TR128

REPUBLIC OF SOUTH AFRICA

M. J. SILBERBAUER

DEPARTMENT OF WATER AFFAIRS

Landsat water quality surveillance—development of the model CALMCAT

A M Howman   P L Kempster

REPUBLIC OF SOUTH AFRICA

DEPARTMENT OF WATER AFFAIRS

TR 128

Landsat water quality surveillance— development of the model CALMCAT

A M Howman P L Kempster

DEPARTMENT OF WATER AFFAIRS
Hydrological Research Institute

Technical Report No. TR 128

LANDSAT WATER QUALITY SURVEILLANCE— DEVELOPMENT OF THE MODEL CALMCAT

by

A. M. Howman and P. L. Kempster

DW 145/1

DEPARTMENT OF WATER AFFAIRS

AHowman P.L. Kempster

With the

Compliments

of the

Department of Water Affairs

July 1986

Department of Water Affairs
Private Bag X313
Pretoria
0001

ISBN 0 621 10318 7

Printed by the Government Printer, Bosman Street, Private Bag X85, Pretoria, 0001
Gedruk deur die Staatsdrukker, Bosmanstraat, Privaatsak X85, Pretoria, 0001

NOTE

The methodology that has been proposed in this report is new and unorthodox. It is presented as a way of assisting the quest for quantifying real world situations using remotely sensed data. Controversy over the method is expected and welcomed. This work will have served its purpose if it can aid in the successful monitoring of our environment and if it has raised more questions than it has answered.

> "The best way to summarize a mass of multifactor data is by a simple equation or set of equations. The data, however, must be studied critically, and here the standard texts give little guidance beyond stern warnings to be cautious. Routine use of standard computer programs to fit equations to data does not usually succeed. A large proportion of the failures is due, not to the programs, computers or data, but to the analyst's approach".

C. DANIEL and F.S. WOOD (1971)
in <u>Fitting Equations to Data</u>,
Wiley Interscience.

The methodology that has been proposed in this report is new and unorthodox. It is presented as a way of assisting the quest for quantifying real world situations using remotely sensed data. Controversy over the method is expected and welcomed. This work will have served its purpose if it can aid in the successful monitoring of our environment and if it has raised more questions than it has answered.

"The best way to summarize a mass of multifactor data is by a simple equation or set of equations. The data, however, must be studied critically, and here the standard texts give little guidance beyond stern warnings t o be cautious. Routine use of standard computer programs to fit equations to data does not usually succeed. A large proportion of the failures is due, not to the programs, computers or data, but to the analyst's approach".

C. DANIEL and F.S. WOOD (1971)
in *Fitting Equations to Data*,
Wiley Interscience.

ACKNOWLEDGEMENTS

The Project was a success and a pleasure to undertake thanks to the cooperation and assistance received from many sources. We would like to thank the following people for the part they played in the development of the model CALMCAT*.

Mr. H. Bosman of the Hydrological Research Institute (HRI), who assisted with all aspects of photography and with the printing of the report.
Mr. T. Boyle of the Satellite Remote Sensing Centre (SRSC), who went out of his way to process the Computer Compatible Tapes (CCT) and to obtain the digital data.
Mr. E. Braune, Director of the HRI for his support and advice.
Dr. T. Crowe of the Percy Fitzpatrick Institute for African Ornithology of the University of Cape Town (UCT), for suggesting the use of the Stepwise Discriminant Analysis.
The Foundation for Research Development of the Council for Scientific and Industrial Research (CSIR), for supporting the project and providing the satellite imagery.
Ms. E. Davies (HRI) who helped with the sampling and the analysis of the water quality samples.
Dr. J. Day of the Department of Zoology, UCT, who gave advice on aspects of limnology.
The Department of Water Affairs for permitting the project to be carried out under its auspices.
Mr. S. de Wet (HRI) for casually suggesting that a fully operational research project was feasible and thus leading to intensive sampling operations.
Mrs. C. Edwards (HRI) who cheerfully transformed a scrawl and many more scrawls into an excellently typed final document.
Mrs. A. Faber (HRI) who assisted with the technical aspects.
Mrs. E. Gerber who standardised and skillfully reproduced all of the figures.
Mr. F. Gerber (HRI) for literally enabling the project to get off the ground, organising boats, equipment and aiding in every sort of emergency.
Dr. D. Grobler of the National Institute for Water Research (CSIR) for his useful comments.
Prof. C.T. Haan of the Agricultural Engineering Department, Oklahoma State University, for the useful discussion held on the problem of non-normal data and the concept of representativeness.
Dr. W. Hattingh, former director of the HRI, for supporting the project in its initial stages.
Mr. L. Herbst of the Agricultural Technical Services Computer Section, for helping with the loading of the CCT's.
Mr. U. Looser (HRI) who helped with photography and reproducing the plates.
Mr. H. Maaren (HRI) who supported the project throughout its duration and who wrote many of the subroutines required to analyse the data, in particular the final model CALMCAT.

---

* CALMCAT - Canonical Analysis Landsat Model of Chlorophyll *a* and Turbidity.

Dr. O.G. Malan of the National Physics Research Laboratory of the CSIR, who originally gave us the Bloemhof Dam image, assisted with the original motivations of the project and helped to read the CCT's.
Mr. I. Marais of the SRSC who organised the CCT's and the satellite imagery.
Mrs. J. Meyer of the Department of Statistics, University of the Witwatersrand, for the useful discussions on the statistical analysis of the data.
Mr. A. Moathlodi (HRI) who helped organise the sampling network and assisted in the sampling expeditions.
Mr. S. Piper of the Department of Surveying and Mapping, University of Natal, for his advice on the statistical approach.
Dr. P.R. Roberts, Manager Scientific Services, Department of Water Affairs is thanked for his useful comments and suggestions.
Dr. H. Rüther of the Department of Surveying, UCT, who gave advice on aspects of remote sensing.
The sampling team consisted of many members of the HRI staff to whom we are most thankful for their time and willing cooperation.
Mr. D. Sartory (HRI), for his comments on chlorophyll and for assisting with the sampling.
Mr. I. Schoonraad (HRI), who proved to be a patient teacher and converted Filliben's R and Grubb's t test from Basic into Fortran IV. He also assisted with the 'LINREG' program.
Mr. A. Seed (HRI), who arrived just in time to set up P.I.P.S. (Portable Image Processing Suite) and who helped to produce the final maps of distribution and concentration.
Mr. M. Silberbauer (HRI), who assisted with sampling, many computer programs and who was a source of novel ideas.
Mr. J.P. Trichardt of Computer Systems, Department of Water Affairs who wrote the program to read the CCT's.
Mrs. E. Truter (HRI), who assisted with the sampling operations and expertly analysed the water samples.
Dr. H.R. Van Vliet (HRI), for his moral support throughout the project.
Mr. R. Wixley of the National Research Institute of Mathematical Sciences (CSIR), for originally suggesting the use of the Canonical Correlation Analysis.
Mrs. S. Young (HRI), for her help with analysing the water quality samples and advising on all aspects of chlorophyll.
Mrs. D. Zietsman (HRI) for providing invaluable assistance in organising the data and helping with the computer programs.

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

UNITS OF MEASUREMENT

Throughout this report the units of measurement used for chlorophyll a and turbidity are microgramme per litre (μg/ℓ) and nephelometric turbidity units (NTU) respectively.

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF PLATES

LIST OF APPENDICES

LIST OF APPENDICES

ABSTRACT

The need for accurate, synoptic, up to date information, concerning the quality of South African impoundments, prompted an investigation into the potential and limitations of Landsat reflectance data for assessing chlorophyll *a* and turbidity in Roodeplaat Dam.

Surface and integrated chlorophyll *a* as well as surface and integrated turbidity were collected simultaneously with the satellite's overpass from 32 sampling sites on the impoundment. Six days, between 1981.10.14 and 1982.11.16, were cloud free and the data were analysed in order to establish the relationship between the specific water quality conditions and the satellite reflectance data.

Prior to the analysis certain factors required attention. Firstly, it was important to accurately align the sampling sites with their corresponding Landsat pixels. Secondly, the satellite reflectance data were corrected for Influences of haze and the angle of the sun. Thirdly, the requirements that the water quality surface reference data be representative of the range of conditions in the impoundment and that data be normally distributed, and that outliers excluded from the data set, were recognised. Lastly, the interrelationship between chlorophyll *a* and turbidly and the multicollinearity evident between the four reflectance bands, demanded that a multi-variate statistical technique be Implemented, in order to adequately analyse the available data.

The Canonical Correlation multi-variate regression analysis was chosen to investigate the relationship between the surface reference data and the four Landsat wavebands. Canonical Correlations (r) ranged from 0.95 to 0.79 and the Canonical Coefficients enabled characteristics of the relationship between the variables to be established. As a general trend, surface chlorophyll *a* showed correlation with all of the wavebands, whereas integrated chlorophyll corresponded with bands 6 and 7. Surface turbidity mainly related to bands 4 and 5, but also at times to bands 6 and 7, while integrated turbidity related to bands 4 and 5. The trends varied between overpasses however indicating that the relationship was complex and unique to each specific overpass.

In addition to the use of the Canonical Correlation Analysis, the unsupervised classification technique and colour coding assisted in the interpretation of the conditions within the impoundment.

From the coefficients obtained in the Canonical Correlation Analysis, with the help of linear regression analysis, a set of simultaneous equations was established which described the relationship between the surface reference data and the satellite reflectance data. Explicit solution of these equations allowed the model CALMCAT* to be produced with which chlorophyll *a* and turbidity could be simulated from the satellite reflectance data.

For three of the days tested the accuracy of CALMCAT simulations ranged from 0.4% to 26% relative error for chlorophyll *a* and 2% to 20% for turbidity. Of overriding importance to the application of the model is the representativeness of the surface reference data set.

Incorporating the entire surface of the impoundment into the model provided synoptic and quantitative information of the distributions and concentrations of chlorophyll *a* and turbidity in the impoundment unlike any other presently available data source.

---

* CALMCAT - Canonical Analysis Landsat Model of Chlorophyll *a* and turbidity

ABSTRACT

The need for accurate, synoptic, up to date information, concerning the quality of South African impoundments, prompted an investigation into the potential and limitations of Landsat reflectance data for assessing chlorophyll a and turbidity in Roodeplaat Dam.

Surface and integrated chlorophyll a as well as surface and integrated turbidity were collected simultaneously with the satellite's overpass, from 32 sampling sites on the impoundment. Six days, between 81.10.14 and 82.11.16, were cloud free and the data were analysed in order to establish the relationship between the specific water quality conditions and the satellite reflectance data.

Prior to the analysis certain factors required attention. Firstly, it was important to accurately align the sampling sites with their corresponding Landsat pixels. Secondly, the satellite reflectance data were corrected for influences of haze and the angle of the sun. Thirdly, the requirements that the water quality surface reference data be representative of the range of conditions in the impoundment and that data be normally distributed, and that outliers excluded from the data set, were recognised. Lastly, the interrelationship between chlorophyll a and turbidity and the multicollinearity evident between the four reflectance bands, demanded that a multi-variate statistical technique be implemented, in order to adequately analyse the available data.

The Canonical Correlation multi-variate regression analysis was chosen to investigate the relationship between the surface reference data and the four Landsat wavebands. Canonical Correlations (r) ranged from 0,95 to 0,79, and the Canonical Coefficients enabled characteristics of the relationship between the variables to be established. As a general trend, surface chlorophyll a showed correlation with all of the wavebands, whereas integrated chlorophyll a corresponded with bands 6 and 7. Surface turbidity mainly related to bands 4 and 5, but also at times to bands 6 and 7, while integrated turbidity related to bands 4 and 5. The trends varied between overpasses however, indicating that the relationship was complex and unique to each specific overpass.

In addition to the use of the Canonical Correlation Analysis, the unsupervised classification technique and colour coding assisted in the interpretation of the conditions within the impoundment.

/From the ....

From the coefficients obtained in the Canonical Correlation Analysis, with the help of linear regression analysis, a set of simultaneous equations was established which described the relationship between the surface reference data and the satellite reflectance data. Explicit solution of these equations allowed the model CALMCAT* to be produced with which chlorophyll a and turbidity could be simulated from the satellite reflectance data.

For three of the days tested the accuracy of CALMCAT simulations ranged from 0,4% to 26% relative error for chlorophyll a and 2% to 20% for turbidity. Of overriding importance to the application of the model is the representativeness of the surface reference data set.

Incorporating the entire surface of the impoundment into the model provided synoptic and quantitative information of the distributions and concentrations of chlorophyll a and turbidity in the impoundment unlike any other presently available data source.

---

* CALMCAT - Canonical Analysis Landsat Model of Chlorophyll a and Turbidity

OPSOMMING

Die behoefte aan akkurate, sinoptiese en opdatum inligting oor die kwaliteit van Suid-Afrikaanse damme het tot 'n ondersoek na die potensiaal en beperkings van Landsat weerkaatsingsdata gelei as 'n metode om chlorofiel a en turbiditeit in Roodeplaatdam te skat.

Oppervlak en geïntegreede chlorofiel a asook oppervlak en geïntegreede turbiditeit is gelyktydig met die satelliet oorvlug, by 32 monsterpunte op die dam gemonster. Ses dae, tussen 81.10.14 en 82.11.16, was wolk- en probleemvry en die data is ontleed om die verwantskap tussen spesifieke waterkwaliteitstoestande en die satelliet weerkaatsingsdata vas te stel.

Sekere faktore het aandag vereis alvorens die ontleding uitgevoer kon word. Eerstens, was dit belangrik om die monsterpunte met ooreenstemende Landsat 'pixels' te rig. Tweedens was die satelliet weerkaatsingsdata vir die invloed van dynserigheid en sonshoek gekorrigeer. Derdens was die vereistes dat die waterkwaliteit-oppervlakverwysingsdata verteenwoordigend moet wees van die verskeidenheid van kondisies in die dam, dat die data normaalversprei moet wees en dat uitskieters verwyder moet word uit datastel, erken. Tenslotte, het die inter-verwantskap tussen chlorofiel a en turbiditeit en die multi-kolineariteit wat tussen die vier weerkaatsings golflengte-gebiede bestaan, die implimentering van 'n veel-veranderlike statistiese tegniek genoodsaak, om sodoende die beskikbare data doeltreffend te kan ontleed.

Die Canoniese-korrelasie veelveranderlike regressie-analise is gekies om die verwantskap tussen die oppervlak verwysingsdata en die vier Landsat golflengte-bande te ondersoek. Canoniese-korrelasies (r) het gestrek vanaf 0,95 tot 0,79, en die Canoniese-koëffisiënte het dit moontlik gemaak om die karakter van die verwantskap tussen die veranderlikes vas te stel. As 'n algemene verskynsel, het chlorofiel a 'n korrelasie getoon met al die golflengtes, terwyl geïntegreerde chlorofiel a met golflengte-bande 6 en 7 ooreengestem het.

Oppervlak turbiditeit het hoofsaklik met bande 4 en 5 ooreengestem, maar ook met tye met bande 6 en 7 terwyl geïntegreerde turbiditeit 'n verwantskap getoon het met bande 4 en 5. Die karakter van die verwantskap het egter verskil tussen oorvlugte, wat aangetoon het dat die verwantskap kompleks en uniek is vir elke spesifieke oorvlug.

Bo en behalwe die gebruik van Canoniese-korrelasie analise, was die nie-toesighoudklassifikasie-tegniek en kleurkodering waardevol in die interpretasie van toestande in die dam.

/Van die ....

Van die koëffisiënte verkry in die Canoniese korrelasie analise, met behulp van lineêreregressieanalise, is 'n stel gelyktydige vergelykings opgestel wat die verhouding tussen die oppervlak verwysingsdata in die satelliet weerkaatsingsdata beskryf het. Eksplisiete oplossing van die vergelykings het gelei tot die model CALMCAT* waarmee chlorofiel a en turbiditeit gesimuleer kan word met behulp van satelliet weerkaatsingsdata.

Vir drie van die dae getoets het die akkuraatheid van die simulasies met behulp van CALMCAT gestrek vanaf 0,4% tot 26% relatiewe-fout vir chlorofiel a en vanaf 2% tot 20% vir turbiditeit. In die toepassing van die model is die vereiste dat die oppervlakverwysingsdata verteenwoordigend moet wees, van primêre belang.

Inlywing van die hele oppervlak van die dam in die model het sinoptiese en kwalitatiewe inligting van die distribusie en konsentrasie van chlorofiel a en turbiditeit in die dam verskaf anders as enige huidig beskikbare data bron.

---

* CALMCAT - Canonical Analysis Landsat Model of Chlorophyll a and Turbidity

DEPARTMENT OF WATER AFFAIRS
HYDROLOGICAL RESEARCH INSTITUTE

Technical Report No. TR 128

LANDSAT WATER QUALITY SURVEILLANCE - DEVELOPMENT OF THE MODEL CALMCAT

by

ALISON HOWMAN and PHILLIP L. KEMPSTER
1986

EXECUTIVE SUMMARY

The objectives of the Landsat water quality surveillance project were as follows:

(1) To show that remotely sensed data could be used in the evaluation of chlorophyll a and turbidity in impoundments. Algal blooms for example show up as red and turbidity as whitish blue on false colour composites.

(2) To extract detail, not visible on the false colour composites, from the satellite digital information by colour coding of the digital data and thereby showing up differences in water quality.

(3) To show that it is possible to calibrate a satellite image to obtain quantitative measurements (simulations) of surface chlorophyll a and turbidity and integrated chlorophyll a and turbidity (integrated to secchi disc depth) for each 80 x 80 metre pixel within the impoundment. From this data, the entire surface area of the impoundment can firstly, be classified into areas of various concentration classes and secondly, be graphically contoured with isolines of chlorophyll a and turbidity concentrations. This objective involved developing the CALMCAT* model to calibrate the digital satellite data in terms of chlorophyll a and turbidity concentrations.

Fulfilling these objectives made it possible to obtain information on specific water quality conditions in impoundments unlike any information obtained to date.

The use of remotely sensed data for chlorophyll a and turbidity estimation is limited by the following:

(1) The problem of obtaining simultaneous surface reference data concurrent with the satellite's overflight. The simultaneous collection of surface reference data and satellite reflectance data is desirable in order to overcome the problem of variability in atmospheric transparency as well as in changes in conditions in the impoundment. Non-concurrent surface reference data may be used in the CALMCAT model for

*CALMCAT - Canonical Analysis Landsat Model of Chlorophyll a and Turbidity.

simulating chlorophyll a and turbidity from satellite digital data collected at other periods with, however, a decrease in accuracy.

(2) The need to have a surface reference data set which is representative of the full range of chlorophyll a and turbidity values in the impoundment is essential for the accurate calibration of the CALMCAT model. Non- representative data results in calibration parameters which have no reality to the underlying condition in the impoundment.

(3) Where only one of the four parameters is of interest to the user i.e. surface chlorophyll a, it is still necessary to measure all four variables viz., surface and integrated chlorophyll a and surface and integrated turbidity, in the surface reference data set in order to calibrate the CALMCAT model.

The CALMCAT model achieved the following accuracy for the days on which surface reference data was collected concurrently with the satellite reflectance data. The error in the simulated mean chlorophyll a varied between 2 to 9 µg/ℓ chlorophyll a, a percentage relative error of 0,4% to 26%. The error in simulated mean turbidity varied between 0,2 and 1,0 NTU, a percentage relative error of 2% to 20%. For satellite remotely sensed data collected non-concurrently with the surface reference data, i.e. where a calibration was extrapolated to other overpasses, the error in simulated surface and integrated turbidity mean values varied between 0,2 and 1,0 NTU, a percentage relative error of 5% to 20%. The error in mean simulated surface chlorophyll a was 14 µg/ℓ (51% relative error) while integrated chlorophyll a could not be quantified.

A three tiered approach to the evaluation of chlorophyll a and turbidity using remotely sensed data is examined. The first tier being a quantitative evaluation of areas of chlorophyll a and turbidity from a false colour composite, the second tier being a semi-quantitative evaluation using colour coding of the digital data, and the third tier being quantitative estimation of chlorophyll a and turbidity using the model CALMCAT together with the surface reference data calibration set.

The three tiered approach to evaluate chlorophyll a and turbidity using satellite reflectance data has application to a number of practical limnological problems.

(1) The suitability of the siting of existing sampling positions can be evaluated. Planning the distribution of sampling sites in an impoundment so as to be representative of prevailing conditions can be assisted.

(2) The synoptic information on chlorophyll a and turbidity distributions may assist in the siting of withdrawal points for water abstraction, as well as in the siting of recreational facilities.

(3) The ability of satellite remote sensing to detect sources of nutrient pollution leading to localised algal blooms can assist in studying the extent to which such pollution is dispersed, together with circulation patterns, in the water body. This aspect is of relevance to the siting of sewage outfalls.

(4) The synoptic data provided by CALMCAT may assist limnologists in studying the relationship between water quality conditions and nutrient inputs, in verifying and calibrating water quality models, and in evaluating the validity of assumptions.

(5) By running CALMCAT on historical Landsat images, using current calibration data, historical estimates of chlorophyll a and turbidity in impoundments may be obtained. This may assist in the detection of trends in water quality conditions.

Up to now limnologists have relied upon point measurements of chlorophyll a and turbidity in order to obtain information on these variables in an impoundment. The use of the CALMCAT model together with remotely sensed data now makes it possible for limnologists to obtain chlorophyll a and turbidity values for the entire surface of the impoundment. This should enable the behaviour of chlorophyll a and turbidity to be established with greater certainty than was previously possible .

CHAPTER 1

BACKGROUND INFORMATION

1.1 INTRODUCTION

The space race and the decision to get a man on the moon by the end of the 1960's started a trend in technology which has since proved to be an invaluable source of data of the Earth's resources. The space race stimulated the science of remote sensing defined as "the science and/or technique used in gaining information about material objects by means of measurements made over a distance without physical contact" (Liebenberg, 1977). The first space images showed how solar energy reflected by objects on the earth's surface could be measured and registered by remotely placed sensors, namely satellites. Together with the revolutionary progress in the field of spectroscopy and electronics, the sensors were not confined to capturing data in the visible spectrum and extended further into the infra-red range of the electro-magnetic spectrum. The value of satellite imagery was recognised and rapid growth took place in the field of remote sensing.

In 1972, Landsat 1, the first of the more important Earth resources satellite series, was launched (Curran, 1984). Subsequently, Landsats 2, 3, 4 and 5 have been put into operation providing near world wide coverage of the earth and its resources. Investigations have indicated a wide range of applications for which Landsat imagery can be used (Ackermann, 1974; House of Lords Select Committee on Science and Technology, 1983).

This report deals specifically with Landsat's application in the field of water quality, in particular, the detection and quantification of specific water quality conditions in impoundments. The critical nature of South Africa's water resources, Landsat's unique monitoring ability and the direct reception of Landsat data in South Africa, provided the impetus for the study.

1.2 SOUTH AFRICA'S WATER PROBLEM - A BRIEF REVIEW

South African water resource managers and planners face the problem of a severe shortfall of water by the year 2 000 (Draft Report on the Management of South African Water Resources, 1985). The concern is directly related to South Africa's position in the drought belt of the globe and its socio-economic standing as a fast developing country.

The high population growth rate, increasing urbanization and industrialization, rising expectations and standards of living, present a grave picture when combined with the scarcity and variability of the rainfall, which is the major source of water in South Africa (Whitmore, 1978).

It is therefore of major importance that the water resources of South Africa be managed and developed with maximum efficiency and speed. This entails maintaining the quality of established and new water supplies, developing new sources of water, and being able to quantify the water resources available at any one time. It is at this point that the lack of accurate, up to date data impedes efficient management. Researchers have suggested that the answer to the problem lies in the use of satellite imagery with its regularly recorded, accurate, synoptic data, available at low cost and in quantifiable terms (Kendrick, 1976; Malan, 1976; Skibitzke, 1976; Reed, 1978; Croteau, 1979).

Research in America and Europe has indicated that Landsat has been successfully used in water quality surveillance, particularly in detecting chlorophyll a (algal pigment) and turbidity (suspended solids) in water bodies (Bukata and Bruton, 1974; Moore, 1980; Lindell, 1981). An image received of an impoundment in South Africa, Bloemhof Dam, showed evidence of the above-mentioned water quality conditions. Plate 1.1 shows firstly, an algal bloom in the southern arm of the impoundment indicated by red patches visible on the surface of the water, and secondly, suspended solids are visible in the northern arm of the impoundment, identified by a bluish-white colour. The fact that the two conditions are visible on the image raises the questions; to what extent can quantitative information of water quality conditions be gained from satellite imagery? In addition, if one can quantify the distribution of chlorophyll a and turbidity, how is this information to be used in the management of the impoundment? In an attempt to throw further light on the subject, it was decided to investigate Landsat's potential for monitoring the possible deterioration of a South African impoundment by pollution in the form of sediment and nutrient-containing effluent from urban, industrial and farming sources.

PLATE 1.1: FALSE COLOUR IMAGE OF BLOEMHOF DAM SHOWING AN ALGAL BLOOM IN THE LOWER (SOUTHERN) ARM AND SUSPENDED SOLIDS IN THE UPPER (NORTHERN) ARM OF THE IMPOUNDMENT.

In 1981, the South African tracking and receiving station, the Satellite Remote Sensing Centre (SRSC) began receiving data direct from Landsat. The water quality project on Roodeplaat Dam was initiated and this report presents the methodology and results obtained.

1.3 LANDSAT

The Landsat series of satellites have a number of characteristics which has made them invaluable data captors (Lillesand and Kiefer, 1979). Positioned in a sun synchronous, polar orbit, and flying at a fixed altitude, varying between 920 kilometres (km) and 700 km depending on the specific Landsat concerned, Landsat has been equipped with a Multi-Spectral Scanner (MSS) which records energy returns of radiance from the earth in four spectral bands. These bands correspond to wavelengths in the visible, green and red and two bands in the near infra-red spectral regions : Band 4 = 0,5 to 0,6 µm*; Band 5 = 0,6 to 0,7 µm; Band 6 = 0,7 to 0,8 µm and Band 7 = 0,8 to 1,1 µm. The multiwaveband data are recorded in digital format of integer values, 0 to 255 inclusive.

The satellites provide synoptic views of the earth's surface. Each image covers an area of 180 km by 180 km and the same area can be imaged every 18 days. Regular monitoring can be carried out and the resolution of each picture element (pixel) on the earth's surface is 80 metres (m) by 80 metres. Thus for the first time rapid, regular, synoptic, quantifiable, low cost** data can be obtained of the everchanging features of the earth's surface. This means that objective, spatial comparisons can be made, inaccessible areas reached, and manpower, time and money saved. The major disadvantages are:

(1) the 18 day delay period between coverage,
(2) the fact that cloud cover obstructs Landsat images and
(3) that the imagery needs specific image processing equipment and expertise to take full advantage of the images' potential.

In the field of water quality Landsat has been used in a number of different applications (Rodda, 1976; Munday et al, 1980; Hill and Graham, 1980; Moore, 1980; Muralikrishna and Rao, 1982; Thiruvengadachari et al, 1983). A summary of some of the major applications are presented in Table 1.1.

---

* µm = micrometer.
** The cost of the computer compatible tapes (CCT) used in the study was ca., R365,00.

TABLE 1.1: APPLICATIONS OF LANDSAT-DERIVED INFORMATION IN THE FIELD OF WATER QUALITY

1. The measuring of and delineation in impoundments of:
   1.1 Particulate contaminants
   1.2 Chlorophyll concentration levels
   1.3 Turbidity concentration levels/suspended solids
   1.4 Circulation features
2. Assessing discharge plumes and adequacy of sampling point siting
3. Constructing and calibrating water quality models
4. Seasonal monitoring of impoundments
5. Regulatory permit monitoring

1.4 SPECIFIC WATER QUALITY CONDITIONS: CHLOROPHYLL a AND TURBIDITY

The two water quality conditions chosen for examination using Landsat data were chlorophyll a (algal pigment) and turbidity (suspended solids).

1.4.1 Chlorophyll a

Chlorophyll a is generally considered to be the most reliable measure of an impoundment's response to eutrophication (Lambou *et al*, 1982; Sartory, 1982). Chlorophyll is the primary green photosynthetic pigment present in algae and in all oxygen-evolving photosynthetic organisms (Wetzel, 1983). It is the algal plant pigment, chlorophyll a that the satellite detects and not algal biomass *per se*. The presence of chlorophyll elicits the red pseudo colouring seen on satellite images.

For the purposes of this report the occurrence of chlorophyll a is considered to be synonymous with the presence of algae.

Algae are microscopic aquatic organisms that grow extremely fast in the presence of plant nutrients such as phosphorus and nitrogen. Excessive algal growth is considered to be a major water quality problem (Toerien, 1975, 1977). The clogging of filters, flow meters, valves and irrigation canals may occur. Tastes and odours can be unpleasant and foul smelling scums on water surfaces are not conducive to recreational activities. Certain algae, under specific conditions, release toxins that can poison livestock (Powling, 1977). "For water management purposes, it is of value to have some means of predicting the degree of nuisance conditions that might be expected" (Walmsley, 1984).

It is therefore important to try and quantify chlorophyll a concentrations in an impoundment. Previously, estimations of chlorophyll a concentrations have been carried out using point source measurements and it has been recognised that satellite derived data, with synoptic and quantifiable advantages, can be of aid in determining the distribution of chlorophyll a concentrations with greater efficiency (Bukata and Bruton, 1974; Sydor et al, 1978; Welby et al, 1980; Canfield, 1983).

1.4.2 Turbidity

Turbidity is determined by the concentration, size, shape and refractive index of suspended particles (including chlorophyll a) which increase the amount of energy backscattered in water bodies (Moore, 1980). The presence of suspensoids is determined by the turbidity of the water which is recognised as bluish-white on satellite images. Turbidity may be associated with a number of effects. For instance, decreased light penetration can occur, therefore decreasing light in the photic zone which may inhibit rooted plant growth and algal productivity. On the other hand nutrients are associated with suspended sediments which, depending on the availability, serve as a food source and a stimulus for algal productivity.

Sediment laden waters also affect the treatability of water, sometimes blocking filters, pipelines and tunnels, while attempts to flocculate certain types of sediment can be expensive and difficult.

Investigations into sediment transport are important in understanding the hydro-dynamics of a water body and for the purposes of modelling the system.

Satellite imagery's synoptic and quantifiable data could be advantageous in assessing the turbidity in an impoundment.

1.5 OBJECTIVE OF THE PROJECT

The major objective of the project was to determine the potential and limitations for quantitative measurement of the distribution of turbidity (suspensoids) and chlorophyll a (algae) in a specific water body using Landsat data.

This meant that the relationship between the two water quality conditions and the satellites reflectance data had to be established. The aim being to assess the possibility of obtaining reasonable estimates of chlorophyll a and turbidity, using satellite derived data. The remote sensing technique provides a potential method of obtaining synoptic data on chlorophyll a and turbidity in impoundments not readily obtainable by other methods.

1.6 OVERVIEW OF THE REPORT

An evaluation of information and literature relevant to the topic is presented in Chapter 2. Chapter 3 presents the methods of analysis used to obtain accurate surface reference and satellite reflectance data and the statistical techniques used to analyse the data sets. The results of initial statistical analyses utilizing the Stepwise Discriminant Analysis and Canonical Correlation Analysis are presented, and an interpretion of results is attempted in Chapter 4. The problem of using statistical information derived from the Canonical Correlation Analysis and Linear Regression analysis for simulative purposes is looked at in Chapter 5 and the accuracy of the results obtained from the simulative equations are examined. A model, designed to estimate the distribution of water quality conditions in an impoundment using satellite reflectance data, and the synoptic view of conditions which is then available is presented in Chapter 6. Chapter 7 examines questions concerning the number of sampling sites required to be sampled in order to obtain reasonable results, and the possibility of extrapolating information from one day to another. Chapter 8 concludes the report, emphasising the most important aspects revealed in the study.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

"Monitoring is at the heart of the nation's water quality management effort; without it, enforcement and clean up programs can be of only limited effectiveness" (Sayers, 1971).

Investigation into the spatial and temporal nature of satellite data has shown that satellite imagery is a valuable monitoring tool for assessing and evaluating water resources (McGinnis et al, 1980). Table 2.1 presents a few of the fields of study in which satellite imagery has been applied. The potential applications would be greatly increased if the data detected by the satellite could be accurately quantified. The methodologies for quantification need to be based on sound reasoning, with an acknowledgement of the practical real life problems involved, a recognition of the limitations of statistics while adhering as much as possible to statistical assumptions, and attempting to obtain reliable, reproducible, predictive criteria that can satisfy the critical eyes of the scientific and legal communities (Latin et al, 1976; Lins, 1979).

This review presents some of the factors that remote sensing researchers, working in the field of water quality, have recognised as requiring attention if the quantification of water quality conditions using satellite imagery is to be successfully achieved.

2.2 SURFACE REFERENCE DATA CONSIDERATIONS

Landsat's ability to detect water quality conditions in an impoundment can only be accurately assessed when used in conjunction and calibrated with water quality data obtained simultaneously with the satellite's overpass (Anderson, 1979; Schaeffer et al, 1979; LeCroy, 1982; Whitlock et al, 1982; Khorram and Cheshire, 1983; Thiruvengadachari et al, 1983). Without accurate surface reference data, the relationship between water quality conditions and satellite reflectance data cannot be adequately calibrated and any inaccuracies in the data collection or analysis will result in erroneous conclusions.

2.2.1 Concurrent Collection of Surface and Satellite Data

The collection of water quality data carried out simultaneously with the satellite overflight is essential to avoid distortions due to significant atmospheric, hydraulic and solar influences (LeCroy, 1982; Whitlock et al, 1982). Some researchers have found this requirement to be physically and economically unachievable and yet procedures to correct for the time lapse between overflights and sampling have not been discussed in the available literature (Kuo and Blair, 1976). As short a time lapse as possible between the overpass and data sampling has been recommended by many investigators.

TABLE 2.1: WATER RESOURCES INVESTIGATIONS CARRIED OUT USING SATELLITE DERIVED DATA

INVESTIGATIONS INTO WATER BODY DYNAMICS

Reservoir monitoring
Seasonal changes
Circulation patterns
Establish current conditions of lake
Monitor nature, extent and source of possible changes
Set up, calibrate and verify real time estuarine water quality models
Mixing between fresh and sea water

HAVE BEEN RESEARCHED BY:
Gupta and Dodechtel, (1982)
Burgy, (1973)
Shih and Gervin, (1980)
Bukata et al, (1975(b), 1978)
Hill and Graham, (1980)
Rango et al, (1983).

INVESTIGATIONS INTO ENVIRONMENTAL PROBLEMS

Siltation/sedimentation
Distribution and transport of sediment
Trophic status/eutrophication
Biomass energy balance
Reliable alarm facility
Contaminants due to erosion, run-off and industrial discharge

HAVE BEEN RESEARCHED BY:
Boland, (1976)
Muralikrishna and Rao, (1982)
Moore, (1980)
Thiruvengadachari et al, (1980)
Munday et al, (1979)
Gupta and Bodechtel, (1982)
Welby et al, (1980)
Herschy, (1980)
Sydor et al, (1978)
Witzig and Whitehurst, (1981)
Verdin, (1985).

INVESTIGATIONS INTO RESOURCE MANAGEMENT

Tourism/recreation
Commercial
Agriculture
Irrigation
Water quality
Simultaneous view of other water bodies
Comprehensive data base
Planning and evaluating the results of water management activities
Urban water resources planning.

HAVE BEEN RESEARCHED BY:
Scarpace et al, (1979)
Khorram, (1981)
Thiruvengadachari et al, (1980)
Carpenter, (1982)
Jackson and Ragan, (1977).

OTHER INVESTIGATIONS

Cheaper alternatives for limnological surveys
Flood control
Ground water recharge
Drainage networks
Establishing and enforcement of regulations
Understanding the system
Environmental impact of land use practices within surrounding environment
Relaying hydrological data.

HAVE BEEN RESEARCHED BY:
Lindell, (1981)
Thiruvengadachari et al, (1983)
Jarman, (1973)
Khorram, (1981)
Paulson, (1974)
Morgan et al, (1981).

2.2.2 Sampling Depth

The problem of depth is considered to be very important where bottom depth in shallow waters can influence relationships and reflectance values (McCluney, 1974; Whitlock et al, 1978; Khorram, 1981; Lillesand et al, 1983). Alternatively sampling to secchi disc depth (m) has been considered appropriate (Thiruvengadachari et al, 1983). Secchi disc measurements give an indication of water clarity but are criticised because of their crudity and the fact the measurement is dependent on cloud cover and an individual's acuity of vision (Scarpace et al, 1979; Lillesand et al, 1983).

2.2.3 Identification of Sampling Sites

The identification of sampling sites is a problem faced by many researchers, particularly when repetitive coverage from the same sampling point is required. Vandalism often prevents the use of marker buoys and determining sampling positions from landmarks can be difficult, especially when sampling is carried out on large water bodies. The importance though, of accurately locating sampling sites in order to relate to specific satellite pixels, cannot be over-emphasized and is discussed in Section 2.4.4.

2.2.4 Sampling Design

Daniel and Wood (1971); Harris et al (1976); Boland et al (1979); Carpenter (1982); Mace (1983) and Van Belle and Hughes (1983) all point out that the design of the sampling program is crucial. A major source of error can be introduced into data if the surface reference data obtained is not representative of the whole range of water quality conditions present in the impoundment at that time. Prior knowledge of the idiosyncrasies of the impoundment is highly advantageous. Mace (1983), suggested that sampling points should be located "to minimise the number of points necessary to characterise lake water quality and to ensure that their distribution matches both the variability inherent in the water and the resolution of the remote sensing system". Thornton et al (1982), point out that the "sample design should allow the characterization of the system as well as permit comparative evaluation through time and/or across systems". Whitlock et al (1982), are even more specific by establishing that sampling points should be located in such a way that Daniel and Woods' (1971) criterion is satisfied for all bands in the regression equation. Details on this issue are discussed in Section 2.5.2.

2.2.5 Preservation, Storage and Analysis of Surface Reference Data

Thiruvengadachari et al (1983) point out that a fundamental step in the collection of water quality samples is the preservation, storage and analysis of the data. Analyses should be consistent and accurate and should be performed as soon as possible after data collection (Whitlock et al, 1978). Turbidity and chlorophyll a samples, in particular, change irreversibly due to inadequate preservation and storage (Sartory, 1982).

2.3 WATER QUALITY PARAMETERS

In an attempt to determine the range of variables that can be detected by remote sensing satellites, a number of water quality parameters have been investigated for possible correlation with reflectance data. Chlorophyll a and turbidity are two water quality conditions that can be directly sensed by Landsat (Harris et al, 1976; Carpenter, 1982; Ulbricht, 1983), refer to Section 1.4. Parameters which directly affect reflectance values in one or all of the four wavebands can be directly measured when calibrated with surface reference data. For the purpose of this report these parameters will be called 'direct' parameters. The parameters which do not themselves affect reflectance values, but instead do so via one of the direct parameters will be called 'indirect' parameters (Iwanski et al, 1980). Table 2.2 presents a list of parameters that have been investigated.

2.3.1 Water Quality Conditions as Sensed via Indirect Parameters

Of all of the parameters investigated most do not reflect light in the range measured by the satellite. Indirect relationships between some of the variables, e.g. the presence of chlorophyll a (direct parameter) and the presence of phosphorus and nitrogen (indirect parameters) can indicate correlations with the reflectance bands, but correlations can also be spurious. For instance, Khorram and Cheshire (1983), in their work on the Neuse River Estuary in North Carolina, U.S.A. indicate that reflectance data was significantly correlated with salinity. It is unlikely, however, that salinity can be detected by the satellite, instead it may be a variation of chlorophyll a associated with variations in salinity that the satellite is measuring. In addition, an indirect relationship such as this can really only be applied in a steady state condition where the situation is localised.

Grimshaw et al (1980), report that variables such as log total orthosphosphate and log total alkalinity did not contribute to the multiple regression in a highly significant manner. Similarly Shimoda et al (1984), found that there was no correlation between the indirect variables of oxygen saturation, acid soluble calcium concentrations and acid soluble magnesium concentrations and satellite reflectance data.

2.3.2 Direct Water Quality Conditions: Chlorophyll a and Turbidity

The presence of chlorophyll a and turbidity in impoundments has long been recognised on satellite imagery and attempts have been made to quantify these conditions (Yarger et al, 1973; Bukata and Bruton, 1974; Kritikos et al, 1974; Bukata et al, 1975(a); Rogers et al, 1975; Harris et al, 1976; McHenry et al, 1976; Ritchie et al, 1976; Stortz et al, 1976; Chagarlamudi et al, 1979; Munday et al, 1979; Moore, 1980; Welby et al, 1980; Lindell, 1981; Carpenter, 1982; Ulbricht, 1983; Hilton, 1984). The results of Scarpace et al, 1979; Schaeffer et al, 1979; Sheng and Lick, 1979; Hill and Graham, 1980; Iwanski et al, 1980; have all shown that Landsat does indeed measure chlorophyll a and turbidity.

TABLE 2.2: EXAMPLES OF WATER QUALITY PARAMETERS SO FAR INVESTIGATED

(a) Parameters with direct correlation with reflectance intensity.

Chlorophyll a/algae
Suspended sediments
Temperature (sensor dependent)
Secchi disc depth
Light transmission
Colour
Sediment transport and circulation patterns
Particulate organic carbon
Phaeopigments
Chlorophyll c and b
Iron
Tannin and lignin

(b) Parameters with indirect correlation with reflectance intensity

Conductivity
Salinity
Dissolved oxygen
Alkalinity
Calcium
Nitrite
Nitrate
Magnesium
Total organic carbon
Dissolved organic carbon
Total phosphorus
Total nitrogen
Ammonia
Kjeldahl nitrogen
Dissolved and total orthophosphate
Wind speed
Filtered and unfiltered water

Bukata et al (1974), determined that band 4 clearly delineated the bottom contours of the impoundment, if surface turbidity was relatively low and the maximum optical penetration was over 14 metres. Band 5 was found to have a linear correlation with turbidity while bands 6 and 7 measured surface chlorophyll a for concentrations of 4 mg/m^3 or more. Muraliskrishna and Rao (1982), indicated that bands 6 and 7 correspond to surface features whereas bands 4 and 5 offer information on subsurface features. Further research has confirmed these claims (Harris et al, 1976; Bartolucci et al, 1977; Scarpace et al, 1979; LeCroy, 1982).

2.3.3 Algae

Lindell (1981), analysed the reflectance signals with respect to different algal species dominating different parts of a lake and indicated that the different types of algae may emit different reflectances. Hilton (1984), reports that limited work has been done on obtaining equivalent multispectral scanner spectral intensities of the major algal groups and the research that has been carried out is usually under laboratory conditions or from aircraft remote sensing (Lekan and Coney, 1982). Hilton (1984) suggests that "the use of sensors with more channels could allow spectral signatures of different groups of algae to be typed and mapped. It is unlikely that remote sensing will ever get down to genus level let alone species level but it could be useful in improving sampling strategy".

Yentsch and Phinney (1982) and Carpenter (1982), make the point that the measurement of chlorophyll a is a combination of degradation products, phaeopigments, detrital material and dissolved fluorescent material, and that the variability in accessory pigmentation can be a major source of error in any analysis. The variations in chlorophyll that occur from lake to lake and within lakes may provide a key to the assessment of satellite data (Carpenter, 1982). On this issue little research has been carried out (Witzig and Whitehurst, 1981; Carpenter, 1982; Hilton, 1984).

2.3.4 Vertical Migration of Algae

Algae migrate vertically in the water column in a light orientated response (Wetzel, 1983). This vertical migration contributes significantly to the changing conditions in the impoundment and can cause a distinct error in the surface reference data collected hours or days after the satellite overpass (Klemas, 1976). Harris et al (1976), report that "it is possible that surface values of chlorophyll may be nowhere similar to the samples taken due to the microstratification of phytoplankton". Ulbricht (1983), expresses the view that algae can only be detected by satellites due to their presence near or on the surface of the water by bands 6 and 7, and just under the surface by bands 4 and 5.

2.3.5 Horizontal Movement of Algae

Mainly as a result of wind and water movement, horizontal variations in algae are fairly significant, particularly in close proximity to the littorial zone (Wetzel, 1983). Wind has a great impact on algal movement and on the resuspension of algae. For example, diatoms are particularly heavy and therefore usually only mix and move in response to wind or current action (Mrs. S. Young* - pers. comm.). Inaccuracies therefore in the detection and quantification of chlorophyll a could result from horizontal variations.

2.3.6 Activity Stages of Algae

Algae cells increase, sink, rest, metabolize and decay and depending on the proportion of cells in the various physio- logical stages, the rate of growth and size of the population will differ (Wetzel, 1983). Scarpace et al (1979), report that a eutrophic lake could be classified as being oligotrophic if it were sampled the day after a large algal bloom had died off. Phaeopigments are the products of chlorophyll breakdown, in other words, of decaying algae. The relative proportions of phaeopigment to the total chlorophyll a pigment is an indication of the health of the algae in an impoundment. The detection of phaeopigments using satellite data is something that has not been pursued in any great detail but it is possible that differences in reflectance between healthy and decaying algae may exist. This matter is briefly examined in Section 4.3.

2.3.7 The Relationship between Algae and Turbidity

The remote sensing researcher should be aware of the fact that sunlight, as well as nutrients such as nitrogen and phosphorus, is required for algae to grow. The presence therefore of suspended sediments can have two conflicting effects on algal growth. Firstly, the prevention of light penetration by sediments in suspension inhibits algal productivity. Secondly, "phosphate adsorbed onto sediments can make up a large proportion of the total phosphate available for algal growth in an impoundment" (Grobler and Davies, 1981). Harris et al (1976), report that the presence of suspended solids undoubtedly interferes with the reflectance values of low chlorophyll levels and may overwhelm small to moderate values of surface chlorophyll. These researchers indicate that if there is low radiance°, surface algae will show up in bands 4 and 5 whereas if there is high radiance it will show up in bands 6 and 7.

---

* Mrs S. Young - Hydrological Research Institute, Department of Water Affairs.

° The terms radiance and reflectance are used interchangeably in the literature although radiance does imply a change in wavelength reflectance.

In addition turbidity is a measure of light penetration and absorbance and therefore inorganic as well as organic suspended solids will be included in the turbidity measured by nephelometry. A problem can arise therefore in situations when there is a low turbidity and high chlorophyll a (Holmquist, 1977; Hilton, 1984; Verdin, 1985).

It becomes apparent that the relationship between turbidity and algae is complex and highly interrelated. This problem has been recognised by scientists and the consequences of this interrelatedness will be discussed in Section 2.4.7.

2.4 SATELLITE REFLECTANCE DATA CONSIDERATIONS

Landsat's sensors (refer to Section 1.3) were specifically designed and optimised for observations of land cover and terrestrial resources rather than for water resources. Therefore the data received from the satellite have not been considered to be well suited to aquatic applications (Carpenter, 1982; Hilton, 1984). Nevertheless, the potential offered by such a data source could not be overlooked and, despite the difficulties, a great deal of research has been undertaken in the water resources field (Ackermann, 1974; Skibitzke, 1976).

The major factors to be considered when using satellite reflectance data in the field of water quality determination are discussed below.

2.4.1 Corrections

In the process of capturing and transmitting data from and to earth, Landsat MSS data can be distorted, mainly due to satellite or terrestrial effects or limitations in the sensor systems. "Radiometrically, the digital numbers do not always accurately relate to scene energy levels; geometrically, image positions of features do not accurately relate to map positions" (Lillesand and Kiefer, 1979).

2.4.2 Geometric Corrections

In order to obtain quantitative results and to enable precise registration of an image with reference points, the correction of major distortions inherent in Landsat MSS data are considered to be necessary by most researchers (Schaeffer et al, 1979).

The distortions are mainly a result of, firstly, the satellites variation in altitude, attitude and velocity. Secondly, of the sensors detectors, optics and scan mechanism and lastly, variations in terrain, perspective and map projection (Palmer, 1981). Standard geometric transformations are usually applied to correct the data.

2.4.3 Radiometric Corrections

The radiance values obtained from MSS data are not always equivalent to ground reflectance values due to atmospheric attenuation, haze and the angle of the sun. These three factors are a major source of error and many algorithms have been suggested in order to correct images (Bukata et al, 1974; Holmquist, 1977; Welby et al, 1980;

Aranuvachapun and Le Blond, 1981; MacFarlane and Robinson, 1984). Atmospheric effects and haze have been estimated by researchers using radiation values obtained from airports and clear lakes in an attempt to standardise reflectances (Holmquist, 1977; Scarpace et al, 1979; Lillesand et al, 1983; Verdin, 1985). Problems occur though, where features of this nature, are not present or are too small to recognise.

The position of the sun at the time of image capture has a strong influence on reflectance values (Ritchie et al, 1976). Carpenter (1982), proposes that the sun's elevation is an important predictor in models of turbidity and chlorophyll pigments. Oppositely Munday et al (1979), report that the solar angle has a negligible effect on their regression analysis and instead use a method of data reduction known as chromaticity analysis, which permits the adjustment of atmospheric variation between dates. The Munday et al noise correction technique "suppresses noise when all bands suffer radiance changes in the same proportion while leaving spectral properties of the data unaffected". These researchers claim that their technique can be applied to new data that lack surface calibration for standardizing the data. LeCroy (1982), recommends that "variations in solar zenith angle should be normalized or accounted for in the data reduction process as well as atmospheric effects".

Verdin (1985) proposes that "Failure to account for atmospheric effects when working with multidate imagery can potentially lead to erroneous assessments of reservoir trophic state".

Sometimes a malfunctioning detector may cause image lines to be defective thereby resulting in a striping or banding effect on the image. Striping in the reflectance data can have a big influence on water quality monitoring due to the low reflectances of water (Shimoda et al, 1984). In order to establish a greater degree of uniformity, researchers have recalibrated the data to improve homogeneity as in the case of Carpenter (1982), or have used statistical techniques such as mean and standard deviation matching, histogram equalization and random noise additions as proposed by Shimoda et al (1984).

2.4.4 Sampling Site/Pixel Registration

Substantial errors can be introduced into an analysis of the relationship between water quality conditions and satellite reflectance data if sampling positions are inaccurately located and registered with the satellites pixels values (Carpenter, 1982; Lillesand et al, 1983; Mace, 1983; Verdin, 1985). Attempts to overcome this problem of misregistration have been tackled in various ways. Munday et al (1979), and Grimshaw et al (1980), undertook surveys of all their sampling positions. Khorram and Cheshire (1983), located sites on a nautical chart. Many other researchers have used a pixel averaging system so that the effect of possible inaccuracies of locating a sampling site would be minimised (Shih and Gervin, 1980). Pixel averaging is a widely recognised technique used to smooth data. Averaging effectively increases the size of the resolution element (thereby introducing new errors) and supposedly removes random errors and noise without substantially degrading the imagery (Whitlock et al, 1982; Mace, 1983). Variations between 36 pixel averages (Bukata et al, 1975(a)) and 3 x 3 pixel windows have been used (LeCroy, 1982; Lillesand et al, 1983; Shimoda et al, 1984).

In most cases research has shown that the use of more than one pixel value is necessary and averaging of the values helps minimise the uncertainty and possible spurious variations due to the inexact location of sites.

2.4.5 Water/Land Delineation

Although water bodies are usually easily recognisable on a Landsat image, delineation of the water/land boundary is sometimes not clear. This indistinction can be due to reed or swamp beds growing along the edges of the impoundment, the presence of algae or turbidity, or the 80 m resolution of the satellite which can pick up mixels. Mixels, in this instance, are a mixture of land and water pixels.

In order to delineate water surfaces, band 7, which shows up the greatest difference between land and water, is often used to delimit the boundary. Schaeffer et al (1979), Thiruvengadachari et al (1980), Lindell (1981), Hilton (1984), Khorram and Cheshire (1983), Lillesand et al (1983) and Mace (1983) are a few of the researchers who have used band 7 as a means of separating water from land.

Supervised classification is a technique whereby a researcher identifies surface cover categories visible on the image, either by the chromatic signature or by surface reference data, and uses a computer based routine to convert reflectance data into sets of specific, discrete classes (Lillesand and Kiefer, 1979). Supervised classification has also been used to determine the land/water boundary (Bukata et al, 1974; Muraliskrishna and Rao, 1982; Graham and Hill, 1983).

2.4.6 Colour Coding

The interpretation of relative differences of water quality conditions in an impoundment can be improved by certain image enhancement techniques. A technique frequently used is colour coding which represents reflectance values as colours, with a colour code representing a concentration scale. Therefore, different colours correspond to designated reflectance values and in turn interpretations can be made on the basis of relative amounts of each colour present (Iwanski et al, 1980; Khorram, 1981; Shimoda et al, 1984). Experience is required to classify subtle differences in colour (Holmquist, 1977; Lindell, 1981), but colour coding can often give an immediately discernable picture of differences in an impoundment.

2.4.7 Multicollinearity

A very important feature of Landsat's MSS, with respect to examining water quality conditions, is the multicollinearity of the reflectance data (Shih and Gervin, 1980). Some researchers have established that Landsat reflectances vary nonlinearly with suspended solids concentrations (Munday et al, 1979), and that the conditions can be detected in all 4 wavebands (LeCroy, 1980). An inspection of colour coded images indicates that water quality conditions are usually visible in at least three of the four bands depending on the concentrations. Bands 4, 5 and 6 are often correlated to turbidity while band 7 usually correlates with high concentrations of chlorophyll (Bukata et al, 1975(a); Harris et al, 1976; Holmquist, 1977; Boland et al, 1979; Munday et al, 1979; Muraliskrishna and Rao, 1982).

It has been established that information from more than one band width is required to predict water quality conditions with any reasonable degree of accuracy (Grimshaw et al, 1980).

2.4.8 General

Finally, on a general note, many researchers mention difficulties associated with unpredictable circumstances such as satellite failure, uncoordinated overflight/sampling operations and the inability to read computer compatible tapes.

2.5 STATISTICAL CONSIDERATIONS

2.5.1 Introduction

The ultimate objective of researchers in the field of water quality research using remote sensing, has been to determine the best techniques available for quantifying water quality conditions, with the aid of satellite reflectance data. Statistical analyses in the form of multi-variate statistical techniques are necessary to define and calibrate data analysis algorithms (Boland et al, 1979; Scarpace et al, 1979; Shih and Gervin, 1980; Lindell, 1981; Whitlock et al, 1982; Khorram, 1981; Lillesand et al, 1983; Mace, 1983; Shimoda et al, 1984).

Certain statistical assumptions form the basis of the least squares procedure in multi-linear regression analysis. Daniel and Wood (1971), discuss the necessary assumptions. Firstly, the suitable form of the equation should be chosen, namely that the independent variables are in linear relationship with the reflectance bands (Grant, 1983). Secondly, the data should be representative of the whole range of combined environmental conditions and variables under investigation. Thirdly, the surface reference data should be uncorrelated and statistically independent. Fourthly, all observations should have the same variance. Fifthly, all the conditions should be defined with as small an error as possible, and lastly, if uncontrollable error occurs, then the distribution of such error should be a normal one (Whitlock et al, 1982; Van Belle and Hughes, 1984). It becomes clear that there are some important issues associated with the statistical representation of the relationship between water quality variables and satellite reflectance data.

2.5.2 Representativeness

The poor fit sometimes achieved between water quality variables and reflectance data in regression equations has been of vital concern to many researchers. This problem has been attributed to an inadequate range of water quality data used to obtain the regression equations (Boland et al, 1979; Carpenter, 1982). A requirement of regression analysis, if it is to provide useful simulative or predictive models, is that the parameters should cover a representative range of values (Carpenter, 1982). Over a restricted range of values, a statistical relationship can only be valid if conditions are constant. The greater the range of data obtained covering the full complement of conditions, the more successful the equation or model will be. In an effort to space data evenly along the full range of values, sampling sites have been chosen specifically to include a representative range of problem areas (LeCroy, 1982). Lillesand et al (1983), have gone to the other extreme and avoid any apparent extraneous scene element such as algal blooms that could cause anomalies in the relationship. This approach can be criticised in that deliberate exclusion of high concentrations of algae will cause inaccurate calibration of water quality conditions. In addition, a standard practice is to transform to logarithm turbidity and chlorophyll in an attempt to reduce the variance caused by larger values. Log transformations have been used by numerous researchers (Munday et al, 1979; Grimshaw et al, 1980; Aranuvachapun and Le Blond, 1981; Carpenter, 1982).

It is, therefore, imperative that information on the full range of values present in the water body at the time of sampling, be included in the statistical analysis in order to ensure statistical representativeness of the relevant conditions (Boland et al, 1979; LeCroy, 1982).

In analysis, the data may form clusters and one cluster may often be equivalent to only one point in the determination of a regression slope (Daniel and Wood, 1971). A spuriously high value of r might result due merely to the heterogeneity of the data and the coefficient of determination will mean nothing because the data are not uniformly distributed (Kenney and Keeping, 1954). Mace (1983), failed to produce significant results possibly due to the fact that the sampling points were clustered in one area.

Whitlock et al (1982), recommended a Rule of Thumb Criterion suggested by Daniel and Wood (1971), for use in regression analysis for remote sensing of sediments and pollutants. Daniel and Wood's Criterion is based on variances such that:

> "the variance of radiance about the mean for the ground truth locations (σ^2RADi) be at least 10 times the variance of data noise (σ^2Ni)" (Whitlock et al, 1982).

This criterion implies that a great number of clustered data points can be excluded. While this may be the approach to use where there is uncontrolled noise, it can be criticised because there is always error in ground truth/surface reference data.

Boland et al (1979), report that they made an effort to obtain a normal distribution of surface reference data but not the reflectance data. Exactly how this was attempted is not clear. LeCroy (1982), specifically chose sample sites to include problem areas, in an effort to space the data evenly between extremes for more accurate statistical representation. Statistical normalisation was used on the data to eliminate noise due to atmospheric, solar and system variability.

2.5.3 Homogeneity

A second factor very closely related to the statistical problem of representativeness is that of non-homogeneity. The inhomogeneous distribution of water quality conditions, inherent in all data collections, is a source of error in statistical analysis (Munday et al, 1979). Often more than one statistical population of a variable is present in an impoundment. This can be due to many factors, for example;

(1) a variety of algal species;
(2) differing activity stages of algal species;
(3) changes in temperature, currents and wind;
(4) the presence of suspended sediments;
(5) complex mixing, scattering and absorption processes in the water column;
(6) morphometric differences in the water body.

2.5.4 Outliers

The presence of outliers, often single data points, can seriously bias the estimated slope of the regression line. (Draper and Smith, 1966; Daniel and Wood, 1971; Munday et al, 1979; Whitlock et al, 1982; Mace, 1983). Munday et al (1979), recognised the possible skewness and possible non-linearity in their data. Therefore they removed data, one set at a time, where suspended sediment values were in excess of 1000 µg/ℓ. Munday et al, report that even after removing data their r values were still high, which indicated that neither high suspended sediment values nor any single data point had an unduly large influence on the results.

Thiruvengedachari et al (1983), made sure that their sample size was large and a factor of redundancy was built in so that "unsuitable" sampling points could be edited out. Mace (1983), although acknowledging the problem that "regression equations are heavily dependent upon a single data point", included outliers and therefore the results are questionable.

