Cybernetics in Identification Modelling and Prediction of Flows and Water Quality of Non-Tidal Rivers

THESIS

Submitted for the Degree of

Doctor of Philosophy (Engineering)

UNIVERSITY OF NORTH BENGAL

NORTH BENGAL
UNIVERSITY LIBRARY
RAJA RAMMOHUNPUR

Subir Kumar Sen Gupta

JALPAIGURI GOVERNMENT ENGINEERING COLLEGE

JALPAIGURI : WEST BENGAL

1983

ST - VERF

STOCK TAKING - 2011

Ref
621.27
S 476c

83987

14 FEB 1984

PREFACE

Our two main resources are land and water. The great majority of the people are involved in farm related activities. Our engineering skills will be more beneficial to the country if these are directed to designing water resources systems for flood control, irrigation, power generation and for providing clean drinking water. The economic growth of this country is inextricably linked to the water resources development. And it is on the river basins that the people through the ages, noticed or un-noticed, build homes, make love, raise children sing songs, write poetry and even whittle statues.

The present research effort as contained in this report is a humble contribution of this investigator to the water resources development.

ACKNOWLEDGEMENT

The investigater expresses has deep sense of gratitude and heartfelt thanks to Dr. A. S. Chaudhuri for kindly suggesting the problem and his invaluable help and encouragement throughout the investigation.

Extensive help received from Prof. M. N. Mallik at every stage in carrying out this research work is greatfully acknowledged.

The investigater also expresses his deep sense of gratitude to Dr. N. C. Sinha Professor and Head of the Department of Civil Engineering, Jalpaiguri Government Engineering College for his co-operation and encouragement throughout the execution of this research project.

The investigater is indebted to Prof. N. Majumder, Professor and Head of the Department of the Electrical Engineering for providing the Computer Laboratory facilities.

The investigater expresses his sincere thanks to Dr. H. Bhaumik Principal, Jalpaiguri Government Engineering College for providing necessary working facilities to carry out this research work at Jalpaiguri Government Engineering College.

Sincere thanks are extended to Dr. M.B. Beck of International Institute of Applied System Analysis Laxenberg, Vienna, Dt. Peter Young of the Australian National University and Dr. Lennart Ljung of Linkoping University who kindly helped by sending Large number of their research papers on recursive estimation techniques.

The investigator takes this opportunity to thank Mr. D. P. Mukherjee and Mr. H. Das of Central Water Power Commission who also extended their help and Co-operation in carrying out this research work.

Thanks are also expressed to Mr. Alok Kumar Das who inspite of his heavy assignment timely completed the typing of this report.

(Subir Kumar Sengupta)

A B S T R A C T

Cybernetics deals with the purposeful analysis of complex system with a view to implement control mechanism of the system with optimally required human decision. Cybernetic methods are important tools which may be effectively used in the development and in the forecasting, planning and control of water resources systems.

The present investigation deals with two broad aspects of a water resource system of non-tidal rivers : (1) the real time simulation of hourly and daily river flow for on-line monitoring and (2) real time simulation of the water quality, the dissolved oxygen and the B.O.D of non-tidal rivers.

In the first case the outcome of the present investigation is the development of real time simulation techniques with the cybernetic methods of multilayer and combinatorial group method of data handling algorithms, and recursive instrument variable algorithms with on line adaptiveness in parameter variation. The aforesaid cybernetic methods are modified and improved to make them amenable to the peculiarities of the flows of the hilly rivers of North Bengal. Effectiveness of the developed techniques have been demonstrated with field data observed at different gauging stations of the river Teesta in North Bengal.

In the second case with the help of cybernetic methods of combinatorial group method of Data Handling Learning Identification Algorithms, the real time recursive simulation algorithms with parameter optimisation and with the on-line adaptive instrument variable algorithms, the Biochemical Oxygen Demand and Dissolved Oxygen levels of non-tidal rivers have been simulated, verified and validated with prediction optimisation. The existing methods are

modified and improved to take into account the peculiarities of the water quality behaviour. Applicability of the methods for water quality forecasting purpose has been demonstrated with field data observed in a non-tidal river passing through a highly industrialized zone. Fuzzy control algorithms of technical cybernetics have been suggested to implement reservoir control in accordance with adaptivity determined on on-line and optimally regularised decision of water quality management.

It is hoped that when in near future there will be sophistication in the water resources management in the form of using control centre with computers linked through microwave with on-line water flow and water quality monitoring points, this investigation will be very helpful. The computational procedure to use improved and modified cybernetic methods for flow and water quality simulation have been developed in well organised programme packages in high-level software.

C O N T E N T S

		Page
Preface	... ...	i
Acknowledgement	... ...	ii
Abstract	... ...	iv
Contents	... ...	vi
List of Tables	... ...	
List of Figures	... ...	

CHAPTER I

INTRODUCTION ... ... 1 - 17

1.0	Introduction	1
1.1	Scope of the Work ...	2
1.2	The Teesta, its Catchment and the Observation Sites.	5
1.3	The River Ganga and its Basin ...	17
1.4	Sources of Data ...	17

CHAPTER II

SURVEY OF THE EXISTING LITERATURE AND THE STATE OF ART 18 - 59

2.1	Introduction	18
2.2	Mathematical Models of Water Resources Systems	19
2.2.1	Box-Jenkins Family of Stochastic Models	19
2.3	Recursive Estimation Technique ...	21
2.3.1	Linear Least Square Regression Analysis	22
2.3.2	Least Square Estimation of Parameters	23

		Page
2.3.3.	Recursive Least Square Estimation of Parameters	24
2.3.4	Recursive Instrument Variable Algorithm	28
2.3.5	Exponential Weighting of Past Data ...	30
2.3.6	Recursive Time Varying Parameter Estimation Algorithm	31
2.4	Multilayer Group Method of Data Handling Algorithms	34
2.4.1	Brief Description of GMDH ...	36
2.5	Combinatorial Group Method of Data Handling Algorithm ...	44
2.5.1	Process Equation ...	45
2.6	Real Time Recursive Parameter Tracking Algorithms	47
2.6.1	Processing of the Statistical Data ...	47
2.6.2	Development of the Non-stationary Process Model	49
2.7	Concept of Fuzzy Control and Suggestive Examination of its Practical Implementation	57
2.7.1	Approach of Fuzzy set Theory ...	57
2.8	Conclusion	59

CHAPTER III

FORECASTING OF RIVER FLOWS DURING A SHORT SPAN STORM PERIOD.		60 - 88
3.0	Abstract	60
3.1	Introduction	60
3.2	Part One : Combinatorial Group Method of Data Handling Algorithm in Formulating Six Hourly Flow Forecasting Model ...	62

		Page
3.3	Part Two : Use of Multilayer GMDH in Formulating the Six hourly Flow Forecasting Model of the River Teesta at Domohani Road Bridge Point Near Jalpaiguri Town ...	75
3.4	Real Time Estimation of River Flow During a Short Span Storm Period ...	81
3.5	Conclusion	86

CHAPTER VI

RECURSIVE ESTIMATION OF RIVER FLOWS DURING A SHORT SPAN STORM PERIOD ... ... 89 - 97

4.0	Abstract	89
4.1	Introduction	89
4.2	Development of Non Stationary Process Model ...	90
4.3	Development of Recursive Parameter Estimation Algorithm	92
4.4	Illustration	93
4.5	Conclusion	97

CHAPTER V

REAL TIME FORECASTING OF DAILY FLOWS OF A NON TIDAL RIVER BY INSTRUMENT VARIABLE ALGORITHM ... 98 - 108

5.0	Abstract	98
5.1	Introduction	98
5.2	Development of the Real Time Description of the Daily River Flow Process ...	99
5.3	Illustration	105
5.4	Conclusion	105

Page

CHAPTER VI

FORECASTING OF DAILY DISSOLVED OXYGEN LEVELS OF A NON TIDAL RIVER BY SELF ORGANISATION OF MATHEMATICAL MODELS ... ... 109 - 121

6.0 Abstract 109

6.1 Introduction 109

6.2 The Concept of Self Organisation ... 110

6.3 Illustration 115

6.4 Conclusion 117

CHAPTER VII

MODELLING OF BIOCHEMICAL OXYGEN DEMAND (B.O.D) OF A NON TIDAL RIVER BY A LEARNING ALGORITHM AND RECURSIVE INSTRUMENT VARIABLE TECHNIQUE ... 122 - 146

7.0 Abstract 122

7.1 Introduction 122

7.2 Modelling of B.O.D. By Combinatorial Group Method of Heuristic Learning Identification Algorithm 124

7.3 Illustration 129

7.4 Modelling of B.O.D. By Recursive Instrument Variable Algorithm ... 130

7.5 Illustration 143

7.6 Conclusion 144

Page

CHAPTER VIII

REAL TIME FORECASTING OF DISSOLVED OXYGEN LEVEL OF A NON TIDAL RIVER ... 147 - 157

8.0 Abstract 147

8.1 Introduction 147

8.2 Processing of the Statistical Data ... 148

8.3 Development of the Non-Stationary Process Model ... 149

8.4 Development of the Real Time Recursive Least Square Prediction Algorithm ... 151

8.5 Illustration 153

8.6 Conclusion 154

CHAPTER IX

RECURSIVE PREDICTION OF DISSOLVED OXYGEN LEVEL OF A NON TIDAL RIVER ... ... 158 - 177

9.0 Abstract 158

9.1 Introduction 158

9.2 Development of the Real Time Model ... 160

9.3 Illustration 171

9.4 Conclusion 172

CHAPTER X

FORECASTING OF DAILY DISSOLVED OXYGEN LEVELS OF A NON TIDAL RIVER BY DYNAMIC LEAST SQUARE RECURSIVE ESTIMATION TECHNIQUE ... ... 178 - 192

10.0 Abstract 178

10.1 Introduction 178

10.2 Development of the Real Time Model ... 179

		Page
10.3	Illustration	188
10.4	Conclusion	189

CHAPTER XI

LEARNING IDENTIFICATION DYNAMIC ALGORITHMS FOR MODELLING OF DISSOLVED OXYGEN INCORPORATING INTERACTION PARAMETERS ... ... 193 - 212

11.0	Abstract	193
11.1	Introduction	193
11.2	Multilayer Group Method of Data Handling Algorithm	194
11.3	Illustrations	197
11.4	Dynamic Least Square Identification Model ...	198
11.5	Illustrations	206
11.6	Conclusion	206

CHAPTER XII

CONCEPT OF FUZZY CONTROL AND A SUGGESTIVE EVALUATION OF ITS APPLICATION IN THE WATER QUALITY MANAGEMENT OF A NON TIDAL RIVER ... ... 213 - 221

12.0	Abstract	213
12.1	Introduction	213
12.2	The Concept of Fuzzy Control in Water Quality Management	216
12.3	Conclusion	221

Page

CHAPTER XIII

CONCLUSION AND SCOPE OF FURTHER WORK ... 222

BIBLIOGRAPHY ... 227

LIST OF FIGURES

			Page
1.2.1	Teesta River and its Catchment	..	7
2.3.0	Steps of Development of the Recursive Parameter Estimation Algorithm	..	40
2.4.1	Concept of Feed Forward Transformation	..	46
2.4.2	Flow Chart of Multilayer GMDH Algorithms	..	48
3.2.1	Teesta River showing gauging Stations	..	67
3.2.2	Observed Flow at Domohoni	..	68
3.2.3a	Correlation Coefficient — Domohoni and Coronation Bridge	..	69
3.2.3b	Correlation Coefficient - Domohoni and Teestabazar	..	69
3.2.3c	Correlation Coefficient — Domohoni and Rangpo	..	70
3.2.3d	Correlation Coefficient — Domohoni and Great Rangit	..	70
3.2.3e	Correlation Coefficient Domohoni and Sankalan	..	70
3.2.4a	Observed Flow at Sankalan	..	71
3.2.4b	Observed Flow at Great Rangit	..	72
3.2.4c	Observed Flow at Rangpo	..	72
3.2.4d	Observed Flow at Teesta Bazar	..	73
3.2.4e	Observed Flow at Coronation Bridge	..	73
3.2.5	Observed and Modelled Values of Flow of Teesta River at Domohoni	..	74
3.3.1	Integral Square Error at Different Layers For Model Flow at Domohoni	..	79

		Page
3.3.2	Hourly Flow of River Teesta at Domohani Road Bridge ..	80
3.4.1	Flows of River Teesta at Domohani — observed and Modelled ..	87
4.4.1	Teesta River Showing Gauging Stations ..	95
4.4.2	Observed and Modelled Values of Flow of Teesta River ..	96
5.3.1	Observed Daily Flow and Error of Prediction ..	106
5.3.2	Correlation Coefficients Versus Shift of Instants in Days for Data Set ..	107
5.3.3	Correlation Coefficient Versus Shift of Time Instances in Days of the Error Sequence ..	108
6.3.1	Observed and Modelled Values of D.O. Level ..	118
6.3.2	Power Density Spectra of D.O. Level ..	119
6.3.3	Correlation Coefficient Versus Shift of Time Instances of D.O. Level ..	120
6.3.4	Correlation Coefficient Versus Shift of Time for Error ..	121
7.3.1	Observed Values of B.O.D in mg/Litre and Error of Prediction ..	131
7.3.2	Power Density Spectra Versus Cycles Per Day for B.O.D. ..	132
7.3.3	Correlation Coefficient Versus Shift of Instances (Days) for B.O.D ..	133
7.3.4	Correlation Coefficient Versus Shift of Instances of Time in Days for Errors of Prediction for B.O.D. ..	134
7.5.1	Observed Values of B.O.D and Predicted Values of B.O.D ..	145
7.5.2	Correlation Coefficient VS Shift of Instances of Time of the Error of Prediction of Daily Levels of B.O.D ..	146

		Page
8.5.1	Observed D.O. Level and Error of Prediction ..	155
8.5.2	Power Density Spectra of D.O. Level ..	156
8.5.3	Correlation Coefficient Versus Shift of Time Instances of D.O. Level ..	157
9.3.1	Observed D.O. and Modelled Values of D.O. With Instrument Variable Algorithm ..	173
9.3.2	Power Density Spectra of D.O. Level ..	174
9.3.3	Correlation Coefficient Versus Shift of Time Instances ..	175
9.3.4	Observed D.O. and Instrument Variables in Interpolation Region ..	176
9.3.5	Coefficient of Correlation of the Error Sequence for Different Shift of Instances of Time ..	177
10.3.1	Observed and Modelled Values of D.O. Level ..	190
10.3.2	Correlation Coefficient Versus Shift of Time Instances ..	191
10.3.3	Coefficient of Correlation Versus Shift of Instances of Time for the Error of Prediction ..	192
11.3.1	Input Data for Modelled Values of D.O.Level ..	207
11.3.2	Integral Square Error in Different Layers ..	208
11.3.3	Observed and Modelled Values of D.O. Level (Up Stream end) ..	209
11.3.4	Observed and Modelled Values of D.O. Level (Down stream end) ..	210
11.5.1	Observed and Modelled Values of D.O. Level (Up Stream end) ..	211
11.5.2	Observed and Modelled Values of D.O. Level (Down Stream end) ..	212
12.2.1	Dissolved Oxygen Concentration and Membership Function ..	217
12.2.2	A River System ..	217
12.2.3	Fuzzy Control Algorithm ..	218

LIST OF TABLES

			Page
2.5.1	Polynomials of Gradually Increasing Complecity of Five Variables	..	65
6.2.1	Polynomials of Gradually Increasing Complecity of Eight Variable	..	116
12.2.1	Dissolved Oxygen Concentration	..	219
12.2.2	Fuzzy Set Defination for Reservoir Release	..	219

CHAPTER I

INTRODUCTION

CHAPTER 1

1.0 Introduction

The present investigations deals with the two broad aspects of water resources system of non tidal rivers (i) the flows in a short span monsoon storm period and in normal times, (ii) dissolved oxygen and biochemical oxygen demand of a river passing through the highly industrialised zone.

The first part describes the flows of river Teesta passing through the Himalayan and Sub-Himalayan regions of North Bengal in monsoon storm period as well as in normal times in the form of well defined mathematical relationships for real time simulation and on line monitoring.

The second part describes the dissolved oxygen level and biochemical oxygen demand of the river Can in eastern England in the form of well defined stochastic difference equations for real time simulation amenable to on line adaptiveness. Fuzzy control mechanism has been suggested for on line reservoir control operation with optimally regularised decision for the indices of the water quality management.

The practical implementability of different methods of applied cybernetics and pattern recognition such as multilayer and combinatorial group method of data handling algorithms of self organisation technique, learning identification technique, least square recursive algorithms with on line adaptiveness in parameter variations, recursive instrument variable algorithms, dynamic recursive least square recursive algorithms and fuzzy control algorithms, have been demonstrated. Wherever necessary modification and improvement of the existing analytical techniques have been suggested for practical implementations. The computational procedures have been developed in well organised programme packages in high level software.

The report contains thirteen chapters. The contents and the scope of the individual chapter are briefly described below :

1.1. Scope of the Work:

1.1.1 Chapter I

The chapter I deals with the introduction and the problem of investigation. It contains the scope of the work, description of the river systems and the sources of data.

1.1.2 Chapter II

The chapter II deals extensively with the survey of the existing literature and the state of -the-art. It discusses the effectiveness and shortcomings of the previous works. It exposes gradually the harmonious evolution of the scientific thoughts pertaining to the present investigation.

1.1.3 Chapter III

The chapter III deals with the real time monitoring the river flows during a short span monsoon storm period for flood warning as well as for planning and design of analytical on line controller for the hydraulic structures of the respective water resources systems. The importance of learning identification algorithms of combinatorial and multilayer group method of data handling techniques has been successfully demonstrated for recursive forecasting of hourly flows of the hilly river Teesta flowing through the Himalayan region during a short span monsoon flood period. The chapter observes that for real time application the recursive instrument variable algorithms are satisfactory.

1.1.4. Chapter IV

The chapter IV develops a non stationary recursive parameter estimation technique minimizing an objective function with a view to implement on line adoptiveness in process parameter variations for recursive forecasting of the hourly flows of the river Teesta during

a monsoon flood period. When subjected to test with field data the model of the river flow process thus obtained is found to simulate adequately the major variations of flows observed during the storm period.

1.1.5 Chapter V

The chapter V deals with the problems of identifying the parameters of the daily flow process of the river Teesta described by linear stochastic difference equations. The parameters are evaluated on real time basis with recursive instrument variables algorithms. The dynamic model developed in this chapter is observed to simulate closely the major variations found in the field measurements of the daily flows of the river Teesta at Domohani Road Bridge point near Jalpaiguri town in North Bengal.

1.1.6 Chapter VI

In this chapter a simple dynamic model of the dissolved oxygen levels of a non-tidal river has been developed by a heuristic self-organisation technique. The model structure has been identified with the help of a learning identification algorithm known as combinatorial group method of data handling technique. The model has been verified by simulation against field data. It is found that the major variations observed in the measured data have been adequately represented in the model.

1.1.7 Chapter VII

This chapter deals with the combinatorial group method of heuristic learning identification algorithm and recursive instrument variable technique to obtain a realistic mathematical description of the biochemical oxygen demands of a non tidal river flowing through a highly industrialised zone. The simple lumped parameter dynamic models developed in this chapter highlight the potential possibilities of implementation of automatic control mechanism at the effluent discharge points to regulate water quality in accordance with a predetermined decision methodology. The field measurements of the biochemical oxygen demand collected over a 30-day period from the river Cam in eastern England are observed to be adequately represented

in the dynamic models.

1.1.8 Chapter VIII

In this chapter recursive least square non stationary time series analysis technique of cybernetics has been put to good use in on line forecasting of daily dissolved oxygen concentration of a non tidal river. When tested with field data collected over a 30-day period from the river Cam in eastern England, the developed dynamic model is found to simulate closely the major variations of dissolved oxygen levels observed in the field measurements.

1.1.9. Chapter IX

This chapter deals with the recursive instrument variable forecasting technique in conjunction with the least square recursive parameter tracking algorithms of cybernetics for real time prediction of the levels of dissolved oxygen of a non tidal river passing through a highly industrialised region. A simple dynamic model has been obtained. When tested with field data the dynamic model is found to simulate adequately the major variations of dissolved oxygen observed in the field measurements. The techniques developed in this chapter highlight the possibilities of the implementation of the on line regulatory mechanism for controlling water quality of a river.

1.1.10 Chapter X

This chapter deals with an application of the dynamic least square estimation algorithms for on line modelling of dissolved oxygen levels of a non tidal river passing through a highly industrialised region. The mathematical description of the dissolved oxygen levels allows the real time monitoring of water quality. Actual data are used to calibrate the real time dynamic model. Agreement between the observed and the predicted values of the levels of dissolved oxygen illustrates the potential use of the dynamic least square recursive estimation technique for on line control of water quality in accordance with predetermined decision strategies.

1.1.11 Chapter XI

In chapter XI the multilayer group method of data handling algorithm and dynamic recursive least square technique of applied cybernetics have been used for obtaining dynamic descriptions of daily dissolved oxygen level of a non tidal river correlating the interactions of daily biochemical oxygen demand, sustained sunlight and volumetric flow rate. When tested with field data the models are found to simulate adequately the major variations of the levels of dissolved oxygen observed in the field measurements.

1.1.12 Chapter XII

This chapter offers a conceptual examination of the possibilities of fuzzy control application for water quality management of a non tidal river. In the present work a systematic approach of fuzzy control has been introduced rather as an analytical tool than as an immediately applicable closed loop control strategy.

1.1.13 Chapter XIII

This chapter concludes the report. The specific areas of further research are suggested. At the end a list of referenced bibliography is enclosed.

Emphasis has been given to make the individual chapter self - contained so far as its presentation of idea, development of illustrations are concerned.

1.2 The Teesta, its catchment and the observation sites

The river Teesta rising from the Himalayan ranges in north Sikkim and passing through the deep gorges for nearly 138 Kms. debouches upon the plain of West Bengal near Sevoke. The Teesta is a very fast flowing river. Its average velocity is 6.2 metre per second. In winter its water is seagreen. In summer and in rainy

season when the ice in the glacier melts quickly and when its catchment is bathed in torrential rains the milky white water flows through the river surging its narrow Himalayan fluvial course.

The accompanying map, Fig 1.2.1 depicts the Teesta river and its catchment. A brief description of the river is given. The river Lohnak originates from the snow line in North Sikkim at a height of about 6401 metres. The river Polichu originates from the Zemu glacier at a height of 4968 metres. These two rivers combine at Lachen, after which it is known as the Zemuchu river. At Chungthang Lachen Chu at down stream of Chungthang to form the river Teesta. Thus the Teesta, in Bengali language called Trisrota meaning that three flows have combined together, is formed by the rivers Zemu Chu, Lachen Chu and the Lachung Chu. At Sankalan the river Talung Chu originating from the Talung grasier in north western Sikkim at a height of about 5873 metres, joins the river Teesta. Upto Sankalan length of the river from the origin is about 70 k.m and the catchment area is about 4200 sq. k.m.

From Sankalan the river Teesta flows through the narrow Himalayan gorges and comes to Singtam. About 15 k.m up-stream of Singtam the river Dik Chu joins the Teesta. At Singtak from the easternside the river Rongni Chu joins the Teesta. Upto Kantitar the length of the river is approximate 114 K.m from the origin and the catchment area is approximate 4874 sq. k.m from the origin.

At Rongpo the river Rongpo Chu from the eastern catchment region joins with the Teesta. The length of the river from the origin upto Rongpo is approximate 116 k.m and the total catchment area of the river including its tributaries upto Rongpo is approximately 5405 sq. k.m from the origin.

Near Singlabazar the river Great Rangit combines with the river Ramam and the river Little Rangit and flows as the Great Rangit river. This combination of three rivers being in an addition of about 1956 sq. k.m. of catchment area to the Teesta catchment.

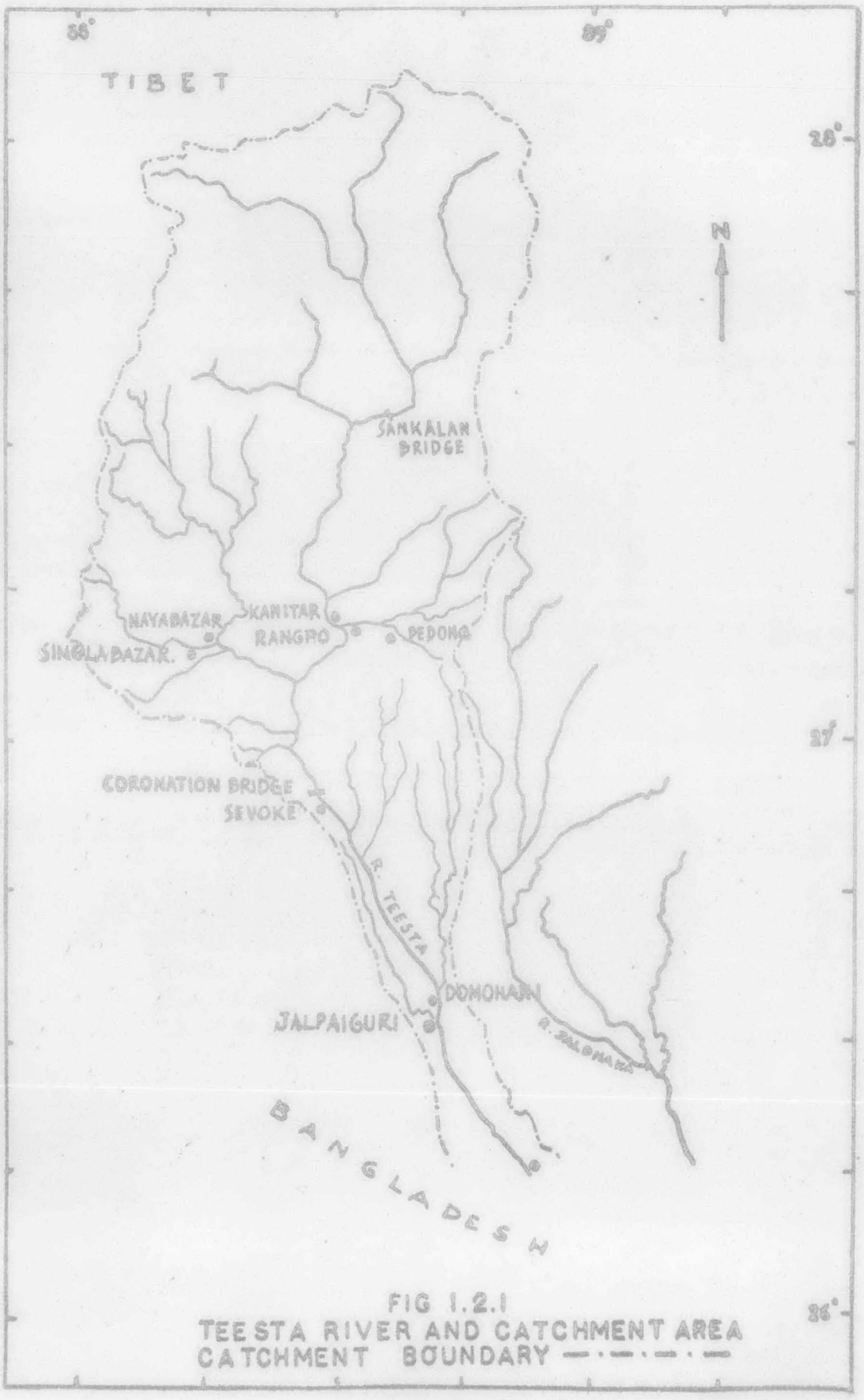


FIG 1.2.1
TEESTA RIVER AND CATCHMENT AREA
CATCHMENT BOUNDARY — · — · — · —

At about 3 k.m up-stream of Teestabazar the Great Rangit joins the Teesta. The confluence of the Great Rangit and the Teesta is unforgettable. The clear sea green water of the Great Rangit mixes with the milky white water of the Teesta and thus creating a wonderful cocktail of nature. The length of the river from the origin upto Teestabazar is about 134 k.m and the catchment is (approx.) 7714 sq. k.m.

Upto Coronation Bridge the length of the Teesta from the origin is (approx) 158 k.m and the catchment area is about 8147 sq. k.m. Upto Sevoke its length from the origin is (approx) 160 k.m and the catchment area is about 8179 sq.k.m.

In the plains the important tributaries of the Teesta are the Lish, Ghish, Chel, Neora and the Karla. Upto Domohani Road Bridge the length of the Teesta from the origin is (approx) 206 k.m and the catchment area is (approx) 9432 sq. k.m.

The Teesta mixes with the Brahmaputra.

1.2.1 The main observation stations.

The main observation stations from where the data for this investigation were collected have been shown in the map. A brief description of the stations are given below.

Sitewise brief note of various observation stations.

1) SANKALAL BRIDGE

a) Location : Lat. - 27° 30' N, Long - 88° 32' E, on river Teesta in hilly terrain of North Sikkim down-stream of confluence of Lachenchu, Lachungchu and Talungchi. National Highway 31 A is about 3 k.m from the site.

b)	Nature of Station :	Gauge & Discharge with 15 Watt H.F. Wireless facilities.
c)	Mode of observation :	Hourly gauge observed round the clock during monsoon. Discharge observation thrice a day at 0800, 1200 and 1600 hrs. using wooden float and a float run of 30 m. Cross-section taken once a monthe using sounding of 15 Kg. wt.
d)	Length of river from origin upto site :	70 K.M (approx.)
e)	Catchment area upto site from origin :	4200 sq. K.M (approx.)
f)	Date of commencement of gauge/discharge :	12.10.72 / 2.12.72
g)	Maxm. observed gauge/ Discharge during monsoon of 1979 :	759 m on 2.7.79 (1900 Hrs) / 1433.64 cumecs on 12.7.79 (1600 hrs)
h)	Maximum ever recorded gauge/ Discharge :	762.30 m on 18.6.73 (0100 hrs.) / 1739.07 cumecs on 18.7.73 (0700 hor)
i)	Type of Raingauge :	NIL

2) KHANITAR

a)	Location :	Lat 27° 10.5' N, long - 88° 30' E in Sikkim on Teesta HH-31 A is about 1 K.M from the site.
b)	Nature of station :	G & D with 15 Watt. H.F wireless facilities.
c)	Mode of observation :	Hourly gauge observed roung the clock during monsoon. Discharge observation thrice a day at 0800, 1200 and 1600 hrs. using Wooden float and a float run of 30 M. Cross-section taken once a week using sounding of 25 kg wt.

d) Length of river from origin upto site : 114 K.M (approx.)

e) Catchment area upto site from origin : 4874 Sq. K.M (approx.)

f) Date of commencement of observation of Gauge/Discharge : 12.6.70 / 12/6.70

g) Maxm. observed gauge/ Discharge during monsoon of 1979 : 396.79 m on 2.7.79 (2000 hrs.)
1784.49 cumecs on 23.7.79 (0700 hrs)

h) Maxm. ever recorded Gauge/Discharge : 397.50m on 13.10.73 (0500 hrs)
2197.74 cumecs on 18.6.79 (1200 hrs)

i) Type of raingauge : One S.R. Raingauge and one ordinary raingauge.

3) ROMPO

(i) ON RIVER RANGPOCHU :

(a) Location : Lat. 27° 10' N, long - 88° 32' E, on river Rongpochu i.e. tributary of river Teesta near the Junction of L.R.P Road at NEL-31 A is about 1 K.M from the site.

(b) Nature of station : G & D with 16 Watt. H.F. Wireless facilities.

(c) Mode of observation : Hourly Gauge observed round the clock during monsoon. Discharge observation thrice a day at 0800, 1200 and 1600 hrs. using wooden float and a float run of 30 m Cross-section taken once a week using sounding of 20 Kg. Wt.

(d) Date of commencement of observation of Gauge/Discharge : 5.6.70 / 1.7.70

(e) Maxm. observed Gauge Discharge during monsoon of 1979 : 304.608 Mtr. on 3.8.79 (0300 hrs.)
603.25 cumecs on 23.7.79 (1200 hrs)

(f) Maxm. ever recorded Gauge/Discharge : 304.608 Mtr. on 3.8.79 (0300 hrs)
732.016 cumecs on 26.9.72 (1200 hrs)

(g) Type of Raingauge : One ordinary raingauge.

4) SINGLABAZAR

(i) ON RIVER GREAT RANGIT

(a) Location : Lat 27° 07' N and Long 88° 14' E on river Great Rangit i.e. tributary of river Teesta, is 2 K.M from the site.

(b) Nature of station : Gauge with 15 watt. H.F Wireless facilities.

(c) Mode of observation : Hourly gauge observed round the clock during the monsoon.

(d) Date of commencement of observation of Gauge/Discharge : 11.12.69

(e) Maxm. observed Gauge/ Discharge during monsoon '79 : 310.00 Mtr. on 24.7.69 (0300 hrs)

(f) Maxm. ever recorded Gauge / Discharge : 310.35 Mtr on 12.10.73 (2100 hrs.)

(g) Type of raingauge : One S.R. Raingauge and one ordinary raingauge.

(ii) ON RIVER RAMAM

(a) Location : Lat. 27° 7.5N, Long 88° 15.5'E on river Ramam i.e. tributary of river Teesta, is 1 K.M from the site.

(b) Nature of station : Gauge & Discharge.

(c) Mode of observation : Hourly gauge observed round the clock during monsoon. Discharge observation thrice a day at 0300, 1200 and 1600 hrs. using wooden float and a float run of 30 Mtr. Cross section taken once a week using sounding of 26 Kg.Wt.

(d) Length of river from origin upto site : 32 K.M. (approx.)

(e) Catchment area upto site from origin : 385 Sq. K.M (approx.)

(f) Date of commencement or observation of Gauge/Discharge : 11.12.69 / 1.5.73

(g) Mean. observed Gauge/ Discharge during monsoon of 1979 : 320.00 Mtr. on 29.7.79 (100 hrs)
140.63 cumecs on 29.7.79 (1200 hrs)

(h) Mean. ever recorded Gauge/Discharge : 320.15 Mtr. on 12.10.73 (2200 hrs)
421.63 cumecs on 16.8.78 (0700 hrs)

(i) Type of Raingauge : NIL

(iii) ON RIVER LITTLE RANGIT

(a) Location : Lat. 27^{0} 5 $\frac{5}{6}$ 'N, Long 88^{0} 15 $\frac{1}{2}$ 'E on river Little Rangit i.e. tributary of river Teesta, is about 3 K.M from the site.

(b) Nature of station : G & D with Non-exchange telephone facilities with Singlehaser.

(c) Mode of observation : Hourly gauge observed round the clock during monsoon. Discharge observation thrice a day at 0800, 1200 and 1600 hrs. using wooden float and a float run of 30 Mtr Cross section taken once a day using sounding of 25 Kg. Wt.

(d) Length of river from origin upto site : 35 K.M (approx.)

(e) Catchment area upto site from origin : 184 Sq. K.M. (approx)

(f) Date of commencement of observation of Gauge/Discharge : 21.6.73 / 1.7.73

(g) Mean. observed Gauge/ Discharge during monsoon of 1979 : 321.53 Mtr. on 8.8.79 (2400 hrs)
83.60 cumecs on 30.9.79 (0700 hrs)

(h) Mean. ever recorded Gauge/Discharge : 314.30 Mtr on 12.10.73 (1800 hrs)
268.57 cumecs on 7.9.75 (1700 hrs)

(i) Type of Raingauge : NIL

5) TISTABAZAR

(a) Location : Lat 27° 03' N, Long 88° 25' E on river Teesta on NH - 31.

(b) Nature of Station : G & D with 15 Watt Wireless facilities and Non-Exchange telephone line between Camp shed and bridge point and camp shed to Sevoke site.

(c) Mode of observation : Hourly gauge observed round the clock during monsoon. Discharge observation thrice a day at 0800, 1200 and 1600 hrs using wooden float and a float run of 70 Mtr. Cross section taken once a month through sounding of 25 kg. Wt.

(d) Length of river from origin upto site : 134 K.M. (approx.)

(e) Catchment area upto site from origin : 7714 Sq. K.M (approx.)

(f) Date of Commentement of observation of Gauge / Discharge : 1.5.69 / 22.8.74

(g) Maxm. observed Gauge/ Discharge during monsoon of 1979 : 211.80 Mtr. on 24.7.79 (0400 hrs)
2906.17 cumecs on 24.7.79 (0700 hrs)

(h) Maxm. ever recorded Gauge/Discharge : 217.00 Mtr on 13.10.73 (0300 hrs)
7642.45 cumecs on 13.10.73 (0600 hrs)

(i) Type of Raingauge : One S.R. Raingauge and one ordinary raingauge.

6) CORONATION

(a) Location : An important forecasting point on river Teesta on NH - 31.

(b) Nature of Station : G & D with 15 Watt Wireless facilities and Non-Exchange telephone line between Camp shed and bridge point and Camp shed to Sevoke site.

(c) Mode of observation : Hourly gauge observed round the clock during monsoon. Discharge observation thrice a day at 0800, 1200 and 1600 hrs using wooden float and a float run of 70 Mtr. Cross - section taken once a month through sounding of 25 kg. Wt.

(d) Length of river from origin upto site : 158 M.M (approx.)

(e) Catchment area upto site from origin : 8147 Sq. K.M (approx.)

NORTH BENGAL
UNIVERSITY LIBRARY
RAJA RAMMOHUNPUR

(f) Date of commencement of observation of Gauge/Discharge : 10.9.74 / 1.8.74

(g) Mean. observed Gauge/Discharge during monsoon of 1979 : 151.40 Mtr. on 24.7.79 (0600 hrs)
2751.36 cumecs on 24.7.79 (0700 hrs)

(h) Mean. ever recorded Gauge/Discharge : 156.500 Mtr. on 12.10.73 (2030 hrs)
5090.00 cumecs on 13.10.73 (0700 hrs)

(i) Type of Raingauge : NIL.

7) DOMOHANI

(a) Location : This is an important forecasting stn. on Teesta. Lat 26° 32' 44" N, Long 88° 45' 42" E. on river Teesta. The observations being done on Rd. Bridge over river Teesta. Jalpaigri town is about 8 K.M. from site.

(b) Nature of station : G & D with V.H.F Wireless facilities.

(c) Mode of observation : Hourly gauge observed round the clock during monsoon. Discharge observation tries a day at 0800, 1200 and 1600 hrs. using current meter.
Cross-Section taken once daily using 25 Kg. Wts.

(d) Length of river upto site from the origin : 206 K.M (approx.)

(e) Catchment area upto site from origin : 9432 Sq. K.M. (approx.)

(f) Date of commencement of observation of Gauge/Discharge : 1.6.72 / 1.6.72

(g) Mean. observed Gauge/Discharge during monsoon of 1979 : 85.67 Mtr. on 24.7.79 (1100 hrs)
5609.47 cumecs on 24.7.79 (1200 hrs)

83987
14 FEB 1984

(h) Maxm. ever recorded Gauge / Discharge : 86.78 Mtr. on 13.6.72 (1400 hrs)

10,935.97 cumecs on 28.7.72 (0700 hrs).

(i) Type of Raingauge : One S.R. Raingauge and one ordinary raingauge and one ordinary raingauge.

Table showing the distance of various gauge observation in various basin from the forecasting stations and their travel time.

Sl. No:	Base Station :	F/C Station:	River distance between base station to F/C station in K.M	Approx. Travel time :
	TEESTA CATCHMENT			
1.	Samdong Bridge	Coronation Bdg.	84	10 Hrs
2.	Khanitar	Coronation Bdg.	44	4 Hrs
3.	Rongpo	Coronation Bdg.	42	4 Hrs
4.	Singlebazar	Coronation Bdg.	54	4 Hrs
5.	Naynbazar	Coronation Bdg.	54	4 Hrs
6.	Toastabazar	Coronation Bdg.	22	2 Hrs
1.	Samdong Bridge	Domohani Rd. Bdg.	132	16 Hrs
2.	Khanitar	Domohani Rd. Bdg.	92	10 Hrs
3.	Rongpo	Domohani Rd. Bdg.	90	10 Hrs
4.	Singlebazar	Domohani Rd. Bdg.	102	10 Hrs
5.	Naynbazar	Domohani Rd. Bdg.	102	10 Hrs.
6.	Toastabazar	Domohani Rd. Bdg.	70	8 Hrs.
7.	Coronation Bridge	Domohani Rd. Bdg.	48	6 Hrs.

1.3. The River Cam and its Basin

The river Cam flows through Cambridgeshire, an eastern country of England, bounded north by Lincolnshire, east by Norfolk and Suffolk, south by Essex and Hertfordshire and west by Bedfordshire, Huntingdonshire and Northamptonshire. The area of the geographical country, which includes the Isle of Ely, is 867.4 square miles.

All but the southernmost part of the country lies within the fen district, a flat alluvial tract, only a few feet above sea level, intersected by innumerable drainage channels.

The two principal rivers are the Ouse and its tributaries including the river Cam and the Nene. For the greater part of their courses within the country these rivers flow slowly between raised banks, the most spectacular of which are two great parallel cuts, the Old and the New Bedford rivers, running northeast for 30 miles from Erith to Denver Sluice in Norfolk. The area between them acts as a flood reservoir if need be. The ouse is tidal below Denver. The dry chalk up lands east and south east of Cambridge carry belts and clumps of beech, seen at their best in the area around Newmarket health. Further west the clays and greensands flanking the valleys of the Bourn brook and in the upper reaches of the Cam favour ash and elm and a variety of shrubs. Willows are the characteristic trees. The river Cam joins the Ouse 4 miles above Ely. Paper, cement, light engineering item and printing materials are the most important industry of this region.

1.4. Sources of Data

Data pertaining to Tawta river system have been obtained from the Jalpaiguri Field Division of the Central Water Commission, Government of India. Dr. M. B. Beck of the International Institute of Applied Systems Analysis, Laxenburg, Austria, has kindly sent us the data of water quality of the river Cam in eastern England.

CHAPTER II

SURVEY OF THE EXISTING LITERATURE AND THE STATE OF ART

CHAPTER II

THE SURVEY OF THE EXISTING LITERATURE AND THE STATE - OF - ART

2.1 Introduction

The first part of the present investigation identifies the physical processes of the surface water hydrology of the Tawata catchment area in the Himalayan and sub-Himalayan region during a monsoon flood period and in normal times in well defined mathematical description for hourly monsoon flood flows and daily normal time flows. The mathematical descriptions of the river flows give a simulation model amenable to real time computation with on line adaptiveness.

In the second part the dissolved oxygen levels and biochemical oxygen demand of non-tidal river have been simulated, verified and validated with prediction optimization with field data observed in the river Cam in eastern England. For the analysis of the above physical processes the different methods of cybernetics have been used. Cybernetics deals with the purposeful analysis of complex system with optimally required human decision. Cybernetics method are important tools which may be effectively used in the development and in the forecasting, planning and control of water resources systems.

The out come of the present investigation is the development of real time simulation techniques with the methods of cybernetics of recursive instrumental variable algorithms with on line adaptiveness in parameter variations, and multilayer and combinatorial group method of data handling algorithms of self organisation of mathematical models. Fuzzy control algorithms of technical cybernetics have been suggested to implement reservoir control in accordance with adaptively determined on line and optimally regularised decision of water quality management.

2.2 Mathematical Models of Water Resources Systems.

The water resources literature abounds with numerous types of mathematical models for surface water hydrology, dissolved oxygen levels and biochemical oxygen demands. The methods which are relevant to the present investigation are discussed below with the associated state-of-the-art.

2.2.1 Box-Jenkins family of stochastic models.

Box-Jenkins family of stochastic models [1] is considered to be of significant importance in water resources. Box-Jenkin's work is actually a culmination of the research of many prominent statisticians starting with the pioneering work of Yule [2]. Recently, new innovations in Box-Jenkin's models have been developed for applying to engineering problems [3,4,5].

Historically, hydrologists have often assumed non-seasonal data to be either Markov [6], a Box-Jenkin's model with one autoregressive (AR) and one moving average (MA) term [7], fractional Gaussian noise [8] or broken line process [9]. Markov model as is used in hydrology may be treated as a subset of Box-Jenkin's models [10]. Method of fractional Gaussian noise and broken line process lack suitable estimation procedure [11] for the model parameters for general simulation studies in hydrology.

Hydrological processes are non-stationary in nature. One of the basic conditions for the ARMA process to be applicable to time series is that the time series be stationary [12]. To overcome the problem of non-stationarity, researchers often transfrom the historical data by a suitable deterministic transformation to remove seasonality. Various types of deseasonalisation transformation have been employed in hydrology. Clarke [13] provides a comprehensive description of wide range of techniques for removing seasonality. Since the hydrological systems are complex in nature the conditions of non-stationary or the ARMA processes themselves should be relaxed or modified respectively, in order to make the processes useful in hydrology.

In the present investigation a modified version of [14] Box-Jenkin's models has been used. In the ensuing section the basic feature of Box-Jenkins modal have been described.

Let $Z_1, Z_2 \ldots, Z_{t-1}, Z_t, Z_{t+1}, \ldots, Z_N$ be a discrete non-seasonal time series measured at equal time intervals.

Suppose Z_t can be described as a autoregressive integrated moving average modal (ARIMA) as a generating process of the form

$$\phi(B) \left\{ [(1-B)^d Z_t^{(\lambda)}] - \mu \right\} = \theta(B)\, a_t \quad \ldots$$

in which $Z_t^{(\lambda)}$ = some appropriate transformation of Z_t such as Box-Cox [15] transformation. It is imperative that the historical record be suitably transformed to fit a normal distribution as far as possible [16]. No transformation is also a possible option;
t = discrete time, B = backward shift operator defined by

$$B^k \, Z_t^{(\lambda)} = Z_{t-k}^{(\lambda)}$$

for K = 1, 2 ...,

a_t = normally independently distributed white noise residual with mean 0 and variance (homoscedastic) $\sigma^2 a$; $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 \ldots \phi_p B^p$ is the non-seasonal autoregressive operator of order p and ϕ_i, i = 1, 2 ...., p are the non-seasonal autoregressive parameters; $(1-B)^d = \nabla^d$ is the non-seasonal differencing operator of order d to produce non-seasonal stationarity of the d th difference, usually d = 0, 1 or 2; $w_t = \nabla^d Z_t^{(\lambda)}$ is the series formed by differencing the $Z_t^{(\lambda)}$ series and n = N-d is the number of terms in the w_t series; μ = mean level of the process (if d > 0 often $\mu = 0$ and when d = 0 then μ is usually assumed the mean value of $Z_t^{(\lambda)}$ series); and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots \theta_q B^q$ is the non-seasonal moving average operator of order q and θ_i, i=1,2, ... , q are the non-seasonal moving average parameters.

Box and Jenkin's [1] use the notation (p,d,q) to list the orders of ARIMA model given in equation (1). When a non-seasonal model is stationary and requires no differencing (i.e. d = 0) it is often referred to as simply an autoregressive moving average process (ARMA) . The notation ARMA (p, q) is used to represent the orders of AR and MA operators respectively. A non-seasonal ARMA (p,q) without data transformation can be described as

$$z_t = \sum_{k=1}^{p} \phi_k z_{t-k} - \sum_{k=1}^{q} \theta_k a_{t-k} + a_t \qquad \ldots (1)$$

ϕ and θ co-efficients may be estimated by Yule-Walker equations, the least square method or the maximum likelihood method.

For the purpose of comparing the merits of ARMA models with different orders, it is necessary to have measures of how well they fit the time series.

2.3 Recursive Estimation Technique.

There are many ways of obtaining recursive algorithms. Some of the early references on recursive identification methods are given in [65] . It is not attempted to present all its variants in their wide spectrum of use. The discussion is limited to that part which is relevant to the present investigation and it deals with more than just the subject of estimation algorithm : it treats also the subjects of system identification and forecasting. This is due to the fact that the techniques of estimation derive in part from the broader field of system identification which in-corporates estimation with modal structure identification, modal verification and modal validation. The investigator is heavily debted to Beck [17] for his excellent treatment of the subject in a tutorial manner in a fashion highly understandable for practical use. Excellent treatment of the time series by Young [18] has acted as a guide.

In the present investigation black box models have been assumed and therefore only such models are discussed. This type models is often encountered in physical system. When suitably transformed the model becomes amenable to recursive techniques. Soderstrom et al [19] and Ljung [20,21,22,23,24] have given a good coverage on recursive identification methods.

Recursive technique has been defined by Young [25] as " a technique in which an estimate is updated on receipt of fresh informations ". Steps of development of the recursive parameter estimation algorithms has been depicted in Fig 2.3.0.

2.3.1. Linear Least Square Regression Analysis.

Without making any assumption about the physical phenomena of the system which are thought to govern the system's dynamic behaviour the system is often viewed as a black box. Identification means that a model is fitted to measured U_i [i=1, 2 .... n] of the input signal and y_i [i = 1,2 ..... n] of the output signal. The analysis of time series suggests that there is a probability that the current values of the output $Y(t_k)$ is a function of the previous output observations, the autoregressive terms $Y(t_{k-1})$, $Y(t_{k-2})$ ... ; and the past observations of inputs $U_1(t_{k-1})$, $U_2(t_{k-1})$, $U_1(t_{k-2})$ together with the current unknown realistic noise process $V(t_k)$. Therefore the system may be assumed as

$$Y(t_k) = \sum_{i=1}^{n} \beta_i \; Y(t_{k-i}) + \sum_{j=1}^{m} \sum_{i=1}^{n} \delta_{ji} \; U_j(t_{k-1}) + V(t_k) \quad \ldots (2.1)$$

Determination of n is known as model order determination. Chaudhuri [26] has suggested output-output and output-input correlation as an intutive consideration for model order determination which is also the model structure identification.

In the polynomials of back ward shift operator, the equation (2.1) can be rearranged as

$$A\left(q^{-1}\right) Y(t_k) = \sum_{j=1}^{m} B_j\left(q^{-1}\right) U_j(t_k) + V(t_k) \quad \ldots(2.2)$$

where the backward shift operator q^{-1} is defined by

$$q^{-1} Y(t_k) = Y(t_{k-1}) \quad \ldots(2.3)$$

and

$$A\left(q^{-1}\right) = 1 - \beta_1 q^{-1} - \beta_2 q^{-2} \ldots - \beta_n q^{-n} \quad \ldots(2.4)$$

$$B_j\left(q^{-1}\right) = \delta_{j1} q^{-1} + \delta_{j2} q^{-2} \ldots + \delta_{jn} q^{-n} \quad \ldots(2.5)$$

This model is quite flexible since it requires that the equation be linear in parameters [27].

Equation(2.2) can be represented as

$$Y(t_k) = Z^T(t_k)\alpha + V(t_k) \quad \ldots 2.6)$$

Where

$$Z^T(t_k) = [\, Y(t_{k-1}) \ldots Y(t_{k-n}),\ U_1(t_{k-1}) \ldots U_1(t_{k-n}) \ldots U_m(t_{k-1}) \ldots U_m(t_{k-n}) \,] \quad \ldots(2.7)$$

and

$$\alpha = [\, \beta_1 \ldots \beta_n,\ \delta_{11} \ldots \delta_{1n} \ldots \delta_{m1} \ldots \delta_{mn} \,] \quad \ldots(2.8)$$

2.3.2. Least square estimation of parameters

Least square estimate of the parameters is obtained by minimizing the loss function defined as the sum of the squared errors,

$$J \triangleq \sum_{k=1}^{N} \left(Y(t_k) - Z^T(t_k)\,\hat{\alpha} \right)^2 \quad \ldots(2.9)$$

in which the estimates $\hat{\alpha}$ of α that minimizes J are called the

least square estimates. The modal response errors $\hat{V}(t_k) = Y(t_k) - Z^T(t_k)\hat{\alpha}$ are not in general identical with $V(t_k)$ but converges to $V(t_k)$ as $\hat{\alpha}$ converges to true value of α.

Differentiating J with respect to parameter vector $\hat{\alpha}$ and then setting the vector of derivatives equal to zero we have the well known equations for the least square parameter estimates,

$$\hat{\alpha} = \left[\sum_{k=1}^{N} Z(t_k) Z^T(t_k) \right] \left[\sum_{k=1}^{N} Z(t_k) Y(t_k) \right] \quad \ldots (2.10)$$

2.3.3. Recursive least square estimation of parameters

Recursive form of least square estimation of parameters is an elegant way of updating estimates $\hat{\alpha}$ which changes as it converges to true value α.

Consider the equation

$$Y(t_k) = Z^T(t_k)\alpha + V(t_k)$$

The least square estimate $\hat{\alpha}$ of α is given by the equation(2.10). With direct analogy from equation(2.10) we may write.

$$\hat{\alpha}(t_k) = \left[\sum_{j=1}^{k} Z(t_k) Z^T(t_j) \right] \left[\sum_{j=1}^{k} Z(t_j) Y(t_j) \right] \quad \ldots (2.11)$$

Equation (2.11) can be written in coincise form as

$$\hat{\alpha}(t_k) = P(t_k)\, b(t_k) \quad \ldots (2.12)$$

where

$$P(t_k) \triangleq \left[\sum_{j=1}^{k} Z(t_j) Z^T(t_j) \right]$$

and

$$b(t_k) \triangleq \left[\sum_{j=1}^{k} Z(t_j) Y(t_j) \right]$$

Recursive relationship for P (.) and b (.) can be set as

$$\left[P(t_k)\right]^{-1} = \sum_{j=1}^{k} Z(t_j)\, Z^T(t_j)$$

$$= \left[P(t_{k-1})\right]^{-1} + Z(t_k)\, Z^T(t_k) \qquad \ldots (2.13)$$

Similarly

$$b(t_k) = b(t_{k-1}) + Z(t_k) * Y(t_k) \qquad \ldots (2.14)$$

Premultiplying by $P(t_k)$ and post-multiplying by $P(t_{k-1})$ we get from eqn. (2.13)

$$P(t_{k-1}) = P(t_k) + P(t_k)\, Z(t_k)\, Z^T(t_k)\, P(t_{k-1}) \qquad \ldots (2.15)$$

Post multiplying by $Z(t_k)$ equation (2.15) given,

$$P(t_{k-1})\, Z(t_k) = P(t_k)\left[1 + Z^T(t_k)\, P(t_{k-1})\, Z(t_k)\right] \qquad \ldots (2.16)$$

Post multiplying by $\left[1+Z^T(t_k)\, P(t_{k-1})\, Z(t_k)\right]^{-1} Z^T(t_k)\, P(t_{k-1})$

equation (2.16) gives.

$$P(t_{k-1})\, Z(t_k)\left[1+Z^T(t_k)P(t_{k-1})Z(t_k)\right]^{-1} Z^T(t_k)\, P(t_{k-1})$$

$$= P(t_k)\, Z(t_k) Z^T(t_k) P(t_{k-1}) \qquad \ldots (2.17)$$

From eqn. (2.15) we get

$$P(t_k) = P(t_{k-1}) - P(t_k) Z(t_k) Z^T(t_k)\, P(t_{k-1}) \qquad \ldots (2.18)$$

Finally substitution of eqn. (2.17) in eqn. (2.18) gives,

$$P(t_k) = P(t_{k-1}) - P(t_{k-1})\, Z(t_k)\left[1+Z^T(t_k)P(t_{k-1})Z(t_k)\right]^{-1}$$

$$Z^T(t_k) P(t_{k-1}) \qquad \ldots (2.19)$$

Now from eqn. (2.12)

$$\hat{\alpha}(t_k) = P(t_k).\ b(t_k)$$

i.e.

$$\hat{\alpha}(t_k) = \left[P(t_{k-1}) - P(t_{k-1}) Z(t_k)\left[1+Z^T(t_k)P(t_{k-1})Z(t_k)\right]^{-1}\right.$$

$$\left. Z^T(t_k)\, P(t_{k-1})\right]\ b(t_{k-1}) + Z(t_k) Y(t_k)$$

and since $P(t_{k-1})\, b(t_{k-1}) = \hat{\alpha}(t_{k-1})$

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) - P(t_{k-1})Z(t_k)\left[1+Z^T(t_k)P(t_{k-1})Z(t_k)\right]^{-1} Z^T(t_k)\hat{\alpha}(t_{k-1})$$
$$+ P(t_{k-1})Z(t_k)Y(t_k) - P(t_{k-1})Z(t_k)\left[1+Z^T(t_k)P(t_{k-1})Z(t_k)\right]^{-1}$$
$$Z^T(t_k)P(t_{k-1})Z(t_k)Y(t_k)$$

Since $Z^T(t_k)\, P(t_{k-1})Z(t_k)$ is scalar and hence $\left[1+Z^T(t_k)P(t_{k-1})Z(t_k)\right]^{-1}$ is also scalar. Therefore

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) - P(t_{k-1})Z(t_k)\left[1+Z^T(t_k)P(t_{k-1}Z(t_k)\right]^{-1} Z^T(t_k)\hat{\alpha}(t_{k-1})$$
$$+ P(t_{k-1})Z(t_k)\left[1+Z^T(t_k)P(t_{k-1})Z(t_k)\right]^{-1}Y(t_k)$$

And hence we get the recursive least square parameter estimation algorithm as

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) + P(t_{k-1})Z(t_k)\left[1+Z^T(t_k)P(t_{k-1})Z(t_k)\right]^{-1}$$
$$\left[Y(t_k) - Z^T(t_k)\,\hat{\alpha}(t_{k-1})\right] \qquad \ldots (2.20)$$

where

$$P(t_k) = P(t_{k-1}) - P(t_{k-1})Z(t_k)\left[1+Z^T(t_k)P(t_{k-1})Z(t_k)\right]^{-1}$$
$$Z^T(t_k)P(t_{k-1}) \qquad \ldots (2.21)$$

with

$$P(t_{k-1}) = \left[\sum_{j=1}^{k-1} Z(t_j)Z^T(t_j)\right] \qquad \ldots (2.22)$$

Least square technique is of great importance for parameter estimation [26]. But this does not overcome the problem of bias.

The problem of bias is discussed below

The non-recursive least square estimation of parameter is given by

$$\hat{\alpha}(t_k) = \left[\sum_{j=1}^{k} Z(t_j)Z^T(t_j)\right]^{-1}\left[\sum_{j=1}^{k} Z(t_j)Y(t_j)\right] \quad \ldots (2.23)$$

where the system equation is taken as

$$Y(t_j) = Z^T(t_j)\alpha + V(t_j) \quad \ldots (2.24)$$

Upon substitution of equation (2.24) in equation (2.23) we have

$$\hat{\alpha}(t_k) = \left[\sum_{j=1}^{k} Z(t_j)Z^T(t_j)\right]^{-1}\left[\sum_{j=1}^{k} Z(t_j)\left[Z^T(t_j)\alpha + V(t_j)\right]\right]$$

$$= \left[\sum_{j=1}^{k} Z(t_j)Z^T(t_j)\right]^{-1}\left[\sum_{j=1}^{k} Z(t_j)Z^T(t_j)\alpha\right]$$

$$+ \left[\sum_{j=1}^{k} Z(t_j)Z^T(t_j)\right]\left[\sum_{j=1}^{k} Z(t_j)V(t_j)\right]$$

Therefore

$$\hat{\alpha}(t_k) = \alpha + \left[\sum_{j=1}^{k} Z(t_j)Z^T(t_j)\right]\left[\sum_{j=1}^{k} Z(t_j)V(t_j)\right] \quad \ldots (2.25)$$

Equation (2.25) shows that for estimate $\hat{\alpha}(t_k)$ to be unbiased the following condition must hold

$$E\left\{Z(t_j)\,V(t_j)\right\} = 0 \quad \text{for all } K \quad \ldots (2.26)$$

This is only possible if $V(t_k) = e(t_k)$ where $e(t_k)$ is a white noise sequence [17].

As pointed out by Beck [29] for most practical cases of interest $V(t_k)$ is not a white gaussian sequence and the estimate is not unbiased. To overcome the problem of bias many variants of recursive parameter estimation algorithms have been suggested of which recursive instrument variable algorithm of Young [25] and generalised least square formulation of Hastings et al and James [30] are important. The essential components of these algorithms are similar. The recursive instrument variable algorithm is easily amenable to computation and as observed by Young [31, 32, 33, 34, 35, 36] it may well offer an unified and comprehensive approach to system identification.

2.3.4. Recursive instrument variable algorithm

Most likely source of biased estimate is the presence of auto correlated noise process $E\left\{ V(t_k)\ V(t_{k-1}) \right\} = 0$ for all k which implies $E\left\{ Y(t_{k-1})\ V(t_k) \right\} \neq 0$, i.e. there is a significant correlation between the noise sequence and the past values of output. Referring to eqn. (2) and with suitable estimate of the parameters in $A(q^{-1})$ and $B_j(q^{-1})$ as $A(q^{-1})$ and $B_j(q^{-1})$ respectively a deterministic time series denoted as auxiliary model can be computed as

$$Y(t_k) = \left[1 - A(q^{-1}) \right] Y(t_k) + \sum_{j=1}^{m} B_j(q^{-1})\ U_j(t_k) \qquad \ldots(2.27)$$

Equation (2.2) and 2.27) suggest that (i) variation in $Y(t_k)$ should be strongly correlated with variations in the noise corrupted output observations $Y(t_k)$ but (ii) these variations in $Y(t_k)$ should be uncorrelated with $V(t_k)$ provided $V(t_k)$ is not correlated with the measured input sequence $U_j(t_k)$ i.e. $E\left\{ U_j(t_k)\ V(t_l) \right\} = 0$ for all j, k, l.

Sequence of $Y(t_k)$ is called the sequence of instrumental variables. Consequently the vector $Z(t_k)$ is modified as $\hat{Z}(t_k)$ defined by

$$\hat{Z}(t_k) = \left[\hat{Y}(t_{k-1}), \ldots \hat{Y}(t_{k-n}), U_1(t_{k-1}), \ldots U_1(t_{k-n}), \ldots U_m(t_{k-n}) \right]^T \qquad \ldots(2.28)$$

Conditions of unbiased estimates are modified as

$$E\left\{ Z(t_k)\ V(t_k) \right\} = 0 \quad \text{for all } k$$

Replacing $Z(t_k)$ by $\hat{Z}(t_k)$ and not $Z^T(t_k)$ by $\hat{Z}^T(t_k)$ heuristically, Recursive Instrument variable Algorithm is given by.

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) + \hat{P}(t_{k-1})\hat{Z}(t_k) \left[1+Z^T(t_k)\hat{P}(t_{k-1})\hat{Z}(t_k)\right]^{-1} \left\{ Y(t_k) - Z^T(t_k)\hat{\alpha}(t_{k-1}) \right\} \quad \ldots \quad (2.29)$$

$$\hat{P}(t_k) = \hat{P}(t_{k-1}) - \hat{P}(t_{k-1})\hat{Z}(t_k) \left[1+Z^T(t_k)\hat{P}(t_{k-1})\hat{Z}(t_k)\right]^{-1} Z^T(t_k)\ \hat{P}(t_{k-1}) \quad \ldots \quad (2.30)$$

with

$$\hat{P}(t_k) = \left[\sum_{j=1}^{k} Z(t_j)Z^T(t_j) \right]^{-1} \quad \ldots \quad (2.31)$$

This recursive algorithm is used to estimate the parameters of a regression relationship which varies with time by passing through time series data and attempting to track the parameter variations. The data may be processed iteratively, each time using a data set in order to further refine the estimates to obtain better statistical efficiency. For a given block of N data elements the recursive method terminate after N steps where as the iterative procedure continues until parameters no longer changes with further iteration. For on line process this can also be used as the basis for continuous updating of the auxiliary model parameters [37].

