIMA (0,1,1) model only achieves a 0.7% reduction in residual variance. This gain is very small and in the interests of model parsimony the ARIMA (2,1,3) is rejected and the simple ARIMA (0,1,1) retained.

6.5.1 Original Series of Barometric Pressure

It was not possible to build a time series model for the raw series of barometric pressure. Figure 6.19 shows a plot of two days worth of data. The series was characterized by periods of rising and falling barometric pressure with periods where the pressure was constant. The correlations of the series indicated that one degree of non-seasonal differencing was necessary and the differenced autocorrelations are shown in Figure 6.20. The first three values are zero and mean that the series is most likely random. Large spikes are then seen at lags 5, 8, 10, 12, 15, 16 which is very suggestive of some form of seasonal behaviour. A priori, this is completely unexpected since the weather cycle over the month of April should not result in the barometric pressure series containing a seasonal component. In fact, from a priori considerations it is not unreasonable to expect that the series be completely random. It may be possible to assign some degree of cyclical behaviour to atmospheric pressure when considered on a monthly basis from year to year but not during any particular month on its own. Thus the seemingly seasonal behaviour of the series was attributed to the method of pressure monitoring. No electronic instrument was available to monitor barometric pressure and the series was created by entering points from the barograph charts into the computer. Over periods when the pressure was constant this was satisfactory but difficulties arose when the pressure was changing. The reading could be read from the chart to a resolution of 0.5 mb but over periods of steady changes only the beginning and end values and times were read. The intervening data points were then calculated and entered into the computer in increments of 0.25 mb. Typically, the pressure gradients resulted in each 0.25 increment lasting for 5, 8, 10 or 12 values. These 'periods' coincide with the seasonal behaviour apparent in the auto and partial autocorrelations. Ideally, no stepping of the values should be present in which case it is likely that the complex seasonal components would be missing making the modelling of the series possible and this would entail using an electronic monitoring device.

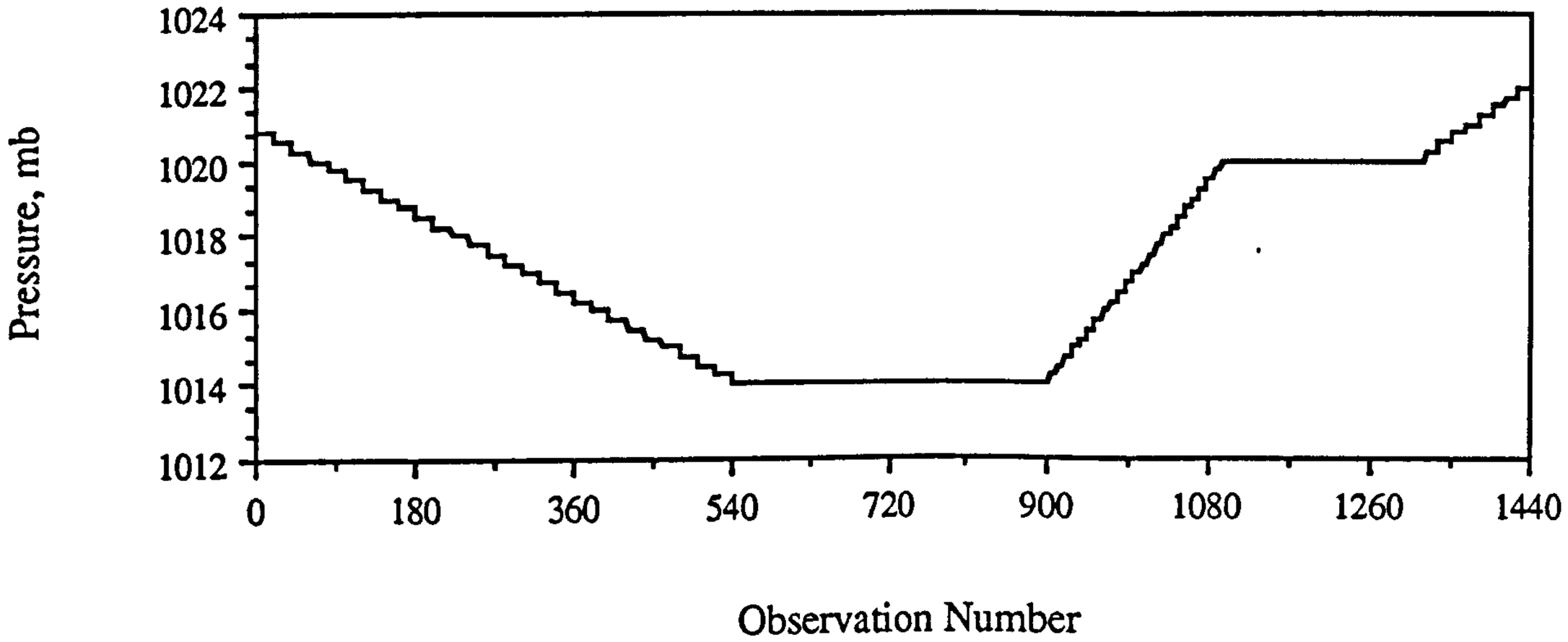


Figure 6.19 Plot of Selected Raw Barometric Pressure Series.

If only one seasonal cycle had been evident a model could have been built but SPSS-X Trends™ does not allow ARIMA modelling of series with multi-seasonality. In any case such a model, although representative of the time series, would not be representative of the underlying process that generated it.

6.5.2 Hourly Average Series of Barometric Pressure

The effect of producing hourly averages of the raw series removed all traces of seasonal behaviour. The plot of the complete series can be seen in Figure 6.21. The correlations of the series indicated that it was not random but it was impossible to attribute how much of the correlation was due to the original series and it was decided to go ahead and model the time series. The model building procedure was straightforward and the best fit to the series was an ARIMA ([1,7,8],2,2) given by:

$$(1 + \underset{(-5.0)}{0.85}B + \underset{(-3.4)}{0.16}B^{7} + \underset{(-2.1)}{0.13}B^{8})\nabla^{2} B_t = (1 + \underset{(-6.0)}{1.02}B + \underset{(-5.4)}{0.21}B^{2})\,\varepsilon_t \qquad [6.11]$$

Autocorrelations: DAT
Transformations: difference (1)

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	.000	.007	*	.000	1.000
2	.000	.007	*	.000	1.000
3	.000	.007	*	.000	1.000
4	.043	.007	.*	40.195	.000
5	.148	.007	.***	513.679	.000
6	.053	.007	.*	574.229	.000
7	.032	.007	.*	596.840	.000
8	.162	.007	.***	1162.277	.000
9	.105	.007	.**	1400.066	.000
10	.183	.007	.****	2126.410	.000
11	.027	.007	.*	2142.691	.000
12	.259	.007	.*****	3590.506	.000
13	.015	.007	*	3595.178	.000
14	.042	.007	.*	3633.585	.000
15	.157	.007	.***	4165.444	.000
16	.149	.007	.***	4645.470	.000
17	.036	.007	.*	4673.911	.000
18	.129	.007	.***	5035.956	.000
19	.019	.007	*	5043.455	.000
20	.253	.007	.*****	6426.724	.000
21	.034	.007	.*	6452.178	.000
22	.051	.007	.*	6508.368	.000
23	.013	.007	*	6511.879	.000
24	.348	.007	.*******	9131.322	.000
25	.168	.007	.***	9739.116	.000
26	.036	.007	.*	9767.569	.000
27	.086	.007	.**	9928.542	.000
28	.061	.007	.*	10008.448	.000
29	.022	.007	*	10018.508	.000
30	.252	.007	.*****	11391.716	.000
31	.018	.007	*	11398.450	.000
32	.134	.007	.***	11788.702	.000
33	.044	.007	.*	11830.805	.000
34	.051	.007	.*	11887.029	.000
35	.163	.007	.***	12460.067	.000
36	.293	.007	.******	14319.310	.000
37	.024	.007	*	14331.287	.000
38	.012	.007	*	14334.281	.000
39	.023	.007	*	14345.281	.000
40	.290	.007	.******	16167.746	.000
41	.045	.007	.*	16211.758	.000
42	.082	.007	.**	16358.535	.000
43	.029	.007	.*	16377.255	.000
44	.091	.007	.**	16557.187	.000
45	.201	.007	.****	17431.528	.000
46	.018	.007	*	17438.268	.000
47	.029	.007	.*	17456.992	.000
48	.286	.007	.******	19231.197	.000
49	.081	.007	.**	19374.548	.000
50	.162	.007	.***	19941.103	.000

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 21600 Computable first lags after differencing: 21598

Figure 6.20 Autocorrelations of Differenced Original Barometric Pressure Series.

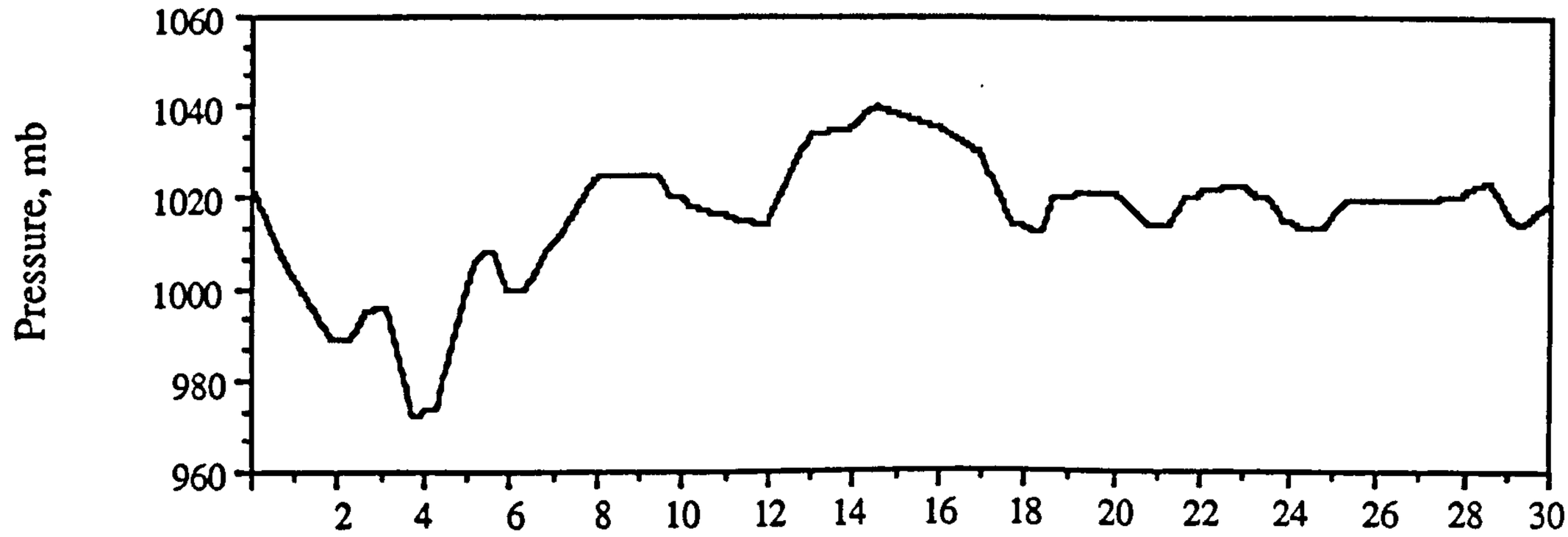


Figure 6.21 Hourly Average Plot of Barometric Pressure.

6.5.3 10-Minute Average Series of Barometric Pressure

The series for 10-minute averages of barometric pressure contained a seasonal component with a period of 12. Again this was most likely because of the format of the original series but two degrees of non-seasonal differencing were necessary to make it stationary and no seasonal differences were needed. The best fit to the series was the ARIMA $([1,5],2,2)(2,0,0)^{12}$ given by:

$$\underset{(-6.1)\ (8.8)}{(1 + 0.41B - 0.14B^5)}\underset{(15.8)\ \ (13.6)}{(1 - 0.26B^{12} - 0.23B^{24})}\ \nabla^2 B_t = \underset{(6.8)\ \ (3.1)}{(1 - 0.48B - 0.47B^2)}\ \varepsilon_t \qquad [6.12]$$

6.5.4 One Day Series of Barometric Pressure

The plot of one days worth of barometric pressure is shown in Figure 6.22. For almost half of the day the pressure is constant until it falls steadily by 5mb after which it is constant again. Seasonality was present in the original autocorrelations and the best fit model was an unusual ARIMA $(0,1,[8,20])(2,0,0)^{12}$ given by:

$$(1 - 0.72B^{12} - 0.19B^{24})\ \nabla^1 B_t = (1 - 0.24B^{8} + 0.23B^{20})\ \varepsilon_t \qquad [6.13]$$
$$\text{(18.7)} \quad \text{(5.0)} \qquad\qquad \text{(6.5)} \quad \text{(-6.3)}$$

The addition of the moving average parameters at lags 8 and 20 resulted in low value residual autocorrelations with low Box-Ljung values and high probabilities, indicating that the model fitted the series very well. However, this example demonstrates that even though an ARIMA model can be built to represent the time series, its usefulness must always be questioned inlight of knowledge that is available concerning the original time series which in this case is for one days worth of data. Allowing for the fact that the seasonal component is due to the method of data collection, the moving average parameters at lags 8 and 20 cannot be explained. Again, however, the model was built for use in determining the effect of barometric pressure on methane emission (in chapter 7) and the question of its representativeness is ignored at this point.

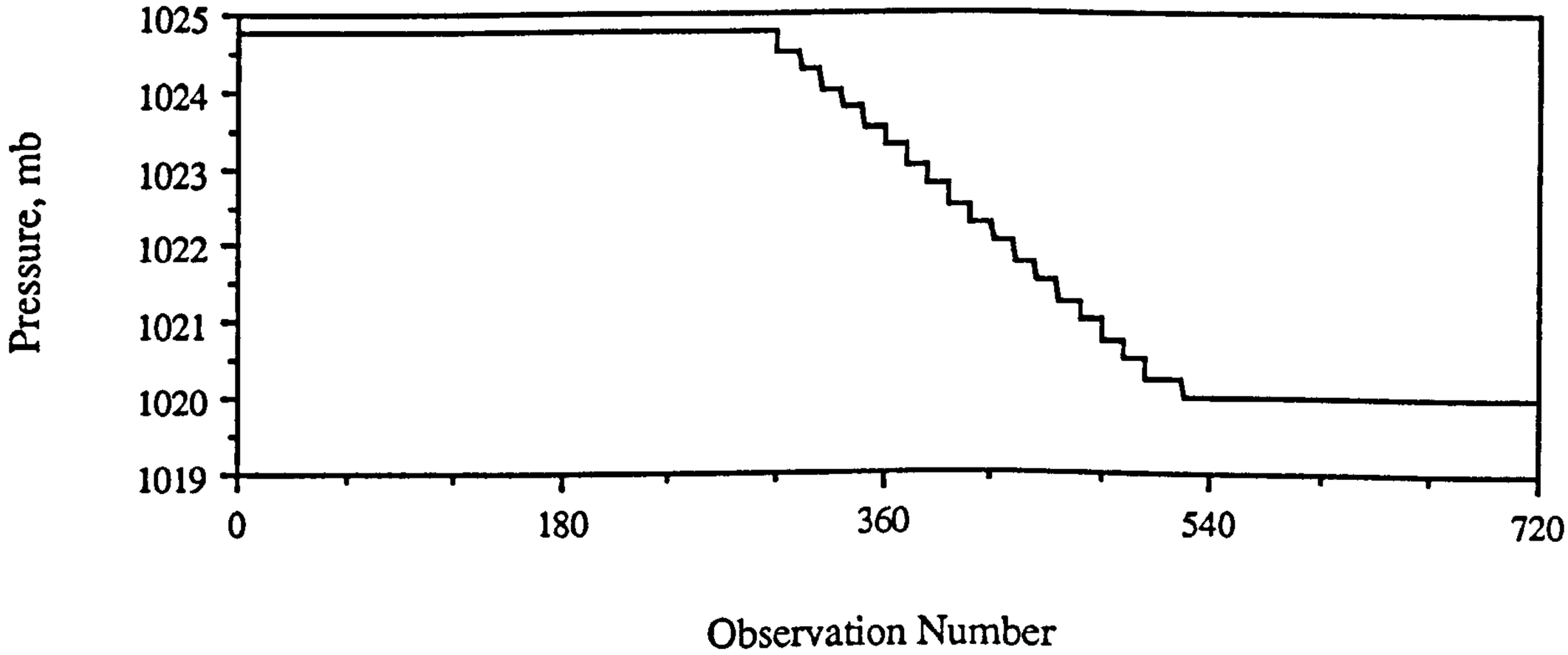


Figure 6.22 Plot of One Day Series of Barometric Pressure.

6.6 Coal Production

The univariate models for coal production were built from data relating to the average cutting speed of the coal shearer for one complete strip or partial strip. Figure 6.23, a plot of the production from two typical weekdays where each bar represents one strip or

part of a strip, its height relating to the average speed of the cut. A typical shift cut 4 to 5 strips and during the week 3 shifts per day were worked. Some production took place over the weekends but this period was usually kept free for maintenance.

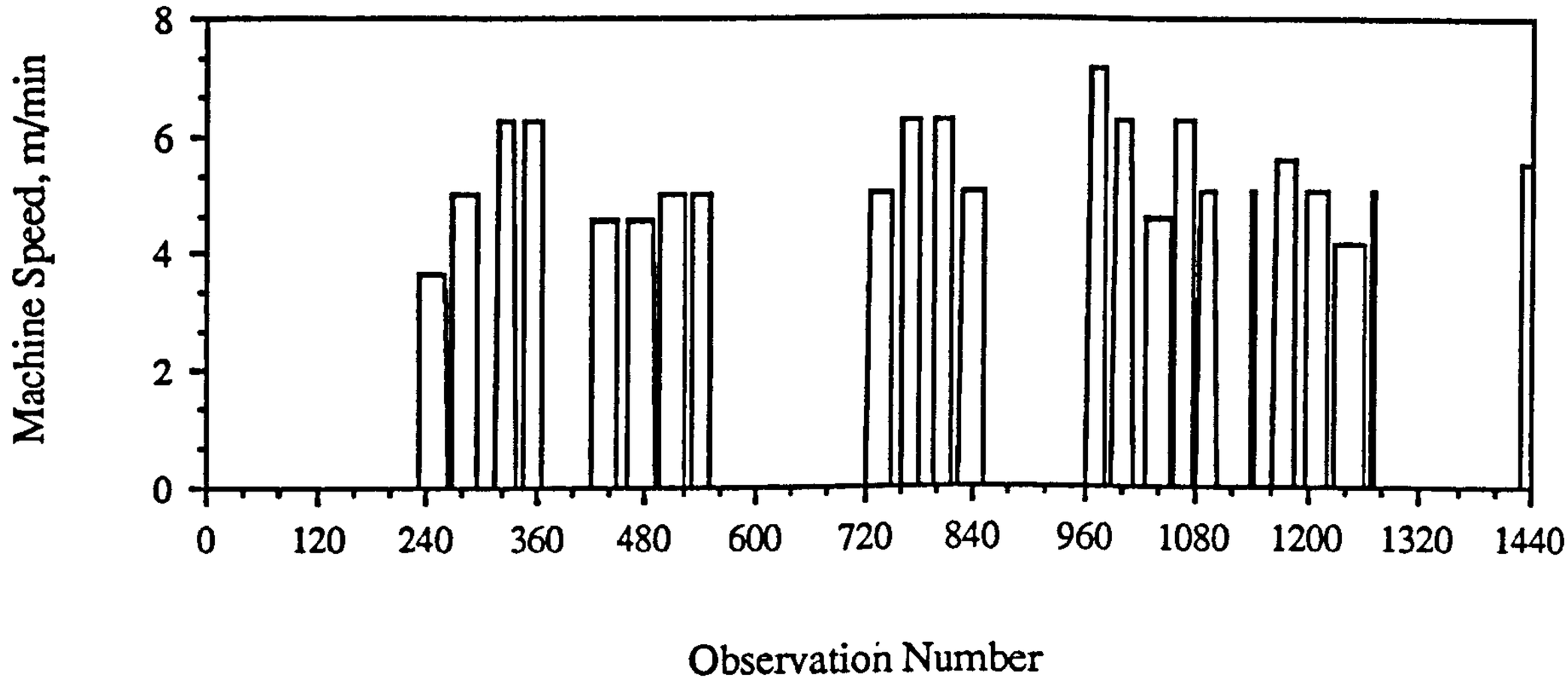


Figure 6.23 Plot of Typical Original Production Series.

Estimating the actual coal production is not an easy task due to the number of factors that contribute to the true amount mined. Although the density of coal was probably fairly constant the volume of coal per strip was not. Also, the level of strata disturbance is related to the total volume of material removed which is not the same as the volume of coal mined. Perhaps the most likely indicator of potential methane emission is face advance rate, as used by other methods for methane prediction but this is not suitable for use over a short time period. Thus, because of the need to obtain an initial production series of one observation every 2 minutes, the machine speed was selected as the potential production indicator of methane emission.

6.6.1 Original Series of Production

The plot of the auto and partial autocorrelations of the original production series were complex and much trial and error was necessary before a suitable fit could be found. There appeared to be evidence of multi-seasonality with a dominant seasonal component

of 10 and smaller periods of 5 and perhaps 20. A final estimation of the model with an SAR(2) component of period 10 was sufficient to remove the dominant seasonal component and reduce the values of the correlations at the minor periodic lags. Fitting AR parameters to lags 2, 3 and 5 also served to lower the Box-Ljung values of the residuals at low lags but they remained high at higher ones and no amount of searching and overfitting produced a model that performed better in terms of diagnostic checks. The values of the AR parameters were also small but their t-values were large confirming their validity. The final form of the model was an ARIMA $([2,3,5],1,0)(2,0,0)^{10}$ given by:

$$\underset{(5.5)}{(1 + 0.04B^2} + \underset{(-6.5)}{0.04B^3} + \underset{(-9.2)}{0.06B^5)}(1 + \underset{(-20.4)}{0.14B^{10}} - \underset{(22.2)}{0.15B^{20}})\ \nabla^1 P_t = \varepsilon_t \qquad [6.14]$$

6.6.2 Hourly Average Series of Production

The hourly average series was identified as containing a seasonal component of period 24. This was consistant with the knowledge that production periods followed a daily pattern, or rather there were definite time periods when the machine would not be cutting (shift change overs) and times when it would. This did not apply so much to the production over the weekend and it was thought that a second weekly cycle could also be present in the series, i.e., the production at a cetain hour on a Monday would be linked to the production at the same hour on all Mondays. This implies a seasonal component of period 144 and although the correlations were plotted they did not reveal its presence. A further attempt to see if a weekly cycle was present amounted to fitting seasonal parameters with a period of 144 but unfortunately, this was beyond the capabilities of SPSS-X Trends™ due to problems of mainframe computer memory availability. The plot of the complete series is shown in Figure 6.24 and illustrates the possible weekly cycle.

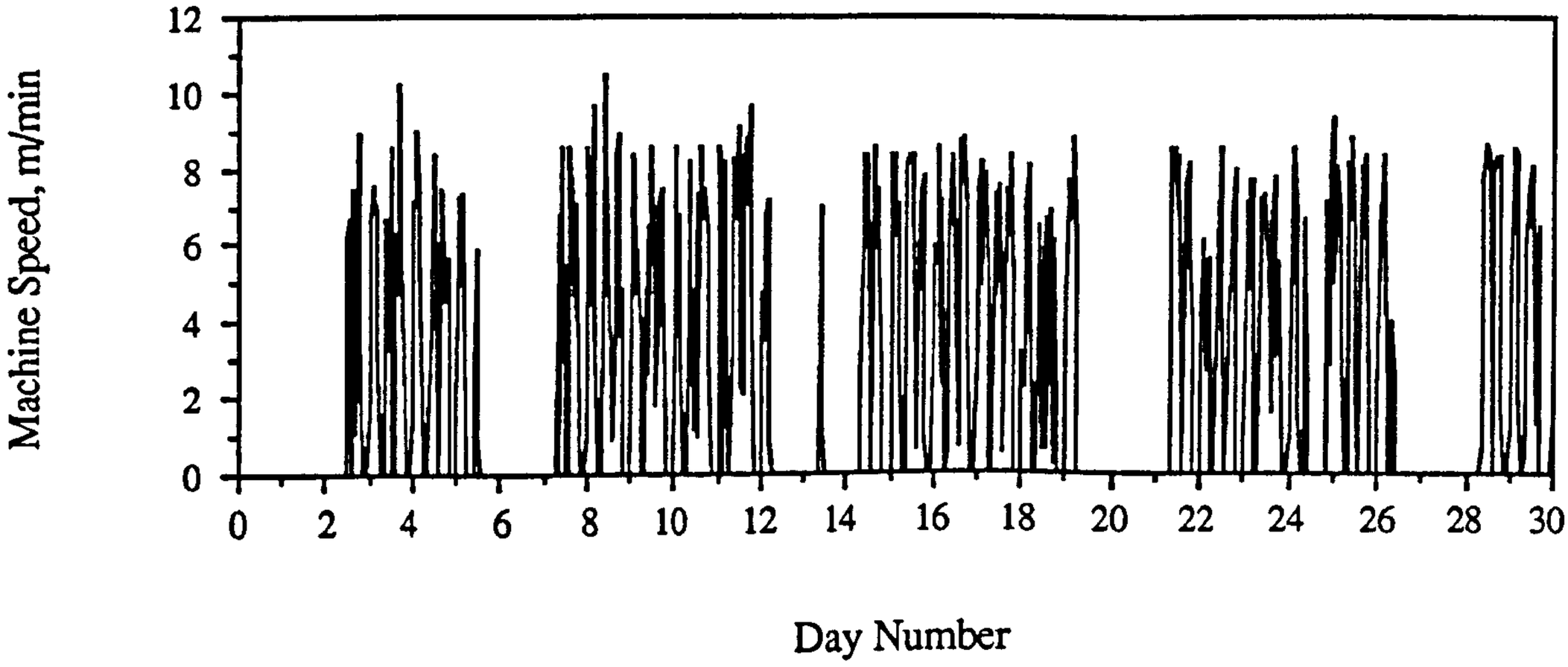


Figure 6.24 Hourly Average of Production Series.

The best fit to the series was obtained from the ARIMA $(4,1,4)(2,0,0)^{24}$ given by:

$$\underset{(5.5)}{(1 - 0.96B} \underset{(-5.4)}{+ 0.93B^2} \underset{(3.4)}{- 0.57B^3} \underset{(-5.5)}{+ 0.35B^4)(1} \underset{(7.2)}{- 0.29B^{12}} \underset{(2.9)}{- 0.12B^{24})} \nabla^1 P_t =$$

$$\underset{(8.1)}{(1 - 1.44B} \underset{(-4.7)}{+ 1.24B^2} \underset{(3.9)}{- 0.94B^3} \underset{(-2.1)}{+ 0.29B^4)} \varepsilon_t \qquad [6.15]$$

6.6.3 10-Minute Average Series of Production

No seasonal components were observed in the correlations of the 10-minute average series and the building of a representative model was quite straightforward. The best fit model was an ARIMA (4,1,[1,2,4]) given by:

$$\underset{(12.8)}{(1 - 0.75B} \underset{(-13.4)}{+ 0.75B^2} \underset{(4.4)}{- 0.12B^3} \underset{(-8.2)}{+ 0.23B^4)} \nabla^1 P_t = \underset{(12.3)}{(1 - 0.72B} \underset{(-6.7)}{+ 0.36B^2} \underset{(6.2)}{- 0.17B^4)} \varepsilon_t \qquad [6.16]$$

The AR and MA parameters fitted at lag 1 have similar values indicating that they might be superfluous but re-estimation with their removal resulted in a worse fit so they were retained. This model is similar to the one fitted to the hourly average series of methane concentration but without the seasonal component.

6.6.4 One Day Series of Production

The autocorrelations of the one days production series were small until lag 20. This was due to the length of time taken to cut one strip which on average was 40 minutes long or 20 observations. It was found that the series could be fitted by an ARIMA ([20],1,0), that is an AR parameter fitted to lag 20 only and subsequent fitting of the residual series indicated that this was appropriate. Alternatively, the series could be represented by an SARIMA $(0,1,0)(1,0,0)^{20}$ which is infact the same as the former model. In terms of simplicity the best model is the ARIMA ([20],1,0) given by:

$$(1 + 0.21B^{20})\ \nabla^1 P_t = \varepsilon_t \qquad [6.17]$$

(-5.6)

6.7 Modelling of the Methane Drainage Range Variables

The methane drainage range variables monitored were drainage methane concentration, static pressure and differential pressure. The drainage static pressure provides an indication of the suction available at the borehole which determines the quantity of gas that is drawn into the drainage range from the borehole. The volume of total gas flow in the drainage range is determined by the differential pressure. At this stage the objective was only to build time series models, leaving the finding of a possible relationship between the drainage range parameters and production to the next chapter.

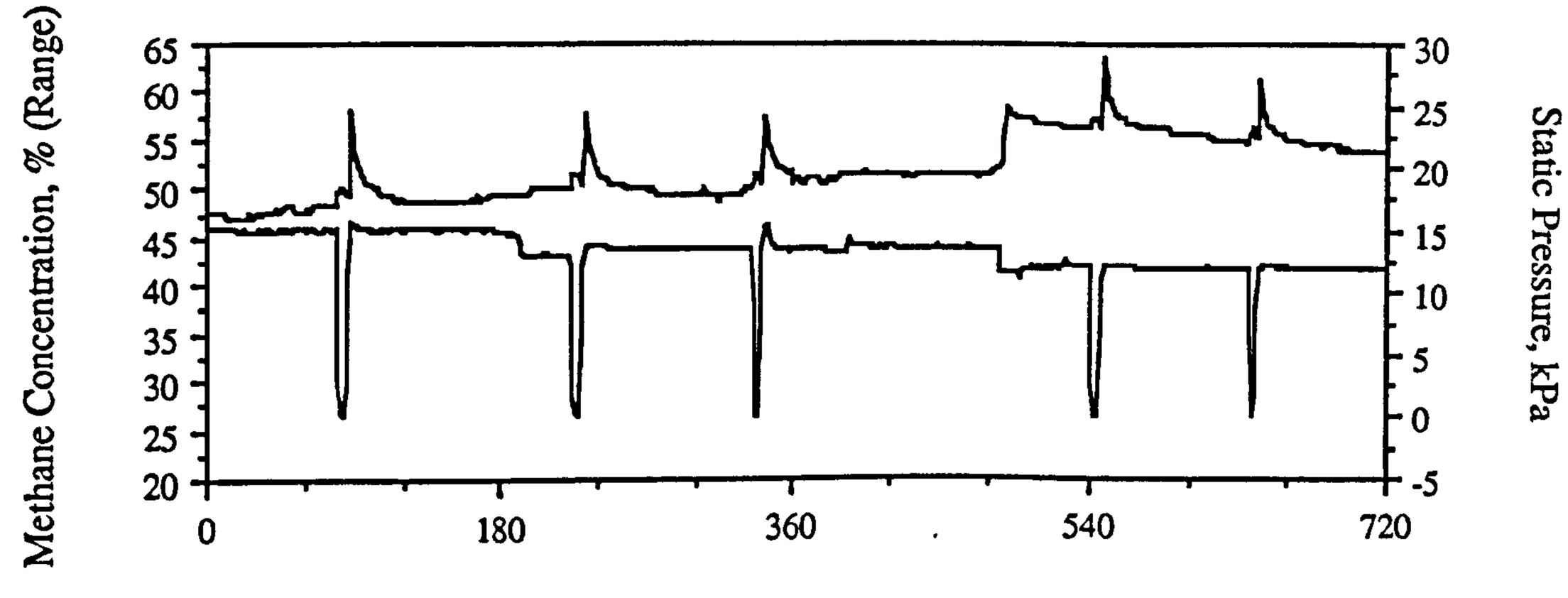


Figure 6.25 Plot of Drainage Methane Concentration and Static Pressure.

6.7.1 Original Series of Methane Drainage Range Variables

Plots of a typical portion of the original three series are shown in Figures 6.25 and 6.26. In Figure 6.25, drainage methane concentration is plotted against static pressure whilst in Figure 6.26 drainage methane concentration is plotted against differential pressure. It is seen that the behaviour of the static and differential pressure is identical since if the pumps are not working the pressure drops and there is a corresponding rise in drainage methane concentration. The pumps were usually stopped for a very short period of time for the purpose of topping up their sumps. The increase in methane purity is not immediate once the pumps have stopped and occurs after a short time lag of a few minutes. Once the pumps are restarted the purity trace falls at a slower rate than it increased. Since the three parameters behave similarly it was anticipated that their ARIMA models would have a very similar structure and this proved to be true in some cases.