2.5.5 Interdependence

The interdependence between the four reflectance bands and between the surface reference variables amounts to dealing with a high level of multicollinearity (Grimshaw et al, 1980; Whitlock et al, 1982; Khorram and Cheshire, 1983;). Some researchers have ignored this factor and have excluded reflectance bands with the idea in mind that the band excluded contributes very little to the relationship or does not contain satisfactory information (Carpenter, 1982; Mace, 1983; Verdin, 1985). Carpenter decided that band 7 data contributed very little to the relationship and therefore excluded band 7 from his models. Boland et al (1979), determined that although band 7 had the "poorest discrimination and the lowest information content of any of the bands, it weighed heavily and consistently in all models. When this band was excluded, the resulting models were statistically unsatisfactory". It was also recognised that by using ratios of bands in a model, the ratios "contributed nonlinear components into the regressions" (Boland et al, 1979).

Other researchers have employed linear regression techniques that regress one Y variable with multiple X variables (Holmquist, 1977; Aranuvachapun et LeBlond, 1981; LeCroy, 1982). Shih and Gervin (1980), used ridge regression analysis to eliminate multicollinearity.

Some researchers have recognised the importance of the multicollinearity evident between water quality variables and reflectance bands, and attempts have been made to examine all of the statistical parameters involved (Witzig and Whitehurst, 1981; Whitlock et al, 1982; Khorram and Cheshire, 1983). Carpenter (1982), suggested that canonical variate analysis could be undertaken to establish quantitative relationships.

2.5.6 Statistical Techniques

A variety of statistical techniques have been used in attempts to calibrate reflectance values with surface reference data. The variety and number of methods illustrates the difficulties experienced in determining the relationship and emphasises the point that more than one approach can be used to analyse the data. Often, the types of analyses used by researchers are not clearly recognisable and information has to be gleaned from abbreviated reports. Table 2.3 presents information obtained from a few of the literature reports in an attempt to give some idea of the techniques used. The groupings are not mutually exclusive, the various analyses and list of researchers are by no means exhaustive and are given by way of illustration. Classification procedures, enhancement techniques and pattern recognition have been used by many other researchers in place of, or in conjunction with, statistical techniques (Jarman, 1973; Kritikos et al, 1974; Bartolucci et al, 1977; Holmquist, 1977; Fisher et al, 1979; Gupta and Bodechtel, 1982; Lekan and Coney, 1982; Muralikrishna and Rao, 1982; Ulbricht, 1983). The comments are made by the authors and may be of value to the experienced investigator in this field of study.

2.5.7 The Modelling of Water Quality Conditions using Satellite Reflectance Data

Multiple linear regression models have been produced to assess the applicability of the regression relationships that have been established between water quality conditions and satellite reflectance values.

Boland et al (1979), undertook an extensive comprehensive study of selected Illinois water bodies. Multiple regression models were produced and the difficulties encountered were:

(1) the non-normality of the ground truth and MSS data and
(2) the limited range of the digital values in the MSS bands.

TABLE 2.3: STATISTICAL TECHNIQUES ILLUSTRATING THE APPROACHES USED BY OTHER RESEARCHERS

TECHNIQUE	REFERENCE	COMMENTS
Multivariate Statistics		
Multiple linear regression with ratios of various bands	Grimshaw et al, 1980	Predominantly bands 4 and 5 for chlorophyll and bands 4, 5 and 7 for turbidity.
Multiple linear regression with bands, squares of reflectances and ratios	Lillesand et al, 1983	Used Trophic State Indices transforms, complicated ratios and squares.
Multiple linear regression with ratios and squares of reflectance	Khorram and Cheshire, 1983	Complicated signatures Chlorophyll associated with band 4.
Multiple linear regression with mean reflectance band values	Khorram, 1981	Suspended solids model excludes band 4. Band 7 excluded in most instances.
Multiple linear regression with ratio and quadratics	LeCroy, 1982	Best results obtained using all four reflectance bands
Multiple linear regression	Shimoda et al, 1984	Image signatures and indirect parameters
Multiple linear regression	Whitlock et al, 1982	Includes inherent upwelled radiance. Comprehensive with considerations well discussed.
Best subsets multiple linear regression	Mace, 1983	Mean radiance values excluding band 7.
Multiple linear regression Cluster analysis Principle Components Analysis	Schaeffer et al, 1979	Spectral ranking of many lakes. Spectral signatures developed.
Multiple linear regression Principle Components Analysis	Carpenter, 1982	Complex but comprehensive. Excluded band 7 but included sun angle and time of data collection.

Continued ....

TECHNIQUE	REFERENCE	COMMENTS
Stepwise multiple linear regression Principle Components Analysis Cluster Analysis Ratios	Boland _et al_, 1979	Comprehensive and detailed analysis.
Stepwise Multiple linear regression	Thiruvengadachari _et al_, 1983	Detailed discussion of considerations.
Stepwise Multiple linear regression	Iwanski _et al_, 1980	Included indirect parameters.
Ridge Regression Analysis	Shih and Gervin, 1980	Includes all reflectance bands. Comprehensive analysis.

<u>Linear Regression</u>

Linear regression	Bukata _et al_, 1974	Linear relationships determined.
Linear regression	Harris _et al_, 1976	Atmospheric conditions included.
Linear regression	Ritchie _et al_, 1976	Included suns angle and solar radiation.
Linear regression	Stortz _et al_, 1976	Band 5 related to turbidity.
Linear regression	Chagarlamudi _et al_, 1979	Ignored band 7. Used quantative brightness (QB), QB = $\sqrt{\text{Standardised bands 4, 5, 6}}$ and secchi disc.
Linear regression	Scarpace _et al_, 1979	Complicated by the introduction of atmospheric corrections.
Linear regression	Shih and Gervin, 1980	Comprehensive comparison.
Linear regression and time series analysis	Holmquist, 1977	Atmospheric corrections included
Linear regression and chromaticity analysis	Lindell, 1981	Qualitative
Linear regression and chromaticity analysis	Munday _et al_, 1979	Comprehensive analysis

The models that were developed provided relative estimates of chlorophyll and secchi disc depth (as well as other factors) and were used to develop generalised rankings of trophic levels of lakes.

Carpenter (1982), investigated lakes in Australia and attempted to model one reservoir using six days of data. In order to generate models that were not date specific, two significant predictions were included; the sine of the sun's elevation angle, in an attempt to account for variation in scene brightness between different dates; and the time of sampling, as a predictor for pigment (Carpenter and Carpenter, 1983).

Carpenter reported that the models were successful, modelling both turbidity and pigment very accurately. When testing the models on data from other lakes in the close vicinity, turbidity was successfully predicted but not pigment, thereby indicating that the turbidity model could be extended to other lakes. Carpenter produced a map of predicted turbidity distribution for the whole lake which he felt was accurate but added a note of caution that "if the point of interest is representative of a new regime not encountered in the generation of the model then the prediction may not be at all accurate" (Carpenter, 1982).

Khorram and Cheshire (1983), undertook a study of the Neuse River Estuary, North Carolina. They investigated the use of models previously developed in other areas, to see if they applied to their own geographical area, without success. These researchers then attempted to determine the parameters for the models by carrying out correlation analysis between all water quality parameters, the four MSS bands and their "typical band combinations and ratios". The regressions which best represented the relationship between water quality measurements and mean MSS data, were chosen to be verified with data not previously used in the model development. Khorram and Cheshire report that the R^2 values for chlorophyll a and turbidity were low.

In the search for reasonably accurate models to predict water quality conditions using satellite data it becomes apparent that the following factors are important.

(1) It is important to choose models that make physical sense or else the data are forced to fit models where there may be no physical reality.

(2) A good statistical fit does not guarantee a model which will give a good predictability because the fit considers only the set of data used to derive the relationship (Lillesand et al, 1983). Therefore outside the range of data for which it was developed the model can be expected to have poor applicability.

(3) Multi-variate multiple linear regression statistical techniques are required which include all of the reflectance bands and the water quality parameters.

(4) Corrections for sun's angle effects and haze are advisable.

CHAPTER 3

METHOD OF ANALYSIS

3.1 POINTS OF INTEREST

Quantitative analysis of water quality conditions in an impoundment using satellite derived data requires that the relationship between the water quality variables and satellite reflectance values be calibrated. Factors that can affect the relationship need to be recognised and taken into consideration. Some of the more pertinent factors have been recognised and will be discussed in an attempt to clarify and justify the method of analysis that has been used in this report.

3.1.1 An Ideal Experiment

In order to obtain the ideal calibration between surface reference data and satellite reflectance data one requires an ideal experiment. For example, this experiment would consist of a number of experimental tanks, 200 m x 200 m in size, each of which should contain a specific concentration of chlorophyll/turbidity, from low to high concentrations covering the full range of possible concentration values. Therefore at the times of the satellite overflights the complete range of chlorophyll/ turbidity values could be identified and the reflectance values obtained.

The ideal, however, is practically and economically not feasible. Instead a dynamic water body serves as the 'tank' or rather, 'range of tanks' and the problems associated with the real world situation have to be faced. Some of these problems are found in the form of assumptions.

It is generally assumed that the data collected in a sampling situation are representative of the conditions present at that time. This is not necessarily true and certain questions can be asked. Firstly, can it be assumed that a one litre sample of water taken from the side of a boat is representative of the 6400 m^2 area around it and the possible variations within that area? Secondly, can one pixel be representative and relatively homogeneous of the conditions surrounding it? Thirdly, can the presence of outliers and conversely, the presence of many values clustered into one specific value range, bias the analysis? Lastly, can it be assumed that the impoundment will contain the full range of possible water quality conditions, at the time of the satellite overpass?

The problems arise for two main reasons. Firstly, due to the very nature of the sampling operation, the surface reference sampling sites are usually established on a random basis before the satellite overpass and therefore do not always necessarily represent the distribution of conditions in the impoundment. The distribution of water quality conditions is not stable and often results in scattered and noisy pixels. For instance, the chlorophyll on the surface of the impoundment can vary greatly depending on winds, currents, algal species, temperature etc. Secondly, there may be marked gradients in the distribution of the water quality conditions due to an impoundment having different nutrient inputs. Consequently, this can result in an area of the impoundment, with the higher nutrient input, having higher chlorophyll values than the rest of the impoundment. As a result of these two factors it is not possible to know the actual distribution of water quality conditions in an impoundment until after the satellite overpass and the sampling operation. In essence, therefore, the successful calibration of the relationship between specific water quality conditions and satellite reflectance data depends on:

(1) the representativeness of the data set,
(2) the homogeneity of the data set and the exclusion of outliers,
(3) the assumption of one statistical or biological population.

These points are discussed more fully below.

3.1.2 The Representativeness of the Data

It is important to note that "representativeness" of a range of chlorophyll/turbidity values and "representativeness" of the surface area of the impoundment are two entirely different concepts. Representativeness of the chlorophyll/turbidity range means that the full range of water quality from low to high concentration values need to be obtained and therefore their corresponding low to high satellite reflectance values. Representativeness of the surface area of the impoundment is a geometric concern where the impoundment is divided up into equal areas which may or may not give a representative sample of water quality conditions. The prime concern of this work is with the former in terms of chlorophyll a and turbidity. The crux of the matter revolves around ensuring that one's data adequately represents all ranges of conditions present in the impoundment in an unbiased way.

For example, the situation may arise whereby, only one or two sampling points represent a large area of a particular chlorophyll a range, and conversely a large number of points may represent a smaller area of another chlorophyll a range. Therefore when analysing the data, the larger number of points will weight the equation incorrectly, a spuriously high value of r might result and while true for the sample set, there may be no physical reality or meaning for the parent population (Kenney and Keeping, 1954; Witzig and Whitehurst, 1981; Whitlock et al, 1982). An analogy may be seen in the situation when investigating land use types using Landsat data. It is considered necessary to sample a number of fields of each land use type in order to classify the image.

Therefore it is conventional to plan the position and number of sampling points prior to sampling in order to obtain representative sampling. If one land use type is sampled intensively to the exclusion of the others, then an inaccurate classification of the image will result (Prof. C. Haan* - pers. comm.).

*Prof. C. Haan - Agricultural Engineering Department, Oklahoma State Univeristy.

3.1.3 Homogeneity, the Assumption of One Statistical or Biological Population and Outliers

There are indications that one cannot assume that waterbodies are homogeneous or consist of one population, statistically or biologically. Algae by their very nature may migrate in the water column depending on temperature, currents and wind. In addition, there are various types of algae (blue- greens, greens and diatoms) which have their own coloration, any of which may dominate in an impoundment at different times.

To confuse the issue even further, suspended sediments also differ in composition depending on their source. The morphometry, geology and microclimate of an impoundment also affects the assumption of one population.

The problem of outliers is best explained by way of an example. For instance, if polluted areas are confined to small areas and are sampled by few points, these points are often considered to be outliers as they represent a different population to the rest of the impoundment. It is important to recognise this fact because if a linear regression analysis is carried out using all of the data collected, the least squares method would minimize the sum of the squares of the distance from the regression line for the points, and the few points of high concentration (outliers) will pull the regression towards themselves. The many points of low concentration will bias the lower end of the regression equation in their favour. Again the result will be that a mean slope is determined - a compromise regression equation which may have no predictive capabilities either for the low or the high ranges of concentration.

3.2 SURFACE REFERENCE DATA

In order to achieve the objective of the project, accurate information of the conditions in the water body were required. The collection of surface reference data was therefore of prime importance.

It is necessary to point out differences in terminology here. Many studies have been undertaken in which the term "ground truth data" appears. This term was unsuitable for our purposes considering, firstly, that we were dealing with a water surface and not with the ground. Secondly, one cannot expect a single measurement to represent the 'truth' of a whole 80 m x 80 m pixel. The term 'surface reference data' was preferred because it acknowledges the limitations of a single reference point in time and space.

3.2.1 The Sampling Network

In order to assess Landsat as an aid to water quality surveillance, Roodeplaat Dam, situated 30 km north east of Pretoria and covering an area of 398 hectares, was chosen as the site for an extensive sampling program (Figure 3.1). Tables 3.1 and 3.2 present the characteristics of Roodeplaat Dam and its catchment.

Roodeplaat Dam has two arms, the western arm is long and fairly narrow while the eastern arm is fairly broad and open. The major

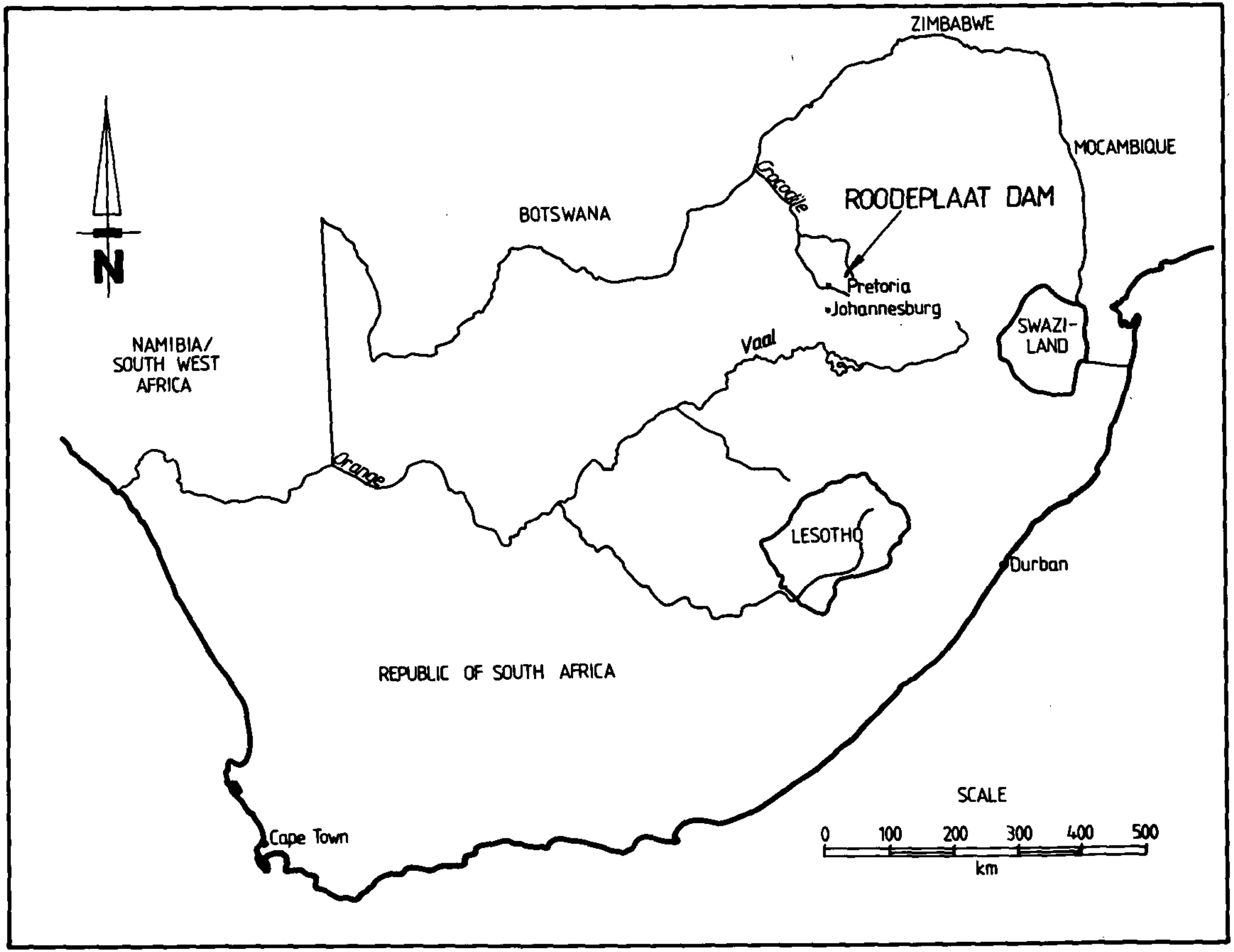


FIGURE 3.1 : **LOCATION OF ROODEPLAAT DAM**

TABLE 3.1: CHARACTERISTICS OF ROODEPLAAT DAM AND ITS CATCHMENT+

Geographical location	25° 37'S; 28° 22'E
Magisterial district	Pretoria
Catchment type	Urban/industrial, farmland, mines
Usage of dam	Recreation, potable water
Catchment area	668 km^2
Inflowing rivers	Pienaars River, Hartbeesspruit, Edendalespruit.
Dam wall completed	1959
*F.S.L. volume	41,907 x 10^6 m^3
F.S.L. area	3,96 km^2
F.S.L. maximum depth	43 m
F.S.L mean depth	10,6 m

*F.S.L = full supply level.

TABLE 3.2: AVERAGE TERM ANNUAL HYDROLOGICAL CHARACTERISTICS OF ROODEPLAAT DAM +

	*Average mean	*C.V. %
Volume x 10^6 m^3	41,425	3,3
Area km^2	3,898	2,6
Mean depth	10,57	0,7
Annual inflow x 10^6 m^3	59,01 (a)	
Annual outflow x 10^6 m^3	55,68 (b)	
Retention time a	0,70	

*Average mean is based on monthly values and an annual cycle; Period: January to December (1970-1978); C.V. = coefficient of variation.

+ From: Pieterse and Bruwer, 1980.

Note:
According to the reservoir records of the Department of Water Affairs, the annual inflow (a) is 49,2 x $10^6 m^3$ and the annual outflow (b) is 43,5 x $10^6 m^3$ (Mr. J. Schutte[o] pers. comm.).

[o] Mr. J. Schutte - Directorate of Hydrology, Department of Water Affairs.

rivers flowing into the impoundment are the Hartbeesspruit and the Pienaars River, which enter the impoundment at the southern end of the western arm and the Edendalespruit which enters at the eastern side of the impoundment. Most pollution enters the impoundment from the Hartbeesspruit and Pienaars River which flow through Pretoria's eastern suburbs. Therefore it is in the western arm of the impoundment where concentrations of suspended sediment and chlorophyll a are highest. The water in the eastern main body of the impoundment generally has lower chlorophyll a and turbidity values.

Roodeplaat Dam had a previously existing network of seven sampling sites and it was decided in September 1981 to increase the number of sites to 32, including the existing sampling sites. The sampling network was established on a randomly distributed basis giving good coverage to all areas within the impoundment (Figure 3.2). Where available, existing stabilised platforms or buoys were used to mark the sampling positions.

For the first four months of this project, two separate sets of samples were taken, 50 metres apart, at each of the 32 sampling sites, in order to examine the representativeness of the sampled data. It was then decided to increase the number of sampling sites, to include a transect along the western arm of the impoundment where the most significant differences in water quality had been noticed. The number of sampling sites was increased to 55 and due to the logistics of the sampling operation only one set of samples were collected at each site (Figure 3.3).

In the following analysis data collected from the 32 sampling points were examined. The additional 23 sampling sites were used for testing the accuracy of the calibrated regression equations. (Refer to Section 5.4).

3.2.2 Sampling undertaken concurrently with the Satellites Overflight

LANDSAT passes over the Johannesburg/Pretoria area at approximately 09h25 and records the whole scene in approximately 27 seconds. Weather conditions in the area fluctuate, with cloudy conditions around noon being apparent in summer, and hazy conditions due to dust etc., being manifest in winter. It was therefore fortunate that the satellite overflights occurred early in the day. It was imperative that the surface reference data were obtained as near as possible to the time of overflight and a routine became established in which, one hour before each overflight, two HRI boats and well instructed personnel would be on the water collecting their first water samples. Generally all of the samples were collected within two hours. It also became apparent that the sampling team had to be prepared to sample at every opportunity as the chances of completely cloud free skies were low.

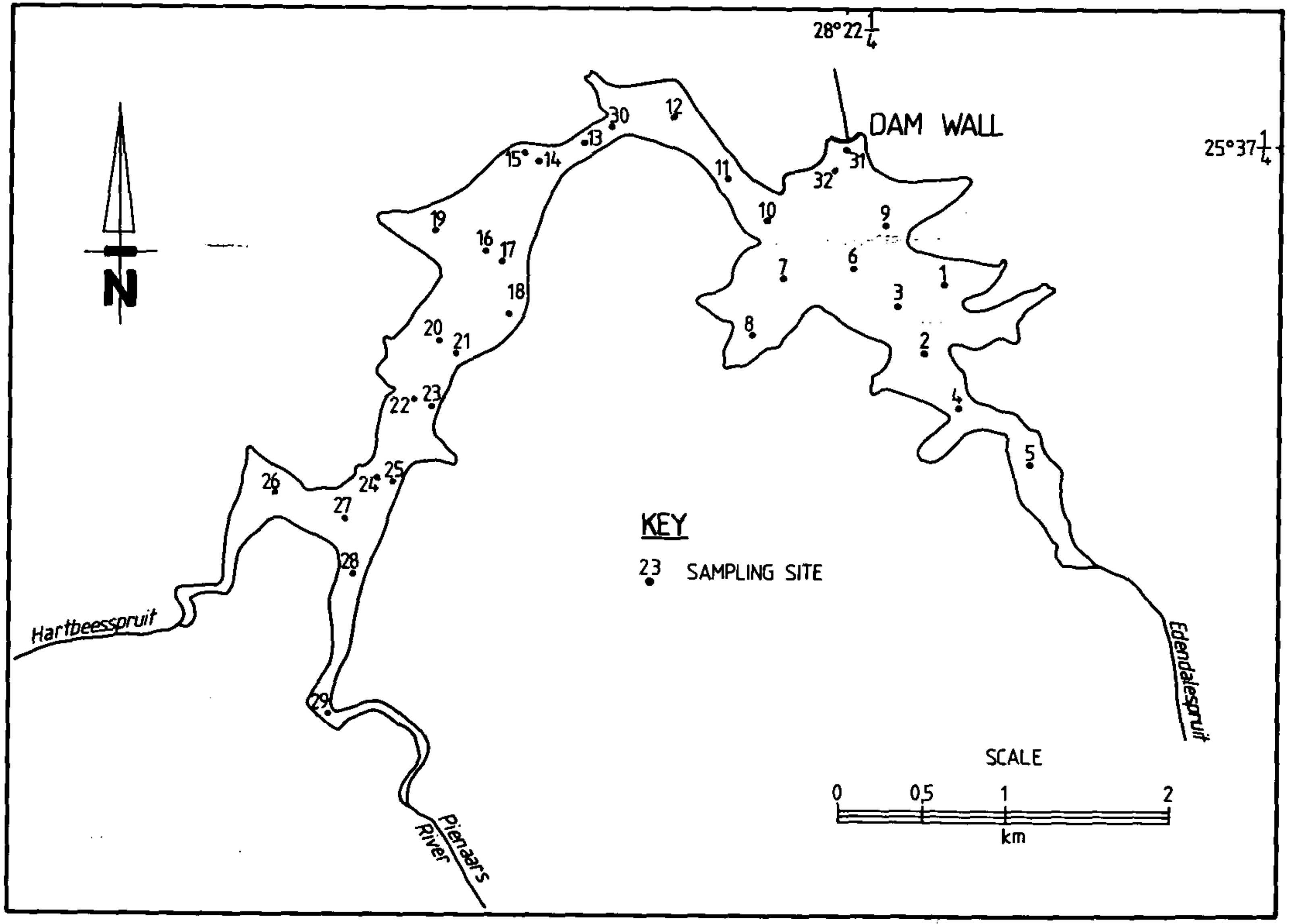


FIGURE 3.2 : ROODEPLAAT DAM SHOWING SAMPLING SITES: 32 POINTS

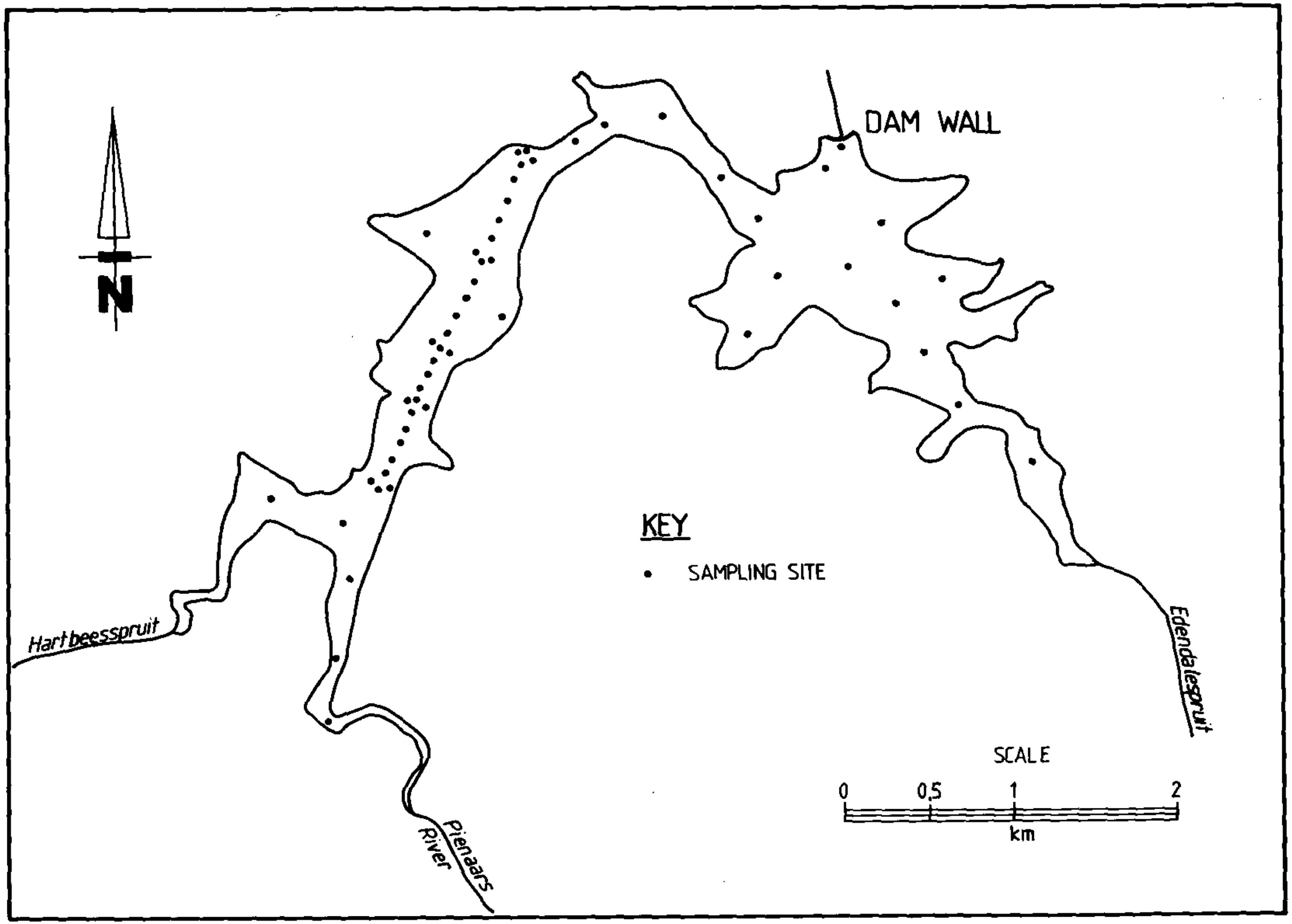


FIGURE 3.3: ROODEPLAAT DAM SHOWING SAMPLING SITES: 55 POINTS

3.2.3 Water Quality Variables

The water quality data collected for analysis during each overflight were as follows:

(1) Surface chlorophyll a (μg/ℓ)
(2) Integrated chlorophyll a (μg/ℓ)
(3) Secchi disc depth (m)
(4) Surface turbidity (NTU)
(5) Integrated turbidity (NTU)
(6) Surface water temperature and sunshine conditions.

Climatic and chemical data for other determinands were also obtained but are not directly relevant to this report. The data are reported by Howman and Kempster (1983(a)).

3.2.4 Sampling Techniques and Equipment

Hydrological Research Institute (HRI) personnel used standard sampling techniques to collect the water quality samples. Surface samples of chlorophyll and turbidity were taken directly from the surface of the water using buckets and transferred to 1 litre plastic bottles.

Integrated samples were obtained using hosepipe sampling. A 1.9 cm diameter hosepipe with a weighted end was lowered into the water as far as the secchi disc visibility depth. The weighted end was then raised to the surface capturing the water column in the pipe.

Secchi disc depths were determined by using standard black and white, 30 cm diameter, secchi discs suspended from the shaded side of the boat. The secchi depth was used to indicate the depth to which water had to be sampled, when taking the integrated samples, as secchi depth quickly and easily determined the approximate depth of light penetration in the visible range (Boland et al, 1979).

Mercury thermometers were used to take the surface water temperatures. The people collecting the samples estimated sunshine conditions as clear, medium or overcast on a subjective basis.

3.2.5 Analysis of Water Samples

The water quality samples were analysed by the staff of the Chemical and Biological Analytical Service (HRI) on the same day, immediately after the sampling operation. The chlorophyll a samples were analysed by the method described in Appendix A (Truter, 1981; Sartory, 1982). Turbidity analyses were carried out using a Hach Turbidimeter and measured in nephelometric turbidity units (NTU).

3.2.6 Storage of data

The data were then recorded on data coding forms, punched onto computer cards and stored on data files in the format given in Appendix B. The surface reference data for each of the six days under analysis in this report are presented in Appendix C.

3.3 SATELLITE REFLECTANCE DATA

3.3.1 Introduction

The 14th October 1981 saw the first simultaneous Landsat overflight/water quality sampling operation take place on Roodeplaat Dam. It was the first of many attempts to obtain data but the efforts of the research team were continually thwarted by cloud and rain and by the breakdown of Landsat 2 in February 1982. Eventually a total of six attempts proved to be successful throughout the period October 1981 to November 1982 and these will be discussed in detail.

3.3.2 The Computer Compatible Tapes

The Computer Compatible Tapes (CCT's), were obtained from the Satellite Remote Sensing Centre (SRSC) at Hartbeesthoek (Howman, 1984). All of the tapes had been corrected for sun's angle and were dehazed in a standard manner at Hartbeesthoek in order to maintain uniformity. There was one problem. Landsat 2 broke down in February 1982, Landsat 3 quickly took over but there were difficulties in data capture and eventually, by the end of the project, Landsat 4 was in operation. This meant that data were collected by three different satellites and therefore were subject to different data processing. Without adequate image processing facilities or expertise, it was considered to be impossible for the researcher to take into consideration these differences, be they large or small, and therefore this problem was ignored.

The CCT's obtained for the Landsat Water Quality Project for Roodeplaat Dam are given in Table 3.3:

TABLE 3.3: INFORMATION ON THE COMPUTER COMPATIBLE TAPES USED IN THE LANDSAT WATER QUALITY PROJECT

WRS*	DATE	SUN'S ANGLE	IDENTITY NO.+
182/78	81.10.14	48° 36'	22457-07143
182/78	81.11.01	51° 80'	22475-07150
182/78	81.12.07	52° 71'	22511-07162
170/78	82.09.13	42° 34'	40058-07293
182/78	82.09.30	46° 48'	31670-07231
170/78	82.11.16	50° 40'	40122-07302

* World Reference System
\+ The Individual Landsats are identified by the first digit of the identity number

The digital data from the tapes were stored on the Departmental mainframe computer system and any further manipulations and statistical analysis were carried out using the digital data.

The data were accessed by an image processing system originally called CATNIPS (Cape Town Image Processing Suite) and modified for use on the Departmental mainframe system (Maaren, 1981).

3.3.3 Colour Coding of the Reflectance Data

The initial analysis of the satellite reflectance data was carried out on the image processing system at the SRSC. Roodeplaat Dam was located and the surrounding land areas were masked out using band 7 values as the land/water delimiter (refer to Section 2.4.5). The reflectance values within the impoundment in each reflectance band were then colour coded (refer to Section 2.4.6) using a predetermined coded pseudo colour bar (Plate 3.1).

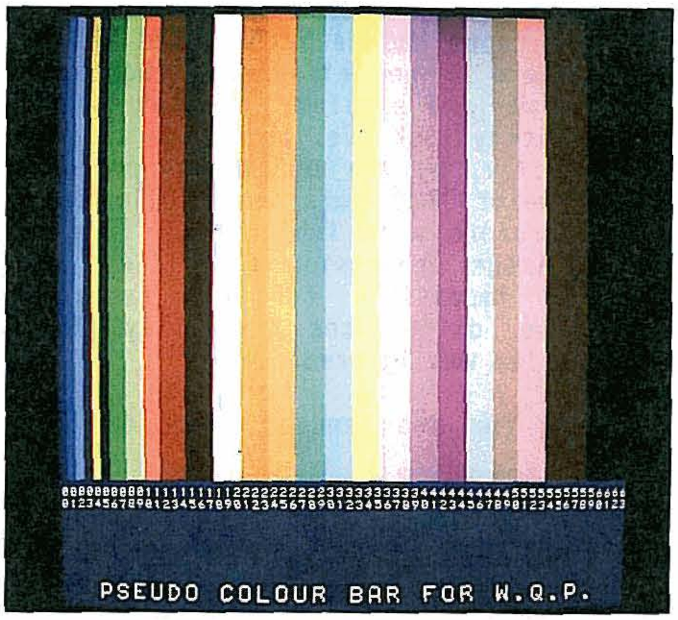


PLATE 3.1: COLOUR CODED BAR SHOWING THE FULL RANGE OF 0 - 255 REFLECTANCE VALUES DIVIDED INTO 25 COLOUR CLASSES

The satellite reflectance values are indicated on the horizontal axis of the bar. The full range of 0-255 reflectance values were divided into 25 classes. Each colour on the bar represented an actual reflectance interval of 4 x n reflectance units where n = 1 to 5. For example, from left to right, the darkest blue colour labelled 0 represented reflectance values of 0 to 3, therefore the lowest values recorded by the satellite. The light blue colour labelled 2 represented reflectance values of 8 to 11 and the yellow (4) represented 16 to 19 digital reflectance values. The green shade, labelled 6 and 7 (with n = 2) represented digital reflectance values 24 to 31.

The colour coded images in each spectral band for three of the days analysed, presented on Plates 4.1 to 4.12, provided a visual impression of reflectance conditions in the impoundment at a glance (refer to Section 4.4).

3.3.4 Unsupervised Classification of the Reflectance Data

The visual data, although helpful, were considered to provide insufficient quantifiable information and the image processing system CATNIPS was used to undertake any further classification and analysis.

Unsupervised classification, or "numerical taxonomy", splits pixels into groups or clusters in feature space "such that the distances between points within a cluster are a miniumum while the distances between clusters are a maximum" (Piper, 1981). An image processing computer program interprets digital data into categories. Unsupervised classification using all four reflectance bands was undertaken and the results are discussed in Section 4.5.

3.3.5 The Alignment of Reflectance Data with Surface Reference Data

A problem arose with the digital data when trying to accurately align pixel values with their corresponding surface reference sampling points. This matter of cartographic registration could greatly affect results and therefore a mathematical method for estimating pixel position was investigated (refer to Section 2.4.4).

The most accurate means of achieving this alignment obviously would be by obtaining the geometric latitude and longitude of both sampling position and pixel position. Unfortunately geometric positioning was not included with the satellite data and manually assessing latitudes and longitudes for both the satellite data and the sampling sites was time consuming. Therefore alternative methods had to be investigated which are discussed in Appendix D. The decision was taken to use a weighted 3 x 3 pixel window, at each sampling site, in order to determine the most accurate pixel value to be used in the analysis. The program which undertakes this task is given in Appendix E. An example of the results of this subroutine are given in Appendix F.

3.4 STATISTICAL ANALYSIS

3.4.1 Introduction

Calibrating the relationship between chlorophyll/turbidity and reflectance values and not simply describing one's sample set per se was the crux of the investigation.

At the outset it must be stressed that, judging by the literature and the visual satellite imagery, it was assumed that a relationship between specific water quality parameters and satellite reflectance data does exist.

3.4.2 To Establish the Representativeness and Accuracy of the Surface Reference Data

In the initial stages of the project at each sampling site, two sets of surface reference data were collected, 50 metres apart from each other. The reason for this was to give an indication of the representativeness and accuracy of the bucket samples collected within a 6400m^2 area. Simple linear correlation analysis was carried out on all of the 32 duplicate data sets. The statistics for the individual data sets for one day's data (81.12.07) are given in Table 3.4.

TABLE 3.4: SUMMARY OF STATISTICS FOR SURFACE REFERENCE DATA SETS, ROODEPLAAT DAM 81.12.07

Variable (1)	Mean	Standard deviation	Coefficient of variation	Smallest value	Largest value	Smallest standard score	Largest standard score	Skewness	Kurtosis
SUCOB	20,7	14,3	0,69	4,60	44,20	-1,13	1,64	0,29	-1,68
SUCOA	21,0	15,1	0,72	2,00	45,80	-1,26	1,64	0,32	-1,67
INCOB	16,6	13,7	0,82	4,30	42,70	-0,90	1,90	0,85	-1,05
INCOA	21,4	16,6	0,8	2,30	43,30	-1,15	1,35	0,04	-2,00
SUTURA	5,6	4,3	0,77	1,10	13,00	-1,04	1,73	0,45	-1,47
SUTURB	5,7	4,5	0,78	1,00	14,00	-1,06	1,85	0,47	-1,32
INTURA	5,8	4,6	0,79	0,90	14,00	-1,07	1,80	0,47	-1,36
INTURB	5,3	3,9	0,74	1,30	13,00	-1,02	1,95	0,68	-1,13

(1) Code: SU = Surface; IN = Integrated; CO = Chlorophyll; TUR = Turbidity; A = 1st data set;
B = Duplicate data set (2nd)

TABLE 3.5: LINEAR CORRELATION COEFFICIENTS (r) OF TWO SETS OF SURFACE REFERENCE DATA FOR 81.12.07

	SUCOB	SUCOA	INCOB	INCOA	SUTURB	INTURA	INTURB	SUTURA
SUCOB	1,00							
SUCOA	0,94	1,00						
INCOB	0,83	0,76	1,00					
INCOA	0,83	0,90	0,78	1,00				
SUTURB	0,95	0,96	0,83	0,85	1,00			
INTURA	0,94	0,96	0,83	0,83	0,99	1,00		
INTURB	0,94	0,94	0,91	0,88	0,97	0,97	1,00	
SUTURA	0,95	0,98	0,83	0,89	0,99	0,99	0,97	1,00

Code: SU = Surface; IN = Integrated; CO = Chlorophyll; TUR = Turbidity; A = 1st data set;

B = Duplicate data set (2nd)

r = Correlation coefficient where r = 1 indicates a perfect correlation and r = 0 indicates no correlation at all.

Table 3.5 shows the results of the linear correlation analysis between the duplicate samples at each point for surface chlorophyll _a_ (r = 0,94), surface turbidity (r = 0,99), integrated chlorophyll _a_ (r = 0,78) and integrated turbidity (r = 0,97). The results provided evidence that the surface reference data were representative of the pixel

TABLE 3.6: LINEAR CORRELATION COEFFICIENTS OF LOG TRANSFORMED DATA (81.12.07)

	SUCOL	INCOL	INTUL	SUTUL
SUCOL	1,00			
INCOL	0,73	1,00		
INTUL	0,77	0,83	1,00	
SUTUL	0,82	0,88	0,94	1,00

and of the situation within their respective $6400m^2$ pixel area and were therefore adequate for analysis. In addition, the efficiency of the analytical methods, used in the water quality analysis, was illustrated by the agreement achieved between the duplicate sampling points.

To deal with any possible lack of linearity, log transformation of the surface reference data was used. As seen in Table 3.4 the data in this instance proved to be positively skewed with a negative kurtosis, thereby requiring a transform.

The means of the duplicate samples were then established and a linear correlation analysis of the logs of the four prime independent variables showed some important relationships. Table 3.6 gives the statistical characteristics of the log transformed data sets. Table 3.6 shows that the variables were correlated with one another. The correlation r = 0,82 between surface chlorophyll _a_ and surface turbidity and r = 0,83 between integrated chlorophyll _a_ and integrated turbidity, indicates that chlorophyll _a_ and turbidity are virtually indistinguishable from one another.

3.4.3 To Determine the Homogeneity of the Data and the Existence of more than One Statistical Population

The Stepwise Discriminant Analysis procedure is a statistical test that best characterizes differences between groups and gives "a good visual representation of how distinct the groups are (for two groups, the points (cases) are projected onto a line where the groups are further apart)...", (Dixon and Brown, 1979). The Stepwise Discriminant Analysis demands that potentially separable groups be delineated by the researcher and the analysis illustrates agreement or disagreement of the groupings (Dixon and Brown, 1979). The categorisation of sampling points for the Stepwise Discriminant Analysis on Roodeplaat Dam is illustrated on Figure 3.4.

Four main groups were specified. P depicts the points situated near the Pienaars and Hartbeesspruit Rivers entrance into the impoundment. The highest concentrations of chlorophyll and turbidity are found in this area. C represents sampling points along the canoe lanes on the western arm of the impoundment. B represents those points in the northern part of the impoundment along the connecting limb between the two arms. D group represents the large body of the impoundment - the eastern arm.

The analysis was used to determine the existence of one or more populations in the impoundment.

The results of the analysis indicated clearly, in all cases examined, that at least two distinctly different populations were present in the impoundment (refer to Section 4.2 and Appendix K). As can be seen from Figure 3.5, the D population is substantially different to the C and P populations and the outlier may bias the analysis. Care, therefore, was required in order to obtain the necessary representative ranges of concentration values.

3.4.4 Determination of the Existence of Outliers and the Problem of Non-Normality

The Stepwise Discriminant Analysis showed that an outlier was present in the data (see Figure 3.5) and its presence could bias the regression equation. Furthermore the predictive ability of the regression equation would be highly suspect. The method of analysis was chosen so that if two separate populations were indicated, Fillibens Probability Plot Correlation Coefficient test for normality and Grubbs t test for detecting outliers would be used (Wainwright and Gilbert, 1981).

Any problem concerning the subjective exclusion of outliers was overcome by using Grubbs' t test for rejecting both very low and very high values together with Fillibens R test for normality.

Filliben introduced a test which depended on "the linearity of normal order probability plots with normally distributed data" (Wainwright and Gilbert, 1981). Considered to be equally as powerful as Shapiro Wilk's statistic W, this test is seen to have the added advantage of being more simple due to its emphasis on linearity and its incorporation of the product moment correlation coefficient (Filliben, 1975). It is a test that is readily extendable to testing non-normal distributions, is easily implemented on a computer (Dhanoa, 1982), is less limited by the sample size and has been tested favourably with other normal test statistics (Filliben, 1975; Wainwright and Gilbert, 1981).

Values of Filliben's R at the 0,05 probability level are tabulated against sample size (Appendix G) and for data to be normally distributed "the computer value of R must equal or exceed the tabulated value".

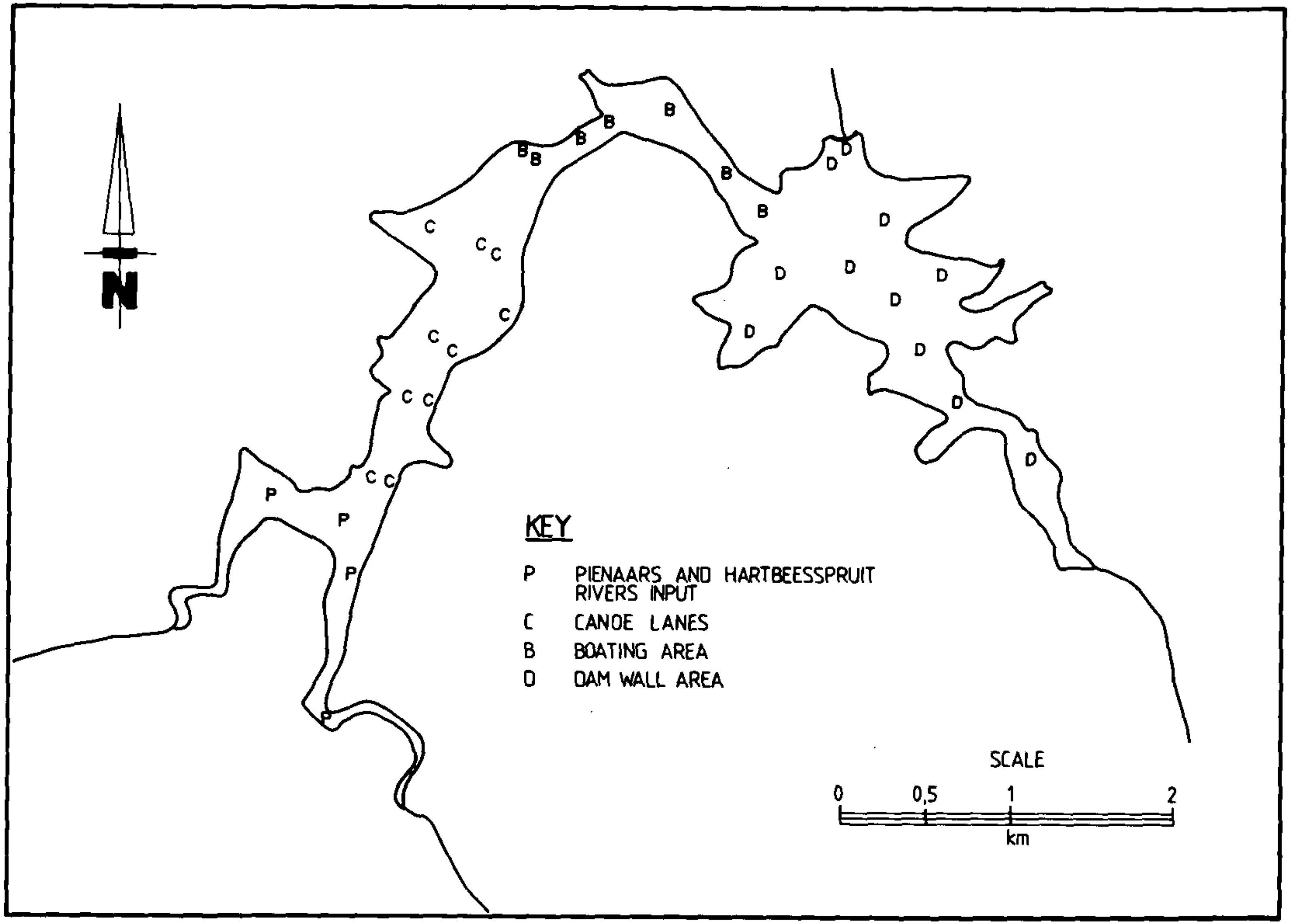


FIGURE 3.4: CATEGORIES FOR THE STEPWISE DISCRIMINANT ANALYSIS

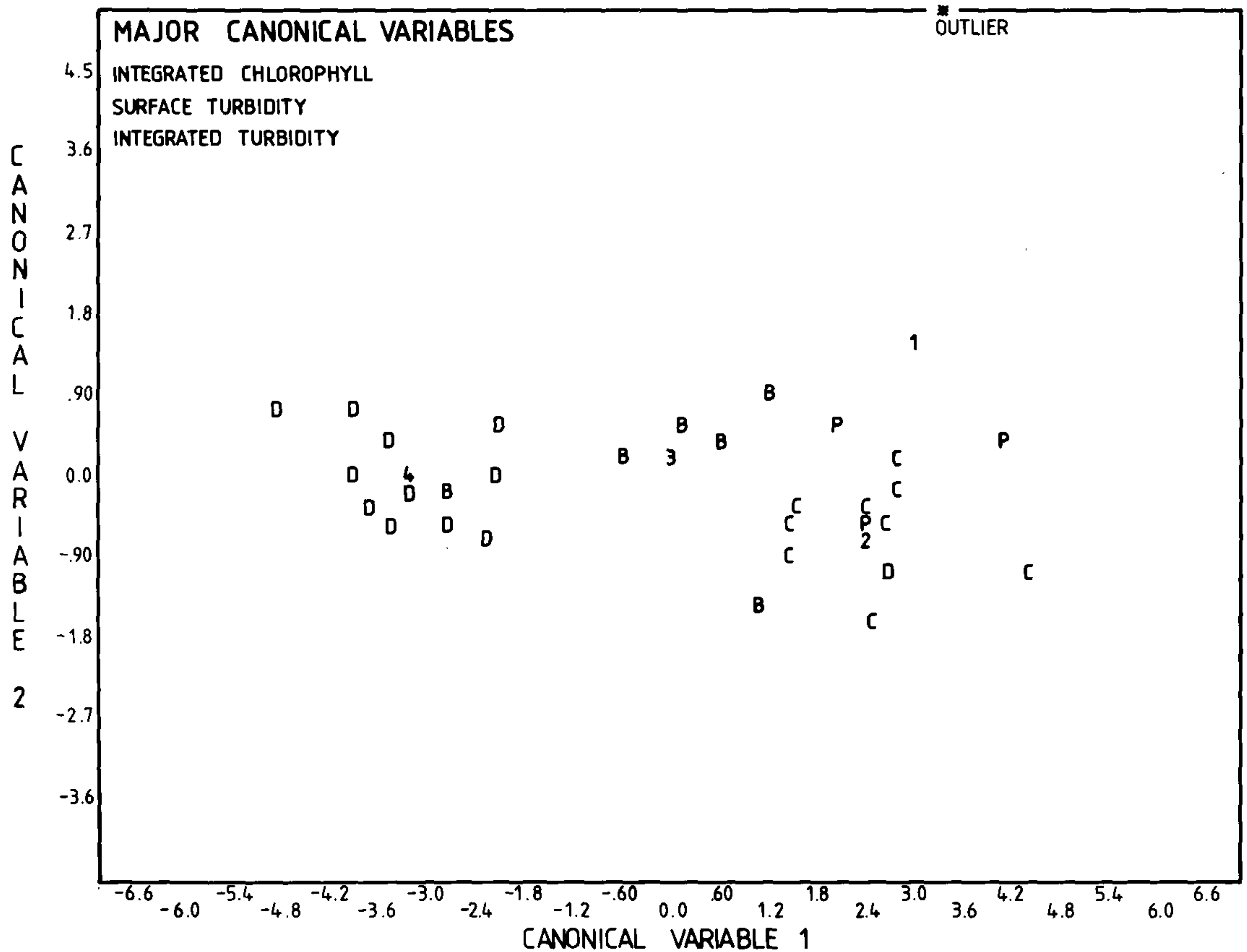


FIGURE 3.5: STEPWISE DISCRIMINANT ANALYSIS SCATTER PLOT FOR 81·10·14

Using Appendix G, the computed value of Grubb's t had to equal or exceed the tabulated value for a value to be considered to be an outlier at the 0,05 level of probability (Grubbs and Beck, 1972; Wainwright and Gilbert, 1981).

Both the Grubbs t test for outliers and Filliben's Probability Plot Correlation Coefficient test for normality were acquired in BASIC and were translated into Fortran IV. The program is given in Appendix H. An example of the results obtained from Program 'Filli' is given in Appendix I.

If the Filliben's R and Grubbs t tests indicated that the data were normal and without outliers the data were considered to have fulfilled the necessary requirements for normality and therefore representativeness, and were used in further tests. If the tests indicated the presence of one or more outliers, then outliers were removed one at a time and if the data were then normal, further statistical analysis would be performed.

If Filliben's Probability Plot Correlation Coefficient test (Wainwright and Gilbert, 1981) indicated that the data set was not normally distributed even after outliers were removed, a procedure had to be found to ensure the normality of the data. Looking for a computerised function to do just this proved to be painstakingly difficult and, after receiving different opinions from statisticians in South Africa and overseas, the matter still has not been resolved satisfactorily.

Data which are not normally distributed may be an indication that some of the data could be clustered around a point. In order to remove the unbalanced cluster and obtain a representative range of data, a few options were suggested.

Firstly, in areas where clustering was not evident, it was thought that perhaps duplicating data points and including them into the data set would increase the weight of the data points in the range by a factor greater than one and that this would establish an adequate range of values. This option, however, did not work as the Canonical Correlation Analysis rejected the data and the matter was not pursued further.

Statisticians then suggested that the log transformed data should be logged a second time. This log log transformation is considered to be a standard practice (Mrs. J. Meyer* - pers. comm.).

Log log data should be normal but in this instance the logs were negative and as this is mathematically not acceptable this option was abandoned. Another option - root squaring of residual values was suggested but this was considered to be beyond the scope of this work and can only be suggested as a future possibility to be tackled by interested statisticians.

---

* Mrs. J. Meyer - Department of Statistics, University of Witwatersrand.

No other recommendations were forthcoming, people held their hands up in horror and despair, others said "Ignore the problem". No one has so far been able to help with the problem of how to deliberately exclude data points from a non-normal distribution using a blind standard procedure, i.e., an explicit model, and using as much of the available data as possible. This is a problem which often confronts a researcher in the field of water quality because of the difficulties of obtaining a representative sample (see Section 3.1.2).

So in this dilemma, with many criticisms but no help forthcoming, it was decided to make the best of a bad situation and choose a model, however imperfect, that would help normalise the data and thereby obtain a representative subset for analysis. It was decided that a simple procedure would be attempted to exclude a few data points, where excessive clustering occurred.

3.4.5 Normalising the Data

The approach for the selection of data points was based on the shape of the normal distribution using the area under segments of the normal curve. This procedure is presented in Appendix J.

Effectively the test for normality lifted any possible bias from the data and the exclusion of outliers removed the problem of having two possibly separate populations. In addition, the test proved to be easily duplicated and was as objective as possible, under the circumstances.

3.4.6 The Canonical Correlation Analysis

The interdependency between both the water quality conditions (chlorophyll a and turbidity) and the four reflectance bands, meant that a statistical test was required that would take into account the interrelatedness.
A multi-variate multiple regression analysis technique was required and the use of Canonical Correlation Analysis was recommended.

Howard Hotelling, the originator of the Canonical Correlation analysis in 1936, described the concept behind his work as follows:

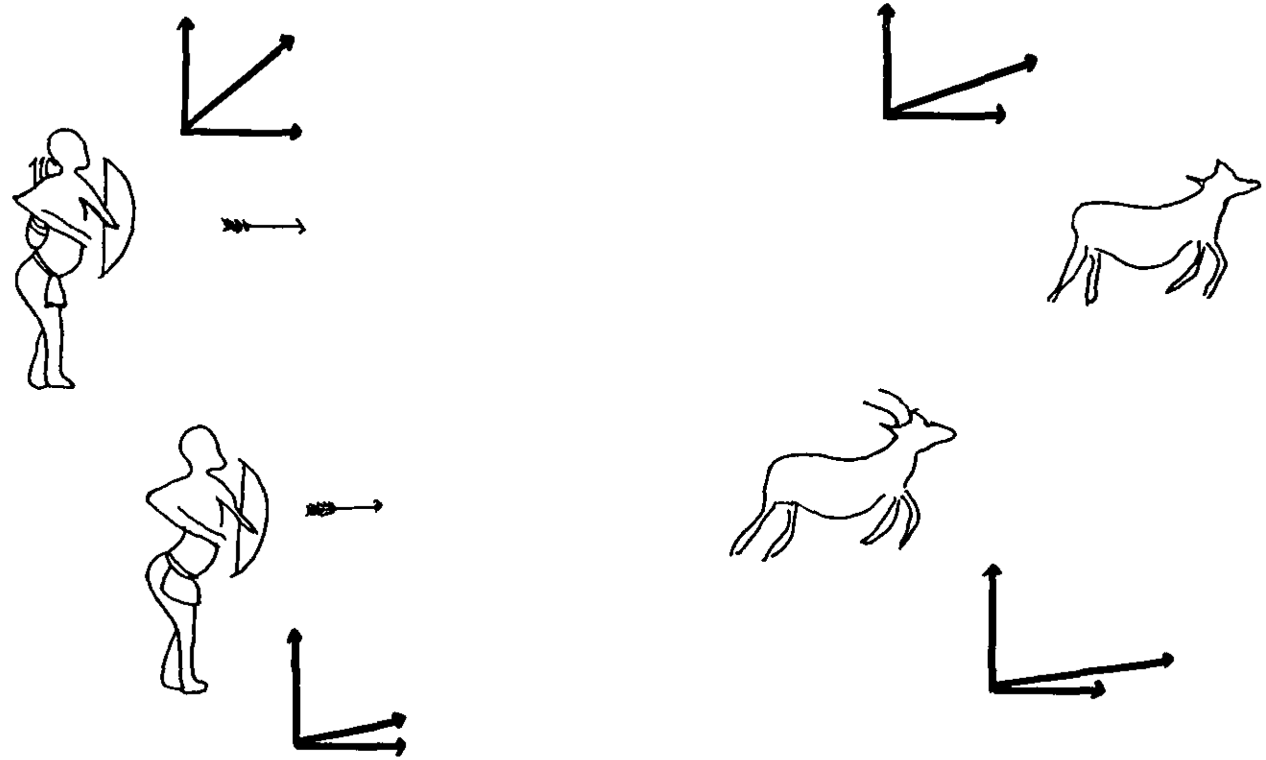
"Marksmen side by side firing simultaneous shots at targets so that the deviations are in part due to independent individual errors and in part to common causes such as wind, provide a familiar introduction to the theory of correlation; but only the correlation of the horizontal components is ordinarily discussed, whereas the complex consisting of horizontal and vertical deviations may be even more interesting. The wind at two places may be compared, using both components of the velocity in each place. A fluctuating vector is thus matched at each moment with another fluctuating vector." (Hotelling, 1936).

A concept of the Canonical Correlation Analysis as envisaged by the authors is shown on Figure 3.6.