Sometimes large errors are found between the predicted and the observed output. This is not so much a consequence of spurious errors in the measurement but are due primarily to changing values of the model parameters. In the following section methods for estimating slowly varying parameters are discussed.

2.3.5. Exponential weighting of past data

When the observed process behaviour is changing with time one common approach is to estimate the parameters on the basis of recent past observations and not on distant past observations. In the method of exponential weighting of past data the memory of the estimator is curtailed by decreasing the weight of the past data exponentially instead of giving them equal weightage. In this case the loss function to be minimised is

$$J_{ew}(t_k) = \sum_{j=1}^{k} \mu^{k-j} \left(Y(t_j) - Z^T(t_j) \hat{x}\ (t_k) \right)^2; \quad 0 \ll \mu \ll 1 \qquad \ldots (2.32)$$

This indicates that most recent errors are weighted preferentially and they will thus dominate the computation of $\hat{x}\ (t_k)$.

The choice of μ is dependent on the rate of change of the parameters. Choice of lower value of μ allow faster forgetting of the past data. Following the similar procedure as in the case of development of recursive least square algorithm, recursive exponential weighting of past data algorithm is obtained as

$$\hat{x}\ (t_k) = \hat{x}\ (t_{k-1}) + P'\ (t_{k-1})\ Z(t_k) \left[\mu + Z^T\ (t_k)\ P'\ (t_{k-1})\ Z(t_k) \right]^{-1}$$

$$\left[Y(t_k) - Z^T\ (t_k)\ \hat{x}\ (t_{k-1}) \right] \qquad \ldots (2.33)$$

$$P'\ (t_k) = \frac{1}{\mu} \Big\{ P'\ (t_{k-1}) - P'\ (t_{k-1}) Z(t_k) \left[\mu + Z^T\ (t_k) P'\ (t_{k-1}) Z(t_k) \right]^{-1}$$

$$Z^T\ (t_k) P'\ (t_{k-1}) \Big\} \qquad \ldots (2.34)$$

in which

$$P'(t_k) \triangleq \left[\sum_{j=1}^{k} \mu^{k-j}\ Z(t_j)\ Z^T(t_j) \right] \qquad \ldots (2.35)$$

Recursive exponential weighting of past data model is one method for allowing recursive estimation of time varying parameters. But there are two basic limitations on the use of this method. Firstly it does not allow any prior selection between different expected rates of change of parameter. Secondly this method is appropriate for slowly varying parameters. There are many physical systems where relatively fast disturbance of the system may occure. In the following section a rather different method for estimation of time varying parameters is presented.

2.3.6. Recursive time varying parameter estimation algorithm.

Covariance matrix of least square parameter estimation is given by

$$P^{e}(t_k) = E\left\{ \left[\left(\hat{\alpha}(t_k) - \alpha \right) \left(\hat{\alpha}(t_k) - \alpha \right)^{T} \right] \right\} \quad \ldots \quad (2.36)$$

which is related to the matrix $P(t_k)$ by the simple relation

$$P^{e}(t_k) = \sigma^{2} P(t_k) \quad \ldots \quad (2.37)$$

where $$= E\left\{ V(t_k)\ V(t_k) \right\} \quad \ldots \quad (2.38)$$

In practice σ^{2} would not be known but can be estimated as the variance of the residual error sequence.

$E\left\{ V(t_k) \right\} = Y(t_k) - Z(t_k)\hat{\alpha}$ when $\hat{\alpha}$ have achieved satisfactory convergence. Substituting $P(t_k)$ by $P^{e}(t_k) / \sigma^{2}$ in eqns. (2.2) and (2.21) we get

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) + P^{e}(t_{k-1}) Z(t_k) \left[\sigma^{2} + Z^{T}(t_k) P^{e}(t_{k-1}) Z(t_k) \right]^{-1}$$

$$\left[Y(t_k) - Z^{T}(t_k)\ \hat{\alpha}(t_{k-1}) \right] \quad \ldots \quad (2.39)$$

$$P^{e}(t_k) = P^{e}(t_{k-1}) - P^{e}(t_{k-1}) Z(t_k) \left[\sigma^{2} + Z^{T}(t_k) P^{e}(t_{k-1}) Z(t_k) \right]^{-1}$$

$$Z^{T}(t_k) P^{e}(t_{k-1}) \quad \ldots \quad (2.40)$$

in which

$$P^{-1}(t_k) = \sigma^2 \sum_{j=1}^{k} Z(t_j)\, Z^T(t_j) \qquad \ldots \quad (2.41)$$

Assuming parameters $\alpha(t_k)$ undergoes a general random walk [38] as $\alpha(t_k) = \alpha(t_{k-1}) + \xi(t_k)$ where $\xi(.)$ is a sequence of mutually independent identically distributed random variables. We define a positive matrix D usually diagonal as

$$D = E\left\{ \xi(t_k) \cdot \xi^T(t_k) \right\} \qquad \ldots \quad (2.42)$$

Then for parameter estimation we have,

Prediction :

$$\hat{\alpha}(t_k/t_{k-1}) = \hat{\alpha}(t_{k-1}/t_{k-1}) \qquad \ldots \quad (2.43)$$

$$P(t_k/t_{k-1}) = P(t_{k-1}/t_{k-1}) + D \qquad \ldots \quad (2.44)$$

By suitable substitution in eqns. (2.29) and (2.40) we have correction :

$$\hat{\alpha}(t_k/t_k) = \hat{\alpha}(t_k/t_{k-1}) + P(t_k/t_{k-1})Z(t_k) \left[\sigma^2 + Z^T(t_k)P(t_k/t_{k-1})Z(t_k) \right]^{-1} \left\{ Y(t_k) - Z^T(t_k)\, \hat{\alpha}(t_k/t_{k-1}) \right\} \qquad \ldots \quad (2.45)$$

$$P(t_k/t_k) = P(t_k/t_{k-1}) - P(t_k/t_{k-1})Z(t_k) \left[\sigma^2 + Z^T(t_k)P(t_k/t_{k-1})Z(t_k) \right]^{-1} Z^T(t_k)P(t_k/t_{k-1}) \qquad \ldots \quad (2.46)$$

Combining equations (2.43) and (2.44) with equations (2.45) and (2.46) respectively we get the recursive least square estimation of time varying parameters as

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) + \left\{ P(t_{k-1}) + D \right\} Z(t_k) \left[\sigma^2 + Z^T(t_k) \left\{ P(t_{k-1}) + D \right\} Z(t_k) \right]^{-1} \left[Y(t_k) - Z^T(t_k)\, \hat{\alpha}(t_{k-1}) \right] \qquad \ldots \quad (2.47)$$

$$P(t_k) = \left\{P(t_{k-1})+D\right\} - \left\{P(t_{k-1})+D\right\} Z(t_k) \left[\sigma^2 + Z^T(t_k) \left\{P(t_{k-1})+D\right\} Z(t_k)\right]^{-1} Z^T(t_k) \left\{P(t_{k-1})+D\right\} \quad \ldots (2.48)$$

Specification of D remains a problem for practical use of this sophisticated algorithm.

In the foregoing discussion a comprehensive methodologies of recursive estimation have been presented. Whitehead and Young [39] have demonstrated the feasibility of constructing realistic dynamic stochastic water quality (BOD - DO) models for non-tidal river systems. The models are satisfactorily identified and statistically validated by reference to practical field data of flow BOD-DO in a 55 Km stretch of Bedford Ouse River System in England. Whitehead and Young [40], Young, Whitehead and Beck [41] have been able to demonstrate the particular utility of recursive methods of time series analysis both for identification and estimation of water resources system models. The recursive algorithms have proved to be valuable aids for obtaining relatively efficient estimates of various model parameters in a straight forward and simple manner. Recursive algorithms provide a powerful general methods of data processing well suited to the modeling problem of water resources systems.

Beck [42] has studied the feasibility of using on-line adaptive predictor in the form of recursive exponential weighting of past data algorithm for obtaining the advance estimate of the qualities of the effluent flow from the sewer network of Stockholm, Sweden to the Reppala, Stockholm waste water treatment plant. The input raw material to a waste water treatment plant exhibits large, and generally poorly quantified, variation with time. In particular, rainfall run off can cause gross overloading of the treatment process of the plant. Procedure of constructing on-line adaptive predictor model has two steps :

(i) the parameters of a multiple inputs/single output time series model are recursively estimated at each time step by recursive algorithm of exponential weighting of past data ;

(ii) a forecast of the plant influent flow is then made on the basis of newly updated prediction model. Beck has observed that the one-step ahead forecasts of the plant influent flow obtained from the recursive adaptive predictor closely approach the observed field data. It seems, there, that a stochastic model derived from recursive time series analysis yields useable results. Such an input/output black box model for the relationship between Ui and Y assumes little or no a priori knowledge of the physical laws of the system and takes a relatively macroscopic view of the cause effect relationships involved.

2.4 Multilayer Group Method of Data Handling Algorithm

Introduction

Ivakhnenko's [58] multilayer group method of data handling is a heuristic method of self-organisation of different partial models. This method involves the generation and comparison of all possible combinations of input output and to select the best possible ones according to the criterion of integral square error.

Water quality in rivers is usually measured in terms of dissolved oxygen levels and biochemical oxygen demand loading. Extensive work has been done to develop mathematical models describing variations of D.O. and B.O.D on the basis of mass balances equations and the significance of the different terms appearing in these balance equations has been considered in details in literature [43,44,45,46,47,48]. Very recently schemes of parameter estimation of such models have gained enormous importance [49,50,51,52,53, 54,55,56].

First Koivo and Phillips [51] analysed the problem of estimating the deoxygenation and reoxygenation rates of the classical Streeter-Phelps model [47] from BOD and DO measurements. They also considered the possibility of estimating the deoxygenation rate when only DO measurements are available [52]. Rinaldi and Soncini-Sessa [54] obtained modification of this method to give an estimate of the reoxygenation rate based on the knowledge of the DO data. It is difficult to use all these estimation techniques in practical situation because they require the measurement points to be distributed uniformly along the river, a condition which is hardly possible to satisfy. Moreover, the techniques of parameter estimation of Koivo and Phillips [51,52], Koivo and Koivo [53], Rinaldi, Romano and Soncini-Sessa [54] and Rinaldi et al [55] were verified with "synthetic data" i.e. data generated with random errors (white noise) to the exact solution of the model. Very recently Rinaldi and Soncini-Sessa [67] have presented a new scheme for estimation of the deoxygenation and reoxygenation rates of Streeter-Phelps' model [47]. The scheme is derived from the least squares method and its dimentionality is reduced as much as possible by taking advantages of the mathematical structure of the model. The reliability of the estimation scheme is considered on a data set observed on Bormida river in Northern Italy.

More sophisticated schemes, like Kalman filtering extended Kalman filtering [49, 50] and Instrument variable have been used for parameter estimation of relatively complex waterquality model of a river. In these methods the estimation is carried out in a recursive manner adaptively updating at each iteration the estimate of the parameters on the basis of the last set of data made available.

2.4.1 Brief description of multilayer GMDH

In multilayer group method of data handling algorithms, polynomials are used as the basic means of investigation of complex dynamical systems. The polynomials of prediction are regression equations which connect the current values of output with the current and/or past values of input variables. Regression analysis allows to evaluate the co-efficients of the polynomial by criterion of minimum mean square error. Then the polynomials are treated in the same manner as are seeds in the agricultural selection, an unique mathematical concept propagated and established by Academician A.G. Ivakhnenko and his co-workes [59,60,61] of the Institute of Cybernetics, Kiev, USSR.

Volterra series [62] introduced to non-linear system analysis by Wiener [63], learning filter of Gabor, Wilby and Woodcock [64] and the perception of Rosenblatt [66] has provided the conceptual basis for multilayer GMDH. Astrom and Eykhoff [65] pointed out that problems may arise with the use of volterra series or high degree polynomial to approximate non-linear functions because of the fact that there are many co-efficients to estimate, many data are needed and the computation with the resulting large matrices may be prohibitive. Ivakhnenko's multilayer GMDH algorithms are free of these problems. He models the input output relationships of complex processes using multilayer network structure of Rosenblatt's [67] perception type, who designed the model of brain's perception.

Salient features of multilayer GMDH as applicable in multilayer selection process which is used in the present investigation are briefly described here.

Each output element in the network implements a non-linear function of its inputs. The function implemented is usually a second order polynomial of its inputs. Since each element generally takes

two inputs, the implemented function by an element in one of the layers is given by

$$Y = A_2(x) = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4x_1^2 + a_5x_2^2 \qquad (2.49)$$

Only those elements whose performance indices exceed the thresh hold at that layer are allowed to pass to the next layer. Therefore, the network represents a feed forward transformation whereby each succeeding layer in the network increases by two the degree of the multipolynomial fit to the input properties of x_i. Figure 2.4.1 depicts the concept of feed forward transformation.

The selection hypothesis employed by Ivakhnenko to select the elements to be used in the succeeding layers involves two basic conclusions; the composite character of a system must be based on the use of the signals which control the totality of the elements of the system, and the long history of the art of selection as observed in the case of plants and animals can be successfully extended to the science of engineering cybernetics.

Let us explain the two conditions. To get, say, plants in the agricultural sense with certain specific properties, a large number of plants are sown which may have these properties, and the plants are crossed. From the harvest of the first generation, the plants are chosen which better our requirement (the first self selection) as compared to others. The seeds of the selected plants are again sown and crossed. From the second harvest we select certain seeds and the seeds are sown, and so on. Rules employed in the process of mass selection are as follows :

i) For each generation certain optimal number of seeds are sown.

ii) The selection process cannot be completed in a single generation (at least 3 to 4 generations are needed).

Perception algorithm [66] of Rosenblatt duplicate the above mentioned process. Perception can be used for identification of extremal processes, in control theory sense. The complex surface of extremal heap is approximated by polynomials. The signals applied to the perception input contains information about the surface of interest to us. The surface is usually described by a number of experimental points and simple function of their co-ordinates. In accordance with the selection hypothesis, the simple polynomials of second degree that are easiest to inscribe in the surfaces are taken. The combination of data are subjected to the first threshold selection, in accordance with the integral square error criterion on a separate checking set. Only some of the polynomials which fit best into the sought surface are allowed to pass into the second layer where they form more complex combinations of polynomials of fourth degree. From the second layer again the polynomials which fit best into the sought surface are singled out and are allowed to pass into the third layer and so on. The process continues so long as minimum of a selection criterion is obtained. This constitutes Ivakhnenko's multilayer group method of data handling algorithms [68,69,70].

The co-efficients of each layer in the network are determined in the following manner.

Consider one element in the first layer. It implements the function $A_2(x)$ shown in equation (2.49). The data are divided into two sets — training and checking sets. Assume that there are N-input vectors in the training set each one of them is composed of property values.

$$X_n = (x_{n1}, x_{n2} \ldots, x_{np})$$

$$n = 1,2 \ldots, N$$

Denote the n-th desired output as ϕ_n. A set of six co-efficients for the element (which has inputs as x_{ni} and x_{nj}) must be obtained so that the integral square error between the outputs of this element Y_n and the true outputs ϕ_n on a separate checking

set is minimized. The co-efficients are obtained by solving Gauss Normal equations. The system of equation are written as

$$\begin{aligned} \phi_1 &= a_0 + a_1 x_{1i} + a_2 x_{1j} + a_3 x_{1i} x_{1j} + a_4 x_{1i}^2 + a_4 x_{1i}^2 + a_5 x_{1j}^2 \\ &\cdot \\ &\cdot \\ \phi_N &= a_0 + a_1 x_{Ni} + a_2 x_{Nj} + a_3 x_{Ni} x_{Nj} + a_4 x_{Ni}^2 + a_5 x_{Nj}^2 \end{aligned} \tag{2.50}$$

in the matrix form $\phi = XA$ where ϕ, X and A are N x 1, N x 6 and 6 x 1 matrices respectively (the first element of each row of x matrix is unity). Vector A contains a set of six co-efficients which enables this element to be approximated to the true outputs in accordance with the method of least square. While estimating the co-efficients it has been intuitively assumed that the equation error is a white noise process with zero mean, constant variance and uncorrelated with inputs, and it is significantly small. The second assumption is that the inputs and outputs are exactly known without any measurement error [71].

This process is repeated for each element in the first layer with the exponents in matrix x changing each time depending on the identity of two inputs to the particular elements. The same technique is repeated to find the six co-efficients of each element in the succeeding layers. After the values of the co-efficients are obtained based on the training set data, the performance index of a given element is determined by computing the integral square error between the output of each element and the true output on the basis of an independent checking set. Only those elements whose performance index are satisfactory are allowed to pass to the next layer. The use of an independent checking set helps filter out the noisy portion of the data set. Figure 2.42 shows a flow chart of multilayer GMDH algorithm - the Ivakhnenko's theory of self-organisation.

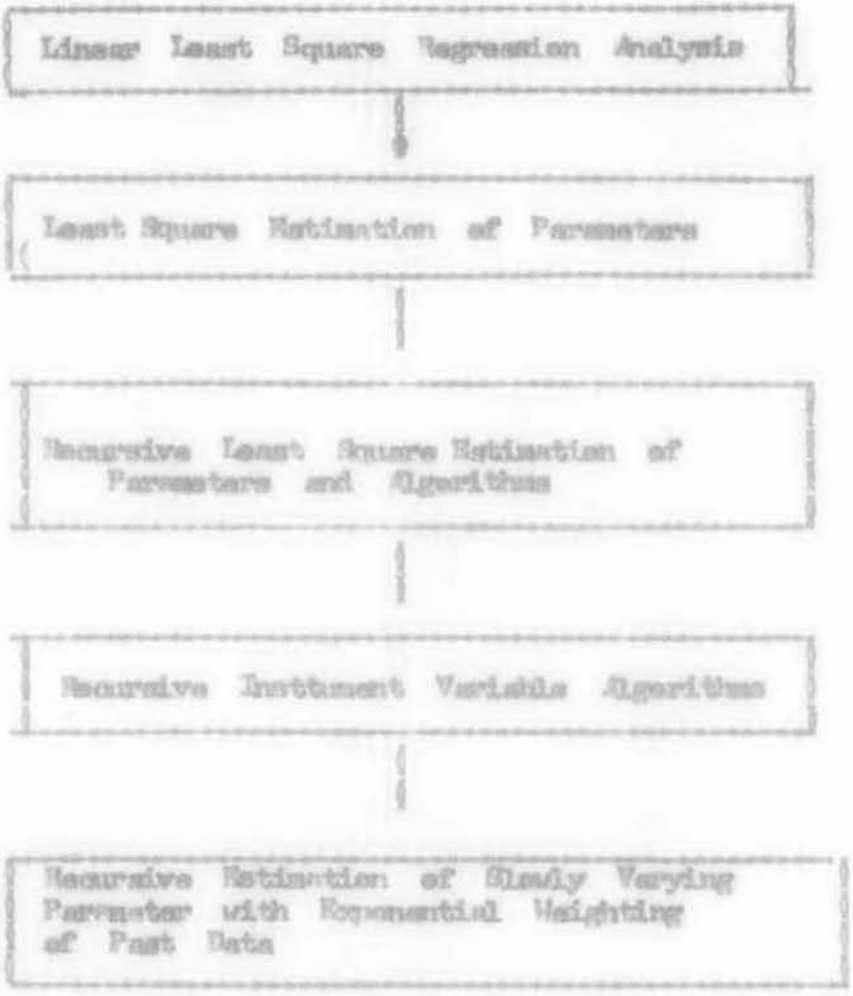


Figure 2.3.0 Steps of Development of the Recursive Parameter Estimation Algorithm.

Computational method for multilayer GMDH algorithm has been briefly described in the ensuing section.

The complete description of any process is given by

$$\phi = f(x_1, x_2, x_3, \ldots, x_n) \qquad \ldots (2.51)$$

The process is to be constructed of several layers of partial description of two input variables taken at a time.

$$y_1 = f(x_1, x_2),\ y_2 = f(x_3, x_4) \ldots y_m = f(x_{n-1}, x_n)$$

$$m = \frac{n!}{2!\,(n-2)!} \qquad \ldots (2.52)$$

$$z_1 = f(y_1, y_2),\ z_2 = f(y_3, y_4) \ldots z_p = f(y_{m-1}, y_m)$$

$$p = \frac{m!}{2!\,(m-2)!}$$

and so on, where m and p are the number of pairwise combinations of first and second layer respectively.

Inputs which have strong correlation with the output are selected. Correlation functions are defined as

$$k_{yy}(\lambda) = \frac{\sum_{i=1}^{N-\lambda} (y(i) - \bar{y})\ (Y(i+\lambda) - \bar{y})}{\left[\sum_{i=1}^{N-\lambda} (y(i) - \bar{Y})^2 \sum_{j=1+\lambda}^{N} (y(j) - \bar{Y})^2\right]^{\frac{1}{2}}} \qquad \ldots (2.53)$$

$$K_{yx}(\lambda) = \frac{\sum_{i=1}^{N-\lambda} (Y(i) - \bar{y})\ (x(i+\lambda) - \bar{x})}{\left[\sum_{i=1}^{N-\lambda} (Y(i) - \bar{y})^2 \sum_{j=1+\lambda}^{N} (x(j) - \bar{x})^2\right]^{\frac{1}{2}}} \qquad \ldots (2.54)$$

where $K_{yy}(.)$ and $K_{yx}(.)$ are autocorrelation and cross correlation of output and output-input respectively for different lag λ, $\lambda = 0,1,2, \ldots M$; N = number of data points.

Data are rationalised in the form

$$x(k) = \frac{X(k) - X(min)}{X(max) - X(min)} \qquad \ldots \quad (2.55)$$

where X(k) is the actual value of data at the k-th instant of time. All the experimental data are divided into two sets in ratio of approximately 2:1 in training and checking sets respectively. Data with higher values of variations are kept for training set and those with lower values for checking set. The variation is defined as

$$D^2 = \frac{1}{n}\left[x_1^2 + x_2^2 + \ldots + x_n^2 \right] \qquad \ldots \quad (2.56)$$

The co-efficients of the first layer of partial description to given as shown in the equation.

$$y_a = a_{oa} + a_{1a}x_b + a_{2a}x_c + a_{3a}x_b x_c + a_{4a}x_b^2 + a_{5a}x_c^2 \qquad \ldots \quad (2.57)$$

where a_a is the number of combinations and b, c are indices of combinations of input variables taken two at a time. The co-efficients are computed by solving a system of normal Gaussian equations. The left hand sides of the equation are set equal to the values of output at every points. After finding the values of the co-efficients the values of the intermediate variables are obtained. Then using the data of the checking set the integral square error between the intermediate variables and the true output (from checking set) is determined. Only the variables which give low error are selected for subsequent use. Those variables are retained in the training and checking set. Variables with high error figure are discarded. The number of intermediate variables should be kept same as the number of input variables. In the second layer of selection the co-efficients of the partial description.

$$z_a = b_{oa} + b_{1a}y_b + b_{2a}y_c + b_{3a}y_b y_c + b_{4a}y_b^2 + b_{5a}y_c^2 \qquad \ldots \quad (2.58)$$

of the layer are calculated and the accuracy is checked again to select the accurate intermediate variables of the layer. The process of selection continues so long as the integral square error on a checking set comes to a minimum and in the next layer starts increasing. Thus multilayer GMDH comes to practical convergency.

The integral square error criterion is defined as

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N1} (Yi\ (\text{checking}) - Yi(\text{model}))^2}{\sum_{i=1}^{N1} (Yi(\text{checking}))^2} \quad \ldots \quad (2.59)$$

Every intermediate variable is examined for its effect on prediction accuracy. The training set is used for finding the co-efficients of the partial description, whereas the checking set is used to evaluate the quality of partial description. Thus multilayer GMDH has inherent decision regularisation.

Polynomial description of the process is obtained in the form of partial description of intermediate variables of different layers. Eliminating the intermediate variables the complete polynomial description of the process is obtained in the form of Gabor - Kolmogorov type of polynomial as

$$Y = a_0 + \sum_{i=1}^{N} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k}^{n} a_i a_j a_k x_i x_j x_k \quad \ldots \quad (2.60)$$

With the help of multilayer GMDH algorithms Ivakhnenko obtained the polynomial description of British economy for prediction and control on the basis of characteristic variables established by Parks and Pyatt [72]. Heuristic self-organisation method

proposed by Ivakhnenko in GMDH algorithms has been used in a modified form by Ikeda, Ochiai and Sawaragi [73, 74] for developing a non-linear river flow model from the available data of river flows and mean areal precipitation. It is observed from the numerical comparisons made between the prediction model by GMDH and by elaborate hydraulic methods, that there are significant improvements in the heuristic prediction algorithms for real time computation. Tamura and Kondo [75] have used GMDH algorithms for identifying spatial pattern of air pollution concentration in a large area. The heuristic GMDH algorithms have been used by Duffy and Franklin [76, 77] to model an environmental system producing high nitrate level in agricultural drain water in the corn belt in the United States. The method amounts to fitting a polynomial to the multiinput single output response surface. They observed that the GMDH is advantageous with systems characterised by complexity with many variables and parameters, ill defined mathematical structures and limited data. These algorithms are useful for empirically generating hypothesis about which relatively little is known. R.K. Mehra [78] has employed GMDH for forecasting wheat crop yield using weather data. A comparison of the results with Baier [79] shows that crop prediction using GMDH compares favourably with the results obtained using theoretical - empirical models based on over ten years of research. The structural information obtained from GMDH as to which input variables have significant effect on wheat crop yield is also quite significant. Mehra has suggested the use of all data points alternatively as training and checking sets. This technique is expected to give good results. Chaudhuri [80] has used GMDH for identification of the interactions of meteorological processes on monthly tea crop production. It is observed that multilayer GMDH gives good prediction results, identifies the significant variables, and gives an insight into the controlling aspects to adhere to a desired level of tea crop production.

2.5 Combinational Group Method of Data Handling Algorithm

With the theory of self-organisation [81, 82] using heuristic learning algorithm commonly known as group method of data handling it has been possible to formulate mathematical models for complex processes with prediction optimization.

The concept of self-organization can be illustrated as follows. When the model complexity gradually increases the computer finds by shifting the different models, the minimum of a selection criterion which the computer has been ordered to look for. Thus the computer indicates to the operator the model of optimum complexity.

2.5.1 Process equation

The physical process involved in a storm period is stochastic in nature. The process can be represented in the form of a finite order stochastic difference equation of the type as

$$y_{(k)} = f\ (x_1(k-n),\ x_2(k-m),\ x_3(k-p),\ \dots) \qquad \dots\ (2.61)$$

$n, m, p, \dots$, are the instants of the
$x_1, x_2, x_3, \dots$, respectively

which have highest correlation with $y_{(k)}$. We write

$y_{(k)} = y$, flow at the k-th instant

$x_1(k-n) = x_1$, flow at a up-stream gauging station 1 which has highest correlation at lag instant n

$x_2(k-m) = x_2$, flow at a up-stream gauging station 2 which has highest correlation at lag instant m, and so on.

Let us assume

$$x_i(k-n) \text{ as } i = 1,2,3,4,5;\ j = n, m, p, q, r$$

so the process equation becomes

$$y=f(x_1, x_2, x_3, x_4, x_5) \qquad \dots\ (2.62)$$

The function f(.) in equation (2.62) is sought in the class of quadratic polynomials on the basis of a table of polynomials of gradually increasing complexity of five variables as shown in table 2.5.1 with the theory of self-organization of different mathematical models.

The model of optimum complexity is selected on the basis of minimum of integral square criterion. Integral square error is defined as

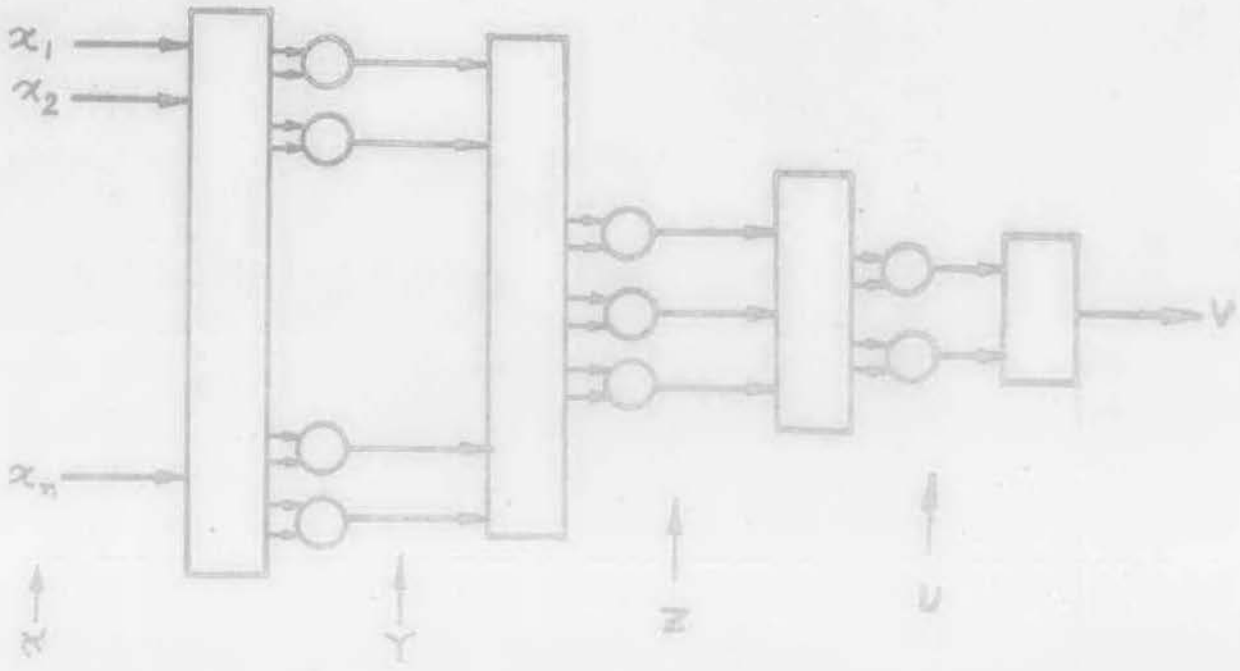


FIG. 2.4.1

Concept of feed forward transformation.

$$r^2 = \frac{\sum_{i=1}^{N} \left(y_{tab}(i) - y_{d.m}(i) \right)^2}{\sum_{i=1}^{N} \left(y_{tab}(i) \right)^2} \qquad \ldots (2.63)$$

where $y_{tab}(i)$, i=1, 2 ..., N hours, are the tabulated values of the output variable in the interpolation region and $y_{d.m}(i)$ are the values of the variable obtained from the model.

2.6. Real Time Recursive Parameter Tracking algorithms.

In the preceding section a comprehensive mathematical model for a non-stationary process has been developed in line with the procedures suggested by Box and Jenkins [1], Jenkins and Watts [83], Dixon [84], Kashyap and Rao [85, 86] and Chaudhuri [87]. The process parameters are evaluated with an on-line tracking algorithm by minimising a quadratic cost function.

2.6.1 Processing of the Statistical Data

The observed data of a physical process are P(i), i=1,2,.. N1 the time instants.

The mean value is

$$\bar{P}(N1) = \frac{1}{N1} \sum_{i=1}^{N1} P(i) \qquad \ldots (2.64)$$

Auto covariance of the data at lag instant k is given by

$$CA(k) = \frac{1}{N1-k} \sum_{i=1}^{N1-k} (P(i)-\bar{P}(N1)) (P(i+k) - \bar{P}(N1)) \qquad \ldots (2.65)$$

Where $k = 0, 1, 2 \ldots, M$; $M < \frac{1}{4} N1$

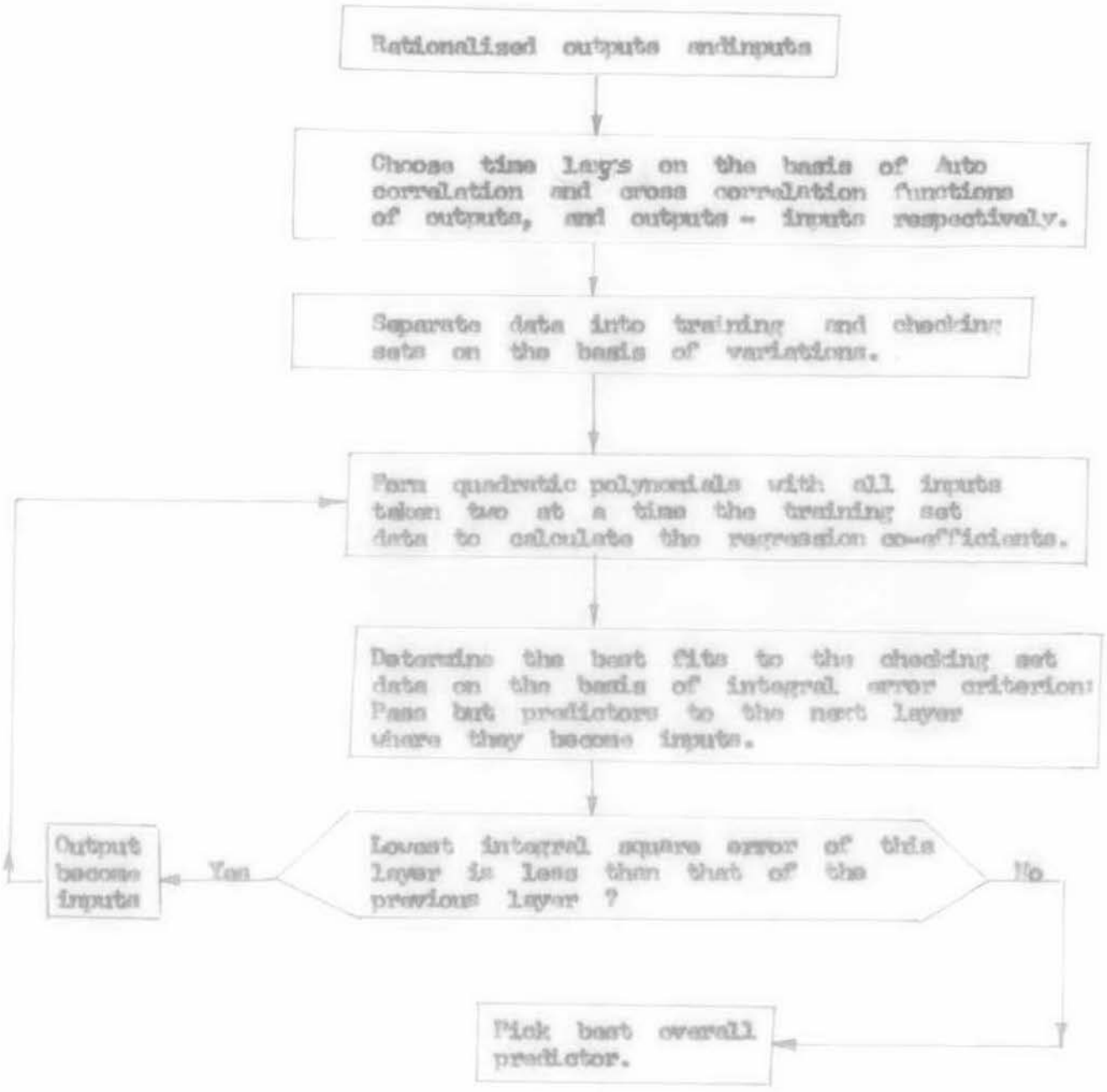


Figure 2.42 - Flow chart of multilayer GMDH algorithm.

Normalised covariance co-efficient is

$$RA(k) = \frac{CA(k)}{CA(o)} \qquad \ldots \qquad (2.66)$$

Estimate of the normalised power density spectra for the data are given by,

$$PS(\omega_h) = \frac{2}{M} \sum_{k=o}^{M} E_k \, RA(k) \cos \omega_h k \qquad \ldots \qquad (2.67)$$

Where $\omega_h = 2\pi f_h$ and $0 \leq f_h \leq 0.5$, $h = 0, 1, 2, \ldots, M$

E_k is defined as weight for window correction

$$E_k = \begin{cases} 1.0 \; ; & 0 < k < M \\ 0.5 \; ; & k = 0, M \end{cases} \qquad \ldots \qquad (2.68)$$

These raw estimate of power spectral density are smoothed by using Hanning Window to obtain the final estimates of the power spectrum. The smoothed estimates of the ordinates of power spectrum are given by

$$\begin{aligned} h = 0; \quad & S(\omega_o) = 0.54 \, PS(\omega_o) + 0.46 \, PS(\omega_1) \\ 0 < h < M, \quad & S(\omega_h) = 0.23 \, PS(\omega_{h-1}) + 0.54 \, PS(\omega_h) + 0.23 PS(\omega_{h+1}) \\ h = M, \quad & S(\omega_h) = 0.54 \, PS(\omega_M) + 0.46 \, PS(\omega_{M-1}) \end{aligned} \qquad \ldots \qquad (2.69)$$

Periodicity in terms of fundamental and its harmonics can be estimated from the power spectral density vs frequency characteristics.

2.6.2. Development of the Non-stationary process model.

Non-stationary process with deterministic periodicity can be expressed as

$$P(k) = U(k) + V(k) \qquad \ldots \qquad (2.70)$$

where

U(k) is the deterministic trend component,

V(k) is the stationary stochastic component,

k = 1,2 ... are time instances.

Deterministic periodic component can be represented as,

$$U(k) \simeq \sum_{l=0}^{L} \left(a_l \cos\left(l\, 2\pi\, f_0 k \right) + b_l \sin\left(l\, 2\pi\, f_0 k \right) \right) \quad \ldots \quad (2.71)$$

where f_0 is the fundamental cycle per instance of time.
Stochastic component V(k) can be expressed in the autoregressive form of,

$$V(k) \simeq \sum_{n=1}^{N} C_{k-n}\, V(k-n) + \eta_k \quad \ldots \quad (2.72)$$

V(k) can also be expressed as,

$$V(k-n) = P(k-n) - U(k-n) \quad \ldots \quad (2.73)$$

Hence, equation (2.72) can be rewritten as,

$$V(k) \simeq \sum_{n=1}^{N} C_{k-n}\, P(k-n) - \sum_{n=1}^{N} C_{k-n}\, U(k-n) + \eta_k \quad \ldots \quad (2.74)$$

It follows from equation (2.70) and (2.71)

$$P(k) \simeq \sum_{n=1}^{N} C_{k-n} P(k-n) + \sum_{n=0}^{N} \sum_{l=0}^{L} \left[a_{k-n}^{l} \cos\left(l\, 2\pi f_0 (k-n) \right) \right.$$

$$\left. + b_{k-n}^{l} \sin\left(l 2\pi f_0 (k-n) \right) \right] + \eta_k \quad \ldots \quad (2.75)$$

In matrix form process P(k) can be expressed as

$$P(k) = a^{T}(k-1)\; Z(k-1) + \eta_k \quad \ldots \quad (2.76)$$

where

$$a^T(k-1) = \left[C_{k-1}, \ldots, C_{k-n}, \sum_{n=0}^{N} (d^o_{k-n}, \ldots, d^L_{k-n}, \ldots, g^L_{k-n}) \right]$$

and

$$Z(k-1) = \left[P(k-1), \ldots, P(k-n), 1 \ldots, \right.$$

$$\left. \mathrm{Cos}\,(\ell\,(2\pi\, f_o(k-n)), \ldots, \mathrm{Sin}\,(\ell\, 2\pi\, f_o(k-n)) \right]^T$$

$\hat{P}(k)$, the estimate of $P(k)$ can be very clearly written as

$$\hat{P}(k) = \begin{bmatrix} a_o \\ c_{k-1}^1 \\ c_{k-2}^2 \\ \cdot \\ \cdot \\ \cdot \\ d_{k-1}^1 \\ d_{k-1}^1 \\ \cdot \\ \cdot \\ \cdot \\ d_{k-1}^2 \\ d_{k-1}^2 \\ g_k^1 \\ g_{k-1}^1 \\ \cdot \\ \cdot \\ \cdot \\ g_k^2 \\ g_{k-1}^2 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}^T \begin{bmatrix} 1 \\ P(k-1) \\ P(k-2) \\ \cdot \\ \cdot \\ \cdot \\ \text{Sin } w_o(k) \\ \text{Sin } w_o(k-1) \\ \cdot \\ \cdot \\ \cdot \\ \text{Sin } 2w_o(k) \\ \text{Sin } 2w_o(k-1) \\ \text{Cos } w_o(k) \\ \text{Cos } w_o(k-1) \\ \cdot \\ \cdot \\ \cdot \\ \text{Cos } 2w_o(k) \\ \text{Cos } 2w_o(k-1) \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \quad \ldots(2.77)$$

As suggested by Eykhoff [27] the co-efficient vector 'a' can be estimated minimising the quadratic function $J_k(a)$,

$$J_k(a) = \sum_{j=1}^{k} (P(i)-a^T x(j-1))^2 + (a-a(o))^T S^{-1}(o) (a-a(o)) \quad \ldots (2.78)$$

where a(o) is the available a prior estimate of the co-efficient vector a and S(o) is the positive definite weighting matrix of the order of mxm, where

$$m = (N+1)(2L+1) \quad \ldots (2.79)$$

N is the predictory interval

and L is the number of harmonics including the fundamental.

For minimisation,

$$\frac{\partial J_k(a)}{\partial a} = -2 \sum_{j=1}^{k} x(j-1)(P(j)-a^T x(j-1)) + 2S^{-1}(o)(a-a(o)) \quad \ldots (2.80)$$

Since we are seeking for a minimum,

$$\frac{\partial J_k(a)}{\partial a} = 0.0 \quad \ldots (2.81)$$

It follows from equation (2.81)

$$\sum_{j=1}^{k} x(j-1)P(j) + S^{-1}(o)a(o) = \sum_{j=1}^{k} x(j-1)x^T(j-1)a + S^{-1}(o)a \quad \ldots (2.82)$$

Let $$S^{-1}(k) = \sum_{j=1}^{k} x(j-1)x^T(j-1) + S^{-1}(o) \quad \ldots (2.83)$$

and

$$d(k) = \sum_{j=1}^{k} x(j-1)P(j) + S^{-1}(o)\, a(o) \quad \ldots (2.84)$$

Denoting the estimate of the co-efficient vector 'a' as a(k) at time instant k,

$$S^{-1}(k)\, a(k) = d(k)$$

$$a(k) = S(k)\, d(k) \qquad \ldots (2.85)$$

From equations (2.83) and (2.84) the following recursive equations are obtained,

$$S^{-1}(k+1) = S^{-1}(k) + x(k)\, x^T(k) \qquad \ldots (2.86)$$

$$d(k+1) = d(k) + x(k)\, P(k+1) \qquad \ldots (2.87)$$

A matrix M has been defined as

$$M = S(k) - S(k)x(k)x^T(k)S(k)(1+x^T(k)S(k)x(k))^{-1} \qquad \ldots (2.88)$$

Therefore $M\, S^{-1}(k+1) = I+S(k)x(k)x^T(k) - [S(k)x(k)x^T(k) + S(k)x(k)x^T(k)S(k)x(k)x^T(k) \times \{1+x^T(k)S(k)x(k)\}^{-1}]$ ... (2.89)

or $MS^{-1}(k+1) = I$ ... (2.90)

or $M = S(k+1)$ ... (2.91)

So $S(k+1) = S(k)-S(k)Z(k)Z^T(k)S(k)(1+Z^T(k)S(k)Z(k))^{-1}$ ... (2.92)

From equation (2.85) it follows

$$a(k+1) = S(k+1)\, d(k+1) \qquad \ldots (2.93)$$

Substituting equations (2.87) and (2.92) in equation (2.93)

$$a(k+1) = [S(k)-S(k)x(k)x^T(k)S(k)\ (1+x^T(k)S(k)x(k))^{-1} (d(k)+x(k)P(k+1))] \qquad \ldots (2.94)$$

It follows from equation (2.94)

$$a(k+1) = a(k)+S(k)x(k)\,[1+x^T(k)S(k)x(k)]^{-1} \times [P(k+1) - x^T(k)\, a(k)] \qquad \ldots (2.95)$$

Post multiplication of equation (2.92) by Z(k) gives,

$$S(k+1)\,Z(k) = S(k)\,Z(k) - \frac{S(k)Z(k)Z^T(k)S(k)Z(k)}{1+Z^T(k)S(k)Z(k)} \quad \ldots \quad (2.96)$$

$$\text{i.e. } S(k+1)\,Z(k) = \frac{S(k)\,Z(k)}{1+Z^T(k)S(k)Z(k)} \quad \ldots \quad (2.97)$$

So, after substitution of equation (2.97) in equation (2.95) we get,

$$a(k+1) = a(k) + S(k+1)Z(k)\left[P(k+1) - a^T(k)Z(k)\right] \quad \ldots \quad (2.98)$$

Following equations are the recursive algorithm for estimating the parameters of the non-stationary model.

$$a(k+1) = a(k) + S(k+1)Z(k)\left[P(k+1) - a^T(k)Z(k)\right] \quad \ldots \quad (2.99a)$$

$$S(k+1) = S(k) - S(k)Z(k)Z^T(k)S(k)\left[1+Z^T(k)S(k)Z(k)\right]^{-1} \quad \ldots \quad (2.99b)$$

Periodicity L as suggested by Chaudhuri [87] may be estimated from the power density spectra versus cycle per time instance. Initial value of the predictory interval N is estimated from the correlation function. The correlation function is difined as

$$\phi_{pp}(\lambda) = \frac{\sum_{i=1}^{N-\lambda} (P(i)-\bar{P}(N))(P(i+\lambda)-\bar{P}(N))}{\left[\sum_{i=1}^{N-\lambda} (P(i)-\bar{P}(N))^2 \sum_{j=1+\lambda}^{N} (P(j)-\bar{P}(N))^2\right]^{\frac{1}{2}}} \quad \ldots \quad (2.100)$$

where λ is the shift of instance of time.

For estimating the parameters the algorithms are initialised as, for $k = 0$

$$S(o) = I$$

$$a(o) = 0$$

and

$$P(j) = 0, \text{ for } j = o, -1, -2, \ldots$$

The non-stationary model of optimum complexity is obtained be comparing the criterion of integral square error between the predicted values and the observed values in the interpolation region for different values of the prehistory interval N; though the initial value of N is estimated from the plot of correlation function versus shift of instances of time. Chaudhuri [87] has observed that the optimum value of N is closer to that estimated from the correlation function when the correlation function approaches zero. Criterion of the integral square error is defined as

$$\sigma^2_{\hat{P}} = \frac{\sum_{i=1}^{M} (P(i) - \hat{P}(i))^2}{\sum_{i=1}^{M} (P(i))^2}$$

Assuming some amount for heuristicity Kashyap and Rao [88, 89, 90] have used similar type of algorithms for estimating the parameters of stochastic models of river flows and ground water levels. In their work Kashyap and Rao have postulated stochastic models with sets of undetermined parameters and these parameters are estimated in real-time using available data of river flows and ground water level. The predictors are recursive and yield the predicted flows and ground water level at one step ahead instance of time based upon all presently available information. In his work Chaudhuri [87] has followed the recursive algorithm of Kashyap and Rao for estimating the parameters of non-stationary process of monthly tea crop production. Their techniques have been further improved by Chaudhuri without assuming any heuristicity. In his work Chaudhuri has suggested a new method for ascertaining the optimum number of terms incorporating the values of the past instances of time in the autoregressive portion of the model of the monthly tea crop production. Thus this method determines the model of optimum complexity.

2.7 Concept of Fuzzy Control and Suggestive Examination of its Practical Implementation.

Fuzzy set theory [91,92,93,94,95,96,97], originally developed by Zadeh, deals with the process containing uncertainties only qualitatively described. The proposed model to describe such processes is based on the association of a membership function (A) to each set A of imprecisely defined quantities and the use of logic operations on the measures of membership for fuzzy reasoning. Because of the qualitative nature of the processes involved, linguistic methods developed for advanced software are also applicable to fuzzy analysis and decision making. Its adaptability and simplicity of implementations on a digital computer using the linguistic version of fuzzy set theory make it a very promising approach to the problem of control with uncertainties.

2.7.1 Approach of Fuzzy set theory.

In the absence of clearly defined objectives or imprecisely defined objectives the conventional quantitative technique of applied cybernetics are unsuitable or are unable to give any tangible result. As the complexity of a system increases the imprecision in objective creeps in. And this is defined as the fuzzyness in the system's behaviour.

Viewed in this perspective, the traditional techniques of system analysis are not well suited for dealing with complex system with imprecisely defined objectives because they fail to come to the grips of the fuzziness of the system's behaviour. To deal with such system in realistic sense we need approaches which employ methodological framework quite tolerant of imprecision in objectives, rather than rigorous mathematical formulation.

Three main distinguishing features which define the fuzzy approach are :- (i) Use of linguistic variables in place of or in addition to numerical variables, (ii) Characterization of simple

relations between variables by conditional fuzzy statements, and (iii) characterisation of complex relation by fuzzy algorithms.

A linguistic variable is defined as a variable whose values are a collection of words in a natural or artificial language. To illustrate, the values of the fuzzy variable height might be expressed as tall, not tall, etc. Thus height as a linguistic variable. The linguistic variables provide systematic means for an approximate characterisation of complex system with ill defined objectives.

In quantitative approach to system theory, a dependence between two numerically valued variables x and y is usually characterised by a conditional statement as IF x is 5 THEN y is 8, IF x is 10 THEN y is 16 etc. If x and y are fuzzy variables the fuzzy conditional statements are of the form IF x is LARGE THEN y is Small.

A fuzzy algorithm is an ordered sequence of instructions which may contain fuzzy conditional statements, e.g.
Reduce x slightly if y is large.
Increase x very slightly if y is not very large and very small
If x is small then stop; otherwise increase x by 2.
The execution of such instruction is governed by the compositional rule of inference.

Let X be a space of points with generic elements of X denoted by x, $X = \{x\}$

A fuzzy set A in X is characterised by a membership function $f_A(x)$ which associates with each point in X a real number in the interval $[0, 1]$ with the value of $f_A(x)$ at x representing the grade of membership of x in A.

Union of two fuzzy sets A and B with respective membership functions $f_A(x)$ and $f_B(x)$ is fuzzy set C, written as $C = A \cup B$ whose membership function is related to those of A and B by

$$f_C(x) = \text{Max}\left[f_A(x), \quad f_B(x) \right], \quad x \in X$$

or, in abbreviated from

$$f_C \;\&\; f_A \;\vee\; f_B$$

Intersection of two fuzzy sets A & B with respective membership function $f_A(x)$ and $f_B(x)$ is a fuzzy set C, written as $C = A \cap B$, whose membership function is related to those of A & B by $f_C(x) = \min \left[f_A(x), f_B(x) \right], \quad x \in X$

or, in abbreviated form

$$f_C = f_A \wedge f_B$$

other properties of fuzzy control have been discussed in references 91 to 96.

A few cases of real time control of water quality of non tidal river have been demonstrated in practice. It is due to the lack of clearly defined objectives for the management of water quality in a river system and also a lack of precisely specified standards of stream quality to be maintained by the application of real time control. Beck [98] has suggested the use of a Fuzzy Controller Synthesising different Fuzzy Control algorithms for practical implementation of water quality management of a river. Tong [99] has emphasised that further computational efforts are needed to implement Fuzzy Control algorithm in practice to synthesise a Fuzzy Controller.

2.8. Conclusion

The investigator has given an exhaustive exposition of the state-of-the-art of the different methods of applied cybernetics which have either been used or have the potential application in water resources system.

In the chapters which follow the methods of cybernetics are used for identification, modelling and prediction of flows and water quality of non-tidal rivers. Wherever necessary analytical modification and computational improvement are incorporated.

CHAPTER XXI

FORECASTING OF RIVER FLOWS DURING A SHORT SPAN STORM PERIOD

CHAPTER III

FORECASTING OF RIVER FLOWS DURING A SHORT SPAN STORM PERIOD

3.0 Abstract:

Monitoring of river flows during a short-span storm period is necessary for flood warning as well as for planning and control of water resources systems. The importance of learning identification algorithms of combinatorial and multilayer group method of data handling techniques has been successfully demonstrated for forecasting of hourly flows of a hilly river of the Himalayan region during a short-span monsoon storm period. For real time application the recursive instrument variable algorithm has been found to be highly satisfactory.

3.1 Introduction :

The present work uses methods of Cybernetics in identification modelling and prediction of river flows during a short span storm period.

Cybernetics deals with the purposeful analysis of complex systems with a view to implement system's control mechanism with optimally required human decision. Cybernetic methods are important tools which may be effectively used in the development and in the forecasting, planning and control of water resources systems.

The present section deals with the monitoring of river flows during a short span storm period [100]. The work consists of three parts.

In part one a dynamic model for hourly river flows with a minimum of sixth lag instance in the measurement of up-stream flow using a learning identification algorithm has been presented.

With the theory of self-organisation using heuristic learning algorithm commonly known as combinational group method of data handling it has been possible to formulate mathematical models for complex processes with prediction optimisation. The concept of self-organisation can be illustrated as follows : when the model complexity gradually increases the computer finds by shifting the different models, the minimum of a selection criterion which the computer has been ordered to look for. Thus the computer indicates to the operator the model of optimum complexity. The purpose of this part of the work is to obtain an easily usuable model of optimum complexity for hourly prediction of the flow of river Teesta at Domohani Road Bridge at Jalpaiguri in North Bengal for a storm period on the basis of the up-stream flow at five gauging stations in the Himalayan region with a minimum lag of six hourly time instants.

In part two of the work a dynamic model for hourly river flows by a learning identification algorithm commonly known as multilayer group method of data handling algorithm, correlating lagged instant flows at different up-stream gauging stations has been presented. The model has been verified by field measurements.

In part three of the work real-time estimation of river flow during a short span storm period has been presented. The operation of the large scale water resources systems demands that the real-time monitoring of river flow during a storm period and the non-line control of river through different hydraulic structures requires a consistent and reliable set of data representing the predicted flow on the basis of part measurements and the system parameters. This paper presents in a comprehensive and unified manner the role of recursive estimation techniques for real-time forecasting of river flows during a storm period. The effectiveness of the recursive estimation algorithms have been investigated for real-time prediction of flows of the river Teesta in North Bengal during a short span storm period.

In the design of large scale water resources system needed to meet the rapid expansion of irrigation and hydropower generation, the trend is towards more advanced engineering operation in terms of monitoring, security and control of river flows during a short span storm period. This requirements, so necessary to provide a reliable decision methodology, have placed greater demands on the quantity and quality of the real-time prediction of river flows.

The basic function of an estimation program is to convert the raw data on measurements and status of the system into a reliable real-time information regarding the state and the status of the system to provide the means for a more reliable and effective closed loop decision algorithms for controlling the different hydraulic structures of a large scale water resources system. In the present work highly sophisticated recursive algorithms for process parameter estimation have been used for predicting hourly flow of the river Teesta at Domohani Road Bridge for a short span storm period.

3.2. Part One : Combinatorial Group Method of Data Handling Algorithm in Formulating Six-hourly Flow Forecasting Model

Process equation

The physical process involved in a storm period is stochastic in nature. The process can be represented in the form of a finite order stochastic difference equation of the type as

$$y_{(k)} = f(x_1(k-n), x_2(k-m), x_3(k-p), \ldots) \quad \ldots (3.1)$$

$n, m, p, \ldots,$ are the instants of the

$x_1, x_2, x_3, \ldots,$ respectively

which have highest correlation with $y_{(k)}$.

The arguments having a strong correlation with y_k are then selected for inclusion in the process equation on the basis of the correlation functions for the time shift λ defined as

$$K_{yx}(\lambda) = \frac{\sum_{i=1}^{N-\lambda} (y(i)-\bar{y})\,(x(i+\lambda)-\bar{x})}{\left[\sum_{i=1}^{N-\lambda} (y(i)-\bar{y})^2 \sum_{j=1+\lambda}^{N} (x(j)-\bar{x})^2\right]^{\frac{1}{2}}} \qquad \ldots \quad (3.2)$$

where (bar) indicates the mean value and N is the number of data points.

$y_{(k)} = y_k$ flow at Dombund at the k-th instant

$x_1(k-n) = x_{1k}$ flow at a up-stream gauging station 1 which has highest correlation at lag instant n.

$x_2(k-m) = x_{2k}$ flow at a up-stream gauging station 2 which has highest correlation at lag instant m_k and so on.

so the process equation becomes

$$y = f(x_1, x_2, x_3, x_4, x_5) \qquad \ldots \quad (3.3)$$

The function f (.) in equation (3.2) is sought in the class of quadratic polynomials on the basis of a table of polynomials of gradually increasing complexity of five variables as shown in Table 2.5.1 with the theory of self-organization of different mathematical models.

The model of optimum complexity is selected on the basis of minimum of integral square criterion. Integral square error is defined as

$$r^2 = \frac{\sum_{i=1}^{N} (y_{tab}(i) - y_{d.m}(i))^2}{\sum_{i=1}^{N} (y_{tab}(i))^2} \quad \ldots \quad (3.4)$$

where $y_{tab}(i)$, i=1,2, ..., N hours, are the tabulated values of the output variable in the interpolation region and $y_{d.m}(i)$ are the values of the variable obtained from the model.

Illustration

Figure 3.2.1 shows the map of the Teesta basin region under study. The gauging stations are Domohani, Coronation Bridge, Teestabazar, Tangpo (Teesta), Great Rangeet (Sivalabazar) and Sankalan denoted respectively by y, x_1, x_2, x_3, x_4 and x_5 . The storm period from 16.00 hours on 22.7.79 to 16.00 hours on 25.7.79 was considered for the study.

The hourly data were rationalised as

$$x_{(k)} = \frac{X_{(k)} - X_{min}}{X_{max} - X_{min}} \quad \ldots \quad (3.5)$$

where X's are the observed values in actual units. Rationalised values of observed data are shown in Fig 3.2.2. Co-efficients of correlation for different shifts of time instances are shown in Fig.3.2.3 (a to e). Input data are shown in Fig 3.2.4 (a to e).

Table 2.5.1

Gradually increasing complexity of polynomials of five variables.

General form of Polynomial :

$$y=a_0+a_1x_1+a_2x_2+a_3x_1x_2+a_4x_3+a_5x_1x_3+a_6x_2x_3+a_7x_4+a_8x_1x_4+a_9x_2x_4+a_{10}x_3x_4+a_{11}x_5+a_{12}x_1x_5+a_{13}x_2x_5+a_{14}x_3x_5+a_{15}x_4x_5$$

x_1	x_2	x_1x_2	x_3	x_1x_3	...	x_3x_4	...	x_4x_5
$y_1=a_0+a_1x_1$	$y_2=a_0+a_1x_2$	$y_4=a_0+a_1x_1x_2$	$y_8=a_0+a_1x_3$	$y_{16}=a_0+a_2x_1x_3$	...	$y_{512}=a_0+$	...	.....
	$y_3=a_1+a_2x_1$	$y_5=a_0+a_1x_1+$	$y_9=a_0+a_1x_1$	$y_{17}=a_0+$		$a_1x_3x_4$		
	$+a_2x_2$	$+a_2x_1x_2$	$+a_2x_3$			$y_{511}=a_0+a_1x_1$		
		$y_6=a_0+a_1x_2$				$+a_2x_3x_4$		
		$+a_2x_1x_2$	$y_{15}=a_0+a_1x_1$					
		$y_7=a_0+a_1x_1+$	$+a_2x_2+a_3x_1x_2$			$y_{1023}=a_0+a_1x_1$		$y_{32867}=a_0+$
		$a_2x_x+a_3x_1x_2$	$+a_4x_3$			$+a_2x_2+a_3x_1x_2$		$+a_1x_1+a_2x_2+$
						$+a_4x_3+a_5x_1x_3$		$+a_3x_1x_2+...$
						$+...+a_{10}x_3x_4$		$+a_{15}x_4x_5$

The General rule of total number of combination is 2^n-1 where n is the total number of terms containing the variables in the general form of the polynomial.

A selected number of polynomials were tested and the polynomial of optimum complexity was obtained as

$$y = 0.11321307 + 0.27777937 + 0.51539645\, x_2$$
$$+1.19813443\, x_3 + 0.64339720\, x_4 + 0.34528273\, x_5$$
$$+3.63166863\, x_1x_4 - 0.45064139\, x_1x_3$$
$$-3.05256229\, x_1x_4 - 0.62043912\, x_1x_5 - 2.50910875 x_2x_3 + 1.39236393 x_2x_4$$
$$-6.06417210\, x_2x_5 - 0.17344234\, x_3\, x_4 \qquad \ldots (3.6)$$
$$+4.42314422\, x_3x_5 + 3.10848409 x_4x_5$$

So in finite difference equation the hourly flow at Dowlaiswaram during a storm period is given as

$$\begin{aligned} y_{(k)} = {} & -0.11321207 + 0.27777937\, x_1(k-6) \\ & -0.51539645\, x_2(k-7) + 1.98113443\, x_3(k-9) \\ & +0.64339780\, x_4(k-8) + 0.34528273\, x_5(k-10) \\ & +3.63166863\, x_1(k-6)\, x_2(k-7) - 0.45064139\, x \\ & x_1(k-6)\, x_3(k-9) - 3.05256229\, x_1(k-6)\, x_4(k-8) \\ & -0.62043912\, x_1(k-6)\, x_5(k-10) - 2.50910875\, x \\ & x_2(k-7)\, x_3(k-9) + 1.39236393\, x_2(k-7)\, x_4(k-8) \\ & -6.06417210\, x_2(k-7)\, x_5(k-10) - 0.17344234\, x \qquad \ldots (3.7) \\ & x_3(k-9)\, x_4(k-8) + 4.42314422\, x_3(k-9)\, x_5(k-10) \\ & + 3.10848409\, x_4(k-8)\, x_5(k-10) \end{aligned}$$

The corresponding value of the integral square error is 0.009426. The figure 3.2.5 shows the observed and the modelled values.

The forecasting model has a minimum time lag of sixth instance. When tested against field data the model is found to simulate adequately the major variations observed in the measured data [101].