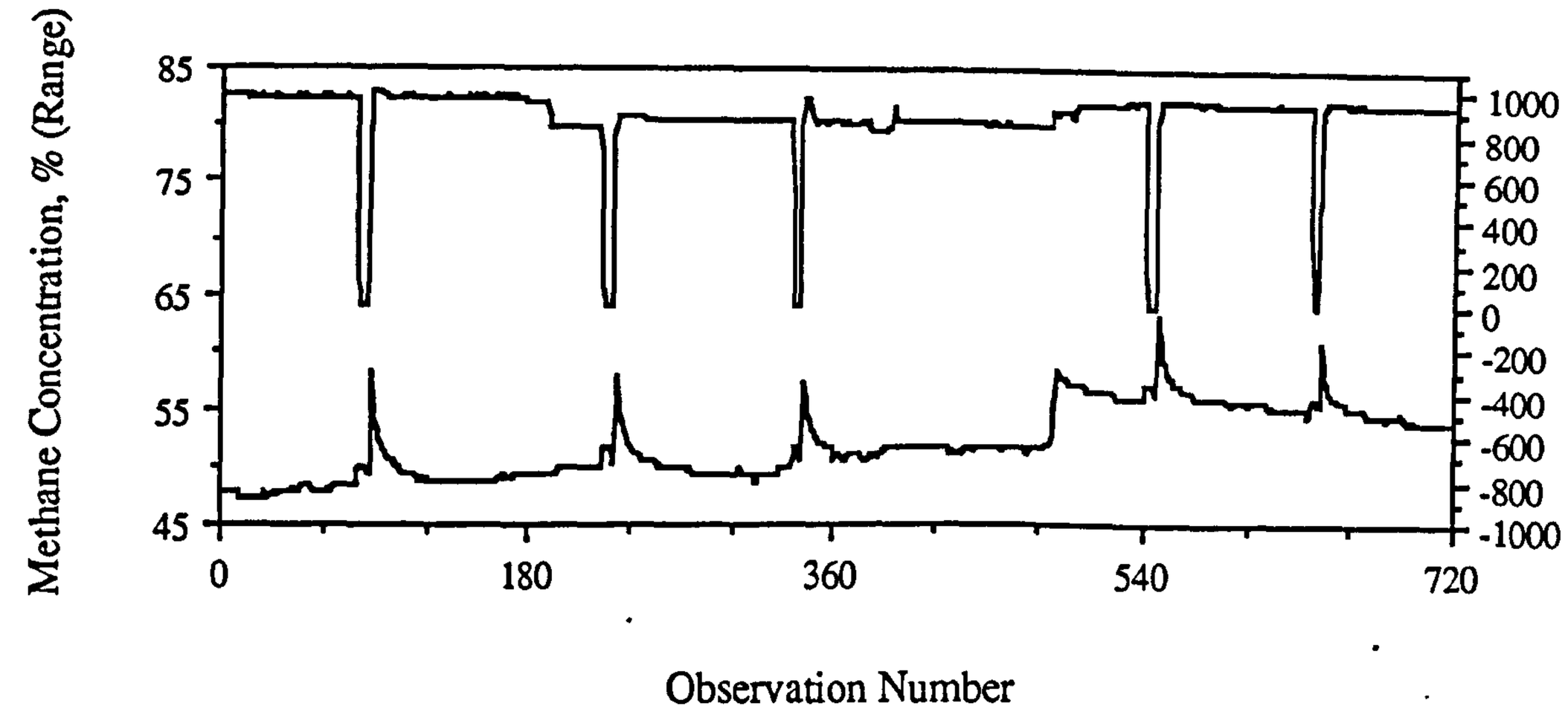


Figure 6.26 Plot of Drainage Methane Concentration and Differential Pressure.

For the original three series the best models consisted purely of autoregressive terms that contained a high number of parameters. The best models for each of the series were;

Drainage Methane Concentration: ARIMA ([2,3,4],1,0)

$$(1 + \underset{(-20.5)}{0.14}B^2 + \underset{(-13.7)}{0.09}B^3 + \underset{(-9.6)}{0.06}B^4)\nabla^1 Q_t = \varepsilon_t \qquad [6.18]$$

Static Pressure: ARIMA (7,1,0)

$$(1 - \underset{(63.3)}{0.43}B + \underset{(-22.5)}{0.17}B^2 - \underset{(3.4)}{0.03}B^3 + \underset{(-15.6)}{0.12}B^4 + \underset{(-9.6)}{0.07}B^5 + \underset{(-15.5)}{0.11}B^6 + \underset{(-4.8)}{0.03}B^7)\nabla^1 S_t = \varepsilon_t \qquad [6.19]$$

Differential Pressure: ARIMA ([1,2,4,5,6,7],1,0)

$$(1 - \underset{(33.2)}{0.23}B + \underset{(-11.8)}{0.08}B^2 + \underset{(-11.8)}{0.08}B^4 + \underset{(-16.8)}{0.12}B^5 + \underset{(-20.4)}{0.14}B^6 + \underset{(-7.7)}{0.05}B^7)\nabla^1 D_t = \varepsilon_t \qquad [6.20]$$

It is seen that most of the values of the AR parameters from lags three onwards are small but their inclusion is warranted because of their significant t-values and their ability to lower the Box-Ljung values of the residuals. All of the models proved to be good fits of their respective series. It is interesting to note that the model for methane purity did not need an AR parameter at lag 1 which is probably due to the delay caused by changes in the drainage pressure.

6.7.2 Hourly Average Series of Methane Drainage Range Variables

The models built from hourly averages of the original series did not require any seasonal components. Some seasonal effect in all three series had been anticipated by an a priori consideration of the frequent stoppage of the methane exhausters but their stoppage occurred at irregular intervals and so had no noticeable periodic affect. Since there was an absence of a seasonal affect due to the methane exhausters it was feasible that the methane purity series contained a seasonal component due to the production cycle. This was also not present and suggests that a univariate model for drainage methane concentration would be unaffected by the cyclical nature of coal production. It was also found that the three models had similar forms in that they were best represented by only AR parameters. No problems occurred in the building of the three time series average models and they are;

Average Drainage Methane Concentration: ARIMA ([1,2,11],1,0)

$$(1 - \underset{(3.7)}{0.14}B + \underset{(-3.3)}{0.13}B^2 + \underset{(-5.3)}{0.20}B^{11})\nabla^1 Q_t = \varepsilon_t \qquad [6.21]$$

Average Static Pressure: ARIMA (4,1,0)

$$(1 + \underset{(-10.8)}{0.42}B + \underset{(-6.5)}{0.27}B^2 + \underset{(-4.7)}{0.19}B^3 + \underset{(-3.9)}{0.15}B^4)\nabla^1 S_t = \varepsilon_t \qquad [6.22]$$

Average Differential Pressure: ARIMA (4,1,0)

$$(1 - 0.49B - 0.34B^2 + 0.23B^3 + 0.16B^4)\nabla^1 D_t = \varepsilon_t \qquad [6.23]$$
$$\quad (-12.6) \quad (-8.1) \quad (-5.3) \quad (-4.0)$$

The models for average static and differential pressure have the same form and very similar parameters values which is not surprising since they are functions of each other.

6.7.3 10-Minute Average Series of Methane Drainage Range Variables

The transformation of the three original series into 10-minute average values brought about a change in their model forms. Again, the two models for the static and differential pressure were very similar but curiously the model for methane purity retained the same form as the methane purity models for the original and hourly average series in that it consisted purely of AR parameters. Restimating the pressure models with purely AR parameters produced models with a similar residual variance to the final ones but the Box-Ljung values of the residuals were higher indicating their unsuitability. The final models were:

10-Minute Average Drainage Methane Concentration: ARIMA ([1,2,6],1,0)

$$(1 - 0.33B + 0.21B^2 + 0.13B^6)\nabla^1 Q_t = \varepsilon_t \qquad [6.24]$$
$$\quad (20.0) \quad (-12.8) \quad (-8.4)$$

10-Minute Average Static Pressure: ARIMA (1,1,2)

$$(1 - 0.16B)\nabla^1 S_t = (1 - 0.50B - 0.34B^2)\ \varepsilon_t \qquad [6.25]$$
$$\quad (4.2) \qquad\qquad (12.8) \quad (12.7)$$

10-Minute Average Differential Pressure: ARIMA ([2],1,2)

$$(1 - 0.10B^2)\nabla^1 D_t = (1 - 0.37B - 0.50B^2)\ \varepsilon_t \qquad [6.26]$$
$$\quad (4.8) \qquad\qquad\quad (25.6) \quad (28.0)$$

6.7.4 One Day Series of Methane Drainage Range Variables

The final models built for the drainage range parameters were those from one complete day of monitoring. Each series was 720 observations in length and it was to be expected that the models would be similar to those fitted to the original series. This turned out to be true and very good fits were obtained by only using AR parameters. The models were:

One Days Drainage Methane Concentration: ARIMA ([2,3],1,0)

$$(1 + 0.20B^2 + 0.09B^3)\nabla^1 Q_t = \varepsilon_t \qquad [6.27]$$
$$\quad (-5.2) \quad\ (-2.5)$$

One Days Static Pressure: ARIMA ([1,2,5],1,0)

$$(1 - 0.39B + 0.12B^2 + 0.29B^5)\nabla^1 S_t = \varepsilon_t \qquad [6.28]$$
$$\quad (10.4) \quad (-3.1) \quad (-8.5)$$

One Days Differential Pressure: ARIMA ([1,5,6],1,0)

$$(1 - 0.10B + 0.14B^5 + 0.23B^6)\nabla^1 D_t = \varepsilon_t \qquad [6.29]$$
$$\quad (2.71) \quad (-3.8) \quad (-5.9)$$

6.8 Conclusion

This chapter has followed the Box-Jenkins time series methodology outlined in chapter 4 to build univariate statistical time series models for various mine environmental and production variables. The purpose of this is two-fold. The first is to test them in the next chapter to investigate whether the supposed influence of certain variables on methane emission can be confirmed by time series analysis and the second is to compare forecasts from representative multivariate models for methane concentration to those obtained from univariate ones.

With the exception of the original series of barometric pressure it was possible to build time series models for all of the monitored and transformed data. The majority of the models were straightforward to build but a difficulty arose in the inability to reduce the Box-Ljung values of the residual autocorrelations at high lags for models built from very long time series. However, this was not a serious problem and was unlikely to influence the forecasting performance or useability of the models. Some of the series contained missing data that was replaced by likely values but these did not cause any problems during model building and this is one of the advantages of using long time series for this purpose.

It was seen that although a model could be built with no prior knowledge of the variable concerned it was in fact necessary to examine the form of the final model to see whether from a priori considerations it was suitable. The most obvious example of this important point is the original series for barometric pressure. A model could not be built from this data because the series displayed multi-seasonality caused by the method of data recording and transference. Models were built from the transformed series but their validity is questionable. It remains to be seen whether they could be of use in multivariate analysis.

Only the models built from methane concentration data are of use for univariate forecasts since this is the dependent variable. No importance is attached to the forecasting ability of the other models built for the independent variables and no forecasts from these will be obtained. They are not of a forecasting interest in their own right and their main purpose is to provide an indication of their effect on methane concentration.

CHAPTER SEVEN

MULTIVARIATE MODELS FOR METHANE PREDICTION

7.1 Introduction

This chapter describes investigative work to determine the correlation between methane concentration and its chosen explanatory variables, air velocity (quantity), coal production and barometric pressure. The univariate models built in the previous chapter according to the Box-Jenkins univariate model building methodology are used to attempt to derive models relating two or more variables in a multivariate analysis which follows an extension of the Box-Jenkins univariate methodology.

The purpose of this chapter is to develop a multivariate model for methane concentration that can be used for forecasting. The forecasts will then be compared to the univariate ones for methane concentration in the next chapter.

An analysis of the correlation between the methane drainage variables and production is also included in this chapter.

7.2 Multivariate Modelling

Multivariate analysis of time series is very much related to univariate analysis in that the univariate models are used to relate the behaviour of two or more time series. True multivariate analysis or modelling allows for complete causal relationships where, if it is appropriate, both series are allowed to cause changes in each other. Such modelling is very difficult to perform and fortunately for the time series modelled in this thesis the assumption that methane concentration is the only dependent variable simplifies the analysis. The relationship that is used to describe the behaviour of the dependent variable in terms of independent ones is known as the transfer function, which in this case is restricted to a single output model. In relating a dependent variable to a number of independent ones it is rarely possible to completely account for its behaviour in terms of

the independent ones. This difference or unaccountability is referred to as noise and is present because there are usually factors which contribute to the behaviour of the dependent variable that for some reason cannot be incorporated in the modelling process. Depending on the importance of the explanatory variables this difference may be very small whilst at other times it may be large and autocorrelated. The transfer function plus noise model is therefore a series of transfer functions accounting for the relationship between the dependent and independent variables and a univariate ARIMA model that describes the accountable error or noise. For the time series presented in this thesis the transfer function plus noise relationship is illustrated in Figure 7.1.

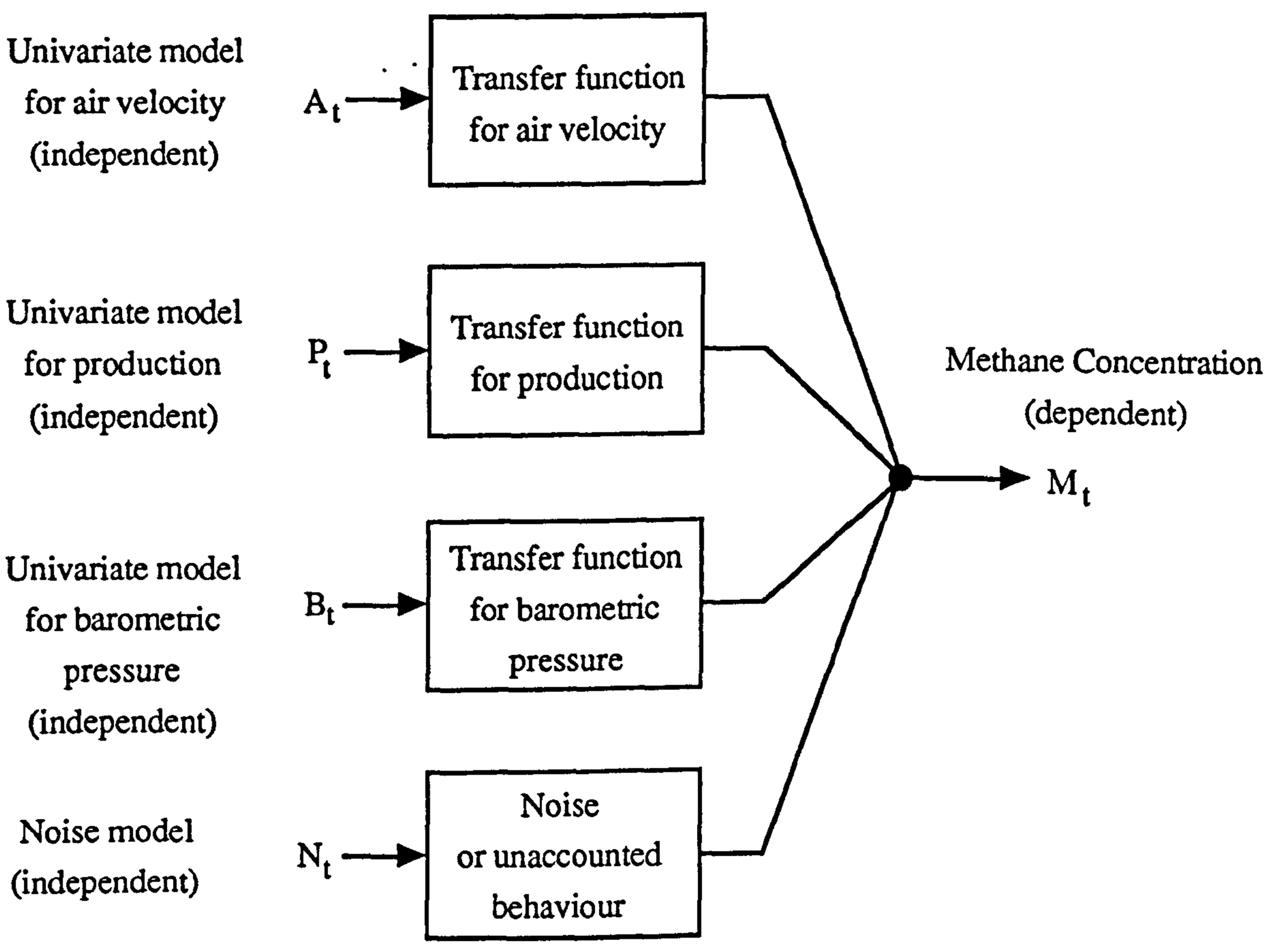


Figure 7.1 Transfer Function Plus Noise Relationship - A Conceptual Model.

As well as building the transfer function model for a forecasting purpose it is often the case that the analysis will verify or dismiss factors that were included in the original conceptual model. The conceptual model for methane emission was presented in chapter 6 and appears in a more relevant form in Figure 7.1. Ignoring the drainage range

variables, the selected independent variables deemed to influence methane concentration are air velocity, production and barometric pressure. Without prior knowledge a statistician may well decide that each variable affects methane emission in equal proportions and this, of course, would be incorrect. It has been explained earlier in chapter 5 and is worth repeating here, that air velocity (hence quantity) affects methane concentration immediately. Production contains two affecting components, the first is gas emission from the face that causes a rise in concentration in the short-term and the second is strata gas emission which occurs after an unknown time lag. The affect of barometric pressure is more difficult to quantify except but that it is well known that in exhaust ventilation systems a fall in barometric pressure causes additional methane to be released. Thus, each of the three explanatory variables possess different degrees of influence on methane emission and serve to demonstrate that an adequate knowledge of the physical system is necessary if the statistical analysis is to be carried out intelligently and effectively. Some degree of expert knowledge is therefore required to decide when the independent variables are introduced into the model. The first variable to be introduced should be the one that is believed to have the most effect which in this case would be air velocity. Other variables are then introduced one by one as appropriate.

The experience gained when building the univariate models in the previous chapter and obtaining forecasts from them proved to be invaluable during the course of the multivariate analysis presented in this chapter. In particular it was discovered that the univariate forecasts, which are to be shown and discussed in the next chapter, were not highly dependent on the complexity of the univariate model. Without this knowledge it would have been difficult to carry out the multivariate analysis according to the relationship between methane concentration and its explanatory variables presented in the conceptual model. It was in fact necessary to use this knowledge so that less complex models for the series could be obtained and enable the multivariate models to be built.

The multivariate methodology follows the three stages of model identification, estimation and diagnostic checking that was used to build the univariate models. The identification stage makes use of a procedure to estimate the cross-correlation between the residual series after univariate modelling has been performed. The form of the cross-correlations are then used to infer a possible relationship between the two residual series. After a relationship has been identified estimation of the model parameters is performed by using a non-linear regression procedure within SPSS-X™. These first two parts of the modelling process make use of procedures that are different from those used for

univariate ARIMA modelling and will be explained fully in section 7.2. Once a suitable model has been found its appropriateness is checked by examining its residual series and the significance of its parameters.

In order to infer a relationship between two or more variables use can be made of the original series without first building univariate models for them. This is quite similar to regression analysis where the influence of independent variables on a dependent variable can be found. Unfortunately it is often the case that time series contain time trends and when they are correlated it is possible that a high degree of correlation will exist between them for no reason what-so-ever. Such instances are the bane of any statistical analysis and unless scrupulous consideration of a priori knowledge is undertaken this will lead to incorrect relationships and conclusions. Spurious correlations are therefore a problem but one that can be minimized by transforming the variables using a technique known as 'pre-whitening'. As an example consider the two time series X_t and Y_t. A bivariate approach is detailed here but a complete discussion of the mathematics for complete multivariate analysis can be found in Newbold and Granger [53]. Two, separate univariate models are built for X_t and Y_t and assuming they have been correctly identified their residuals should exhibit white-noise proprieties and therefore contain no trends. This is useful because the residual series can be used as a substitute for the relevant time series and by using them to cross-correlate, the likelihood of spurious correlation is reduced. Thus pre-whitening is simply using the residual series of the univariate models for cross-correlation. Building the two univariate models for series X_t and Y_t results in models of the form:

$$\phi_x(B)\ X_t = \theta_x(B)\ \varepsilon_{xt} \qquad [7.1]$$

and

$$\phi_y(B)\ Y_t = \theta_y(B)\ \varepsilon_{yt} \qquad [7.2]$$

These can be re-arranged to give expressions for the residual series ε_{xt} and ε_{yt}, hence:

$$\varepsilon_{xt} = \frac{\phi_x(B)}{\theta_x(B)}\ X_t \qquad [7.3]$$

and

$$\varepsilon_{yt} = \frac{\phi_y(B)}{\theta_y(B)} Y_t \qquad [7.4]$$

The two residual series are cross-correlated and values are determined for both positive and negative lags. Where there are significant correlation values at negative lags this means that feedback is present or the independent series is affected by the dependent series. Clearly, this can be a problem but where the possibility of feedback is not considered possible any significant correlation values at negative lags can be conveniently ignored. Any correlation larger than the 2sd mark is considered to be significant and it is possible that a number of correlations may be so. For example, if the cross-correlation between e_{xt} and e_{yt-j} were significant this would imply that e_{yt} has a causal effect on e_{xt} with a lag of j time intervals. More importantly, such an occurrence would imply that Y_t has a causal effect on X_t with the same time interval. Thus, the correct use of the cross-correlations are paramount to identifying the relationship between X_t and Y_t.

After identifying which correlations are significant for positive lags a general unidirectional causality model between the two pre-whitened residual series can be identified. Such a model has the form:

$$\varepsilon_{xt} = \frac{\omega_2(B)}{\omega_1(B)} \varepsilon_{yt} + \frac{\omega_3(B)}{\omega_1(B)} \xi_{xt} \qquad [7.5]$$

where

ξ_{xt} = white noise error term,

$\omega_1(B)$ = a finite polynomial,

$\omega_2(B)$ and $\omega_3(B)$ = polynomials of order defined by appropriate cross-correlations.

This equation represents a transfer function for the residual series and can be translated into a model in terms of X_t and Y_t. This is done by substituting for ε_{xt} and ε_{yt} in equation 7.5 using equations 7.3 and 7.4 and obtaining the transfer function plus noise model:

$$\frac{\phi_x(B)}{\theta_x(B)} X_t = \frac{\omega_2(B)}{\omega_1(B)} \frac{\phi_y(B)}{\theta_y(B)} Y_t + \frac{\omega_3(B)}{\omega_1(B)} \xi_{xt} \qquad [7.6]$$

This is further re-arranged by multiplying both sides of the equation by $\phi_x(B)/\theta_x(B)$ to give a transfer function plus noise model for the dependent variable X_t in terms of Y_t and the error term ξ_{xt}, given by:

$$X_t = \frac{\theta_x(B)}{\phi_x(B)} \frac{\omega_2(B)}{\omega_1(B)} \frac{\phi_y(B)}{\theta_y(B)} Y_t + \frac{\theta_x(B)}{\phi_x(B)} \frac{\omega_3(B)}{\omega_1(B)} \xi_{xt} \qquad [7.7]$$

This equation looks rather complicated and consists of seven polynomials but in practice most of these will not be needed as they are dependent on the complexity of the original univariate models and the identified residual relationship.

After identifying the form of the transfer function plus noise model its parameters are estimated by the use of a non-linear regression package. Once the parameters have been estimated the model is scrutinized for inadequacies in a similar way to the univariate models. The two most important diagnostic checks are the t-values of the parameters and the examination of the residual or error series.

The theory outlined in this section will now be applied to the modelling of the environmental and production time series. Only one complete example of the transfer function building process will be described and this will be for the hourly average series to model the relationship between hourly average methane concentration, air velocity, production and barometric pressure. As in chapter 6, for the other models only items of interest are noted, in addition to the report of the final form of their models.

7.3 A Multivariate Analysis for Methane Concentration

Following the same outline as the previous chapter, the univariate models built for methane concentration, air velocity, production and barometric pressure were analysed.

The first series to be modelled were the hourly average ones. The main reason for this choice was that they were amongst the shortest in length and thus model building would take less time, at least from a computational point of view. Also, at this stage it was suspected that any models that could be produced by such a multivariate analysis of hourly average values would be the most useful, both in terms of forecasting ability and potential production control. The model building procedure is essentially the same regardless of the series time interval so only a detailed account of the procedure is given for the hourly average data. For the other series points of interest are noted but the model building procedure is only outlined.

7.3.1 A Multivariate Analysis for Hourly Average Methane Concentration

The hypotheses presented in chapter 5, section 5.7 of the effect of air velocity on methane concentration was scrutinized here. If a methane concentration trace is studied it can often be fluctuating rapidly and it is quite possible that the rate of methane emission could be relatively constant and the fluctuations in the methane concentration could be purely due to changes in airflow. For series that contain spot values such as those where data was recorded with an original time interval of 2 minutes, it is possible to convert a methane concentration value to an emission rate if the air velocity (hence quantity) was also known at that instant. However, when air velocity data is averaged it no longer becomes appropriate to use it to convert the methane concentration value to an emission rate because neither of the variables are spot values. For this reason it was decided not to use the air velocity variable in this model.

Now that air velocity had been rejected as being unsuitable the next variable to consider was production. A priori expectations are that production contains two components that cause methane emission. The first is an immediate one, whose effect should occur at lag 0 while the second should appear after an unknown number of time lags, as the strata gas release is observed. Thus, a cross-correlation between the methane concentration and production variables should contain a significant value at lag 0 and significant values at unknown higher lags. The first task is to obtain the residual series from the two models for methane concentration and production and then cross-correlate them. The identified models for hourly average methane concentration and production were:

$$\nabla^1\nabla^{24}M_t = (1 - 0.15B + 0.10B^5)(1 - 0.94B^{24})\ M\varepsilon_t \qquad [7.8]$$

and

$$\underset{(5.5)}{(1 - 0.96B} + \underset{(-5.4)}{0.93B^2} - \underset{(3.4)}{0.57B^3} + \underset{(-5.5)}{0.35B^4})(1 - \underset{(7.2)}{0.29B^{12}} - \underset{(2.9)}{0.12B^{24}})\ \nabla^1 P_t =$$

$$\underset{(8.1)}{(1 - 1.44B} + \underset{(-4.7)}{1.24B^2} - \underset{(3.9)}{0.94B^3} + \underset{(-2.1)}{0.29B^4})\ P\varepsilon_t \qquad [7.9]$$

where

$M\varepsilon_t$ = error term for the methane concentration model,

$P\varepsilon_t$ = error term for the air velocity model.

By re-arranging equations 7.8 and 7.9, expressions for the residual series were obtained, given by:

$$M\varepsilon_t = \frac{\nabla^1\nabla^{24}M_t}{(1 - 0.15B + 0.10B^5)(1 - 0.94B^{24})} \qquad [7.10]$$

and

$$P\varepsilon_t = \frac{(1 - 0.96B + 0.93B^2 - 0.57B^3 + 0.35B^4)(1 - 0.29B^{12} - 0.12B^{24})}{(1 - 1.44B + 1.24B^2 - 0.94B^3 + 0.29B^4)}\nabla^1 P_t \qquad [7.11]$$

These two residual series $M\varepsilon_t$ and $P\varepsilon_t$ are then cross-correlated, the results of which are shown in Figure 7.2.

```
Cross Correlations: RESIDUALPAV   Error for PAV from ARIMA, MOD_2 NOCON
                    RESIDUALMAV   Error for MAV from ARIMA, MOD_1 NOCON

      Cross Stand.

Lag   Corr.    Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1
                     +----+----+----+----+----+----+----+----+
  0    .571    .038                    . |*.*********
  1    .267    .038                    . |*.***
  2    .251    .038                    . |*.***
  3    .213    .038                    . |*.**
  4    .137    .038                    . |*.*
  5    .104    .038                    . |**
  6    .173    .038                    . |*.*
  7    .222    .038                    . |*.**
  8    .208    .038                    . |*.**
  9    .164    .038                    . |*.*
 10    .191    .038                    . |*.**
 11    .224    .038                    . |*.**
 12    .155    .038                    . |*.*
 13    .150    .038                    . |*.*
 14    .162    .038                    . |*.*
 15    .127    .038                    . |*.*
 16    .201    .038                    . |*.**
 17    .182    .038                    . |*.**
 18    .178    .038                    . |*.**
 19    .207    .038                    . |*.**
 20    .149    .038                    . |*.*
 21    .129    .039                    . |*.*
 22    .163    .039                    . |*.*
 23    .165    .039                    . |*.*
 24    .165    .039                    . |*.*
 25    .162    .039                    . |*.*

Plot Symbols:       Autocorrelations *     Two Standard Error Limits .
Total cases:  720     Computable 0-order correlations:  695
```

Figure 7.2 Cross-correlations of Methane Concentration and Production Residuals.

The cross-correlations show a very large value at lag 0 that confirms the hypothesis that there is an immediate effect of production on methane concentration. The value is positive and means that an increase in production causes an increase in methane concentration and a decrease in production causes a decrease in methane concentration. Lags 1 to 28 also have significant values and this can be interpreted as a gradual release of strata gas over this long period of time. To obtain a relationship between the two error terms it was necessary to group together the effect of the largest lag groups so that equal weights were placed on each of the production variables lagged 1 to 8 hours and 24 to 26 hours. The first attempt at estimation took the following form:

$$M\varepsilon_t = \alpha_1 P\varepsilon_t + \alpha_2 (0.125\ P\varepsilon_{t-1,t-8}) + \alpha_3 (0.333\ P\varepsilon_{t-24,t-26}) + e_t \qquad [7.12]$$

where

$$P\varepsilon_{t-1,t-8} = P\varepsilon_{t-1} + P\varepsilon_{t-2} + P\varepsilon_{t-3} + P\varepsilon_{t-4} + P\varepsilon_{t-5} + P\varepsilon_{t-6} + P\varepsilon_{t-7} + P\varepsilon_{t-8}$$

$$P\varepsilon_{t-24,t-26} = P\varepsilon_{t-24} + P\varepsilon_{t-25} + P\varepsilon_{t-26}$$

The estimation resulted in coefficient values that had significant t-values and the relationship was given by:

$$M\varepsilon_t = \underset{(14.5)}{0.015}\, P\varepsilon_t + \underset{(2.1)}{0.004}\,(0.125\, P\varepsilon_{t-1,t-8}) + \underset{(4.1)}{0.007}\,(0.333\, P\varepsilon_{t-24,t-26}) + e_t \qquad [7.13]$$

The next stage was then to obtain the relationship between the actual series for methane concentration and this was obtained in the following manner. The model for the hourly average production series was given by:

$$(1 - \underset{(5.5)}{0.96}B + \underset{(-5.4)}{0.93}B^2 - \underset{(3.4)}{0.57}B^3 + \underset{(-5.5)}{0.35}B^4)(1 - \underset{(7.2)}{0.29}B^{12} - \underset{(2.9)}{0.12}B^{24})\, \nabla^1 P_t =$$

$$(1 - \underset{(8.1)}{1.44}B + \underset{(-4.7)}{1.24}B^2 - \underset{(3.9)}{0.94}B^3 + \underset{(-2.1)}{0.29}B^4)\, P\varepsilon_t \qquad [7.14]$$

The model for hourly average methane concentration was given in equation 7.8 and is repeated here:

$$\nabla^1\nabla^{24}M_t = (1 - 0.15B + 0.10B^5)(1 - 0.94B^{24})\, M\varepsilon_t \qquad [7.15]$$

By substituting equation 7.13 for the error relationship into equation 7.15 for methane concentration and replacing the coefficient values as α variables, the resulting equation is:

$$\nabla^1\nabla^{24}M_t = (1 - \alpha_1 B - \alpha_2 B^5)(1 - \alpha_3 B^{24})$$
$$\text{x}\left[\alpha_1\, P\varepsilon_t + \alpha_2\,(0.125\, P\varepsilon_{t-1,t-8}) + \alpha_3\,(0.333\, P\varepsilon_{t-24,t-26}) + e_t\right] \qquad [7.16]$$

The next step is to substitute equation 7.11, the expression for $P\varepsilon_t$ in terms of P_t and replacing its coefficient values as α variables, into 7.16 to obtain an equation for methane concentration in terms of only the actual series for production, P_t, given by:

$$M_t = X.Y\ \alpha_4 P_t + X.Y\ \alpha_5\ (0.125 P_{t-1,t-8}) + X.Y\ \alpha_6\ (0.333 P_{t-24,t-26}) + X\ e_t \qquad [7.17]$$

where

$$X = (1 - \alpha_1 B - \alpha_2 B^5)(1 - \alpha_3 B^{24})$$

$$Y = \frac{(1 - \alpha_7 B - \alpha_8 B^2 - \alpha_9 B^3 - \alpha_{10} B^4)(1 - \alpha_{11} B^{12} - \alpha_{12} B^{24})}{(1 - \alpha_{13} B - \alpha_{14} B^2 - \alpha_{15} B^3 - \alpha_{16} B^4)}$$

Unfortunately this complex regression equation proved impossible for the mainframe computer to estimate. This was because of three reasons. The first is due to the large number of coefficients to be estimated (16 in total) and the second is the complex lag structure which is itself due to the complex form of the two univariate models for methane concentration and production. Theoretically these reasons should have posed no real problem for the non-linear estimation procedure but practically it was unable to perform the necessary regression and this was the third and main reason why the equation could not be regressed. In order to overcome this problem it was necessary to refer back to the univariate model building process and obtain simpler ARIMA models for methane concentration and production. A possible difficulty that might be envisaged at this point would be the question of whether it would be appropriate to use simpler univariate models for the multivariate analysis since forecasting accuracy might suffer. Chapter 8 shows that in practical terms using a simpler model does not result in a significant loss of forecasting accuracy and may even result in better parameter estimation within the multivariate analysis. The most important point to consider is the form of the cross-correlations between the two residual series. The effect of the independent variable on the dependent variable is determined in this way and if the cross-correlations of the residual series from the simpler univariate models revealed similar results as those from the best fit models then the simpler models can be used as satisfactory proxies of the best fit ones for the purpose of multivariate analysis. Also the production model is of no interest for univariate forecasts and so some accuracy of fit can be lost without causing a problem. The forecasts of methane concentration from a multivariate model in terms of production are calculated by feeding the relevant equation the known values of production which are

assumed to be at the model builders discretion.

```
Cross Correlations:   RESIDUALPAV   Error for PAV from ARIMA, MOD_4 NOCON
                      RESIDUALMAV   Error for MAV from ARIMA, MOD_3 NOCON

      Cross Stand.

Lag   Corr.    Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1
                   +----+----+----+----+----+----+----+----+
  0    .489    .038                     . |*.******** 
  1    .075    .038                     . |*.
  2   -.002    .038                     . * .
  3   -.032    .038                     .*| .
  4   -.082    .038                     **| .
  5   -.073    .038                     .*| .
  6   -.016    .038                     . * .
  7    .066    .038                     . |*.
  8    .071    .038                     . |*.
  9    .013    .038                     . * .
 10    .013    .038                     . * .
 11    .065    .038                     . |*.
 12   -.027    .038                     .*| .
 13   -.025    .038                     . * .
 14    .024    .038                     . * .
 15    .003    .038                     . * .
 16    .105    .038                     . |**
 17    .063    .038                     . |*.
 18    .017    .038                     . * .
 19    .009    .038                     . * .
 20   -.063    .038                     .*| .
 21   -.052    .039                     .*| .
 22    .009    .039                     . * .
 23    .033    .039                     . |*.
 24    .127    .039                     . |*.*
 25    .023    .039                     . * .

Plot Symbols:       Autocorrelations *     Two Standard Error Limits .
Total cases:  720     Computable 0-order correlations:  695
```

Figure 7.3 Cross-correlations of Modified Methane Concentration and Production Residuals.