Hotelling developed the technique to extract suitable descriptive functions from a multiplicity of correlations in psychological testing. Since then the Canonical Correlation Analysis has been used to study the correlation structure between two sets of variables

A CONCEPT OF THE CANONICAL CORRELATION ANALYSIS

STRENGTH
ARROW SIZE
WIND
VELOCITY
SPIN
DISTANCE



STRENGTH
SIZE
TERRAIN
SPEED

OPTIMISES COMPLEX RELATIONSHIPS SIMULTANEOUSLY INTO A SINGLE FUNCTION IN N DIMENSIONAL SPACE.

(Haan, 1977) and "can be viewed as extension of multiple regression analysis" (Dixon and Brown, 1979). There are usually sets of dependent Y variables (in this instance reflectance bands 4, 5, 6 and 7) as well as sets of independent X variables (surface chlorophyll a and turbidity, integrated chlorophyll a and turbidity). "The problem is to find a linear combination of the X variables that has maximum correlation with a linear combination of the Y variables" (Dixon and Brown, 1979).

The computerised Canonical Correlation Analysis used is part of the BMDP Biomedical Computer Program P-series (Dixon and Brown, 1979).

Canonical Correlation analysis was carried out between the following sets of data:

(1) Log surface chlorophyll a (SUCOL) and log surface turbidity (SUTUL) with reflectance bands 4, 5, 6 and 7.

(2) Log integrated chlorophyll a (INCOL) and log integrated turbidity (INTUL) with reflectance bands 4, 5, 6 and 7.

In the statistical analysis, the independent data sets were split up owing to the fact that the presence of too many mutual correlations within the independent data set resulted in singularity (Gittins, 1979). The splitting up of the data set also simplified the interpretation of results.

The Canonical Correlation produced two correlations: the maximum and the second highest correlation possible between the variates. Analysis indicated that the second coefficient was of little value for the requirements of this research and therefore the analysis was confined to the first and highest correlation coefficient.

The Canonical Coefficients can be presented in an untransformed or a standardised manner. It was decided to use the untransformed form, to facilitate use of the coefficients in a model to calculate absolute values of the variables, as the standardised form, involves standard deviation units, which could vary between overpasses.

Concerning the Canonical Coefficients the following is important: if the logs of the surface reference data values are similar and in the same order of magnitude, then a comparison of coefficients provides a direct indication of the relative contribution of each variable to the Y variates. If the numerical values are not the same order of magnitude then the coefficients cannot be directly compared. The comparison, in the latter case, was done between the absolute value of the product of a Canonical Coefficient and the mean value of the variable concerned. In this case the data were within the same order of magnitude and therefore the coefficients did indicate relative magnitudes of importance. It was therefore possible to ascertain the relationships between the individual water quality variables and the separate reflectance bands.

Canonical Correlations and Coefficients for the variables, were obtained for each day's data. Each variable's percentage contribution to the relationship was obtained using the Canonical Coefficient and the mean of the data set. The negative sign of the coefficient was not included in this calculation. The following formula was used:

To determine the percentage contribution of surface chlorophyll a:

$$\frac{(\text{SUCOL (Coeff)} \times \text{SUCOL (Mean)})}{(\text{SUCOL(Coeff)}\times\text{SUCOL(Mean)})+(\text{SUTUL(Coeff)}\times(\text{SUTUL(Mean)})} \times 100 = \%$$

To determine the percentage contribution of BAND 4:

$$\frac{(\text{BAND 4 (Coeff)} \times \text{BAND 4 (Mean)})}{(\text{B4(Coeff)}\times\text{B4(Mean)})+(\text{B5(Coeff)}\times\text{B5(Mean)})+(\text{B6(Coeff)}\times\text{B6(Mean)})+(\text{B7(Coeff)}\times\text{B7(Mean)})} \times 100 = \%$$

.... Equation 1

Finally an underlying assumption of the Canonical Correlation Analysis was that the data should be normally distributed in order to be able to provide a reliable interpretation of the analysis (Gittins, 1979). Gittins further elaborates "Nonlinearity, heterogeneity and the presence of deviant observations (outliers) can largely nullify a canonical analysis. Thus validation is directed largely to the detection of these features." This was one of the reasons for undertaking the test for normality and normalising the data where it did not satisfy this criterion.

3.4.7 Discussion of the Statistical Analysis

Many problems and suggested solutions to the problems have been presented. A statistical analysis capable of undertaking the analysis was established and according to Murphy's law another problem emerged. The point of view of whether or not it was necessary to exclude outliers and normalise the data set was the next problem to be encountered.

Three major and differing points of view came to the fore. Firstly, one opinion said that under no circumstances should one eliminate any data because each sample is a valid contributor to the data set. This opinion will henceforth be termed 'Including All Data'.

A second viewpoint recognised the problem pointed out in Section 3.1.3 that few points of high concentration could bias the regression equation thereby invalidating the equations simulative or predictive capabilities. This standpoint demanded that outliers should be excluded from the data set and analysed separately, henceforth termed 'Excluding Outliers'.

Thirdly, in the face of the problems discussed in Section 3.1.1, 3.4.4 and 3.4.6 it was recognised that one should attempt to fulfil the implicit assumptions of the regression analysis technique employed, as well as overcoming the problem of non-representativeness and non-uniformity of the data. This point of view, that outliers should be removed and analysed separately, and that the data set should be of a normal distribution, will henceforth be termed "Normalised Data".

The controversy between these viewpoints and the desire of the authors to obtain as accurate answers as possible to fulfil the original objective, led to an in depth analysis being carried out. Each point of view was examined according to its own precondition and the results illustrating the differences, if any, are presented.

CHAPTER 4

RESULTS AND DISCUSSION OF INITIAL ANALYSIS

4.1 INTRODUCTION

Statistical analysis relating the surface reference data and satellite reflectance data for six days of data was carried out. A brief review of the methods of analysis follows. Except in severe cases, the log transformation applied to the surface reference data*, smoothed the data sufficiently to comply with the requirement of normality. The Stepwise Discriminant Analysis test, was used to ascertain the existence or not, of more than one population. In addition, Filliben's Probability Plot Correlation Coefficient (R) test for normality, incorporating Grubbs t test for detecting outliers was used to establish the representativeness of the data set. The Canonical Correlation multivariate regression analysis was then applied to the data in order to determine the correlation between the surface reference data and satellite reflectance data (refer to Section 3.4.6).

Unsupervised classification of the satellite reflectance data was computed (refer to Section 3.3.4) and the colour coded imagery was discussed (refer to Section 3.3.3).

Three separate statistical avenues were used throughout the investigation. Firstly, entitled 'Including All Data', analysis of all data, irrespective of the presence of more than one population or any outliers, was undertaken. Secondly, using Filliben's R and Grubb's t test as a basis, only outliers were excluded from the data set and analyses were carried out entitled 'Excluding Outliers'. Thirdly, both Filliben's R and Grubb's t test, together with the normalising procedure described in Appendix J, were used to normalise the data where log transformation had not adequately smoothed the data. Analyses carried out by this method were called 'Normalised Data'.

Results of these analyses in the triple avenue approach are presented.

4.2 STEPWISE DISCRIMINANT ANALYSIS

The Stepwise Discriminant Analysis was carried out on each of the days data and the results are presented in Appendix K. The results indicated that for each of the days there were two, if not more, distinct populations present in the impoundment.

* It is important to note that in some situations it may be necessary to log both surface reference data and the satellite reflectance data.

4.3 CANONICAL CORRELATION ANALYSIS

A Canonical Correlation Analysis from the BMDP - 79 Biomedical Computer Programs, P-series, (Dixon and Brown, 1979) was chosen to examine the problems of correlating satellite reflectance data with monitored water quality data. Analysis was carried out between the following data:

(1) Log surface chlorophyll a (SUCOL) and log surface turbidity (SUTUL) with reflectance bands 4, 5, 6 and 7, (BAND 4, BAND 5, BAND 6, BAND 7).

(2) Log integrated chlorophyll a (INCOL) and log integrated turbidity (INTUL) with reflectance bands 4, 5, 6, and 7, (BAND 4, BAND 5, BAND 6, BAND 7).

The Canonical Correlations for all 6 days as shown in Table 4.1, indicate that the r values are relatively high in each instance. The data for both 81.11.01 and 81.12.07 were normally distributed without any outliers and the day 82.11.16 was normal after outliers were excluded, hence the bracketed duplication of results in Table 4.1 indicates that it was not necessary to carry the normalisation process any further.

Differences in r values for the 3 options can be observed but there is not a consistent trend. The Canonical Correlation is, by definition, the best possible linear polynomial correlation between variables. The r correlations are, however, a function of the data set which may or may not be representative of the real underlying correlation for the parent population. Therefore the high r values are not sufficient evidence of stable correlation. Examination of the Canonical Coefficients and the percentage contribution of each variable to the relationship (refer to Section 3.4.6), brought to light more information (Tables 4.2 to 4.13; Howman and Kempster 1983(b)).

The standard linear regression equation $Y = MX + K$ can aid in understanding the results of the Canonical Correlation Analysis:

If Y represents the dependent variables, in this instance reflectance bands 4, 5, 6 and 7, and
X represents the independent variables, surface and integrated chlorophyll a and surface and integrated turbidity,
M is the slope of the regression line and
K is the intercept on the Y axis.

TABLE 4.1: CANONICAL CORRELATIONS (r)

DATE	INCLUDING ALL DATA		EXCLUDING OUTLIERS		NORMALISED DATA	
	SUCOL/ SUTUL	INCOL/ INTUL	SUCOL/ SUTUL	INCOL/ INTUL	SUCOL/ SUTUL	INCOL/ INTUL
81.10.14	0,88	0,89	0,85	0,87	0,87	0,87
81.11.01	0,79	0,93	(0,79)*	(0,93)	(0,79)	(0,93)
81.12.07	0,94	0,95	(0,94)	(0,95)	(0,94)	(0,95)
82.09.13	0,87	0,86	0,78	0,76	0,79	0,80
82.09.30	0,90	0,92	0,89	0,92	0,83	0,92
82.11.16	0,95	0,95	0,91	0,90	(0,91)	(0,90)

*( ) The brackets imply that the distribution passes the test for normality at the previous unbracketed stage, and that the normalising procedure therefore ceased at that stage.

TABLE 4.2: RESULTS OF THE CANONICAL CORRELATION ANALYSIS FOR SURFACE CHLOROPHYLL a/SURFACE TURBIDITY AND SATELLITE REFLECTANCE BANDS

81.10.14.

		INCLUDING ALL DATA				EXCLUDING OUTLIERS				NORMALISED DATA			
DATE: 81.10.14		n 32	CC	Mean	%	n 31	CC	MEAN	%	n 26	CC	MEAN	%
INDEPENDENT X VARIABLES	SUCOL		0,91	1,46	23		-0,98	1,44	22		-0,03	1,45	1
	SUTUL		6,43	0,7	77		7,39	0,69	78		7,14	0,69	99
DEPENDENT Y VARIABLES	BAND 4		0,39	5,59	64		0,41	5,52	64		0,35	5,31	51
	BAND 5		0,08	6,41	15		0,16	6,19	28		0,21	6,27	36
	BAND 6		0,03	8,81	8		-0,02	8,45	6		-0,05	8,69	13
	BAND 7		0,06	7,25	13		-0,01	6,9	2		0	7,19	0
CANONICAL CORRELATION	r	0,88				0,85				0,86			
TAIL PROBABILITY		0,0000				0,0000				0,0000			

CC = CANONICAL COEFFICIENT
MEAN = MEAN OF DATA SET
SUCOL = SURFACE CHLOROPHYLL a
n = NUMBER OF SAMPLING POINTS
% = PERCENTAGE CONTRIBUTION
SUTUL = SURFACE TURBIDITY

As an example the polynomial function obtained for the 81.10.14 overpass for surface chlorophyll a and turbidity, shown in Table 4.2 under the 'Including All Data' Option may be written as follows:

BAND 4 (0,39) + BAND 5 (0,08) + BAND 6 (0,03) + BAND 7 (0,06) = M SUCOL (0,91) + SUTUL (6,43) + K.

Major points in the interpretation of this equation are:

(1) Surface chlorophyll a with a coefficient of 0,91 contributes 23% of the relationship to the independent variable.

(2) Surface turbidity is the major independent variable representing 77% of the relationship.

(3) Band 4 with a coefficient of 0,39 is seen to be the most important dependent variable (64%).

(4) Bands 6 + 7 jointly represent 21% of the relationship contributed by the dependent variables.

(5) The highest independent coefficient may be directly related to the highest dependent coefficient, thereby connecting surface turbidity with band 4. The Canonical Correlation mainly represents a relationship between surface turbidity and band 4 since the contribution of surface chlorophyll a to the relationship is only 23%.

The 'Excluding Outliers' and 'Normalised Data' approaches can be interpreted in a similar manner.

The polynomial function for the integrated chlorophyll a and turbidity data for the option 'Including All Data' (Table 4.3.) is:

BAND 4 (0,32) + BAND 5 (0,18) + BAND 6 (0,02) + BAND 7 (-0,01) = M INCOL (-0,99) + INTUL (7,15) + K

This equation suggests the following:

(1) Integrated turbidity is the prime independent variable contributing 79% of the relationship.

(2) Band 4 is the prime dependent variable (56%).

(3) Band 4 is linked to integrated turbidity.

(4) Bands 6 and 7 have less significance (8%).

The results of the 'Including All Data' option for 81.10.14 indicate that Landsat detects suspended solids (turbidity) in Roodeplaat Dam. Both surface and integrated turbidity results are highly correlated with band 4 and to a lesser extent with band 5: this supports the established theory that bands 4 and 5 show up suspended solids (Bukata and Bruton, 1974; Moore, 1980; Lindell, 1981). A low band 7 contribution suggests that there are no high concentrations of algae.

TABLE 4.3: RESULTS OF THE CANONICAL CORRELATION ANALYSIS FOR INTEGRATED CHLOROPHYLL a/INTEGRATED TURBIDITY AND SATELLITE REFLECTANCE BANDS

81.10.14.

		INCLUDING ALL DATA				EXCLUDING OUTLIERS				NORMALISED DATA			
DATE: 81.10.14		n 32	CC	Mean	%	n 31	CC	MEAN	%	n 26	CC	MEAN	%
INDEPENDENT X VARIABLES	INCOL		-0,99	1,46	21		-2,53	1,45	41		-2,67	1,45	44
	INTUL		7,15	0,74	79		7,18	0,73	59		6,78	0,72	56
DEPENDENT Y VARIABLES	BAND 4		0,32	5,59	56		0,34	5,52	49		0,3	5,31	37
	BAND 5		0,18	6,41	36		0,23	6,19	37		0,29	6,27	42
	BAND 6		0,02	8,81	6		-0,02	8,45	4		-0,08	8,69	16
	BAND 7		-0,01	7,25	2		-0,05	6,9	10		-0,03	7,19	5
CANONICAL CORRELATION	r		0,89					0,87				0,87	
TAIL PROBABILITY			0,0000					0,0000				0,0000	

CC = CANONICAL COEFFICIENT
MEAN = MEAN OF DATA SET
INCOL = INTEGRATED CHLOROPHYLL a
n = NUMBER OF SAMPLING POINTS
% = PERCENTAGE CONTRIBUTION
INTUL = INTEGRATED TURBIDITY

TABLE 4.4: RESULTS OF THE CANONICAL CORRELATION ANALYSIS FOR SURFACE CHLOROPHYLL a/SURFACE TURBIDITY AND SATELLITE REFLECTANCE BANDS

81.11.01.

		INCLUDING ALL DATA				EXCLUDING OUTLIERS				NORMALISED DATA			
DATE: 81.11.01		n 32	CC	Mean	%	n 32	CC	MEAN	%	n 32	CC	MEAN	%
INDEPENDENT X VARIABLES	SUCOL		-4,23	1,51	50								
	SUTUL		7,87	0,8	50								
DEPENDENT Y VARIABLES	BAND 4		0,2	6,19	23								
	BAND 5		0,08	7,25	11								
	BAND 6		0,28	8,56	44								
	BAND 7		0,2	6,0	22								
CANONICAL CORRELATION	r			0,79									
TAIL PROBABILITY				0,0001									

CC = CANONICAL COEFFICIENT
MEAN = MEAN OF DATA SET
SUCOL = SURFACE CHLOROPHYLL a
n = NUMBER OF SAMPLING POINTS
% = PERCENTAGE CONTRIBUTION
SUTUL = SURFACE TURBIDITY

TABLE 4.5: RESULTS OF THE CANONICAL CORRELATION ANALYSIS FOR INTEGRATED CHLOROPHYLL a/INTEGRATED TURBIDITY AND SATELLITE REFLECTANCE BANDS

81.11.01.

		INCLUDING ALL DATA				EXCLUDING OUTLIERS				NORMALISED DATA			
DATE: 81.11.01		n 32	CC	Mean	%	n 32	CC	MEAN	%	n 32	CC	MEAN	%
INDEPENDENT X VARIABLES	INCOL		1,29	1,53	31								
	INTUL		5,57	0,80	69								
DEPENDENT Y VARIABLES	BAND 4		0,15	6,19	21								
	BAND 5		0,27	7,25	43								
	BAND 6		0,07	8,56	13								
	BAND 7		-0,17	6,0	23								
CANONICAL CORRELATION	r		0,93										
TAIL PROBABILITY			0,0000										

CC = CANONICAL COEFFICIENT
MEAN = MEAN OF DATA SET
INCOL = INTEGRATED CHLOROPHYLL a
n = NUMBER OF SAMPLING POINTS
% = PERCENTAGE CONTRIBUTION
INTUL = INTEGRATED TURBIDITY

TABLE 4.6: RESULTS OF THE CANONICAL CORRELATION ANALYSIS FOR SURFACE CHLOROPHYLL a/SURFACE TURBIDITY AND SATELLITE REFLECTANCE BANDS

81.12.07.

		INCLUDING ALL DATA				EXCLUDING OUTLIERS				NORMALISED DATA			
DATE: 81.12.07		n 28	CC	Mean	%	n 28	CC	MEAN	%	n 28	CC	MEAN	%
INDEPENDENT X VARIABLES	SUCOL		0,02	1,14	2								
	SUTUL		2,62	0,54	98								
DEPENDENT Y VARIABLES	BAND 4		0,05	11,93	21								
	BAND 5		0,06	10,93	23								
	BAND 6		0,1	8,46	30								
	BAND 7		-0,12	6,25	26								
CANONICAL CORRELATION	r		0,94										
TAIL PROBABILITY			0,0000										

CC = CANONICAL COEFFICIENT
MEAN = MEAN OF DATA SET
SUCOL = SURFACE CHLOROPHYLL a
n = NUMBER OF SAMPLING POINTS
% = PERCENTAGE CONTRIBUTION
SUTUL = SURFACE TURBIDITY

TABLE 4.7: RESULTS OF THE CANONICAL CORRELATION ANALYSIS FOR INTEGRATED CHLOROPHYLL a/INTEGRATED TURBIDITY AND SATELLITE *REFLECTANCE BANDS*

81.12.07.

		INCLUDING ALL DATA				EXCLUDING OUTLIERS				NORMALISED DATA			
DATE: 81.12.07		n 28	CC	Mean	%	n 28	CC	MEAN	%	n 28	CC	MEAN	%
INDEPENDENT X VARIABLES	INCOL		0,63	1,13	35								
	INTUL		2,38	0,56	65								
DEPENDENT Y VARIABLES	BAND 4		0,06	11,93	41								
	BAND 5		0,05	10,93	31								
	BAND 6		0,05	8,46	24								
	BAND 7		-0,01	6,25	4								
CANONICAL CORRELATION	r		0,95										
TAIL PROBABILITY			0,0000										

CC = CANONICAL COEFFICIENT
MEAN = MEAN OF DATA SET
INTOL = INTEGRATED CHLOROPHYLL a
n = NUMBER OF SAMPLING POINTS
% = PERCENTAGE CONTRIBUTION
INTUL = INTEGRATED TURBIDIY

TABLE 4.8: RESULTS OF THE CANONICAL CORRELATION ANALYSIS FOR SURFACE CHLOROPHYLL a/SURFACE TURBIDITY AND SATELLITE REFLECTANCE BANDS

82.09.13.

		INCLUDING ALL DATA				EXCLUDING OUTLIERS				NORMALISED DATA			
DATE: 82.09.13		n 31	CC	Mean	%	n 30	CC	MEAN	%	n 25	CC	MEAN	%
INDEPENDENT X VARIABLES	SUCOL		0,43	1,25	11		0,57	1,23	10		0,79	1,24	14
	SUTUL		6,73	0,70	89		9,34	0,68	90		8,56	0,69	86
DEPENDENT Y VARIABLES	BAND 4		0,05	14,52	13		0,17	14,47	35		0,23	14,88	49
	BAND 5		0,36	9,16	59		0,38	8,87	48		0,36	9,08	48
	BAND 6		0,10	7,81	14		0,08	7,5	8		-0,01	7,72	2
	BAND 7		-0,10	7,77	14		-0,08	7,5	9		0,0	7,92	0,4
CANONICAL CORRELATION	r		0,87				0,78				0,79		
TAIL PROBABILITY			0,0000				0,0014				0,0063		

CC = CANONICAL COEFFICIENT
MEAN = MEAN OF DATA SET
SUCOL = SURFACE CHLOROPHYLL a
n = NUMBER OF SAMPLING POINTS
% = PERCENTAGE CONTRIBUTION
SUTUL = SURFACE TURBIDITY

TABLE 4.9: RESULTS OF THE CANONICAL CORRELATION ANALYSIS FOR INTEGRATED CHLOROPHYLL a/INTEGRATED TURBIDITY AND SATELLITE REFLECTANCE BANDS

82.09.13.

		INCLUDING ALL DATA				EXCLUDING OUTLIERS				NORMALISED DATA			
DATE: 82.09.13		n 31	CC	Mean	%	n 30	CC	MEAN	%	n 25	CC	MEAN	%
INDEPENDENT X VARIABLES	INCOL		-0,64	1,23	13		-1,23	1,22	17		-1,52	1,21	20
	INTUL		7,23	0,74	87		10,54	0,72	83		10,01	0,73	80
DEPENDENT Y VARIABLES	BAND 4		0,03	14,52	8		0,16	14,47	34		0,26	14,88	50
	BAND 5		0,34	9,2	58		0,34	8,87	45		0,31	9,08	36
	BAND 6		0,14	7,81	20		0,12	7,5	13		-0,05	7,72	5
	BAND 7		-0,10	7,77	14		-0,07	7,5	8		0,09	7,92	9
CANONICAL CORRELATION	r		0,86				0,76				0,80		
TAIL PROBABILITY			0,0000				0,0015				0,0025		

CC = CANONICAL COEFFICIENT
MEAN = MEAN OF DATA SET
INCOL = INTEGRATED CHLOROPHYLL a
n = NUMBER OF SAMPLING POINTS
% = PERCENTAGE CONTRIBUTION
INTUL = INTEGRATED TURBIDITY

TABLE 4.10: RESULTS OF THE CANONICAL CORRELATION ANALYSIS FOR SURFACE CHLOROPHYLL a/SURFACE TURBIDITY AND SATELLITE REFLECTANCE BANDS

82.09.30.

		INCLUDING ALL DATA				EXCLUDING OUTLIERS				NORMALISED DATA			
DATE: 82.09.30		n 32	CC	Mean	%	n 30	CC	MEAN	%	n 17	CC	MEAN	%
INDEPENDENT X VARIABLES	SUCOL		3,62	1,3	97		2,45	1,28	65		3,35	1,29	78
	SUTUL		0,24	0,71	3		2,6	0,66	35		1,83	0,66	22
DEPENDENT Y VARIABLES	BAND 4		0,31	21,3	79		0,33	21,2	78		0,45	21,24	77
	BAND 5		0,06	14,13	10		0,1	13,87	16		0,08	13,71	9
	BAND 6		0,05	12,78	8		0,01	12,33	1		-0,11	12,65	11
	BAND 7		-0,02	10,78	3		-0,04	10,53	5		0,03	10,53	3
CANONICAL CORRELATION	r		0,90				0,89				0,83		
TAIL PROBABILITY			0,0000				0,0000				0,0300		

CC = CANONICAL COEFFICIENT
MEAN = MEAN OF DATA SET
SUCOL = SURFACE CHLOROPHYLL a
n = NUMBER OF SAMPLING POINTS
% = PERCENTAGE CONTRIBUTION
SUTUL = SURFACE TURBIDITY

TABLE 4.11: RESULTS OF THE CANONICAL CORRELATION ANALYSIS FOR INTEGRATED CHLOROPHYLL a/INTEGRATED TURBIDITY AND SATELLITE REFLECTANCE BANDS

82.09.30.

		INCLUDING ALL DATA				EXCLUDING OUTLIERS				NORMALISED DATA			
DATE: 82.09.30		n 32	CC	Mean	%	n 30	CC	MEAN	%	n 17	CC	MEAN	%
INDEPENDENT X VARIABLES	INCOL		-0,66	1,35	17		-0,2	1,33	5		-7,12	1,34	43
	INTUL		6,26	0,71	83		6,75	0,69	95		18,45	0,69	57
DEPENDENT Y VARIABLES	BAND 4		0,14	21,31	41		0,23	21,2	55		0,01	21,24	3
	BAND 5		0,11	14,13	22		0,17	13,87	27		0,24	13,71	42
	BAND 6		0,14	12,78	25		0,04	12,33	6		0,18	12,65	29
	BAND 7		-0,08	10,78	12		-0,1	10,53	12		-0,19	10,53	26
CANONICAL CORRELATION	r		0,92				0,92				0,92		
TAIL PROBABILITY			0,0000				0,0000				0,0000		

CC = CANONICAL COEFFICIENT
MEAN = MEAN OF DATA SET
INCOL = INTEGRATED CHLOROPHYLL a
n = NUMBER OF SAMPLING POINTS
% = PERCENTAGE CONTRIBUTION
INTUL = INTEGRATED TURBIDITY

TABLE 4.12: RESULTS OF THE CANONICAL CORRELATION ANALYSIS FOR SURFACE CHLOROPHYLL a/SURFACE TURBIDITY AND SATELLITE REFLECTANCE BANDS

82.11.16.

		INCLUDING ALL DATA				EXCLUDING OUTLIERS				NORMALISED DATA			
DATE: 82.11.16		n 28	CC	Mean	%	n 24	CC	MEAN	%	n	CC	MEAN	%
INDEPENDENT X VARIABLES	SUCOL		2,67	1,4	89		4,28	1,34	88				
	SUTUL		0,58	0,78	11		1,02	0,75	12				
DEPENDENT Y VARIABLES	BAND 4		0,3	21,5	42		0,38	21,38	30				
	BAND 5		-0,05	22,39	8		-0,04	22,17	3				
	BAND 6		0,24	20,79	32		0,52	19,46	38				
	BAND 7		-0,13	21,79	18		-0,38	20,63	29				
CANONICAL CORRELATION	r		0,95				0,91						
TAIL PROBABILITY			0,0000				0,0000						

CC = CANONICAL COEFFICIENT
MEAN = MEAN OF DATA SET
SUCOL = SURFACE CHLOROPHYLL a
n = NUMBER OF SAMPLING POINTS
% = PERCENTAGE CONTRIBUTION
SUTUL = SURFACE TURBIDITY

TABLE 4.13: RESULTS OF THE CANONICAL CORRELATION ANALYSIS FOR INTEGRATED CHLOROPHYLL a/INTEGRATED TURBIDITY AND SATELLITE REFLECTANCE BANDS

82.11.16.

		INCLUDING ALL DATA				EXCLUDING OUTLIERS				NORMALISED DATA			
DATE: 82.11.16		n 28	CC	Mean	%	n 24	CC	MEAN	%	n 24	CC	MEAN	%
INDEPENDENT X VARIABLES	INCOL		2,65	1,39	87		4,25	1,33	87				
	INTUL		0,68	0,82	13		1,10	0,79	13				
DEPENDENT Y VARIABLES	BAND 4		0,29	21, 5	41		0,35	21,38	28				
	BAND 5		-0,05	22,39	7		-0,02	22,17	2				
	BAND 6		0,24	20,79	32		0,53	19,46	40				
	BAND 7		-0,14	21,79	20		-0,38	20,63	30				
CANONICAL CORRELATION	r		0,95				0,90						
TAIL PROBABILITY			0,0000				0,0001						

CC = CANONICAL COEFFICIENT
MEAN = MEAN OF DATA SET
INCOL = INTEGRATED CHLOROPHYLL a
n = NUMBER OF SAMPLING POINTS
% = PERCENTAGE CONTRIBUTION
INTUL = INTEGRATED TURBIDITY

When extending the interpretation of results to the other two options 'Excluding Outliers' and 'Normalised Data' it becomes apparent that the trend of both surface and integrated turbidity relating to bands 4 and 5 is repeated. Both the independent and dependent variables appear to become more polarised with the normalisation procedure. Only for one other day's data, 82.09.13 (Tables 4.8 and 4.9), does the abovementioned trend follow in a similar vein. On two occasions, 81.12.07 (Table 4.7) and 82.09.30 (Table 4.11), integrated turbidity is related to band 4. Conversely though, on 81.12.07 (Table 4.6) surface turbidity appears to be related to band 6. On 82.09.30 (Table 4.10) surface chlorophyll a is connected with band 4. Further discrepancies are found on 82.11.16 (Tables 4.12 and 4.13) where both surface and integrated chlorophyll a appear to be related to band 4. The existence of the abovementioned discrepancies is not unexpected and emphasises the need to examine all of the different approaches. There are many factors that affect the relationship between water quality conditions and satellite reflectance values and either of the approaches may be applicable depending on the situation.

A major reason for the inconsistencies could be the presence of different algal species in the impoundment at the time of the overflights (refer to Section 2.3.3). Figure 4.1 gives an idea of the distribution of the main algal genera present at specific points in the impoundment (refer to Figure 3.2). It is apparent that the proportion of the genera is fairly constant at each site. Equally relevant is the fact that the genera change from month to month. The green algae (*Oöcystis* sp, *Cryptomonas* sp and *Pediastrum* sp) are largely evident in October 1981 and September 1982. The blue-greens (*Microcystis* sp, *Anabeana* sp and *Chroöccus* sp) appear in December 1981 and November 1982. The diatoms (*Melosira* sp) appear in November and December 1981.

Overall the green algal species appear to be the most prevalent. The possiblity that blue-green algae reflect light differently from green algae is a question beyond the scope of this study.

Figure 4.2 illustrates the seasonal cycle of major algal species in Roodeplaat Dam over the period of the Landsat Water Quality Project (Young and Silberbauer, 1984). Unfortunately surface reference data and satellite reflectance data were not obtained when high concentrations of algae were present.

The relative stages of algal growth also might be important in understanding the inconsistencies in the results.

The presence of phaeopigments was examined to determine a possible influence, if any, on the data (refer to Section 2.3.6).

Canonical Correlation Analysis using phaeopigment data was attempted but the complexity of the results and the obvious intercorrelations made it difficult to come to a definite conclusion. In an attempt to assess the influence of phaeopigment on the Canonical Correlation Analysis of chlorophyll a, a simple linear regression analysis was carried out. All six day's data were analysed using the 'Including All Data' coefficients for the individual reflectance bands. Table 4.14 lists the results.

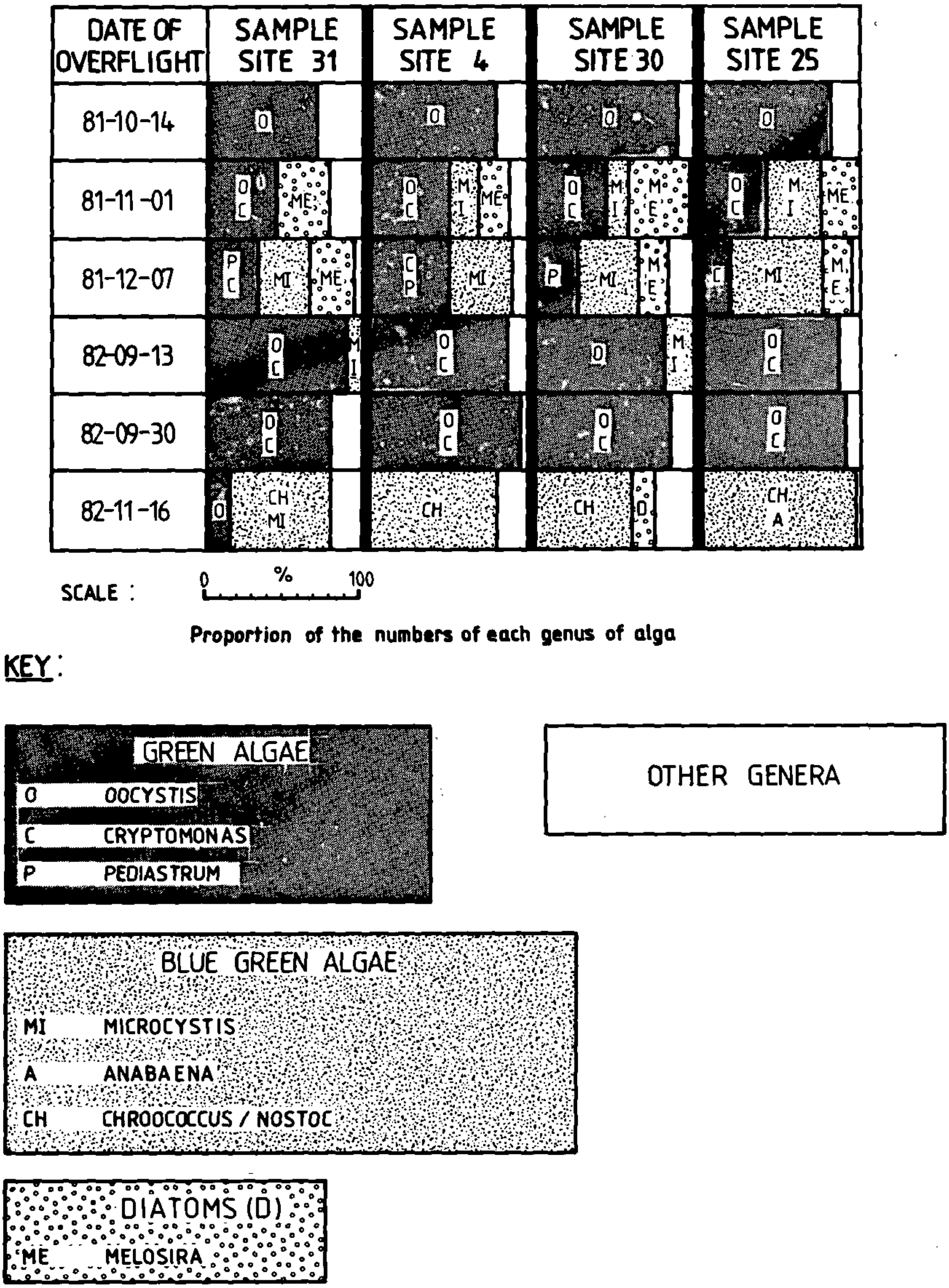


FIGURE 4.1: MAJOR ALGAL GENERA PRESENT AT FOUR SAMPLING SITES ON ROODEPLAAT DAM AT THE TIME OF THE SATELLITE OVERFLIGHTS

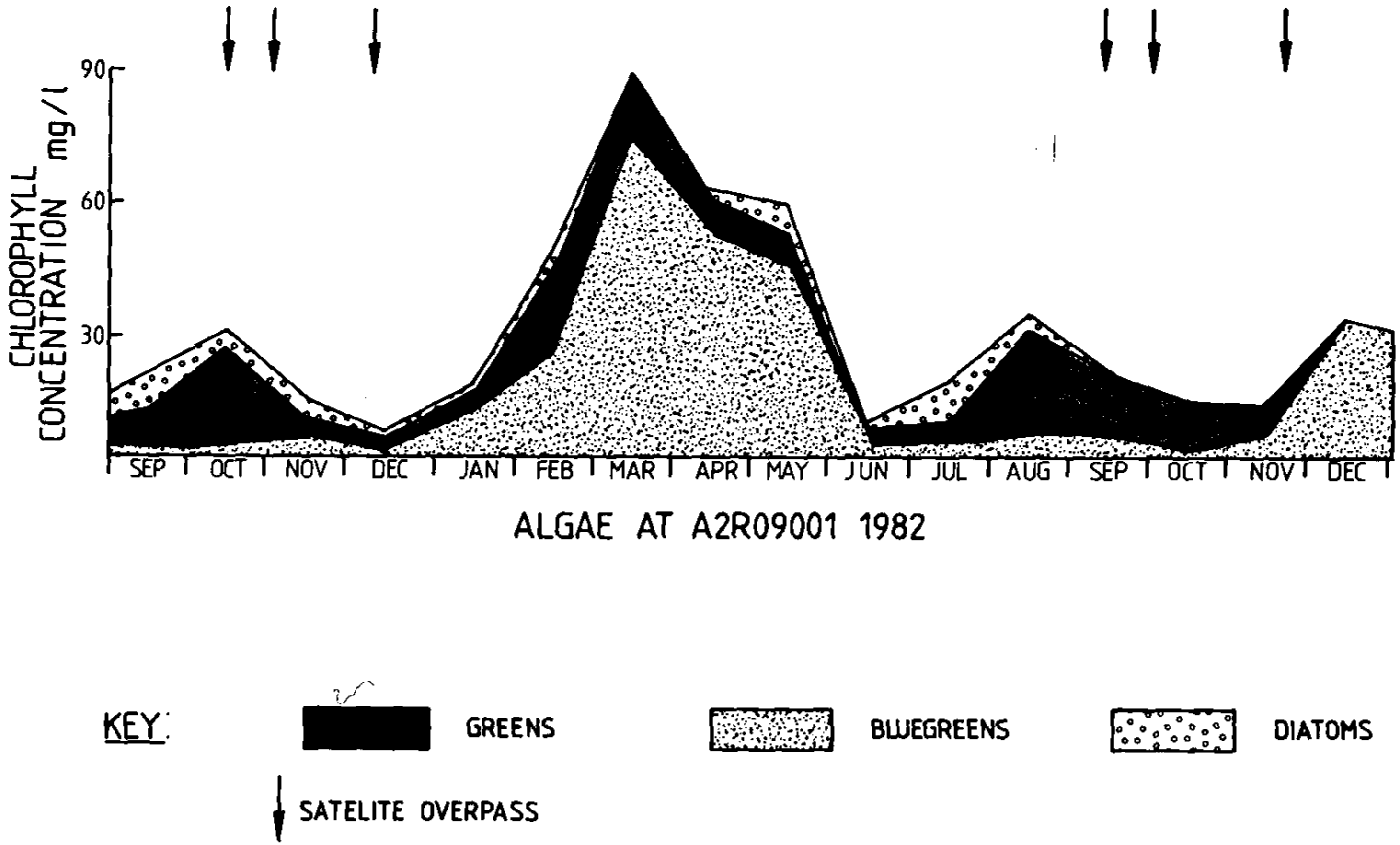


FIGURE 4.2: SEASONALITY IN ROODEPLAAT DAM (YOUNG AND SILBERBAUER, 1984)

TABLE 4.14: LINEAR REGRESSION ANALYSIS OF PHAEOPIGMENT WITH THE CANONICAL COEFFICIENTS OF THE REFLECTANCE BANDS

DATE	RATIO OF PHAEOPIGMENT/ TOTAL PIGMENT	CANONICAL COEFFICIENTS BAND 7	BAND 6	BAND 5	BAND 4
81.10.14	0,129	0,06	0,025	0,084	0,387
81.11.01	0,219	-0,198	0,28	0,082	0,195
81.12.07	0,264	-0,121	0,096	0,055	0,048
82.09.13	0,164	-0,1	0,102	0,364	0,052
82.09.30	0,164	-0,024	0,05	0,062	0,313
82.11.16	0,110	-0,131	0,238	-0,053	0,301
r		-0,47	0,082	0,089	-0,70
Y intercept		0,045	0,106	0,614	0,52
Slope		-0,007	0,001	0,002	-0,017

The results of this brief analysis indicated very low correlation coefficients between phaeopigment/total pigment ratio and the Canonical Correlations of the reflectance bands 5 and 6 (r = 0,089 and r = 0,082 respectively), but a negative correlation between the Canonical Coefficients for band 7 (r = -0,47) and band 4 (r = -0,70). This shows that the state of health of the algae, as reflected in the amount of phaeopigment present, has a noticeable effect, particularly for the Canonical Coefficient of band 4. The influence of phaeopigments on the calibration of surface reference data relative to satellite reflectance data, is something that would be worth pursuing but is beyond the scope of this study.

4.4 AN INTERPRETATION OF THE COLOUR CODING

The digital reflectance data in colour coded format provided visual impressions of conditions in the impoundment at different times (refer to Section 3.3.3).

Data for the 82.09.30, illustrated on Plates 4.1 to 4.4 indicate relatively high reflectance values all over the impoundment in bands 4 and 5. Band 6 shows a heterogeneous range of values while band 7 has fairly low reflectances. This could indicate the presence of more turbid than chlorophyll laden water due to the high values recognisable in bands 4 and 5.

Tables 4.10 and 4.11, which present the Canonical Coefficients and the percentage contribution of each variable to the relationship on the 82.09.30 indicate that surface chlorophyll a and integrated turbidity are both related to band 4, a contradictory picture.

Plates 4.5 to 4.8 of the day 82.11.16, illustrate an error that can occur with satellite data. A radiometric error has caused some data to be lost. Plates 4.5 and 4.6 show bands 4 and 5 to have high reflectance values all over the impoundment. Bands 6 and 7 also have relatively high values indicating the possible presence of chlorophyll. Tables 4.12 and 4.13 support this observation with surface and integrated chlorophyll a relating to band 4 on 82.11.16.

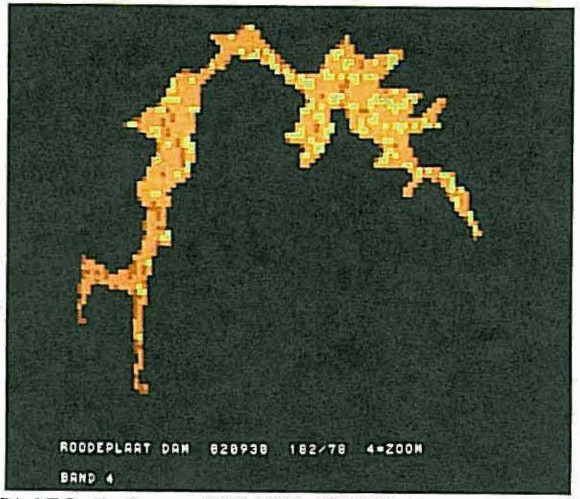


PLATE 4.1: COLOUR CODED REFLECTANCE
BAND 4 - 82.09.30

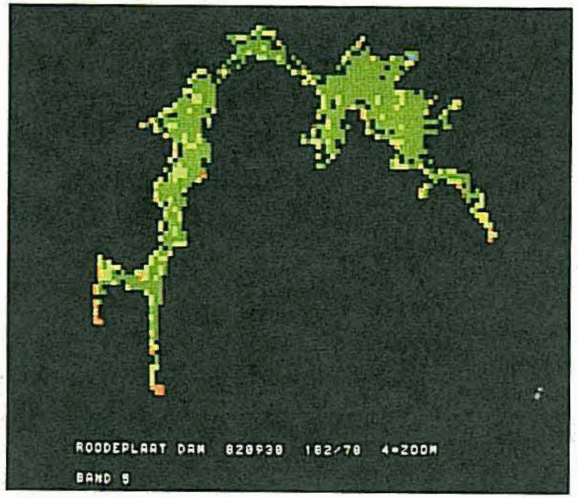


PLATE 4.2: COLOUR CODED REFLECTANCE
BAND 5 - 82.09.30

PLATE 4.3: COLOUR CODED REFLECTANCE
BAND 6 - 82.09.30

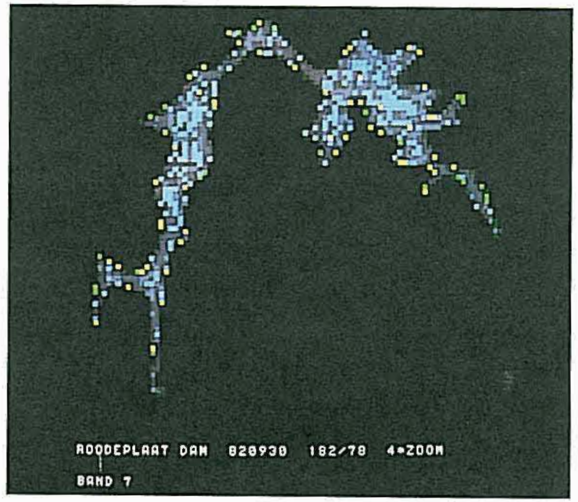


PLATE 4.4: COLOUR CODED REFLECTANCE
BAND 7 - 82.09.30

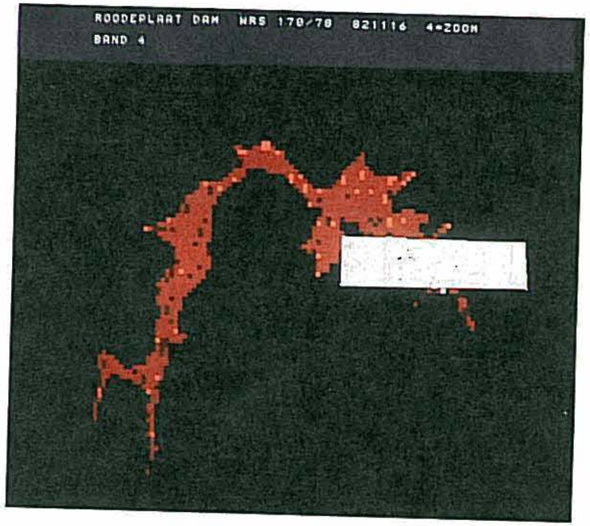


PLATE 4.5: COLOUR CODED REFLECTANCE
BAND 4 - 82.11.16

PLATE 4.6: COLOUR CODED REFLECTANCE
BAND 5 - 82.11.16

PLATE 4.7: COLOUR CODED REFLECTANCE
BAND 6 - 82.11.16

PLATE 4.8: COLOUR CODED REFLECTANCE
BAND 7 - 82.11.16

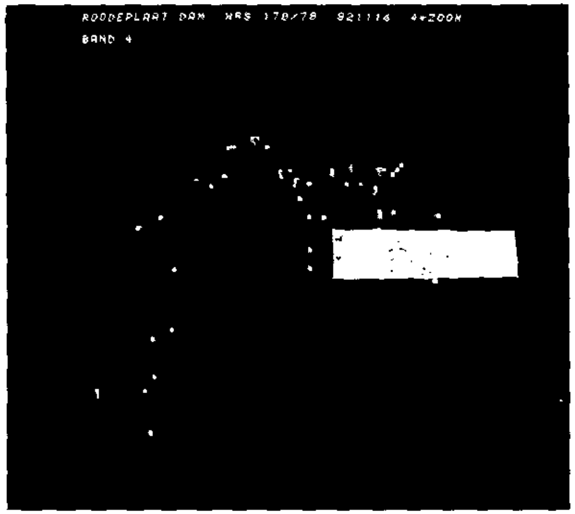


PLATE 4.5: COLOUR CODED REFLECTANCE
BAND 4 - 82.11.16

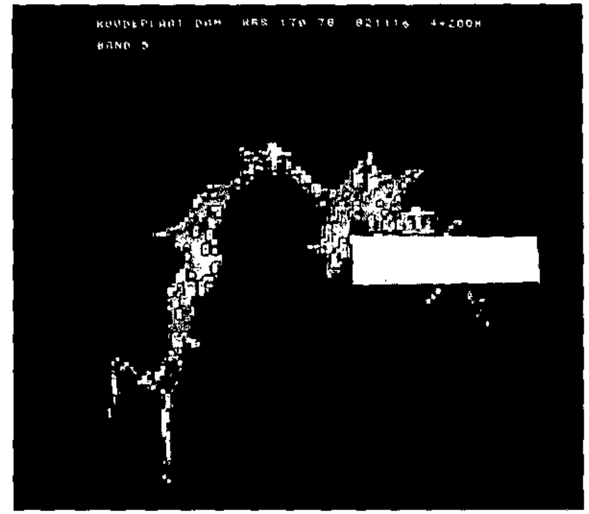


PLATE 4.6: COLOUR CODED REFLECTANCE
BAND 5 - 82.11.16

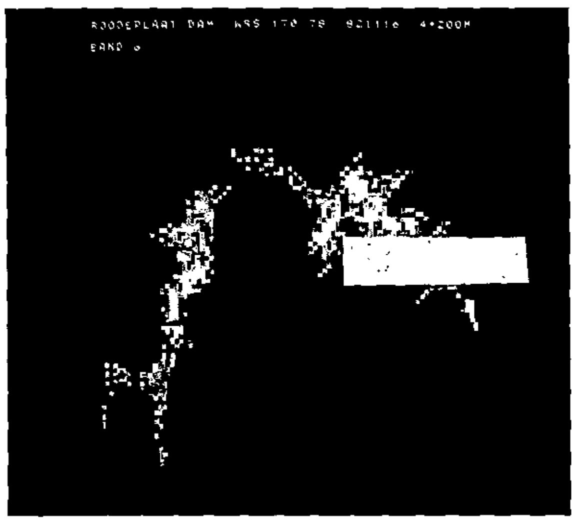


PLATE 4.7: COLOUR CODED REFLECTANCE
BAND 6 - 82.11.16

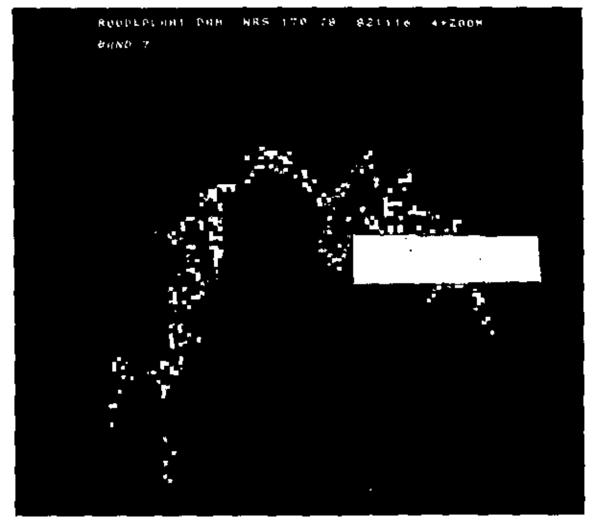


PLATE 4.8: COLOUR CODED REFLECTANCE
BAND 7 - 82.11.16

Colour coding therefore enables qualitative and comparative observations to be made. The degree of heterogeneity can be assessed, but the distinction between chlorophyll a and turbidity distributions is not always obvious, particularly when low concentrations of both chlorophyll and turbidity are present. Finally, a problem associated with colour coding is the photographic process involved, which can cause variation between the colour distributions.

4.5 UNSUPERVISED CLASSIFICATION

Using a modified image processing program (Modified CATNIPS) an unsupervised classification of 4 wave bands for each day's data produced classified digital images of Roodeplaat Dam. The different signatures on the classified image were represented by symbols to enhance the visual effect. One image for 81.12.07 showed outstanding differences in the classification (Figure 4.3). The statistics performed on the data (Table 4.15) indicate that bands 4 and 5 mainly account for 3 of the 4 reflectance classes. The two more significant classes ($,M) evident along the western arm of the impoundment indicate different water quality conditions. The relationship to bands 4 and 5 suggests the presence of suspended sediments (mean reflectance units of 15,621 and 20,979 for band 4 and 12,807 and 22,643 for band 5.

The 4th class (-), registering a relatively high value in band 7 (16,669) and found only along the shoreline can be considered to be mixels (mixed pixels of water and vegetation).

Plate 4.13 represents the unsupervised classification for 81.12.07 obtained from the SRSC. The image shows four distinct classified regions, the areas which they cover and the turbidity concentrations that have been attributed to each class.

TABLE 4.15: STATISTICS OF THE UNSUPERVISED CLASSIFICATION FOR 81.12.07

CLUSTER	NUMBER OF PIXELS	R MEAN	R SIGMA
M	140	12,93	1,16
$	140	14,90	3,26
.	568	7,34	2,93
-	133	10,16	3,80

		MEANS	4 by 20	
BANDS	M	$	.	-
4	15,621	20,979	6,845	8,414
5	12,807	22,643	4,607	8,902
6	8,221	17,086	3,968	16,218
7	4,871	9,693	3,928	16,669

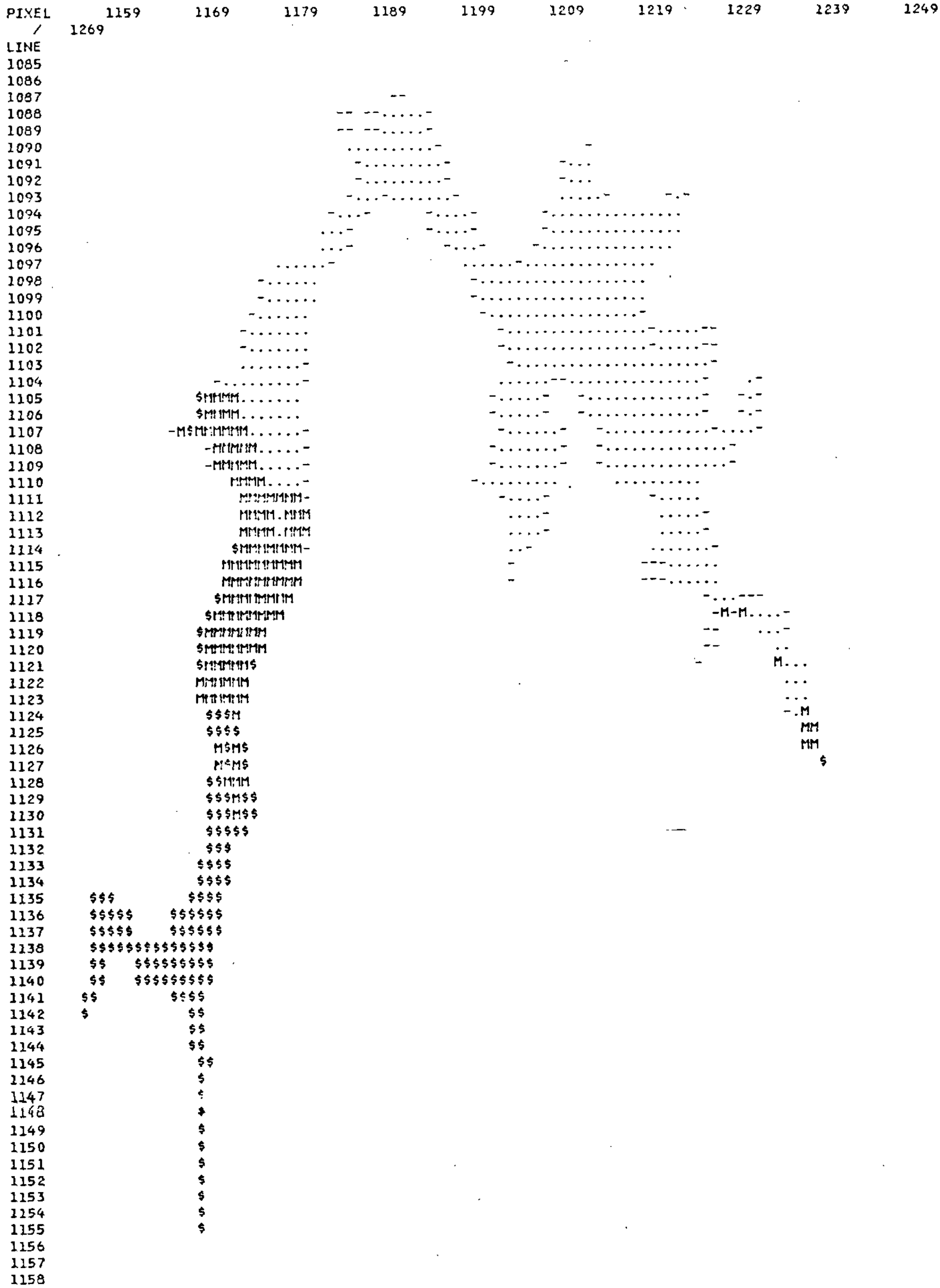


FIGURE 4.3: UNSUPERVISED CLASSIFICATION OF ROODEPLAAT DAM 81-12-07

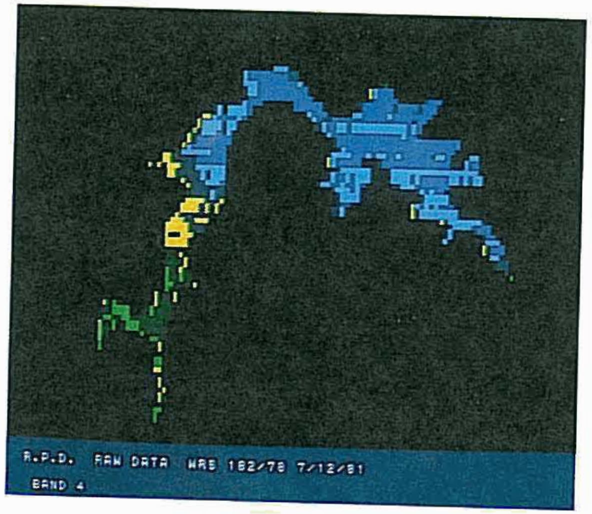


PLATE 4.9: COLOUR CODED REFLECTANCE BAND 4 - 81.12.07

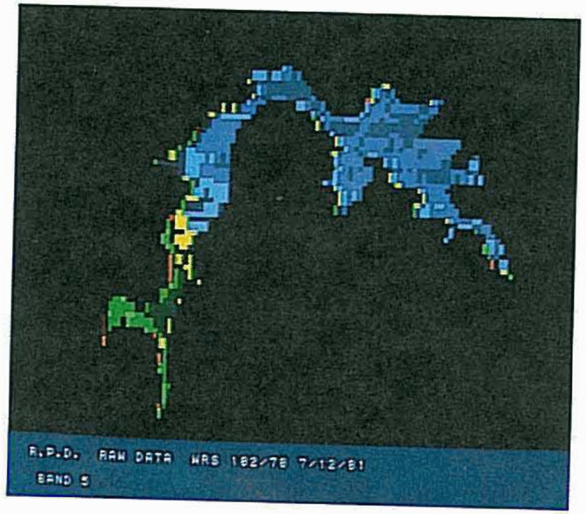


PLATE 4.10: COLOUR CODED REFLECTANCE BAND 5 - 81.12.07

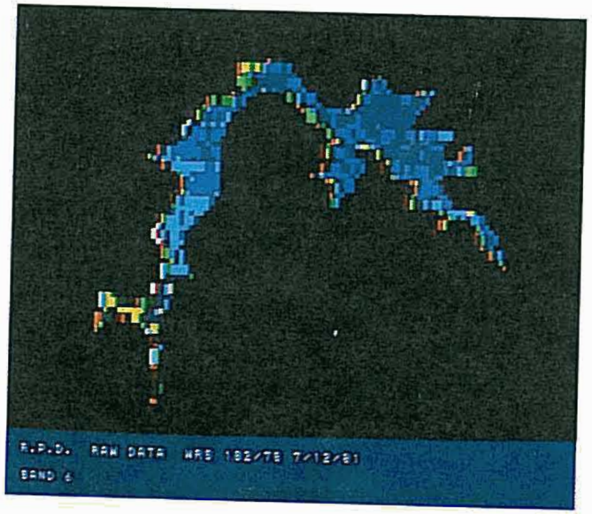


PLATE 4.11: COLOUR CODED REFLECTANCE BAND 6 - 81.12.07

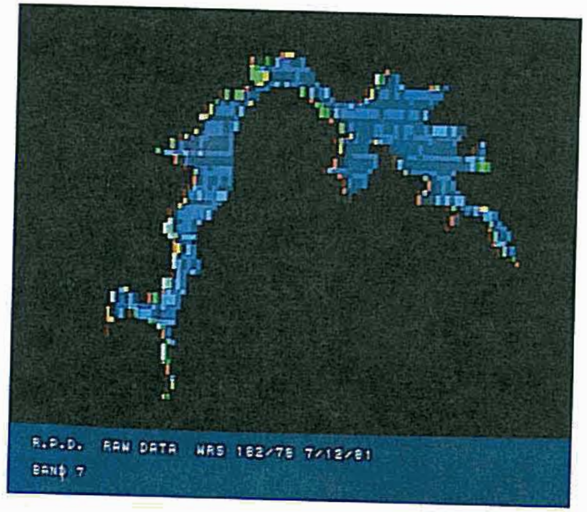


PLATE 4.12: COLOUR CODED REFLECTANCE BAND 7 - 81.12.07

4.6 OVERVIEW OF ONE DAY'S DATA 81.12.07

It is necessary to collate information for one day's data in order to obtain a perspective of the issues involved. The data for 81.12.07 was chosen for examination because the data contained the widest range of distributed water quality conditions as judged from the colour coded images.