FIG- 1

TEESTA
AND ITS TRIBUTARIES.
SCALE - 1 INCH = 16 MILES

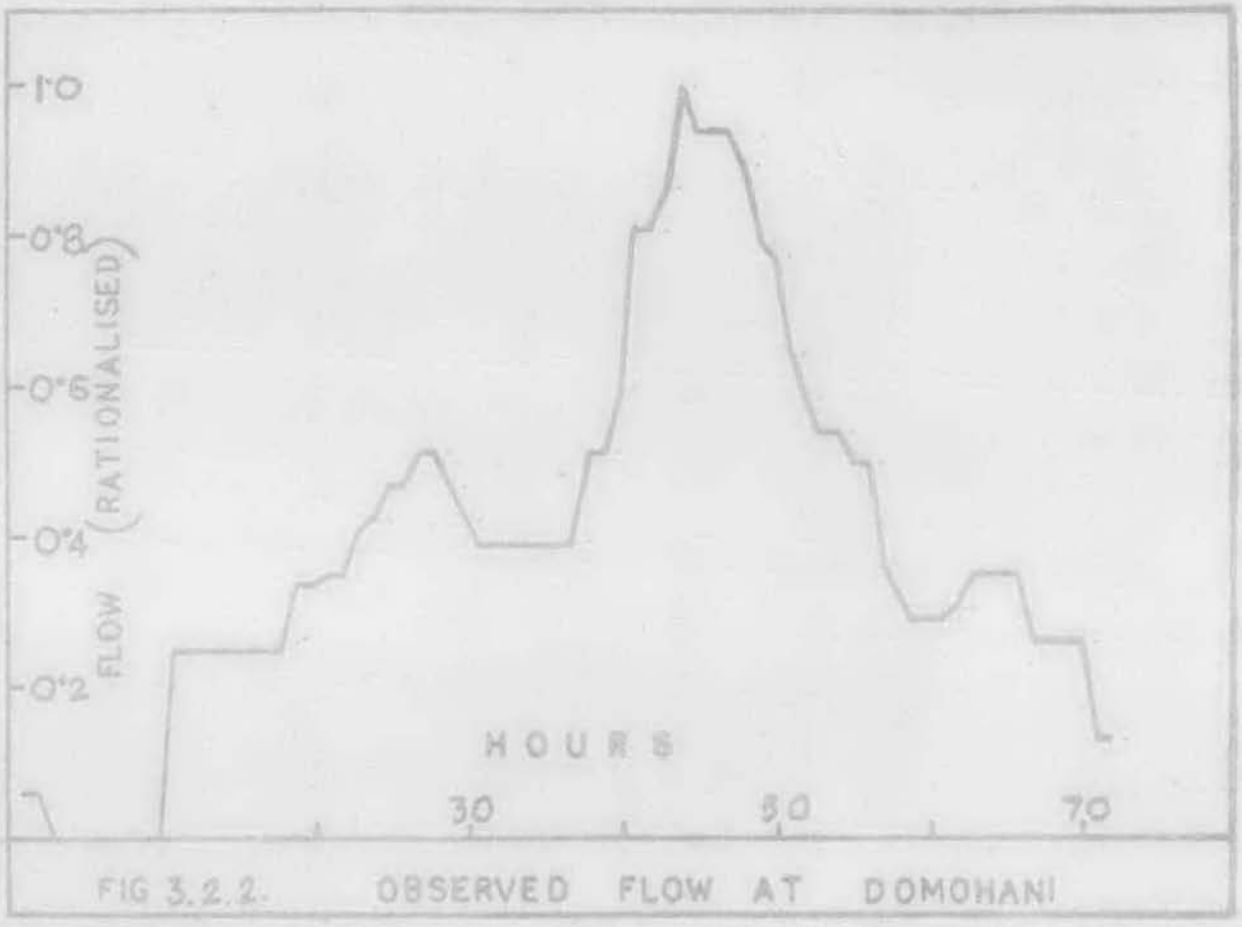


FIG 3.2.2. OBSERVED FLOW AT DOMOHANI

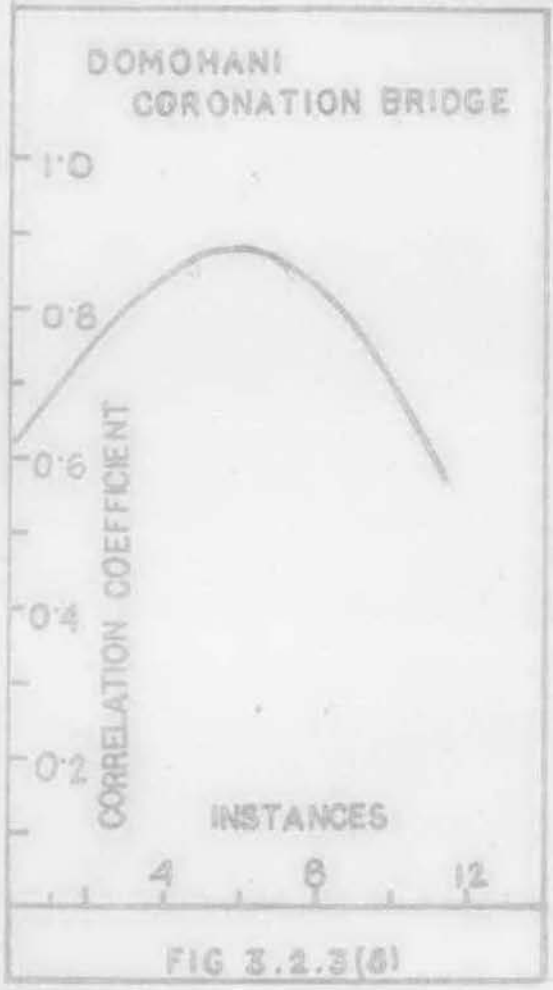


FIG 3.2.3(a)

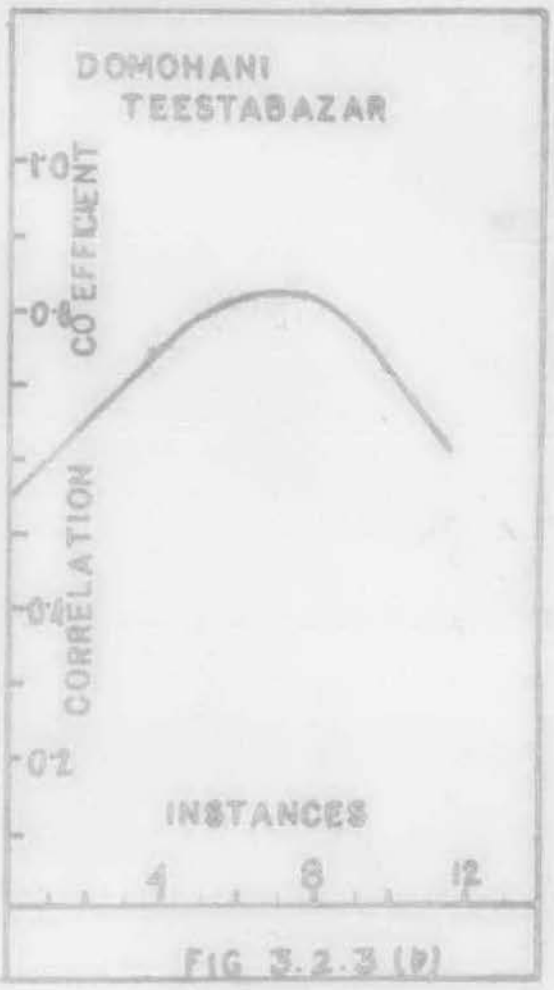


FIG 3.2.3 (b)

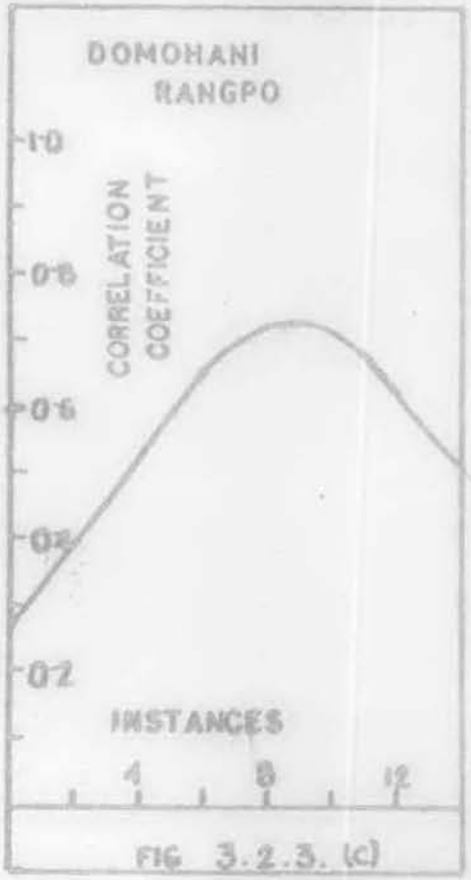


FIG 3.2.3. (c)

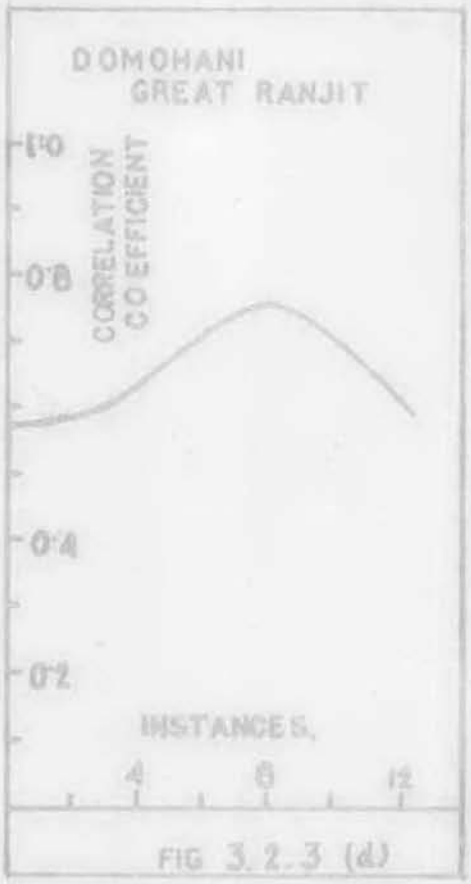


FIG 3.2.3 (d)

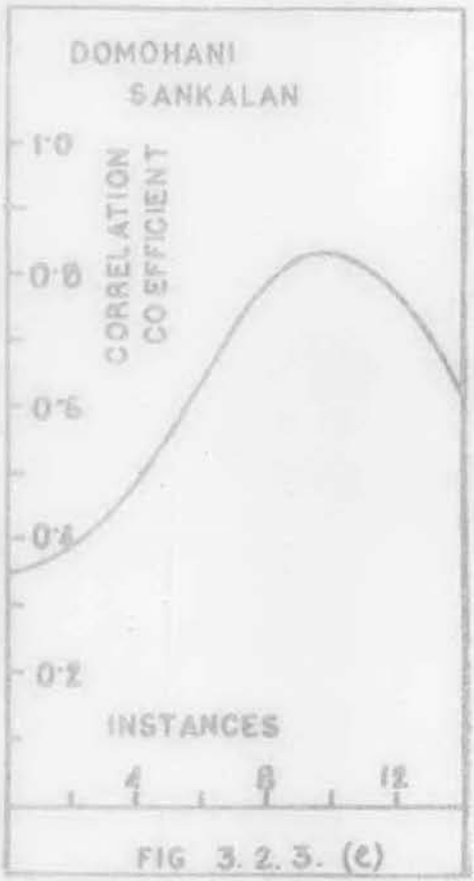


FIG 3.2.3. (e)

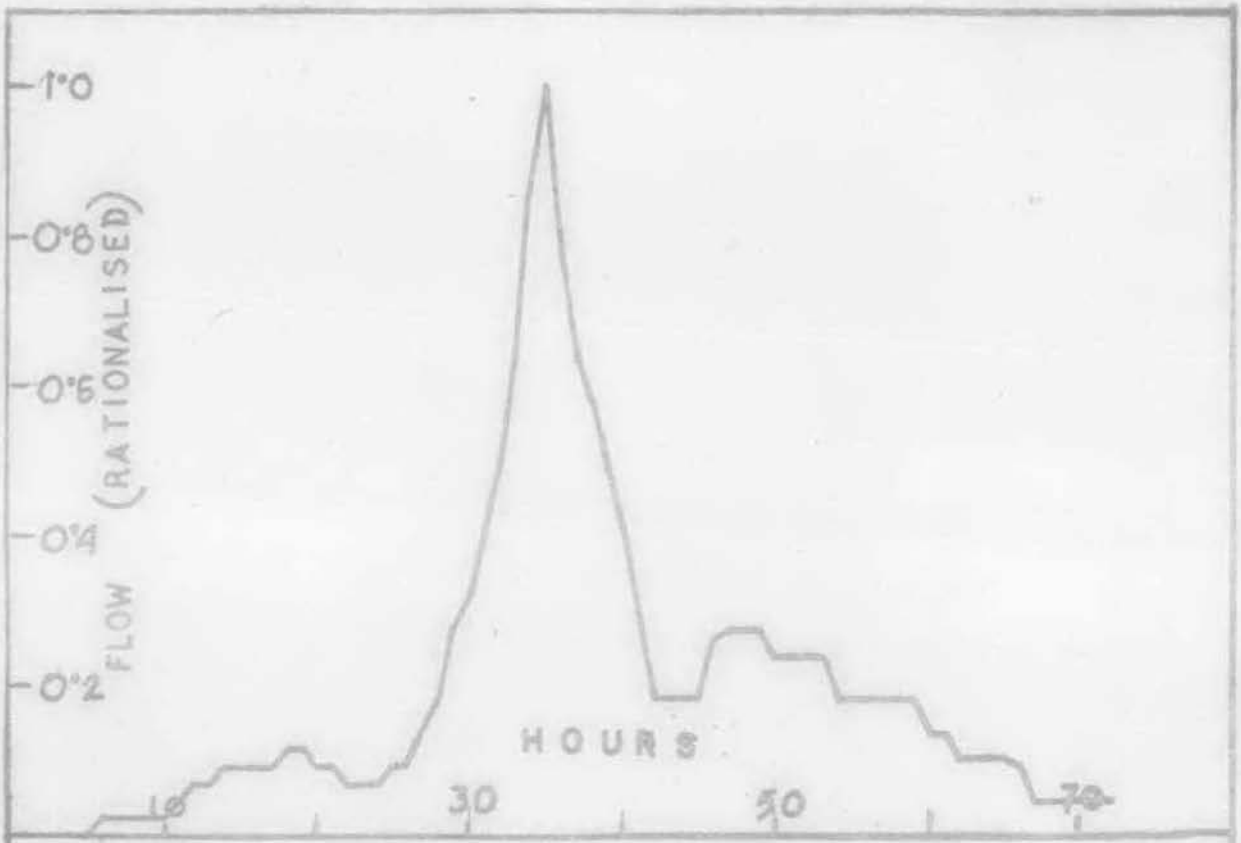


FIG 3.2.4.8 OBSERVED FLOW AT SANKALAN

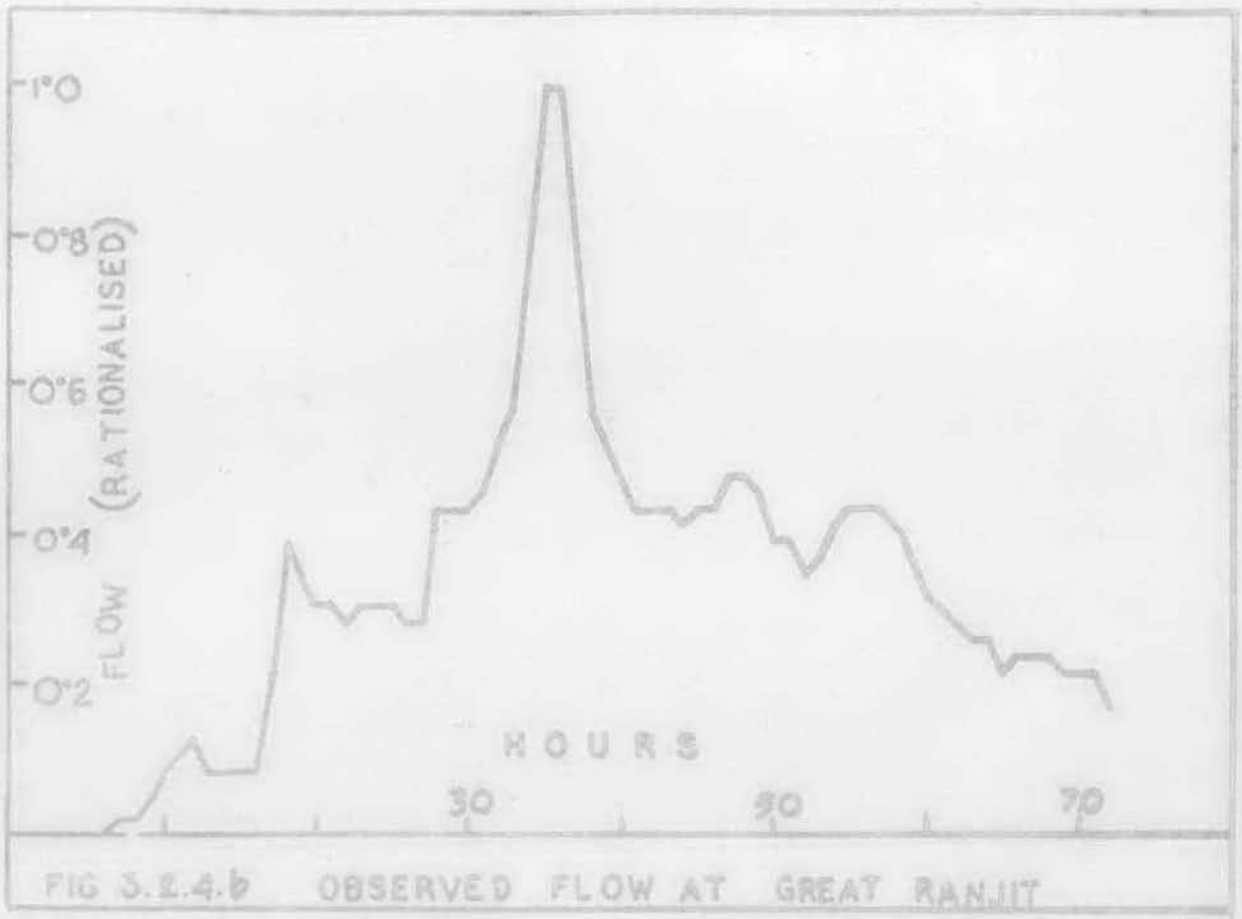


FIG 3.2.4.b OBSERVED FLOW AT GREAT RANJIT

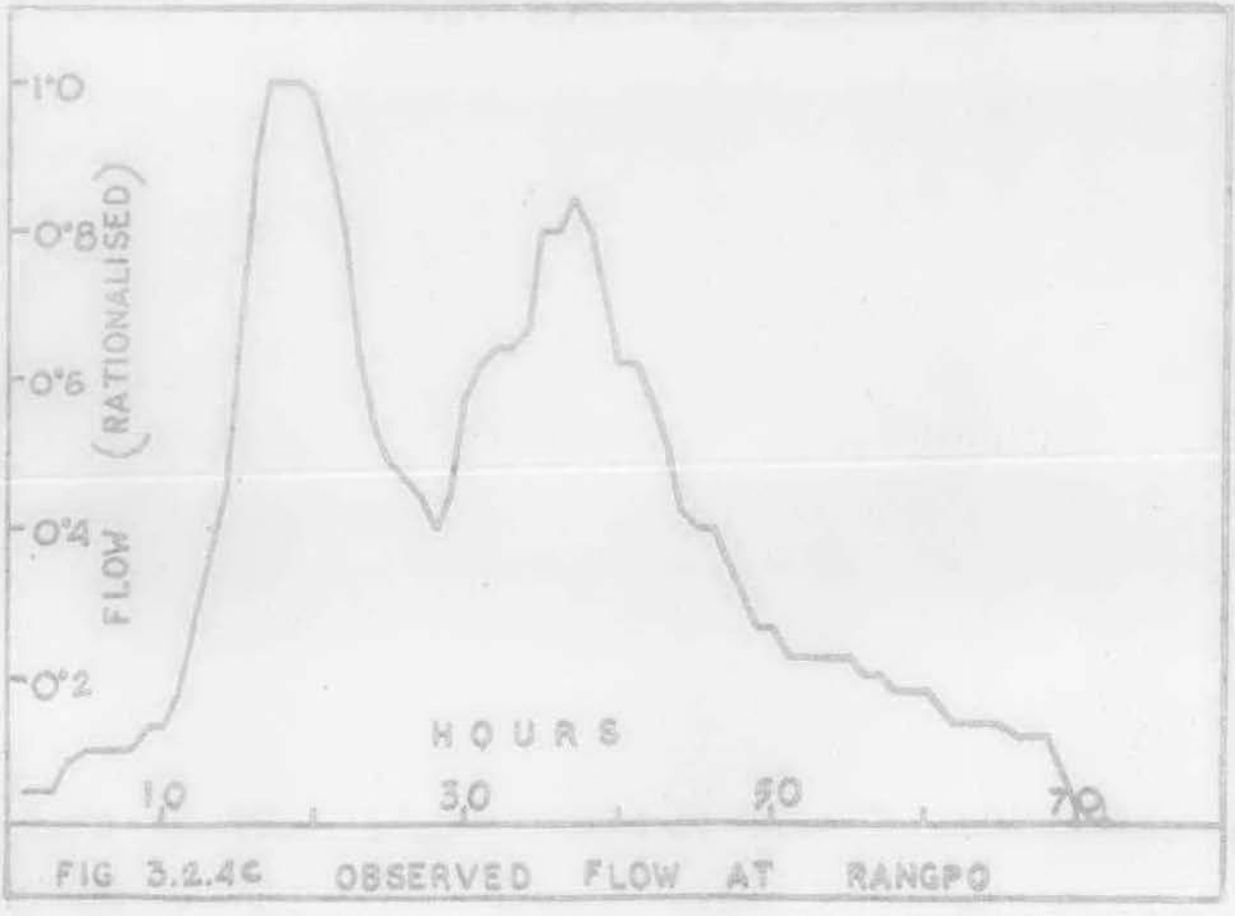


FIG 3.2.4c OBSERVED FLOW AT RANGPO

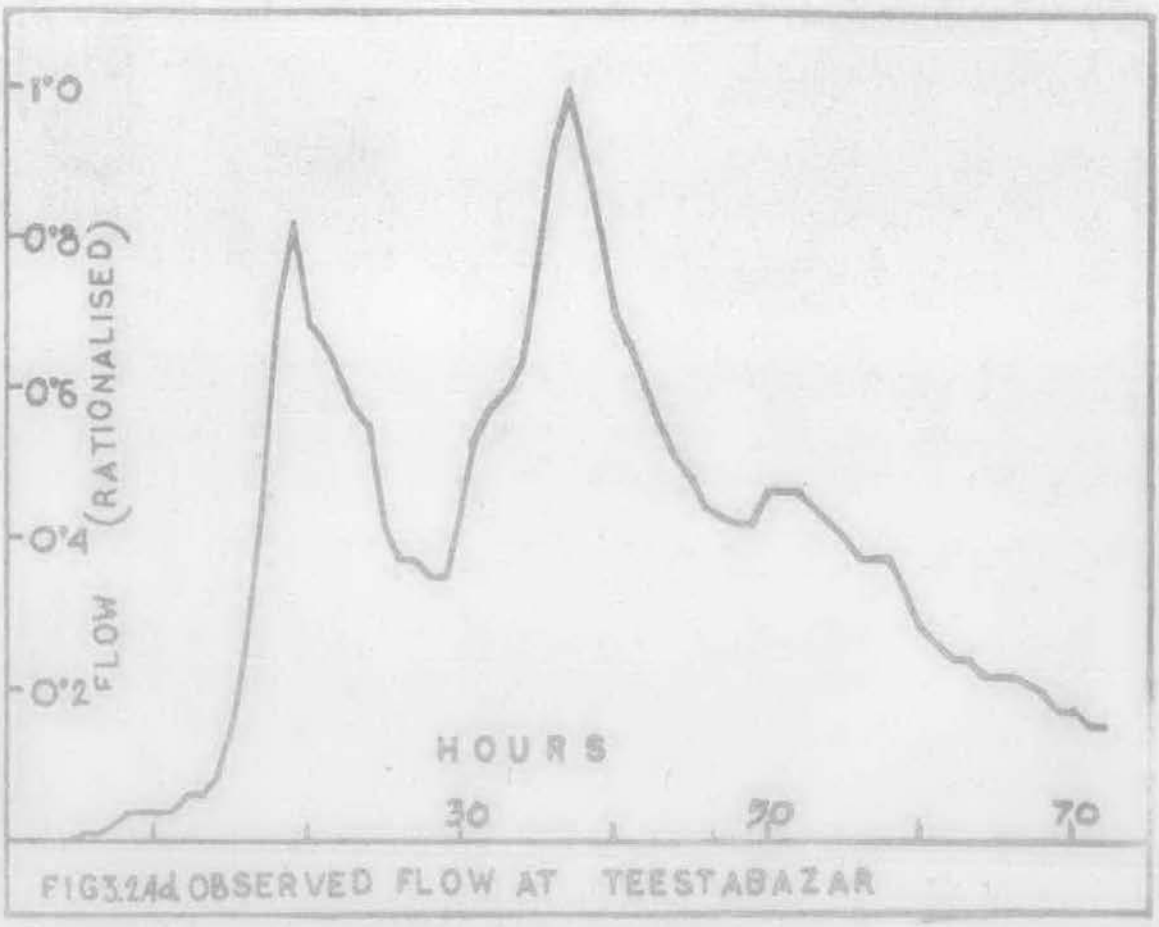


FIG 3.2.4d OBSERVED FLOW AT TEESTABAZAR

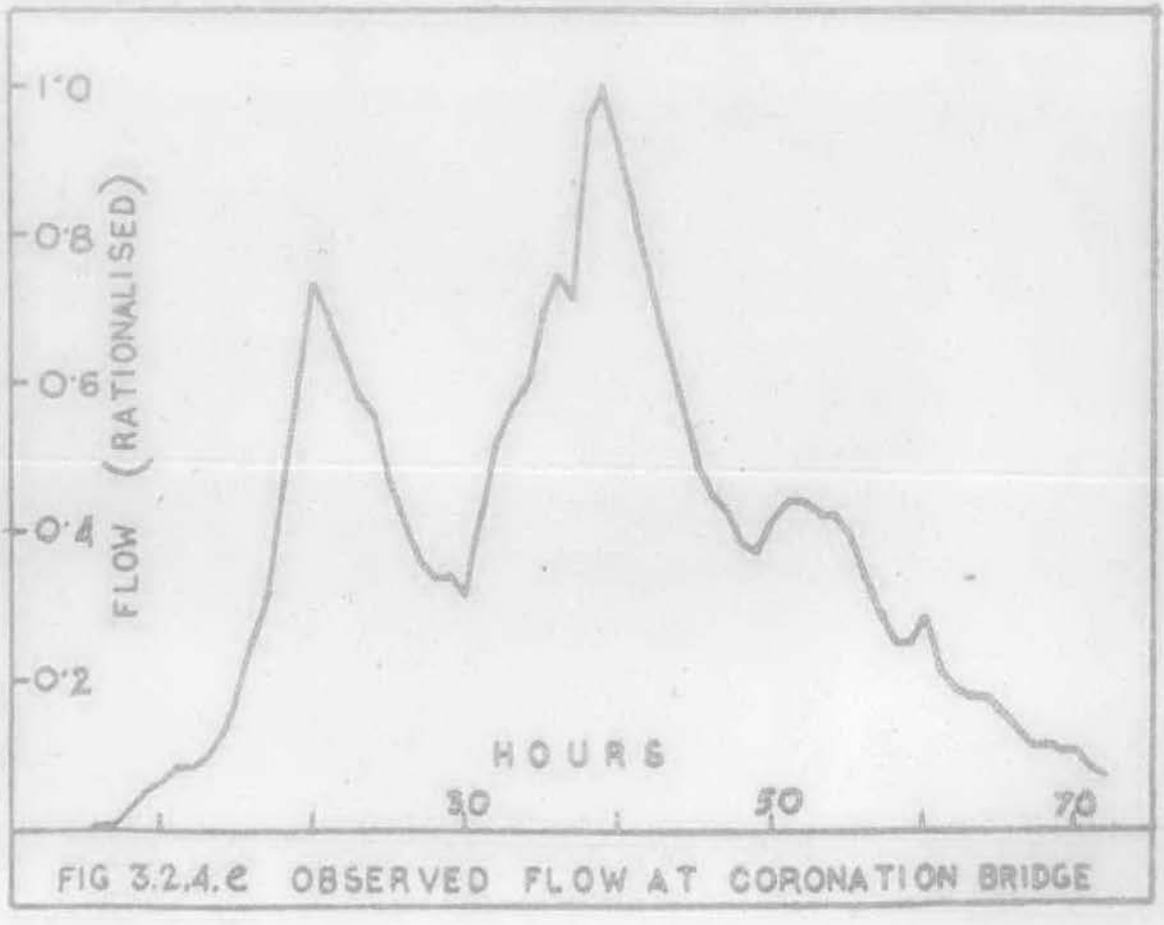


FIG 3.2.4.e OBSERVED FLOW AT CORONATION BRIDGE

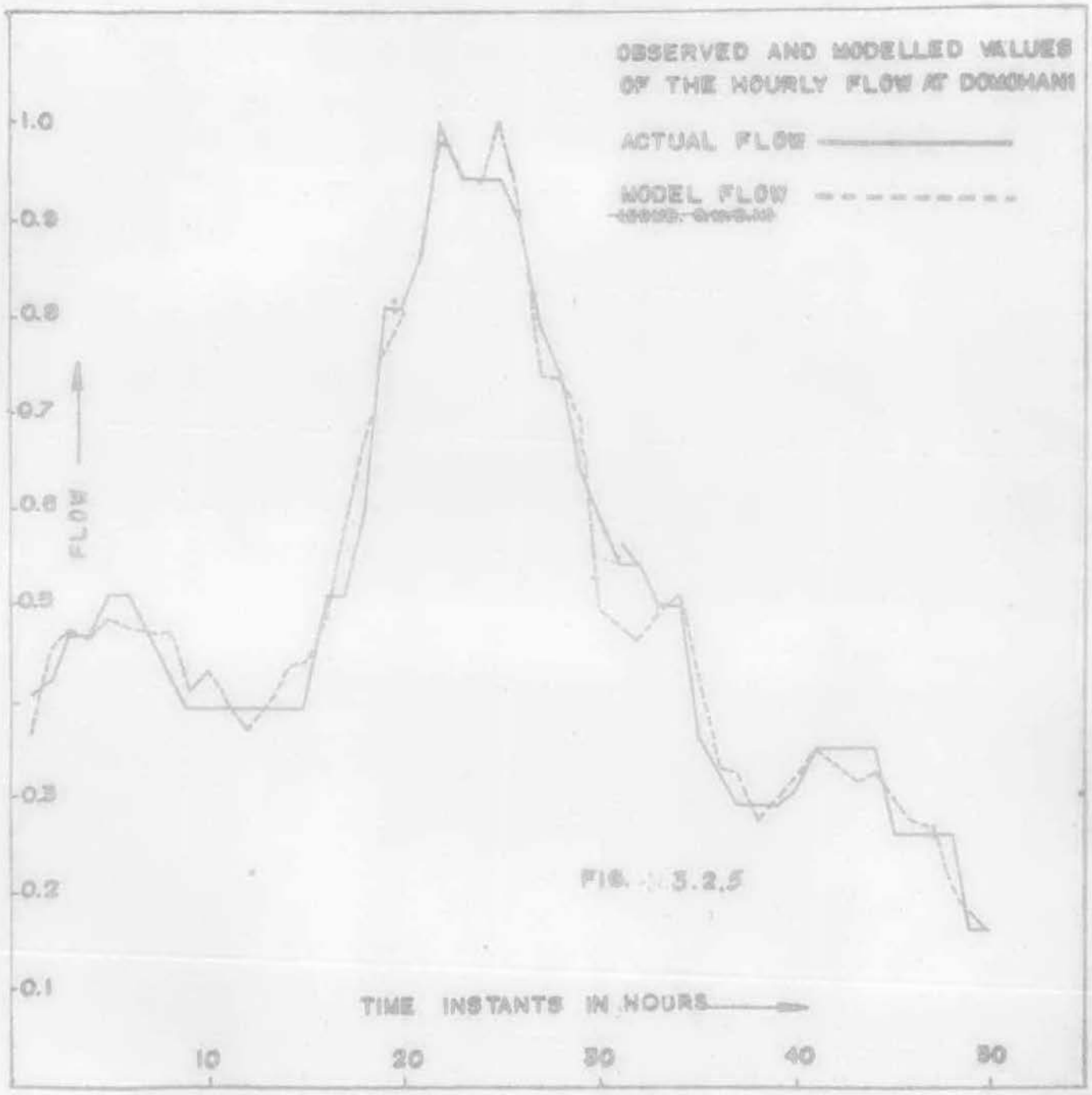

OBSERVED AND MODELLED VALUES
OF THE HOURLY FLOW AT DOMOHANI
ACTUAL FLOW
MODEL FLOW
1.0
0.9
0.8
0.7
0.6
0.5
0.3
0.2
0.1
FLOW
TIME INSTANTS IN HOURS
10
20
30
40
50

FIG. 5.2.5

3.3 Part Two : Use of multilayer GMDH in formulating the six-hourly flow forecasting model of the river Teesta at Domohani Road Bridge Point near Jalpaiguri town.

The river Teesta rising from Himalayan ranges in north Sikkim and passing through the deep gorges for nearly 138 Kms. arrives at the plain of West Bengal near Sevoke in Darjeeling district. The Teesta is a very fast, flowing river, its average velocity being 6.2 metres per second.

Data were taken for a storm period commencing from 16.00 hours on the 22nd. July, 1979 to 15.00 hours on 25th. July, 1979 for the gauging stations at Domohani Road Bridge, Coronation Bridge, Teestabazar, Rongpo (Teesta), Great Rangeet at Singlabazar and at Sankalan. The stations are shown in Fig 3.2.1.

Raw data were transformed as

$$x_{(k)} = \frac{X_{(k)} - X_{min}}{X_{max} - X_{min}} \qquad \ldots \quad (3.8)$$

where $X_{(k)}$ is the observed and X_{min} and X_{max} are the minimum and maximum values of the data.

Formulation of the process equation

Hourly flow process at Domohani Road Bridge can be represented by a generaly form of process equation as

$$Y_{(k)} = f(Y(k-1), Y(k-2), \ldots, X_1(k-1), X_1(k-2) \ldots X_2(k-1), X_2(k-2) \ldots) \quad (3.9)$$

the subscript k, k-1, k-2, ..., refer respectively to the current hour, an hour preceding the current hour and two hours preceding the current hour and so on. $y(.)$, $x_1(.)$, $x_2(.)$, $x_3(.)$, $x_4(.)$. and $x_5(.)$ are the transformed value of the hourly gauge reading at Domohani Road Bridge, Coronation Bridge, Teestabazar, Rongpo (Teesta) Great Rangeet at Singlabazar and at Sankalan respectively.

The arguments having a strong correlation with $y_{(k)}$ are then selected for inclusion in the process equation on the basis of the correlation functions for the time shift λ defined as

$$\phi_{yx}(\lambda) = \frac{\sum_{i=1}^{N-\lambda} \left(y(i) - \frac{1}{N}\sum_{k=1}^{N} Y(k)\right)\left(x(i+\lambda) - \frac{1}{N}\sum_{k=1}^{N} x(k)\right)}{\left[\sum_{i=1}^{N-\lambda} \left(Y(i) - \frac{1}{N}\sum_{k=1}^{N} Y(k)\right)^2 \sum_{j=1+\lambda}^{N} \left(x(j) - \frac{1}{N}\sum_{k=1}^{N} x(k)\right)^2\right]^{\frac{1}{2}}} \quad \ldots \quad (3.9)$$

where N is the number of data points.

After such selection of arguments as having strong correlation with hourly flow at Deschnel Road Bridge, the process equation becomes :

$$y_{(k)} = f(x_1(k-6), x_2(k-7), x_3(k-9), x_4(k-8), x_5(k-10)) \quad \ldots \quad (3.10)$$

or denoting these respective arguments as

$$y(k) = y,\ x_1(k-6) = x_1;\ x_2(k-7) = x_2,\ x_3(k-9) = x_3$$

$$x_4(k-8) = x_4 \text{ and } x_5(k-10) = x_5$$

the process equation becomes,

$$y = f(x'_1, x'_2, x'_3, x'_4, x'_5) \quad \ldots \quad (3.11)$$

First layer of selection

There are now $\binom{5}{2} = 10$ possible combinations of selecting 2 arguments at a time out of 5. For every such combination, the partial regression equation is written in the form,

$$y_a = \alpha_{0a} + \alpha_{1a} x'_b + \alpha_{2a} x'_c + \alpha_{3a} x'_b x'_c + \alpha_{4a} x'^2_b + \alpha_{5a} x'^2_c \quad \ldots (3.12)$$

where $a = 1, 2, \ldots, 10$ while b and c are indices for all 10 combinations. And these, therefore lead to 10 systems of normal Gaussian equations with matrices of the order 6 x 6. The co-efficients α's are then estimated by solving the normal equation systems constructed from the data set. For estimating the co-efficients it is assumed that the equation error is very small being distributed with zero mean, constant variance and also uncorrelated with the inputs. The second assumption is that for the construction of the model the inputs and outputs are known exactly without any measurement error.

The accuracy of every variable y_a is calculated by using the whole data set. From all the variables we choose 5 more accurate ones which give low values of integral square error criterion as defined as :

$$ISE = \frac{\sum_{i=1}^{N} (y_i(\text{observed}) - y_i(\text{modelled}))^2}{\sum_{i=1}^{N} (y_i(\text{observed}))^2} \quad \ldots \quad (3.13)$$

Other layers selection

5 intermediate variables of y-layer chosen from the first layer give 10 combinations of two arguments of y-layer. Again this in the second layer these become

$$z_a = \beta_{oa} + \beta_{1a} y_b + \beta_{2a} y_c + \beta_{3a} y_b y_c + \beta_{4a} y_b^2 + \beta_{5a} y_c^2 \quad \ldots \quad (3.14)$$

where $a = 1, 2 \ldots,$ 10 while b and c indices of all 10 combinations. Calculation of the co-efficients β's and estimation of the accuracy of the variables z_a are repeated as in case of y_a.

The 5 z_a variables are then chosen for the next layer u_a.

$$u_a = \gamma_{oa} + \gamma_{1a} z_a + \gamma_{2a} z_b + \gamma_{3a} z_a z_b + \gamma_{4a} z_a^2 + \gamma_{5a} z_b^2 \quad \ldots (3.15)$$

In this way layer after layer is tested for accuracy by using the whole data set and on the basis of minimum integral square criterion as explained earlier. For all layers variables on the left hand side of the equations are kept equal to the value of the output variable.

It has been observed for the present process that as the layer increases the integral square errors converge asymptotically to a very small value. To make the model amenable to practical application the integral square error of 7.308044×10^{-3} at z_2 has been considered for point of termination for formulation of the process equation. The changes of minimum integral square for different layers are shown in Fig 3.3.1.

Result

The flow model of the river Teesta at Domohani Road Bridge for a storm period with a minimum of sixth lagged measurement at Coronation Bridge Point has been identified by the polynomials as shown below :

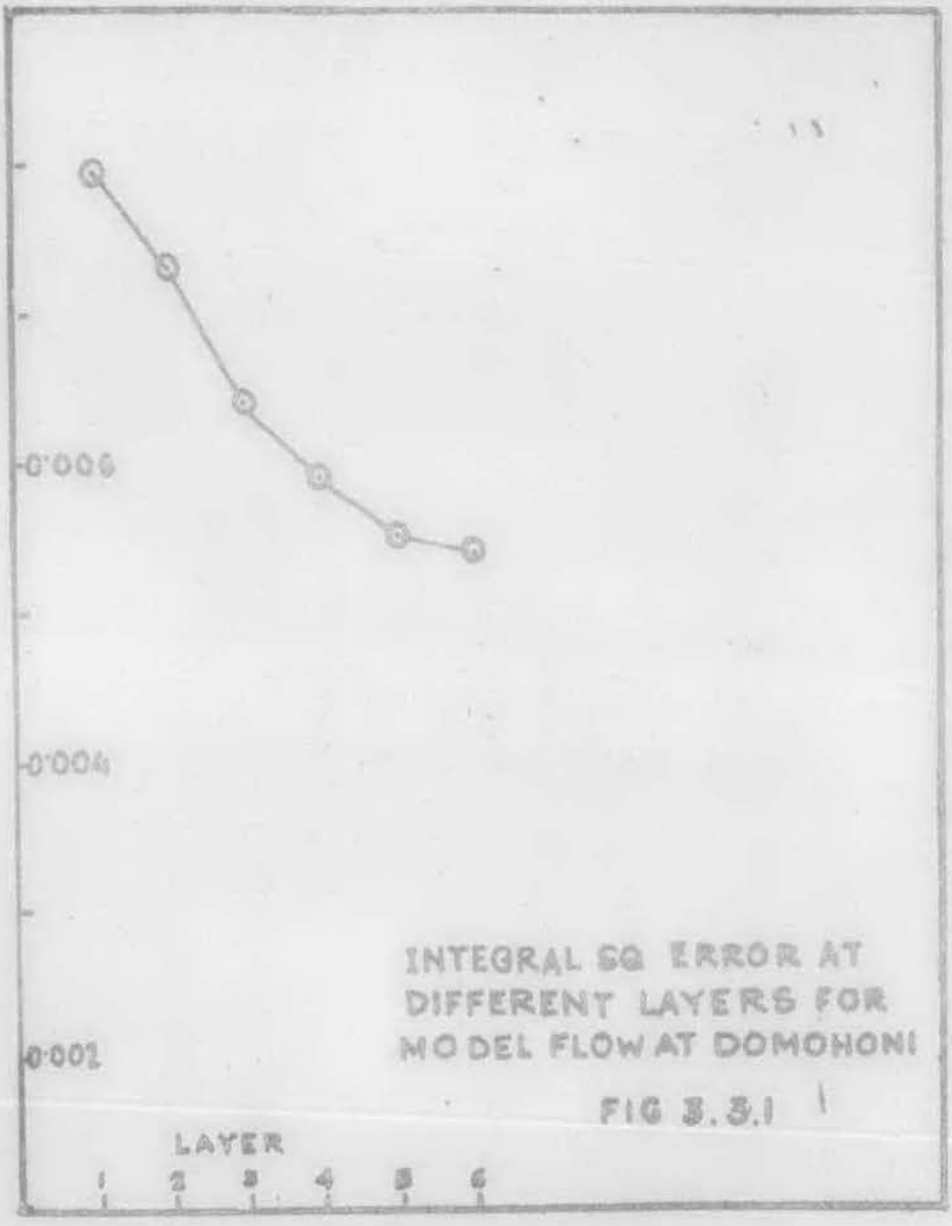
$$y = z_2$$

$$z_2 = -0.0146179 + 1.19980778\, y_2 - 0.10964442\, y_4 + 5.54703006\, y_2 y_4 - 2.90983183\, y_2^2 - 2.71481457\, y_4^2$$

$$y_2 = 0.0444305 + 0.22436070\, x_2 + 0.89651073\, x_4 + 3.90684696\, x_2 x_4 + 1.48099172\, x_2^2 - 2.54885663\, x_4^2$$

$$y_4 = 0.01748533 + 0.92202678\, x_4 + 0.22256408\, x_5 + 1.85165843\, x_4 x_5 - 1.03780842\, x_4^2 - 0.75291287\, x_5^2$$

Fig. 3.3.2 shows the observed and the modelled values of hourly flow of the river Teesta at Domohani Road Bridge on the basis of flow measurements with a minimum sixth lagged instant at Coronation Bridge Point. [102, 103]



INTEGRAL SQ ERROR AT DIFFERENT LAYERS FOR MODEL FLOW AT DOMOHONI

FIG 3.3.1

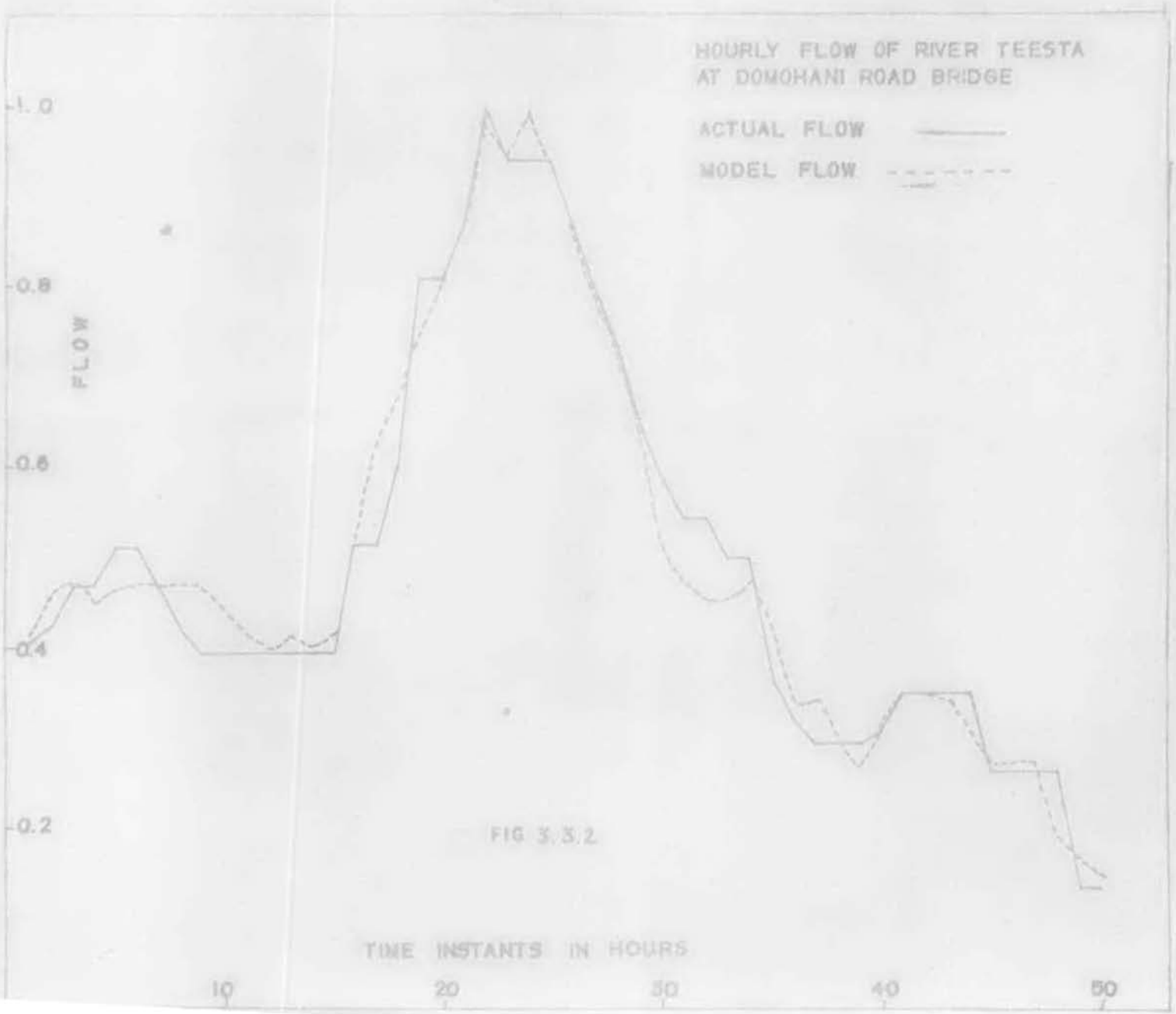

HOURLY FLOW OF RIVER TEESTA
AT DOMOHANI ROAD BRIDGE
ACTUAL FLOW
MODEL FLOW
1.0
0.8
0.6
0.4
0.2
FLOW
TIME INSTANTS IN HOURS
10
20
30
40
50

FIG 3.3.2

3.4 Part three : Real Time Estimation of river flow during a short span storm period

Development of the recursive algorithms

Recursive technique has been defined as " a technique in which estimate is updated on receipt of fresh information ".

Description of system's dynamic equation

Strong correlation of the down-stream flows with those at upstream flows particularly at the confluence of the tributaries suggests the process may be represented in the form of a non-stationary time series with a probability that the current value of the output $Y_{(tk)}$ is a function of the previous output observations, the autoregressive terms $Y(t_{k-1})$, $Y(t_{k-2})$, ..., the past values of the highly correlated deterministic inputs.

$U_1(t_{k-r1})$, $U_2(t_{k-r_2})$, ..., $U_1(t_{k-r1-1})$, $U_2(t_{k-r_2-1})$, ...,

together with a current unknown realisation of the noise process $\gamma(tk)$. The /xxx process can be represented as

$$Y(t_k) = \sum_{i=1}^{n} \beta_i Y(t_{k-1}) + \sum_{j=1}^{n} \sum_{i=0} \delta_{ji} U_j (t_{k-rj-i}) + \gamma(t_k) \quad \ldots \quad (3.16)$$

Determination of n is known as the model order determination. It is suggested that the output correlation may be intuitively considered for model order determination which is also the model structure identification.

With backward shift operator $q^{-1}(.)$ defined as

$q^{-1} Y(t_k) = Y(t_{k-1})$, the eqn. (3.16) transforms to,

$$Y(t_k) = \left[\sum_{i=1}^{n} \beta_i q^{-1} \right] \Big/ \; Y(t_k) + \sum_{j=1}^{m} \left[\sum_{i=0}^{n} \delta_{ji} q^{-1} \right] \Big/ U_j(t_{k-r_{j-1}})$$

$$+ \gamma(t_k) \qquad \ldots \quad (3.17)$$

$\gamma(t_k)$ may be expressed in a moving average sequence as

$$\gamma(t_k) = \sum_{p=1}^{p} \gamma_p \gamma(t_{k-p}) + v(t_k) \qquad \ldots \quad (3.18)$$

where $v(t_k)$ is a white noise innovation process with

$$E\left\{ v(t_k)\, v(t_j) \right\} = 0 \quad \text{for} \quad j \neq k \qquad \ldots \quad (3.19)$$

and

$$E\left\{ v(t_k)\, v(t_j) \right\} = \sigma^2 \quad \text{for} \quad j = k$$

This model is quite flexible since it requires that the equations be linear in parameters.

Eqn.(3.17) may be represented in the form as

$$Y(t_k) = z^T(t_k) \alpha + \gamma(t_k) \qquad \ldots \quad (3.20)$$

where α is the parameter vector with the property of slowly varying with time and amenable to recursive adaptiveness, and

$$z^T(t_k) = \Big[Y(t_{k-1}), \ldots, Y(t_{k-n}), U_1(t_{k-r1}), \ldots,$$

$$U_1(t_{k-r1-n}), \ldots, U_m(t_{k-rm}), \ldots, U_m(t_{k-rm-n}) \Big] \qquad \ldots \quad (3.21)$$

where r_j, $j = 1, 2, \ldots, n$ stands for the lag time instant of upstream which have the strongest correlation with the down-stream flow, $Y(t_k)$.

Least square estimation of parameters

The parameter values of equation (3.20) are estimated by minimising a loss function defined as the sum of the squared errors as

$$J \triangleq \sum_{k=1}^{N} \| Y(t_k) - z^T(t_k) \hat{\alpha} \|^2 \qquad \ldots \quad (3.22)$$

The estimates $\hat{\alpha}$ of α that minimize J are called least squares estimates.

So, for minimization,

$$\frac{\partial J}{\partial \hat{\alpha}} = 0$$

$$\text{or } 2\left[\sum_{k=1}^{N} z(t_k)z^T(t_k)\right]\hat{\alpha} - 2\left[\sum_{k=1}^{N} z(t_k)\, Y(t_k)\right] = 0$$

And hence we have the well-known equations for least squares parameter estimation as

$$\hat{\alpha} = \left[\sum_{k=1}^{N} z(t_k)\, z^T(t_k)\right]^{-1}\left[\sum_{k=1}^{N} z(t_k)\, Y(t_k)\right] \qquad \ldots (3.23)$$

Recursive instrument variable algorithms for parameter estimation.

The most likely source of biased estimate is the presence of significant noise sequence correlation between noise sequence and the past values of the output.

Referring to equation (3.17) with suitable estimates of the parameters an auxiliary time series model can be computed as

$$Y(t_k) = \left[\sum_{i=1}^{n} \delta_{i1} q^{-1}\right] Y(t_k) + \sum_{j=1}^{m}\left[\sum_{i=0}^{n} \delta_{ji} q^{-1}\right] U_j(t_{k-rj-i}) \quad \ldots (3.24)$$

Equations (3.17) and (3.18) suggest (i) variation in $\hat{Y}(t_k)$ should be strongly correlated with variations in the noise corrupted output observations $Y(t_k)$, but (ii) these variations in $\hat{Y}(t_k)$ should be uncorrelated with (t_k) provided (t_k) is not correlated with the measured input sequences.

$U_j(t_{k-r_j})$ i.e., $E\left\{U_j(t_{k-r_j}) \quad \eta(t_i)\right\} = 0$ for all j, K and i.

Sequence $\hat{Y}(t_k)$ is called the sequence of instrument variables. Consequently, the vector $z^T(t_k)$ is modified as $\hat{z}^T(t_k)$, defined by

$$\hat{z}^T(t_k) = \left[\hat{Y}(t_{k-1}), \ldots, \hat{Y}(t_{k-n}), U_1(t_{k-r1}), \ldots, U_1(t_{k-r1-n}), \right.$$
$$\left. \ldots, U_m(t_{k-r_m}), \ldots, U_m(t_{k-rm-n}) \right]$$

The conditions of unbiased estimates are modified as

$$E\left\{z(t_k) \quad \eta(t_k)\right\} = 0 \text{ for all } k$$

with
$$E\left\{v(t_k) \quad v(t_j)\right\} = 0 \text{ for } k = j$$
$$= \sigma^2 \text{ for } k = j$$

The recursive instrument variable algorithms are given as

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) + \hat{P}^e(t_{k-1}) \hat{z}(t_k) \left[1 + z^T(t_k) \hat{P}^e(t_{k-1}) \hat{z}(t_k)\right]^{-1}$$
$$\left\{Y(t_k) - z^T(t_k) \hat{\alpha}(t_{k-1})\right\} \quad \ldots (3.25)$$

$$\hat{P}^e(t_k) = \hat{P}^e(t_{k-1}) - \hat{P}^e(t_{k-1}) \hat{z}(t_k) \left[1 + z^T(t_k) \hat{P}^e(t_{k-1}) \hat{z}(t_k)\right]^{-1}$$
$$z^T(t_k) \hat{P}^e(t_{k-1}) \quad \ldots (3.26)$$

with

$$\hat{Y}(t_k) = \left[\sum_{i=1}^{n} \hat{\beta}_i q^{-1}\right] \hat{Y}(t_k) + \sum_{j=1}^{m} \left[\sum_{i=0}^{n} \delta_{ji} q^{-1}\right] U_j(t_{k-r_j-i})$$

and $\hat{P}^e(t_k) \triangleq \left[\sum_{j=1}^{k} z(t_j) \quad z^T(t_j)\right]^{-1}$

$\hat{\alpha}(.)$ and $\hat{P}^e(.)$ have been initialised by least square method using the whole data set or a block of data.

Results of investigation

Data for a short span storm period from 16.00 hours on 23.7.79 to 15.00 hours on 25.7.79 were observed at the gauging stations at Sankalan, Rongpo (Teesta), Great Rangeet (Singlabazar), Teestabazar, Coronation Bridge Point and at Domohani Road Bridge of the river Teesta in North Bengal.

Input variables are selected on the basis of the strongest co-efficient of correlation with flows at Domohani. The input output variables are rationalised in accordance with $x(k) = \frac{X(k) - X_{min}}{X_{max} - X_{min}}$ (ii) where X(.) is the observed value and X_{max} and X_{min} is the maximum and the minimum value of the data sequence. The co-efficient of correlation has been defined as

$$\phi_{yx}(\lambda) = \frac{\sum_{i=1}^{N-\lambda} (y(i) - \frac{1}{N}\sum_{i=1}^{N} Y(i))(x(i+\lambda) - \frac{1}{N}\sum_{i=1}^{N} x(i))}{\left[\sum_{i=1}^{N-\lambda} (y(i) - \frac{1}{N}\sum_{i=1}^{N} y(i))^2 \sum_{j=1+\lambda}^{N} (x(j) - \sum_{i=1}^{N} x(i))^2\right]^{\frac{1}{2}}} \quad \ldots (3.27)$$

Using recursive instrument variable algorithms the minimum mean error and the minimum integral square error were found to be 4.9725513933 E - 05 and 5.8441299733 E-03. Fig 3.4.1a and 3.4.1b show the observed and the modelled values of the rationalised flow of the river Teesta at the Domohani Road Bridge for a short span storm period from 21.00 hours on 23.7.79 to 15.00 hours on 25.7.79. $\nu(k)$ is found to be almost uncorrelated with $\nu(j)$ for $k \neq j$ and 1.0 for $k = j$.

$$E\{\nu(k)\,\nu(j)\} = 0.99981 \quad \text{for} \quad j = k.$$

$\eta(k)$ has been found to have a moving average sequence of prehistory interval, P as 4 for a minimum mean error of the

recursive instrument variable algorithms have been found to simulate extremely well the observed hourly flow during a short span storm period for real-time operation with recursive instrument variable algorithm mean error is almost absent and $\gamma(k) = \nu(k)$ [104_7].

3.5 Conclusion

In part one a simple dynamic model of hourly flow of the river Teesta at Domohani Road Bridge Point has been presented with a minimum of sixth lag instance in the measurement of up-stream flows using a heuristic learning identification technique known as combinatorial group method of data handling algorithms. The model has been verified by simulation against field data observed during a short span storm period. Beauty of this model lies in its simplicity. It is observed that this model is more accurate than the model obtained through the complex process of decision regularisation as in the case of multilayer group method of data handling algorithms. To make this model general in application further investigation is needed to be carried out with data for different storm periods covering a number of years so as to bring out the time varying characteristics of the parameters of the polynomial description of the process of river flow.

In part two the capability of multilayer GMDH to forecast river flows quite sufficient ahead of time of occurrence has been demonstrated. It is possible to predict flow of the river Teesta at Domohani Road Bridge near Jalpaiguri town on the basis of a sixth hourly minimum gauged readings at the up-stream region at the Coronation Bridge Point and other gauging stations.

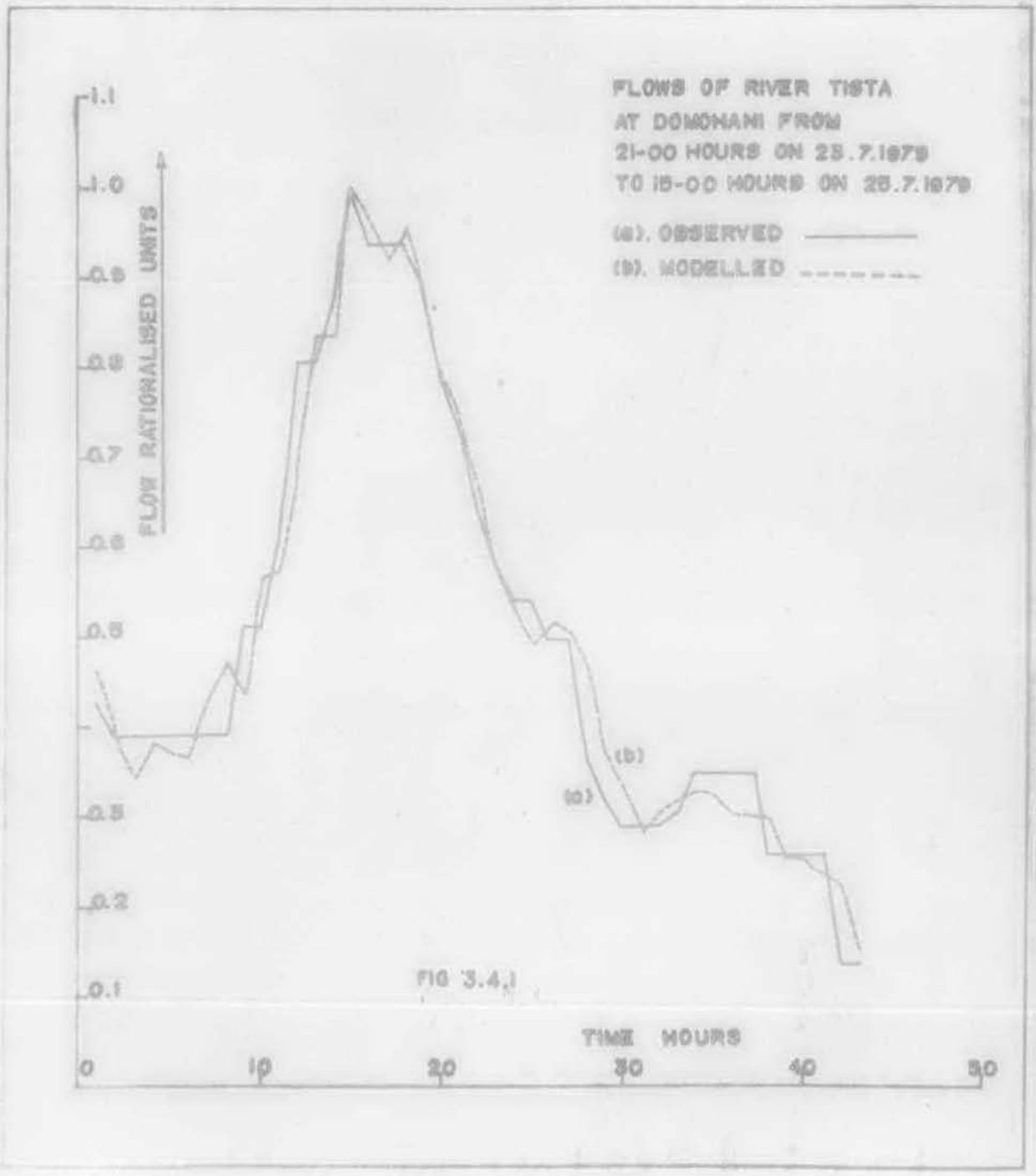

FLOWS OF RIVER TISTA
AT DOMOHANI FROM
21-00 HOURS ON 25.7.1979
TO 15-00 HOURS ON 28.7.1979
(a). OBSERVED
(b). MODELLED
FLOW RATIONALISED UNITS
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.2
0.1
(b)
(a)
TIME HOURS
0
10
20
30
40
50

FIG 3.4.1

In part three the effectiveness of the recursive estimation algorithms have been investigated for real-time prediction of the flows of the river Teesta at Domohani Road Bridge near Jalpaiguri town during a short span storm period. In the design of large scale water resources systems needed to meet the rapid expansion of irrigation and hydropower generation the trend is towards more advanced engineering operation in terms of monitoring, security and control of river flows during a short span storm period. This requirement so necessary to provide a reliable decision methodology have placed greater demands on the quantity and quality of real-time prediction of river flows. The very accurate nature of the prediction of river flows by recursive estimation algorithm has demonstrated that the demands are adequately met. Further investigational efforts are needed to obtain real-time prediction with other versions of recursive estimation techniques, e.g., exponential weighting of past data and Kalman filtering techniques. From the work presented in this part it is claimed that on-line flood warning and control measures can be implemented provided adequate real-time information processors are installed at the gauging stations with closed loop decision regularisation for operation and control of hydraulic structures of the water resources systems.

The present investigation has planted the seedling for on-line and real-time operation of water resources system and is addressed to the water resources engineers planning system installation and manufacturing personnel providing necessary hard ware and soft ware support.

CHAPTER IV

RECURSIVE ESTIMATION OF RIVER FLOWS DURING A SHORT SPAN STORM PERIOD

CHAPTER IV

RECURSIVE ESTIMATION OF RIVER FLOWS DURING A SHORT SPAN STORM PERIOD.