After referring back to the model building attempts that were used for the production series a suitable proxy was found to be the ARIMA $(0,1,2)(1,0,0)^{24}$. The residual standard error for this model was 1.224 as compared to the originals 1.154 and amounts to an increase of 5.37% which is not very great. The residuals from this model had larger Box-Ljung values but were not too different to those from the original model. A suitable proxy for the original methane concentration model was found to be the ARIMA $([1,5],1,0)(0,1,1)^{24}$. The difference in residual standard error between this model and the original one was less than 0.1% and the residual autocorrelations were very similar to

those obtained from the original models fit. Forecasts were also obtained from this proxy model so that they could be compared in the next chapter to the univariate forecasts from the original hourly average methane concentration model.

Cross-correlating the residuals from the two proxy univariate models for methane concentration and production it was found that they did indeed have a similar form to those from the earlier production model. The new cross-correlations are shown in Figure 7.3. It is seen that there is a large positive spike at lag 0 while at other lags the values are very much smaller than previously. However, although most of the lags have values that do not appear to be significant according to the 2sd significance criteria they can still be used as a proxy and any inadequacies will become evident when the error relationship is regressed. The form of the cross-correlations led to the identification of a new relationship between the two error series where equal weights were assignend to lags 1 to 13 and lags 14 to 26 and the regression equation took the same form as the earlier one. After regression the equation was:

$$M\varepsilon_t = \underset{(11.5)}{0.014}\, P\varepsilon_t + \underset{(2.1)}{0.009}\,(0.077\, P\varepsilon_{t-1,t-13}) + \underset{(5.1)}{0.023}\,(0.077\, P\varepsilon_{t-14,t-26}) + e_t \qquad [7.18]$$

This is very similar to the expression for the previous relationship between the error series for methane concentration and production (equation 7.13) and confirms the suitability of the two proxy univariate models for methane concentration and production. An important requirement before model building can proceed is whether the error series from the regression of 7.18 demonstrates white-noise properties. If they do not then another expression between $M\varepsilon_t$ and $P\varepsilon_t$ needs to be sought. They are shown in Figure 7.4 and indicate that the white-noise condition has been satisfied, confirming 7.18 as correct. The simpler univariate models for these were:

$$\underset{(-3.6)}{(1 + 0.14B} + \underset{(-2.8)}{0.10B^5})\nabla^1\nabla^{24}M_t = \underset{(25.6)}{(1 - 0.94B^{24})}\, M\varepsilon_t \qquad [7.19]$$

and

$$(1 - 0.39B^{24})\nabla^{1}P_t = (1 - 0.48B - 0.32B^{2})\ P\varepsilon_t \qquad [7.20]$$

(11.0) (9.1) (13.6)

By substituting equation 7.18 for the error relationship into equation 7.19 for methane concentration and replacing the coefficient values as α variables, the resulting equation is:

$$(1 - \alpha_1 B - \alpha_2 B^{5})\nabla^{1}\nabla^{24}M_t = (1 - \alpha_3 B^{24})$$

$$x\left[\alpha_1\ P\varepsilon_t + \alpha_2\ (0.077\ P\varepsilon_{t-1,t-13}) + \alpha_3\ (0.077\ P\varepsilon_{t-14,t-26}) + e_t\right] \qquad [7.21]$$

Equation 7.20 is re-arranged to express $P\varepsilon_t$ in terms of P_t to give,

$$P\varepsilon_t = \frac{(1 - \alpha_6 B^{24})}{(1 - \alpha_4 B - \alpha_5 B^{2})}\nabla^{1}P_t \qquad [7.22]$$

The final step is to substitute equation 7.22 into 7.21 to obtain an equation for methane concentration in terms of the actual series for production, P_t, given by:

$$M_t = X.Y\ \alpha_7\ P_t + X.Y\ \alpha_8\ (0.077\ P_{t-1,t-13}) + X.Y\ \alpha_9\ (0.077P_{t-14,t-26}) + X\ e_t \qquad [7.23]$$

where

$$X = \frac{(1 - \alpha_3 B^{24})}{(1 - \alpha_1 B - \alpha_2 B^{5})}$$

$$Y = \frac{(1 - \alpha_6 B^{24})}{(1 - \alpha_4 B - \alpha_5 B^{2})}$$

This represents the final regression equation that was used to estimate the α coefficients. Again, problems were experienced in estimating the coefficients and the regression had to be performed in stages. This was primarily due to the limit set on mainframe computer time where the regression exceeded the maximum permissible time allowed for a single batch job. After a programme run had used all of its allowed time the final coefficient values were used as starting values for the next stage of estimation until the reduction between successive residual sums of squares was less than $1x10^{-8}$. The final estimation resulted in the achievement of very satisfactory coefficient values and the final equation was:

$$M_t = \text{Y. } 3.126\, P_t - \text{Y. } 1.021\, (0.077\, P_{t-1,t-13}) - \text{Y. } 1.11\, (0.077 P_{t-14,t-26}) + \text{X } e_t \qquad [7.24]$$

where

$$X = \frac{(1 + 0.03B^{24})}{(1 - 0.09B - 0.11B^{5})}$$

$$Y = \frac{(1 - 0.02B^{24})(1 + 0.03B^{24})}{(1 - 0.09B - 0.11B^{5})(1 - 0.12B - 0.09B^{2})}$$

The t-values of the coefficients were; $\alpha_1 = 259.7$, $\alpha_2 = 465.1$, $\alpha_3 = -15.8$, $\alpha_4 = 10.3$, $\alpha_5 = 9.8$, $\alpha_6 = 3.8$, $\alpha_7 = 17.2$, $\alpha_8 = -6.3$, $\alpha_9 = -4.1$ and these are all satisfactory. The final test of the models suitability is to see whether the residuals autocorrelations demonstrate white noise properties and they are shown in Figure 7.5. Both their autocorrelation and Box-Ljung values are all below significance and indicate that the white noise condition has been satisfied.

So far, a model has been built that is theoretically capable of predicting methane concentration values purely from a knowledge of production. It is seen that the model is comprised of four components. The first is due to the effect of coal cutting that releases methane into the air at lag 0 which with a data time interval of 1 hour primarily accounts for methane released from the actual seam being worked. However, it is possible that strata gas could also be appearing in the airstream within this hour. The second and third components have been identified as due to methane which is gradually released over a

Autocorrelations: RESIDUAL FOR R1

Lag	Auto-Corr.	Stand. Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	.005	.036	. |*.	1.346	.895
2	-.009	.036	. * .	1.548	.813
3	.010	.036	. |*.	3.554	.638
4	.011	.036	. |*.	4.251	.543
5	.002	.036	. * .	5.879	.510
6	-.004	.036	. * .	6.274	.486
7	.001	.036	. * .	7.109	.411
8	.011	.036	. |*.	7.977	.379
9	.010	.036	. |*.	8.557	.327
10	.005	.035	. * .	10.552	.476
11	-.012	.035	. * .	12.671	.501
12	.013	.035	. * .	13.824	.451
13	.005	.035	. * .	14.759	.407
14	.013	.035	. * .	16.544	.354
15	-.007	.035	. * .	18.631	.297
16	.013	.035	. | .	19.490	.258
17	.011	.035	. * .	20.301	.237
18	.021	.035	. |*.	21.127	.219
19	.010	.035	. * .	21.452	.247
20	.012	.035	. * .	21.894	.211

Figure 7.4 Residual Autocorrelations for Regression Error.

Autocorrelations: RESIDUAL FOR R2

Lag	Auto-Corr.	Stand. Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	.006	.040	. |*.	2.947	.567
2	-.010	.040	. * .	4.008	.418
3	.016	.040	. |*.	7.818	.529
4	-.013	.040	.*| .	11.522	.042
5	.001	.040	. * .	11.523	.073
6	.007	.040	. * .	11.556	.116
7	.001	.040	. * .	11.557	.172
8	.022	.040	. * .	11.857	.221
9	-.019	.040	.*| .	16.919	.076
10	-.035	.040	.*| .	17.676	.089
11	-.015	.040	. * .	17.826	.121
12	-.022	.040	. * .	18.128	.153
13	.025	.040	. * .	18.519	.184
14	.013	.040	. * .	18.626	.231
15	.007	.040	. * .	18.659	.287
16	.033	.040	. |*.	19.359	.308
17	-.010	.039	. * .	19.420	.366
18	-.053	.039	.*| .	21.256	.323
19	.020	.039	. * .	21.508	.368
20	.011	.039	. * .	21.583	.424

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 720 Computable first lags: 621

Figure 7.5 Final Model Residuals for Hourly Average Methane Concentration and Production.

long period of time from strata above and below the working horizon. The fourth component is an error term which at this stage represents unknown factors and this has been accounted for by its own univariate term so that the residuals of the whole model exhibit white-noise properties.

```
Cross Correlations:  RESIDUALBAV  Error for BAV from ARIMA, MOD_2 NOCON
                     RESIDUALMAV  Error for MAV from ARIMA, MOD_1 NOCON

 Transformations:  difference (1)
      Cross Stand.

Lag  Corr.   Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1
                   +----+----+----+----+----+----+----+----+----+
  0   .048   .038                      . |*.
  1  -.112   .038                     **| .
  2   .079   .038                     . |**
  3  -.049   .038                     .*| .
  4   .023   .038                     . * .
  5  -.028   .038                     .*| .
  6  -.052   .038                     .*| .
  7   .087   .038                     . |**
  8  -.049   .038                     .*| .
  9   .019   .038                     . * .
 10  -.048   .038                     .*| .
 11   .117   .038                     . |**
 12  -.124   .038                     **| .
 13   .055   .038                     . |*.
 14  -.030   .038                     .*| .
 15   .027   .038                     . |*.
 16   .031   .038                     . |*.
 17  -.051   .038                     .*| .
 18   .031   .038                     . |*.
 19   .035   .038                     . |*.
 20  -.046   .039                     .*| .
 21   .082   .039                     . |**
 22  -.026   .039                     .*| .
 23   .015   .039                     . * .
 24   .008   .039                     . * .
 25   .006   .039                     . * .

Plot Symbols:       Autocorrelations *     Two Standard Error Limits .
Totalcases:720 Computable 0-order correlations afterdifferencing: 694
```

Figure 7.6 Cross-correlations of Modified Methane Concentration and Barometric Pressure Residuals.

The next variable to be introduced to the model was barometric pressure. The procedure for introducing additional variables is identical to that used to obtain a model for methane concentration in terms of production in that the first thing to do is cross-correlate the error series of the univariate models for methane concentration and the additional variable, which in this case is barometric pressure. The results of these cross-correlations are

shown in Figure 7.6. It has already been decided that production would be by far the most significant indicator of methane emission and once a model for methane concentration in terms of a lesser variable has been built the two are combined and all of the models coefficients re-estimated. The form of the error term of the transfer function model is kept as that which was used to account for the noise left unexplained by the most important variable.

The residual cross-correlations of methane concentration and barometric pressure did have significant values for a large number of lags. This was considered to be suspect since the data for barometric pressure was flawed and the significant cross-correlations were most probably incorrect. However, this in itself would not have prevented a model for methane concentration in terms of barometric pressure from being built. Practically though, there would be small gain in doing this and more importantly it was thought that any model for methane concentration in terms of production and barometric pressure would have proved impossible for the computer to estimate. Furthermore, the noise term in the bivariate model for methane concentration (with production) was quite small and this indicated that there would have been little gain in having a model in terms of both production and barometric pressure.

It has been demonstrated that although the procedure for multivariate modelling is relatively straightforward, in practice a number of very real constraints are placed upon the model builder. The most significant are the limitations imposed by computer time and software which are due mainly to the number of data points to be analysed. Another difficulty is the suitability of the original data which is very important if a correct model is to be built. Thus from the conceptual model for methane concentration where three variables were intended to be used to account fully for the behaviour of methane concentration it has only been possible to use production as the indicator variable.

7.3.2 A Multivariate Analysis for 10-Minute Average Methane Concentration

The hourly average multivariate model for methane concentration contained parameters to explain the contribution of both coal front (including gas from conveyed coal) and strata gas. As the time interval between observations decreases so does the ability of the model

to represent gas that appears after a long period of time, i.e. strata gas. The 10-minute average series was chosen because it had the potential to produce useful forecasts which was due to its potential ability to represent coal front gas and possibly strata gas. The 10-minute averages of methane concentration were not spot values and therefore, it was not appropriate to use air velocity as an indicator variable. Neither was it appropriate to use barometric pressure because of the suspect nature of the barometric pressure model.

The 10-minute average series each contained 4320 data points and the first difficulty that presented itself was the likelihood that any regression equation that contained a large number of parameters would be almost impossible to estimate because of this large amount of data. The problem was overcome by selecting less complex univariate models for methane concentration and production. Two satisfactory models were the ARIMA (2,1,0) as opposed to (2,1,3) for methane concentration and the ARIMA (2,1,1) as opposed to (4,1,[1,2,4]) for production. In each case the increase in residual standard error was small and amounted to increases of 1% and 3.5% respectively. Forecasts were also obtained from the less complex model for methane concentration so they could be compared to those from the original model in the next chapter.

The residual error cross-correlations showed large significant values at lag 0 and lag 1 with smaller although still significant values at higher lags. This behaviour was thought to be mainly due to coal front gas. As the methane was released from the coal during cutting it began to appear on the methane monitor at lag 0, or rather during the first 10 minutes after release. Methane recorded at lag 1 represents that which was recorded between 10 to 20 minutes after it was emitted and was attributed to both coal front gas and gas that was released from the coal while it was on the AFC and belt conveyors. At higher lags, the effect of coal front gas lessens as seen by the smaller values that these lags take and the methane is most likely to be mainly that which is released from conveyed coal. At lags higher than 6, which represents a time span of over 60 minutes, strata gas appears (as confirmed by the residual cross-correlation of the hourly average series) and the contribution of emission sources becomes less clear.

Remembering that the simpler yet suitable univariate models were chosen so that the final regression model would not be beyond the capabilities of the computer, the relationship between $M\varepsilon_t$ and $P\varepsilon_t$ was also simplified. Accordingly, a separate term was used to describe the behaviour at lag 0 whilst the effect of lags 1 to 7 were added together but

with greater weighting attached to the production variable at lag 1. After estimation the relationship between the error series for methane concentration and lagged variables of the error series for production was:

$$M\varepsilon_t = 0.0045\ P\varepsilon_t + 0.0042\ P\varepsilon_{(lagged)} + e_t \qquad [7.25]$$

where

$P\varepsilon_{(lagged)} = 0.5P\varepsilon_{t-1} + 0.5(0.167P\varepsilon_{t-2,t-7})$

This relationship proved to be highly appropriate with each of the coefficients having significant t-values and the residuals left after estimation displaying no signs of autocorrelation.

After obtaining the new equations for the simpler univariate models for methane concentration and production these were rearranged to provide a regression equation for methane concentration in terms of lagged variables of actual production given by:

$$M_t = X.Y\ \alpha_6\ P_t + X.Y\ \alpha_7\ P_{(lagged)} + Y\ e_t \qquad [7.26]$$

where

$P_{(lagged)} = 0.3P_{t-1} + 0.167P_{t-2,t-7}$

$$X = \frac{(1 - \alpha_3 B - \alpha_4 B^2)}{(1 - \alpha_5 B)}$$

$$Y = \frac{1}{(1 - \alpha_1 B - \alpha_2 B^2)}$$

The regression of this equation proved to be time consuming and the best fit model was given by:

$$M_t = X.Y\ 0.57\ P_t + X.Y\ 0.31\ P_{(lagged)} + Y\ e_t \qquad [7.27]$$

where

$$X = \frac{(1 - 0.21B + 0.05B^2)}{(1 - 0.12B)}$$

$$Y = \frac{1}{(1 + 0.09B - 0.02B^2)}$$

The t-values of the coefficients were; $\alpha_1 = -223.6$, $\alpha_2 = 49.2$, $\alpha_3 = 11.1$, $\alpha_4 = -9.2$, $\alpha_5 = 29.8$, $\alpha_6 = 6.2$, $\alpha_7 = 5.8$ and these are all satisfactory. The residual autocorrelations are shown in Figure 7.7 and they demonstrate excellent white-noise characteristics, indicating that the estimated model and its parameters are appropriate.

A satisfactory model for 10-minute average methane concentration in terms of production has been built, albeit with some degree of difficulty. The large number of data points meant that the coefficient regression had to be carried out a number of times. Although the final coefficient values had significant t-values and the model residuals were white-noise a series of sensitivity tests were carried out to ensure that the results were correct and could be reproduced. These tests consisted of using different starting values for each of the regression stages. It was found that the same results were obtained regardless of the coefficient starting values and only the number of regression stages to obtain the final coefficient values varied.

The bivariate model for 10-minute average methane concentration consists of 2 components. The first is a term to account for the methane released within 10 minutes of the cut and contains methane released from the coal face and that released during transportation. The second component accounts for methane released between 10 and 70 minutes after cutting. This term consisted of lagged variables to account for both coal face and conveyor methane and possibly the appearance of strata gas. A greater weighting was applied to methane appearing on the monitor at a time between 10 and 20 minutes.

Autocorrelations: RESIDUALMPAV10

Lag	Auto-Corr.	Stand. Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	.001	.015	.*.	.007	.996
2	.001	.015	.*.	.013	1.000
3	.000	.015	.*.	.013	1.000
4	.000	.015	.*.	.014	1.000
5	-.001	.015	.*.	.015	1.000
6	.004	.015	.*.	.069	1.000
7	.000	.015	.*.	.070	1.000
8	.019	.015	.*.	1.623	.996
9	-.071	.015	*|.	23.612	.009
10	-.023	.015	.*.	25.924	.007
11	.000	.015	.*.	25.924	.011
12	-.002	.015	.*.	25.939	.017
13	.001	.015	.*.	25.941	.026
14	-.018	.015	.*.	27.420	.025
15	.000	.015	.*.	27.420	.037
16	-.001	.015	.*.	27.424	.052
17	.001	.015	.*.	27.543	.064
18	.010	.015	.*.	28.127	.067
19	.000	.015	.*.	28.537	.045
20	.001	.015	.*.	28.861	.051

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 4320 Computable first lags: 4314

Figure 7.7 Final Model Residual Autocorrelations for MAV10 Model.

7.3.3 A Multivariate Analysis for Original Methane Concentration

A multivariate analysis of the data recorded over the original timescale was found to be impractical and could not be carried out. This was due to restrictions imposed by the computational procedures that were used to perform the regression analysis and the availability of mainframe computer memory. Each original series consisted of 21,000 observations and this was too large for the computer. However, when the univariate models for methane concentration that were built from the complete original series are compared to those built from only one days data it is found that they are similar. For example, the best model to represent methane concentration was the ARIMA (2,1,3) but the simpler model ARIMA (1,1,0) was also found to be satisfactory. The best model for methane concentration built from only one days data was also an ARIMA (1,1,0). This is no great surprise and in effect confirms that methane concentration models built from data recorded over the original timescale and from only one days worth of data results in a satisfactory and representative model.

This is not the case though for models built for production and barometric pressure which necessarily rely on the need for a longer series so that the patterns within them are discernible. Thus the models for production from the original series and one days data are completely different and are given by the ARIMA ([2,3,5],1,0)(2,0,0)^{10} and the ARIMA ([20],1,0) respectively. This therefore, immediately limits the potential usefulness of such models. On one hand it is inconvenient to record data over a long period of time but on the other, recording over a shorter period of time results in models which are not true representations of the underlying processes. The univariate models for barometric pressure and production were not built for the purpose of prediction. They are only of use in a multivariate analysis where the objective is to identify whether they cause changes in methane concentration.

The inclusion of a term for air velocity was given thought during the analysis of one days data. Originally, it was intended to correct the methane concentration value for air quantity so that actual methane emission could be found and the univariate models for air velocity in chapter 6 were built for this purpose. With the exception of the model for hourly average air velocity none of these models contained a seasonal component and although the behaviour was far from being random it would be impossible to obtain forecasts that coincided with the actual air velocity at a time t. This means that unlike production which is or could be a known variable the air velocity is not and any model that included a term for air velocity would be worthless. To put matters into perspective, consider the following example. The cross-sectional area of the roadway at the measuring station at Thoresby 119's was 7.4 m^2 and with typical values of air velocity of 2.8 m/s and a constant methane emission rate of 9000 l/s the general body methane concentration would be 0.72%. If the air velocity had been 3.0 m/s at that instant the general body concentration would have been 0.67%. This is a difference of 7.5% and it will be seen in chapter 8 that the forecasting accuracy is within this level of magnitude. Thus, since a forecast of a value for air velocity can only be an approximation of the real value and it is unlikely that the two will coincide there is no gain in obtaining a relationship for methane emission. The best that can be done is to forecast methane concentration to a high degree of accuracy so that an indication of the most likely value is obtained.

Even so, the whole idea of the analysis was to verify if such assumptions would be correct and a subsequent cross-correlation of the residuals from the models for ones days methane concentration and air velocity produced surprising results. It was found that

there were no significant cross-correlations at all and this in fact means that the information contained within the methane concentration value contains a component due to the effect of air velocity and so no further gain is achieved by the consideration of air velocity. The same result was obtained from the residual cross-correlations from the models of the complete original series for methane concentration and air velocity. Both cross-correlation plots are given in Figures A2.1 and A2.2, Appendix 2.

```
Cross Correlations:  RESIDUALPD   Error for PD from ARIMA, MOD_2 NOCON
                     RESIDUALMD   Error for MD from ARIMA, MOD_1 NOCON

     Cross Stand.

Lag  Corr.   Err. -1  -.75  -.5 -.25   0   .25  .5   .75  1
                  +----+----+----+----+----+----+----+----+----+
  0  -.038   .037                      *|.
  1   .012   .037                      .*.
  2   .029   .037                      .|*
  3   .044   .037                      .|*
  4   .059   .037                      .|*
  5   .098   .037                      .|.*
  6   .031   .037                      .|*
  7   .037   .037                      .|*
  8   .136   .038                     . |*.*
  9   .037   .038                     . |*.
 10   .022   .038                     . * .
 11   .014   .038                     . * .
 12  -.007   .038                     . * .
 13  -.004   .038                     . * .
 14  -.025   .038                     .*| .
 15  -.041   .038                     .*| .
 16  -.009   .038                     . * .
 17  -.044   .038                     .*| .
 18   .012   .038                     . * .
 19   .001   .038                     . * .
 20  -.067   .038                     .*| .
 21  -.022   .038                     . * .
 22   .057   .038                     . |*.
 23  -.065   .038                     .*| .
 24  -.016   .038                     . * .
 25  -.043   .038                     .*| .

Plot Symbols:       Autocorrelations *     Two Standard Error Limits .
Total cases:  720     Computable 0-order correlations:  719
```

Figure 7.8 Cross-correlations of One Day Methane Concentration and Production Residuals.

The effect of barometric pressure was also investigated. The residual cross-correlations of one days methane concentration and barometric pressure did not result in any significant values. Even allowing for the fact that the model for barometric pressure was

not probably correct it is unlikely that any change in barometric pressure could effect methane emission over such a short time interval. The residual cross-correlation plot is shown in Figure A2.3, Appendix 2. Occasionally, however, very rapid falls in barometric pressure of around 30mb in 1 hour occur and these are known to cause increases in underground methane emission. In fact, collieries are normally given warnings of such pressure drops by weather centres. If such large pressure drops were to occur it is unlikely that a multivariate model with a data time interval of 2 minutes would be able to account for the drop in pressure, mainly because the lag between pressure drop and methane emission would still be large. It is possible though that when either 10-minute or hourly average values are used, a significant fall in barometric pressure and its effect on methane emission could be determined and compensated for if an appropriate model was available.

The last indicating parameter to investigate was coal production. The residual cross-correlations of methane concentration and production are shown in Figure 7.8. Only the values at lags 5 and 8 are larger than the 2sd mark while lags 3 and 4 are larger than 1sd and the values at lags 0, 1 and 2 are smaller still. The explanation for this behaviour is that once the face coal has been cut and gas begins to be released it takes between 6 and 16 minutes before it is monitored by the methane monitor. The gas that appears after a short time (between 6 and 10 minutes) can be attributed to that which is released by the cutting operation. Gas appearing later is from coal on the AFC and belt conveyor. No significant values are seen at lags greater than 8 and thus it will be expected that any model will only be able to account for methane concentration in terms of coal front gas only. The contribution of strata gas over a short time of 16 or so minutes is negligible and because of the error that is attributed to cross-correlation values at high lags it is very difficult to discern the strata gas contribution at such high lags. This explanation of the residual cross-correlation led to the identification of the following residual relationship given by:

$$M\varepsilon_t = \underset{(2.1)}{0.001\ P\varepsilon_{t-3}} + \underset{(2.8)}{0.002\ P\varepsilon_{t-5}} + \underset{(3.7)}{0.002\ P\varepsilon_{t-8}} + e_t \qquad [7.28]$$

The one day univariate models for methane concentration and production were:

$$(1 - \alpha_4 B)\nabla^1 M_t = M\varepsilon_t \qquad [7.29]$$

and

$$(1 - \alpha_5 B^{20})\nabla^1 P_t = P\varepsilon_t \qquad [7.30]$$

and after rearranging and substitution the transfer function equation for methane concentration in terms of production was:

$$M_t = \alpha_1 \frac{(1 - \alpha_4 B^{20})}{(1 - \alpha_5 B)} P_{t-3} + \alpha_2 \frac{(1 - \alpha_4 B^{20})}{(1 - \alpha_5 B)} P_{t-5}$$

$$+ \alpha_3 \frac{(1 - \alpha_4 B^{20})}{(1 - \alpha_5 B)} P_{t-8} + \frac{1}{(1 - \alpha_5 B)} e_t \qquad [7.31]$$

The above equation, although containing six coefficients that needed to be estimated, was very much simpler than the other regression equations encountered so far. This was mainly due to the simple nature of the univariate models. Unfortunately, even though a number of varied attempts were made at trying to estimate the coefficient equation 7.31, none were successful. No solution could be found where the coefficients had sensible or significant values. This was unusual as the residual cross-correlations of 7.29 and 7.30, shown in Figure 7.8 produced results that were conversant with practical knowledge of coal face gas emission.

The inability to obtain correct coefficient values was most probably due to two reasons. The first and most significant was the production series itself. It was thought that only one days production data was insufficient to allow the regression procedure to estimate sensible values. This is evident from a comparison of the univariate models for the one day series and complete original series of production. They are very much different and basically the one day model does not contain enough information about the true nature or

pattern of production. Also, it would be inappropriate to use the original more complex model to represent the one day series of production since the regression would be looking for relationships that were not evident in the one day production series itself. The second reason was attributed to the non-linear regression procedure. Normally, the procedure was not sensitive to coefficient starting values but at times the estimation resulted in either no solution or a false solution. Occasionally poor starting values caused the procedure to look for a local rather than a global solution to the coefficient estimates. This problem is exacerbated if the regression parameter i.e. the production series contains insufficient information to allow the procedure to find the right estimate. Thus, it was not possible to obtain a relationship between one days methane concentration and production.

On reflection, however, a transfer function model on the original timescale between methane and production would be of very limited use because production had been identified as only having a short-term effect on methane concentration, at the chosen time interval of one observation every 2-minutes. Furthermore, any increase in forecasting accuracy that could be achieved by such a model could easily be lost because of the transient nature of the air velocity.

7.4 A Multivariate Analysis of the Methane Drainage Variables

It was seen in chapter 6, section 6.6 that the univariate models for the methane drainage parameters all had very similar forms. With the exception of the models for 10-minute average static and differential pressures, that included MA parameters, the univariate models were best represented by AR parameters only. This similarity provides a hint as to the possible behaviour of cross-correlating the model residuals. All of the residual cross-correlation plots can be found in Appendix 2.

The residual cross-correlations between the original series of drainage concentration and static pressure showed significant values at lags 0, 2, 3, 7, 8 and 9 only (Figure A2.4). This means that a change in static pressure affects the methane concentration over a time period of up to 18 minutes. The residual cross-correlations of the one day series of drainage concentration and static pressure showed significant values at lags 0, 1, 2, 3, 6, 7, 8, 9 (Figure A2.5). A comparison between the two reveals that the signs of the lags are the same and indicate that the relationship between concentration and static pressure is

evident from consideration of only one days worth of data. A significant negative value at lag 0 means that a decrease in static pressure causes an increase in methane concentration for the first 2 minutes. Lag 2 has a positive value which can be interpreted as a slight increase in static pressure that momentarily decreases methane concentration. This could be caused by the effect of acoustic mass in the drainage system. The other significant lags are all negative and confirm that as the static pressure drop affects the whole drainage system, methane concentration continues to increase until it levels out after 18 minutes.

The residual cross-correlations between methane concentration and differential pressure for the original series (Figure A2.6) do not show significant values at any lags while the ones for the one day series showed significant lags at 0, 1, 2, 5, 6, 7, 8 (Figure A2.7). No reason could be thought of to account for this difference. However, it is appropriate that the cross-correlations of static and differential pressure with methane concentration should be similar since they are components of the total pressure generated by the methane pumps.

It was found that over the data collection period which ran from February to September 1991, the methane drainage pumps were in continual operation and only stopped for either short periods for maintenance or very rarely because of a breakdown. During a maintenance period the pumps were only stopped for a few minutes so the methane concentration was only affected for a short period of time. This consistency of operation means that there is no real need for a model for methane concentration in terms of static and differential pressure since if they are only varying moderately and known stoppages are only for a very short period of time the best forecasts of concentration will be from the concentration model itself. However, if a situation exists where the performance of the drainage pumps is not constant it may be necessary to account for rapid and significant frequent changes in system pressure.

The residual cross-correlation plots from the 10-minute average models validate the explanations given for the behaviour of those for the series recorded over the original time interval. They can be seen in Figures A2.8 and A2.9. The cross-correlations of drainage methane concentration against static pressure and drainage methane concentration against differential pressure both show significant values at lags 0, 1 and 2. This means that changes in the drainage pressure cause changes in the concentration for up to 30 minutes

since each lag represents a time interval of 10-minutes. This time is longer than that indicated by the original residual cross-correlations and is most probably due to the 10-minute averaging of the values.

The residual cross-correlations from the hourly average models, seen in Figures A2.10 and A2.11, of drainage methane concentration with static and differential pressures show significant values at lags 0 and 1. This is interpreted as a period of influence of up to 2 hours and again this is probably due to the hourly meaning of the values. Thus, the effect of meaning is to incorrectly attach importance to values in the original series at high lags which is synonymous with the original series residual cross-correlation values occurring at high lags. From practical considerations there is no reason why these values should mean anything, although when the series is hourly averaged they do. This would result in an incorrect relationship being surmized if the real meaning of the residual cross-correlations were not realized.

The drainage methane concentration is thus influenced by the drainage pressure and for the same reasons given as to why air velocity (hence quantity) would be inappropriate for inclusion in a multivariate model for the prediction of general body methane concentration so the inclusion of either static and differential pressure would also be inappropriate.

The main point of interest and that which would be most important for inclusion in a possible multivariate model for drainage methane concentration prediction was the effect of production. As a matter of course residual cross-correlations between drainage concentration and production were calculated for the 4 different model time intervals. These are shown in Figures A2.12, A2.13, A2.14 and A2.15. Even before the residual cross-correlation plots are examined the results can be anticipated. It is highly improbable that production could influence the quantity of methane appearing in the drainage range over the original monitoring time interval of 2-minutes and this is confirmed by the residual cross-correlations shown in Figure A2.9. This confirmation is reassuring in that it indicates that the two univariate models for drainage concentration and production are correct. Conversely, the residual cross-correlations between the univariate models for one days data, shown in Figure A2.10, reveal significant values at lags 2 and 3. These values are spurious and indicate that both or either of the one days data models for methane concentration and production are incorrect and indeed from the previous section it was noted that the model for production was unsuitable.