To reconstruct a picture of the data already presented:

Plates 4.9 to 4.12 illustrate the colour coded reflectance data for 81.12.07.
Figure K.2 in Appendix K presents the Stepwise Discriminant Analysis for 81.12.07.
Tables 4.6 and 4.7 disclose the Canonical Correlation, Canonical Coefficients and each variables percentage contribution to the relationship for 81.12.07.
Figure 4.3, Table 4.15 and Plate 4.13 represent information concerning the unsupervised classification of the day's data.

Plates 4.9 and 4.10, the colour coded images for bands 4 and 5 clearly indicate the presence of water quality conditions along the southern part of the left arm of Roodeplaat Dam. Bands 6 and 7 (Plates 4.11 and 4.12) also indicate differing conditions but to a lesser extent. The remainder of the impoundment appears to be relatively homogeneous.

Figure K.2 illustrates that potentially 3 different populations are present in the impoundment: P, representing the southern most polluted arm of the impoundment; C, the canoe lanes along the western arm and B and D, the low reflectances, depicting clear water of the main body of the impoundment. The one Canonical variable that dominates the analysis is surface turbidity.

Table 4.6 reveals that for this overpass surface turbidity is by far the dominant variable (98%) and that all of the bands contribute fairly equally to the relationship. Band 6 shows a slight head (30%). The Canonical Correlation of 0,94 is high. It is noteworthy that there are no outliers in the data for this image and there was no necessity to normalise the data.

Table 4.7 affirms the presence of turbid water with a fairly high contribution (65%) of integrated turbidity. Band 4 (41%) in particular appears to be related to the integrated turbidity. The high Canonical Correlation of 0,95 suggests a good relationship.

PLATE 4.9: COLOUR CODED REFLECTANCE BAND 4 - 81.12.07

PLATE 4.10: COLOUR CODED REFLECTANCE BAND 5 - 81.12.07

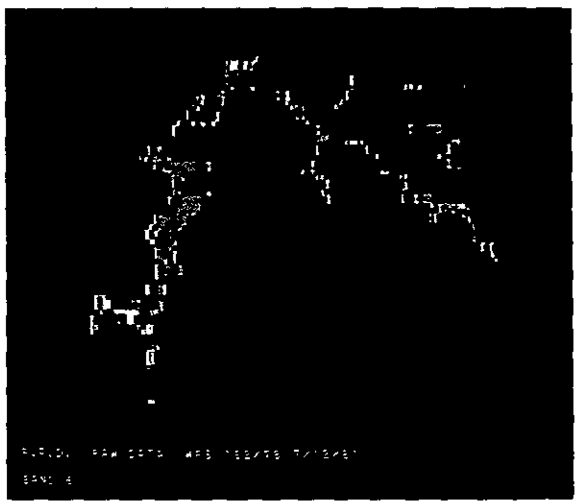


PLATE 4.11: COLOUR CODED REFLECTANCE BAND 6 - 81.12.07

PLATE 4.12: COLOUR CODED REFLECTANCE BAND 7 - 81.12.07

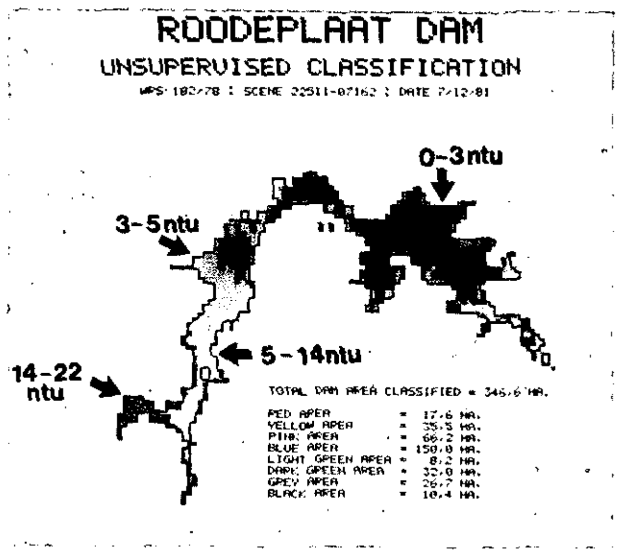


PLATE 4.13: UNSUPERVISED CLASSIFICATION OF ROODEPLAAT DAM - 81.12.07

The unsupervised classification and colour coded images were further reinforced by a classification of the 81.12.07 image being produced by an image processing system at Hartbeesthoek, Plate 4.13. The classification identified 8 classes, 4 which could be considered to be border classes indicating mixed areas of vegetation and water. The remaining 4 classes distinguished different water quality conditions.

From the abovementioned results, due to the obvious weight in favour of turbidity, the image classes were compared with the surface reference data for surface and integrated turbidity (Figure 4.4). Five turbidity categories became apparent (Figure 4.5).

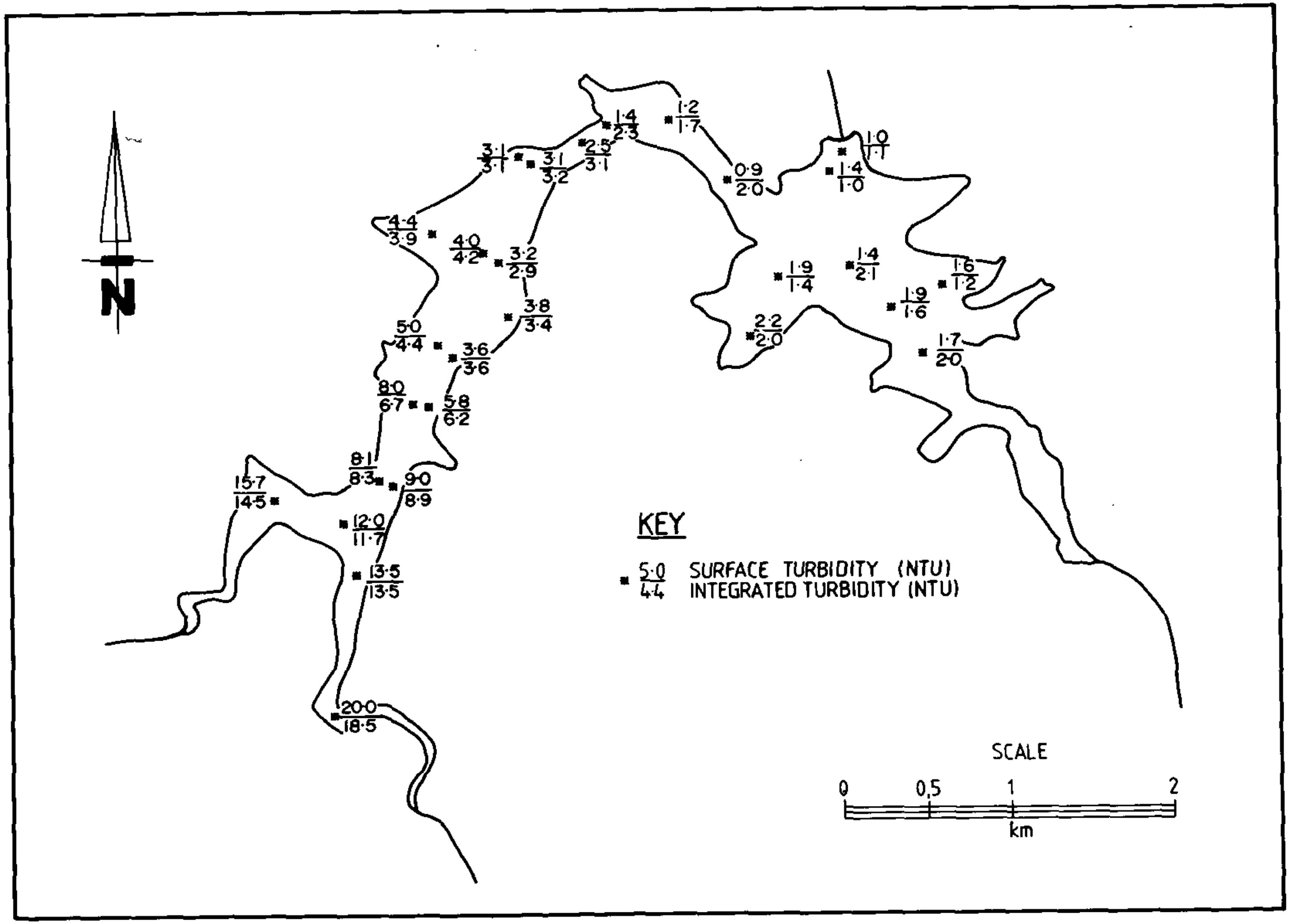


FIGURE 4.4: SURFACE AND INTEGRATED TURBIDITY SURFACE REFERENCE DATA FOR 81·12·07

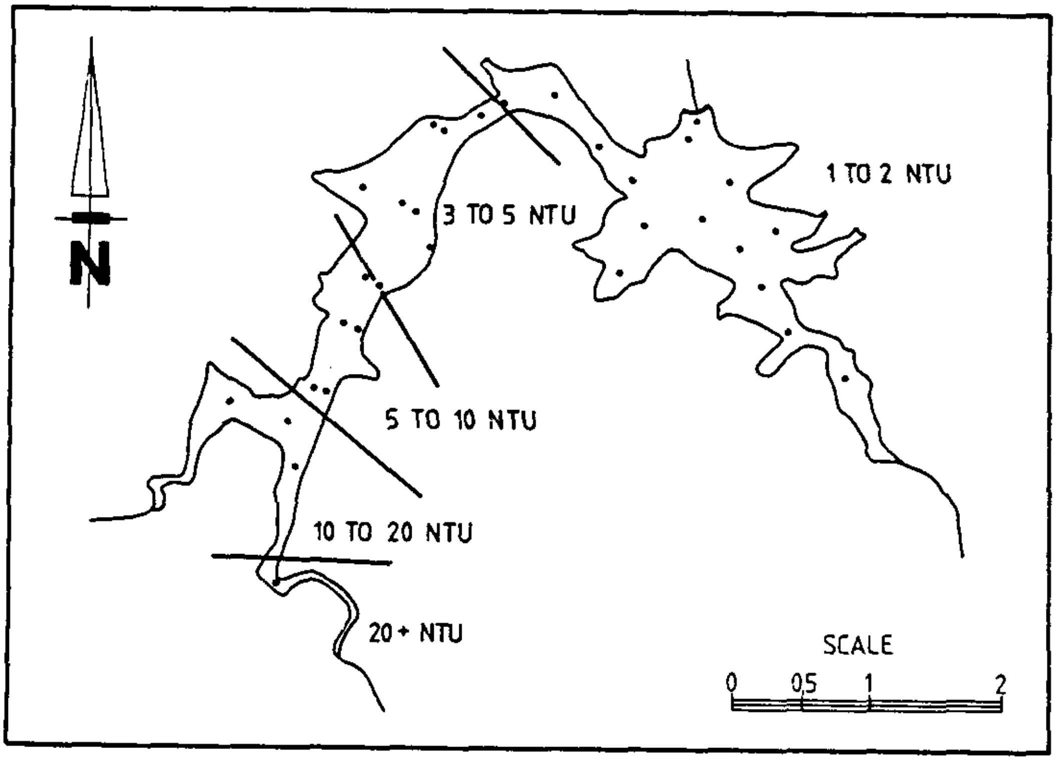


FIGURE 4.5: TURBIDITY CLASSES FOR ROODEPLAAT DAM - 81·12·07

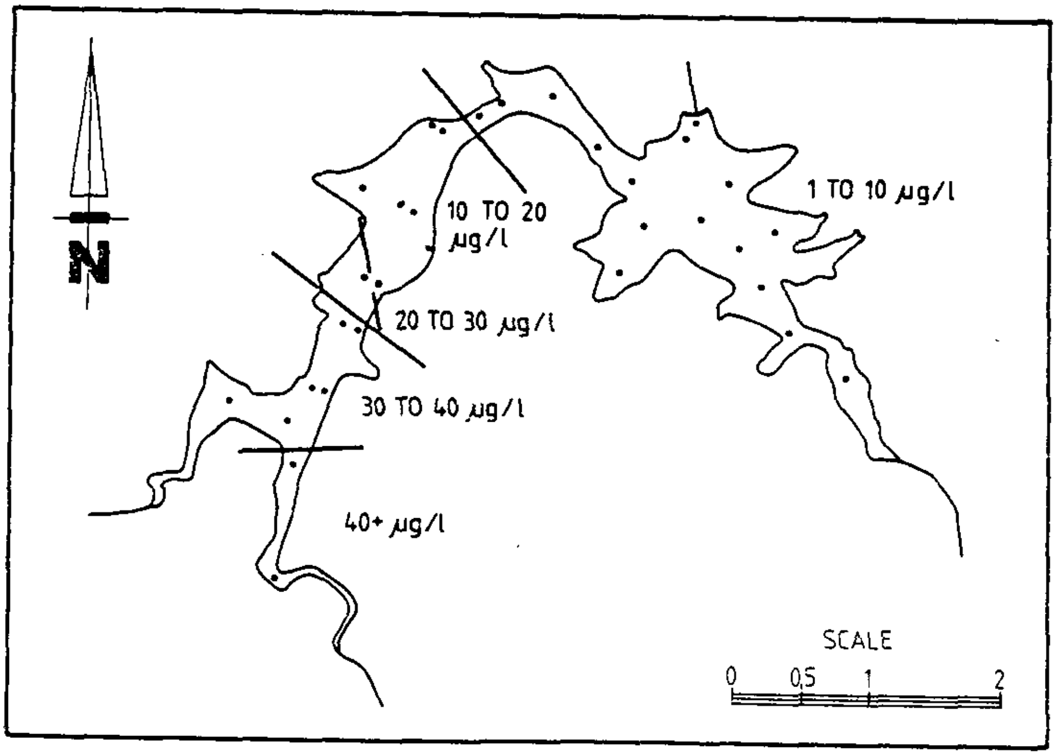


FIGURE 4.7: CHLOROPHYLL CLASSES FCR ROODEPLAAT DAM - 81·12·07

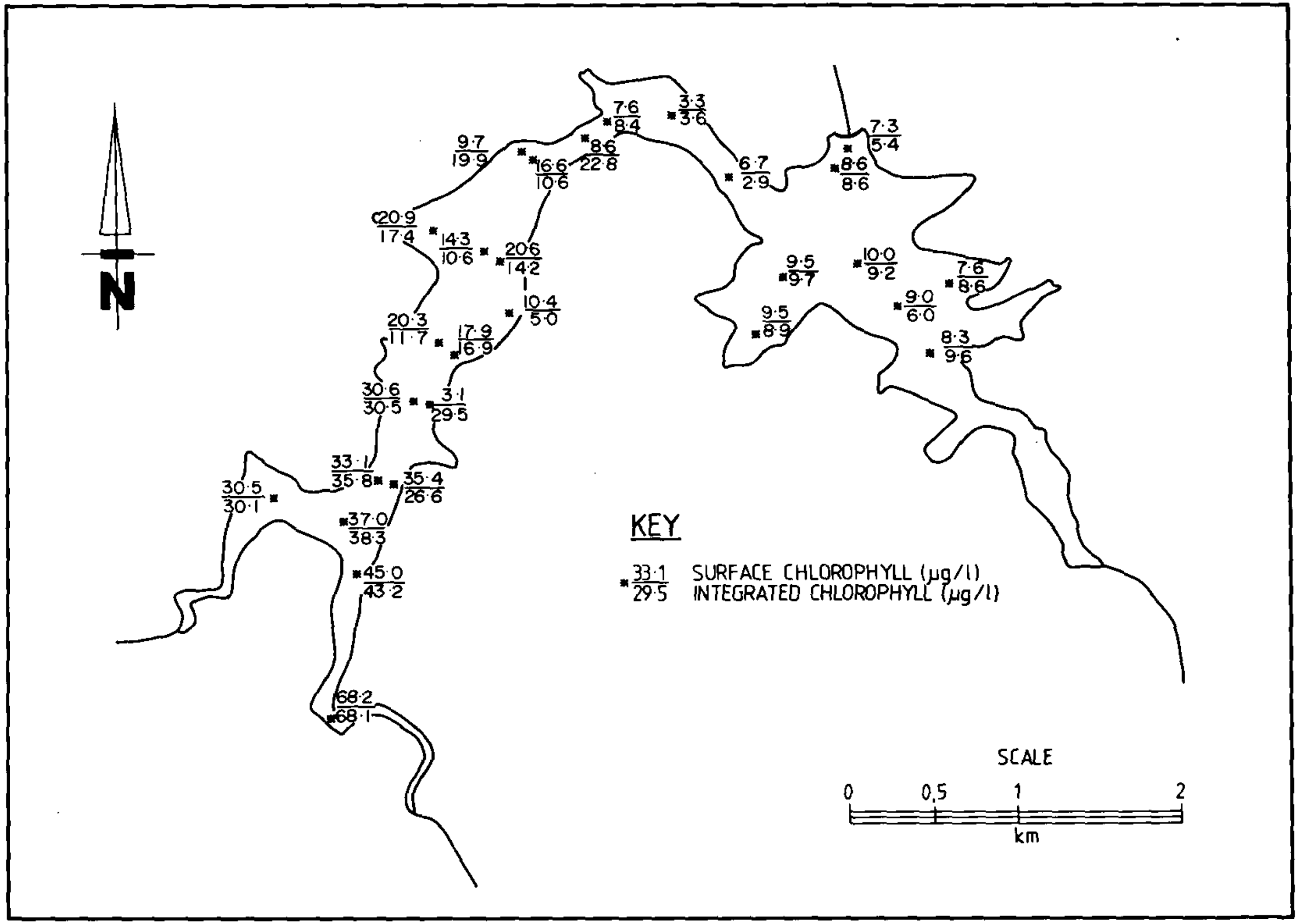


FIGURE 4.6: SURFACE AND INTEGRATED CHLOROPHYLL SURFACE REFERENCE DATA FOR 81·12·07

A query arose as to the importance of bands 6 and 7 which had fairly high percentage contributions in the relationship (30 and 26% respectively). Plate 4.13 classes were compared with surface and integrated chlorophyll a data (Figure 4.6). The results given in Figure 4.7 illustrate the high contribution chlorophyll a maintains in the overall context.

This information immediately identifies a major problem associated with distinguishing the difference between chlorophyll a and turbidity. Turbidity was previously identified into classes of 0-3, 3-5, 5-14 and 14-22 NTU (refer to Plate 4.13). In similar areas, chlorophyll a was also present at concentrations of approximately 0-10 μg/ℓ, 10-20 μg/ℓ, 20-30 μg/ℓ and 30 + μg/ℓ respectively.

All of the problems discussed in Sections 2 and 3 suddenly become highly relevant. The necessity for applying multi-variate analysis to the data, the instability of the Canonical Correlation Analysis and the multi-collinearity of the surface reference data and satellite reflectance data were all made very apparent.

4.7 SUMMARY

The initial results of the investigation have highlighted some important points. Firstly, there is a distinct correlation between specific water quality conditions and satellite reflectance data. Secondly, the relationship between the dependent and independent data sets is a complex one and it is difficult to isolate individual relationships. Thirdly, the question of which statistical approach should be followed in order to obtain the most accurate results is very difficult to ascertain. Fourthly, chlorophyll a and turbidity, particularly at low concentrations, are interrelated.

In order to gain quantitative results it is therefore essential to build on the basis of multi-variate analysis, incorporating the problem of multicollinearity of the data set, and attempt to establish a model with which to simulate water quality conditions from satellite-derived information. A model has been attempted and will be discussed in the following chapters.

CHAPTER 5

USE OF THE CANONICAL COEFFICIENTS FOR SIMULATIVE PURPOSES

5.1 INTRODUCTION

Canonical Correlation Analysis, provides a set of multi-variate coefficients and correlation coefficients, which represent, in this instance, the relationship between surface reference data and satellite reflectance data.

A problem inherent in Canonical Correlation Analysis is that the equations produced are in their implicit form (refer to Section 4.3) which means that the equations require solution before they can be understood and used for simulative purposes.

Therefore a method was established which could incorporate the Canonical Coefficients and the multi-linear relationship, in such a way that quantifiable and interpretable results could be acquired for simulative purposes and the most appropriate of the three different approaches 'Including All Data', 'Excluding Outliers' and 'Normalised Data' could be assessed.

5.2 OBTAINING THE LINEAR REGRESSION EQUATION

The Canonical Correlation Analysis is an extension of linear regression analysis and a linear regression equation in the form of Y = MX + K can be obtained, for which Y and X are linear polynomial variables. The coefficients for each variable in the polynomials are determined in the canonical analysis, the slope (M) of the regression line and the Y intercept (K) need to be established, however. This was carried out by inserting the Canonical Coefficients, together with the respective surface reference and satellite reflectance data, into a linear regression program. The linear regression program 'LINREG' is presented in Appendix L. In this way the slope (M) of the regression line and the Y intercept (K) were obtained and the linear regression line for each set of data for each day was acquired. An example of results obtained from program 'LINREG' is given in Appendix M. Values obtained for M and K for each day and for each of the three approaches are presented in Appendix N. The M and K values differed for each day and with each different approach. This information suggested that the relationship, between the surface reference data and the satellite reflectance data, was unique to each specific overpass.

It must be pointed out that values for the combination of variables, surface chlorophyll a/integrated chlorophyll a and surface turbidity/integrated turbidity are included in Appendix N. As there are four surface reference data unknowns it is essential to have four simultaneous equations in order to solve explicitly for the surface reference variables. The statistical assumption requiring interdependency between variables was ignored.

The four equations to be solved are in the form

$$Y = MX + K \text{ as discussed in Section 4.3}$$

The solution of the four simultaneous equations with 4 known and 4 unknown variables made it possible for the model CALMCAT* to be set up which could simulate water quality conditions (refer to Section 6.1).

5.3 SOLVING THE SIMULTANEOUS EQUATIONS

The solution of the four simultaneous equations (Appendix O) enabled a model to be produced. By substituting the respective Canonical Coefficients and M and K values into the model the calibration equations for each day and for each of the three approaches were obtained. Values could be calculated for the four water quality variables for each pixel of the impoundment, by entering the corresponding reflectance values of the four wavebands. This made possible the determination of chlorophyll a and turbidity concentrations at specific sites in the impoundment.

Appendix P presents the subroutine used to determine concentrations at specific sites. Appendix Q is an example of the calibration data set required to run Subroutine "Convrt".

5.4 TESTING THE ACCURACY OF THE CALIBRATION EQUATIONS

To determine the accuracy of CALMCAT and the calibration equations obtained from the Canonical Correlation Analysis, the linear regression program and the solving of the 4 simultaneous equations, it was necessary to test the equations. For three of the sampling days (82.09.13, 82.09.30 and 82.11.16), 55 sampling points had been sampled, but only 32, or fewer depending on the alternative applied, had been used in the calibration of the models for each specific day i.e., the establishment of the Canonical Coefficients and Canonical relationship. Therefore data from 23 sampling points on 2 of the days and 22 sampling points on one day were available to test the accuracy of the models+ (refer to Sections 3.2.1 and 6.1). These sampling points, which had not been used in the Canonical Analysis were termed the verification data set. The concentrations of each water quality variable at the verification data sites were simulated using the model CALMCAT and the simulated values were compared with the observed values.

Two indicators were used to assess the performance of the models and the calibration equations. Firstly a coefficient of efficiency of model performance was used to examine the accuracy of the calibration equations on both the original calibrated data set, as well as the verification data set that had not previously been used in the model development or calibration thereof. Secondly the Student's t test and the percentage relative error, between the simulated and the observed mean values of the verification data were determined.

---

* CALMCAT - Canonical Analysis Landsat Model of Chlorophyll a and Turbidity

\+ CALMCAT has three variants i.e. Including All Data, Excluding Outliers, Normalised Data

In this instance, 'simulated' values are concentrations calculated for pixels from the reflectance values by the CALMCAT model, and therefore they represent the simulated ambient water quality conditions present in the impoundment.

5.4.1 The Coefficient of Efficiency of Model Performance

The coefficient of efficiency of model performance (Nash and Sutcliffe, 1970) is "an index of one to one correspondence that is sensitive to systematic errors in the model output" (Roberts, 1978). The statistic has the form

$$\text{Coefficient of Efficiency} = \left(\frac{\Sigma(O_i - MO_i)^2 - \Sigma(O_i - S_i)^2}{\Sigma(O_i - MO_i)^2} \right)$$

where O and S represent observed and simulated data respectively and M represents the mean.

The coefficient of efficiency essentially determines the closeness of the observed versus simulated data to the 45° line* on a graph. The closer the regression line lies to
the 45° line the higher the coefficient of efficiency. Used in conjunction with the coefficient of determination (R^2) "the value of the coefficient of efficiency will be lower than the coefficient ofdetermination if the results from the model are highly correlated but biased" (Aitken, 1973). Ideally, the coefficient of efficiency values should approximate to 1,0, with intercept values of 0 and slope of 1,0.

5.4.2 The Student's t Test

The Student's t test from the SPSS Statistical Package for the Social Sciences (SPSS), (Nie et al, 1975), was used to test the similarity between the observed and the simulated data set means. The equation for the t test is;

$$\bar{td} = \left(\frac{(\bar{x}_1 - \bar{x}_2)}{s\bar{d}} \right) \quad \text{with } (n_1 + n_2 - 2) \text{ degrees of freedom,}$$

where $(\bar{x}_1 - \bar{x}_2)$ is the difference between the two means and

where

$$s\bar{d} = s^2/n_1 + s^2/n_2$$

is the pooled variance and n_1 and n_2 are the respective sample sizes of the two groups.

---

* Equal value line

The t test gives an indication of the significance of the difference between the means. The closer the t value to zero, the better the fit, whereas the larger the value, (sign ignored) the poorer the simulation.

If the absolute value of t (the sign ignored) is greater than the critical value of t obtained from a table of the t distribution then there is a significant difference between the two means. The critical value for 44 degrees of freedom at the 5% two tailed level of significance is 2,02.

The data were antilogged before the t test analysis.

The percentage relative error between the simulated and the observed mean values were calculated using the following equation:

$$\left(\frac{\text{Simulated Mean} - \text{Observed Mean}}{\text{Observed Mean}} \times 100\right) = \text{Percentage relative error}$$

5.5 RESULTS AND DISCUSSION OF RESULTS

The results of the coefficient of efficiency analysis for the calibration data set, the results of the t test analysis, the percentage relative error and the coefficient of efficiency analysis for the verification data set, for each option and for each day are presented and are discussed.

5.5.1 Overpass of 82.09.30

The coefficient of efficiency analysis of the Calibration Data set for the 'Including All Data' Option, 82.09.30, is shown in Table 5.1. The mean and standard deviation values between the observed and simulated data are comparable. The coefficients of efficiency and determination are all greater than 0,73 with the exception of surface turbidity. The reason for surface turbidity having such low coefficients, while integrated turbidity indicates good correspondence levels between observed and simulated values, may lie in the noise of the observed data. Of the four variables; surface turbidity is the only variable showing a standard deviation for the observed data greater than the observed mean. The results indicate that the calibration equations provide an acceptable fit with the possible exception of surface turbidity.

Table 5.2 listing the results of the verification data 'Including All Data' option for day 82.09.30, shows that the mean values of the observed and the simulated data, for each water quality variable, are very similar. The t value for surface chlorophyll a, of -0,06, suggested a good simulation for mean values, whereas integrated chlorophyll a had a high t value, of 3,38, indicating a poorer simulation. The t values for the turbidity variables are below the critical limit of 2,02 and are therefore acceptable. The percentage relative error for all of the variables are below 9% with a low error accredited to the simulation of surface chlorophyll a.

TABLE 5.1: COEFFICIENT OF EFFICIENCY ANALYSIS FOR THE CALIBRATION DATA SET,

'INCLUDING ALL DATA' OPTION FOR 82.09.30

	SUCOL	INCOL	SUTUL	INTUL
Number of samples	32	32	32	32
Mean of observed data	24,27	27,11	6,46	5,65
Mean of simulated data	25,72	27,65	5,47	5,73
Std. dev. of observed data	17,06	20,15	7,02	3,36
Std. dev. of simulated data	21,45	21,04	2,56	3,36
Regression intercept	5,4	3,65	1,10	0,36
Regression slope	0,73	0,85	0,98	0,92
Coeff. of determination R^2	0,85	0,79	0,13	0,86
Coeff. of efficiency	0,73	0,79	0,11	0,85

In contrast to the calibration data set which generally showed high coefficients of efficiency (Table 5.1), the coefficients of efficiency for the verification data set were poor. The reason for this discrepancy probably lies in the inherent noise in the surface reference data which places a fundamental limitation on the precision which may be achieved.

While an overall trend is observable for plots of simulated versus observed data (Figure 5.1 to 5.4) the relationship shows a high degree of scatter particularly over small ranges of the data (see Figure 5.1). It is the latter phenomenon which is responsible for the poor coefficients of efficiency for the verification data set as this data set only represents a small region of the total range in contrast to the calibration data set. Figure 5.3 indicates the reason for the low coefficients of efficiency obtained for the surface turbidity calibration data set (Table 5.1), as an extreme outlier with an observed surface turbidity of 41 NTU, not lying on the equal value line, is present.

The results of the 'Excluding Outliers' Option on the Calibrated Data set of 82.09.30, Table 5.3, presented similar mean and standard deviation values. The coefficient of determination for all of the variables including surface turbidity were acceptable but the coefficients of efficiency were less than 0,5, except for integrated turbidity where a value of 0,78 was found.

TABLE 5.2: ANALYSIS OF ACCURACY OF THE VERIFICATION DATA SET FOR 82.09.30, 'INCLUDING ALL DATA' OPTION

82.09.30	INCLUDING ALL DATA									
Water Quality Variables 23 Cases		Mean	Std. Dev.	Diff. Mean	t Test	% Relative Error	Regression Intercept	Regression Slope	R^2	Coeff. of Efficiency
Surface Chlorophyll a µg/ℓ	Observed	27,17	10,59	-0,11	-0,06	0,4	11	0,59	0,62	0,33
	Simulated	27,28	14,12							
Integrated Chlorophyll a	Observed	30,61	8,43	0,87	3,38	3	19,23	0,38	0,44	-0,72
	Simulated	29,75	14,68							
Surface Turbidity NTU	Observed	5,67	1,69	-0,52	-1,96	9	1,9	0,61	0,72	0,32
	Simulated	6,19	2,36							
Integrated Turbidity	Observed	5,86	1,59	-0,29	-1,01	5	2,48	0,55	0,75	0,22
	Simulated	6,15	2,50							

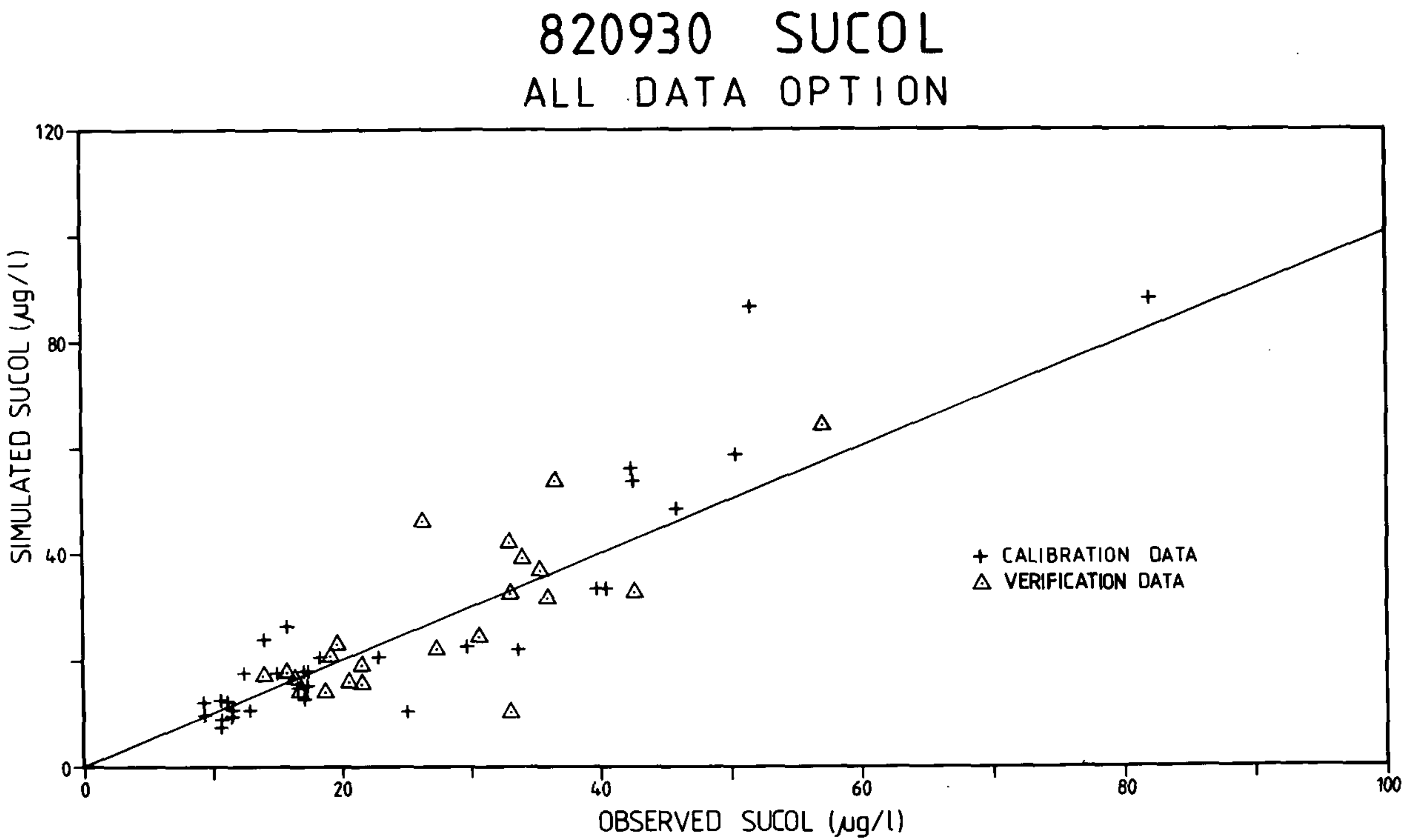


FIGURE 5.1 OBSERVED VERSUS SIMULATED SURFACE CHLOROPHYLL FOR 82·09·30

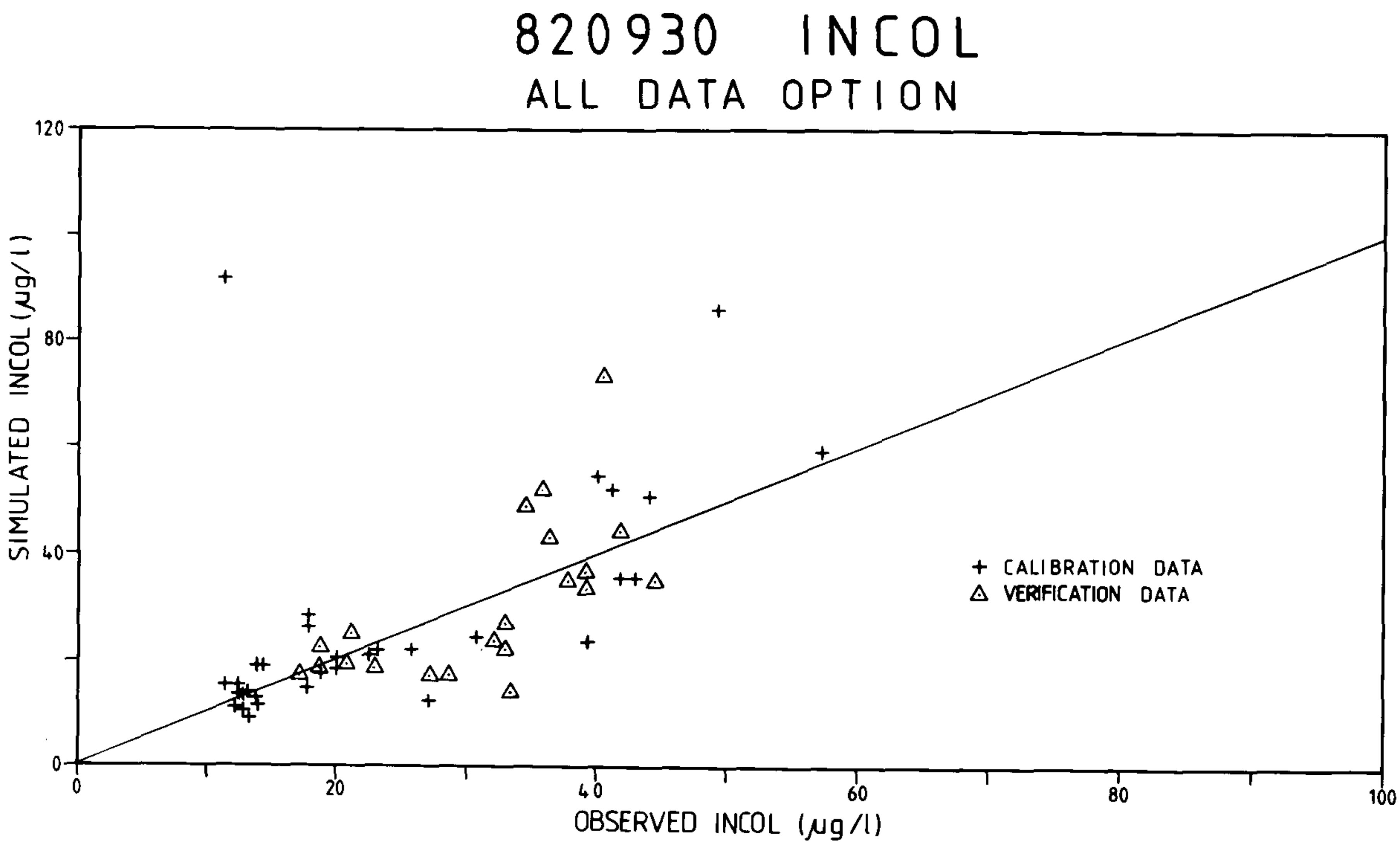


FIGURE 5.2 OBSERVED VERSUS SIMULATED INTEGRATED CHLOROPHYLL FOR 82·09·30

87

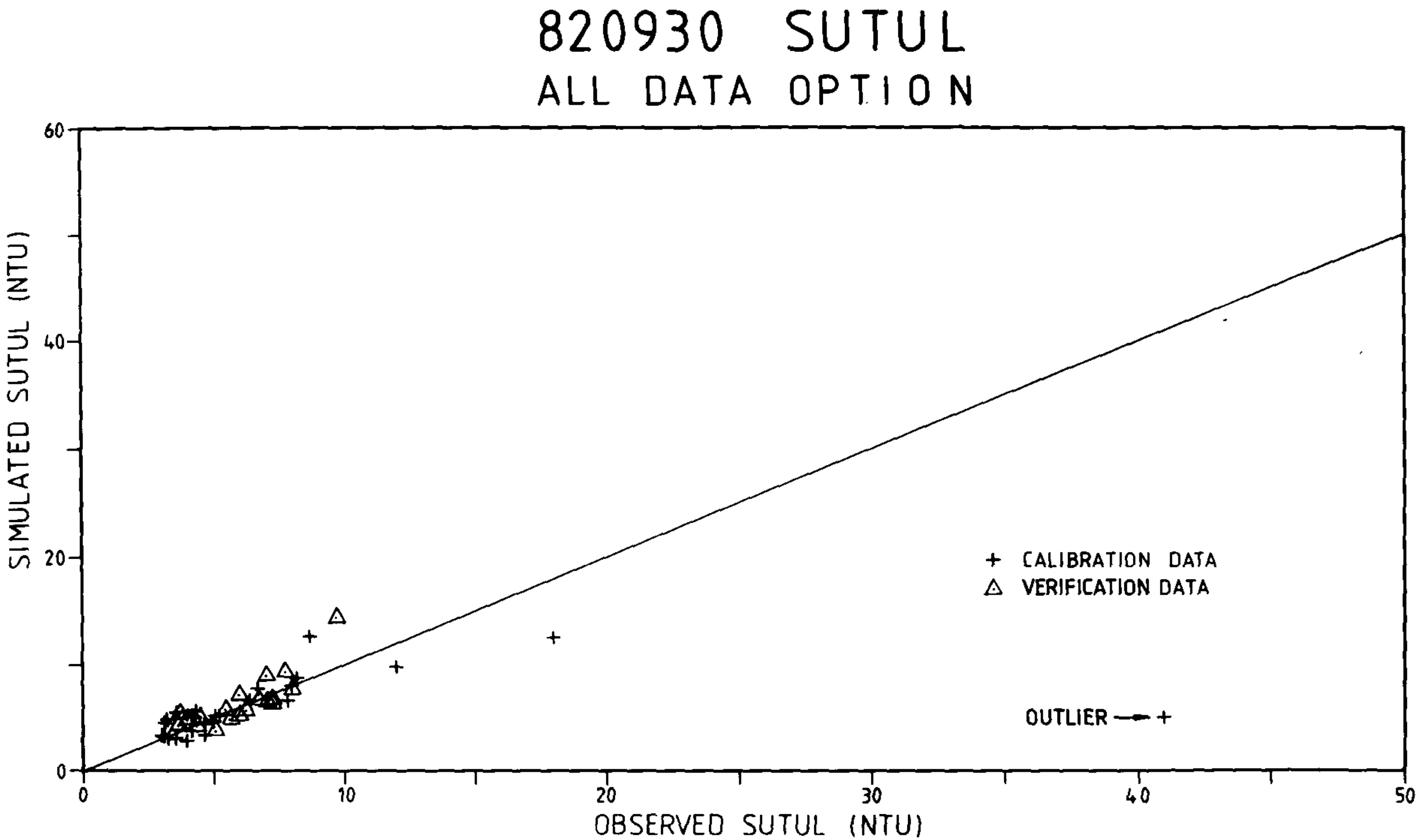


FIGURE 5.3 OBSERVED VERSUS SIMULATED SURFACE TURBIDITY FOR 82·09·30

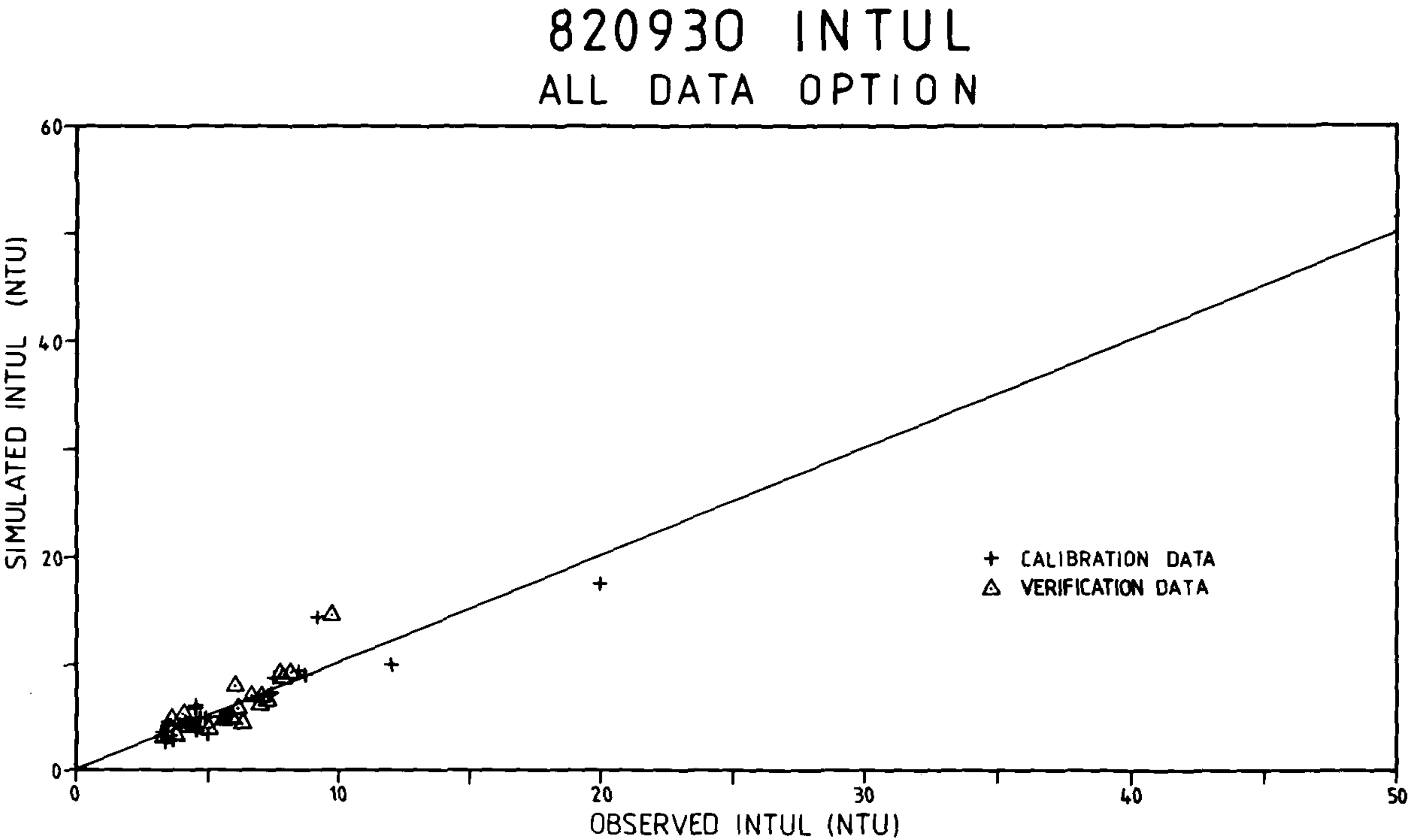


FIGURE 5.4 OBSERVED VERSUS SIMULATED INTEGRATED TURBIDITY FOR 82·09·30

TABLE 5.3: COEFFICIENT OF EFFICIENCY ANALYSIS FOR THE CALIBRATION DATA SET,

'EXCLUDING OUTLIERS' OPTION FOR 82.09.30

	SUCOL	INCOL	SUTUL	INTUL
Number of samples	30	30	30	30
Mean of observed data	22,57	24,54	4,92	5,22
Mean of simulated data	24,40	26,10	5,09	5,30
Std. dev. of observed data	13,84	13,47	2,12	2,16
Std. dev. of simulated data	19,73	17,68	2,46	2,39
Regression intercept	8,21	7,83	1,52	0,88
Regression slope	0,59	0,64	0,67	0,82
Coeff. of determination R^2	0,70	0,71	0,6	0,82
Coeff. of efficiency	0,34	0,47	0,44	0,78

The verification data set (Table 5.4) showed poor chlorophyll a coefficients of determination and efficiency and 8% to 12% relative error in their estimates. The t test results indicated acceptable values at the 5% two tailed level of probability and good predictions for integrated turbidity. Overall the 'Excluding Outliers' Option produced acceptable results for integrated turbidity.

Table 5.5 showing the results of the 'Normalised Data' Option indicated good mean and standard deviation values between observed and simulated data, reasonable coefficients of determination all above 0,72 and an acceptable coefficient of efficiency for integrated turbidity of 0,78. The coefficients of efficiency for the remaining variables were between 0,59 and 0,61. The verification data, Table 5.6, showed small t values and low percent relative errors for all of the variables, low coefficients of determination and efficiency for chlorophyll a and acceptable coefficients for integrated turbidity.

TABLE 5.4: ANALYSIS OF ACCURACY OF THE VERIFICATION SET FOR 82.09.30, 'EXCLUDING OUTLIERS' OPTION

82.09.30		EXCLUDING OUTLIERS								
Water Quality Variables 23 Cases		Mean	Std. Dev.	Diff. Mean	t Test	% Relative Error	Regression Intercept	Regression Slope	R^2	Coeff. of Efficiency
Surface Chlorophyll a µg/ℓ	Observed	27,17	10,59	-3,18	-1,48	12	11,48	0,52	0,4	-0,04
	Simulated	30,35	12,96							
Integrated Chlorophyll a	Observed	30,61	8,43	2,32	1,08	8	18,94	0,41	0,47	-0,56
	Simulated	28,30	14,04							
Surface Turbidity NTU	Observed	5,67	1,69	0,38	1,72	7	2,01	0,69	0,77	0,56
	Simulated	5,29	2,15							
Integrated Turbidity	Observed	5,86	1,59	-0,12	-1,55	2	1,91	0,66	0,78	0,57
	Simulated	5,98	2,13							

TABLE 5.5: COEFFICIENT OF EFFICIENCY ANALYSIS FOR THE CALIBRATION DATA SET,

'NORMALISED DATA' OPTION FOR 82.09.30

	SUCOL	INCOL	SUTUL	INTUL
Number of samples	17	17	17	17
Mean of observed data	22,55	24,49	4,75	5,06
Mean of simulated data	23,93	25,63	4,84	5,14
Std. dev. of observed data	12,79	12,15	1,58	1,60
Std. dev. of simulated data	16,65	15,15	1,86	1,91
Regression intercept	6,34	6,56	1,24	1,07
Regression slope	0,68	0,70	0,72	0,78
Coeff. of determination R^2	0,78	0,76	0,72	0,86
Coeff. of efficiency	0,59	0,61	0,61	0,78

In summary, the three options for the day 82.09.30 indicated that the 'Including All Data' Option obtained both the best (0,85 for integrated turbidity) and the worst (0,11 for surface turbidity) coefficients of efficiency. The best overall coefficients of efficiency were provided by the 'Normalised Data' Option where the coefficients lay between 0,59 and 0,78.

The best overall coefficients of determination were obtained by the 'Normalised Data' Option, 0,72 to 0,86 for the four variables.

The t test showed relative errors lay in the range 0,4% to 9% for the 'Including All Data' Option; between 2% to 12% for the 'Excluding Outliers' Option; and between 2% to 5% for the 'Normalised Data' Option.

Considered synoptically, the 'Normalised Data' Option provided the best calibration for the overpass of 82.09.30. The calibration equations of the 'Normalised Data' model are used in the model to be discussed in Chapter 6.

While there was good simulation between the means of the observed versus simulated data sets, there was not always such a good fit for individual pixels. Tables 5.7 to 5.9 indicate that in some individual cases the values varied fairly significantly. This can be expected due to the fact that the surface reference data for each pixel represents a sample of 1 000 mℓ taken within a pixel and cannot be expected to be exactly equal to the average water quality conditions as seen by the satellite over an 80 m by 80 m area i.e. noisy data.

TABLE 5.6: ANALYSIS OF ACCURACY OF THE VERIFICATION DATA SET FOR 82.09.30, 'NORMALISED DATA' OPTION

82.09.30	NORMALISED DATA									
Water Quality Variables 23 Cases		Mean	Std. Dev.	Diff. Mean	t Test	% Relative Error	Regression Intercept	Regression Slope	R^2	Coeff. of Efficiency
Surface Chlorophyll *a* µg/ℓ	Observed	27,17	10,59	-1,43	-0,63	5	13,44	0,48	0,33	-0,08
	Simulated	28,60	12,94							
Integrated Chlorophyll *a*	Observed	30,61	8,43	-1,21	-0,56	4	18,42	0,38	0,33	-0,55
	Simulated	31,82	12,71							
Surface Turbidity NTU	Observed	5,67	1,69	0,16	0,71	3	0,56	0,93	0,61	0,60
	Simulated	5,51	1,42							
Integrated Turbidity	Observed	6,00	1,52	-0,14	-0,81	2	0,50	0,89	0,73	0,71
	Simulated	6,00	1,52							

TABLE 5.7: OBSERVED VERSUS SIMULATED WATER QUALITY DATA FOR INDIVIDUAL SAMPLING POINTS 82.09.30, 'INCLUDING ALL DATA' OPTION

SAMPLING POINT	SURFACE CHLOROPHYLL a µg/ℓ		INTEGRATED CHLOROPHYLL a µg/ℓ		SURFACE TURBIDITY NTU		INTEGRATED TURBIDITY NTU	
NO.	OBS.	SIM.	OBS.	SIM.	OBS.	SIM.	OBS.	SIM.
33	57,00	64,12	40,50	73,45	9,70	14,49	9,70	14,62
34	36,50	53,58	35,80	52,12	8,00	7,85	7,80	8,75
35	34,00	39,17	41,90	44,36	7,70	9,48	8,10	9,18
36	26,30	46,03	34,50	48,87	7,00	9,08	7,70	9,78
37	35,40	36,81	39,20	36,73	7,20	6,37	7,00	6,32
38	35,90	31,62	39,20	33,65	7,00	6,58	7,20	6,61
39	42,60	32,66	44,50	34,99	7,20	6,79	7,00	7,00
40	33,00	32,66	37,80	34,99	6,70	6,79	6,60	7,00
41	33,00	42,07	36,40	42,95	6,00	7,26	6,00	8,00
42	30,60	24,43	33,00	27,10	6,20	5,81	6,10	5,82
43	33,00	10,33	33,50	14,32	5,90	5,30	6,30	4,49
44	27,30	22,23	32,10	23,88	5,60	4,98	5,60	4,82
45	21,50	19,10	33,00	22,39	5,40	5,75	5,70	5,06
46	20,60	16,03	28,70	17,58	5,00	3,84	5,00	4,06
47	21,50	15,81	27,30	17,54	4,40	4,11	5,00	3,94
48	18,70	14,09	27,30	17,30	4,40	5,20	4,90	4,36
49	16,30	16,71	23,00	18,79	4,10	4,45	4,40	4,25
50	19,10	20,89	18,70	22,54	3,70	4,81	4,30	4,83
51	15,69	17,78	20,10	20,04	3,60	4,48	3,60	4,81
52	13,90	17,30	20,60	19,54	3,70	4,59	4,00	4,50
53	16,70	16,71	18,70	18,79	3,90	4,45	4,40	4,25
54	16,70	14,35	17,50	17,22	4,30	4,58	4,40	4,43
55	19,60	22,91	21,10	25,00	3,70	5,43	4,00	5,18

TABLE 5.8: OBSERVED VERSUS SIMULATED WATER QUALITY DATA FOR INDIVIDUAL SAMPLING POINTS 82.09.30, 'EXCLUDING OUTLIERS' OPTION

SAMPLING POINT	SURFACE CHLOROPHYLL a µg/ℓ		INTEGRATED CHLOROPHYLL a µg/ℓ		SURFACE TURBIDITY NTU		INTEGRATED TURBIDITY NTU	
NO.	OBS.	SIM.	OBS.	SIM.	OBS.	SIM.	OBS.	SIM.
33	57,00	48,64	40,50	66,07	9,70	12,25	9,70	12,65
34	36,50	54,58	35,80	54,08	8,00	8,30	7,80	8,65
35	34,00	41,88	41,90	38,46	7,70	7,16	8,10	8,39
36	26,30	56,23	34,50	42,07	7,00	7,00	7,70	8,49
37	35,40	44,57	39,20	42,76	7,20	6,46	7,00	7,05
38	35,90	35,32	39,20	33,57	7,00	5,96	7,20	6,64
39	42,60	35,40	44,50	32,89	7,20	6,01	7,00	6,73
40	33,00	35,40	37,80	32,89	6,70	6,01	6,60	6,73
41	33,00	45,39	36,40	39,08	6,00	6,71	6,00	7,45
42	30,60	26,24	33,00	25,82	6,20	5,13	6,10	5,69
43	33,00	11,22	33,50	11,09	5,90	3,24	6,30	4,01
44	27,30	20,94	32,10	30,62	5,60	5,69	5,60	5,52
45	21,50	21,53	33,00	23,77	5,40	4,80	5,70	5,41
46	20,60	18,45	28,70	15,92	5,00	3,48	5,00	3,93
47	21,50	21,53	27,30	16,75	4,40	3,39	5,00	4,08
48	18,70	23,33	27,30	13,90	4,40	3,01	4,90	4,27
49	16,30	20,42	23,00	18,62	4,10	3,83	4,40	4,41
50	19,10	34,20	18,70	17,58	3,70	3,34	4,30	4,53
51	15,69	19,63	20,10	16,29	3,70	3,74	3,60	4,31
52	13,90	20,51	20,60	18,24	3,70	3,86	4,00	4,47
53	16,70	20,42	18,70	18,62	3,90	3,83	4,40	4,41
54	16,70	16,18	17,20	14,66	4,30	3,53	4,40	4,15
55	19,60	26,06	21,10	27,04	3,70	5,02	4,00	5,53

TABLE 5.9: OBSERVED VERSUS SIMULATED WATER QUALITY DATA FOR INDIVIDUAL SAMPLING POINTS 82.09.30, 'NORMALISED DATA' OPTION

SAMPLING POINT	SURFACE CHLOROPHYLL a µg/ℓ		INTEGRATED CHLOROPHYLL a µg/ℓ		SURFACE TURBIDITY NTU		INTEGRATED TURBIDITY NTU	
NO.	OBS.	SIM.	OBS.	SIM.	OBS.	SIM.	OBS.	SIM.
33	57,00	38,46	40,50	38,82	9,70	7,80	9,70	9,48
34	36,50	59,16	35,80	57,41	8,00	8,24	7,80	8,26
35	34,00	32,36	41,90	35,65	7,70	6,55	8,10	7,69
36	26,30	43,85	34,50	46,13	7,00	7,41	7,70	8,26
37	35,40	55,72	39,20	63,39	7,20	7,62	7,00	7,71
38	35,90	35,48	39,20	39,17	7,00	6,30	7,20	6,71
39	42,60	32,81	44,50	34,99	7,20	6,14	7,00	6,59
40	33,00	32,81	37,80	34,99	6,70	6,74	6,60	6,49
41	33,00	40,93	36,40	40,09	6,00	6,89	6,00	7,13
42	30,60	25,18	33,00	28,05	6,20	5,27	6,10	5,70
43	33,00	7,48	33,50	10,00	5,90	3,01	6,30	3,91
44	27,30	30,83	32,10	37,07	5,60	5,57	5,60	5,74
45	21,50	22,86	33,00	29,79	5,40	5,00	5,70	5,74
46	20,60	17,62	28,70	19,23	5,00	4,13	5,00	4,23
47	21,50	21,09	27,30	25,64	4,40	4,50	5,00	4,76
48	18,70	16,37	27,30	22,34	4,40	4,20	4,90	5,08
49	16,30	20,94	23,00	25,41	4,10	4,57	4,40	4,90
50	19,10	25,00	18,70	27,99	3,70	5,02	4,30	5,33
51	15,69	15,85	20,10	16,60	3,60	4,07	3,60	4,34
52	13,90	19,36	20,60	22,70	3,70	4,46	4,00	4,81
53	16,70	20,94	18,70	25,41	3,90	4,57	4,40	4,90
54	16,70	13,34	17,20	15,74	4,30	3,78	4,40	4,27
55	19,60	29,38	21,10	35,32	3,70	5,55	4,00	5,93

5.5.2 Overpass of 82.09.13

The results of the overpass for 82.09.13 are presented in Tables 5.10 to 5.15. All three options, 'Including All Data'; 'Excluding Outliers' and 'Normalised Data' gave negative coefficients of efficiency for integrated chlorophyll a. The coefficient of determination for integrated chlorophyll a was poor for the 'Including All Data' Option; whereas the coefficients of determination for surface and integrated chlorophyll a were poor for the 'Excluding Outliers' and 'Normalised Data' Options.