4.0 Abstract :

A recursive non-stationary parameter estimation technique has been developed and used to forecast on on-line basis the hourly flows of a non-tidal river during a short span storm period. The model is simple and requires no elaborate initialisation. The model has been subjected to test with field data observed at different gauging stations of the river Teesta in North Bengal during a short span storm period. Predicted data are found to simulate closely the variation of flows observed during the storm period.

4.1 Introduction :

The real time monitoring of river flow during a storm period and the on-line control of river flow through different hydraulic structures require a consistent and reliable set of data representing the predicted flow on the basis of past measurements and a suitable parameter tracking algorithm for the flow process. This work introduces some of the fundamental techniques and principles of control theory broadly classified as cybernetics and indicates how they might be used for forecasting of hourly flow of non-tidal river during a short span storm period. This forecasting model would help in the subsequent development of the methodologies for on-line operation of complex water resources control systems.

In the present work the recursive non-stationary time series analysis technique has been used for on-line forecasting of hourly river flow during a storm period. The recursive parameter estimation algorithms have the advantages of minimal computational requirements with ability to track the time varying parameters. On the basis of these information processing algorithms a real time dynamic model of the river flows during a short span storm period has been obtained to forecast the hourly flows at Domohani Road pak Bridge point of the river Teesta near Jalpaiguri town in North Bengal.

4.2. Development of non-stationary process model

Non-stationary process can be expressed as,

$$Y(k) = U(k) + V(k) \qquad \ldots (4.1)$$

where $U(k)$ is the deterministic input component,
$V(k)$ is the stationary stochastic component,
$k = 1, 2, \ldots,$ are time instances.

The strong correlation of down-stream flows with those at up-stream flows particularly at the confluence of tributaries suggests that the process may be represented in the form of a non-stationary time series with a probability that the current value of the output $Y(k)$ is a function of the previous output observations, the auto regressive terms $Y(k-1)$, $Y(k-2)$, ..., and the highly correlated deterministic inputs, the up-stream flows.

$U1(k-r_1)$, $U_2(k-r_2)$, ..., $U_1(k-r_1-1)$, $U_2(k-r_2-1)$, ... ,

together with the current unknown realisation of the noise process

$\gamma(k)$.

The process can be represented as

$$Y(k) = \sum_{i=1}^{n} \beta(k-i)Y(k-i) + \sum_{j=1}^{n} \sum_{i=0}^{n} \alpha(k-ji)U_j(k-r_j-i) + \eta(k) \quad \ldots (4.2)$$

where the deterministic input components

$$U(k) = \sum_{j=1}^{n} \sum_{i=0}^{n} \alpha(k-j)U_j(k-r_j-i)$$

and the stationary stochastic component

$$V(k) = \sum_{i=1}^{n} \beta(k-i)\, Y(k-i) + \eta(k)$$

$\eta(k)$ is modelled as

$$\eta(k) = \sum_{p=1}^{r} \gamma(k-p)\, \eta(k-p) + \nu(k) \qquad \ldots \quad (4.3)$$

Where $\nu(k)$ is a white noise innovation process with

$$E\{\nu(k) \quad \nu(j)\} = 0 \quad \text{for } j \neq k$$

and $$E\{\nu(k) \quad \nu(j)\} = 1 \quad \text{for } j \neq k$$

The determination of n is known as the model order determination. The output correlation has been suggested as an intuitive consideration for model order determination which is also known as model order identification. The determination of r in autoregressive error sequence $\eta(k)$ depends on the conditions when residuals of the estimate of $\eta(k)$ i.e., $\eta(k) = \hat{\eta}(k)$ becomes a white noise process.

The model is depicted in equation (4.2) is quite flexible since it requires that the equation be linear in parameters.

The equation (4.2) may be represented in the form as

$$Y(k) = a^T(k-1)\, Z(k-1) + \gamma(k) \quad \ldots (4.4)$$

where

$$a^T(k-1) = \Big[\, \beta(k-1),\ \beta(k-2),\ \ldots,\ \beta(k-n) \sum_{j=1}^{m} \Big[\, \alpha(k-j_0),\ \alpha(k-j_1),\ \ldots,\ \alpha(k-j_n) \,\Big],\ \gamma(k-1)$$

$$\gamma(k-2),\ \ldots,\ \gamma(k-r) \,\Big] \quad \ldots (4.5)$$

and

$$Z(k-1) = \Big[\, Y(k-1),\ Y(k2),\ \ldots,\ Y(k-n) \sum_{j=1}^{m} \Big[\, U_j(k),\ U_j(k-1),\ \ldots,\ U_j(k-n) \,\Big]$$

$$\gamma(k-1),\ \gamma(k-2),\ \ldots,\ \gamma(k-\theta) \,\Big]^T \quad \ldots (4.6)$$

The estimate of $Y(k)$ is written as

$$\hat{Y}(k) = a^T(k-1)\, Z(k-1) \quad \ldots (4.7)$$

4.3 Development of recursive parameter estimation algorithm

The coefficient vector 'a' can be estimated by minimising the quadratic function $J_k(a)$,

$$J_k(a) = \sum_{j=1}^{k} (Y(j) - a^T Z(j-1))^2 + (a-a(o))^T S^{-1}(o) (a-a(o)) \quad \ldots (4.8)$$

where $a(o)$ is the available a priori estimate of the coefficient vector 'a' and $S(o)$ is the positive definite weighting matrix of the order $q \times q$, q being $n + (n+1)\, m + r$.

For minimization

$$\frac{\partial J_k(a)}{\partial a} = 2 \sum_{j=1}^{k} Z(j-1)Y(j) - a^T Z(j-1) + 2S^{-1}(o)(a-a(o)) \quad \ldots (4.9)$$

since we are seeking for a minimum

$$\frac{\partial J_k(a)}{\partial a} = 0 \quad \ldots (4.10)$$

By matrix inversion lemma and by denoting the estimate of 'a' as 'a (k)' at the time instant k the recursive parameter estimation algorithm becomes,

$$a(k+1) = a(k) + S(k+1)Z(k) \left[Y(k+1) - a^T(k)Z(k) \right] \quad \ldots (4.11a)$$

$$S(k+1) = S(k) - S(k)Z(k)Z^T(k)S(k) \left[1 + Z^T(k)S(k)Z(k) \right]^{-1}$$

The algorithms have been initialised with

$k = o$, $S(o) \hat{=} I$ (unit matrix), $a(o) \hat{=} o$ and

$Y(j) = o$, for $j = o, -1, -2, \ldots$

4.4. Illustration.

Data for a short span storm period from 16.00 hours on 23.7.79 to 15.00 hours on 25.7.79 were observed at the gauging stations at Sankalan, Rongpo (Teesta), Great Rangeet (Singlabazar), Teestabazar, Coronation Bridge and Domohani Road Bridge (Jalpaiguri town) of the river Teesta in North Bengal. The gauging stations are shown in Fig 4.4.1.

The input and the output variables are rationalised in accordance with

$$x(k) = \frac{X(k) - X_{min}}{X_{max} - X_{min}} \quad \ldots (4.12)$$

where $x(.)$ is the observed value at the k-th instant and X_{min} and X_{max} are the minimum and maximum values of the data sequence. The input variables are selected on the basis of the strongest coefficient of correlation with the flows at Domohani. The coefficient of correlation has been defined as,

$$\phi_{yx}(\lambda) = \frac{\sum_{i=1}^{N-\lambda}\left(Y(i) - \frac{1}{N}\sum_{i=1}^{N} y(i)\right)\left(x(i+\lambda) - \frac{1}{N}\sum_{i=1}^{N} x(i)\right)}{\left[\sum_{i=1}^{N-\lambda}\left(y(i) - \frac{1}{N}\sum_{i=1}^{N} y(i)\right)^2 \sum_{j=1+\lambda}^{N}\left(x(j) - \frac{1}{N}\sum_{i=1}^{N} x(i)\right)^2\right]^{\frac{1}{2}}} \quad \ldots (4.13)$$

where is the shift of instances of time. On the basis of the coefficients of correlation the hourly flows at Domohani Road Bridge in rationalised units may be expressed as

$$\bar{y}^{(k)} = f(y(k-1), y(k-2), x_1(k-6), x_2(k-7), x_3(k-9), x_4(k-8), x_5(k-10) \quad \ldots (4.14)$$

where $x_1(.)$, $x_2(.)$, $x_3(.)$, $x_4(.)$ and $x_5(.)$ are respectively the hourly flows at Coronation Bridge, Teesta Bazar, Great Rangeet (Singlabazar), Rongpo (Teesta) and Sankalan respectively.

With output prehistory interval = 2 and the moving average terms r = 2, the integral square error defined as

$$ISE = \frac{\sum_{i=1}^{N} \left[y(i) - y(i) \right]^2}{\sum_{i=1}^{N} (y(i))^2} \quad \ldots (4.15)$$

has been found as 0.017584 where $y(.)$ is the modelled output. Fig 4.4.2a and Fig 4.4.2b show the observed and the modelled values. The model is found to adhere very closely to the observed data. Algorithms are simple and require no elaborate initialisation [105].

FIG- 1

TEESTA

AND ITS TRIBUTARIES.

SCALE - 1 INCH = 16 MILES

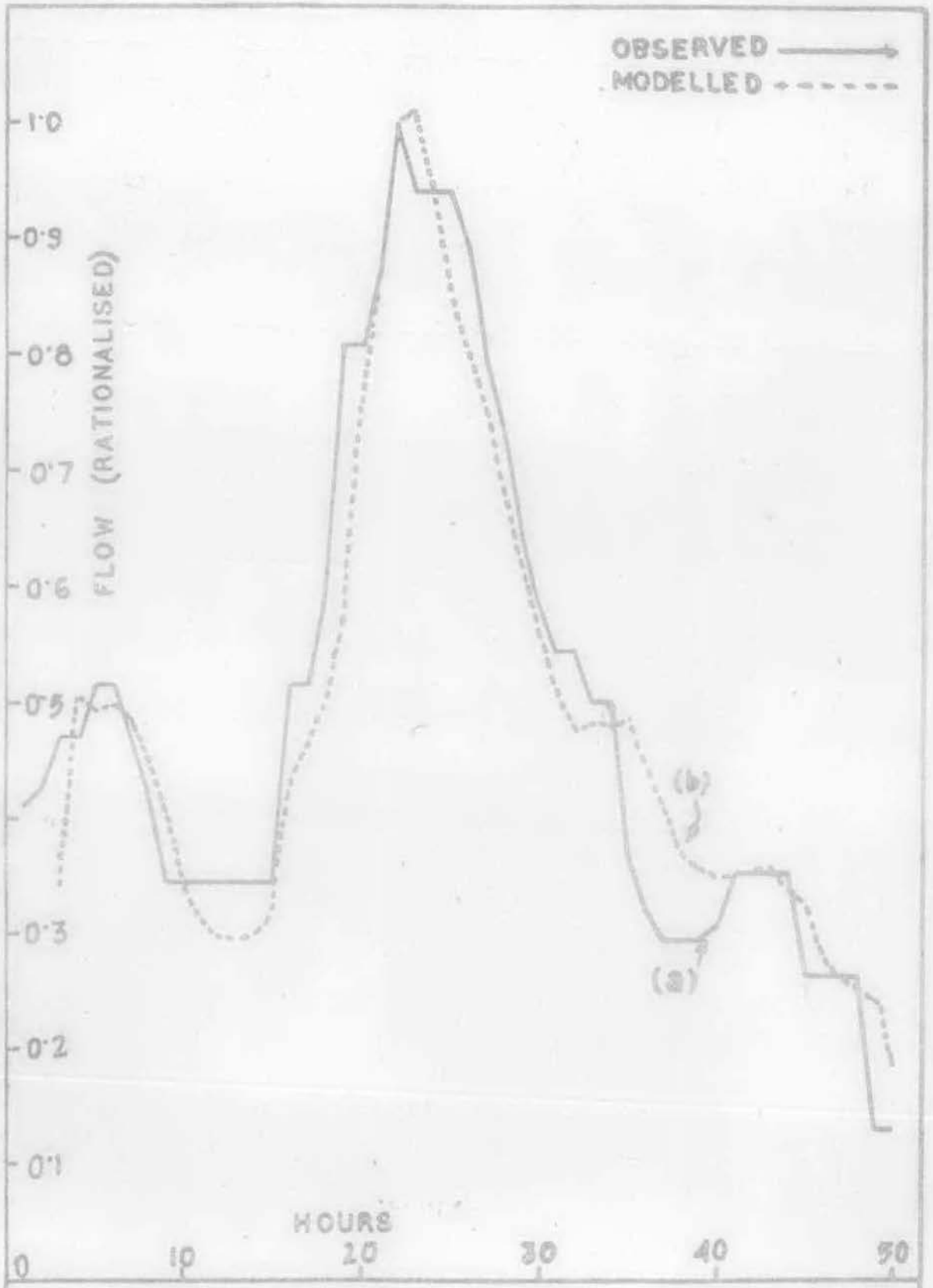


FIG 4.4.2 OBSERVED AND MODELLED VALUES OF FLOW

4.5 Conclusion

From the work presented here it is claimed that on-line flood warning and control measures can be implemented provided adequate real time information processors are installed at the gauging stations with closed loop decision regularisation for operation and control of hydraulic structures of the complex water resources systems.

CHAPTER V

REAL TIME FORECASTING OF DAILY FLOWS OF A NON TIDAL RIVER BY INSTRUMENT VARIABLE ALGORITHM

CHAPTER V

REAL TIME FORECASTING OF DAILY FLOWS OF A NON TIDAL RIVER BY INSTRUMENT VARIABLES ALGORITHMS

5.0 Abstract

A recursive procedure is presented for real time forecasting of daily flows of a non tidal river. The dynamic model is found to simulate adequately the major variations observed in the field measurements of daily flows of the river Teesta gauged at Domohani Road Bridge near Jalpaiguri town in North Bengal.

5.1 Introduction

The paper deals with the problem of identifying the parameters described by linear stochastic difference equations. The technique includes application of recursive instrument variable algorithms. From the knowledge of the available data, a recursive procedure is presented for real time forecasting of daily flows of a non tidal river to best fit and also to track adaptively the observed data in the least squared sense to give one step ahead prediction. When tested with field data the mathematical description of the daily river flow process is found to simulate adequately the major variations observed in the field measurements collected over a 5-year period from the 1st. January, 1975 to the 31st. December, 1979 of the river Teesta in North Bengal at Domohani gauging station near Jalpaiguri town. The technique of real time forecasting discussed in this paper highlight the possibilities of implementation of on-line regulatory mechanism for controlling daily flow of a non tidal river in accordance with a predetermined decision methodology.

The recursive instrument variable algorithms have the advantages of minimum computational requirements and they have the ability to adaptively update the slowly varying parameters of the physical process.

5.2. Development of the real time mathematical description of the daily river flow process.

In recursive estimation algorithm the estimate of the output at a future instant is constantly updated on receipt of fresh information. The presence of significance co-efficient of correlation of daily river flow at past instants with the current instant clearly suggests that the process may be represented in the form of a time series with the current value of the output $p(k)$ as a function of the previous output observations, the auto-regressive terms $P(k-1)$, $P(k-2)$, ... , together with the current unknown realisation of the noise process. The process thus, can be represented as

$$P(k) = \left[\sum_{i=1}^{N} \beta_i q^{-1} \right] P(k) + \eta(k) \qquad \ldots \quad (5.1)$$

where q^{-1} is the backward shift operator defined as $q^{-1} P(k) = P(k-1)$. The determination of N is known as the model order determination which is also the model order identification. The number of the significant instants in the output correlation may be intuitively considered for model order determination.

$\eta(k)$ may be expressed in a moving average sequence as

$$\eta(k) = \sum_{p=1}^{P} \gamma_p \eta(k-p) + \upsilon(k) \qquad \ldots \quad (5.2)$$

Where $v(.)$ is a white noise innovation process with

$$E\{v(k)\ v(j)\} = 0 \quad \text{for } j \neq k \text{ and}$$
$$E\{v(k)\ v(j)\} = \rho^2 \quad \text{for } j = k \qquad \ldots (5.3)$$

The equation (5.1) can be written as

$$P(k) = s^T(k)\alpha + \gamma(k) \qquad \ldots (5.4)$$

where α is the parameter vector with property of slowly varying with time and amenable to recursive adaptiveness, and

$$s^T(k) = [\, P(k-1),\ P(k-2),\ \ldots,\ P(k-N) \,]$$

The least square estimation of the process parameters can be obtained by minimising a loss function defined as the sum of the squared errors as

$$J \triangleq \sum_{k=1}^{N_k} (P(k) - s^T(k)\hat{\alpha})^2 \qquad \ldots (5.5)$$

The estimates $\hat{\alpha}$ of α that minimise J are called least square estimates and $\hat{\alpha}$ is defined as

$$\hat{\alpha} = \left[\sum_{k=1}^{N_1} s(k)\, s^T(k) \right]^{-1} \left[\sum_{k=1}^{N_1} s(k)\, P(k) \right] \qquad \ldots (5.6)$$

Let $R^*(k) \triangleq \left[\sum_{j=1}^{k} s(j)\, s^T(j) \right]^{-1}$

and

$$S(k) = \left[\sum_{j=1}^{k} s(j)\, P(j) \right]$$

So equation (5.6) becomes

$$\hat{\alpha} = R^{*}(k)\ S(k) \qquad \ldots \qquad (5.7)$$

The recursive relationship for $R^{*}(.)$ and $S(.)$ can be set as

$$\left[R^{*}(k) \right]^{-1} = \sum_{j=1}^{k=k} x(j)\, x^{T}(j) + x(k)\ x^{T}(k)$$

$$= \left[R^{*}(k-1) \right]^{-1} + x(k)\ x^{T}(k) \qquad \ldots \qquad (5.8)$$

Similarly,

$$S(k) = S(k-1) + x(k)\ y(k) \qquad \ldots \qquad (5.9)$$

Premultiplying by $R^{*}(k)$ we get from equation (5.8)

$$I = R^{*}(k) \left[R^{*}(k-1) \right]^{-1} + R^{*}(k)x(k)x^{T}(k) \qquad \ldots \qquad (5.10)$$

Postmultiplication of equation (5.10) by $R^{*}(k-1)$

$$R^{*}(k-1) = R^{*}(k) + R^{*}(k)\, x(k)\, x^{T}(k)\, R^{*}(k-1) \qquad \ldots \qquad (5.11)$$

Postmultiplying equation (5.11) by $x(k)$.

$$R^{*}(k-1)\, x(k) = R^{*}(k)x(k) + R^{*}(k)x(k)x^{T}(k)\, R^{*}(k-1)x(k)$$

So we get,

$$R^{*}(k-1)x(k) = R^{*}(k)x(k) \left[1+x^{T}(k)\, R^{*}(k-1)\,(k) \right] \qquad \ldots \qquad (5.12)$$

Post multiplication by $\left[1+x^{T}(k)\ R^{*}(k-1)x(k) \right]^{-1} x^{T}(k)\, R^{*}(k-1)$ the equation (5.12) gives,

$$R^{*}(k-1)x(k) \left[1+x^{T}(k)\, R^{*}(k-1)x(k) \right]^{-1} x^{T}(k)\, R^{*}(k-1) = R^{*}(k)x(k)$$

$$x^{T}(k)\, R^{*}(k-1) \qquad \ldots \qquad (5.13)$$

On substitution of equation (5.11) in equation (5.13)

$$R^*(k) = R^*(k-1) - R^*(k-1)\, s(k) \left[1+s^T(k)R^*(k-1)s(k) \right]^{-1} \times s^T(k)\ R^*(k-1) \quad \ldots \quad (5.14)$$

From equation (5.7) denoting the estimate $\hat{\alpha}$ at time instant k as $\hat{\alpha}(k)$

$$\hat{\alpha}(k) = R^*(k)\ S(k)$$

Combining equations (5.9) and (5.14)

$$\hat{\alpha}(k) = \left[R^*(k-1) - R^*(k-1)\, s(k) \left[1+s^T(k)\ R^*(k-1)\, s(k) \right]^{-1} \times s^T(k)\ R^*(k-1) \right] \left\{ S(k-1) + s(k)\ P(k) \right\} \quad \ldots \quad (5.15)$$

since $R^*(k-1)\ S(k-1) = \hat{\alpha}(k-1)$

The equation (5.15)

$$\hat{\alpha}(k) = \hat{\alpha}(k-1) - R^*(k-1)\, s(k) \left[1+s^T(k)\ R^*(k-1)s(k) \right]^{-1} \times s^T(k)\ \hat{\alpha}(k-1) + R^*(k-1)\, s(k)\, P(k) - R^*(k-1)\, s(k) \left[1+s^T(k)\ R^*(k-1)\, s(k) \right]^{-1} \times \left\{ Z^T(k)\ R^*(k-1)\ Z(k)\ P(k) \right\} \quad (5.16)$$

since $s^T(k)\ R^*(k-1)\ s(k)$ is scalar the equation (5.6) becomes

$$\hat{\alpha}(k) = \hat{\alpha}(k-1) + R^*(k-1)\, s(k) \left[1 + s^T(k)\ R^*(k-1)\, s(k) \right]^{-1} \times \left[P(k) - s^T(k)\ \hat{\alpha}(k-1) \right] \quad \ldots \quad (5.17)$$

The equations (5.17), (5.14) with

$$R^*(k) \triangleq \left[\sum_{j=1}^{k} s(j)\ s^T(j) \right]^{-1}$$

are the recursive least square algorithms. But these algorithms do not overcome the problem of bias. The technique to minimise the bias is stated as follows.

From equation (5.6) the non recursive estimation of parameters at instant k is

$$\hat{\alpha}(k) = \left[\sum_{j=1}^{k} x(j)\, x^T(j) \right]^{-1} \left[\sum_{j=1}^{k} x(j)\, P(j) \right] \quad \ldots (5.18)$$

and the process equation is,

$$P(j) = x^T(j)\alpha + \gamma(j) \quad \ldots (5.19)$$

Upon substitution of equation (5.19) in equation (5.18)

$$\hat{\alpha}(k) = \left[\sum_{j=1}^{k} x(j)x^T(j) \right]^{-1} \left[\sum_{j=1}^{k} x(j) \left[x^T(j)\alpha + \gamma(j) \right] \right]$$

$$= \left[\sum_{j=1}^{k} x(j)x^T(j) \right]^{-1} \left[\sum_{j=1}^{k} x(j)x^T(j) \right] \alpha$$

$$+ \left[\sum_{j=1}^{k} x(j)x^T(j) \right]^{-1} \left[\sum_{j=1}^{k} x(j)\, \gamma(j) \right] \quad \ldots (5.20)$$

So,

$$\hat{\alpha}(k) = \alpha + \left[\sum_{j=1}^{k} x(j)\, x^T(j) \right]^{-1} x \left[\sum_{j=1}^{k} x(j)\, \gamma(j) \right]$$

For the estimate of $\hat{\alpha}(k)$ to converge to

$$E\left\{ x(j)\, \gamma(j) \right\} = 0 \quad \text{for all } j$$

and

$$E\left\{ \gamma(j)\, \gamma(k) \right\} = 0 \quad \text{for } j \neq k$$

i.e. $\gamma(j)$ converge to white noise process $v(j)$. Under these

conditions the estimates will be unbiased. However, these conditions are difficult to achieve in practice. And consequently the vector $z(k)$ has been modified incorporating the instrument variable $\hat{P}(.)$ as $\hat{z}(k)$ defined as

$$\hat{z}(k) = \left[\hat{P}(k-1),\ \hat{P}(k-2),\ldots,\ \hat{P}(k-n) \right]^{T} \quad \ldots \quad (5.21)$$

where $\hat{P}(.)$ is the recursive least square estimation of $P(.)$. Since $\hat{P}(.)$ sequence in equation (5.21) are already estimates, in a sense optimum, they are almost uncorrelated with $\eta(k)$. Thus the conditions of unbiased estimates are modified as

$$E\left\{ \hat{z}(k) \quad \eta(k) \right\} = 0 \quad \text{for all } k$$

Replacing $z(k)$ by $\hat{z}(k)$ rather heuristically and not $z^{T}(k)$ by $\hat{z}^{T}(k)$, because of appearance in the interpolation region, in equations (5.14) and (5.17) the recursive instrument variable algorithm becomes,

$$\hat{\alpha}(k) = \hat{\alpha}(k-1) + \hat{R}^{*}(k-1)\hat{z}(k) \left[1+z^{T}(k)\hat{R}^{*}(k-1)z(k) \right]^{-1} \times \left\{ P(k) - z^{T}(k) \times \hat{\alpha}(k-1) \right\} \quad \ldots \quad (5.22a)$$

$$\hat{R}^{*}(k) = \hat{R}^{*}(k-1) - \hat{R}^{*}(k-1)\hat{z}(k) \left[1+z^{T}(k)\hat{R}^{*}(k-1)\hat{z}(k) \right]^{-1} \times z^{T}(k) \quad \hat{R}^{*}(k-1) \quad \ldots \quad (5.22b)$$

$$R^{*}(k) = \left[\sum_{j=1}^{k} \hat{z}(j) \quad \hat{z}(j) \right]^{-1} \quad \ldots \quad (5.22c)$$

The algorithm may be initialised with $\hat{\alpha}(.)$ as a least square estimate for a block of data in the interpolation region and $\hat{R}^{*}(.)$ for the same block of data.

5.3. Illustration.

A representative portion of the data set i.e. the daily flow of the river Teesta gauged at Domohani near Jalpaiguri town, from the 20th. January, 1978 to the 31st. March, 1978 is shown in Fig 5.3.1(a). The correlation co-efficient versus shift of instances of time of the data set from the 1st. January, 1978 to the 31st. March, 1978 is shown in Fig 5.3.2 . It is evident that the process has a significant prehistory interval N = 8. Fig 5.3.1(b) shows the errors of prediction. The values of the integral squared errors and the mean error are 0.027353 and 0.007206 cubic metre per second respectively. It is observed from the correlation co-efficient versus shift of time instances of the error sequence shown in Fig 5.3.3 that the errors of prediction are almost uncorrelated [106, 107] .

5.4. Conclusion.

A dynamic model for real time simulation of daily river flows has been presented. The model gives one step ahead on-line prediction of daily flows on the basis of the currently available information. It is hoped that the real time recursive prediction of daily river flows would assist the designer and operator of the large scale water resources systems involving control, utilisation and disposition of river water.

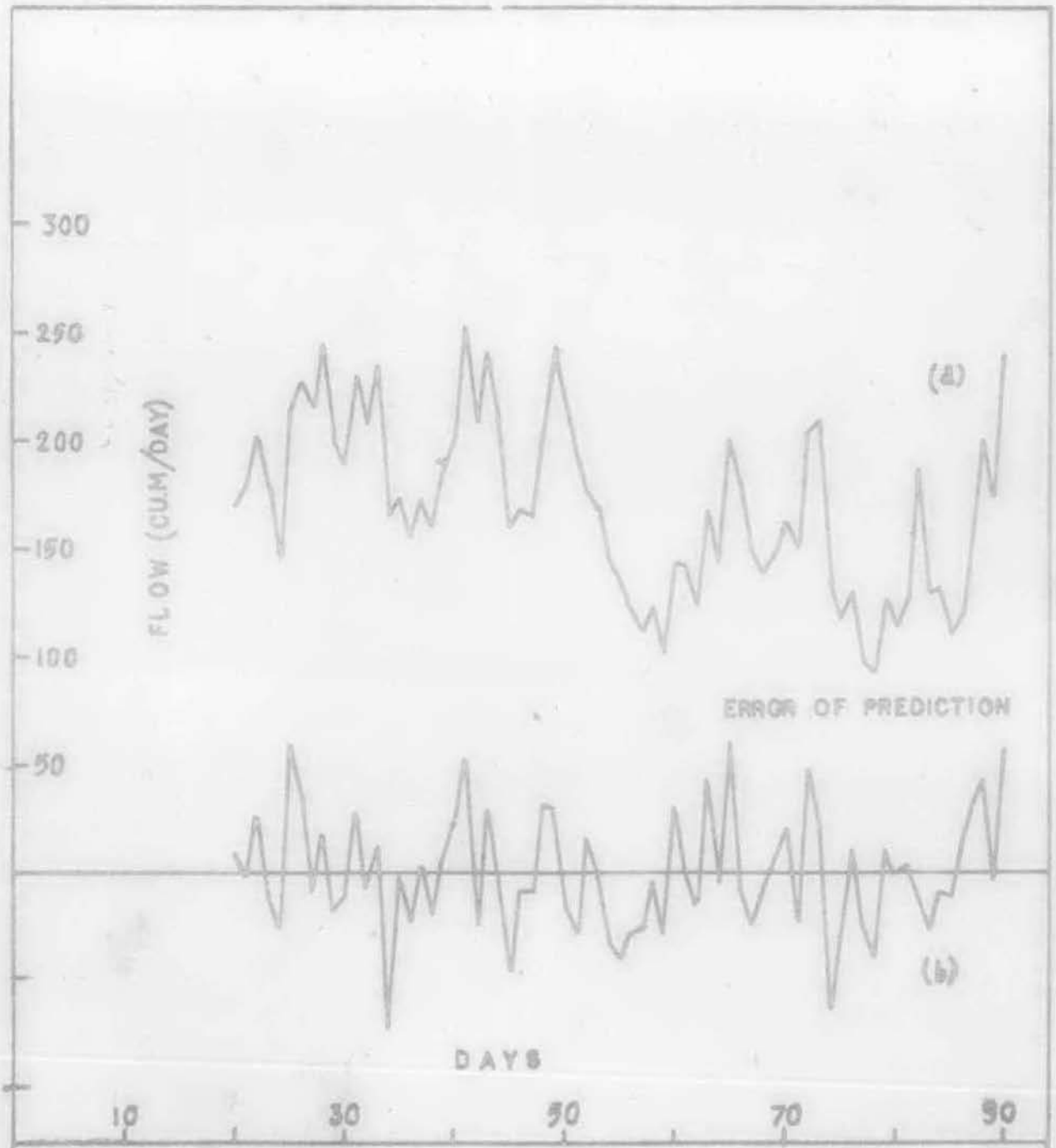


FIG 5.3.1 OBSERVED DAILY FLOW AND ERROR OF PREDICTION

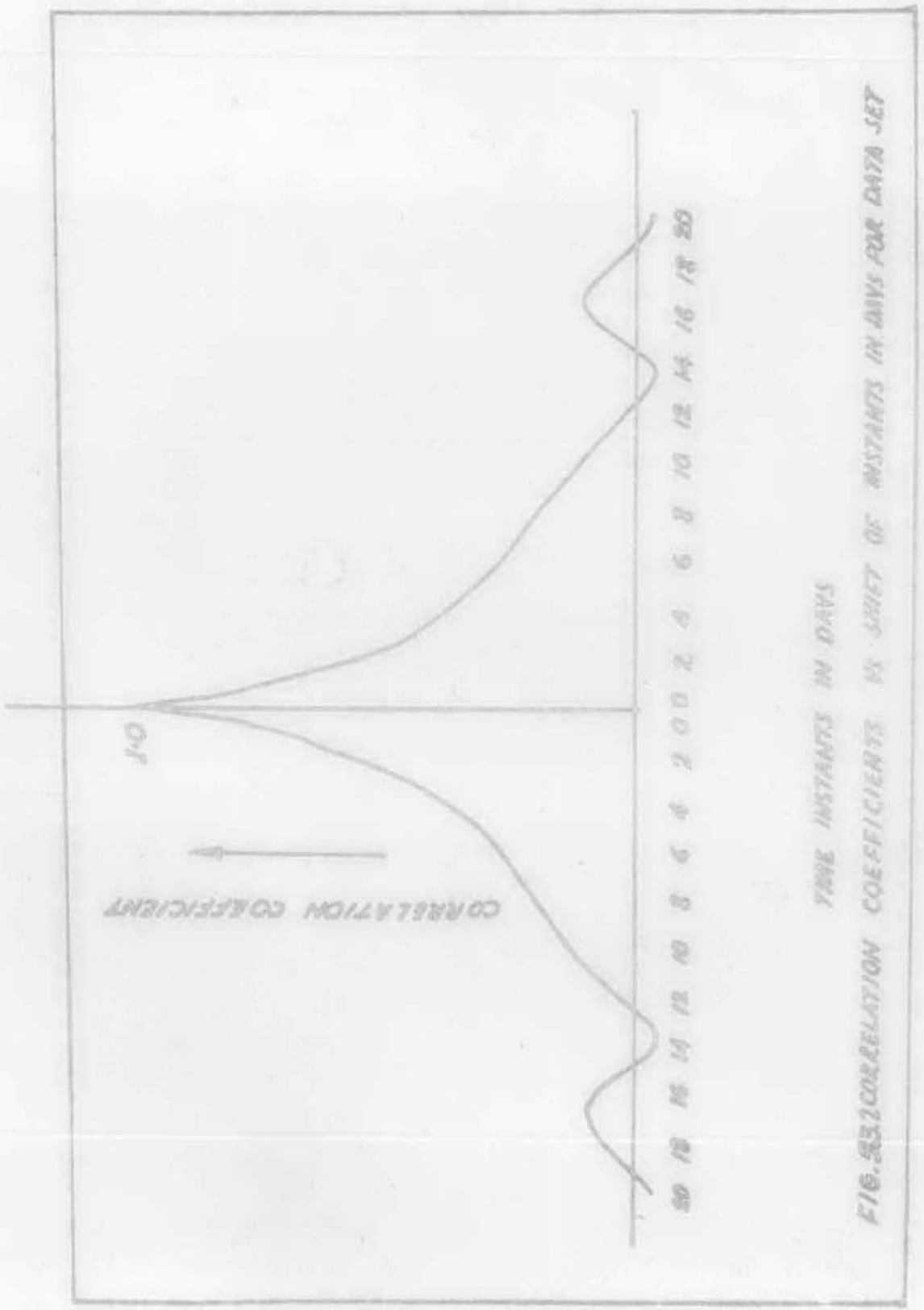


FIG. 5.3.1 CORRELATION COEFFICIENTS VS SHIFT OF INSTANTS IN DAYS FOR DATA SET

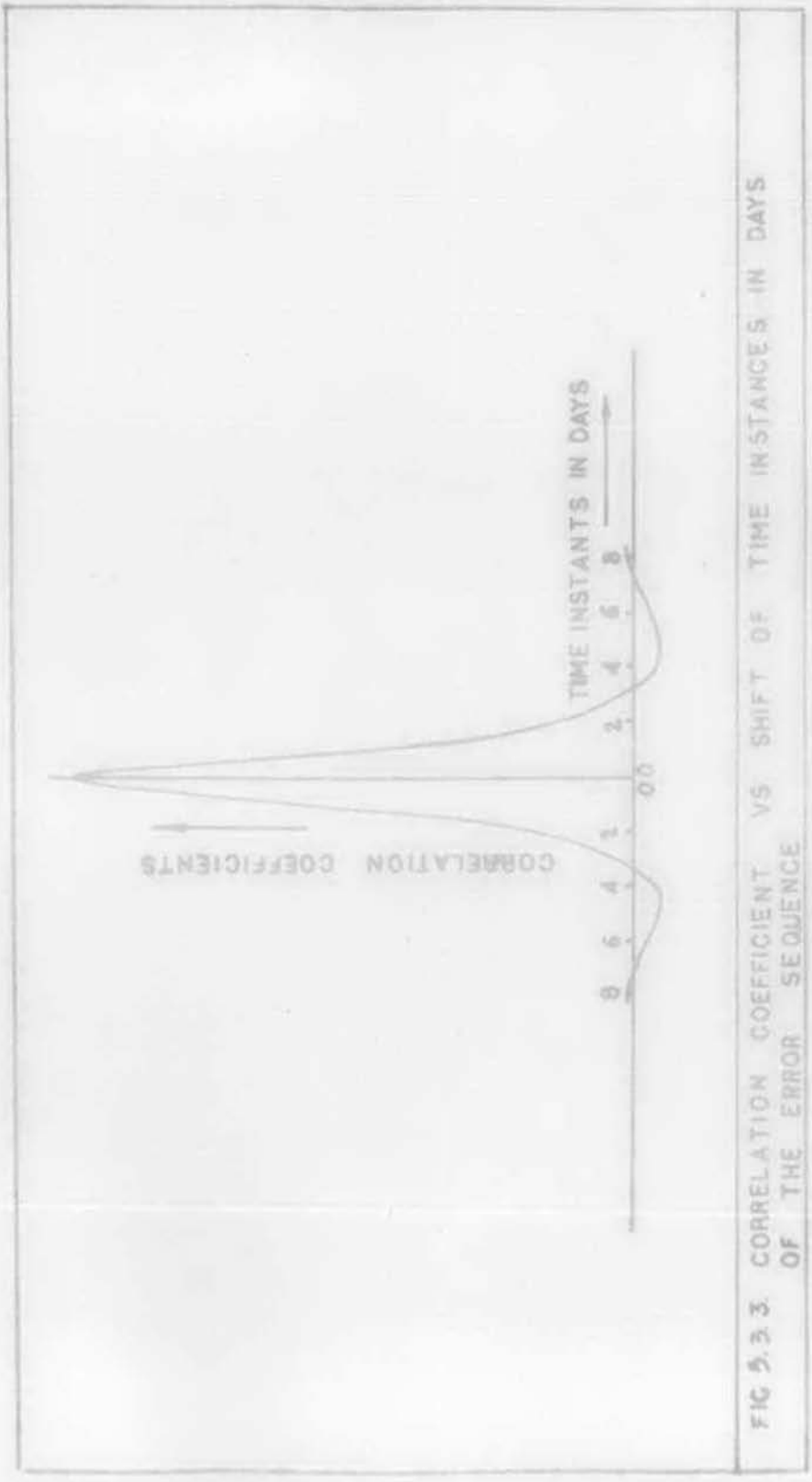


FIG 5.3.3. CORRELATION COEFFICIENT VS SHIFT OF TIME INSTANCES IN DAYS OF THE ERROR SEQUENCE

CHAPTER VI

FORECASTING OF DAILY DISSOLVED OXYGEN LEVELS OF A NON TIDAL RIVER BY SELF ORGANISATION OF MATHEMATICAL MODELS.

CHAPTER VI

FORECASTING OF DAILY DISSOLVED OXYGEN LEVELS OF A NON TIDAL RIVER BY SELF ORGANISATION OF MATHEMATICAL MODELS.

6.0 Abstract :

A simple dynamic model of the dissolved oxygen levels of a nontidal river has been obtained. The model structure has been identified with the help of a learning identification algorithm known as combinatorial group method of data handling technique. The model has been verified by simulation against field data. It is found that the major variations observed in the measured data have been adequately represented in the model.

6.1. Introduction :

The realistic modelling of dissolved oxygen levels (D.O) is important for controlling the water quality of a river. This paper presents a simple model for forecasting daily D.O. levels of a nontidal river. The D.O. levels are a major index of the general state of the aquatic environment. The important engineering decisions in water quality control relate to the determination, regulation, maintenance and control of the level of D.O. that is consistent with the multiple uses of the natural water bodies. Furthermore, it has been observed that even the temporary fluctuations of the D.O. level below the critical values can cause a long term and possibly irreversible damage to the ecosystem.

This work introduces a technique of technical cybernetics classified as the theory of self organisation of mathematical models and commonly known as the combinatorial group method of data handling

algorithms. With the help of this method it has been possible to formulate mathematical models of complex processes with prediction optimisation.

6.2. The Concept of Self-Organisation :

When the model complexity gradually increases the computer finds by shifting the different models, the minimum of a selection criterion which the computer has been ordered to look for. Thus the computer indicates to the operator the model of optimum complexity.

The purpose of this part of the work is to obtain an easily applicable model of optimum complexity for daily forecasting of D.O. levels of a nontidal river. The model has been verified by field measurement of dissolved oxygen collected over a 80-day period from the river Cam in Eastern England.

Model Development :

The daily measured data for D.O. concentration of a stream are

$P(i)$ $i = 1, 2 \ldots , N1$, i being the time instants in days

The mean value is given by,

$$\bar{P}(N1) = \frac{1}{N1} \sum_{i=1}^{N1} p(i) \qquad \ldots \quad (6.1)$$

Auto covariance of the data at lag instant k is given by

$$CA(k) = \frac{1}{N1-k} \sum_{i=1}^{N1-k} \left[p(i) - \bar{p}(N1) \right] \left[P(i+k) - \bar{p}(N1) \right] \quad \ldots (6.2)$$

Where $k = 0, 1, 2, \ldots, M$ and $M < \frac{1}{4} N1$

The normalised co-efficients of covariance are given by

$$RA(k) = \frac{CA(k)}{CA(0)} \qquad \ldots \quad (6.3)$$

The estimate of the normalised power density spectra for the data are given by,

$$PS(w_h) = \frac{2}{\pi} \sum_{k=0}^{M} E_k R_A(k) \cos w_h k \quad \ldots \quad (6.4)$$

where $w_h = 2\pi f_h$ and $f_h = \frac{h}{2M}$, $0 \leq f_h \leq 0.5$

and $h = 0, 1, 2, \ldots, M$

E_k has been defined as the weight for window correction and may be taken as

$$E_k = \begin{cases} 1.0 & \text{for } 0 < k < M \\ 0.5 & \text{for } k = 0, M \end{cases} \quad \ldots \quad (6.5)$$

These raw estimate of power spectral density are smothed by using Hanning Window to obtain the final estimates of the power spectrum. The smoothed estimates of the ordinates of the power spectrum are given by

$$\begin{aligned} S(w_h) &= 0.54\ PS(w_h) + 0.46\ PS(w_1) \quad \text{for } h = 0 \\ S(w_h) &= 0.23\ PS(w_{h-1}) + 0.54\ PS(w_h) + 0.23\ PS(w_{h+1}) \\ &\qquad \text{for } 0 < h < M \quad \ldots \quad (6.6) \\ S(w_h) &= 0.54\ PS(w_h) + 0.46\ PS(w_{h-1}) \quad \text{for } h = M \end{aligned}$$

The periodicity in terms of fundamental and its harmonics can be estimated from the power spectral density versus frequency characteristics.

The non-stationary process model with deterministic periodicity can be expressed as

$$P(k) = U(k) + V(k) \quad \ldots\ldots \quad (6.7)$$

Where

$U(k)$ is the deterministic trend component,

$V(k)$ is the stationary stochastic components

$k = 1, 2, \ldots,$ are time instances.

The deterministic periodic components can be represented as

$$U(k) = \sum_{r}^{R} \left(a_r \cos(r 2\pi f_0 k) + b_r \sin(r 2\pi f_0 k) \right) \quad \ldots (6.8)$$

Where f_o is the fundamental cycle per instance of time, the stochastic components V(k) can be expressed in autoregressive form as

$$V(k) = \sum_{n=1}^{N} C_{k-n} \; V(k-n) + \eta(k) \qquad \ldots \quad (6.9)$$

Where $n = 1, 2 \ldots, N$ the prehistory interval and $\eta(k)$ the error sequence. The error sequence $\eta(k)$ is modelled in the moving average form as,

$$\eta(k) = \sum_{j=1}^{m} q_{k-j} \; \eta(k-j) + \nu(k) \qquad \ldots \quad (6.10)$$

Where m is the prehistory interval of the moving average terms of $\eta(k)$. $\nu(k)$ resembles to a white noise process with zero mean and constant variance

V(k) can be written as

$$V(k-n) = \beta(k-n) - U(k-n) \qquad \ldots \quad (6.11)$$

Hence equation (6.9) can be written as

$$V(k) \sum_{n=1}^{N} C_{k-n} \; P(k-n) - \sum_{n=1}^{N} C_{k-n} \; U(k-n) + \eta(k) \qquad \ldots \quad (6.12)$$

From equations (6.7) and (6.8) it follows that,

$$P(k) = \sum_{n=1}^{N} C_{k-n} \; P(k-n) + \sum_{n=0}^{N} \sum_{r=0}^{R} \left[d^{r}_{k-n} \cos\left(r 2\pi f_o (k-n)\right) \right.$$
$$\left. + g^{r}_{k-n} \sin\left(r 2\pi f_o (k-n)\right) \right] + \eta(k) \qquad \ldots \quad (6.13)$$

In matrix from the process P(k) can be expressed as

$$P(k) = a^{T}(k-1) \; Z(k-1) + \eta(k) \qquad \ldots \quad (6.14)$$

Where $a^{T}(k-1) = \left[C_{k-1}, \ldots, C_{k-n}, \sum_{n=0}^{N} d^{0}_{k-n} \ldots, d^{R}_{k-n}, \right.$

$$\left. g^{0}_{k-n} \ldots, \; g^{R}_{k-n} \right) \Big]^{T} \ldots (6.15)$$

and

$$Z(k-1) = [P(k-1), \ldots, P(k-N), 1, \ldots, \sin(R2\pi f_0(k-N)), \ldots, \cos(R2\pi f_0(k-N))]^T \quad \ldots (6.16)$$

In functional form

$$P(k) = f\big(P(k-1), P(k-2), \ldots P(k-N), \sin(2\pi f_0 k), \cos(2\pi f_0 k), \sin(2\pi f_0(k-1)), \cos(2\pi f_0(k-1)) \ldots \sin(2\pi f_0(k-N)), \cos(2\pi f_0(k-N)) \ldots \sin(R2\pi f_0(k-N)), \cos(R2\pi f_0(k-N))\big) + \eta(k) \quad \ldots (6.17)$$

For simplicity let us assume that N = 2 and R = 1

Then

$$P(k) = f\big(P(k-1), P(k-2), \sin(2\pi f_0 k), \cos(2\pi f_0 k), \sin(2\pi f_0(k-1)), \cos(2\pi f_0(k-1)), \sin(2\pi f_0(k-2)), \cos(2\pi f_0(k-2))\big) + \eta(k) \quad (6.18)$$

Let $P(k) = y$, $P(k-1) = x_1$, $P(k-2) = x_2$, $\sin(2\pi f_0 k) = x_3$, $\cos(2\pi f_0 k) = x_4$, $\sin(2\pi f_0(k-1)) = x_5$, $\cos(2\pi f_0(k-1)) = x_6$, $\sin(2\pi f_0(k-2)) = x_7$ and $\cos(2\pi f_0(k-2)) = x_8$

and $\eta(k)$ as residual ξ

The equation (6.18) becomes

$$\text{Or} \quad Y = f(x_1, x_2, \ldots, x_8) + \xi \quad \ldots (6.19)$$

The estimate of y as $\hat{y}$ is defined as

$$\hat{y} = f(x_1, x_2, \ldots\ldots, x_8) \quad \ldots (6.20)$$

The function f(.) in equation (6.20) is sought in the class of quadratic polynomials on the basis of a table of polynomials of gradually increasing complexity of eight variables as shown in table 6.2.1 with the theory of self organisation of mathematical models.

The model of optimum complexity is selected on the basis of minimum of integral square error criterion. The integral square error is defined as

$$ISE = \frac{\sum_{i=1}^{N1} (y_o(i) - y_m(i))^2}{\sum_{i=1}^{N1} (y_o(i))^2} \qquad \ldots \quad (6.21)$$

Where $y_o(i)$, i=1, 2, ..., N = days, are the observed values of the output variable in the interpolation region and $Y_m(i)$ are the values of the variables obtained from the model. For the model of optimum complexity the residual has almost the property of a random variable with correlation co-efficient.

$\phi_{\xi\xi}(\lambda) = 1.0$ for $\lambda = 0$ | where λ is the
$\phi_{\xi\xi}(\lambda) = 0$ for $\lambda \neq 0$ | shift of instance

The correlation co-efficient is defined as

$$\phi_{\xi\xi}(\lambda) = \frac{\sum_{i=1}^{N1-\lambda} (P(i) - \frac{1}{N1} \sum_{i=1}^{N1} P(i)) \; (P(i+\lambda) - \frac{1}{N1} \sum_{i=1}^{N1} P(i))}{\left[\sum_{i=1}^{N1-\lambda} (P(i) - \frac{1}{N1} \sum_{i=1}^{N1} P(i))^2 \sum_{J=1+\lambda}^{N1} (P(j))^2 - \frac{1}{N1} \sum_{i=1}^{N1} P(i) \right]^{\frac{1}{2}}} \qquad \ldots (6.22)$$

6.3 Illustration :

Fig 6.3.1a shows the D.O. level at the down stream end of 4.7 km stretch of the River Can in eastern England, collected over 80 days period during the summer of 1972. Fig 6.3.2 shows the power density spectra versus cycle per month characteristics of the observed data. It is evident that the fundamental component of the periodicity, is 0.03125 cycle per day and the harmonics are negligible.

From the correlation co-efficient versus shift of instances of time as shown in Fig 6.3.3 it is obvious that the process follows a maximum prehistory interval of 5 times instances in days of which the most significant prehistory interval N = 2.

A selected number of polynomials were tested and the polynomial of optimum complexity was selected as

$$Y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + a_4 x_3 + a_5 x_1 x_3 + a_6 x_2 x_3 + a_7 x_4 + a_8 x_1 x_4 + a_9 x_2 x_4 + a \ldots\ldots \quad (6.23)$$

In finite difference form with periodicity the equation (6.23) can be expressed as

$$\begin{aligned} P(k) = {} & 0.229857 + 0.892091\, P(k-1) + 0.344082\, P(k-2) - 1.751167\, P(k-1) \\ & P(k-2) + 0.078086 \times \sin 2\pi (0.03125)\, k \\ & - 0.039061\, P(k-1) \sin 2\pi (0.03125)\, k + 0.290350\, p(k-2) \\ & \sin 2\pi (0.03125)\, k + 0.007663 \cos 2\pi (0.03125)\, k \\ & - 0.021738\, P(k-1) \cos 2\pi (0.03125)\, k \\ & - 0.001492\, P(k-2) \cos 2\pi (0.03125)\, k \end{aligned} \quad \ldots (6.24)$$

Table 6.2.1

Gradually increasing complexity of polynomials of eight variables.

General from of Polynomial :

$$y=a_0+a_1x_1+a_2x_2+a_3x_1x_2+a_4x_3+a_5x_1x_3+a_6x_2x_3+a_7x_4+a_8x_1x_4+a_9x_2x_4+a_{10}x_3x_4+a_{11}x_5+a_{12}x_1x_5+a_{13}x_2x_5+$$
$$+a_{14}x_3x_5+a_{15}x_4x_5+a_{16}x_6+a_{17}x_1x_6+a_{18}x_2x_6+a_{19}x_3x_6+a_{20}x_4x_6+a_{21}x_5x_6+a_{22}x_7+a_{23}x_1x_7+a_{24}x_2x_7+$$
$$+a_{25}x_3x_7+a_{25}x_4x_7+a_{27}x_5x_7+a_{28}x_6x_7+a_{29}x_8+a_{30}x_1x_8+a_{31}x_2x_8+a_{32}x_3x_8+a_{33}x_4x_3+a_{34}x_5x_8+a_{35}x_6x_8+a_{36}x_7x_8.$$

x_1	x_2	x_1x_2	x_3	x_1x_3	...	x_3x_4	...	x_7x_8
$y_1=a_0+a_1x_1$	$y_2=a_0+a_1x_2$	$y_4=a_0+a_1x_1x_2$	$y_8=a_0+a_1x_3$	$y_{16}=a_0+$	...	$y_{512}=$	...	.....
	$y_3=a_1+a_1x_1$	$y_5=a_0+a_1x_1+$	$y_9=a_0+a_1x_1$	$a_1x_1x_3$		$a_0+a_1x_3x_4$		
	$+a_2x_2$	$+a_2x_1x_2$	$+a_2x_3$	$a_{17}=a_0+\cdots$		$y_{513}=a_0+a_1x_1$		
		$y_6=a_0+a_1x_2+$				$+a_2x_3x_4$		
		$+a_2x_1x_2$	$y_{15}=a_0+a_1x_1$					
		$y_7=a_0+a_1x_1$	$+a_2x_2+a_3x_1x_2$			$y_{1023}=a_0+a_1x_1+$		$y_{36}=a_0+$
		$+a_2x_2+a_3x_1x_2$	$+a_4x_3$			$+a_2x_2+a_3x_1x_2+$		$(2-1)$
						$a_4x_3+a_5x_1x_3+$		$+a_1x_1+a_2x_2+$
						$\cdots+a_{10}x_3x_4$		$+a_3x_1x_2+\cdots$
								$+a_{36}x_7x_8$

The General rule of total number of combination is 2^n-1 where n is the total number of terms containing the variables in the general form of the polynomial.

The model has integral square error as 6.6325 E - 03 and mean error as 2.8280 E - 03. Fig 6.3.3 shows the error of prediction. Fig 6.3.4 shows the correlation co-efficient versus shift of instances of time of the errors of prediction. It is observed that errors are almost uncorrelated for $\lambda \neq 0$.

6.4 Conclusion :

The forecasting model has been found to simulate adequately the major variations of the dissolved oxygen levels observed in the measured data.

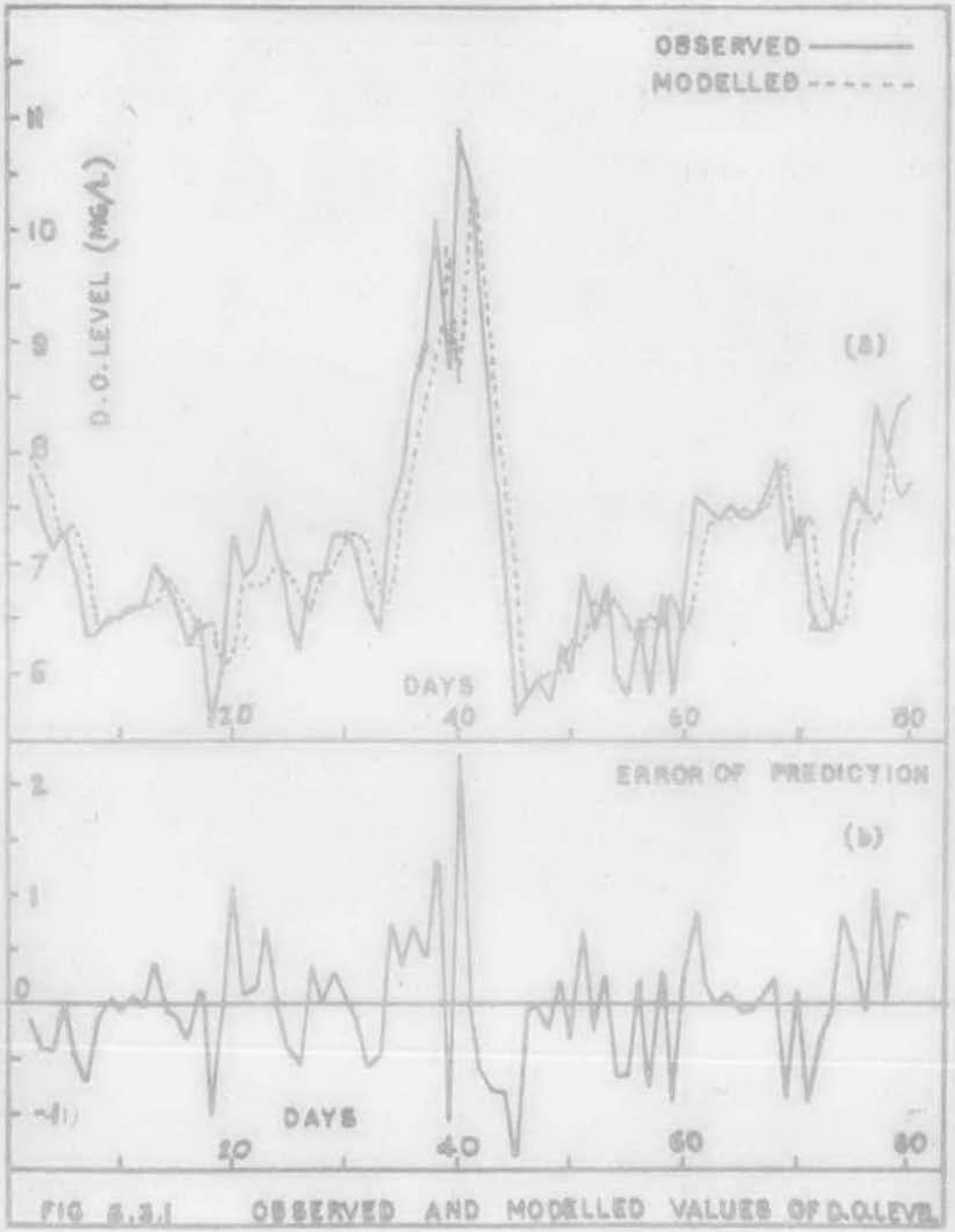
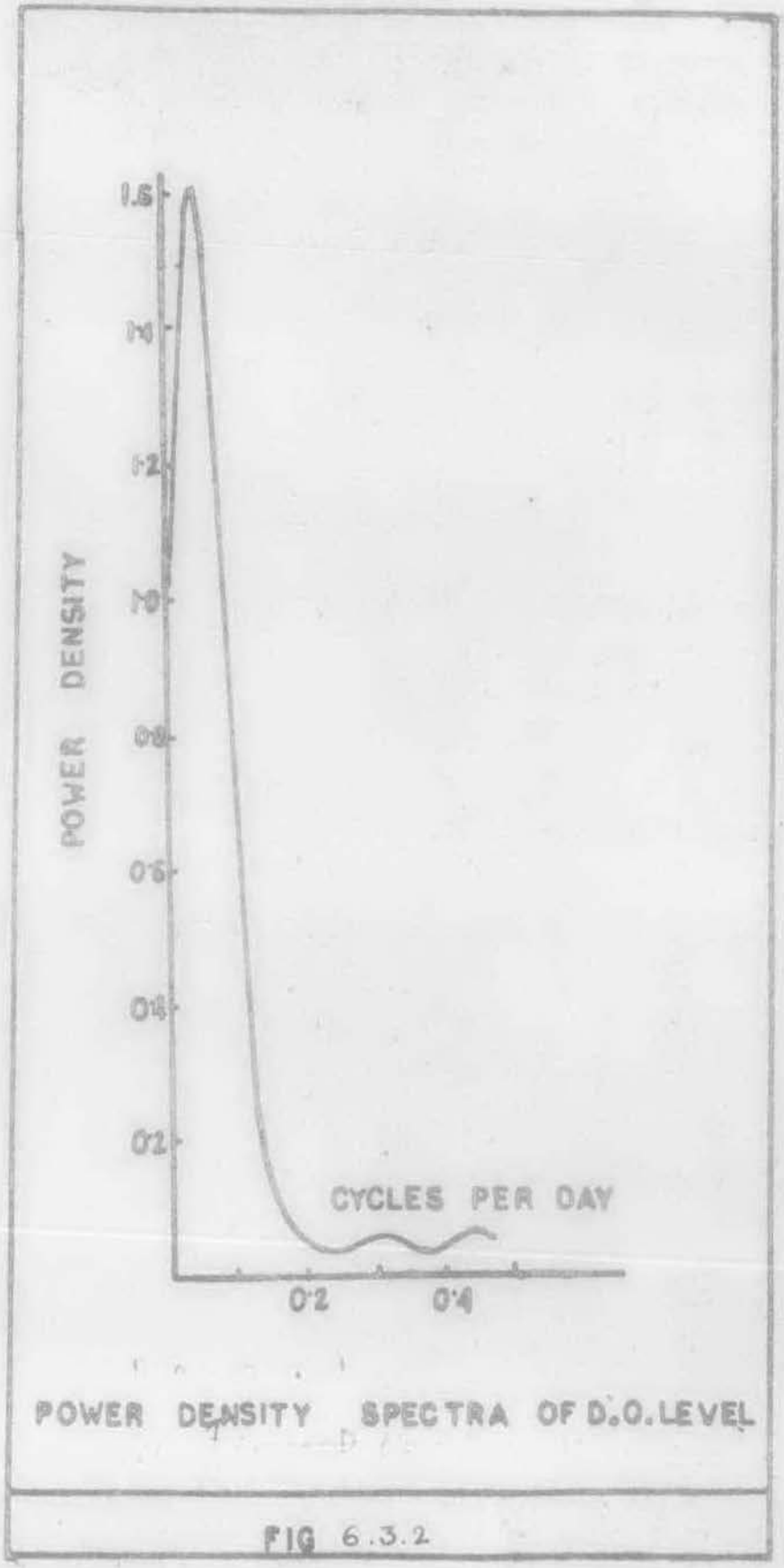


FIG 6.3.1 OBSERVED AND MODELLED VALUES OF D.O. LEVEL



POWER DENSITY SPECTRA OF D.O. LEVEL

FIG 6.3.2

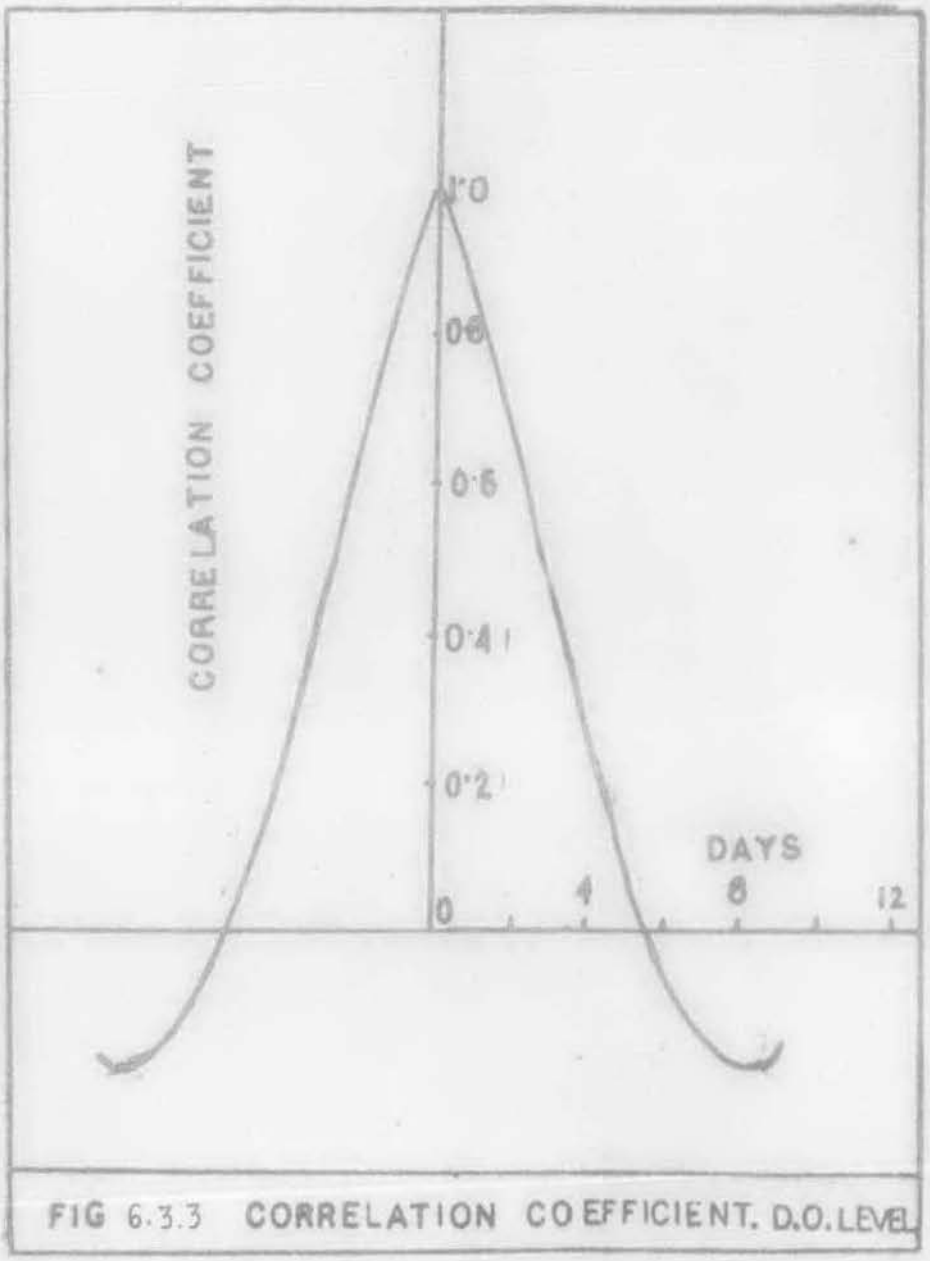


FIG 6.3.3 CORRELATION COEFFICIENT. D.O. LEVEL

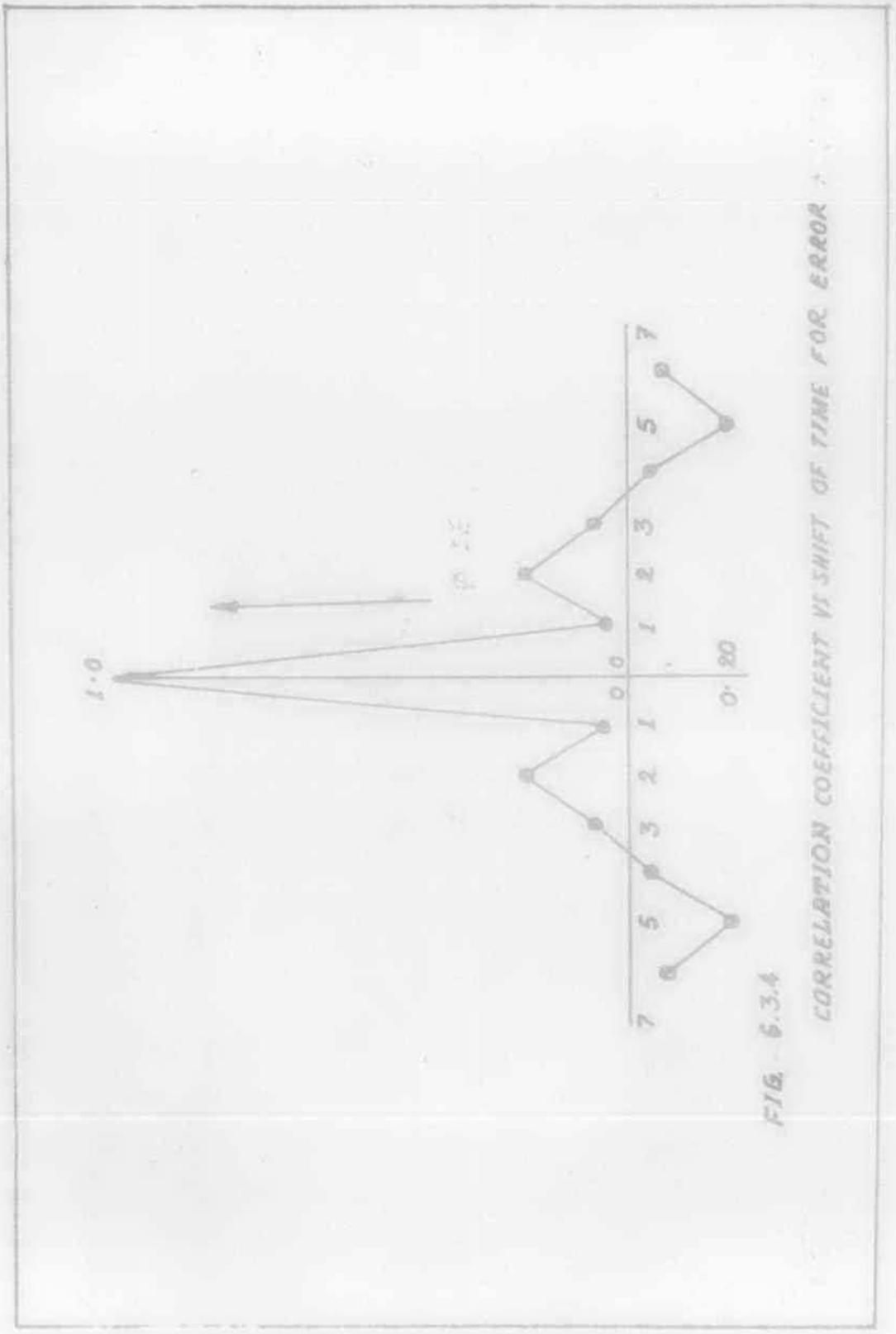


FIG. 6.3.4

CORRELATION COEFFICIENT VS SHIFT OF TIME FOR ERROR

CHAPTER VII

MODELLING OF BIOCHEMICAL OXYGEN DEMAND (B.O.D) OF A NON TIDAL RIVER BY A LEARNING ALGORITHM AND RECURSIVE INSTRUMENT VARIABLE TECHNIQUE.