The residual cross-correlation plots from the 10-minute average models for drainage methane concentration and production, shown by Figure A2.14, do not have any values that are larger than the 2sd significance mark. Thus, it would appear that if there is to be any link between production and methane concentration it most likely occurs only after a large time lag. This assumption is verified by the hourly average residual cross-correlations obtained from the original hourly average model for drainage methane concentration and the simpler model for production, obtained in section 7.2.1 and shown by Figure A2.15. Only one value is significant at the 2sd mark and this occurs at lag 12. Many lags have values that are larger than 1sd but less than 2sd. This suggests that the effect of production on drainage methane concentration is very subtle and acts over a long period of time. This slight relationship between the residuals causes a difficulty in estimating an expression for the drainage methane concentration error in terms of the production error as a number of lags have to be considered as a group. This technique of assigning importance to a group of lags was used earlier but in this case the lags only have very small values. Lags 1, 2 and 3 all have values greater than 1sd so they were considered as a group. The selection of other lags to group was more difficult and only lag 12 was selected as it was the only lag to have a value larger than 2sd. Thus the relationship between the error series was:

$$M\varepsilon_t = \alpha_1 P\varepsilon_{t-1,t-3} + \alpha_2 P\varepsilon_{t-12} + e_t \qquad [7.32]$$

where

$P\varepsilon_{t-1,t-3} = P\varepsilon_{t-1} + P\varepsilon_{t-2} + P\varepsilon_{t-3}$

After estimation it was found that this relationship was satisfactory with both the α coefficients having significant t-values and the residuals, shown in Figure 7.9, demonstrating white-noise properties.

The hourly average univariate models for drainage methane concentration and production were:

$$(1 - \underset{(3.7)}{0.14B} + \underset{(-3.3)}{0.13B^2} + \underset{(-5.3)}{0.20B^{11}})\nabla^1 Q_t = \varepsilon_t \qquad [7.33]$$

and

$$(1 - 0.39B^{24})\nabla^{1}P_t = (1 - 0.48B - 0.32B^2)\ P\varepsilon_t \qquad [7.34]$$

(11.0) (9.1) (13.6)

After rearranging and substitution the regression equation for drainage methane concentration in terms of production was:

$$(1 - \alpha_1 B - \alpha_2 B^2 - \alpha_3 B^{11})\ Q_t = \alpha_7 \frac{(1 - \alpha_4 B^{24})}{(1 - \alpha_5 B - \alpha_6 B^2)} P_{t-1,t-3}$$

$$+ \alpha_8 \frac{(1 - \alpha_4 B^{24})}{(1 - \alpha_5 B - \alpha_6 B^2)} P_{t-12} + e_t \qquad [7.35]$$

The regression of this equation produced coefficient values that were all significant, the final model being:

$$(1 - 0.17B - 0.14B^2 - 0.02B^{11})\ Q_t = 0.84 \frac{(1 + 0.002B^{24})}{(1 - 0.01B - 0.19B^2)} P_{t-1,t-3} +$$

$$0.65 \frac{(1 + 0.002B^{24})}{(1 - 0.01B - 0.19B^2)} P_{t-12} + e_t \qquad [7.36]$$

The t-values of the coefficients were; $\alpha_1 = 8.4$, $\alpha_2 = 8.9$, $\alpha_3 = 22.3$, $\alpha_4 = -3.2$, $\alpha_5 = 2.4$, $\alpha_6 = 11.6$, $\alpha_7 = 3.5$, $\alpha_8 = 3.1$ and these are all satisfactory. The residual autocorrelations are shown in Figure 7.10 and they demonstrate satisfactory white-noise characteristics, indicating that the estimated model and its parameters are appropriate.

Autocorrelations: RESID

Lag	Corr.	Auto- Err.	Stand. -1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	-.014	.038	. * .	.147	.929
2	-.024	.038	. * .	.558	.906
3	.024	.037	.*.	.984	.912
4	-.031	.037	*|.	1.648	.895
5	-.034	.037	*|.	2.453	.874
6	.005	.037	.*.	2.472	.929
7	.008	.037	.*.	2.520	.961
8	-.003	.037	.*.	2.526	.980
9	-.010	.037	.*.	2.604	.989
10	-.068	.037	*|.	5.907	.880
11	-.026	.037	*|.	6.386	.895
12	-.051	.037	*|.	8.238	.828
13	-.068	.037	*|.	11.600	.638
14	.017	.037	.*.	11.799	.694
15	-.002	.037	.*.	11.801	.758
16	.129	.037	.|.**	23.805	.125
17	.015	.037	.*.	23.961	.156
18	-.068	.037	*|.	27.370	.096
19	-.058	.037	*|.	29.811	.073
20	.014	.037	.*.	29.959	.093

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 720 Computable first lags: 706

Figure 7.9 Residual Autocorrelations for Drainage Methane Concentration and Production Error Relationship.

Autocorrelations: RESIDUAL FOR R7

Lag	Corr.	Auto- Err.	Stand. -1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	-.002	.038	. * .	0.528	.965
2	-.005	.038	. * .	1.443	.858
3	.007	.038	. * .	2.367	.728
4	-.011	.038	.*| .	2.977	.693
5	.003	.038	. * .	3.888	.601
6	.010	.038	. |*.	4.521	.704
7	.002	.038	. * .	5.657	.862
8	-.012	.038	.*| .	7.133	.577
9	-.010	.038	.*| .	8.438	.455
10	-.007	.038	. * .	9.146	.322
11	.012	.038	. * .	10.431	.275
12	-.009	.038	. * .	14.824	.451
13	.005	.038	. * .	15.784	.387
14	-.008	.038	. * .	17.661	.246
15	.007	.038	. * .	19.115	.147
16	-.003	.038	. * .	21.233	.276
17	.001	.038	. * .	22.365	.341
18	.006	.038	. * .	23.555	.253
19	-.007	.038	. * .	24.671	.214
20	-.005	.038	. * .	25.916	.201

Figure 7.10 Final Model Residual Autocorrelations for Drainage Methane Concentration and Production Regression Error.

For completeness, residual cross-correlation plots between drainage methane concentration and barometric pressure were obtained. The residual series from the one days model of barometric pressure was not used since it is unlikely that even a very abrupt change in barometric pressure could cause either an increase or a decrease in the drainage methane concentration. The residual cross-correlation plots for drainage methane concentration with 10-minute average and hourly average barometric pressure are illustrated by Figures A2.16 and A2.17 respectively. Both show no significant correlation values at any lag. Even allowing for the fact that the two univariate barometric models are not ideal this result is not unexpected. A possible effect of a change in barometric pressure of 40mb over a period of hours is completely lost due to the transient nature of the drainage system.

A model for hourly average drainage methane concentration in terms of production has been successfully built and consists of two components. The residual cross-correlations revealed that no effect of production was evident until lag 1, or after one hour of production had elapsed. Also, the effect of production was subtle and even though the model is appropriate in terms of the diagnostic checks employed it remains to be seen whether it will be capable of forecasting to any degree of accuracy.

7.5 Conclusion

This chapter has presented the results of a multivariate analysis to determine the correlation between methane concentration and its chosen explanatory variables. From the simple conceptual model to account for methane concentration in terms of a number of explanatory variables it was found that only production could be used to build a model expressing methane concentration in terms of production. Barometric pressure and air velocity were found not to be suitable. In the case of barometric pressure, the univariate models were not representative of actual barometric pressure because of the inaccuracy of the method of monitoring. Air velocity was not included because as a spot value there was no correlation between itself and methane concentration while as an average value it would have been appropriate to convert the average value for methane concentration into a methane flowrate.

The theory for multivariate analysis allows true causal (two-way) relationships to be determined but the models were built on the assumption that the dependent variable (methane concentration) did not influence the independent variable of production and this was found to be satisfactory. The actual model building process was limited by computational difficulties but in most cases these only served to lengthen the time taken to arrive at a suitable model. Where models were not built this was because of other factors such as incorrect residual cross-correlations caused by inappropriate univariate models. The regression was found to be very dependent on the length of the time series.

Three models were built in this chapter. A model for hourly average methane concentration and production was built first, followed by a model for 10-minute average methane concentration and production. The methane drainage range parameters were also analysed and a third model was built for hourly average drainage methane concentration and production. It was not possible to build a model for either one days or the complete original series of methane concentration and production because of their length and the inadequacy of the model to represent one days production. However, it was suggested that even if such models could have been built their forecasting usefulness would not have been very great. Also, a decision not to use air velocity was taken because it cannot be exactly predicted and so a model for actual methane emission would be inappropriate. It was also seen that proxy univariate models could be substituted for the original ones to enable the multivariate models to be built.

Each multivariate model identified production at various lags as components within the model. The model for hourly average methane concentration contained parameters to account for both coal front and strata gas emission. This means that the model should be capable of producing accurate long-term i.e. up to 48 hours, forecasts. The 10-minute average model contained parameters to account mainly for coal front gas emission and so would likely be best for short-term in-shift forecasts. The model for hourly average drainage methane concentration identified a subtle relationship between itself and production and because of its time scale is better suited to long-term forecasts.

In the next chapter the forecasts obtained from both the multivariate and univariate (original and proxy) models for methane concentration and drainage methane concentration are compared and the suitability of the models to mining process control is also examined.

CHAPTER EIGHT

AN ANALYSIS OF THE UNIVARIATE AND MULTIVARIATE FORECASTS

8.1 Introduction

In the previous two chapters it was demonstrated that it was possible to build time series models for mine environmental and production data. In this chapter the usefulness of these models is discussed with particular reference to their forecasting abilities and their application to mining process control.

Although it was seen in chapters 6 and 7 that with the exception of models for barometric pressure, appropriate time series models could be built for the monitored data, their ultimate usefulness is determined by their forecasting ability. According to the diagnostic checks performed during the model building process the models proved to be good fits to the particular time series data. However, the statistical complexity of any model is of no real benefit if the model produces forecasts that are either inaccurate or useless. The simplest measure of forecasting accuracy is to compute the forecasts for the out-of-sample periods and then compare them with actual data. Other measures of forecasting accuracy involve the comparison of the mean square forecasting errors between models but this is not applicable where there is only a single forecasting model. All of the forecasts were generated within the SPSS-X Trends™ package.

8.2 Forecasting Criteria

It is not sufficient just to build models and obtain forecasts from them, some criteria needs to be set down so that the forecasts can be correctly understood. An important point to note is that accuracy of the forecasts depends on how well the original univariate model fitted the time series. This is especially important when such models are used to build a multivariate model. The forecasts lend themselves to planning and control and these three items can be defined as:

1. a **forecast** is simply a series of future values,

2. a **plan** allows action to be taken whereby the forecasts can be changed,

3. **control** involves using the forecasts or the forecasting errors to take corrective action according to certain control criteria.

The aim of this thesis is to develop models that can be used to satisfy all three of these items. The models must be capable of producing accurate forecasts so that confidence in their ability to produce them can be gained. The forecasts themselves represent the possible outcome according to changes in the models influencing variables. For application in mining this means the determination of methane concentration levels due to coal production. If the forecasts reveal that at a certain time a problem will occur then a plan can be made and further forecasts made to see whether the actions contained within the plan were sufficient to solve the problem. If a different number of scenarios are analysed then the forecasts and hence the plan can be developed into a control method that could be completely automated.

8.3 Forecasts of the Original Series for Methane Concentration

In chapter 6 two univariate models were built to represent the original series for methane concentration. They were built from 16,000 data points with a time interval between each observation of 2 minutes. The main reason for building two models was to demonstrate that the choice of the final form of a model does not greatly influence its forecasting performance so long as a minimum degree of diagnostic checks have been satisfied. Essentially this means that the fitting of a small number of parameters to a time series will reduce the residual standard error to a value that is only further reduced slightly by the addition of extra model parameters. In the case of the first model which was an ARIMA (2,1,2) and proved to be the best fit, its residual standard error was 0.0121831. The second model chosen was the very simple ARIMA (1,1,0) whose residual standard error was 0.0121997. Thus the addition of the extra parameters to produce the best fit model resulted in a 0.14% decrease in residual standard error. However, the more complicated model was deemed to be the best fit because its residuals demonstrated good white-noise properties whereas the simple model did not. From a statistical point of view the best model should be the best fit one but from a practical point of view it is interesting to

compare the actual forecasts of these models to determine how important the fit of the model is in obtaining accurate forecasts.

There are two components to the forecasts obtained from the models. The first are in-series forecasts where values within the historical period are forecasted. If the model has been correctly specified then the predicted values will tend to follow the original ones closely. The second and the most interesting forecasts are those that are made into the validation period A hint as to how well the model will forecast into the validation period is given by examining the residual autocorrelations after the model has been applied to the complete series, i.e. both the historical and validation parts. If there is only a small change in the residual autocorrelation values it is likely that the model will produce better forecasts.

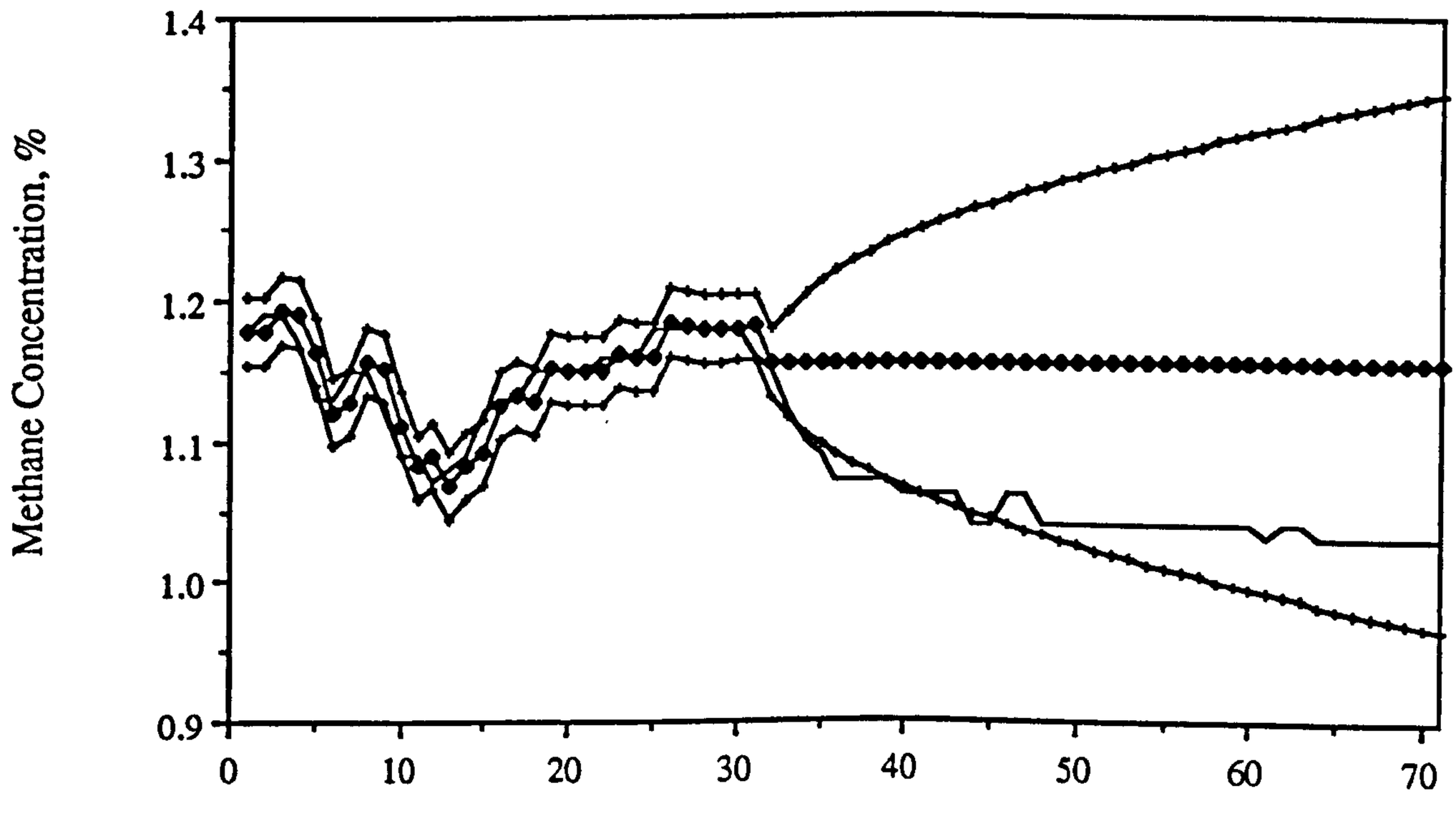


Figure 8.1 Methane Concentration Forecasts for ARIMA (2,1,2) Model.

The forecasts of both of the univariate models for the original methane concentration series are illustrated by Figures 8.1 and 8.2 respectively. Referring to Figure 8.1, the first 30 forecasts are in-series forecasts and the forecast values follow the actual methane concentrations closely indicating that the original univariate model (an ARIMA (2,1,2)) fitted the series well. The forecasts from value 31 to 71 are forecasts into the validation period. It is seen that only the forecast values 31 and 32 coincide with the actual value for methane concentration and all subsequent forecasts stay at the forecast value equal to 1.15% at data point 32. This behaviour is typical of univariate forecasts where no seasonal or periodic effects are evident and the forecasts simply converge to a horizontal unchanging level. Meanwhile the actual values for methane concentration slowly decrease before stabilizing at around 1.03%. Where no seasonal effects are evident in the series over which the model was estimated it is not possible to obtain useful forecasts and so such models are extremely limited in their forecasting ability.

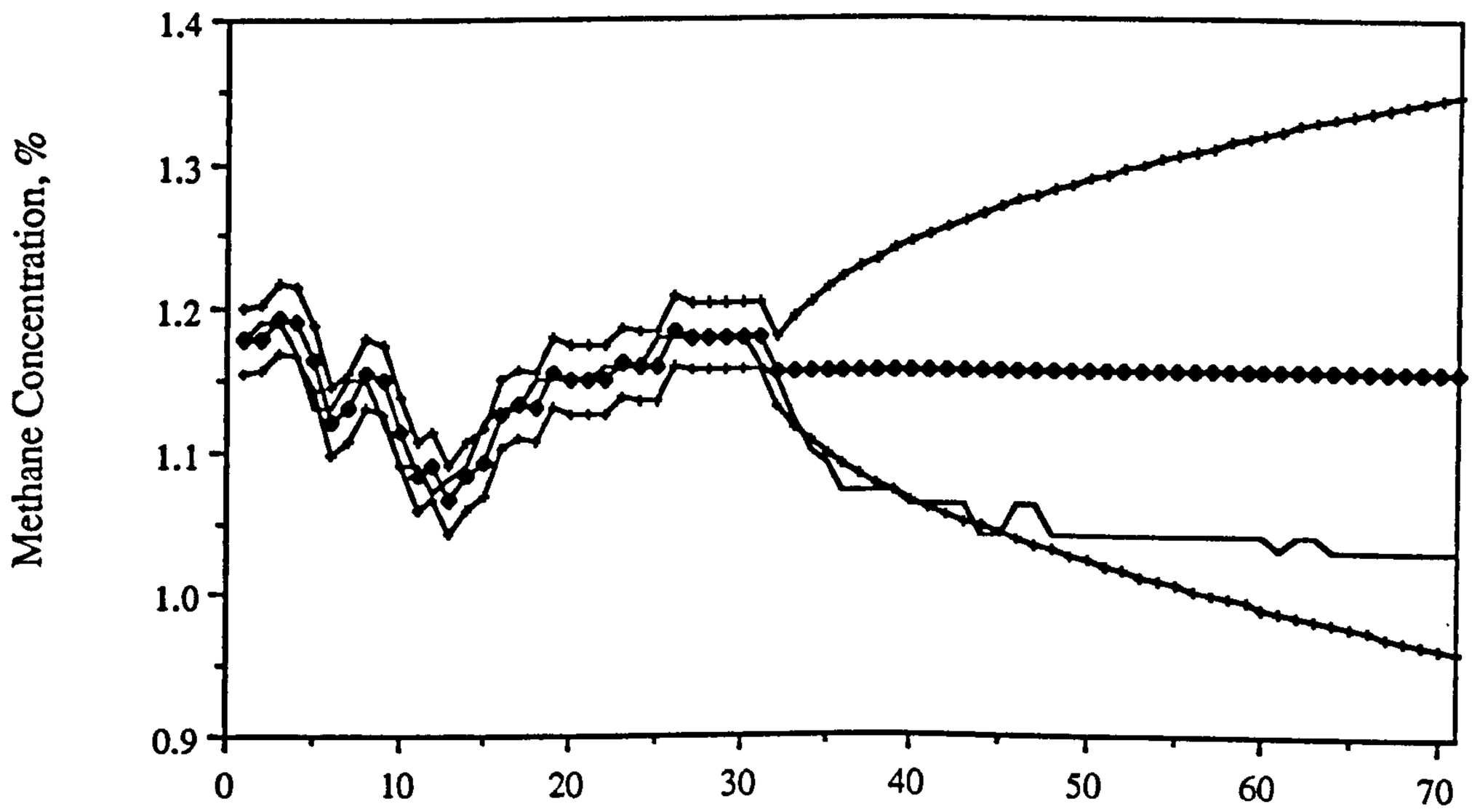


Figure 8.2 Methane Concentration Forecasts for ARIMA (1,1,0) Model.

The validation forecasts presented in Figure 8.1 are unconditional in that they are n-step ahead forecasts. This means that the model uses forecasted values to compute subsequent forecasts. The in-series forecasts are conditional in that they are all 1-step ahead forecasts, i.e. the forecast of n+1 takes account of the actual value at n and any other values necessary for the model to consider. It is possible to compute conditional validation forecasts but these would give a false impression of the models forecasting ability. Within mining it is unconditional forecasts that are of interest primarily because the objective is to forecast likely methane concentration values some time into the future. It is also seen that as the forecast period lengthens so do the forecast 95% confidence limits and although the model quickly reached a stagnant forecast of 1.15% most of the actual methane concentration values were within the confidence limits and the forecasting accuracy was around 15%.

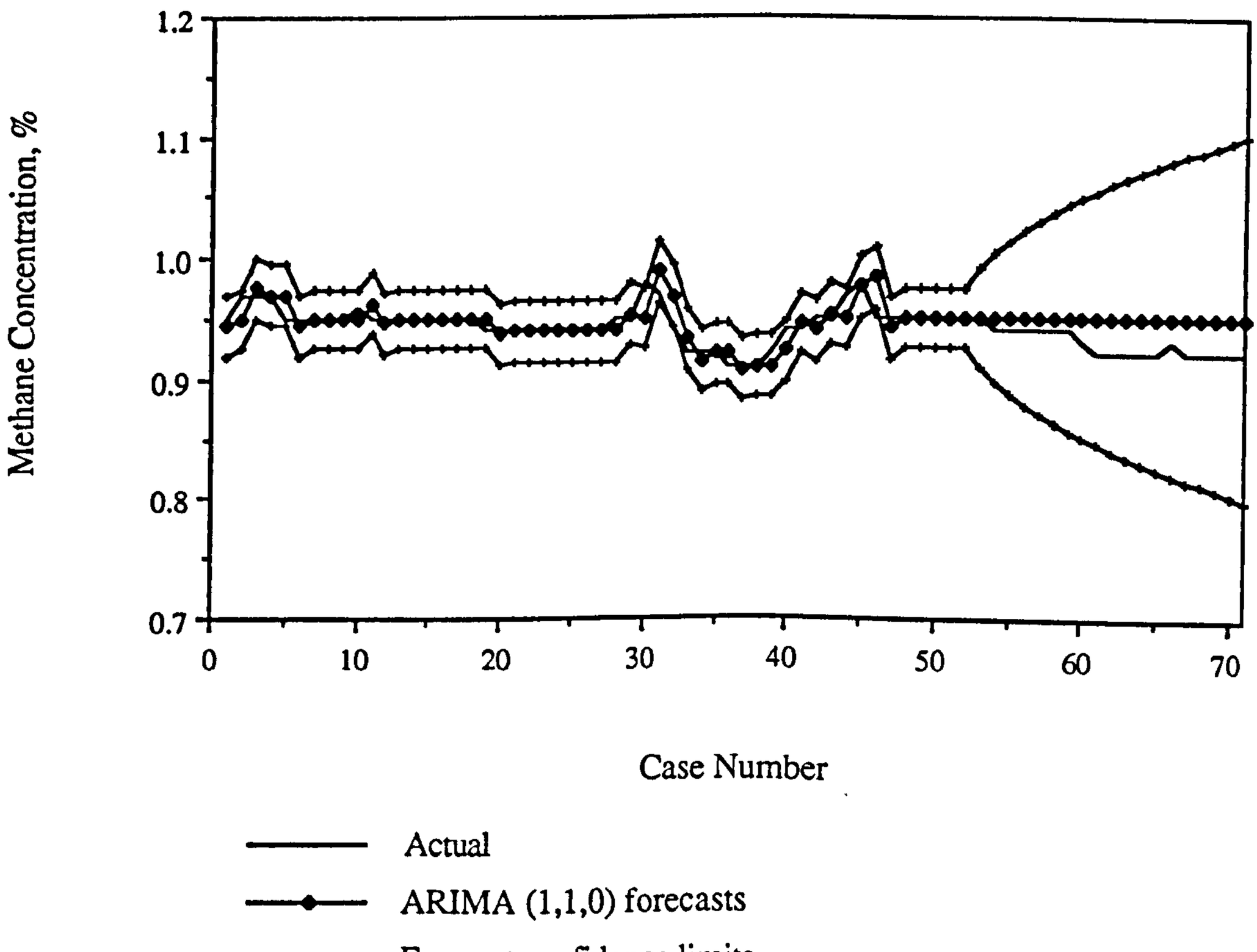


Figure 8.3 One Day Methane Concentration Forecasts for ARIMA (1,1,0) Model.

Figure 8.2 shows the forecasts obtained for the simpler univariate model for methane concentration, the ARIMA (1,1,0). To 2 decimal places all of the forecasts (both in-series and validation) are the same as for the more complex univariate model. Therefore, for this example at least, it appears that a simpler univariate model is quite suitable for obtaining representative forecasts as compared to a more complicated model. A further test of model suitability on the grounds of forecasting ability is to compare the Mean Square Forecasting Errors or MSFE's. The MSFE for the more complex model was 0.0493 while for the simpler model the MSFE was 0.0491. Although the difference between them is small it appears that are far as the MSFE is concerned the best forecasting model is the simple ARIMA (1,1,0). On the grounds of parsimony the more complicated model should be rejected primarily because if the univariate model was used to obtain a multivariate relationship models with the fewest number of parameters tend to perform best.

The actual values of methane concentration are gradually declining and in reality the values occurred over a period from 03:40 to 06:00 hours on Monday 22nd April 1991. No production took place during this time and the last full strip was cut at 03:50 hours on Saturday 20th April. The forecasts were not able to follow the actual methane concentration values when they were declining. If the forecasts are to be of any real use then they must be capable of providing an indication of when a turning point will occur. In general, unconditional forecasts from a univariate model that contains no seasonal parameters are unsuited to providing this sort of information which for application in mining is of paramount importance. Had the forecasts been obtained for a period of normal weekday production they would have not been able to predict when either an increase or a decrease in methane concentration would occur.

8.4 Forecasts of the One Day Series for Methane Concentration

The univariate model for the one day series of methane concentration with a time interval between observations of 2 minutes was a simple ARIMA (1,1,0). The in-series and validation forecasts can be seen in Figure 8.3. The in-series forecasts follow the actual series well which is itself fairly steady. The forecasts are from case 30 to 70 and are similar to the forecasts in the previous section in that they have a constant value. The forecasts are for the period 20:00 to 22:20 hours on Wednesday 10th April 1991. No production took place during this period and the last strip was cut at 17:15. The actual

methane concentration was steadily declining and production did not recommence until 00:10 the next day. Thus, in such cases the forecasts appear to be adequate and are very close to the actual values.

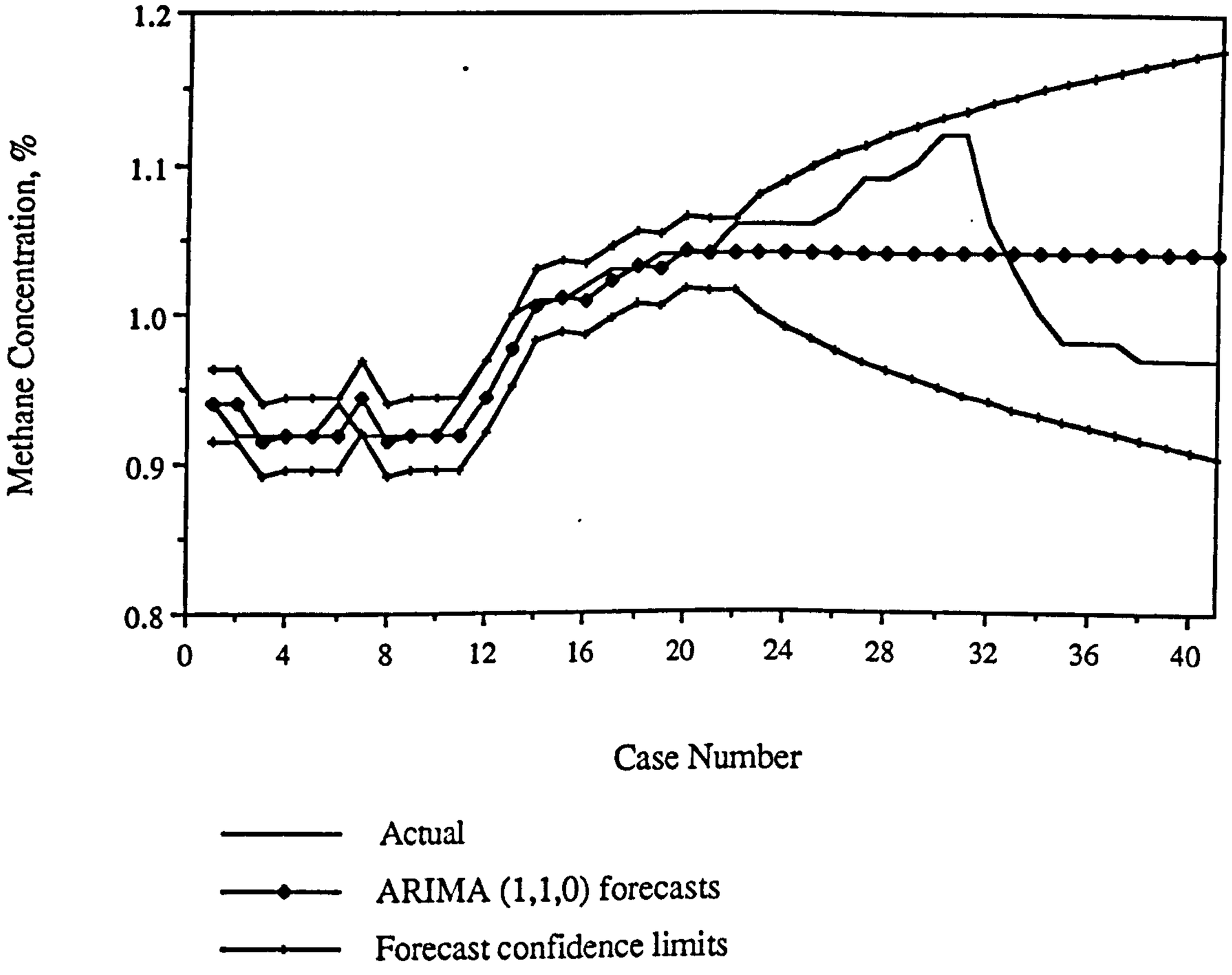


Figure 8.4 One Day Extended Methane Concentration Forecasts for ARIMA (1,1,0) Model.

To demonstrate how important it is for the model to be capable of providing forecasts that show when methane levels either decrease or increase the one days series was lengthened so that forecasts could be made in a period when the machine was cutting. The ARIMA (1,1,0) was fitted to the longer series and the residual standard error decreased by 4.5% indicating that the autoregressive parameter fitted well. The forecasting period was chosen as 12:00 to 13:20 the next day. Case 1 in Figure 8.4 corresponds to 12:00 and the in-series forecasts are for the period 12:00 to 12:40. During this time production commenced at 12:10 with the cutting of a complete strip which finished at 12:55. The

shuffle was carried out quickly and production recommenced at 13:05. Previous to the cutting of the first complete strip of the shift the machine had cut a 5m shuffle at 11:50. The actual methane concentration values are fairly steady from case 1 to 10 before starting to increase noticeably. This increase is most likely due to coal front gas appearing a few minutes after the start of the cutting at 12:10. Case 28 corresponds to 12:55 when the machine completed the first strip but the methane level continued to increase until 13:04 after which it begins to decrease. The forecasts are from case 21 onwards and immediately converge to a value of 1.04% that does not change throughout the forecasting period. Thus, the univariate forecasts are unable to follow the change in methane concentration due to the short-term effects of production released coal front gas.

Although it was not possible to build a multivariate model for one day methane concentration in terms of production it is envisaged that such a model would be able to predict more relevant values for methane concentration. Short-term forecasts from a multivariate model rely on a representative univariate model for production which over a short time period is difficult to acquire. It was seen that a simple univariate model could be used to describe the behaviour of methane concentration both over a long period of time (i.e. the complete month of original data) and over only one days data. Where production is concerned it would be inappropriate to use the complicated univariate model fitted to the complete series as a representation of the one days production since the seasonal components would not be evident.