Despite the inability to simulate individual data points as shown by the unacceptable coefficients of efficiency and determination, the t tests on the verification data sets showed that the mean values were acceptably simulated with relative errors of less than 10%, except for integrated chlorophyll a where relative errors of between 14% and 16% for the three options were found.

TABLE 5.10: COEFFICIENT OF EFFICIENCY ANALYSIS FOR THE CALIBRATION DATA SET,

'INCLUDING ALL DATA' OPTION FOR 82.09.13

	SUCOL	INCOL	SUTUL	INTUL
Number of samples	31	31	31	31
Mean of observed data	20,65	20,05	5,33	5,85
Mean of simulated data	21,23	24,80	5,38	6,08
Std. dev. of observed data	9,10	9,43	2,48	3,08
Std. dev. of simulated data	19,01	39,38	2,54	3,88
Regression intercept	12,19	16,04	0,38	1,25
Regression slope	0,40	0,16	0,92	0,76
Coeff. of determination R^2	0,69	0,46	0,89	0,91
Coeff. of efficiency	-0,89	-12,07	0,88	0,81

TABLE 5.11: ANALYSIS OF ACCURACY OF THE VERIFICATION DATA SET FOR 82.09.13, 'INCLUDING ALL DATA' OPTION

82.09.13		INCLUDING ALL DATA								
Water Quality Variables 22 Cases		Mean	Std. Dev.	Diff. Mean	t Test	% Relative Error	Regression Intercept	Regression Slope	R^2	Coeff. of Efficiency
Surface Chlorophyll a µg/ℓ	Observed	20,80	4,58	1,1	2,00	5	13,69	0,38	0,26	-0,67
	Simulated	19,69	4,09							
Integrated Chlorophyll a	Observed	21,15	3,94	3,43	3,11	16	14,37	0,38	0,45	-1,52
	Simulated	17,72	6,92							
Surface Turbidity NTU	Observed	5,10	2,06	-0,09	-0,24	2	1,24	0,74	0,19	0,16
	Simulated	5,19	1,21							
Integrated Turbidity	Observed	5,44	1,38	-0,16	-0,57	3	2,44	0,53	0,32	0,06
	Simulated	5,59	1,45							

TABLE 5.12: COEFFICIENT OF EFFICIENCY ANALYSIS FOR THE CALIBRATION DATA SET,

'EXCLUDING OUTLIERS' OPTION FOR 82.09.13

	SUCOL	INCOL	SUTUL	INTUL
Number of samples	30	30	30	30
Mean of observed data	19,49	19,04	4,94	5,35
Mean of simulated data	19,81	24,22	5,01	5,52
Std. dev. of observed data	6,56	7,69	1,24	1,28
Std. dev. of simulated data	7,37	19,87	1,49	1,90
Regression intercept	7,54	12,39	1,59	2,45
Regression slope	0,60	0,27	0,67	0,53
Coeff. of determination R^2	0,46	0,50	0,64	0,61
Coeff. of efficiency	0,26	-3,48	0,48	0,1

*

TABLE 5.14: COEFFICIENT OF EFFICIENCY ANALYSIS FOR THE CALIBRATION DATA SET,

'NORMALISED DATA' OPTION FOR 82.09.13

	SUCOL	INCOL	SUTUL	INTUL
Number of samples	25	25	25	25
Mean of observed data	20,00	19,41	5,08	5,48
Mean of simulated data	19,84	24,63	5,13	5,62
Std. dev. of observed data	6,96	8,27	1,31	1,36
Std. dev. of simulated data	6,37	20,43	1,47	1,88
Regression intercept	5,50	11,99	1,40	2,23
Regression slope	0,73	0,30	0,72	0,58
Coeff. of determination R^2	0,45	0,55	0,64	0,63
Coeff. of efficiency	0,39	-2,84	0,54	0,29

* Table 5.13 is on the next page

TABLE 5.13: ANALYSIS OF ACCURACY OF THE VERIFICATION DATA SET FOR 82.09.13, 'EXCLUDING OUTLIERS' OPTION

82.09.13		EXCLUDING OUTLIERS								
Water Quality Variables 22 Cases		Mean	Std. Dev.	Diff. Mean	t Test	% Relative Error	Regression Intercept	Regression Slope	R^2	Coeff. of Efficiency
Surface Chlorophyll a µg/ℓ	Observed	20,80	4,58	0,16	0,15	0,8	11,91	0,43	0,35	- 0,27
	Simulated	20,63	6,33							
Integrated Chlorophyll a	Observed	21,15	3,94	-3,17	-0,06	15	17,78	0,14	0,31	-12.45
	Simulated	24,32	3,94							
Surface Turbidity NTU	Observed	5,10	2,06	-0,19	-0,47	4	1,16	0,75	0,24	0,20
	Simulated	5,28	1,35							
Integrated Turbidity	Observed	5,44	1,38	-0,32	-0,12	6	2,39	0,53	0,40	0,03
	Simulated	5,75	1,65							

TABLE 5.15: ANALYSIS OF ACCURACY OF THE VERIFICATION DATA SET FOR 82.09.13,

'NORMALISED DATA' OPTION

82.09.13	NORMALISED DATA									
Water Quality Variables 22 Cases		Mean	Std. Dev.	Diff. Mean	t Test	% Relative Error	Regression Intercept	Regression Slope	R^2	Coeff. of Efficiency
Surface Chlorophyll a µg/ℓ	Observed	20,80	4,58	0,87	0,87	4	11,18	0,48	0,34	- 0,09
	Simulated	19,93	5,52							
Integrated Chlorophyll a	Observed	21,15	3,94	-2,99	-0,84	14	18,35	0,12	0,30	-17,38
	Simulated	24,14	18,40							
Surface Turbidity NTU	Observed	5,10	2,06	-0,10	-0,27	2	1,12	0,76	0,24	0,22
	Simulated	5,20	1,33							
Integrated Turbidity	Observed	5,44	1,38	-0,23	-0,78	4	2,57	0,51	0,39	- 0,01
	Simulated	5,66	1,70							

5.5.3 Overpass of 82.11.16

The results for the 82.11.16 overpass are presented in Tables 5.16 to 5.21. For all three options of the calibration data, the coefficients of determination and efficiency were high for surface and integrated chlorophyll a, between 0,79 and 0,96. This was not the case, however, for turbidity and negative or zero coefficients of efficiency were obtained for integrated turbidity in all three options, indicating a problem in the calibration of turbidity for this overpass.

The t test on the verification data set showed poor accuracy for the modelling of all variables except surface and integrated turbidity using the 'Including All Data' Option. The error of as much as 26% for the 'Excluding Outlier' approach for chlorophyll a occurred despite the high coefficients of efficiency found for the calibration data set as discussed above.

TABLE 5.16: COEFFICIENT OF EFFICIENCY ANALYSIS FOR THE CALIBRATION DATA SET,

'INCLUDING ALL DATA' OPTION FOR 82.11.16

	SUCOL	INCOL	SUTUL	INTUL
Number of samples	28	28	28	28
Mean of observed data	39,47	37,20	7,23	7,83
Mean of simulated data	41,60	37,15	7,25	9,08
Std. dev. of observed data	68,26	60,11	5,40	5,98
Std. dev. of simulated data	73,98	57,90	5,87	11,58
Regression intercept	1,90	- 0,37	1,64	3,82
Regression slope	0,90	1,01	0,77	0,44
Coeff. of determination R^2	0,96	0,95	0,71	0,73
Coeff. of efficiency	0,95	0,95	0,64	-0,49

TABLE 5.17: ANALYSIS OF ACCURACY OF THE VERIFICATION DATA SET FOR 82.11.16, 'INCLUDING ALL DATA' OPTION

82.11.16	INCLUDING ALL DATA									
Water Quality Variables 23 Cases		Mean	Std. Dev.	Diff. Mean	t Test	% Relative Error	Regression Intercept	Regression Slope	R^2	Coeff. of Efficiency
Surface Chlorophyll a µg/ℓ	Observed	34,95	24,56	8,02	2,28	23	3,02	1,19	0,54	0,42
	Simulated	26,93	15,24							
Integrated Chlorophyll a	Observed	34,20	21,89	8,35	2,71	24	1,69	1,26	0,57	0,39
	Simulated	25,85	13,12							
Surface Turbidity NTU	Observed	6,58	2,75	0,30	0,72	5	0,60	0,95	0,48	0,46
	Simulated	6,28	2,00							
Integrated Turbidity	Observed	7,07	3,05	0,16	0,32	2	2,17	0,71	0,53	0,43
	Simulated	6,91	3,12							

TABLE 5.18: COEFFICIENT OF EFFICIENCY ANALYSIS FOR THE CALIBRATION DATA SET,

'EXCLUDING OUTLIERS' OPTION FOR 82.11.16

	SUCOL	INCOL	SUTUL	INTUL
Number of samples	24	24	24	24
Mean of observed data	24,35	23,88	6,28	6,76
Mean of simulated data	24,94	23,82	5,98	7,11
Std. dev. of observed data	12,24	11,89	3,26	3,57
Std. dev. of simulated data	12,75	10,87	1,79	3,96
Regression intercept	2,75	0,76	0,91	3,40
Regression slope	0,87	0,97	0,90	0,47
Coeff. of determination R^2	0,81	0,79	0,24	0,27
Coeff. of efficiency	0,79	0,79	0,23	-0,08

*

TABLE 5.20: COEFFICIENT OF EFFICIENCY ANALYSIS FOR THE CALIBRATION DATA SET,

'NORMALISED DATA' OPTION FOR 82.11.16

	SUCOL	INCOL	SUTUL	INTUL
Number of samples	24	24	24	24
Mean of observed data	24,35	23,88	6,28	6,76
Mean of simulated data	24,77	23,81	5,92	6,96
Std. dev. of observed data	12,24	11,89	3,26	3,57
Std. dev. of simulated data	12,70	11,01	1,74	3,73
Regression intercept	2,71	0,92	0,82	3,27
Regression slope	0,87	0,96	0,92	0,50
Coeff. of determination R^2	0,82	0,80	0,24	0,27
Coeff. of efficiency	0,80	0,80	0,23	0,00

Table 5.19 is on the next page

TABLE 5.19: ANALYSIS OF ACCURACY OF THE VERIFICATION DATA SET FOR 82.11.16, 'EXCLUDING OUTLIERS' OPTION

82.11.16	EXCLUDING OUTLIERS									
Water Quality Variables 23 Cases		Mean	Std. Dev.	Diff. Mean	t Test	% Relative Error	Regression Intercept	Regression Slope	R^2	Coeff. of Efficiency
Surface Chlorophyll a µg/ℓ	Observed	34,95	24,56	7,75	2,31	22	-0,33	1,30	0,6	0,46
	Simulated	27,20	14,66							
Integrated Chlorophyll a	Observed	34,20	21,89	8,94	2,75	26	-4,88	1,55	0,56	0,32
	Simulated	25,26	10,63							
Surface Turbidity NTU	Observed	6,58	2,75	0,31	0,87	5	0,36	0,99	0,61	0,60
	Simulated	6,27	2,17							
Integrated Turbidity	Observed	7,07	3,05	-1,35	-1,42	19	3,83	0,38	0,78	-1,42
	Simulated	8,42	7,00							

TABLE 5.21: ANALYSIS OF ACCURACY OF THE VERIFICATION DATA SET FOR 82.11.16, 'NORMALISED DATA' OPTION

82.11.16	NORMALISED DATA									
Water Quality Variables 23 Cases		Mean	Std. Dev.	Diff. Mean	t Test	% Relative Error	Regression Intercept	Regression Slope	R^2	Coeff. of Efficiency
Surface Chlorophyll a µg/ℓ	Observed	34,95	24,56	7,84	2,38	22	0,69	1,26	0,61	0,48
	Simulated	27,11	15,19							
Integrated Chlorophyll a	Observed	34,20	21,89	8,78	2,79	26	-3,38	1,48	0,59	0,36
	Simulated	25,42	11,35							
Surface Turbidity NTU	Observed	6,58	2,75	0,39	1,08	6	0,28	1,02	0,61	0,58
	Simulated	6,19	2,10							
Integrated Turbidity	Observed	7,07	3,05	-1,17	-1,37	17	3,75	0,41	0,78	-0,93
	Simulated	8,24	6,64							

5.6 SUMMARY

The simulations of point source concentrations of chlorophyll a and turbidity were made possible by a model, CALMCAT, obtained using Canonical Correlation Coefficients and Linear Regression Analysis. Point source data not previously incorporated in the establishment of the model CALMCAT were used to test the accuracy of the simulated results. Calibration equations provided relatively accurate simulations with percent relative errors ranging from 0,4% to 26% for three of the days tested. Good simulations were found between the mean values of observed versus simulated data sets, however, the simulation for individual pixels demonstrated considerable noise in the system.

The overpass of the 82.11.16 did not provide an acceptable calibration for all four variables. This may be due to a number of factors:

(1) The difficulty of obtaining a representative data set considering the noise level in the data.
(2) The assumption of a linear relationship between the four water quality variables and the four reflectance bands.

If the latter assumption is not satisfied then the Canonical Correlation Analysis would be unable to provide an adequate calibration relationship even with a fully representative data set.

CHAPTER 6

THE MODEL CALMCAT*

6.1 INTRODUCTION

The model CALMCAT for the simulation of chlorophyll and turbidity from Landsat reflectance data is essentially the procedure consisting of firstly, the selection of a representative, surface reference data set, secondly, the establishment of simulative equations through the use of the Canonical Correlation Analysis and Linear Regression techniques and thirdly, the calculation of surface and integrated chlorophyll and turbidity values for each pixel with the subroutine DAMLOD. The menu for the model CALMCAT is summarized in Appendix S.

The model CALMCAT has three variants. The first is the model which includes all data in the calibration procedure. The second variant is the model which excludes outliers in the calibration set. The third variant termed the 'Normalised Data' model is where the calibration data set has been normalised prior to the application of the Canonical procedure. The concentration and the overall percentage distribution of surface and integrated chlorophyll a and surface and integrated turbidity was determined for each of the three options. Walmsley and Butty (1979) separated chlorophyll a data into specific ranges and ascribed nuisance values to each range (μg/ℓ), namely

" 0 - 10	No problem encountered
10 - 20	Algal scums evident
20 - 30	Nuisance conditions encountered
>30	Severe nuisance conditions encountered"

(Walmsley, 1984).

These same value ranges were used for fine class interval classifications, in the following analysis.

The model for simulating water quality conditions over the entire impoundment using satellite reflectance data is given in Appendix R.

The results of the model for each option for the overflight pass on 82.09.30 are presented on Tables 6.1 to 6.3.

6.2 RESULTS

Table 6.1 presents mean, maximum and minimum values and distributional estimates of concentrations, as simulated by CALMCAT for the 'Normalised Data' calibration equation.

---

* CALMCAT - Canonical Analysis Landsat Model of Chlorophyll a and Turbidity.

TABLE 6.1: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF THE WATER QUALITY VARIABLES IN ROODEPLAAT DAM USING SATELLITE REFLECTANCE DATA,

NORMALISED DATA MODEL for 82.09.30

Water/Land limit = 25

Numbers of pixels in impoundment = 849

SURFACE CHLOROPHYLL RESULTS
µg/ℓ

MEAN = 27,47
MAX = 430,09
MIN = 0,81

CLASS RANGE	PERCENTAGE AREA
0,00 - 9,99	21,44
10,00 - 19,99	36,16
20,00 - 29,99	18,96
30,00 - 39,99	7,66
40,00 - 49,99	2,83
50,00 - 59,99	2,59
60,00 - 69,99	1,88
70,00 - 79,99	1,53
80,00 - 89,99	2,12
90,00 - 99,99	1,06
100,00 - 109,99	1,41
110,00 - 439,99	2,37

SURFACE TURBIDITY RESULTS
NTU

MEAN = 4,87
MAX = 22,52
MIN = 1,02

CLASS RANGE	PERCENTAGE AREA
0,00 - 1,99	6,60
2,00 - 3,99	38,28
4,00 - 5,99	35,10
6,00 - 7,99	7,89
8,00 - 9,99	6,83
10,00 - 11,99	3,06
12,00 - 13,99	1,41
14,00 - 15,99	0,35
20,00 - 21,99	0,24
22,00 - 23,99	0,24

INTEGRATED CHLOROPHYLL RESULTS
µg/ℓ

MEAN = 28,41
MAX = 344,05
MIN = 1,42

CLASS RANGE	PERCENTAGE AREA
0,00 - 9,99	17,67
10,00 - 19,99	31,57
20,00 - 29,99	22,50
30,00 - 39,99	10,84
40,00 - 49,99	4,95
50,00 - 59,99	2,83
60,00 - 69,99	1,88
70,00 - 79,99	1,53
80,00 - 89,99	1,77
90,00 - 99,99	1,53
100,00 - 109,99	0,47
110,00 - 349,99	2,49

INTEGRATED TURBIDITY RESULTS
NTU

MEAN = 5,14
MAX = 21,06
MIN = 1,84

CLASS RANGE	PERCENTAGE AREA
0,00 - 1,99	0,47
2,00 - 3,99	35,92
4,00 - 5,99	39,34
6,00 - 7,99	12,84
8,00 - 9,99	6,60
10,00 - 11,99	2,71
12,00 - 13,99	0,94
14,00 - 15,99	0,94
20,00 - 21,99	0,24

Surface chlorophyll a in the range between 1 and 29,99 µg/ℓ was found to cover 77% of the impoundment. An area of 20% was shown as having between 30 to 99,99 µg/ℓ and 3% of the area had over 100 µg/ℓ. The maximum simulated value of surface chlorophyll was 430 µg/ℓ but it is highly likely that the high values over 100 µg/ℓ are due to mixels, mixed land and water pixels. An area of 72% of the impoundment was simulated as containing between 1 to 30 µg/ℓ in the water column (integrated chlorophyll a).

Turbidity values between 1 to 7,99 NTU were found to cover 88% of the impoundment whereas integrated turbidity for the same value range was found to cover 89% of the area.

Tables 6.2 and 6.3 are similar to Table 6.1. The maximum values simulated by each option differ, but the distribution trend remains fairly constant as shown in Table 6.4. For example, when comparing the three options, integrated chlorophyll a between the range of 0-30 µg/ℓ varies in percentage area covered between 77% and 72% - a difference of 5%. The higher range of concentrations between 30 and 100 µg/ℓ varied by 5%. These results illustrate that, despite differences, the three options, nevertheless indicate the same distribution pattern. This is best illustrated by comparing the simulated distributions for surface turbidity using the 'Normalised' model and the 'Including All Data' model. This comparison is shown in Table 6.5. In Chapter 5 where the adequacy of the calibration was discussed, the 'Including All Data' model for 82.09.30 gave low coefficients of efficiency and determination, 0,11 and 0,13 respectively (refer to Table 5.1), whereas the 'Normalised Data' model gave coefficients of efficiency and determination of 0,61 and 0,72 respectively (refer to Table 5.5). An inspection of Table 6.5 shows that the distribution pattern for surface turbidity obtained by the 'Including All Data' model provides much the same pattern as that obtained for the 'Normalised' model. The percentage area covered by surface turbidity in the 0 to 1,99 NTU range, for the 'Normalised Data' model is higher (6,60%) than the 'Including All Data' model for the same range; whereas the 'Including All Data' model indicates slightly higher percentage areas covered in the 2 to 7,99 NTU range and in the 26 to 27,99 NTU range.

Thus it appears that the 'Including All Data' model, despite being considered to be largely unacceptable in Chapter 5, nevertheless provides an adequate synoptic quantification of surface turbidity as indicated by the comparison in Table 6.5, where the researcher is interested in a class interval classification of chlorophyll and turbidity rather than in quantitative exact accuracy.

TABLE 6.2: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF THE WATER QUALITY VARIABLES IN ROODEPLAAT DAM USING SATELLITE REFLECTANCE DATA,

EXCLUDING OUTLIERS MODEL for 82.09.30

Water/Land limit = 25

Number of pixels in impoundment = 849

SURFACE CHLOROPHYLL RESULTS
µg/ℓ

MEAN = 25,20
MAX = 273,15
MIN = 1,83

CLASS RANGE	PERCENTAGE AREA
0,00 - 9,99	26,03
10,00 - 19,99	32,51
20,00 - 29,99	18,61
30,00 - 39,99	5,89
40,00 - 49,99	5,77
50,00 - 59,99	4,00
60,00 - 69,99	2,00
70,00 - 79,99	1,06
80,00 - 89,99	0,82
90,00 - 99,99	0,59
100,00 - 109,99	0,71
110,00 - 279,99	2,03

SURFACE TURBIDITY RESULTS
NTU

MEAN = 5,81
MAX = 51,08
MIN = 0,78

CLASS RANGE	PERCENTAGE AREA
0,00 - 1,99	7,30
2,00 - 3,99	42,87
4,00 - 5,99	20,85
6,00 - 7,99	12,37
8,00 - 9,99	5,42
10,00 - 11,99	2,83
12,00 - 13,99	2,12
14,00 - 15,99	2,00
16,00 - 17,99	0,24
20,00 - 21,99	0,94
24,00 - 51,99	3,09

INTEGRATED CHLOROPHYLL RESULTS
µg/ℓ

MEAN = 31,23
MAX = 504,87
MIN = 1,39

CLASS RANGE	PERCENTAGE AREA
0,00 - 9,99	20,26
10,00 - 19,99	34,04
20,00 - 29,99	20,02
30,00 - 39,99	7,77
40,00 - 49,99	4,00
50,00 - 59,99	3,30
60,00 - 69,99	2,47
70,00 - 79,99	1,18
80,00 - 89,99	1,06
90,00 - 99,99	1,18
100,00 - 109,99	0,35
110,00 - 509,99	4,39

INTEGRATED TURBIDITY RESULTS
NTU

MEAN = 5,47
MAX = 34,22
MIN = 1,63

CLASS RANGE	PERCENTAGE AREA
0,00 - 1,99	0,47
2,00 - 3,99	42,64
4,00 - 5,99	30,39
6,00 - 7,99	12,13
8,00 - 9,99	5,30
10,00 - 11,99	2,83
12,00 - 13,99	2,36
14,00 - 15,99	1,30
16,00 - 17,99	1,30
18,00 - 19,99	0,24
20,00 - 21,99	0,35
24,00 - 35,99	0,72

TABLE 6.3: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF THE WATER QUALITY VARIABLES IN ROODEPLAAT DAM USING SATELLITE REFLECTANCE DATA,

INCLUDING ALL DATA MODEL for 82.09.30

Water/Land limit = 25

Number of pixels in impoundment = 849

SURFACE CHLOROPHYLL RESULTS
µg/ℓ

MEAN = 26,98
MAX = 384,01
MIN = 1,94

CLASS RANGE	PERCENTAGE AREA
0,00 - 9,99	19,79
10,00 - 19,99	43,82
20,00 - 29,99	15,55
30,00 - 39,99	5,06
40,00 - 49,99	3,65
50,00 - 59,99	1,53
60,00 - 69,99	2,47
70,00 - 79,99	1,77
80,00 - 89,99	1,53
90,00 - 99,99	1,41
100,00 - 109,99	0,35
110,00 - 389,99	3,08

SURFACE TURBIDITY RESULTS
NTU

MEAN = 5,39
MAX = 27,35
MIN = 2,05

CLASS RANGE	PERCENTAGE AREA
0,00 - 1,99	0, 0
2,00 - 3,99	39,22
4,00 - 5,99	37,57
6,00 - 7,99	10,25
8,00 - 9,99	4,71
10,00 - 11,99	4,24
12,00 - 13,99	1,41
14,00 - 15,99	1,30
16,00 - 17,99	0,47
18,00 - 19,99	0,12
20,00 - 21,99	0,24
24,00 - 27,99	0,48

INTEGRATED CHLOROPHYLL RESULTS
µg/ℓ

MEAN = 27,94
MAX = 313,94
MIN = 3,91

CLASS RANGE	PERCENTAGE AREA
0,00 - 9,99	14,72
10,00 - 19,99	41,22
20,00 - 29,99	20,97
30,00 - 39,99	6,60
40,00 - 49,99	3,89
50,00 - 59,99	3,42
60,00 - 69,99	1,30
70,00 - 79,99	1,88
80,00 - 89,99	1,41
90,00 - 99,99	1,06
100,00 - 109,99	1,18
110,00 - 319,99	2,37

INTEGRATED TURBIDITY RESULTS
NTU

MEAN = 5,56
MAX = 34,66
MIN = 2,01

CLASS RANGE	PERCENTAGE AREA
0,00 - 1,99	0,00
2,00 - 3,99	41,46
4,00 - 5,99	32,74
6,00 - 7,99	10,84
8,00 - 9,99	4,00
10,00 - 11,99	3,89
12,00 - 13,99	2,12
14,00 - 15,99	2,94
16,00 - 17,99	0,35
18,00 - 19,99	0,71
20,00 - 21,99	0,24
32,00 - 35,99	0,71

TABLE 6.4: DISTRIBUTIONAL TREND OF PERCENTAGE AREA COVERED FOR EACH OF THE THREE OPTIONS FOR 82.09.30

VARIABLES	SUCOL µg/ℓ		INCOL µg/ℓ		SUTUL NTU		INTUL NTU	
RANGE	0-30	30-100	0-30	30-100	0-8	8-20	0-8	8-20
OPTION								
NORMALISED DATA	77%	20%	72%	25%	88%	12%	89%	11%
EXCLUDING OUTLIERS	77%	20%	74%	21%	83%	13%	86%	13%
INCLUDING ALL DATA	79%	17%	77%	20%	87%	12%	85%	14%

TABLE 6.5: SIMULATED CONCENTRATIONS AND DISTRIBUTION OF SURFACE TURBIDITY FOR THE 'NORMALISED DATA' AND 'INCLUDING ALL DATA' OPTIONS FOR THE OVERPASS OF 82.09.30

CLASS RANGE NTU	PERCENTAGE AREA FOR THE 'NORMALISED DATA' OPTION	PERCENTAGE AREA FOR THE 'INCLUDING ALL DATA' OPTION	DIFFERENCE
0,00 - 1,99	6,60	0,00	6,60
2,00 - 2,99	38,28	39,22	-0,94
4,00 - 5,99	35,10	37,57	-2,47
6,00 - 7,99	7,89	10,25	-2,36
8,00 - 9,99	6,83	4,71	2,12
10,00 - 11,99	3,06	4,24	-1,18
12,00 - 13,99	1,41	1,41	0,00
14,00 - 15,99	0,35	1,30	-0,95
16,00 - 17,99	0,00	0,47	-0,47
18,00 - 19,99	0,00	0,12	-0,12
20,00 - 21,99	0,24	0,24	0,00
22,00 - 23,99	0,24	0,00	0,24
24,00 - 27,99	0,00	0,48	-0,48
MEAN	4,87	5,39	
MINIMUM	1,02	2,05	
MAXIMUM	22,52	27,35	
COEFF. OF EFFICIENCY	0,61	0,11	
COEFF. OF DETERMINATION	0,72	0,13	

6.3 A SYNOPTIC VIEW

The value of satellite reflectance data lies mainly in its synoptic view. Concentrations and areal estimates can be enhanced by the spatial characteristics of satellite data to provide researchers with data concerning the location, concentration and areal coverage of a specific water quality condition. Mapping the distribution of water quality conditions using values determined by CALMCAT was a necessary step.

Maps of simulated water quality conditions were produced using the simulated data and P.I.P.S, a portable image processing suite for remote sensing and geographic information systems, obtained from the Department of Surveying and Mapping, University of Natal, (O'Donoghue et al, 1983). Plates 6.1 to 6.4 illustrate the results showing concentration contours of chlorophyll a and turbidity as determined by the Canonical Correlations Analysis and the satellite reflectance data for the 'Normalised Data' calibration option. Surface chlorophyll a results shown on Plate 6.1 indicate that chlorophyll a concentrations are highest along the western arm where the Pienaars River and Hartbeesspruit enter the impoundment. It is also evident that concentrations of chlorophyll a are found along the edge of the impoundment, where the greatest productivity can be expected, and along the northern shoreline perhaps due to wind action. The distribution of turbidity (Plates 6.3 and 6.4) appears to be similar to that of the chlorophyll a distribution illustrating the interrelationship between chlorophyll a and turbidity. With greater knowledge of prevailing conditions it may be possible to infer current circulation and wind movement.

6.4 SUMMARY

It is evident that satellite reflectance data provides spatial and quantifiable information unlike any other data source yet available in the field of water resources. The extrapolation of point source data to that of synoptic data is a step forward for limnological and hydrological research. Quantitative, reasonably accurate information showing the position and concentration of specific water quality conditions may, for example, assist with a number of practical limnological problems. Firstly, the suitability of the positioning of existing sampling sites can be evaluated, and planning the distribution of sampling sites in an impoundment so as to be representative of prevailing conditions can be assisted. Secondly, the synoptic information on chlorophyll a and turbidity may be used to assist in the siting of withdrawal points for water abstraction, as well as in the siting of recreational facilities. Thirdly, the ability of satellite remote sensing to detect sources of nutrient pollution leading to localised algal blooms can assist in studying the extent to which such pollution is dispersed together with circulation patterns in the water body. This aspect is of relevance to the siting of sewage outfalls. Fourthly, the synoptic chlorophyll a and turbidity data provided by CALMCAT may assist limnologists in studying the relationship between water quality conditions and nutrient inputs, in verifying and calibrating water quality models, and in evaluating the validity of assumptions.

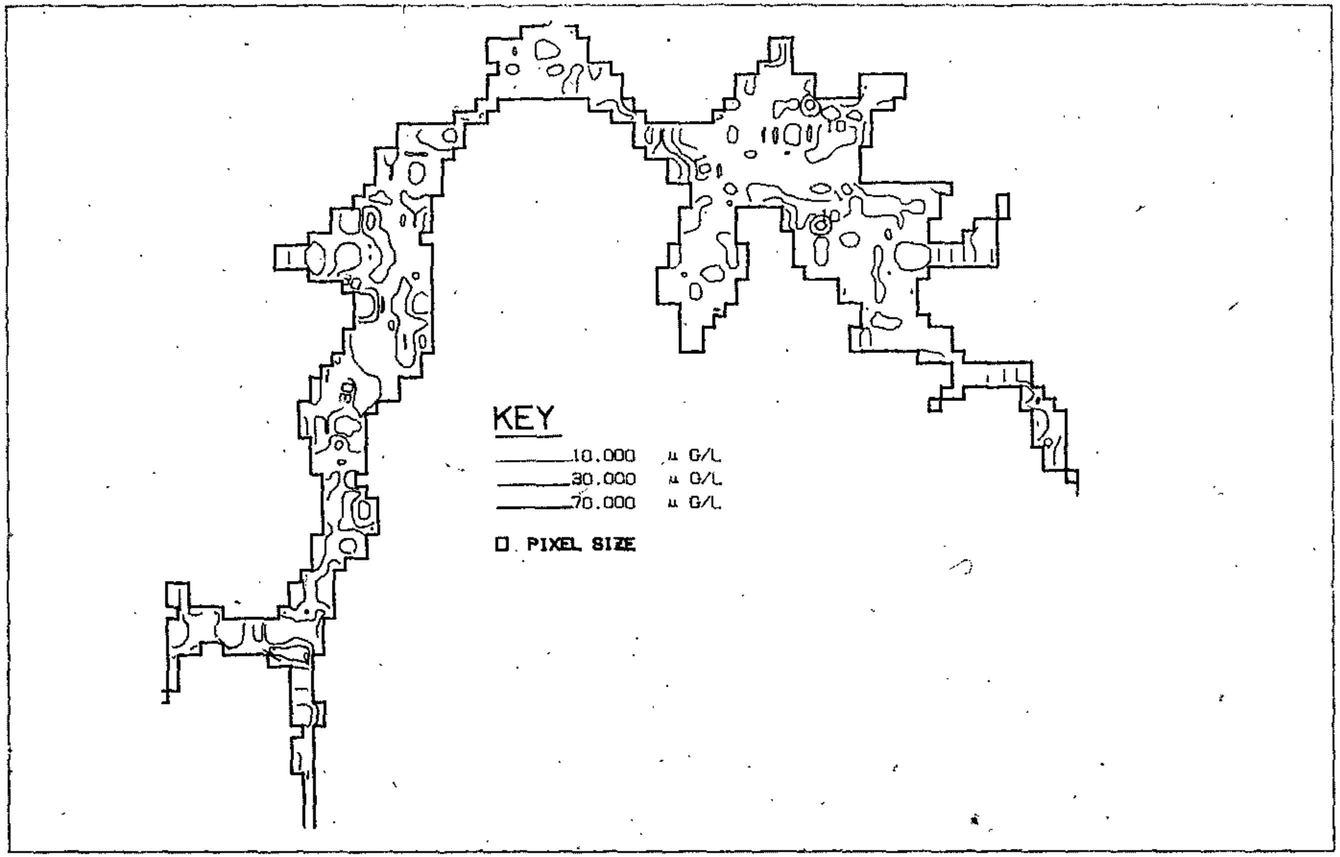


PLATE 6.1: SIMULATED SURFACE CHLOROPHYLL FOR 82.09.30

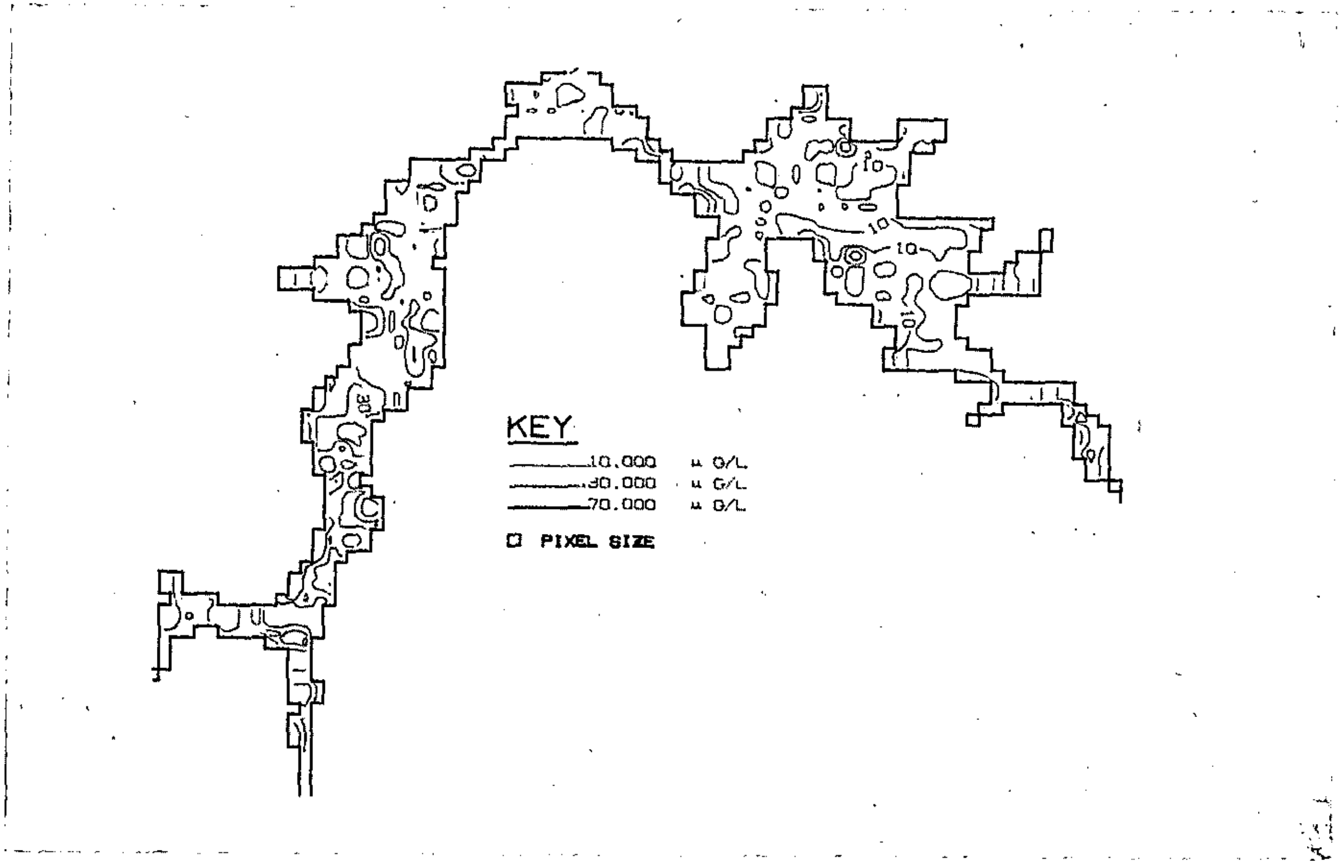


PLATE 6.2: SIMULATED INTEGRATED CHLOROPHYLL FOR 82.09.30

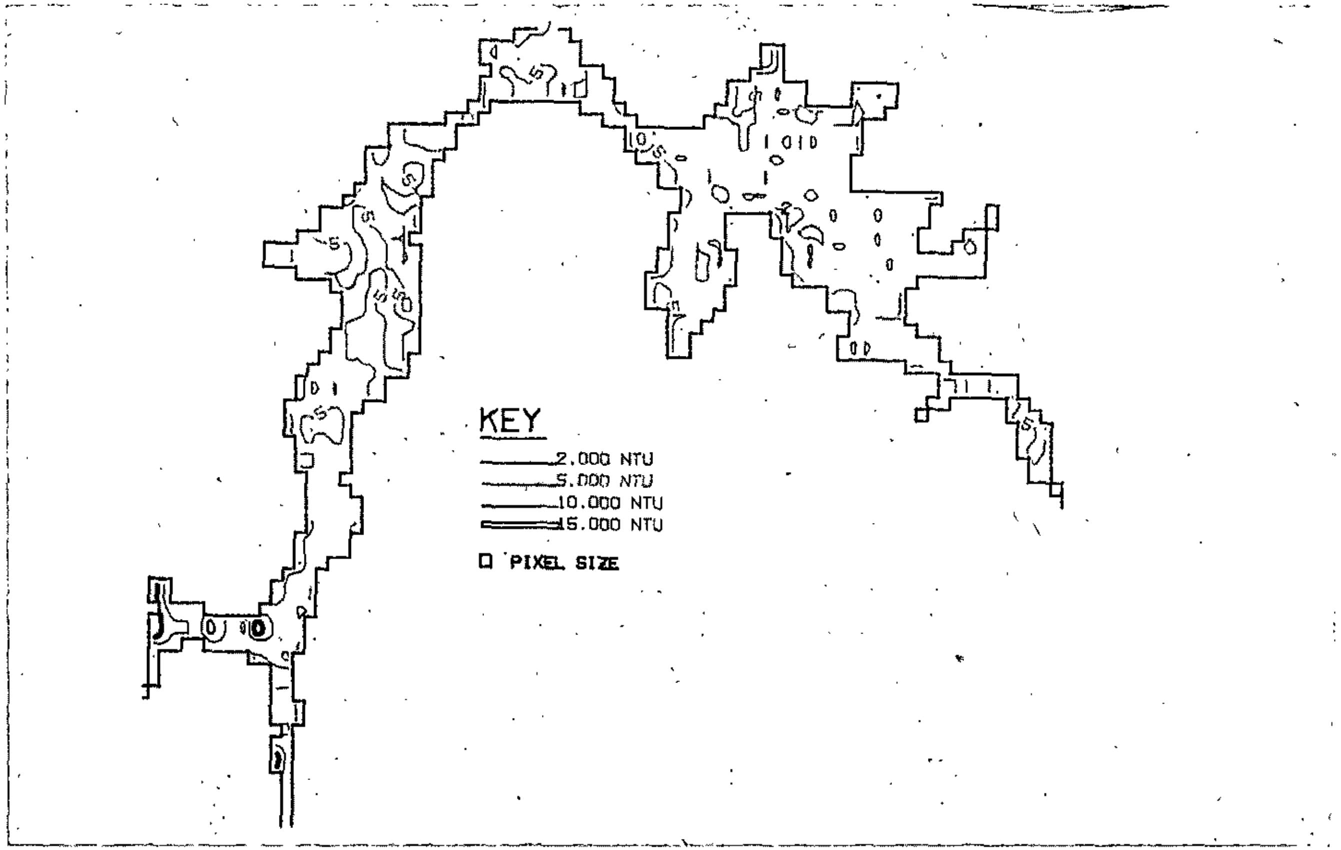


PLATE 6.3: SIMULATED SURFACE TURBIDITY FOR 82.09.30

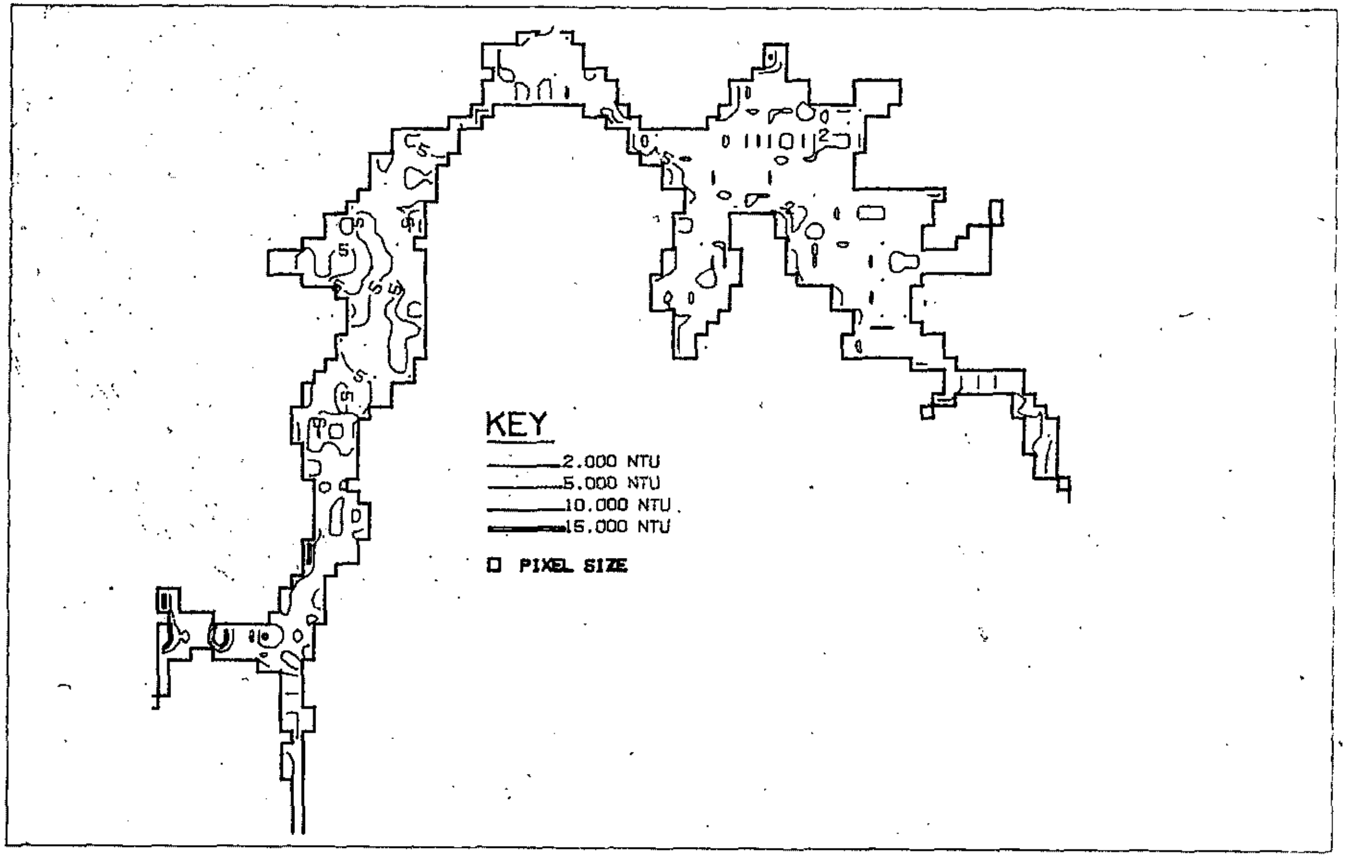


PLATE 6.4: SIMULATED INTEGRATED TURBIDITY FOR 82.09.30

The method of analysis and the resulting model discussed in this report can be applied to any impoundment visible to the satellite, providing that adequate care is taken to ensure representative surface reference data and providing that the basic and the surface reference data in the impoundment can be approximated by a linear model.

Only when limnologists and hydrologists realise the potential of satellite imagery for aiding water resources management and serious attempts are made to utilize this vast source of information, will the value of satellite reflectance data really be determined.

CHAPTER 7

QUESTIONS WHICH ARISE IN THE PRACTICAL APPLICATION OF CALMCAT

7.1 INTRODUCTION

The value of Landsat reflectance data in detecting specific water quality conditions has been reported and yet, before Landsat data can be utilized on an operational basis for water quality purposes, some outstanding questions need to be answered. Some of the more pertinent questions that have been posed are as follows:

(1) How many sampling points on a water body are required to adequately calibrate the satellite reflectance data with the surface reference data?

(2) Can calibration equations obtained from seven sampling points on one day be extrapolated to another day?

(3) Can a single set of calibration equations be generated from combining all of the days seven historical sampling points together?

(4) Can a single set of calibration equations be generated from measurements obtained on several different occasions?

Where possible these questions were examined in order to obtain some idea of the limitations. The model discussed in this report was applied to specific situations.

7.2 How many sampling points on a water body are required to adequately calibrate the satellite reflectance data with the surface reference data?

The aim of the analysis was to see how effective a calibration would be using a relatively few number of points. Taking into consideration the four reflectance bands and the two water quality variables used in the Canonical Correlation Analysis, a minimum number of 7 sampling points were examined.

Seven historical sampling points on Roodeplaat Dam, officially demarcated by the Hydrological Research Institute of the Department of Water Affairs for determining water quality conditions on the impoundment, were examined. The seven historical sampling points are the points numbered 4, 15, 16, 24, 29, 30 and 31 shown on Figure 3.2. Sampling points were positioned near the two major inputs, the main output and at sites where variation could be expected. The 'educated guess' which helped determine where the sampling sites should be placed was aimed at establishing representative sampling sites of the water quality conditions present in the impoundment.

The CALMCAT model was undertaken using the 82.09.30 data for the seven points. The model was run using the coefficients determined from the surface reference data and the satellite reflectance data of the 7 historical sampling points and the simulated values were compared with the observed verification data (23 data pairs) previously used to test the accuracy of the model. The observed

TABLE 7.1: OBSERVED VERSUS SIMULATED WATER QUALITY DATA USING DATA FROM SEVEN HISTORICAL SAMPLING POINTS

ROODEPLAAT DAM DATE: 82.09.30

Sampling Point	Surface Chlorophyll a µg/ℓ		Integrated Chlorophyll a µg/ℓ		Surface Turbidity NTU		Integrated Turbidity NTU	
No.	OBS.	SIM.	OBS.	SIM.	OBS	SIM.	OBS.	SIM.
33	57,00	51,52	40,50	67,76	9,70	6,67	9,70	9,75
34	36,50	47,42	35,80	48,53	8,00	7,38	7,80	8,09
35	34,00	33,81	41,90	39,63	7,70	4,93	8,10	6,50
36	26,30	40,27	34,50	44,67	7,00	5,71	7,70	7,08
37	35,40	31,26	39,20	29,65	7,20	4,57	7,00	5,14
38	35,90	28,58	39,20	29,44	7,00	4,80	7,20	5,47
39	42,60	30,27	44,50	32,06	7,20	5,31	7,00	6,01
40	33,00	30,27	37,80	32,06	6,70	5,31	6,60	6,01
41	33,00	39,45	36,40	41,78	6,00	6,90	6,00	7,52
42	30,60	23,23	33,00	24,32	6,20	4,50	6,10	5,00
43	33,00	10,38	33,50	12,45	5,90	2,38	6,30	3,05
44	27,30	19,77	32,10	18,92	5,60	3,66	5,60	4,00
45	21,50	16,79	33,00	17,42	5,40	2,84	5,70	3,51
46	20,60	17,34	28,70	16,94	5,00	4,54	5,00	4,29
47	21,50	15,78	27,30	15,07	4,40	3,36	5,00	3,48
48	18,70	13,37	27,30	14,06	4,40	2,37	4,90	2,94
49	16,30	16,29	23,00	16,00	4,10	3,40	4,40	3,62
50	19,10	21,53	18,70	21,23	3,70	4,35	4,30	4,49
51	15,69	19,45	20,10	20,37	3,60	5,08	3,60	5,00
52	13,90	17,26	20,60	17,42	3,70	3,76	4,00	3,98
53	16,70	16,29	18,70	16,00	3,90	3,40	4,40	3,62
54	16,70	14,89	17,20	15,89	4,30	3,56	4,40	3,85
55	19,60	20,70	21,10	20,51	3,70	3,66	4,00	4,13

versus simulated values for individual sampling sites are given in Table 7.1. The t test determined whether the difference between the means of the simulated and the observed surface reference data was significant or not (Table 7.2).

Individual sampling points show discrepancies between observed and simulated data (Table 7.1), a fact not surprising in view of the noise in the surface reference data. The results of the t test indicated that the means of surface and integrated chlorophyll a concentrations were acceptably simulated with t values below the critical 2,02 value and with percent relative errors ranging from 8% to 13% (refer to Table 7.2). Surface and integrated turbidity, however, had t values of 4,44 and 3,38 respectively, indicating that there were significant differences between observed and simulated turbidity mean values at the 5% two tailed level of significance. The percentage relative error ranged from 21,5% to 13,5% respectively.

TABLE 7.2: t TEST ANALYSIS BETWEEN OBSERVED AND SIMULATED WATER QUALITY MEASUREMENTS OBTAINED FROM THE CALIBRATION OF SEVEN HISTORICAL SAMPLING POINTS

Water Quality Variables 23 Cases		Mean	Std. Dev.	Diff Mean	t Test	% Relative Error
SURFACE CHLOROPHYLL a µg/ℓ	Observed	27,17	10,59	2,13	1,35	8
	Simulated	25,04	11,28			
INTEGRATED CHLOROPHYLL a	Observed	30,61	8,43	3,99	1,79	13
	Simulated	26,62	13,88			
SURFACE TURBIDITY NTU	Observed	5,67	1,69	1,22	4,44	21,5
	Simulated	4,45	1,36			
INTEGRATED TURBIDITY	Observed	5,86	1,59	0,79	3,38	13,5
	Simulated	5,07	1,76			

A comparison was made between the seven point calibration simulated concentration and distribution values for the whole impoundment, and the simulated values obtained using the 'Normalised Data' calibration discussed in Section 5.5.1 and Section 6.1 (Table 6.2). The 'Normalised Data' calibration for 82.09.30 was considered to be accurate and therefore is used as a basis for comparison. Both a fine and a coarse class interval division was used for this comparison. The results of the comparison are given in Tables 7.3 to 7.6.

For the fine class interval classification of the simulated data values, it was evident from Table 7.3 that surface chlorophyll a values obtained using the seven point calibration model were comparable with the values obtained using the 'Normalised Data' calibration model. Only in the 10 to 19,99 µg/ℓ fine class range was there a 9% difference between the two simulations. Using the coarse class intervals, the two calibrations presented similar distributions.

Table 7.4 of integrated chlorophyll a indicated a difference between the distribution of the two models. There was a 13% integrated chlorophyll a distribution difference in the fine class range of 10 to 19,99 µg/ℓ and a 7% to 8% areal difference in the coarse class range between the seven point calibration and the 'Normalised Data' calibration.

Surface and integrated turbidity values (Tables 7.5 and 7.6) indicated discrepancies in the 0% to 1,99% and the 4% to 5,99% fine class intervals, and yet the coarse class intervals of the 0 to 7,99 NTU range indicated acceptably similar values.

These results illustrate the point that the accuracy depends on the class intervals chosen as well as the number of data points and the representative properties of the sample set. Ballpark estimates were obtained by the seven point calibration which, in some circumstances, may prove to be acceptable, if viewed in terms of time and money spent on obtaining a higher degree of accuracy. The authors caution, however, that non-representative samples may result in completely inaccurate estimates. The greater the number of samples taken, though, the greater the chance of obtaining accurate and trusted results.

A factor that should not be forgotten is that of the overall size and morphometry of the impoundment. The larger the impoundment the larger the sample set should be. Unfortunately this factor could not be investigated. Particular care should be taken when dealing with large impoundments. It may be necessary to divide the impoundment into more than one sample set. For example, as shown in Plate 1.1 of Bloemhof Dam, two entirely different water conditions are evident. The calibration of this impoundment may involve dividing the impoundment into two separate sample sets and undertaking two separate Canonical procedures.

In summary, the sampling of a water body is expensive and manpower intensive and therefore, it is important to obtain the necessary accuracy with as few sampling points as possible. The number of sampling points required to obtain reasonably accurate calibration

TABLE 7.3: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF SURFACE CHLOROPHYLL a FOR 82.09.30 COMPARING THE 'NORMALISED DATA' AND 'SEVEN POINT CALIBRATION' MODELS SHOWING COMPARABILITY

	'NORMALISED DATA' MODEL FOR 82.09.30	'SEVEN POINT CALIBRATION' MODEL FOR 82.09.30
FINE CLASS INTERVAL µg/ℓ	PERCENTAGE AREA	PERCENTAGE AREA
0,00 - 9,99	21,44	18,85
10,00 - 19,99	36,16	45,47
20,00 - 29,99	18,96	15,90
30,00 - 39,99	7,66	6,12
40,00 - 49,99	2,83	2,71
50,00 - 59,99	2,59	2,83
60,00 - 69,99	1,88	2,00
70,00 - 79,99	1,53	1,18
80,00 - 89,99	2,12	2,00
90,00 - 99,99	1,06	0,12
100,00 - 109,99	1,41	0,47
110,00 - 439,99	2,37	2,37
COARSE CLASS INTERVAL µg/ℓ		
0,00 - 29,99	76,56	80,22
30,00 - 99,99	19,67	16,96
100,00 +	3,78	2,84
MEAN =	27,47	24,86
MAX =	430,09	248,41
MIN =	0,81	2,47

TABLE 7.4: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF INTEGRATED CHLOROPHYLL a FOR 82.09.30 COMPARING THE 'NORMALISED DATA' AND 'SEVEN POINT CALIBRATION' MODELS SHOWING COMPARABILITY IN THE COARSE CLASS INTERVAL RANGE

	'NORMALISED DATA' MODEL FOR 82.09.30	'SEVEN POINT CALIBRATION' MODEL FOR 82.09.30
FINE CLASS INTERVAL µg/ℓ	PERCENTAGE AREA	PERCENTAGE AREA
0,00 - 9,99	17,67	18, 4
10,00 - 19,99	31,57	44,64
20,00 - 29,99	22,50	15,90
30,00 - 39,99	10,84	5,65
40,00 - 49,99	4,95	3,42
50,00 - 59,99	2,83	3,30
60,00 - 69,99	1,88	1,18
70,00 - 79,99	1,53	1,41
80,00 - 89,99	1,77	1,88
90,00 - 99,99	1,53	0,59
100,00 - 109,99	0,47	1,41
110,00 - 349,99	2,49	2,48
COARSE CLASS INTERVAL µg/ℓ		
0,00 - 29,99	71,74	78,94
30,00 - 99,99	25,33	17,43
100,00 +	2,96	3,89
MEAN =	28,41	26,72
MAX =	344,05	313,09
MIN =	1,42	3,63

TABLE 7.5: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF SURFACE TURBIDITY FOR 82.09.30 COMPARING THE 'NORMALISED DATA' AND 'SEVEN POINT CALIBRATION' MODELS SHOWING COMPARABILITY IN THE COARSE CLASS INTERVAL RANGE

	'NORMALISED' DATA MODEL FOR 82.09.30	'SEVEN POINT CALIBRATION' MODEL FOR 82.09.30
FINE CLASS INTERVAL NTU	PERCENTAGE AREA	PERCENTAGE AREA
0,00 - 1,99	6,60	12,96
2,00 - 3,99	38,28	37,34
4,00 - 5,99	35,10	25,91
6,00 - 7,99	7,89	10,01
8,00 - 9,99	6,83	6,12
10,00 - 11,99	3,06	0,71
12,00 - 13,99	1,41	2,59
14,00 - 15,99	0,35	0,71
16,00 - 17,99	0,00	1,53
18,00 - 19,99	0,00	0,71
20,00 - 21,99	0,24	0,71
22,00 - 23,99	0,24	0,00
24,00 - 45,99	0,00	0,72
COARSE CLASS INTERVAL NTU		
0,00 - 7,99	87,87	86,22
8,00 - 19,99	11,65	12,37
20,00 +	0,48	1,43
MEAN =	4,87	5,09
MAX =	22,52	44,86
MIN =	1,02	0,83

TABLE 7.6: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF INTEGRATED TURBIDITY FOR 82.09.30 COMPARING THE 'NORMALISED DATA' MODEL AND 'SEVEN POINT CALIBRATION' MODELS SHOWING COMPARABILITY IN THE COARSE CLASS INTERVAL RANGE

	'NORMALISED DATA' MODEL FOR 82.09.30	'SEVEN POINT CALIBRATION' MODEL FOR 82.09.30
FINE CLASS INTERVAL NTU	PERCENTAGE AREA	PERCENTAGE AREA
0,00 - 1,99	0,47	8,36
2,00 - 3,99	35,92	40,28
4,00 - 5,99	39,34	28,03
6,00 - 7,99	12,84	9,07
8,00 - 9,99	6,60	4,71
10,00 - 11,99	2,71	2,93
12,00 - 13,99	0,94	1,65
14,00 - 15,99	0,94	1,53
16,00 - 17,99	0,00	1,06
18,00 - 19,99	0,00	0,82
20,00 - 21,99	0,24	0,59
22,00 - 23,99	0,24	0,35
24,00 - 49,99	0,00	0,72
COARSE CLASS INTERVAL NTU		
0,00 - 7,99	88,87	85,74
8,00 - 19,99	11,19	12,6
20,00 +	0,24	1,66
MEAN =	5,14	5,32
MAX =	21,06	48,25
MIN =	1,84	1,18

relationships between surface reference data and satellite reflectance data should not be fewer than the number of variables used in the multiple regression analysis, but more importantly the sampling sites should, as much as possible, be representative of conditions in the impoundment. The seven point calibration model did provide comparable simulations of concentrations, particularly when coarse class interval ranges were used. It is important to note that the calibration was specific to that day of sampling.