CHAPTER VII

" MODELLING OF BIOCHEMICAL OXYGEN DEMAND (B.O.D.) OF A NON TIDAL RIVER BY A LEARNING ALGORITHM AND RECURSIVE INSTRUMENT VARIABLE TECHNIQUE"

7.0 Abstract

The combinatorial group method of heuristic learning identification algorithm and recursive instrument variable technique are used to obtain realistic mathematical description of biochemical oxygen demand of a non-tidal river flowing through a highly industrialised region. The simple lumped parameter dynamic models developed in this report highlight the potential possibilities of implementation of automatic control mechanism at the effluent discharge points to regulate water quality in accordance with a predetermined decision methodology.

When tested with field data the models are found to simulate adequately the major variations of biochemical oxygen demand observed in the field measurements.

7.1 Introduction.

This work offers scope for useful applications of group method of data handling algorithm as well as recursive instrument variable technique for modelling of biochemical oxygen demand of a non-tidal river. By an intelligent regulation of waste discharges it has been possible to control water quality in a natural body of water such as river. The successful control of water quality relates to the determination of a consistent level of waste treatment

facility. The mathematical model for B.O.D. level of a water body reflects the response of the treated or untreated waste discharge in a specified length of the river. The mathematical models of biochemical oxygen demand of a natural water body helps in the selection of size and location of waste treatment plant and in the determination of cost benefit effects of existing or proposed installation of waste water plants.

There is a continuous and increasing awareness about the water quality in a river system, particularly, when the river passes through a thickly populated and highly industrialised region and the river is subjected to effluent discharges. It is required that the pollution levels of the water body should be maintained within reasonable limits consistent with the needs of community and also for maintaining an overall ecological balance of the aquatic ecosystem.

The quality of water in a river depends on many factors — a few of these are — the levels of dissolved oxygen and biochemical oxygen demand, the presence of nitrates and chlorides, phosphates, the presence of bacteria and so on. However, the dissolved oxygen levels and biochemical oxygen demand (D.O. and B.O.D) are considered two principal measures of water quality which indicate the general health of the aquatic environment of a river.

The dissolved oxygen level is expressed in mg/litre and can be measured by an in-stream probe. The biochemical oxygen demand which is, in fact, an aggregate of the capacities of absorption of oxygen by materials present in river water, is usually expressed in mg/litre and can be measured from a sample of river water. The sample is to be kept in a controlled environment for a period of about 5 days.

The important engineering decisions in water quality control relate to the determination, regulation and maintenance of the levels of dissolved oxygen and biochemical oxygen demand those are consistent with the multiple uses of natural water bodies. The D.O.

and B.O.D concentrations are important aspects of water quality of rivers since the temporary fluctuations of the levels of the dissolved oxygen and biochemical oxygen demand can cause long-term and possibly irreversible damage to the ecosystem.

In Chapter (6) author has suggested method for modelling of dissolved oxygen levels of non tidal rivers. The model is found to simulate adequately the major variations of dissolved oxygen levels as observed in the field measurements. The present work introduces some of the techniques and principles of modern control theory broadly classified as cybernetics and indicate how these can be used in modelling of biochemical oxygen demand of non tidal rivers.

It is believed that models of biochemical oxygen demand developed in this paper would subsequently help in the formulation of methodologies for decision regularisation to effect an optimal water quality control mechanism. In the present work dynamic models of the levels of B.O.D for a length of a river have been obtained through two different techniques of cybernetics; one, the combinatorial group method of heuristic learning identification algorithm and the other, the recursive instrument variable technique. The recursive instrument variable technique is amenable to on line computation provided adequate real time data are available in time. The models have been verified by simulation against field measurement collected over a 79 day period in summer in 1972 from the river Can in Eastern England.

7.2. Modelling of B.O.D by Combinatorial Group Method of Heuristic Learning Identification Algorithm.

The combinatorial group method of heuristic learning identification algorithm can be classified as the theory of self organisation of mathematical models. With the help of this method it has been possible to formulate mathematical models of complex processes.

The concept of self organisation can be illustrated as follows :

When the model complexity gradually increases the computer finds by shifting the different models, the minimum of a selection criterion which the computer has been ordered to look for. Thus the computer indicates to the operator the model of optimum complexity.

The purpose of this part of the work is to obtain a mathematical description of daily biochemical oxygen demand of a non tidal river by the combinatorial group method of heuristic learning identification algorithm.

Model Development :

The observed data are processed according to the method already developed in Chapter (6). However, for the sake of completeness we are reproducing the relevant portion.

The daily measured data for B.O.D. concentration a river are

$P(i)$, $i = 1, 2 \ldots, N1$, i being the time instants in days

The mean value is given by

$$\bar{P}(N1) = \frac{1}{N1} \sum_{i=1}^{N1} P(i) \qquad \ldots \quad (7.1)$$

Auto covariance of the data at lag instant k is given by

$$CA(k) = \frac{1}{N1-k} \sum_{i=1}^{N1-k} \left[P(i) - \bar{P}(N1) \right] \left[P(i+k) - \bar{P}(N1) \right] \qquad \ldots \quad (7.2)$$

where $k = 0, 1, 2, \ldots, M$ and $M < \frac{1}{4} N1$

The normalised co-efficient of covariance are given by

$$RA(k) = \frac{CA(k)}{CA(o)} \qquad \ldots \quad (7.3)$$

The estimates of the normalised power density spectra for the data are given by,

$$PS(w_h) = \frac{2}{\Pi} \sum_{k=0}^{M} E_k \, RA(k) \cos w_h k \quad \ldots \quad (7.4)$$

where $w_h = 2 \Pi f_h$ and $f_h = \frac{h}{2M}$; $0 \leq f_h \leq 0.5$

and $h = 0, 1, 2, \ldots, M$

E_k has been defined as the weight for window correction and may be taken as

$$E_k = \begin{cases} 1.0 & \text{for } 0 < k < M \\ 0.5 & k = 0, M \end{cases} \quad \ldots \quad (7.5)$$

These raw estimates of power spectral density are smoothed by using Hanning Window to obtain the final estimates of the power spectrum. The smoothed estimates of the ordinates of the power spectrum are given by

$$S(w_h) = 0.54 \, PS(w_h) + 0.46 \, PS(w_h + 1) \text{ for } h = 0$$

$$S(w_h) = 0.23 \, PS(w_{h-1}) + 0.54 \, PS(w_h) + 0.23 \, PS(w_{h+1}) \text{ for } 0 < h < M$$

$$S(w_h) = 0.54 \, PS(w_h) + 0.46 \, PS(w_{h-1}) \text{ for } h = M \quad \ldots \quad (7.6)$$

The periodicity in terms of fundamental and its harmonics can be estimated from the power spectral density versus frequency characteristics. We are developing a generalised procedure of analysis. For some physical systems periodicity may be totally absent. In that case power density spectra monotonically decreases with the increase of frequency.

The non-stationary process model with the deterministic periodicity can be expressed as

$$P(k) = U(k) + V(k) \quad \ldots \quad (7.7)$$

Where U (k) is the deterministic trend component,

V (k) is the stationary stochastic component

k = 1, 2, ..., are time instances.

The deterministic periodic components can be represented as

$$U(k) = \sum_{r=0}^{R} (a_r \cos (r\ 2\ \Pi\ f_o k) + b_r \sin (r\ 2\Pi f_o k)) \quad \ldots \quad (7.8)$$

where f_o is the fundamental cycle per instance of time. The stochastic components V(k) can be expressed in autoregressive form as

$$V(k) = \sum_{n=1}^{N} C_{k-n} V_{(k-n)} + \eta(k) \quad \ldots \quad (7.9)$$

where n = 1, 2, ... N, the prehistory interval and $\eta(k)$ the error sequence. The error sequence $\eta(k)$ modelled in the moving average form as,

$$\eta(k) = \sum_{j=1}^{n} a_{k-j} \eta(k-j) + \upsilon(k) \quad \ldots \quad (7.10)$$

where n is the prehistory interval of the moving average terms of $\eta(k)$. $\upsilon(k)$ resembles to a white noise process with zero mean and constant variance.

V(k) can be written as,

$$V(k-n) = P(k-n) - U(k-n) \quad \ldots \quad (7.11)$$

Hence the equation (7.9) can be written as

$$V(k) = \sum_{n=1}^{N} C_{k-n} P(k-n) - \sum_{n=1}^{N} C_{k-n} U(k-n) + \eta(k) \quad \ldots \quad (7.12)$$

From equations (7.9) and (7.8) it follows that

$$P(k) = \sum_{n=1}^{N} C_{k-n} P(k-n) + \sum_{n=0}^{N} \sum_{r=0}^{R} \left[d^{r}_{k-n} \cos (r2\Pi f_o(k-n)) + g^{r}_{k-n} \sin (r2\Pi f_o(k-n)) \right] + \gamma(k) \quad \ldots (7.13)$$

In functional form

$$P(k) = f(P(k-1), P(k-2), \ldots, P(k-N) \sin (2\Pi f_o k), \cos (2\Pi f_o k), \sin (2\Pi f_o(k-1)), \cos (2\Pi f_o(k-1)), \ldots, \sin (2\Pi f_o(k-N), \cos (2\Pi f_o(K-N)), \ldots, \sin (R2\Pi f_o(K-N), \cos (R2\Pi f_o(K-N))) \quad \ldots (7.14)$$

Let $P(k) = y$; $P(k-1) = x_1, \ldots, P(k-N) = x_N$

$\sin (2\Pi f_o k) = x_{N+1}$; $\cos (2\Pi f_o k) = x_{N+2}, \ldots,$

$\cos (R2\Pi f_o(k-n)) = x_{3N+2}$ and $\gamma(k)$ as residual ξ

The equation (7.14) then becomes

$$y = f (x_1, x_2, \ldots, x_{3N+2}) + \xi \quad \ldots (7.15)$$

The estimate of y as $\hat{y}$ is defined as

$$\hat{y} = f (x_1, x_2, \ldots, x_{3N+2}) \quad \ldots (7.16)$$

The function $f(.)$ in equation (7.16) is sought in a class of quadratic polynomials of gradually increasing complexity on the basis of minima of a selection criterion, namely, the integral square error criterion.

Polynomials of gradually increasing complexity of five variables has been shown in Table 2.5.1.

The selection criterion, the integral square error is defined as,

$$I.S.E. = \frac{\sum_{i=1}^{N1} (y_o(i) - y_m(i))^2}{\sum_{i=1}^{N1} (y_o(i))^2} \qquad \ldots \quad (7.17)$$

where $y_o(i)$, $i = 1, 2, \ldots,$ N1 days, are the observed values of the output variable in the interpolation region and $y_m(i)$ are the values of the variables obtained from the model. For the model of optimum complexity the residual has almost the property of a random variable with correlation co-efficient

$$\phi_{\xi\xi}(\lambda) = 1.0 \quad \text{for} \quad \lambda = 0$$

and

$$\phi_{\xi\xi}(\lambda) = 0.0 \quad \text{for} \quad \lambda \neq 0$$

where λ is the shift of instances of time.

The correlation co-efficient is defined as

$$\frac{\sum_{i=1}^{N1-\lambda} \left(\xi(i) - \frac{1}{N1}\sum_{i=1}^{N1}\xi(i)\right)\left(\xi(i+\lambda) - \frac{1}{N1}\sum_{i=1}^{N1}\xi(i)\right)}{\left[\sum_{i=1}^{N1-\lambda}\left(\xi(i) - \frac{1}{N1}\sum_{i=1}^{N1}\xi(i)\right)^2 \sum_{i=1+\lambda}^{N1}\left(\xi(i) - \frac{1}{N1}\sum_{i=1}^{N1}\xi(i)\right)^2\right]^{\frac{1}{2}}} \qquad \ldots \quad (7.18)$$

7.3 Illustration

Fig 7.3.1a shows the levels of biochemical oxygen demand at the down stream end of 4.7 Km stretch of the river Cam in Eastern England collected over 80 days period during the summer of 1972. Fig 7.3.2 shows the power density spectra versus cycle per day characteristics of the observed data. It is evident from the

monotonically decreasing power density spectra when frequency increases that that process of biochemical oxygen demand does not show any periodicity. From the correlation co-efficient versus shift of instances time in days as shown in Fig 7.3.3. It is obvious that the process follows a maximum prehistory interval of 8 time instances in days of which the significant prehistory interval N = 5. A selected number of polynomials were tested and the polynomial of optimum complexity was obtained as

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4x_3 + a_5x_1x_3 + a_6x_2x_3 + a_7x_4 + a_8x_1x_4 + a_9x_2x_4 + a_{10}x_3x_4 + a_{11}x_5 + a_{12}x_1x_5 + a_{13}x_2x_5 + a_{14}x_3x_5 \quad \ldots (7.19)$$

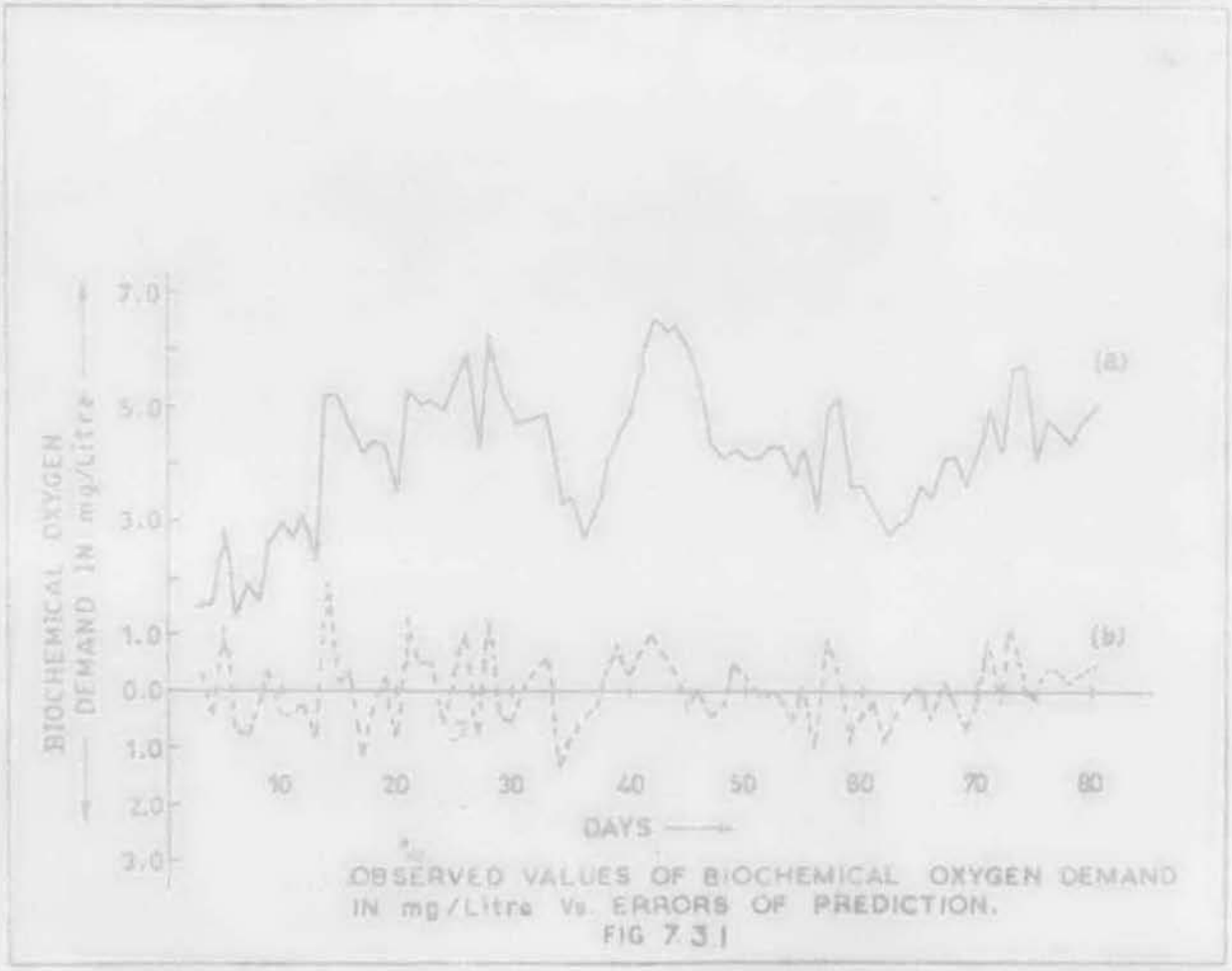
In finite difference form the equation (7.19) can be expressed as

$$\begin{aligned}\hat{P}(k) = {} & -0.148860 + 0.810478\,P(k-1) + 0.899741\,P(k-2) \\ & - 0.628953\,P(k-1)\,P(k-2) - 0.365301\,P(k-3) \\ & + 0.777806\,P(k-1)\,P(k-3) - 0.224127\,P(k-2)\,P(k-3) \\ & + 0.448458\,P(k-4) - 0.344826\,P(k-1)\,P(k-4) \\ & + 0.022652\,P(k-2)\,P(k-4) - 0.152537\,P(k-3)\,P(k-4) \\ & + 0.485092\,P(k-5) - 0.226081\,P(k-1)\,P(k-5) \\ & - 0.108033\,P(k-2)\,P(k-5) - 0.001700\,P(k-3)\,P(k-5) \quad \ldots \quad (7.20)\end{aligned}$$

The model has an integral error as 0.022332 and mean error as $-5.9086172 - 0.4$. Fig 7.3.1(b) shows the error of prediction. Fig 7.3.4 shows the correlation co-efficient versus shift of instances of time of the errors of prediction. It is observed that the errors of prediction are almost uncorrelated for $\lambda \neq 0$.

7.4 Modelling of B.O.D. by recursive Instrument Variable Algorithm.

In recursive estimation algorithm the estimate of the output at a future instants is constantly updated on receipt of fresh



OBSERVED VALUES OF BIOCHEMICAL OXYGEN DEMAND IN mg/Litre Vs. ERRORS OF PREDICTION.

FIG 7.3.1

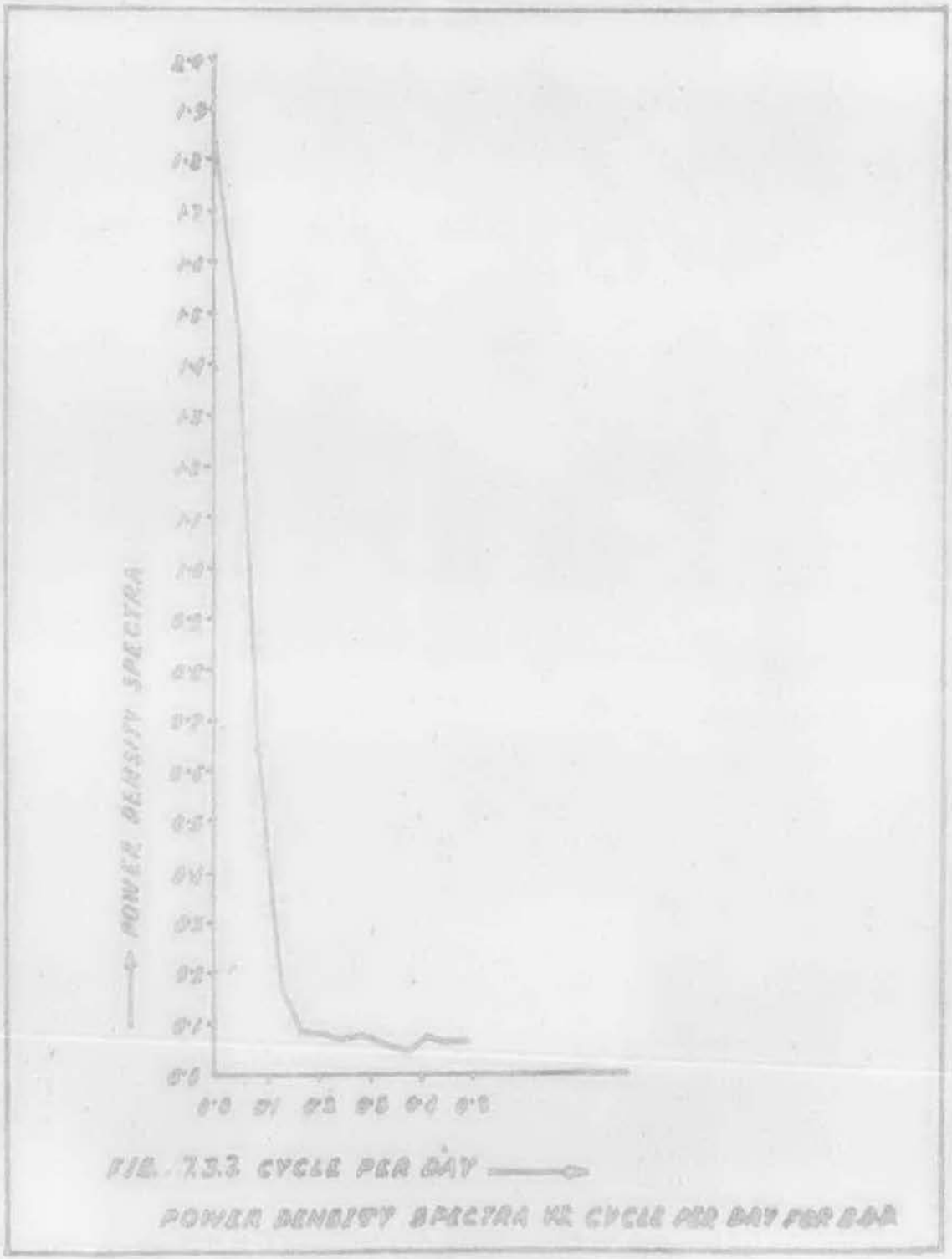


FIG. 7.3.3 POWER DENSITY SPECTRA VS CYCLE PER DAY FOR BAR

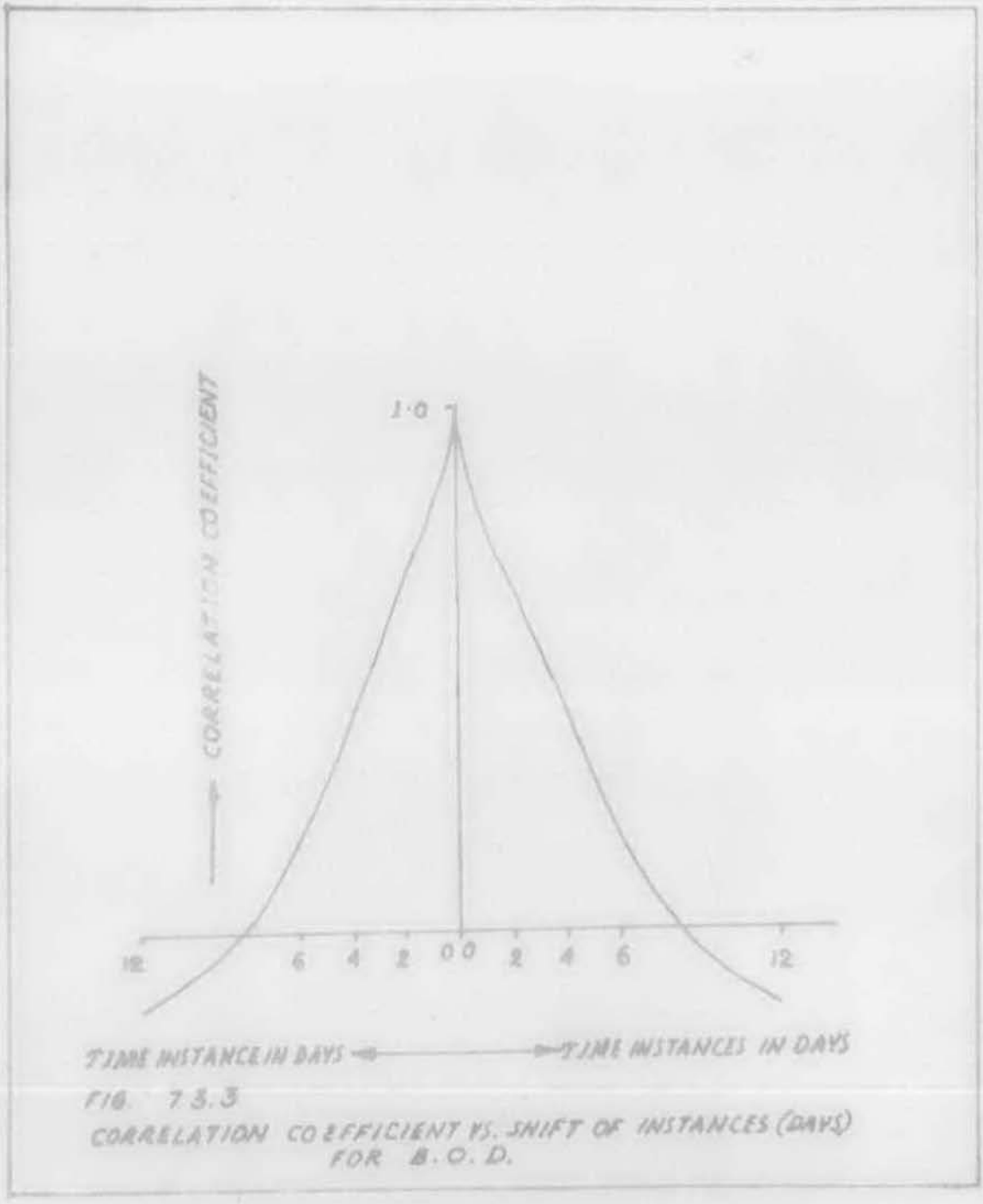


FIG. 7.3.3
CORRELATION COEFFICIENT VS. SHIFT OF INSTANCES (DAYS) FOR B.O.D.

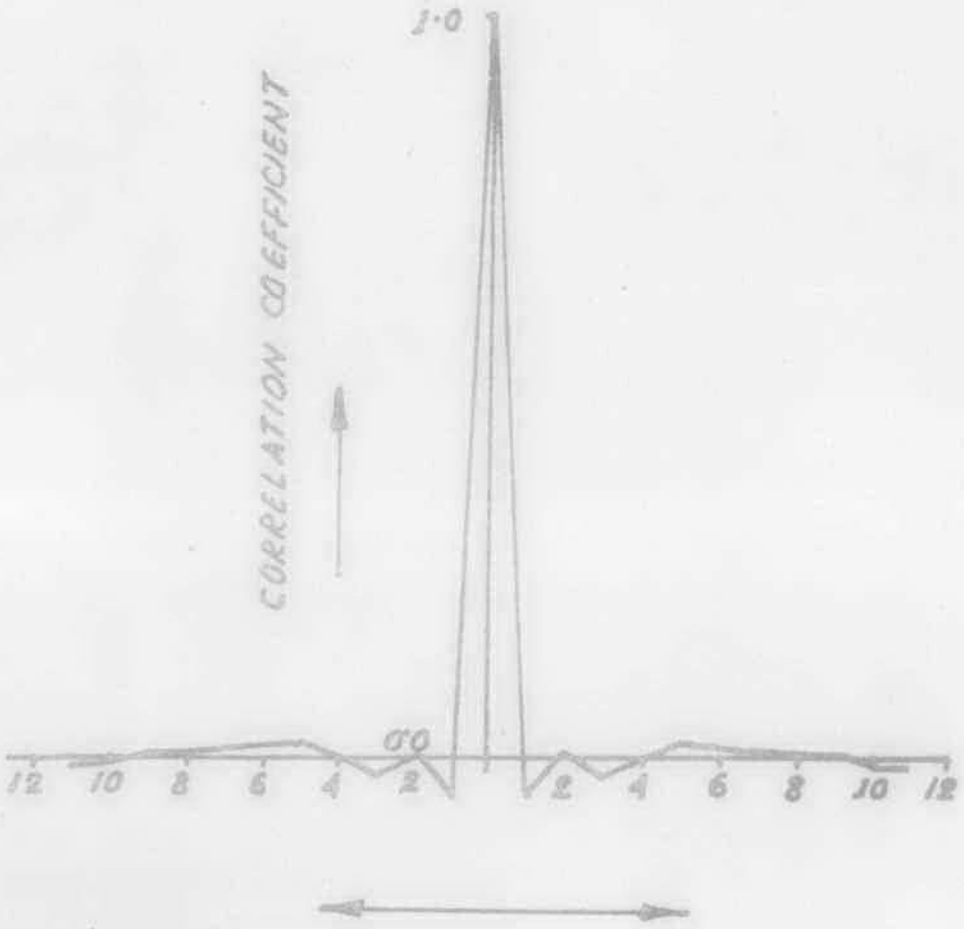


FIG. 7.3.4

CORRELATION CO-EFFICIENT VS. SHIFT OF INSTANCES OF TIME IN DAYS FOR ERRORS OF PREDICTION FOR BIOCHEMICAL OXYGEN DEMAND

information. The presence of significance co-efficient of correlation with the levels of biochemical oxygen demand at past instants with the current instant clearly suggests that the process may be represented in the form of a stationary time series with a probability that the current value of the output P(k) is a function of the previous output observations, the auto-regressive terms P(k-2), ..., together with the current unknown realisation of the noise process. The process thus, can be represented as

$$P(k) = \left[\sum_{i=1}^{N} \beta_i q^{-1} \right] P(k) + \eta(k) \qquad \ldots(7.21)$$

where q^{-1} is the backward shift operator defined as $q^{-1}\beta(k) = P(k-1)$. The determination of N is known as the model order determination which is also the model order identification. The number of the significant instants in the output correlation may be intuitively considered for model order determination.

$\eta(k)$ may be expressed in a moving average sequence as

$$\eta(k) = \sum_{p=1}^{P} \gamma_p \, \eta(k-p) + \vartheta(k) \qquad \ldots (7.22)$$

where $\vartheta(.)$ is a white noise innovation process with

$$E\{\vartheta(k) \; \vartheta(j)\} = 0 \text{ for } j \neq k \text{ and}$$
$$E\{\vartheta(k) \; \vartheta(j)\} = \sigma^2 \text{ for } j = k \qquad \ldots (7.23)$$

The model is quite flexible since it requires that the equations be linear in parameters.

The equation (7.21) can be written as

$$P(k) = z^T(k)\alpha + \eta(k) \qquad \ldots (7.24)$$

where ' α ' is the parameter vector with property of slowly varying with time and amenable to recursive adaptiveness, and

$$z^T(k) = [\, P(k-1), \; P(k-2), \ldots, \; P(k-N) \,]$$

The least square estimation of the process parameters can be obtained by minimizing a loss function defined as the sum of the squared errors as

$$J = \sum_{k=1}^{N1} (P(k) - z^T(k) \hat{\alpha})^2 \qquad \ldots \quad (7.25)$$

The estimates $\hat{\alpha}$ of α that minimise J are called least square estimates and $\hat{\alpha}$ is defined as

$$\hat{\alpha} = \left[\sum_{k=1}^{N1} z(k) \; z^T(k) \right]^{-1} \left[\sum_{k=1}^{N1} z(k) P(k) \right] \qquad \ldots \quad (7.26)$$

Let

$$R^*(k) \triangleq \left[\sum_{j=1}^{k} z(j) \; z^T(j) \right]^{-1}$$

and

$$S(k) \triangleq \left[\sum_{j=1}^{k} z(j) \, P(j) \right]$$

So equation (7.26) becomes,

$$\hat{\alpha}(k) = R^*(k) \; S(k) \qquad \ldots \quad (7.27)$$

The recursive relationship for $R^*(.)$ and $S(.)$ can be set as

$$\left[R^*(k) \right]^{-1} \sum_{j=1}^{k-1} z(j) z^T(j) + z(k) \; z^T(k)$$

$$= \left[R^*(k-1) \right]^{-1} + z(k) \; z^T(k) \qquad \ldots \quad (7.28)$$

Similarly,

$$S(k) = S(k-1) + z(k) \; P(k) \qquad \ldots \quad (7.29)$$

Premultiplying by $R^*(k)$ we get from equation (7.28)

$$I = R^*(k) \left[R^*(k-1) \right]^{-1} + R^*(k) x(k) x^T(k) \quad \ldots \quad (7.30)$$

Postmultiplication of equation (7.30) by $R^*(k-1)$

$$R^*(k-1) = R^*(k) + R^*(k)\, x(k)\, x^T(k)\, R^*(k-1)$$

$$R^*(k-1) = R^*(k) + R^*(k)\, x(k)\, x^T(k)\, R^*(k-1) \quad \ldots \quad (7.31)$$

Post multiplying equation (7.31) by $x(k)$

$$R^*(k-1)\, x(k) = R^*(k)\, x(k) + R^*(k)\, x(k) x^T(k)\, R^*(k-1) x(k)$$

So we get,

$$R^*(k-1)\, x(k) = R^*(k) x(k) \left[1 + x^T(k)\, R^*(k-1)\, x(k) \right] \quad \ldots \quad (7.32)$$

Post multiplication by $\left[1 + x^T(k)\, R^*(k-1) x(k) \right]^{-1} x^T(k) R^*(k-1)$

the equation (7.32) gives,

$$R^*(k-1)\, x(k) \left[1 + x^T(k) R^*(k-1)\, x(k) \right] x^T(k) R^*(k-1) = R^*(k) x(k) x^T(k) R^*(k-1)$$

$$\ldots \quad (7.33)$$

Substitution of equation (7.31) in equation (7.33)

$$R^*(k) = R^*(k-1) - R^*(k-1) x(k) \left[1 + x^T(k)\, R^*(k-1)\, x(k) \right]^{-1} \times x^T(k)\, R^*(k-1) \quad \ldots \quad (7.34)$$

From equation (7.27) denoting the estimate α at time instant k as $\hat{\alpha}(k)$

$$\hat{\alpha}(k) = R^*(k) S(k)$$

Combining equations (7.29) and (7.34)

$$\hat{\alpha}(k) = \left[R^*(k-1) - R^*(k-1)\, x(k) \left[1 + x^T(k)\, R^*(k-1) x(k) \right]^{-1} \times x^T(k)\, R^*(k-1) \right] \left\{ S(k-1) + x(k)\, P(k) \right\} \quad \ldots \quad (7.35)$$

Since $R^*(k-1)\, S(k-1) = \hat{\alpha}(k-1)$

The equation (7.35) can be written as

$$\hat{\alpha}(k) = \hat{\alpha}(k-1) - R^*(k-1)\, z(k) \left[1+z^T(k) R^*(k-1) z(k)\right]^{-1} \times z^T(k) \hat{\alpha}(k-1) + R^*(k-1)\, z(k) P(k) - R^*(k-1)\, z(k) \left[1+z^T(k) R^*(k-1) z(k)\right]^{-1} \times \left\{ z^T(k) R^*(k-1) z(k) P(k) \right\} \quad \ldots (7.36)$$

Since

$z^T(k) R^*(k-1) z(k)$ is scalar the equation (7.36) becomes

$$\hat{\alpha}(k) = \hat{\alpha}(k-1) + R^*(k-1) z(k) \left[1+z^T(k) R^*(k-1) z(k)\right]^{-1} \times \left[P(k) - z^T(k) \hat{\alpha}(k-1) \right] \quad \ldots (7.37)$$

The equations (7.37), (7.34) with

$$R^*(k) \triangleq \left[\sum_{j=1}^{k} z(j)\, z^T(j) \right]^{-1}$$

are the recursive least square algorithms. But these algorithms do not overcome the problem of bias. The technique to minimise the bias is stated as follows. From equation (7.26) the non recursive estimation of parameters at instant k is

$$\hat{\alpha}(k) \left[\sum_{j=1}^{k} z(j) z^T(j) \right]^{-1} \left[\sum_{j=1}^{k} z(j) P(j) \right] \quad \ldots (7.38)$$

and the process equation is,

$$P(j) = z^T(j) \alpha + \eta(j) \quad \ldots (7.39)$$

Upon substitution of equation (7.39) in equation (7.38)

$$\hat{\alpha}(k) = \left[\sum_{j=1}^{k} z(j) z^T(j) \right]^{-1} \left[\sum_{j=1}^{k} z(j) \left[z^T(j) \hat{\alpha} + \eta(j) \right] \right]$$

$$= \left[\sum_{j=1}^{k} x(j)x^T(j) \right]^{-1} \left[\sum_{j=1}^{k} x(j)\, x^T(j) \right] \alpha$$

$$+ \left[\sum_{j=1}^{k} x(j)x^T(j) \right]^{-1} \left[\sum_{j=1}^{k} x(j)\ \gamma(j) \right] \quad \ldots (7.40)$$

So,

$$\hat{\alpha}(k) = \alpha + \left[\sum_{j=1}^{k} x(j)x^T(j) \right]^{-1} x \left[\sum_{j=1}^{k} x(j)\ \gamma(j) \right]$$

For the estimation of $\hat{\alpha}(k)$ to converge to α

$$E\left\{ x(j)\ \gamma(j) \right\} = 0 \quad \text{for all } j$$

and

$$E\left\{ \gamma(j)\ \gamma(k) \right\} = 0 \quad \text{for } j \neq k$$

i.e. $\gamma(j)$ converges to white noise process $v(j)$.
Under these conditions the estimates will be unbiased. However, these conditions are difficult to achieve in practice. And consequently the vector $x(k)$ has been modified incorporating the instrument variable $\hat{P}(.)$ as $\hat{x}(k)$ defined as

$$\hat{x}(k) = \left[\hat{P}(k-1),\ \hat{P}(k-2),\ \ldots,\ \hat{P}(k-N) \right]^T \quad \ldots (7.41)$$

where $\hat{P}(.)$ is the recursive least square estimation of $P(.)$. Since $\hat{P}(.)$ sequence in equation (7.41) are already estimates, in a sense optimum, they are almost uncorrelated with $\gamma(k)$. Thus the conditions of unbiased estimates are modified as

$$E\left\{ \hat{x}(k)\ \gamma(k) \right\} = 0 \quad \text{for all } k.$$

Replacing $x(k)$ by $\hat{x}(k)$ rather heuristically and not $x^T(k)$ by $\hat{x}^T(k)$, because of experience in the interpolation region, in equations (7.34) and (7.37) the recursive instrument variable algorithms become,

$$\hat{\alpha}(k) = \hat{\alpha}(k-1) + \hat{R}^{*}(k-1)z(k)\left[1+a^{T}(k)\,\hat{R}^{*}(k-1)\,z(k)\right]^{-1} \times \left\{P(k) - a^{T}(k)\,\hat{\alpha}(k-1)\right\} \quad \ldots (7.42a)$$

$$\hat{R}^{*}(k) = \hat{R}^{*}(k-1) - \hat{R}^{*}(k-1)\,z(k)\left[1+a^{T}(k)\,\hat{R}^{*}(k-1)\,z(k)\right]^{-1} \times a^{T}(k)\,\hat{R}^{*}(k-1) \quad \ldots (7.42b)$$

$$\hat{R}^{*}(k) = \left[\sum_{j=1}^{k} z(j)\,z(j)\right]^{-1} \quad \ldots (7.42c)$$

The algorithms may be initialised with $\hat{\alpha}(.)$ as a least square estimate for a block of data in the interpolation region and $R^{*}(.)$ for the same block of data. The instrument variables z (.) also obtained through a separate parameter tracking algorithm as detailed below.

A sequence of instrument variables P(.) are obtained from a stationary time series model of the observed data.

$$P(t_k) = V(t_k) \quad \ldots (7.43)$$

where $V(t_k)$ is a stationary stochastic component and $t_k = 1,2 \ldots,$ are the time instances $V(t_k)$ may be expressed in autoregressive from as,

$$V(t_k) = \sum_{n=1}^{N} C_{k-n}\,V(t_{k-n}) + \hat{\varepsilon}(t_k) \quad \ldots (7.44)$$

where $n = 1, 2 \ldots,$ N the prehistory interval.

N may be intuitively determined from the co-affident of correlation versus shift of instances of time.

The output error sequence $\hat{\varepsilon}(t_k)$ may be modelled as

$$\hat{\varepsilon}(t_k) = \sum_{j=1}^{m} q_{k-j}\left(P(t_{k-j}) - \hat{P}(t_{k-j})\right) \quad \ldots (7.45)$$

where $\hat{P}(t_{k-j})$ is the estimate of the output on the basis of observation upto t_{k-j-1} instant. The prehistory n of the moving average terms of the output error sequence may be determined for a particular type of process to adhere to a minimum of integral square error criterion defined as,

$$ISE = \frac{\sum_{i=1}^{N1} P(t_i) - \hat{P}(t_k))^2}{\sum_{i=1}^{N1} (P(t_i))^2} \quad \ldots \quad (7.46)$$

From equation (7.43)

$$P(t_{k-n}) = V(t_{k-n}) \quad \ldots \quad (7.47)$$

Thus,

$$P(t_k) = \sum_{n=1}^{N} C_{k-n} P(t_{k-n}) \qquad \hat{} (t_k)$$

In matrix form

$$\hat{P}(t_k) = a^T(t_{k-1})\, s(t_{k-1}) \quad \ldots \quad (7.48)$$

where $a^T(t_{k-1}) = [\, C_{k-1}, C_{k-2}, \ldots, C_{k-N} \,]$

and

$$s(t_{k-1}) = [\, P(t_{k-1}), P(t_{k-2}), \ldots, P(t_{k-N}) \,]^T$$

The co-efficient vector 'a' can be estimated by minimising the quadratic performance criterion $J_k(a)$, defined as,

$$J_k(a) = \sum_{j=1}^{k} (P(t_j) - a^T s(t_{j-1}))^2 + (a - a(t_o))^T a^{-1}(t_o)(a - a(t_o)) \quad \ldots (7.49)$$

where $a(t_o)$ is the available a priori estimate of the co-efficient vector 'a' and $S(t_o)$ is the positive definite weighting matrix of the order $m1 \times m1$, where $m1 = M1$.

For minimisation,

$$\frac{\partial J_k(a)}{\partial a} = 0$$

$$\text{or } -2 \sum_{j=1}^{k} \left[x(t_{j-1}) \{ P(t_j) - a^T x(t_{j-1}) \} \right] + 2S^{-1}(t_o)(a - a(t_o)) = 0 \quad \ldots \quad (7.50)$$

It follows from equation (7.50)

$$\sum_{j=1}^{k} x(t_{j-1}) P(t_j) + S^{-1}(t_o) a(t_o) = \sum_{j=1}^{k} x(t_{j-1}) x^T(t_{j-1}) + S^{-1}(t_o) a \quad \ldots (7.51)$$

Let

$$S^{-1}(t_k) = \sum_{j=1}^{k} x(t_{j-1}) x^T(t_{j-1}) + S^{-1}(t_o) \quad \ldots \quad (7.52)$$

and

$$d(t_k) = \sum_{j=1}^{k} x(t_{j-1}) P(t_j) + S^{-1}(t_o) a(t_o) \quad \ldots \quad (7.53)$$

Denoting the co-efficient vector 'a' as $a(t_k)$ at the time instant t_k,

$$S^{-1}(t_k)\, a(t_k) = d(t_k)$$

or

$$a(t_k) = S(t_k)\, d(t_k) \quad \ldots \quad (7.54)$$

From equations (7.52) and (7.53) the following recursive equations are obtained,

$$S^{-1}(t_{k+1}) = S^{-1}(t_k) + x(t_k) x^T(t_k) \quad \ldots \quad (7.55)$$

$$d(t_{k+1}) = d(t_k) + z(t_k)\ P(t_{k+1}) \qquad \ldots \quad (7.56)$$

By matrix inversion lemma the recursive parameter estimation algorithms to obtain the instrument variables $P(t_k)$ are,

$$a(t_{k+1}) = a(t_k) + S(t_{k+1})z(t_k)\left[P(t_{k+1}) - a^T(t_k)z(t_k)\right] \qquad \ldots \quad (7.57a)$$

$$S(t_{k+1}) = S(t_k) - S(t_k)z(t_k)z(t_k)z^T(t_k)S(t_k)\left[1+z^T(t_k)S(t_k)z(t_k)\right]^{-1}$$

$$\ldots \quad (7.57b)$$

The algorithms are initialised with

$S(t_0) \hat{=} I$ (unit matrix); $a(t_0) \hat{=} 0$

$P(t_j) = 0$ for $j = 0, -1, -2, \ldots,$

and $\hat{\delta}(t_j) = 0$ for $j = 0, -1, -2, \ldots$).

7.5 Illustration :

Fig 7.5.1(a) shows the level of daily biochemical oxygen demand of 4.7km stretch of the river Cam in Eastern England collected over 79 day period in summer of 1972.

Fig 7.5.1(b) shows the output obtained with recursive instrument variable algorithms with output prehistory interval $N = 5$ and integral square error = 0.0355 and mean error = 7.2993E-03. It is observed that the errors are almost uncorrelated [109], as is evident from Fig 7.5.2. The prehistory interval of the error sequence $\hat{\delta}(.)$ for instrument variables, n is taken as 2. It is evident that there is considerable improvement of integral square error for combinatorial group method of heuristic learning identification algorithms over recursive instrument variable technique for the same output prehistory interval. However, the recursive instrument variable technique is amenable to possible on line control.

7.6 Conclusion :

The present study relating to a realistic modelling of the biochemical oxygen demand of a non tidal river is rather preliminary and tentative in nature. However, this study particularly the use of recursive instrument variable, algorithm for modelling of B.O.D. highlights the possibility of real time implementation of automatic control of water quality in accordance with a predetermined level of B.O.D. of river water.

There is a high degree of coupling between different biochemical transformation processes in river water. The spatial variations of biochemical oxygen demand occur. This necessitates a systematic collection of water quality data with spatial distribution of collecting points. In that case it will be possible to have a distributed parameter mathematical description of the biochemical oxygen demand and it is hoped that the distributed parameter model would further increase our understanding of the complex biochemical transformation processes. Nevertheless, the simple lumped parameter dynamic models obtained in the present study appear to provide a potentially adequate description of the biochemical oxygen demand supplying practical information for implementation of operational control mechanism for water quality of a natural water body.

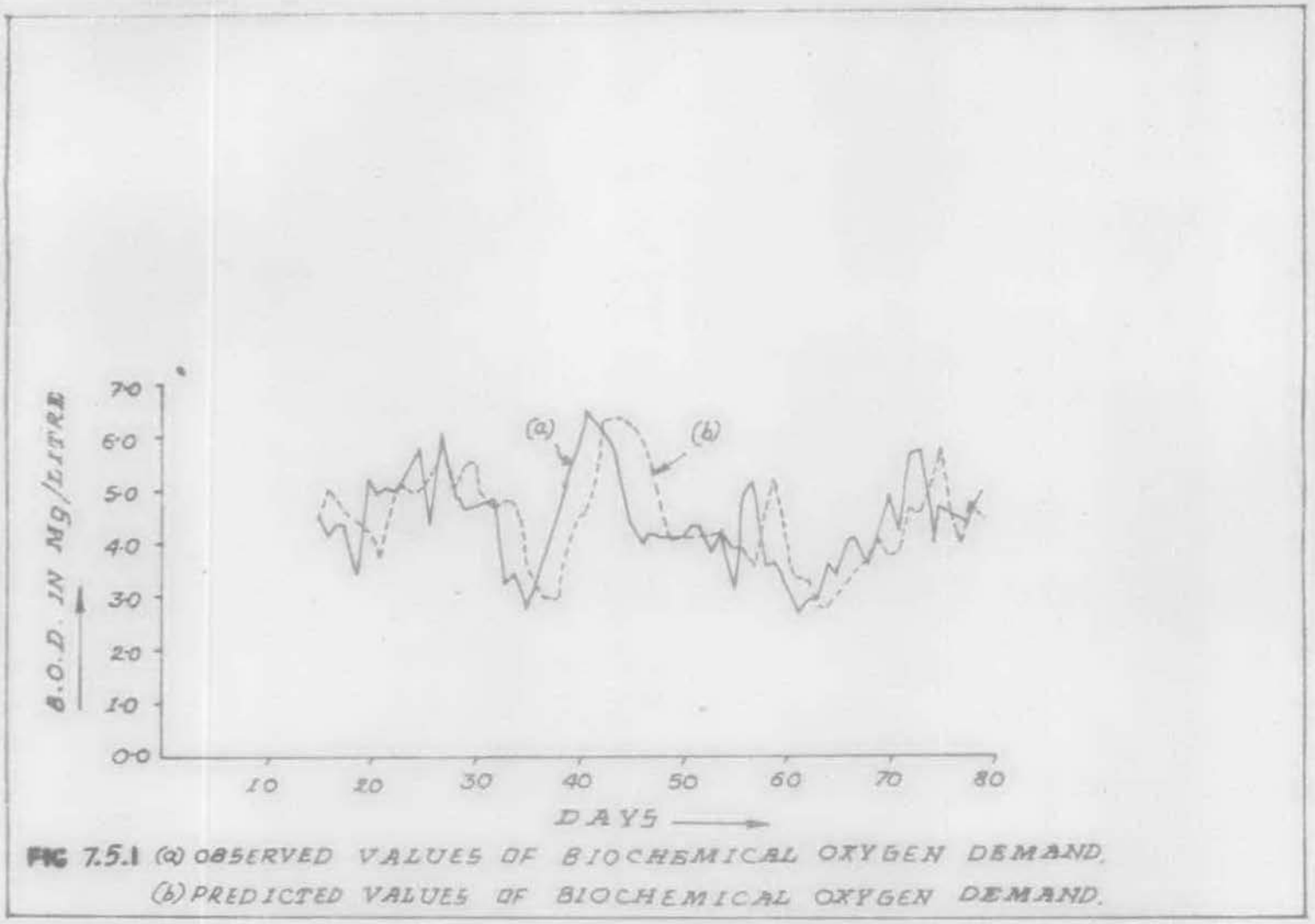
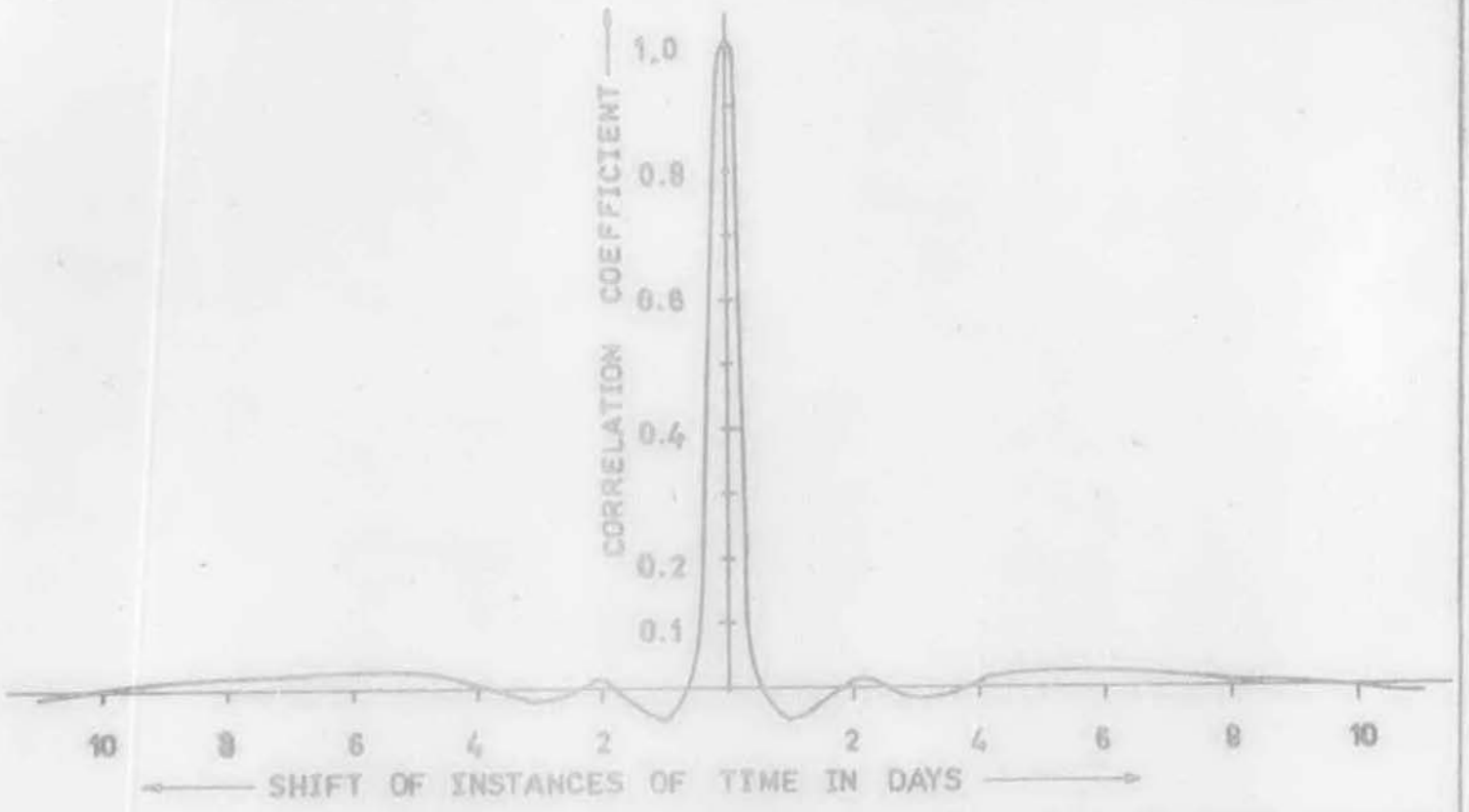


FIG 7.5.1 (a) OBSERVED VALUES OF BIOCHEMICAL OXYGEN DEMAND.
(b) PREDICTED VALUES OF BIOCHEMICAL OXYGEN DEMAND.



CORRELATION COEFFICIENTS Vs. SHIFT OF INSTANCES OF TIME OF THE ERROR OF PREDICTION OF DAILY LEVELS OF B.O.D.

FIG 7.5.2

CHAPTER VIII

REAL TIME FORECASTING OF DISSOLVED OXYGEN LEVEL OF A NON TIDAL RIVER

CHAPTER VIII

REAL TIME FORECASTING OF DISSOLVED OXYGEN LEVEL OF A NON-TIDAL RIVER.

8.0 Abstract :

Recursive least square non-stationary time series analysis techniques of cybernetics has been put to good use in on-line forecasting of daily dissolved oxygen concentrations of a non-tidal stream. A simple dynamic model is obtained. When tested with field data the model is shown to simulate adequately the major variations of D.O. levels observed in the field measurements.

8.1 Introduction :

This work introduces some of the fundamental techniques and principles of control theory broadly classified as cybernetics and indicates how they might be put to good use in the real time forecasting of the water quality of a non-tidal stream. This forecasting model would help in the subsequent development of the methodologies for realtime controlling of water quality. For on line (Real time) information processing the basis of the processing mechanism is a recursive least square algorithm for process parameter estimation. Recursive parameter estimation algorithms have the advantages of minimal computational requirements and their ability to track time varying parameters. In the present work dynamic model for dissolved oxygen in a reach of river has been obtained. The model has been verified by simulation against field measurement collected over a 80-day period from the River Cam in Eastern England [108].

The dissolved oxygen level is a major index of the general state of an aquatic environment. The important engineering decisions in water quality control relate to the determination, regulation and maintenance of the level of D.O. that is consistent with the multiple uses of the natural water bodies. The D.O. concentrations are important aspect of water quality in streams since the temporary fluctuations of D.O. level below critical values can cause long term and possibly irreversible damage to the ecosystem [107].

8.2. Processing of the Statistical Data :

The daily measured data for D.O. concentration of a stream are $P(i)$, $i = 1, 2, \ldots, N1$.

the time instants (days).

The mean value is

$$\bar{P}(N1) = \frac{1}{N1} \sum_{i=1}^{N1} P(i) \qquad \ldots \quad (8.1)$$

Auto covariance of the data at lag instant k is given by

$$CA(k) = \frac{1}{N1-k} \sum_{i=1}^{N1-k} (P(i) - \bar{P}(N1)) (P(i+k) - \bar{P}(N1)) \qquad (8.2)$$

where $k = 0, 1, 2, \ldots, M \qquad M < \frac{1}{4} N1$

Normalised co-efficients of covariance is

$$RA(k) = \frac{CA(k)}{CA(o)} \qquad \ldots \quad (8.3)$$

Estimate of the normalised power density spectra for the data are given as,

$$PS(W_h) = \frac{2}{\pi} \sum_{k=0}^{M} E_k \; RA(k) \; \cos W_h k \qquad (8.4)$$

where $w_h = 2\pi f_h$; $f_h = \frac{h}{2M}$; $0 \leq f_h \leq 0.5$,

$h = 0, 1, 2, \ldots, M$

E_k has been defined as the weight for window correction,

$$E_k = \begin{cases} 1.0, & 0 < k < M \\ 0.5, & K = 0, \quad M \end{cases} \qquad \ldots (8.5)$$

These raw estimate of power spectral density are smoothed by using Hanning Window to obtain the final estimates of the power spectream. The smoothed estimates of the ordinates of power spectrum are given by

$$h = 0, \quad S(w_0) = 0.54\ PS(w_0) + 0.46\ PS(w_{h+1})$$

$$0 < h < M; \quad S(w_h) = 0.23\ PS(w_{h-1}) + 0.54\ PS(w_h) + 0.23\ PS(w_{h+1}).$$

$$h = M; \quad S(w_h) = 0.54\ PS(w_M) + 0.46\ PS(w_{M-1}) \ldots (8.6)$$

Periodicity in terms of fundamental and its harmonics can be estimated from the power spectral density Vs frequency characteristics.

8.3. Development of the Non-Stationary Process Model :

Non-stationary process with deterministic periodicity can be expressed as

$$P(k) = U(k) + V(k) \qquad \ldots (8.7)$$

Where U(k) is the deterministic trend components,

V(K) is the stationary stochastic component,

$k = 1, 2, \ldots,$ are time instances,

Deterministic periodic component can be represented as

$$U(k) = \sum_{i=1}^{L} (a_i \cos(i\ 2\pi f_0 k) + b_i \sin(i\ 2\pi f_0 k)) \qquad \ldots (8.8)$$

where fo is the fundamental cycle per instance of time. Stochastic component V(k) can be expressed in the autoregressive form as,

$$V(k) = \sum_{n=1}^{N} C_{k-n} V(k-n) + \eta_{(k)} \quad \ldots (8.9)$$

where $n = 1, 2, \ldots, N$, the prehistory interval and $\eta_{(k)}$ is modelled in moving average form

as $$\eta_{(k)} = \sum_{j=1}^{m} C_{k-j} \eta_{(k-j)} + v_{(k)} \quad \ldots (8.10)$$

where m is the prehistory interval of the moving average terms and v(k) resembles to a white noise process with zero mean and constant variance.