8.5 Forecasts of the Hourly Average Series for Methane Concentration

8.5.1 Univariate Forecasts

Forecasts were obtained for both of the univariate models for hourly average methane concentration and for the multivariate model for methane concentration in terms of production. Figures 8.5, 8.6, 8.7, 8.8, 8.9 and 8.10 these respectively.

Figure 8.5 illustrates the in-series and validation forecasts of the ARIMA $(0,1,[1,5])(0,1,1)^{24}$ for hourly average methane concentration. This was the model identified as the best fit but one that could not be used to obtain a relationship between

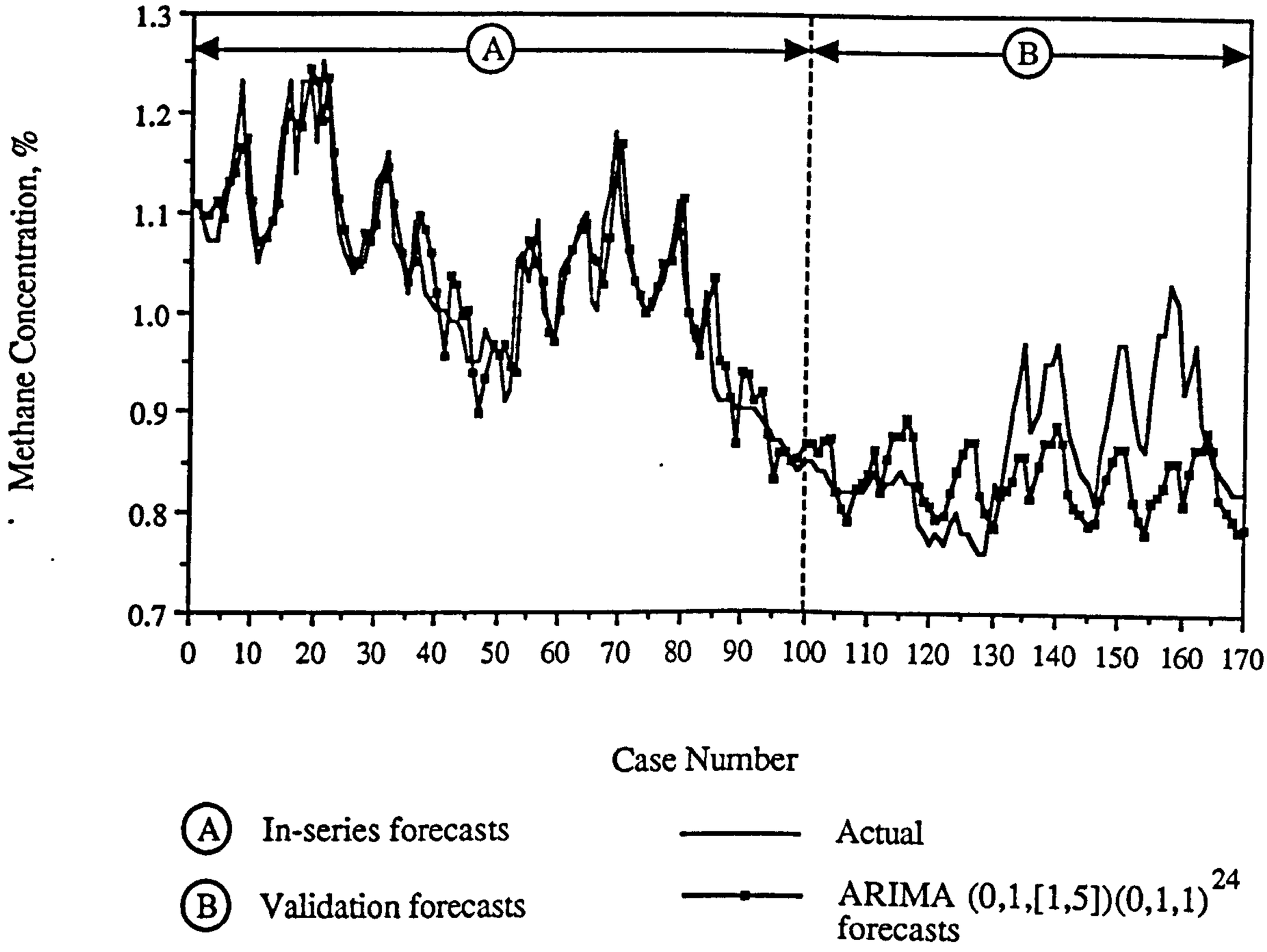


Figure 8.5 Hourly Average Methane Concentration Forecasts for ARIMA $(0,1,[1,5])(0,1,1)^{24}$ Model.

methane concentration and production. The first 100 cases (the in-series forecasts) demonstrate how well the model fits the series. Case 1 corresponds to an hourly average value at time 22:00 on the Tuesday 23rd April 1991, while case 100 occurs at time 02:00 on Sunday 28th April. Over this period of 4 days the actual methane concentration peaks a number of times and these peaks occur for each working shift. Overall the general trend is decreasing and the in-series forecasts follow the actual values closely. From case 85 (Saturday, 10:00) onwards the actual methane concentration values begin to decrease. Coal production had stopped early on in the morning and it is seen that the actual methane concentration responds to the halt of production quickly. This behaviour was due to one main reason and was explained by examining the lag structure of the univariate model. Aside from the 24 hour seasonal component the non-seasonal component consisted of MA parameters at lags 1 and 5. This means that after a period of 5 hours of no production the value for methane concentration rapidly falls, and it is likely that the effect of producing hourly averages exacerbated this. However, the main concern is whether

the univariate model can produce accurate forecasts, but it is still important to be able to account for actual behaviour by examining the structure of the univariate model.

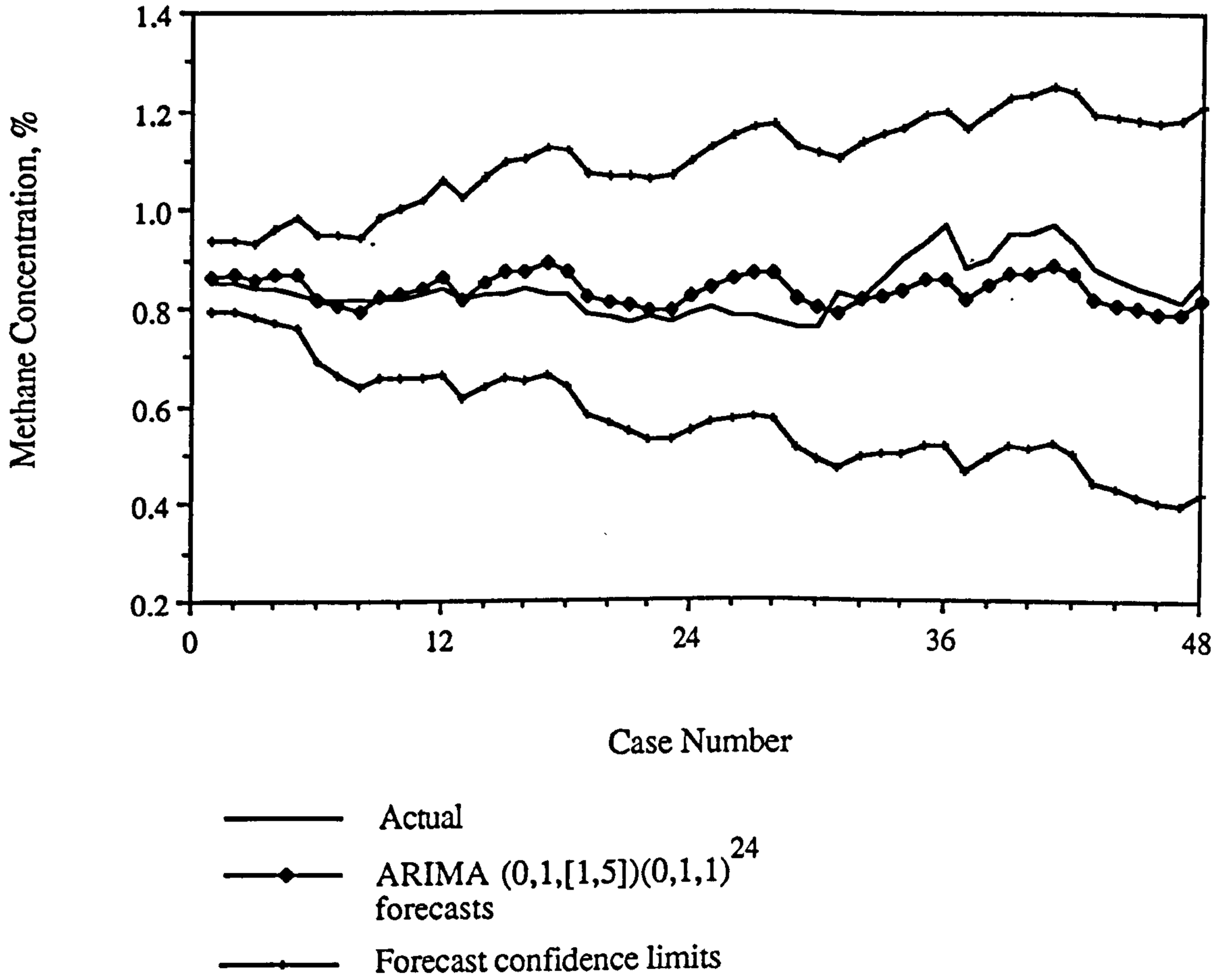


Figure 8.6 Validation Forecasts for ARIMA $(0,1,[1,5])(0,1,1)^{24}$ Model.

The validation forecasts (from case 100 onwards) are better illustrated by Figure 8.6. This figure shows the forecasts and their 95% confidence limits for a period of 48 hours. It is seen that they follow the actual trace quite well but are generally above the actual values for the first 30 forecast cases and are then below the actual value. In fact they appear to move slightly upwards and downwards about a steady level. The forecasts beyond 48 hours (case 150 onwards) are seen in Figure 8.5 and they do not follow the actual trace well. From case 130 onwards which corresponds to 08:00 on Monday 29th April the actual methane concentration increases due to the resuming of production. The forecasts however, do not respond and will continue to oscillate until they stabilize at a steady value, regardless of the true amount of methane emission. Thus, although the

initial forecasts appear to agree well with the actual values of methane concentration they cannot be used to give accurate predictions for longer than 24 hours. Even so, the overall accuracy of the univariate forecasts is approximately 10% and this is acceptable considering that factors such as air quantity and barometric pressure where large deviations from normal may well have caused the methane concentration to change unduly. A further inadequacy of the univariate model is its inability to account for the long-term contribution of strata gas emission. The univariate model does not respond to the contribution of strata gas which tends to be overshadowed by the effect of coal front gas.

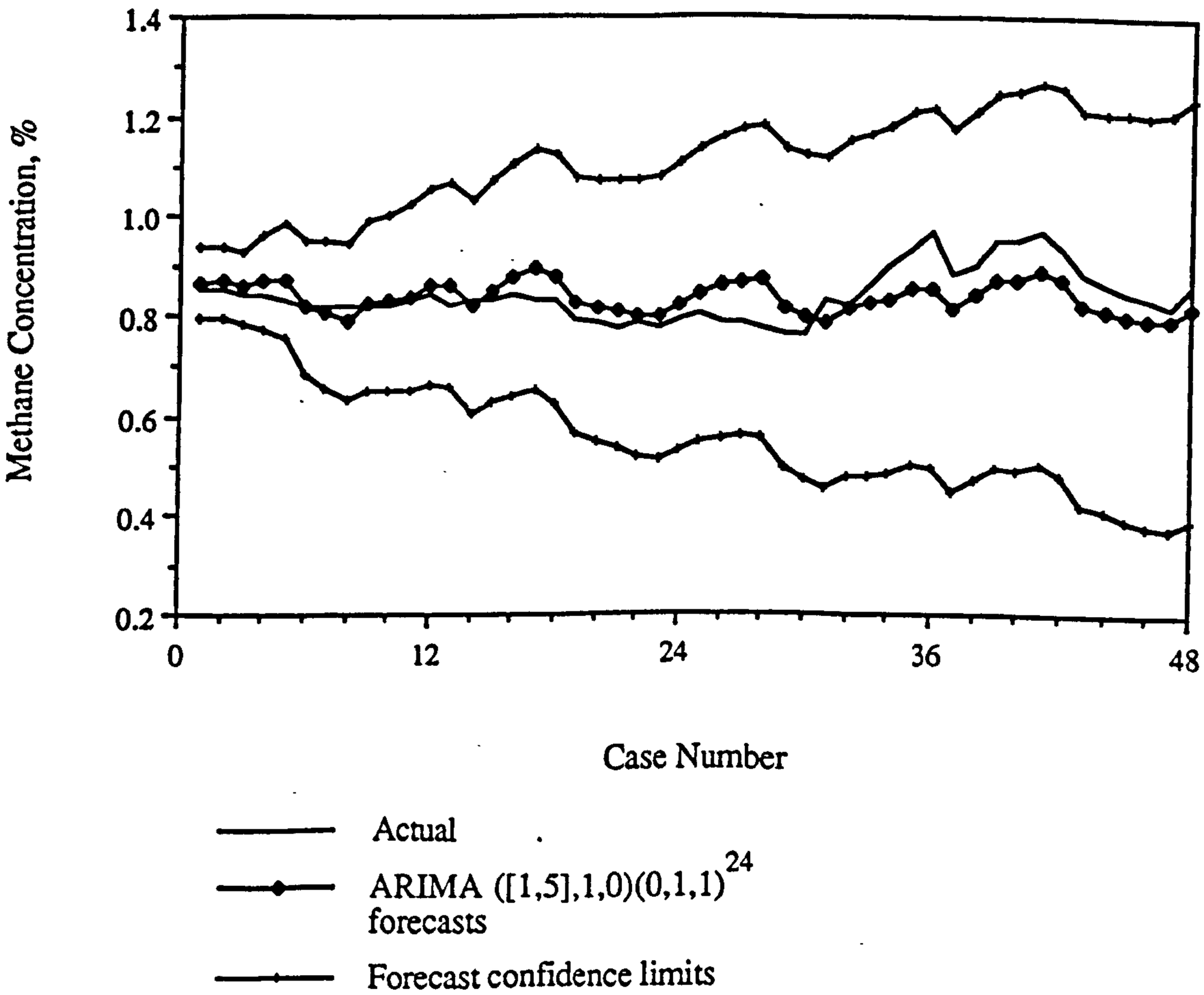


Figure 8.7 Validation Forecasts for ARIMA ([1,5],1,0)$(0,1,1)^{24}$ Model.

The validation forecasts obtained from the ARIMA ([1,5],1,0)$(0,1,1)^{24}$ are shown in Figure 8.7 and they are almost identical to those from the original model for hourly average methane concentration, as illustrated by Figure 8.6. The simpler model was

chosen so that a model for methane concentration in terms of production could be built and both the in-series and validation forecasts confirm its suitability. An examination of the two models MSFE's reveals values of 0.0050 and 0.0052 for the original and simpler models respectively. This increase in MSFE is small and indicates that the simpler model for hourly average methane concentration was suitable for use in the multivariate analysis that was performed in the previous chapter. A comparison of the validation forecasts from the two models can be seen in Figure 8.8.

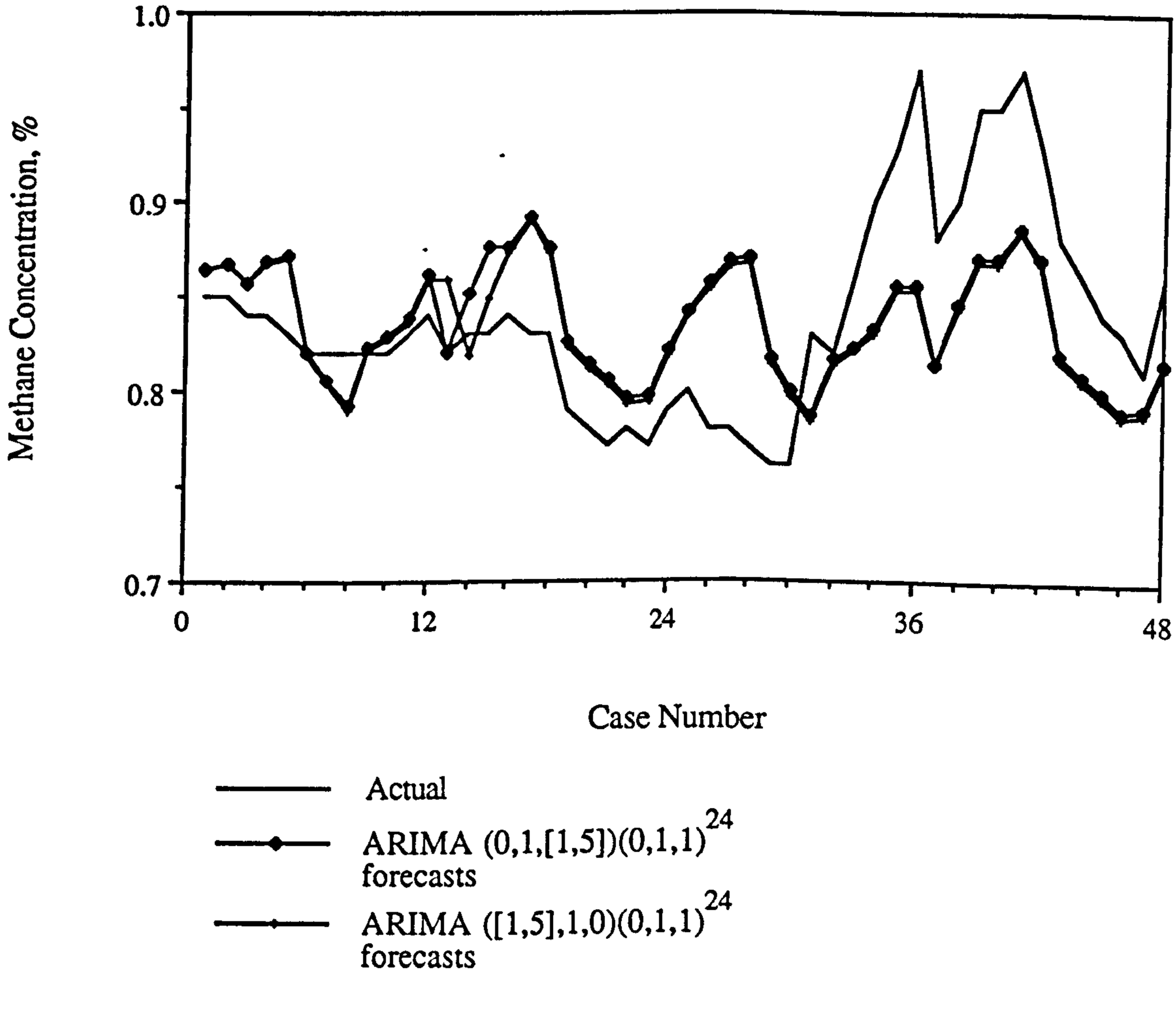


Figure 8.8 Comparison of Validation Forecasts for ARIMA $([1,5],1,0)(0,1,1)^{24}$ and ARIMA $(0,1,[1,5])(0,1,1)^{24}$ Models.

8.5.2 Multivariate Forecasts

The multivariate model for hourly average methane concentration was only able to take account of one influencing variable and that was coal production. The building of the model was carried out with adherence to strict diagnostic criteria and the model should, in theory, be capable of better forecasts. It was seen that the multivariate model contained components to account for both coal front gas (including conveyor gas) and strata gas. The forecasts from the model are shown by Figure 8.9 and it can be immediately seen that they are an improvement on the univariate ones. The in-series forecasts are similar to the ones obtained by the univariate model and this is expected. The validation forecasts can be more clearly seen in Figure 8.10. To obtain forecasts it was necessary to supply the model with values for production and so the model was supplied with the actual production that took place so that the forecasts could be compared with the actual values for methane concentration.

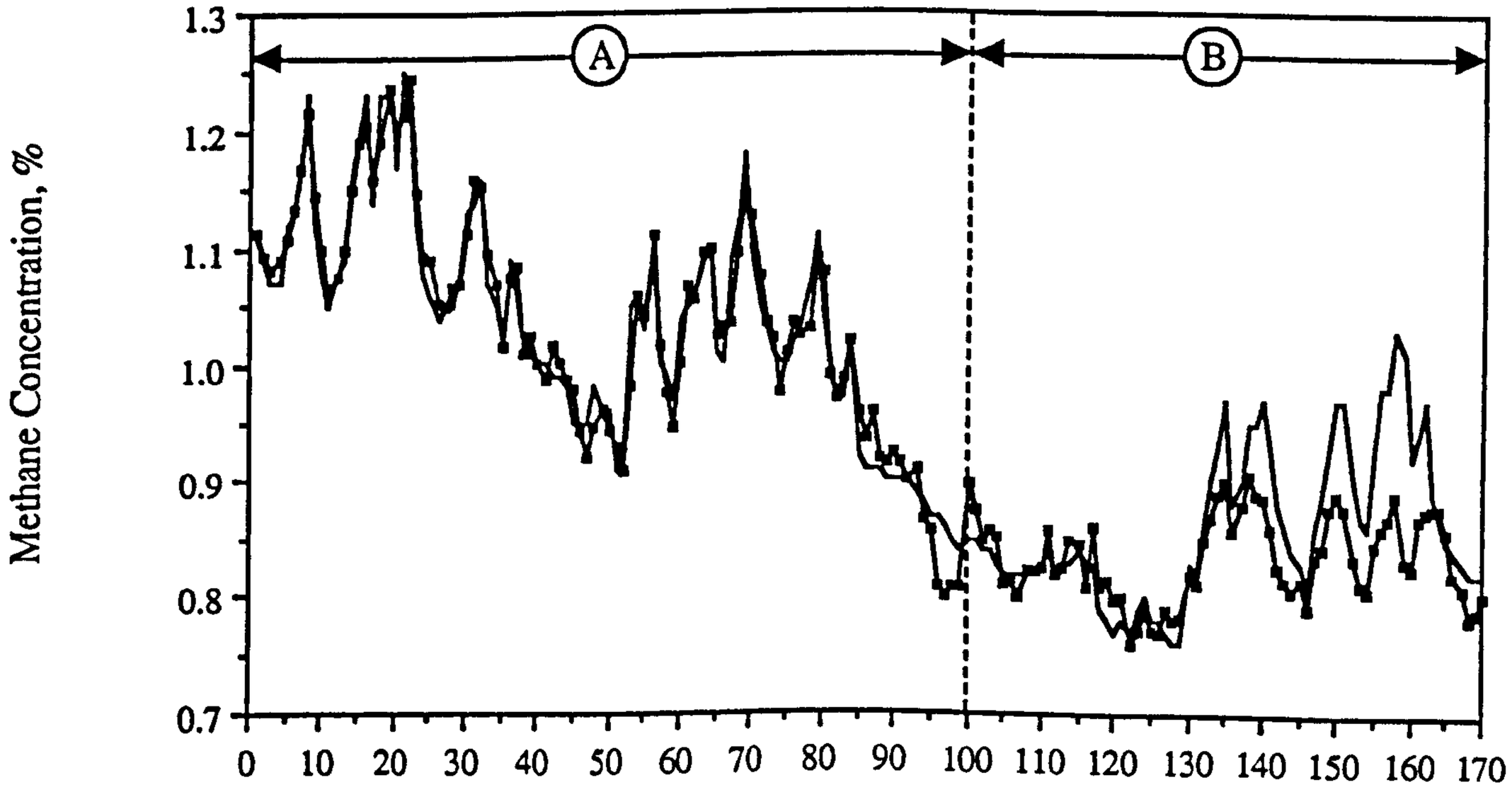


Figure 8.9 Multivariate Forecasts for Hourly Average Methane Concentration.

Initially the production values were zero since the validation forecasting period was from Sunday morning to Tuesday evening. Referring to Figure 8.10 which shows forecasts over a period of 48 hours from Sunday morning to Monday evening, the forecasts follow the actual methane concentration values well during the production free Sunday. During this period the multivariate model only considers lagged values of production to account for emissions of strata gas. Production resumed at 08:00 on the Monday morning and this corresponds to case 32 on the graph. The first peak at case 36 was predicted by the model and so was the subsequent fall in concentration explained by the production free period before the next shift commenced cutting. The second peak was also predicted but it is noticeable that the forecasts are lower than the actual methane concentration values. This is corroborated by Figure 8.9 where the forecasts beyond case 130 are usually lower than the actual values. Never-the-less, what is of importance is that the turning points or times when the methane concentration level changes are predicted accurately. Another important point to note here is that even though the multivariate model accounts for strata gas emission it is apparent that coal front gas emission is the most significant contributor to hourly average methane concentration values.

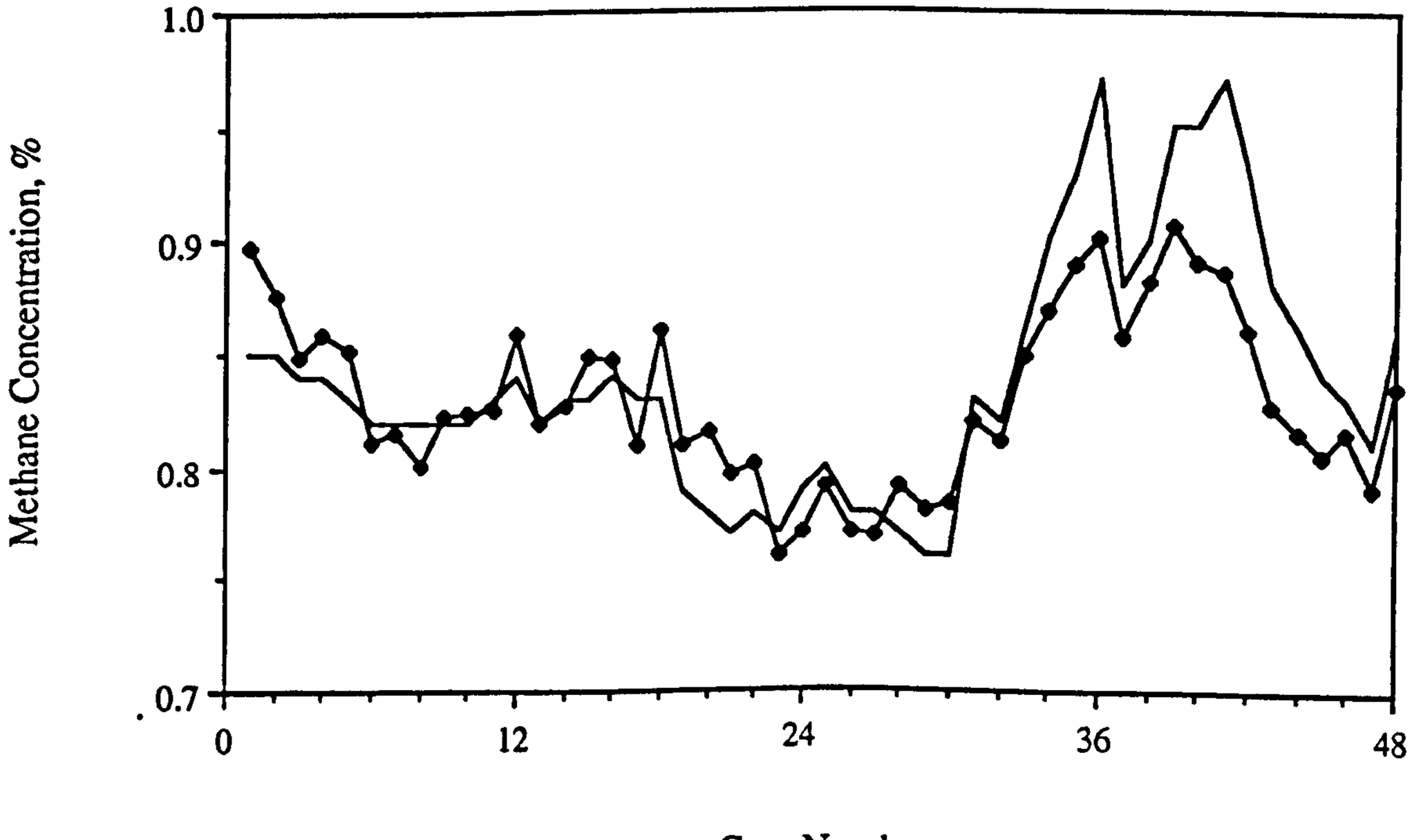


Figure 8.10 Multivariate Validation Forecasts for Hourly Average Methane Concentration.

Whereas the univariate forecasts appear to converge to a slightly oscillating level the multivariate ones are capable of accounting for coal production. The univariate models are limited in that they can only forecast future values with no regard to what might happen in this period. The forecasts illustrated in Figure 8.9 show two important features. The first is that when there is a period of no production the multivariate model begins to take less account of coal front gas and more of strata gas and this results in a realistic decay of methane concentration. The second is that the model is able to respond to a change in the production rate and in periods of variable production the model is more dependent on recent production and hence coal front gas. These two factors mean that the model could be used to plan production so that the forecasted values will be different and also to control both production and methane concentration according to a control policy.

8.6 Forecasts of the 10-Minute Average Series for Methane Concentration

8.6.1 Univariate Forecasts

The univariate forecasts for 10-minute average methane concentration can be seen in Figures 8.11 and 8.12. Figure 8.11 illustrates the forecasts for the original model for 10-minute average methane concentration, the ARIMA (2,1,3). The in-series forecasts from case 1 to case 70 follow the actual methane concentration values well but the validation forecasts are disappointing. However, they are no great surprise since the ARIMA model does not contain any seasonal component and the forecasts quickly reach a steady value.

The 10-minute univariate model is only capable of responding to coal front gas emission which over a short timescale is the most significant contributor to the methane concentration value. Case 1 corresponds to 19:40 on Wednesday 24th April 1991 and case 70 to 07:20 the next day. The afternoon shift had completed its last cut at 18:40 with a 150m strip and this accounts for the decline in methane concentration evident on the graph. The night shift commenced production at 23:40 (case 25) with a partial strip of 100m that included a stoppage from 23:45 to 00:10. The methane concentration trace shows 2 smaller peaks instead of 1 whole one and this accounts both for the reduced amount of coal cut and the stoppage. The first full strip was cut at 00:55 and the peaks (from case 35 onwards) correspond to each cut. The night shift stopped cutting at 4:10

and the methane concentration begins to decline. The forecasts are from case 70 onwards and are not able to predict the rise in methane concentration due to the start of the day shift.

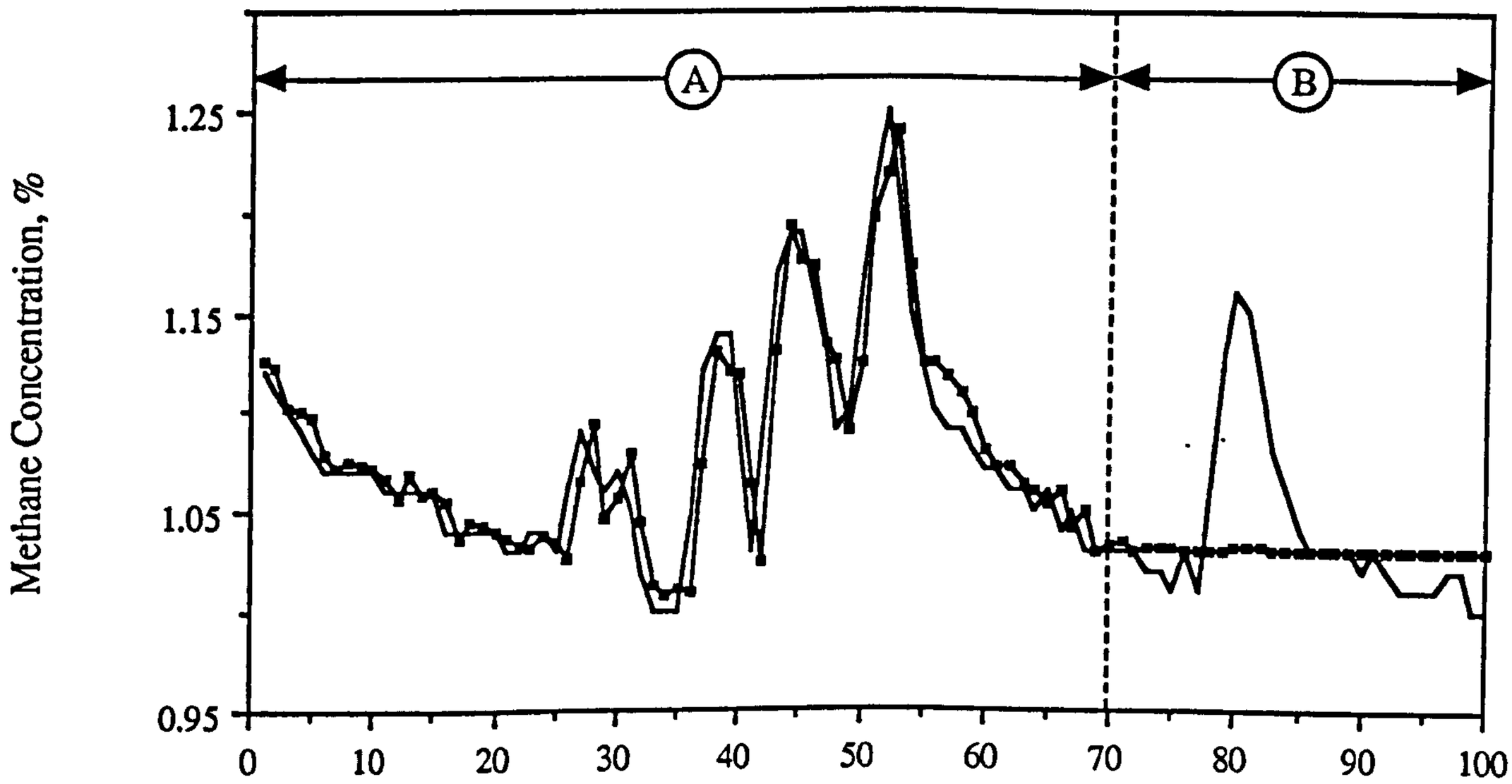


Figure 8.11 10-Minute Average Methane Concentration Forecasts for ARIMA (2,1,3) Model.