7.3 <u>Can calibration equations, obtained from using seven sampling points on one day, be extrapolated to another day?</u>

In order to investigate this question, data and the resulting calibration equations from the seven historical sampling points on Roodeplaat (as discussed in Section 7.2) for the 81.12.07 overpass were used to simulate values for the overpass of the 82.09.30.

Simulated concentrations and distributions of the four water quality variables for both the fine and coarse class intervals were determined and are presented in Tables 7.7 to 7.10. It is evident from Tables 7.7 and 7.8 that there are large discrepancies between the 'Normalised Data' Model values for 82.09.30 and the 'Seven Point December Calibration' Model values for surface and integrated chlorophyll <u>a</u>.

Each water quality variable showed large differences between the simulated values in the fine class interval ranges.

Tables 7.9 and 7.10 present the surface and integrated turbidity results which show that the coarse class interval values between the two models are comparable. It can be postulated that the reason why the turbidity results show higher accuracy can be related to results described in Section 4.6. The overpass of 81.12.07 shows surface turbidity to be the dominant variable in the Canonical Analysis and therefore it is likely that calibration equations obtained from the 81.12.07 data will simulate turbidities more accurately than chlorophyll <u>a</u>.

The results of this analysis indicate that the extrapolation of a seven sampling point calibration from one day to another did not produce accurate results with the exception of turbidity in the coarse class range.

TABLE 7.7: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF SURFACE CHLOROPHYLL a FOR 82.09.30 COMPARING THE 'NORMALISED DATA' MODEL FOR THIS DATE AND THE 'SEVEN POINT DECEMBER CALIBRATION' MODEL SHOWING LARGE DISCREPANCIES

FINE CLASS INTERVAL µg/ℓ	'NORMALISED DATA MODEL' FOR 82.09.30 PERCENTAGE AREA	'SEVEN POINT DECEMBER CALIBRATION' MODEL FOR 82.09.30 PERCENTAGE AREA
0,00 - 9,99	21,44	41,93
10,00 - 19,99	36,16	43,23
20,00 - 29,99	18,96	11,43
30,00 - 39,99	7,66	1,77
40,00 - 49,99	2,83	0,71
50,00 - 59,99	2,59	0,12
60,00 - 69,99	1,88	0,24
70,00 - 79,99	1,53	0,35
80,00 - 89,99	2,12	0,00
90,00 - 99,99	1,06	0,00
100,00 - 109,99	1,47	0,00
110,00 - 439,99	2,37	0,24
COARSE CLASS INTERVAL µg/ℓ		
0,00 - 29,99	76,56	96,59
30,00 - 99,99	19,67	3,19
100,00 +	3,78	0,24
MEAN =	27,47	13,44
MAX =	430,09	111,68
MIN =	0,81	2,27

TABLE 7.8: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF INTEGRATED CHLOROPHYLL a FOR 82.09.30 COMPARING THE 'NORMALISED DATA' MODEL FOR THIS DATE AND THE 'SEVEN POINT DECEMBER CALIBRATION' MODEL SHOWING LARGE DISCREPANCIES

FINE CLASS INTERVAL µg/ℓ	'NORMALISED' DATA MODEL FOR 82.09.30 PERCENTAGE AREA	'SEVEN POINT DECEMBER CALIBRATION' MODEL FOR 82.09.30 PERCENTAGE AREA
0,00 - 9,99	17,67	1,18
10,00 - 19,99	31,57	6,60
20,00 - 29,99	22,50	6,71
30,00 - 39,99	10,84	8,24
40,00 - 49,99	4,95	8,13
50,00 - 59,99	2,83	6,95
60,00 - 69,99	1,88	5,89
70,00 - 79,99	1,53	4,36
80,00 - 89,99	1,77	6,01
90,00 - 99,99	1,53	3,78
100,00 - 109,99	0,47	5,42
110,00 - 349,99	2,49	37,39
COARSE CLASS INTERVAL µg/ℓ		
0,00 - 29,99	71,74	14,49
30,00 - 99,99	25,33	42,76
100,00 +	2,96	42,81
MEAN =	28,41	144,74
MAX =	344,05	3253,60
MIN =	1,42	2,69

TABLE 7.9: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF SURFACE TURBIDITY FOR 82.09.30 COMPARING THE 'NORMALISED DATA' MODEL FOR THIS DATE AND THE 'SEVEN POINT DECEMBER CALIBRATION' MODEL SHOWING COMPARABILITY IN THE COARSE CLASS INTERVAL RANGE

	'NORMALISED DATA' MODEL FOR 82.09.30	'SEVEN POINT DECEMBER CALIBRATION' MODEL FOR 82.09.30
FINE CLASS INTERVAL NTU	PERCENTAGE AREA	PERCENTAGE AREA
0,00 - 1,99	6,60	19,2
2,00 - 3,99	38,28	31,33
4,00 - 5,99	35,10	25,32
6,00 - 7,99	7,89	10,13
8,00 - 9,99	6,83	5,06
10,00 - 11,99	3,06	2,83
12,00 - 13,99	1,41	0,82
14,00 - 15,99	0,35	1,77
16,00 - 17,99	0,00	0,94
18,00 - 19,99	0,00	0,59
20,00 - 21,99	0,24	0,35
22,00 - 23,99	0,24	0,12
24,00 - 48,00	0,00	1,55
COARSE CLASS INTERVAL NTU		
0,00 - 7,99	87,87	85,98
8,00 - 19,99	11,65	12,01
20,00 +	0,48	2,02
MEAN =	4,87	5,1
MAX =	22,52	47,79
MIN =	1,02	0,51

TABLE 7.10: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF INTEGRATED TURBIDITY FOR 82.09.30 COMPARING THE 'NORMALISED DATA' MODEL FOR THIS DATE AND THE 'SEVEN POINT DECEMBER CALIBRATION' MODEL SHOWING COMPARABILITY IN THE COARSE CLASS INTERVALS RANGE

	'NORMALISED DATA' MODEL FOR 82.09.30	'SEVEN POINT DECEMBER CALIBRATION' MODEL FOR 82.09.30
FINE CLASS INTERVAL NTU	PERCENTAGE AREA	PERCENTAGE AREA
0,00 - 1,99	0,47	9,42
2,00 - 3,99	35,92	43,11
4,00 - 5,99	39,34	26,97
6,00 - 7,99	12,84	11,66
8,00 - 9,99	6,60	4,00
10,00 - 11,99	2,71	1,88
12,00 - 13,99	0,94	1,53
14,00 - 15,99	0,94	0,47
16,00 - 17,99	0,00	0,24
18,00 - 19,99	0,00	0,24
20,00 - 21,99	0,24	0,24
22,00 - 23,99	0,24	0,24
COARSE CLASS INTERVAL NTU		
0,00 - 7,99	88,87	91,16
8,00 - 19,99	11,19	8,36
20,00 +	0,24	0,48
MEAN =	5,14	4,55
MAX =	21,06	22,86
MIN =	1,84	0,92

7.4 Can a single set of calibration equations be generated from combining all of the days seven historical sampling points together?

Using the seven historical sampling points from each of the days a single set of calibration equations was determined termed the 'All Seven Point Calibration'. Once again the values obtained from the 'Normalised Data' Model for 82.09.30 were used as a basis for accuracy and comparisons were made with the 'All Seven Point Calibration' simulation. The results are shown in Tables 7.11 to 7.14.

For each water quality variable, the fine class interval simulated concentrations and distributions were not comparable. Only the coarse class interval range for surface chlorophyll a (Table 7.11) showed a reasonable similarity. These results indicate that a general calibration of all of the seven historical points produced comparable accuracies of concentrations and distributions of surface chlorophyll a using a coarse class interval range for the day tested.

TABLE 7.11: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF SURFACE CHLOROPHYLL a FOR 82.09.30 COMPARING THE 'NORMALISED DATA' MODEL FOR THIS DATE AND 'ALL SEVEN POINT CALIBRATION' MODEL SHOWING SIMILARITY IN THE COARSE CLASS INTERVALS RANGE

	'NORMALISED DATA' MODEL FOR 82.09.30	'ALL SEVEN POINT CALIBRATION' MODEL FOR 82.09.30
FINE CLASS INTERVAL µg/ℓ	PERCENTAGE AREA	PERCENTAGE AREA
0,00 - 9,99	21,44	43,70
10,00 - 19,99	36,16	26,27
20,00 - 29,99	18,96	8,72
30,00 - 39,99	7,66	5,89
40,00 - 49,99	2,83	5,42
50,00 - 59,99	2,59	1,30
60,00 - 69,99	1,88	1,30
70,00 - 79,99	1,53	2,12
80,00 - 89,99	2,12	1,41
90,00 - 99,99	1,06	1,06
100,00 - 109,99	1,41	0,35
110,00 - 439,99	2,37	2,49
COARSE CLASS INTERVAL µg/ℓ		
0,00 - 29,99	76,56	78,69
30,00 - 99,99	19,67	18,5
100,00 +	3,78	2,84
MEAN =	27,47	24,44
MAX =	430,09	704,59
MIN =	0,81	0,57

TABLE 7.12: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF INTEGRATED CHLOROPHYLL a FOR 82.09.30 COMPARING THE 'NORMALISED DATA' MODEL FOR THIS DATE AND 'ALL SEVEN POINT CALIBRATION' MODEL SHOWING LARGE DISCREPANCIES

	'NORMALISED DATA' MODEL FOR 82.09.30	'ALL SEVEN POINT CALIBRATION' MODEL FOR 82.09.30
FINE CLASS INTERVAL µg/ℓ	PERCENTAGE AREA	PERCENTAGE AREA
0,00 - 9,99	17,67	44,41
10,00 - 19,99	31,57	27,33
20,00 - 29,99	22,50	9,07
30,00 - 39,99	10,84	5,54
40,00 - 49,99	4,95	4,00
50,00 - 59,99	2,83	2,36
60,00 - 69,99	1,88	2,12
70,00 - 79,99	1,53	1,53
80,00 - 89,99	1,77	0,47
90,00 - 99,99	1,53	0,35
100,00 - 109,99	0,47	0,94
110,00 - 349,99	2,49	1,91
COARSE CLASS INTERVAL µg/ℓ		
0,00 - 29,99	71,74	80,81
30,00 - 99,99	25,33	16,37
100,00 +	2,96	2,85
MEAN =	28,41	21,76
MAX =	344,05	451,63
MIN =	1,42	0,77

TABLE 7.13: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF SURFACE TURBIDITY FOR 82.09.30 COMPARING THE 'NORMALISED DATA' MODEL FOR THIS DATE AND 'ALL SEVEN POINT CALIBRATION' MODEL SHOWING LARGE DISCREPANCIES

	'NORMALISED DATA' MODEL FOR 82.09.30	'ALL SEVEN POINT CALIBRATION' MODEL FOR 82.09.30
FINE CLASS INTERVAL NTU	PERCENTAGE AREA	PERCENTAGE AREA
0,00 - 1,99	6,60	12,60
2,00 - 3,99	38,28	23,79
4,00 - 5,99	35,10	15,08
6,00 - 7,99	7,89	12,60
8,00 - 9,99	6,83	8,24
10,00 - 11,99	3,06	2,59
12,00 - 13,99	1,41	3,18
14,00 - 15,99	0,35	2,47
16,00 - 17,99	0,00	3,65
18,00 - 19,99	0,00	2,00
20,00 - 21,99	0,24	1,65
22,00 - 23,99	0,24	2,00
24,00 - 45,99	0,00	10,15
COARSE CLASS INTERVAL NTU		
0,00 - 7,99	87,87	64,07
8,00 - 19,99	11,65	22,13
20,00 +	0,48	13,80
MEAN =	4,87	12,26
MAX =	22,52	310,54
MIN =	1,02	0,34

TABLE 7.14: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF INTEGRATED TURBIDITY FOR 82.09.30 COMPARING THE 'NORMALISED DATA' MODEL FOR THIS DATE AND THE 'ALL SEVEN POINT CALIBRATION' MODEL SHOWING DISCREPANCIES

	'NORMALISED DATA' MODEL FOR 82.09.30	'ALL SEVEN POINT CALIBRATION' MODEL FOR 82.09.30
FINE CLASS INTERVAL NTU	PERCENTAGE AREA	PERCENTAGE AREA
0,00 - 1,99	0,47	13,73
2,00 - 3,99	35,92	26,50
4,00 - 5,99	39,34	19,08
6,00 - 7,99	12,84	11,43
8,00 - 9,99	6,60	5,77
10,00 - 11,99	2,71	5,06
12,00 - 13,99	0,94	2,83
14,00 - 15,99	0,94	2,47
16,00 - 17,99	0,00	2,59
18,00 - 19,99	0,00	0,94
20,00 - 21,99	0,24	2,47
22,00 - 23,99	0,24	1,30
24,00 - 120,99	0,00	5,81
COARSE CLASS INTERVAL NTU		
0,00 - 7,99	88,87	70,79
8,00 - 19,99	11,19	19,66
20,00 +	0,24	9,58
MEAN =	5,14	8,44
MAX =	21,06	118,27
MIN =	1,84	0,62

7.5 Can a single set of calibration equations be generated from measurements obtained on several different occasions?

In order to investigate this question the surface reference data obtained from 5 days of sampling i.e. 81.10.14, 81.11.01, 81.12.07, 82.09.13 and 82.11.16 were combined. Data for the day 82.09.30 was left out in order not to bias the analysis. The data obtained for 82.09.30 using the 'Normalised Data' model, as discussed in Section 6.2 were considered to be accurate and therefore could be used to test the accuracy of the 'Five Day Calibration' Models. The CALMCAT model was carried out using two of the options discussed in Section 3.4.7. The first option tested included all of the five days data, including outliers and ignoring the assumption of normality - hence named the 'Five Day Calibration Including All Data' Model. The second option followed the assumption of normality, excluding outliers and removing clustering, therefore the normalised data sets for each of the days were combined - hence named the 'Five Day Calibration Normalised Data' Model (see Appendix T). The two options were tested in order to obtain some idea of how outliers and non-normal data would affect a generalised calibration.

Tables 7.15 to 7.18 present a comparison of the distribution trend between the 'Normalised Data' calibration results and those for the 'Five Day Calibration' Models. Table 7.15 indicates that the 'Five Day Calibration Including All Data' Model has 36% of the impoundment area containing between 0 to 9,99 µg/ℓ of surface chlorophyll a in comparison with 21% using the 'Normalised Data' Model and 22% using the 'Five Day Calibration 'Normalised Data' Model. All of the remaining interval ranges both coarse and fine have comparable distributions for surface chlorophyll a. The 'Five Day Calibration 'Normalised Data' Model has slightly closer values to the 'Normalised Data' than the 'Five Day Calibration Including All Data' Model.

Table 7.16 presents the integrated chlorophyll a results and shows distribution differences in the fine class interval ranges of 10,00 to 29,99 µg/ℓ and over 110 µg/ℓ. Although the results show inaccuracies the 'Five Day Calibration Including All Data' Option gives slightly better results than the 'Five Day Calibration Normalised Data' Model. Both surface and integrated turbidity distributions (Tables 7.17 and 7.18) for the 'Five Day Calibrations' show large discrepancies in comparison to the 'Normalised Data' Model distributions in both the fine and coarse interval ranges. Surprisingly both of the 'Five Day Calibration' Models show very similar results.

The results from this analysis indicate that surface chlorophyll a is the one water quality variable that has been reasonably accurately simulated using a generalised 'Five Day Calibration' Model. The data so far collected in this project cannot be considered to be sufficient to generate a single set of calibration equations that can accurately simulate integrated chlorophyll a and surface and integrated turbidity. Neither concentrations nor distributions for these conditions were adequately simulated using the calibration models established from a combination of different days data.

TABLE 7.15: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF SURFACE CHLOROPHYLL a FOR 82.09.30 COMPARING THE 'NORMALISED DATA' MODEL FOR THIS DATE AND THE TWO 'FIVE DAY CALIBRATION' MODELS SHOWING REASONABLE COMPARABILITY

	'NORMALISED DATA' MODEL FOR 82.09.30	'FIVE DAY CALIBRATION INCLUDING ALL DATA' MODEL	'FIVE DAY CALIBRATION NORMALISED DATA' MODEL
FINE CLASS INTERVAL µg/ℓ	PERCENTAGE AREA	PERCENTAGE AREA	PERCENTAGE AREA
0,00 - 9,99	21,44	35,77	22,31
10,00 - 19,99	36,16	30,81	40,02
20,00 - 29,99	18,96	13,93	16,88
30 00 - 39,99	7,66	6,49	7,91
40,00 - 49,99	2,83	4,25	4,72
50,00 - 59,99	2,59	1,30	2,36
60,00 - 69,99	1,88	2,01	1,77
70,00 - 79,99	1,53	0,47	1,30
80,00 - 89,99	2,12	1,77	1,30
90,00 - 99,99	1,06	0,59	0,71
100,00 - 109,99	1,41	0,35	0,12
110,00 - 439,99	2,37	2,24	0,60
COARSE CLASS INTERVAL µg/ℓ			
0,00 - 29,99	76,56	80,51	79,21
30,00 - 99,99	19,67	16,88	20,07
100,00 +	3,78	2,59	0,71
MEAN =	27,47	22,03	22,49
MAX =	430,09	257,88	197,86
MIN =	0,81	0,70	1,61

TABLE 7.16: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF INTEGRATED CHLOROPHYLL a FOR 82.09.30 COMPARING THE 'NORMALISED DATA' MODEL FOR THIS DATE AND THE TWO 'FIVE DAY CALIBRATION' MODELS SHOWING POOR COMPARABILITY

	'NORMALISED DATA' MODEL FOR 82.09.30	'FIVE DAY CALIBRATION INCLUDING ALL DATA' MODEL	'FIVE DAY CALIBRATION NORMALISED DATA' MODEL
FINE CLASS INTERVAL µg/ℓ	PERCENTAGE AREA	PERCENTAGE AREA	PERCENTAGE AREA
0,00 - 9,99	17,67	22,79	39,9
10,00 - 19,99	31,57	21,25	17,59
20,00 - 29,99	22,50	12,51	13,11
30,00 - 39,99	10,84	8,74	2,83
40,00 - 49,99	4,95	5,90	3,19
50,00 - 59,99	2,83	4,84	2,60
60,00 - 69,99	1,88	1,53	2,13
70,00 - 79,99	1,53	1,77	2,01
80,00 - 89,99	1,77	1,18	2,24
90,00 - 99,99	1,53	3,19	1,18
100,00 - 109,99	0,49	0,94	1,06
110,00 - 4090,00	2,49	15,40	12,15
COARSE CLASS INTERVAL µg/ℓ			
0,00 - 29,99	71,74	56,55	70,61
30,00 - 99,99	25,33	27,15	16,18
100,00 +	2,96	16,34	13,21
MEAN =	28,41	78,45	64,90
MAX =	344,05	2956,08	4089,00
MIN =	1,42	0,33	0,13

TABLE 7.17: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF SURFACE TURBIDITY FOR 82.09.30 COMPARING THE 'NORMALISED DATA' MODEL FOR THIS DATE AND THE TWO 'FIVE DAY CALIBRATION' MODELS SHOWING POOR COMPARABILITY

	'NORMALISED DATA' MODEL FOR 82.09.30	'FIVE DAY CALIBRATION INCLUDING ALL DATA' MODEL	'FIVE DAY CALIBRATION NORMALISED DATA' MODEL
FINE CLASS INTERVAL NTU	PERCENTAGE AREA	PERCENTAGE AREA	PERCENTAGE AREA
0,00 - 1,99	6,60	4,96	4,96
2,00 - 3,99	38,28	17,47	16,77
4,00 - 5,99	35,10	17,24	18,18
6,00 - 7,99	7,89	8,03	7,67
8,00 - 9,99	6,83	10,51	11,57
10,00 - 11,99	3,06	8,03	7,91
12,00 - 13,99	1,41	5,90	4,01
14,00 - 15,99	0,35	2,95	3,07
16,00 - 17,99	0,00	2,83	6,26
18,00 - 19,99	0,00	3,54	1,53
20,00 - 21,99	0,24	1,42	1,65
22,00 - 23,99	0,24	1,53	2,48
24,00 - 262,00	0,00	15,61	13,94
COARSE CLASS INTERVAL NTU			
0,00 - 7,99	87,87	47,7	47,58
8,00 - 19,99	11,65	33,76	34,35
20,00 +	0,48	18,56	18,07
MEAN =	4,87	15,92	15,50
MAX =	22,52	260,28	237,48
MIN =	1,02	0,43	0,46

TABLE 7.18: SIMULATED CONCENTRATIONS AND DISTRIBUTIONS OF INTEGRATED TURBIDITY FOR 82.09.30 COMPARING THE 'NORMALISED DATA' MODEL FOR THIS DATE AND THE TWO 'FIVE DAY CALIBRATION' MODELS SHOWING POOR COMPARABILITY

	'NORMALISED DATA' MODEL FOR 82.09.30	'FIVE DAY CALIBRATION INCLUDING ALL DATA' MODEL	'FIVE DAY CALIBRATION NORMALISED DATA' MODEL
FINE CLASS INTERVAL NTU	PERCENTAGE AREA	PERCENTAGE AREA	PERCENTAGE AREA
0,00 - 1,99	0,47	2,60	3,78
2,00 - 3,99	35,92	16,17	17,47
4,00 - 5,99	39,34	20,66	17,59
6,00 - 7,99	12,84	12,87	7,91
8,00 - 9,99	6,60	11,57	10,63
10,00 - 11,99	2,71	8,26	8,62
12,00 - 13,99	0,94	4,60	5,55
14,00 - 15,99	0,94	4,37	3,19
16,00 - 17,99	0,00	2,60	3,66
18,00 - 19,99	0,00	2,83	2,48
20,00 - 21,99	0,24	2,13	3,07
22,00 - 23,99	0,24	0,71	0,59
24,00 - 250,00	0,00	10,63	15,46
COARSE CLASS INTERVAL NTU			
0,00 - 7,99	88,87	52,30	46,75
8,00 - 19,99	11,19	34,23	34,13
20,00 +	0,24	13,47	19,12
MEAN =	5,14	12,04	15,78
MAX =	21,06	134,75	248,47
MIN =	1,84	0,67	0,51

More data needs to be collected to establish firstly, the reasons for the differences between overpasses and secondly, to establish if there are seasonal patterns in the calibration data. This is a field of enquiry that requires attention.

7.6 OUTLIERS

A question which arises from the analysis discussed in this report is: If outliers are present in the data, their presence could be indicative of a pollution source, for example, that needs to be detected and quantified. If so, how can outliers be analysed?

"The treatment of these outliers is an unresolved and controversial question" (Haan, 1977). Often outliers represent conditions that are physically out of the ordinary. It can be postulated that in the data relevant to this report, some outliers determined by Filliben's R and Grubbs t test could be mixels (mixed land and water pixels) if they lie near the edge of the impoundment.

Outliers have the potential to invalidate results as discussed in Section 2.5.4, but this is not always the case. The outliers may be part of the linear function, but are excluded because not all of the ranges of data between the high and the low values have been adequately represented, and they are seen as being out of bounds.

Equally, outliers are known to stabilise regressions where the bulk of the data contains noise and has poor correlations. It is a difficult situation to assess.

In order to analyse outliers a number of outlier points are required. The decision was made to combine all the outliers, for the 6 days, into one set of data, in order to determine whether or not they represented one population.

A Stepwise Discriminant Analysis was carried out and the results, given on Figure 7.1 showed that the outliers were of different populations and therefore could not be analysed together. The fact that the outliers were different may indicate that they were most likely caused by different influences.

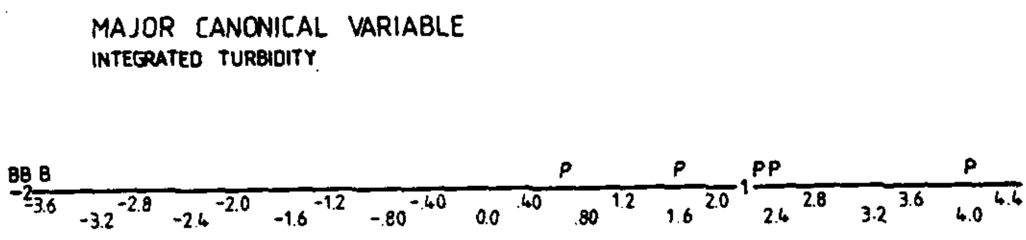


FIGURE 7.1: STEPWISE DISCRIMINANT ANALYSIS HISTOGRAM FOR THE OUTLIERS

7.7 SUMMARY

It is practically desirable, in terms of sampling costs, that the number of sampling points needed to calibrate a satellite image of a water body be kept to a minimum. It is essential to ensure, however, that the sampling points be representative of conditions in the impoundment and that there are more sampling points than the number of variables used in the statistical analysis. For one of the days examined seven representative sampling sites on the impoundment provided acceptable simulations. The error in the mean simulated chlorophyll a < 5μg/ℓ and the error in the mean simulated turbidity < 2 NTU (Table 7.2).

Researchers may also attempt to use calibration equations obtained for one day to extrapolate information to another day. Such extrapolations, while providing poor results in terms of detail are, however, often acceptable for turbidity where the researcher is only interested in a coarse class interval classification i.e., a limited extrapolation is possible. The error in areal estimates of turbidity, in the coarse class range were < 3% (Tables 7.9 and 7.10). Extrapolation of chlorophyll however was not so successful as areal estimates of surface chlorophyll a were out by 20% (Table 7.7) and > 50% in the case of integrated chlorophyll (Table 7.8).

Generalised calibration equations obtained from combining all of the different days data did not produce accurate results with one exception viz., surface chlorophyll a, where a coarse class interval areal estimate was < 4% (Tables 7.11 and 7.15).

In conclusion, the extrapolation of calibration equations is problematical and the accuracy depends on the representativeness and the variation of conditions in impoundments from day to day.

CHAPTER 8

8. CONCLUSION, PRECONDITIONS AND RECOMMENDATIONS

The major objective of the Landsat Water Quality Surveillance Project of Roodeplaat Dam was to determine the potential and limitation for quantitative measurement of the distribution of chlorophyll a and turbidity using Landsat data. This objective has been achieved. Landsat data, when calibrated against surface reference data, can provide acceptably accurate simulations of water quality conditions in an impoundment providing that an attempt is made to adhere to certain preconditions.

8.1 PRECONDITIONS

The preconditions for accurate analysis to be accomplished are as follows:

(1) The sampling of a water body should be undertaken concurrently with the satellite overflight.

(2) The sampling network should be set up to ensure that the entire range of different water quality conditions within the water body are monitored.

(3) The alignment of the sampling position with its relevant pixel on a Landsat scene should be as accurate as possible.

(4) The water quality conditions that are being investigated should be visible to the satellite. This implies that water quality conditions without 'colour' in the spectral region 0,5 µm to 1,1 µm cannot be directly monitored.

(5) The analysis of the water quality samples should be undertaken as soon as possible after the sampling operation.

(6) The multicollinearity and interrelatedness of both the reflectance bands and the water quality conditions should be recognised, and in order to take this factor into account, some type of multivariate statistical analysis should be used.

(7) The statistical assumptions of the multivariate analysis should be adhered to as far as possible.

(8) Cognisance should be taken of the fact that water bodies are non-homogeneous and therefore more than one statistical population may be present in the water quality data, with consequent implications to the statistical analysis and interpretation.

(9) The presence of outliers in the surface reference data should be determined, if indicative of another population as described by the Stepwise Discriminant Analysis, and may need to be excluded prior to analysis.

(10) The water quality data sampled may not necessarily represent conditions in the impoundment and therefore attempts should be made to obtain a representative subset of the conditions present.

(11) The simulative equations that result from a multivariate analysis should be tested with data not previously used in the development of the model.

(12) Where only one of the four parameters is of interest to the user, e.g. surface chlorophyll a, it is still necessary to measure all four variables viz., surface and integrated chlorophyll and surface and integrated turbidity in order to calibrate the CALMCAT model.

8.2 THE RECOMMENDED METHOD

In this study, an attempt to observe the abovementioned conditions has been made, and the relationship between chlorophyll a, turbidity, and the four MSS reflectance bands using 6 different days of data has been determined.

(1) The false colour and colour coded images of Roodeplaat Dam indicated that differences in chlorophyll a and turbidity were recognisable.

(2) The Stepwise Discriminant Analysis was used to indicate the existence or not or more than one population in the impoundment.

(3) Grubbs t test was used for detecting outliers.

(4) Fillibens R Probability Plot Correlation Coefficient test was used to determine the normality of the data set. A normalisation procedure, using areas under the normal curve, was used to detect clustering and obtain a representative subset of data when the logged data was not normal.

(5) A multivariate statistical analysis i.e., the Canonical Correlation Analysis, was used to correlate water quality conditions with the 4 reflectance bands.

(6) The Canonical Correlation Analysis was limited to selected pairs of water quality variables to avoid singularity due to the high intercorrelations between the surface reference data variables.

(7) A linear regression analysis, and the Canonical Correlation Coefficients were used to obtain calibration equations of the relationship between surface reference data and satellite reflectance data. The model CALMCAT was established from the combined analysis.

(8) In order to test the accuracy of the model, the simulations of the calibration equations were tested against data not previously used in the analysis, by means of the coefficient of efficiency and the Student's t test.

(9) The calibrated model was used to calculate the chlorophyll a and turbidity values for each pixel over the entire area of the impoundment thus enabling a synoptic view of these parameters to be obtained.

The menu for the analysis of surface reference data and satellite reflectance data is given in Appendix S.

8.3 THE RESULTS

When the abovementioned preconditions and methods were observed, the mean water quality conditions within Roodeplaat Dam were calculated with an accuracy of ≤ 9 µg/ℓ for chlorophyll and ≤ 1 NTU for turbidity.

Simulated versus observed chlorophyll a values at individual sampling sites varied to a much greater extent due to noise and the non homogeneity of the chlorophyll a. Turbidity values at individual sampling sites were fairly accurately simulated.

The utilization of the model CALMCAT to calculate chlorophyll a and turbidity values for the entire water body, provided synoptic, accurate information of the distribution and concentration of these variables.

For Roodeplaat Dam it was found that each calibration equation was unique for that day and that each calibration differed from one overpass to the next.

Attempts to determine the least numbers of sampling points required to obtain accurate calibration of satellite reflectance data with surface reference data, indicated that it is not the number of sampling points that is important, but that the sampling points be as representative as possible of the full range of water quality conditions present in the impoundment.

8.4 SUMMARY

The relationship between specific water quality conditions for chlorophyll a and turbidity and Landsat MSS bands was determined through the careful selection of a representative subset of water quality data and the use of the Canonical Correlation Analysis. Development of the model CALMCAT made use of the established relationship in a simulative fashion. Accurate estimates of distributions and concentrations of chlorophyll a and turbidity in an impoundment were gained. Information of this nature may potentially complement and enhance point source information presently applied to water resources investigations. Withdrawal points for water purification, the siting of recreational facilities, matters relating to sewage disposal and the relationship between water quality conditions and nutrient inputs are a few of the fields of study that may benefit from the comprehensive information that can be obtained from satellite derived information. Landsat data can be effectively used to produce practical, quantitative information of the water quality conditions in impoundments.

REFERENCES

ACKERMANN, W.C. (1974) Water Management - A General Report, Hydrological Sciences Bulletin XIX, 1, 3, 119-130.

AITKEN, A.P. (1973) Assessing Systematic Errors in Rainfall-Runoff Models, Journal of Hydrology, 20, 121-136.

ANDERSON, D.G. (1979) Satellite Versus Conventional Methods in Hydrology, Satellite Hydrology, (Eds. Deutsch, M., Wiesnet, D.R., Rango, A), Proc. of the Fifth Annual William T. Pecora Memorial Symposium on Remote Sensing, Sioux Falls, South Dakota, June 10-15, American Water Resources Association, Minneapolis, Minnesota, 33-36.

ARANUVACHAPUN, S. and LeBLOND, P.H. (1981) Turbidity of Coastal Water Determined from Landsat, Remote Sensing of Environment, 84, 113-132.

BARTOLUCCI, L.A., ROBINSON, B.F. and SILVA, L.F. (1977) Field Measurements of the Spectral Response of Natural Waters, Photogrammetric Engineering and Remote Sensing, 43, 5, 595-598.

BOLAND, D.H.P. (1976) Trophic classification of Lakes using Landsat -1. (ERTS -1) Multispectral Scanner Data, U.S. Environmental Protection Agency, Corvallis Environmental Research Laboratory, Corvallis, Oregon. EPA-600/3-76-037, 1-245.

BOLAND, D.H.P., SCHAEFFER, D.J., SEFTON, D.F., CLARKE, R.P., and BLACKWELL, R.J. (1979) Trophic Classification of Selected Illinois Water Bodies: Lake Classification Through Amalgamation of LANDSAT Multispectral Scanner and Contact- Sensed Data, U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada. EPA-600/3-79-123, 1-225.

BUKATA, R.P. (1978) Physical Interpretation of Synoptic Patterns Observed over the Great Lakes from Satellite Altitudes, Water Quality Bulletin, 3, 4, 12-13, 22-23.

BUKATA, R.P. and BRUTON, J.E. (1974) The Application of Telemetered ERTS Data to Lakes Research, Proc. of the Second Canadian Symposium on Remote Sensing, University of Guelph, Guelph, Ontario, April 29 - May 1, 495-506.

BUKATA, R.P., BRUTON, J.E., JEROME, J.H. BOBBA, A.G. and HARRIS, G.P. (1975(a)) The Application of Landsat-1 Digital Data to a Study of Coastal Hydrography, Journal Source unknown, Canada Centre for Inland Waters, Burlington, Ontario, 331-348.

BUKATA, R.P., HARAS, W.S., and BRUTON, J.E. (1975(b)) The application of ERTS-1 digital data to water transport phenomena in the Point Pelee-Rondeau area, Verh.Internat. Verein. Limnol, 19, 168-178.

BUKATA, R.P., HARRIS, G.P. and BRUTON, J.E. (1974). The Detection of Suspended Solids and Chlorophyll 'A' utilizing Digital Multispectral ERTS-1 Data, Proc. of the Second Canadian Symposium on Remote Sensing, University of Guelph, Guelph, Ontario. April 1-May 1, 551-564.

BURGY, R.H. (1973) Application of ERTS-1 Data to Aid in Solving Water Resources Management Problems in the State of California, Proc. Symposium on Significant Results obtained from the Earth Resources Technology Satellite-1, 2, Summary of Results, NASA, Greenbelt, Maryland, 151-166.

CANFIELD, D.E. (1983) Prediction of Chlorophyll a Concentrations in Florida Lakes: The Importance of Phosphorus and Nitrogen, Water Resources Bulletin, 19, 2, 255-262.

CARPENTER, D.J. (1982) An Assessment of Landsat MSS Satellite Data for Inland Water Resources Application, Department of National Development and Energy, Australian Water Resources Council, Technical Paper No. 72, Australian Government Publishing Service, Canberra.

CARPENTER, D.J and CARPENTER, S.M. (1983) Modeling Inland Water Quality Using Landsat Data, Remote Sensing of Environment, 13, 4, 345-352.

CHAGARLAMUDI, P., HECKY, R.E., and SCHUBERT, J.S. (1979) Quantitative Monitoring of Sediment Levels in Freshwater Lakes from Landsat, COSPAR Symposium on the Contribution of Space Observations to Water Resources Studies, Bangalore, India, 29 May-2 June.

CROTEAU, C. (1979) Remote Sensing, Water Spectrum, 11, 3, 46-53.

CURRAN, P.J. (1984) Remote Sensing from Earth Resources Satellites - Review and Preview, Satellite Remote Sensing - Review and Preview, Proc. of the Tenth Anniversary International Conference, University of Reading, Reading, 18-21 September, The Remote Sensing Society, 1-11

DANIEL, C and WOOD, F.S.(1971) Fitting Equations to Data, Computer Analysis of Multifactor Data for Scientists and Engineers, Wiley Interscience, John Wiley and Sons Inc, 1-342.

DEPARTMENT OF WATER AFFAIRS. (1985) Draft Report on the Management of South African Water Resources, Pretoria, RSA.

DHANOA, M.S. (1982) A BASIC computer program which tests for normality and the presence of outliers: some modifications, Laboratory Practice, April, 330-331.

DIXON, W.J. and BROWN, M.B. (Eds) (1979) BMDP-79 Biomedical Computer Programs P-series, University of California Press, Berkley, California.

DRAPER, N.P. and SMITH, H. (1966) Applied Regression Analysis, John Wiley and Sons Inc.

FILLIBEN, J.J. (1975) The Probability Plot Correlation Coefficient Test for Normality, Technometrics, 17, 1, 111-117.

FISHER, L.T., SCARPACE, F.L., and THOMPSEN, R.G. (1979) Multidate Landsat Lake Quality Monitoring Program, Photogrammetric Engineering and Remote Sensing, 45, 4, 519-527.

GITTINS, R. (1979) Ecological Applications of Canonical Analysis, Multivariate Methods in Ecological Work, (Eds. Orloci. L., Rao. C.R., Stiteler W.M.) Statistical Ecology, Vol 7. International Co-operative Publishing House, Fairland, Maryland, 309-535.

GRAHAM, D.S. and HILL, J.M. (1983) Landsat and Water Quality Model Verification, Journal of Transportation Engineering, Proc. of the American Society of Civil Engineers, 109, 5, 640-650.

GRANT, A. (1983) Multiple Regression with non-linear data structures, Laboratory Practice, 32, 2, 79.

GRIMSHAW, H.J., TORRANS, S.M. and LERA, T. (1980) Regression Analysis of Reservoir Water Quality Parameters with Digital Satellite Reflectance Data, Restoration of Lakes and Inland Waters, International Symposium on Inland Waters and Lake Restoration, Sept. 8-12, Portland, Maine, 222-225.

GROBLER, D.C and DAVIES, E. (1981) Sediments as a Source of Phosphate: A Study of 38 Impoundments, Water S.A., 7, 1, 54-60.

GRUBBS, F.E. and BECK, G. (1972) Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations, Technometrics, 14, 4, 847-854.

GUPTA, R.P. and BODECHTEL, J. (1982) Geotechnical Applications of Landsat Image Analysis of Bhakra Dam Reservoir, India, Remote Sensing of Environment, 12, 1, 3-13.

HAAN, C.T. (1977) Statistical Methods in Hydrology, The Iowa State University Press, Ames, Iowa.

HARRIS, G.P., BUKATA, R.P. and BRUTON, J.E. (1976) Satellite Observations of Water Quality, Transportation Engineering Journal, Proc. American Society of Civil Engineers, 102, TE3, 537-554.

HERSCHY, R.W. (1980) Hydrological data collection by satellite, Proc. Instn. Civ. Engrs. Part 1, 68, 759-771.

HILL, J.M. and GRAHAM, D.S. (1980) Using Enhanced Landsat Images for Calibrating Real-Time Estuarine Water-Quality Models, Water Quality Bulletin, Satellite Hydrology, 5, 1, 20-23.

HILTON, J. (1984) Airborne Remote Sensing for Freshwater and Estuarine Monitoring, Water Research, 18, 10, 1195-1223.

HOLMQUIST, K.W. (1977) The LANDSAT Lake Eutrophication Study, Wisconsin University, Madison, MSc. Thesis, Department of Civil and Environmental Engineering, PB-290678, 1-94.

HOTELLING, H. (1936) Relations Between Two Sets Of Variates, Biometrika, 28, 321-377.

HOUSE OF LORDS SELECT COMMITTEE ON SCIENCE AND TECHNOLOGY (1983) Remote Sensing and Digital Mapping, 1, Her Majesty's Stationery Office, London.

HOWMAN, A.M. and KEMPSTER, P.L. (1983(a)) Surface Reference Date Collection, Internal Progress Report No. 1, Hydrological Research Institute, Department of Water Affairs, Pretoria, RSA.

HOWMAN, A.M. and KEMPSTER, P.L. (1983(b)) Landsat-Water Quality Surveillance of Roodeplaat Dam, EDIS 83 - Earth Data Information Systems Symposium, 19-20 September, Pretoria, RSA.

HOWMAN, A.M. (1984) The Story of the Computer Compatible Tapes for the Landsat Water Quality Project, Internal Progress Report No. 3, Hydrological Research Institute, Department of Water Affairs, Pretoria, RSA.

IWANSKI, M.L., HIGGINS, J.M., KIM, B.R. and YOUNG, R.C. (1980) Factors Affecting Water Quality in Cherokee Reservoir, Tennessee Valley Authority, Division of Water Resources, Chattanooga, Tennessee, 37401, 63-67.

JACKSON, T.J. and RAGAN, R.M. (1977) Value of Landsat in Urban Water Resources Planning, Journal of the Water Resources Planning and Management Division, Proc. of the American Society of Civil Engineers, 103, WR1, 33-46.

JARMAN, J.W. (1973) The Earth Resources Program of the Corps of Engineers, Symposium of Significant Results Obtained From Earth Resources Technology Satellite-1, 2, Summary of Results, NASA, Greenbelt, Maryland, 127-141.

KENDRICK, P.J. (1976) Remote Sensing and Water Quality, Journal WPCF, 48, 10, 2243-2246.

KENNEY, J.K. and KEEPING, E.S. (1954) Mathematics of Statistics, Part One, Third Edition, D. Van Nostrand Company Inc, Princetown, New Jersey, 252-285.

KHORRAM, S. (1981) Water quality mapping from Landsat digital data, International Journal of Remote Sensing, 2, 2, 145-153.

KHORRAM, S. and CHESHIRE, H.M. (1983) The Use of Landsat MSS Digital Data in Water Quality Mapping of the Neuse River Estuary, N.C., Water Resources Research Institute of the University of North Carolina, Report No. 193, Raleigh, North Carolina, 1-29.

KLEMAS, V. (1976) Remote Sensing of Coastal Wetland Vegetation and Estuarine Water Properties, Estuarine Processes, 2, Academic Press, New York, 381-403.

KRITIKOS, H., YORINKS, L. and SMITH, H. (1974) Suspended Solids Analysis Using ERTS-A Data, Remote Sensing of Environment, 3, 69-78.

KUO, C.Y. and BLAIR, C.H. (1976) Rational method for correcting sea truth of suspended sediment concentration related to remote sensing, Technical Report 76-C3, Old Dominion University, Norfolk, V.A.

LAMBOU, V.W., HERN, S.C., TAYLOR, W.D. and WILLIAMS, L.R. (1982) Chlorophyll, Phosphorus, Secchi Disk and Trophic State, Water Resources Bulletin, 18, 5, 807-813.

LATIN, H.A., TANNEHILL, G.W. and WHITE, R.E. (1976) Remote Sensing Evidence and Environmental Law, California Law Review, 64, 6, 1300-1446.

LeCROY, S.R. (1982) Kerr Reservoir Landsat Experiment Analysis for November 1980, Kentron Technical Center, Hampton, Virginia, NASA Contractor Report 165924.

LEKAN, J.F. and CONEY, T.A. (1982) The Use of Remote Sensing to Map the Areal Distribution of Cladophora Glomerata at a Site in Lake Huron, Journal of Great Lakes Research, 8, 1, 144-152.

LIEBENBERG, E.C. (1977) Geography GGR. 103/001 Guide 1, Department of Geography, University of South Africa, Pretoria.

LILLESAND, T. M. and KIEFER, R.W. (1979) Remote Sensing and Image Interpretation, John Wiley and Sons Ltd. New York.

LILLESAND, T.M., JOHNSON, W.L., DEUELL, R.L., LINDSTROM, O.M and MEISNER, D.E. (1983) Use of Landsat Data to Predict the Trophic State of Minnesota Lakes, Photogrammetric Engineering and Remote Sensing, 49, 2, 219-229.

LINDELL, L.T. (1981) Experiences from correlations of Landsat data versus transmission of light and chlorophyll a, Verh. Internat. Verein. Limnol., 21, 1, 438-441.

LINS, H.F. (1979) Some Legal Considerations in Remote Sensing, Photogrammetric Engineering and Remote Sensing, 45, 6, 741-748.

MAAREN, H. (1981) Digital Processing Landsat MSS Data, Operational Manual, Progress Report No 14, Hydrological Research Institute, Department of Water Affairs, Pretoria, RSA.

MACE, T.H. (1983) Use of Archival Landsat and Near-Shore Sampling Data for Water Quality Mapping in Lake Ontario, EPA 63-03-3049, 1-10.

MacFARLANE, N. and ROBINSON, I.S. (1984) Atmospheric correction of LANDSAT MSS data for a multidate suspended sediment algorithm, Int. J. of Remote Sensing, 5, 3, 561-576.

MALAN, O.G. (1976) Oë in die Ruim, Scientiae, 17, 4.

McCLUNEY, W.R. (1974) Estimation of sunlight penetration in the sea for remote sensing, Technical Memorandum X-70643, National Aeronautics and Space Administration.

McGINNIS, D.F., SCOFIELD, R.A., SCHNEIDER, S.R. and BERG, C.P. (1980) Satellites as Aid to Water Resources Managers, Journal of the Water Resources Planning and Management Division, Proc. of the American Society of Civil Engineers, 106, WR1, 1-19.

McHENRY, J.R., RITCHIE, J.C. and SCHIEBE, F.R. (1976) Estimating the Suspended Sediment Load in Reservoirs, Water Resources Bulletin, 12, 6, 81-92.

MOORE, G.K. (1980) Satellite remote sensing of water turbidity, Hydrological Sciences Bulletin, 25, 4, 407-421.

MORGAN, E.L., EAGLESON, K.W., HERMANN, R. and McCOLLOUGH, N.D. (1981) New Developments in Automated Biosensing from Remote Water Quality Stations and Satellite Data Retrieval for Resources Management, Journal of Hydrology, 51, 339-345.

MUNDAY, J.C., ALFÖLDI, T.T. and AMOS, C.L. (1979) Bay of Fundy Verification of a System for Multidate Landsat Measurement of Suspended Sediment, Satellite Hydrology, (Eds. Deutsch, M., Wiesnet, D.R., Rango, A.), Proc. of the Fifth Annual William T. Pecora Memorial Symposium on Remote Sensing, Sioux Falls, South Dakota, June 10-15, American Water Resources Association, Minneapolis, Minnesota, 622-640.

MUNDAY, J.C., ALFÖLDI, T.T. and AMOS, C.L. (1980) Application of a System for Automated Multidate Landsat Measurement of Suspended Sediment, Water Quality Bulletin, Satellite Hydrology, 5, 1, 6-10.

MURALIKRISHNA, I.V. and RAO, K.R. (1982) Landsat Application for Studying Aquatic Suspended Sediment Patchiness, Journal of the Institution of Engineers, India, 62, EN 3. 115-116.

NASH, J.E. and SUTCLIFFE, J.V. (1970) River Flow Forecasting Through Conceptual Models. Part 1 - A Discussion of Principles, Journal of Hydrology, 10, 282-290.

NIE, N.H., HADLAI HULL, C., JENKINS, J.G., STEINBRENNER, K., BENT D.H.(1975) SPSS Statistical Package for the Social Sciences, 2nd Ed, McGraw Hill Inc.

O'DONOGHUE, D., FORBES, A.M. and PIPER, S.E. (1983) The P.I.P.S. Manual, Department of Surveying and Mapping, University of Natal, Unpublished manuscript, Durban, R.S.A.

PAULSON, R.W. (1974) A test of an experimental polar- orbiting satellite system for relaying hydrological data, International Association of Hydrological Sciences Publication No. 112, Proceedings of the Paris Symposium, September, 23-28.

PALMER, D.W. (1981) Introductory Notes, Course in Remote Sensing, Department of Surveying and Mapping, University of Natal, Durban, R.S.A.

PIETERSE, A.J.H., and BRUWER, C.A. (1980) Roodeplaat Dam in Limnology of some Selected South African Impoundments, Water Research Commission and National Institute for Water Research (CSIR), Pretoria, R.S.A., 119-129.

PIPER, S.E. (1981) Introductory Notes, Course in Remote Sensing, Department of Surveying and Mapping, University of Natal, Durban, R.S.A.

POWLING, I.J. (1977) Algae in Farm Water Supplies, Water Talk, 39, State Rivers and Water Supply Commission, Australia, 13-15.

RANGO, A., FELDMAN, A., GEORGE III, T.S. and RAGAN, R.M. (1983) Effective use of Landsat Data in Hydrologic Models, Water Resources Bulletin, 19, 2, 165-174.

REED, G.D. (1978) Water Characteristics, Journal WPCF, 50, 6, 1022-1025.

RITCHIE, J.C., SCHIEBE, F.R. and McHENRY, J.R. (1976) Remote Sensing of Suspended Sediments in Surface Waters, Photogrammetric Engineering and Remote Sensing, 42, 12, 1539-1545.

ROBERTS, P.J.T. (1978) A comparison of the performance of selected conceptual models of the rainfall-runoff process in semi-arid catchments near Grahamstown, Report No. 1/78, Hydrological Research Unit, Department of Geography, Rhodes University, Ph.D Thesis.

RODDA, J.C. (1976) Facets of Hydrology, John Wiley and Sons, New York.

ROGERS, R.H. et al (1975) Application of Landsat to the surveillance and control of eutrophication in Saginaw Bay, Proc. of the Tenth International Symposium on Remote Sensing of Environment, ERIM, Ann. Arbor, MI, 437-446.

SARTORY, D.P. (1982) Spectrophotometric analysis of chlorophyll a in freshwater phytoplankton, Department of Environment Affairs, Technical Report TR115, Pretoria, RSA, 1-163.

SAYERS, W.T. (1971) Water Quality Surveillance, Environmental Science and Technology, 5, 2, 114-119.

SCARPACE, F.L., HOLMQUIST, K.W. and FISHER, L.T. (1979) Landsat Analysis of Lake Quality, Photogrammetric Engineering and Remote Sensing, 45, 5, 623-633.

SCHAEFFER, D.J., CLARKE, R.P., SEFTON, D.F. and BOLAND, D.H.P. (1979) Assessment and Classification of Selected Illinois Lakes through the Application of Space Technology, Satellite Hydrology, (Eds. Deutsch, M., Wiesnet, D.R., Rango, A.), Proc. of the Fifth Annual William T. Pecora Memorial Symposium on Remote Sensing, Sioux Falls, South Dakota, June 10-15, American Water Resources Association, Minneapolis, Minnesota, 641-648.

SHENG, Y.P. and LICK, W. (1979) The Transport and Resuspension of Sediments in a Shallow Lake, Journal of Geophysical Research, 84, C4, 1809-1826.

SHIH, S.F. and GERVIN, J.C. (1980) Ridge Regression Techniques Applied to Landsat Investigation of Water Quality in Lake Okeechobee, Water Resources Bulletin, 16, 5, 790-796.

SHIMODA, H., ETAYA, M., SAKATA, T., GODA, L. and STELCZER, K. (1984) Water Quality Monitoring of Lake Balaton Using Landsat MSS Data, Proc. of the Eighteenth International Symposium of Remote Sensing of Environment, Paris, France, Oct. 1-5.

SKIBITZKE, H.E. (1976) Remote Sensing for Water Resources, Remote Sensing of Environment, (Eds. Lintz, J. and Simonett, D.S.), Addison Wesley Publishing Company Inc, 572-592.

STORTZ, K., CLAPPER, R.and SYDOR, M. (1976) Turbidity Sources in Lake Superior, Journal of Great Lakes Research, 2, 2, 393-401.

SYDOR, M., STORTZ, K.R. and SWAIN, W.R. (1978) Identification of Contaminants in Lake Superior Through Landsat 1 Data, Journal of Great Lakes Research, 4, 2, 142-148.

THIRUVENGADACHARI, S., SUBBA RAO, P. and RAMACHANDRA RAO, K. (1980) Surface Water Inventory Through Satellite Sensing, Journal of the Water Resources Planning and Management Division, Proc. of the American Society of Civil Engineers, 106, WR2, 493-502.

THIRUVENGADACHARI, S., INAMDAR, N.G., TAMBRALLI, H.G., KOTBAGI, S.V. and DEEKSHATULU, B.L. (1983) Some ground truth considerations in inland water quality surveys, International Journal of Remote Sensing, 4, 3, 537-544.

THORNTON, K.W., KENNEDY, R.H., MAGOUN, A.D. and SAUL, G.E. (1982) Reservoir Water Quality Sampling Design, Water Resources Bulletin, 18, 3, 471-480.

TOERIEN, D.F. (1975) South African eutrophication problems: A perspective, Wat. Pollut. Contr., 74, 134-142.

TOERIEN, D.F. (1977) A review of eutrophication and guidelines for its control in South Africa, CSIR Special Report WAT 48, Council for Scientific and Industrial Research, Pretoria, R.S.A.

TRUTER, E. (1981) Manual for Biological Analysis of Freshwater Samples, Task 1, Chemical/Biological Analytical Services Project N3/0501 (not published), Department of Water Affairs, Pretoria, R.S.A.

ULBRICHT, K.A. (1983) LANDSAT image of blue green algae in the Baltic sea, International Journal of Remote Sensing, 4, 4, 801-802.

VAN BELLE, G. and HUGHES, J.P. (1983) Monitoring for water quality: fixed stations versus intensive surveys, Journal WPCF, 55, 4, 400-404.

VAN BELLE, G. and HUGHES, J.P. (1984) Nonparametric Tests for Trend in Water Quality, Water Resources Research, 20, 1, 127-136.

VERDIN, J.P. (1985) Monitoring Water Quality Conditions in a Large Western Reservoir with Landsat Imagery, Photogrammetric Engineering and Remote Sensing, 51, 3, 343-353.

WAINWRIGHT, S.J. and GILBERT, R.I. (1981) A BASIC computer program which tests for normality and the presence of outliers in data, Laboratory Practice, 30, 5, 467.

WALMSLEY, R.D. (1984) A Chlorophyll a Trophic Status Classification System for South African Impoundments, Journal of Environmental Quality, 13, 1, 97-104.

WALMSLEY, R.D., and BUTTY, M. (1979) Eutrophication of Rivers and Dams, Part 6. An investigation of the chlorophyll-nutrient relationships for 21 South African impoundments, Contract Report, Water Research Commission, Pretoria, South Africa.

WELBY, C.W., WITHERSPOON, A.M. and HOLMAN III, R.E. (1980) Trophic State Determination for Shallow Coastal Lakes from Landsat Imagery, Water Quality Bulletin, Satellite Hydrology, 5, 1, 11-14, 24-25.

WETZEL, R.G. (1983) Limnology, 2nd Edition, Saunders College Publishing.

WHITLOCK, C.H., KUO, C.Y. and LeCROY, S.R. (1982) Criteria for the Use of Regression Analysis for Remote Sensing of Sediment and Pollutants, Remote Sensing of Environment, 12, 151-168.

WHITLOCK, C.H., WITTE, W.G., USRY, J.W and GURGANUS, E.A. (1978) Penetration Depth at Green Wavelengths in Turbid Waters, Photogrammetric Engineering and Remote Sensing, 44, 11, 1405-1410.

WHITMORE, J.S. (1978) South Africa's Water Potential - A Review, The Road Ahead, International Futures Conference, July.

WITZIG, A.S., and WHITEHURST, C.A. (1981) Current Use and Technology of Landsat MSS Data for Lake Trophic Classification, Water Resources Bulletin, 17, 6, 962-970.

YARGER, H.L., McCAULEY, J.R., JAMES, G.W., MAGNUSON, L.M. and MARZOLF, G.R. (1973) Water Turbidity Detection Using ERTS-1 Imagery, Proc. of Symposium on Significant Results obtained from the Earth Resources Technology Satellite-1. 1, Technical Presentations Paper E12, NASA Contract NAS5 -21822, Greenbelt, Maryland, 651-658.

YENTSCH, C.S. and PHINNEY, D.A. (1982) The use of the attenuation of light by particulate matter for the estimate of phytoplankton chlorophyll with reference to the Coastal Zone Colour Scanner, Journal of Plankton Research, 4, 1, 93-102.

YOUNG, S. and SILBERBAUER, M.J. (1984) Seasonality in Roodeplaat Dam, Poster presented at the Perspectives in Southern Hemisphere Limnology Symposium, Wilderness, South Africa, 3 to 13 July.

APPENDIX A: ANALYSIS OF CHLOROPHYLL *a* IN FRESHWATER PHYTOPLANKTON*

1. Apparatus

(a) Filter Apparatus e.g. Millipore vacuum/pressure pump 115 v 50 Hz (xx60 110 50) with a 6 place filter holder manifold (xx25 047 00)

(b) Spectrophotometer: e.g. Varian Techtron UV-Vis Model 635.

(c) Water bath with contact thermometer.

(d) Centrifuge.

(e) Test tubes with screw caps.

(f) Centrifuge tubes with caps.

(g) Glass-fibre filter membranes: e.g. Sartorius SM 134000 or Whatman GF/C.

2. Method

(i) Filter a known volume of sample through a glass-fibre filter, allow to suck dry.

(ii) Roll up filter with the entrapped algae and place in a screw capped test tube.

(iii) Add 9.8 mℓ 91.8% ethanol. As the glass-fibre filters retain on average 0.2 mℓ of water, this gives a final concentration of 10 mℓ 90% ethanol. Mark the final volume level.

(iv) Place in water bath at 78°C and allow to boil for 5 min. Make sure that the screw caps are not too loose as the ethanol will evaporate off. If any loss is noted after boiling, make up to volume mark with 90% ethanol.

(v) Allow to stand in the dark at room temperature for 1 h to 24 hours. If room temperature is high (> 30°C) place in a refrigerator.

(vi) After extraction decant extract into a centrifuge tube and cap. The tube must be capped as ethanol will evaporate from an open tube during centrifugation.

(vii) Centrifuge at 4 000 rpm for 5 min.

(viii) Decant 4 mℓ of the sample into a 1 cm pathlength spectrophotometer cuvette.

(ix) Read the absorbance at 665 nm and 750 nm, using 90% ethanol as the reference blank.

* Truter, 1981.

(x) Add 100 mℓ of a 0.3 moles/ℓ HCL solution, shake well and allow to stand for 2 min.

(xi) Reread the "665 nm" absorbance, scanning for the Absorbance peak between 665 nm and 666.5 nm. Reread the absorbance at 750 nm.

3. Calculation

(i) Values from step 9 are $D_{665}b$ and $D_{750}b$

Values from step 11 are $D_{665}a$ and $D_{750}a$

(ii) Subtract the 750 nm readings from the 665 nm readings.

e.g. $D_{665}b - D_{750}b = E_{665}b$

$D_{665}a - D_{750}a = E_{665}a$.

(iii) Insert values into the formula:

$$\text{chlorophyll } \underline{a} \text{ (mg/ℓ extract)} = \frac{(E_{665}b - E_{665}a)\times(R/R-1)\times K}{L}$$

where (R/R-1) = 2.39 (R = the "acid factor" 1.72)

K = 11.99 (specific absorption coefficient of chlorophyll a in 90% ethanol = 83.4)

L = Pathlength of cuvette in centimeters (= 1 cm).

The Equation is thus:

Chlorophyll a (mg/ℓ extract) = $(E_{665}b - E_{665}a)\times 28.66$

If 10 cm^3 extract is used, then for the final answer multiply the mg/ℓ by the following amounts depending on the original volume filtered. The final answer is in µg/ℓ.

Formula: $$\frac{C_a \times V}{V_1}$$

where

C_a = Concentration of chlorophyll a in mg/ℓ in the extract.