V(k) can also be expressed as,

$$V(k-n) = P(k-n) - U(k-n) \quad \ldots (8.11)$$

Hence equation (8.9) can be written as

$$V(k) = \sum_{n=1}^{N} C_{k-n} P_{(k-n)} - \sum_{n=1}^{N} C_{k-n} U_{(k-n)} + \eta_{(k)} \quad \ldots (8.12)$$

It follows from equations (8.7) and (8.8)

$$P(k) = \sum_{n=1}^{N} C_{k-n} P_{(k-n)} + \sum_{n=0}^{N} \sum_{i=0}^{L} \left[d^{i}_{k-n} \cos\left(i 2\pi fo\,(k-n)\right) + g^{i}_{k-n} \sin\left(i\, 2\pi fo\,(k-n)\right) \right] + \eta(k) \quad \ldots (8.13)$$

In matrix form the process P(k) can be expressed as

$$P(k) = a^{T}(k-1)\, Z(k-1) + \eta(k) \quad \ldots (8.14)$$

where $a^T(k-1) = [C_{k-1}, \ldots, C_{k-n}, \sum_{n=0}^{N} (d^0_{k-n}), \ldots, g^{\frac{1}{2}}_{k-n}]$ ... (8.15)

And

$$Z(k-1) = [P(k-1), \ldots, P(k-n), 1, \ldots, \text{Cos } i(2\lambda f_0 (k-n)), \ldots, \text{Sin } i(2\lambda f_0 (k-n))]^T \quad \ldots \quad (8.16)$$

$\hat{P}(k)$, the estimate of $P(k)$ can be very clearly written as

$$\hat{P}(k) = a^T (k-1) \; Z(k-1) \quad \ldots \quad (8.17)$$

8.4 Development of the Real Time Recursive Least Square Prediction Algorithm :

The co-efficient vector a can be estimated by minimising the quadratic function $J_k(a)$,

$$J_k(a) = \sum_{j=1}^{k} (P(i) - a^T Z(j-1))^2 + (a-a(o))^T S^{-1}(o) (a-a(o)) \quad (8.18)$$

where a(o) is the available a priori estimate of the co-efficient vector 'a' and S(o) is the positive definite weighting matrix of the order mxm, where

$$m = (N+1)(2L+1) \quad \ldots \quad (8.19)$$

and N is the preliminary internal

and L is the number of harmonics including the fundamental,

For minimization,

$$\frac{\partial J_k(a)}{\partial a} = -2 \sum_{j=1}^{k} Z(j-1)(P(j)-a^T Z(j-1)) + 2S^{-1}(o)(a - a(o)) \quad \ldots \quad (8.20)$$

Since we are seeking for a minima,

$$\frac{\partial J_k(a)}{\partial a} = 0 \qquad \ldots \quad (8.21)$$

It follows from equation (8.21)

$$\sum_{j=1}^{k} Z(j-1)P(j) + S^{-1}(0)\, a(o) = \sum_{j=1}^{k} Z(j-1)Z^T(j-1)a + S^{-1}(o)a \qquad (8.22)$$

$$\text{Let } S^{-1}(k) = \sum_{j=1}^{k} Z(j-1)Z^T(j-1) + S^{-1}(o) \qquad \ldots \quad (8.23)$$

$$\text{and} \quad d(k) = \sum_{j=1}^{k} Z(j-1)\, P(j) + S^{-1}\, a(o) \qquad \ldots \quad (8.24)$$

Denoting the estimate of the co-efficient vector 'a' as a(k) at time instant k,

$$S^{-1}(k)\; a(k) = d(k) \qquad \ldots \quad (8.25)$$

$$a(k) = S(k)\; d(k)$$

From equations (8.23) and (8.24) the following recursive equations are obtained,

$$S^{-1}(k+1) = S^{-1}(k) + Z(k)\, Z^T(k) \qquad \ldots \quad (8.26)$$

$$d(k+1) = d(k) + Z(k)\, P(K+1) \qquad \ldots \quad (8.27)$$

By matrix inversion lemma the recursive least square parameter estimation algorithms are

$$a(k+1) = a(k) + S(k+1)Z(k)\left[P(k+1) - a^T(k)\, Z(k)\right] \qquad \ldots \quad (8.28a)$$

$$S(K+1) = S(k) - S(k)Z(k)Z^T(k)S(k)\left[1+Z^T(k)S(k)Z(k)\right] \ldots \quad (8.28b)$$

Periodicity L may be estimated from the power density spectra versus cycle per time instance. Initial value of the prehistory interval N is estimated from the correlation function. The correlation function is defined as

$$\phi_{pp}(\lambda) = \frac{\sum_{i=1}^{N1-\lambda} (P(i)-\bar{P}(N1))\,(P(i+\lambda)-\bar{P}(N1))}{\left[\sum_{i=1}^{N1-\lambda} (P(i)-\bar{P}(N1))^2 \sum_{j=1+\lambda}^{N1} (P(j)-\bar{P}(N))^2\right]^{\frac{1}{2}}} \quad \ldots (8.29)$$

where λ is the shift of instances of time. For estimation of the parameters of the moving average components of equation (8.10) recursive algorithms (8.28a) and (8.28b) have been followed. The algorithms have been initialised with $k = 0$; $S(0) = I$ (unit matrix), $a(0) = 0$ and $P(j) = 0$, for $j = u, -1, -2, \ldots$

The non-stationary model of optimum complexity is obtained by comparing the criterion of integral square error between the predicted values and the observed values in the interpolation region for the different values of the prehistory interval N.

Criterion for the integral square error is defined as

$$\sigma_P^2 = \frac{\sum_{i=1}^{N1} (P(i) - \hat{P}(i))^2}{\sum_{i=1}^{N1} (P(i))^2} \quad \ldots (8.30)$$

8.5 Illustration :

Figure 8.5.1a shows the D.O. level at down stream end of 4.7 Km stretch of the River Cam in Eastern England, collected over an 80 day period during summer of 1972. Figure 8.5.2 shows the power density spectra versus cycle per day characteristic. It is evident that the fundamental component of the periodicity has 0.03125 cycle

per day and the harmonics are negligible. From the correlation coefficient versus shift of time as shown in figure 8.5.3 it is obvious the process follows a maximum prehistory interval of 5 time instances (days).

Minimum value of the integral square error was obtained as 0.0102826 with prehistory interval N = 2. The innovation process v(k) has been identified to have a mean of 0.077871 with variance 1.06121 for moving average terms in corporating N = 2. Figure 8.5.1b shows the error of prediction.

8.6. Conclusion :

A simple dynamic model is presented for the description of D.O. level in a non-tidal stream. When tested with field data the model is shown to simulate adequately the major variations observed in the field.

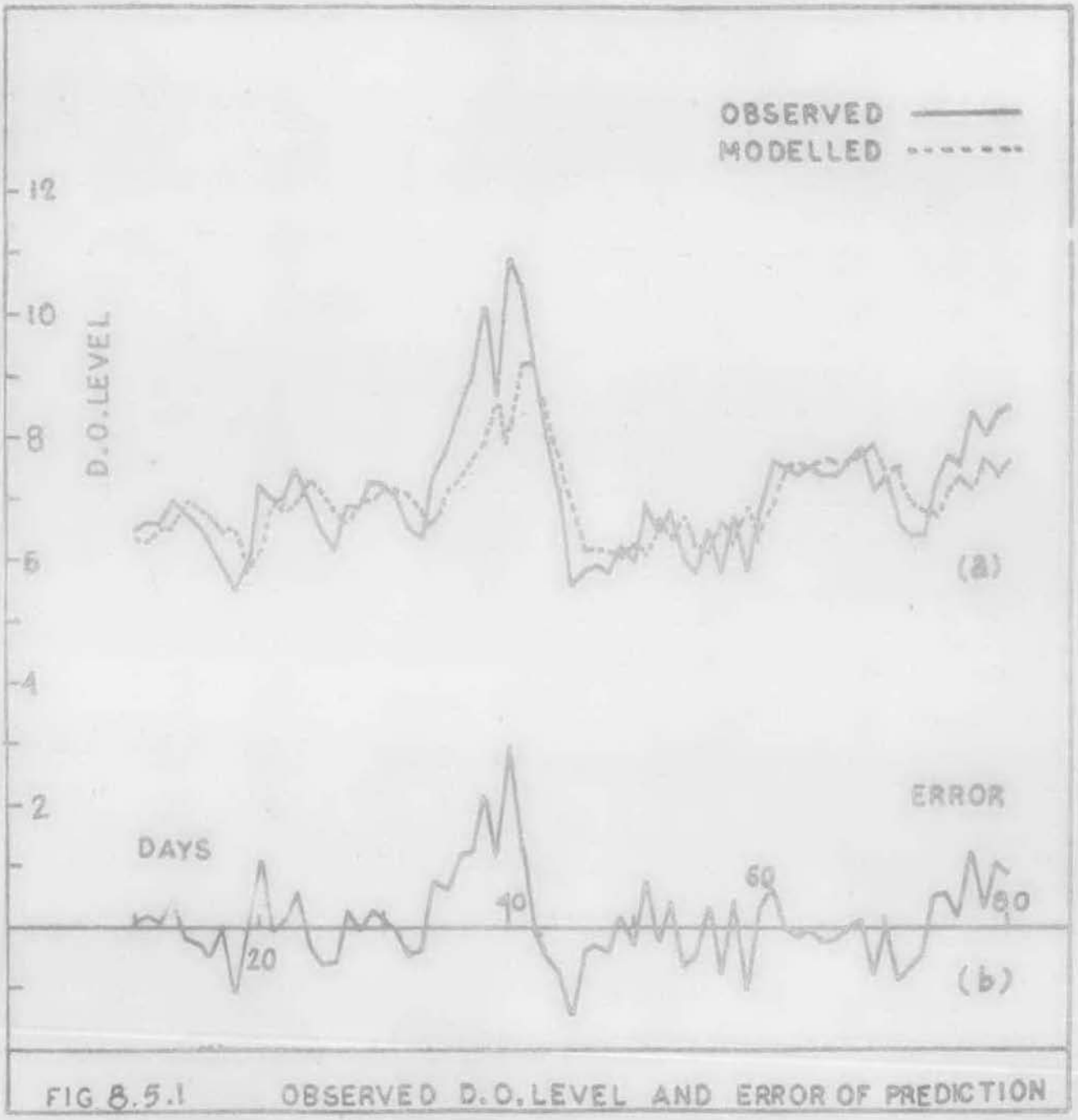
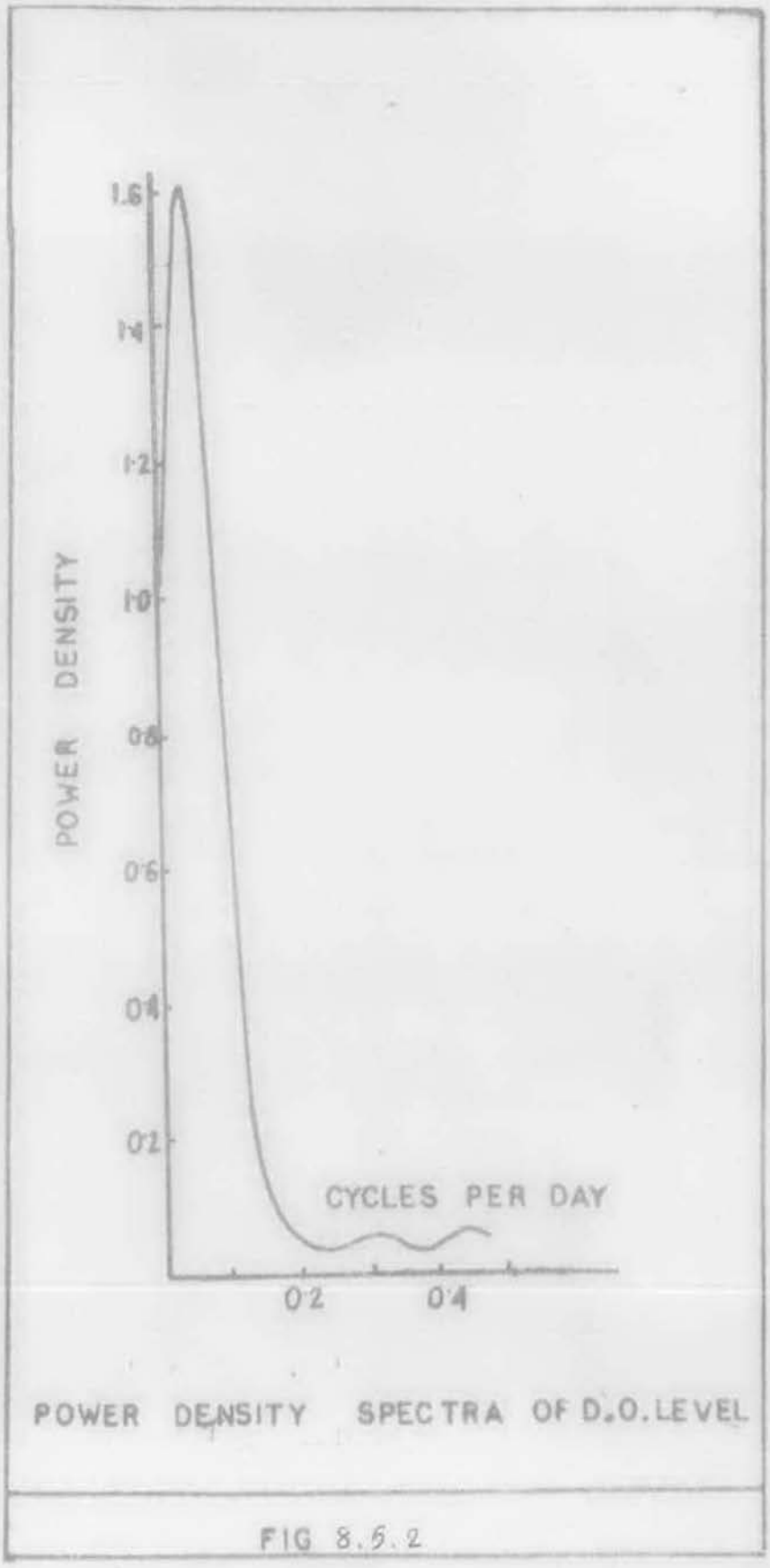


FIG 8.5.1 OBSERVED D.O. LEVEL AND ERROR OF PREDICTION



POWER DENSITY SPECTRA OF D.O. LEVEL

FIG 8.5.2

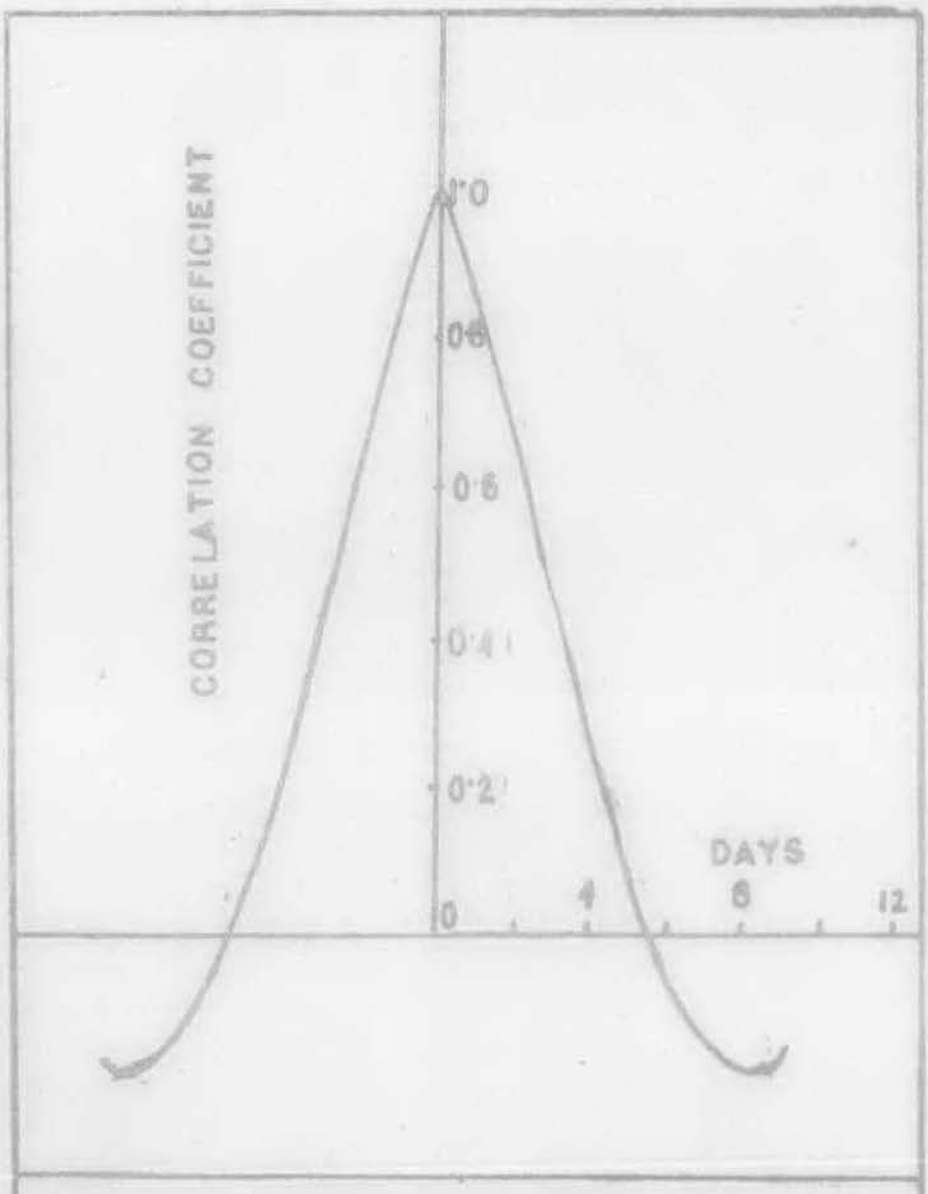


FIG 6.53 CORRELATION COEFFICIENT. D.O. LEVEL

CHAPTER IX

RECURSIVE PREDICTION OF DISSOLVED OXYGEN LEVEL OF A NON TIDAL RIVER

CHAPTER IX

RECURSIVE PREDICTION OF DISSOLVED OXYGEN LEVEL OF A REAL TIDAL RIVER.

9.0 Abstract :

Recursive instrument variable forecasting technique in conjunction with least square recursive parameter tracking algorithms of cybernetics has been used for real time prediction of the levels of the dissolved oxygen of a nontidal river passing through a highly industrialised region. A simple dynamic model has been obtained. When tested with field data the model is found to simulate adequately the major variation of the levels of dissolved oxygen observed in the field measurements. The techniques developed in this paper highlight the possibilities of the implementation of the on line regulatory mechanism for controlling water quality of a river.

9.1 Introduction :

This work offers scope for an useful application of recursive parameter tracking algorithms for real time forecasting of dissolved oxygen levels of a non tidal river.

There is a continuous and increasing awareness about the water quality in a river system, particularly, when the river passes through a thickly populated and highly industrialised region and the river is subjected to effluent discharges. It is required that the pollution levels of the water body should be maintained within reasonable limits consistent with the needs of the community and also for maintaing an overall ecological balance.

The quality of water in a river depends on many factors a few of these are - the level of dissolved oxygen, the biochemical oxygen demand, the presence of nitrates and chlorides, phosphates, the presence of bacteria and so on. However, the dissolved oxygen level and biochemical oxygen demand (D.O. and B.O.D) are considered two principal measures of water quality which indicate the general health of the river.

The dissolved oxygen level is expressed in mg/litre and can be measured by an in-stream probe. The biochemical oxygen demand which is, in fact, an aggregate of the capacities of the absorption of oxygen by the materials present in river water and usually expressed also in mg/litre, can be measured from a sample of river water. The sample is to be kept in a controlled environment for a period of about 5 days.

Determination of the level of dissolved oxygen is rather not difficult and that is why it is considered a major index of the general health of an aquatic environment. The important engineering decisions in water quality control relate to the determination, regulation and maintenance of the level of the dissolved oxygen that is consistent with the multiple uses of the natural water bodies. The D.O. concentrations are important aspects of water quality of rivers since the temporary fluctions of the levels of the dissolved oxygen below critical values can cause long term and possibly irreversible damge to the ecosystem.

This work introduces some of the techniques and principles of modern control theory broadly classified as cybernetics and indicates how these can be used in the real time forecasting of the level of dissolved oxygen of a non-tidal river. It is believed that the real time forecasting model of the level of dissolved oxygen developed in this paper would subsequently help in the formulation of methodologies for decision regulatisation to effect an on line water quality control mechanism. The basis of the information processing mechanism is a recursive instrument variable state estimation technique in conjunction with recursive least square parameter tracking algorithm. The recursive parameter estimation

algorithms have the advantages of minimal computational requirements and their ability to track time varying parameters. In the present work a dynamic real time model for the level of dissolved oxygen for a length of a river has been obtained. The model has been verified by simulation against field measurement collected over a 80 day period in summer in 1972 from the river Cam in Eastern England.

9.2. Development of the Real Time Model :

In recursive estimation algorithms the estimate of the output at a future instant is constantly updated on receipt of fresh information. The presence of significant coefficient of correlation with the levels of dissolved oxygen at past instants with the current instant clearly suggests that the process may be represented in the form of a non-stationary time series with a probability that the current value of the $Y(t_k)$ is a function of the previous output observations, the auto-regressive terms $Y(t_{k-1}), Y(t_{k-2}) \ldots$, together with a current unknown realisation of the noise process $\eta(t_k)$. The process, therefore, can be represented as

$$Y(t_k) = \left[\sum_{i=1}^{N1} \beta_i q^{-1} \right] Y(t_k) + \eta(t_k) \qquad \ldots \quad (9.1)$$

Where $q^{-1}(.)$ is the backward shift operator defined as $q^{-1} Y(t_k) = Y(t_{k-1})$. The determination of N1 is known as the model order determination which is also the modal structure identification. The number of the significant instants in the output correlation may be intuitively considered for modal order determination. The coefficient of correlation may be defined as,

$$\phi_{yy}(\lambda) = \frac{\sum_{i=1}^{N-\lambda} \left(Y(i) - \frac{1}{N} \sum_{i=1}^{N} Y(i)\right)\left(Y(i+\lambda) - \frac{1}{N} \sum_{i=1}^{N} Y(i)\right)}{\left[\sum_{i=1}^{N-\lambda} \left(Y(i) - \frac{1}{N} \sum_{i=1}^{N} Y(i)\right)^2 \sum_{j=1+\lambda}^{N} \left(Y(j) - \frac{1}{N} \sum_{i=1}^{N} Y(i)\right)^2 \right]^{\frac{1}{2}}}$$

Where N is the number of data points and λ are the shift of instances of time.

$\eta(t_k)$ may be expressed in a moving average sequence as

$$\eta(t_k) = \sum_{p=1}^{P} \gamma_p \, n(t_{k-p}) + \nu(t_k) \qquad \ldots \quad (9.3)$$

Where $\nu(.)$ is a white noise innovation process with

$$E\left\{\nu(t_k)\ \nu(t_j)\right\} = 0 \quad \text{for } j \neq k \qquad \ldots \quad (9.4)$$

and

$$E\left\{\nu(t_k)\ \nu(t_j)\right\} = \rho^2 \text{ for } j = k$$

This model is quite flexible since it requires that the equations be linear in parameters.

The equation (9.1) can be expressed as

$$Y(t_k) = z^T(t_k)\alpha + \eta(t_k) \qquad \ldots \quad (9.5)$$

Where α is the parameter vector with property of slowly varying with time and amenable to recursive adaptiveness, and

$$z^T(t_k) = \left[\, Y(t_{k-1}),\ Y(t_{k-2}),\ \ldots,\ Y(t_{k-n}) \,\right]$$

The least square estimation of the process parameters can be obtained by minimising a loss function defined as the sum of the squared errors used as

$$J \triangleq \sum_{k=1}^{N} \left(Y(t_k) - z^T(t_k)\,\hat{\alpha} \right)^2 \qquad \ldots \quad (9.6)$$

The estimates $\hat{\alpha}$ of α that minimise J are called least square estimates and $\hat{\alpha}$ is obtained as

$$\hat{\alpha} = \left[\sum_{k=1}^{N} z(t_k)\, z^T(t_k) \right]^{-1} \left[\sum_{k=1}^{N} z(t_k)\, Y(t_k) \right] \qquad \ldots \quad (9.7)$$

The most likely source of biased estimates is the presence of significant noise sequence correlation between the noise sequence and the past values of the output. To make the estimates unbiased as far as possible a section of output sequence of past observation is computed as a sequence of instrument variables.

Estimates of the sequence of Instrument variables.

A sequence of instrument variable $Y(t_k)$ is obtained from a non stationary time series model of the observed data

$$Y(t_k) = U(t_k) + V(t_k) \quad \ldots \quad (9.8)$$

Where U(.) is the deterministic trend component
V(.) is the stationary stochastic component
K = 1, 2 ..., are the time instances.

The determinixtic periodic component U(k) can be represented as,

$$U(t_k) = \sum_{r=0}^{R} a_r \mathrm{Cos}(r\,2\,f_o k) + b_r \mathrm{Sin}(r2\,f_o k) \quad \ldots \quad (9.9)$$

Where f_o is the fundamental cycle per instance of time and R are the number of frequencies including the fundamental. The number of frequencies is determined by a statistical data processing techniques as briefly described below.

Statistical characteristics of the observed data.

The daily measured data for D.O. concentration of a river are $Y(t_i)$ i = 1,2 ..., N, i being time instants in days.

The mean value is given by

$$Y(t_N) = \frac{1}{N} \sum_{i=1}^{N} Y(t_i) \quad \ldots \quad (9.10)$$

Auto covariance of the data at lag instant k is given by

$$CA(k) = \frac{1}{NK} \sum_{i=1}^{N-k} \left[Y(t_i) - Y(t_N) \right] \left[Y(t_{i+k}) - Y(t_N) \right] \quad \ldots \quad (9.11)$$

Where $k = 0\ 1,2, \ldots, M$ and $M < \frac{1}{4} N$

Normalised coefficients of covariance are given by

$$RA(k) = \frac{CA(k)}{CA(o)} \qquad \ldots \quad (9.12)$$

The estimates for the normalised power density spectra for the data are given as

$$PS(w_h) = \frac{2}{\pi} \sum_{k=o}^{M} E_k \, RA(k) \, \text{Cos} \, w_h \, k \qquad \ldots \quad (9.13)$$

Where $w_h = 2\pi f_h$; $fh = \frac{h}{2M}$ $h = 0, 1, 2 \ldots, M$

and $0.0 \leqslant f_h \leqslant 0.5$

E_k has been defined as the weight for window correction,

$$E_k = \begin{cases} 1.0, & 0 < k < M \\ 0.5; & k = 0, M \end{cases} \qquad \ldots \quad (9.14)$$

These raw estimates of power density spectra are smoothed to obtain the final estimates of the power spectrum. The smoothed estimates of the ordinates of the power spectrus are given by

$$h = 0; \; S(w_h) = 0.54 \; PS(w_h) + 0.46 \; PS(w_{h+1})$$

$$0 \; h \; M; \; S(w_h) = 0.23 \; PS(w_{h-1}) + 0.54 \; PS(w_h) + 0.23 \; PS(w_{h+1})$$

$$h = M; \; S(w_h) = 0.54 \; PS(w_h) + 0.46 \; PS(w_{h-1}) \qquad \ldots \quad (9.15)$$

Periodicity in terms of fundmental and its harmonics can be estimated from the power density spectra versus frequency characteristics.

The stochastic component $V(t_k)$ can be expressed in the autoregressive form as,

$$V(t_k) = \sum_{n=1}^{N1} C_{k-n} \, V(t_{k-n}) + \hat{\delta}(t_k) \qquad \ldots \quad (9.16)$$

Where n 1, 2, ..., N 1 the prehistory interval. N 1 may be intuitively determined from the coefficient of correlation versus shift off instances of time.

The output error sequence $\hat{\delta}(t_k)$ may be modelled as

$$\hat{\delta}(t_k) = \sum_{j=1}^{m} q_{k-j} \left(Y(t_{k-j}) - \hat{Y}(t_{k-j}) \right) \quad \ldots \quad (9.17)$$

Where $\hat{Y}(t_{k-j})$ is the estimate of output on the basis of the observation upto t_{k-j-1} instant.

The prehistory m of the moving average terms of the output error sequence may be determined for a particular type of process to adhere to a minimum of the integral square error criterion defined as,

$$ISE = \frac{\sum_{i=1}^{N} \left(Y(t_i) - \hat{Y}(t_i) \right)^2}{\sum_{i=1}^{N} \left(Y(t_i) \right)^2} \quad \ldots \quad (9.18)$$

$V(t_k)$ can be expressed as

$$V(t_{k-n}) = Y(t_{k-n}) - U(t_{k-n})$$

Hence the equation (9.16) can be written as,

$$V(t_{k-n}) = \sum_{n=1}^{N1} C_{k-n} Y(t_{k-n}) - \sum_{n=1}^{N1} C_{k-n} U(t_{k-n}) + \hat{\delta}(t_k) \quad (9.19)$$

It follows from equations (9.8) and (9.9)

$$V(t_k) = \sum_{n=1}^{N1} C_{k-n} Y(t_{k-n}) + \sum_{n=0}^{N1} \sum_{r=0}^{R} \left[d_{k-n}^{r} \operatorname{Cos}(r 2\pi f_o(k-n) + g_{k-n}^{r} \operatorname{Sin}(r 2\pi f_o(k-n)) \right] + \hat{\delta}(t_k) \quad \ldots \quad (9.20)$$

In matrix form,

$$Y(t_k) = a^T(t_{k-1})\ Z(t_{k-1}) \qquad \ldots \qquad (9.21)$$

Where

$$a^T(t_{k-1}) = \underline{\quad} C_{k-1}, \ldots, C_{k,n}, \sum_{n=0}^{N} d^0_{k-n}, \ldots, d^R_{k-n}, g^0_{k-n1} \ldots g^R_{k-n} + \sum_{j=1}^{m} q_{k-1} \underline{\quad}$$

and

$$Z(t_{k-1}) = \underline{\quad} Y(t_{k-1}), \ldots Y(t_{k-M1}), 1, \ldots, \text{Sin}(R2\pi f_0(k-M1)),$$
$$\text{Cos}(R\,1\pi f_0(k-M1)), (Y(t_{k-1}) - \hat{Y}(t_{k-1})), \ldots$$
$$Y(t_{k-m}) - \hat{Y}(t_{k-m}) \underline{\quad}^T \qquad \ldots \qquad (9.22)$$

The coefficient vector 'a' can be estimated by minimizing the quadratic function $J_k(a)$, defined as,

$$J_k(a) = \sum_{j=1}^{k} (Y(t_j) - a^T Z(t_{j-1}))^2 + (a-a(t_0))^T S^{-1}(t_0)(a-a(t_0)) \qquad \ldots \qquad (9.23)$$

Where $a(t_0)$ is the available a priori estimate of the coefficient vector 'a' and $S(t_0)$ is the positive definite weighting matrix of the order $m_1 \times m_1$, where.

$$m_1 = (N1+1)\ (2R+1) \qquad \ldots \qquad (9.24)$$

For minimization,

$$\frac{\partial J_k(a)}{\partial a} = -2 \sum_{j=1}^{k} Z(t_{j-1})(Y(t_j) - a^T Z(t_{j-1})) + 2S^{-1}(t_0)(a-a(t_0)) = 0 \qquad \ldots \qquad (9.25)$$

It follows from equation (9.25)

$$\sum_{j=1}^{k} Z(t_{j-1})Y(t_j)+S^{-1}(t_o)a(t_o) = \sum_{j=1}^{k} Z(t_{j-1})Z^T(t_{j-1})+S^{-1}(t_o)\ a \quad \ldots \quad (9.26)$$

$$\text{Let } S^{-1}(t_k) = \sum_{j=1}^{k} Z(t_{j-1})Z^T(t_{j-1})+S^{-1}(t_o) \quad \ldots \quad (9.27)$$

and

$$d(t_k) = \sum_{j=1}^{k} Z(t_{j-1})Y(t_j) + S^{-1}(t_o)a(t_o) \quad \ldots \quad (9.28)$$

Denoting the coefficient vector 'a' as $a(t_k)$ at time instant t_k

$$S^{-1}(t_k)\ a(t_k) = d(t_k) \quad \ldots \quad (9.29)$$

$$a(t_k) = S(t_k)\ d(t_k)$$

From equations (9.27) and (9.28) the following recursive equations are obtained.

$$S^{-1}(t_{k+1}) = S^{-1}(t_k)+Z(t_k)Z^T(t_k) \quad \ldots \quad (9.30)$$

$$d(t_{k+1}) = d(t_k) + Z(t_k)\ Y(t_{k+1}) \quad \ldots \quad (9.31)$$

By matrix inversion lemma the recursive parameter estimation algorithms to obtain the instrument variables $Y(t_k)$ are

$$a(t_{k+1}) = a(t_k)+S(t_{k+1})Z(t_k)\left[Y(t_{k+1}) - a^T(t_k)Z(t_k)\right] \quad \ldots \quad (9.32a)$$

$$S(t_{k+1}) = S(t_k) - S(t_k)Z(t_k)Z^T(t_k)S(t_k)\left[1+Z^T(t_k)S(t_k)Z(t_k)\right] \quad \ldots \quad (9.32b)$$

The algorithms are initialised with

$S(t_o) \triangleq I$ (Unit Matrix); $a(t_o) \triangleq 0$; $Y(t_j) = 0$ for $j = 0, -1, -2,$ and $\hat{\delta}(t_j)=0$ for $j = 0, -1, -2.$

Referring to equation (9.7)

Let

$$P^{*}(t_k) \triangleq \left[\sum_{j=1}^{k} Z(t_j) Z^{T}(t_j) \right]$$

and

$$b(t_k) \triangleq \left[\sum_{j=1}^{k} Z(t_j)\, Y(t_j) \right]$$

So equation (9.7) becomes.

$$\hat{\alpha} = P^{*}(t_k)\, b(t_k) \qquad \ldots \quad (9.33)$$

The recursive relationship for $P^{*}(.)$ and $b(.)$ can be set as

$$\left[P^{*}(t_k) \right]^{-1} = \left[\sum_{j=1}^{k} Z(t_j) Z^{T}(t_j) \right] = \sum_{j=1}^{k-1} Z(t_j) Z^{T}(t_j) + Z(t_k) Z^{T}(t_k)$$

$$= \left[P^{*}(t_{k-1}) \right]^{-1} + Z(t_k) Z^{T}(t_k) \qquad \ldots \quad (9.34)$$

Similarly,

$$b(t_k) = b(t_{k-1}) + Z(t_k)\, Y(t_k) \qquad \ldots \quad (9.35)$$

Premultiplying by $P^{*}(t_k)$ and post. The equation (9.34) becomes

$$I = P^{*}(t_k) \left[P^{*}(t_{k-1}) \right]^{-1} + P^{*}(t_k) Z(t_k) Z^{T}(t_k) \qquad \ldots \quad (9.36)$$

Post multiplying equation (9.36) by $P^{*}(t_{k-1})$

$$P^{*}(t_{k-1}) = P^{*}(t_k) + P^{*}(t_k)\, Z(t_k) Z^{T}(t_k)\, P^{*}(t_{k-1}) \qquad \ldots \quad (9.37)$$

Post multiplying equations (9.37) by $Z(t_k)$

$$P^{*}(t_{k-1})\, Z(t_k) = P^{*}(t_k)\, Z(t_k) + P^{*}(t_k)\, Z(t_k) Z^{T}(t_k) P^{*}(t_{k-1}) Z(t_k)$$

So we get,

$$P^{*}(t_{k-1}) Z(t_k) = P^{*}(t_k) Z(t_k) \left[1 + Z^{T}(t_k) P^{*}(t_{k-1}) Z(t_k) \right] \qquad \ldots \quad (9.38)$$

Post multiplication by

$$\left[\right.$$

$$\left[1+Z^T(t_k)P^e(t_{k-1})Z(t_k)\right]^{-1} Z^T(t_k)P^e(t_{k-1})$$

the equation (9.38) given,

$$P^e(t_{k-1})Z(t_k)\left[1+Z^T(t_k)\,P^e(t_{k-1})Z(t_k)\right]^{-1} Z^T(t_k)\,P^e(t_{k-1})$$

$$= P^e(t_k)\,Z(t_k)Z^T(t_k)\,P^e(t_{k-1}) \qquad \ldots \quad (9.39)$$

Substitution of equation (9.37) in equation (9.39) gives,

$$P^e(t_k) = P^e(t_{k-1}) - P^e(t_{k-1})Z(t_k)\left[1+Z^T(t_k)P^e(t_{k-1})Z(t_k)\right]^{-1}$$

$$Z^T(t_k)\,P^e(t_{k-1}) \qquad \ldots \quad (9.40)$$

From equation (9.33)

$$\hat{\alpha}\,(t_k) = P^e(t_k)\;b\,(t_k)$$

Combining equation (9.35) and (9.40)

$$\hat{\alpha}(t_k) = \left[P^e(t_{k-1}) - P^e(t_{k-1})\,Z(t_k)\left[1+Z^T(t_k)\,P^e(t_{k-1})\,Z(t_k)\right]^{-1}\right.$$

$$\left.\times Z^T(t_k)\,P^e(t_{k-1})\right]\left\{b\,(t_{k-1}) + Z(t_k)\,Y(t_k)\right\} \qquad (9.41)$$

Since $P^e(t_{k-1})\,b\,(t_{k-1}) = \hat{\alpha}\,(t_{k-1})$

The equation (9.41) becomes,

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) - P^e(t_{k-1})\;Z(t_k)\left[1+Z^T(t_k)\,P^e(t_{k-1})\,Z(t_k)\right]^{-1}$$

$$\times Z^T(t_k)\,\hat{\alpha}\,(t_{k-1}) + P^e(t_{k-1})\;Z(t_k)\;Y(t_k) \ldots$$

$$- P^e(t_{k-1})\,Z\,(t_k)\left[1+Z^T\,(t_k)\,P^e\,(t_{k-1})\,Z(t_k)\right]^{-1} \ldots \qquad (9.42)$$

$Z^T(t_k)\,P^e(t_{k-1})\,Z(t_k)\,Y(t_k)$

Since $Z^T\,(t_k)\;P^e\;(t_k)\;Z(t_k)$ is scalar

The equation (9.42) becomes

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) + P^*(t_{k-1})\, Z(t_k) \left[1+Z^T(t_k)\, P^*(t_{k-1})\, Z(t_k)\right]^{-1} \times \left[Y(t_k) - Z^T(t_k)\ \hat{\alpha}(t_{k-1})\right] \quad \ldots \quad (9.43)$$

The equations (9.43) and (9.40) with

$$P^*(t_k) \triangleq \sum_{j=1}^{k} Z(t_j)\ Z^T(t_j)\ ]^{-1}$$

are the recursive least square algorithms. But these algorithms do not overcome the problem of bias. The technique to minimise the bias is stated as follows.

From equation (9.7) the non recursive estimation of parameters at instant t_k is

$$\hat{\alpha}(t_k) = \left[\sum_{j=1}^{k} Z(t_j)\, Z^T(t_j)\right]^{-1} \left[\sum_{j=1}^{k} Z(t_j)\, Y(t_j)\right] \quad \ldots \quad (9.44)$$

and the process equation is,

$$Y(t_j) = Z^T(t_j)\,\alpha + \eta(t_j) \quad \ldots \quad (9.45)$$

Upon substitution of equation (9.45) in equation (9.44)

$$\hat{\alpha}(t_k) = \left[\sum_{j=1}^{k} Z(t_j)Z^T(t_j)\right]^{-1} \left[\sum_{j=1}^{k} Z(t_j)\left[Z^T(t_j) + \eta(t_j)\right]\right]$$

$$= \left[\sum_{j=1}^{k} Z(t_j)Z^T(t_j)\right]^{-1} \left[\sum_{j=1}^{k} Z(t_j)Z^T(t_j)\right]$$

$$+ \left[\sum_{j=1}^{k} Z(t_j)\, Z^T(t_j)\right]^{-1} \left[\sum_{j=1}^{k} Z(t_j)\, \eta(t_j)\right] \quad (9.46)$$

So,

$$\hat{\alpha}(t_k) = \alpha + \left[\sum_{j=1}^{k} Z(t_j)\, z^T(t_j) \right]^{-1} \left[\sum_{j=1}^{k} Z(t_j)\, \eta(t_j) \right] \quad (9.47)$$

For the estimation $\hat{\alpha}(t_k)$ to converges to

$$E\left\{ Z(t_j)\ \eta(t_j) \right\} = 0 \quad \text{for all } j$$

and

$$E\left\{ \eta(t_j)\, \eta(t_k) \right\} = 0 \quad \text{for } j \neq k$$
$$= \rho^2 \quad \text{for } j = k \quad \ldots \quad (9.48)$$

i.e. $\eta(t_j)$ converges to white noise process $v(t_j)$. Under these conditions the estimates will be unbiased. However, these conditions are difficult to achieve in practice. Since $\hat{Y}(t_k)$ sequence in equation (9.21) are already estimates, in a sense optimum, they are almost uncorrelated with $\eta(t_k)$. And consequently the vector $z(t_k)$ has been modified incorporating the instrument variables $\hat{Y}(.)$ as

$$\hat{z}(t_k) = \left[\hat{Y}(t_{k-1}), \hat{Y}(t_{k-2}), \ldots, \hat{Y}(t_{k-n}) \right] \quad \ldots \quad (9.49)$$

Thus the conditions of unbiased estimates are modified as

$$E\left\{ \hat{Z}(t_k)\ \eta(t_k) \right\} = 0 \quad \text{for all } k$$

Replacing $Z(t_k)$ by $\hat{Z}(t_k)$ and rather heuristically not $Z^T(t_k)$ by $\hat{Z}^T(t_k)$ because of experience in the interpolation region, in the equation (9.43) and (9.49) the recursive instrument variable algorithms become.

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) + \hat{P}^{\theta}(t_{k-1})\, \hat{Z}(t_k) \left[1 + Z^T(t_k)\, \hat{P}^{\theta}(t_{k-1})\, \hat{Z}(t_k) \right]^{-1}$$
$$\left\{ Y(t_k) - Z^T(t_k)\, \hat{\alpha}(t_{k-1}) \right\} \quad \ldots \quad (9.50a)$$

$$\hat{P}^{\theta}(t_k) = \hat{P}^{\theta}(t_{k-1}) - \hat{P}^{\theta}(t_{k-1})\, \hat{Z}(t_k) \left[1 + Z^T(t_k) \hat{P}^{\theta}(t_{k-1})\, \hat{Z}(t_k) \right]^{-1}$$
$$Z^T(t_k)\ \hat{P}^{\theta}(t_{k-1}) \quad \ldots \quad (9.50b)$$

$$\hat{P}^{*}(t_k) = \left[\sum_{j=1}^{k} Z(t_j) Z^{T}(t_j) \right]^{-1} \qquad \ldots \quad (9.50a)$$

The algorithms may be initialised with $\hat{\alpha}(.)$ as a least square estimate for a block of data in the interpolation region and $\hat{P}^{*}(.)$ for the same block of data.

9.3. Illustration :

Fig 9.3.1a shows the level of the dissolved oxygen at down stream and of 4.7 K.M stretch of the River Cam in Eastern England, collected over 80 day period during summer of 1972. Fig 9.3.2. shows the power density spectra versus cycle per day characteristics. It is evident that the fundamental component of periodicity has 0.03125 cycle per day and the harmonics are negligible. Fig 9.3.3. shows the coefficient of correlation versus shift of instances of time. Fig 9.3.4 shows the instrument variables in the interpolation region for output prehistory interval $N1 = 2$ and error sequence pre history interval $P = 2$. The sequence of instrument variables has an integral square error of 0.02892 and mean error 0.274885. Fig 9.3.1c shows the output obtained with recursive instrument variable algorithms with output prehistory interval $N1=2$ and integral square error = 8.820266 E - 03 and mean error 0.034068. Fig 9.3.5 shows the coefficients of correlation of the error sequence with recursive instrument variable algorithms for the different shift of instances of time. It is observed that the errors are almost uncorrelated [111].

9.4 Conclusion :

A simple dynamic model is presented for the real time forecasting of the level of the dissolved oxygen in a non-tidal river. When tested with field data the model is found to simulate adequately the major variations observed in the field. At the present moment automatic monitoring and control of river pollution is almost non existent. The present work regarding the real time forecasting of the level of dissolved oxygen is preliminary. However, this study, nevertheless analytical in nature, highlights the possibilities of the implementation of the real time and on line regulatory mechanism for controlling water quality of a river. The real time forecasting techniques as applied here would suggest the location of effluent treatment installations for maintaining a good regulation of dissolved oxygen about pre-specified levels, when the forecasting patterns are studied and adequately recognised for the different reaches of the river.

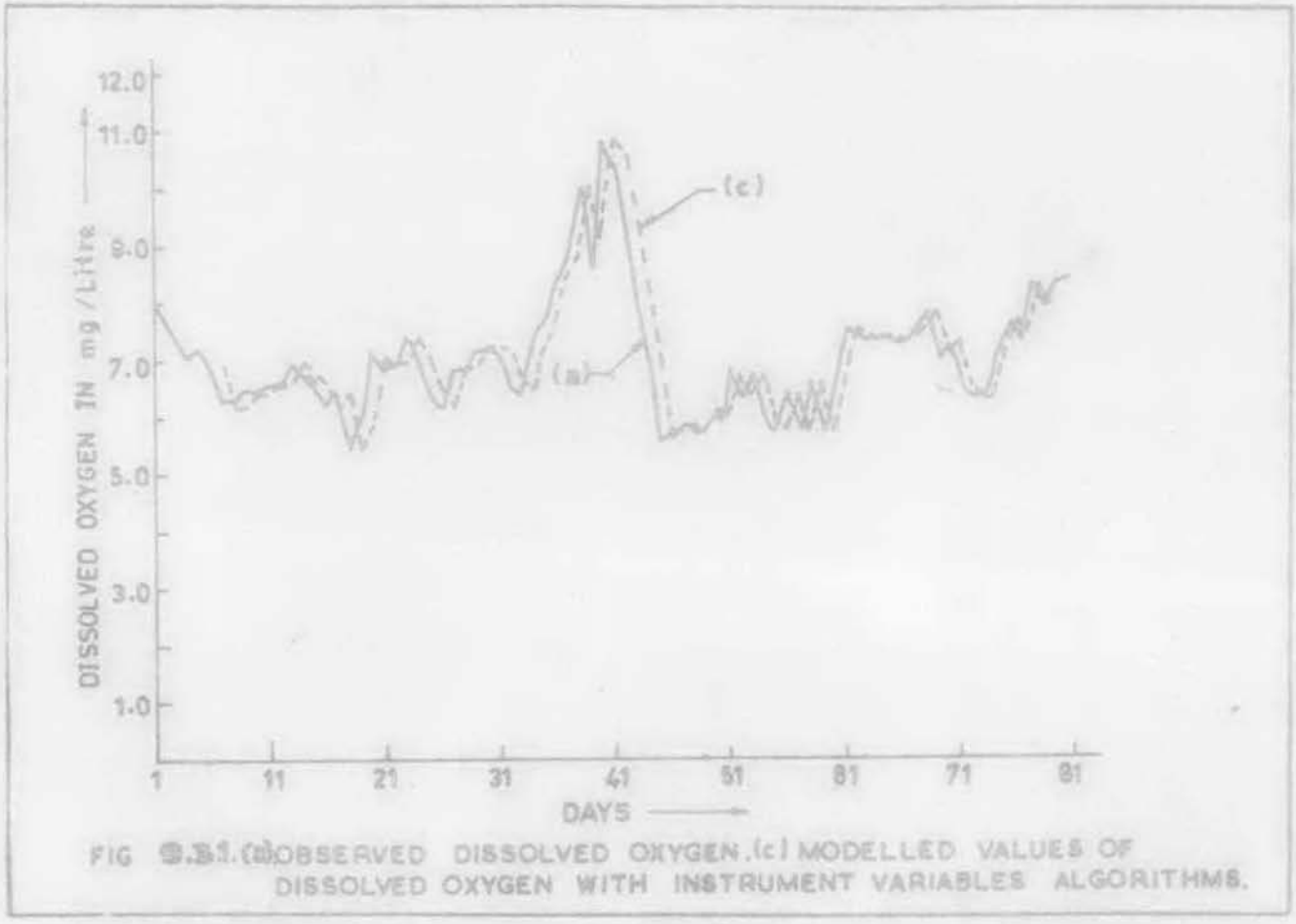
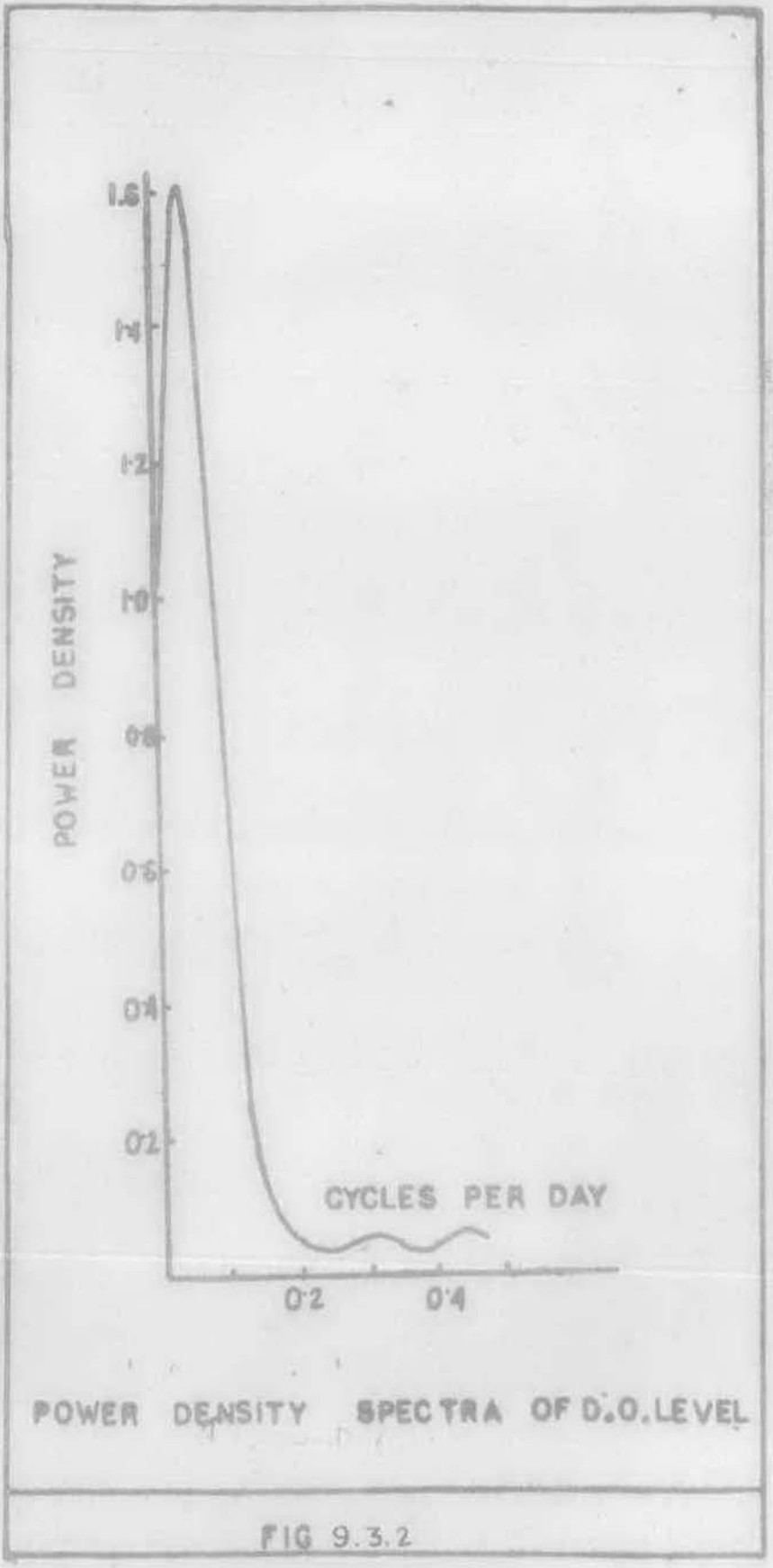


FIG 9.31. (a) OBSERVED DISSOLVED OXYGEN, (c) MODELLED VALUES OF DISSOLVED OXYGEN WITH INSTRUMENT VARIABLES ALGORITHMS.



POWER DENSITY SPECTRA OF D.O. LEVEL

FIG 9.3.2

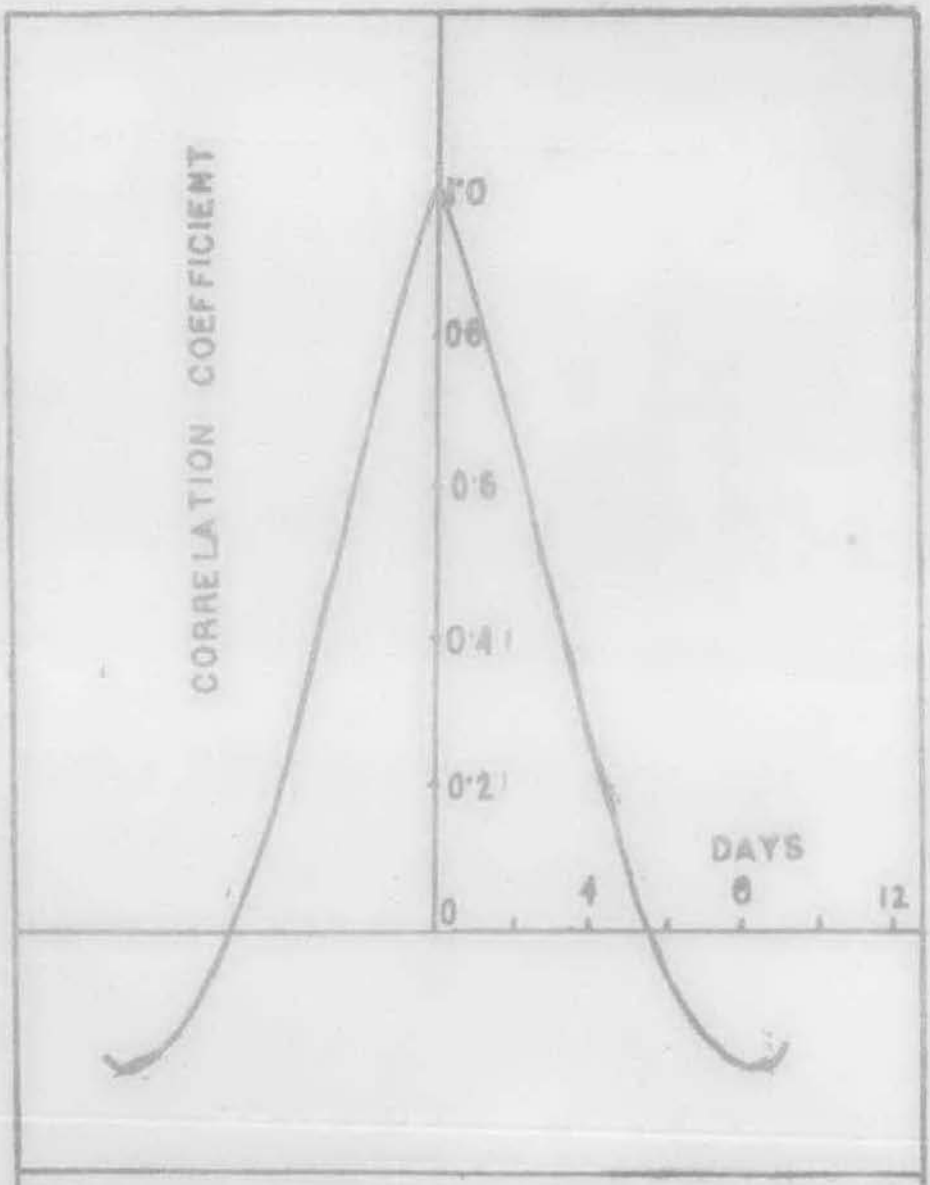
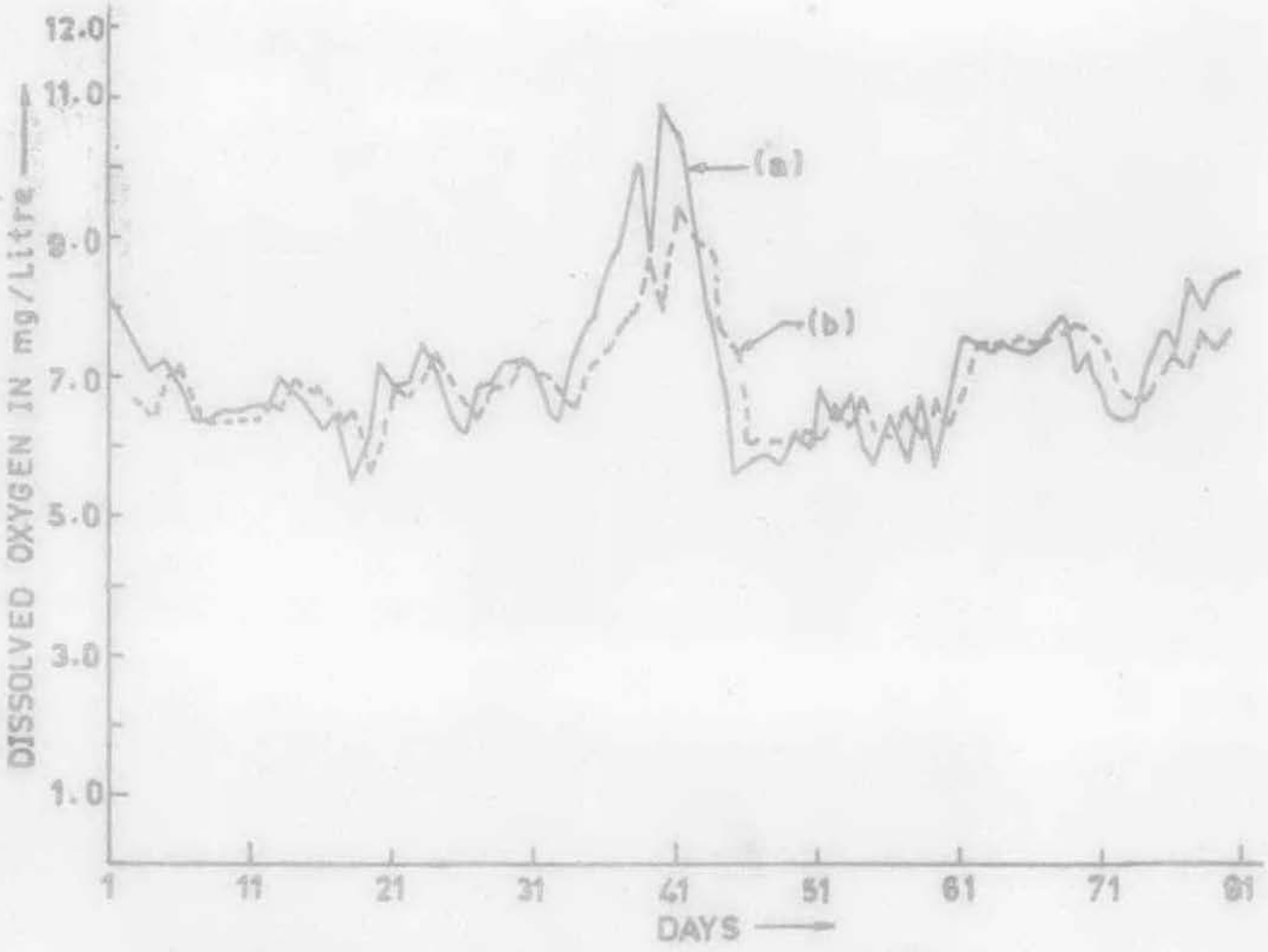
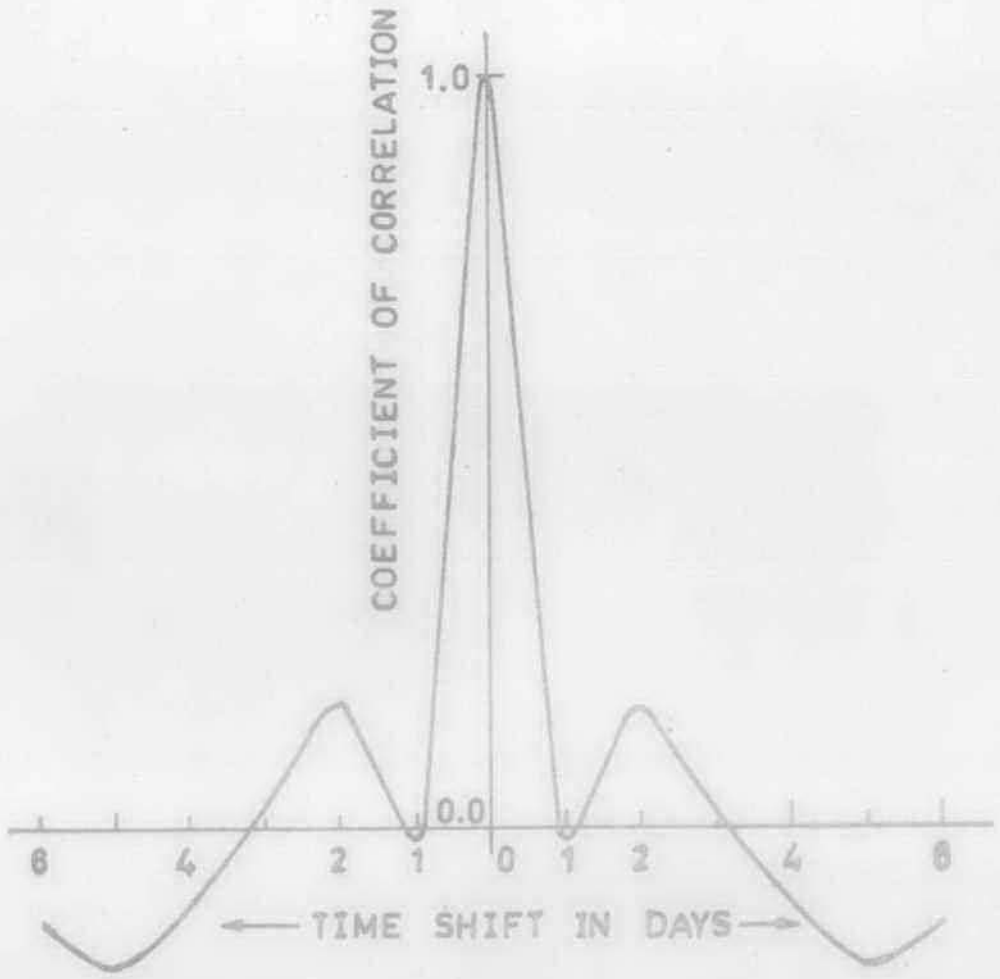


FIG 9.3.3 CORRELATION COEFFICIENT. D.O. LEVEL



(a).OBSERVED DISSOLVED OXYGEN (b). INSTRUMENT VARIABLES IN INTERPOLATION REGION.