Figure 8.12 is a comparison of the validation forecasts for the original ARIMA (2,1,3) and simpler ARIMA (2,1,0) models. It shows that the validation forecasts for the models are almost identical. The MSFE for the ARIMA (2,1,3) was 0.0222 while for the ARIMA (2,1,0) it was 0.0218. Therefore, the simpler univariate model was appropriate to be used to obtain a multivariate relationship for 10-minute average methane concentration.

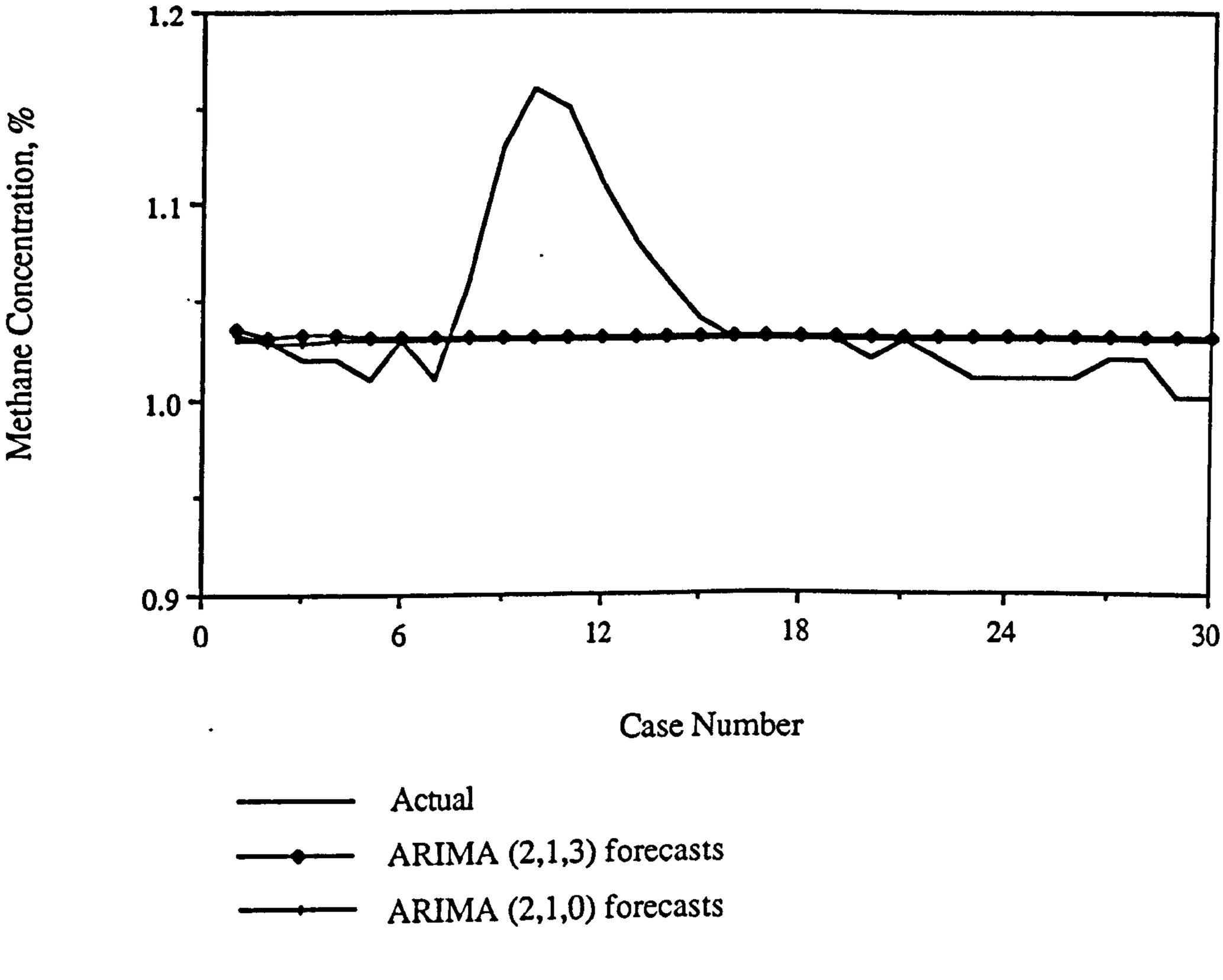


Figure 8.12 Comparison of Validation Forecasts for ARIMA (2,1,3) and ARIMA (2,1,0) Models.

8.6.2 Multivariate Forecasts

The multivariate model for 10-minute average methane concentration was identified as consisting of two components. The first was a term to account for the methane released within 10 minutes of the cut and contains methane released from the coal face and that released during transportation. The second component accounts for methane released between 10 and 70 minutes after cutting. This term consisted of lagged variables to account for both coal face and conveyor methane and possibly the appearance of strata gas. During the development of the model it was necessary to aggregate the contribution of methane released between 10 and 70 minutes after cutting and a greater weighting was applied to methane appearing on the monitor at a time between 10 and 20 minutes.

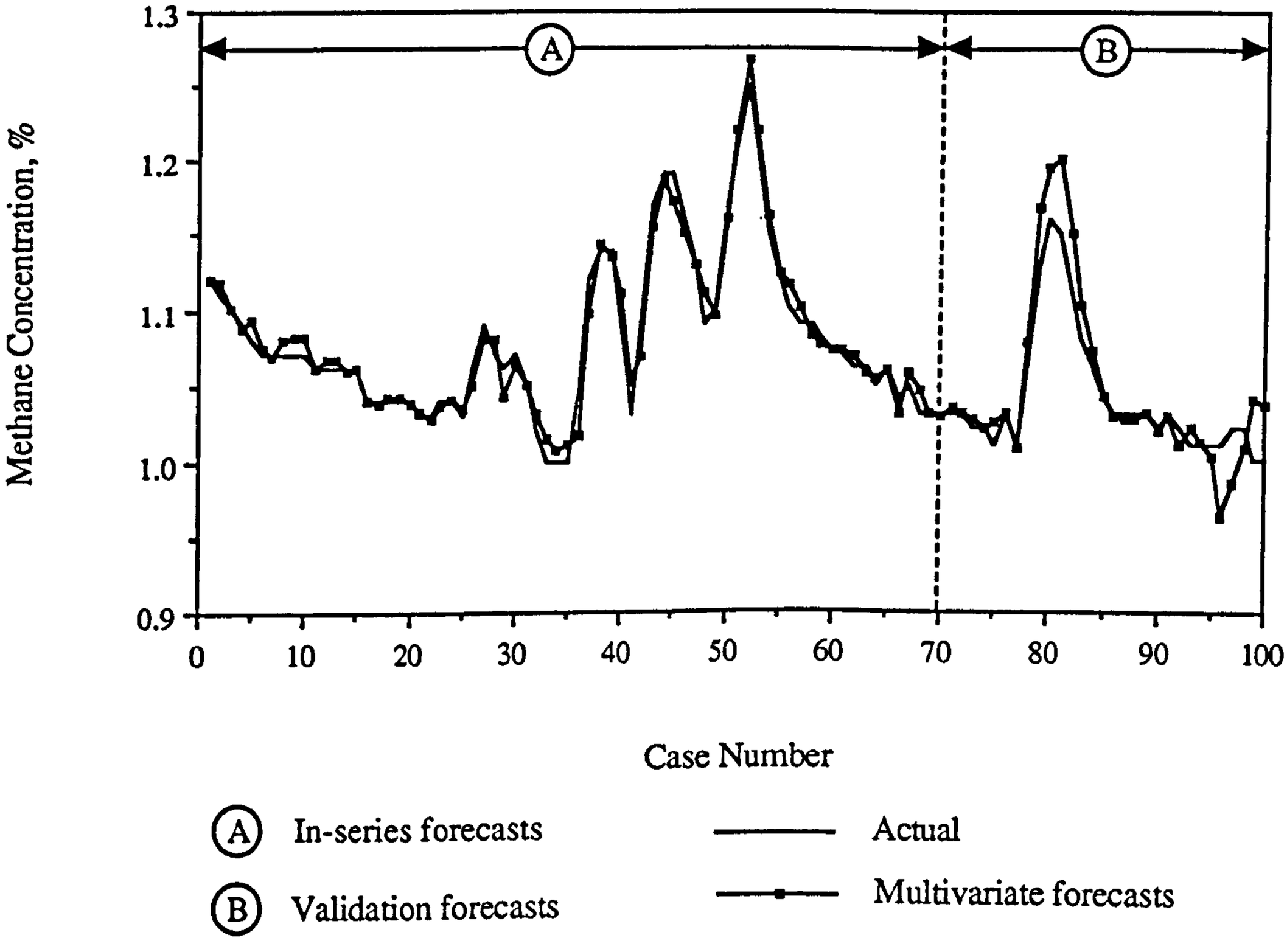


Figure 8.13 Multivariate Forecasts for 10-Minute Average Methane Concentration.

The multivariate forecasts are shown by Figure 8.13. The in-series forecasts follow the actual methane concentration values well. The validation forecasts are an improvement on the univariate ones in that they are able to predict methane concentration according to the rate of coal production. Figure 8.14 is a more detailed plot of the forecasts. Case 1 on this plot corresponds to 07:20 and production did not start until 08:15. The multivariate forecasts predict the rise in methane concentration due to the start of the day shift cutting and then predicts a decrease in methane due to discontinued production. If a univariate model had been identified that contained a seasonal component it would have predicted methane peaks for nonexistent strips. The production series that was used to obtain the forecasts was real and after the first strip of the day no more were cut. In fact the machine did start another cut but after 30 m the AFC suffered a break-down which took until 14:45 to repair. Thus, the model was able to predict representative values of methane concentration when no further production took place.

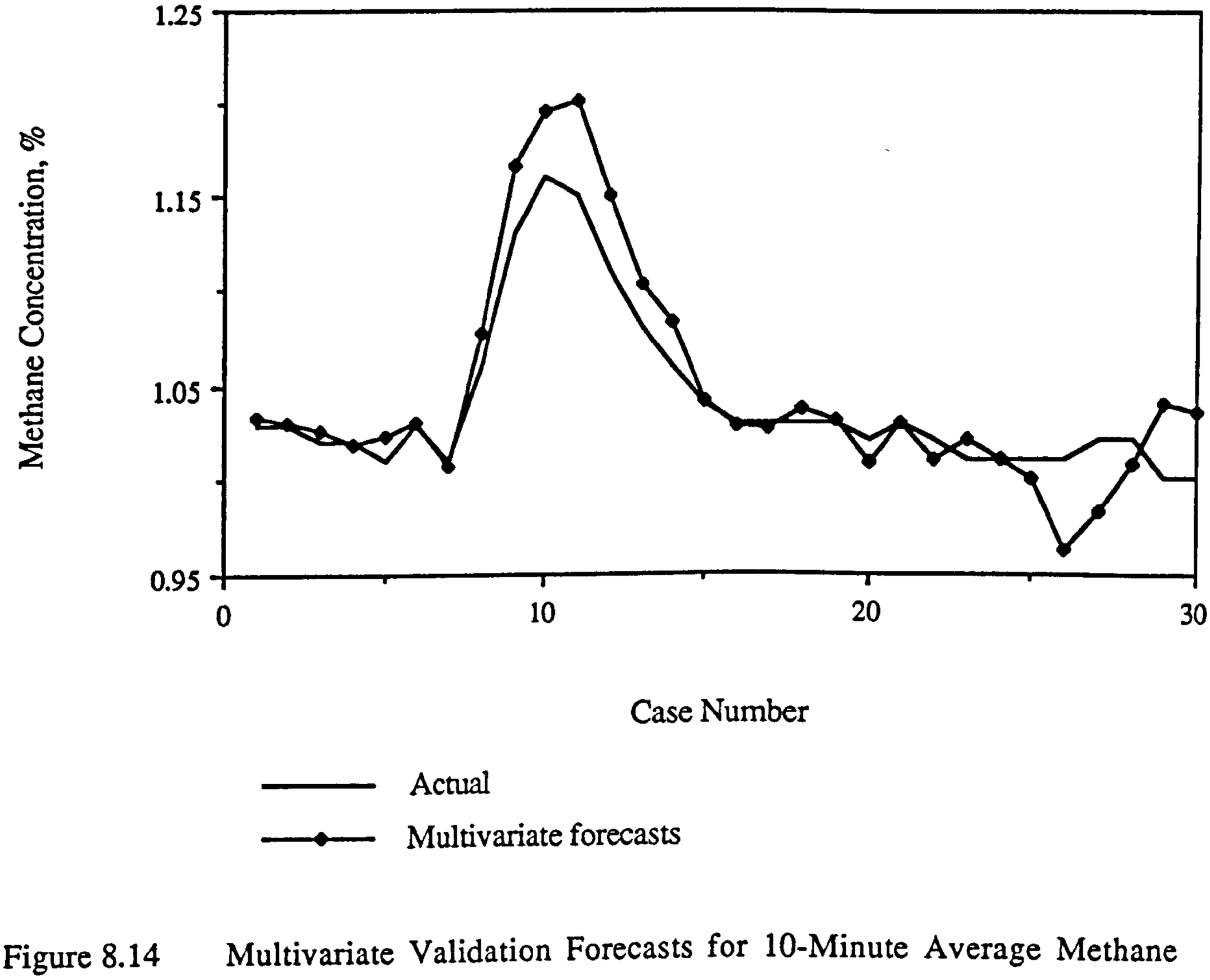


Figure 8.14 Multivariate Validation Forecasts for 10-Minute Average Methane Concentration.

8.7 Forecasts of the Methane Drainage Range Concentration

8.7.1 Univariate Forecasts

The univariate models for drainage methane concentration were all similar in that they consisted only of AR parameters. Such models are very limited in their forecasting ability and in general are only suitable for short-term forecasts of a few time intervals or steps ahead. From a planning point of view these limitations cause a number of difficulties. The drainage methane concentration was not subject to as much fluctuation as the general air body methane concentration and in fact, was relatively constant over substantial periods of time when the pumps were running. Thus, for a level value, univariate forecasts can be quite useful, but of course they are not able to respond to changes in level unless they contain some seasonal component. The inability to predict such changes

means that the forecasts have little worth.

Methane drainage is practised in most UK coal mines to some degree and many use the gas to heat water for bathing and heating. Some mines are planning to generate electricity for their own and community use and others sell the gas to industry. The quantity of gas exhausted is not under any direct control by the colliery and at times situations occur when a knowledge of a decrease in the amount of drained gas would be useful. For example, Parkside Colliery sells methane to a nearby chemical plant and they are obliged to give notice of a decrease in methane flowrate and quality so that the plant has time to switch to gas supplied by British Gas.

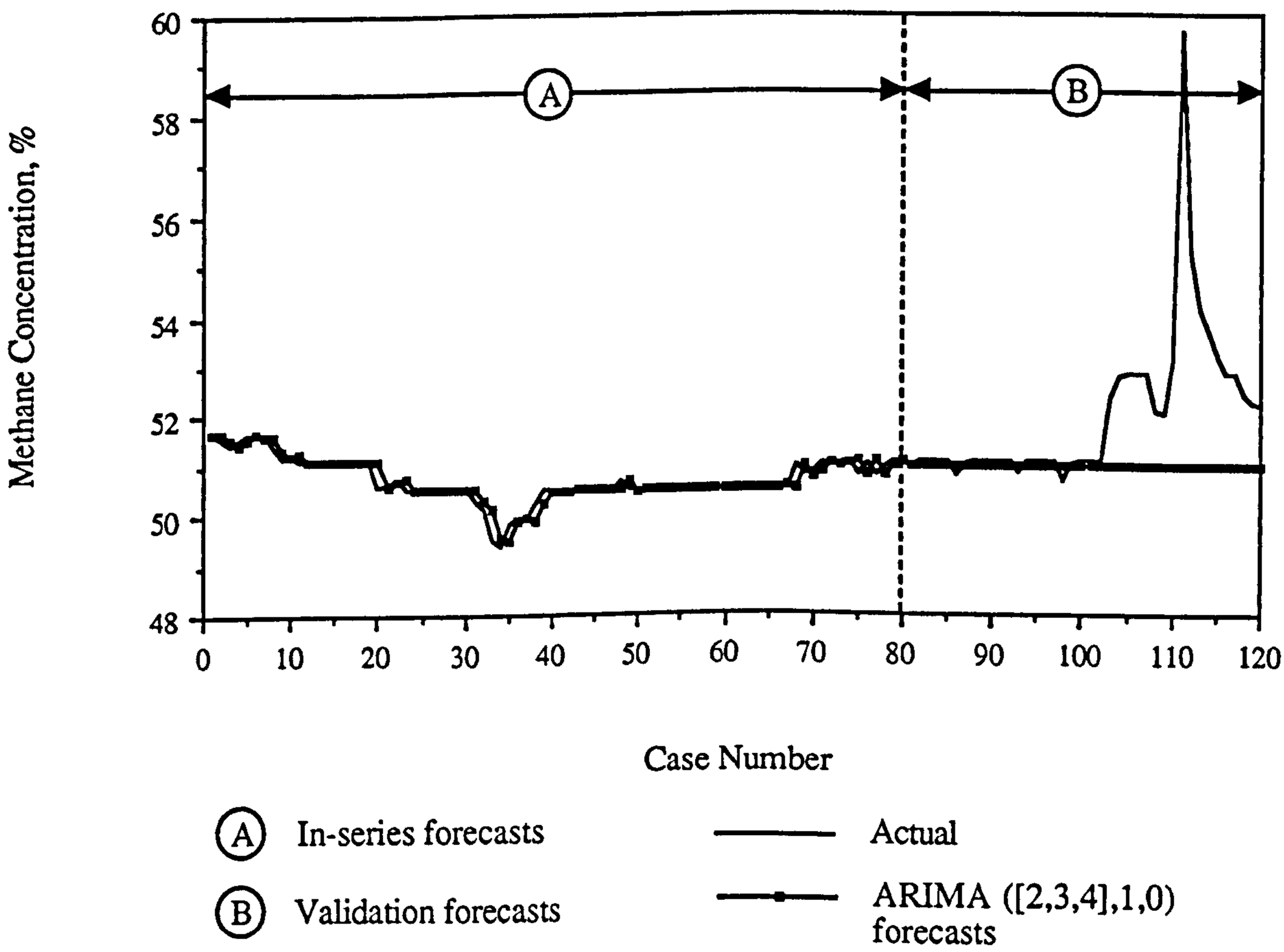


Figure 8.15 Drainage Methane Concentration Forecasts for ARIMA ([2,3,4],1,0) Model.

Thus the most useful forecasts are those to enable the colliery to plan a course of action rather than attempt to control the quantity of drained methane. The univariate forecasts for the original drainage methane concentration are shown in Figure 8.15. As expected the in-series forecasts are good while the validation forecasts although constant are satisfactory until a rise in concentration occurs due to a stoppage of the pumps. The univariate forecasts quickly reach a steady value which over a short time interval are adequate, even when the pumps are stopped for a short period of time. In practice, however, there is little gain from either having such forecasts or a model for the drainage methane concentration for short time intervals since the colliery would not be concerned with short-term variations in the quantity of drained gas.

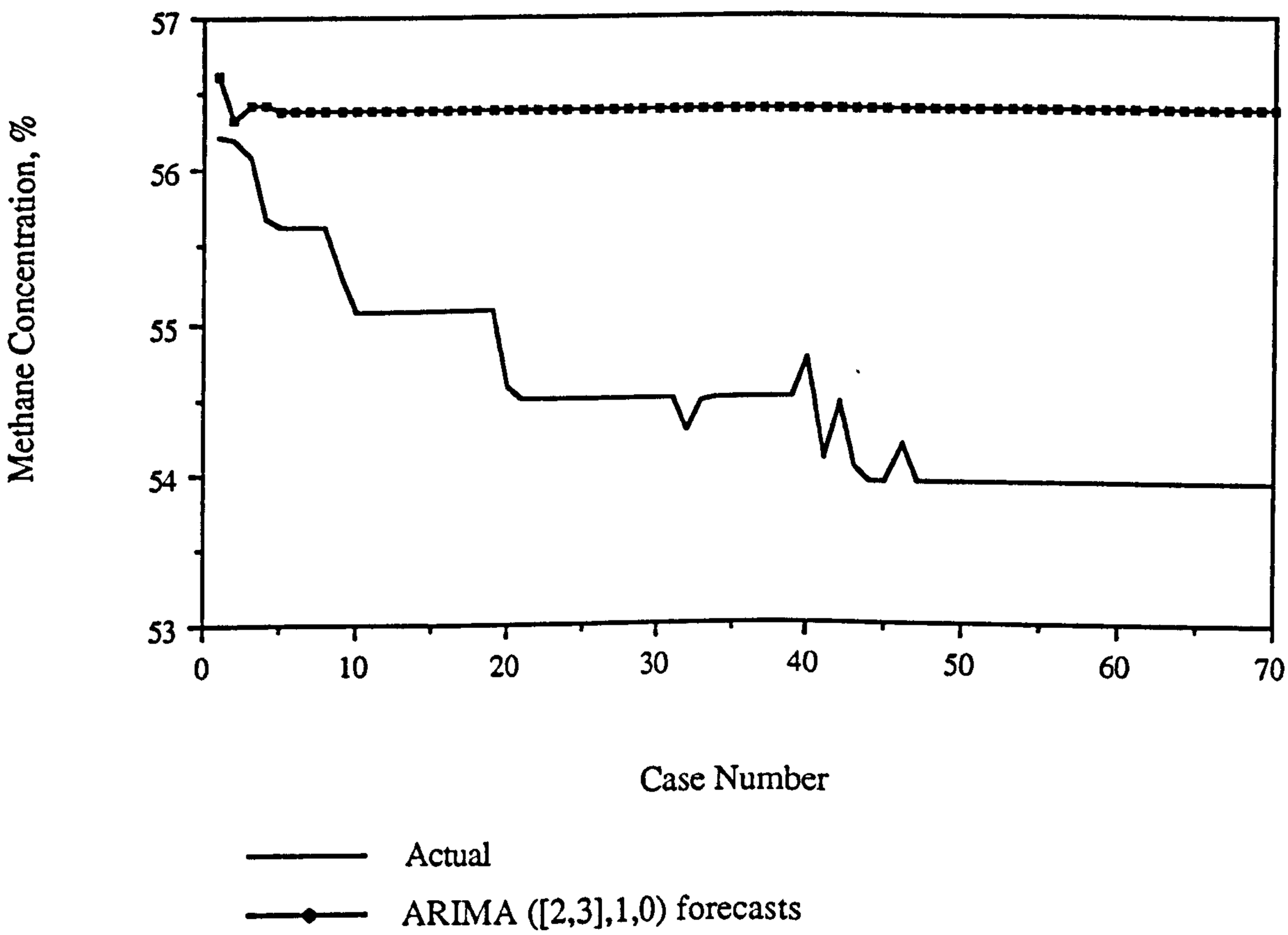


Figure 8.16 One Day Drainage Methane Concentration Validation Forecasts for ARIMA ([2,3],1,0) Model.

The validation forecasts from the one day series of drainage methane concentration are illustrated by Figure 8.16. The forecasts reach a steady level after 4 time periods (8

minutes) and do not correspond to the actual values that decrease before becoming constant.

The forecasts from the 10-minute average series behave in much the same way and are shown by Figure 8.17. It is more noticeable that as the time interval increases so does the potential for a change in level and also the potential for practical usage. Figure 8.18 shows the forecasts for hourly average drainage methane concentration. No seasonal components were evident in the univariate model but the model contained an AR parameter at lag 11. This means that for some reason the value 11 hours previously had significance but it is not easy to ascribe an appropriate reason for this. The validation forecasts quickly reach a steady level while the actual values fall and rise.

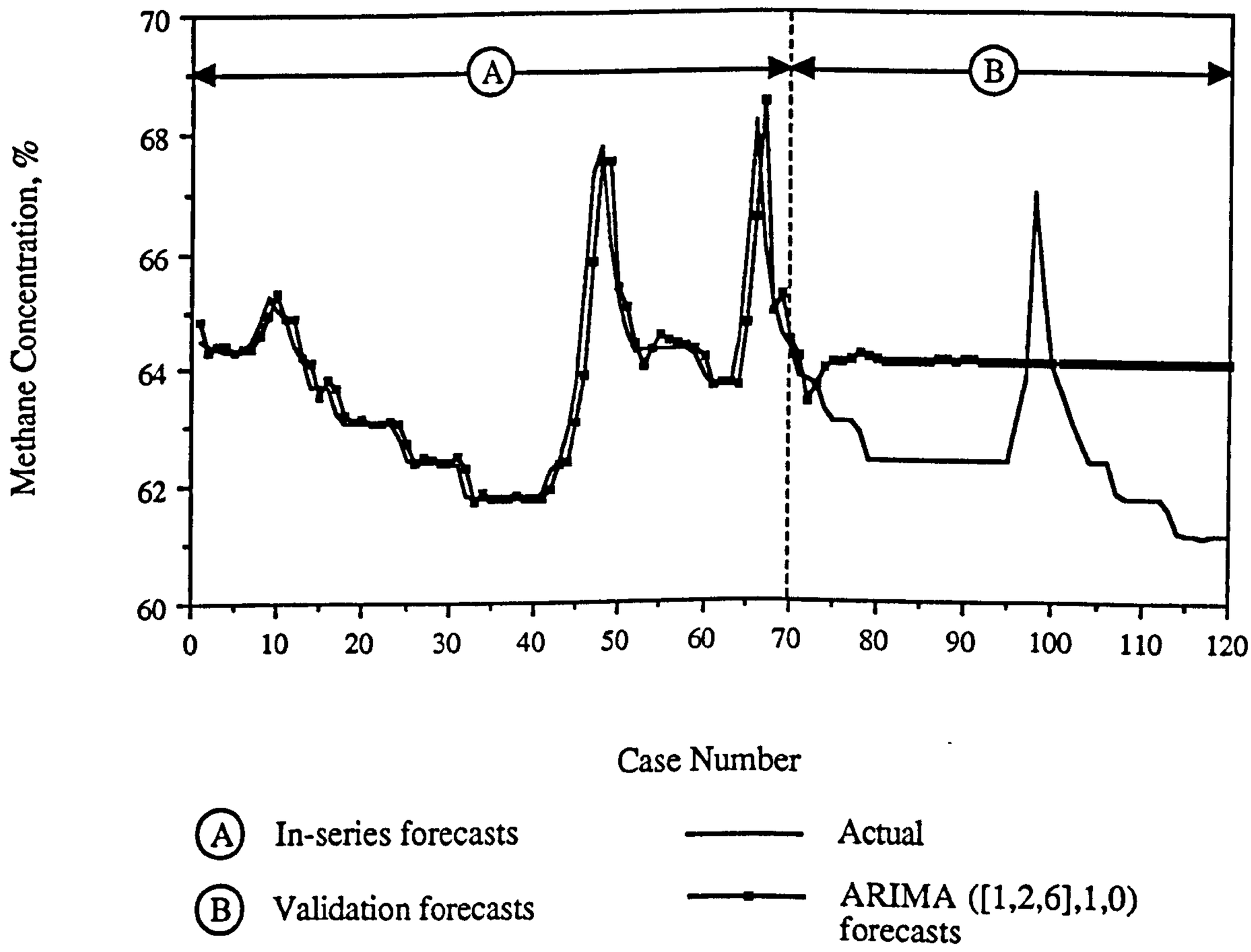


Figure 8.17 10-Minute Average Drainage Methane Concentration Forecasts for ARIMA ([1,2,6],1,0) Model.

The univariate models have all demonstrated that their forecasts are of limited use. The reason for this is not because of the inadequacy of the univariate models but rather because of the nature of long-term univariate forecasts which very quickly reach a steady level. For use in planning therefore, they have limited application. For a model to be of practical use it must be capable of accurately forecasting a number of hours into the future and only a model for hourly average drainage methane concentration has this potential.

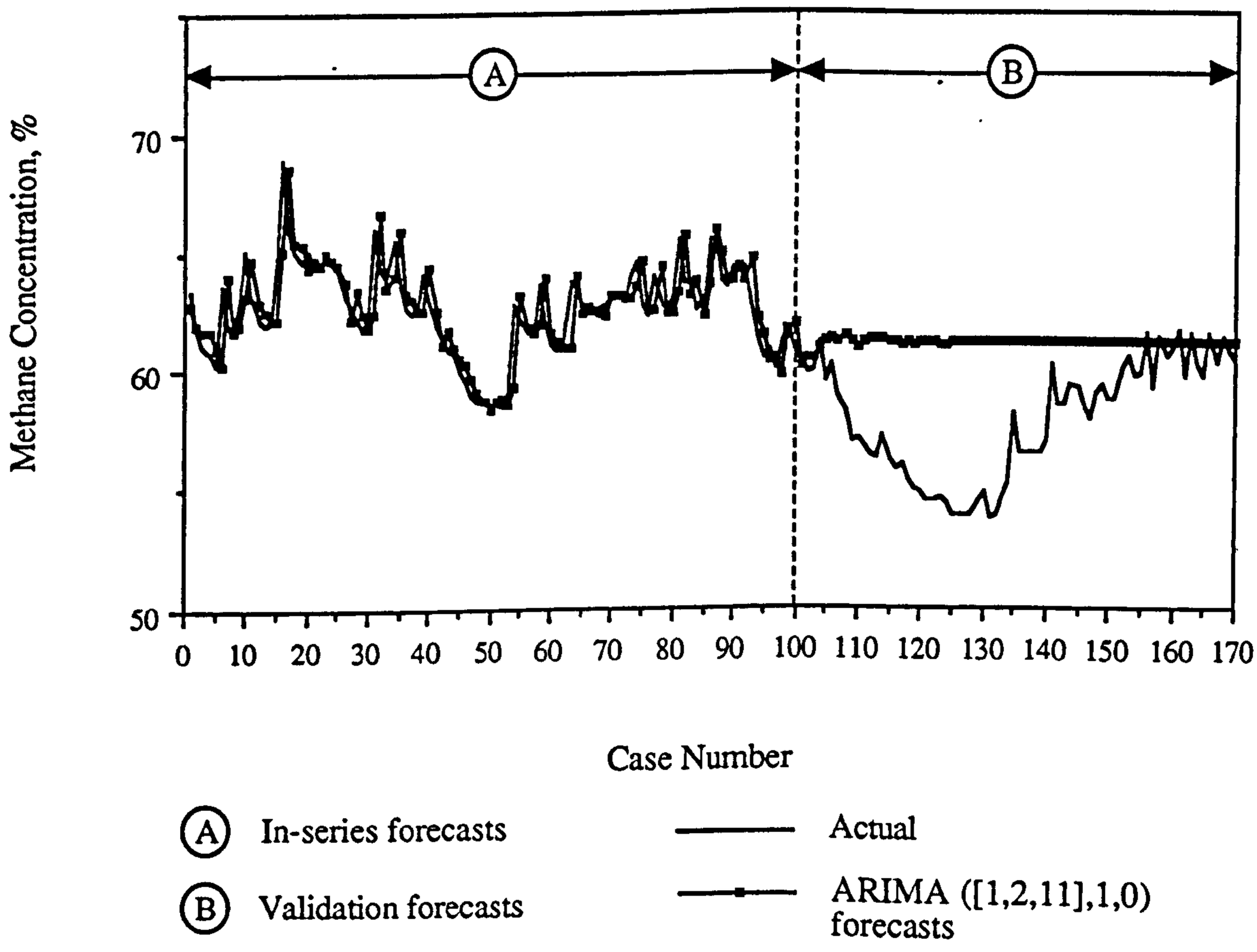


Figure 8.18 Hourly Average Drainage Methane Concentration Forecasts for ARIMA ([1,2,11],1,0) Model.

8.7.2 Multivariate Forecasts

In the previous chapter no correlation between drainage methane concentration and coal production was identified for models other than the hourly average ones. This was not unusual since drainage methane concentration values monitored at a frequency of 2-

minutes can in no way be affected by recent coal production. It is conceivable that there could be a link but this is easily obscured by the large amount of data that would be needed to identify such a relationship. The multivariate analysis identified a weak residual relationship between drainage methane concentration and production that signified a subtle relationship between them. The subsequent model estimation resulted in an appropriate model that identified a link between production and drainage methane concentration at 1 to 3 hours and 12 hours previously. This is consistant with the univariate model that consisted of parameters at lags 1, 2 and 11. It would seem therefore, that production has a rapid and lagged effect on the quantity of drained gas.