V = Volume of extract in mℓ.

V_1 = Volume of sample filtered in litres.

APPENDIX B: LANDSAT/WATER QUALITY SURFACE REFERENCE DATA SAMPLING FORM

SITE 1 – 10

DATE 11 – 16

SUNSHINE CONDITIONS 17

1	CLEAR
2	MEDIUM
3	OVERCAST

TIME 18 – 22

SAMPLE POINT 23 24

SAMPLE A

SURFACE CHLOROPHYLL
μg/ℓ — 25 – 29

INTEGRATED CHLOROPHYLL
μg/ℓ — 30 – 34

SECCHI DISC
m — 35 – 38

SURFACE TURBIDITY
NTU — 39 – 42

INTEGRATED TURBIDITY
NTU — 43 – 46

WIND SPEED
m/sec — 47 – 50

WIND DIRECTION 51 52

AIR TEMPERATURE
°C — 53 – 56

CARD NUMBER 80

APPENDIX C: SURFACE REFERENCE DATA COLLECTED FOR ROODEPLAAT DAM, CONCURRENTLY WITH THE SATELLITE OVERFLIGHTS

81.10.14
81.11.01
81.12.07
82.09.13
82.09.30
82.11.16

SURFACE REFERENCE DATA, MEAN OF DUPLICATES 81.10.14

SAMPLING POINT NO.	SURFACE CHLOROPHYLL a µg/ℓ	INTEGRATED CHLOROPHYLL a µg/ℓ	SURFACE TURBIDITY NTU	INTEGRATED TURBIDITY NTU
1	25,3	33,0	3,0	3,6
2	28,0	34,1	3,4	3,9
3	22,8	29,3	3,1	3,5
4	33,3	39,0	3,9	3,9
5	32,2	34,7	4,2	4,6
6	32,8	39,3	3,2	3,5
7	26,1	26,7	3,3	4,0
8	29,3	28,8	3,6	3,6
9	21,2	27,5	3,3	3,8
10	24,9	32,1	3,7	5,8
11	30,4	30,7	3,7	4,0
12	29,9	30,4	5,9	5,2
13	29,3	33,9	5,6	6,8
14	26,9	20,5	5,4	5,3
15	27,2	32,7	5,4	5,9
16	22,8	25,3	5,9	6,2
17	22,6	24,7	5,5	6,1
18	42,0	18,3	7,3	6,0
19	29,0	22,4	8,0	5,8
20	27,7	19,6	6,7	7,8
21	23,0	20,6	7,0	6,4
22	27,4	24,0	6,3	6,9
23	24,8	24,9	6,6	6,8
24	33,4	25,2	6,3	7,2
25	28,0	27,9	6,6	7,5
26	25,3	27,7	7,8	8,8
27	33,4	23,4	6,0	6,6
28	33,7	30,9	6,8	7,3
29	107,6	82,0	10,9	13,5
30	23,1	31,0	5,5	6,1
31	30,2	31,3	3,8	4,4
32	26,1	30,4	3,8	3,8

SURFACE REFERENCE DATA, MEAN OF DUPLICATES 81.11.01

SAMPLING POINT NO.	SURFACE CHLOROPHYLL a µg/ℓ	INTEGRATED CHLOROPHYLL a µg/ℓ	SURFACE TURBIDITY NTU	INTEGRATED TURBIDITY NTU
1	16,8	32,7	3,3	4,2
2	39,4	35,5	4,5	4,5
3	29,7	32,0	4,0	4,2
4	33,8	32,0	4,0	4,4
5	34,7	30,6	4,9	5,4
6	21,9	35,3	4,0	4,9
7	84,2	26,9	17,2	4,1
8	18,7	25,7	4,0	4,8
9	18,3	18,9	3,8	4,3
10	22,8	23,8	4,1	4,7
11	26,1	25,4	4,1	4,2
12	24,4	28,9	4,3	4,2
13	62,2	40,7	10,4	6,1
14	37,7	41,0	6,3	7,7
15	36,3	40,4	7,4	7,0
16	37,5	35,0	6,7	7,8
17	45,7	42,2	7,9	7,5
18	29,8	39,6	6,3	6,8
19	47,4	37,1	8,2	7,1
20	44,5	45,9	7,8	8,3
21	52,5	41,0	10,2	7,5
22	29,1	40,0	7,7	10,0
23	32,8	46,7	7,8	8,8
24	27,9	29,9	7,4	11,2
25	22,8	38,1	6,8	9,1
26	29,3	42,4	8,9	10,4
27	33,6	35,5	8,5	9,7
28	31,4	40,8	9,8	12,5
29	33,5	31,8	10,5	12,5
30	107,8	46,7	18,0	6,7
31	18,2	27,7	3,1	3,5
32	20,1	26,7	3,6	3,9

SURFACE REFERENCE DATA, MEAN OF DUPLICATES 81.12.07

SAMPLING POINT NO.	SURFACE CHLOROPHYLL a µg/ℓ	INTEGRATED CHLOROPHYLL a µg/ℓ	SURFACE TURBIDITY NTU	INTEGRATED TURBIDITY NTU
1	7,6	8,6	1,6	1,2
2	8,3	9,6	1,7	2,0
3	9,0	6,0	1,9	1,6
6	10,0	9,2	1,4	2,1
7	9,5	9,7	1,9	1,4
8	9,5	8,9	2,2	2,0
11	6,7	2,9	0,9	2,0
12	3,3	3,6	1,2	1,7
13	8,6	22,8	2,5	3,1
14	16,6	10,6	3,1	3,2
15	9,7	19,9	3,1	3,1
16	14,3	10,6	4,0	4,2
17	20,6	14,2	3,2	2,9
18	10,4	5,0	3,8	3,4
19	20,9	17,4	4,4	3,9
20	20,3	11,7	5,0	4,4
21	17,9	16,9	3,6	3,6
22	30,6	30,5	8,0	6,7
23	3,1	29,5	5,8	6,2
24	33,1	35,8	8,1	8,3
25	35,4	26,6	9,0	8,9
26	30,5	30,1	15,7	14,5
27	37,0	38,3	12,0	11,7
28	45,0	43,2	13,5	13,5
29	68,2	68,1	20,0	18,5
30	7,6	8,4	1,4	2,3
31	7,3	5,4	1,0	1,1
32	8,6	8,6	1,4	1,0

SURFACE REFERENCE DATA 82.09.13

SAMPLING POINT NO.	SURFACE CHLOROPHYLL a µg/ℓ	INTEGRATED CHLOROPHYLL a µg/ℓ	SURFACE TURBIDITY NTU	INTEGRATED TURBIDITY NTU
1	15,2	17,5	4,3	4,9
2	18,1	17,9	4,2	4,9
3	15,0	16,1	4,0	4,7
4	21,8	22,6	4,4	5,2
5	25,4	29,7	5,4	6,2
6	16,5	13,6	4,3	4,1
7	14,0	12,9	3,7	4,1
8	13,9	10,7	4,3	3,7
9	12,3	12,7	3,6	3,8
10	11,9	12,7	3,7	3,7
11	11,5	10,7	3,6	4,4
12	13,1	14,3	3,9	4,5
13	16,4	14,3	4,1	4,4
15	22,9	18,8	4,9	5,4
16	20,9	24,2	5,0	5,5
17	21,3	18,8	4,4	4,6
18	18,8	17,2	4,5	5,3
19	16,8	22,5	5,3	5,5
20	20,1	22,9	5,5	6,2
21	22,9	21,7	5,0	5,8
22	24,2	12,3	6,1	6,7
23	22,5	20,1	6,7	6,8
24	27,0	23,4	6,7	7,4
25	27,0	27,5	6,8	6,9
26	29,5	29,5	7,2	7,7
27	35,7	26,6	7,1	7,4
28	32,8	45,5	7,7	8,0
29	55,3	50,4	17,0	21,0
30	11,5	11,9	4,2	4,2
31	13,1	12,7	3,9	4,5
32	12,7	9,8	3,6	4,0
34	22,5	25,4	8,1	8,0
35	30,3	27,9	7,1	7,5
36	24,6	24,2	6,9	7,5
37	25,8	25,0	6,6	7,6
38	25,4	24,3	12,0	6,6
39	27,0	23,8	4,0	6,9
40	25,0	23,4	4,2	6,3
41	22,1	20,9	6,0	5,4
42	16,8	22,1	5,0	5,5
43	24,6	23,4	3,7	4,5
44	20,9	22,1	5,5	4,3
45	20,9	25,0	4,2	4,5
46	21,7	19,7	5,0	5,2
47	21,3	23,8	4,3	5,5
48	18,8	20,9	3,5	5,2
49	17,2	17,6	4,0	4,8
50	16,8	17,6	4,4	4,0
51	13,5	16,8	3,4	4,3
52	16,0	16,0	3,6	4,5
53	16,0	13,1	3,7	3,4
54	14,3	15,6	3,3	4,3
55	16,0	16,8	3,6	3,8

SURFACE REFERENCE DATA 82.09.30

SAMPLING POINT NO.	SURFACE CHLOROPHYLL a µg/ℓ	INTEGRATED CHLOROPHYLL a µg/ℓ	SURFACE TURBIDITY NTU	INTEGRATED TURBIDITY NTU
1	10,7	13,3	4,0	3,4
2	11,5	12,9	3,7	4,0
3	11,5	12,2	3,3	3,5
4	11,1	11,5	3,2	3,5
5	14,0	17,9	3,6	4,5
6	10,7	12,9	3,6	3,7
7	10,7	13,3	3,1	3,3
8	10,7	12,5	3,2	3,3
9	9,3	14,0	3,2	3,4
10	12,9	14,0	3,8	3,9
11	12,5	14,3	3,6	3,5
12	17,2	17,9	4,2	4,6
13	15,8	17,9	4,2	4,5
14	17,5	20,1	4,1	4,2
15	17,2	20,1	4,0	4,3
16	22,9	23,3	4,7	4,7
17	18,3	25,8	4,2	4,9
18	25,1	27,2	4,7	5,0
19	17,2	22,6	4,3	4,7
20	33,7	39,4	5,3	5,8
21	29,7	30,8	5,1	5,5
22	39,8	41,9	6,4	7,4
23	40,5	43,0	7,8	7,3
24	42,6	41,2	6,7	7,5
25	45.9	44,1	8,2	8,5
26	51,6	49,3	8,7	9,2
27	50,4	57,3	12,0	12,0
28	42,4	40,1	8,0	8,7
29	82,0	111,2	18,0	20,0
30	16,8	18,9	4,0	4,5
31	15,0	14,0	3,4	3,5
32	9,3	12,5	3,4	4,1
33	57,0	40,5	9,7	9,7
34	36,5	35,8	8,0	7,8
35	34,0	41,9	7,7	8,1
36	26,3	34,5	7,0	7,7
37	35,4	39,2	7,2	7,0
38	35,9	39,2	7,0	7,2
39	42,6	44,5	7,2	7,0
40	33,0	37,8	6,7	6,6
41	33,0	36,4	6,0	6,0
42	30,6	33,4	6,2	6,1
43	33,0	33,5	5,9	6,3
44	27,3	32,1	5,6	5,6
45	21,5	33,0	5,4	5,7
46	20,6	28,7	5,0	5,0
47	21,5	27,3	4,4	5,0
48	18,7	27,3	4,4	4,9
49	16,3	23,0	4,1	4,4
50	19,1	18,7	3,7	4,3
51	15,6	20,1	3,6	3,6
52	13,9	20,6	3,7	4,0
53	16,7	18,7	3,9	4,4
54	16,7	17,2	4,3	4,4
55	19,6	21,1	3,7	4,0

SURFACE REFERENCE DATA 82.11.16

SAMPLING POINT NO.	SURFACE CHLOROPHYLL a µg/ℓ	INTEGRATED CHLOROPHYLL a µg/ℓ	SURFACE TURBIDITY NTU	INTEGRATED TURBIDITY NTU
1	9,7	11,8	3,0	3,1
2	12,6	14,3	3,4	3,9
3	13,2	13,2	3,6	3,6
4	11,8	11,2	3,4	3,6
5	13,2	12,6	3,1	4,0
6	13,8	14,6	4,1	4,6
7	14,9	13,5	3,6	4,1
8	16,6	18,1	4,2	4,4
9	13,2	10,9	3,1	3,4
10	14,0	14,0	3,8	3,9
11	12,9	11,8	4,2	4,4
12	14,0	13,2	4,4	4,5
13	19,5	18,9	4,6	4,7
14	20,1	19,2	5,2	5,5
15	21,5	22,1	5,0	5,5
16	24,1	21,2	6,0	6,1
17	23,5	24,7	5,5	6,4
18	21,3	20,7	6,0	6,2
19	22,1	21,8	5,0	5,6
20	27,4	26,1	6,4	6,5
21	32,1	30,1	5,5	6,5
22	33,0	32,2	8,3	8,0
23	33,7	33,7	7,3	7,4
24	48,3	45,5	9,8	13,0
25	42,2	40,6	7,5	7,8
26	34,8	38,1	13,0	13,5
27	57,4	56,0	8,4	9,5
28	121,0	114,6	15,5	17,0
29	369,7	325,3	28,0	31,0
30	17,2	16,8	4,0	4,5
31	13,0	13,5	3,9	4,0
32	14,0	12,6	17,0	18,0
33	143,2	129,6	18,0	20,0
34	45,9	45,9	8,7	9,3
35	49,6	48,0	8,6	8,8
36	35,7	38,1	7,3	7,6
37	31,5	34,4	6,6	7,3
38	30,8	33,7	5,9	7,3
39	33,0	30,4	6,5	7,0
40	33,7	25,4	6,8	7,6
41	29,3	31,5	6,4	6,7
42	30,8	30,1	6,4	7,0
43	33,0	33,1	7,1	7,1
44	29,7	29,7	7,2	6,5
45	29,6	27,4	5,3	5,4
46	27,0	24,8	4,9	5,4
47	27,4	27,4	4,9	5,2
48	25,8	25,1	5,0	5,2
49	25,1	25,1	5,2	5,8
50	25,2	23,8	5,2	5,5
51	24,1	22,9	5,2	5,4
52	22,9	22,9	5,2	5,5
53	23,5	24,4	4,9	5,6
54	23,5	28,4	4,6	5,8
55	23,5	24,1	5,4	5,6

APPENDIX D: THE ALIGNMENT OF REFLECTANCE DATA WITH SURFACE REFERENCE DATA

Sampling sites were identified on the impoundment using suitable landmarks, which in turn, could be recognised on the satellite image. In order to identify the pixel corresponding to the sampling position three methods were used. The first method named 'Chance a Pixel' simply meant pinpointing the sampling point on a map of the impoundment (Figure D.1). This technique required a good knowledge of the research area. Although fairly accurate it was decided to improve on the reliability by using Method 2.

Method 2, or 'Average Pixel', involved using the pinpointed pixel of Method 1 and averaging its value together with those of the surrounding pixels (Figure D.1). Method 3 or 'Weighted Pixel' used the 'Chance a Pixel' (Method 1) and weighted it and the surrounding eight values in the following manner. The 'Chance a Pixel' was given a weighting of 4, all crosswise pixels were weighted by 2 and diagonal pixels by 1. The sum of the pixel values was divided by the sum of all the weights of the pixel values used in the summation. The program which carries out this task is given in Appendix E and an example of the results is given in Appendix F.

It is important to note here that due to the fact that pixel reflectance values vary with individual wave bands and because of the possibility that some sampling points may be positioned near land areas, band 7 values were used to determine the water/land value limit. The computer program was written in such a way that the water/land boundary as determined by band 7 would set the limits for the remaining 3 bands. Any values falling outside the limit would not be included in the estimation and the weightings would be affected accordingly.

In order to determine which of the three methods was best, 'Chance a Pixel', 'Average Pixel' and 'Weighted Pixel' values for two different data sets (81.10.14 and 81.12.07) were examined using the Canonical Correlation Analysis (see Section 3.4.6). The results are shown in Table D.1.

Table D.1. indicated that the 'Weighted Pixel' method produced the best overall results and the decision was made to use the 'Weighted Pixel' method only for subsequent analysis.

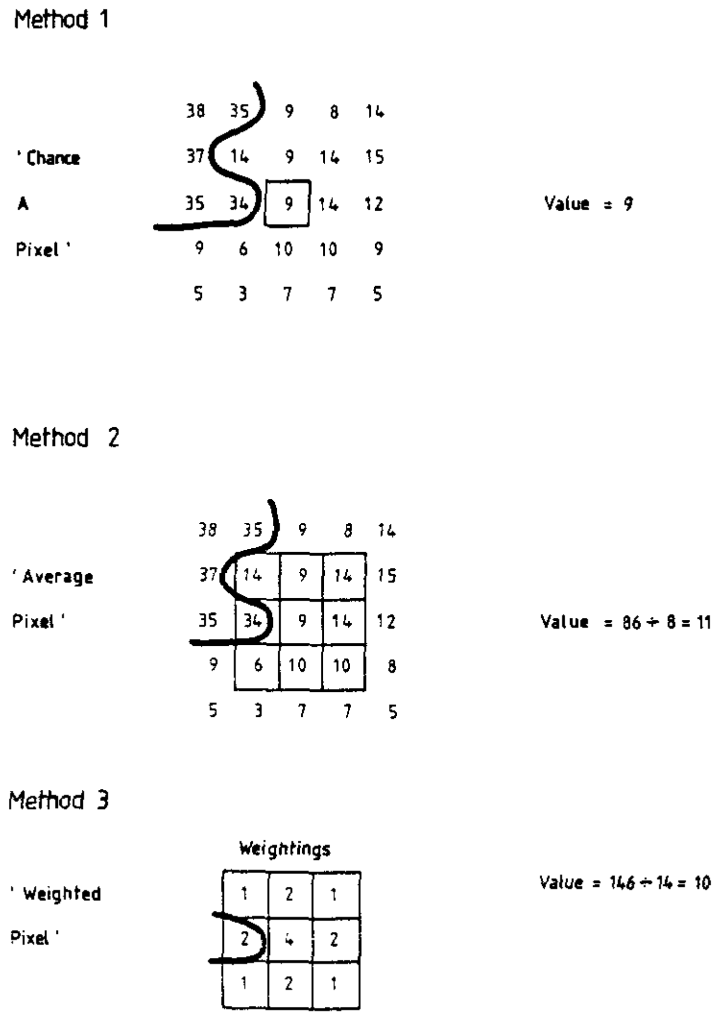


FIGURE D.1: THREE DIFFERENT METHODS USED TO DETERMINE PIXEL REFLECTANCE VALUES FOR THE SAMPLING SITES

TABLE D.1: CANONICAL CORRELATIONS (r) OBTAINED FOR ROODEPLAAT DAM, USING THREE DIFFERENT PIXEL ALIGNMENT METHODS.

Variables	'Chance a Pixel'	'Average Pixel'	'Weighted Pixel'	Date
Surface chlorophyll a and	0,76	0,82	0,84	81.10.14
Surface turbidity	0,95	0,93	0,93	81.12.07
Integrated chlorophyll a and	0,77	0,83	0,86	81.10.14
Integrated turbidity	0,95	0,95	0,95	81.12.07

APPENDIX E: SUBROUTINE "WEIGHT" WHICH CALCULATES WEIGHTED MEAN PIXEL VALUES AT SPECIFIC SAMPLING SITES

```
      SUBROUTINE WEIGHT
C
C     THIS SUBROUTINE CALCULATES WEIGHTED MEAN REFLECTANCES
C     AROUND A GIVEN PIXEL OF A AFFINED IMAGE
C
C         THE FIRST IMAGE READ
C         MUST BE BAND 7 DATA............
C
      IMPLICIT INTEGER(A-Z)
      DIMENSION INPUT(1000),ID(100,9,4),NPL(100),NPS(100),
     *NDIVW(100),IP(100,4),NPNN(100),MM(100),IOUT(100),TYPE(4)
      REAL XX,YY,XW
      COMMON/PIC1/NL,NS,BAND,SLDF,SSDF,LLDF,LSDF
      COMMON/PIC2/SL,SS,NLL,NSS,ID1,ID2
      COMMON/FILES/NUMFI(10),NUMFO(10),NEXT,NO,NINT
      COMMON/EXEC/EX,PROCES
      EX=1
C
      DATA NDIVW/100*16/
C
C     INPUT DATA CARDS TAKE ONE POINT PER RECORD;
C     FORMAT IS:29 X,2I4 FOR LINE AND PIXEL.
C
      READ(5,7050) HEAD,(TYPE(J),J=1,4)
 7050 FORMAT(20X,I6,/,20X,4A6)
      WRITE(6,7065) HEAD,TYPE
 7065 FORMAT(20X,'DATE OF ANALYSIS = ROODEPLAAT - ',I6,/,
     *20X,'TYPE OF ANALYSIS - ',4A6,/,/,/)
      READ(5,7040) NLO,NSO,LIM
 7040 FORMAT(3I4)
      WRITE(6,7000) NLO,NSO,LIM
 7000 FORMAT(5X,"PIXEL OFFSET LIMIT = ",I4,/,
     *5X,"SAMPLE OFFSET LIMIT = ",I4,/,
     *5X,"LAND-WATER LIMIT = ",I4,/)
      K=0
  720 K=K +1
      READ(13,7001,END=760) NPOINT,NNL,NNS
      NPL(K)=NNL+NLO
      NPS(K)=NNS+NSO
      NPNN(K)=NPOINT
      GO TO 720
 7001 FORMAT(16X,I2,11X,2I4)
  760 EX=1
      NPNT=K-1
C
C     CHECK THE DISK FILE TO BE OKE
C
      CALL DISKSZ
      NOS= LSDF-SSDF+1
      IF(NOS .GT.1000)CALL PRINT(1,6,24,' DISPLY BUFFER TOO SMALL')
      IF(NS .GT. 1000)CALL PRINT(1,6,26,'*** INPUT BUFFER TOO SMALL')
      NIP=NEXT+NINT
C
C     START READING THE APPROPRIATE REFLECTANCE VALUES
C
      DO 725 N=1,NPNT
      IREC=NPL(N)-SLDF
      ISKIP=NPS(N)-SSDF-1
      DO 732 J=1,9
  732 IOUT(J)=0
      DO 733 J=1,4
  733 MM(J)=0
      DO 726 JF=1,3
      IREC=IREC+1
      IF(NEXT .EQ. 0) GO TO 730
      DO 7010 IN=1,NIP
      IST=(IN-1)*3+1
```

```
      CALL ORIGIN(NUMFI(IN),IREC,ISKIP,3,INPUT(IST))
 7010 CONTINUE
      GO TO 735
      DO 8000 IN=1,NIP
      IST=(IN-1)*3+1
  730 CALL READ(NUMFI(IN),IREC,NOS,INPUT(IST))
      IF(ISKIP .EQ. 0) GO TO 8000
      DO 736 JR=IST,IST+2
      INPUT(JR) = INPUT(JR+ISKIP)
  736 CONTINUE
 8000 CONTINUE
  735 DO 745 I=1,4
      IST= (I-1)*3+1
      DO 740 KN=IST,IST+2
      MM(I)=MM(I)+1
      M=MM(I)
      IF(I.NE.1)GO TO 741
      IF (INPUT(KN).LE.LIM)GO TO 737
      IOUT(M)=1
      GO TO 742
  741 IF (IOUT(M).EQ.0)GO TO 737
      GO TO 734
  742 IF((M.EQ.1).OR.(M.EQ.3).OR.(M.EQ.7).OR.(M.EQ.9))NDIVW(N)=NDIVW
     *(N)-1
      IF((M.EQ.2).OR.(M.EQ.4).OR.(M.EQ.6).OR.(M.EQ.8)) NDIVW(N)=
     *NDIVW(N)-2
      IF(M.EQ.5) NDIVW(N)=NDIVW(N)-4
  734 ID(N,M,I)=0
      GO TO 740
  737 ID(N,M,I)=INPUT(KN)
  740 CONTINUE
  745 CONTINUE
  726 CONTINUE
  725 CONTINUE
C
C     WRITE OUT THE HEADING OF THE OUTPUT PAGE
C
      WRITE(6,7020)
 7020 FORMAT(/5X,'POINT NO. LINE SAMPLE       ACTUAL         WEIGHTED'
     *,/,30X,'     MEAN                 MEAN ',/,25X,'BANDS 7  6  5  4',8X,'
     *7  6  5  4')
      DO 750 K=1,NPNT
      DO 755 I=1,4
      MW=ID(K,1,I)+ID(K,3,I)+ID(K,7,I)+ID(K,9,I)+2*(ID(K,2,I)+ID(K,4,I)
     *+ID(K,6,I)+ID(K,8,I))+4*ID(K,5,I)
      IF (NDIVW(K).EQ.0) GO TO 755
      XX=FLOAT(MW)
      YY=FLOAT(NDIVW(K))
      XW=FLOAT(XX/YY)
      IP(K,I)=IFIX(XW+0.5)
  755 CONTINUE
  756 WRITE(6,7030)NPNN(K),NPL(K),NPS(K),(ID(K,5,I),I=1,4),(IP(K,I),
     *I=1,4)
 7030 FORMAT(/,8X,I3,3X,I4,2X,I4,5X,(4I3,6X,4I3))
  750 CONTINUE
      RETURN
      END
```

APPENDIX F: EXAMPLE OF RESULTS OBTAINED FROM SUBROUTINE "WEIGHT"

```
DATE OF ANALYSIS = ROODEPLAAT - 820913
TYPE OF ANALYSIS - WEIGHTS

PIXEL OFFSET LIMIT =     0
SAMPLE OFFSET LIMIT =     0
LAND-WATER LIMIT =   30

K=1, Numf!=, <EXP>=50, IMAG1=-4017975963580.0, ID1=-4017975963580.0, IMAG2=-3881569865620.
THE AREA CONSIDERED IN WEIGHT IS: SL=  380 SS= 1350 NL=  105 NS=  105 ID=  40 58- 07293
                                      BAND =       7
K=2, Numf!=, <EXP>=49, IMAG1=-4017975963580.0, ID1=-4017975963580.0, IMAG2=-3881569865620.
THE AREA CONSIDERED IN WEIGHT IS: SL=  380 SS= 1350 NL=  105 NS=  105 ID=  40 58- 07293
                                      BAND =       6
K=3, Numf!=, <EXP>=48, IMAG1=-4017975963580.0, ID1=-4017975963580.0, IMAG2=-3881569865620.
THE AREA CONSIDERED IN WEIGHT IS: SL=  380 SS= 1350 NL=  105 NS=  105 ID=  40 58- 07293
                                      BAND =       5
K=4, Numf!=, <EXP>=47, IMAG1=-4017975963580.0, ID1=-4017975963580.0, IMAG2=-3881569865620.
THE AREA CONSIDERED IN WEIGHT IS: SL=  380 SS= 1350 NL=  105 NS=  105 ID=  40 58- 07293
                                      BAND =       4
```

POINT NO.	LINE	SAMPLE	ACTUAL MEAN				WEIGHTED MEAN			
		BANDS	7	6	5	4	7	6	5	4
1	415	1428	1	7	7	9	4	6	7	12
2	423	1427	1	7	7	9	3	6	6	11
3	416	1424	6	2	2	12	6	4	5	13
4	428	1432	7	6	10	15	11	10	10	12
5	433	1441	6	10	6	12	11	12	9	14
6	413	1417	7	2	11	14	5	4	9	14
7	415	1411	1	5	7	9	5	5	6	12
8	421	1409	7	5	6	16	7	6	7	14
9	408	1419	7	3	10	16	6	3	8	14
10	410	1409	7	6	6	15	6	5	8	13
11	407	1405	6	7	2	12	10	9	9	15
12	401	1397	1	9	10	15	4	6	7	13
13	407	1390	6	10	6	18	5	9	6	18
15	407	1385	14	15	9	12	15	13	10	12
16	418	1382	7	1	6	16	6	4	8	14
17	418	1383	7	3	10	10	7	4	9	11
18	421	1385	20	18	15	16	15	13	11	16
19	418	1377	1	9	6	19	4	8	9	18
20	427	1380	7	3	12	15	6	6	10	15
21	427	1381	1	9	12	15	5	7	11	16
22	436	1377	7	9	10	13	7	8	10	17
23	436	1378	7	6	12	15	12	12	11	16
24	447	1374	7	7	9	16	7	8	10	17
25	447	1375	7	7	12	16	8	9	12	17
26	448	1364	7	7	7	18	7	7	10	18
27	450	1372	7	10	7	11	7	8	10	15
28	457	1374	7	7	11	18	14	12	13	17
29	469	1373	0	0	0	0	16	17	18	16
30	404	1392	1	5	11	14	7	8	10	14
31	403	1417	7	7	9	14	9	8	9	14
32	406	1415	7	5	7	11	6	5	6	12

APPENDIX G: CRITICAL VALUES OF FILLIBEN'S R AND GRUBB'S t TEST AT THE 0,05 LEVEL OF PROBABILITY *
(WHNI/D/AH/NOROUT)

Sample size:

n	t	R	n	t	R
3	0,879	1,153	54	0,977	2,956
4	0,868	1,463	55	0,978	2,992
5	0,879	1,672	56	0,978	2,992
6	0,890	1,822	57	0,978	2,992
7	0,899	1,938	58	0,978	2,992
8	0,905	2,032	59	0,978	2,992
9	0,912	2,110	60	0,980	3,025
10	0,917	2,176	61	0,980	3,025
11	0,922	2,234	62	0,980	3,025
12	0,926	2,285	63	0,980	3,025
13	0,931	2,331	64	0,980	3,025
14	0,934	2,371	65	0,981	3,055
15	0,937	2,409	66	0,981	3,055
16	0,940	2,443	67	0,981	3,055
17	0,942	2,475	68	0,981	3,055
18	0,945	2,504	69	0,981	3,055
19	0,947	2,532	70	0,982	3,082
20	0,950	2,557	71	0,982	3,082
21	0,952	2,580	72	0,982	3,082
22	0,954	2,603	73	0,982	3,082
23	0,955	2,624	74	0,982	3,082
24	0,957	2,644	75	0,983	3,107
25	0,958	2,663	76	0,983	3,107
26	0,959	2,681	77	0,983	3,107
27	0,960	2,698	78	0,983	3,107
28	0,962	2,714	79	0,983	3,107
29	0,962	2,730	80	0,984	3,130
30	0,964	2,745	81	0,984	3,130
31	0,965	2,757	82	0,984	3,130
32	0,966	2,773	83	0,984	3,130
33	0,967	2,786	84	0,984	3,130
34	0,967	2,799	85	0,985	3,151
35	0,968	2,811	86	0,985	3,151
36	0,968	2,823	87	0,985	3,151
37	0,969	2,835	88	0,985	3,151
38	0,970	2,846	89	0,985	3,151
39	0,971	2,857	90	0,985	3,171
40	0,972	2,866	91	0,985	3,171
41	0,972	2,877	92	0,985	3,171
42	0,973	2,887	93	0,985	3,171
43	0,973	2,896	94	0,985	3,171
44	0,973	2,905	95	0,986	3,189
45	0,974	2,914	96	0,986	3,189
46	0,974	2,923	97	0,986	3,189
47	0,974	2,931	98	0,986	3,189
48	0,975	2,940	99	0,986	3,189
49	0,975	2,948	100	0,987	3,207
50	0,977	2,956			
51	0,977	2,956			
52	0,977	2,956			
53	0,977	2,956			

*From Wainwright and Gilbert, 1981.

APPENDIX H: PROGRAM "FILLI" FILLIBEN'S R AND GRUBB'S t TEST FOR NORMALITY AND OUTLIERS. FORTRAN IV

```
$RESET FREE
FILE   5(KIND  =  READER)
FILE   6(KIND  =  PRINTER)
FILE   7(KIND  =  DISK)
FILE  10(KIND = DISK,  TITLE="WHNI/D/AH/NOROUT ON W36",  FILETYPE=7)
C=====================================================================
C
C      PROGRAM "FILLI"
C
C=====================================================================
C
C      FILLIBEN'S TEST FOR NORMALITY AND GRUBB'S TEST FOR OUTLIERS
C      (P = 0.05)"
C
C=====================================================================
C
C      ADAPTED FROM :-
C      WAINWRIGHT, S.J.   AND    GILBERT, R.I. ;
C      LABORATORY. PRACTISE,  VOL.30,  NO.5,  MAY 1981,  P467;
C      ( MODIFIED BY I. SCHOONRAAD AND A.HOWMAN ,  FEB.1983 )
C
C=====================================================================
C
C
C      IMPORTANT VARIABLES ( OPTIONS ) ARE: -
C           SUCOL  =  SURFACE CHLOROPHYLL VALUES  - LOGS
C           INCOL  =  INTEGRATED CHLOROPHYLL VALUES  - LOGS
C           SECDL  =  SECCHI DISC VALUES  - LOGS
C           SUTUL  =  SURFACE TURBIDITY VALUES  - LOGS
C           INTUL  =  INTEGRATED TURBIDITY VALUES  - LOGS
C
C
C      IN ORDER TO RUN THIS PROGRAM THE FOLLOWING JOB DECK
C        INFORMATION IS REQUIRED: -
C
C          FILE 10 = WHNI/D/AH/NOROUT
C          ( CRITICAL VALUES OF FILLIBENS R AND GRUBBS T AT THE 0.05
C            PROBABILITY LEVEL. )
C          FILE 7 = WATER QUALITY AND REFLECTANCE DATA FILE
C          (* MISSING DATA SHOULD BE INDICATED BY THE VALUE " 99.9 " .)
C          NUMBER OF SETS OF DATA TO BE ANALYSED IN COLUMNS 1, 2
C          VARIABLE OPTIONS REQUIRED - IN (A6, 1X) FORMAT.
C
C
C=====================================================================
C
C
       DOUBLE PRECISION  NAME
       REAL      OPTION(5), SPOINT(100), IOPT(5)
       INTEGER   NTOTAL, ICOUNT, DATE, OSPNT(100)
       REAL SUCOL(100), INCOL(100), SECDL(100), SUTUL(100), INTUL(100),
     *  T(100), GRUBTS, E1, E2, GRUBTL
       DATA OPTION/"SUCOL", "INCOL", "SECDL", "SUTUL", "INTUL"/
       READ (5, 140) NTOTAL
   140 FORMAT(I3)
       READ (7, 110) NAME, DATE, OSPNT(1), SUCOL(1), INCOL(1), SECDL(1),
     *  SUTUL(1), INTUL(1)
   110 FORMAT(A10, A6, I2, 1X, 5F7.4)
       WRITE (6, 141) NAME , DATE
   141 FORMAT (10 (1X, A6))
       DO 1000 I = 2, NTOTAL
```

```
        READ (7, 120)OSPNT(I), SUCOL(I), INCOL(I), SECDL(I),
     *                 SUTUL(I), INTUL(I)
 120    FORMAT(16X, I2, 1X, 5F7.4)
1000 CONTINUE
     READ (5, 130)(IOPT(I), I = 1, 5)
 130 FORMAT (5(A6, 1X))
     DO 1300 K = 1, 5
      DO 1100 J = 1, 5
       IF(IOPT(K).EQ.OPTION(J)) GO TO (1200, 1210, 1220, 1230, 1240), J
       GO TO 1100
1200   WRITE (6, 100)
 100 FORMAT("1",24X,"FILLIBEN'S TEST FOR NORMALITY AND GRUBB'S TEST FOR
     * OUTLIERS",/,23X,64("*"),/)
 980   FORMAT(7X, I4,11X, F8.4)
       WRITE(6, 150) NAME, DATE, IOPT(K)
 150   FORMAT (10X,A10,5X,A6,/,8X,14("*"),1X,10("*"),/,10X,"DATA
     * OPTION  TESTED IS ", A6, /, 8X, 42("*"), /)
       DO 1010 IRE = 1, 100
1010   SPOINT(IRE) = OSPNT(IRE)
       ICOUNT = NTOTAL
       CALL FILL (SUCOL, SPOINT, ICOUNT)
       WRITE(6,777)
 777   FORMAT (1X,"SAMPLE POINT NO.    VALUE (LOG)",/)
       WRITE(6, 980)((SPOINT(J), SUCOL(J)), J = 1, ICOUNT)
       GO TO 1300
1210   WRITE (6, 100)
       WRITE(6, 150) NAME, DATE, IOPT(K)
       DO 1020 IRE = 1, 100
1020   SPOINT(IRE) = OSPNT(IRE)
       ICOUNT = NTOTAL
       CALL FILL (INCOL, SPOINT, ICOUNT)
       WRITE(6,777)
       WRITE(6, 980)((SPOINT(J), INCOL(J)), J = 1, ICOUNT)
       GO TO 1300
1220   WRITE (6, 100)
       WRITE(6, 150) NAME, DATE, IOPT(K)
       DO 1030 IRE = 1, 100
1030   SPOINT(IRE) = OSPNT(IRE)
       ICOUNT = NTOTAL
       CALL FILL (SECDL, SPOINT, ICOUNT)
       WRITE(6,777)
       WRITE(6, 980)((SPOINT(J), SECDL(J)), J = 1, ICOUNT)
       GO TO 1300
1230   WRITE (6, 100)
       WRITE(6, 150) NAME, DATE, IOPT(K)
       DO 1040 IRE = 1, 100
1040   SPOINT(IRE) = OSPNT(IRE)
       ICOUNT = NTOTAL
       CALL FILL (SUTUL, SPOINT, ICOUNT)
       WRITE(6,777)
       WRITE(6, 980)((SPOINT(J), SUTUL(J)), J = 1, ICOUNT)
       GO TO 1300
1240   WRITE (6, 100)
       WRITE(6, 150) NAME, DATE, IOPT(K)
       DO 1050 IRE = 1, 100
1050   SPOINT(IRE) = OSPNT(IRE)
       ICOUNT = NTOTAL
       CALL FILL (INTUL, SPOINT, ICOUNT)
       WRITE(6,777)
       WRITE(6, 980)((SPOINT(J), INTUL(J)), J = 1, ICOUNT)
1100  CONTINUE


1300 CONTINUE
     END
```

```
3
C======================================================================
C
      SUBROUTINE FILL (VAL, SPOINT, ICOUNT)
C
C======================================================================
C
      INTEGER ICOUNT, SPOINT(100)
      REAL VAL(100), DEV(100), MEAN1, MEAN, SUM, SQDEV, CONST(100),
     *  TSTAT(100), X(100), T(100)
 1010 Z1 = 0.0
      Z2 = 0.0
      Z3 = 0.0
      Z4 = 0.0
      SQDEV = 0.0
      SUM = 0.0
      SUMDEV = 0.0
      MEAN1 = 0.0
C
C     EXCLUDE MISSING DATA
C
      IE = 1
 1020 IF (VAL(IE).EQ. 99.9) GO TO 1030
      IE = IE + 1
      IF(IE.GT.ICOUNT) GO TO 1060
      IF(IE.EQ.ICOUNT) IFLAG =  - 1
      GO TO 1020
 1030 WRITE(6, 100)SPOINT(IE)
  100 FORMAT(30X, "MISSING DATA FOUND AT SAMPLING POINT NO.  ", I3, /)
      ICOUNT = ICOUNT - 1
      IF(IFLAG.EQ. - 1) GO TO 1060
      DO 1040 JE = IE, ICOUNT
      VAL(JE) = VAL(JE + 1)
      SPOINT(JE) = SPOINT(JE + 1)
 1040 CONTINUE
      GO TO 1020
 1060 DO 1000 J = 1, ICOUNT
 1000 MEAN1 = MEAN1 + VAL(J)/ICOUNT
      DO 1050 J = 1, ICOUNT
      SUM = SUM + VAL(J)
      DEV(J) = VAL(J) - MEAN1
      SQDEV = SQDEV + ((VAL(J) - MEAN1) * (VAL(J) - MEAN1))
 1050 CONTINUE
C
C     CALCULATE MEAN
C
      MEAN = SUM/ICOUNT
      Y = FLOAT(ICOUNT)
      CONST(ICOUNT) = 0.5 *  * (1.0/Y)
      CONST(1) = 1.0 - CONST(ICOUNT)
      DO 1100 I = 2, ICOUNT - 1
      Y1 = FLOAT(I - 0.3175)
      Y2 = FLOAT(ICOUNT + 0.365)
      CONST(I) = (Y1/Y2)
 1100 CONTINUE
      CONST1 = 2.515517
      CONST2 = 0.802853
      CONST3 = 0.010328
      DONST1 = 1.432788
      DONST2 = 0.189269
      DONST3 = 0.001308
```

```
      DO 1200 I = 1, ICOUNT
C
C     SUM OF DEVIATIONS
C
      SUMDEV = SUMDEV + (VAL(I) - MEAN) * (VAL(I) - MEAN)
      IF(CONST(I).GT.0.5) GO TO 1500
 1480 PROB = CONST(I)
      GO TO 1550
 1500 PROB = 1.0 - CONST(I)
 1550 TSTAT(I) = SQRT(ALOG(1.0/(PROB * PROB)))
 1200 CONTINUE
C
C     STANDARD DEVIATION
C
      STDEV = SQRT(SUMDEV/(ICOUNT - 1))
      DO 1750 I = 1, ICOUNT
      X(I) = TSTAT(I)-(CONST1 + CONST2 * TSTAT(I) + CONST3 * TSTAT(I) *
     *TSTAT(I))/(1+DONST1*TSTAT(I)+DONST2*TSTAT(I)*TSTAT(I)+DONST3*
     *  TSTAT(I) * TSTAT(I) * TSTAT(I))
 1750 CONTINUE
      DO 1900 I = 1, ICOUNT
      IF(CONST(I).LT.0.5) X(I) =  - X(I)
 1900 CONTINUE
      DO 2300 I = 1, ICOUNT - 1
      DO 2250 J = I + 1, ICOUNT
      IF(DEV(J).GE.DEV(I)) GO TO 2250
      DEV1 = DEV(I)
      DEV(I) = DEV(J)
      DEV(J) = DEV1
 2250 CONTINUE
 2300 CONTINUE
      DO 2600 I = 1, ICOUNT
      Z1 = Z1 + DEV(I) * X(I)
      Z2 = Z2 + DEV(I) * DEV(I)
      Z3 = Z3 + X(I)
      Z4 = Z4 + X(I) * X(I)
 2600 CONTINUE
      WRITE(6, 30)
   30 FORMAT(25X, "FILLIBEN'S TEST FOR NORMALITY", /, 24X, 31("-"), /)
      R1 = Z1/SQRT(Z2 * (Z4 - Z3 * Z3/ICOUNT))
      WRITE (6, 600) R1, MEAN, STDEV
  600 FORMAT(30X,"FILLIBEN'S R =  ",F5.2,/,30X,"MEAN =  ",F10.2,/,30X,
     *  "STANDARD DEVIATION =  "F10.2)
C
      CALL NORM(R1, T, ICOUNT, D9)
C
      WRITE (6, 700)
  700 FORMAT (25X, "GRUBB'S TEST FOR OUTLIERS ", /, 24X, 27("-"), /)
      DO 3250 I = 1, ICOUNT - I
      DO 3200 J = I + 1, ICOUNT
      IF(VAL(J).GT.VAL(I)) GO TO 3200
      VAL1 = VAL(I)
      ITEMP = SPOINT(I)
      VAL(I) = VAL(J)
      SPOINT(I) = SPOINT(J)
      VAL(J) = VAL1
      SPOINT(J) = ITEMP
 3200 CONTINUE
 3250 CONTINUE
      U = SQRT(SQDEV/(ICOUNT - 1))
      WRITE(6, 800)VAL(I)
```

```
  800 FORMAT(30X, "THE SMALLEST VALUE IS  =  ", F7.2)
      GRUBTS = (MEAN1 - VAL(1))/U
      WRITE(6, 900)GRUBTS
  900 FORMAT(30X, "GRUBB'S 'T' VALUE FOR THIS VALUE IS  =  ", F7.2)
C
      CALL SOUT(VAL, GRUBTS, T, E1, E2, ICOUNT, SPOINT)
C
      WRITE(6, 955)VAL(ICOUNT)
  955 FORMAT(30X, "THE LARGEST VALUE IS  =  ", F7.2)
      GRUBTL = (VAL(ICOUNT) - MEAN1)/U
      WRITE(6, 958) GRUBTL
  958 FORMAT(30X, "GRUBB'S 'T' FOR THIS VALUE IS  = ", F6.2)
C
      CALL LOUT(VAL, GRUBTL, ICOUNT, E8, E9, T, SPOINT)
C
      IF(D9.EQ.2.AND.E2.EQ. - 1.AND.E8.EQ. - 1) GO TO 3650
      IF(D9.EQ.1.AND.E2.EQ. - 1.AND.E8.EQ. - 1) GO TO 4000
      GO TO 4080
 4000 WRITE(6, 960)
  960 FORMAT(30X, "DATA NOT NORMALLY DISTRIBUTED,  BUT NO OUTLIERS")
      GO TO 3650
 4080 F =  - 1
      IF(E2.EQ.1) F = 1
      IF(E8.EQ.1) F = 1
      IF(F.EQ.1) CALL EXOUT(VAL, E1, E9, ICOUNT, MEAN1, SPOINT)
      IF(F.EQ. - 1) GO TO 4200
      GO TO 1010
 4200 WRITE(6, 975)
  975 FORMAT(5X, "LOGIC ERROR -  PROGRAM SHOULD NOT ARRIVE AT THIS
     *  STATEMENT")
 3650 RETURN
      END

C
C==================================================================
C
      SUBROUTINE LOUT(VAL, GRUBTL, ICOUNT, E8, E9, T, SPOINT)
C
C     THIS SUBROUTINE TESTS FOR THE LARGEST OUTLIER VALUE.
C==================================================================
C
      INTEGER  SPOINT(100)
      REAL GRUBTL, VAL(100), E8, E9, T(100)
      IF(GRUBTL.GE.T(ICOUNT)) GO TO 900
      WRITE(6, 100)
  100 FORMAT(30X, "THE LARGEST VALUE IS NOT AN OUTLIER")
      E8 =  - 1
      E9 =  - 999
      GO TO 930
  900 WRITE(6, 150)VAL(ICOUNT), SPOINT(ICOUNT)
  150 FORMAT(30X, "THE LARGEST VALUE ", F7.2, " IS AN OUTLIER", /,
     *  30X, "THE SAMPLING POINT NUMBER OF THE OUTLIER IS ", I3, /, /)
C     ALL OUTLIERS (VAL(ICOUNT)) MUST BE EXCLUDED FROM RECALCULATION.
      E9 = VAL(ICOUNT)
      E8 = 1
  930 RETURN
      END
```

```
C
C=====================================================================
C
      SUBROUTINE SOUT (VAL, GRUBTS, T, E1, E2, ICOUNT, SPOINT)
C
C     THIS SUBROUTINE TESTS FOR THE SMALLEST OUTLIER VALUE.
C
C=====================================================================
C
      INTEGER  SPOINT(100)
      REAL VAL(100), GRUBTS, T(100), E1, E2
      IF(GRUBTS.GE.T(ICOUNT)) GO TO 700
      WRITE (6, 100)
  100 FORMAT (30X, "THE SMALLEST VALUE IS NOT AN OUTLIER")
      E2 =  - 1
      E1 =  - 999
      GO TO 730
  700 WRITE (6, 110) VAL(1), SPOINT(1)
  110 FORMAT (30X, "THE VALUE", F7.2, "IS AN OUTLIER", /,
     *  30X, "THE SAMPLING POINT NUMBER OF THE OUTLIER IS ", I3, /)
C     ALL OUTLIERS (VAL(1)) MUST BE EXCLUDED FROM RECALCULATION.
C
      E1 = VAL(1)
      E2 = 1
  730 RETURN
      END

C
C=====================================================================
C
      SUBROUTINE NORM (R1, T, ICOUNT, D9)
C
C     THIS SUBROUTINE TESTS FOR NORMALITY.
C
C=====================================================================
C
      REAL R(100), T(100), R1
      DO 1000 K = 3, 100
      I = K - 2
 1000 READ(10 = K, 100, END = 1100) R(I), T(I)
  100 FORMAT (9X, F6.3, 5X, F6.3)
 1100 IF (R1.GE.R(ICOUNT)) GO TO 5450
      WRITE (6, 110)
  110 FORMAT (30X, "THE DATA IS NOT NORMALLY DISTRIBUTED, P = 0.05", /)
      D9 = 1
      GO TO 5600
 5450 WRITE (6, 120)
  120 FORMAT (30X, "THE DATA IS NORMALLY DISTRIBUTED,  P = 0.05", /)
      D9 = 2
 5600 RETURN
      END
```

```
C
C=====================================================================
C
      SUBROUTINE EXOUT(VAL, E1, E9, ICOUNT, MEAN1, SPOINT)
C
C     THIS SUBROUTINE EXCLUDES OUTLIERS BEFORE RECALCULATION.
C
C=====================================================================
C
      INTEGER SPOINT(100), NSPT(100)
      REAL VAL(100), H(100)
      DO 1035 I = 1, ICOUNT
      IF (VAL(I).EQ.E1) VAL(I) =  - 10
      IF (VAL(I).EQ.E9) VAL(I) =  - 10
 1035 CONTINUE
      DO 1040 I = 1, 100
 1040 H(I) = 0.0
      J = 0
      DO 1060 I = 1, ICOUNT
      IF(VAL(I).EQ. - 10) GO TO 1060
      J = J + 1
      H(J) = VAL(I)
      NSPT(J) = SPOINT(I)
      NO = J
 1060 CONTINUE
      DO 1130 J = 1, 100
 1130 VAL(J) = 0.0
      DO 1140 L = 1, NO
      SPOINT(L) = NSPT(L)
 1140 VAL(L) = H(L)
      ICOUNT = NO
      IF(ICOUNT.GE.3) GO TO 1150
      WRITE(6, 100)
  100 FORMAT(30X, "LESS THAN THREE VALUES LEFT AFTER OUTLIERS EXCLUDED")
      GO TO 1190
 1150 MEAN1 = 0
      DO 1180 I = 1, ICOUNT
 1180 MEAN1 = MEAN1 + VAL(I)/ICOUNT
      WRITE(6, 110)
  110 FORMAT(25X, "RECALCULATION OF FILLIBEN'S TEST FOR NORMALITY", /,
     *  24X,48("-"),/25X,"AFTER THE EXCLUSION OF THE ABOVE OUTLIERS",/,
     *  24X, 42("-"), /)
 1190 RETURN
      END
```

APPENDIX I: EXAMPLE OF RESULTS OBTAINED FROM PROGRAM "FILLI"

```
                    FILLIBEN'S TEST FOR NORMALITY AND GRUBB'S TEST FOR  OUTLIERS
                    ****************************************************************

        ROODEPLAAT     820913
        ************** **********
        DATA           OPTION  TESTED IS SUCOL
        ******************************************

                    FILLIBEN'S TEST FOR NORMALITY
                    -----------------------------

                         FILLIBEN'S R =   0.98
                         MEAN =          1.28
                         STANDARD DEVIATION =         0.16
                         THE DATA IS NORMALLY DISTRIBUTED,  P = 0.05

                    GRUBB'S TEST FOR OUTLIERS
                    -------------------------

                         THE SMALLEST VALUE IS  =      1.06
                         GRUBB'S 'T' VALUE FOR THIS VALUE IS  =      1.35
                         THE SMALLEST VALUE IS NOT AN OUTLIER
                         THE LARGEST VALUE IS  =     1.74
                         GRUBB'S 'T' FOR THIS VALUE IS  =   2.81
                         THE LARGEST VALUE     1.74 IS AN OUTLIER
                         THE SAMPLING POINT NUMBER OF THE OUTLIER IS  29


                    RECALCULATION OF FILLIBEN'S TEST FOR NORMALITY
                    ----------------------------------------------
                    AFTER THE EXCLUSION OF THE ABOVE OUTLIERS
                    -----------------------------------------

                    FILLIBEN'S TEST FOR NORMALITY
                    -----------------------------

                         FILLIBEN'S R =   0.99
                         MEAN =          1.27
                         STANDARD DEVIATION =         0.14
                         THE DATA IS NORMALLY DISTRIBUTED,  P = 0.05

                    GRUBB'S TEST FOR OUTLIERS
                    -------------------------

                         THE SMALLEST VALUE IS  =      1.06
                         GRUBB'S 'T' VALUE FOR THIS VALUE IS  =      1.45
                         THE SMALLEST VALUE IS NOT AN OUTLIER
                         THE LARGEST VALUE IS  =     1.55
                         GRUBB'S 'T' FOR THIS VALUE IS  =   2.01
                         THE LARGEST VALUE IS NOT AN OUTLIER
```

SAMPLE POINT NO.	VALUE (LOG)
11	1.0607
30	1.0607
10	1.0755
9	1.0897
32	1.1038
12	1.1173
31	1.1173
8	1.1430
7	1.1461
3	1.1761
1	1.1818
13	1.2148
6	1.2175
19	1.2253
2	1.2577
18	1.2742
20	1.3032
16	1.3201
17	1.3284
4	1.3385
23	1.3522
15	1.3598
21	1.3598
22	1.3838
5	1.4048
24	1.4314
25	1.4314
26	1.4698
28	1.5159
27	1.5527

FILLIBEN'S TEST FOR NORMALITY AND GRUBB'S TEST FOR OUTLIERS

ROODEPLAAT 820913

DATA OPTION TESTED IS INCOL

FILLIBEN'S TEST FOR NORMALITY

FILLIBEN'S R = 0.98
MEAN = 1.26
STANDARD DEVIATION = 0.18
THE DATA IS NORMALLY DISTRIBUTED. P = 0.05

GRUBB'S TEST FOR OUTLIERS

THE SMALLEST VALUE IS = 0.99
GRUBB'S 'T' VALUE FOR THIS VALUE IS = 1.54
THE SMALLEST VALUE IS NOT AN OUTLIER
THE LARGEST VALUE IS = 1.70
GRUBB'S 'T' FOR THIS VALUE IS = 2.47
THE LARGEST VALUE IS NOT AN OUTLIER

SAMPLE POINT NO.	VALUE (LOG)
32	0.9912
11	1.0294
8	1.0294
30	1.0755
22	1.0899
31	1.1038
10	1.1038
9	1.1038
7	1.1106
6	1.1335
13	1.1553
12	1.1553
3	1.2068
18	1.2355
1	1.2430
2	1.2529
17	1.2742
15	1.2742
23	1.3032
21	1.3365
19	1.3522
4	1.3541
20	1.3598
24	1.3692
16	1.3838
27	1.4249
25	1.4393
26	1.4698
5	1.4728
28	1.6580
29	1.7024

FILLIBEN'S TEST FOR NORMALITY AND GRUBB'S TEST FOR OUTLIERS
****

ROODEPLAAT 820913
************* **********
DATA OPTION TESTED IS SUTUL
****

FILLIBEN'S TEST FOR NORMALITY

FILLIBEN'S R = 0.90
MEAN = 0.70
STANDARD DEVIATION = 0.14
THE DATA IS NOT NORMALLY DISTRIBUTED, P = 0.05

GRUBB'S TEST FOR OUTLIERS

THE SMALLEST VALUE IS = 0.56
GRUBB'S 'T' VALUE FOR THIS VALUE IS = 1.01
THE SMALLEST VALUE IS NOT AN OUTLIER
THE LARGEST VALUE IS = 1.23
GRUBB'S 'T' FOR THIS VALUE IS = 3.77
THE LARGEST VALUE 1.23 IS AN OUTLIER
THE SAMPLING POINT NUMBER OF THE OUTLIER IS 29

RECALCULATION OF FILLIBEN'S TEST FOR NORMALITY

AFTER THE EXCLUSION OF THE ABOVE OUTLIERS

FILLIBEN'S TEST FOR NORMALITY

FILLIBEN'S R = 0.96
MEAN = 0.68
STANDARD DEVIATION = 0.10
THE DATA IS NOT NORMALLY DISTRIBUTED, P = 0.05

GRUBB'S TEST FOR OUTLIERS

THE SMALLEST VALUE IS = 0.56
GRUBB'S 'T' VALUE FOR THIS VALUE IS = 1.22
THE SMALLEST VALUE IS NOT AN OUTLIER
THE LARGEST VALUE IS = 0.89
GRUBB'S 'T' FOR THIS VALUE IS = 2.01
THE LARGEST VALUE IS NOT AN OUTLIER
DATA NOT NORMALLY DISTRIBUTED, BUT NO OUTLIERS

SAMPLE POINT NO.	VALUE (LOG)
9	0.5563
11	0.5563
32	0.5563
7	0.5682
10	0.5682
12	0.5911
31	0.5911
3	0.6021
13	0.6128
2	0.6232
30	0.6232
1	0.6335
6	0.6335
8	0.6335
4	0.6435
17	0.6435
18	0.6532
15	0.6902
16	0.6990
21	0.6990
19	0.7243
5	0.7324
20	0.7404
22	0.7853
23	0.8261
24	0.8261
25	0.8325
27	0.8513
26	0.8573
28	0.8865

FILLIBEN'S TEST FOR NORMALITY AND GRUBB'S TEST FOR OUTLIERS
****

ROODEPLAAT 820913
************** **********
DATA OPTION TESTED IS INTUL
****

FILLIBEN'S TEST FOR NORMALITY

FILLIBEN'S R = 0.90
MEAN = 0.74
STANDARD DEVIATION = 0.15
THE DATA IS NOT NORMALLY DISTRIBUTED, P = 0.05

GRUBB'S TEST FOR OUTLIERS

THE SMALLEST VALUE IS = 0.57
GRUBB'S 'T' VALUE FOR THIS VALUE IS = 1.14
THE SMALLEST VALUE IS NOT AN OUTLIER
THE LARGEST VALUE IS = 1.32
GRUBB'S 'T' FOR THIS VALUE IS = 3.98
THE LARGEST VALUE 1.32 IS AN OUTLIER
THE SAMPLING POINT NUMBER OF THE OUTLIER IS 29

RECALCULATION OF FILLIBEN'S TEST FOR NORMALITY

AFTER THE EXCLUSION OF THE ABOVE OUTLIERS

FILLIBEN'S TEST FOR NORMALITY

FILLIBEN'S R = 0.98
MEAN = 0.72
STANDARD DEVIATION = 0.10
THE DATA IS NORMALLY DISTRIBUTED, P = 0.05

GRUBB'S TEST FOR OUTLIERS

THE SMALLEST VALUE IS = 0.57
GRUBB'S 'T' VALUE FOR THIS VALUE IS = 1.47
THE SMALLEST VALUE IS NOT AN OUTLIER
THE LARGEST VALUE IS = 0.90
GRUBB'S 'T' FOR THIS VALUE IS = 1.84
THE LARGEST VALUE IS NOT AN OUTLIER

SAMPLE POINT NO.	VALUE (LOG)
8	0.5682
10	0.5682
9	0.5798
32	0.6021
6	0.6128
7	0.6128
30	0.6232
11	0.6435
13	0.6435
12	0.6532
31	0.6532
17	0.6628
3	0.6721
1	0.6902
2	0.6902
4	0.7160
18	0.7243
15	0.7324
16	0.7404
19	0.7404
21	0.7634
5	0.7924
20	0.7924
22	0.8261
23	0.8325
25	0.8388
24	0.8692
27	0.8692
26	0.8865
28	0.9031

APPENDIX J: A PROCEDURE TO NORMALISE THE DATA

The approach for the selection of data points was based on the shape of the normal distribution using the area under segments of the normal curve. This is given in Figure J.1. where the normal distribution has been divided into 7 class intervals. The class intervals are indicated in standard deviation units on the abscissa, with the fraction of the total number of data points lying in each class interval being given as a percent value.