FIG 9.3.4



COEFFICIENT OF CORRELATION OF THE ERROR SEQUENCE FOR DIFFERENT SHIFT OF INSTANCES OF TIME.

FIG 9.3.5

CHAPTER X

FORECASTING OF DAILY DISSOLVED OXYGEN LEVELS OF A NON TIDAL RIVER BY DYNAMIC LEAST SQUARE RECURSIVE ESTIMATION TECHNIQUE

CHAPTER X

FORECASTING OF DAILY DISSOLVED OXYGEN LEVELS OF A NON TIDAL RIVER BY DYNAMIC LEAST SQUARE RECURSIVE ESTIMATION TECHNIQUE.

10.0 Abstract :

This work deals with an application of the dynamic least square estimation algorithms for on-line modelling of dissolved oxygen levels of a non tidal river passing through a highly industrialised region. The mathematical description of the dissolved oxygen levels allows for the real time monitoring of water quality. Actual data are used to calibrate the real time dynamic model. Agreement between the observed and the predicted values of the levels of dissolved oxygen illustrates the potential use of the dynamic least square recursive estimation technique for on-line control of water quality in accordance with predetermined decision strategies.

10.1. Introduction :

The objective of this paper is to develop a dynamic model of the daily levels of the dissolved oxygen concentration of a non tidal river for the purpose of forecasting of the levels of dissolved oxygen on real time basis. The most important variable of interest in water quality is the concentration of dissolved oxygen, usually expressed in mg/litre, in the natural water body. Even the temporary fluctions of the concentration of the dissolved oxygen in a water body below the critical values may cause a long-term and possible irreversible damage to the aquatic ecosystem.

The improvement and consequent control of water quality of a river can be achieved by intelligent regulation of water discharges as to volume of discharge and the point of discharge. The important engineering decisions in water quality control relate to the determination of the concentration of the dissolved oxygen that is consistent with the multiple uses of the natural water bodies. This implies that the mathematical simulation models must have the ability to forecast the concentration of dissolved oxygen levels as closely as possible on real time basis.

This paper closely examines the development of dynamic least square recursive technique for forecasting the dissolved oxygen concentration on real time basis. The model has been verified by field measurement of the concentration of dissolved oxygen collected over a 80 day period in summer of 1972 from the river Cam in Eastern England.

For the sake of clarity of presentation the development of the dynamic least square recursive estimation technique has been given as a complete exposition.

10.2. Development of the Real Time Model :

In recursive estimation algorithms the estimate of the output at a future instant is constantly updated on receipt of fresh information. The presence of significant co-efficient of correlation with the levels of dissolved oxygen at past instants with the current instant clearly suggests that the process may be represented in the form of a stationary time series with a probability that the current value of the output, $y(t_k)$ is a function of the previous output observations, the auto-regressive terms $y(t_{k-1})$, $y(t_{k-2})$, ..., together with a current unknown realisation of the noise process.

The process, thus, can be represented as

$$y(t_k) = \left[\sum_{i=1}^{N1} \beta_i q^{-1} \right] y(t_k) + \eta(t_k) \quad \ldots \quad (10.1)$$

where q^{-1} is the backward shift operator defined as $q^{-1}y(t_k) = y(t_{k-1})$. The determination of N1 is unknown as the model order determination which is also the model structure identification. The number of the significant instants in the output correlation may be intuitively considered for model order determination. The co-efficient of correlation may be defined as,

$$\phi_{yy}(\lambda) = \frac{\sum_{i=1}^{N-\lambda} (y(t_i) - \frac{1}{N} \sum_{i=1}^{N} y(t_i))(y(t_{i+\lambda}) - \frac{1}{N} \sum_{i=1}^{N} y(t_i))}{\left[\sum_{i=1}^{N-\lambda} (y(t_i) - \frac{1}{N} \sum_{i=1}^{N} y(t_i))^2 \sum_{j=1+\lambda}^{N} (y(t_j) - \frac{1}{N} \sum_{i=1}^{N} y(t_i))^2 \right]^{\frac{1}{2}}}$$

$$\ldots \quad (10.2)$$

where N is the number of data points and λ are shift of instances of time.

The equation (10.1) can be expressed as

$$y(t_k) = x^T(t_k)\alpha + \eta(t_k) \quad \ldots \quad (10.3)$$

where 'α' is the parameter vector with property of slowly varying with time and amenable to recursive adaptiveness, and

$$x^T(t_k) = [\, y(t_{k-1}), y(t_{k-2}), \ldots, y(t_{k-N1}) \,]$$

The least square estimation of the process parameters can be obtained by minimising a loss function defined as the sum of the squared errors as

$$J \triangleq \sum_{k=1}^{N} (y(t_k) - x^T(t_k) \alpha \quad)^2 \qquad \ldots \quad (10.4)$$

The estimates $\hat{\alpha}$ of α that minimise J are called least square estimates and $\hat{\alpha}$ is obtained as

$$\hat{\alpha} = \left[\sum_{k=1}^{N} x(t_k) \; x^T(t_k) \right]^{-1} \left[\sum_{k=1}^{N} x(t_k) \; y(t_k) \right] \ldots (10.5)$$

Let

$$p^*(t_k) \triangleq \left[\sum_{j=1}^{k} x(t_j) \; x^T(t_j) \right]^{-1}$$

and

$$b(t_k) \triangleq \left[\sum_{j=1}^{k} x(t_j) \; y(t_j) \right]$$

So the equation (10.5) becomes,

$$\hat{\alpha} = p^*(t_k) \; b(t_k) \qquad \ldots \qquad (10.6)$$

The recursive relationship for $p^*(.)$ and $b(.)$ can be set as

$$\left[p^*(t_k) \right]^{-1} = \sum_{j=1}^{k-1} x(t_j) \; x^T(t_j) + x(t_k) x^T(t_k)$$

$$= \left[p^*(t_{k-1}) \right]^{-1} + x(t_k) \; x^T(t_k) \qquad \ldots \qquad (10.7)$$

Similarly,

$$b(t_k) = b(t_{k-1}) + s(t_k)\, y(t_k) \qquad \ldots \qquad (10.8)$$

Premultiplying by $P^a(t_k)$ we get from equation (10.7)

$$I = P^a(t_k) \left[P^a(t_{k-1})\right]^{-1} + P^a(t_k)\, s(t_k)\, s^T(t_k) \qquad \ldots \qquad (10.9)$$

Post multiplication of equation (10.9) by $P^a(t_{k-1})$

$$P^a(t_{k-1}) = P^a(t_k) + P^a(t_k)\, s(t_k)\, s^T(t_k)\, P^a(t_{k-1}) \qquad \ldots \qquad (10.10)$$

Post multiplication of equation (10.10) by $s(t_k)$

$$P^a(t_{k-1})s(t_k) = P^a(t_k)s(t_k) + P^a(t_k)s(t_k)s^T(t_k)P^a(t_{k-1})s(t_k)$$

So we get,

$$P^a(t_{k-1})s(t_k) = P^a(t_k)s(t_k)\left[1+s^T(t_k)P^a(t_{k-1})s(t_k)\right] \qquad \ldots \qquad (10.11)$$

Post multiplication by $\left[1+s^T(t_k)P^a(t_{k-1})s(t_k)\right]^{-1} s^T(t_k)P^a(t_{k-1})$ the equation (10.10) becomes,

$$P^a(t_{k-1})s(t_k)\left[1+s^T(t_k)P^a(t_{k-1})s(t_k)\right]^{-1} s^T(t_k)P^a(t_{k-1}) = P^a(t_k)s(t_k)s^T(t_k)P^a(t_{k-1}) \qquad \ldots \qquad (10.12)$$

Substitution of equation (10.10) in equation (10.12) gives,

$$P^a(t_k) = P^a(t_{k-1}) - P^a(t_{k-1})s(t_k)\left[1+s^T(t_k)P^a(t_{k-1})s(t_k)\right]^{-1} s^T(t_k)\, P^a(t_{k-1}) \qquad \ldots \qquad (10.13)$$

From equation (10.6)

$$\hat{\alpha}(t_k) = P^a(t_k)\, b(t_k)$$

Combining equations (10.8) and (10.13)

$$\hat{\alpha}(t_k) = \left[P^*(t_{k-1}) - P^*(t_{k-1}) z(t_k) \left[1 + z^T(t_k) P^*(t_{k-1}) z(t_k) \right]^{-1} \right.$$
$$\left. \times z^T(t_k) P^*(t_{k-1}) \right] \left\{ b(t_{k-1}) + z(t_k)\, y(t_k) \right\} \quad \ldots \quad (10.14)$$

Since

$$P^*(t_{k-1})\, b(t_{k-1}) = \hat{\alpha}(t_{k-1})$$

The equation (10.14) becomes,

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) - P^*(t_{k-1})\, z(t_k) \left[1 + z^T(t_k) P^*(t_{k-1}) z(t_k) \right]^{-1}$$
$$\times z^T(t_k)\, \hat{\alpha}(t_{k-1}) + P^*(t_{k-1})\, z(t_k)\, y(t_k)$$
$$- P^*(t_{k-1}) z(t_k) \left[1 + z^T(t_k) P^*(t_{k-1}) z(t_k) \right]^{-1} \quad \ldots \quad (10.15)$$

$z^T(k)\ P^*(t_{k-1})\ z(t_k)\ .\ Y(t_k)$

Since $z^T(t_k)\ P^*(t_{k-1})\ z(t_k)$ is scalar the equation (10.15) becomes,

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) + P^*(t_{k-1}) z(t_k) \left[1 + z^T(t_k) P^*(t_{k-1}) z(t_k) \right]^{-1}$$
$$\times \left[y(t_k) - z^T(t_k)\ \hat{\alpha}(t_{k-1}) \right] \quad \ldots \quad (10.16)$$

The equations (10.16) and (10.13) with

$$P^*(t_k) \triangleq \left[\sum_{i=1}^{k} z(t_i)\ z^T(t_i) \right]^{-1}$$

are the recursive least square algorithms.

The recursive least square algorithms do not overcome the problem of bias. The problem of bias in the parameter estimates may be defined as follows.

Recalling equation (10.5)

$$\hat{\alpha}(t_k) = \left[\sum_{j=1}^{k} x(t_j) x^T(t_j) \right]^{-1} \left[\sum_{j=1}^{k} x(t_j) y(t_j) \right]$$

Upon substitution of the process equation

$$y(t_j) = x^T(t_j)\alpha + \eta(t_j) \text{ in equation (10.5)}$$

$$\hat{\alpha}(t_k) = \alpha + \left[\sum_{j=1}^{k} x(t_j) x^T(t_j) \right]^{-1} \left[\sum_{j=1}^{k} x(t_j) \eta(t_j) \right] \quad \ldots \quad (10.17)$$

for (t_k) to be unbiased

$$E\left\{ x(t_j) \quad \eta(t_j) \right\} = 0 \text{ for all } j$$

and

$$E\left\{ \eta(t_k) \quad \eta(t_j) \right\} \quad 0 \text{ for } k \neq j$$

and

$$E\left\{ \eta(t_k) \quad \eta(t_k) \right\} = \sigma^2$$

i.e. (t_k) approaches the white noise sequence (t_k).

Let us define the covariance matrix of the least square parameter estimation as $P(t_k)$

$$P(t_k) = \text{Cov}\left\{ \tilde{\alpha} \quad (t_k) \right\} = E\left(\hat{\alpha}(t_k) - \alpha \right)\left(\hat{\alpha}(t_k) - \alpha \right)^T$$

$$= P^*(t_k) E\left\{ \left[\sum_{j=1}^{k} x(t_j) \quad (t_j) \right] \left[\sum_{j=1}^{k} (t_j) x^T(t_j) \right] \right\} P^{*T}(t_k)$$

$$P(t_k) = \sigma^2 p_e(t_k) \qquad \ldots \qquad (10.18)$$

σ^2 may be estimated as the variance of the residual error sequence $E(t_k) = y(t_k) - z^T(t_k)\ (t_k)$ after the estimates have achieved satisfactory convergence. Least square estimates of α may be considered for initialisation. Thus for almost unbiased estimate the least square recursive algorithms of (10.3) and (10.16) have been transformed by making substitution of equation (10.18) as

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) + P(t_{k-1})z(t_k) \Big/ \left[\sigma^2 + z^T(t_k)P(t_{k-1})z(t_k)\right] \Big/$$

$$\left\{ y(t_k) - z^T(t_k)\hat{\alpha}(t_{k-1}) \right\} \qquad \ldots \qquad (10.19)$$

$$P(t_k) = P(t_{k-1}) - P(t_{k-1})z(t_k) \Big/ \left[\sigma^2 + z^T(t_k)P(t_{k-1})z(t_k)\right] \Big/ z(t_k)P(t_{k-1}) \quad \ldots (10.20)$$

In case of time varying parameters the large errors between the observed and the modelled outputs obtained from the algorithms (10.19) and (10.20) are mainly due to the error for changing values of the parameters $\hat{\alpha}$ of the model. If the parameters variation laws are known a model for the parameter variation can be obtained and the algorithms of equation (10.19) and (10.20) can be updated accordingly. In the absence of any specific parameter variation law a random walk model may be postulated as

$$\alpha(t_k/t_{k-1}) = \alpha(t_{k-1}/t_{k-1}) + \zeta(t_k/t_{k-1}) \qquad \ldots \qquad (10.21)$$

Recalling equation (10.18)

$$P(t_k/t_{k-1}) = \mathrm{Cov}\left\{ \tilde{\alpha}(t_k/t_{k-1}) \right\}$$

$$= E\left\{ \alpha(t_k/t_{k-1}) - \alpha)(\alpha(t_k/t_{k-1}) - \alpha)^T \right\}$$

Hence, from equation (10.21)

$$P(t_k/t_{k-1}) = E\left\{ \alpha(t_{k-1}/t_{k-1}) + \zeta(t_k/t_{k-1}) - \alpha \right) \times \left(\alpha(t_{k-1}/t_{k-1}) + \zeta(t_k/t_{k-1}) - \alpha \right)^T \right\} \quad \ldots (10.22)$$

From equation (10.17)

$$P(t_k/t_{k-1}) = E\Big\{ \Big(\int \sum_{j=1}^{k-1} x(t_j)x^T(t_j) \int^{-1} \int \sum_{j=1}^{k-1} x(t_j)\ \eta(t_j) \int + \zeta(t_k/t_{k-1}) \Big)$$

$$\Big(\int \sum_{j=1}^{k-1} x(t_j)x^T(t_j) \int^{-1} \int \sum_{j=1}^{k-1} x(t_j)\ \eta(t_j) \int + \zeta(t_k/t_{k-1}) \Big)^T \Big\}$$

since $\zeta(t_k/t_{k-1})$ is a white noise process it is uncorrelated with $x(.)$, and $\eta(t_j)$ is scalar,

$$P(t_k/t_{k-1}) = \int \sum_{j=1}^{k-1} x(t_j)x^T(t_j) \int^{-1} \int \sum_{j=1}^{k-1} x(t_j)\, E\Big\{ \sum_{j=1}^{k} \eta(t_j)\ \eta^T(t_j) \Big\} \sum_{j=1}^{k-1} x^T(t_j) \int \times$$

$$\int \Big[\sum_{j=1}^{k} x(t_j)x^T(t_j) \Big]^{-1} \int^T + E\left\{ \zeta(t_k/t_{k-1})\ \zeta^T(t_k/t_{k-1}) \right\} \quad \ldots (10.23)$$

Considering $D = E\left\{ \zeta(t_k/t_{k-1}) \; \zeta^T(t_k/t_{k-1}) \right\}$ ... (10.24)

$$P(t_k/t_{k-1}) = \sigma^2 P^*(t_{k-1}/t_{k-1}) + D \quad \text{...} \quad (10.25)$$

$$P(t_k/t_{k-1}) = P(t_{k-1}/t_{k-1}) + D \quad \text{...} \quad (10.26)$$

and

$$\hat{\alpha}(t_k/t_{k+1}) = \hat{\alpha}(t_{k-1}/t_{k-1}) \quad \text{...} \quad (10.27)$$

Algorithms (10.19) and (10.20) are updated as

$$\hat{\alpha}(t_k/t_k) = \hat{\alpha}(t_k/t_{k-1}) + P(t_k/t_{k-1}) z(t_k) \left[\sigma^2 + z^T(t_k) P(t_{k-1}) z(t_k) \right]^{-1} \times \left\{ y(t_k) - z^T(t_k) \; \hat{\alpha}(t_k/t_{k-1}) \right\} \quad \text{...} \quad (10.28)$$

$$P(t_k/t_k) = P(t_k/t_{k-1}) - P(t_k/t_{k-1}) z(t_k) \left[\sigma^2 + z^T(t_k) P(t_k/t_{k-1}) z(t_k) \right]^{-1} \times z^T(t_k) \; P(t_k/t_{k-1}) \quad \text{...} \quad (10.29)$$

Combining equations (10.26) and (10.27) with (10.28) and (10.29) the dynamic recursive least square algorithms are

$$\hat{\alpha}(t_k/t_k) = \hat{\alpha}(t_{k-1}/t_{k-1}) + \left\{ P(t_{k-1}/t_{k-1}) + D \right\} z(t_k) \left[\sigma^2 + z^T(t_k) \left\{ P(t_{k-1}/t_{k-1}) + D \right\} z(t_k) \right]^{-1} \times \left\{ y(t_k) - z^T(t_k) \; \hat{\alpha}(t_{k-1}/t_{k-1}) \right\} \quad \text{...} \quad (10.30)$$

$$P(t_k/t_k) = \left\{ P(t_{k-1}/t_{k-1}) + D \right\} - \left\{ P(t_{k-1}/t_{k-1}) + D \right\} z(t_k) \times \left[\sigma^2 + z^T(t_k) \left\{ P(t_{k-1}/t_{k-1}) + D \right\} z(t_k) \right]^{-1} z^T(t_k) \left\{ P(t_{k-1}/t_{k-1}) + D \right\} \quad \text{...} \quad (10.31)$$

Retaining the simpler notations the equations (10.30) and (10.31) become

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) + \left\{ P(t_{k-1}) + D \right\} s(t_k) \left[\sigma^2 + s^T(t_k) \right.$$

$$\left\{ P(t_{k-1}) + D \right\} s(t_k) \left. \right]^{-1}$$

$$\times \left\{ y(t_k) - s^T(t_k) \hat{\alpha}(t_{k-1}) \right\} \quad \ldots \quad (10.32)$$

$$P(t_k) = \left\{ P(t_{k-1}) + D \right\} - \left\{ P(t_{k-1}) + D \right\} s(t_k) \left[\sigma^2 + s^T(t_k) \right.$$

$$\left\{ P(t_{k-1}) + D \right\} s(t_k) \left. \right]^{-1}$$

$$\times s^T(t_k) \left\{ P(t_{k-1}) + D \right\} \quad \ldots \quad (10.33)$$

The dynamic recursive least square algorithms are initialised with least square estimate of parameter vector $\hat{\alpha}(.)$.

10.3 Illustration :

Fig 10.3.1(a) shows the dissolved oxygen level at down stream end of the river Cam in Eastern England, collected over a 80 day period in summer of 1972. Fig 10.3.1(b) shows the predicted values of D.O. levels (dissolved oxygen levels) obtained with dynamic least square recursive estimation algorithms for optimum prehistory interval N1=7. The corresponding integral square error defined as

$$\text{ESE} = \frac{\sum_{i=1}^{N} (y(t_i) - y_o(t_i))^2}{\sum_{i=1}^{N} (y(t_i))^2}$$

is 8.1793-03 and the mean error is 0.0179, where $y_o(.)$ are the predicted values of dissolved oxygen levels. Fig 10.3.2. shows the correlation co-efficient versus shift of instances of time for the dissolved oxygen levels.

Fig 10.3.3. shows the correlation co-efficient versus shift of instances of time for the errors of prediction. It is found that the errors are almost uncorrelated. [112].

10.4 Conclusion :

The close agreement between the observed and the computed values of the levels of dissolved oxygen clearly demonstrates the potential application of dynamic least square recursive estimation algorithms for real time control of water quality of a non tidal river.

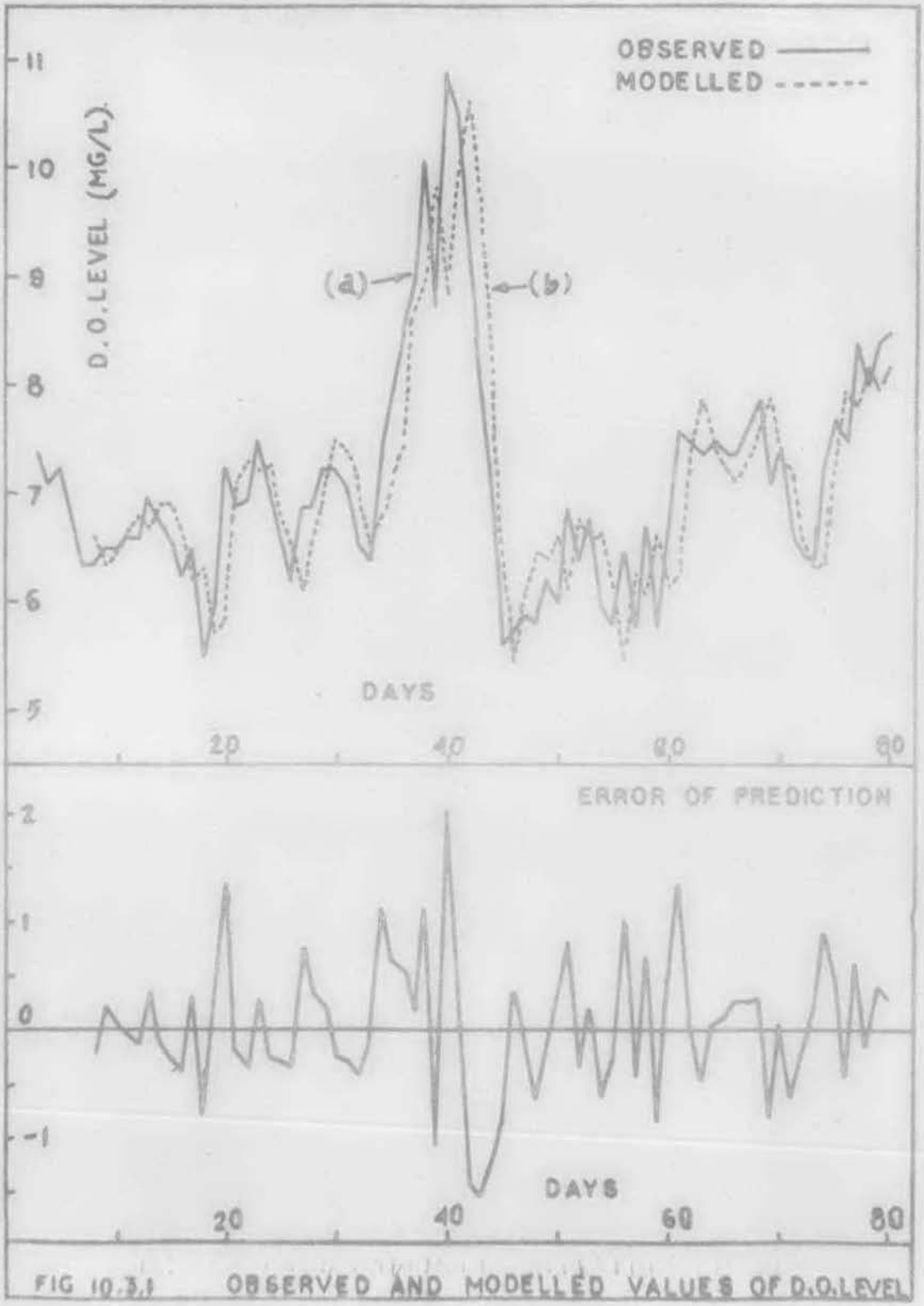


FIG 10.3.1 OBSERVED AND MODELLED VALUES OF D.O.LEVEL

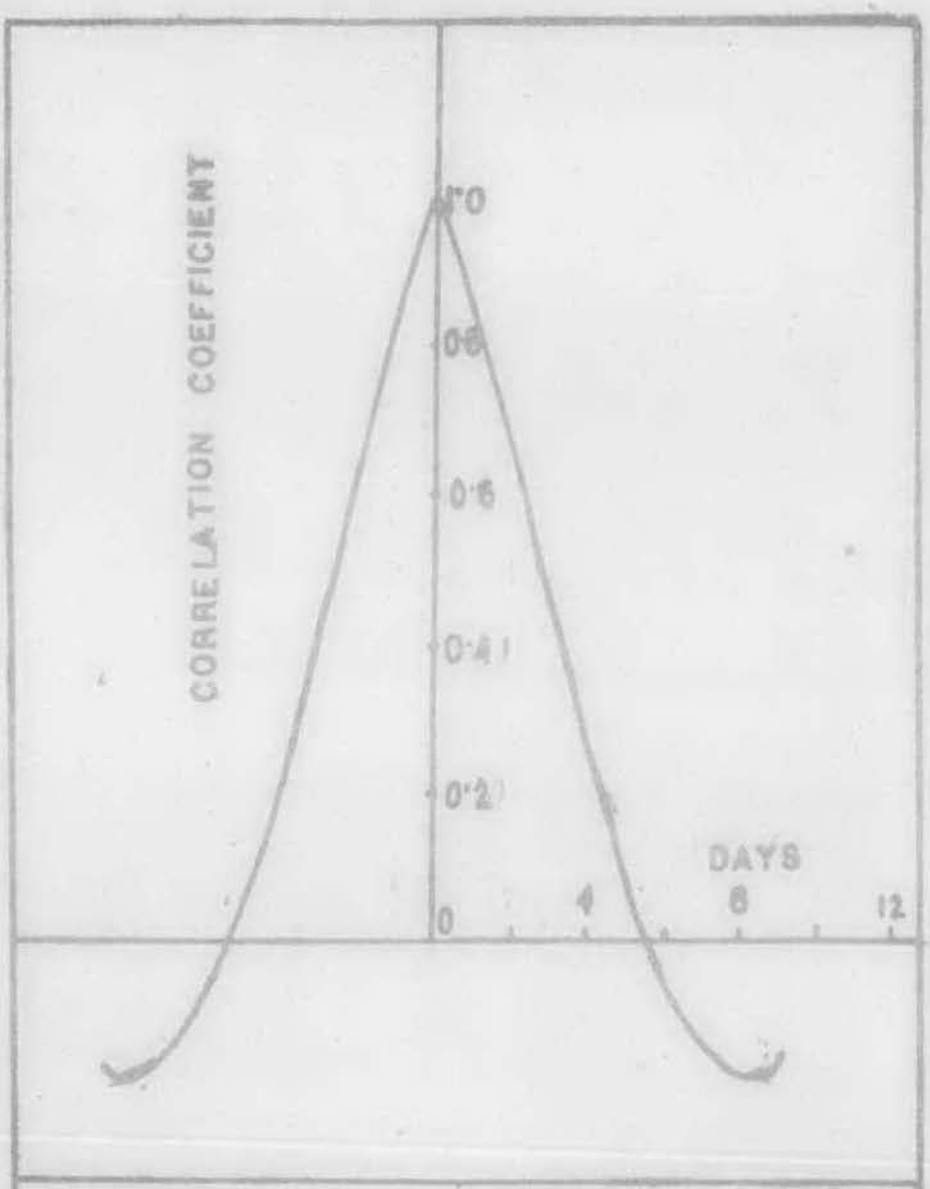


FIG 10.3.2 CORRELATION COEFFICIENT. D.O. LEVEL

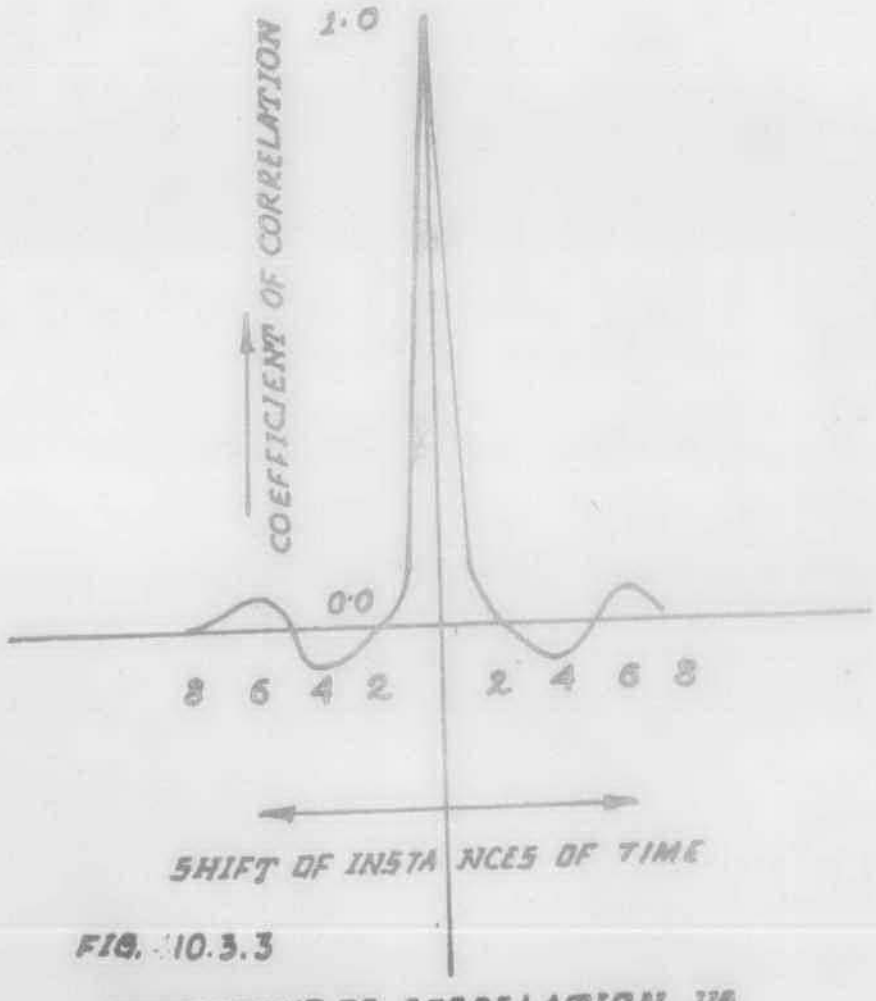


FIG. 10.3.3

COEFFICIENT OF CORRELATION VS SHIFT OF INSTANCES OF TIME FOR THE ERROR OF PREDICTION.

CHAPTER XI

LEARNING IDENTIFICATION DYNAMIC ALGORITHMS FOR MODELLING OF DISSOLVED OXYGEN INCORPORATING INTERACTION PARAMETERS

CHAPTER XI

LEARNING IDENTIFICATION DYNAMIC ALGORITHMS FOR MODELLING OF DISSOLVED OXYGEN INCORPORATING DISTRACTION PARAMETERS.

11.0 Abstract :

In this work multilayer group method of data handling algorithm and dynamic recursive least square technique of applied cybernetics have been used for obtaining dynamic descriptions of dissolved oxygen level of a non-tidal river flowing through a highly industrialised region. When tested with field data the models are found to simulate adequately the major variations of the levels of dissolved oxygen observed in the field measurements.

11.1 Introduction :

It is understandable that significant interaction exists between dissolved oxygen (DO), biochemical oxygen demand (BOD), sustained sunlight and volumetric flow rate in a river system [50, 113]. It is important, therefore to describe explicitly the effect of BOD, sustained sunlight and volumetric flow rate on the DO levels of a non tidal river. This work consists of two parts. In the first part multilayer group method of data handling algorithms have been used to obtain the mathematical description of DO. The second part demonstrates the application of dynamic least square recursive technique for modelling of dissolved oxygen levels. Two methods are discussed briefly and are duly illustrated from the data collected on the relevant variables over an 80 day period from the 6th. June 1972 to the 25th. August, 1972 from a 4.7 KM stretch of the River Cam in Eastern England at the upstream end of the stretch and at the down stream end.

11.2. Part One : Multilayer Group Method of Data Handling Algorithms :

Formation of the Process Equation :

The daily dissolved oxygen levels at the upstream end of the 4.7 KM stretch of the River Can can be represented by a general from of process equation as

$$y(k) = f(y(k-1), y(k-2), \ldots, x_1(k), x_1(k-1), \ldots, x_2(k), x_2(k-1), \ldots, x_3(k), x_3(k-1), \ldots,) \quad \ldots (11.1)$$

the subscript k, k-1, k-2, ..., refer respectively to the current day, one day preceding the current day, two days preceding the current day and so on $y(.)$, $x_1(.)$, $x_2(.)$ and $x_3(.)$ are the daily levels of DO, BOD sustained sunlight hours and the volumetric flow rate of the river respectively. The sustained sunlight effects

$I(t_k)$ in hours is defined as

$$I(t_k) = I(t_{k-1}) + \frac{1}{\tau} \left[u_3(t_k) \frac{\theta(t_k) - \bar{\theta}}{\bar{\theta}} - I(t_{k-1}) \right] \quad \ldots (11.2)$$

where $u_3(t_k)$ is the sunlight incident on the system during day t_k (hour day $^{-1}$), $\bar{\theta}$ is an arbitrary mean river water temperature, in degrees celsius (may be considered as the mean daily temperature for the duration of experiment), τ is the time constant in days (obtained as the significant instant of cross correlation with BO), $I(t_k)$ may be initialised as $\bar{I}$, the mean hourly incident of sunlight for the period of experiment.

The arguments having correlations with y(k) are then selected for inclusion in the process equation on the basis of the correlation functions for the time shift λ in days, defined as

$$\phi_{yx}(\lambda) = \frac{\sum_{i=1}^{N-\lambda} \left(y(i) - \frac{1}{N}\sum_{k=1}^{N} y(k)\right)\left(x(i+\lambda) - \frac{1}{N}\sum_{k=1}^{N} x(k)\right)}{\left[\sum_{i=1}^{N-\lambda}\left(y(i) - \frac{1}{N}\sum_{k=1}^{N} y(k)\right)^2 \sum_{j=1+\lambda}^{N}\left(x(j) - \frac{1}{N}\sum_{k=1}^{N} x(k)^2\right)\right]^{\frac{1}{2}}} \quad \ldots \quad (11.3)$$

where N is the number of data points.

After such selection of arguments as having correlation with daily dissolved oxygen levels the process equation becomes.

$$y(k) = f(y(k-1),\ y(k-2),\ x_1(k),\ x_2(k),\ x_3(k)\ ) \quad \ldots \quad (11.4)$$

or denoting these respective arguments as $y(k) = y$,

$y(k-1) = x'_1$, $y(k-2) = x'_2$ $x_1(k) = x_3'$, $x_2(k) = x_4'$

$x_3(k) = x_5'$ the process equation becomes,

$$y = f(x_1', x_2', x_3', x_4', x_5') \quad \ldots \quad (11.5)$$

First Layer of Selection :

There are now $\binom{5}{2} = 10$ possible combinations of selecting two arguments at a time out of 5. For every such combination, the partial regression equation is written as

$$y_a = \alpha_{oa} + \alpha_{1a} x'_b + \alpha_{2a} x_c' + \alpha_{3a} x_b' x_c' + \alpha_{4a} x_b'^2 + \alpha_{5a} x_c^2 \quad \ldots \quad (11.6)$$

Where a = 1, 2, ..., 10, while b and c are indices for all 10 combinations. And these lead to 10 systems of normal Gaussian equations with matrices of the order 6 x 6. The coefficients 's are then estimated by solving normal equation systems constructed from the data set. For estimating the coefficients it is assumed that the equation error is very small being distributed with zero mean, constant variance and also uncorrelated with the inputs. The Second assumption is that for the construction of the model the inputs and output are known without any measurement error.

The accuracy of every variable y_a is calculated by using the whole data set. From all variable five more accurate ones are chosen which give low values of integral square error criterion as

$$ISE = \frac{\sum_{i=1}^{N} (\text{y observed (i)} - \text{y modelled (i)})^2}{\sum_{i=1}^{N} (\text{y observed (i)})^2} \quad \ldots \quad (11.7)$$

Selection of Other Layers :

Five intermediate variables of y_a - layer chosen from the first layer give 10 combinations of two arguments of y_a - layer. Again in the second layer these become

$$z_a = \beta_{oa} + \beta_{1a} y'_b + \beta_{2a} y'_c + \beta_{3a} y'_b y'_c + \beta_{4a} y'^2_b + \beta_{5a} y'^2_c \quad \ldots \quad (11.8)$$

Where a = 1, 2, ..., 10 while b and c are indices of all 10 combinations. Calculation of the coefficients of β and estimation of the accuracy of the variables z_a are repeated as in the case of y_a.

The five z_a variables are then chosen for the next layer u_a.

$$u_a = \gamma_{oa} + \gamma_{1a} z_a' + \gamma_{2a} z_b' + \gamma_{3a} z_a' z_b' + \gamma_{4a} z_a'^2 + \gamma_{5a} z_b'^2 \quad \ldots \quad (11.9)$$

In this way each layer is tested for accuracy by using the entire data set and on the basis of minimum integral square criterion explained earlier. For all layers, variables on the left hand side of the equations are kept equal to the value of the output variable.

11.3. Illustrations :

It was observed that as the layer increases the integral square errors converge asymptotically to a very small value. To make the model amenable to practical application the integral square error of 1.621198×10^{-3} at z_2 was considered for the point of termination for formulation of the process equation. The changes of minimum integral square error for different layers are shown in Fig 11.3.2.

The DB model of River Cam at the beginning of 4.7 KM stretch has been identified by the polynomials as

$$y = z_2$$

$$z_2 = -13473993 - 8.0053672\ y_4 + 0.3045025\ y_5 + 20.0247866\ y_4 y_5 - 9.5061438\ y_4^2 - 10.5344898\ y_5^2.$$

$$y_4 = 4.9836030 - 0.6698284\ x_2' x_3' - 0.0752754\ x_2'^2 - 0.2076490\ x_3'^2$$

$$y_5 = -0.5975669 + 1.4993823\ x_3' + 0.0497549\ x'_5 + 0.0924680\ x'_3 x'_5 - 0.0806891\ x'^2_3 - 0.0727057\ x'^2_5 \ldots (11.10)$$

Fig 11.3.1. (a,b & c) show the input data. Fig 11.3.3 shows the observed and modelled values of the DO levels at the upstream end of 4.7 KM stretch of river Cam. The daily dissolved oxygen Levels at the down stream end of the 4.7 KM stretch of the river Cam was modelled in a similar way in which case the process equation was represented by

$$y(k) = f(y(k-1), y(k-2), x_1(k), x_2(k), x_3(k), x_4(k), x_4(k-1)) \quad \ldots \quad (11.10a)$$

where $x_1(.)$, $x_2(.)$, $x_3(.)$ and $x_4(.)$ are daily D.O. level at the down stream and daily D.O. level and daily BOD level. Levels at upstream end, volumetric flow and sustained sunlight hours respectively. Observed and modelled values of the D.O. Level at the down stream end is shown in Fig 11.3.4 which was obtained at the second layer with minimum integral square error of 4.754965×10^{-3}.

11.4 Part Two : Dynamic Least Squares Identification Model :

It has been observed in Part One that DO levels $y_{(k)}$ is a function of the previous output observations the autoregressive terms $y(k-1)$, $y(k-2)$ and the correlated deterministic inputs. $x_1(k)$, $x_2(k)$, $x_3(k)$.

Together with a current unknown realisation of the noise $\eta(k)$, the process equation can be represented as

$$y(k) = \beta_1 y(k-1) + \beta_2 y(k-2) + \delta_1 x_1(k) + \delta_2 x_2(k) + \delta_3 x_3(k) + \eta(k) \quad \ldots \quad (11.11)$$

The equation (11.11) can be expressed as

$$y(t_k) = z^T(t_k) \alpha + \eta(k) \quad \ldots \quad (11.12)$$

where 'α' is the parameter vector with property of slowly varying with time and amenable to recursive adaptiveness, and

$$z^T(t_k) = \left[y(t_{k-1}),\ y(t_{k-2}),\ x_1(t_k),\ x_2(t_k),\ x_3(t_k) \right]$$

The least square estimation of the process parameters can be obtained by minimising a loss function defined as the sum of the squared error as

$$J \triangleq \sum_{k=1}^{N} \left(y(t_k) - z^T(t_k)\ \alpha \right)^2 \qquad \ldots \quad (11.13)$$

The estimates $\hat{\alpha}$ of α that minimise J are called least square estimates and $\hat{\alpha}$ is obtained as

$$\hat{\alpha} = \left[\sum_{k=1}^{N} z(t_k)\, z^T(t_k) \right]^{-1} \left[\sum_{k=1}^{N} z(t_k) y(t_k) \right] \qquad (11.14)$$

Let

$$p^*(t_k) \triangleq \left[\sum_{j=1}^{k} z(t_j)\, z^T(t_j) \right]$$

and

$$b(t_k) \triangleq \left[\sum_{j=1}^{k} z(t_j)\ y(t_j) \right]$$

So the equation (11.14) becomes,

$$\hat{\alpha} = p^*(t_k)\ b(t_k) \qquad \ldots \quad (11.15)$$

The recursive relationship for p^* and $b(.)$ can be set as

$$\left[p^*(t_k) \right]^{-1} = \sum_{j=1}^{k-1} z(t_j)\, z^T(t_j) + z(t_k)\, z^T(t_k)$$

$$= \left[p^*(t_{k-1}) \right]^{-1} + z(t_k)\, z^T(t_k) \qquad \ldots \quad (11.16)$$

Similarly,

$$b(t_k)=b(t_{k-1})+z(t_k)y(t_k) \qquad \ldots \quad (11.17)$$

Premultiplying by $P^*(t_k)$ we get from equation (11.16)

$$I=P^*(t_k)\left[P^*(t_{k-1})\right]^{-1}+P^*(t_k)\,z(t_k)z^T(t_k) \qquad \ldots \quad (11.18)$$

Post multiplication of equation (11.18) by $P^*(t_{k-1})$

$$P^*(t_{k-1})=P^*(t_k)+P^*(t_k)z(t_k)z^T(t_k)P^*(t_{k-1}) \qquad \ldots \quad (11.19)$$

Post multiplication of equation (11.19) by $z(t_k)$

$$P^*(t_{k-1})z(t_k) = P^*(t_k)z(t_k)z(t_k)+P^*(t_k)z^T(t_k)P^*(t_{k-1})z(t_k)$$

So we get,

$$P^*(t_{k-1})z(t_k)=P^*(t_k)z(t_k)\left[1+z^T(t_k)P^*(t_{k-1})z(t_k)\right]^{-1}z^T(t_k)P^*(t_{k-1})$$

The equation (11.19) becomes,

$$P^*(t_{k-1})z(t_k)\left[1+z^T(t_k)P^*(t_{k-1})z(t_k)\right]^{-1}z^T(t_k)P^*(t_{k-1})$$

$$= P^*(t_k)z(t_k)z^T(t_k)P^*(t_{k-1}) \qquad \ldots \quad (11.21)$$

Substitution of equation (11.19) in equation (11.21) gives,

$$P^*(t_k)=P^*(t_{k-1})-P^*(t_{k-1})z(t_k)\left[1+z^T(t_k)P^*(t_{k-1})z(t_k)\right]^{-1}z^T(t_k)P^*(t_{k-1})$$

$$\ldots \quad (11.22)$$

From equation (11.15)

$$\hat{\alpha}(t_k) = P^*(t_k)\,b(t_k)$$

Combining equations (11.17) and (11.22)

$$\hat{\alpha}(t_k)=\left[P^*(t_{k-1})-P^*(t_{k-1})z(t_k)\left[1+z^T(t_k)P^*(t_{k-1})z(t_k)\right]^{-1}\right.$$

$$\left.\times z^T(t_k)P^*(t_{k-1})\right]\left\{b(t_{k-1})+z(t_k)y(t_k)\right\} \qquad \ldots \quad (11.23)$$

Since

$$P^{e}(t_{k-1})\, b(t_{k-1}) = \hat{\alpha}(t_{k-1})$$

The equation (11.23) becomes,

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) - P^{e}(t_{k-1})\, z(t_k) \left[1 + z^{T}(t_k) P^{e}(t_{k-1}) z(t_k) \right]^{-1}$$
$$\times z^{T}(t_k) \hat{\alpha}(t_{k-1}) + P^{e}(t_{k-1})\, z(t_k)\, y(t_k)$$
$$- P^{e}(t_{k-1})\, z(t_k) \left[1 + z^{T}(t_k) P^{e}(t_{k-1}) z(t_k) \right]^{-1} \ldots (11.24)$$
$$z^{T}(t_k) P^{e}(t_{k-1}) z(t_k) \; y(t_k)$$

Since $z^{T}(t_k)\, P^{e}(t_{k-1})\, z(t_k)$ is scalar the equation (11.24) be comes,

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) + P^{e}(t_{k-1})\, z(t_k) \left[1 + z^{T}(t_k) P^{e}(t_{k-1}) z(t_k) \right]^{-1}$$
$$\times \left[y(t_k) - z^{T}(t_k) \hat{\alpha}(t_{k-1}) \right] \quad \ldots \quad (11.25)$$

The equations (11.25) and (11.22) with

$$P^{e}(t_k) = \left[\sum_{j=1}^{k} z(t_j)\, z^{T}(t_j) \right]^{-1}$$

are the recursive least square algorithms

The recursive least square algorithms do not overcome problem of bias. The problem of bias in the parameter estimates may be defined as follows.

Recalling equation (11.14)

$$\hat{\alpha}(t_k) = \left[\sum_{j=1}^{k} z(t_j) z^{T}(t_j) \right]^{-1} \left[\sum_{j=1}^{k} z(t_j) y(t_j) \right]$$

Upon substitution of the process equation.

$$y(t_j) = z^{T}(t_j)\, \alpha + \eta(t_j) \quad \text{in equation (11.14)}$$

$$\hat{\alpha}(t_k) = \alpha + \left[\sum_{j=1}^{k} s(t_j)\, s^T(t_j)\right]^{-1} \left[\sum_{j=1}^{k} s(t_j)\, \gamma(t_j)\right] \quad \ldots (11.26)$$

for $\hat{\alpha}(t_k)$ to be unbiased

$$E\left\{s(t_j)\ \gamma(t_j)\right\} \simeq 0 \quad \text{for all } j$$

and

$$E\left\{\gamma(t_k)\ \gamma(t_j)\right\} \simeq 0 \quad \text{for } k \neq j$$

and

$$E\left\{\gamma(t_k)\ \gamma(t_k)\right\} = \sigma^2$$

i.e. $\gamma(t_k)$ approaches the white noise sequence $v(t_k)$

Let us define the covariance matrix of the least square parameter estimation as $P(t_k)$

$$P(t_k) = \text{Cov}\left\{\tilde{\alpha}(t_k)\right\} = E\left\{(\hat{\alpha}(t_k) - \alpha)(\hat{\alpha}(t_k) - \alpha)^T\right\}$$

$$= P_\theta(t_k)\left[E\left\{\left[\sum_{j=1}^{k} s(t_j)\ \gamma(t_j)\right]\left[\gamma(t_j)\, s^T(t_j)\right]\right\}\right] P_\theta^T(t_k)$$

or $$P_\theta(t_k) = \sigma^2\, P_\theta(t_k) \quad \ldots (11.27)$$

σ^2 may be estimated as the variance of the residual error sequence $E(t_k) = y(t_k) - s^T(t_k)\ \hat{\alpha}(t_k)$ after the estimates have achieved satisfactory convergence. Least square estimates of α may be considered for initialisation. Thus for almost unbiased estimates the least square recursive algorithms of (11.22) and (11.25) have been transformed by making substitution of equation (11.27) as

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) + P(t_{k-1})z(t_k)\left[\sigma^2 + z^T(t_k)P(t_{k-1})z(t_k)\right]^{-1}$$

$$\left\{ y(t_k) - z^T(t_k)\ \hat{\alpha}(t_{k-1}) \right\} \qquad \ldots \quad (11.28)$$

$$P(t_k) = P(t_{k-1}) - P(t_{k-1})\, z(t_k)\left[\sigma^2 + z^T(t_k)P(t_{k-1})\, z(t_k)\right]^{-1}$$

$$\times\ z(t_k)\, P(t_{k-1}) \qquad \ldots \quad (11.29)$$

In case of time varying parameters the large errors between the observed and the modelled outputs obtained from the algorithms (11.28) and (11.29) are mainly due to the error in choosing values of the parameters $\hat{\alpha}$ of the model. If the parameters variation laws are known a model for the parameter variation can be obtained and the algorithms of equations (11.28) and (11.29) can be updated accordingly. In the absence of any specific parameter variation law a random walk model may be postulated as

$$\alpha(t_k/t_{k-1}) = \alpha(t_{k-1}/t_{k-1}) + \zeta(t_k/t_{k-1}) \qquad \ldots \quad (11.30)$$

Recalling equation (11.27)

$$P(t_k/t_{k-1}) = \mathrm{Cov}\left\{ \tilde{\alpha}(t_k/t_{k-1}) \right\}$$

$$= E\left\{ \left(\alpha(t_k/t_{k-1}) - \alpha\right)\left(\alpha(t_k/t_{k-1}) - \alpha\right)^T \right\}$$

Hence, from equation (11.30)

$$P(t_k/t_{k-1}) = E\left\{ \left(\alpha(t_{k-1}/t_{k-1}) + \zeta(t_k/t_{k-1}) - \alpha\right)^T \right. \qquad (11.31)$$

$$\left. \left(\alpha(t_{k-1}/t_{k-1}) + \zeta(t_{k-1}/t_{k-1}) - \alpha\right)^T \right\}$$

From equation (11.26)

$$P(t_k/t_{k-1}) = E\left\{ \left(\left[\sum_{j=1}^{k-1} z(t_j) z^T(t_j) \right]^{-1} \left[\sum_{j=1}^{k-1} z(t_j)\, \eta(t_j) \right] \right. \right.$$

$$\left. \left. + \zeta(t_{k-1}/t_{k-1}) \right) \times \right.$$

$$\left(\left[\sum_{j=1}^{k-1} x(t_j)x^T(t_j) \right]^{-1} \left[\sum_{j=1}^{k-1} x(t_j)\ \gamma(t_j) \right]^{-1} + \zeta(t_k/t_{k-1}) \right)^T \Bigg\}$$

since $\zeta(t_k/t_{k-1})$ is a white noise process it is uncorrelated with $x(.)$, and $\gamma(t_j)$ is scalar,

$$P(t_k/t_{k-1}) = \left[\sum_{j=1}^{k-1} x(t_j)x^T(t_j) \right]^{-1} \left[\sum_{j=1}^{k-1} x(t_j) E\left\{ \sum_{j=1}^{k} \gamma(t_j)\ \gamma^T(t_j) \right\} \right.$$

$$\left. \sum_{j=1}^{k-1} x^T(t_j) \right]^{-1} \times$$

$$\left[\left[\sum_{j=1}^{k} x(t_j)x^T(t_j) \right]^{-1} \right]^T + E\left\{ \zeta(t_k/t_{k-1})\ \zeta^T(t_k/t_{k-1}) \right\} \qquad (11.32)$$

Considering $D = E\left\{ \zeta(t_k/t_{k-1})\ \zeta^T(t_k/t_{k-1}) \right\}$ ... (11.33)

$$P(t_k/t_{k-1}) = \sigma^2 P x(t_{k-1}/t_{k-1}) + D \qquad \ldots \quad (11.34)$$

$$P(t_k/t_{k-1}) = P(t_{k-1}/t_{k-1}) + D \qquad \ldots \quad (11.35)$$

and

$$\hat{\alpha}(t_k/t_{k+1}) = \hat{\alpha}(t_{k-1}/t_{k-1}) \qquad \ldots \quad (11.36)$$

Algorithms (11.28) and (11.29) are updated as

$$\hat{\alpha}(t_k/t_k) = \hat{\alpha}(t_k/t_{k-1}) + P(t_k/t_{k-1})\, x(t_k) \left[\sigma^2 + x^T(t_k) P(t_k/t_{k-1}) \right.$$

$$\left. x(t_k) \right]^{-1} \times \left\{ y(t_k) - x^T(t_k)\ \hat{\alpha}(t_k/t_{k-1}) \right\} \qquad \ldots \quad (11.37)$$

$$P(t_k/t_k) = P(t_k/t_{k-1}) - P(t_k/t_{k-1}) x(t_k) \left[\sigma^2 + x^T(t_k) P(t_k/t_{k-1}) x(t_k) \right]^{-1}$$

$$\times\ x^T(t_k)\, P(t_k/t_{k-1}) \qquad \ldots \quad (11.38)$$

Combining equations (11.35) and (11.36) with (11.37) and (11.38) the dynamic recursive least square algorithms are

$$\hat{\alpha}(t_k/t_k) = \hat{\alpha}(t_{k-1}/t_{k-1}) + \left\{ P(t_{k-1}/t_{k-1}) + D \right\} \quad x(t_k)$$

$$\left[\sigma^2 + x^T(t_k) \left\{ P(t_{k-1}/t_{k-1}) + D \right\} x(t_k) \right]^{-1}$$

$$x \left\{ y(t_k) - x^T(t_k) \hat{\alpha}(t_{k-1}/t_{k-1}) \right\} \quad \ldots \quad (11.39)$$

$$P(t_k/t_k) = \left\{ P(t_k/t_{k-1}) + D \right\} - \left\{ P(t_{k-1}/t_{j-1}) + D \right\} \quad x(t_k)$$

$$x \left[\sigma^2 + x^T(t_k) \left\{ P(t_{k-1}/t_{k-1}) + D \right\} x(t_k) \right]^{-1} x^T(t_k) \left\{ P(t_{k-1}/t_{k-1}) + D \right\}$$

$$\ldots \quad (11.40)$$

Retaining the simpler notations the equations (11.39) and (11.40) become

$$\hat{\alpha}(t_k) = \hat{\alpha}(t_{k-1}) + \left\{ P(t_{k-1}) + D \right\} x(t_k) \left[\sigma^2 + x^T(t_k) \left\{ P(t_{k-1}) + D \right\} x(t_k) \right]^{-1}$$

$$x \left\{ y(t_k) - x^T(t_k) \hat{\alpha}(t_{k-1}) \right\} \quad \ldots \quad (11.41)$$

$$P(t_k) = \left\{ P(t_{k-1}) + D \right\} - \left\{ P(t_{k-1}) + D \right\} x(t_k) \left[\sigma^2 + x^T(t_k) \left\{ P(t_{k-1}) + D \right\} x(t_k) \right]^{-1}$$

$$x \; x^T(t_k) \left\{ P(t_{k-1}) + D \right\} \quad \ldots \quad (11.42)$$

The dynamic recursive least square algorithms are initialised with least square estimates of parameter vector $\alpha(.)$.

11.5 Illustration :

Fig 11.5.1a shows the dissolved oxygen level at upstream end of 4.7 KM stretch of the river Cam in Eastern England collected over a 80 day period in summer of 1972. Fig 11.5.1b shows the predicted values of D.O. levels (dissolved oxygen levels) obtained with dynamic least square recursive estimation algorithms. The corresponding integral square error defined as

$$ISE = \frac{\sum_{i=1}^{N} (y(t_i) - y_o(t_i))^2}{\sum_{i=1}^{N} (y(t_i))^2}$$

is 2.739772671 E-03 and the mean error is 3.59892609 E-03, where $y_o(.)$ are the modelled values of dissolved oxygen levels. The D.O. level y(k) at the down stream end was taken as a function of the previous output observations y(k-1), y(k-2) and the correlated deterministic inputs $x_1(k)$, $x_2(k)$, $x_3(k)$, $x_4(k)$ and $x_4(k-1)$ as given in equ. 11.11 Fig 11.5.2a shows dissolved oxygen but at the down stream end of 4.7 KM stretch of the river Cam in Eastern England . Collected over 80 day period in Summer 1972. Fig 11.5.2b shows the predicted values of D.O. level obtained by dynamic least square error.

11.6 Conclusion :

In both the methods the modelled data are found in close agreement with the observed ones. The mathematical models based on learning identification algorithms and dynamic recursive least square technique highlight the possibilities of using real time regulatory mechanism for controlling water quality of a non-tidal river.