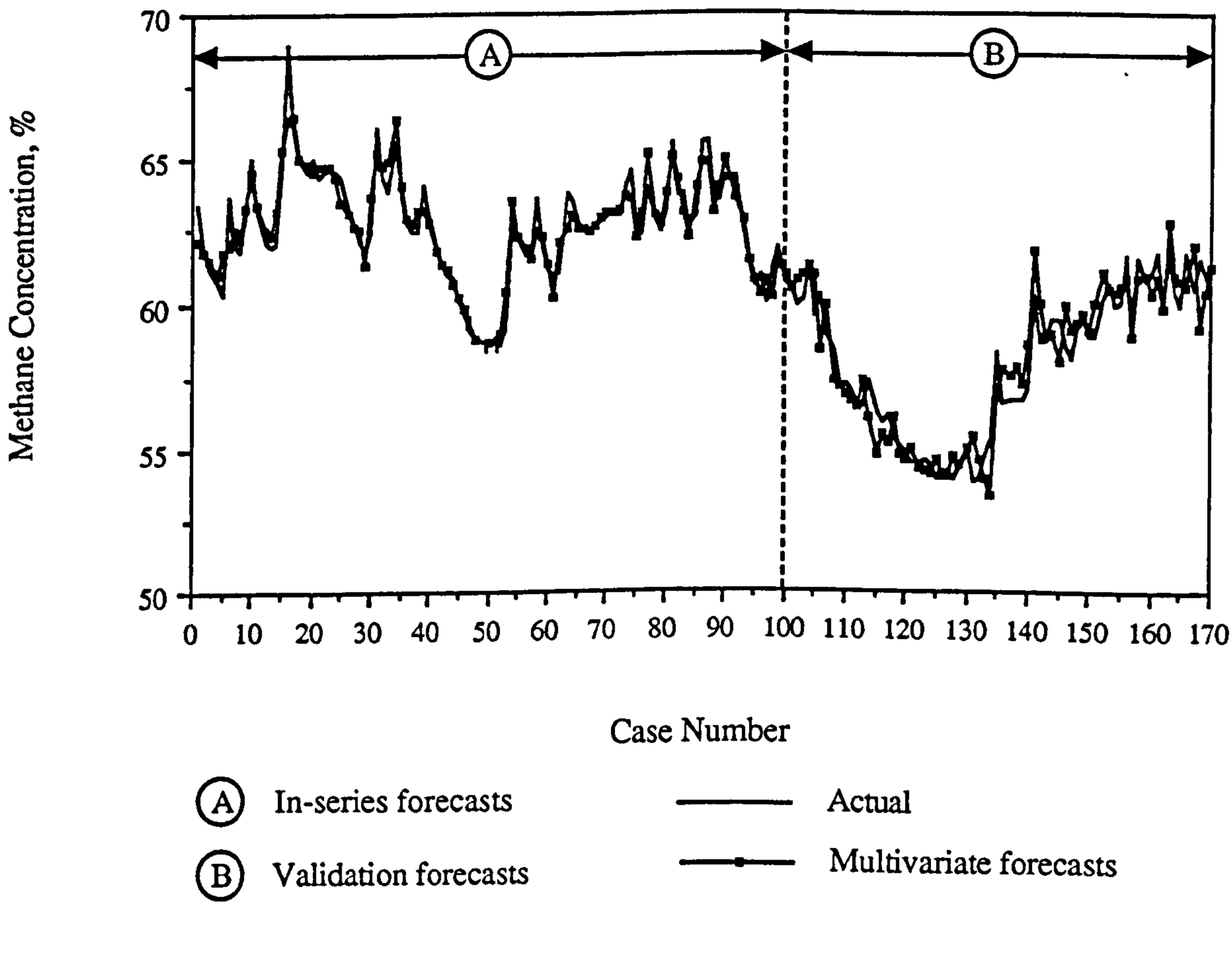


Figure 8.19 Multivariate Forecasts for Hourly Average Drainage Methane Concentration.

In building the hourly average model for general air body methane concentration in terms of production it was found that strata gas took up to 26 hours to appear in the airstream. This helps to explain the correlation between drainage methane concentration and

production since the methane drainage boreholes are targeted into high gas content strata and so gas being drawn towards them would take less time to appear.

The forecasts from the model are illustrated by Figure 8.19. The forecasts are for the same time period as those for the hourly average general body methane concentration. The in-series forecasts are good while the validation forecasts manage to predict the actual values well, albeit with some fluctuation. A drop in concentration occurs around case 50 which corresponds to early on the Friday morning and is due to production having stopped at 04:20. Production recommenced at 08:15 and an increase in concentration is seen because of this. Concentration is then seen to fluctuate and each peak corresponds to each shift of coaling. Production stopped for the weekend break early on the Saturday morning (case 75 onwards) and after a time delay of a few hours the concentration is seen to decrease noticeably, before rising again due to the restarting of production on the Monday morning (case 130 onwards). The good correlation between the behaviour of the actual concentration values and production verifies the suitability of the multivariate model for drainage methane concentration. This model identified that production caused a change in concentration after 1 hour and that production up to 12 hours previously, contributed to the gas present in the drainage system.

Figure 8.20 Multivariate Validation Forecasts for Hourly Average Drainage Methane Concentration.

The validation forecasts can be more clearly seen in Figure 8.20. These forecasts are for 48 steps ahead i.e. 48 hours and they manage to follow the actual values quite well. The forecasts are accurate to within 10% of the actual value.

During the course of the research only the methane drainage variables of 119's were monitored and in practice a colliery would be more concerned with the total amount of gas present in the whole drainage system. It would still be necessary, however, to monitor each district individually so that a true indication of the likely concentration due to production could be found. The forecasts from each district could then be aggregated to arrive at a total amount for the whole mine.

8.8 Application of the Models to Process Control

Now that the models have been used to obtain forecasts and the forecasts have been examined according to their ability to predict changes in the level of methane concentration it is necessary to discuss how they might be used when applied to process control. Process control is a means whereby the forecasts are used to provide either a computer or human operator with information on which they may or may not act. A good example of this are the forecasts for methane concentration. If a model were working within a method for production control and the forecasts indicated that at a desired rate of production a methane concentration in excess of the statutory 1.25% would be exceeded at a certain time two courses of action are possible. The first, and least attractive would be to do nothing, while the second involves taking action to lower the impending high methane concentration levels. Lowering of the methane concentration could be accomplished by either revising the planned production schedule or introducing additional airflow at the required time. Depending on how far the statutory methane limit was to be exceeded it may be appropriate to use a combination of these two factors.

For a model to be used in such instances it must be capable of predicting changes in the methane concentration and this immediately excludes the use of all the univariate models for methane concentration except the one for the hourly average series. However, it was seen that although this model contained a seasonal component its forecasts are based entirely on the assumption that future production will be of a similar pattern to that that took place during the historical portion that was used to estimate the model. Such a model would not produce accurate forecasts for widely differing production schedules.

Only the multivariate models for methane concentration were capable of predicting accurate values for methane concentration according to changes in coal production. The two models were for hourly average and 10-minute average methane concentration and both demonstrate particular forecasting abilities. These are defined by the time interval between observations and the effect of production on methane emission during a particular models timescale.

8.8.1 Forecasting Ability According to Time Interval Between Observations

It was observed in sections 8.3 and 8.4, which dealt with univariate forecasts of methane concentration from both the complete original series and one day series of methane concentration, that the forecasts reached a steady level after a short number of forecasted time periods. It was not possible to build a multivariate model for methane concentration on the original timescale but if one had been available it would only have been capable of forecasting methane concentration which would have mainly been due to coal front gas emission. To use the forecasts for control would be difficult as it is unrealistic to vary coal production over such a short timescale and impractical to vary the air flowrate as a means to lower the methane concentration.

The forecasts of the 10-minute average univariate models are similar to those of original timescale in that they to quickly reached a steady level. The multivariate forecasts are able to predict the changes in methane concentration due to production and the forecasts are accurate for about 25 time intervals or around 4 hours. The value for methane concentration is composed mainly of coal front and conveyer gas and the multivariate model contains components to account for gas released after 10 minutes of cutting and a term encompassing gas released between 10 and 70 minutes after cutting. Thus the model is better suited to providing an indication of how a change in production in the short-term would effect methane concentration.

The hourly average multivariate model for methane concentration was able to predict accurately for 48 time intervals or 48 hours and thus provides a good indication of likely values of methane concentration according to the level of production. This model contains terms for coal front and strata gas and would provide an ideal indication of methane concentration due to both of these factors. The hourly average model for drainage methane concentration also produced accurate forecasts for 48 time intervals and contained terms to account for recent and lagged production.

8.8.2 Forecasting Ability for Control Purposes

For the purpose of providing a control mechanism for methane concentration it is proposed that a combination of the two multivariate models for methane concentration could be used. The 10-minute model is capable of accurate forecasts of methane concentration due almost entirely to coal face and conveyer gas and if forecasts indicated that the statutory methane level was to be exceeded action could be taken to prevent this. These 'short-term' forecasts are possibly suited to short-term ventilation solutions whereby coal production remains unaltered and extra airflow is created to reduce the general body methane concentration. Such a solution might be appropriate where the methane concentration lies close to the 1.25% limit and the peaks from each shift are sufficient to raise the concentration above this level. Ideally the whole coal cutting operation would have to be in complete computer control for this scenario to work.

The hourly average model is capable of a much longer forecasting period and could be used to plan production so that maximum use of the working time was realized. For example, from an arbitrary start time it is intended to cut coal without reaching the 1.25% limit for a non-stop period of 24 hours after which there is a scheduled maintenance period of 4 hours. The effects of various production rates on methane concentration could be determined and a number of production options could be available. One of these may be to produce coal at such a rate that the statutory methane limit is exceeded at a time as close to the start of the maintenance period as possible in the supposed knowledge that the 4 hour stoppage would be sufficient to allow methane levels to decrease so that production could be restarted at the scheduled time. The forecasts from the 10-minute average model could be used in conjunction with the hourly average model in this particular example as they might provide an early warning of the possibility of greater or lower methane levels than expected. If this model forecasts lower concentrations production could be stepped up. If higher concentrations are predicted extra airflow could be provided as a short-term solution leaving production unaltered.

The model for hourly average drainage methane concentration could be very useful to a colliery that is dependent on its drained gas for power generation or commercial revenue. The model is capable of predicting likely values of drainage methane concentration that can be converted to a flowrate using an average value for the flowrate of gas in the range. A colliery would be able to predict short-term variations in quantity due to the models

term accounting for production 1 to 3 hours previously and longer-term forecasts due a lagged production effect of 11 hours. It is doubtful that the forecasts would be used for control purposes because a colliery is more concerned with producing coal rather than gas. Instead they would provide a warning of whether the availability of gas would cause a problem and if so appropriate action could be taken.

8.8.3 Model and Forecast Monitoring

The forecasting models can either be used by themselves whereby the predictions they produce are acted upon by a control room operator or by a computer within an expert system. How much control such a person would have is dependent upon the type of control desired. In a simple situation the control room operator may only be able to inform the face team or manager of likely problems caused by over production, leaving further decision to those with greater ability or superiority. A second and much more complex application of the models would be their inclusion within a dedicated computer control package capable of complete coal face environmental and production control. However the models are used it will prove necessary to monitor them so that they continue to produce accurate forecasts and the nature of monitoring and diagnostic checking will depend on the complexity of model use.

The univariate models that are used to produce a multivariate model are based upon the statistical behaviour of past values and in producing univariate forecasts assumes that the future will behave in much the same way. The multivariate models also rely on this assumption even though they are capable of predicting methane concentration in terms of coal production. There may come a time when some future events are untypical of past behaviour and if these continue it is likely that the multivariate model will cease to produce accurate forecasts. It then becomes vitally important to be able to correct for such instances.

The on-going suitability of a model can become inadequate for two basic reasons. Firstly, the dependent variable which in this case is methane concentration can be disturbed in the short-term, i.e. by very large fluctuations in air velocity but the underlying behaviour may be largely unaffected. Secondly, the underlying behaviour of the dependent variable may change in which case the model may need substantial

revision. Such an occasion could occur if the coal face moves into an area of altered geology. This could be sufficient to completely change the time for strata gas to appear due to changes in permeability and stress fields that are themselves caused by production.

A number of monitoring tests are available to detect changes that may involve the altering of a model and these can also be used to improve upon the original model specification. The tests are basically the same regardless of how the models are utilized. The monitoring tests can only be applied to models that are an adequate representation of the data and so care must be taken during the model building process to ensure that this is so. The tests are then applied to new data, forecasts and model residuals.

There are three main tests used in monitoring a multivariate model [54]. These are:

1. single anomalous residuals,
2. changes in the model parameters,
3. change in the model structure.

The first of these can occur if the model encounters a dramatic change in the behaviour of the dependent methane concentration variable that is not typical of its previous historical behaviour. Applying the model to the new augmented series it is possible that this behaviour will result in a few abnormally large residuals. A change in variable behaviour could be for two reasons. Firstly, although the airflow was not included in the multivariate models, the methane concentration values are highly dependent on the airflow variable. The air velocity was found to fluctuate by approximately 10% over the data monitoring period but if in the future it fluctuated at 20% then the relationship between methane concentration and production would be altered. Secondly, a major change in the production variable would also result in model inadequacy. At present production at Thoresby seldom took place over the weekend, but if production was under complete computer control and the shift pattern was changed substantially the model may need to be respecified to account for this change in production pattern.

It is possible for the model parameters to change over time without there being a change in the model structure. As new data becomes available the univariate models used to

obtain the multivariate relationship should be re-fitted to the augmented series without a change in the model structure. If there has been little change in the behaviour of the variables this will generally result in a confirmation of the models suitability or even an improvement in the appropriateness of the models parameters. However, if this is not true then the whole univariate model building process will need to be carried out so that new univariate models are available. For whatever reason causes the change in the model parameters the multivariate analysis would also need to be performed again.

A change in the model structure is identified by examining the model residuals. If the model is correct the residuals should demonstrate white-noise characteristics. If the residuals show autocorrelation then this may be evidence of whether structural inadequacy or incorrect parameter values. Usually, if the models parameters are noticeably changed after re-fitting then it is necessary to re-specify the model structure.

8.9 Conclusion

This analysis presented in this chapter has been concerned with a critical assessment of the forecasting performances of the univariate and multivariate models for general air body and drainage methane concentration. The forecasting ability of each of the models was also examined from the point of view of their ability to forecast correct levels of methane concentration. A discussion on how the models could be incorporated into a control policy was also entered into and lastly aspects of model performance monitoring were detailed.

The univariate validation forecasts of methane concentration all highlighted the type of forecasts that such models are capable of producing. For univariate models that contained no seasonal component the forecasts were seen to very quickly reach a steady or constant value. The only exception was the hourly average model for methane concentration. The forecasts from this model predicted higher concentrations than actual during the production free weekend and were unable to forecast the increase of methane concentration due to the recommencing of production at the beginning of the week. In general, univariate models are only capable of producing representative forecasts when the future behaviour of a variable is similar to its past behaviour and in the prediction of methane concentration this is not usually true.

Forecasts were also obtained from the proxy univariate models for the original series of methane concentration, hourly average methane concentration and 10-minute average methane concentration. These models were used in chapter 7 to obtain a multivariate model between methane concentration and coal production. According to appropriate diagnostic checks these models appeared to be adequate and their suitability was further confirmed by the forecasts obtained from them. In all cases the forecasts were almost identical to those from the correct models and this means that so long as a minimum degree of diagnostic checks have been performed, the model forecasts are not too dependent on the structure of the model.

The univariate models were seen to be incapable of forecasting methane emission according to changing levels of production and this contrasts noticeably with the performance of the multivariate models. The multivariate models for hourly average and 10-minute average general body methane concentration and hourly average drainage methane concentration were all able to predict accurate values for methane concentration due to changing levels of production. The two models for general body methane concentration were able to predict values that consisted of both coal front and strata gas components. The forecasts for drainage methane concentration identified how recent and lagged production influenced the concentration of methane in the range.

The time interval between observations was seen to be a decisive factor in determining how useful the forecasts would be for planning and control. Models built from series with a short time interval between observations are only capable of a few useful step ahead forecasts at best. This is true for both univariate and multivariate models. Although a multivariate model for methane concentration data with a time interval of 2-minutes was not built, it is doubtful whether the forecasts could be of use. The forecasts would only show short-term variations in methane concentration and in practice it would be difficult to provide responsive action if an excessive level of methane concentration was anticipated. The two multivariate models for general body methane concentration demonstrated different forecasting abilities according to their time scales. The model for 10-minute average methane concentration was able to forecasts short-term methane values due mainly to the release of coal front gas for each strip. The hourly average model was able to forecast accurately for 48 hours into the future and included terms to account for coal front and strata gas emission.

The good forecasting ability of the models for general body methane concentration enables them to be used within a planning and control strategy. Two brief scenarios were discussed and each model would be capable of providing answers in the form of representative predictions to these and other scenarios. The hourly average model is appropriate for providing long-term forecasts while the 10-minute average model is more suited to providing short-term indications of methane concentration. If the models were used for planning and control it is believed that they should be used in conjunction with each other. Planning and control are general terms that will depend on how the models are to be used. The models are quite capable of inclusion within a sophisticated control system that could be used for compete coal face environmental and production control.

Where the models are being used for control purposes it is very important to monitor their performance. In an on-going control situation it is almost certain that the model parameter values will alter and the model structure may also need revision. Checks for model inadequacies will have to be carried out either by a qualified person or ideally by a control mechanism itself thereby ensuring that the models are performing at their best.

CHAPTER NINE

CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER WORK

9.1 Summary of the Research

This thesis describes an investigation into the application of a statistical method for the prediction of methane concentration in longwall coal districts. This method of methane prediction was achieved by using the Box-Jenkins time series analysis method to obtain univariate and multivariate models for methane concentration and other variables. The aim of the research was to demonstrate that such models could be used to accurately predict methane concentration from a consideration of already existing data.

In chapter 2 a concise account was given of the origin of methane and the mechanisms which cause its release and flow into mine workings. Chapter 3 was a review of previous methods of methane prediction. In this chapter a statistical method of methane prediction was presented as being distinct from other empirical or numerical ones. The reason for this treatment was whereas truly empirical methods of methane prediction rely on a knowledge of quantitative data of many physical parameters, a purely statistical method such as time series analysis does not.

Chapter 4 provided an introduction to the basic philosophy of time series analysis and explains the three stages involved in building a univariate model. Time series analysis allows a model to be built purely on a statistical basis and then seeks to use expert knowledge concerning the variables to explain the structure of the univariate model. In this way the best possible model can be built to describe the behaviour of the time series. The structure of the univariate model was shown to be dependent on the frequency and quantity of the time series data used in the model building process. For example, in the analysis of the hourly average series for methane concentration a visual inspection of the data revealed the possibility of two seasonal components, a daily cycle and a weekly cycle. The analysis, however, was only able to discern a daily cycle of 24 observations and not the weekly cycle of 144 observations. In general, the smaller the time interval between observations the more difficult it is to discern cycles with long periods because too much data tends to obscure the patterns within it.

Aspects of underground environmental monitoring were examined in chapter 5. Detailed descriptions were given of the environmental monitors used at Thoresby Colliery to monitor the data. Reference was made to ensuring that the monitors were in their normal state of calibration and maintenance so that typical data was recorded. All of the variables used to build the time series models were already monitored or recorded by the colliery and the data analysis required no extra environmental variables to be monitored. Severe problems were initially encountered due to the difficulty of extracting the data from the colliery information system computer. These were solved by the acquisition of a communication package that was used to transfer data from the colliery computer to a PC. After transfer considerable effort was required to transform the raw data into a format suitable for time series analysis.

Univariate models of methane concentration and other variables for data with four different timescales were built in chapter 6. It was proposed that methane concentration was the only dependent variable and that it could be caused or influenced by coal production, barometric pressure and air velocity. The univariate models were built for two main reasons. Firstly, the intention of chapter 7 was to investigate the possibility of a multivariate relationship between methane concentration and its explanatory variables but before this could be done it was necessary to build univariate models for each variable. Secondly, the univariate models for methane concentration were to be used to generate forecasts to compare with multivariate ones. Models were built for all of the variables except barometric pressure. These models were flawed because of the inadequacy of the method of barometric pressure monitoring and transfer of data to the computer. A further problem was the inability to reduce the Box-Ljung values of the residual autocorrelations at high lags for models built from very long time series. It was shown, however, that a simpler model could be obtained with only a small increase in residual standard error as penalty. An important conclusion of this chapter was that once a suitable model had been arrived at, reference was made to a priori knowledge in an attempt to explain its structure and hence verify its suitability.

Chapter 7 was concerned with a multivariate analysis of the variables. A conceptual model was presented in which methane concentration was linked to causal variables. An outline of the procedure for multivariate analysis was also provided. The analysis was based on the correlation between methane concentration and other variables such as production and barometric pressure. For averaged models it was inappropriate to use air velocity so that a model for actual methane emission could be built, while for spot values

there were no correlations between methane concentration and air velocity. Production was identified as the only variable capable of being used to build a multivariate model for both general body and drainage range methane concentration. Multivariate models were built for hourly average general body and drainage range methane concentration and 10-minute average general body methane concentration but for some of the models it was necessary to use simpler univariate proxies to enable them to be built. A number of computational difficulties were experienced that proved to be a limiting factor on the type of model that could be built. Essentially, models containing a large number of model coefficients and series that were long in length, were beyond the capabilities of the software used to perform the analysis. It was seen that the structure of the multivariate models were conversant with practical knowledge of the likely effect of production on methane concentration.

In chapter 8 univariate and multivariate forecasts for methane concentration were illustrated and compared. Forecasts of univariate models were characterized by their rapid convergence to a steady value and their inability to predict turning points in methane concentration. The multivariate models were shown to be capable of predicting when methane concentration levels would fall and rise and are thus able to forecast methane concentration according to the rate of production. The forecasting ability of the multivariate models were found to be dependent on the timescale of the models. Hourly average models were able to account for both coal front and strata gas emission while the 10-minute average model was better at predicting coal front gas emission. For the most beneficial use of the forecasting models it is envisaged that they should be used in conjunction with each other. There is no doubt that the models lend themselves to use for planning and control and if used for such purposes it will be necessary to monitor their performance.

9.2 Main Conclusions

The most important conclusion of the research is that the statistical method of time series analysis is capable of producing statistical models for the prediction of methane concentration. Univariate models do not produce useful forecasts but they are necessary to infer a multivariate relationship between methane concentration and explanatory variables. It was found that coal production was the most important variable to consider and subsequent multivariate models for methane concentration in terms of production

produced forecasts that were capable of predicting changes in the level of methane emission.

The analysis was carried out by making use of environmental data that was already in existence. This is an important point because the prediction method does not entail any additional effort and expense to acquire physical data such as strata gas contents, and alleviates the need for an accurate understanding of strata stress/permeability relationships. Unlike analytical models that usually have to be modified for application to slightly different circumstances statistical ones have the benefit of being appropriate for each individual situation. This means that they are capable of producing an optimum model for each new situation.

The analysis has resulted in models that clarify the relationship between methane concentration and its explanatory variables. If a model is built for an average series it is inappropriate to correct for air velocity and obtain actual methane emission. This is only appropriate where spot values are concerned but such models are not particularly suited to planning and control due to the short-term nature of their forecasts. Barometric pressure is only likely to cause evident changes in methane emission if the rate of pressure change is large and if the timescale of the model is lengthy. Otherwise any observed effect will not be evident in the time series analysis. Production was identified as the most important variable and its effect on methane concentration was consistant with practical knowledge and theory of production related methane emission. It was also seen that the models were capable of predicting directional changes in methane concentration according to changes in the rate of coal production.

The multivariate models are suitable for use in planning and control. The forecasts of methane concentration according to changes in production can be used to plan production so that the statutory methane limit is not exceeded. If forecasts indicate possible problems the models can consider different production rates so that a number of solutions could be found. It is also possible that the models could be contained within a dedicated computer programme for complete environmental and production control of a completely automated coal face. For application within a control system the models could be used for both short and long term predictions, i.e., by combining the predictions from models with various timescales to produce a control strategy. It may also be possible to combine production control with short-term ventilation control, i.e. by increasing ventilation

quantity to the district if a model predicts that the statutory methane limit will be exceeded for a short period of time.

9.3 Recommendations for Further Research

Three major difficulties were experienced during the course of the research. They were related to data transference and analysis and these were identified as areas necessary for further research. Current versions of the software used in colliery MINOS systems do not allow easy transfer of data from one computer to another. This poses a serious problem to the availability of data that can be used for research purposes and future environmental control. After data was transferred considerable effort was needed to transform the data into a format suitable for time series analysis and this needs to be improved for serious use of environmental data. For the purpose of time series analysis it is desirable to have data that is in a digital format, i.e. that has been monitored by an electronic device, therefore, barometric pressure should be so monitored and a method for electronically monitoring production devised. The software used to perform the analysis was adequate for the building of univariate models but had difficulty in coping with lengthy time series and large numbers of model coefficients for multivariate analysis. Investigations should be made to obtain software more dedicated to the determination of multivariate transfer function models.

The time series data was from one colliery and although in theory the statistical technique can be applied to any circumstance it is necessary to obtain data from other collieries to confirm that time series techniques are suitable for general application. This will help potential users of environmental and production control to gain confidence in the ability of time series analysis to produce accurate predictions.

This thesis has considered the popular Box-Jenkins approach to time series analysis but in recent years statistical science has developed newer techniques that are beginning to be applied to practical problems. These new methods such as full information maximum likelihood model estimation and Neural Networks should be examined for their ease of use and suitability to planning and control. The details of how time series models could be implemented is left to future researchers.

ACKNOWLEDGEMENT

The author would like to thank:

Eur Ing Professor D. J. Hodges and Eur Ing Professor D. Potts for the provision of the facilities of the Department of Mining Engineering at the University of Nottingham.

Ian Longson and Mick Tuck, my supervisors, for their help, guidance and patience.

Dr. J. R. Brown for the use of his Macintosh SE computer and printer.

Mr. K. Fidler, Manager of Thoresby Colliery, Nottinghamshire Group for his permission to use colliery data and photographs. Grateful thanks also to Mr. Day, Deputy Manager, Mr. K. Mosely, Ventilation Officer and Mr. T. Clarke, Control Room Officer.

Mr. A. J. Pickering and his staff of the Ventilation Department, Nottinghamshire Area, Edwinstowe and Mr. P. Smith of HQTD, British Coal Corporation, Bretby.

Mr. P. Riley and other staff of the Cripps Computing Centre for their assistance with computing problems.

Dr. Göktay Ediz for his help with the initial chapters.

Mr. D. De Barr and the other technical staff of the Mining Department who have helped me considerably over the years on items of a practical nature.

SERC for providing the financial support needed to carry out the research.

Thanks also to Suzanne and Josephine.

REFERENCES

1. MINING JOURNAL, London, Vol. 318, No. 8163, March 6, 1992, ISSN 0026-5225.

2. THE MINE VENTILATION SOCIETY OF SOUTH AFRICA, " Environmental Engineering in South African Mines", Cape Town, 1982, ISBN 0-620-06258-4.

3. COWARD, H. F., JONES, G. W., " Limits of Flammability of Gases and Vapours ", US Bureau of Mines, Bulletin No. 503, 1952.

4. HARGRAVES, A.J., " Keynote Address: Background to Seam Gas Drainage in Australia ", Proceedings of the Symposium on Seam Gas Drainage with Particular Reference to the Working Seam, Australasian Institute of Mining and Metallurgy, pp. 21, May 1982.

5. RICE, D.D., CLAYPOOL, G.E., " Generation, Accumulation and Resource Potential of Biogenic Gas ", American Association of Petroleum Geologists, Vol. 65, pp. 5-25, 1981.

6. EDWARDS, J.S., WHITTAKER, B.N., DURUCAN, S., " Methane Hazards in Tunnelling Operations ", Proceedings of the 5th International Symposium, Tunnelling '88, Institute of Mining and Metallurgy, p. 98, April 1988.

7. HEDBURG, H.D., " Methane Generation and Petroleum Migration ", In: Problems of Petroleum Migration, Eds. Roberts, W.H. and Cordell, R.J., The American Association of Petroleum Geologists, Studies in Geology No. 10, pp. 179-206, 1980.

8. CREEDY, D.P., " Geological Sources of Methane In Relation to Surface and Underground Hazards ", Symposium: Methane - Facing the Problems, pp. 1.4.1-1.4.8., September 1989.

9. RICHARDS, M.J., " An Investigation into the Movement of Firedamp from the Strata in the Region of Working Longwall Faces ", Ph.D. Thesis, University of Nottingham, Department of Mining Engineering, 1975.

10. DUNMORE, R., " Gas Flow Through Underground Strata ", Mining Engineer, Vol. 128, No. 100, pp. 193-199, January 1969.

11. GUNTHER, J., " Investigation of the Relationship Between Coal and the Gas Contained in It ", Revue de L'industrie Minerale, Vol. 47, No.10, p.693, 1965.

12. CURL, S.J., " Methane Prediction in Coal Mines ", IEA COAL RESEARCH, p. 79, 1978, ISBN 92-9029-018-8.

13. CREEDY, D.P., KERSHAW, S., " Firedamp Prediction - A Pocket Calculator Solution ", Mining Engineer, Vol. 147, No. 317, pp. 377-379, February 1988.

14. YEREBASMAZ, G., " An Investigation into the Flow of Methane Through Coal Samples ", M.Phil. Thesis, University of Nottingham, Department of Mining Engineering, 1981.

15. PATCHING, T.H., " The Retention and Release of Gas in Coal - A Review ", CIM Bulletin, Vol. 63, No. 703, pp. 1302-1308, 1970.

16. JOLLY, D.C., MORRIS, L.H., HINSLEY, F.B., " An Investigation into the Relationship Between the Methane Sorption Capacity of Coal and Gas Pressure", Mining Engineer, Vol. 127, No. 94, pp. 539-548, July 1968.

17. BIELICKI, R.J., PERKINS, J.H., KISSELL, F.N., " Methane Diffusion Parameters for Sized Coal Particles. A Measuring Apparatus and Some Provisional Results ", US Bureau of Mines, Report of Investigation, RI 7697, p. 15, 1972.

18. LANGMUIR, I., " The Constitution and Fundamental Properties of Solids and Liquids ", Journal of the American Chemical Society, Vol. 38, pp. 2221-2295, 1916.

19. BRUNNAUER, S., EMMET, P.H., TELLER, E., " Adsorption of Gases in Multi-Molecular Layers ", Journal of the American Chemical Society, Vol. 60, pp. 309-319, 1938.

20. KEEN, T.F., " The Simulation of Methane Flow in Carboniferous Strata ", Ph.D. Thesis, University of Nottingham, Department of Mining Engineering, October 1977.

21. MOTT, R.A., " The Origin and Composition of Coals ", Fuel (London), Vol. 22, pp. 20-27, 1943.

22. KIM, A.G., " Estimating Methane Content of Bituminous Coalbeds from Adsorption Data ", US Bureau of Mines, Report of Investigation, RI 8245, p. 22, 1977.

23. JOUBERT, J.I., GREIN, C.T., BIENSTOCK, D., " Effect of Moisture on the Methane Capacity of American Coals ", Fuel, Vol. 53, No. 3, pp. 186-191, July 1974.

24. HARPALANI, S., McPHERSON, M.J., " An Experimental Investigation to Evaluate Gas Flow Characteristics of Coal ", 4th International Mine Ventilation Congress, Queensland, pp. 175-182, July 1988.

25. DUNMORE, R., " A Theory of Emission of a Mixture of Methane and Ethane from Coal ", International Mine ventilation Congress, p. 7, Johannesburg, 1975.

26. KISSELL, F.N., EDWARDS, J.C., " Two-Phase Flow in Coalbeds ", US Bureau of Mines, Report of Investigation, RI 8066, p. 22, 1975.

27. OWILI-EGER, A.S.C., RAMANI, R.V., " Mathematical Modelling of Methane Flow in Mines ", Proceedings of the 12th International Symposium on the Application of Computing and Mathematics in the Mining Industry, Colorado School of Mines, Vol. 2, Section G, pp. 218-229, 1974.

28. VUTUKURI, V.S., LAMA, R.D., " Environmental Engineering in Mines ", Cambridge University Press, p. 504, 1986, ISBN 0-521-246059.

29. MORDECAI, M., MORRIS, L.H., " The Effect of Stress on the Flow of Gas Through Coal Measure Strata ", Transactions of the Institute of Mining Engineers, Vol.133, pp. 435-443, 1974.

30. MINES AND QUARRIES ACTS (1954), " The Law Relating to Safety and Health in Mines and Quarries. Part 1 The Act, Part 2 Regulations " HMSO, 1954.

31. KISSELL, F.N., McCULLOCH, C.M., ELDER, C.H., " The Direct Method of Determining Methane Content of Coalbeds for Ventilation Design ", US Bureau of Mines, Report of Investigation, RI 7767, p. 22, May 1973.

32. FISECKI, M.Y., BARRON, K., " Methane Pressure and Flow Measurements in Coal and Surrounding Strata ", CIM Bulletin, October 1975.

33. GLUCKAUF, " Firedamp Drainage ", pp. 39, 1980, ISBN 3-7739-0254-9.

34. MCPHERSON, M.J., " The Occurrence of Methane in Mine Workings ", Journal of the Mine Ventilation Society of South Africa, Vol. 28, No. 8, pp. 118-128, 1975.

35. MORDECAI, M., " An Investigation into the Effect of Stress on Permeability of Rock Taken from Carboniferous Strata ", Ph.D. Thesis, University of Nottingham, Department of Mining Engineering, 1971.