The data were ranked from largest to smallest and the actual fraction of data points in each class interval calculated (Table J.1). Where a given class interval contained far more data points than the expected fraction for the normal distribution (Table J.2) then duplicated values were excluded from the class interval one at a time. After each exclusion, Filliben's R and t test were rerun and so the procedure continued until the subset satisfied the test for normality.

An example of the weighting procedure is illustrated below:

STEP 1: Determine the boundary values for the data set using the values for areas under the normal curve (Figure J.1). The following equation was used:

$$\text{MEAN of data set} + \left(\text{CLASS INTERVAL BOUNDARY VALUE FOR THE NORMAL CURVE UNITS} \times \text{STANDARD DEVIATION OF DATA SET} \right) = \text{BOUNDARY VALUE}$$

e.g.:
Surface Turbidity
Mean = 0,6813
Standard Deviation = 0,1023

Table J.1 illustrates the results.

STEP 2: Rank the data set and determine the number of sampling points falling within the specified boundary limits. Table J.2 illustrates the procedure.

STEP 3: Run Filliben's R and Grubb's t test.

STEP 4: Remove outliers.

STEP 5: Rerun 'Filli' if data are not normally distributed and show areas of clustering.

STEP 6: Remove duplicate data values one by one, as illustrated in Table J.2.

STEP 7: Run 'Filli' the test for normality on the remaining data.

STEP 8: If the data was still not normal start at Step 5 again and repeat.

TABLE J.1: DETERMINING BOUNDARY VALUES USING AREAS UNDER THE NORMAL CURVE

CLASS BOUNDARY OF NORMAL CURVE (σ)	% AREA UNDER CURVE	DETERMINED BOUNDARY FOR SURFACE TURBIDITY
	5%	
-1,645		0,5131
	12,5%	
-0,935		0,5857
	20%	
-0,326		0,6480
	25%	
0,326		0,7147
	20%	
0,935		0,7769
	12,5%	
1,645		1,8495
	5%	

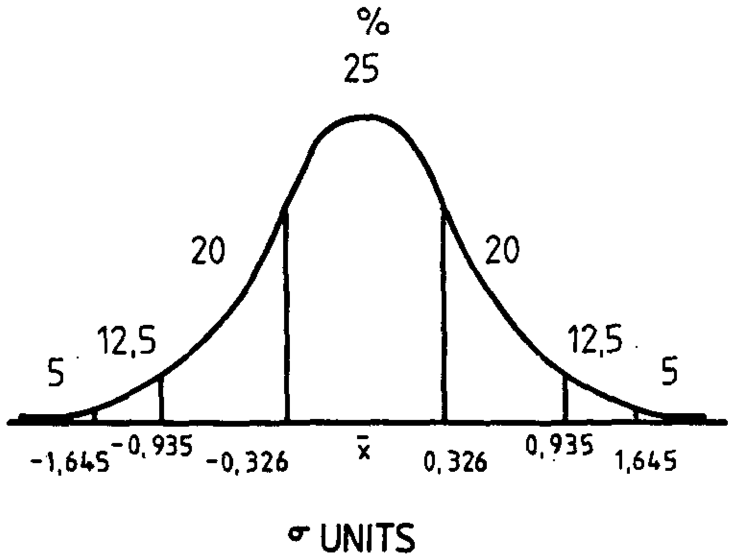


FIGURE J.1: NORMAL DISTRIBUTION CURVE SHOWING CLASS INTERVALS AND AREAS UNDER THE CURVE USED FOR DETERMINING CLUSTERING

TABLE J.2: WEIGHTING PROCEDURE FOR ROODEPLAAT DAM 82.09.13

SURFACE TURBIDITY

Original Log data		Ranked after outliers excluded		Limit. % under CURVE. No. of points.	Duplicates removed	After outliers and duplicates excluded		
				0-0,5131				
				5%				n = 0
				n = 0				
01	0,6335	9	0,5563	0,5131-0,5857		9	0,5563	
02	0,6232	11	0,5563	12,5%		11	0,5563	
03	0,6021	32	0,5563	n = 5		32	0,5563	n = 5
04	0,6435	7	0,5682			7	0,5582	
05	0,7324	10	0,5682			10	0,5682	
06	0,6335							
07	0,5682	12	0,5911					
08	0,6335	31	0,5911		(12) 0,5911	31	0,5911	
09	0,5563	3	0,6021	0,5857-0,6480	(2) 0,6232	3	0,6021	
10	0,5682	13	0,6128	20%	(1) 0,6335	13	0,6128	n = 6
11	0,5563	2	0,6232	n = 11	(6) 0,6335	30	0,6232	
12	0,5911	30	0,6232		(4) 0,6435	8	0,6335	
13	0,6128	1	0,6335			17	0,6435	
15	0,6902	6	0,6335					
16	0,6990	8	0,6335					
17	0,6435	4	0,6435					
18	0,6532	17	0,6435					
19	0,7243							
20	0,7404	18	0,6532	0,6480-0,7147		18	0,6532	
21	0,6990	15	0,6902	25%		15	0,6902	n = 4
22	0,7853	16	0,6990	n = 4		16	0,6990	
23	0,8261	21	0,6990			21	0,6990	
24	0,8261							
25	0,8325	19	0,7243	0,7147-0,7769		19	0,7243	
26	0,8573	5	0,7324	20%		5	0,7324	n = 3
27	0,8513	20	0,7404	n = 3		20	0,7404	
28	0,8865							
29	1,2304	22	0,7853	0,7769-0,8495		22	0,7853	
30	0,6232	23	0,8261	12,5%		23	0,8261	n = 4
31	0,5911	24	0,8261	n = 4		24	0,8261	
32	0,5563	25	0,8325			25	0,8325	
		27	0,8513	0,8495 +		27	0,8513	
		26	0,8573	5%		26	0,8573	n = 3
		28	0,8865	n = 3		28	0,8865	

Any continuous and symmetrical distribution would serve as an adequate aid to selecting a sub-set from the sample set to achieve a reasonably uniform distribution of chlorophyll/turbidity and thus attempt an approach to the ideal experiment.

Effectively the test for normality lifted any possible bias from the data and the exclusion of outliers removed the problem of having two possibly separate populations. In addition the test proved to be easily duplicated and was as objective as possible, under the circumstances.

APPENDIX K: STEPWISE DISCRIMINANT ANALYSIS

The results of the Stepwise Discriminant Analysis test are presented in Figures K.1 to K.5 (refer to Section 3.4.3). The D category, in all instances, is shown to be present at the lower end of the value range, followed in succession by categories B and C, with the P category situated at the opposite end of the value range. The sampling points in classes D and B could be considered to be one class, as often the central point location of their individual groups lie very close together (Figures K.2 and K.3). It is apparent that there are misclassifications of sampling points in the boundary classes chosen. In some instances category C points could have been classified as P points (Figures K.1 and K.5). Nonetheless the analysis, for each of the 6 days's data showed that two, if not more, distinct populations were present in the impoundment (Figures 3.6; K.1 to K.5). In some instances histograms and not full scattergrams of the analyses were produced by the Stepwise Discriminant Analysis Program. The reason for this is that "the group means and all cases are plotted in a scatterplot. The axes are the first two canonical variables. If there is only one canonical variable a histogram is plotted" (Dixon and Brown, 1979). In effect this means that one of the variables may be outstandingly dominant in the relationship under examination. The dominant Canonical variable in the relationship is indicated on the Figures.

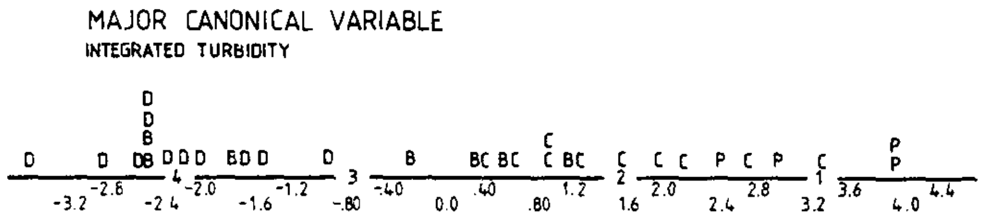


FIGURE K1: STEPWISE DISCRIMINANT ANALYSIS HISTOGRAM FOR 81·11·01

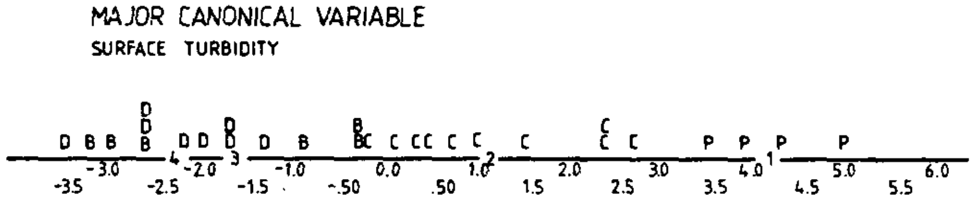


FIGURE K.2: STEPWISE DISCRIMINANT ANALYSIS HISTOGRAM FOR 81·12·07

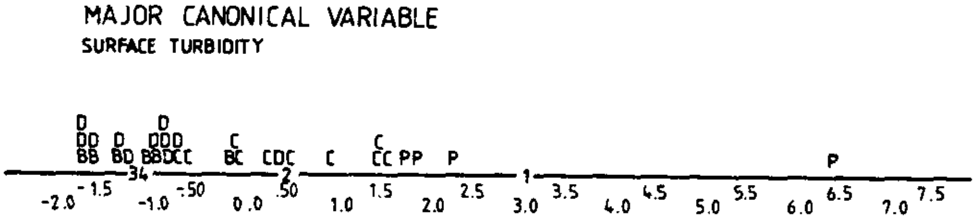


FIGURE K.3: STEPWISE DISCRIMINANT ANALYSIS HISTOGRAM FOR 82·09·13

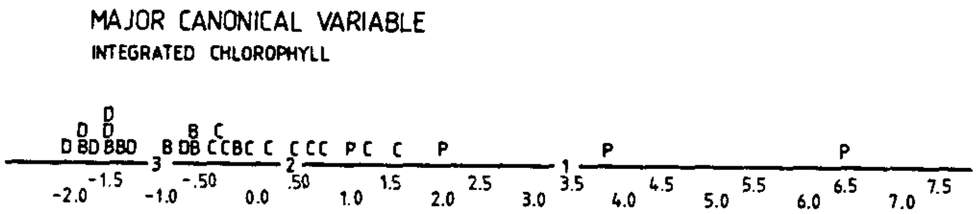


FIGURE K.4: STEPWISE DISCRIMINANT ANALYSIS HISTOGRAM FOR 82·11·16

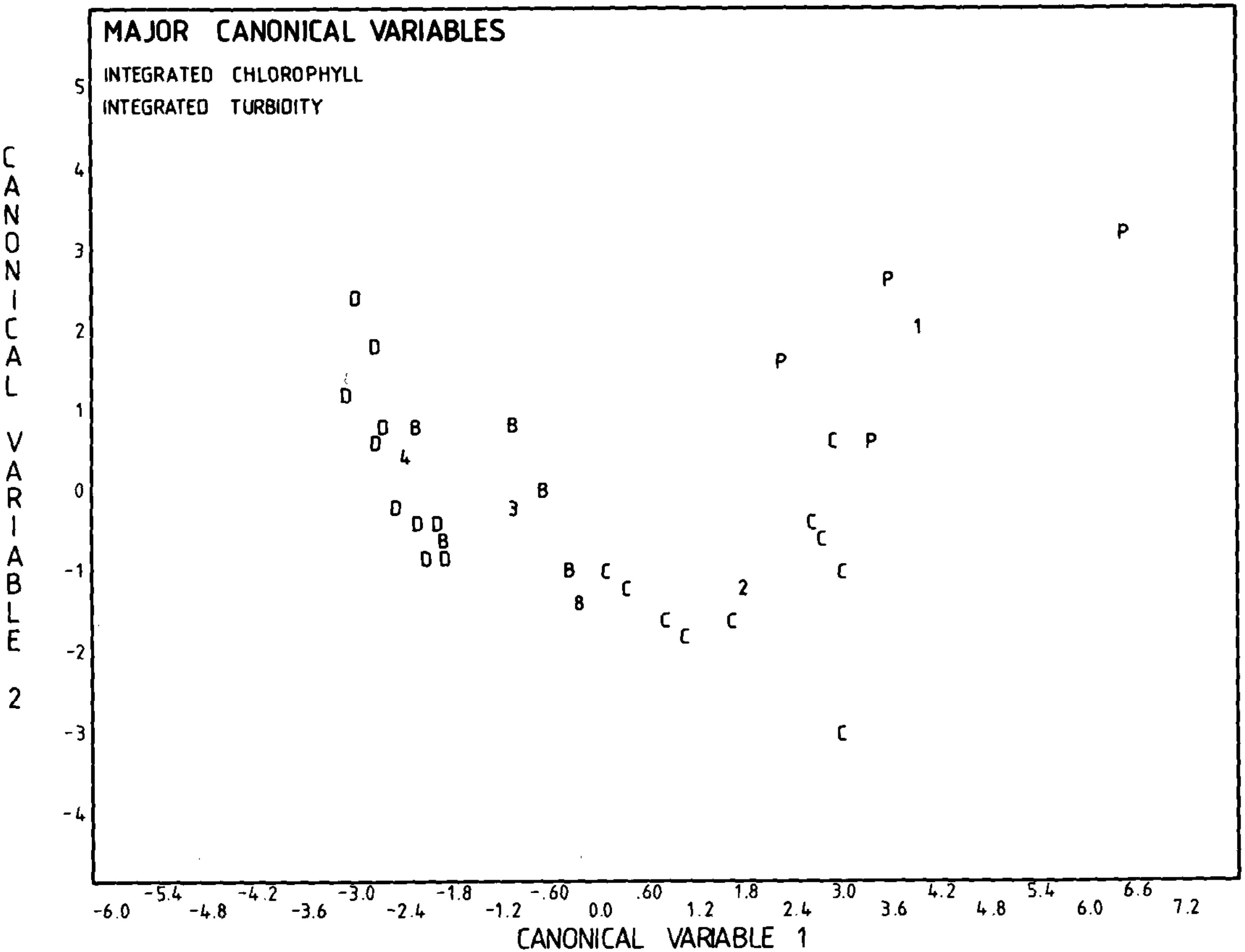


FIGURE K.5: STEPWISE DISCRIMINANT ANALYSIS SCATTERPLOT FOR 82·09·30

APPENDIX L: PROGRAM "LINREG" A LINEAR REGRESSION PROGRAM TO OBTAIN THE SLOPE AND INTERCEPT TERMS USING THE CANONICAL COEFFICIENTS OBTAINED FROM THE CANONICAL CORRELATION ANALYSIS. FORTRAN IV

==

```
$RESET FREE
$ SET SEQ
$ SET LINEINFO
$ RESET LIST
C
C================================================================
C
C     PROGRAM "LINREG"
C
C================================================================
C
C     "LINREG" DETERMINES  :-
C          THE SLOPE "M" OF THE REGRESSION LINE AND
C          THE ORDINATE INTERCEPT "K", FOR THE FITTED LINEAR
C          FUNCTION    Y = MX + K.
C     IN ORDER TO OBTAIN THE LINEAR REGRESSION EQUATION FOR A SET
C      OF DATA USING THE CANONICAL COEFFICIENTS AS THE CORRESPONDING
C      "X" AND "Y" LINEAR POLYNOMIAL FUNCTIONS.
C
C================================================================
C
C     IMPORTANT INDEPENDENT VARIABLE PAIRS ARE:-
C       SUCOL / SUTUL (SURFACE CHLOROPHYLL AND TURBIDITY)
C       INCOL / INTUL (INTEGRATED CHLOROPHYLL AND TURBIDITY)
C       SUCOL / INCOL (SURFACE AND INTEGRATED CHLOROPHYLL)
C       SUTUL / INTUL (SURFACE AND INTEGRATED TURBIDITY)
C
C
C     IN ORDER TO RUN THIS PROGRAM THE FOLLOWING JOB DECK
C       INFORMATION IS REQUIRED: -
C
C         FILE 7 = WATER QUALITY AND REFLECTANCE DATA FILE.
C         THE DATE OF ANALYSIS...(I6).
C         TITLE OF THE ANALYSIS...
C         THE INDEPENDENT VARIABLES NAMES ...(2A5).
C         THE CORRESPONDING FORMAT (I3) ACCORDING TO THE
C          FOLLOWING ....
C                 SUCOL / SUTUL = 1
C                 INCOL / INTUL = 2
C                 SUCOL / INCOL = 3
C                 SUTUL / INTUL = 4
C         THE CORRESPONDING CANONICAL COEFFICIENTS...(2F7.3).
C         THE NUMBER OF PAIRS OF DATA...(I2).
C         THE DEPENDENT VARIABLES NAMES...(4A5).
C         THE CORRESPONDING CANONICAL COEFFICIENTS...(4F7.3).
C         THE NUMBER OF SETS OF DATA...(I2).
C
C
C==============================================================
C
C       THE PROGRAM WAS WRITTEN BY I.SCHOONRAAD AND A.HOWMAN
C        OF THE HYDROLOGICAL RESEARCH INSTITUTE, FEB.1983.
C
C==============================================================
C
C     HOW1
C     THIS PROGRAM CALCULATES THE LINEAR COMBINATION OF THE
C     CANONICAL INDEPENDENT VARIABLE PAIRS.
C
C==============================================================
$SET AUTOBIND
```

```
$BIND = FROM (LSTATS)P/IMSL/ = ON STATS
          DIMENSION A(200),B(200),S(200)
          DIMENSION IMAG4(5151),RANGE(4),ITITLE(144),ICHAR(10)
          DO 5 I = 1,200
             A(I) = 0
             B(I) = 0
             S(I) = 0
    5     CONTINUE
          M = 0
          E = 0
          R = 0
          WRITE (6,7)
    7     FORMAT (2X," LINEAR REGRESSION ANALYSIS USING CANONICAL COEFFICIEN
         *TS ",/,2X,57("*"),/)
          READ (5,15) DATE
   15     FORMAT (A6)
          WRITE (6,23) DATE
   23     FORMAT ("  DATE ANALYSED - ROODEPLAAT =",A6)
          READ (5,17) STAT
   17     FORMAT (A10)
          WRITE (6,19) STAT
   19     FORMAT ("  STATISTICAL ANALYSIS = ",A10,/)
          WRITE (6,10)
   10     FORMAT (" ENTER INDEPENDANT VARIABLE NAMES AS 5 CHARS")
          READ (5,20) V1,V2
   20     FORMAT (2A5)
          READ(5,25) IFORM
   25     FORMAT(I3)
          WRITE (6,30) V1,V2
   30     FORMAT (3X, " VAR1=",A5,2X," VAR2=",A5)
          WRITE (6,40)
   40     FORMAT ( " ENTER COEFFICIENTS FOR VAR1 AND VAR2 AS C1,C2 ")
          READ (5,50) C1,C2
   50     FORMAT (2F7.3)
          WRITE (6,60)C1,C2
   60     FORMAT (3X, " C1=",F7.3,2X,"C2=",F7.3,I2)
          READ (5,80) N
   80     FORMAT (I3)
          WRITE (6,70) N
   70     FORMAT ( " ENTER NUMBER OF PAIRS OF DATA = ",I3)
          DO 110 I = 1,N,1
          GOTO( 1, 2, 3, 4) IFORM
    1     READ(7,100) A(I), B(I)
  100     FORMAT (19X,F7.4,14X,F7.4)
          GOTO 6
    2     READ(7,101) A(I), B(I)
  101     FORMAT (26X,F7.4,14X,F7.4)
          GOTO 6
    3     READ(7,102) A(I), B(I)
  102     FORMAT(19X,2F7.4)
          GOTO 6
    4     READ(7,103) A(I), B(I)
  103     FORMAT(40X,2F7.4)
          GOTO 6
    6     CONTINUE
  110     CONTINUE
          S1 = 0
  185     DO 190 I = 1,N,1
          S(I) = (C1 * A(I)) + (C2 * B(I))
          S1 = S1 + S(I)
  190     CONTINUE
```

```
      S2 = S1 / N
      WRITE (6,200) S2
200   FORMAT ( " MEAN=",F7.3)
      WRITE (6,230) V1,V2
230   FORMAT (/, 3X,A5,10X,A5,11X,"SUM",5X,"SEQUENCE NUMBER",
     */,3X,5("-"),10X,5("-"),11X,3("-"),5X,15("-"))
      DO 250 I = 1,N,1
      WRITE (6,240) A(I),B(I),S(I),I
240   FORMAT ( 1X,F7.3,8X,F7.3,8X,F7.3,8X,I3)
250   CONTINUE
C
C
C======================================================================
C
C     HOW2
C     THIS PART OF THE PROGRAM CALCULATES THE LIN COMB. OF CANONICAL
C     DEP VARIABLE PAIRS
C
C======================================================================
      INTEGER C6
      DIMENSION C(200),D(200),T(200)
      DO 305 I = 1,200
         A(I) = 0
         B(I) = 0
         C(I) = 0
         D(I) = 0
         T(I) = 0
305   CONTINUE
      WRITE (6,310)
310   FORMAT (/," ENTER DEPENDANT VARIABLE NAMES AS 5 CHARS")
      READ (5,320) V1,V2,V3,V4
320   FORMAT (4A5)
      WRITE (6,330) V1,V2,V3,V4
330   FORMAT (3X, " VAR1=",A5,2X,"VAR2=",A5,2X,"VAR3=",A5,2X,"VAR4=",A5)
      WRITE (6,340)
340   FORMAT ( " ENTER COEFFICIENTS FOR VAR1 TO VAR4 AS C1,C2,C3,C4")
      READ (5,350) C1,C2,C3,C4
350   FORMAT (4F7.3)
      WRITE (6,360) C1,C2,C3,C4
360   FORMAT (3X, " C1=",F7.3,4X,"C2=",F7.3,4X,"C3=",F7.3,4X,"C4=",F7.3)
      READ (5,380) N
380   FORMAT (I3)
      WRITE (6,370) N
370   FORMAT ( " ENTER NUMBER OF SETS OF DATA = ",I3)
      REWIND 7
      DO 410 I =1,N,1
      READ (7,400) A(I),B(I),C(I),D(I)
400   FORMAT (54X,I3,3I4)
410   CONTINUE
      T1 = 0
490   DO 500 I = 1,N,1
      T(I) = (C1*A(I)) + (C2*B(I)) + (C3*C(I)) + (C4*D(I))
      T1 = T1 + T(I)
500   CONTINUE
      T2 = T1/N
      WRITE (6,510) T2
510   FORMAT ( " MEAN=",2X,F7.3,/)
      WRITE (6,520) V1,V2,V3,V4
520   FORMAT ( 2X,A5,8X,A5,8X,A5,8X,A5,10X,"SUM",5X,
     *"SEQUENCE NUMBER",/,2X,5("-"),8X,5("-"),8X,5("-"),8X,5("-"),10X,
     *3("-"),5X,15("-"))
```

```
      DO 540 I = 1,N,1
      WRITE (6,530) A(I),B(I),C(I),D(I),T(I),I
530   FORMAT ( 1X,F5.1,8X,F5.1,8X,F5.1,8X,F5.1,8X,F7.3,8X,I3)
540   CONTINUE
      LOW = 1
      HIGH = N
      CALL LINREG (S,T,LOW,HIGH,M,E,R)
      NOFY = 1
      IY = 200
      ITITLE(1) = 0
      INC = 1
      RANGE(1) = 0
      RANGE(2) = 0
      RANGE(3) = 0
      RANGE(4) = 0
      ICHAR(1) = "*"
      IOPT = 0
      STOP
      END
C
C=================================================================
C
      SUBROUTINE LINREG (X,Y,N1,NN,M,E,R)
C
C     N1 IS THE LOWEST VALUE OF THE X,Y ARRAYS TO BE USED AND NN IS
C     THE HIGHEST
C
C=================================================================
C
      REAL X(N1:NN),Y(N1:NN),M,E,R
      INTEGER I
C
C     INITIALIZE SUMS TO ZERO
C
      SUMX = 0
      SUMY = 0
      SUMXY = 0
      SUMXSQ = 0
      SUMYSQ = 0
C
      N = NN - N1 + 1
      DO 560 I = N1,NN,1
      XY = X(I) * Y(I)
      XSQ = X(I) * X(I)
      YSQ = Y(I) * Y(I)
      SUMX = SUMX + X(I)
      SUMY = SUMY + Y(I)
      SUMXY = SUMXY + XY
      SUMXSQ = SUMXSQ + XSQ
      SUMYSQ = SUMYSQ + YSQ
560   CONTINUE
      SQSUMX = SUMX * SUMX
      SQSUMY = SUMY * SUMY
      DENOM = SQSUMX - (N*SUMXSQ)
      M = (SUMX * SUMY - N * SUMXY) / DENOM
      E = (SUMX * SUMXY - SUMXSQ * SUMY) / DENOM
      RDENOM = SQRT((N * SUMXSQ - SQSUMX) * (N * SUMYSQ - SQSUMY))
      R = (N * SUMXY - SUMX * SUMY) / RDENOM
      WRITE (6,570) M,E
570   FORMAT (/,6X, " Y = MX + K  : ",F7.2," X + ",F7.2,/,6X,35("*"),/)
      WRITE (6,580) R
580   FORMAT (6X, " CORRELATION COEFFICIENT = ",F7.2,/,6X,35("*"))
      RETURN
      END
```

APPENDIX M: EXAMPLE OF RESULTS OBTAINED FROM PROGRAM "LINREG"

```
LINEAR REGRESSION ANALYSIS USING CANONICAL COEFFICIENTS
*******************************************************

DATE ANALYSED - ROODEPLAAT =820913
STATISTICAL ANALYSIS =      DATA

ENTER INDEPENDANT VARIABLE NAMES AS 5 CHARS
   VAR1=SUTUL   VAR2=INTUL
ENTER COEFFICIENTS FOR VAR1 AND VAR2 AS C1,C2
   C1= 5.369  C2= 1.688
ENTER NUMBER OF PAIRS OF DATA =  31
MEAN=  4.996
```

SUTUL	INTUL	SUM	SEQUENCE NUMBER
0.634	0.690	4.566	1
0.623	0.690	4.511	2
0.602	0.672	4.367	3
0.644	0.716	4.664	4
0.732	0.792	5.270	5
0.634	0.613	4.436	6
0.568	0.613	4.085	7
0.634	0.568	4.360	8
0.556	0.580	3.965	9
0.568	0.568	4.010	10
0.556	0.644	4.073	11
0.591	0.653	4.276	12
0.613	0.644	4.376	13
0.690	0.732	4.942	14
0.699	0.740	5.003	15
0.644	0.663	4.574	16
0.653	0.724	4.730	17
0.724	0.740	5.139	18
0.740	0.792	5.313	19
0.699	0.763	5.042	20
0.785	0.826	5.611	21
0.826	0.833	5.841	22
0.826	0.869	5.903	23
0.833	0.839	5.886	24
0.857	0.887	6.099	25
0.851	0.869	6.038	26
0.887	0.903	6.284	27
1.230	1.322	8.838	28
0.623	0.623	4.398	29
0.591	0.653	4.276	30
0.556	0.602	4.003	31

```
ENTER DEPENDANT VARIABLE NAMES AS 5 CHARS
   VAR1=BAND4  VAR2=BAND5  VAR3=BAND6  VAR4=BAND7
ENTER COEFFICIENTS FOR VAR1 TO VAR4 AS C1,C2,C3,C4
   C1=  0.042     C2=  0.364     C3=  0.117     C4= -0.112
ENTER NUMBER OF SETS OF DATA =  31
MEAN=    3.987
```

BAND4	BAND5	BAND6	BAND7	SUM	SEQUENCE NUMBER
12.0	7.0	6.0	4.0	3.306	1
11.0	6.0	6.0	3.0	3.012	2
13.0	5.0	4.0	6.0	2.162	3
12.0	10.0	10.0	11.0	4.082	4
14.0	9.0	12.0	11.0	4.036	5
14.0	9.0	4.0	5.0	3.772	6
12.0	6.0	5.0	5.0	2.713	7
14.0	7.0	6.0	7.0	3.054	8
14.0	8.0	3.0	6.0	3.179	9
13.0	8.0	5.0	6.0	3.371	10
15.0	9.0	9.0	10.0	3.839	11
13.0	7.0	6.0	4.0	3.348	12
18.0	6.0	9.0	5.0	3.433	13
12.0	10.0	13.0	15.0	3.985	14
14.0	8.0	4.0	6.0	3.296	15
11.0	9.0	4.0	7.0	3.422	16
16.0	11.0	13.0	15.0	4.517	17
18.0	9.0	8.0	4.0	4.520	18
15.0	10.0	6.0	6.0	4.300	19
16.0	11.0	7.0	5.0	4.935	20
17.0	10.0	8.0	7.0	4.506	21
16.0	11.0	12.0	12.0	4.736	22
17.0	10.0	8.0	7.0	4.506	23
17.0	12.0	9.0	8.0	5.239	24
18.0	10.0	7.0	7.0	4.431	25
15.0	10.0	8.0	7.0	4.422	26
17.0	13.0	12.0	14.0	5.282	27
16.0	18.0	17.0	16.0	7.421	28
14.0	10.0	8.0	7.0	4.380	29
14.0	9.0	8.0	9.0	3.792	30
12.0	6.0	5.0	6.0	2.601	31

```
   Y = MX + K  :    0.87 X +    -0.36
  **********************************

   CORRELATION COEFFICIENT =     0.87
  **********************************
```

APPENDIX N: VALUES FOR THE SLOPE (M) OF THE REGRESSION LINE AND THE INTERCEPT (K) ON THE Y AXIS

VALUES FOR THE SLOPE (M) OF THE REGRESSION LINE AND THE INTERCEPT (K) ON THE Y AXIS AS OBTAINED FROM THE CANONICAL COEFFICIENTS AND THE LINEAR REGRESSION PROGRAM 'INCLUDING ALL DATA' OPTION.

'INCLUDING ALL DATA'	SUCOL/SUTUL		INCOL/INTUL		SUCOL/INCOL		SUTUL/INTUL	
DATE	M	K	M	K	M	K	M	K
81.10.14	0,88	-1,78	0,89	-0,30	0,80	08,31	0,90	-1,40
81.11.01	0,79	3,08	0,93	-3,44	0,73	-8,73	0,93	-2,03
81.12.07	0,94	-0,11	0,95	-0,41	0,87	-1,80	0,96	-0,12
82.09.13	0,87	0,72	0,86	-0,54	0,76	-2,23	0,87	-0,36
82.09.30	0,90	3,54	0,92	2,24	0,90	3,46	0,92	2,21
82.11.16	0,95	3,39	0,95	3,31	0,95	3,50	0,83	4,30

VALUES FOR THE SLOPE (M) OF THE REGRESSION LINE AND THE INTERCEPT (K) ON THE Y AXIS AS OBTAINED FROM THE CANONICAL COEFFICIENTS AND THE LINEAR REGRESSION PROGRAM 'EXCLUDING OUTLIERS' OPTION.

'EXCLUDING OUTLIERS	SUCOL/SUTUL		INCOL/INTUL		SUCOL/INCOL		SUTUL/INTUL	
DATE	M	K	M	K	M	K	M	K
81.10.14	0,85	-1,21	0,87	1,43	0,70	-15,85	0,88	-1,91
81.11.01	0,79	3,08	0,93	-3,44	0,73	-8,73	0,93	-2,03
81.12.07	0,94	-0,11	0,95	-0,41	0,87	-1,80	0,96	-0,12
82.09.13	0,81	2,54	0,75	0,50	0,66	-1,02	0,78	0,57
82.09.30	0,89	3,80	0,94	3,45	0,89	3,47	0,93	2,10
82.11.16	0,91	3,76	0,90	3,48	0,90	3,72	0,69	8,45

VALUES FOR THE SLOPE (M) OF THE REGRESSION LINE AND THE INTERCEPT (K) ON THE Y AXIS AS OBTAINED FROM THE CANONICAL COEFFICIENTS AND THE LINEAR REGRESSION PROGRAM 'NORMALISED DATA' OPTION.

'NORMALISED'	SUCOL/SUTUL		INCOL/INTUL		SUCOL/INCOL		SUTUL/INTUL	
DATE	M	K	M	K	M	K	M	K
81.10.14	0,87	-1,43	0,87	1,57	0,73	-14,91	0,88	-1,96
81.11.01	0,79	3,08	0,93	-3,44	0,73	-8,73	0,93	-2,03
81.12.07	0,94	-0,11	0,95	-0,41	0,87	-1,80	0,96	-0,12
82.09.13	0,81	3,11	0,78	1,39	0,66	0,38	0,79	1,19
82.09.30	0,83	5,03	0,91	0,93	0,86	2,57	0,90	1,72
82.11.16	0,91	3,85	0,90	3,61	0,90	3,84	0,69	8,63

Y = MX + K

M = slope of the line

K = intercept on Y-axis

APPENDIX O: SOLVING THE SIMULTANEOUS EQUATIONS

Terms used:

Surface Reference Data Variables -
- Surface Chlorophyll = SC
- Integrated Chlorophyll = IC
- Surface Turbidity = ST
- Integrated Turbidity = IT

Satellite Reflectance Data Variables -
- Band 4 = B4
- Band 5 = B5
- Band 6 = B6
- Band 7 = B7

Values for D, E, F and G are corresponding Canonical Coefficients expressed as D1, D2, D3, D4, E1, E2 etc. for each of the respective equations.

The following simultaneous equations are obtained from the Canonical Correlation Analysis.

$$Y1 = M1.X+K1 \quad \ldots\ldots \quad (1)$$

where $Y1 = D1(B4)+D2(B5)+D3(B6)+D4(B7)$
and $X = N1(SC)+N2(ST)$;

$$Y2 = M2.X+K2 \quad \ldots\ldots \quad (2)$$

where $Y2 = E1(B4)+E2(B5)+E3(B6)+E4(B7)$
and $X = O1(IC)+O2(IT)$;

$$Y3 = M3.X+K3 \quad \ldots\ldots \quad (3)$$

where $Y3 = F1(B4)+F2(B5)+F3(B6)+F4(B7)$
and $X = P1(SC)+P2(IC)$;

$$Y4 = M4.X+K4 \quad \ldots\ldots \quad (4)$$

where $Y4 = G1(B4)+G2(B5)+G3(B6)+G4(B7)$
and $X = Q1(ST)+Q2(IT)$;

From Equation (1):

$$Y1 = (M1.N1.SC)+(M1.N2.ST)+K1$$

i.e.

$$ST = \left\{\frac{Y1-K1-(M1.N1.SC)}{M1.N2}\right\} \quad \ldots\ldots \quad (5)$$

Substitute for ST from Equation (5) into Equation (4):

i.e. $Y4 = M4.Q1.ST + M4.Q2.IT+K4$

i.e.

$$Y4 = M4.Q1.\left\{\frac{Y1-K1-M1.N1.SC}{M1.N2}\right\} + M4.Q2.IT+K4$$

Multiply by (M1.N2):

i.e.

$$M1.N2.Y4 = M4.Q1.Y1-M4.Q1.K1-M4.Q1.M1.N1.SC+M4.Q2.M1.N2.IT+K4.M1.N2$$

Solve for IT:

$$IT = \left\{\frac{M1.N2.Y4-M4.Q1.Y1+M4.Q1.K1+M4.Q1.M1.N1.SC-K4.M1.N2}{M4.Q2.M1.N2}\right\} \quad (6)$$

There are two equations in the explicit form for ST and IT provided SC is known (i.e. Equations (5) and (6) above.)

For ease of operation, define new variables to simplify Equations (5) and (6) as follows:
Let U = (Y1-K1)

Equation (5) then becomes:

$$ST = \left\{ \frac{U-M1.N1.SC}{M1.N2} \right\} \qquad (7)$$

Likewise let V = M1.N2.Y4-M4.Q1.Y1+M4.Q1.K1-K4.M1.N2
and let W = (M4.Q2.M1.N2)

Equation (6) then becomes:

$$IT = \left\{ \frac{V+M4.Q1.M1.N1.SC}{W} \right\} \qquad (8)$$

Now substitute for IT from Equation (8) into Equation (2), which then becomes:

$$Y2 = M2.O1.IC+M2.O2. \left\{ \frac{V+M4.Q1.M1.N1.SC}{W} \right\} +K2$$

Solve for IC:

$$IC = \left\{ \frac{W.Y2-W.K2-M2.O2.(V+M4.Q1.M1.N1.SC)}{W.M2.O1} \right\} \qquad (9)$$

Expand Equation (9):

$$IC = \left\{ \frac{W.Y2-W.K2-M2.O2.V-M2.O2.M4.Q1.M1.N1.SC}{W.M2.O1} \right\}$$

Define new variables for ease of operation as follows:
Let H1 = (W.Y2-W.K2-M2.O2.V)
Let H2 = (M2.O2.M4.Q1.M1.N1)
Let H3 = (W.M2.O1)

The expanded form of Equation (9) then becomes:

$$IC = \left\{ \frac{H1-H2.SC}{H3} \right\} \qquad (10)$$

Substitute for IC from Equation (10) into Equation (3):

$$Y3 = M3.P1.SC + M3.P2. \left\{ \frac{H1-H2.SC}{H3} \right\} +K3$$

i.e. $H3.Y3 = H3.M3.P1.SC+M3.P2.H1-M3.P2.H2.SC+K3.H3$ .... (11)

and solve for SC:

$$SC = \left\{ \frac{(H3.Y3-M3.P2.H1-K3.H3)}{(H3.M3.P1-M3.P2.H2)} \right\} \qquad (12)$$

Equation (12) is an explicit solution for SC.
To obtain IC insert value for SC into Equation (10)
To obtain IT insert value for SC into Equation (8)
To obtain ST insert value for SC into Equation (7)

APPENDIX P: SUBROUTINE "CONVERT" THE MODEL FOR SIMULATING CONCENTRATIONS OF WATER QUALITY CONDITIONS AT SPECIFIC SITES. FORTRAN IV

```
      SUBROUTINE CONVRT (IP,NPN,NPNN,NDAM,LIM,LN,OP)
C
      COMMON/SUCOL/D1,D2,D3,D4,N1,N2,M1,K1
      COMMON/INCOL/E1,E2,E3,E4,O1,O2,M2,K2
      COMMON/SUTUL/F1,F2,F3,F4,P1,P2,M3,K3
      COMMON/INTUL/G1,G2,G3,G4,Q1,Q2,M4,K4
      DIMENSION IP(NPN,4),NPNN(100),OP(NPN,4)
C
C  THIS SUBROUTINE CONVERTS A FOUR BAND INPUT TO A FOUR
C    DIMENSIONSAL OUTPUT. EACH DIMENSION REPRESENTING SUCOL
C    INCOL,SUTUL,INTUL.
C
      REAL IC,IT,N1,N2,M1,M2,M3,M4,K1,K2,K3,K4
C       OBTAIN THE KNOWN DATA FROM WHNI/D/AH/LOAD/DATA
C           THE SUCOL SUTUL DATA
      IF (LN.GT.0) GO TO 1
      WRITE (6,225)
  225 FORMAT (/,10X,'CALIBRATION DATA - THE CANONICAL COEFFICIENTS',
     */,9X,47('-')/)
      WRITE (6,222)
  222 FORMAT(/,20X,'THE SUCOL SUTUL DATA')
      READ (5,100) D1,D2,D3,D4,N1,N2,M1,K1
      WRITE(6,100) D1,D2,D3,D4,N1,N2,M1,K1
  100 FORMAT(4(5X,F7.3),/,2(5X,F7.3),/,(5X,F7.3),/,(5X,F7.3))
C           THE INCOL INTUL DATA
      WRITE (6,333)
  333 FORMAT(/,20X,'THE INCOL INTUL DATA')
      READ (5,100) E1,E2,E3,E4,O1,O2,M2,K2
      WRITE(6,100) E1,E2,E3,E4,O1,O2,M2,K2
C           THE SUCOL INCOL DATA
      WRITE (6,120)
  120 FORMAT(/,20X,'THE SUCOL INCOL DATA')
      READ (5,100) F1,F2,F3,F4,P1,P2,M3,K3
      WRITE(6,100) F1,F2,F3,F4,P1,P2,M3,K3
C           THE SUTUL INTUL DATA
      WRITE (6,444)
  444 FORMAT(/,20X,'THE SUTUL INTUL DATA')
      READ (5,100) G1,G2,G3,G4,Q1,Q2,M4,K4
      WRITE(6,100) G1,G2,G3,G4,Q1,Q2,M4,K4
C
C        WRITE THE HEADING.
      IF(LN.LT.0) WRITE (6,2001)
 2001 FORMAT (/,5X,"POINT NO.",6X,2X,"SUCOL",5X,
     *"INCOL",5X,"SUTUL",5X,"INTUL")
C     INITIAL STEPS TO SOLVING THE EQUATIONS
    1 NDAM = 0
      DO 150 K=1,NPN
      B4=FLOAT(IP(K,4))
      B5=FLOAT(IP(K,3))
      B6=FLOAT(IP(K,2))
      B7=FLOAT(IP(K,1))
      IF (B7.GE.LIM) GO TO 151
      NDAM=NDAM+1
      Y1=((D1*B4)+(D2*B5)+(D3*B6)+(D4*B7))
      Y2=((E1*B4)+(E2*B5)+(E3*B6)+(E4*B7))
      Y3=((F1*B4)+(F2*B5)+(F3*B6)+(F4*B7))
      Y4=((G1*B4)+(G2*B5)+(G3*B6)+(G4*B7))
C
      U=(Y1-K1)
C
      V=((M1*N2*Y4)-(M4*Q1*Y1)+(M4*Q1*K1)-(K4*M1*N2))
```

```
C
      W=(M4*Q2*M1*N2)
C
      H1=((W*Y2)-(W*K2)-(M2*O2*V))
      H2=(M2*O2*M4*Q1*M1*N1)
      H3=(W*M2*O1)
C     TO SOLVE FOR SUCOL
      SC=((H3*Y3)-(M3*P2*H1)-(K3*H3))/((H3*M3*P1)-(M3*P2*H2))
      OP(K,1)=SC
C
C     TO SOLVE FOR INCOL
      IC=(H1-(H2*SC))/(H3)
      OP(K,2)=IC
C
C     TO SOLVE FOR SUTUL
      ST=(U-(M1*N1*SC))/(M1*N2)
      OP(K,3)=ST
C
C     TO SOLVE FOR INTUL
      IT=(V+(M4*Q1*M1*N1*SC))/ W
      OP(K,4)=IT
C       WRITE THE RESULTS.
      IF(LN.LT.0) WRITE(6,2000) NPNN(K),SC,IC,ST,IT
 2000 FORMAT (/,8X,I3,6X,4F10.3)
      GO TO 150
  151 OP(K,1)=-99.
  150 CONTINUE
      RETURN
      END
```

APPENDIX Q: CALIBRATION DATA FOR 82.09.30, INCLUDING ALL DATA OPTION - AN EXAMPLE

SURFACE CHLOROPHYLL a/SURFACE TURBIDITY CALIBRATION DATA:

COEFFICIENTS FOR THE DEPENDENT DATA (BANDS)
D1= 0,313 D2= 0,062 D3= 0,050 D4= -0,024
COEFFICIENTS FOR THE INDEPENDENT DATA (WATER QUALITY)
N1= 3,622 N2= 0,237
SLOPE OF THE LINEAR REGRESSION EQUATION
M1= 0,900
INTERCEPT OF THE LINEAR REGRESSION EQUATION
K1= 3,537

INTEGRATED CHLOROPHYLL a/INTEGRATED TURBIDITY CALIBRATION DATA:

COEFFICIENTS FOR THE DEPENDENT DATA (BANDS)
E1= 0,142 E2= 0,114 E3= 0,136 E4= -0,083
COEFFICIENTS FOR THE INDEPENDENT DATA (WATER QUALITY)
O1= -0,664 O2= 6,262
SLOPE OF THE LINEAR REGRESSION EQUATION
M2= 0,919
INTERCEPT OF THE LINEAR REGRESSION EQUATION
K2= 2,242

SURFACE AND INTEGRATED CHLOROPHYLL a CALIBRATION DATA:

COEFFICIENTS FOR THE DEPENDENT DATA (BANDS)
F1= 0,315 F2= 0,060 F3= 0,051 F4= -0,023
COEFFICIENTS FOR THE INDEPENDENT DATA (WATER QUALITY)
P1= 3,241 P2= 0,575
SLOPE OF THE LINEAR REGRESSION EQUATION
M3= 0,902
INTERCEPT OF THE LINEAR REGRESSION EQUATION
K3= 3,461

SURFACE AND INTEGRATED TURBIDITY CALIBRATION DATA:

COEFFICIENTS FOR THE DEPENDENT DATA (BANDS)
G1= 0,158 G2= 0,114 G3= 0,125 G4= -0,078
COEFFICIENTS FOR THE INDEPENDENT DATA (WATER QUALITY)
Q1= 0,193 Q2= 5,227
SLOPE OF THE LINEAR REGRESSION EQUATION
M4= 0,921
INTERCEPT OF THE LINEAR REGRESSION EQUATION
K4= 2,211

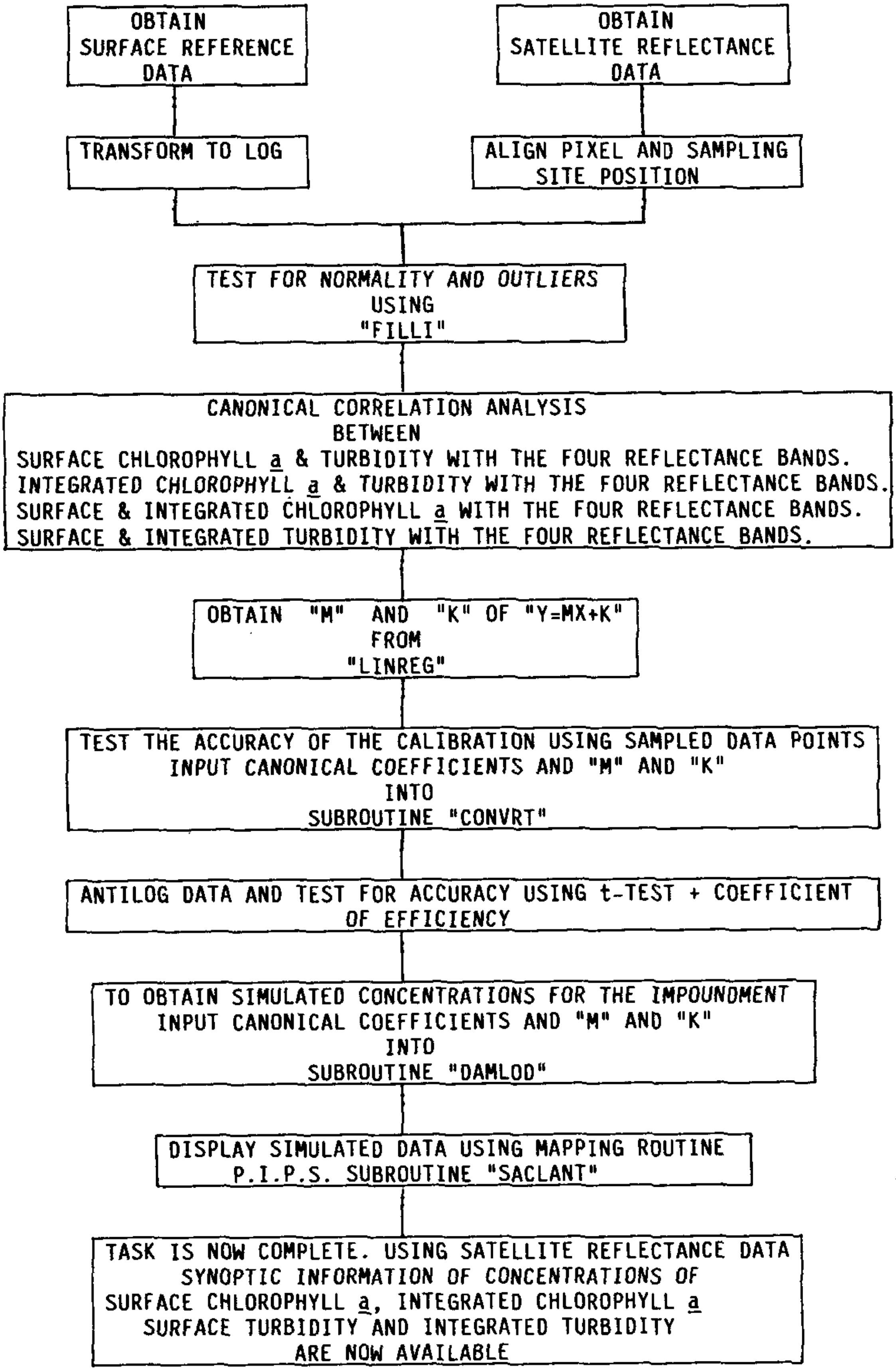
APPENDIX R: SUBROUTINE "DAMLOD" THE MODEL FOR SIMULATING CONCENTRATIONS OF WATER QUALITY CONDITIONS OVER THE ENTIRE IMPOUNDMENT. FORTRAN IV

```
      SUBROUTINE DAMLOD
C     THIS SUBOUTINE CONVERTS A DAM WATER SURFACE TO PREDICTED VALUES OF
C     SUCOL,INCOL,SUTUL AND INTUL. OUTPUT IS A FREQUENCY DISTRIBUTION
C     OF ALL PIXELS CONSIDERED TO BE WATER AND PROVISION HAS BEEN MADE
C     FOR 50 CLASSES.
C     INPUT FROM 4 BANDS,BAND7,6,5 AND 4 L FILE1,2,3 AND 4 RESPECTIVELY|
C

      INTEGER SL, SS, SSDF,SLDF,EX,SAFE,BAND,NI(4),NIMAX (4)
      DIMENSION INPUT (4000),OUTP(4000),FREQ(4,5000),VAR(4),
     *VMIN(4), VMAX(4),NPNN(100)
      REAL IC,IT,LOWLIM
      COMMON/PIC1/NL,NS,BAND,SLDF,SSDF,LLDF,LSDF
      COMMON/PIC2/SL,SS,NLL,NSS,ID1,ID2
      COMMON/FILES/NUMFI(10),NUMFO(10),NEXT,NO,NINT
      COMMON/EXEC/EX,PROCES
      EX=1
C
C     READ THE LOCAL PARAMETERS
C     DELTA = CLASS INTERVAL FOR CHLOROPHYL
C     DIFT = CLASS INTERVAL FOR TURBIDITY
      WRITE (6,777)
  777 FORMAT (5X,'LOAD DATA FOR ROODEPLAAT DAM',/,3X,32('*'),/)
C
      READ(5,888) DATE,TYPE
  888 FORMAT(10X,I6,2X,A5)
      WRITE(6,999) DATE,TYPE
  999 FORMAT(7X,"DATE=" I6,5X,"TYPE="A5,/)
      READ (5,1111) LIM,DELTA,DIFT
 1111 FORMAT(10X,3(I3))
      WRITE (6,2000)LIM,DELTA,DIFT
 2000 FORMAT(5X,"LAND-WATER LIMIT(BAND 7)="I3,/,
     *5X,"CHLOROPHYLL STEP="I3,/,5X,"TURBIDITY STEP="I3,/)
C
C     CHECK THE DISK FILES FOR SIZE
C
      CALL DISKSZ
      NOS = LSDF-SSDF+1
      IF (NOS .GT. 1000)CALL PRINT (1,6,24,'XXX INPUT BUFFER TOO SMALL')
      IF (NS .GT. 1000)CALL PRINT (1,6,26,'XXX INPUT BUFFER TOO SMALL')
      NIP= NEXT+NINT
      NTOT=0
      NOUT=0
C
C     START PROCESSING THE DATA LINE BY LINE
C
      DO 50 KR=1,4
      VMAX(KR)=0
      VAR(KR)=0.
      VMIN(KR)=100
      DO 50 J=1,500
   50 FREQ(KR,J)=0
      ISKIP=SS-SSDF
      DO 1000 LIN = SL,NLL
      IREC = LIN-SLDF+2
      DO 1010 IN=1,NIP
      IST = (IN-1) * NS+1
```

```
      CALL ORIGIN (NUMFI(IN),IREC,ISKIP,NS,INPUT(IST))
 1010 CONTINUE
C
C     NOW CONVERT THE LINE JUST READ
C
      LINN = LIN-SL
      CALL CONVRT (INPUT,NS,NPNN,NDAM,LIM,LINN,OUTP)
C
C     TABULATE THE VALUES AND CONVERT TO ANTI-LOGS;DISCARD AREA
C     OF LAND
C
      DO 120 JR=1,NS
      IF(OUTP(JR).EQ.-99)GO TO 120
      DO 110 K=1,4
      DD=DELTA
      IF (K.GT.2) DD =DIFT
      IIST=(K-1)*NS
      OP=10** OUTP(IIST+JR)
      IF (OP.GT.499*DD) GO TO 100
      WRITE(18+K,1)LIN,JR,OP
    1 FORMAT(3I4)
  110 CONTINUE
      DO 130 K=1,4
      DD=DELTA
      IF(K.GT.2) DD=DIFT
      IIST=(K-1)*NS
      OP = 10**OUTP(IIST+JR)
      VAR (K)= VAR(K)+ OP
      IF (OP .GT. VMAX(K))  VMAX(K) =OP
      IF (OP.LT. VMIN(K))  VMIN(K) =OP
      NI(K)=OP/DD +1
      OUTP(IIST+JR)=0.
      IF (NI(K) .GT.NIMAX(K)) NIMAX(K) = NI(K)
      FREQ (K,NI(K)) = FREQ (K,NI(K)) +1
  130 CONTINUE
      GO TO 120
  100 NOUT=NOUT+1
C     IF (K.EQ.1) WRITE  (6,2020)
C     IF (K.EQ.2) WRITE  (6,2025)
C     IF (K.EQ.3) WRITE  (6,2030)
C     IF (K.EQ.4) WRITE  (6,2035)
C     WRITE (6,2060) LIN,JR,OP
C2060 FORMAT(/,10X,'AT LINE',I5,'SAMPLE ' ,I5,'EXCESSIVE OUTPUT='F12.1)
C     WRITE (6,2065) INPUT(JR),INPUT(JR+NS),INPUT(JR+2.*NS),
C    *INPUT(3.*NS+JR)
C2065 FORMAT (10X,'INPUT BAND 7,6,5,4 =  ',4(X,I4))

  120 CONTINUE
      NTOT =NTOT +NDAM
 1000 CONTINUE
      NTOT=NTOT-NOUT
      WRITE(6,2040) NTOT,NOUT
C
C     NOW THE ENTIRE DAM SURFACE IS PROCESSED
C
C     CONVERT FREQUENCIES TO PERCENTAGE OF DAM SURFACE
      DO 125 J=1,4
      DO 125 KR=1,NIMAX(J)
  125 FREQ (J,KR)=FREQ(J,KR)/FLOAT (NTOT) *100
      DO 150  K=1,4
      NC =  NIMAX(K)
```

```
      VMEAN =VAR(K) /FLOAT(NTOT)
      VMX = VMAX(K)
      VMN = VMIN(K)
      DD = DELTA
      IF(K.EQ.1) WRITE (6,2020)
      IF(K.EQ.2) WRITE (6,2025)
      IF(K.EQ.3) WRITE (6,2030)
      IF(K.EQ.4) WRITE (6,2035)
      IF (K.GT.2) DD= DIFT
      WRITE (6,2010) VMEAN, VMX, VMN
 2010 FORMAT(5X,'MEAN=',F10.2,/,5X,'MAX=',F10.2,/,5X,'MIN=',F10.2,/,
     */,7X,'CLASS RANGE ',4X,'PERCENTAGE AREA',/)
      DO 160 J=1,NC
      IF (FREQ(K,J).EQ.0.) GO TO 160
      VLOWL = (J-1) * DD
      UPLIM =VLOWL + DD
      WRITE (6,2050) VLOWL,UPLIM, FREQ(K,J)
  160 CONTINUE
 2050 FORMAT (5X,3(2X,F8.2))
  150 CONTINUE
      RETURN
 2020 FORMAT(/,10X,'SURFACE CHLOROPHYLL RESULTS - ug/l',/,9X,36('-'),/)
 2025 FORMAT(/,10X,'INTEGRATED CHLOROPHYLL RESULTS - ug/l',/,9X,39('-')
     *,/)
 2030 FORMAT(/,10X,'SURFACE TURBIDITY RESULTS - NTU',/,9X,33('-'),/)
 2035 FORMAT(/,10X,'INTEGRATED TURBIDITY RESULTS - NTU',/,9X,36('-'),/)
 2040 FORMAT(/,10X,'NUMBER OF PIXELS IN DAM = ',I10,/,9X,38('-'),
     */,10X,'NUMBER OF PIXELS WITH EXCESSIVE VALUES = ',I10,/)
      END
```

APPENDIX S: CALMCAT MENU FOR THE ANALYSIS OF SURFACE REFERENCE DATA AND SATELLITE REFLECTANCE DATA IN ORDER TO OBTAIN WATER QUALITY INFORMATION



APPENDIX T: GENERAL CALIBRATION DATA FOR ROODEPLAAT DAM - USING FIVE DAYS OF DATA - 'NORMALISED DATA'OPTION

==

SURFACE CHLOROPHYLL a/SURFACE TURBIDITY CALIBRATION DATA:

COEFFICIENTS FOR THE DEPENDENT DATA (BANDS)
D1= 0,060 D2= 0,005 D3= 0,328 D4= -0,262
COEFFICIENTS FOR THE INDEPENDENT DATA (WATER QUALITY)
N1= -1,433 N2= 5,145
SLOPE OF THE LINEAR REGRESSION EQUATION
M1= 0,690
INTERCEPT OF THE LINEAR REGRESSION EQUATION
K1= 0,580

INTEGRATED CHLOROPHYLL a/INTEGRATED TURBIDITY CALIBRATION DATA:

COEFFICIENTS FOR THE DEPENDENT DATA (BANDS)
E1= 0,026 E2= 0,023 E3= 0,334 E4= -0,254
COEFFICIENTS FOR THE INDEPENDENT DATA (WATER QUALITY)
O1= -0,117 O2= 4,487
SLOPE OF THE LINEAR REGRESSION EQUATION
M2= 0,680
INTERCEPT OF THE LINEAR REGRESSION EQUATION
K2= -0,410

SURFACE AND INTEGRATED CHLOROPHYLL a CALIBRATION DATA:

COEFFICIENTS FOR THE DEPENDENT DATA (BANDS)
F1= -0,103 F2= 0,037 F3= 0,389 F4= -0,260
COEFFICIENTS FOR THE INDEPENDENT DATA (WATER QUALITY)
P1= 1,664 P2= 2,576
SLOPE OF THE LINEAR REGRESSION EQUATION
M3= 0,610
INTERCEPT OF THE LINEAR REGRESSION EQUATION
K3= -2,630

SURFACE AND INTEGRATED TURBIDITY CALIBRATION DATA:

COEFFICIENTS FOR THE DEPENDENT DATA (BANDS)
G1= 0,027 G2= 0,012 G3= 0,353 G4= -0,267
COEFFICIENTS FOR THE INDEPENDENT DATA (WATER QUALITY)
Q1= 2,382 Q2= 1,952
SLOPE OF THE LINEAR REGRESSION EQUATION
M4= 0,700
INTERCEPT OF THE LINEAR REGRESSION EQUATION
K4= -0,500