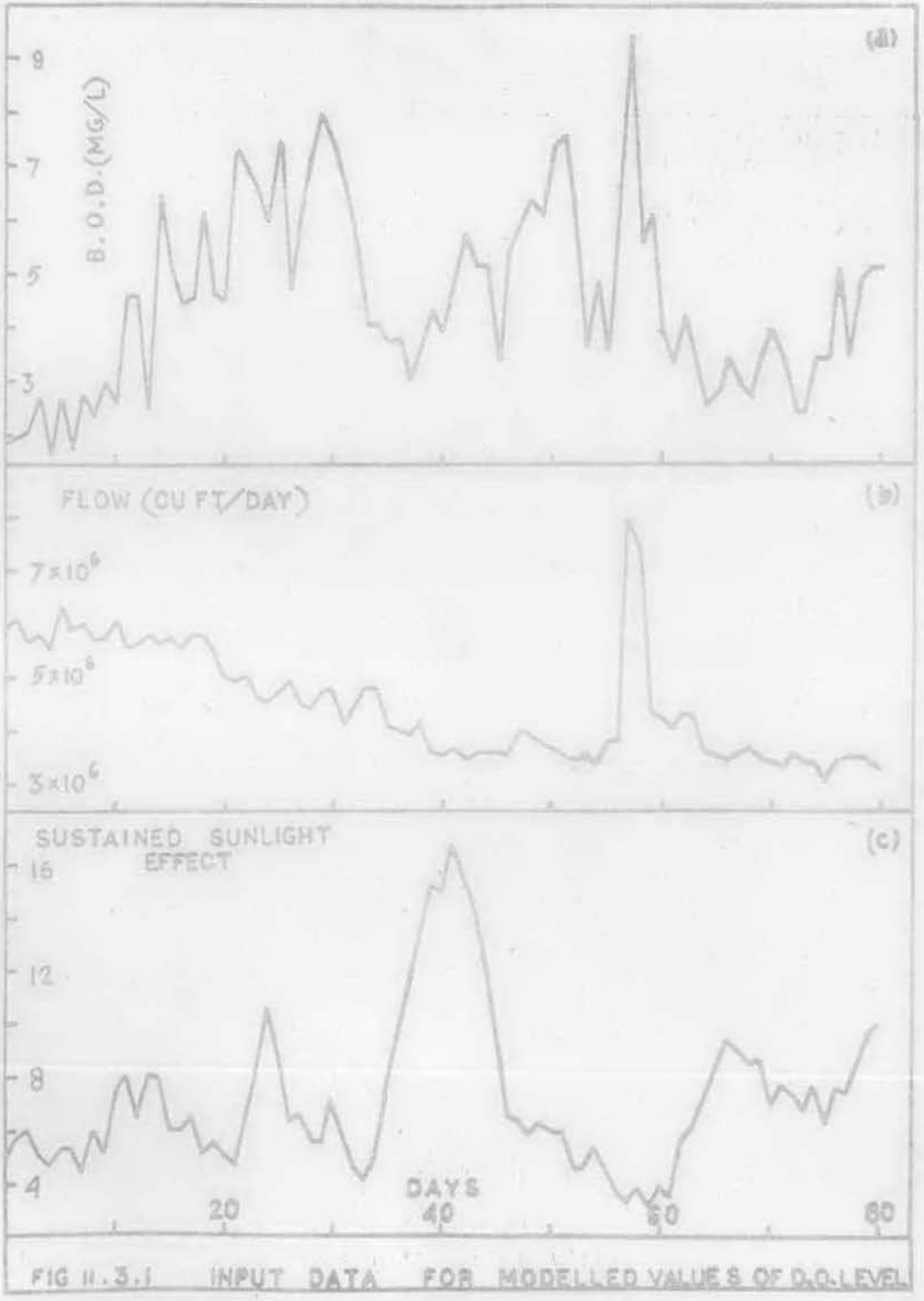
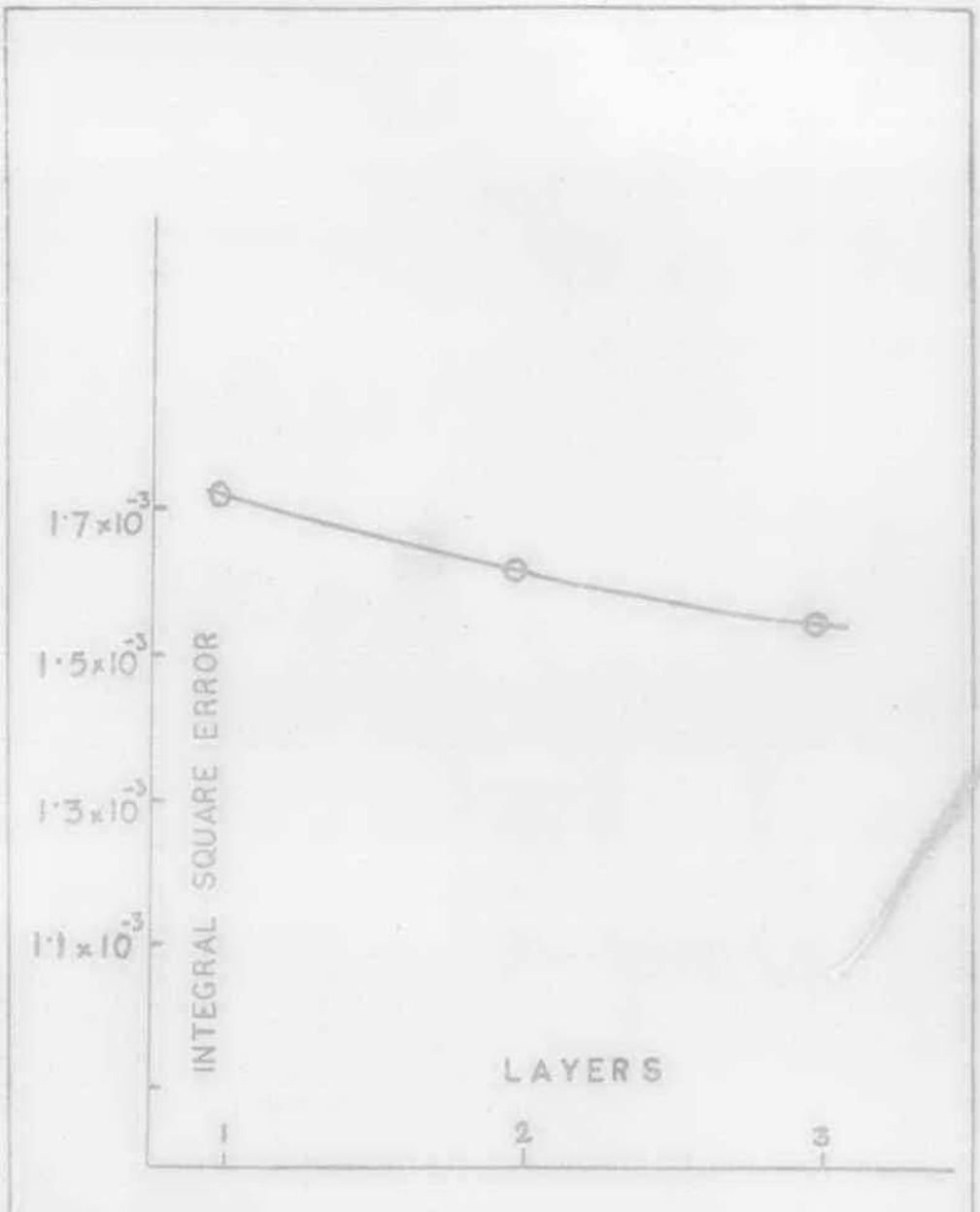


FIG II.3.1 INPUT DATA FOR MODELLED VALUES OF D.O. LEVEL



INTEGRAL SQUARE ERROR IN DIFFERENT LAYERS

FIG 11.3.2

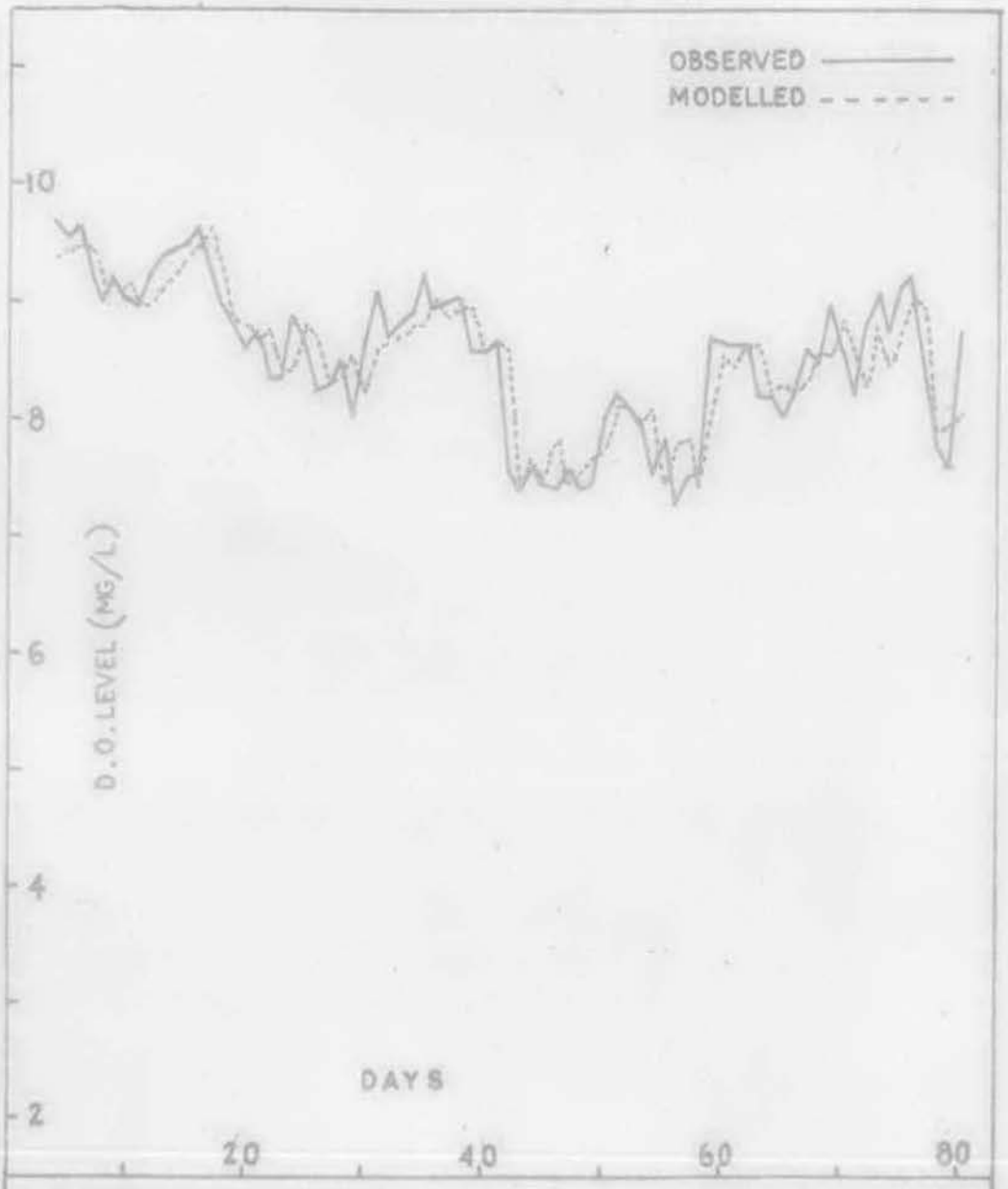


FIG II.3.3 OBSERVED AND MODELLED VALUES OF D.O.LEVEL

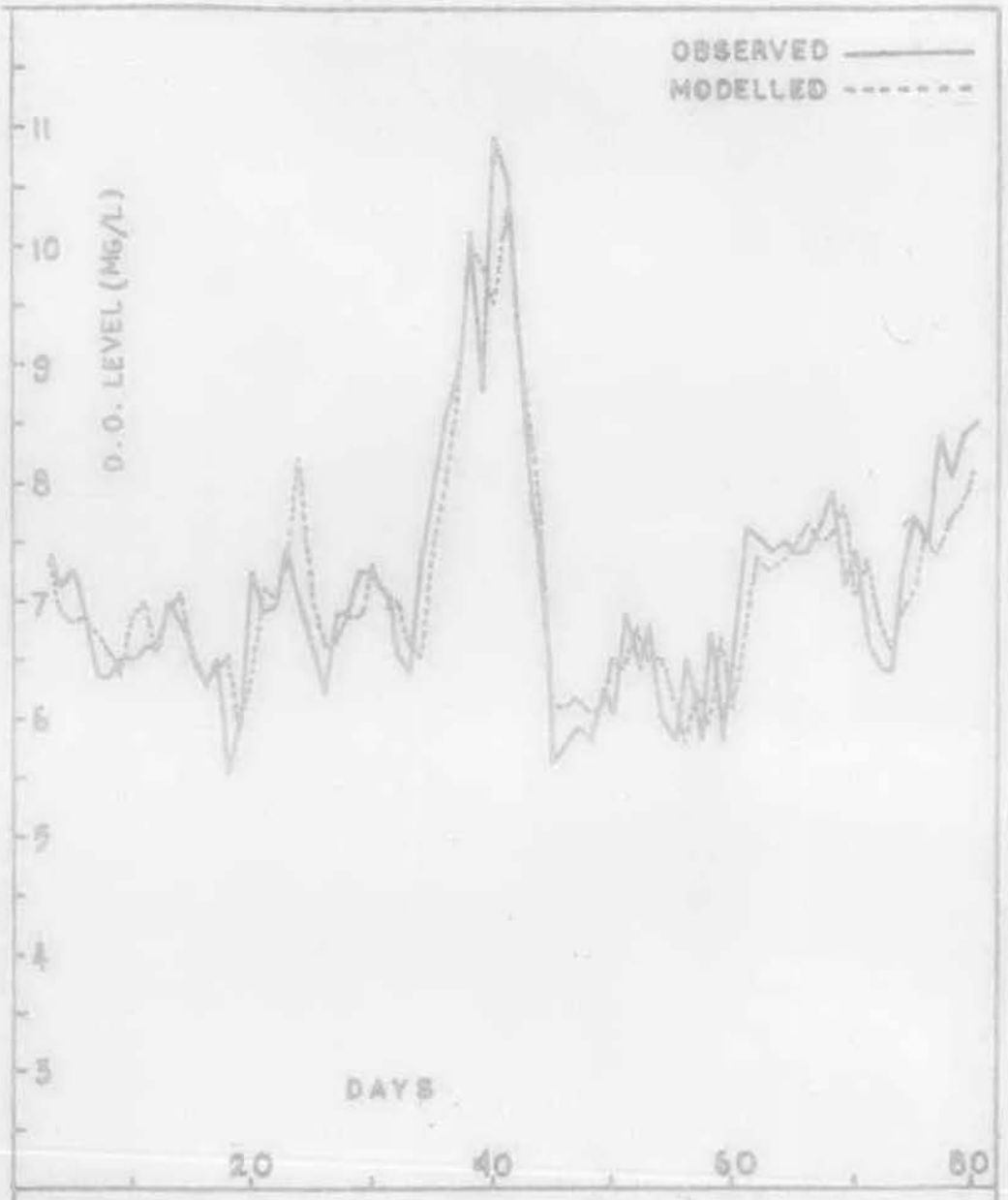


FIG 11.3.4 OBSERVED AND MODELLED VALUES OF D.O. LEVEL

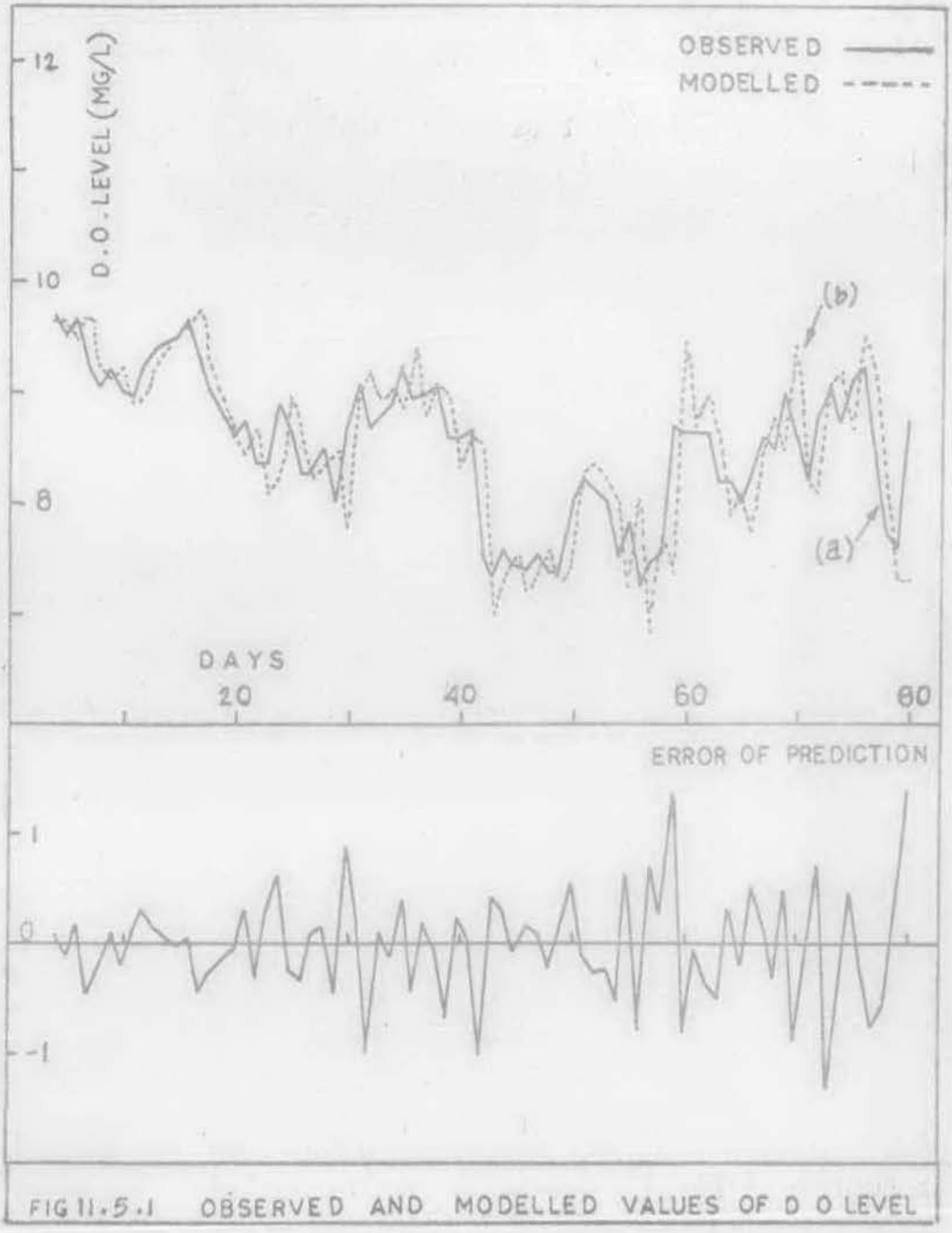


FIG 11.5.1 OBSERVED AND MODELLED VALUES OF D O LEVEL

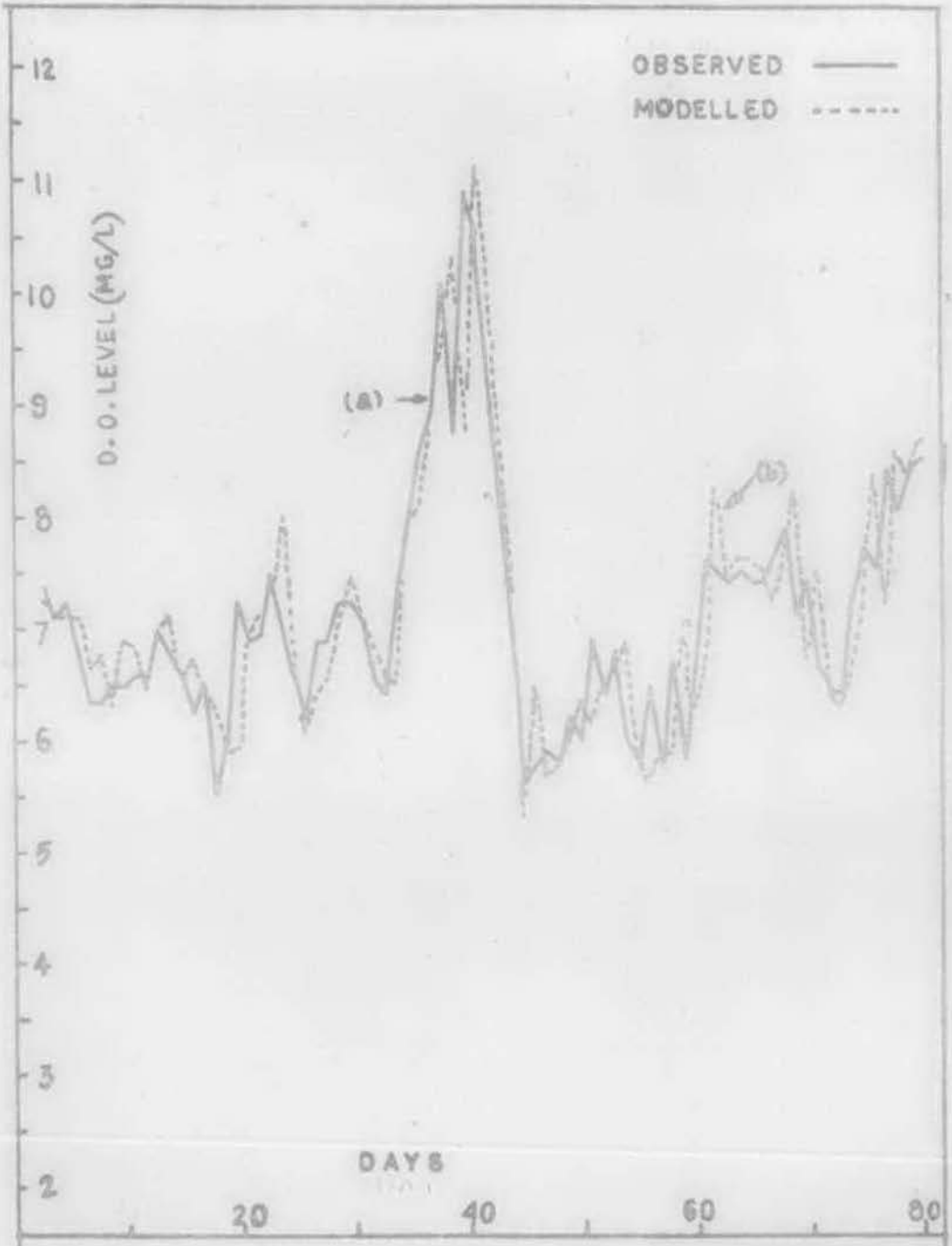


FIG 11.5.2 OBSERVED AND MODELLED VALUES OF D.O.LEVEL

CHAPTER XII

CONCEPT OF FUZZY CONTROL AND A SUGGESTIVE EXAMINATION OF ITS APPLICATION IN THE WATER QUALITY MANAGEMENT OF A NON TIDAL RIVER

CHAPTER XII

CONCEPT OF FUZZY CONTROL AND A SUGGESTIVE EXAMINATION OF ITS APPLICATION IN THE WATER QUALITY MANAGEMENT OF A NON-TIDAL RIVER.

12.0 Abstract :

This chapter offers a conceptual examination of the possibilities of fuzzy control application for water quality management of a non-tidal river. In the present work a systematic approach of fuzzy control has been introduced rather as an analytical tool than as an immediately applicable closed loop control strategy.

12.1 Introduction :

In this work the application of fuzzy control techniques to the problem of combined real time control of river discharge and water quality of a river has been considered. It offers a speculative examination of the possibilities of fuzzy control application in the operation of river system management.

A few cases of real time control of water quality of a non-tidal river have been demonstrated in practice. The most important of the reasons for this lack of practical application is "a lack of clearly defined objectives for the management of water quality in a river basin and a lack of precisely specified standards of stream quality to be maintained by the application of real time control." In the present work a systematic approach incorporating fuzzy control has been introduced rather as an analytical tool than as an immediately applicable closed loop control strategy. The approach has been briefly outlined in the following section.

The success of the mechanistic system to unravel the secrets of the physical world has enabled us to build better and better machines to harness the natural resources for the benefit of mankind. This has been possible because the advent of the computer age has stimulated a rapid expansion in the use of quantitative techniques for the analysis of physical systems. But in the absence of clearly defined objective or of imprecisely defined objectives the conventional quantitative techniques of cybernetics are unsuitable or are unable to give any tangible results. As the complexity of a system increases the imprecision in objective creeps in. And this is defined as the fuzziness in system behaviour.

Viewed in this perspective, the traditional techniques of system analysis are not well suited for dealing with complex system with imprecisely defined objectives because they fail to come to the grips of the fuzziness of the system's behaviour. To deal with such system in realistic sense we need approaches which employ methodological framework quite tolerant of imprecision in objectives, rather than rigorous mathematical formulation.

Three main distinguishing features which define the fuzzy approach are :-

(i) use of linguistic variables in place of or in addition to numerical variables, (ii) characterisation of simple relations between variables by conditional fuzzy statements; and (iii) characterisation of complex relation by fuzzy algorithms.

A linguistic variable is defined as a variable whose values are a collection of words in a natural or artificial language. To illustrate, the values of the fuzzy variable height might be expressed as tall, not tall, somewhat tall, very tall, not very tall, etc. Thus height is a linguistic variable. The linguistic variables provide systematic means for an approximate characterisation of complex system with ill defined objectives.

In quantitative approach to system theory, a dependence between two numerically valued variables x and y is usually characterised by a conditional statement as IF x is 5 THEN y is 8, IF x is 10 THEN y is 16 etc. If x and y are fuzzy variables the fuzzy conditional statements are of the form IF x is LARGE THEN y is Small.

A fuzzy algorithm is an ordered sequence of instructions which may contain fuzzy conditional statements, e.g. Reduce x slightly if y is large. Increase x very slightly if y is not very large and not very small. If x is small then stop; otherwise increase x by 2. The execution of such instruction is governed by the compositional rule of inference.

Let X be a space of points with generic elements of X denoted by x, $X = \{x\}$

A fuzzy set A in X is characterised by a membership function $f_A(x)$ which associates with each point in X a real number in the interval $[0, 1]$ with the value of $f_A(x)$ at x representing the grade of membership of x in A.

Union of two fuzzy sets A and B with respective membership functions $f_A(x)$ and $f_B(x)$ is a fuzzy set C, written as $C = A \cup B$ whose membership function is related to those of A and B

$$f_C(x) = \max [f_A(x), f_B(x)], \quad x \in X$$

or, in abbreviated from

$$f_C = f_A \vee f_B$$

Intersection of two fuzzy sets A & B with respective membership function $f_A(x)$ and $f_B(x)$ is a fuzzy set C, written as $C = A \cap B$, whose membership function is related to those of A & B by $f_C(x) = \min [f_A(x), f_B(x)], \quad x \in X,$ or, in abbreviated from

$$f_c = f_A \wedge f_B$$

Other properties of fuzzy control have been discussed in references 91 to 96.

12.2. The Concept of Fuzzy Control in Water Quality Management :

The idea of using fuzzy control as a means of water quality management of a river system is due to Beak (98).

Let us define three fuzzy sets A, B, C as

A = DO concentration much less than 4 gm^{-3}

B = Satisfactory DO concentrations

C = High DO concentrations

Fig 12.2.1 shows a membership function μ_D (.) which expresses the degree of membership of any given DO concentration in the fuzzy set of DO concentration.

With a suitable discretization of the continuum of DO concentrations the membership functions of Fig 12.2.1 can be approximated as shown in Table 12.2.1.

To illustrate the principal features of fuzzy control algorithm let us consider a river system as shown in Fig 12.2.2. Control decisions can be implemented through the setting of reservoir releases and through effluent discharge manipulation. A suitable fuzzy control algorithm is depicted in Fig 12.2.3. Let us assume the reservoir release set as same as Table 12.2.1. and is depicted in Table 12.2.2.

The effect of reservoir release on the DO concentration is a fuzzy relation $P \triangleq R \circ S$

where $R = \left\{ \mu_D(.) \right\}$ and $S = \left\{ \mu_r(.) \right\}$

The composition of fuzzy relations R and S is fuzzy relation P and is defined as V $\left\{ \wedge \text{ of cartesan product of R and S} \right\}$ of Supremum $\left\{ \text{minimum of cartesan product terms on R and S} \right\}$ i.e.

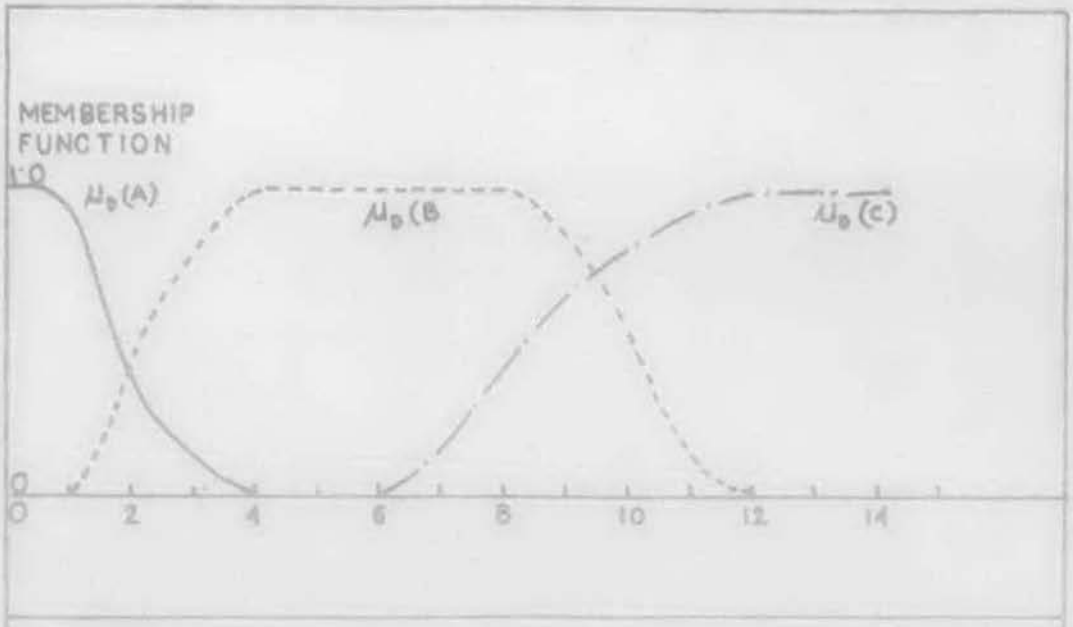


FIG 12.2.1 DISSOLVED OXYGEN CONCENTRATION (GM^{-3})

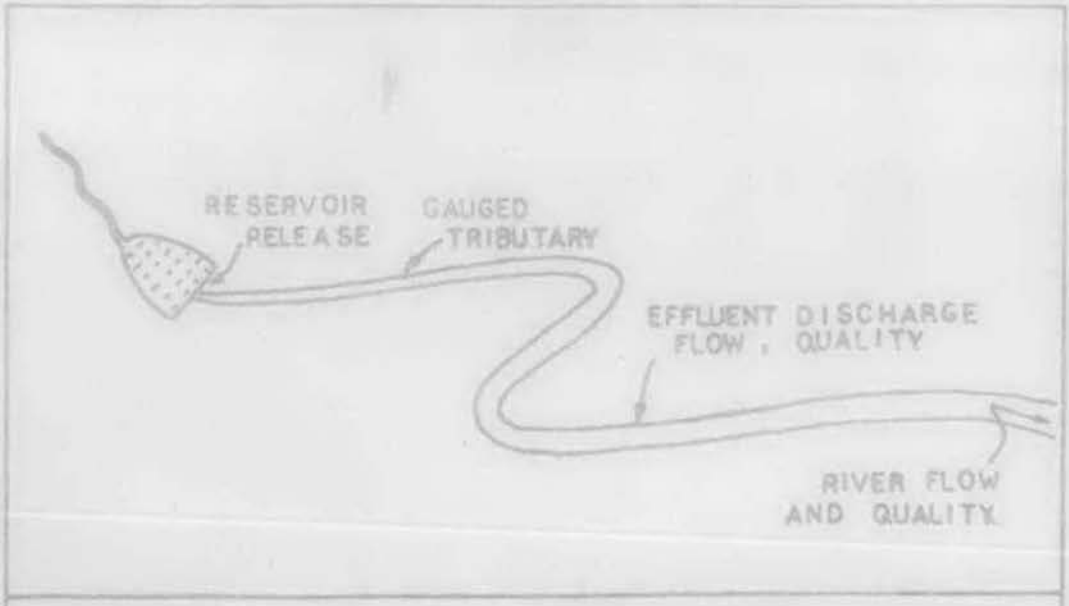


FIG 12.2.2. A RIVER SYSTEM

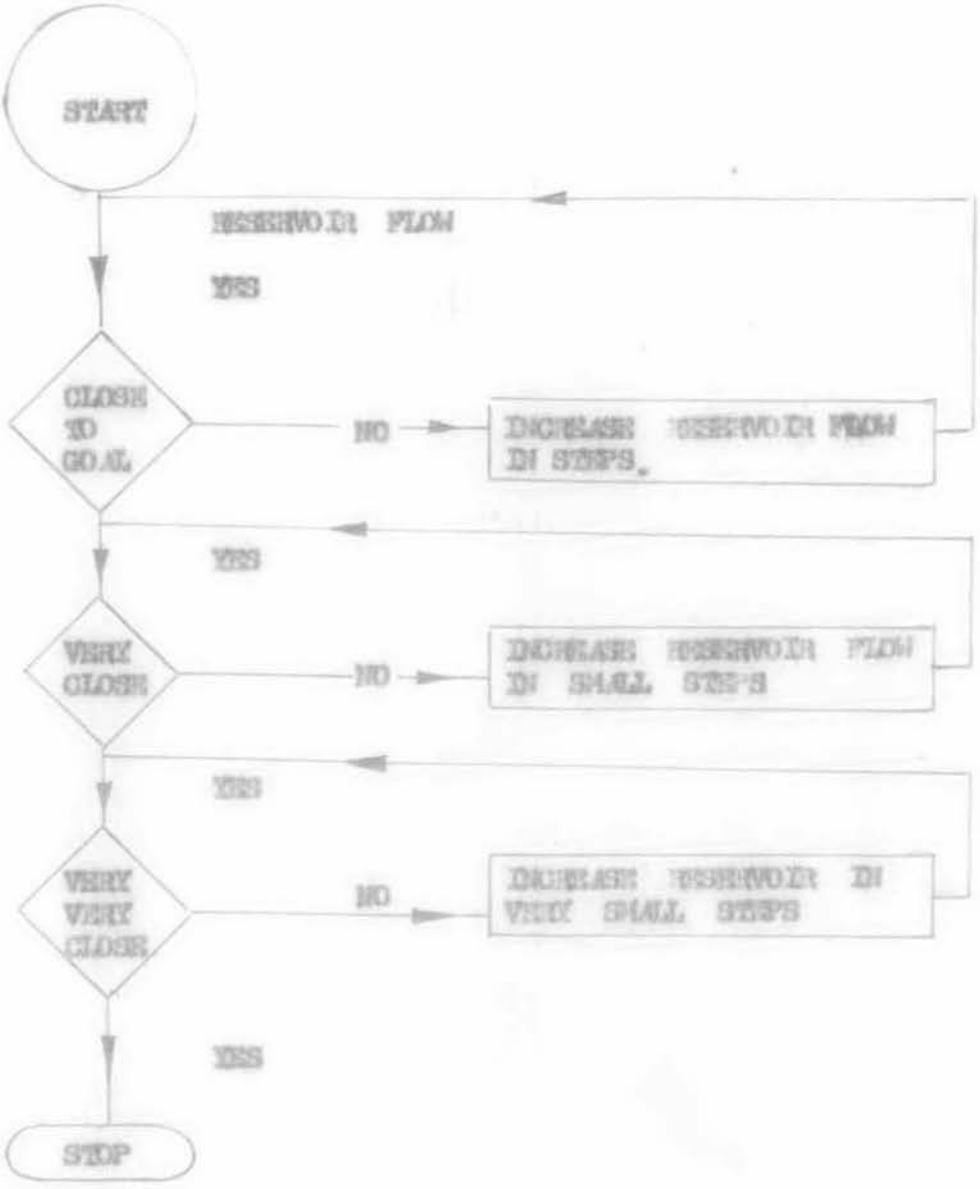


Fig 12.2.3. FUZZY CONTROL ALGORITHM.

Table 12.2.1 Dissolved Oxygen Concentration.

DO concn (gm^{-3})	2.0	2.0-3.0	3.0-4.0	4.0-6.0	6.0-8.0	8.0-10.0	10.0-12.0	12.0
$D^{(A)}$	1.0	0.7	0.2	0	0	0	0	0
$D^{(B)}$	0	0.3	0.6	1.0	1.0	0.7	0.1	0
$D^{(C)}$	0	0	0	0	0.2	0.8	0.9	1.0

Table 12.2.2 Fuzzy set deffination for reservoir release.

Reservoir release $m^3 s^{-1}$	2.0	2.0-3.0	3.0-4.0	4.0-6.0	6.0-8.0	8.0-10.0	10.0-12.0	12.0
$r^{(X)}$	1.0	0.7	0.2	0	0	0	0	0
$r^{(X)}$	0	0.3	0.6	1.0	1.0	0.7	0.1	0
$r^{(Z)}$	0	0	0	0	0.2	0.8	9	1.0

$$\text{Sup. min} \begin{bmatrix} 1.0 & 0.7 & 2.0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0.6 & 1.0 & 1.0 & 0.7 & 0.1 & 0 & \\ 0 & 0 & 0 & 0 & 0.2 & 0.8 & 0.9 & 1.0 & \end{bmatrix} \begin{matrix} 1.0 & 0 & 0 \\ 0.7 & 0.3 & 0 \\ 0.2 & 0.6 & 0 \\ 0 & 1.0 & 0 \\ 0 & 1.0 & 0.2 \\ 0 & 1.7 & 0.8 \\ 0 & 0.1 & 0.9 \\ 0 & 0 & 1.0 \end{matrix}$$

$$= \begin{bmatrix} 1.0 & 0.3 & 0 \\ 0.3 & 1.0 & 0.8 \\ 0 & 0.7 & 1.0 \end{bmatrix}$$

The nine discrete points of fuzzy composition ROS is defined as

ROW	x	COLM	VALUE		
1	x	1	1.0	=	Reservoir release $< 2.0\ m^{3-1}s$ DO $2.0\ gm^{-3}$ in set A.
1	x	2	0.3	=	Reservoir release $3.0 - 4.0\ m^3\ s^{-1}$ DO approximate $3.0\ gm^{-3}$ in set A
1	x	3	0.0	=	Reservoir release $> 4\ m^3\ s^{-1}$ DO $4\ gm^{-3}$ in Set A
2	x	1	0.3	=	Reservoir release $< 2.0\ m^3\ s^{-1}$ DO 2.0 to $3.0\ gm^{-3}$ in set B.
2	x	2	1.0	=	Reservoir release $4.6\ m^3\ s^{-1}$ DO $4.6\ gm^{-3}$ in set B
2	x	3	0.8	=	Reservoir release $8.10\ m^3\ s^{-1}$ DO approx. $8\ gm^{-3}$ in set B

3 x 1	0.8	=	Reservoir release 0 $m^3 s^{-1}$ DO concentration 2.0 - 6.0 gm^{-3}
3 x 2	0.7	=	Reservoir release 8.10 x $m^3 s^{-1}$ DO approx 8 gm^{-3}
3 x 3	1.0	=	Reservoir release $>$ 12 $m^3 s^{-1}$ DO $>$ 12 gm^{-3}

12.3 Conclusion :

In the preceeding discussion an approach has been sketched for water quality management of a river system with imprecise and ill defined objectives. Speculation has been centered on a realistic and optimistic application of fuzzy control. It is emphasized that further investigational efforts are needed to implement fuzzy control algorithms in practice to synthesise a fuzzy controller.

Fuzzy set theory, originally developed by Zadeh, deals with processes containing uncertainties only qualitatively described. The proposed model to describe such processes is based on the association of a membership function $\mu(A)$ to each set A of imprecisely defined quantities and the use of logic operations on the measures of membership for fuzzy reasoning. Because of the qualitative nature of the processes involved, linguistic methods developed for advanced computer software are also applicable to fuzzy analysis and decision making. Its adaptability and simplicity of implementations on a digital computer using the linguistic version of fuzzy set theory make it a very promising approach to the problem of control with uncertainties [113] .

CHAPTER XIII

CONCLUSION AND SCOPE OF FUTURE WORK

CHAPTER XIII

CONCLUSION AND SCOPE OF FURTHER WORK

The present investigation "Cybernetics in Identification, Modelling and Prediction of Flows and Water Quality of Non-tidal Rivers" demonstrates the practical implementability of different methods of applied cybernetics and pattern recognition such as multilayer and combinatorial group methods of data handling algorithms of self-organisation technique, learning identification technique, least square recursive algorithms with on line adaptiveness with parameter variations, recursive instrument variable algorithms, dynamic least square recursive algorithms and fuzzy control technique. Whenever necessary modification and improvement of the existing analytical techniques have been suggested for practical implementation. The computational procedures have been developed in well organised programs packages in high level language.

There are several other methods of pattern recognition of computer sciences to ascribe structure on processed data by decision theoretic or syntactic technique such as trainable controllers, reinforcement learning, Bayesian learning, stochastic automata, stochastic approximation and decision making schemata. The practical implementation of these methods for water resources systems are to be investigated.

The Chapter I deals with the introduction and formulation of the problem of investigation. It contains the scope of work, description of the river systems and the sources of data.

The Chapter II deals extensively with the survey of the existing literature and the state-of-the-art. It discuss the effectiveness and shortcomings of the previous works. It exposes gradually the harmonious evolution of the scientific thoughts pertaining to the present investigation. In group method of data handling algorithms the integral square error criterion has been used for selecting the unique model of optimum complexity. Since by using only one selection criterion the different physical processes of nature can not be modelled, particularly when data are noisy, the unbiasedness criterion for the residuals has been tested. The convergence criterion of the different version of recursive techniques has not been fully exposed in the state-of-the art; because of the fact that the emphasis of the investigation is centered on the implementation of the techniques. The state-of-the-art of the combined parameter and state estimation technique, e.g., extended Kalman filtering has not be discussed; since it has not been used in the present investigation.

In Chapter III the importance of learning identification algorithms of combinatorial and multilayer group method of data handling technique and recursive instrument variable method has been successfully demonstrated for on line forecasting of hourly flow of the hilly river Teesta during a short span monsoon storm period. Implementability of the recursive versions of the dynamic least square and exponential weighting of past data techniques are still to be observed and evaluated.

In Chapter IV an adequately simulated mathematical description has been obtained by using a non-stationary recursive parameter estimation technique minimising an objective function with a view to implement on line adaptiveness in process parameter variation for real time forecasting of the hourly flow of the river Teesta in North Bengal for a monsoon storm period. As an alternative extended

Kalman filtering technique may be of combined state and parameter estimation may be used. It is hoped that extended Kalman filter may give better results. However, the investigator believes that the recursive instrument variable algorithm with on line parameter variations is bound to give highly convergent estimates.

The Chapter V deals with the problem of identifying the parameters of the daily flow process of the river Teesta described by linear stochastic difference equations. The parameters are evaluated on real time basis with recursive instrument variables algorithms. The dynamic model developed in this Chapter is observed to simulate closely the major variations found in the field measurements of the daily flows of the river Teesta at Domohani Road Bridge point near Jalpaiguri town in North Bengal. A more scientific and realistic mathematical description of the daily river flow may be developed if suitable data base is obtained for the different meteorological processes of rainfall, sun shine hour, wind velocity etc. for an evenly distributed gauging stations over the catchment area in the Himalayan region.

In Chapter VI a simple dynamic model of the dissolved oxygen levels of a non-tidal river has been developed by a heuristic self-organisation technique. The model structure has been identified with the help of a learning identification algorithm commonly known as combinatorial group method of data handling technique. The model has been verified and validated with respect to the observed data inaccordance with an integral square error criterion. The presence of noisy data is not ruled out. As such it is suggested to use unbiasedness criterion in addition to the integral square error one. The unbiasedness criterion may be defined as

Minimise

$$\frac{\sum_{t=1}^{N} \left[Y_A(t) - Y_B(t) \right]^2}{\sum_{t=1}^{N} \left[Y_A(t) \right]^2}$$

where $Y_A(t)$ = polynomial model with points with even indices

$Y_B(t)$ = polynomial model with points with odd indices.

The Chapter VII deals with the combinatorial group method of heuristic learning identification algorithm and recursive instrument variable technique to obtain real time mathematical descriptions of biochemical oxygen demand of a non-tidal river. Because of noise in the data set (errors in observation and measurements being forth the noisy data) a combined model selection criterion of unbiasedness and integral square error has been suggested.

In Chapter VIII recursive least square non stationary time series analysis technique of applied cybernetics has been used for on line forecasting of daily dissolved oxygen concentration of a non-tidal river. When tested with field data the developed dynamic model is found to simulate aquately the major variations of D.O. levels obtained in field measurements.

The Chapter IX deals with the recursive instrument variable forecasting technique in conjunction with the least square recursive parameter tracking algorithm for real time prediction of the levels of D.O. of a non-tidal river. The method developed in this Chapter highlight the possibilities of implementation of the on line regulatory mechanism for controlling water quality of a river.

The Chapter X deals with an application of the dynamic least square estimation algorithms for on line modelling of dissolved oxygen levels.

In Chapter XI the multilayer group method of data handling algorithm and dynamic recursive least square technique of applied cybernetics have been used for obtaining dynamic descriptions of daily dissolved oxygen level of a non tidal river correlating the interactions of daily biochemical oxygen demand, sustained sunlight and volumetric flow rate. When tested with field data the models are found to simulate adequately the major variations of the levels of dissolved oxygen observed in the field measurements.

The application of the fuzzy control technique to the combined real time control of discharge and water quality has been considered in the Chapter XII. It offers a speculative and conceptual examination of the possibilities of fuzzy control application in the operation of river system management.

The improvement and control of water quality in a natural body of water such as a river can be achieved by intelligent regulation of municipal and industrial waste discharges. The important engineering decisions in water quality control relate to the determination of the level of waste treatment that is consistent with the multiple uses of the natural water bodies. This implies the ability to forecast in real time and on line basis the levels of dissolved oxygen and biochemical oxygen demand to effect a closed loop regulatory mechanism for waste treatment techniques commensurate in optimal use of waste treatment facilities. Serious research efforts are needed to plan and design closed loop computer controlled regulatory mechanism to control water quality in accordance with a predetermined decision methodology. The present investigation leading to the development of different versions of the mathematical descriptions of water quality processes is a humble contribution to this objective.

It is hoped that when in near future there will be sophistication in the water resources management in the form of control centre with computer linked through microwave with on-line river flow and water quality monitoring points, this investigation will be very helpful.

BIBLIOGRAPHY

B I B I L I O G R A P H Y

[1] Box, G.E.P., and Jenkins, G.M., " Time Series Analysis Forecasting and Control", Holden-Day. Inc., San Francisco, Calif, 1970.

[2] Yule, G.U., "On a method of Investigating Periodicities in Series, with Special Reference to Wolfer Sunspot numbers", Philosophical Transactions of the Royal Society, London, England, A226, 1927, PP 267-298.

[3] Hipel, K.W., McLeod, A.I. and Lennox, W.C., "Advances in Box - Jenkins Modelling, Part One, Model Construction, " Water Resources Research, Vol. 13, No. 3, June 1977, PP 596-575.

[4] ----------, "Advances in Box Jenkins Modelling, Part Two, Applications", Water Resources Research, Vol. 13 No. 3, June 1977, PP 577-586.

[5] McLeod, A.I., " Improved Box-Jenkins Estimates", Biometrika, Vol. 64, No. 3, 1977.

[6] Jettmar, R.U., and Young, G.K. "Hydrologic Estimation and Economic Regret," Water Resources Research, Vol. 11, No. 5, 1975, PP 648-656.

[7] Tao, P.C. and Delleur, J.W., " Seasonal and Non-seasonal ARMA Models in Hydrology," Journal of the Hydraulics Division ASCE, Vol. 102, No. 4710, Octo 1976.

[8] Mandelbrot, B. B. and Wallis, J.R., " Computer Experiments with fractional Gaussian Noises, Part I - Average and Variances," Water Resources Research, Vol. 5, No. 1, Feb, 1969, PP 228-241.

[9] Garcia, L.E., Dandy, D.R. and Matin, J.M., "Long Memory Monthly Stream Flow Simulation by a Broken Line Method," Vol. 8, No. 4, 1972 PP 1100-1105.

[10] Hipel, K.W., McBean, E.A. and McLeod, A.I., "Hydrologic Generating Model Selection," Journal of the Water Resources Planning and Management Division, ASCE, Vol. 105, No. WR2, Sept . 1979.

[11] Lawrance, A.J., and Kottegoda, N.T., "Stochastic Modelling of River Flow Time Series", Journal of the Royal Statistical Society, Series A. 140, Part 1, 1977, PP 1 - 47.

[12] Tao, P.C., and Bulleur, J.W., "Models of Stochastic and Chronologic Structure, Prediction and Simulation of Run-off Sequences - Application to Lower Ohio Basin", Technical Report No. 60, Purdue University Water Resources Research Centre, Lafayette, Ind.1975.

[13] Clarke, R.T., Mathematical Models in Hydrology, Food and Agriculture Organization of the United Nations, Rome, Italy, 1973.

[14] Chaudhuri, A.S., "Prediction of Monthly Tea Crop," Herald of the Kiev Polytechnical Institute, 1979.

[15] Box, G.E.P., and Cox, D.R., "An Analysis of Transformation," Journal of the Royal Statistical Society, Series B, Vol. 26, 1964 , PP. 211 - 252.

[16] Singh, K.P., "Synthetic Hydrologic Transformation — A Case Study," Journal of the Hydraulic Division, ASCE, Vol. 105, No. 4710, Octo 1979.

[17] Beck, M.B., "System Identification, Estimation, and Forecasting of Water Quality Part 1 : Theory "Working Paper, WP-79-31, April, 1979, International Institute for Applied System Analysis, A-2361 Laxenburg, Austria.

[18] Young, P.C., "A Recursive Approach to Time Series Analysis," Bulletin of the Institute of Mathematics and its Application, Vol. 10, PP 209-224, 1974.

[19] Soderstrom, T., L. Ljung and I.Gustavsson, " A Theoretical Analysis of Recursive Identification Methods," Automatica, Vol. 14, PP. 231-244, 1978.

[20] Ljung, L., "On Positive Real Transfer Functions and the Convergence of Some Recursive Schemes," IEEE Trans. Automatic Control, Vol. AC-22, Aug. 1977.

[21] _____," Analysis of Recursive Stochastic Algorithms", IEEE Trans. Automatic Control, Vol. AC-22, Aug. 1977.

[22] _____ " Convergence Analysis of Parameter Identification Methods," IEEE Trans., Vol. AC - 23, No.5, 1978.

[23] _____ " A Theoretical Analysis of Recursive Identification Algorithms," Automatica, Vol. 14, PP 231-244, 1978.

[24] _____ " Asymtotic Behaviour of Extended Kalman Filter as a Parameter Estimator For Linear Systems," IEEE Trans., Vol. AC-24, No. 1, Feb. 1979, PP. 36-50.

[25] Young, P.C., Optimization in the Process of Modns — A Guided Tour", Report Number AS/32 (1975), Centre for Resource and Environmental Studies, Australian National University Canberra.

[26] Chaudhuri, A.S., "System Analysis of Ten Industry : Identification, Modelling and Prediction," Thesis Submitted to the University of Calcutta - Dec. 1979.

[27] Eykhoff, P., System Identification - Parameter and State Estimation, Willey, London, 1974.

[28] Draper, N.R., and Smith, H., "Applied Regression Analysis," Willey, New York, 1966.

[29] Beck, M.B., "Real Time Control of Water Quality and Quantity," Institute of Applied System Analysis, Laxenburg, Austria.

[30] Hastings - James, R. and Sage, M.W., " Recursive Generalised Least square Procedures for on line Identification of Process Parameters," Proceedings of IEE, 116, PP. 2057 - 2062, 1969.

[31] Young, P.C., "Some Observations of Instrumental Variable Methods of Time Series Analysis," International Journal of Control, Vol. 23, PP 593-612, 1976.

[32] Young, P.C., "General Theory of Modelling for Badly defined Systems," Page 103-135, Modelling Identification and Control in Environmental Systems, Edited by G.C. Vansteenkiste, Amsterdam, North-Holland, 1978.

[33] Young, P.C., and Jakeman, A.J., " Refinfored Instrument Variable Methods of Recursive Time Series Analysis Part I : Single Input, Single Output Systems," Report AS/R12 Centre for Resources and Environmental Studies, Australian National University, Canberra, Australia, 1978.

[34] _____, "Refined Instrumental Variable Methods of Recursive Time Series Analysis Part II : Multivariable Systems," Report AS/R13, Centre for Research and Environmental Studies, Australian National University, Canberra, Australia, 1978.

[35] _____, "Refined Instrumental Variable Methods of Recursive Time Series Analysis Part III : Extensions," Report AS/R17, Centre for Resources and Environmental Studies, Australian National University, Canberra, Australia, 1978.

[36] Young, P.C., Jakeman, A.J., and Mc-Murtrie, R.E., "An Instrumental Variable Method for Model Structure Identification," "Report AS/R22, Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia, 1978.

[37] Young, P.C., Shaw-ell, S.H., and Neethling, C.G., " A Recursive approach to time Series Analysis," Technical Report CUED/B-Control/TR16, University Engineering Department, Cambridge, U.K. 1971.

[38] Cox, D.R., and Miller, H.D., "The Theory of Stochastic Processes," Chapman & Hall, London, 1972.

[39] Whitehead, P., and Young, P.C., "A Dynamic-Stochastic Model for Water Quality in Part of the Bedford Ouse River System," Modelling and Simulation of Water Resources System, G.C. Vansteenkiste, (Ed.), North-Holland Publishing Company, 1975.

[40] Whitehead, P., and Young, P.C., "The Bedford Ouse Study Dynamic Model," 2nd. Semi-Annual Report to the Steering Group of the Great Ouse Associated Committee, 1974, Control Group Internal Note CN/74/1, University of Cambridge, U.K.

[41] Young, P.C., Whitehead, P.G., and Beck, M.B., "The Bedford Ouse Study Dynamic Model," 1st Semi-Annual Report to the Steering Group of the Great Ouse Associated Committee, Control Group Internal Note, CN/73/4, 1973, University of Cambridge, U.K.

[42] Beck, M.B., "The Identification and Adaptive Prediction of Urban Sewer Flows," International Journal of Control, 1977, Vol. 25, No.3, PP 425-440.

[43] Camp, T.R., Water and its Impurities, Reinhold Publishing Corporation, New York, N.Y., 1963.

[44] Di Toro, D.M., "Stream Equations and Method of Characteristics," Journal of the Sanitary Engineering Division, ASCE, Vol. 95, No. SA 4, Proc. Paper 6723, Aug. 1969, PP 669-703.

[45] Dobbins, W.E., "BOD and Oxygen relationship in Streams," Journal of the Sanitary Engineering Division, ASCE, Vol. 90, No. SA3, Proc. Paper 3949, June 1964, PP 53-78.

[46] O' Connor, D.J., " The Temporal and Spatial Distribution of Dissolved Oxygen in Streams," Water Resources Research, Vol. 3, No. 1, Feb. 1967, PP 65-79.

[47] Streeter, H.W., and Phelps, E.B., "A Study of the Pollution and Natural Purification of the Ohio River," Public Health Bulletin, U.S. Department of Health Education and Welfare, No. 146, 1925.

[48] Thomann, R.V., " Mathematical Model of Dissolved Oxygen," Journal of the Sanitary Engineering Division, ASCE, Vol. 89, No. SA5, Proc. Paper 3680, Oct., 1966, PP. 1-30.

[49] Beck, M.B., "Maximum Likelihood Identification Applied to DO-BOD Algae Models for a Fresh Water Stream," Report 7431, Lund Institute of Technology, Lund, Sweden, 1974.

[50] Beck, M.B., and Young, P.C., "Systematic Identification of DO-BOD Model Structure," Journal of the Environmental Engineering Division, ASCE, Vol. 102, No. EE5, Proc. Paper 12461, Oct, 1976. PP 909-927.

[51] Koivo, A.J., and Phillips, G.R., "Identification and Mathematical Models for DO and BOD Concentrations in Polluted Streams from Noise Corrupted Measurements," Water Resources Research, Vol. 7, No. 4, Aug., 1971, PP 853-862.

[52] _____, "On Determination of BOD and Parameters in Polluted Stream Models From DO Measurements Only," Water Resource Research, Vol. 8, No. 2, April, 1972, PP 78-486.

[53] Koivo, H.N., and Koivo, A.J., "Optimal Estimation of Polluted Stream Variables," Proceedings, 3rd. IFAC Symposium on Identification and System Parameter Estimation, P.Eykhoff, ed, North Holland/American Elsevier, Amsterdam, The Netherlands, 1973.

[54] Rinaldi, S., Romano, P., and Soncini -- Sessa, R., "Parameter Estimation of a Streeter Philps Type Water Pollution Model," 4th. IFAC Symposium on Identification and System Parameter Estimation, Tbilisi, USSR, Sept., 1976.

[55] Rinaldi, S., et al., "Modelling and Controllof Water Quality" McGraw Hill Book Co., Inc., New York, N.Y., 1978.

[56] Young, P.C., and Whitehead, P.G., "A Numerical Approach to Time Series Analysis for Multivariable Systems," G.C Vansteenkiste ed., North Holland Publishing Co., Amsterdam, The Netherlands, 1975.

[57] Rinaldi, S., Romano, P., and Soncini - Sessa, R., "Parameter Estimation of Streeter Phelps Models," Journal of the Environmental Engineering Division, Proceedings of the American Society of Civil Engineers, Vol. 105, No. EE1 February, 1979, PP 75-88.

[58] Ivakhnenko, A.G., "Heuristic Self-Organization of Problems of Engineering Cybernetics," Automatica, Vol. 6, PP 207-219, 1970.

[59] Ivakhnenko, A.G., and Koppa, Yu. V., "Group Method of Data Handling for Solution of Various Interpolation Problems of Cybernetics," Second IFAC Symposium of Identification and Process Parameter Estimation, Prague, Paper 21, 1970.

[60] Ivakhnenko, A.G., "Polynomial Theory of Complex Systems," Transaction of Systems, Man and Cybernetics of Institution of Electrical and Electronics Engineers, IEEE, Vol. SMC 1, No. 4, October, 1971.

[61] Ivakhnenko, A.G., " Long Term Prediction and Control of Complex Process," Tekhnika, Kiev, 1975 (Book in Russian).

[62] Volterra, V., "Theory of Functional and Integral and Integro Differential Equations," London, Blackie, 1930.

[63] Wiener, N., "Non Linear Problems in Random Theory," New York, Wiley, 1958.

[64] Gabor, D., Wilby, W.P.L., and Wood Cook, R., "A Universal Non-linear Filter Predictor and Simulator which Optimises itself by a Learning Process," Proceedings of the Institution of Electrical Engineers, London, IEE Proc., Vol. 108, Part B, PP. 422 - 433.

[65] Astron, K.J., and Eykhoff, P., "System Identification, A Survey," Automatica, Vol. 7, PP. 123 - 162, March, 1971.

[66] Rosenblatt, F., "Principles of Neurodynamics," Spartan Books, Washington, 1962.

[67] Rosenblatt, F., "Perception : A Probabilistic Model for Information Storage and Organization in the Brain," Phychological Review, Vol. 65, PP 386-408, November, 1958.

[68] Ivakhnenko, A.G., "The Group Method of Data Handling a Rival of the Method of Stochastic Approximation," Soviet Automatic Control, No. 3, 1968.

[69] Ivakhnenko, A.G., et al., "Group Method of Data Handling in the Problem of Pattern Recognition and Decision Making", Soviet Automatic Control, No. 5, 1968.

[70] Ivakhnenko, A.G., et al., "Group Method of Data Handling in the Problem of Identification of Multicentrasal Plant," Soviet Automatic Control No.2, 1969.

[71] Box, G.E.P., "Use and Abuse of Regression," Technometrics, Vol. 8, PP. 625 - 629, Nov., 1966.

[72] Parks, P., and Pyatt, Y., "Controlling the Economy - an Exercise in Collaboration," Electron Power, Vol. 15, August, 1969, PP 270-274.

[73] Kodo, Saburo, et al., "Sequential GMDH Algorithm and its Application to River Flow Predictions," Institution of Electrical and Electronics Engineers, IEEE Transaction on Systems, Man and Cybernetics, Vol. SMC-6, No. 7, July, 1976.

[74] _____, "Non Linear Prediction on Model of River Flow by Self-organization Method," International Journal of System Science, 1976.

[75] Tamura, H., and Kondo, T., "Large Spatial Pattern Identification of Air Pollution by a Combined Model of Source Perceptor Matrix and Revised GMDH," IFAC Symposium on Environmental Systems Planning, Design and Control, Kyoto Japan, Aug, 1-5, 1977.

[76] Duffy, J.J., and Franklin, M.A., "A Case Study of Environmental System Modelling with Group Method of Data Handling," Joint Automatic Control Conference, U.S.A., Proc. PP 101-111, 1973.

[77] _____, "A Learning Identification Algorithm and its Application to an Environmental System," IEEE Trans. Systems, Man and Cybernetics, Vol. SMC-5, No. 2, March, 1975.

[78] Nahra, H.K., "Group Method of Data Handling Algorithms (GMDH) Review and Experience," IEEE Conference on Decision and Control, December 7-9, 1977, New Orleans, U.S.A.

[79] Bader, W., and Williams, G.D.V., "Regional Wheat Yield Prediction from Weather Data in Canada," WMO Symposium on Agrometeorology of the Wheat Crop, Braunschweig, Federal Republic of Germany, Oct. 1973, PP 263-283.

[80] Mustafi, A., and Chaudhuri, A.S., "Use of Multilayer Group Method of Data Handling for Predicting Tea Crop Production," Journal of the Society of Indian Agricultural Statistics, Vol. XXXIII, No. 1, April, 1981.

[81] Ivakhnenko, A.G., "Present State of the Theory of Computer Aided Self-Organisation of Mathematical Model," Soviet Automatic Control, Vo. 8, September-October, 1975.

[82] Chaudhuri, A.S., "Medium-term and Long-term Prediction Model of Annual Tea Crop Production Using Heuristic Learning Algorithm," Journal of the Indian Society of Agricultural Statistics, Vol. XXXIII, No. 2, August, 1981.

[83] Jenkins, G.M., and Watts, D.G., " Spectral Analysis and its Application," Holden Day, 1968.

[84] Dixon, W.J. (Editor), "Biomedical Computer Programmes," University of California Press, Berkley, California, 1970.

[85] Kashyap, R.L., and Rao, A.R., "Real time Recursive Prediction of River Flows," Automatica, Vol. 9, PP 175-183, 1973.

[86] _____, "Analysis, Construction, Validation of Stochastic Models of Monthly River Flows," Hydromechanics CE-HYD-71-1, Purdue University, 1973.

[87] Chaudhuri, A.S., "System Analysis of Indian Tea Industry," Report Submitted to Indian Tea Association, Tocklai, Jorhat, Assam, November, 1979.

[88] Kashyap, R.L., "Estimation of Parameters in a partially Whitened Representation of a Stochastic Process," IEEE Trans. Automatic Control, Vol. A.C. 19, PP 13-21, Feb 1974.

[89] Kashyap, R.L., and Rao, A.R., "Stochastic Modelling of River Flow," IEEE Trans. Automatic Control, Vol. AC-19, No. 6, December, 1974.

[90] Kashyap, R.L., and Rao, A.R., "Dynamic Stochastic Models from Empirical Data," Academic Press, New York, 1976.

[91] Zadeh, L.A., "Fuzzy Sets, "Information & Control, Vol. 8., PP 338-353, 1965.

[92] _____, "Fuzzy Algorithms, " Information and Control, Vol. 12, PP 94-102, 1968.

[93] Santos, E., "Fuzzy Algorithms, " Information and Control, Vol. 17, PP 326-339, 1970.

[94] Bellman, R.E., and Zadeh, L.A., "Decision Making in a Fuzzy Environment," Management Science, Vol. 17, PP B 141-B-164, 1970.

[95] Zadeh, L.A., "On Fuzzy Algorithms," Electronic Research Laboratory, Univ. of California, Berkley, Memo-19-325, 1971.

[96] _____, "Outline of a New Approach to the Analysis of Complex System and Decision Process," IEEE Transaction on System, Man and Cybernetics, Vol. SMC-3, No.1, Jal.1973.

[97] Saridis, G.N., "Application of Pattern Recognition Methods to Control Systems," IEEE Transactions on Automatic Control, Vol. AC-26, No. 3. June, 1981.

[98] Beck, M.B., "Real Time Control of Water Quality and Quantity," International Institute of Applied System Analysis, 2361 Laxenburg, Austria, Research Memorandum, RM-79-19, April, 1978.

[99] Tong, R.M., A Control Engineering Review of Fuzzy Systems," Automatica, 13, PP 559-569, 1977.

[100] Sengupta, S.K., Maulik, M.N, and Chaudhuri, A.S., "Forecasting of River Flows During a Short Span Storm Period," presented to the World Congress on Water Resources, on Water for Human Consumption, Man and His Environment" of International Water Resources Association (IWRA) held in Buenos Aires, Argentina, September 3-11, 1982.

[101] Sengupta, S.K., Maulik, M.N., and Chaudhuri, A.S., "Forecasting of River Flow During Storm Periods by a Learning Identification Algorithm," Journal of Civil Engineering Division of the Institution of Engineers (India), Vol. 63, PT CH 1, July 1982 PP 6-9.

[102] Maulik, M.N., Sengupta, S.K., and Chaudhuri, A.S., "Prediction of Hourly Flow by Computer Aided Self-Organisation of Mathematical Models," accepted for publication in the Journal of Irrigation and Power of the Central Board of Irrigation and Power, Govt. of India.

NORTH BENGAL
UNIVERSITY LIBRARY
RAJA RAMMOHUNPUR

103. Maulik, M.N., Sengupta, S.K., and Chaudhuri, A.S., " Forecasting of hourly River Flow Using Heuristic Learning Algorithm," Journal of the Civil Engineering Division of the Institution of Engineers (India), Vol. 63, Pt CL 1, July, 1982, pp 10-11.

104. Maulik, M.N., Sengupta, S.K., and Chaudhuri, A.S., "Real Time Estimation of River Flow During A short Storm Period," accepted for publication to the Journal of the Civil Engineering Division of the Institution of Engineers (India).

105. Sengupta, S.A., Maulik, M.N., and Chaudhuri, A.S., "Recursive Estimation of River Flows During a Short Span Storm Period," Submitted for publication in Indian Journal of Power and River Valley Development.

106. _____, "Real Time Forecasting of Daily Flows of a Non-tidal River by Instrument Variable Algorithm," submitted for publication to the Journal of the Civil Engineering Division of Institution of Engineers (India).

107. Maulik, M.N., Sengupta, S.K., and Chaudhuri, A.S., "Real Time Forecasting of Daily Flow of a Non-Tidal River," Journal of the Civil Engineering Division of the Institution of Engineers (India), Vol. 63, PT CL 3, November, 1982.

108. Sengupta, S.K., Maulik, M.N., and Chaudhuri, A.S., " Forecasting of Daily Dissolved Oxygen Levels of a Non-Tidal River by Self-Organisation of Mathematical Models," accepted for publication to the Journal of the Environmental Engineering Division of the Institution of Engineers (India).

109. Sengupta, S.K., Maulik, M.N., and Chaudhuri, A.S., " Modelling of Bio-Chemical Oxygen Demand (B.O.D) of a Non-Tidal River by a Learning Algorithm and Recursive Instrument Variable Technique", accepted for publication to the Journal of the Environmental Engineering Division of the Institution of Engineers (India).

110. Sengupta, S.K., Chaudhuri, A.S., "Real Time Forecasting of Dissolved Oxygen Level of a Non-Tidal River," Journal of the Environmental Engineering Division of the Institution of Engineers (India), Vol. 62, PT EN 3, June, 1982, PP 127 - 130.

111. Sengupta, S.K., Maulik, M.N., and Chaudhuri, A.S., " Recursive Prediction of Dissolved Oxygen Level of a Non-Tidal River", submitted for publication to Sen-Khyn of the Indian Statistical Institute, Calcutta.

112. _____, "Forecasting of Daily Dissolved Oxygen Levels of a Non-Tidal River by Dynamic Least Square Recursive Estimation Technique," submitted for publication to the Journal of the Environmental Engineering Division of the Institution of Engineers (India).

113. Beck, M.B., and Young, P.C., "A Dynamic Model for DO-BOD Relationships in a Non-Tidal Stream" - Water Research, Vol. 9 PP 769 to 776, Pergamon Press, 1975.

114. Sengupta, S.K., Maulik, M.N., and Chaudhuri, A.S., " Learning Identification Dynamic Algorithms for Modelling of Dissolved Oxygen Incorporating Interaction Parameters," submitted for publication to the Journal of Water Research of the International Association on Water Pollution Research, England.

115. Sengupta, S.K., Maulik, M.N., Chaudhuri, A.S., " Concept of Fuzzy Control and a Suggestive Examination of its Application in the Water Quality Management of a Non-Tidal River," submitted for publication to the Journal of the Environmental Engineering Division of the Institution of Engineers (India).

116. Maulik, M.N., Sengupta, S.K., and Chaudhuri A.S., " Rainfall-run-off Model of Flows of a River for a short term Storm Period by Learning Algorithms," Indian Journal of Power and River Valley Development, January - February 1982.