36. DURUCAN, S., " An Investigation into the Stress-Permeability Relationships of Coals and Flow Patterns Around Working Longwall Faces ", Ph.D. Thesis, University of Nottingham, Department of Mining Engineering, October 1981.

37. DUNMORE, R., " Towards a Method of Prediction of Firedamp Emission for British Coal Mines ", Proceedings of the 2nd International Mine Ventilation Congress, Reno, Nevada, USA, pp. 354, 1980.

38. AIREY, E.M., " Gas Emission from Broken Coal: An Experimental and Theoretical Investigation ", International Journal of Rock Mechanics and Mineral Science, Vol. 5, pp. 475-494, 1968.

39. DUNMORE, R., " Prediction of Gas Emission from Longwall Faces ", Mining Engineer, Vol. 140, No., 233, pp. 565-572, February 1981.

40. NATIONAL COAL BOARD, " Prediction of Firedamp Emission ", Technical Report, No. 2., Mining Research and Development Establishment, Bretby, ECSC Research Project 7220-AC/801 (October 1976 - March 1977).

41. BRUYET, B., " Variations de Teneur dans les Retours d'Air de Tailles ", Cerchar Documents Techniques, 3, pp. 85-100.

42. NATIONAL COAL BOARD, " Prediction of Firedamp Emission ", Technical Report, No. 4., Mining Research and Development Establishment, Bretby, ECSC Research Project 7220-AC/801 (October 1977 - March 1978).

43. OWILI-EGER, A.S.C., RAMANI, R.V., " Mathematical Modelling of Methane Flow in Mines ", Proceedings of the 12th International Symposium on the Application of Computing and Mathematics in the Mining Industry, Colorado School of Mines, Vol. 2, Section G, pp. 218-229, 1974.

44. O'SHAUGHNESSY, S.M., " The Computer Simulation of Methane Flow Through Strata Adjacent to Working Longwall Coal Faces ", Ph.D. Thesis, University of Nottingham, Department of Mining Engineering, May 1980.

45. KING, G.R., ERTEKIN, T.M., " A Survey of Mathematical Models Related to Methane Production from Coal Seams ", Parts I and II, Proceedings of the 1989 Coalbed Methane Symposium, The University of Alabama/Tuscaloosa, pp. 125-155, April 1989.

46. EDIZ, I.G., EDWARDS, J.S., " A Numerical Prediction Method of Methane Flow Through Strata Adjacent to a Working Longwall Coal Face ", to be published.

47. EDIZ, I.G., " Numerical Simulation of Methane Flow Through Strata Adjacent to a Working Longwall Coal Face ", Ph.D. Thesis, University of Nottingham, Department of Mining Engineering, October 1991.

48. EDIZ, I.G., " Private Communications ", Department of Mining Engineering, University of Nottingham, May 1991.

49. KAFFANKE, H., " Medium Term Prediction of the Methane Content in the Return Ventilation Flow of Workings ", Information Symposium on Methane, Climate, Ventilation in the Coal Mines of the European Communities, Vol. 1, pp. 170-192, Luxembourg, November 1980.

50. O'DONOVAN, T.M., " Short Term Forecasting ", John Wiley, 1983, ISBN 0-471-90013-3.

51. JENKINS, G.M., WATTS, D.G., " Spectral Analysis and Its Applications ", Holden Day, San Francisco, 1968.

52. YULE, G.U., " On A Method of Investigating Periodicities in Disturbed Series, With Special Reference to Wolfer's Sunspot Numbers ", Phil. Trans., A226, pp. 267, 1927.

53. GRANGER, C.W.J., NEWBOLD, P., " Forecasting Economic Time Series ", Academic Press, London, 1977, ISBN 0-12-295150-6.

54. McLEOD, G., " Box Jenkins in Practice ", Vol. 1, Time Series Library, Lancaster, 1983, ISBN 0-9506423-2-0.

55. BOX, G.E.P., JENKINS, G.M., " Time Series Analysis : forecasting and control ", Holden Day, San Francisco, 1970, ISBN 0-8162-1104-3.

56. QUENOUILLE, M.H., " Approximate Tests of Correlation in Time Series ", J. Roy. Stat. Soc., B11, pp. 68 - 84, 1949.

57. SPSS Inc., " SPSS-X Trends™ ", USA, 1988, ISBN 0-918469-52-X.

58. LJUNG, G.M., BOX, G.E.P, " On a Measure of Lack of Fit in Time Series Models ", Biometrika, Vol 65, No. 2, pp. 67 - 72, 1978.

59. NATIONAL COAL BOARD, Mining Environment Bulletin, No. 2, pp. 8, November 1978, ISSN 0140-3451.

60. FINK, Z.J., ADLER, D.T., " Continuous Monitoring System for Mine Gas Concentrations Using Tube-bundles ", US Bureau of Mines, Report of Investigation, RI 8060, 1975.

61. HORMOZDI, I., " A Review of the Developments in Automatic Monitoring and Control of Mine Atmospheric Environment ", Mineral Engineering Department, Pennsylvania State University, USA, 1979.

62. NATIONAL COAL BOARD, " Application of Computer-Based Environmental Monitoring ", Mining Research and Development Establishment, Bretby, ECSC Research Project 7220-AS/807 (January 1978 - December 1982), 1983.

63. KRZYSTOLIK, P., SWIERGOT, F., " Investigations on the Influence of Some Elements Determining the Response Time of CH_4 Detectors ", Proceedings of the 21st International Conference of Safety in Mines Research Institutes, Sydney, Australia, 10/21-25/1985.

64. FISHER, T.J., COHEN, A.F., " Sensor Location Strategy for Minewide Environmental Monitoring ", Proceedings of the 21st International Conference of Safety in Mines Research Institutes, Sydney, Australia, 10/21-25/1985, pp. 393-398.

65. WELSH, J.H., COHEN, A.E., CHILTON, J.E., " Suggested Minimum Performance Specifications for Underground Coal Mine Environmental Monitoring Systems ", US Bureau of Mines, Information Circular 9157, pp. 39, 1987.

66. MORRIS, I.H., GRAY, G.W., " Environmental Monitoring and Control in the United Kingdom ", International Conference on Remote Control & Monitoring in Mining, United Kingdom, pp. 10.1 - 10.16, 1977.

67. SMITH, P., Private Communications, Headquarters Technical Department, British Coal Corporation, Bretby, December 1990.

68. CLARKE, T., Private Communications, Thoresby Colliery, 1989 - 1991.

APPENDIX 1

EXAMPLE OF A CONTROL CENTRE RECORD SHEET

THORESBY COLLIERY — CONTROL CENTRE RECORD

DAY WEDNESDAY
DATE 27-2-91
SHIFT NTS

RECORD OF ALL DELAYS

No	From	To	Mins	Reason for Delays
	3.05	—		CHANGING BOM BELT + PULL
	—	—		IN Y/B + D.NTING
	FINAL	—		CHANGED BOM + PULLED IN
				4m - D.NTED 4m W/SIDE +
				3m T/SIDE (D.NT NOW 10m
				FROM FACE) DROPPED ALL
				LIGHTS IN 9/B SNICKET
	12.30	12.55	25	4/G SHUFFLE + 9/B BINGHAM NET
	—	—	25	M/C ½ SHUFFLED
	—	—	—	4/G READY
	E.O.S.	—		WORN G/S GRADE PAN RACE
	—	—		AT 4/G END - NEW ONE IN ON
	—	—		DAYS
	9.05	9.55	50	4/G SHUFFLE + 9.30 9/B NOT TENS
	1.00	1.20	20	LUMP AT 4/G
	—	—	70	M/C TO SHUFFLE
	—	—	—	9/G READY

Panel & Length	M/c Start Pos'n		M/c CYCLES 1	2	3	4	5	6	M/c Final Pos'n	Shift Mtrs Cut	24 Hr. Mtrs Cut	Shift Strips	Week Strips
114'	AT	Start							AT	—		—	
	Mtrs to	Finish							Mtrs to				
263	4/G	Mtrs							4/G		—		—
		Start											
	Mtrs to	Finish							Mtrs to				
		Mtrs											
119'	50	Start	9.55	10.55	11.50	12.55	2.05	3.10	AT	1450		6	
	Mtrs to	Finish	10.40	11.40	12.30	1.45	2.55	4.00	Mtrs to				
250	4/G	Mtrs	200	250	250	250	250	250	S/G		3620		36
		Start											
	Mtrs to	Finish							Mtrs to				
		Mtrs											
145'	160	Start	8.35	9.55	11.00	12.10	1.40	2.55					
	Mtrs to	Finish	9.05	10.45	11.45	1.25	2.30	3.45	Mtrs to				
250	4/G	Mtrs	90	250	250	250	250	250					
		Start	4.00						AT	1590		7	
	Mtrs to	Finish	4.45						Mtrs to				
		Mtrs	250						4/G		3190		29
									TOTAL	3040	6810		

SHAFT SCORES

Time	11.00	12.00	1.00	2.00	3.00	4.00	5.00	6.00
Cars Wound In No. 1							29	MAT
Cars Wound Out No.1							35	DIRT
Skips Wound	49	49/98	49/147	48/195	49/244	48/292	49/341	49/390

SKIP DAYS

	CH4 119'	145'
10.00	1.26%	0.60%
11.00	1.25%	0.81%
12.00	1.40%	0.78%
1.00	1.32%	0.84%
2.00	1.35%	0.94%
3.00	1.50%	0.79%
4.00	1.52%	0.92%
5.00	1.38%	0.66%
6.00	—	

BELT DAYS

BUNKER LEVELS AND CAPACITY

940

High Hazle	450T	3/4	FULL	FULL	FULL	FULL	FULL	FULL	FULL
Parkgate Vertical	1000T	50%	60%	80%	80%	80%	80%	80%	80%
Parkgate Inseam	450T	—	—	—	—	—	—	—	—
Parkgate Incline	1000T	30%	40%	40%	50%	70%	80%	80%	30%

DEVELOPMENTS

46' M EATON 1.30 SET No 8 H/BACK - C/DOWN FOR No 9 (EXCESSIVE BAGGING) FINAL CUT/SET No 9 - REMOVED LEG & SHEETS - STARTED TO CUT FOR No 10

41' D SHAW 9.45 EXT EIMCO CABLE TO GET SAME TO 4/G & DINT T/E OF BELT FINAL EXT EIMCO CABLE 64m & GOT M/C 30m DOWN 4/G - BOLTED DOWN BELT T/E. TOOK 3L M/R DOWN STARTED TO DINT - GOT 6 PROPS + 3 RAILS FROM 4/G JUNT & LOADED FOR 145 - BUILT W/CH - PANZER PLOUGH & T/COVER REMOVED

40' B VERNON 11.30 CUTTING 1.25 CUT/SET & BOLT 1 RING 3.10 CUT/SET & BOLT 2ND FINAL CUT/SET & BOLT 2 RINGS CUT/SET 3RD RING - STARTED TO BOLT SAME - THIS TO COMP

WI K SHILTON 9.30 REMOVE RING LFT/BRKT OFF SWR K FOR 122' + 9/UNT FOR CH/BASE FINAL GOT CHOCK BASES DOWN & BOTH CHOCK PUT IN POSITION - GOT RING L/BRKT OFF M/C & SENT TO 122'

DRIFT B MELLINS 9.10 RING & SHOULDER TO SHEET & 11.00 CUTTING

115' A PICKERING - 9.15 LEG TO SET AT No 19 + BREAK OUT G/F 1/3 TO PUSH OVER 1.30 SET LEG TO No 19 - 9/OUT & REMOVED LAST 3 GLENIS PINS AT 4/G END - COMP REMOVING BEAM PLATES & BROKE G/FEEDER CHAIN BROKE OUT 4/STN READY TO ADV FINAL SET RAM UP & PUSHED G/FEEDER OVER 12" CHOCK FLEXIS TO EXT - PULLED 4/STN TO No 15 CH LAID 3PR TRACK - 4/STN TO MOVE 2 MORE CHOCKS WHEN AFC PUSHED OVER POWER OFF ALL SHIFT B.M.S TRIP

123' J RYDINGS 10.00 FITTING FLIGHTS 12.00 20 FLIGHTS FITTED - NOW INSTALLING PANS FINAL UNLOADED & FITTED 10 PANS & THREADED BOTTOM CHAIN (TOTAL 128)

122 4/G C BROWN 9.30 CONT TO CUT DOWN 9.40/E.O.S. M/C CENTRE CHAIN OUT OF RACE FINAL DRILLED OUT 8 HOLES IN RACE & TAPPED (6 HOLES EX TO DO)

122 3/4 C BROWN 9.30 CONT TO CUT DOWN 12.00 SET BOW, DOWN TRIMMING FOR LEGS 1.30 M/C IN SWILLEY, UNABLE TO TRACK BACK FINAL M/C STILL IN SWILLEY - 5" RAM REQ - SET 1 LEG TO BEAM

APPENDIX 2

RESIDUAL CROSS-CORRELATIONS OF ENVIRONMENTAL AND PRODUCTION DATA

Cross Correlations: RESIDUALAD Error for AD from ARIMA, MOD_2 NOCON
RESIDUALMD Error for MD from ARIMA, MOD_1 NOCON

Cross Stand.

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
0	-.008	.037	.*.
1	-.021	.037	.*.
2	.005	.037	.*.
3	-.018	.037	.*.
4	-.014	.037	.*.
5	-.002	.037	.*.
6	-.026	.037	*|.
7	-.023	.037	.*.
8	.044	.038	. |*.
9	.005	.038	. * .
10	.045	.038	. |*.
11	.006	.038	. * .
12	-.032	.038	.*| .
13	-.056	.038	.*| .
14	-.020	.038	. * .
15	.013	.038	. * .
16	-.007	.038	. * .
17	.016	.038	. * .
18	.012	.038	. * .
19	.011	.038	. * .
20	-.030	.038	.*| .
21	-.026	.038	.*| .
22	.014	.038	. * .
23	.000	.038	. * .
24	.038	.038	. |*.
25	.015	.038	. * .

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 720 Computable 0-order correlations: 719

Figure A2.1 Residual Cross-correlation of Methane Concentration and Air Velocity (One Day Series).

Cross Correlations: RESIDUALA Error for AIRAPRIL from ARIMA, MOD_2 NOCON
RESIDUALM Error for METHAPRIL from ARIMA, MOD_1 NOCON

Lag	Cross Corr.	Stand. Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
0	.001	.007	*
1	.000	.007	*
2	.001	.007	*
3	-.001	.007	*
4	.002	.007	*
5	.002	.007	*
6	.001	.007	*
7	.001	.007	*
8	.001	.007	*
9	.001	.007	*
10	-.002	.007	*
11	.001	.007	*
12	.000	.007	*
13	.001	.007	*
14	-.001	.007	*
15	-.002	.007	*
16	.000	.007	*
17	-.002	.007	*
18	-.001	.007	*
19	-.003	.007	*
20	.000	.007	*
21	.002	.007	*
22	.001	.007	*
23	.003	.007	*
24	.002	.007	*
25	.003	.007	*

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 21600 Computable 0-order correlations: 21598

Figure A2.2 Residual Cross-correlation of Methane Concentration and Air Velocity (Original Series).

Cross Correlations: RESIDUALBD Error for BARD from ARIMA, MOD_2 NOCON
RESIDUALMD Error for METHD from ARIMA, MOD_1 NOCON

Cross Stand.

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
0	-.045	.037	*|.
1	-.040	.037	*|.
2	-.060	.037	*|.
3	-.057	.037	*|.
4	.028	.037	.|*
5	.034	.037	.|*
6	-.006	.037	.*.
7	.008	.037	.*.
8	-.019	.038	. * .
9	-.014	.038	. * .
10	-.013	.038	. * .
11	.016	.038	. * .
12	.038	.038	. |*.
13	-.031	.038	.*| .
14	-.056	.038	.*| .
15	-.065	.038	.*| .
16	-.045	.038	.*| .
17	-.008	.038	. * .
18	-.003	.038	. * .
19	-.043	.038	.*| .
20	.053	.038	. |*.
21	-.015	.038	. * .
22	.006	.038	. * .
23	-.016	.038	. * .
24	.024	.038	. * .
25	-.020	.038	. * .

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 720 Computable 0-order correlations: 719

Figure A2.3 Residual Cross-correlation of Methane Concentration and Barometric Pressure (One Day Series).

```
Cross Correlations: RESIDUALS  Error for STATAPRIL from ARIMA, MOD_1 NOCON
                    RESIDUALQ  Error for BM2HAPRIL from ARIMA, MOD_1 NOCON

      Cross Stand.

Lag   Corr.    Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1
                    +----+----+----+----+----+----+----+----+----+
  0  -.017    .007                       *
  1  -.008    .007                       *
  2   .028    .007                       .*
  3   .016    .007                       *
  4   .006    .007                       *
  5   .008    .007                       *
  6  -.008    .007                       *
  7  -.023    .007                       *
  8  -.028    .007                      *.
  9  -.012    .007                       *
 10  -.009    .007                       *
 11  -.010    .007                       *
 12  -.005    .007                       *
 13  -.003    .007                       *
 14  -.003    .007                       *
 15  -.005    .007                       *
 16  -.004    .007                       *
 17   .000    .007                       *
 18  -.004    .007                       *
 19  -.002    .007                       *
 20  -.001    .007                       *
 21  -.001    .007                       *
 22  -.002    .007                       *
 23  -.002    .007                       *
 24  -.002    .007                       *
 25  -.001    .007                       *

Plot Symbols: Autocorrelations *  Two Standard Error Limits .
Total cases: 21600  Computable 0-order correlations: 21599
```

Figure A2.4 Residual Cross-correlation of Drainage Methane Concentration and Static Pressure (Original Series).

Cross Correlations: RESIDUALSD Error for STATD from ARIMA, MOD_1 NOCON
RESIDUALQD Error for BM2HD from ARIMA, MOD_1 NOCON

Cross Stand.

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
0	-.186	.037	***.|.
1	-.076	.037	*.|.
2	.361	.037	.|.******
3	-.104	.037	*.|.
4	-.030	.037	*|.
5	-.054	.037	*|.
6	-.119	.037	*.|.
7	-.328	.037	******.|.
8	-.349	.038	*****.*| .
9	.109	.038	. |**
10	-.015	.038	. * .
11	.067	.038	. |*.
12	.046	.038	. |*.
13	.068	.038	. |*.
14	.025	.038	. * .
15	.040	.038	. |*.
16	.003	.038	. * .
17	.008	.038	. * .
18	.025	.038	. |*.
19	-.004	.038	. * .
20	-.006	.038	. * .
21	.027	.038	. |*.
22	-.009	.038	. * .
23	.023	.038	. * .
24	.013	.038	. * .
25	.002	.038	. * .

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 720 Computable 0-order correlations: 719

Figure A2.5 Residual Cross-correlation of Drainage Methane Concentration and Static Pressure (One Day Series).

```
Cross Correlations: RESIDUALD   Error for DIFFAPRIL from ARIMA, MOD_1 NOCON
                   RESIDUALQ   Error for BM2HAPRIL from ARIMA, MOD_2 NOCON

      Cross Stand.

Lag   Corr.    Err. -1   -.75   -.5  -.25    0    .25   .5    .75   1
                    +----+----+----+----+----+----+----+----+----+
  0    .000    .007                        *
  1    .002    .007                        *
  2    .000    .007                        *
  3    .002    .007                        *
  4    .001    .007                        *
  5    .002    .007                        *
  6    .003    .007                        *
  7    .002    .007                        *
  8    .004    .007                        *
  9    .003    .007                        *
 10    .000    .007                        *
 11    .000    .007                        *
 12    .001    .007                        *
 13    .001    .007                        *
 14    .000    .007                        *
 15    .000    .007                        *
 16   -.001    .007                        *
 17   -.002    .007                        *
 18   -.001    .007                        *
 19   -.001    .007                        *
 20    .000    .007                        *
 21    .001    .007                        *
 22    .002    .007                        *
 23    .001    .007                        *
 24    .001    .007                        *
 25    .000    .007                        *

Plot Symbols: Autocorrelations *  Two Standard Error Limits .
Total cases: 21600  Computable 0-order correlations: 21553
```

Figure A2.6 Residual Cross-correlation of Drainage Methane Concentration and Differential Pressure (Original Series).

Cross Correlations: RESIDUALDD Error for DIFFD from ARIMA, MOD_1 NOCON
RESIDUALQD Error for BM2HD from ARIMA, MOD_1 NOCON

Cross Stand.

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
0	-.208	.037	***.|.
1	.551	.037	.|.**********
2	-.192	.037	***.|.
3	-.056	.037	*|.
4	-.006	.037	.*.
5	-.074	.037	*|.
6	-.177	.037	***.|.
7	-.299	.037	*****.|.
8	-.230	.038	***.*| .
9	.068	.038	. |*.
10	.009	.038	. * .
11	.052	.038	. |*.
12	.058	.038	. |*.
13	.058	.038	. |*.
14	.039	.038	. |*.
15	.024	.038	. * .
16	.006	.038	. * .
17	.016	.038	. * .
18	.003	.038	. * .
19	-.001	.038	. * .
20	.013	.038	. * .
21	.008	.038	. * .
22	.011	.038	. * .
23	.016	.038	. * .
24	.005	.038	. * .
25	.015	.038	. * .

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 720 Computable 0-order correlations: 719

Figure A2.7 Residual Cross-correlation of Drainage Methane Concentration and Differential Pressure (One Day Series).

Cross Correlations: RESIDUALS10AV Error for STAT10AV from ARIMA, MOD_1 NOCON
RESIDUALQ10AV Error for BM2H10AV from ARIMA, MOD_1 NOCON

Lag	Cross Corr.	Stand. Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
0	-.139	.015	**.|.
1	-.342	.015	******.|.
2	-.167	.015	**.|.
3	.098	.015	.|.*
4	-.007	.015	.*.
5	.033	.015	.|*
6	.032	.015	.|*
7	-.054	.015	*|.
8	-.019	.015	.*.
9	.024	.015	.*.
10	.023	.015	.*.
11	.026	.015	.|*
12	.019	.015	.*.
13	.012	.015	.*.
14	.002	.015	.*.
15	.022	.015	.*.
16	.008	.015	.*.
17	.026	.015	.|*
18	.007	.015	.*.
19	.000	.015	.*.
20	-.003	.015	.*.
21	.010	.015	.*.
22	.018	.015	.*.
23	-.015	.015	.*.
24	.034	.015	.|*
25	.026	.015	.|*

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 4320 Computable 0-order correlations: 4319

Figure A2.8 Residual Cross-correlation of Drainage Methane Concentration and Static Pressure (10-Minute Average Series).

Cross Correlations: RESIDUALD10AV Error for DIFF10AV from ARIMA, MOD_1 NOCON
RESIDUALQ10AV Error for BM2H10AV from ARIMA, MOD_1 NOCON

Lag	Cross Corr.	Stand. Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
0	-.080	.015	*.|.
1	-.426	.015	********.|.
2	-.114	.015	*.|.
3	.073	.015	.|*
4	-.042	.015	*|.
5	.004	.015	.*.
6	.006	.015	.*.
7	-.046	.015	*|.
8	-.024	.015	.*.
9	.025	.015	.*.
10	.013	.015	.*.
11	.016	.015	.*.
12	.009	.015	.*.
13	.009	.015	.*.
14	-.003	.015	.*.
15	.006	.015	.*.
16	-.004	.015	.*.
17	.012	.015	.*.
18	-.002	.015	.*.
19	-.010	.015	.*.
20	-.011	.015	.*.
21	.003	.015	.*.
22	.000	.015	.*.
23	-.022	.015	.*.
24	.024	.015	.*.
25	.014	.015	.*.

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 4320 Computable 0-order correlations: 4317

Figure A2.9 Residual Cross-correlation of Drainage Methane Concentration and Differential Pressure (10-Minute Average Series).

Cross Correlations: RESIDUALSAV Error for STATAV from ARIMA, MOD_1 NOCON
RESIDUALQAV Error for BM2HAV from ARIMA, MOD_1 NOCON

Cross Stand.

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
			+----+----+----+----+----+----+----+----+
0	-.553	.037	**********.|.
1	-.187	.037	***.|.
2	.038	.037	.|*
3	.042	.037	.|*
4	.052	.037	.|*
5	.043	.037	.|*
6	.014	.037	.*.
7	-.026	.037	*|.
8	.006	.038	. * .
9	-.015	.038	. * .
10	.021	.038	. * .
11	-.002	.038	. * .
12	.062	.038	. |*.
13	.090	.038	. |**
14	-.013	.038	. * .
15	-.027	.038	.*| .
16	-.088	.038	**| .
17	-.040	.038	.*| .
18	.097	.038	. |**
19	.075	.038	. |**
20	-.017	.038	. * .
21	-.008	.038	. * .
22	.134	.038	. |*.*
23	-.027	.038	.*| .
24	-.027	.038	.*| .
25	-.021	.038	. * .

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 720 Computable 0-order correlations: 719

Figure A2.10 Residual Cross-correlation of Drainage Methane Concentration and Static Pressure (Hourly Average Series).

Cross Correlations: RESIDUALDAV Error for DIFFAV from ARIMA, MOD_1 NOCON
RESIDUALQAV Error for BM2HAV from ARIMA, MOD_1 NOCON

Cross Stand.

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
			+----+----+----+----+----+----+----+----+
0	-.517	.037	*********.|.
1	-.207	.037	***.|.
2	.007	.037	.*.
3	-.030	.037	*|.
4	.004	.037	.*.
5	.025	.037	.*.
6	.024	.037	.*.
7	.001	.037	.*.
8	.005	.038	. * .
9	-.010	.038	. * .
10	.013	.038	. * .
11	-.002	.038	. * .
12	.054	.038	. |*.
13	.105	.038	. |**
14	-.056	.038	.*| .
15	-.014	.038	. * .
16	-.037	.038	.*| .
17	-.038	.038	.*| .
18	.059	.038	. |*.
19	.060	.038	. |*.
20	-.045	.038	.*| .
21	.021	.038	. * .
22	.142	.038	. |*.*
23	-.009	.038	. * .
24	-.004	.038	. * .
25	-.048	.038	.*| .

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 720 Computable 0-order correlations: 719

Figure A2.11 Residual Cross-correlation of Drainage Methane Concentration and Differential Pressure (Hourly Average Series).

Cross Correlations: RESIDUALP Error for PRODAPRIL from ARIMA, MOD_2 NOCON
RESIDUALQ Error for BM2HAPRIL from ARIMA, MOD_1 NOCON

Transformations: difference (1)

Lag	Cross Corr.	Stand. Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
0	.011	.007	*
1	-.011	.007	*
2	.003	.007	*
3	.001	.007	*
4	-.001	.007	*
5	-.005	.007	*
6	.005	.007	*
7	.001	.007	*
8	-.007	.007	*
9	.002	.007	*
10	.002	.007	*
11	.013	.007	*
12	-.011	.007	*
13	.005	.007	*
14	.005	.007	*
15	.003	.007	*
16	-.004	.007	*
17	-.001	.007	*
18	-.010	.007	*
19	-.003	.007	*
20	.017	.007	*
21	-.006	.007	*
22	.006	.007	*
23	-.017	.007	*
24	.003	.007	*
25	-.001	.007	*

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 21600 Computable 0-order correlations after differencing: 21598

Figure A2.12 Residual Cross-correlation of Drainage Methane Concentration and Production (Original Series).

Cross Correlations: RESIDUALPD Error for PRODD from ARIMA, MOD_2 NOCON
RESIDUALQD Error for BM2HD from ARIMA, MOD_1 NOCON

Lag	Cross Corr.	Stand. Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
0	.000	.037	.*.
1	-.022	.037	.*.
2	-.075	.037	*.|.
3	.082	.037	.|.*
4	-.004	.037	.*.
5	-.016	.037	.*.
6	.013	.037	.*.
7	-.025	.037	*|.
8	-.016	.038	. * .
9	.033	.038	. |*.
10	.016	.038	. * .
11	.017	.038	. * .
12	-.014	.038	. * .
13	.002	.038	. * .
14	.002	.038	. * .
15	.003	.038	. * .
16	.027	.038	. |*.
17	.057	.038	. |*.
18	-.104	.038	**| .
19	-.009	.038	. * .
20	-.019	.038	. * .
21	.008	.038	. * .
22	.023	.038	. * .
23	-.020	.038	. * .
24	-.039	.038	.*| .
25	-.013	.038	. * .

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 720 Computable 0-order correlations: 719

Figure A2.13 Residual Cross-correlation of Drainage Methane Concentration and Production (One Day Series).

Cross Correlations: RESIDUALP10AV Error for PROD10AV from ARIMA, MOD_2 NOCON
RESIDUALQ10AV Error for BM2H10AV from ARIMA, MOD_1 NOCON

Cross Stand.

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
0	-.004	.015	.*.
1	-.003	.015	.*.
2	.014	.015	.*.
3	.019	.015	.*.
4	-.039	.015	*|.
5	.000	.015	.*.
6	.028	.015	.|*
7	.036	.015	.|*
8	.017	.015	.*.
9	.011	.015	.*.
10	.027	.015	.|*
11	.019	.015	.*.
12	.002	.015	.*.
13	.021	.015	.*.
14	.021	.015	.*.
15	.024	.015	.*.
16	.022	.015	.*.
17	.018	.015	.*.
18	.032	.015	.|*
19	-.005	.015	.*.
20	.016	.015	.*.
21	-.005	.015	.*.
22	-.006	.015	.*.
23	.007	.015	.*.
24	-.004	.015	.*.
25	.016	.015	.*.

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 4320 Computable 0-order correlations: 4319

Figure A2.14 Residual Cross-correlation of Drainage Methane Concentration and Production (10-Minute Average Series).

Cross Correlations: RESIDUALPAV Error for PRODAV from ARIMA, MOD_2 NOCON
RESIDUALQAV Error for BM2HAV from ARIMA, MOD_1 NOCON

Lag	Cross Corr.	Stand. Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
0	-.004	.037	.*.
1	.056	.037	.|*
2	.055	.037	.|*
3	.067	.037	.|*
4	.012	.037	.*.
5	-.010	.037	.*.
6	-.025	.037	*|.
7	-.019	.037	.*.
8	.027	.038	. |*.
9	-.035	.038	.*| .
10	-.008	.038	. * .
11	.045	.038	. |*.
12	.070	.038	. |*.
13	-.013	.038	. * .
14	.035	.038	. |*.
15	-.007	.038	. * .
16	.022	.038	. * .
17	.009	.038	. * .
18	.034	.038	. |*.
19	.002	.038	. * .
20	.040	.038	. |*.
21	.032	.038	. |*.
22	-.067	.038	.*| .
23	.028	.038	. |*.
24	.014	.038	. * .
25	.041	.038	. |*.

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 720 Computable 0-order correlations: 719

Figure A2.15 Residual Cross-correlation of Drainage Methane Concentration and Production (Hourly Average Series).

Cross Correlations: RESIDUALB10AV Error for BAR10AV from ARIMA, MOD_2 NOCON
RESIDUALQ10AV Error for BM2H10AV from ARIMA, MOD_1 NOCON

Cross Stand.

Lag	Corr.	Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
0	-.002	.015	.*.
1	-.001	.015	.*.
2	-.001	.015	.*.
3	-.001	.015	.*.
4	-.002	.015	.*.
5	-.001	.015	.*.
6	-.001	.015	.*.
7	-.001	.015	.*.
8	-.002	.015	.*.
9	-.002	.015	.*.
10	-.001	.015	.*.
11	-.001	.015	.*.
12	-.002	.015	.*.
13	-.002	.015	.*.
14	-.002	.015	.*.
15	-.001	.015	.*.
16	-.002	.015	.*.
17	-.001	.015	.*.
18	-.002	.015	.*.
19	-.001	.015	.*.
20	-.001	.015	.*.
21	-.001	.015	.*.
22	-.002	.015	.*.
23	-.002	.015	.*.
24	-.002	.015	.*.
25	-.002	.015	.*.

Plot Symbols: Autocorrelations * Two Standard Error Limits .
Total cases: 4320 Computable 0-order correlations: 4319

Figure A2.16 Residual Cross-correlation of Drainage Methane Concentration and Barometric Pressure (10-Minute Average Series).

Cross Correlations: RESIDUALBAV Error for BARAV from ARIMA, MOD_2 NOCON
RESIDUALQAV Error for BM2HAV from ARIMA, MOD_1 NOCON

Lag	Cross Corr.	Stand. Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1
0	.007	.037	.*.
1	.000	.037	.*.
2	-.025	.037	.*.
3	.003	.037	.*.
4	-.004	.037	.*.
5	.010	.037	.*.
6	.004	.037	.*.
7	.018	.038	. * .
8	.005	.038	. * .
9	.008	.038	. * .
10	.004	.038	. * .
11	.021	.038	. * .
12	.004	.038	. * .
13	.000	.038	. * .
14	.011	.038	. * .
15	.014	.038	. * .
16	.022	.038	. * .
17	.005	.038	. * .
18	.006	.038	. * .
19	.019	.038	. * .
20	.018	.038	. * .
21	-.002	.038	. * .
22	.013	.038	. * .
23	.011	.038	. * .
24	.011	.038	. * .
25	.001	.038	. * .

Plot Symbols: Autocorrelations * Two Standard Error Limits .

Total cases: 720 Computable 0-order correlations: 718

Figure A2.17 Residual Cross-correlation of Drainage Methane Concentration and Barometric Pressure (Hourly Average Series).